

Imperial College London

Department of Computing

**Optimistic and Pessimistic
Ambiguous Chance Constraints
with Applications**

Vladimir Roitch

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of Imperial College London, and the
Diploma of Imperial College

April 2015

Abstract

In this thesis, we consider optimisation problems which involve ambiguous chance constraints, i.e., probabilistic constraints where the probability distribution of the primitive uncertainties is at least partly unknown. In this case, we can define an ambiguity set that contains all distributions consistent with our prior knowledge of the uncertainty and take either a pessimistic (worst-case) or optimistic (best-case) view of the world. The former view can be used to actively optimise a system whilst guaranteeing some predefined level of safety; being robust even if the worst-case scenario materialises. The latter view can be used to actively optimise a system where it is required to reconstruct realisations of a random variable whose distribution is not known precisely.

We characterise the ambiguity set through generalised moment bounds and structural properties such as symmetry, unimodality, or independence patterns. Sufficient conditions are presented under which the corresponding chance constraints admit equivalent explicit tractable conic reformulations that can be solved with off-the-shelf solvers. However, in general, ambiguous chance constrained problems are provably difficult and we suggest efficiently computable conservative approximations.

To illustrate the effectiveness of our reformulations, we give two detailed and novel examples. First, we consider the pricing problem of a provider of cloud computing services. This provider faces uncertain demand and wishes to maximise profit, whilst maintaining a desired level of quality of service. We show that such a problem naturally fits within the pessimistic ambiguous chance constraint framework. Second, we consider the problem of improving the quality of a photographic image by reconstructing and then removing noise. We show that such a problem can be formulated as an optimistic ambiguous chance constrained program that generalises, and offers new insight to, an existing powerful image denoising approach.

Acknowledgements

My deepest gratitude is to my adviser, Dr Daniel Kuhn, who illuminated the often dark path through this endeavour. Dr Wolfram Wiesemann was also a great inspiration. Both their accessibility, guidance, and support has been greatly appreciated.

I am also grateful to my friends and colleagues at Imperial College London: Angelos Georghiou, Iakovos Kakouris, Michael Hadjiyiannis, Grani Hanasusanto, Duy Luong, Chin Pang Ho, Dimitra Bampou, Christos Gavriel, Phebe Vayanos, Panos Parpas, Paula Rocha, Loizos Markides, Vahan Hovhannisyan, Eva Kalyvianaki, Juan Campos, Vitali Avagyan, and many others. I am especially thankful to Prof. Berç Rustem and Dr Amani El-Kholy for the encouragement, interest, and guidance.

Finally, I wish to thank my parents for their support.

Declaration of Originality

The work presented in this thesis is my own (except where duly acknowledged) and has not been submitted for the award of any degree or diploma in any other tertiary institution.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Contents

1	Introduction	19
1.1	Motivation	19
1.2	Aims of the Thesis	26
1.3	Contributions and Structure	26
1.4	Notation	28
2	Distributionally Robust Uncertainty Quantification and Chance Constraints	31
2.1	Introduction	32
2.2	Ambiguity Sets	36
2.2.1	Nested Moment Ambiguity Sets	37
2.2.2	Structural Ambiguity Sets	38
2.2.3	Examples	40
2.3	Uncertainty Quantification	44
2.4	Chance Constrained Programming	54
2.4.1	Individual Chance Constraints	54
2.5	Conclusion	58
2.6	Appendix	60
3	Ambiguous Chance Constraints with Conic Dispersion Measures	79
3.1	Introduction	79
3.2	Worst-Case Joint Chance Constraints with Left-Hand Side Uncertainty	84
3.2.1	Intractability of Joint Chance Constraints with Left-Hand Side Uncertainty	87

3.2.2	Approximations based on the Conditional Value-at-Risk	89
3.2.3	Approximations based on Sequential Convex Optimisation Algorithm	95
3.3	Worst-Case Joint Chance Constraints with Right-Hand Side Uncertainty	97
3.3.1	Tractable Reformulation for Conic Dispersion Measures	98
3.3.2	Intractability of Joint Chance Constraints with Support Information	101
3.3.3	Intractability of Generic Polyhedral Ambiguity Set	104
3.4	Best-Case Joint Chance Constraints with Conic Dispersion Measures	108
3.5	Conclusion	113
3.6	Appendix	114
4	Worst-Case Ambiguous Chance Constraints: Cloud Computing	123
4.1	Introduction	123
4.2	Cloud Computing	124
4.3	Stochastic Model	125
4.4	Solving the Stochastic Model	130
4.5	Distributionally Robust Model	132
4.5.1	Worst-Case Chance Constraint	133
4.5.2	Worst-Case Expected Revenue	135
4.5.3	Single Product Monopolistic Model	137
4.6	Multi-Product Monopolist	140
4.7	Conclusion	145
4.8	Appendix	147
5	Best-Case Ambiguous Chance Constraints: Image Denoising	161
5.1	Introduction	161
5.2	Deterministic Image Denoising	162
5.2.1	Computational Experiments	167
5.3	Stochastic Image Denoising	169
5.3.1	Pessimistic Chance Constraints	169

5.3.2	Optimistic Chance Constraints	171
5.4	Conclusion	174
5.5	Appendix	175
6	Conclusion and Future Work	193

List of Figures

3.1	Illustration of the i -th constraint inside the probability expression in (3.9) when $x_i = 1$ (left), $x_i = 0$ (middle) and $x_i \in (0, 1)$ (right). The light shaded region denotes the realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$ for which the constraint is satisfied. The contour lines represent the projected density function of a probability distribution $\mathbb{P} \in \mathcal{P}$ which violates the i -th constraint for the dark shaded realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$	88
3.2	Block coordinate descent algorithm for problem (3.22).	96
3.3	Illustration of the i -th constraint inside the probability expression in (3.27) when $x_i = 1$ (left), $x_i = -1$ (middle) and $x_i \in (-1, 1)$ (right). The light shaded region denotes the realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$ for which the constraint is satisfied. The contour lines represent the projected density function of a probability distribution $\mathbb{P} \in \mathcal{P}$ which violates the i -th constraint for the dark shaded realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$	103
4.1	The worst-case probability derived in Lemma 4.1 with support constraints as a function of mean μ and standard deviation σ . The right figure visualises the level sets of the worst-case probability.	134
4.2	The worst-case probability derived in Lemma 4.2 without support constraints as a function of mean μ and standard deviation σ . The right figure shows the difference of the worst-case probabilities corresponding to the models with and without support.	135
4.3	The worst-case expected revenue function (non-negative support) from Lemma 4.3 as a function of price p , with $d(p) = 1 - p$, $\mu = 1.1$, $c = 1$, and $\sigma = 1.6$	137
4.4	Visualisations of the optimal price p and demand as a function of the distribution parameters μ and σ from Lemma 4.4. In both cases $d(p) = 1 - p$, $c = 1$, and $\varepsilon = 1\%$	139

4.5	The optimal worst-case expected revenue function (non-negative support) from Lemma 4.4 as a function of mean μ and standard deviation σ , with $d(p) = 1 - p$, $c = 1$, and $\varepsilon = 1\%$.	139
4.6	Optimal prices (p_1^*, p_2^*) and expected revenue as functions of the cross-price elasticity of demand $-0.1 \leq x \leq 0.1$, for a two-product monopolist.	145
4.7	The three cases. The red dashed line indicates the dual curve $z(\xi)$, while the solid blue line is the primal objective $\mathbb{1}_{\xi \geq 1}$.	151
5.1	Plot of h and f as a function of σ . Image size is 50x50, standard deviation of noise is 0.1. The point where the functions become constant is the upper bound given in Lemma 5.1.	166
5.2	Examples of no noise, Gaussian noise, uniform noise, and salt-and-pepper noise (left to right). The standard deviation of the noise of each individual pixel is set to 0.1. The image is the commonly used “Lena” photograph.	167
5.3	Plot showing how the standard deviation of the noise affects the optimal σ and the quality of the denoised image.	168
5.4	Plot showing how the size of the image (square root of the number of pixels) is related to the optimal σ .	168
5.5	The photo on the left is the original image. The middle photo is noisy image with Gaussian noise and a standard deviation 0.05. The right image is the denoised image $\mathbf{u}^*$ with $\sigma = 3.9115$ and $d(\mathbf{u}^*) = 62\%$. The image is the commonly used “Lena” photograph.	169

List of Tables

2.1	Summary of the results in this chapter. Checks (✓) and crosses (✗) highlight tractable reformulations and intractability results, respectively.	59
3.1	Definitions of χ and ρ for some of the ambiguity sets studied in this chapter.	96
5.1	Average denoising quality $d(\mathbf{u}^*)$ for images of size 100x100 for different values of ε (rows) and δ (columns) with the a priori optimally on average selected $\sigma = 8.73$. The smallest distance to the original image is achieved with $\varepsilon = 0.2$ and $\delta = 0.04$	173
5.2	Average denoising quality $d(\mathbf{u}^*)$ for images of size 100x100 for different values of ε (rows) and δ (columns) with the statistical estimate $\sigma = 9.99$. The smallest distance to the original image is achieved with $\varepsilon = 0.3$ and $\delta = 0.02$	173

Chapter 1

Introduction

1.1 Motivation

Many interesting real-world problems in science, technology, engineering, and business can be posed (without loss of generality) as optimisation problems of the form

$$\begin{aligned} & \text{minimise} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && g(\mathbf{x}, \boldsymbol{\xi}) \leq 0, \end{aligned} \tag{1.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a vector of decision variables whose values are to be selected by the above optimisation problem. Restrictions (usually physical) on possible values of $\mathbf{x}$ are encoded in the set $\mathcal{X} \subseteq \mathbb{R}^n$. The quality of each decision is evaluated with respect to a cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Naturally, we make decisions based on observed data or parameters which are usually outside our control. This enters our optimisation problem through the vector of parameters $\boldsymbol{\xi} \in \mathbb{R}^m$, and we require that additionally the decision $\mathbf{x}$ satisfies $g(\mathbf{x}, \boldsymbol{\xi}) \leq 0$, where $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$.

If $\boldsymbol{\xi}$ is precisely known (deterministic), then the tractability of (1.1) depends on the properties of f , g , and $\mathcal{X}$, such as convexity. For example, if the feasible region after an epigraph transformation of (1.1) can be expressed in the very general form of an intersection of hyperplanes with the semi-definite cone [114], then there are already a dozen off-the-shelf solvers which can be used to solve (1.1) in polynomial

time. Of course the optimal decision $\mathbf{x}^*$ as well as the optimal value $f(\mathbf{x}^*)$ is sensitive to the choice of the input data ξ . If (1.1) was a linear program (i.e. $\mathcal{X}$ is polyhedral and f, g are affine) and g jointly affine in $(\mathbf{x}, \xi)$ then there are powerful techniques for determining how the optimal value would change as ξ varies, but even these are very limited [68, 119, 118]. Example 1.1 illustrates this.

Example 1.1 (Linear Programming Sensitivity Analysis). *A linear program in canonical form has the following structure:*

$$\begin{aligned} f(\mathbf{b}) = \text{minimise} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^n \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{1.2}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of decision variables and $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the constraint coefficient matrix. In this example the aim is to show how the optimal objective value of (1.2) varies as the data vector $\mathbf{b}$ is varied, i.e. $\nabla f(\mathbf{b})$. To do this we will dualise (1.2). The dualisation process allows us to write down an equivalent problem to (1.2) but where the objective function is a linear combination of the constraints of (1.2). A good introduction to convex duality and related techniques can be found in [14]. The first step is to write down the Lagrangian function which is the sum of the objective function plus a linear combination of all the constraints

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) - \boldsymbol{\mu}^\top \mathbf{x} = \mathbf{x}^\top (\mathbf{c} + \mathbf{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu}) - \mathbf{b}^\top \boldsymbol{\lambda},$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are known as dual variables. Strong duality tells us that the optimal value of (1.2) is equivalent to

$$\begin{aligned} \text{maximise} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \in \mathbb{R}_+^m, \boldsymbol{\mu} \in \mathbb{R}_+^n \end{aligned} \tag{1.3}$$

where g is known as the dual function, and defined as

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} \text{minimise} \quad \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) & \text{if } \mathbf{c} + \mathbf{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0} \\ \text{subject to} \quad \mathbf{x} \in \mathbb{R}^n & \text{otherwise.} \end{cases} = \begin{cases} -\mathbf{b}^\top \boldsymbol{\lambda} & \text{if } \mathbf{c} + \mathbf{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0} \\ -\infty & \text{otherwise.} \end{cases}$$

Substituting the definition of g into (1.3) means that the optimal value of (1.2) is equivalent to the optimal value of the following optimisation problem:

$$\begin{aligned} f(\mathbf{b}) = \text{maximise} \quad & \mathbf{b}^\top \boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{\lambda} \in \mathbb{R}^m \\ & \mathbf{A}^\top \boldsymbol{\lambda} \leq \mathbf{c}, \quad \boldsymbol{\lambda} \leq \mathbf{0} \end{aligned} \tag{1.4}$$

Intuitively, one can imagine that for a fixed $\mathbf{b}$, the quantity $\nabla f(\mathbf{b})$ is closely related to $\boldsymbol{\lambda}^*$, the optimal solution of (1.4). Indeed, this is why dual variables are sometimes known as shadow prices, as they indicate how the objective value of (1.1) varies as the constant right-hand-side vector $\mathbf{b}$ is varied. More details on such perturbation and sensitivity analysis techniques can be found in [14, §5.6] and [79].

In reality, since $\boldsymbol{\xi}$ is exogenous, it needs to be measured or forecast before it is used as input to (1.1), and therefore it is unreasonable to assume that it is precisely known. Perhaps we know that all possible values of $\boldsymbol{\xi}$ are contained in some set Ξ and instead consider a more robust optimisation problem

$$\begin{aligned} \text{minimise} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X} \\ & g(\mathbf{x}, \boldsymbol{\xi}) \leq 0 \quad \forall \boldsymbol{\xi} \in \Xi. \end{aligned} \tag{1.5}$$

The optimal value of (1.5) is greater than or equal to the optimal value of (1.1) since (1.5) contains more constraints. If the set Ξ is finite, and does not contain “too many” elements, then the constraint may be explicitly replicated for each $\boldsymbol{\xi} \in \Xi$. Otherwise, further conditions on Ξ and g are required to ensure tractability [6], typically by invoking convex duality theory. For example it is sufficient for tractability if Ξ is a polyhedron and g is bilinear in $\mathbf{x}$ and $\boldsymbol{\xi}$. This classical robust optimisation problem provides robustness for all possible values $\boldsymbol{\xi} \in \Xi$. By solving (1.5) and implementing the optimal decision $\mathbf{x}^*$ the encountered cost is guaranteed not to exceed $f(\mathbf{x}^*)$, independent of the actual value of $\boldsymbol{\xi} \in \Xi$ that materialises. Example 1.2 illustrates the techniques of classical robust optimisation.

Example 1.2 (Classical Robust Optimisation). *In this example we will show how classical robust op-*

timisation techniques can be applied to deal with the following robust constraint:

$$\exists \mathbf{c} \in \mathbb{R}^n : \quad \mathbf{c}^\top \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}, \quad (1.6)$$

where $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ is a polyhedron. Constraint (1.6) ensures that all points in the polyhedron $\mathcal{X}$ are on one side of the hyperplane with coefficients $\mathbf{c}$. The robust constraint (1.6) is equivalent to enforcing that the minimum of $\mathbf{c}^\top \mathbf{x}$ over $\mathbf{x} \in \mathcal{X}$ is not less than zero, i.e.

$$\begin{aligned} \exists \mathbf{c} \in \mathbb{R}^n : \quad \min \quad & \mathbf{c}^\top \mathbf{x} && \geq 0. \\ \text{s.t.} \quad & \mathbf{x} \in \mathbb{R}^n && \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0} && \end{aligned} \quad (1.7)$$

Applying the same dualisation techniques as in Example 1.1 we have the following, equivalent to (1.7), constraint:

$$\begin{aligned} \exists \mathbf{c} \in \mathbb{R}^n : \quad \max \quad & \mathbf{b}^\top \boldsymbol{\lambda} && \geq 0. \\ \text{s.t.} \quad & \boldsymbol{\lambda} \in \mathbb{R}^m && \\ & \mathbf{A}^\top \boldsymbol{\lambda} \leq \mathbf{c}, \quad \boldsymbol{\lambda} \leq \mathbf{0} && \end{aligned} \quad (1.8)$$

Finally, requiring the maximum to be non-negative is the same as finding a feasible solution with non-negative objective value. We therefore arrive at the robust counterpart of (1.6):

$$\exists \mathbf{c} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \mathbb{R}^m : \quad \mathbf{b}^\top \boldsymbol{\lambda} \geq 0, \quad \mathbf{A}^\top \boldsymbol{\lambda} \leq \mathbf{c}, \quad \boldsymbol{\lambda} \leq \mathbf{0}.$$

Using convex duality we have converted a semi-infinite constraint into an equivalent finite constraint. A full treatment of robust optimisation can be found in [6].

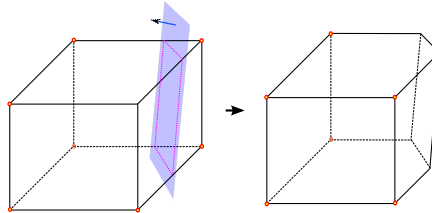
Since Ξ contains all possible values of ξ , the solution of (1.5) may be overly conservative. Indeed, the optimal decision may be to withdraw completely from the enterprise if Ξ contains the most extreme scenarios. To alleviate this, we can assign a probability to each element of Ξ , i.e. a probability

distribution $\mathbb{P}$ over Ξ , and consider instead the following chance constrained optimisation problem

$$\begin{aligned} & \text{minimise} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && \mathbb{P}\left(g(\mathbf{x}, \tilde{\xi}) \leq 0\right) \geq 1 - \varepsilon, \end{aligned} \tag{1.9}$$

where $\tilde{\xi}$ is a random variable, and we require that the constraint holds with high probability (ε is a small value). This allows the constraint to be violated for less likely realisations of $\tilde{\xi}$ which otherwise would have rendered the solution overly conservative. Optimisation problems with chance constraints were first considered in [23, 80, 94]. Unfortunately, these optimisation problems are notoriously difficult. Even for a fixed $\mathbf{x}$ evaluating $\mathbb{P}\left(g(\mathbf{x}, \tilde{\xi}) \leq 0\right)$ involves computing the value of multidimensional integrals, a feat which becomes increasingly difficult as the dimension of $\tilde{\xi}$ increases. Example 1.3 illustrates this. Even if g is convex in $\mathbf{x}$, the implied feasible set of $\mathbf{x}$ is typically non-convex and sometimes even disconnected [87, 94], the former case illustrated by Example 1.4. In the simplest case, if $\mathbb{P}$ was a discrete distribution, the standard reformulation of the chance constraint requires the introduction of binary variables (one for each point of mass). The value of a binary variable corresponding to a particular scenario would be set to 1 if and only if the constraint is satisfied for that scenario. That scenario's probability would then contribute to the total, which is constrained to exceed $1 - \varepsilon$. This approach is illustrated in Example 1.5.

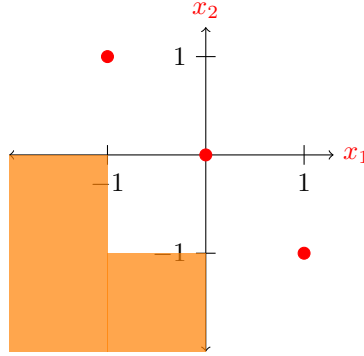
Example 1.3 (Intractability of Uncertainty Quantification). *In this example we show a result from [40]. Suppose $\tilde{\xi}$ has a uniform distribution over the n -dimensional unit cube, i.e. $\tilde{\xi} \sim \mathbb{U}([0, 1]^n)$, then $\mathbb{P}\left(\mathbf{a}^\top \tilde{\xi} \leq b\right)$ is equivalent to computing the volume of the knapsack polytope $[0, 1]^n \cap \{\xi \in \mathbb{R}^n : \mathbf{a}^\top \xi \leq b\}$.*



Computing the volume of the knapsack polytope is known to be a difficult problem in the computational complexity literature, belonging to the class of #P-hard problems [45]. This means that the problem of

computing the volume of a knapsack polytope can be reduced to the problem of uncertainty quantification, thus showing classical chance constraints (even with a single inequality inside the probability expression) are intractable in general.

Example 1.4 (Non-convexity of Chance Constraints). Suppose a two-dimensional random variable $\tilde{\xi}$ has the discrete distribution $\tilde{\xi} \sim \frac{1}{3}\delta_{(-1,1)} + \frac{1}{3}\delta_{(0,0)} + \frac{1}{3}\delta_{(1,-1)}$, i.e. has three masts each with equal probability of $\frac{1}{3}$. Then, the set of (x_1, x_2) satisfying the chance constraint $\mathbb{P}\left(x_1 \leq \tilde{\xi}_1 \wedge x_2 \leq \tilde{\xi}_2\right) \geq \frac{1}{2}$ is indicated by the shaded area in the following diagram:



The intuition behind the non-convexity is that for any feasible (x_1, x_2) we require at least two of the support points to satisfy the inequalities inside the probability expression.

Example 1.5 (Binary Reformulation of Chance Constraints). Take the same discrete distribution as in Example 1.4, i.e. $\tilde{\xi} \sim \frac{1}{3}\delta_{(-1,1)} + \frac{1}{3}\delta_{(0,0)} + \frac{1}{3}\delta_{(1,-1)}$. Then

$$\begin{aligned}
 & \exists \mathbf{x} \in \mathbb{R}^2, \mathbf{b} \in \{0, 1\}^3 : \\
 & \frac{1}{3}b_1 + \frac{1}{3}b_2 + \frac{1}{3}b_3 \geq \frac{1}{2} \\
 \Leftrightarrow & \quad \begin{aligned}
 & \exists \mathbf{x} \in \mathbb{R}^2 : \\
 & P\left(\mathbf{x} \leq \tilde{\xi}\right) \geq \frac{1}{2} \\
 & b_1 = 1 \implies \mathbf{x} \leq [-1, 1]^\top \\
 & b_2 = 1 \implies \mathbf{x} \leq [1, -1]^\top \\
 & b_3 = 1 \implies \mathbf{x} \leq [0, 0]^\top.
 \end{aligned}
 \end{aligned}$$

The implication constraints involving binary variable can be directly reformulated into linear constraints using classical techniques from mixed-integer linear programming, such as big-M. Each binary variable corresponds to a scenario from the discrete distribution. The solver aims to set as many elements of $\mathbf{b}$ to 1 so that the total probability exceeds the required bound. However, setting b_i to 1 now enforces that the feasible $\mathbf{x}$ satisfies all inequality constraints for scenario i .

Another practical burden of (1.9) is that the distribution $\mathbb{P}$ must be known precisely in order to evaluate the left-hand side of the chance constraint. In many practical situations, $\mathbb{P}$ is estimated based on historical observations and is therefore itself subject to uncertainty. We have no hope of accurately describing the true distribution from empirical data as the dimension of $\tilde{\xi}$ increases. Indeed, we return to the situation in (1.1) where an inaccurate measurement of $\mathbb{P}$ will lead to overly-optimistic decisions, and perhaps infeasibility under the true distribution. We can apply the same strategy as we did in problem (1.5), by defining an ambiguity set of distributions $\mathcal{P}$. This ambiguity set contains all distributions which are consistent with the available a priori knowledge about the true distribution. This could include partial moments, support, bounds on measures of dispersion, or perhaps structural properties such as symmetry, unimodality, or independence patterns. We arrive finally at the ambiguous chance constrained problem considered in this thesis:

$$\begin{aligned} & \text{minimise} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && \inf_{\mathbb{P} \in \mathcal{P}} / \sup \mathbb{P} \left(g(\mathbf{x}, \tilde{\xi}) \leq 0 \right) \geq 1 - \varepsilon. \end{aligned} \tag{1.10}$$

If the inner optimisation problem in (1.10) is an infimum over $\mathcal{P}$, then the ambiguous chance constraint is called pessimistic, worst-case, or robust. It is equivalent to

$$\mathbb{P} \left(g(\mathbf{x}, \tilde{\xi}) \leq 0 \right) \geq 1 - \varepsilon \quad \forall \mathbb{P} \in \mathcal{P},$$

which follows the spirit of the robust optimisation problem (1.5). It turns out that distributionally robust chance constraints are much more amenable to equivalent tractable reformulations, than the standard chance constraint in (1.9). The first such result [102] considered a single-product newsvendor problem where $\mathcal{P}$ contains all probability distributions satisfying specified first- and second-order moments. Many other results have followed since, and we refer the reader to [6] for an overview of available tractable reformulations.

If, on the other hand, the inner optimisation problem in (1.10) is a supremum over $\mathcal{P}$ then the

ambiguous chance constraint is called optimistic or best-case. It is equivalent to

$$\exists \mathbb{P} \in \mathcal{P} : \mathbb{P} \left(g(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \varepsilon.$$

To the best of our knowledge, such problems have received much less attention in the literature.

1.2 Aims of the Thesis

This thesis considers optimisation problems of the form (1.10) and explores under what conditions (on f , g , $\mathcal{P}$, etc.) ambiguous chance constraints admit equivalent and tractable reformulations, i.e. can be solved with off-the-shelf optimisation software. We also motivate when and why such constraints should be included by presenting several contrasting examples. We also analyse, and get insight into precisely what the chance constraint is enforcing.

In general problem (1.10) is intractable. We will see that this manifests itself mostly through non-convex bilinearities between decision variables. Either the problem is inherently difficult, or we have just not found a clever equivalent representation of the problem. In the former case we aim to use computational complexity theory to prove that the respective classes of ambiguous chance constrained programs are indeed difficult, i.e. cannot be solved in polynomial time (unless $P=NP$).

1.3 Contributions and Structure

The rest of the thesis is divided into four chapters, which can be summarised as follows.

In Chapter 2 we consider the problem of uncertainty quantification and chance constrained programming in the distributionally robust setting. The aim of uncertainty quantification is to evaluate $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(g(\mathbf{x}, \boldsymbol{\xi}) \leq 0)$ for a fixed decision $\mathbf{x}$. If this value exceeds $1 - \varepsilon$, then we can certify that a given physical, engineering or economic system satisfies multiple safety conditions with high probability. This is the first step to a more ambitious goal of actively influencing the system so as to guarantee and maintain its safety: the chance constrained program. We assume that the parameters of the system are governed by an ambiguous distribution that is only known to belong to an ambiguity set characterised through generalised moment bounds and structural properties such as symmetry, unimodality or independence

patterns. We delineate the watershed between tractability and intractability in ambiguity-averse uncertainty quantification and chance constrained programming. Using tools from distributionally robust optimisation, we derive explicit conic reformulations for tractable problem classes and suggest efficiently computable conservative approximations for intractable ones. The results of this chapter have been published in [58]

- G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann. ‘A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming’. In: *Mathematical Programming B* (2015), pp. 1–28

Chapter 3 contains the core results on joint worst- and best-case ambiguous chance constraints. We consider a general setting where the ambiguity set contains support, mean, and general conic bounds on the dispersion constraints. We distinguish two types of uncertainty: left- and right-hand side. The former is present if there is any product between decision variable $\mathbf{x}$ and uncertainty $\tilde{\xi}$ inside the probability expression. These two classes have starkly different complexity properties. In general, problems with left-hand side uncertainty are NP-hard, even with the absence of the support constraints and with arguably the simplest polyhedral dispersion measure. Problems with right-hand side uncertainty fare better as we find specific classes of problems which yield an equivalent tractable conic representation. Nevertheless, we show that problems with right-hand side uncertainty are also NP-hard in general. Finally, we look at best-case ambiguous chance constraints and show that there are similar classes of problems which have tractable equivalents. For those problems where tractable reformulations are not yet known, we show how they can be conservatively approximated by an alternating convex optimisation procedure, or with a conditional value-at-risk approximation. The results of this chapter are contained in the following working paper [60]

- G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann. *Ambiguous Joint Chance Constraints with Conic Dispersion Measures*. Working Paper. Imperial College London and École Polytechnique Fédérale de Lausanne, 2015

In Chapter 4 we introduce a novel application of worst-case ambiguous chance constraints, where we consider the profit maximisation problem of a cloud computing service provider. The objective is

to select a price that maximises expected profit, while at the same time ensuring an adequate level of quality of service for customers. We assume the demand is uncertain and depends on the selected price. We maximise the worst-case expected profit subject to a distributionally robust chance constraint that guarantees a high level of service. The focus of this chapter is on distilling insights, and we derive closed-form solutions to this problem to gain further insights into the behaviour of the chance constraint. The results of this are contained in the following working paper [100]

- V. Roitch and D. Kuhn. *Distributionally Robust Pricing of Monopolistic Cloud Computing Services with Service Level Agreements*. Working Paper. Imperial College London and École Polytechnique Fédérale de Lausanne, 2015

Finally, in Chapter 5 we motivate the use of best-case ambiguous chance constraints. We consider the problem of image denoising, where, given a photographic image corrupted by noise, the objective is to recover the original image. There are many techniques for doing this, and we show that a well known and powerful approach [53] can be interpreted as a total variation minimisation problem subject to a best-case ambiguous chance constraint, where the noise is modelled as a random vector. We show that further insights can be gained from this chance constrained formulation. Additionally we show that for this type of problem, worst-case ambiguous chance constraints are unsuitable, thereby justifying the proposed approach further. However, considering worst-case ambiguous chance constraints for the image denoising problem is not in vain, since we show that a crucial condition of tractability in Chapter 3 is only sufficient, thus opening up the potential to expand the class of tractable ambiguous chance constraints. The results presented in this chapter are significantly extended and developed from the numerical example contained in [60], which is presented in Chapter 3.

1.4 Notation

Vectors are written with a bold lower-case font, e.g. $\mathbf{x} \in \mathbb{R}^n$. Elements of vectors (and other scalar quantities) are not bold, e.g. $x_i \in \mathbb{R}$. Matrices are written with bold upper-case font, e.g. $\mathbf{X} \in \mathbb{R}^{n \times m}$. Random variables are denoted with a tilde, e.g. $\tilde{\xi}$. By $\mathbb{R}_+$ and $\mathbb{R}_{++}$ we denote the set of non-negative and positive real numbers, respectively. The symbols $\mathbf{1}$ and $\mathbf{0}$ denote constant vectors/matrices of all

1s and 0s respectively, the dimensions of which is ensured to be clear from the context, e.g. in $\mathbf{x}^\top \mathbf{1}$ the dimension of $\mathbf{1}$ is understood to be the same as the dimension of $\mathbf{x}$. All inequalities and operators are element-wise by default, e.g. $\mathbf{x} \geq \mathbf{0}$ indicates that vector $\mathbf{x}$ is element-wise non-negative and $|\mathbf{x}|$ is the vector of absolute values of all elements of $\mathbf{x}$. Whenever an optimisation problem is presented, the decision variables and their domains are written as the first row of the constraints, e.g. in

$$\begin{aligned} & \text{minimise} && \dots \\ & \text{subject to} && \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}_+^m, \mathbf{z} \in \mathcal{Z} \\ & && \dots, \end{aligned}$$

the decision variables are $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. The indicator function of a subset A of a set X is written as $\mathbb{1}_A : X \rightarrow \{0, 1\}$. Inequalities appearing in the following style $\succsim_{\mathcal{K}}$ are reserved for generalised inequalities over cones $\mathcal{K}$, e.g. $\geq$ is equivalent to $\succsim_{\mathbb{R}_+}$. If no cone is explicitly given, then the positive semi-definite cone is assumed.

Chapter 2

Distributionally Robust Uncertainty Quantification and Chance Constraints

The objective of uncertainty quantification is to certify that a given physical, engineering or economic system satisfies multiple safety conditions with high probability. A more ambitious goal is to actively influence the system so as to guarantee and maintain its safety, a scenario which can be modelled through a chance constrained program. In this chapter we assume that the parameters of the system are governed by an ambiguous distribution that is only known to belong to an ambiguity set characterised through generalised moment bounds and structural properties such as symmetry, unimodality or independence patterns. We delineate the watershed between tractability and intractability in ambiguity-averse uncertainty quantification and chance constrained programming. Using tools from distributionally robust optimisation, we derive explicit conic reformulations for tractable problem classes and suggest efficiently computable conservative approximations for intractable ones.

2.1 Introduction

Consider a physical, engineering or economic system and encode its state through a parameter vector $\mathbf{z} \in \mathbb{R}^P$. Suppose that the reliable operation of the system requires that some safety constraints $\mathbf{S}\mathbf{z} \leq \mathbf{t}$ must be satisfied, where $\mathbf{S} \in \mathbb{R}^{J \times P}$ is termed the *technology matrix* and $\mathbf{t} \in \mathbb{R}^J$ the *right-hand side vector*. Due to measurement errors, limited observability and missing data, the parameter vector is uncertain for the vast majority of systems of practical interest and must therefore be modelled as a random variable $\tilde{\mathbf{z}}$ governed by a probability distribution $\mathbb{Q}$. In this situation a problem of great practical importance is to certify that the system is safe with high confidence. Formally, one should ascertain the satisfaction of the inequality $\mathbb{Q}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}] \geq 1 - \varepsilon$, where $\varepsilon \in (0, 1)$ represents a prescribed safety tolerance or violation probability. Examples of societally relevant safety constraints include the prevention of blackouts in electricity grids, the containment of inflation in national economies, the limitation of seismic damages in structural engineering, the assurance of quality of service standards in telecommunication systems, the limitation of the likelihood to develop cancer due to the exposure to a substance etc., see [88].

In most real-life applications, evaluating the exact probability of safe operation is very challenging, if not impossible. On the one hand, the probability distribution $\mathbb{Q}$ is typically unknown as $\mathbb{Q}$ may only be indirectly observable through historical samples, which could be explained by several strikingly different distributions. On the other hand, even if $\mathbb{Q}$ was precisely known, the computation of $\mathbb{Q}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$ would require the evaluation of an integral over a possibly high-dimensional polytope, which is computationally cumbersome.

A remedy for the first difficulty is to adopt a distributionally robust approach and to embrace the fact that $\mathbb{Q}$ is merely known to belong to an *ambiguity set* $\mathcal{P}$. This set is typically defined as the family of all distributions that share certain known moments (mean, variance, covariances, higher-order moments, median-absolute deviation etc.) or known structural properties (symmetry, unimodality, multimodality, independence patterns, tail behaviour etc.) with the otherwise unknown distribution $\mathbb{Q}$. The ambiguity of $\mathbb{Q}$ prompts us to investigate the *uncertainty quantification problem*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}], \quad (2.1)$$

which quantifies the worst-case probability of safe operation with respect to all distributions $\mathbb{P} \in \mathcal{P}$. As $\mathbb{Q} \in \mathcal{P}$ by construction, the optimal value of (2.1) provides a conservative estimate (lower bound) for $\mathbb{Q}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$. In order to certify the safety of the system with respect to the true distribution $\mathbb{Q}$, it is thus sufficient to show that the optimal value of (2.1) exceeds $1 - \varepsilon$. Maybe surprisingly, the distributionally robust approach can also mitigate the intractability of evaluating high-dimensional integrals. Using the duality theory for moment problems in conjunction with the rich arsenal of modern robust optimisation techniques, one can show that worst-case probability problems of the type (2.1) are computationally tractable across a wide variety of relevant ambiguity sets.

Rather than passively certifying the safety of a given system, a more ambitious goal would be to actively influence the system so as to maintain its safety. This scenario can conveniently be captured by a *robust chance constrained program*, where $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ represents the vector of design decisions.

$$\begin{aligned} & \text{minimise} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}(\mathbf{x})\tilde{\mathbf{z}} \leq \mathbf{t}(\mathbf{x})] \geq 1 - \varepsilon \end{aligned} \tag{2.2}$$

Here, the technology matrix $\mathbf{S}(\mathbf{x}) = (\mathbf{s}_1(\mathbf{x}), \dots, \mathbf{s}_J(\mathbf{x}))^\top$ and the right-hand side vector $\mathbf{t}(\mathbf{x}) = (t_1(\mathbf{x}), \dots, t_J(\mathbf{x}))^\top$ may depend on the design decisions in an affine fashion, that is, $\mathbf{s}_j(\mathbf{x}) = \mathbf{S}_j^\top \mathbf{x} + \mathbf{s}_j$ and $t_j(\mathbf{x}) = \mathbf{t}_j^\top \mathbf{x} + t_j$, where $\mathbf{S}_j \in \mathbb{R}^{N \times P}$, $\mathbf{s}_j \in \mathbb{R}^P$, $\mathbf{t}_j \in \mathbb{R}^N$ and $t_j \in \mathbb{R}$ for all $j \in \mathcal{J} = \{1, \dots, J\}$. Moreover, $\mathbf{c} \in \mathbb{R}^N$ encodes the costs of different design decisions. We assume that problem (2.2) is tractable if it is stripped of the probabilistic constraint. This is the case, for instance, if $\mathcal{X}$ is a polytope defined by its facets or vertices.

The choice of the ambiguity set $\mathcal{P}$ should be guided by the following principles: (i) The set $\mathcal{P}$ must contain $\mathbb{Q}$ with certainty (or at least with high confidence). (ii) The structure of $\mathcal{P}$ should facilitate a tractable reformulation (or at least a tractable conservative approximation) of the uncertainty quantification problem (2.1) and the chance constrained program (2.2). (iii) Among all ambiguity sets satisfying the properties (i) and (ii), $\mathcal{P}$ should be chosen as small as possible in the sense of set inclusion. Property (i) enables us to certify the safety of the system at hand under the unknown distribution $\mathbb{Q}$ by solving the uncertainty quantification problem (2.1). Moreover, property (ii) ensures that (2.1) and (2.2) can be

solved efficiently, while property (iii) controls the conservatism of the uncertainty quantification problem (2.1), thereby limiting the risk that a safe system is not recognised as such.

After the fundamental papers [106, 124], various ambiguity sets have been studied in the literature on uncertainty quantification [56, 88] and distributionally robust optimisation [6]. Ambiguity sets of special interest include the *Markov ambiguity set* containing all distributions with known mean and support [121], the *Chebyshev ambiguity set* containing all distributions with known bounds on the first and second-order moments [25, 32, 50, 71, 91, 115, 122, 125, 126], the *Gauss ambiguity set* containing all unimodal distributions from within the Chebyshev ambiguity set [90, 93], various generalised Chebyshev ambiguity sets that specify *asymmetric moments* [25, 26, 85], *higher-order moments* [29, 69, 111] or *marginal moments* [38, 39], the *median-absolute deviation ambiguity set* containing all symmetric distributions with known median and mean absolute deviation [60], the *Huber ambiguity set* containing all distributions with known upper bound on the expected Huber loss function [33, 121], the *Wasserstein ambiguity set* containing all distributions that are close to the empirical distribution with respect to the Wasserstein metric [42, 81, 92], the *Kullback-Leibler divergence ambiguity set* and *likelihood ratio ambiguity set* [19, 62, 63, 71, 120] containing all distributions that are sufficiently likely to have generated a given data set, the *Hoeffding ambiguity set* containing all component-wise independent distributions with a box support [10, 12, 19], the *Bernstein ambiguity set* containing all distributions from within the Hoeffding ambiguity set subject to marginal moment bounds [87], several ϕ -divergence-based ambiguity sets [7, 123] containing all discrete distributions close to a given nominal distribution, *goodness-of-fit ambiguity sets* containing all distributions that pass prescribed statistical tests [11] etc. The proposed terminology associates most ambiguity sets with mathematicians who invented well-known probability inequalities or statistical indicators related to the respective ambiguity sets.

In this chapter we endeavour to

- present a unifying framework for formulating and solving uncertainty quantification problems and robust chance constrained programs;
- demonstrate that many of the ambiguity sets listed above represent special cases of a canonical ambiguity set underlying our unifying framework;
- delineate the watershed between tractability and intractability in uncertainty quantification and

robust chance constrained programming.

In [121] it has been shown that most moment-based ambiguity sets emerge as special cases of a canonical ambiguity set that contains all distributions under which the probabilities of some conic-representable confidence sets fall between prescribed upper and lower bounds, and the mean values of the uncertain parameters satisfy a linear equality constraint. While [121] describes methods for computing worst-case expectations of biconvex loss functions, the focus of the present chapter is to compute worst-case probabilities, that is, worst-case expectations of discontinuous indicator functions. Moreover, while [121] focuses exclusively on moment-based ambiguity sets, the present chapter investigates a much richer class of ambiguity sets characterised both in terms of moment constraints and structural information such as symmetry, unimodality, multimodality, independence patterns etc. In particular, we also show that several classical inequalities of probability theory as well as their multidimensional generalisations emerge as special cases of our unified framework.

The overarching objective of this work is to review broad classes of uncertainty quantification and chance constrained programming problems that are computationally tractable. An intimately related secondary objective is to explore the boundaries of tractability. It is thus natural to focus attention on *linear* safety constraints in the probabilistic expressions of (2.1) and (2.2). Indeed, the uncertainty quantification problem (2.1) becomes intractable already in the presence a single convex quadratic safety constraint, even if the underlying ambiguity set contains all distributions supported on a polytope.

Theorem 2.1. *Evaluating the quadratic uncertainty quantification problem*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} [\|\mathbf{S}\tilde{\mathbf{z}}\|_2 \leq t] \tag{2.3}$$

is strongly NP-hard even if the ambiguity set satisfies

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{P}[\mathbf{C}\tilde{\mathbf{z}} \leq \mathbf{d}] = 1\},$$

where $\mathcal{P}_0(\mathbb{R}^P)$ denotes the set of all probability distributions on $\mathbb{R}^P$.

Proof. See Appendix. □

The remainder of the chapter develops as follows. Section 2.2 discusses the design of ambiguity sets using moment constraints and structural information. Sections 2.3 and 2.4 provide tractable reformulations and complexity results for various uncertainty quantification and chance constrained programming problems, respectively. An efficient approximation algorithm for intractable problems is reported in Section 3.2.3, and a summary of the main results is provided in Section 2.5. All proofs are relegated to an accompanying technical report [59].

Notation. A generalised inequality $\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y}$ with respect to a proper (closed, convex, pointed, solid) cone $\mathcal{K}$ implies that $\mathbf{y} - \mathbf{x} \in \mathcal{K}$. We denote by $\mathbb{S}^P$ ($\mathbb{S}_+^P$) the space (cone) of all symmetric (positive semidefinite) matrices in $\mathbb{R}^{P \times P}$ and use $X \preceq Y$ as a notational shorthand for the matrix inequality $X \preceq_{\mathbb{S}_+^P} Y$ where $X, Y \in \mathbb{S}^P$. The cone dual to a proper cone $\mathcal{K}$ is denoted as $\mathcal{K}^*$. We use $\mathcal{P}_0(B)$ to represent the set of all probability distributions supported on a Borel subset B of $\mathbb{R}^P$. If $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ represents the joint distribution of two random vectors $\tilde{\mathbf{z}} \in \mathbb{R}^P$ and $\tilde{\mathbf{u}} \in \mathbb{R}^Q$, then $\Pi_{\tilde{\mathbf{z}}} \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P)$ denotes the marginal distribution of $\tilde{\mathbf{z}}$ under $\mathbb{P}$. We extend this definition to ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ by setting $\Pi_{\tilde{\mathbf{z}}} \mathcal{P} = \bigcup_{\mathbb{P} \in \mathcal{P}} \{\Pi_{\tilde{\mathbf{z}}} \mathbb{P}\}$. For two sets A and B the relation $A \Subset B$ indicates that A is a subset of the relative interior of B . For a logical expression $\mathcal{E}$, we define $\mathbb{I}_{[\mathcal{E}]} = 1$ if $\mathcal{E}$ is true; $= 0$ otherwise.

2.2 Ambiguity Sets

We first propose a canonical representation of ambiguity sets as intersections of *moment ambiguity sets* (characterising features of $\mathbb{Q}$ such as the mean, variance or median-absolute deviation) and *structural ambiguity sets* (describing symmetry, unimodality or independence properties etc.). We define the moment and structural ambiguity sets of interest in Sections 2.2.1 and 2.2.2, respectively, and we showcase the expressiveness of our framework in Section 2.2.3.

2.2.1 Nested Moment Ambiguity Sets

As in [121], we focus on *nested moment ambiguity sets* of the form

$$\mathcal{P}^n = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \\ \mathbb{P}[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{C}_i] \in [\underline{p}_i, \bar{p}_i] \quad \forall i \in \mathcal{I} \end{array} \right\}, \quad (2.4)$$

where $\mathbb{P}$ is a joint distribution of the random vector $\tilde{\mathbf{z}} \in \mathbb{R}^P$ appearing in the uncertainty quantification and chance constrained programming problems and some auxiliary random vector $\tilde{\mathbf{u}} \in \mathbb{R}^Q$. We assume that $\mathbf{A} \in \mathbb{R}^{K \times P}$, $\mathbf{B} \in \mathbb{R}^{K \times Q}$, $\mathbf{b} \in \mathbb{R}^K$ and $\mathcal{I} = \{1, \dots, I\}$, while the confidence sets $\mathcal{C}_i$ satisfy

$$\mathcal{C}_i = \{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} \preceq_{\mathcal{K}_i} \mathbf{d}_i\} \quad (2.5)$$

with $\mathbf{C}_i \in \mathbb{R}^{L_i \times P}$, $\mathbf{D}_i \in \mathbb{R}^{L_i \times Q}$, $\mathbf{d}_i \in \mathbb{R}^{L_i}$ and $\mathcal{K}_i$ being proper cones. Note that the inclusion of the auxiliary vector $\tilde{\mathbf{u}}$ in $\mathcal{P}^n$ seems redundant as $\tilde{\mathbf{u}}$ could be absorbed in $\tilde{\mathbf{z}}$. In the next section, however, we will impose structural requirements on the marginal distribution of $\tilde{\mathbf{z}}$ that do not affect $\tilde{\mathbf{u}}$. We allow K or Q to be zero, in which case the expectation condition in (2.4) is void or the random vector $\tilde{\mathbf{u}}$ is absent, respectively. We also assume that $\mathcal{C}_i$ is essentially strictly feasible,¹ $\underline{p}_i, \bar{p}_i \in [0, 1]$ and $\underline{p}_i \leq \bar{p}_i$ for all $i \in \mathcal{I}$. Nested moment ambiguity sets of the form (2.4) have been used in [121] to evaluate worst-case expectations of convex functions. Here, we use them to compute worst-case probabilities of polytopes. As we will see shortly, allowing for multiple confidence sets $\mathcal{C}_i$ enables us to model a rich variety of ambiguity sets. By designing the confidence sets $\mathcal{C}_i$ appropriately, for example, we can prescribe the modality structure of $\tilde{\mathbf{z}}$, or we can specify confidence regions for the moments of $\tilde{\mathbf{z}}$ if these moments are estimated from historical samples, see [121].

In the remainder we impose the following regularity conditions.

- (B) The confidence set $\mathcal{C}_I$ has probability one, that is, $\underline{p}_I = \bar{p}_I = 1$, and all other confidence sets $\mathcal{C}_i$, $i = 1, \dots, I-1$, are bounded subsets of $\mathcal{C}_I$.
- (N) For all $i, i' \in \mathcal{I}$, $i \neq i'$, we have either $\mathcal{C}_i \subseteq \mathcal{C}_{i'}$, $\mathcal{C}_{i'} \subseteq \mathcal{C}_i$ or $\mathcal{C}_i \cap \mathcal{C}_{i'} = \emptyset$.

¹ We call $\mathcal{C}_i$ *essentially strictly feasible* if there is $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i$ that satisfies all non-polyhedral constraints in (2.5) strictly, see [8].

The boundedness condition **(B)** implies that the confidence set $\mathcal{C}_I$ contains the support of the joint random vector $(\tilde{\mathbf{z}}, \tilde{\mathbf{u}})$. This does not restrict generality since we are free to choose $\mathcal{C}_I = \mathbb{R}^P \times \mathbb{R}^Q$, but it simplifies some model formulations in later sections. Condition **(B)** also stipulates that the confidence sets $\mathcal{C}_i$, $i = 1, \dots, I - 1$, are bounded. This is necessary to obtain tractable formulations for the uncertainty quantification and chance constrained programming problems. The nesting condition **(N)** imposes a strict partial order on the confidence sets $\mathcal{C}_i$ with respect to the $\Subset$ -relation, and it also requires that incomparable sets are disjoint. We remark that for two sets $\mathcal{C}_i$ and $\mathcal{C}_{i'}$, the relation $\mathcal{C}_i \Subset \mathcal{C}_{i'}$ can be verified efficiently in many cases, for instance if both sets are polyhedral. All examples studied in this chapter satisfy the nesting condition by construction. The importance of the nesting condition is highlighted by the following result, which is proven in [121, Theorem 2].

Theorem 2.2. *Verifying whether the nested moment ambiguity set $\mathcal{P}^n$ defined in (2.4) is empty is strongly NP-hard even if $\mathcal{P}^n$ does not involve any expectation conditions (i.e., $K = 0$) and there are only two bounded (second-order) conic representable confidence sets $\mathcal{C}_1, \mathcal{C}_2$ with $\mathcal{C}_1 \subseteq \mathcal{C}_2$ but $\mathcal{C}_1 \not\Subset \mathcal{C}_2$.*

Proof. See [121, Theorem 2]. □

Theorem 2.2 implies that if the nesting condition **(N)** is violated, then the uncertainty quantification and chance constrained programming problems are strongly NP-hard. Some results in this chapter require that in addition to **(N)**, any set $\mathcal{C}_i$ that contains another set $\mathcal{C}_{i'}$, $\mathcal{C}_{i'} \Subset \mathcal{C}_i$, must have an affine dimension of at least 2. This dimensionality condition will be satisfied by all examples in this chapter, and it certainly holds for most applications of practical interest.

2.2.2 Structural Ambiguity Sets

As highlighted in [121], nested moment ambiguity sets of the form (2.4) allow us to model an abundance of (generalised) moment conditions. However, they fail to capture commonly encountered structural properties of the marginal distribution of $\tilde{\mathbf{z}}$. In the remainder of the chapter, we will thus intersect the nested moment ambiguity set (2.4) with various structural ambiguity sets that capture features such as symmetry, unimodality or independence. We require all structural ambiguity sets to be convex and weakly closed. For ease of exposition, we temporarily set $Q = 0$, thus assuming that there are no auxiliary random variables $\tilde{\mathbf{u}}$, but we will revoke this restriction at the end of this section.

In this work we focus on structural ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^P)$ that possess a *Choquet representation*, whereby every distribution $\mathbb{P} \in \mathcal{P}$ can be written as a mixture (i.e., an infinite convex combination) of extremal distributions of $\mathcal{P}$. Thus, for every Borel set $B \in \mathcal{B}(\mathbb{R}^P)$ we require that

$$\mathbb{P}[B] = \int_{\mathcal{V}} \mathbb{V}_{\mathbf{v}}(B) \mathbb{M}(\mathrm{d}\mathbf{v}), \quad (2.6)$$

where $\mathbb{V}_{\mathbf{v}}$, $\mathbf{v} \in \mathcal{V} \subseteq \mathbb{R}^P$, represents the family of extremal distributions (extreme points) of $\mathcal{P}$, and $\mathbb{M} \in \mathcal{P}_0(\mathcal{V})$ is the mixture distribution generating $\mathbb{P}$. This implies that the family of extreme points admits a finite-dimensional parameterisation. In distributionally robust optimisation such structural ambiguity sets were first studied in [93]. We show next that this abstract framework covers several practically relevant classes of structural ambiguity sets.

Symmetry Let $\mathcal{P}^s$ be the set of all point symmetric distributions on $\mathbb{R}^P$ with centre $\mathbf{m}$. Thus, $\mathbb{P} \in \mathcal{P}^s$ if and only if $\mathbb{P}[B] = \mathbb{P}[2\mathbf{m} - B]$ for all Borel sets $B \in \mathcal{B}(\mathbb{R}^P)$. The extremal distributions of $\mathcal{P}^s$ are

$$\mathbb{V}_{\mathbf{v}} = \frac{1}{2}\delta_{\mathbf{v}} + \frac{1}{2}\delta_{2\mathbf{m}-\mathbf{v}} \quad \text{for } \mathbf{v} \in \mathbb{R}^P,$$

where $\delta_{\mathbf{v}}$ and $\delta_{2\mathbf{m}-\mathbf{v}}$ denote the Dirac distributions that place all probability mass on the points $\mathbf{v}$ and $2\mathbf{m} - \mathbf{v}$, respectively.

Unimodality A distribution $\mathbb{P}$ is called unimodal with centre $\mathbf{m}$ if $\mathbb{P}[t(B - \mathbf{m})]/t^P$ is non-increasing in $t > 0$ for all $B \in \mathcal{B}(\mathbb{R}^P)$, see e.g. [35]. The definition implies that if $\mathbb{P}$ has a continuous density function $\rho(\mathbf{z})$, then $\mathbb{P}$ is unimodal if and only if $\rho(t(\mathbf{z} - \mathbf{m}))$ is non-increasing in $t > 0$ for any $\mathbf{z} \in \mathbb{R}^P$. The extreme points of the set of unimodal distributions are the radial distributions $\mathbb{V}_{\mathbf{v}}$, $\mathbf{v} \in \mathbb{R}^P$, that are supported on line segments from $\mathbf{m}$ to $\mathbf{v}$ and that satisfy

$$\mathbb{V}_{\mathbf{v}}([\mathbf{m}, \mathbf{m} + t(\mathbf{v} - \mathbf{m})]) = t^P \quad \forall t \in [0, 1].$$

α -Unimodality For $\alpha > 0$, let $\mathcal{P}^\alpha$ be the set of α -unimodal distributions with centre $\mathbf{m}$, that is, $\mathbb{P} \in \mathcal{P}^\alpha$ if and only if $\mathbb{P}[t(B - \mathbf{m})]/t^\alpha$ is non-increasing in $t > 0$ for all $B \in \mathcal{B}(\mathbb{R}^P)$, see [35]. Note that

an α -unimodal distribution on $\mathbb{R}^P$ is unimodal in the usual sense if $\alpha = P$. Moreover, if an α -unimodal distribution has a continuous density function $\rho(\mathbf{z})$, then $t^{\alpha-P}\rho(t(\mathbf{z} - \mathbf{m}))$ is non-increasing in $t > 0$ for any $\mathbf{z} \in \mathbb{R}^P$. The extreme points of $\mathcal{P}^\alpha$ are the radial distributions $\mathbb{V}_{\mathbf{v}}$, $\mathbf{v} \in \mathbb{R}^P$, that are supported on line segments $[\mathbf{m}, \mathbf{v}]$ and that satisfy

$$\mathbb{V}_{\mathbf{v}}([\mathbf{m}, \mathbf{m} + t(\mathbf{v} - \mathbf{m})]) = t^\alpha \quad \forall t \in [0, 1].$$

Multimodality Consider the ambiguity set generated by the convex combination $\sum_{r=1}^R \lambda_r \mathcal{P}^\alpha(\mathbf{m}_r)$ of α -unimodal ambiguity sets $\mathcal{P}^\alpha(\mathbf{m}_r)$ with centres $\mathbf{m}_r$, $r = 1, \dots, R$ and $\sum_{r=1}^R \lambda_r = 1$. This ambiguity set contains all α -multimodal distributions with a mode of probability mass $\lambda_r \geq 0$ centred at $\mathbf{m}_r$ for each $r = 1, \dots, R$. The extremal distributions of this ambiguity set are representable as $\mathbb{V}_{\mathbf{v}} = \sum_{r=1}^R \lambda_r \mathbb{V}_{r, \mathbf{v}_r}$ for $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_R) \in \mathbb{R}^{RP}$, where $\mathbb{V}_{r, \mathbf{v}_r}$ for $\mathbf{v}_r \in \mathbb{R}^P$ is any extremal distribution of the set $\mathcal{P}^\alpha(\mathbf{m}_r)$, $r = 1, \dots, R$.

Independence Let $\mathcal{P}^i$ be the set of all distributions on $\mathbb{R}^P$ under which the components of the random vector $\tilde{\mathbf{z}}$ are independent. One readily verifies that $\mathcal{P}^i$ violates our convexity assumption. Indeed, we have $\delta_{\mathbf{0}}, \delta_{\mathbf{e}} \in \mathcal{P}^i$, but the components of $\tilde{\mathbf{z}}$ are perfectly correlated under the mixture distribution $\frac{1}{2}\delta_{\mathbf{0}} + \frac{1}{2}\delta_{\mathbf{e}}$. We will show in Section 2.3 that the uncertainty quantification and chance constrained programming problems are typically intractable for ambiguity sets that impose independence among the components of $\tilde{\mathbf{z}}$.

Remark 2.1. *In the presence of auxiliary random variables $\tilde{\mathbf{u}} \in \mathbb{R}^Q$ with $Q > 0$, the above structural ambiguity sets are redefined as the families of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{P}^P \times \mathbb{R}^Q)$ whose marginal projections $\Pi_{\tilde{\mathbf{z}}} \mathbb{P}$ display the structural properties (e.g., symmetry or α -unimodality etc.) outlined above.*

2.2.3 Examples

The ambiguity sets that can be generated by intersecting a nested moment ambiguity set of the form (2.4) with a Choquet-representable structural ambiguity set display a remarkable diversity. We now show that many ambiguity sets from the recent literature can be expressed as instances of this class.

Example 2.1 (Chebyshev Ambiguity Set). Let $\mathcal{P}$ be the ambiguity set of all distributions on $\mathbb{R}^P$ with mean $\boldsymbol{\mu} \in \mathbb{R}^P$ whose covariance matrix is bounded above by $\boldsymbol{\Sigma} \in \mathbb{S}_+^P$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma} \right\}. \quad (2.7)$$

Consider the following instance of the nested moment ambiguity set (2.4), which involves the auxiliary random matrix $\tilde{\mathbf{U}} \in \mathbb{R}^{P \times P}$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{U}}] = \boldsymbol{\Sigma}, \\ \mathbb{P} \left[\begin{pmatrix} 1 & (\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \\ (\tilde{\mathbf{z}} - \boldsymbol{\mu}) & \tilde{\mathbf{U}} \end{pmatrix} \succcurlyeq \mathbf{0} \right] = 1 \end{array} \right\} \quad (2.8)$$

Since $\mathcal{P}'$ only contains one confidence set with probability bounds $\underline{p}_1 = \bar{p}_1 = 1$, the boundedness condition (B) and the nesting condition (N) are trivially satisfied. Moreover, a Schur complement argument implies that $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$. The Chebyshev ambiguity set has been studied extensively in distributionally robust optimisation [32, 50, 121, 125] and can also be generalised to account for uncertainty in the mean and/or covariance matrix of $\tilde{\mathbf{z}}$, see e.g. [32, 121].

If we intersect the nested moment ambiguity set from Example 2.1 with the structural ambiguity set of all unimodal distributions, then we recover an ambiguity set that has been studied in [90] and is closely related to a tightened Chebyshev-type inequality due to Gauss [47].

Example 2.2 (Gauss Ambiguity Set). Let $\mathcal{P}$ be the ambiguity set of all unimodal distributions on $\mathbb{R}^P$ with centre $\mathbf{m} \in \mathbb{R}^P$ and mean $\boldsymbol{\mu} \in \mathbb{R}^P$ whose covariance matrix is bounded above by $\boldsymbol{\Sigma} \in \mathbb{S}_+^P$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{P} \text{ is unimodal with centre } \mathbf{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma} \end{array} \right\}. \quad (2.9)$$

In analogy to the previous example, the ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \begin{array}{l} \Pi_{\tilde{\mathbf{z}}} \mathbb{P} \text{ is unimodal with centre } \mathbf{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{U}}] = \boldsymbol{\Sigma}, \\ \mathbb{P} \left[\begin{pmatrix} 1 & (\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \\ (\tilde{\mathbf{z}} - \boldsymbol{\mu}) & \tilde{\mathbf{U}} \end{pmatrix} \succcurlyeq \mathbf{0} \right] = 1 \end{array} \right\} \quad (2.10)$$

satisfies **(B)** and **(N)** as well as $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$.

Examples 2.1 and 2.2 rely on classical statistical indicators—the mean and variance—to characterise the unknown distribution $\mathbb{Q}$. In the next two examples, we describe $\mathbb{Q}$ through location and dispersion measures from robust statistics, namely the median, the median-absolute deviation and the Huber loss function. While reminiscent of the mean and variance, these indicators may be more reliable as they are easier to estimate from data, see [20, 121].

Example 2.3 (Median-Absolute Deviation Ambiguity Set). *Let $\mathcal{P}$ be the ambiguity set of all symmetric distributions on $\mathbb{R}^P$ with centre $\mathbf{m} \in \mathbb{R}^P$ whose median absolute deviation is bounded above by $\mathbf{f} \in \mathbb{R}_+^P$, that is,*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{P} \text{ is symmetric with centre } \mathbf{m}, \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\mathbf{z}} - \mathbf{m}|] \leq \mathbf{f} \end{array} \right\}, \quad (2.11)$$

where the absolute value is understood to apply component-wise. Note that for symmetric distributions, the median coincides with the centre. Consider the following instance of (2.4), which involves the auxiliary random vector $\tilde{\mathbf{u}} \in \mathbb{R}^P$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \begin{array}{l} \Pi_{\tilde{\mathbf{z}}} \mathbb{P} \text{ is symmetric with centre } \mathbf{m}, \\ \mathbb{P}[\tilde{\mathbf{u}} \geq \tilde{\mathbf{z}} - \mathbf{m}, \tilde{\mathbf{u}} \geq \mathbf{m} - \tilde{\mathbf{z}}] = 1, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f} \end{array} \right\} \quad (2.12)$$

Since $\mathcal{P}'$ involves only a single confidence set with probability bounds $\underline{p}_1 = \bar{p}_1 = 1$, the boundedness condition **(B)** and the nesting condition **(N)** are trivially satisfied. Moreover, one readily verifies that $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$.

Example 2.4 (Huber Ambiguity Set). *Consider the ambiguity set*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[H_{\beta}(\mathbf{f}^{\top}[\tilde{\mathbf{z}} - \boldsymbol{\mu}])] \leq g \right\},$$

where $\boldsymbol{\mu}, \mathbf{f} \in \mathbb{R}^P$, $g \in \mathbb{R}_+$ and $H_{\beta}(z)$ is the Huber loss function with prescribed robustness parameter $\beta > 0$, that is, $H_{\beta}(z) = \frac{1}{2}z^2$ if $|z| \leq \beta$; $= \beta \left(|z| - \frac{1}{2}\beta \right)$ otherwise, see [65]. The expected Huber loss function represents a robust dispersion measure that generalises the variance (for $\beta \rightarrow \infty$) and the mean absolute deviation (for $\beta \rightarrow 0$). Consider now the following instance of (2.4), which involves the auxiliary random variables $\tilde{u}, \tilde{v}, \tilde{w} \in \mathbb{R}_+$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}_+^3) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[\tilde{w}] = g, \\ \mathbb{P} \left[\frac{1}{2} (\mathbf{f}^{\top} \tilde{\mathbf{z}} + \tilde{u} - \tilde{v})^2 + \beta(\tilde{u} + \tilde{v}) \leq \tilde{w} \right] = 1 \end{array} \right\} \quad (2.13)$$

It has been shown in [121] that $\mathcal{P}'$ satisfies $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$. Since $\mathcal{P}'$ only involves a single confidence set with probability bounds $\underline{p}_1 = \bar{p}_1 = 1$, the boundedness condition (B) and the nesting condition (N) are trivially satisfied.

We close this section by reviewing a data-driven ambiguity set that can be constructed directly from independent samples of the unknown distribution $\mathbb{Q}$. In contrast to the previous examples, this ambiguity set converges to the singleton $\{\mathbb{Q}\}$ as the number of available samples tends to infinity. For a detailed discussion of this ambiguity set we refer to [81].

Example 2.5 (Wasserstein Ambiguity Set). *Suppose that we have observed independent samples $\hat{\mathbf{z}}_i$, $i \in \mathcal{I} = \{1, \dots, I\}$, of an unknown data-generating distribution $\mathbb{Q}$ with bounded support, and assume for ease of exposition that $\hat{\mathbf{z}}_i \neq \hat{\mathbf{z}}_j$ for $i, j \in \mathcal{I}$, $i \neq j$. If we denote by $\hat{\mathbb{P}} = \frac{1}{I} \sum_{i=1}^I \delta_{\hat{\mathbf{z}}_i}$ the empirical distribution, then the Wasserstein ambiguity set of size $r > 0$ is defined as*

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : d_p^w(\mathbb{P}, \hat{\mathbb{P}}) \leq r \} \quad (2.14)$$

and thus contains all distributions that reside within the ball of radius r around $\hat{\mathbb{P}}$ with respect to the Wasserstein metric of order $p \geq 1$. For any (marginal) distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ of two P -dimensional

random vectors $\tilde{\mathbf{z}}_1$ and $\tilde{\mathbf{z}}_2$, respectively, the Wasserstein distance $d_p^w(\mathbb{P}_1, \mathbb{P}_2)$ of order p is defined as

$$\begin{aligned} d_p^w(\mathbb{P}_1, \mathbb{P}_2) = & \inf \quad \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}_2\|_p] \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) \\ & \left. \begin{aligned} \mathbb{P}[\tilde{\mathbf{z}}_1 \in B] &= \mathbb{P}_1[\tilde{\mathbf{z}}_1 \in B] \\ \mathbb{P}[\tilde{\mathbf{z}}_2 \in B] &= \mathbb{P}_2[\tilde{\mathbf{z}}_2 \in B] \end{aligned} \right\} \quad \forall B \in \mathcal{B}(\mathbb{R}^P), \end{aligned} \quad (2.15)$$

see e.g. [96]. Note that (2.15) can be viewed as a transportation problem that minimises the expectation of $\|\tilde{\mathbf{z}}_1 - \tilde{\mathbf{z}}_2\|_p$ over all possible joint distributions $\mathbb{P}$ of $\tilde{\mathbf{z}}_1$ and $\tilde{\mathbf{z}}_2$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. If $\bar{u} \in \mathbb{R}_+$ denotes an upper bound on the Euclidean diameter of the support of $\mathbb{Q}$, then the ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R}) : \begin{aligned} & \mathbb{E}_{\mathbb{P}}[\tilde{u}_0] = r, \quad \mathbb{P}[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}, \tilde{u}_0) \in \mathbb{R}^{2P+1}] = 1, \\ & \mathbb{P} \left[\begin{aligned} \|\tilde{\mathbf{z}} - \tilde{\mathbf{u}}\|_p &\leq \tilde{u}_0, \\ \tilde{u}_0 &\leq \bar{u}, \quad \tilde{\mathbf{u}} = \hat{\mathbf{z}}_i \end{aligned} \right] = 1/I \quad \forall i = 1, \dots, I \end{aligned} \right\} \quad (2.16)$$

constitutes an instance of (2.4) that satisfies the boundedness condition **(B)** and the nesting condition **(N)**. Moreover, we have $\Pi_{\tilde{\mathbf{u}}} \mathcal{P}' = \{\hat{\mathbb{P}}\}$ and $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$.

2.3 Uncertainty Quantification

In this section we develop tractable reformulations for instances of the uncertainty quantification problem (2.1) under the premise that $\mathcal{P}$ constitutes a nested ambiguity set of the form (2.4) or an intersection of (2.4) with the set of symmetric or α -unimodal distributions. We also show that the uncertainty quantification problem is generically intractable for ambiguity sets that impose independence among the components of $\tilde{\mathbf{z}}$. The results of this section generalise the results of [60] to instances of (2.4) with $I > 1$ confidence sets.

Our tractability results rely on an interpretation of the uncertainty quantification problem (2.1) as a generalised moment problem of the type [103, Equation (3.2)] whose semi-infinite dual [103, Equation (3.4)] lends itself for further simplification. Strong duality holds under the Slater condition [103, Equation (3.12)], which we henceforth abbreviate as **(S)**. For non-empty ambiguity sets, this condition

is non-restrictive and can always be enforced by slightly perturbing the parameters $\mathbf{b}$, $\underline{p}_i$ and $\bar{p}_i$, but it is cumbersome to state and verify explicitly for the generic nested ambiguity set (2.4). Many examples considered in this chapter involve only a single confidence set ($I = 1$) and no structural information, in which case the Slater condition (S) simplifies to the requirement that $\mathbf{b}$ belongs to the interior of the convex set $\{\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u} : (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1\}$.

We are now in the position to state our first tractability result.

Theorem 2.3 (Moment Ambiguity Sets). *If $\mathcal{P} = \mathcal{P}^n$ is an instance of (2.4) that satisfies the boundedness condition (B), the nesting condition (N) and the Slater condition (S), then the worst-case probability (2.1) coincides with the optimal value of the conic optimisation problem*

$$\begin{aligned}
& \sup \quad \mathbf{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \bar{p}_i \kappa_i \\
& \text{s.t.} \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I, \quad \boldsymbol{\phi}_i \in \mathcal{K}_i^*, \quad i \in \mathcal{I} \\
& \quad \tau_{ij} \in \mathbb{R}_+, \quad (i, j) \in \mathcal{L}, \quad \boldsymbol{\psi}_{ij} \in \mathcal{K}_i^*, \quad (i, j) \in \mathcal{L} \\
& \quad \left. \begin{aligned} & \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \mathbf{d}_i^\top \boldsymbol{\phi}_i \leq 1 \\ & \mathbf{A}^\top \boldsymbol{\gamma} = \mathbf{C}_i^\top \boldsymbol{\phi}_i, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_i^\top \boldsymbol{\phi}_i \end{aligned} \right\} \quad \forall i \in \mathcal{I} \\
& \quad \left. \begin{aligned} & \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \mathbf{d}_i^\top \boldsymbol{\psi}_{ij} - \tau_{ij} t_j \leq 0 \\ & \mathbf{A}^\top \boldsymbol{\gamma} + \tau_{ij} \mathbf{s}_j = \mathbf{C}_i^\top \boldsymbol{\psi}_{ij}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_i^\top \boldsymbol{\psi}_{ij} \end{aligned} \right\} \quad \forall (i, j) \in \mathcal{L},
\end{aligned} \tag{2.17}$$

where $\mathcal{A}(i) = \{i\} \cup \{i' \in \mathcal{I} : \mathcal{C}_i \subseteq \mathcal{C}_{i'}\}$ represents the index set of all supersets (antecedents) of $\mathcal{C}_i$, while $\mathcal{L} = \{(i, j) \in \mathcal{I} \times \mathcal{J} : \exists (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i \cdot \mathbf{s}_j^\top \mathbf{z} > t_j\}$.

Proof. See Appendix. □

The index set $\mathcal{L}$ contains the pair $(i, j) \in \mathcal{I} \times \mathcal{J}$ if there are realisations $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i$ that violate the j -th constraint in (2.1). One can verify efficiently whether $(i, j) \in \mathcal{L}$ by checking whether the optimal value of the convex optimisation problem $\sup \{\mathbf{s}_j^\top \mathbf{z} : (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i\}$ is strictly larger than t_j .

Problem (2.17) is a conic optimisation problem whose size scales polynomially in the size of $\mathbf{S}$ and $\mathbf{t}$ in problem (2.1) as well as the description of the ambiguity set $\mathcal{P}^n$ in (2.4). Moreover, if all cones $\mathcal{K}_i$ are polyhedral, then (2.17) is a linear program. We illustrate Theorem 2.3 with two examples.

Example 2.6 (Generalised Chebyshev Bounds). *Theorem 2.3 allows us to compute the worst-case probability of the event $\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}$ if the distribution of $\tilde{\mathbf{z}}$ is only known to belong to the Chebyshev ambiguity set (2.7) from Example 2.1. Thereby, we recover a generalised multivariate Chebyshev inequality that was discovered in [115]. As $\mathcal{P} = \Pi_{\tilde{\mathbf{z}}}\mathcal{P}'$, where $\mathcal{P}'$ is defined in (2.8), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$. For $\Sigma \succ \mathbf{0}$, $\mathcal{P}'$ is an instance of (2.4) that satisfies the conditions (B), (S) and (N). Thus, we can use Theorem 2.3 to reformulate the uncertainty quantification problem as the semidefinite program*

$$\begin{aligned} \sup \quad & \beta - \boldsymbol{\mu}^\top \boldsymbol{\gamma} - \langle \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top, \boldsymbol{\Gamma} \rangle \\ \text{s.t.} \quad & \beta \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}^P, \quad \boldsymbol{\Gamma} \in \mathbb{S}_+^P, \quad \tau_j \in \mathbb{R}_+, \quad j \in \mathcal{L} \\ & \begin{pmatrix} 1 - \beta & \frac{1}{2}\boldsymbol{\gamma}^\top \\ \frac{1}{2}\boldsymbol{\gamma} & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}, \quad \begin{pmatrix} \tau_j t_j - \beta & \frac{1}{2}(\boldsymbol{\gamma} + \tau_j \mathbf{s}_j)^\top \\ \frac{1}{2}(\boldsymbol{\gamma} + \tau_j \mathbf{s}_j) & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0} \quad \forall j \in \mathcal{L}, \end{aligned} \quad (2.18)$$

where $\mathcal{L} = \{j \in \mathcal{J} : \mathbf{s}_j \neq \mathbf{0} \vee t_j < 0\}$.

Example 2.7 (Data-Driven Uncertainty Quantification). *Theorem 2.3 further allows us to compute the worst-case probability of the event $\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}$ if the distribution of $\tilde{\mathbf{z}}$ is an element of the Wasserstein ambiguity set (2.14) from Example 2.5. Thereby we recover a data-driven probability inequality first discovered in [81]. As $\mathcal{P} = \Pi_{\tilde{\mathbf{z}}}\mathcal{P}'$, where $\mathcal{P}'$ is defined in (2.16), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$. Moreover, $\mathcal{P}'$ is an instance of (2.4) that satisfies the conditions (B) and (N). Even though $\mathcal{P}'$ fails to satisfy the Slater condition (S), one can show that Theorem 2.3 remains valid [81]. Thus, the uncertainty quantification problem (2.1) can be reformulated as the convex optimisation problem*

$$\begin{aligned} \sup \quad & \frac{1}{I} \sum_{i=1}^I \beta_i - \gamma r \\ \text{s.t.} \quad & \boldsymbol{\beta} \in \mathbb{R}^I, \quad \gamma \in \mathbb{R}_+, \quad \boldsymbol{\tau} \in \mathbb{R}_+^{I \times J} \\ & \left. \begin{aligned} \beta_i &\leq 1 \\ \|\tau_{ij} \mathbf{s}_j\|_q &\leq \gamma \\ \beta_i + \tau_{ij} \mathbf{s}_j^\top \hat{\mathbf{z}}_i &\leq \tau_{ij} t_j \end{aligned} \right\} \forall j \in \mathcal{J} \quad \left. \vphantom{\begin{aligned} \beta_i &\leq 1 \\ \|\tau_{ij} \mathbf{s}_j\|_q &\leq \gamma \\ \beta_i + \tau_{ij} \mathbf{s}_j^\top \hat{\mathbf{z}}_i &\leq \tau_{ij} t_j \end{aligned}} \right\} \forall i = 1, \dots, I, \end{aligned} \quad (2.19)$$

where q is defined through $\frac{1}{p} + \frac{1}{q} = 1$. Note that as $r \rightarrow 0$, the variable γ can be driven to infinity at essentially no cost. Thus, the optimal value of (2.19) converges to the fraction of the samples $\hat{\mathbf{z}}_i$, $i = 1, \dots, I$, that satisfy $\mathbf{S}\hat{\mathbf{z}}_i < \mathbf{t}$. Problem (2.19) reduces to a linear program for $p \in \{1, \infty\}$.

We now consider instances of the uncertainty quantification problem (2.1) where $\mathcal{P}$ emerges from the intersection of a nested moment ambiguity set $\mathcal{P}^n$ of the form (2.4) with the set of all symmetric distributions $\mathcal{P}^s$ centred around $\mathbf{m}$. In order to derive a tractable reformulation for this problem class, we require that $\mathcal{P}^n$ satisfies the following technical dimensionality condition:

- (D) The ambiguity set satisfies $I = 1$ (support only), or it satisfies $Q > 0$ (the vector $\tilde{\mathbf{u}}$ is not absent) and all confidence sets $\mathcal{C}_i$, $i \in \mathcal{I}$, are bounded.

Moreover, for the multi-indices $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ ranging over pairs of confidence sets and $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ indexing pairs of inequalities in the uncertainty quantification problem, we define the set

$$\mathcal{D}_{\mathbf{ij}} = \left\{ (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : \begin{array}{ll} (\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_{i^-} \\ j^+ > 0 \implies \mathbf{s}_{j^+}^\top \mathbf{z} > t_{j^+} \\ j^- > 0 \implies -\mathbf{s}_{j^-}^\top \mathbf{z} > t_{j^-} \end{array} \right\}, \quad (2.20)$$

and we impose the following feasibility condition.

- (F) For any $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ and $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$, if $\mathcal{D}_{\mathbf{ij}} \neq \emptyset$, then $\mathcal{D}_{\mathbf{ij}}$ is essentially strictly feasible.

Condition (F) is a mild technical condition that is satisfied, for example, if the cones $\mathcal{K}_i$, $i \in \mathcal{I}$, are polyhedral.

Theorem 2.4 (Symmetry). *Assume that $\mathcal{P} = \mathcal{P}^n \cap \mathcal{P}^s$, where $\mathcal{P}^n$ is an instance of the nested moment ambiguity set (2.4) that satisfies the boundedness condition (B), the nesting condition (N), the dimensionality condition (D) and the feasibility condition (F), whereas $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ under which $\Pi_{\tilde{\mathbf{z}}} \mathbb{P}$ is point symmetric around $\mathbf{m} = \mathbf{0}$. If the Slater condition (S) holds,*

then the worst-case probability (2.1) coincides with the optimal value of the conic program

$$\begin{aligned}
& \sup \quad \mathbf{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \bar{p}_i \kappa_i \\
& \text{s.t.} \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I \\
& \quad \psi_{\mathbf{ij}}^+ \in \mathcal{K}_{i^+}^*, \quad \psi_{\mathbf{ij}}^- \in \mathcal{K}_{i^-}^*, \quad \chi_{\mathbf{ij}}^+, \chi_{\mathbf{ij}}^- \in \mathbb{R}_+, \quad (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \\
& \quad \left. \begin{aligned} & \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \\ & + \mathbf{d}_{i^+}^\top \psi_{\mathbf{ij}}^+ + \mathbf{d}_{i^-}^\top \psi_{\mathbf{ij}}^- - \chi_{\mathbf{ij}}^+ t_{j^+} - \chi_{\mathbf{ij}}^- t_{j^-} \leq w_{\mathbf{j}} \\ & \mathbf{C}_{i^+}^\top \psi_{\mathbf{ij}}^+ - \mathbf{C}_{i^-}^\top \psi_{\mathbf{ij}}^- = \chi_{\mathbf{ij}}^+ \mathbf{s}_{j^+} - \chi_{\mathbf{ij}}^- \mathbf{s}_{j^-} \\ & \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_{i^+}^\top \psi_{\mathbf{ij}}^+ = \mathbf{D}_{i^-}^\top \psi_{\mathbf{ij}}^- \end{aligned} \right\} \quad \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L}, \tag{2.21}
\end{aligned}$$

where $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ and $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ are multi-indices,

$$\mathcal{L} = \{(\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2 : \mathcal{D}_{\mathbf{ij}} \neq \emptyset\}$$

with $\mathcal{D}_{\mathbf{ij}}$ defined in (2.20), $w_{\mathbf{j}} = \mathbb{I}_{[j^+=0]} + \mathbb{I}_{[j^-=0]}$, $\mathbf{s}_0 = \mathbf{0}$ and $t_0 = 0$.

Proof. See Appendix. □

In analogy to Theorem 2.3, Theorem 2.4 provides a reformulation that scales polynomially in the input data. Note also that (2.21) reduces to a linear program whenever the cones $\mathcal{K}_i$ for $i \in \mathcal{I}$ are polyhedral.

We emphasise that in Theorem 2.4, the centre of symmetry is set to $\mathbf{0}$ merely to simplify the exposition. This is without loss of generality and can always be accomplished by a coordinate shift. Moreover, we note that one can efficiently verify whether $(\mathbf{i}, \mathbf{j}) \in \mathcal{L}$. If $\mathbf{j} > \mathbf{0}$, for example, we have $(\mathbf{i}, \mathbf{j}) \in \mathcal{L}$ if and only if the optimal value of the convex optimisation problem

$$\begin{aligned}
& \sup \quad \min \{ \mathbf{s}_{j^+}^\top \mathbf{z} - t_{j^+}, -\mathbf{s}_{j^-}^\top \mathbf{z} - t_{j^-} \} \\
& \text{s.t.} \quad (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q \\
& \quad (\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_{i^+}, \quad (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_{i^-}
\end{aligned}$$

is strictly positive. We illustrate Theorem 2.4 with two examples.

Example 2.8 (Uncertainty Quantification with Robust Dispersion Measures). *Let $\mathcal{P}$ be the median-absolute deviation ambiguity set (2.11) from Example 2.3. Note that $\mathcal{P} = \Pi_{\mathbf{z}}\mathcal{P}'$, where $\mathcal{P}'$ is defined in (2.12). Moreover, $\mathcal{P}'$ can be viewed as the intersection of a moment ambiguity set $\mathcal{P}^n$ of the form (2.4) with $I = 1$ and a structural ambiguity set $\mathcal{P}^s$ containing all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P)$ under which $\Pi_{\mathbf{z}}\mathbb{P}$ is symmetric around $\mathbf{m}$. For $\mathbf{f} > \mathbf{0}$, $\mathcal{P}'$ satisfies the conditions (B), (N), (D), (F) and (S). Thus, we can employ a coordinate shift and use Theorem 2.4 to reformulate the uncertainty quantification problem (2.1) as the tractable linear program*

$$\begin{aligned} \sup \quad & \alpha - \mathbf{f}^\top \boldsymbol{\gamma} \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^K, \quad \chi_{\mathbf{j}}^+, \chi_{\mathbf{j}}^- \in \mathbb{R}_+, \mathbf{j} \in \mathcal{L} \\ & \left. \begin{aligned} 2\alpha + \chi_{\mathbf{j}}^+ (\mathbf{s}_{j^+}^\top \mathbf{m} - t_{j^+}) + \chi_{\mathbf{j}}^- (\mathbf{s}_{j^-}^\top \mathbf{m} - t_{j^-}) &\leq w_{\mathbf{j}} \\ \chi_{\mathbf{j}}^+ \mathbf{s}_{j^+} - \chi_{\mathbf{j}}^- \mathbf{s}_{j^-} &\leq 2\boldsymbol{\gamma}, \quad \chi_{\mathbf{j}}^- \mathbf{s}_{j^-} - \chi_{\mathbf{j}}^+ \mathbf{s}_{j^+} \leq 2\boldsymbol{\gamma} \end{aligned} \right\} \quad \forall \mathbf{j} \in \mathcal{L}, \end{aligned}$$

where $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ is a multi-index ranging over

$$\mathcal{L} = \left\{ \mathbf{j} \in (\mathcal{J} \cup \{0\})^2 : \left[\begin{array}{ll} \exists \mathbf{z} \in \mathbb{R}^P : & j^+ > 0 \implies \mathbf{s}_{j^+}^\top (\mathbf{m} + \mathbf{z}) > t_{j^+} \\ & j^- > 0 \implies \mathbf{s}_{j^-}^\top (\mathbf{m} - \mathbf{z}) > t_{j^-} \end{array} \right] \right\},$$

$w_{\mathbf{j}} = \mathbb{I}_{[j^+=0]} + \mathbb{I}_{[j^-=0]}$, $\mathbf{s}_0 = \mathbf{0}$ and $t_0 = 0$.

Example 2.9 (Data-Driven Uncertainty Quantification, cont'd). *Let $\mathcal{P} = \mathcal{P}' \cap \mathcal{P}^s$, where $\mathcal{P}'$ is the lifted Wasserstein ambiguity set (2.16) from Example 2.5 and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R})$ with the property that $\Pi_{\mathbf{z}}\mathbb{P}$ is point symmetric around $\mathbf{m}$. If $\bar{u}$ is a strict upper bound on the Euclidean diameter of the support of $\mathbb{Q}$, then the ambiguity set $\mathcal{P}'$ satisfies the conditions (B), (N), (D) and (F). Even though $\mathcal{P}'$ fails to satisfy (S), it can be shown that Theorem 2.4 remains valid if r is sufficiently large. Thus, we can reformulate the uncertainty quantification problem as the convex*

optimisation problem

$$\begin{aligned}
& \sup \quad \frac{1}{I} \sum_{i \in \mathcal{I}} \beta_i - \gamma r \\
& \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{R}^I, \quad \gamma \in \mathbb{R}_+ \\
& \quad \chi_{\mathbf{i}\mathbf{j}}^+, \chi_{\mathbf{i}\mathbf{j}}^- \in \mathbb{R}_+, \quad \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^+, \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^- \in \mathbb{R}^P, \quad (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \\
& \quad \left. \begin{aligned} & \|\boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^+\|_q \leq \gamma, \quad \|\boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^-\|_q \leq \gamma, \quad \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^+ + \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^- = \chi_{\mathbf{i}\mathbf{j}}^+ \mathbf{s}_{j+} - \chi_{\mathbf{i}\mathbf{j}}^- \mathbf{s}_{j-} \\ & \beta_{i+} + \beta_{i-} + \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^{+\top} (\hat{\mathbf{z}}_{i+} - \mathbf{m}) + \boldsymbol{\nu}_{\mathbf{i}\mathbf{j}}^{-\top} (\mathbf{m} - \hat{\mathbf{z}}_{i-}) \\ & \quad + \chi_{\mathbf{i}\mathbf{j}}^+ (\mathbf{s}_{j+}^\top \mathbf{m} - t_{j+}) + \chi_{\mathbf{i}\mathbf{j}}^- (\mathbf{s}_{j-}^\top \mathbf{m} - t_{j-}) \leq w_{\mathbf{j}} \end{aligned} \right\} \quad \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L},
\end{aligned}$$

where $\mathbf{i} = (i^+, i^-)$ and $\mathbf{j} = (j^+, j^-)$ are multi-indices ranging over

$$\mathcal{L} = \left\{ (\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2 : \left[\exists \mathbf{z} \in \mathbb{R}^P : \begin{aligned} j^+ > 0 & \implies \mathbf{s}_{j+}^\top (\mathbf{m} + \mathbf{z}) > t_{j+} \\ j^- > 0 & \implies \mathbf{s}_{j-}^\top (\mathbf{m} - \mathbf{z}) > t_{j-} \end{aligned} \right] \right\},$$

$w_{\mathbf{j}} = \mathbb{I}_{[j^+=0]} + \mathbb{I}_{[j^-=0]}$, $\mathbf{s}_0 = \mathbf{0}$, $t_0 = 0$, and $q \geq 1$ is defined through $\frac{1}{p} + \frac{1}{q} = 1$. Problem (2.19) is a linear program if we choose $p \in \{1, \infty\}$.

Next, we consider instances of the uncertainty quantification problem (2.1) where $\mathcal{P}$ emerges from an intersection of the nested moment ambiguity set (2.4) with the set of all α -unimodal distributions $\mathcal{P}^\alpha$. In order to facilitate tractable reformulations, we restrict our attention to the subclass of *epigraphic moment ambiguity sets*. An epigraphic moment ambiguity set is an instance of (2.4) with $I = 1$ (i.e., it contains no confidence set other than the support), where the interaction between $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{u}}$ is captured through the epigraph constraint in

$$\mathcal{C}_1 = \{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_1 \mathbf{z} \preccurlyeq_{\mathcal{K}_1} \mathbf{d}_1, \mathbf{g}(\mathbf{z}) \preccurlyeq_{\mathcal{K}_2} \mathbf{u}\}, \quad (2.22)$$

where $\mathbf{C}_1 \in \mathbb{R}^{L_1 \times P}$, $\mathbf{d}_1 \in \mathcal{K}_1$ and $\mathcal{K}_1 \subseteq \mathbb{R}^{L_1}$, $\mathcal{K}_2 \subseteq \mathbb{R}^Q$ are proper cones. We require that the function $\mathbf{g} : \mathbb{R}^P \rightarrow \mathbb{R}^Q$ is $\mathcal{K}_2$ -convex [15, § 3.6.2] and that the set $\{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{g}(\mathbf{z}) \preccurlyeq_{\mathcal{K}_2} \mathbf{u}\}$ is representable through conic inequalities and additional auxiliary variables [121]. We remark that any confidence set of the form (2.5) can be expressed as an instance of (2.22) and vice versa. Epigraphic ambiguity sets become a special case of (2.4) with $I = 1$, however, if structural properties are imposed on the marginal

projection $\Pi_{\tilde{\mathbf{z}}}\mathbb{P}$.

Theorem 2.5 (Unimodality). *Assume that $\mathcal{P} = \mathcal{P}^e \cap \mathcal{P}^\alpha$ where $\mathcal{P}^e$ is an epigraphic moment ambiguity set and $\mathcal{P}^\alpha$ is the structural ambiguity set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{\mathbf{z}}}\mathbb{P}$ is α -unimodal around $\mathbf{z} = \mathbf{0}$. If the Slater condition $(\mathbf{S})$ holds and $\mathbf{t} \geq \mathbf{0}$, then the uncertainty quantification problem (2.1) is equivalent to the semi-infinite program*

$$\begin{aligned} & \sup \quad \beta + \mathbf{b}^\top \boldsymbol{\gamma} \\ & \text{s.t.} \quad \beta \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \tau_j \in \mathbb{R}_+, \quad \boldsymbol{\psi}_0, \boldsymbol{\psi}_j \in \mathcal{K}_1^*, \quad j \in \mathcal{L} \\ & \quad \left. \begin{aligned} & -\mathbf{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^* \\ & \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi}_0 \leq 1 \\ & \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi}_j \\ & \quad + \tau_j \mathbf{s}_j^\top \mathbf{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j t_j)^{\frac{\alpha}{\alpha+1}} \leq 0 \quad \forall j \in \mathcal{L} \end{aligned} \right\} \quad \forall \mathbf{z} \in \mathbb{R}^P, \end{aligned} \quad (2.23)$$

where $\mathcal{L} = \{j \in \mathcal{J} : \exists (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha \cdot \mathbf{s}_j^\top \mathbf{z} > 0\}$ and $\mathbf{g}^\alpha(\mathbf{z}) = \int_0^1 \mathbf{g}(t\mathbf{z}) \alpha t^{\alpha-1} dt$.

Proof. See Appendix. □

Unlike the previous results in this section, Theorem 2.5 does not provide a tractable reformulation *per se*. Instead, the tractability of problem (2.23) is determined by the properties of the function $\mathbf{g}^\alpha(\cdot)$ that appears in the semi-infinite constraints of (2.23). In the following, we present two examples for which problem (2.23) has a tractable reformulation.

Example 2.10 (Generalised Gauss Bounds). *Theorem 2.5 allows us to compute the worst-case probability of the event $\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}$ if the distribution of $\tilde{\mathbf{z}}$ belongs to the Gauss ambiguity set $\mathcal{P}$ defined in equation (2.9) of Example 2.2. Since $\mathcal{P} = \Pi_{\tilde{\mathbf{z}}}\mathcal{P}'$ for the ambiguity set $\mathcal{P}'$ defined in (2.10), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$. Moreover, the lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 2.5 if $\boldsymbol{\Sigma} \succ \mathbf{0}$, $\mathbf{S}\mathbf{m} \leq \mathbf{t}$ and $\mathcal{C}_1$ is defined as*

$$\mathcal{C}_1 = \{(\mathbf{z}, \mathbf{U}) \in \mathbb{R}^P \times \mathbb{R}^{P \times P} : \mathbf{g}(\mathbf{z}) \preceq \mathbf{U}\},$$

where $\mathbf{g}(\mathbf{z}) = \mathbf{z}\mathbf{z}^\top$ is $\mathbb{S}_+^P$ -convex. Applying a coordinate shift and employing Theorem 2.5 allows us to

reformulate the uncertainty quantification problem (2.1) over the Gauss ambiguity set as a semi-infinite program. Thereby, we recover a generalised multivariate Gauss inequality that was discovered in [90].

Since $\mathbf{g}^\alpha(\mathbf{z}) = \frac{\alpha}{\alpha+2} \mathbf{z} \mathbf{z}^\top$ can be computed explicitly, we may use standard robust optimisation techniques to further simplify this semi-infinite program to the finite convex program

$$\begin{aligned}
& \sup \quad \beta - (\boldsymbol{\mu} - \mathbf{m})^\top \boldsymbol{\gamma} - \langle \boldsymbol{\Sigma} + (\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^\top, \boldsymbol{\Gamma} \rangle \\
& \text{s.t.} \quad \beta \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}^P, \quad \boldsymbol{\Gamma} \in \mathbb{S}_+^P, \quad \tau_j \in \mathbb{R}_+, \quad j \in \mathcal{L} \\
& \quad \begin{pmatrix} 1 - \beta & \frac{1}{2} \frac{\alpha}{\alpha+1} \boldsymbol{\gamma}^\top \\ \frac{1}{2} \frac{\alpha}{\alpha+1} \boldsymbol{\gamma} & \frac{\alpha}{\alpha+2} \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0} \\
& \quad \begin{pmatrix} \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j (t_j - \mathbf{s}_j^\top \mathbf{m}))^{\frac{\alpha}{\alpha+1}} - \beta & \frac{1}{2} \left(\frac{\alpha}{\alpha+1} \boldsymbol{\gamma} - \tau_j \mathbf{s}_j \right)^\top \\ \frac{1}{2} \left(\frac{\alpha}{\alpha+1} \boldsymbol{\gamma} - \tau_j \mathbf{s}_j \right) & \frac{\alpha}{\alpha+2} \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0} \quad \forall j \in \mathcal{L},
\end{aligned} \tag{2.24}$$

where $\mathcal{L} = \{j \in \mathcal{J} : \mathbf{s}_j \neq \mathbf{0}\}$. Note that (2.24) is equivalent to a tractable semidefinite program if α is rational, in which case the non-linear term in the second matrix inequality can be linearised by using a well-known conic expansion of power functions, see e.g. [8, § 2.3.1].

Example 2.11 (Mean-Absolute Deviation & Unimodality). Let $\mathcal{P}$ be the ambiguity set of all distributions on $\mathbb{R}^P$ that are α -unimodal with mean and centre $\mathbf{m}$ and whose mean-absolute deviation is bounded above by $\mathbf{f}$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{P} \text{ is } \alpha\text{-unimodal with centre } \mathbf{m} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \mathbf{m}, \quad \mathbb{E}_{\mathbb{P}}[|\tilde{\mathbf{z}} - \mathbf{m}|] \leq \mathbf{f} \end{array} \right\}.$$

Since $\Pi_{\tilde{\mathbf{z}}} \mathcal{P}' = \mathcal{P}$ for the ambiguity set $\mathcal{P}'$ defined as

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \begin{array}{l} \Pi_{\tilde{\mathbf{z}}} \mathbb{P} \text{ is } \alpha\text{-unimodal with centre } \mathbf{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \mathbf{m}, \quad \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}, \quad \mathbb{P}[|\tilde{\mathbf{z}} - \mathbf{m}| \leq \tilde{\mathbf{u}}] = 1 \end{array} \right\},$$

we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]$. Moreover, $\mathcal{P}'$ satisfies the conditions of Theorem 2.5 if we define $\mathbf{g}(\mathbf{z}) = |\mathbf{z} - \mathbf{m}|$ and require that $\mathbf{S}\mathbf{m} \leq \mathbf{t}$ and $\mathbf{f} > \mathbf{0}$. Applying a coordinate shift and using Theorem 2.5 in conjunction with standard robust optimisation techniques, we can thus reformulate the

uncertainty quantification problem (2.1) as the finite convex program

$$\begin{aligned}
& \sup \quad \beta - \mathbf{f}^\top \boldsymbol{\eta} \\
& \text{s.t.} \quad \beta \in \mathbb{R}, \quad \boldsymbol{\theta} \in \mathbb{R}^P, \quad \boldsymbol{\eta} \in \mathbb{R}_+^P, \quad \tau_j \in \mathbb{R}_+, \quad j \in \mathcal{L} \\
& \quad \boldsymbol{\eta} \geq \boldsymbol{\theta} \geq -\boldsymbol{\eta}, \quad \beta \leq 1 \\
& \quad \left. \begin{aligned} \frac{\alpha}{\alpha+1} \boldsymbol{\eta} &\geq \frac{\alpha}{\alpha+1} \boldsymbol{\theta} + \tau_j \mathbf{s}_j \geq -\frac{\alpha}{\alpha+1} \boldsymbol{\eta} \\ \beta &\leq \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j (t_j - \mathbf{s}_j^\top \mathbf{m}))^{\frac{\alpha}{\alpha+1}} \end{aligned} \right\} \quad \forall j \in \mathcal{L},
\end{aligned} \tag{2.25}$$

where $\mathcal{L} = \{j \in \mathcal{J} : \mathbf{s}_j \neq \mathbf{0}\}$. Note that (2.25) can be reformulated as a tractable second-order cone program if α is rational, see [8, § 2.3.1].

In addition to evaluating the worst-case probability (2.1), practical applications often require simulation runs under the distribution that attains the worst-case probability. We can construct such worst-case distributions from the dual problems of (2.17), (2.21) and (2.23). In the interest of space, we do not embark on this pathway and instead refer the interested reader to [60, 115].

To conclude this section, we investigate the tractability of the uncertainty quantification problem (2.1) when the components of $\tilde{\mathbf{z}}$ are mutually independent under every distribution within the ambiguity set $\mathcal{P}$.

Theorem 2.6 (Independence). *The uncertainty quantification problem (2.1) over the ambiguity set $\mathcal{P}$ defined as*

$$\mathcal{P} = \mathcal{P}^n \cap \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \text{the components of } \tilde{\mathbf{z}} \text{ are independent under } \mathbb{P}\},$$

where $\mathcal{P}^n$ is the moment ambiguity set (2.4), is strongly NP-hard even if $(\mathbf{B})$ and $(\mathbf{N})$ are satisfied and $K = 0$ (no expectation constraints).

Proof. See Appendix. □

One can show that if in addition to the assumptions of Theorem 2.6, the ambiguity set $\mathcal{P}^n$ satisfies $I = 1$ (that is, there are no confidence sets other than the support), then the uncertainty quantification problem (2.1) reduces to a robust feasibility problem that can be solved efficiently. Also, if the moment

ambiguity set in Theorem 2.6 satisfies $K > 0$ but $I = 1$ and $\mathcal{C}_1 = \mathbb{R}^P \times \mathbb{R}^Q$, then problem (2.1) evaluates to 1 if both $\mathbf{S} = \mathbf{0}$ and $\mathbf{t} \geq \mathbf{0}$ and to 0 otherwise.

2.4 Chance Constrained Programming

We now study chance constrained programs of the form (2.2), where the safety of the underlying system can be actively enhanced by adjusting the design decisions $\mathbf{x} \in \mathbb{R}^N$. It turns out that the tractability of (2.2) is intimately related to the number of rows J of the technology matrix $\mathbf{S}(\mathbf{x})$ and the right-hand side vector $\mathbf{t}(\mathbf{x})$. Hence, Section 2.4.1 is devoted to *individual chance constraints* where $J = 1$, and Chapter 3 studies *joint chance constraints* where $J > 1$.

2.4.1 Individual Chance Constraints

As $J = 1$ throughout this section, we can simplify the notation if we denote the technology matrix by $\mathbf{s}(\mathbf{x})^\top$ and the right-hand side vector by $t(\mathbf{x})$.

We first study instances of the chance constrained program (2.2) where the ambiguity set is of the form (2.4). To derive a tractable reformulation for such problems, we restrict our attention to the subclass of *Markov ambiguity sets*, which are defined as instances of (2.4) with $I = 1$ and $\underline{p}_1 = \bar{p}_1 = 1$. Note that these ambiguity sets involve no confidence sets other than the support. We emphasise that apart from the Wasserstein ambiguity set all other examples of Section 2.2.3 constitute either Markov ambiguity sets or result from intersections of Markov ambiguity sets with structural ambiguity sets. Note that Markov ambiguity sets satisfy the conditions (B) and (N).

Theorem 2.7 (Moment Ambiguity Sets). *If $\mathcal{P}$ is a Markov ambiguity set satisfying (S) and $J = 1$, then the chance constraint in (2.2) is satisfied if and only if there are $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^*$, $\tau \in \mathbb{R}_+$ such that*

$$\begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq (1 - \varepsilon)\tau, & \beta + \mathbf{d}_1^\top \boldsymbol{\phi} &\leq \tau, & \beta + \mathbf{d}_1^\top \boldsymbol{\psi} &\leq t(\mathbf{x}) \\ \mathbf{A}^\top \boldsymbol{\gamma} &= \mathbf{C}_1^\top \boldsymbol{\phi}, & \mathbf{B}^\top \boldsymbol{\gamma} &= \mathbf{D}_1^\top \boldsymbol{\phi} \\ \mathbf{A}^\top \boldsymbol{\gamma} + \mathbf{s}(\mathbf{x}) &= \mathbf{C}_1^\top \boldsymbol{\psi}, & \mathbf{B}^\top \boldsymbol{\gamma} &= \mathbf{D}_1^\top \boldsymbol{\psi}. \end{aligned} \tag{2.26}$$

Proof. See Appendix. □

Note that (2.26) is a system of linear constraints that scales polynomially in the description of problem (2.2). The next result shows that the restriction to Markov ambiguity sets in Theorem 2.7 is necessary.

Theorem 2.8. *If $\mathcal{P}$ is an instance of (2.4) with $I > 1$, then the chance constrained program (2.2) is strongly NP-hard even if $J = 1$, $K = 0$ (no expectation constraints) and the boundedness condition (B) as well as the nesting condition (N) are satisfied, while $\mathcal{X}$ is a polyhedron.*

Proof. See Appendix. □

We illustrate the tractable reformulation of Theorem 2.7 with two examples.

Example 2.12 (Chebyshev Ambiguity Set). *We can use Theorem 2.7 to derive a tractable reformulation for individual chance constraints over the Chebyshev ambiguity set $\mathcal{P}$ defined in Example 2.1 with $\Sigma \succ \mathbf{0}$. We have $\mathcal{P} = \Pi_{\tilde{\mathbf{z}}} \mathcal{P}'$ for the Markov ambiguity set $\mathcal{P}'$ defined in (2.8). Theorem 2.7 thus implies that the individual chance constraint $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{s}(\mathbf{x})^\top \tilde{\mathbf{z}} \leq t(\mathbf{x})]$ is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^P$, $\Gamma \in \mathbb{S}_+^P$ and $\tau \in \mathbb{R}_+$ with*

$$\begin{aligned} & \beta - \boldsymbol{\mu}^\top \gamma - \langle \Sigma + \boldsymbol{\mu} \boldsymbol{\mu}^\top, \Gamma \rangle \geq (1 - \varepsilon) \tau \\ & \begin{pmatrix} \tau - \beta & \frac{1}{2} \gamma^\top \\ \frac{1}{2} \gamma & \Gamma \end{pmatrix} \succcurlyeq \mathbf{0} \\ & \begin{pmatrix} t_1(\mathbf{x}) - \beta & \frac{1}{2} (\gamma - \mathbf{s}(\mathbf{x}))^\top \\ \frac{1}{2} (\gamma - \mathbf{s}(\mathbf{x})) & \Gamma \end{pmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

Moreover, one can show that this constraint system is satisfied if and only if

$$\sqrt{\frac{1 - \varepsilon}{\varepsilon}} \|\Sigma^{\frac{1}{2}} \mathbf{s}(\mathbf{x})\|_2 + \boldsymbol{\mu}^\top \mathbf{s}(\mathbf{x}) \leq t(\mathbf{x}),$$

which is a second-order cone constraint, see [50].

Example 2.13 (Huber Ambiguity Set). *Theorem 2.7 allows us to derive a tractable reformulation for individual chance constraints over the Huber ambiguity set $\mathcal{P}$ defined in Example 2.4, assuming that $\mathbf{f}^\top \boldsymbol{\mu} < g$. We have $\mathcal{P} = \Pi_{\tilde{\mathbf{z}}} \mathcal{P}'$ for the Markov ambiguity set $\mathcal{P}'$ defined in (2.13). Theorem 2.7 thus implies that the individual chance constraint $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{s}(\mathbf{x})^\top \tilde{\mathbf{z}} \leq t(\mathbf{x})]$ is satisfied if and only if there are*

$\alpha, \nu_0, \nu_1 \in \mathbb{R}$, $\boldsymbol{\psi} \in \mathbb{R}^P$ and $\tau, \phi, \lambda_0, \lambda_1 \in \mathbb{R}_+$ that satisfy the following semi-definite constraints.

$$\begin{aligned} \alpha - \phi g &\geq (1 - \varepsilon)\tau \\ \begin{pmatrix} \lambda_0 & \nu_0 \\ \nu_0 & \frac{1}{2}\phi \end{pmatrix} \succcurlyeq \mathbf{0}, \quad \begin{pmatrix} \lambda_1 & \nu_1 \\ \nu_1 & \frac{1}{2}\phi \end{pmatrix} \succcurlyeq \mathbf{0} \\ \alpha + \lambda_0 &\leq \tau, \quad \boldsymbol{\psi} = 2\nu_0 \mathbf{f}, \quad -\beta\phi \leq 2\nu_0 \leq \beta\phi \\ \alpha + \lambda_1 - t(\mathbf{x}) + \mathbf{s}(\mathbf{x})^\top \boldsymbol{\mu} &\leq 0, \quad \boldsymbol{\psi} + \mathbf{s}(\mathbf{x}) = 2\nu_1 \mathbf{f}, \quad -\beta\phi \leq 2\nu_1 \leq \beta\phi \end{aligned}$$

We now consider instances of the chance constrained program (2.2) where $\mathcal{P}$ is generated by intersecting a Markov ambiguity set with the set of all symmetric distributions $\mathcal{P}^s$.

Theorem 2.9 (Symmetry). *Assume that $\mathcal{P} = \mathcal{P}^m \cap \mathcal{P}^s$, where $\mathcal{P}^m$ is a Markov ambiguity set that satisfies the dimensionality condition (D) as well as the feasibility condition (F), and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\mathbf{z}}\mathbb{P}$ is point symmetric around $\mathbf{m} = \mathbf{0}$. If the Slater condition (S) holds, then the chance constraint in (2.2) with $J = 1$ and $\varepsilon \in (0, 1/2)$ is satisfied if and only if there are $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star$ and $\tau \in \mathbb{R}_+$ such that*

$$\begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq (1 - \varepsilon)\tau, & \beta + \mathbf{d}_1^\top \boldsymbol{\phi} &\leq \tau \\ 2\beta + \mathbf{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - t(\mathbf{x}) &\leq \tau, & \mathbf{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) &= \mathbf{s}(\mathbf{x}) \\ \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}^+ = \mathbf{D}_1^\top \boldsymbol{\psi}^- = \mathbf{D}_1^\top \boldsymbol{\phi}. \end{aligned} \tag{2.27}$$

Proof. See Appendix. □

In analogy to Theorem 2.7, Theorem 2.9 reexpresses the individual chance constraint as a finite set of tractable constraints. One can show that a Markov ambiguity set $\mathcal{P}^m$ satisfies the feasibility condition (F) whenever the projected support $\{\mathbf{z} \in \mathbb{R}^P : \exists \mathbf{u} \in \mathbb{R}^Q. (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1\}$ is point symmetric around $\mathbf{m} = \mathbf{0}$. The restriction to Markov ambiguity sets in Theorem 2.9 is again necessary.

Theorem 2.10. *If $\mathcal{P} = \mathcal{P}^n \cap \mathcal{P}^s$ where $\mathcal{P}^n$ is an instance of (2.4) with $I > 1$ and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\mathbf{z}}\mathbb{P}$ is point symmetric around $\mathbf{m} = \mathbf{0}$, then the chance constrained program (2.2) is strongly NP-hard even if $J = 1$, $K = 0$ (no expectation constraints) and the conditions (B), (N), (D) and (F) are satisfied, while $\mathcal{X}$ is a polyhedron.*

Proof. See Appendix. □

We illustrate the tractable reformulation of Theorem 2.9 with an example.

Example 2.14 (Median-Absolute Deviation & Symmetry). *We can use Theorem 2.9 to derive a tractable reformulation for individual chance constraints over the median-absolute deviation ambiguity set $\mathcal{P}$ defined in Example 2.3 with $\mathbf{f} > \mathbf{0}$. Note that we have $\mathcal{P} = \Pi_{\mathbf{z}}\mathcal{P}'$ for $\mathcal{P}'$ defined in (2.12). If we assume that $\varepsilon \in (0, 1/2)$ and apply a coordinate shift, then all conditions of Theorem 2.9 are met and the chance constraint in (2.2) is satisfied if and only if*

$$\frac{1}{2\varepsilon} \mathbf{f}^\top |\mathbf{s}(\mathbf{x})| + \mathbf{m}^\top \mathbf{s}(\mathbf{x}) \leq t(\mathbf{x}),$$

which can be expressed by a system of $2P + 1$ linear inequalities.

Next, we consider instances of the chance constrained program (2.2) where the ambiguity set $\mathcal{P}$ emerges from the intersection of an epigraphic moment ambiguity set with the set of all unimodal distributions $\mathcal{P}^\alpha$.

Theorem 2.11 (Unimodality). *Assume that $\mathcal{P} = \mathcal{P}^e \cap \mathcal{P}^\alpha$ where $\mathcal{P}^e$ is an epigraphic moment ambiguity set, whereas $\mathcal{P}^\alpha$, $\alpha > 1$, is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\mathbf{z}}\mathbb{P}$ is α -unimodal around $\mathbf{z} = \mathbf{0}$. If the Slater condition (S) holds, $J = 1$ and $t(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$, then the chance constraint in (2.2) is satisfied if and only if there are $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^*$ such that*

$$\left. \begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq (1 - \varepsilon)\tau, \quad -\mathbf{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^* \\ \beta + \left(\frac{\alpha}{\alpha + 1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\phi} &\leq \tau \\ \beta + \left(\frac{\alpha}{\alpha + 1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi} &\leq \tau \\ + \mathbf{s}(\mathbf{x})^\top \mathbf{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} \tau^{\frac{1}{\alpha+1}} t(\mathbf{x})^{\frac{\alpha}{\alpha+1}} &\leq 0 \end{aligned} \right\} \quad \forall \mathbf{z} \in \mathbb{R}^P. \quad (2.28)$$

We remark that the term $-\tau^{\frac{1}{\alpha+1}} t(\mathbf{x})^{\frac{\alpha}{\alpha+1}}$ is convex on $\{(\tau, \mathbf{x}) \in \mathbb{R}_+ \times \mathcal{X}\}$ since $\alpha > 0$ and $t(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$. If α is a rational number exceeding 1, that is, if $\alpha = p/q$ for some $p, q \in \mathbb{N}$ with $p \geq q$, then the epigraph of this term can be expressed through $\mathcal{O}(p)$ second-order conic constraints, see e.g. [90, Lemma 4.2]. In analogy to the formulation (2.23) of Theorem 2.5, the tractability of (2.28) depends on

the functional form of $\mathbf{g}^\alpha(\cdot)$. We close this section with an example for which the constraint system (2.28) has a tractable reformulation.

Example 2.15 (Gauss Ambiguity Set). *Theorem 2.11 allows us to derive a tractable reformulation for individual chance constraints over the Gauss ambiguity set $\mathcal{P}$ defined in Example 2.2 whenever $\Sigma \succ \mathbf{0}$. Since $\mathcal{P} = \Pi_{\mathbf{z}}\mathcal{P}'$ for $\mathcal{P}'$ defined in (2.10), we can replace $\mathcal{P}$ in the chance constrained program (2.2) with $\mathcal{P}'$. The lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 2.11 if we set*

$$\mathcal{C}_1 = \{(\mathbf{z}, \mathbf{U}) \in \mathbb{R}^P \times \mathbb{R}^{P \times P} : \mathbf{g}(\mathbf{z}) \preceq \mathbf{U}\},$$

where $\mathbf{g}(\mathbf{z}) = \mathbf{z}\mathbf{z}^\top$ is $\mathbb{S}_+^P$ -convex, and require that the mode $\mathbf{m}$ satisfies $\mathbf{s}(\mathbf{x})^\top \mathbf{m} \leq t(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. Applying a coordinate shift, using Theorem 2.11 and employing standard robust optimisation techniques allows us to conclude that the chance constraint in (2.2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^P$, $\boldsymbol{\Gamma} \in \mathbb{S}_+^P$ and $\tau \in \mathbb{R}_+$ such that

$$\begin{aligned} & \beta - (\boldsymbol{\mu} - \mathbf{m})^\top \boldsymbol{\gamma} - \langle \Sigma + (\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^\top, \boldsymbol{\Gamma} \rangle \geq 1 - \varepsilon \\ & \begin{pmatrix} \tau - \beta & \frac{1}{2} \frac{\alpha}{\alpha + 1} \boldsymbol{\gamma}^\top \\ \frac{1}{2} \frac{\alpha}{\alpha + 1} \boldsymbol{\gamma} & \frac{\alpha}{\alpha + 2} \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0} \\ & \begin{pmatrix} \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} \tau^{\frac{1}{\alpha+1}} (t(\mathbf{x}) - \mathbf{s}(\mathbf{x})^\top \mathbf{m})^{\frac{\alpha}{\alpha+1}} - \beta & \frac{1}{2} \left(\frac{\alpha}{\alpha+1} \boldsymbol{\gamma} - \mathbf{s}(\mathbf{x}) \right)^\top \\ \frac{1}{2} \left(\frac{\alpha}{\alpha+1} \boldsymbol{\gamma} - \mathbf{s}(\mathbf{x}) \right) & \frac{\alpha}{\alpha+2} \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

2.5 Conclusion

Table 2.1 summarises the results of this chapter. We reiterate that conic moment ambiguity sets form a strict subclass of epigraphic moment ambiguity sets and the class of epigraphic moment ambiguity sets coincides with the class of Markov ambiguity sets. Neither statement is true, however, if the corresponding moment ambiguity set is intersected with a structural ambiguity set.

Moment information		Moments + symmetry		Moments + unimodality
Conic ambiguity set $\mathcal{P}^c$ $\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \right.$ with $\mathcal{C}_1 = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_1\mathbf{z} + \mathbf{D}_1\mathbf{u} \preceq_{\kappa_1} \mathbf{0} \right\}$	✓ UQ:	Theorem 2.3	✓ UQ:	Theorem 2.4
	✓ ICC:	Theorem 2.7	✓ ICC:	Theorem 2.9
	✓ JCC/FT:	Theorem 3.5	✗ JCC/FT:	Theorem 2.12
	✗ JCC/RT:	Theorem 3.3	✗ JCC/RT:	Theorem 2.12
Epigraphic ambiguity set $\mathcal{P}^e$ $\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \right.$ with $\mathcal{C}_1 = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_1\mathbf{z} \preceq_{\kappa_1} \mathbf{d}_1, \mathbf{g}(\mathbf{z}) \preceq_{\kappa_2} \mathbf{u} \right\}$	✓ UQ:	Theorem 2.3	✓ UQ:	Theorem 2.4
	✓ ICC:	Theorem 2.7	✓ ICC:	Theorem 2.9
	✗ JCC/FT:	Theorem 3.6	✗ JCC/FT:	Theorem 2.12
	✗ JCC/RT:	Theorem 3.3	✗ JCC/RT:	Theorem 2.12
Markov ambiguity set $\mathcal{P}^m$ $\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \right.$ with $\mathcal{C}_1 = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_1\mathbf{z} + \mathbf{D}_1\mathbf{u} \preceq_{\kappa_1} \mathbf{d}_1 \right\}$	✓ UQ:	Theorem 2.3	✓ UQ:	Theorem 2.4
	✓ ICC:	Theorem 2.7	✓ ICC:	Theorem 2.9
	✗ JCC/FT:	Theorem 3.6	✗ JCC/FT:	Theorem 2.12
	✗ JCC/RT:	Theorem 3.3	✗ JCC/RLHS:	Theorem 2.12
Nested ambiguity set $\mathcal{P}^n$ $\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \right.$ with $\mathcal{C}_i = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_i\mathbf{z} + \mathbf{D}_i\mathbf{u} \preceq_{\kappa_i} \mathbf{d}_i \right\}$	✓ UQ:	Theorem 2.3	✓ UQ:	Theorem 2.4
	✗ ICC:	Theorem 2.8	✗ ICC:	Theorem 2.10
	✗ JCC/FT:	Theorem 3.6	✗ JCC/FT:	Theorem 2.12
	✗ JCC/RT:	Theorem 3.3	✗ JCC/RT:	Theorem 2.12

Abbreviations: UQ = Uncertainty quantification, ICC = Individual chance constraint,
JCC/FT = Joint chance constraint with fixed technology matrix,
JCC/RT = Joint chance constraint with random technology matrix.

Table 2.1: Summary of the results in this chapter. Checks (✓) and crosses (✗) highlight tractable reformulations and intractability results, respectively.

2.6 Appendix

For the proof of Theorem 2.1, we recall that the strongly NP-hard 0/1 Integer Programming (IP) problem [46] is defined as follows.

0/1 INTEGER PROGRAMMING. Instance. Given are $\mathbf{E} \in \mathbb{Z}^{M \times P}$ and $\mathbf{f} \in \mathbb{Z}^M$. Question. Is there a vector $\mathbf{y} \in \{0, 1\}^P$ such that $\mathbf{E}\mathbf{y} \leq \mathbf{f}$?

Assume that $\mathbf{y}' \in [0, 1]^P$ constitutes a fractional vector that satisfies $\mathbf{E}\mathbf{y}' \leq \mathbf{f}$. We now show that we can obtain an integral vector $\mathbf{y} \in \{0, 1\}^P$ that satisfies $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ by rounding $\mathbf{y}'$ if its components are ‘close enough’ to 0 or 1.

Lemma 2.1. *Assume that $\mathbf{y}' \in ([0, \varepsilon) \cup (1 - \varepsilon, 1])^P$ satisfies $\mathbf{E}\mathbf{y}' \leq \mathbf{f}$ for $0 < \varepsilon < \min_m \left\{ \left(\sum_p |E_{mp}| \right)^{-1} \right\}$. Then $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ for $\mathbf{y} \in \{0, 1\}^P$, where $y_p = 1$ if $y'_p > 1 - \varepsilon$ and $y_p = 0$ otherwise.*

Remark 2.2. *A proof of Lemma 2.1 can be found in [120, 121]. To keep the chapter self-contained, we repeat the proof here.*

Proof of Lemma 2.1. We have that $\mathbf{e}_m^\top \mathbf{y} \leq \mathbf{e}_m^\top \mathbf{y}' + \sum_p |E_{mp}| \varepsilon < \mathbf{e}_m^\top \mathbf{y}' + 1 \leq f_m + 1$ for all $m \in \{1, \dots, M\}$, where $\mathbf{e}_m^\top$ denotes the m -th row of $\mathbf{E}$. Due to the integrality of $\mathbf{E}$, $\mathbf{f}$ and $\mathbf{y}$, we thus conclude that $\mathbf{E}\mathbf{y} \leq \mathbf{f}$. □

Proof of Theorem 2.1. Fix an instance $(\mathbf{E}, \mathbf{f})$ of the IP problem and assume without loss of generality that the LP relaxation of $(\mathbf{E}, \mathbf{f})$ has a solution, that is, there is $\mathbf{y} \in [0, 1]^P$ such that $\mathbf{E}\mathbf{y} \leq \mathbf{f}$. Consider the following instance of the quadratic uncertainty quantification problem (2.3), where $\varepsilon \in (0, 1/2)$ is chosen as prescribed in Lemma 2.1:

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left[\left\| \tilde{\mathbf{z}} - \frac{1}{2} \mathbf{e} \right\|_2 \leq \sqrt{\frac{P}{4} - \varepsilon(1 - \varepsilon)} \right], \quad (2.29)$$

where

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{P} \left[\tilde{\mathbf{z}} \in [0, 1]^P, \mathbf{E}\tilde{\mathbf{z}} \leq \mathbf{f} \right] = 1 \right\}.$$

We want to show that (2.29) evaluates to 0 if and only if the IP feasibility problem has an affirmative answer. To this end, assume first that (2.29) evaluates to 0. One readily verifies that the infimum in (2.29) must be attained by a probability distribution $\mathbb{P} \in \mathcal{P}$. By construction, $\mathbb{P}$ places all probability mass on realisations $\mathbf{z} \in \mathbb{R}^P$ that satisfy $\mathbf{z} \in [0, 1]^P$, $\mathbf{E}\mathbf{z} \leq \mathbf{f}$ and $\left\| \mathbf{z} - \frac{1}{2}\mathbf{e} \right\|_2 > \sqrt{P/4 - \varepsilon(1 - \varepsilon)}$, that is, $z_p \in [0, \varepsilon) \cup (1 - \varepsilon, 1]$ for all $p = 1, \dots, P$. We can then use Lemma 2.1 to round any such $\mathbf{z}$ to a solution of the IP problem.

Assume now that the instance of the IP problem is feasible, that is, there is $\mathbf{z} \in \{0, 1\}^P$ such that $\mathbf{E}\mathbf{z} \leq \mathbf{f}$. By construction, $\mathbf{z}$ satisfies $\left\| \mathbf{z} - \frac{1}{2}\mathbf{e} \right\|_2 = \sqrt{P/4} > \sqrt{P/4 - \varepsilon(1 - \varepsilon)}$. Thus, we have $\delta_{\mathbf{z}} \in \mathcal{P}$, where $\delta_{\mathbf{z}}$ represents the Dirac distribution that concentrates unit mass at $\mathbf{z}$, and we observe that

$$\begin{aligned} 0 &\leq \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left[\left\| \tilde{\mathbf{z}} - \frac{1}{2}\mathbf{e} \right\|_2 \leq \sqrt{\frac{P}{4} - \varepsilon(1 - \varepsilon)} \right] \\ &\leq \mathbb{P}_{\delta_{\mathbf{z}}} \left[\left\| \tilde{\mathbf{z}} - \frac{1}{2}\mathbf{e} \right\|_2 \leq \sqrt{\frac{P}{4} - \varepsilon(1 - \varepsilon)} \right] = 0, \end{aligned}$$

that is, problem (2.29) indeed evaluates to 0. □

In the remainder, we denote by $\mathcal{D}(i) = \{i' \in \mathcal{I} : \mathcal{C}_{i'} \Subset \mathcal{C}_i\}$ the set of all strict subsets (descendants) of the confidence set $\mathcal{C}_i$, and we define $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'}$, $i \in \mathcal{I}$ as the confidence set $\mathcal{C}_i$ stripped of all its descendants.

Proof of Theorem 2.3. The uncertainty quantification problem (2.1) over the nested moment ambiguity set (2.4) coincides with the optimal value of the following problem, where $\mathcal{M}_+(\mathbb{R}^P \times \mathbb{R}^Q)$

denotes the cone of all non-negative measures on $\mathbb{R}^P \times \mathbb{R}^Q$.

$$\begin{aligned}
& \inf \quad \int_{\mathcal{C}_I} \mathbb{I}_{[\mathbf{S}\mathbf{z} \leq \mathbf{t}]} d\mu(\mathbf{z}, \mathbf{u}) \\
& \text{s.t.} \quad \mu \in \mathcal{M}_+(\mathbb{R}^P \times \mathbb{R}^Q) \\
& \quad \int_{\mathcal{C}_I} [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}] d\mu(\mathbf{z}, \mathbf{u}) = \mathbf{b} \\
& \quad \left. \begin{aligned} \int_{\mathcal{C}_I} \mathbb{I}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} d\mu(\mathbf{z}, \mathbf{u}) &\geq \underline{p}_i \\ \int_{\mathcal{C}_I} \mathbb{I}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} d\mu(\mathbf{z}, \mathbf{u}) &\leq \bar{p}_i \end{aligned} \right\} \forall i \in \mathcal{I}
\end{aligned} \tag{2.30}$$

Note that $\underline{p}_I = \bar{p}_I = 1$ due to condition **(B)**. Hence, every measure μ feasible in (2.30) constitutes a probability measure $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ supported on $\mathcal{C}_I$.

The moment problem (2.30) admits the following semi-infinite dual problem.

$$\begin{aligned}
& \sup \quad \mathbf{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \bar{p}_i \kappa_i \\
& \text{s.t.} \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I \\
& \quad (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) \mathbb{I}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} \leq \mathbb{I}_{[\mathbf{S}\mathbf{z} \leq \mathbf{t}]} \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_I
\end{aligned} \tag{2.31}$$

Strong duality holds due to the Slater condition **(S)**, see [66, 103]. By expanding the right-hand side indicator function as the pointwise minimum of indicator functions of halfspaces, the semi-infinite constraint in (2.31) can be split up into the following system of semi-infinite constraints.

$$\left. \begin{aligned} (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) \mathbb{I}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} &\leq 1 \\ (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) \mathbb{I}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} &\leq 0 \quad \forall j \in \mathcal{J} : \mathbf{s}_j^\top \mathbf{z} > t_j \end{aligned} \right\} \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_I.$$

The nesting condition **(N)** implies that we can partition the support set $\mathcal{C}_I$ into I non-empty disjoint sets, $i \in \mathcal{I}$. We may thus reformulate the above semi-infinite constraints as

$$\left. \begin{aligned} (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) &\leq 1 \\ (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) &\leq 0 \quad \forall j \in \mathcal{J} : \mathbf{s}_j^\top \mathbf{z} > t_j \end{aligned} \right\} \begin{aligned} &\forall i \in \mathcal{I}, \\ &\forall (\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i. \end{aligned} \tag{2.32}$$

As the constraint left-hand sides are linear in $(\mathbf{z}, \mathbf{u})$, we may replace the sets $\bar{\mathcal{C}}_i$ and $\{(\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i : \mathbf{s}_j^\top \mathbf{z} > t_j\}$ in (2.32) with their convex hull $\text{co}(\bar{\mathcal{C}}_i)$ and $\text{co}(\{(\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i : \mathbf{s}_j^\top \mathbf{z} > t_j\})$. Due to the nesting condition (N), these convex hulls coincide with the sets $\mathcal{C}_i$ and $\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i : \mathbf{s}_j^\top \mathbf{z} > t_j\}$, respectively. By strong conic duality, which holds since $\mathcal{C}_i$ is essentially strictly feasible, the first semi-infinite constraint in (2.32) is then satisfied for $i \in \mathcal{I}$ if and only if there exists $\phi_i \in \mathcal{K}_i^*$ with

$$\sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \mathbf{d}_i^\top \phi_i \leq 1, \quad \mathbf{A}^\top \boldsymbol{\gamma} = \mathbf{C}_i^\top \phi_i, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_i^\top \phi_i. \quad (2.33)$$

Note that for $(i, j) \in \mathcal{I} \times \mathcal{J}$, the second semi-infinite constraint in (2.32) is void and can thus be omitted whenever $(i, j) \notin \mathcal{L}$. For $(i, j) \in \mathcal{L}$, however, we can replace the strict inequality in (2.32) with a weak one since the expression

$$(\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'})$$

is continuous in $(\mathbf{z}, \mathbf{u})$ and the set $\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i : \mathbf{s}_j^\top \mathbf{z} > t_j\}$ is non-empty. By strong conic duality, which holds since $\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i : \mathbf{s}_j^\top \mathbf{z} > t_j\}$ is essentially strictly feasible, we then conclude that the constraint is satisfied if and only if there exist $\psi_{ij} \in \mathcal{K}_i^*$ and $\tau_{ij} \in \mathbb{R}_+$ with

$$\sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \mathbf{d}_i^\top \psi_{ij} - \tau_{ij} t_j \leq 0, \quad \mathbf{A}^\top \boldsymbol{\gamma} + \tau_{ij} \mathbf{s}_j = \mathbf{C}_i^\top \psi_{ij}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_i^\top \psi_{ij}. \quad (2.34)$$

The claim now follows from substituting (2.33) and (2.34) into (2.31). $\square$

The proof of Theorem 2.4 relies on the following two auxiliary results.

Lemma 2.2. *If (B), (D) and (N) hold, then for any $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i$, $i \in \mathcal{I}$, there exist $(\mathbf{z}, \mathbf{u}_1), (\mathbf{z}, \mathbf{u}_2) \in \bar{\mathcal{C}}_i$ and $\lambda \in [0, 1]$ with $(\mathbf{z}, \mathbf{u}) = \lambda(\mathbf{z}, \mathbf{u}_1) + (1 - \lambda)(\mathbf{z}, \mathbf{u}_2)$.*

Proof. Under condition (D), the statement trivially holds if $I = 1$. Assume now that $Q > 0$ and $\mathcal{C}_I$ is bounded, and fix a confidence set $\mathcal{C}_i$, $i \in \mathcal{I}$, and parameter realisation $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i$. Under conditions (B)

and **(D)**, the auxiliary variable u_1 exists and is uniformly bounded across all $(\mathbf{z}', \mathbf{u}') \in \mathcal{C}_i \subseteq \mathcal{C}_I$. Thus,

$$\underline{u}_1 = \max_{\theta \in \mathbb{R}_+} \{\theta : (\mathbf{z}, \mathbf{u} - \theta \mathbf{e}_1) \in \mathcal{C}_i\} \quad \text{and} \quad \bar{u}_1 = \max_{\theta \in \mathbb{R}_+} \{\theta : (\mathbf{z}, \mathbf{u} + \theta \mathbf{e}_1) \in \mathcal{C}_i\}$$

both exist and are finite. The nesting condition **(N)** implies that $\mathbf{u}_1 = (\mathbf{z}, \mathbf{u} - \underline{u}_1 \mathbf{e}_1)$ and $\mathbf{u}_2 = (\mathbf{z}, \mathbf{u} + \bar{u}_1 \mathbf{e}_1)$ belong to the boundary of $\mathcal{C}_i$ and are thus elements of $\bar{\mathcal{C}}_i$. The claim now follows by setting $\lambda = \bar{u}_1 / (\underline{u}_1 + \bar{u}_1)$. $\square$

Lemma 2.3. *Let Z be a convex subset of $\mathbb{R}^P$ and define the sets*

$$\begin{aligned} \mathcal{D}_{\mathbf{i}} &= \{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_{i^-}\}, \\ \bar{\mathcal{D}}_{\mathbf{i}} &= \{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) \in \bar{\mathcal{C}}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \bar{\mathcal{C}}_{i^-}\} \end{aligned}$$

for $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$. If **(B)**, **(D)** and **(N)** hold, then $\text{co}(\bar{\mathcal{D}}_{\mathbf{i}}) = \mathcal{D}_{\mathbf{i}}$, $\mathbf{i} \in \mathcal{I}^2$.

Proof. We note that for a fixed $\mathbf{i} \in \mathcal{I}^2$, we have that

$$\begin{aligned} \text{co}(\bar{\mathcal{D}}_{\mathbf{i}}) &= \text{co}\left(\{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) \in \bar{\mathcal{C}}_{i^+}\} \cap \right. \\ &\quad \left. \{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (-\mathbf{z}, \mathbf{u}^-) \in \bar{\mathcal{C}}_{i^-}\}\right) \\ &\subseteq \text{co}\left(\{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) \in \bar{\mathcal{C}}_{i^+}\} \cap \right. \\ &\quad \left. \text{co}\left(\{(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (-\mathbf{z}, \mathbf{u}^-) \in \bar{\mathcal{C}}_{i^-}\}\right)\right) = \mathcal{D}_{\mathbf{i}}, \end{aligned}$$

where the last inclusion follows from the relation $\text{co}(\bar{\mathcal{C}}_i) = \mathcal{C}_i$ for all $i \in \mathcal{I}$, which is implied by the nesting condition **(N)**.

To prove that $\mathcal{D}_{\mathbf{i}} \subseteq \text{co}(\bar{\mathcal{D}}_{\mathbf{i}})$, fix some $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathcal{D}_{\mathbf{i}}$. Thus, $(\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_{i^+}$ and $(-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_{i^-}$. By Lemma 2.2 there exist $\mathbf{u}_1^+, \mathbf{u}_2^+, \mathbf{u}_1^-, \mathbf{u}_2^- \in \mathbb{R}^Q$ and $\lambda^+, \lambda^- \in [0, 1]$ with $(\mathbf{z}, \mathbf{u}_1^+), (\mathbf{z}, \mathbf{u}_2^+) \in \bar{\mathcal{C}}_{i^+}$ and $(-\mathbf{z}, \mathbf{u}_1^-), (-\mathbf{z}, \mathbf{u}_2^-) \in \bar{\mathcal{C}}_{i^-}$ with

$$\mathbf{u}^+ = \lambda^+ \mathbf{u}_1^+ + (1 - \lambda^+) \mathbf{u}_2^+ \quad \text{and} \quad \mathbf{u}^- = \lambda^- \mathbf{u}_1^- + (1 - \lambda^-) \mathbf{u}_2^-.$$

By construction, the vectors $(\mathbf{z}, \mathbf{u}_1^+, \mathbf{u}_1^-)$, $(\mathbf{z}, \mathbf{u}_1^+, \mathbf{u}_2^-)$, $(\mathbf{z}, \mathbf{u}_2^+, \mathbf{u}_1^-)$ and $(\mathbf{z}, \mathbf{u}_2^+, \mathbf{u}_2^-)$ belong to $\overline{\mathcal{D}}_{\mathbf{i}}$, and we have

$$\begin{aligned} (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) &= \lambda^+ \lambda^- (\mathbf{z}, \mathbf{u}_1^+, \mathbf{u}_1^-) + \lambda^+ (1 - \lambda^-) (\mathbf{z}, \mathbf{u}_1^+, \mathbf{u}_2^-) \\ &\quad + (1 - \lambda^+) \lambda^- (\mathbf{z}, \mathbf{u}_2^+, \mathbf{u}_1^-) + (1 - \lambda^+) (1 - \lambda^-) (\mathbf{z}, \mathbf{u}_2^+, \mathbf{u}_2^-). \end{aligned}$$

This implies that $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \text{co}(\overline{\mathcal{D}}_{\mathbf{i}})$. As $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathcal{D}_{\mathbf{i}}$ was chosen arbitrarily, we have $\mathcal{D}_{\mathbf{i}} \subseteq \text{co}(\overline{\mathcal{D}}_{\mathbf{i}})$.

We may thus conclude that $\text{co}(\overline{\mathcal{D}}_{\mathbf{i}}) = \mathcal{D}_{\mathbf{i}}$. $\square$

Proof of Theorem 2.4. The general duality theory for moment problems allows us to express the worst-case probability (2.1) as the optimal value of

$$\begin{aligned} \sup \quad & \mathbf{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} p_i \lambda_i - \bar{p}_i \kappa_i \\ \text{s.t.} \quad & \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I \\ & \mathbb{E}_{\mathbb{P}} \left[(\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) \mathbb{I}_{[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{C}_i]} \right] \leq \mathbb{E}_{\mathbb{P}} [\mathbb{I}_{[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]}] \quad \forall \mathbb{P} \in \mathcal{P}^s(\mathcal{C}_I), \end{aligned} \tag{2.35}$$

where $\mathcal{P}^s(\mathcal{C}_I)$ contains all distributions $\mathbb{P} \in \mathcal{P}_0(\mathcal{C}_I)$ for which $\Pi_{\tilde{\mathbf{z}}} \mathbb{P}$ is symmetric around $\mathbf{m} = \mathbf{0}$, see [103, § 3] or [93, Theorem 3.1]. Strong duality holds due to the Slater condition (S). The extreme points of the structural ambiguity set $\mathcal{P}^s(\mathcal{C}_I)$ are the discrete distributions $\frac{1}{2} \delta_{(\mathbf{z}, \mathbf{u}^+)} + \frac{1}{2} \delta_{(-\mathbf{z}, \mathbf{u}^-)}$ where $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q$ with $(\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_I$ and $(-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_I$. The semi-infinite constraint in (2.35) is thus equivalent to

$$\begin{aligned} \boldsymbol{\gamma}^\top \mathbf{B}(\mathbf{u}^+ + \mathbf{u}^-) + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) (\mathbb{I}_{[(\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_i]} + \mathbb{I}_{[(-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_i]}) &\leq \mathbb{I}_{[\mathbf{S}\mathbf{z} \leq \mathbf{t}]} + \mathbb{I}_{[-\mathbf{S}\mathbf{z} \leq \mathbf{t}]} \\ \forall (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_I, (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_I. \end{aligned}$$

As the support $\mathcal{C}_I$ can be partitioned into the subsets $\overline{\mathcal{C}}_i$, $i \in \mathcal{I}$, we may reformulate the above semi-infinite

constraint as

$$\begin{aligned} \gamma^\top \mathbf{B}(\mathbf{u}^+ + \mathbf{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) &\leq \mathbb{I}_{[\mathbf{S}\mathbf{z} \leq \mathbf{t}]} + \mathbb{I}_{[-\mathbf{S}\mathbf{z} \leq \mathbf{t}]} \\ \forall \mathbf{i} \in \mathcal{I}^2, \forall (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : (\mathbf{z}, \mathbf{u}^+) &\in \bar{\mathcal{C}}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \bar{\mathcal{C}}_{i^-}. \end{aligned}$$

By expanding the right-hand side indicator functions as the pointwise minima of indicator functions on halfspaces, the semi-infinite constraint reduces to

$$\begin{aligned} \gamma^\top \mathbf{B}(\mathbf{u}^+ + \mathbf{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) &\leq w_j \\ \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2, \forall (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) &\in \bar{\mathcal{D}}_{\mathbf{ij}}, \end{aligned} \quad (2.36)$$

where

$$\bar{\mathcal{D}}_{\mathbf{ij}} = \left\{ (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : \begin{array}{ll} (\mathbf{z}, \mathbf{u}^+) \in \bar{\mathcal{C}}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \bar{\mathcal{C}}_{i^-} & \\ j^+ > 0 \implies \mathbf{s}_{j^+}^\top \mathbf{z} > t_{j^+} & \\ j^- > 0 \implies -\mathbf{s}_{j^-}^\top \mathbf{z} > t_{j^-} & \end{array} \right\}$$

for all $\mathbf{i} \in \mathcal{I}^2$. As the semi-infinite constraint (2.36) is affine in the uncertain parameters $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-)$, we can use Lemma 2.3 to replace the parameter range $\bar{\mathcal{D}}_{\mathbf{ij}}$ with its convex hull $\mathcal{D}_{\mathbf{ij}}$, which is defined in (2.20).

If $\mathcal{D}_{\mathbf{ij}} = \emptyset$, then the subconstraint indexed by $(\mathbf{i}, \mathbf{j})$ is vacuously satisfied. Conversely, if $\mathcal{D}_{\mathbf{ij}} \neq \emptyset$, then we can use a standard continuity argument to replace $\mathcal{D}_{\mathbf{ij}}$ with its closure

$$\text{cl}(\mathcal{D}_{\mathbf{ij}}) = \left\{ (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : \begin{array}{ll} (\mathbf{z}, \mathbf{u}^+) \in \mathcal{C}_{i^+}, (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_{i^-} & \\ j^+ > 0 \implies \mathbf{s}_{j^+}^\top \mathbf{z} \geq t_{j^+} & \\ j^- > 0 \implies -\mathbf{s}_{j^-}^\top \mathbf{z} \geq t_{j^-} & \end{array} \right\}.$$

The feasibility condition **(F)** implies that $\text{cl}(\mathcal{D}_{\mathbf{ij}})$ is essentially strictly feasible, and the conic Farkas

lemma [15, Example 2.26] allows us to reformulate (2.36) as

$$\begin{aligned}
& \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \exists \psi_{\mathbf{ij}}^+ \in \mathcal{K}_{i^+}^*, \psi_{\mathbf{ij}}^- \in \mathcal{K}_{i^-}^*, \chi_{\mathbf{ij}}^+, \chi_{\mathbf{ij}}^- \in \mathbb{R}_+ : \\
& \gamma^\top \mathbf{B}(\mathbf{u}^+ + \mathbf{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \\
& - \psi_{\mathbf{ij}}^{+\top} (\mathbf{C}_{i^+} \mathbf{z} + \mathbf{D}_{i^+} \mathbf{u}^+ - \mathbf{d}_{i^+}) - \psi_{\mathbf{ij}}^{-\top} (-\mathbf{C}_{i^-} \mathbf{z} + \mathbf{D}_{i^-} \mathbf{u}^- - \mathbf{d}_{i^-}) \\
& + \chi_{\mathbf{ij}}^+ (\mathbf{s}_{j^+}^\top \mathbf{z} - t_{j^+}) + \chi_{\mathbf{ij}}^- (-\mathbf{s}_{j^-}^\top \mathbf{z} - t_{j^-}) \leq w_{\mathbf{j}} \quad \forall (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q.
\end{aligned}$$

The claim now follows from the observation that an affine function is bounded from above uniformly on the entire space if and only if it is bounded above at the origin and if its gradient vanishes. $\square$

Proof of Theorem 2.5. The duality theory for moment problems allows us to express the worst-case probability (2.1) as the optimal value of the problem

$$\begin{aligned}
& \sup \quad \beta + \mathbf{b}^\top \boldsymbol{\gamma} \\
& \text{s.t.} \quad \beta \in \mathbb{R}, \boldsymbol{\gamma} \in \mathbb{R}^K \\
& \mathbb{E}_{\mathbb{P}} [\beta + (\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma}] \leq \mathbb{E}_{\mathbb{P}} [\mathbb{I}_{\{\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}\}}] \quad \forall \mathbb{P} \in \mathcal{P}^\alpha(\mathcal{C}_1),
\end{aligned} \tag{2.37}$$

where $\mathcal{P}^\alpha(\mathcal{C}_1)$ contains all distributions $\mathbb{P} \in \mathcal{P}_0(\mathcal{C}_1)$ for which $\Pi_{\tilde{\mathbf{z}}} \mathbb{P}$ is α -unimodal around $\mathbf{m} = \mathbf{0}$. Strong duality holds due to the Slater condition (S), which simplifies in the current setting to the requirement that

$$\mathcal{C}_1^\alpha = \{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_1 \mathbf{z} \preccurlyeq_{\kappa_1} \mathbf{d}_1, \mathbf{g}^\alpha(\mathbf{z}) \preccurlyeq_{\kappa_2} \mathbf{u}\}$$

contains the vector $\mathbf{b}$ in its interior, where $\mathbf{g}^\alpha(\mathbf{z}) = \int_0^1 \mathbf{g}(t\mathbf{z}) \alpha t^{\alpha-1} dt$. Any distribution in $\mathcal{P}^\alpha(\mathcal{C}_1)$ can be represented as a mixture of distributions $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ parameterised by $(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q$ with the following properties:

(P1) $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ is supported on $\mathcal{C}_1$;

(P2) $\Pi_{\tilde{\mathbf{z}}} \mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ is the radial α -unimodal distribution on the interval $[0, \mathbf{z}]$;

(P3) $\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}}[\tilde{\mathbf{u}}] = \mathbf{u}$.

Note that if a distribution $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ satisfies (P1) and (P2) for some $(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q$, then $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha$ as

$\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\mathbf{g}(\tilde{\mathbf{z}})] = \mathbf{g}^\alpha(\mathbf{z})$ due to (P2). Conversely, for every $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha$ there exists a distribution $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ that satisfies (P1) and (P2). The constraint in (2.37) is therefore equivalent to

$$\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\beta + (\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma}] \leq \mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\mathbb{I}_{[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]}] \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha. \quad (2.38)$$

We emphasise that $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ is not uniquely determined by $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha$, but all distributions $\mathbb{V}_{(\mathbf{z}, \mathbf{u})}$ satisfying (P1)–(P3) for a given $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha$ have the same effect on constraint (2.38). Note that $\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\tilde{\mathbf{z}}] = \frac{\alpha}{\alpha+1} \mathbf{z}$ and $\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\tilde{\mathbf{u}}] = \mathbf{u}$ by (P2) and (P3), respectively. As $\mathbf{t} \geq \mathbf{0}$, (P2) further implies

$$\mathbb{E}_{\mathbb{V}_{(\mathbf{z}, \mathbf{u})}} [\mathbb{I}_{[\mathbf{S}\tilde{\mathbf{z}} \leq \mathbf{t}]}] = \int_0^1 \mathbb{I}_{[\mathbf{S}\ell \mathbf{z} \leq \mathbf{t}]} \alpha \ell^{\alpha-1} d\ell = \int_0^{F(\mathbf{z})} \alpha \ell^{\alpha-1} d\ell = F(\mathbf{z})^\alpha,$$

where $F(\mathbf{z}) = \min\{1, \inf\{t_j/\mathbf{s}_j^\top \mathbf{z} : j \in \mathcal{J} : \mathbf{s}_j^\top \mathbf{z} > 0\}\}$. Therefore, constraint (2.38) can be decomposed into the following semi-infinite constraints.

$$\left. \begin{aligned} \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}\right)^\top \boldsymbol{\gamma} &\leq 1 \\ \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}\right)^\top \boldsymbol{\gamma} &\leq (t_j/\mathbf{s}_j^\top \mathbf{z})^\alpha \quad \forall j \in \mathcal{L} : \mathbf{s}_j^\top \mathbf{z} > 0 \end{aligned} \right\} \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1^\alpha \quad (2.39)$$

By the conic Farkas lemma, which applies since $\mathcal{C}_1^\alpha$ inherits essential strict feasibility from $\mathcal{C}_1$, the first constraint in (2.39) holds if and only if there is $\boldsymbol{\phi} \in \mathcal{K}_2^*$, $\boldsymbol{\psi}_0 \in \mathcal{K}_1^*$ with

$$\begin{aligned} \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}\right)^\top \boldsymbol{\gamma} + (\mathbf{u} - \mathbf{g}^\alpha(\mathbf{z}))^\top \boldsymbol{\phi} + (\mathbf{d}_1 - \mathbf{C}_1 \mathbf{z})^\top \boldsymbol{\psi}_0 &\leq 1 \\ \forall (\mathbf{z}, \mathbf{u}) &\in \mathbb{R}^P \times \mathbb{R}^Q, \end{aligned}$$

which is equivalent to the existence of $\boldsymbol{\psi}_0 \in \mathcal{K}_1^*$ with $-\mathbf{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^*$ and

$$\beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z})\right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1 \mathbf{z})^\top \boldsymbol{\psi}_0 \leq 1 \quad \forall \mathbf{z} \in \mathbb{R}^P. \quad (2.40)$$

Note that $\mathbf{g}^\alpha(\mathbf{z})$ inherits $\mathcal{K}_2$ -convexity from $\mathbf{g}(\mathbf{z})$ since $\mathcal{K}_2$ is convex. As $-\mathbf{B}^\top \boldsymbol{\gamma} \in -\mathcal{K}_2^*$, the left-hand side of the above constraint is thus concave in $\mathbf{z}$.

For $j \in \mathcal{L}$ one can use a continuity argument to show that the strict inequality $\mathbf{s}_j^\top \mathbf{z} > 0$ in (2.39) can

be replaced by a weak inequality. By the conic Farkas lemma, which applies as $j \in \mathcal{L}$, the j -th constraint in the second line of (2.39) thus holds if and only if there exists $\boldsymbol{\psi}_j \in \mathcal{K}_1^*$ with $-\mathbf{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^*$ and

$$\beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi}_j \leq (t_j/\mathbf{s}_j^\top \mathbf{z})^\alpha \quad \forall \mathbf{z} \in \mathbb{R}^P : \mathbf{s}_j^\top \mathbf{z} \geq 0.$$

Introducing the concave auxiliary function

$$f_j(\mathbf{z}) = \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi}_j$$

to simplify the notation, this semi-infinite constraint can be reformulated as

$$\begin{aligned} 0 &\geq \sup_{\mathbf{z} \in \mathbb{R}^P, \kappa_j \in \mathbb{R}_+} \inf_{\tau_j \in \mathbb{R}_+} f_j(\mathbf{z}) - (t_j/\kappa_j)^\alpha + \tau_j(\mathbf{s}_j^\top \mathbf{z} - \kappa_j) \\ &= \inf_{\tau_j \in \mathbb{R}_+} \sup_{\mathbf{z} \in \mathbb{R}^P, \kappa_j \in \mathbb{R}_+} f_j(\mathbf{z}) - (t_j/\kappa_j)^\alpha + \tau_j(\mathbf{s}_j^\top \mathbf{z} - \kappa_j) \\ &= \inf_{\tau_j \in \mathbb{R}_+} \sup_{\mathbf{z} \in \mathbb{R}^P} f_j(\mathbf{z}) + \tau_j \mathbf{s}_j^\top \mathbf{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j t_j)^{\frac{\alpha}{\alpha+1}}, \end{aligned}$$

where the inner minimisation over τ_j in the first line enforces the implicit constraint $\mathbf{s}_j^\top \mathbf{z} \geq \kappa_j$ for the outer maximisation, while the interchange of minimisation and maximisation in the second line is justified by the classical saddle point theorem, which applies because $j \in \mathcal{L}$.

In summary, the j -th constraint in the second line of (2.39) holds if and only if there exist $\boldsymbol{\psi}_j \in \mathcal{K}_1^*$ and $\tau_j \in \mathbb{R}_+$ with $-\mathbf{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^*$ and

$$\begin{aligned} \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{g}^\alpha(\mathbf{z}) \right)^\top \boldsymbol{\gamma} + (\mathbf{d}_1 - \mathbf{C}_1\mathbf{z})^\top \boldsymbol{\psi}_j \\ + \tau_j \mathbf{s}_j^\top \mathbf{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j t_j)^{\frac{\alpha}{\alpha+1}} \leq 0 \quad \forall \mathbf{z} \in \mathbb{R}^P. \end{aligned} \tag{2.41}$$

The claim now follows by substituting (2.40) and (2.41) into (2.37). $\square$

In order to prove Theorem 2.6, we first derive the following auxiliary result.

Lemma 2.4. *The ambiguity set*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{P}[\tilde{\mathbf{z}} = \mathbf{0}] = 2^{-P}, \quad \mathbb{P}[\tilde{\mathbf{z}} = \mathbf{e}] = 2^{-P} \\ \text{the components of } \tilde{\mathbf{z}} \text{ are independent} \end{array} \right\} \quad (2.42)$$

satisfies $\mathcal{P} = \{\mathbb{P}^0\}$ for the distribution $\mathbb{P}^0$ which places all probability mass uniformly on the vertices of the unit hypercube in $\mathbb{R}^P$:

$$\mathbb{P}^0 = \frac{1}{2^P} \sum_{\mathbf{x} \in \{0,1\}^P} \delta_{\mathbf{x}}.$$

Proof of Lemma 2.4. By construction, we have $\mathbb{P}^0 \in \mathcal{P}$. To show that $\mathbb{P}^0$ is the only distribution in $\mathcal{P}$, fix any distribution $\mathbb{P} \in \mathcal{P}$ and let $\rho_p^0 = \mathbb{P}[\tilde{z}_p = 0]$ and $\rho_p^1 = \mathbb{P}[\tilde{z}_p = 1]$, $p = 1, \dots, P$. By construction, we have $\mathbf{0} \leq \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \leq \mathbf{e}$. Since $\mathbb{P} \in \mathcal{P}$, the components of $\tilde{\mathbf{z}}$ are independent, and we obtain that

$$\mathbb{P}[\tilde{\mathbf{z}} = \mathbf{0}] = \prod_{p=1}^P \mathbb{P}[\tilde{z}_p = 0] = \prod_{p=1}^P \rho_p^0 = 2^{-P}$$

and

$$\mathbb{P}[\tilde{\mathbf{z}} = \mathbf{e}] = \prod_{p=1}^P \mathbb{P}[\tilde{z}_p = 1] = \prod_{p=1}^P \rho_p^1 = 2^{-P}.$$

We show via induction on P that $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2$ is the unique maximiser to

$$\min \left\{ \prod_{p=1}^P \rho_p^0, \prod_{p=1}^P \rho_p^1 \right\} \quad (2.43)$$

over $\{(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1) \in \mathbb{R}^{2P} : \mathbf{0} \leq \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \leq \mathbf{e}\}$. The statement of the lemma then follows from the fact that

$$\prod_p \rho_p^0 = \prod_p \rho_p^1 = 2^{-P} \text{ for } \boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2.$$

The statement trivially holds for $P = 1$. Assume now that $\boldsymbol{\rho}^0, \boldsymbol{\rho}^1 \in \mathbb{R}^{P+1}$ is a maximiser to (2.43) over $\{(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1) \in \mathbb{R}^{2P+2} : \mathbf{0} \leq \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \leq \mathbf{e}\}$ that does not satisfy $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2$. Without loss of generality, assume that $\rho_{P+1}^0 = \delta < 1/2$ and $\rho_{P+1}^1 = 1 - \delta$. We then have that

$$\prod_{p=1}^{P+1} \rho_p^0 = \delta \cdot \prod_{p=1}^P \rho_p^0 \geq \frac{1}{2^{P+1}} \iff \prod_{p=1}^P \rho_p^0 \geq \frac{1}{2^{P+1}} \cdot \frac{1}{\delta}$$

and

$$\prod_{p=1}^{P+1} \rho_p^1 = (1-\delta) \cdot \prod_{p=1}^P \rho_p^1 \geq \frac{1}{2^{P+1}} \iff \prod_{p=1}^P \rho_p^1 \geq \frac{1}{2^{P+1}} \cdot \frac{1}{1-\delta}.$$

Consider $\hat{\rho}^0$ and $\hat{\rho}^1$ defined through $\hat{\rho}_p^0 = \hat{\rho}_p^1 = \sqrt{\rho_p^0 \rho_p^1}$, $p = 1, \dots, P$. Since $\sqrt{\rho_p^0 \rho_p^1} \leq \sqrt{\rho_p^0(1-\rho_p^0)} \leq 1/2$, we have that $\mathbf{0} \leq \hat{\rho}^0 + \hat{\rho}^1 \leq \mathbf{e}$. We now see that

$$\begin{aligned} \prod_{p=1}^P \hat{\rho}_p^0 &= \prod_{p=1}^P \sqrt{\rho_p^0 \rho_p^1} = \sqrt{\prod_{p=1}^P \rho_p^0} \sqrt{\prod_{p=1}^P \rho_p^1} \geq \sqrt{\frac{1}{2^{P+1}} \cdot \frac{1}{\delta}} \sqrt{\frac{1}{2^{P+1}} \cdot \frac{1}{1-\delta}} \\ &= \frac{1}{2^P} \sqrt{\frac{1}{4\delta(1-\delta)}} > \frac{1}{2^P} \end{aligned}$$

for $\delta < 1/2$ and similarly $\prod_{p=1}^P \hat{\rho}_p^1 > 2^{-P}$, which contradicts the fact that $\rho^0 = \rho^1 = \mathbf{e}/2 \in \mathbb{R}^P$ is the unique maximiser to (2.43) over $\{(\rho^0, \rho^1) \in \mathbb{R}^{2P} : \mathbf{0} \leq \rho^0 + \rho^1 \leq \mathbf{e}\}$. A similar argument can be constructed if $\rho_{P+1}^0 = \delta > 1/2$, which concludes the proof. $\square$

Proof of Theorem 2.6. Fix an instance $(\mathbf{E}, \mathbf{f})$ of the IP problem and consider the following instance of the uncertainty quantification problem (2.1):

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{E}\tilde{\mathbf{z}} \leq \mathbf{f}], \quad (2.44)$$

where the ambiguity set $\mathcal{P}$ is defined as in (2.42). Note that this ambiguity set satisfies the boundedness condition (B) and the nesting condition (N) since the two singleton sets $\{\mathbf{0}\}$ and $\{\mathbf{e}\}$ are disjoint and both contained in the interior of $\mathbb{R}^P$. Moreover, the bit-length of the probabilities $\underline{p}_i = \bar{p}_i = 2^{-P}$ in the ambiguity set $\mathcal{P}$ is bounded from above by the bit-length of $\mathbf{E} \in \mathbb{R}^{M \times P}$. From Lemma 2.4 we conclude that problem (2.44) evaluates to

$$\frac{1}{2^P} \sum_{\mathbf{x} \in \{0,1\}^P} \delta_{\mathbf{x}}(\mathbf{E}\tilde{\mathbf{z}} \leq \mathbf{f}) = \frac{1}{2^P} |\{\mathbf{x} \in \{0,1\}^P : \mathbf{E}\mathbf{x} \leq \mathbf{f}\}|,$$

that is, (2.44) evaluates to 0 if and only if the IP problem has a negative answer. The statement now follows since the IP problem is strongly NP-hard. $\square$

Proof of Theorem 2.7. We can employ Theorem 2.3 to conclude that the worst-case probability of the chance constraint in (2.2) is given by

$$\begin{aligned}
& \sup \quad \beta + \mathbf{b}^\top \boldsymbol{\gamma} \\
& \text{s.t.} \quad \beta \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^*, \quad \tau \in \mathbb{R}_+ \\
& \quad \beta + \mathbf{d}_1^\top \boldsymbol{\phi} \leq 1, \quad \mathbf{A}^\top \boldsymbol{\gamma} = \mathbf{C}_1^\top \boldsymbol{\phi}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\phi} \\
& \quad \beta + \mathbf{d}_1^\top \boldsymbol{\psi} - \tau t(\mathbf{x}) \leq 0 \\
& \quad \mathbf{A}^\top \boldsymbol{\gamma} + \tau \mathbf{s}(\mathbf{x}) = \mathbf{C}_1^\top \boldsymbol{\psi}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}
\end{aligned} \quad \left. \vphantom{\sup} \right\} \text{ if } \mathbf{x} \in \mathcal{X}^+, \tag{2.45}$$

where $\mathcal{X}^+ = \{\mathbf{x} \in \mathcal{X} : \exists (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1, \mathbf{s}(\mathbf{x})^\top \mathbf{z} > t(\mathbf{x})\}$ and where we have substituted $\beta = \lambda_1 - \kappa_1$.

In the remainder we will separately investigate the cases $\mathbf{x} \in \mathcal{X}^+$ and $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$. If $\mathbf{x} \in \mathcal{X}^+$, then the chance constraint in (2.2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^*$ and $\tau \in \mathbb{R}_+$ such that

$$\begin{aligned}
& \beta + \mathbf{b}^\top \boldsymbol{\gamma} \geq 1 - \varepsilon, \quad \beta + \mathbf{d}_1^\top \boldsymbol{\phi} \leq 1, \quad \beta + \mathbf{d}_1^\top \boldsymbol{\psi} - \tau t(\mathbf{x}) \leq 0 \\
& \mathbf{A}^\top \boldsymbol{\gamma} = \mathbf{C}_1^\top \boldsymbol{\phi}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\phi} \\
& \mathbf{A}^\top \boldsymbol{\gamma} + \tau \mathbf{s}(\mathbf{x}) = \mathbf{C}_1^\top \boldsymbol{\psi}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}.
\end{aligned} \tag{2.46}$$

Any τ feasible in (2.46) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. Then all constraints involving $\boldsymbol{\phi}$ are redundant as we may set $\boldsymbol{\phi} = \boldsymbol{\psi}$ without loss of generality. Thus, (2.46) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi} \in \mathcal{K}_1^*$ such that

$$\beta + \mathbf{b}^\top \boldsymbol{\gamma} \geq 1 - \varepsilon, \quad \beta + \mathbf{d}_1^\top \boldsymbol{\psi} \leq 0, \quad \mathbf{A}^\top \boldsymbol{\gamma} = \mathbf{C}_1^\top \boldsymbol{\psi}, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi},$$

and the conic Farkas lemma implies that this set of equations is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi} \in \mathcal{K}_1^*$ such that

$$\beta + \mathbf{b}^\top \boldsymbol{\gamma} \geq 1 - \varepsilon, \quad \beta + (\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u})^\top \boldsymbol{\gamma} \leq 0 \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1.$$

Taking expectation with respect to any probability distribution $\mathbb{P} \in \mathcal{P}$, we thus obtain $\beta + \mathbf{b}^\top \boldsymbol{\gamma} =$

$\mathbb{E}_{\mathbb{P}}[\beta + (\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma}] \leq 0$, which contradicts the first constraint in the above expression. We conclude that any feasible τ must be strictly positive. Dividing all constraints in (2.46) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\tau$, $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/\tau$, $\boldsymbol{\psi} \leftarrow \boldsymbol{\psi}/\tau$ and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2.2) is satisfied if and only if (2.26) holds.

Assume now that $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$, that is, $\mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq t(\mathbf{x})$ for all $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1$. In this case, the worst-case probability in (2.2) evaluates to 1 and the chance constraint is vacuously satisfied. We show that (2.26) is vacuously satisfied as well. Indeed, if we set $(\beta, \boldsymbol{\gamma}, \boldsymbol{\phi}, \tau) = (0, \mathbf{0}, \mathbf{0}, 0)$, then (2.26) reduces to

$$\exists \boldsymbol{\psi} \in \mathcal{K}_1^* : \mathbf{d}_1^\top \boldsymbol{\psi} \leq t(\mathbf{x}), \quad \mathbf{s}(\mathbf{x}) = \mathbf{C}_1^\top \boldsymbol{\psi}, \quad \mathbf{D}_1^\top \boldsymbol{\psi} = \mathbf{0}.$$

The conic Farkas lemma implies that this statement is equivalent to the assertion that $\mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq t(\mathbf{x})$ for all $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1$, which is true since $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$.

In summary, we conclude that the chance constraint in (2.2) is equivalent to (2.26) for all $\mathbf{x} \in \mathcal{X}$. Thus, the claim follows. $\square$

Proof of Theorem 2.8. Fix an instance $(\mathbf{E}, \mathbf{f})$ of the IP problem and consider the following instance of (2.2), which is parameterised in $n = 1, \dots, P$:

$$\begin{aligned} \sup \quad & 0 \\ \text{s.t.} \quad & \mathbf{x} \in [0, 1]^P, \quad \mathbf{E}\mathbf{x} \leq \mathbf{f}, \quad \mathbf{e}^\top \mathbf{x} = n \\ & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\tilde{\mathbf{z}}^\top \mathbf{x} \geq 1] \geq \frac{n}{P}, \end{aligned} \tag{2.47}$$

where the ambiguity set $\mathcal{P}$ satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{P}[\tilde{\mathbf{z}} \in [-2, 2]^P] = 1, \quad \mathbb{P}[\tilde{\mathbf{z}} = \mathbf{e}_p] = \frac{1}{P} \quad \forall p = 1, \dots, P \right\}.$$

Here, $\mathbf{e}$ and $\mathbf{e}_p$ denote the vector of all ones in $\mathbb{R}^P$ and the p -th unit basis vector in $\mathbb{R}^P$, respectively. The ambiguity set $\mathcal{P}$ is an instance of (2.4) that satisfies the conditions (B) and (N) since the confidence sets $\{\mathbf{e}_p\}$, $p = 1, \dots, P$, are pairwise disjoint subsets of the interior of the support $[-2, 2]^P$.

We want to show that problem (2.47) is feasible if and only if the IP problem has a solution $\mathbf{x}$ with exactly n components being one. Indeed, we observe that $\mathcal{P} = \{\mathbb{P}_U\}$, where $\mathbb{P}_U$ denotes the uniform distribution over the vertices of the probability simplex in $\mathbb{R}^P$. Thus, one readily verifies that if there is a solution $\mathbf{x}$ to the IP problem with $\mathbf{e}^\top \mathbf{x} = n$, then $\mathbf{x}$ is a feasible solution to problem (2.47).

Conversely, assume that $\mathbf{x}$ solves (2.47). The box constraint $\mathbf{x} \in [0, 1]^P$ and the chance constraint in (2.47) then imply that at least n components of $\mathbf{x}$ are one. From the constraint $\mathbf{e}^\top \mathbf{x} = n$ we then conclude that exactly n components of $\mathbf{x}$ are one and all other components are zero, that is, $\mathbf{x}$ is a solution of the IP problem with exactly n components being one.

In conclusion, problem (2.47) is feasible if and only if the IP problem has a solution $\mathbf{x}$ with exactly n components being one. We can therefore decide the IP problem by checking whether $\mathbf{E}\mathbf{0} \leq \mathbf{f}$ and otherwise sequentially verifying the feasibility of problem (2.47) for $n = 1, \dots, P$. $\square$

Proof of Theorem 2.9. We can employ Theorem 2.4 and conduct some simple algebraic manipulations to conclude that the worst-case probability of the chance constraint in (2.2) is given by

$$\begin{aligned} & \sup \quad \beta + \mathbf{b}^\top \boldsymbol{\gamma} \\ & \text{s.t.} \quad \beta \in \mathbb{R}, \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \quad \boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^*, \quad \tau \in \mathbb{R}_+ \\ & \left. \begin{aligned} & \beta + \mathbf{d}_1^\top \boldsymbol{\phi} \leq 1, \quad \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\phi} \\ & 2\beta + \mathbf{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - \tau t(\mathbf{x}) \leq 1 \\ & \mathbf{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \tau \mathbf{s}(\mathbf{x}) \\ & \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}^+ = \mathbf{D}_1^\top \boldsymbol{\psi}^- \end{aligned} \right\} \quad \text{if } \mathbf{x} \in \mathcal{X}^\pm, \end{aligned}$$

where $\mathcal{X}^\pm = \{\mathbf{x} \in \mathcal{X} : \exists (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1 . \mathbf{s}(\mathbf{x})^\top \mathbf{z} < -t(\mathbf{x}) \vee \mathbf{s}(\mathbf{x})^\top \mathbf{z} > t(\mathbf{x})\}$.

In the remainder we will separately investigate the cases $\mathbf{x} \in \mathcal{X}^\pm$ and $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$. If $\mathbf{x} \in \mathcal{X}^\pm$, then the chance constraint in (2.2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^*$ and

$\tau \in \mathbb{R}_+$ such that

$$\begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq 1 - \varepsilon, & \beta + \mathbf{d}_1^\top \boldsymbol{\phi} &\leq 1 \\ 2\beta + \mathbf{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - \tau t(\mathbf{x}) &\leq 1, & \mathbf{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) &= \tau \mathbf{s}(\mathbf{x}) \\ \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}^+ &= \mathbf{D}_1^\top \boldsymbol{\psi}^- = \mathbf{D}_1^\top \boldsymbol{\phi} \end{aligned} \quad (2.48)$$

Any τ feasible in (2.48) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. We can then conclude that all constraints involving $\boldsymbol{\phi}$ are redundant as we may set $\boldsymbol{\phi} = (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-)/2$ without loss of generality. Thus, the constraint system (2.48) is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^*$ such that

$$\begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq 1 - \varepsilon, & 2\beta + \mathbf{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) &\leq 1 \\ \mathbf{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) &= \mathbf{0}, & \mathbf{B}^\top \boldsymbol{\gamma} = \mathbf{D}_1^\top \boldsymbol{\psi}^+ &= \mathbf{D}_1^\top \boldsymbol{\psi}^-, \end{aligned}$$

and the conic Farkas lemma implies that this set of equations is satisfiable if and only if there exist $\beta \in \mathbb{R}$ and $\boldsymbol{\gamma} \in \mathbb{R}^K$ such that

$$\begin{aligned} \beta + \mathbf{b}^\top \boldsymbol{\gamma} &\geq 1 - \varepsilon \\ 2\beta + (\mathbf{B}\mathbf{u}^+ + \mathbf{B}\mathbf{u}^-)^\top \boldsymbol{\gamma} &\leq 1 \quad \forall (\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : \\ &(\mathbf{z}, \mathbf{u}^+), (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_1. \end{aligned} \quad (2.49)$$

Note that every distribution $\mathbb{P} \in \mathcal{P}$ is a mixture of extremal distributions of the form $\frac{1}{2}\delta_{(\mathbf{z}, \mathbf{u}^+)} + \frac{1}{2}\delta_{(-\mathbf{z}, \mathbf{u}^-)}$, where $(\mathbf{z}, \mathbf{u}^+), (-\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_1$. The semi-infinite constraint in (2.49) implies that $\mathbb{E}_{\mathbb{P}}[2\beta + 2(\mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma}] \leq 1$ for all extremal distributions $\mathbb{P} \in \mathcal{P}$, and we thus obtain that $\beta + \mathbf{b}^\top \boldsymbol{\gamma} = \mathbb{E}_{\mathbb{P}}[\beta + (\mathbf{B}\tilde{\mathbf{u}})^\top \boldsymbol{\gamma}] \leq 1/2$, which contradicts the first constraint in (2.49) since $\varepsilon < 1/2$. We thus conclude that any feasible τ must be strictly positive. Dividing all constraints in (2.48) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\tau$, $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/\tau$, $\boldsymbol{\psi}^+ \leftarrow \boldsymbol{\psi}^+/\tau$, $\boldsymbol{\psi}^- \leftarrow \boldsymbol{\psi}^-/\tau$ and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2.2) is satisfied if and only if (2.27) holds.

Assume now that $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$, that is, $-t(\mathbf{x}) \leq \mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq t(\mathbf{x})$ for all $(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_1$. In this case, the worst-case probability in (2.2) evaluates to 1 and the chance constraint is vacuously satisfied. We show

that (2.27) is vacuously satisfied as well. Indeed, if we set $(\beta, \gamma, \phi, \tau) = (0, \mathbf{0}, \mathbf{0}, 0)$, then (2.27) becomes

$$\exists \psi^+, \psi^- \in \mathcal{K}_1^* : \begin{cases} \mathbf{d}_1^\top (\psi^+ + \psi^-) \leq t(\mathbf{x}), \\ \mathbf{C}_1^\top (\psi^+ - \psi^-) = \mathbf{s}(\mathbf{x}), \\ \mathbf{D}_1^\top \psi^+ = \mathbf{D}_1^\top \psi^- = \mathbf{0}. \end{cases}$$

The conic Farkas lemma implies that this statement is equivalent to the assertion that $\mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq t(\mathbf{x})$ for all $(\mathbf{z}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q$ such that $(\mathbf{z}, \mathbf{u}^+), (\mathbf{z}, \mathbf{u}^-) \in \mathcal{C}_1$, which is true since $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$.

In summary, we conclude that the chance constraint in (2.2) is equivalent to (2.27) for all $\mathbf{x} \in \mathcal{X}$.

Thus, the claim follows. $\square$

Proof of Theorem 2.10. The proof follows the same argument as the proof of Theorem 2.8. We only need to replace $\mathcal{P}$ with the symmetric ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}) : \begin{array}{l} \mathbb{P} \left[\begin{pmatrix} \tilde{z} \\ \tilde{u} \end{pmatrix} \in [-2, 2]^P \times \mathbb{R} \right] = 1, \\ \mathbb{P} \left[\begin{pmatrix} \tilde{z} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_p \\ 0 \end{pmatrix} \right] = \frac{1}{2P} \\ \mathbb{P} \left[\begin{pmatrix} \tilde{z} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} -\mathbf{e}_p \\ 0 \end{pmatrix} \right] = \frac{1}{2P} \end{array} \right\} \quad \forall p = 1, \dots, P$$

and set the right-hand side of the chance constraint in (2.47) to $n/(2P)$. $\square$

Proof of Theorem 2.11. By Theorem 2.5, the chance constraint in (2.2) is satisfied if and only

if there exist $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\phi, \psi \in \mathcal{K}_1^*$ with

$$\left. \begin{aligned} \beta + \mathbf{b}^\top \gamma &\geq 1 - \varepsilon, \quad -\mathbf{B}^\top \gamma \in \mathcal{K}_2^* \\ \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}z + \mathbf{B}g^\alpha(z) \right)^\top \gamma + (\mathbf{d}_1 - \mathbf{C}_1 z)^\top \phi &\leq 1 \\ \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}z + \mathbf{B}g^\alpha(z) \right)^\top \gamma + (\mathbf{d}_1 - \mathbf{C}_1 z)^\top \psi \\ + \tau \mathbf{s}(\mathbf{x})^\top z - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau t(\mathbf{x}))^{\frac{\alpha}{\alpha+1}} &\leq 0 \quad \text{if } \mathbf{x} \in \mathcal{X}^+ \end{aligned} \right\} \quad \forall z \in \mathbb{R}^P, \quad (2.50)$$

where $\mathcal{X}^+ = \{\mathbf{x} \in \mathcal{X} : \exists (z, u) \in \mathcal{C}_1^\alpha, \mathbf{s}_1(\mathbf{x})^\top z > 0\}$.

In the remainder, we will separately investigate the cases $\mathbf{x} \in \mathcal{X}^+$ and $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$. If $\mathbf{x} \in \mathcal{X}^+$, then any τ feasible in (2.50) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. Then (2.50) is satisfiable by any $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\phi, \psi \in \mathcal{K}_1^*$ if and only if it is satisfiable by any $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\phi, \psi \in \mathcal{K}_1^*$ with $\phi = \psi$. We may thus set $\phi = \psi$, in which case the first semi-infinite constraint becomes redundant. Thus, (2.50) is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$, $\psi \in \mathcal{K}_1^*$ with

$$\begin{aligned} \beta + \mathbf{b}^\top \gamma &\geq 1 - \varepsilon, \quad -\mathbf{B}^\top \gamma \in \mathcal{K}_2^* \\ \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}z + \mathbf{B}g^\alpha(z) \right)^\top \gamma + (\mathbf{d}_1 - \mathbf{C}_1 z)^\top \psi &\leq 0 \quad \forall z \in \mathbb{R}^P. \end{aligned}$$

Using the conic Farkas lemma, one can show that these constraints hold if and only if there exists $\beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}^K$ with

$$\begin{aligned} \beta + \mathbf{b}^\top \gamma &\geq 1 - \varepsilon \\ \beta + \left(\frac{\alpha}{\alpha+1} \mathbf{A}z + \mathbf{B}u \right)^\top \gamma &\leq 0 \quad \forall (z, u) \in \mathcal{C}_1^\alpha. \end{aligned}$$

Taking the expectation of the second constraint with respect to any $\mathbb{P} \in \mathcal{P}$ implies that $\beta + \mathbf{b}^\top \gamma = \mathbb{E}_{\mathbb{P}}[\beta + (\mathbf{A}\tilde{z} + \mathbf{B}\tilde{u})^\top \gamma] \leq 0$, which contradicts the first constraint since $\varepsilon < 1$. We thus conclude that any τ feasible in (2.50) must be strictly positive. Dividing all constraints in (2.50) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\gamma \leftarrow \gamma/\tau$, $\phi \leftarrow \phi/\tau$, $\psi \leftarrow \psi/\tau$, and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2.2) is satisfied if and only if (2.28) holds.

Assume next that $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$, which implies that $\mathbf{s}(\mathbf{x})^\top z \leq 0 \leq t(\mathbf{x})$ for all $z \in \{z \in \mathbb{R}^P : \mathbf{C}_1 z \preceq_{\mathcal{K}_1} \mathbf{d}_1\}$. In this case, the chance constraint is vacuously satisfied as $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{s}(\mathbf{x})^\top \tilde{z} \leq t(\mathbf{x})] = 1$. We show that (2.28) is vacuously satisfied as well. Indeed, if we set $(\beta, \gamma, \phi, \tau) = (0, \mathbf{0}, \mathbf{0}, 0)$, then (2.28) reduces

to

$$\exists \boldsymbol{\psi} \in \mathcal{K}_1^* : (\mathbf{d}_1 - \mathbf{C}_1 \mathbf{z})^\top \boldsymbol{\psi} + \mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq 0 \quad \forall \mathbf{z} \in \mathbb{R}^P.$$

Strong conic duality implies that this statement is equivalent to the assertion that $\mathbf{s}(\mathbf{x})^\top \mathbf{z} \leq 0$ for all $\mathbf{z} \in \{\mathbf{z} \in \mathbb{R}^P : \mathbf{C}_1 \mathbf{z} \preceq_{\mathcal{K}_1} \mathbf{d}_1\}$, which is true since $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}^+$. In summary, we conclude that the chance constraint in (2.2) is equivalent to (2.28) for all $\mathbf{x} \in \mathcal{X}$. Thus, the claim follows. $\square$

Theorem 2.12. *If the ambiguity set $\mathcal{P}$ satisfies $\mathcal{P} = \mathcal{P}^c \cap \mathcal{P}^s$ where $\mathcal{P}^c$ is a conic moment ambiguity set and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\mathbf{z}} \mathbb{P}$ is point symmetric around $\mathbf{m} = \mathbf{0}$, then the joint chance constrained program (2.2) is strongly NP-hard even if $\mathbf{S}(\mathbf{x}) = \mathbf{S}$.*

Finally, if the ambiguity set is restricted to contain only distributions $\mathbb{P} \in \mathcal{P}$ under which $\Pi_{\mathbf{z}} \mathbb{P}$ is α -unimodal, then the intractability results of Theorems 3.6 and 3.3 remain valid since any distribution can be approximated arbitrarily well—in the weak sense—by an α -unimodal distribution with sufficiently large $\alpha > 0$, see e.g. [90]. To our best knowledge, the complexity of the joint chance constrained program (2.2) has not been settled for $\mathbf{S}(\mathbf{x}) = \mathbf{S}$ and for $\mathcal{P}$ generated by the intersection of a conic moment ambiguity set and the structural ambiguity set $\mathcal{P}^\alpha$ of all unimodal distributions.

Chapter 3

Ambiguous Chance Constraints with Conic Dispersion Measures

3.1 Introduction

Since its inception in the 1950s [5, 30], stochastic programming has evolved into a dominant methodology for modeling and solving optimisation problems involving uncertainty. A fundamental class of stochastic programming problems are linear *chance constrained programs* of the form

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \\ & \mathbb{Q} \left(\mathbf{T}(\tilde{\boldsymbol{\xi}}) \mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}}) \right) \geq 1 - \varepsilon. \end{aligned} \tag{3.1}$$

Here, the goal is to select a vector of decision variables $\mathbf{x}$ from within the polyhedron $\mathcal{X} \subseteq \mathbb{R}^n$ such that $\mathbf{x}$ minimises the linear cost function $\mathbf{c}^\top \mathbf{x}$, $\mathbf{c} \in \mathbb{R}^n$, and satisfies a set of uncertainty-affected inequalities $\mathbf{T}(\tilde{\boldsymbol{\xi}}) \mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}})$ with probability $1 - \varepsilon$, where the tolerance $\varepsilon \in [0, 1)$ reflects the risk attitude of the decision maker. Both the left-hand side coefficient matrix $\mathbf{T}(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}^{m \times n}$ and the right-hand side vector $\mathbf{u}(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}^m$ in (3.1) depend affinely on a random vector $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^k$ that is governed by the known probability distribution $\mathbb{Q}$. Thus, we can assume that $\mathbf{T}(\tilde{\boldsymbol{\xi}})$ is defined through the row vectors $\mathbf{t}_i(\tilde{\boldsymbol{\xi}})^\top = \tilde{\boldsymbol{\xi}}^\top \mathbf{T}_i + \mathbf{s}_i^\top$, $\mathbf{T}_i \in \mathbb{R}^{k \times n}$ and $\mathbf{s}_i \in \mathbb{R}^n$, $i = 1, \dots, m$, while $\mathbf{u}(\tilde{\boldsymbol{\xi}})$ satisfies $\mathbf{u}(\tilde{\boldsymbol{\xi}}) = \mathbf{U} \tilde{\boldsymbol{\xi}} + \mathbf{v}$, $\mathbf{U} \in \mathbb{R}^{m \times k}$ and $\mathbf{v} \in \mathbb{R}^m$.

The probabilistic constraint is referred to as an *individual chance constraint* if $m = 1$ and a *joint chance constraint* if $m > 1$, respectively.

Initiated by the seminal works of Charnes et al. [23] and Charnes and Cooper [21], chance constrained programs have been employed in numerous application domains ranging from logistics [28], finance [34], project management [107] and network design [117] to emissions control [3], design optimisation [13] and call center staffing [55]. Despite their wide-spread use, chance constrained programs suffer from two major shortcomings: they require an exact specification of the distribution $\mathbb{Q}$, and they typically result in computationally challenging optimisation problems.

Similar to the parameters in a deterministic optimisation problem, which typically represent point estimates of uncertain quantities, the probability distribution $\mathbb{Q}$ in problem (3.1) often constitutes a ‘best guess’ based on historical data or expert judgment. It is well-known, however, that even in seemingly benign chance constrained problems, small estimation errors in the probability distribution can lead to large deviations in the optimal value and the optimal solution. This is particularly worrying as many practical optimisation models contain high-dimensional random vectors for which the estimation of the underlying probability distribution becomes extremely challenging. For a detailed discussion of this issue, we refer to [104] and the references therein.

Even if a sufficiently accurate description of the true probability distribution $\mathbb{Q}$ is available, problem (3.1) remains a challenging optimisation problem. In fact, despite a number of notable exceptions [95, 105], the inclusion of chance constraints typically leads to non-smooth, non-convex and possibly even discontinuous optimisation problems. Moreover, verifying whether a given solution $\mathbf{x}$ satisfies a chance constraint requires multi-dimensional integration, which itself becomes computationally burdensome for random vectors of dimension $k = 4$ and higher.

Both disadvantages of chance constrained programming are alleviated by *distributionally robust chance constrained programs*, which can be defined as follows.

$$\begin{aligned}
\min \quad & \mathbf{c}^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \\
& \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{T}(\tilde{\xi}) \mathbf{x} \leq \mathbf{u}(\tilde{\xi}) \right) \geq 1 - \varepsilon
\end{aligned} \tag{3.2}$$

In problem (3.2), any feasible solution $\mathbf{x}$ satisfies the chance constraint for every distribution $\mathbb{P}$ in the *ambiguity set* $\mathcal{P}$. The ambiguity set is chosen so as to contain all distributions that are consistent with the available *a priori* information about the true distribution $\mathbb{Q}$, such as structural characteristics (e.g. support, unimodality or symmetry) or statistical properties (e.g. confidence intervals or moment bounds). We call the chance constraint in (3.2) *distributionally robust* (or simply *robust*).

By considering all probability distributions that are deemed possible under the available information, problem (3.2) mitigates the reliance on a precise characterisation of the true distribution $\mathbb{Q}$. Since the problem hedges against the worst probability distribution within the ambiguity set, the model reflects a strict aversion towards distributional ambiguity. Ambiguity aversion enjoys strong justification from decision theory [41, 51], and it has proved instrumental to explain various real-world phenomena such as incomplete contracts, stock market volatility and voting behavior [83, 84]. Note that problem (3.2) constitutes a conservative approximation to the classical chance constrained program (3.1) whenever the true distribution $\mathbb{Q}$ is contained in the ambiguity set $\mathcal{P}$.

Perhaps surprisingly, robust chance constraints are often computationally tractable. In fact, for several classes of ambiguity sets $\mathcal{P}$, one can employ classical probability inequalities to equivalently reformulate or conservatively approximate robust individual chance constraints in a tractable way. This has first been observed by Ben-Tal and Nemirovski [10] and Bertsimas and Sim [12], who use Hoeffding’s inequality to derive tractable reformulations of robust individual chance constraints when the components of $\tilde{\xi}$ are independent, symmetric and bounded random variables. Chen et al. [26] also employ the Hoeffding inequality to approximate robust individual chance constraints where the ambiguity set captures asymmetries in the distribution of $\tilde{\xi}$ via forward and backward deviation bounds. Assuming that the components of $\tilde{\xi}$ are independent random variables whose joint distribution belongs to a given convex and compact set, Nemirovski and Shapiro [87] use large deviation-type Bernstein bounds to approximate robust individual chance constraints. Calafiore and El Ghaoui [19] employ various statistical bounds to approximate robust individual chance constraints where the ambiguity set specifies structural properties such as radial symmetry, unimodality or independence. Bertsimas et al. [11] use statistical hypothesis tests to approximate robust individual chance constraints when the ambiguity set has to be estimated from samples of the unknown distribution $\mathbb{Q}$. Xu et al. [122] employ a generalised Chebyshev inequality

ity to derive tractable reformulations of problems with probabilistic envelope constraints, which enforce robust chance constraints at all tolerance levels $\varepsilon \in [0, 1)$. Instead of relying on statistical results, one can also employ duality of moment problems [29] to derive tractable reformulations of robust individual chance constraints. This has first been proposed by El Ghaoui et al. [50] for Chebyshev ambiguity sets which specify the expected value and the covariance matrix of the random parameter vector $\tilde{\xi}$.

The distributionally robust chance constrained program (3.2) becomes considerably more challenging if $m > 1$, that is, if it involves a joint chance constraint. Nemirovski and Shapiro [87] employ Bonferroni’s inequality to conservatively approximate a robust joint chance constraint with violation probability ε by m robust individual chance constraints whose violation probabilities sum up to ε . Although this approach has a long tradition in chance constrained programming [95], Chen et al. [25] demonstrate that the quality of this approximation can decrease substantially with m if the constraint functions are positively correlated. Instead, they propose to approximate the robust joint chance constraint by a robust conditional value-at-risk (CVaR) constraint. For Chebyshev ambiguity sets, they subsequently approximate the robust CVaR constraint using a classical result in order statistics. Zymmler et al. [125] prove that the CVaR approximation in [25] becomes exact if certain scaling parameters are chosen optimally. Unfortunately, optimising simultaneously over the decisions $\mathbf{x}$ and the scaling parameters seems to be difficult. The authors also propose an exact reformulation of the emerging robust CVaR constraint using moment duality. Van Parys et al. [91] use the results of [125] to solve chance constrained finite and infinite horizon control problems. Erdoğan and Iyengar [42] study robust joint chance constraints where the ambiguity set contains all distributions that are within a certain distance of a nominal distribution (in terms of the Prohorov metric). They derive a conservative approximation by sampling from the nominal distribution and enforcing the constraints for all values of ξ that are ‘close’ to any of the samples. In a similar spirit, Hu and Hong [62] and Hu et al. [63] consider ambiguity sets that contain all distributions within a certain distance of a nominal distribution, where the distance is measured in terms of the Kullback-Leibler divergence or the likelihood ratio. They show that in their setting, the robust chance constraint reduces to a classical chance constraint with an updated violation probability. Jiang and Guan [71] and Yanıkoğlu and den Hertog [123] study data-driven versions of robust chance constraints using the ambiguity sets from [62, 125] and [7], respectively, where the ambiguity sets have to

be estimated from samples. Conservative approximations to robust chance constraints involving linear matrix inequalities, which generalise the robust joint chance constraint in (3.2), have been studied by Ben-Tal and Nemirovski [9] and Cheung et al. [27]. There is also a rich literature on distributionally robust combinatorial and mixed-integer programming, see Li et al. [74]. Here, a major goal is to calculate the probability that a binary decision variable adopts the value 1 in the optimal solution. For more detailed surveys of distributionally robust chance constrained programs, we refer to Ben-Tal et al. [6] and Nemirovski [86].

Despite intensive research efforts over the last two decades, we are not aware of any tractable and provably optimal solution scheme for robust *joint* chance constrained problems, nor do we know of any hardness results that preclude the existence of such schemes. In fact, to our best knowledge all existing contributions either propose tractable but conservative approximations, or they develop exact but non-convex reformulations which are subsequently solved to partial optimality.¹

In this chapter, we take a first step towards exploring the tractability of distributionally robust joint chance constraints. We study instances of problem (3.2) where the ambiguity set satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \mathbb{E}_{\mathbb{P}}[g_s(\tilde{\xi})] \leq d_s \quad \forall s \in \mathcal{S} \right\} \quad (3.3)$$

for a support set of the form $\Xi = \{\xi \in \mathbb{R}^k : \mathbf{W}\xi \leq \mathbf{w}\}$, $\mathbf{W} \in \mathbb{R}^{l \times k}$ and $\mathbf{w} \in \mathbb{R}^l$, and polyhedral dispersion measures $g_s(\xi) = \max_{t \in \mathcal{T}} \mathbf{g}_{st}^\top(\xi - \mu) + h_{st}$, $s \in \mathcal{S} = \{1, \dots, S\}$, $\mathcal{T} = \{1, \dots, T\}$, $\mathbf{g}_{st} \in \mathbb{R}^k$, $h_{st} \in \mathbb{R}$ and $\mathbf{d} \in \mathbb{R}_+^S$.

Whenever the ambiguity set $\mathcal{P}$ in (3.3) is not empty, it contains the Dirac distribution δ_μ that places all probability mass on μ whenever $\mathcal{P}$ is non-empty. This follows from the convexity of Ξ and Jensen's inequality, which implies that $\mathbb{E}_{\delta_\mu}[\mathbf{g}(\tilde{\xi})] = \mathbf{g}(\mu) = \mathbf{g}(\mathbb{E}_{\mathbb{P}}[\tilde{\xi}]) \leq \mathbb{E}_{\mathbb{P}}[\mathbf{g}(\tilde{\xi})]$ for all $\mathbb{P} \in \mathcal{P}$. In the remainder of the chapter, we assume that δ_μ satisfies the moment bound in (3.3) *strictly*. This non-restrictive assumption allows us to employ the strong duality results from [66, 103].

Ambiguity sets of the form (3.3) are interesting for several reasons. Firstly, the class of polyhedral dispersion measures covers several variability indicators from robust statistics that are theoretically

¹An exception are instances of (3.2) where only the right-hand side coefficients $\mathbf{u}(\tilde{\xi})$ are uncertain and the ambiguity set bounds a ϕ -divergence to a log-concave distribution. While it follows from [62, 63, 95] that those instances are *convex*, they may not be *practically tractable* as they do not possess a reformulation as a conic optimisation problem.

appealing and practically meaningful, such as the mean absolute deviation and the mean semi-deviation. Compared to classical indicators such as the variance, robust dispersion measures are less affected by outliers and deviations from the model assumptions [20] and are thus well-suited for data-driven problems with sparse historical information [33, 121]. Secondly, we will show that problems involving ambiguity sets of the form (3.3) often possess reformulations as linear or conic-quadratic programs. This is in contrast to Chebyshev or divergence-based ambiguity sets, which typically lead to considerably more challenging semidefinite programs or generic convex optimisation problems.

3.2 Worst-Case Joint Chance Constraints with Left-Hand Side Uncertainty

In this section, we investigate generic problems of the form (3.2) where the chance constraint displays *left-hand side uncertainty*, that is, at least one of the matrices T_i is non-zero and the left-hand side coefficient matrix in (3.2) thus depends non-trivially on $\tilde{\xi}$. The right-hand side vector $\mathbf{u}(\tilde{\xi})$, on the other hand, may or may not be constant in $\tilde{\xi}$. In fact, uncertainty in the coefficient matrix encapsulates uncertainty in the right-hand side vector as a special case since we are free to multiply $\mathbf{u}(\tilde{\xi})$ with a dummy decision variable x_{n+1} that is fixed to 1.

We will demonstrate that for any fixed $\mathbf{x} \in \mathcal{X}$ we can efficiently evaluate the worst-case probability $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\mathbf{T}(\tilde{\xi})\mathbf{x} \leq \mathbf{u}(\tilde{\xi}))$ and construct a sequence of extremal distributions that attain this worst case in the limit. In sharp contrast to classical chance constrained programming, we can thus efficiently check the feasibility of any fixed $\mathbf{x}$ in (3.2). We will then prove that problem (3.2) is nevertheless intractable but admits a family of conservative CVaR approximations. Maybe surprisingly, the best approximation within this family is equivalent to the original problem. As this optimal approximation is hard to find, however, we propose a sequential convex optimisation algorithm that efficiently computes high-quality but suboptimal solutions.

The results of this section can be formulated more concisely by using the notational shorthand

$$\mathcal{I}(\mathbf{x}) = \{i \in \mathcal{I} : \{\xi \in \Xi : \xi^\top T_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} > \mathbf{u}_i^\top \xi + v_i\} \neq \emptyset\}. \quad (3.4)$$

For any fixed $\mathbf{x} \in \mathcal{X}$, $\mathcal{I}(\mathbf{x})$ represents the index set of all constraints that are violated by at least one realisation $\boldsymbol{\xi} \in \Xi$. Note that $\mathcal{I}(\mathbf{x})$ can be determined efficiently by solving m tractable linear programs (LPs). The following proposition asserts that for any fixed $\mathbf{x} \in \mathcal{X}$ the chance constraint's worst-case violation probability can be computed efficiently.

Theorem 3.1 (Worst-Case Probability). *For any fixed decision $\mathbf{x} \in \mathcal{X}$, the worst-case probability $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\mathbf{T}(\tilde{\boldsymbol{\xi}})\mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}}))$ coincides with the optimal value of the following tractable LP.*

$$\begin{aligned}
\max \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \mathbf{d}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathbb{R}_+^S, \boldsymbol{\theta}_0, \boldsymbol{\theta}_i \in \mathbb{R}_+^{S \times T}, \boldsymbol{\phi}_0, \boldsymbol{\phi}_i \in \mathbb{R}_+^l, \tau_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{I}(\mathbf{x}) \\
& \alpha + \mathbf{w}^\top \boldsymbol{\phi}_0 + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{0st} (\mathbf{g}_{st}^\top \boldsymbol{\mu} - h_{st}) \leq 1 \\
& \boldsymbol{\beta} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{0st} \mathbf{g}_{st} + \mathbf{W}^\top \boldsymbol{\phi}_0 \\
& \gamma_s = \sum_{t \in \mathcal{T}} \theta_{0st}, \quad \gamma_s = \sum_{t \in \mathcal{T}} \theta_{ist} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{I}(\mathbf{x}) \\
& \alpha - \tau_i v_i + \tau_i \mathbf{s}_i^\top \mathbf{x} + \mathbf{w}^\top \boldsymbol{\phi}_i + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{ist} (\mathbf{g}_{st}^\top \boldsymbol{\mu} - h_{st}) \leq 0 \quad \forall i \in \mathcal{I}(\mathbf{x}) \\
& \boldsymbol{\beta} + \tau_i \mathbf{T}_i \mathbf{x} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i + \mathbf{W}^\top \boldsymbol{\phi}_i \quad \forall i \in \mathcal{I}(\mathbf{x})
\end{aligned} \tag{3.5}$$

Proof. See Appendix. □

Theorem 3.1 offers an efficient method to check whether a given $\mathbf{x} \in \mathcal{X}$ is feasible in (3.2).

Corollary 3.1. *The feasibility of any fixed decision $\mathbf{x} \in \mathcal{X}$ in the distributionally robust chance constrained program (3.2) can be verified in polynomial time.*

Proof. The decision $\mathbf{x} \in \mathcal{X}$ is feasible in (3.2) if and only if the optimal value of (3.5) exceeds $1 - \varepsilon$. Moreover, (3.5) constitutes a tractable LP that can be solved in polynomial time. □

Note that checking whether a given $\mathbf{x} \in \mathcal{X}$ satisfies the *classical* chance constraint in (3.1) is NP-hard even if $m = 1$ (individual chance constraint), $\mathbf{T}(\tilde{\boldsymbol{\xi}}) = \tilde{\boldsymbol{\xi}}^\top$ and $\mathbb{Q}$ is the uniform distribution on the unit box of $\mathbb{R}^k$ —three benign-looking conditions which guarantee that the feasible set of this chance constraint is actually convex [87]. Corollary 3.1 contrasts sharply with this intractability result for classical chance constraints.

In addition to evaluating the worst-case violation probability of a given decision, practical applications often require simulation runs under the distribution that attains the worst-case probability. In the following, we will thus turn our attention to the construction of such a worst-case distribution. Since the infimum of the chance constraint in problem (3.2) may not be attained by any $\mathbb{P} \in \mathcal{P}$, we will consider a family of worst-case distributions for the perturbed chance constraints $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{T}(\tilde{\xi})\mathbf{x} < \mathbf{u}(\tilde{\xi}) + \delta \mathbf{e} \right)$, $\delta > 0$. It turns out that the worst-case probabilities of these perturbed chance constraints are attained by distributions $\mathbb{P}_\delta^* \in \mathcal{P}$, and that these worst-case probabilities converge to the worst-case probability of the unperturbed chance constraint as δ approaches zero. For notational convenience, we introduce the shorthand notation $\mathcal{I}_0(\mathbf{x}) = \mathcal{I}(\mathbf{x}) \cup \{0\}$.

Theorem 3.2 (Worst-Case Probability Distribution). *For a feasible decision $\mathbf{x}$ in problem (3.2) and a precision parameter $\delta > 0$, let $(\{\lambda_i^*\}_{i \in \mathcal{I}_0(\mathbf{x})}, \{\boldsymbol{\chi}_i^*\}_{i \in \mathcal{I}_0(\mathbf{x})})$ be an optimal solution to the LP*

$$\begin{aligned}
\min \quad & \lambda_0 \\
\text{s.t.} \quad & \lambda_i \in \mathbb{R}_+, \boldsymbol{\chi}_i \in \mathbb{R}^k & \forall i \in \mathcal{I}_0(\mathbf{x}) \\
& \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i = 1, \quad \sum_{i \in \mathcal{I}_0(\mathbf{x})} \boldsymbol{\chi}_i = \boldsymbol{\mu} \\
& d_s \geq \sum_{i \in \mathcal{I}_0(\mathbf{x})} \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\chi}_i - \lambda_i \boldsymbol{\mu}) + \lambda_i h_{st} \} & \forall s \in \mathcal{S} \\
& \boldsymbol{\chi}_i^\top \mathbf{T}_i \mathbf{x} + \lambda_i \mathbf{s}_i^\top \mathbf{x} \geq \mathbf{u}_i^\top \boldsymbol{\chi}_i + \lambda_i (v_i + \delta) & \forall i \in \mathcal{I}(\mathbf{x}) \\
& \mathbf{W} \boldsymbol{\chi}_i \leq \lambda_i \mathbf{w} & \forall i \in \mathcal{I}_0(\mathbf{x}).
\end{aligned} \tag{3.6}$$

We then have that $\mathbb{P}_\delta^* \in \mathcal{P}$ for the distribution $\mathbb{P}_\delta^*$ defined through

$$\mathbb{P}_\delta^*(\tilde{\xi} = \xi) = \begin{cases} \lambda_i^* & \text{for } \xi = \boldsymbol{\chi}_i^* / \lambda_i^*, \text{ where } i \in \mathcal{I}(\mathbf{x}) \text{ such that } \lambda_i^* > 0 \\ \lambda_0^* & \text{for } \xi = (\boldsymbol{\chi}_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}): \lambda_i^* = 0} \boldsymbol{\chi}_i^*) / \lambda_0^*, \end{cases} \tag{3.7}$$

and the family of distributions $\{\mathbb{P}_\delta^*\}_{\delta > 0}$ satisfies

$$\lim_{\delta \downarrow 0} \mathbb{P}_\delta^* \left(\mathbf{T}(\tilde{\xi})\mathbf{x} \leq \mathbf{u}(\tilde{\xi}) \right) = \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{T}(\tilde{\xi})\mathbf{x} \leq \mathbf{u}(\tilde{\xi}) \right). \tag{3.8}$$

Proof. See Appendix. □

3.2.1 Intractability of Joint Chance Constraints with Left-Hand Side Uncertainty

In spite of the positive results reported in Theorems 3.1 and 3.2, we will now prove that the distributionally robust chance constrained program (3.2) is strongly NP-hard even in the absence of support constraints and even if the mean absolute deviation—arguably the simplest polyhedral dispersion measure—is used to quantify the distributions’ spread. To this end, we recall that the strongly NP-hard 0/1 Integer Programming (IP) feasibility problem is defined as follows [46].

0/1 INTEGER PROGRAMMING.
Instance. Given are $\mathbf{A} \in \mathbb{Z}^{m \times n}$ and $\mathbf{b} \in \mathbb{Z}^m$.
Question. Is there a vector $\mathbf{x} \in \{0, 1\}^n$ such that $\mathbf{Ax} \leq \mathbf{b}$?

We aim to reduce the IP feasibility problem to the following instance of the distributionally robust chance constrained program (3.2) with left-hand side uncertainty

$$\begin{aligned}
& \inf \quad 0 \\
& \text{s.t.} \quad \mathbf{x} \in [0, 1]^n \\
& \quad \mathbf{Ax} \leq \mathbf{b} \\
& \quad \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\tilde{\xi}_i x_i \geq 0, \tilde{\xi}_{n+i}(1 - x_i) \geq 0 \quad \forall i \in \mathcal{I} \right) \geq 1 - \varepsilon,
\end{aligned} \tag{3.9}$$

where $\varepsilon \in (0, 1)$, and the matrix $\mathbf{A} \in \mathbb{Z}^{m \times n}$ and the vector $\mathbf{b} \in \mathbb{Z}^m$ correspond to the respective input parameters of the IP feasibility problem. To this end, we assume that the dispersion of the distributions in $\mathcal{P}$ is controlled through an upper bound on their mean absolute deviation, i.e.,

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{\xi}} - \boldsymbol{\mu}|] \leq \mathbf{d}\}, \tag{3.10}$$

where $k = 2n$, $\Xi = \mathbb{R}^{2n}$, $\boldsymbol{\mu} = \mathbf{e} \in \mathbb{R}^{2n}$, $\mathbf{d} = 2\varepsilon\boldsymbol{\mu}/n$ and the modulus operator is understood to apply component-wise. Problem (3.9) is a feasibility problem that evaluates to 0 if it has a feasible solution and to $+\infty$ otherwise. To prove NP-hardness of the chance-constrained program (3.9) (and, *a fortiori*, problem (3.2)), we will show that problem (3.9) is feasible if and only if the corresponding IP feasibility

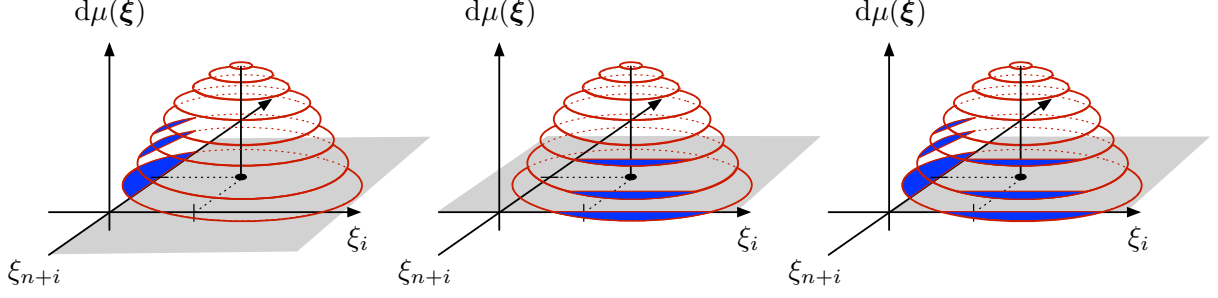


Figure 3.1: Illustration of the i -th constraint inside the probability expression in (3.9) when $x_i = 1$ (left), $x_i = 0$ (middle) and $x_i \in (0, 1)$ (right). The light shaded region denotes the realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$ for which the constraint is satisfied. The contour lines represent the projected density function of a probability distribution $\mathbb{P} \in \mathcal{P}$ which violates the i -th constraint for the dark shaded realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$.

problem has an affirmative answer. To this end, we first derive an analytic expression for the worst-case probability of the chance constraint in (3.9).

Lemma 3.1. *For any fixed $\mathbf{x} \in [0, 1]^n$ the worst-case probability problem on the left-hand side of the chance constraint in (3.9) has the following analytic solution:*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\tilde{\xi}_i x_i \geq 0, \tilde{\xi}_{n+i}(1 - x_i) \geq 0 \quad \forall i \in \mathcal{I} \right) = \max \left\{ 0, 1 - \varepsilon \left(2 - \frac{|\{i \in \mathcal{I} : x_i \in \{0, 1\}\}|}{n} \right) \right\}. \quad (3.11)$$

Proof. See Appendix. □

Lemma 3.1 implies that the worst-case probability of the chance constraint in problem (3.9) is greater than or equal to $1 - \varepsilon$ if and only if $\mathbf{x} \in \{0, 1\}^n$. We explain the intuition behind Lemma 3.1 for the special case where $\varepsilon < 1/2$. By construction of the ambiguity set (3.10), the distributions $\mathbb{P} \in \mathcal{P}$ can place a probability of at most ε/n on parameter realisations of $\tilde{\boldsymbol{\xi}}$ where the i -th component $\tilde{\xi}_i$ is strictly negative, $i = 1, \dots, 2n$. Figure 3.1 illustrates that the i -th constraint inside the probability expression in (3.9) requires $\tilde{\xi}_i$ or $\tilde{\xi}_{n+i}$ to be non-negative whenever $x_i = 1$ or $x_i = 0$, respectively. Hence, for $\mathbf{x} \in \{0, 1\}^n$ each constraint inside the probability expression can be violated with a probability of at most ε/n , that is, the overall violation probability of the chance constraint in (3.9) is bounded above by ε . If $x_i \in (0, 1)$ for some $i = 1, \dots, n$, on the other hand, then the i -th constraint inside the probability expression requires both $\tilde{\xi}_i$ and $\tilde{\xi}_{n+i}$ to be non-negative. In that case, there are distributions $\mathbb{P} \in \mathcal{P}$ that place a probability of $2\varepsilon/n$ on realisations of $\tilde{\boldsymbol{\xi}}$ where either $\tilde{\xi}_i < 0$ or $\tilde{\xi}_{n+i} < 0$. Thus, even if all other components of $\mathbf{x}$ are binary, there are distributions $\mathbb{P} \in \mathcal{P}$ for which the chance constraint is violated

with probability $2\varepsilon/n + (n-1)\varepsilon/n = (n+1)\varepsilon/n > \varepsilon$.

The insight from Lemma 3.1 enables us to prove the postulated intractability result.

Theorem 3.3. *The chance-constrained program (3.2) with the ambiguity set (3.3) is strongly NP-hard for any fixed $\varepsilon \in (0, 1)$ even if $\Xi = \mathbb{R}^{2k}$ and even if the mean absolute deviation (that is, the simplest conceivable polyhedral dispersion measure) is used in the specification of the ambiguity set $\mathcal{P}$.*

Proof. We show that problem (3.9) is feasible if and only if the IP feasibility problem is. To this end, suppose first that there is $\mathbf{x} \in \{0, 1\}^n$ with $\mathbf{Ax} \leq \mathbf{b}$. Lemma 3.1 then implies that $\mathbf{x}$ satisfies the chance constraint in (3.9), which in turn implies that (3.9) is feasible. Assume to the contrary that (3.9) has a feasible solution $\mathbf{x}$. Since $\mathbf{x}$ satisfies the chance constraint in (3.9), Lemma 3.1 implies that $\mathbf{x}$ is binary. As $\mathbf{x}$ also satisfies $\mathbf{Ax} \leq \mathbf{b}$, we conclude that $\mathbf{x}$ solves the IP feasibility problem. $\square$

3.2.2 Approximations based on the Conditional Value-at-Risk

Theorem 3.3 motivates us to seek approximations for the chance constraint in (3.2) that are (i) *computationally tractable* and (ii) *conservative* or *safe* in the sense that any $\mathbf{x}$ satisfying the approximate constraint should also be feasible in the original chance constraint. Approximations with these properties can be obtained by leveraging the rich theory of financial risk measures. Two risk measures of particular relevance for the further development are the *value-at-risk* (VAR) and the *conditional value-at-risk* (CVaR). For any real-valued measurable function $L(\tilde{\xi})$ representing some form of ‘loss’, probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^k)$ and violation probability $\varepsilon \in (0, 1)$, the VaR of the random loss $L(\tilde{\xi})$ at level ε with respect to $\mathbb{P}$ is defined as

$$\mathbb{P}\text{-VaR}_\varepsilon \left(L(\tilde{\xi}) \right) = \inf_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathbb{P} \left(L(\tilde{\xi}) \leq \gamma \right) \geq 1 - \varepsilon \right\}.$$

Note that the VaR coincides with the left $(1 - \varepsilon)$ -quantile of the loss distribution and is generically non-convex in $L(\tilde{\xi})$. Moreover, the CVaR of $L(\tilde{\xi})$ at level ε with respect to $\mathbb{P}$ is defined as

$$\mathbb{P}\text{-CVaR}_\varepsilon \left(L(\tilde{\xi}) \right) = \inf_{\kappa \in \mathbb{R}} \left\{ \kappa + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}} \left((L(\tilde{\xi}) - \kappa)^+ \right) \right\}.$$

One can show that the CVaR is convex and non-decreasing in $L(\tilde{\xi})$ and that $\kappa^* = \mathbb{P}\text{-VaR}_\varepsilon(L(\tilde{\xi}))$ solves the minimisation problem in its definition. If the distribution of $L(\tilde{\xi})$ is absolutely continuous, one can further show that the CVaR coincides with the conditional expectation of the losses exceeding the VaR. For a more detailed discussion of these risk measures we refer to [99].

Due to its convexity, the CVaR lends itself to constructing convex conservative approximations for classical as well as distributionally robust chance constraints. This construction relies on a parametric family of loss functions of the form

$$L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) = \max_{i \in \mathcal{I}} \left\{ \tau_i \left(\mathbf{t}_i(\tilde{\xi})^\top \mathbf{x} - u_i(\tilde{\xi}) \right) \right\},$$

where $\boldsymbol{\tau} \in \Delta_{++}^m = \{\boldsymbol{\tau}' \in \mathbb{R}^m : \boldsymbol{\tau}' > \mathbf{0}, \mathbf{e}^\top \boldsymbol{\tau}' = 1\}$ represents a vector of scaling factors belonging to the relative interior of the probability simplex in $\mathbb{R}^m$. The following observation is then an immediate consequence of the elementary properties of VaR and CVaR.

Observation 3.1. *For any $\mathbf{x} \in \mathcal{X}$, $\boldsymbol{\tau} \in \Delta_{++}^m$ and $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^k)$ we have:*

$$\begin{aligned} (i) \quad & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{T}(\tilde{\xi})\mathbf{x} \leq \mathbf{u}(\tilde{\xi}) \right) \geq 1 - \varepsilon \iff \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \right) \leq 0, \\ (ii) \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \right) \leq 0 \implies \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \right) \leq 0. \end{aligned}$$

Proof. By construction of the loss functions, we have $\mathbf{T}(\tilde{\xi})\mathbf{x} \leq \mathbf{u}(\tilde{\xi})$ if and only if $L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \leq 0$. Assertion (i) thus follows from the proof of Theorem 2.2 in [125], and assertion (ii) holds because

$$\mathbb{P}\text{-CVaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \right) \geq \mathbb{P}\text{-VaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}) \right)$$

for every fixed $\mathbb{P} \in \mathcal{P}$, see [98, Corollary 7]. □

Observation 3.1 implies that the distributionally robust chance constraint in problem (3.2) admits an *equivalent reformulation* in terms of a *worst-case VaR constraint* on the parametric loss function. This constraint requires that the VaR at level ε of $L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi})$ be bounded above by 0 for *all* distributions $\mathbb{P} \in \mathcal{P}$. Observation 3.1 further implies that the distributionally robust chance constraint admits a *conservative approximation* in terms of a *worst-case CVaR constraint*. Note that $\mathbb{P}\text{-CVaR}_\varepsilon(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\xi}))$

is convex in $\mathbf{x}$ for any fixed $\mathbb{P} \in \mathcal{P}$ since the loss function constitutes a maximum of affine functions in $\mathbf{x}$, while the CVaR constitutes a convex and non-decreasing risk measure. As convexity is preserved under maximisation over $\mathbb{P} \in \mathcal{P}$, the worst-case CVaR is in fact convex in $\mathbf{x}$. Thus, the worst-case CVaR constraint has a convex feasible set.

We emphasise that the sublevel sets $\{\boldsymbol{\xi} \in \mathbb{R}^k : L(\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\xi}) \leq 0\}$ are equal across all $\boldsymbol{\tau} \in \Delta_{++}^m$, and thus the choice of the scaling factors has no impact on the worst-case VaR constraint. In contrast, different values of $\boldsymbol{\tau} \in \Delta_{++}^m$ generically result in different worst-case CVaR constraints, all of which provide convex conservative approximations for the distributionally robust chance constraint. This scaling freedom is valuable as it allows us to control the approximation quality. If $\boldsymbol{\tau} \in \Delta_{++}^m$ is treated as a decision variable, we obtain the best possible CVaR approximation:

$$\exists \boldsymbol{\tau} \in \Delta_{++}^m : \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) \right) \leq 0 \quad (3.12)$$

By Observation 3.1, the *optimally scaled* CVaR constraint (3.12) still constitutes a conservative approximation for the distributionally robust chance constraint. The following proposition asserts that (3.12) is equivalent to a finite system of linear and bilinear constraints.

Proposition 3.1. *If the ambiguity set $\mathcal{P}$ is defined as in (3.3), then (3.12) holds if and only if there exist $\alpha, \kappa \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{\gamma} \in \mathbb{R}_+^S$, $\boldsymbol{\tau} \in \Delta_{++}^m$, $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_i \in \mathbb{R}_+^{S \times T}$ and $\boldsymbol{\phi}_0, \boldsymbol{\phi}_i \in \mathbb{R}_+^l, i \in \mathcal{I}$, with*

$$\begin{aligned} \kappa + \frac{1}{\varepsilon}(\alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \mathbf{d}^\top \boldsymbol{\gamma}) &\leq 0 \\ \alpha - \mathbf{w}^\top \boldsymbol{\phi}_0 - \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} (\mathbf{g}_{st}^\top \boldsymbol{\mu} - h_{st}) &\geq 0 \\ \boldsymbol{\beta} + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} \mathbf{g}_{st} + \mathbf{W}^\top \boldsymbol{\phi}_0 &= \mathbf{0} \\ \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{0st}, \quad \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{ist} &\quad \forall i \in \mathcal{I}, \forall s \in \mathcal{S} \\ \alpha + \kappa + \tau_i v_i - \tau_i \mathbf{s}_i^\top \mathbf{x} - \mathbf{w}^\top \boldsymbol{\phi}_i - \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} (\mathbf{g}_{st}^\top \boldsymbol{\mu} - h_{st}) &\geq 0 \quad \forall i \in \mathcal{I} \\ \boldsymbol{\beta} + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i + \mathbf{W}^\top \boldsymbol{\phi}_i = \tau_i \mathbf{T}_i \mathbf{x} &\quad \forall i \in \mathcal{I}. \end{aligned} \quad (3.13)$$

Proof. For any fixed $\boldsymbol{\tau} \in \Delta_{++}^m$ we have

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\varepsilon \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) \right) &= \sup_{\mathbb{P} \in \mathcal{P}} \inf_{\kappa \in \mathbb{R}} \kappa + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}} \left[\left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) - \kappa \right)^+ \right] \\ &= \inf_{\kappa \in \mathbb{R}} \kappa + \frac{1}{\varepsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) - \kappa \right)^+ \right], \end{aligned} \quad (3.14)$$

where the first equality is based on the definition of the CVaR, and the second equality follows from a generalised saddle point theorem [106]. The worst-case expectation in the second line of the above expression corresponds to the optimal value of the moment problem

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) - \kappa \right)^+ \right] &= \sup \int \left(L(\mathbf{x}, \boldsymbol{\tau}, \tilde{\boldsymbol{\xi}}) - \kappa \right)^+ d\mu(\boldsymbol{\xi}) \\ \text{s.t. } &\mu \in \mathcal{M}_+(\Xi), \kappa \in \mathbb{R} \\ &\int d\mu(\boldsymbol{\xi}) = 1 \\ &\int \boldsymbol{\xi} d\mu(\boldsymbol{\xi}) = \boldsymbol{\mu} \\ &\int \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} d\mu(\boldsymbol{\xi}) \leq d_s \quad \forall s \in \mathcal{S}. \end{aligned}$$

Dualising this moment problem results in the following semi-infinite program.

$$\begin{aligned} \inf \quad &\alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \mathbf{d}^\top \boldsymbol{\gamma} \\ \text{s.t. } \quad &\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathbb{R}_+^S \\ &\alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} + \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} \geq (L(\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\xi}) - \kappa)^+ \quad \forall \boldsymbol{\xi} \in \Xi \end{aligned}$$

Strong duality holds as the Dirac distribution $\delta_{\boldsymbol{\mu}}$ is strictly feasible in the primal problem [103]. Substituting the dual problem into (3.14) and decomposing the semi-infinite constraint into $m + 1$ simpler semi-infinite constraints, one for each linear piece of the ‘max’-function $(L(\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\xi}) - \kappa)^+$, shows that (3.12) is satisfied if and only if there exists $\alpha, \kappa \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{\gamma} \in \mathbb{R}_+^S$ and $\boldsymbol{\tau} \in \Delta_{++}^m$ with

$$\begin{aligned} \kappa + \frac{1}{\varepsilon} (\alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \mathbf{d}^\top \boldsymbol{\gamma}) &\leq 0 \\ \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} + \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\geq 0 \quad \forall \boldsymbol{\xi} \in \Xi \\ \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} + \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\geq \tau_i(t_i(\boldsymbol{\xi})^\top \mathbf{x} - u_i(\boldsymbol{\xi})) - \kappa \quad \forall \boldsymbol{\xi} \in \Xi, \forall i \in \mathcal{I}. \end{aligned} \quad (3.15)$$

By using an epigraph formulation for the ‘max’-functions, which is applicable because $\gamma \geq \mathbf{0}$, the first semi-infinite constraint in (3.15) can be expressed as

$$\begin{aligned} 0 &\leq \min \quad \alpha + \beta^\top \xi + \gamma^\top \rho \\ \text{s.t.} \quad &\xi \in \mathbb{R}^k, \rho \in \mathbb{R}^S \\ &\mathbf{W}\xi \leq \mathbf{w}, \rho_s \geq \mathbf{g}_{st}^\top (\xi - \mu) + h_{st} \quad \forall s \in \mathcal{S}, \forall t \in \mathcal{T}. \end{aligned} \quad (3.16)$$

Note that the LP in (3.16) is feasible because the support set Ξ is non-empty. Strong LP duality thus implies that (3.16) is equivalent to the system of linear constraints

$$\exists \Theta_0 \in \mathbb{R}_+^{S \times T}, \phi_0 \in \mathbb{R}_+^l : \begin{cases} \alpha - \mathbf{w}^\top \phi_0 - \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} (\mathbf{g}_{st}^\top \mu - h_{st}) \geq 0 \\ \beta + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} \mathbf{g}_{st} + \mathbf{W}^\top \phi_0 = \mathbf{0}, \quad \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{0st} \quad \forall s \in \mathcal{S}. \end{cases} \quad (3.17)$$

Employing an analogous argument that requires no new ideas and recalling the definitions $\mathbf{t}_i(\tilde{\xi})^\top = \tilde{\xi}^\top \mathbf{T}_i + \mathbf{s}_i^\top$ and $\mathbf{u}(\tilde{\xi}) = \mathbf{U}\tilde{\xi} + \mathbf{v}$ allows us to reformulate the second semi-infinite constraint in (3.15) as

$$\exists \Theta_i \in \mathbb{R}_+^{S \times T}, \phi_i \in \mathbb{R}_+^l : \begin{cases} \alpha + \kappa + \tau_i v_i - \tau_i \mathbf{s}_i^\top \mathbf{x} - \mathbf{w}^\top \phi_i - \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} (\mathbf{g}_{st}^\top \mu - h_{st}) \geq 0 \\ \beta + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i + \mathbf{W}^\top \phi_i = \tau_i \mathbf{T}_i \mathbf{x} \\ \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{ist} \quad \forall s \in \mathcal{S}. \end{cases} \quad (3.18)$$

The claim then follows by substituting (3.17) and (3.18) into (3.15). $\square$

The constraints (3.13) are bilinear in the original decision variables $\mathbf{x}$, the scaling factors τ and the auxiliary variables introduced through dual representations of semi-infinite constraints. Moreover, (3.13) reduces to a simple linear constraint system if either $\mathbf{x}$ or τ is kept fixed. In the following we will show that the optimally scaled CVaR constraint (3.12) is typically (in a sense that will be made precise in the next theorem) equivalent to the distributionally robust chance constraint in (3.2), that is, the CVaR approximation is typically *exact* if we can optimise over τ .

Theorem 3.4. *If $\mathcal{I}(\mathbf{x}) = \mathcal{I}$, then the optimally scaled chance constraint (3.12) is equivalent to the distributionally robust chance constraint in (3.2).*

Proof. From Observation 3.1 we know that (3.12) is sufficient for the chance constraint in (3.2) to hold. In order to show necessity when $\mathcal{I}(\mathbf{x}) = \mathcal{I}$, we recall from the proof of Theorem 3.1 that the chance constraint in (3.2) holds if and only if there exist $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^k$ and $\gamma \in \mathbb{R}_+^S$ with

$$\begin{aligned} \alpha + \mu^\top \beta - d^\top \gamma &\geq 1 - \varepsilon \\ \alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\} &\leq 1 \quad \forall \xi \in \Xi \\ \alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\} &\leq 0 \quad \forall i \in \mathcal{I}, \forall \xi \in \Xi : t_i(\xi)^\top \mathbf{x} - u_i(\xi) > 0. \end{aligned} \quad (3.19)$$

As the closure of the set $\{\xi \in \Xi : t_i(\xi)^\top \mathbf{x} - u_i(\xi) > 0\}$ is given by $\{\xi \in \Xi : t_i(\xi)^\top \mathbf{x} - u_i(\xi) \geq 0\}$ for all $i \in \mathcal{I}(\mathbf{x}) = \mathcal{I}$, the strict inequality in the last line of (3.19) can be replaced by a weak equality via a standard continuity argument. Thus, we may use a non-linear Farkas' Lemma (see, e.g., [97, Theorem 21.2]) to reformulate the i -th semi-infinite constraint in the last line of (3.19) as

$$\exists \tau_i \in \mathbb{R}_+ : \alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\} + \tau_i(t_i(\xi)^\top \mathbf{x} - u_i(\xi)) \leq 0 \quad \forall \xi \in \Xi.$$

If $\tau_i = 0$, then the function $\alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\}$ is non-positive over Ξ , which implies that

$$\alpha + \mu^\top \beta - d^\top \gamma \leq \mathbb{E}_{\mathbb{P}} \left[\alpha + \tilde{\xi}^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\tilde{\xi} - \mu) + h_{st}\} \right] \leq 0 < 1 - \varepsilon$$

for any distribution $\mathbb{P} \in \mathcal{P}$. As the resulting inequality is in conflict with the first constraint in (3.19), we conclude that τ_i must in fact be strictly positive. Thus, the chance constraint in (3.2) holds if and only if there exist $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^k$, $\gamma \in \mathbb{R}_+^S$ and $\tau \in \mathbb{R}_{++}^m$ with

$$\begin{aligned} \alpha + \mu^\top \beta - d^\top \gamma &\geq 1 - \varepsilon \\ \alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\} &\leq 1 \quad \forall \xi \in \Xi \\ \alpha + \xi^\top \beta - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{g_{st}^\top(\xi - \mu) + h_{st}\} + \tau_i(t_i(\xi)^\top \mathbf{x} - u_i(\xi)) &\leq 0 \quad \forall \xi \in \Xi, \forall i \in \mathcal{I}. \end{aligned} \quad (3.20)$$

However, any solution $(\alpha^*, \beta^*, \gamma^*, \tau^*)$ that is feasible in (3.20) gives rise to a solution

$$(\alpha, \kappa, \beta, \gamma, \tau) = \frac{1}{\mathbf{e}^\top \tau^*} (1 - \alpha^*, -1, -\beta^*, \gamma^*, \tau^*)$$

that is feasible in (3.15). Thus the claim follows. $\square$

The overall chance constrained problem becomes

$$\begin{aligned} & \text{minimise} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}, \tau \in \Delta_{++}^m \\ & && \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\varepsilon \left(L(\mathbf{x}, \tau, \tilde{\xi}) \right) \leq 0 \end{aligned} \tag{3.21}$$

where we apply Proposition (3.1) to reformulate the worst-case CVaR constraint. Due to the bilinearity between $\mathbf{x}$ and τ we propose an alternating convex optimisation scheme. Starting with a feasible solution for all decision variables, we fix τ and optimise over $\mathbf{x}$. Then we alternate, fixing $\mathbf{x}$ and optimising over τ until convergence. As mentioned previously, although there are not guarantees on even local optimality, we find that this procedure results in good quality solutions.

3.2.3 Approximations based on Sequential Convex Optimisation Algorithm

We have seen in Chapter 2 that while the uncertainty quantification problems can be solved efficiently for a broad range of ambiguity sets, the associated chance constrained programs frequently become intractable. In this section, we therefore report a heuristic for chance constrained programs that determines ‘good’ but in general suboptimal decisions even if the associated instance is intractable. The key idea is to decompose problem into an *uncertainty quantification problem* that evaluates the worst-case probability of the chance constraint in for a fixed decision $\mathbf{x}$ and a *policy improvement problem* that aims to improve the current decision $\mathbf{x}$.

To this end, we introduce a unified notation for uncertainty quantification problems. We denote the objective function of the unified uncertainty quantification problem by $Q(\psi)$, where we combine all decision variables to a single vector ψ . Likewise, we represent the constraints of the unified uncertainty quantification problem as $\psi \in \mathcal{Q}(\mathbf{x})$, where we replace the coefficient matrix $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_J)^\top$ with

Ambiguity set	Coupled decisions χ	Uncoupled decisions ρ
Moment information (Theorem 2.3)	$(\{\tau_{ij}\}_{ij})$	$(\gamma, \lambda, \kappa, \{\phi_i\}_i, \{\psi_{ij}\}_{ij})$
Moments + symmetry (Theorem 2.4)	$(\{\chi_{ij}^+\}_{ij}, \{\chi_{ij}^-\}_{ij})$	$(\gamma, \lambda, \kappa, \{\psi_{ij}^+\}_{ij}, \{\psi_{ij}^-\}_{ij})$
Moments + unimodality (Example 2.10)	$(\{\tau_j\}_j)$	(β, γ, Γ)
Moments + unimodality (Example 2.11)	$(\{\tau_j\}_j)$	(β, θ, η)

Table 3.1: Definitions of χ and ρ for some of the ambiguity sets studied in this chapter.

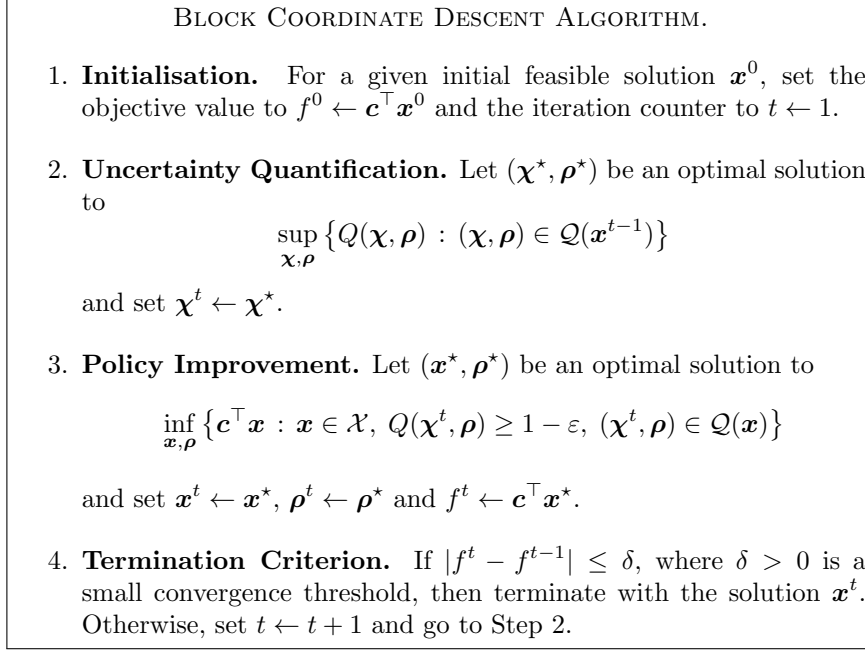


Figure 3.2: Block coordinate descent algorithm for problem (3.22).

$\mathbf{S}(\mathbf{x}) = (\mathbf{s}_1(\mathbf{x}), \dots, \mathbf{s}_J(\mathbf{x}))^\top$ and the right-hand side vector $\mathbf{t}$ with $\mathbf{t}(\mathbf{x}) = (t_1(\mathbf{x}), \dots, t_J(\mathbf{x}))^\top$. Thus, a decision $\mathbf{x} \in \mathcal{X}$ is feasible in the chance constrained program if and only if there is $\psi \in \mathcal{Q}(\mathbf{x})$ such that $Q(\psi) \geq 1 - \varepsilon$. This chance constrained program is non-convex in general as the constraint system $\psi \in \mathcal{Q}(\mathbf{x})$ may involve bilinear couplings between $\mathbf{x}$ and some components of ψ . We thus decompose ψ into a subvector χ that contains all variables which are coupled with $\mathbf{x}$ and the subvector ρ that contains the remaining variables. The chance constrained program can then be formulated as follows.

$$\begin{aligned}
& \text{minimise} && \mathbf{c}^\top \mathbf{x} \\
& \text{subject to} && \mathbf{x} \in \mathcal{X}, (\chi, \rho) \in \mathcal{Q}(\mathbf{x}) \\
& && Q(\chi, \rho) \geq 1 - \varepsilon
\end{aligned} \tag{3.22}$$

Table 3.1 exemplifies the definitions of χ and ρ for some popular ambiguity sets, and Fig. 3.2 presents

a block coordinate descent algorithm for solving (3.22). The algorithm requires a feasible point $\mathbf{x}^0$ as input. Note that the optimisation problems solved in Steps 2 and 3 of the algorithm are convex and can thus be solved efficiently. For any threshold $\delta > 0$, the algorithm terminates after finitely many iterations to a partial optimum of the chance constrained program (3.22), that is, a feasible point $(\mathbf{x}^*, \boldsymbol{\chi}^*, \boldsymbol{\rho}^*)$ where $(\mathbf{x}^*, \boldsymbol{\rho}^*)$ maximises (3.22) for fixed $\boldsymbol{\chi}^*$ and $Q(\boldsymbol{\chi}^*, \boldsymbol{\rho}^*)$ represents the worst-case probability that the system is safe under the fixed decision $\mathbf{x}^*$. For a convergence proof we refer to [73].

3.3 Worst-Case Joint Chance Constraints with Right-Hand Side Uncertainty

We now study the special case of problem (3.2) where only the right-hand side vector $\mathbf{u}(\tilde{\boldsymbol{\xi}})$ is affected by uncertainty, whereas the left-hand side coefficient matrix $\mathbf{T}(\tilde{\boldsymbol{\xi}})$ does not depend on $\tilde{\boldsymbol{\xi}}$. In other words, we assume in this section that $\mathbf{T}_i = \mathbf{0}$ for all $i \in \mathcal{I}$ and hence $\mathbf{T}(\tilde{\boldsymbol{\xi}}) = \mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_m]^\top$, that is, the chance constraint in (3.2) does not contain any bilinear coupling between $\tilde{\boldsymbol{\xi}}$ and $\mathbf{x}$. In this case, we say that the robust chance constraint in (3.2) exhibits *right-hand side uncertainty*.

We first derive a tractable reformulation for robust chance constraints with right-hand side uncertainty under the additional assumption that the ambiguity set $\mathcal{P}$ satisfies $\Xi = \mathbb{R}^k$ (no support) and $\mathbf{g}(\tilde{\boldsymbol{\xi}})$ is a *conic dispersion measure*, that is, the intercepts h_{st} vanish for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$. Examples of conic dispersion measures are the mean absolute deviation and the mean semi-deviation. Counterexamples are the variance and the Huber loss function, which can be interpreted as limiting cases of polyhedral dispersion measures with infinitely many hyperplanes $t \in \mathcal{T}$, but which do not admit representations as conic dispersion measures. Rather surprisingly, our reformulation for the robust chance constraint takes the form of a conic-quadratic optimisation problem, despite the polyhedral nature of the input data. To our best knowledge, we derive the first exact reformulation of a robust joint chance constraint as a tractable conic optimisation problem.

We close the section by showing that our tractability result is sharp in the following sense. If the ambiguity set $\mathcal{P}$ includes support information or polyhedral dispersion measures where not all of the intercepts h_{st} vanish, then the robust chance constrained program with right-hand side uncertainty

may become strongly NP-hard. In those cases, we can use the previously described sequential convex optimisation scheme to obtain high-quality but in general suboptimal solutions.

3.3.1 Tractable Reformulation for Conic Dispersion Measures

We now assume that the robust chance constrained program (3.2) displays right-hand side uncertainty, the ambiguity set does not specify any support information, that is, $\Xi = \mathbb{R}^k$, and $\mathbf{g}(\tilde{\xi})$ is a conic dispersion measure, that is, $h_{st} = 0$ for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$. To simplify the exposition, we homogenise problem (3.2) by replacing $\tilde{\xi}$ with a zero-mean random vector $\tilde{\zeta} = \tilde{\xi} - \boldsymbol{\mu}$ and introducing the auxiliary variables $\mathbf{y} = \mathbf{S}\mathbf{x} - \mathbf{U}\boldsymbol{\mu} - \mathbf{v}$. The robust chance constrained program then becomes

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_-^m, \mathbf{y} = \mathbf{S}\mathbf{x} - \mathbf{U}\boldsymbol{\mu} - \mathbf{v} \\ & \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}(\mathbf{U}\tilde{\zeta} \geq \mathbf{y}) \geq 1 - \varepsilon, \end{aligned} \tag{3.23}$$

where the ambiguity set is defined as

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^k) : \mathbb{E}_{\mathbb{P}}[\tilde{\zeta}] = \mathbf{0}, \mathbb{E}_{\mathbb{P}}[g_s(\tilde{\zeta})] \leq d_s \quad \forall s \in \mathcal{S} \right\} \tag{3.24}$$

and the dispersion measure satisfies $g_s(\zeta) = \max_{t \in \mathcal{T}} \mathbf{g}_{st}^\top \zeta$. In the following, we denote by $\mathbf{u}_i^\top$ the i -th row of the matrix $\mathbf{U}$. Without loss of generality, we can assume that $\mathbf{u}_i \neq \mathbf{0}$ for all $i \in \mathcal{I}$, for otherwise the i -th row of the chance constraint is deterministic and can be absorbed in the definition of the set $\mathcal{X}$. Similarly, the chance constraint in (3.23) is violated for the Dirac distribution $\delta_{\mathbf{0}} \in \mathcal{P}'$ whenever $\mathbf{y} \not\leq \mathbf{0}$. Without loss of generality, can thus impose the requirement that $\mathbf{y} \in \mathbb{R}_-^m$ in problem (3.23). Theorem 3.1 then implies that for a fixed $\mathbf{y} \in \mathbb{R}_-^m$, the worst-case probability of the chance constraint

in (3.23) coincides with the optimal value of the following LP.

$$\begin{aligned}
\max \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \mathbf{d}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathbb{R}_+^S, \boldsymbol{\Theta}_0, \boldsymbol{\Theta}_i \in \mathbb{R}_+^{S \times T}, \tau_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{I}(\mathbf{y}) \\
& \alpha \leq 1, \quad \alpha + \tau_i y_i \leq 0 \quad \forall i \in \mathcal{I}(\mathbf{y}) \\
& \boldsymbol{\beta} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} \mathbf{g}_{st}, \quad \boldsymbol{\beta} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i \quad \forall i \in \mathcal{I}(\mathbf{y}) \\
& \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{0st}, \quad \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{ist} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{I}(\mathbf{y})
\end{aligned} \tag{3.25}$$

Note that we cannot simply replace the worst-case probability in (3.23) with (3.25) and jointly optimise over $(\mathbf{x}, \mathbf{y})$ and $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Theta}_0, \{\boldsymbol{\Theta}_i\}_i, \{\tau_i\}_i)$ for two reasons. Firstly, the second constraint in (3.25) appears to be non-convex as it involves a product of the two decision variables τ_i and y_i . Secondly, the constraints in (3.25) are parameterised in the sets $\mathcal{I}(\mathbf{y})$ which themselves depend on the decision vector $\mathbf{y}$. In the following, we derive a reformulation of problem (3.25) that avoids both problems.

We first argue that we can fix $\alpha = 1$ in problem (3.25). To this end, denote by (3.25') the variant of (3.25) where we strengthen the inequality $\alpha \leq 1$ to an equality. Clearly, the optimal value of (3.25') is bounded above by the optimal value of (3.25). We want to show that both optimal values coincide whenever the optimal value of (3.25) is greater than or equal to $1 - \varepsilon$, that is, whenever $\mathbf{y}$ is feasible in (3.23). Note that the objective function of the dual of problem (3.25) only contains the dual variable corresponding to the inequality $\alpha \leq 1$ in (3.25) because no other constraint in (3.25) has a non-zero right-hand side constant. Strong LP duality implies that the objective values of problem (3.25) and its dual coincide. Since both values are greater than or equal to $1 - \varepsilon$ and thus strictly positive, we conclude that the dual variable corresponding to the inequality $\alpha \leq 1$ in (3.25) is strictly positive. A standard complementary slackness argument then implies that the inequality $\alpha \leq 1$ in (3.25) is binding. Hence, the optimal values of (3.25) and (3.25') indeed coincide whenever the optimal value of (3.25) is greater than or equal to $1 - \varepsilon$, which allows us to fix $\alpha = 1$ in (3.25). We can use this insight to reformulate the seemingly non-convex second constraint in (3.25) as a conic-quadratic constraint [1]:

$$1 + \tau_i y_i \leq 0 \quad \Longleftrightarrow \quad \sqrt{4 + (\tau_i + y_i)^2} \leq \tau_i - y_i \tag{3.26}$$

We now argue that $\mathcal{I}(\mathbf{y}) = \mathcal{I}$ in problem (3.25) for all $\mathbf{y} \in \mathbb{R}_-^m$. Indeed, under the assumptions of this section, the definition of the set $\mathcal{I}(\mathbf{y})$ in (3.4) simplifies to

$$\mathcal{I}(\mathbf{y}) = \{i \in \mathcal{I} : \{\boldsymbol{\xi} \in \mathbb{R}^k : \mathbf{s}_i^\top \mathbf{y} > \mathbf{u}_i^\top \boldsymbol{\xi}\} \neq \emptyset\}.$$

Since $\mathbf{u}_i \neq \mathbf{0}$, there is always $\boldsymbol{\xi} \in \mathbb{R}^k$ such that $\mathbf{u}_i^\top \boldsymbol{\xi} < \mathbf{s}_i^\top \mathbf{y}$, that is, we have $\mathcal{I}(\mathbf{y}) = \mathcal{I}$ for any $\mathbf{y} \in \mathbb{R}_-^m$.

We can thus replace the worst-case probability in (3.23) with (3.25), fix $\alpha = 1$, rewrite the second constraint in (3.25) according to (3.26) and replace $\mathcal{I}(\mathbf{y})$ with $\mathcal{I}$ to arrive at the following result.

Theorem 3.5. *If the robust chance constrained program (3.2) displays right-hand side uncertainty, the ambiguity set does not specify any support information, $\mathbf{g}(\tilde{\boldsymbol{\xi}})$ is a conic dispersion measure and $\mathbf{u}_i \neq \mathbf{0}$ for all $i \in \mathcal{I}$, then (3.2) is equivalent to the tractable conic-quadratic program*

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_-^m, \mathbf{y} = \mathbf{S}\mathbf{x} - \mathbf{U}\boldsymbol{\mu} - \mathbf{v} \\ & \boldsymbol{\gamma} \in \mathbb{R}_+^S, \boldsymbol{\Theta}_0, \boldsymbol{\Theta}_i \in \mathbb{R}_+^{S \times T}, \tau_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{I} \\ & \mathbf{d}^\top \boldsymbol{\gamma} - \boldsymbol{\mu}^\top \boldsymbol{\beta} \leq \varepsilon \\ & \sqrt{4 + (\tau_i + y_i)^2} \leq \tau_i - y_i \quad \forall i \in \mathcal{I} \\ & \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{0st} \mathbf{g}_{st} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i \quad \forall i \in \mathcal{I} \\ & \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{0st}, \quad \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{ist} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{I}. \end{aligned}$$

Proof. The proof is given within the text of the current section, §3.3.1. □

We end this section with some specific, widely-used, ambiguity sets which we are able to reformulate exactly. Chance constraints with fixed technology matrices arise naturally, for example, in inventory control, cash matching, unit commitment or airline revenue management problems etc. We illustrate the tractable reformulation offered by Theorem 3.5 with two examples.

Example 3.1 (Mean-Absolute Deviation). *Theorem 3.5 allows us to derive a tractable reformulation of*

joint chance constraints over mean-absolute deviation ambiguity sets $\mathcal{P}$ of the form

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[|\tilde{\mathbf{z}} - \boldsymbol{\mu}|] \leq \mathbf{f} \},$$

where the absolute value is understood to apply component-wise and $\mathbf{f} > \mathbf{0}$. The joint chance constraint is satisfied if and only if there exist $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^P$ and $\tau_j \in \mathbb{R}_+$, $j \in \mathcal{L}$, with

$$\left. \begin{aligned} 1 + \mathbf{f}^\top \boldsymbol{\gamma} &\geq 1 - \varepsilon, \quad -\boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq \boldsymbol{\gamma} \\ -\boldsymbol{\gamma} &\geq \boldsymbol{\beta} + \tau_j \mathbf{s}_j \geq \boldsymbol{\gamma} \\ \left\| \begin{bmatrix} 2 \\ \tau_j - t_j(\mathbf{x}) + \mathbf{s}_j^\top \boldsymbol{\mu} \end{bmatrix} \right\| &\leq \tau_j + t_j(\mathbf{x}) - \mathbf{s}_j^\top \boldsymbol{\mu} \end{aligned} \right\} \quad \forall j \in \mathcal{J}.$$

Example 3.2 (Mean Semi-Deviation). *Theorem 3.5 also allows us to derive a tractable reformulation of joint chance constraints over more general mean semi-deviation ambiguity sets $\mathcal{P}$ of the form*

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}} - \boldsymbol{\mu}]_+ \leq \mathbf{f}^+, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\mu} - \tilde{\mathbf{z}}]_+ \leq \mathbf{f}^- \},$$

where the operator $[\cdot]_+ = \max\{\cdot, 0\}$ applies component-wise and $\mathbf{f}^+, \mathbf{f}^- > \mathbf{0}$. The joint chance constraint is satisfied if and only if there exist $\boldsymbol{\beta}, \boldsymbol{\gamma}^+, \boldsymbol{\gamma}^- \in \mathbb{R}^P$, $\boldsymbol{\theta}_j, \boldsymbol{\eta}_j \in \mathbb{R}_+^P$, $j \in \mathcal{J} \cup \{0\}$, and $\tau_j \in \mathbb{R}_+$, $j \in \mathcal{J}$, with

$$\left. \begin{aligned} 1 + (\mathbf{f}^+)^\top \boldsymbol{\gamma}^+ + (\mathbf{f}^-)^\top \boldsymbol{\gamma}^- &\geq 1 - \varepsilon \\ \boldsymbol{\beta} = \boldsymbol{\theta}_0 - \boldsymbol{\eta}_0, \quad \boldsymbol{\gamma}^+ &\leq -\boldsymbol{\theta}_0, \boldsymbol{\gamma}^- \leq -\boldsymbol{\eta}_0 \\ \boldsymbol{\beta} + \tau_j \mathbf{s}_j = \boldsymbol{\theta}_j - \boldsymbol{\eta}_j, \quad \boldsymbol{\gamma}^+ &\leq -\boldsymbol{\theta}_j, \boldsymbol{\gamma}^- \leq -\boldsymbol{\eta}_j \\ \left\| \begin{bmatrix} 2 \\ \tau_j - t_j(\mathbf{x}) + \mathbf{s}_j^\top \boldsymbol{\mu} \end{bmatrix} \right\| &\leq \tau_j + t_j(\mathbf{x}) - \mathbf{s}_j^\top \boldsymbol{\mu} \end{aligned} \right\} \quad \forall j \in \mathcal{J}.$$

3.3.2 Intractability of Joint Chance Constraints with Support Information

We now consider distributionally robust chance constrained programs with right-hand side uncertainty where the support Ξ specified in the ambiguity set $\mathcal{P}$ does not fill out the entire space $\mathbb{R}^k$. We will show that such problems are strongly NP-hard even in the absence of any information about the distributions'

dispersion and even if the support Ξ is a hyperrectangle. To this end, we aim to reduce the IP feasibility problem from Section 3.2.1 to the following instance of the distributionally robust chance constrained program (3.2) with right-hand side uncertainty:

$$\begin{aligned}
& \min \quad 0 \\
& \text{s.t.} \quad \mathbf{x} \in [-1, 1]^n \\
& \quad \mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b} \\
& \quad \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(x_i \leq \tilde{\xi}_i, \quad -x_i \leq \tilde{\xi}_{n+i} \quad \forall i \in \mathcal{I} \right) \geq 1 - \varepsilon
\end{aligned} \tag{3.27}$$

Here we assume that $\varepsilon \in (0, 1)$, the matrix $\mathbf{A} \in \mathbb{Z}^{m \times n}$ and the vector $\mathbf{b} \in \mathbb{Z}^m$ correspond to the respective input parameters of the IP feasibility problem and the ambiguity set satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = 1 + c \left(1 - \frac{\varepsilon}{n} \right) \mathbf{e} \right\}, \tag{3.28}$$

where $k = 2n$ and $\Xi = [-1, 1 + c]^{2n}$. Problem (3.27) is a feasibility problem that evaluates to 0 if it has a feasible solution and to $+\infty$ otherwise. We will show that the problem is feasible if and only if the IP feasibility problem has an affirmative answer. To this end, we derive the following analytic expression for the worst-case probability of the chance constraint in (3.27).

Lemma 3.2. *For any fixed $\mathbf{x} \in [-1, 1]^{2n}$ the worst-case probability problem on the left-hand side of the chance constraint in (3.27) has the following analytic solution:*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(x_i \leq \tilde{\xi}_i, \quad -x_i \leq \tilde{\xi}_{n+i} \quad \forall i \in \mathcal{I} \right) = \max \left\{ 0, 1 - \varepsilon - \frac{2c\varepsilon}{n} \left(\sum_{i: x_i \in (-1, 1)} \frac{1 + c}{(1 + c)^2 - x_i^2} \right) \right\} \tag{3.29}$$

Proof. See Appendix. □

Since each term inside the inner parentheses in (3.29) is strictly positive, Lemma 3.2 implies that the left-hand side of the chance constraint in (3.27) is greater than or equal to $1 - \varepsilon$ if and only if $\mathbf{x} \in \{-1, 1\}^n$. The intuition behind Lemma 3.2 is given in Figure 3.3. If $x_i = 1$, then the i -th constraint inside the probability expression in (3.27) requires that $\tilde{\xi}_i \geq 1$, whereas there is no restriction on $\tilde{\xi}_{n+i}$. The expectation and support constraints in the ambiguity set (3.28) imply, however, that all probability

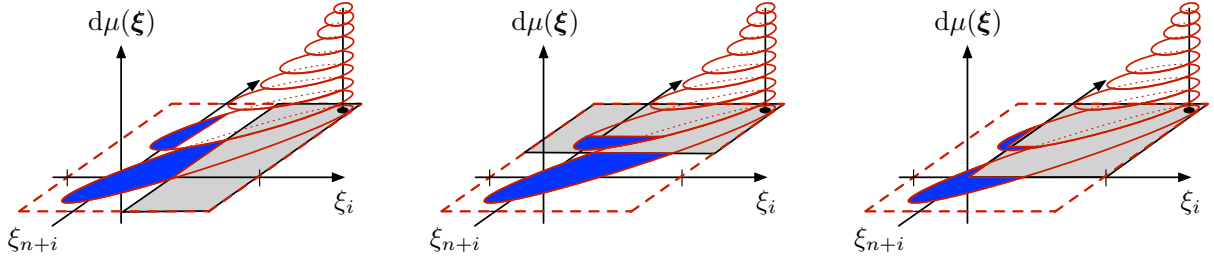


Figure 3.3: Illustration of the i -th constraint inside the probability expression in (3.27) when $x_i = 1$ (left), $x_i = -1$ (middle) and $x_i \in (-1, 1)$ (right). The light shaded region denotes the realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$ for which the constraint is satisfied. The contour lines represent the projected density function of a probability distribution $\mathbb{P} \in \mathcal{P}$ which violates the i -th constraint for the dark shaded realisations of $(\tilde{\xi}_i, \tilde{\xi}_{n+i})$.

distributions $\mathbb{P} \in \mathcal{P}$ place a probability mass of less than ε/n on realisations of $\tilde{\xi}$ where $\xi_i < 1$. Likewise, if $x_i = -1$, then the i -th constraint inside the probability expression in (3.27) requires that $\tilde{\xi}_{n+i} \geq 1$, and the constraints in the ambiguity set imply that all probability distributions $\mathbb{P} \in \mathcal{P}$ place a probability mass of less than ε/n on realisations of $\tilde{\xi}$ where $\xi_{n+i} < 1$. Thus, the overall violation probability of the chance constraint in (3.27) is bounded above by ε if $\mathbf{x} \in \{-1, 1\}^n$. If $x_i \in (-1, 1)$, on the other hand, then the i -th constraint inside the probability expression requires that $\tilde{\xi}_i \geq x_i$ and $\tilde{\xi}_{n+i} \geq -x_i$. By construction, there are distributions $\mathbb{P} \in \mathcal{P}$ that place a probability mass of $2c\varepsilon/n(c+2)$ on realisations of $\tilde{\xi}$ where either $(\tilde{\xi}_i, \tilde{\xi}_{n+i}) = (-1, 1+c)$ or $(\tilde{\xi}_i, \tilde{\xi}_{n+i}) = (1+c, -1)$. Thus, even if $x_j \in \{-1, 1\}$ for all $j \neq i$, there are distributions $\mathbb{P} \in \mathcal{P}$ for which the chance constraint is violated with probability $2c\varepsilon/n(c+2) + (n-1)\varepsilon/n = (2c/[c+2] + n-1)(\varepsilon/n) > \varepsilon$.

Equipped with the insight from Lemma 3.2, we can now prove the postulated intractability result.

Theorem 3.6. *The chance constrained program (3.2) with the ambiguity set (3.3) is strongly NP-hard for any fixed $\varepsilon \in (0, 1)$ even if $S = 0$ (that is, there is no information about the distributions' dispersion) and the support set Ξ is a hyperrectangle.*

Proof. We show that problem (3.27) is feasible if and only if the IP feasibility problem is. To this end, suppose first that there is $\mathbf{y} \in \{0, 1\}^n$ with $\mathbf{A}\mathbf{y} \leq \mathbf{b}$. We then have that $\mathbf{x} = 2\mathbf{y} - \mathbf{e} \in \{-1, 1\}^n$, $\mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b}$, and Lemma 3.2 implies that the chance constraint in (3.27) is satisfied, that is, problem (3.27) is feasible. Assume on the other hand that problem (3.27) has a feasible solution $\mathbf{x}$. By construction, we have $\mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b}$, and Lemma 3.2 implies that $\mathbf{x} \in \{-1, 1\}^n$. Thus, $\mathbf{y} = (\mathbf{x} + \mathbf{e})/2$ is binary and satisfies $\mathbf{A}\mathbf{y} \leq \mathbf{b}$, that is, it solves the IP feasibility problem. $\square$

3.3.3 Intractability of Generic Polyhedral Ambiguity Set

In this section, we briefly consider the more general polyhedral ambiguity set

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^k) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[g_s(\tilde{\xi})] \leq d_s \quad \forall s \in \mathcal{S}\}, \quad (3.30)$$

where the function $\mathbf{g} : \mathbb{R}^k \rightarrow \mathbb{R}^S$ is defined as

$$g_s(\boldsymbol{\xi}) = \max_{t \in \mathcal{T}} \mathbf{g}_{st}^\top (\boldsymbol{\mu} - \boldsymbol{\xi}) + h_{st} \quad \forall s \in \mathcal{S} = \{1, \dots, S\}. \quad (3.31)$$

In the following, we will show that unless $h_{st} = 0, s \in \mathcal{S}, t \in \mathcal{T}$, optimisation problems involving the chance constraint are intractable in general. We aim to reduce the IP feasibility problem to the following feasibility problem.

$$\begin{aligned} \min \quad & 0 \\ \text{s.t.} \quad & \mathbf{x} \in [-1, 1]^n \\ & \mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b} \\ & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(-(c-1)\mathbf{e} \leq \mathbf{x} + \tilde{\xi} \leq (c-1)\mathbf{e} \right) \geq 1 - \varepsilon \end{aligned} \quad (3.32)$$

Here, we consider the ambiguity set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^k) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mathbf{0}, \mathbb{E}_{\mathbb{P}}[|\tilde{\xi}|] \leq \frac{(2c-4)(\varepsilon-\delta)}{n} \mathbf{e}, \mathbb{E}_{\mathbb{P}}[\mathbf{g}(\tilde{\xi})] \leq \frac{\delta^2}{n} \mathbf{e} \right\}, \quad (3.33)$$

where $\mathbf{g}(\boldsymbol{\xi}) = \max(\boldsymbol{\xi} - (c-\delta)\mathbf{e}, \mathbf{0}, -\boldsymbol{\xi} - (c-\delta)\mathbf{e}), \forall \boldsymbol{\xi} \in \mathbb{R}^n$. Here, c is any constant greater than 4 and $\delta \leq \frac{c-3}{4(c-2)}\varepsilon$ is a small positive number whose purpose will become obvious later. The last constraint in (3.33) is designed such that most of the mass of distributions $\mathbb{P} \in \mathcal{P}$ is supported on $[-(c-\delta), c-\delta]^n$.

Lemma 3.3. *If $\mathbf{x} \in \{-1, 1\}^n$, then the worst-case probability on the left-hand side of the chance constraint in (3.32) amounts to at least $1 - \varepsilon$.*

Proof. We first argue that the worst-case probability on the left-hand side of the chance constraint in (3.32) is identical for all $\mathbf{x} \in \{-1, 1\}^n$. To see this, choose any $\mathbf{x}, \mathbf{x}' \in \{-1, 1\}^n$. Then for any $\mathbb{P} \in \mathcal{P}$,

we have

$$\mathbb{P}\left(-(c-1)\mathbf{e} \leq \mathbf{x} + \tilde{\boldsymbol{\xi}} \leq (c-1)\mathbf{e}\right) = \mathbb{Q}\left(-(c-1)\mathbf{e} \leq \mathbf{x}' + \tilde{\boldsymbol{\xi}} \leq (c-1)\mathbf{e}\right)$$

for the distribution $\mathbb{Q} \in \mathcal{P}$ that satisfies $\mathbb{Q}(\tilde{\boldsymbol{\xi}} \in A) = \mathbb{P}(f(\tilde{\boldsymbol{\xi}}) \in A)$ for all Borel-measurable sets $A \subseteq \mathbb{R}^n$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined through $f_i(\xi_i) = \xi_i$ if $x_i = x'_i$; $= -\xi_i$ otherwise. Thus, in the remainder of the proof we assume without loss of generality that $\mathbf{x} = -\mathbf{e}$, in which case the worst-case probability on the left-hand side of the chance constraint in (3.32) simplifies to

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(-(c-2)\mathbf{e} \leq \tilde{\boldsymbol{\xi}} \leq c\mathbf{e}\right).$$

We can reformulate this worst-case probability problem as the optimal value of the moment problem

$$\begin{aligned} \min \quad & \int \mathbb{I}_{[-(c-2)\mathbf{e} \leq \boldsymbol{\xi} \leq c\mathbf{e}]} d\mu(\boldsymbol{\xi}) \\ \text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbb{R}^n) \\ & \int d\mu(\boldsymbol{\xi}) = 1 \\ & \int \boldsymbol{\xi} d\mu(\boldsymbol{\xi}) = \mathbf{0} \\ & \int |\boldsymbol{\xi}| d\mu(\boldsymbol{\xi}) \leq \frac{(2c-4)(\varepsilon-\delta)}{n} \mathbf{e} \\ & \int \mathbf{g}(\boldsymbol{\xi}) d\mu(\boldsymbol{\xi}) \leq \frac{\delta^2}{n} \mathbf{e}. \end{aligned}$$

The dual to this problem is given by the semi-infinite LP

$$\begin{aligned} \max \quad & \alpha - \frac{(2c-4)(\varepsilon-\delta)}{n} \boldsymbol{\gamma}^\top \mathbf{e} - \frac{\delta^2}{n} \boldsymbol{\zeta}^\top \mathbf{e} \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^n, \quad \boldsymbol{\gamma}, \boldsymbol{\zeta} \in \mathbb{R}_+^n \\ & \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} - \boldsymbol{\gamma}^\top |\boldsymbol{\xi}| - \boldsymbol{\zeta}^\top \mathbf{g}(\boldsymbol{\xi}) \leq 1 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n \\ & \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} - \boldsymbol{\gamma}^\top |\boldsymbol{\xi}| - \boldsymbol{\zeta}^\top \mathbf{g}(\boldsymbol{\xi}) \leq 0 \quad \forall i \in \mathcal{I} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n : \xi_i > c \\ & \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} - \boldsymbol{\gamma}^\top |\boldsymbol{\xi}| - \boldsymbol{\zeta}^\top \mathbf{g}(\boldsymbol{\xi}) \leq 0 \quad \forall i \in \mathcal{I} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n : \xi_i < -(c-2). \end{aligned} \tag{3.34}$$

Strong conic duality holds for the above primal and dual pair as the Dirac distribution $\delta_{\mathbf{0}}$ satisfies the bounds in the primal problem strictly [103]. Consider now the solution $\alpha = 1$, $\boldsymbol{\beta} = \boldsymbol{\gamma} = \mathbf{e}/(2c-4)$ and $\boldsymbol{\zeta} = \mathbf{e}/\delta$ to problem (3.34). For this solution, the function $\alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} - \boldsymbol{\gamma}^\top |\boldsymbol{\xi}| - \boldsymbol{\zeta}^\top \mathbf{g}(\boldsymbol{\xi})$ appearing on the

left-hand sides of the constraints in (3.34) satisfies

$$\alpha + \beta^\top \xi - \gamma^\top |\xi| - \zeta^\top g(\xi) = 1 + \frac{1}{2c-4} \mathbf{e}^\top (\xi - |\xi|) - \frac{1}{\delta} \mathbf{e}^\top \max(\xi - (c-\delta)\mathbf{e}, \mathbf{0}, -\xi - (c-\delta)\mathbf{e}).$$

Over $\mathbb{R}^n$, this function attains its maximum value of 1 at any point $\xi^* \in [0, c-\delta]^n$. Hence, $(\alpha, \beta, \gamma, \zeta)$ satisfies the first semi-infinite constraint in (3.34).

In view of the second semi-infinite constraint in (3.34), we observe that for each $i \in \mathcal{I}$, we have

$$\begin{aligned} & \sup \left\{ 1 + \frac{1}{2c-4} \mathbf{e}^\top (\xi - |\xi|) - \frac{1}{\delta} \mathbf{e}^\top \max(\xi - (c-\delta)\mathbf{e}, \mathbf{0}, -\xi - (c-\delta)\mathbf{e}) : \xi \in \mathbb{R}^n, \xi_i > c \right\} \\ & \leq \max \left\{ 1 + \frac{1}{2c-4} \mathbf{e}^\top (\xi - |\xi|) - \frac{1}{\delta} \mathbf{e}^\top \max(\xi - (c-\delta)\mathbf{e}, \mathbf{0}, -\xi - (c-\delta)\mathbf{e}) : \xi \in \mathbb{R}^n, \xi_i \geq c \right\} \\ & \leq \min \left\{ \begin{array}{l} \tau \in \mathbb{R}_+, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}_+^n \\ 1 + (c-\delta) \mathbf{e}^\top (\boldsymbol{\phi} + \boldsymbol{\psi}) - \tau c : \mathbf{e}/(2c-4) - \boldsymbol{\phi} + \boldsymbol{\psi} + \tau \mathbf{e}_i = \boldsymbol{\theta} - \boldsymbol{\eta} \\ \boldsymbol{\theta} + \boldsymbol{\eta} = \mathbf{e}/(2c-4), \boldsymbol{\phi} + \boldsymbol{\psi} \leq \mathbf{e}/\delta \end{array} \right\} \leq 0. \end{aligned}$$

Here, the first inequality holds because we replace the strict inequality inside the supremum with a weak one, while the second inequality follows from weak LP duality. The last inequality holds since the solution $\tau = 1/\delta$, $\boldsymbol{\theta} = \mathbf{e}/(2c-4)$, $\boldsymbol{\eta} = \mathbf{0}$, $\boldsymbol{\phi} = \mathbf{e}_i/\delta$, and $\boldsymbol{\psi} = \mathbf{0}$ is feasible in the minimisation problem and attains an objective value of 0.

In view of the third semi-infinite constraint in (3.34), we observe that for each $i \in \mathcal{I}$, we have

$$\begin{aligned} & \sup \left\{ 1 + \frac{1}{2c-4} \mathbf{e}^\top (\xi - |\xi|) - \frac{1}{\delta} \mathbf{e}^\top \max(\xi - (c-\delta)\mathbf{e}, \mathbf{0}, -\xi - (c-\delta)\mathbf{e}) : \xi \in \mathbb{R}^n, \xi_i < -c+2 \right\} \\ & \leq \max \left\{ 1 + \frac{1}{2c-4} \mathbf{e}^\top (\xi - |\xi|) - \frac{1}{\delta} \mathbf{e}^\top \max(\xi - (c-\delta)\mathbf{e}, \mathbf{0}, -\xi - (c-\delta)\mathbf{e}) : \xi \in \mathbb{R}^n, \xi_i \leq -c+2 \right\} \\ & \leq \min \left\{ \begin{array}{l} \tau \in \mathbb{R}_+, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}_+^n \\ 1 + (c-\delta) \mathbf{e}^\top (\boldsymbol{\phi} + \boldsymbol{\psi}) - \tau(c-2) : \mathbf{e}/(2c-4) - \boldsymbol{\phi} + \boldsymbol{\psi} - \tau \mathbf{e}_i = \boldsymbol{\theta} - \boldsymbol{\eta} \\ \boldsymbol{\theta} + \boldsymbol{\eta} = \mathbf{e}/(2c-4), \boldsymbol{\phi} + \boldsymbol{\psi} \leq \mathbf{e}/\delta \end{array} \right\} \leq 0. \end{aligned}$$

Here, the first inequality holds because we replace the strict inequality inside the supremum with a weak one, while the second inequality follows from weak LP duality. The last inequality holds since the solution $\tau = 1/(c-2)$, $\boldsymbol{\theta} = (\mathbf{e} - \mathbf{e}_i)/(2c-4)$, $\boldsymbol{\eta} = \mathbf{e}_i/(2c-4)$, $\boldsymbol{\phi} = \mathbf{0}$, and $\boldsymbol{\psi} = \mathbf{0}$ is feasible in the minimisation problem and attains an objective value of 0.

We thus conclude that $(\alpha, \beta, \gamma, \zeta)$ is feasible in (3.34). Since $(\alpha, \beta, \gamma, \zeta)$ attains the objective value $1 - \varepsilon$ in (3.34), our previous duality argument implies that the worst-case probability on the left-hand side of the chance constraint in (3.32) must be at least $1 - \varepsilon$. $\square$

Lemma 3.4. *If $\mathbf{x} \in [-1, 1]^n \setminus ([-1, -1 + \delta] \cup (1 - \delta, 1])^n$, then the worst-case probability on the left-hand side of the chance constraint in (3.32) is strictly less than $1 - \varepsilon$.*

Proof. By virtue of Lemma 3.3, the worst-case probability on the left-hand side of the chance constraint in (3.32) is identical to

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(-(c-1) \mathbf{e} < \mathbf{x} + \tilde{\boldsymbol{\xi}} < (c-1) \mathbf{e} \right). \quad (3.35)$$

We denote $\mathcal{I}(\mathbf{x}) = \{i \in \mathcal{I} : x_i \in [-1, -1 + \delta] \cup (1 - \delta, 1]\}$ the components of $\mathbf{x}$ that are close to the integer values $\{-1, 1\}^n$. Consider now the discrete distribution $\mathbb{P}^*$ that satisfies

$$\mathbb{P}^*(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}) = \begin{cases} \kappa & \text{for } \boldsymbol{\xi} = \boldsymbol{\chi} \\ \frac{(\varepsilon - \delta)(c-2)}{n(c+x_i-1)} & \text{for } \boldsymbol{\xi} = (-c+1-x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \\ \frac{(x_i+1)\varepsilon + \delta(c-2)}{n(c+x_i-1)} & \text{for } \boldsymbol{\xi} = (c-1-x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \\ \frac{(\varepsilon - \delta)(c-2)(c-x_i-1)}{n(c-1)^2} & \text{for } \boldsymbol{\xi} = (-c+1-x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}) \\ \frac{(\varepsilon - \delta)(c-2)(c+x_i-1)}{n(c-1)^2} & \text{for } \boldsymbol{\xi} = (c-1-x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}), \end{cases} \quad (3.36)$$

where $\kappa = 1 - \sum_{i \in \mathcal{I}(\mathbf{x})} \frac{\varepsilon}{n} - \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \frac{(\varepsilon - \delta)(2c-4)}{n(c-1)}$ and the components of $\boldsymbol{\chi}$ are defined as

$$\chi_i = \frac{2(c-1)(c-2)(\varepsilon - \delta) + \varepsilon(x_i^2 - (c-1)^2)}{\kappa n(c+x_i-1)},$$

if $i \in \mathcal{I}(\mathbf{x})$; $= 0$ otherwise. By construction, we have $\mathbb{P}^* \in \mathcal{P}$ and

$$\mathbb{P}^* \left(-(c-1) \mathbf{e} < \mathbf{x} + \tilde{\boldsymbol{\xi}} < (c-1) \mathbf{e} \right) = \kappa < 1 - \varepsilon,$$

since $\frac{\varepsilon}{n} < \frac{(\varepsilon - \delta)(2c-4)}{n(c-1)}$ following our assumption that $\delta \leq \frac{c-3}{4(c-2)}\varepsilon$ and the realisations $\boldsymbol{\xi} \in \{(-c+1-x_1)\mathbf{e}_1, (c-1-x_1)\mathbf{e}_1, \dots, (-c+1-x_n)\mathbf{e}_n, (c-1-x_n)\mathbf{e}_n\}$ all violate the constraints inside the probability expression. Thus, the worst-case probability (3.36) is indeed strictly less than $1 - \varepsilon$. $\square$

Lemma 3.5. Let $\delta = \frac{1}{2n \max_{j \in \{1, \dots, m\}} \|\mathbf{a}_j\|_\infty}$. Assume that $\mathbf{y}' \in ([0, \delta) \cup (1 - \delta, 1])^n$ satisfies $\mathbf{A}\mathbf{y}' \leq \mathbf{b}$. Then $\mathbf{A}\mathbf{y} \leq \mathbf{b}$, where $\mathbf{y} \in \{0, 1\}^n$ is a rounding of $\mathbf{y}'$ to the nearest binary vector.

Proof. For any $j \in \{1, \dots, m\}$ we have

$$\mathbf{a}_j^\top (\mathbf{y} - \mathbf{y}') \leq |\mathbf{a}_j^\top (\mathbf{y} - \mathbf{y}')| \leq \|\mathbf{a}_j\|_\infty \|\mathbf{y} - \mathbf{y}'\|_1 \leq \|\mathbf{a}_j\|_\infty n \delta.$$

Following the definition of δ , we obtain $\mathbf{a}_j^\top (\mathbf{y} - \mathbf{y}') \leq \frac{1}{2}$, hence

$$\mathbf{a}_j^\top \mathbf{y} \leq \frac{1}{2} + \mathbf{a}_j^\top \mathbf{y}' \leq \frac{1}{2} + b_j.$$

Since both $\mathbf{a}_j^\top \mathbf{y}$ and b_j are integers, we conclude that $\mathbf{a}_j^\top \mathbf{y} \leq b_j$. Applying this result for all $j \in \{1, \dots, m\}$, we obtain that $\mathbf{A}\mathbf{y} \leq \mathbf{b}$. Thus the claim follows. $\square$

Theorem 3.7. The chance constrained program (3.32) is strongly NP-hard.

Proof. We set $\delta = \min \left\{ \frac{c-3}{4(c-2)}\varepsilon, \frac{1}{n \max_{j \in \{1, \dots, m\}} \|\mathbf{a}_j\|_\infty} \right\}$ in the definition of the ambiguity set (3.33). We show that problem (3.32) is feasible if and only if the IP feasibility problem is. To this end, suppose that there exists $\mathbf{y} \in \{0, 1\}^n$ with $\mathbf{A}\mathbf{y} \leq \mathbf{b}$. Lemma 3.3 then implies that the distributionally robust joint chance constraint in (3.32) is satisfied by $\mathbf{x} = 2\mathbf{y} - \mathbf{e}$, that is, problem (3.32) is feasible.

Assume, on the other hand that there is no binary solution to the IP feasibility problem. The contraposition of Lemma 3.5 shows that there is no $\mathbf{y} \in ([0, \frac{\delta}{2}) \cup (1 - \frac{\delta}{2}, 1])^n$ that satisfies $\mathbf{A}\mathbf{y} \leq \mathbf{b}$. We thus consider $\mathbf{y} \in [0, 1]^n \setminus ([0, \frac{\delta}{2}) \cup (1 - \frac{\delta}{2}, 1])^n$. Lemma 3.4, however, implies that $\mathbf{x} = 2\mathbf{y} - \mathbf{e}$ violates the chance constraint in (3.32). We thus conclude that problem (3.32) is also infeasible. $\square$

3.4 Best-Case Joint Chance Constraints with Conic Dispersion Measures

We can apply the same techniques introduced so far in this chapter to ambiguous chance constraint where the supremum (rather than the infimum) of the probability over the ambiguity set is required. To the best of our knowledge, such problems have not been studied in the literature with as much detail

as the worst-case ambiguous chance constraints. Nevertheless, we consider them here and show their relevance in Chapter 5 with a novel application. We will see that best-case ambiguous chance constraints are better behaved than their worst-case counterparts, but require very similar assumptions to guarantee tractability. To this end, we consider a generalisation of the chance constraint in (3.23),

$$\begin{aligned} \sup \quad & \mathbb{P} \left(s(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \varepsilon \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}, \end{aligned}$$

where $\mathcal{P}$ is the following generalisation of ambiguity set (3.24):

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[g(\tilde{\boldsymbol{\xi}})] \preceq_{\mathcal{K}} \mathbf{d} \right\},$$

and where $\mathbf{x} \in \mathbb{R}^n$, $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^P$, $\Xi \subseteq \mathbb{R}^n$, $\mathbf{d} \in \mathbb{R}^Q$, $g : \mathbb{R}^P \rightarrow \mathbb{R}^Q$, $s : \mathbb{R}^n \times \mathbb{R}^P \rightarrow \mathbb{R}$. We shall require the following assumptions about the parameters in the chance constraint.

(M) The point $(1, \boldsymbol{\mu}, \mathbf{d})$ is in the interior of

$$\mathcal{V} = \left\{ (a, \mathbf{b}, \mathbf{c}) \in \mathbb{R} \times \mathbb{R}^P \times \mathbb{R}^Q : \exists \mu \in \mathcal{M}_+(\Xi) \text{ such that } \begin{aligned} & \int \mu(d\boldsymbol{\xi}) = a, \\ & \int \boldsymbol{\xi} \mu(d\boldsymbol{\xi}) = \mathbf{b}, \\ & \int g(\boldsymbol{\xi}) \mu(d\boldsymbol{\xi}) \preceq_{\mathcal{K}} \mathbf{c} \end{aligned} \right\}.$$

As before, this allows us to invoke the strong duality results for moment problems and enables us to equivalently write the dual of the best-case probability problem as a semi-infinite problem. Additionally, this assumption ensures that the set of optimal solutions to the dual problem is non-empty and bounded, allowing the application of a minimax theorem, as we will see.

(S) For every $\mathbf{x}$ there is $\boldsymbol{\xi} \in \mathbb{R}^P$ such that $s(\mathbf{x}, \boldsymbol{\xi}) < 0$, $\mathcal{K}$ is a proper cone with non-empty interior, and Ξ is a convex set with non-empty interior. These Slater-type conditions will be required in a second dualisation step, again to ensure strong duality.

(C) The functions s and g are convex and $\mathcal{K}$ -convex in $\boldsymbol{\xi}$ respectively. This condition is again required for strong duality.

Theorem 3.8. *If the assumptions (M), (S), and (C) are satisfied, then*

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(s(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0) &\iff \sup_{\lambda \in \mathbb{R}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \Xi} 1 - \lambda \\
\text{s.t. } &\lambda \in \mathbb{R}, \quad \boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \Xi \\
&\lambda \boldsymbol{\xi}_1 + (1 - \lambda) \boldsymbol{\xi}_2 = \boldsymbol{\mu} \\
&\lambda g(\boldsymbol{\xi}_1) + (1 - \lambda) g(\boldsymbol{\xi}_2) \preccurlyeq_{\mathcal{K}} \mathbf{d} \\
&s(\mathbf{x}, \boldsymbol{\xi}_2) \leq 0 \\
&0 \leq \lambda \leq 1.
\end{aligned} \tag{3.37}$$

Proof. We dualise the best-case probability problem in the usual manner. Due to (M) and (S), the dual is equivalent to the following semi-infinite problem:

$$\begin{aligned}
\inf_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^P, \boldsymbol{\gamma} \in \mathbb{R}^Q} \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \boldsymbol{\gamma}^\top \mathbf{d} &= \inf_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^P, \boldsymbol{\gamma} \in \mathbb{R}^Q} \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \boldsymbol{\gamma}^\top \mathbf{d} \\
\text{s.t. } \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \Xi &\quad \text{s.t. } \alpha + h_0(\boldsymbol{\beta}, \boldsymbol{\gamma}) \geq 0 \\
\alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}) \geq 1 \quad \forall \boldsymbol{\xi} \in \Xi : s(\mathbf{x}, \boldsymbol{\xi}) \leq 0 &\quad \alpha + h_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}) \geq 1 \\
\boldsymbol{\gamma} \in \mathcal{K}^* &\quad \boldsymbol{\gamma} \in \mathcal{K}^*,
\end{aligned} \tag{3.38}$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$, and the equality follows from re-writing the semi-infinite constraints as minimisations over $\boldsymbol{\xi}$ with the following auxiliary functions

$$\begin{aligned}
h_0(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \inf_{\boldsymbol{\xi} \in \Xi} \boldsymbol{\beta}^\top \boldsymbol{\xi} + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}) \quad \text{and} \quad h_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}) = \inf_{\boldsymbol{\xi} \in \Xi} \boldsymbol{\beta}^\top \boldsymbol{\xi} + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}) \\
\text{s.t. } \boldsymbol{\xi} &\in \Xi \\
s(\mathbf{x}, \boldsymbol{\xi}) &\leq 0.
\end{aligned}$$

If we were to proceed in the usual manner, the next step would be to convert the semi-infinite constraints into a finite set of constraints by dualising the infimum problems in h_0 and h_1 . We delay this step and instead dualise the outer problem (3.38). We note that both h_0 and h_1 take finite values and are jointly concave in $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ since they are minimums of linear functions. Therefore, we may construct a strictly

feasible solution by selecting α large enough. The dual problem of (3.38) is therefore

$$\sup \inf \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \boldsymbol{\gamma}^\top \mathbf{d} - \lambda_1(\alpha + h_0(\boldsymbol{\beta}, \boldsymbol{\gamma})) - \lambda_2(\alpha + h_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}) - 1) - \bar{\boldsymbol{\gamma}}^\top \boldsymbol{\gamma},$$

where the supremum is over the dual variables $\lambda_1 \in \mathbb{R}_+$, $\lambda_2 \in \mathbb{R}_+$, and $\bar{\boldsymbol{\gamma}} \in \mathcal{K}$, while the infimum is over the primal variables $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^P$, and $\boldsymbol{\gamma} \in \mathbb{R}^Q$. We can directly perform the unconstrained infimum over α . Then, remembering our definition of h_0 and h_1 , and the fact that dual variables λ_1 and λ_2 are non-negative, we can equivalently write the above problem as

$$\sup \inf \sup \boldsymbol{\beta}^\top \boldsymbol{\mu} + \boldsymbol{\gamma}^\top \mathbf{d} - \lambda(\boldsymbol{\beta}^\top \boldsymbol{\xi}_1 + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}_1)) - (1 - \lambda)(\boldsymbol{\beta}^\top \boldsymbol{\xi}_2 + \boldsymbol{\gamma}^\top g(\boldsymbol{\xi}_2) - 1) - \bar{\boldsymbol{\gamma}}^\top \boldsymbol{\gamma},$$

where the outer supremum is over $\lambda \in \mathbb{R} : 0 \leq \lambda \leq 1$ and $\bar{\boldsymbol{\gamma}} \in \mathcal{K}$, while the infimum is over $\boldsymbol{\beta} \in \mathbb{R}^P$ and $\boldsymbol{\gamma} \in \mathbb{R}^Q$. The inner supremum is over $\boldsymbol{\xi}_1 \in \Xi$ and $\boldsymbol{\xi}_2 \in \Xi : s(\mathbf{x}, \boldsymbol{\xi}_2) \leq 0$. The function above is convex in $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and concave in $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, and the set of optimal solutions for the middle infimum problem is non-empty and bounded due to (M) and [103, Proposition 3.4]. We also use (S), which ensures that the feasible set for the inner supremum is convex, non-empty, and has an interior point. Therefore, the infimum and inner supremum can be interchanged by Sion's minimax theorem [110], which yields

$$\sup \inf (1 - \lambda) + \boldsymbol{\beta}^\top (\boldsymbol{\mu} - \lambda \boldsymbol{\xi}_1 - (1 - \lambda) \boldsymbol{\xi}_2) + \boldsymbol{\gamma}^\top (\mathbf{d} - \bar{\boldsymbol{\gamma}} - \lambda g(\boldsymbol{\xi}_1) - (1 - \lambda) g(\boldsymbol{\xi}_2)),$$

where the supremum is over $\lambda \in \mathbb{R} : 0 \leq \lambda \leq 1$, $\bar{\boldsymbol{\gamma}} \in \mathcal{K}$, $\boldsymbol{\xi}_1 \in \Xi$, and $\boldsymbol{\xi}_2 \in \Xi : s(\mathbf{x}, \boldsymbol{\xi}_2) \leq 0$, while the infimum is over $\boldsymbol{\beta} \in \mathbb{R}^P$ and $\boldsymbol{\gamma} \in \mathbb{R}^Q$. The claim now follows by performing the unconstrained infimum and eliminating $\bar{\boldsymbol{\gamma}}$. $\square$

The equivalence in Lemma 3.8 is comparable to the bidual of the worst-case probability problem from which the worst-case distributions are extracted. Problems of this structure are sometimes known as Haar's dual problems [22]. The best-case uncertainty quantification problem requires the solution of the non-convex problem (3.37). However, since λ is one-dimensional and bounded by the unit interval, for practical purposes the interval $[0, 1]$ may be discretised into $1/\delta$ points to guarantee that the absolute difference between the optimal λ of the discretised problem and the non-convex problem is at most δ .

More progress can be made if we impose the following (still reasonable) assumptions

- (H) The function g is positive homogeneous in its argument $\boldsymbol{\xi}$, i.e. $\lambda g(\boldsymbol{\xi}) = g(\lambda \boldsymbol{\xi})$ for all $\lambda \geq 0$.
- (D) The dispersion measure g is $\mathcal{K}$ -non-negative over Ξ , i.e. $g(\boldsymbol{\xi}) \succ_{\mathcal{K}} \mathbf{0}$ for all $\boldsymbol{\xi} \in \Xi$, and symmetric around the mean, i.e. $g(\boldsymbol{\mu} + \boldsymbol{\xi}) = g(\boldsymbol{\mu} - \boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \Xi$.
- (P) The mean is at the origin, i.e. $\boldsymbol{\mu} = \mathbf{0}$. We have seen previously that this transformation can be applied without loss of generality. To simplify the exposition, we also assume that Ξ is polyhedral, i.e., can be represented as a set of linear inequalities $\mathbf{W}\boldsymbol{\xi} + \mathbf{w} \geq 0$, and that $\boldsymbol{\mu} \in \Xi$, implying $\mathbf{w} \geq 0$.

Theorem 3.9. *If the assumptions (M), (S), (C), (H), (D), and (P) are satisfied, then*

$$\begin{aligned}
 \sup_{\text{s.t. } \mathbb{P} \in \mathcal{P}} \mathbb{P}(s(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0) &\geq 1 - \varepsilon &\iff \exists \boldsymbol{\xi} \in \mathbb{R}^P : \\
 &2(1 - \varepsilon)g(\boldsymbol{\xi}) \preccurlyeq_{\mathcal{K}} \mathbf{d} \\
 &s(\mathbf{x}, \boldsymbol{\xi}) \leq 0 \\
 &\varepsilon \mathbf{w} \leq \mathbf{W}\boldsymbol{\xi} \leq (1 + \varepsilon)\mathbf{w}.
 \end{aligned} \tag{3.39}$$

Proof. We start with (3.37) and first substitute out $\boldsymbol{\xi}_1$ using the equality constraint $\boldsymbol{\xi}_1 = -\frac{1-\lambda}{\lambda}\boldsymbol{\xi}_2$:

$$\begin{aligned}
 \sup \quad &1 - \lambda &&\geq 1 - \varepsilon &\iff \exists \lambda \in \mathbb{R}, \quad \boldsymbol{\xi} \in \mathbb{R}^P : \\
 \text{s.t.} \quad &\lambda \in \mathbb{R}, \quad \boldsymbol{\xi} \in \mathbb{R}^P &&2(1 - \lambda)g(\boldsymbol{\xi}) \preccurlyeq_{\mathcal{K}} \mathbf{d} \\
 &\lambda g\left(-\frac{1-\lambda}{\lambda}\boldsymbol{\xi}\right) + (1 - \lambda)g(\boldsymbol{\xi}) \preccurlyeq_{\mathcal{K}} \mathbf{d} &&s(\mathbf{x}, \boldsymbol{\xi}) \leq 0 \\
 &s(\mathbf{x}, \boldsymbol{\xi}) \leq 0 &&-\mathbf{w} \leq \mathbf{W}\boldsymbol{\xi} \leq \frac{\lambda}{1-\lambda}\mathbf{w} \\
 &-\frac{1-\lambda}{\lambda}\mathbf{W}\boldsymbol{\xi} + \mathbf{w} \geq \mathbf{0} \quad \mathbf{W}\boldsymbol{\xi} + \mathbf{w} \geq \mathbf{0} &&0 \leq \lambda \leq \varepsilon \\
 &0 \leq \lambda \leq 1
 \end{aligned}$$

The above equivalence is by positive homogeneity and symmetry of g around $\mathbf{0}$. We now show that given a feasible solution with $\lambda < \varepsilon$, we may construct another feasible solution where $\lambda = \varepsilon$. The claim follows since the feasible set expands monotonically as λ is increased (due to $\mathcal{K}$ -non-negativity and non-negativity of g and $\mathbf{w}$ respectively) until $\lambda = \varepsilon$. $\square$

Under these three additional assumptions, Lemma 3.9 shows that for a fixed $\mathbf{x}$, we can determine if it satisfies the chance constraint by solving a convex feasibility problem. Furthermore, if the function s is jointly convex in $(\mathbf{x}, \boldsymbol{\xi})$, then the feasible set of $\mathbf{x}$ for the chance constraint is convex as well. The best-case ambiguous chance constraint will therefore be equivalent to a convex conic constraint, the tractability of which will depend on the structure of $\mathcal{K}$ and s .

3.5 Conclusion

In this chapter, we studied optimisation problems involving ambiguous chance constraints in detail. We characterised them into two distinct families — best-case and worst-case. Both types have different application areas, but the techniques used to reformulate them are similar.

We delineated the watershed between tractability and non-tractability. In the former we gave concrete, and most importantly, exact reformulations for specific classes of problems. In the latter we showed that in general these problems, although very useful in practice, become NP-hard very quickly. For these hard problems we show how the well-known technique of alternating convex optimisation can be applied.

3.6 Appendix

Proof of Theorem 3.1. For any fixed $\mathbf{x} \in \mathcal{X}$ we have

$$\begin{aligned}
\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{T}(\tilde{\boldsymbol{\xi}}) \mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}}) \right) &= \inf \int \mathbb{I}_{[\mathbf{T}(\boldsymbol{\xi}) \mathbf{x} \leq \mathbf{u}(\boldsymbol{\xi})]} d\mu(\boldsymbol{\xi}) \\
\text{s.t. } \mu &\in \mathcal{M}_+(\Xi) \\
\int d\mu(\boldsymbol{\xi}) &= 1 \\
\int \boldsymbol{\xi} d\mu(\boldsymbol{\xi}) &= \boldsymbol{\mu} \\
\int \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} d\mu(\boldsymbol{\xi}) &\leq d_s \quad \forall s \in \mathcal{S}.
\end{aligned}$$

This infinite-dimensional LP over non-negative measures has a semi-infinite dual LP of the form

$$\begin{aligned}
\sup \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \mathbf{d}^\top \boldsymbol{\gamma} \\
\text{s.t. } \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^S \\
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} \leq \mathbb{I}_{[\mathbf{T}(\boldsymbol{\xi}) \mathbf{x} \leq \mathbf{u}(\boldsymbol{\xi})]} \quad \forall \boldsymbol{\xi} \in \Xi.
\end{aligned} \tag{3.40}$$

Recalling the definitions of $\mathbf{T}(\boldsymbol{\xi})$ and $\mathbf{u}(\boldsymbol{\xi})$, the semi-infinite constraint in (3.40) can be reformulated as

$$\begin{aligned}
\alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\leq 1 \quad \forall \boldsymbol{\xi} \in \Xi \\
\alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\leq 0 \quad \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \boldsymbol{\xi}^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} > \mathbf{u}_i^\top \boldsymbol{\xi} + v_i.
\end{aligned}$$

The parameter sets $\{ \boldsymbol{\xi} \in \Xi : \boldsymbol{\xi}^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} > \mathbf{u}_i^\top \boldsymbol{\xi} + v_i \}$ are empty for all $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})$, and thus the semi-infinite constraint in (3.40) holds if and only if

$$\begin{aligned}
\alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\leq 1 \quad \forall \boldsymbol{\xi} \in \Xi \\
\alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \sum_{s \in \mathcal{S}} \gamma_s \max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} &\leq 0 \quad \forall i \in \mathcal{I}(\mathbf{x}), \forall \boldsymbol{\xi} \in \Xi : \boldsymbol{\xi}^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} > \mathbf{u}_i^\top \boldsymbol{\xi} + v_i.
\end{aligned} \tag{3.41}$$

By using an epigraph formulation for the ‘max’-functions, which is applicable because $\gamma \geq \mathbf{0}$, the first constraint in (3.41) can be expressed as

$$\begin{aligned} 1 &\geq \max \quad \alpha + \beta^\top \xi - \gamma^\top \rho \\ \text{s.t.} \quad &\xi \in \mathbb{R}^k, \rho \in \mathbb{R}^S \\ &\mathbf{W}\xi \leq \mathbf{w}, \quad \rho_s \geq \mathbf{g}_{st}^\top (\xi - \mu) + h_{st} \quad \forall s \in \mathcal{S}, \forall t \in \mathcal{T}. \end{aligned} \quad (3.42)$$

Note that the LP in (3.42) is feasible because the support set Ξ is non-empty. Strong LP duality thus implies that (3.42) is equivalent to the system of linear constraints

$$\exists \theta_0 \in \mathbb{R}_+^{S \times T}, \phi_0 \in \mathbb{R}_+^l : \begin{cases} 1 \geq \alpha + \mathbf{w}^\top \phi_0 + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{0st} (\mathbf{g}_{st}^\top \mu - h_{st}) \\ \beta = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \theta_{0st} \mathbf{g}_{st} + \mathbf{W}^\top \phi_0, \quad \gamma_s = \sum_{t \in \mathcal{T}} \theta_{0st} \quad \forall s \in \mathcal{S}. \end{cases} \quad (3.43)$$

Next, consider any semi-infinite constraint in the second line of (3.41). Without loss of generality, we can replace the parameter set $\{\xi \in \Xi : \xi^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} > \mathbf{u}_i^\top \xi + v_i\}$ by its closure as the corresponding constraint function is continuous in ξ . Note that this parameter set is non-empty because $i \in \mathcal{I}(\mathbf{x})$. Thus, its closure is obtained by replacing the strict inequality in its definition with a weak one.

Using an epigraph reformulation as before, one can then show that the i -th constraint in the second line of (3.41) is satisfied if and only if

$$\begin{aligned} 0 &\geq \max \quad \alpha + \beta^\top \xi - \gamma^\top \rho \\ \text{s.t.} \quad &\xi \in \mathbb{R}^k, \rho \in \mathbb{R}^S \\ &\mathbf{W}\xi \leq \mathbf{w}, \quad \rho_s \geq \mathbf{g}_{st}^\top (\xi - \mu) + h_{st} \quad \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \\ &\xi^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} \geq \mathbf{u}_i^\top \xi + v_i. \end{aligned}$$

By strong LP duality, which holds due to the definition of $\mathcal{I}(\mathbf{x})$, this condition is equivalent to the

following system of linear constraints.

$$\exists \boldsymbol{\Theta}_i \in \mathbb{R}_+^{S \times T}, \boldsymbol{\phi}_i \in \mathbb{R}_+^l, \tau_i \in \mathbb{R}_+ : \begin{cases} 0 \geq \alpha - \tau_i v_i + \tau_i \mathbf{s}_i^\top \mathbf{x} + \mathbf{w}^\top \boldsymbol{\phi}_i + \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} (\mathbf{g}_{st}^\top \boldsymbol{\mu} - h_{st}) \\ \boldsymbol{\beta} + \tau_i \mathbf{T}_i \mathbf{x} = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \Theta_{ist} \mathbf{g}_{st} + \tau_i \mathbf{u}_i + \mathbf{W}^\top \boldsymbol{\phi}_i \\ \gamma_s = \sum_{t \in \mathcal{T}} \Theta_{ist} \quad \forall s \in \mathcal{S} \end{cases} \quad (3.44)$$

The claim then follows by substituting (3.43) and (3.44) into (3.40). $\square$

Proof of Theorem 3.2. We prove the statement in three steps. We first show that λ_0^* bounds the worst-case probability of the chance constraint in (3.2) from above for any $\delta > 0$, and that it converges to that worst-case probability as δ approaches zero. Afterwards, we prove that for any $\delta > 0$, the distribution $\mathbb{P}_\delta^*$ defined in (3.7) is an element of the ambiguity set $\mathcal{P}$. Finally, we show that the satisfaction probability of the chance constraint under distribution $\mathbb{P}_\delta^*$ is bounded above by λ_0^* . The statement of the theorem then follows from these three assertions.

In view of the first step, we note that problem (3.6) is the dual of a perturbed version of problem (3.5) in the statement of Theorem 3.1 where we replace the right-hand side intercepts $\mathbf{v}$ with $\mathbf{v} + \delta \mathbf{e}$. Since strong duality holds, λ_0^* is equal to the worst-case probability of the perturbed chance constraint $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\mathbf{T}(\tilde{\boldsymbol{\xi}}) \mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}}) + \delta \mathbf{e})$, that is, λ_0^* bounds the worst-case probability of the unperturbed chance constraint in (3.2) from above for any $\delta > 0$. For later reference, we also remark that $\lambda_0^* > 0$ since $\mathbf{x}$ is feasible in problem (3.2) and $\varepsilon < 1$. We now show that λ_0^* converges to the worst-case probability of the unperturbed chance constraint in (3.2) as δ approaches zero. To this end, consider the family of optimal values $\{\lambda_0^*(\delta)\}_{\delta > 0}$ of problem (3.6) for different values of δ . By construction, $\lambda_0^*(\delta)$ is upper semi-continuous as it constitutes the partial infimum over upper semi-continuous functions $\delta \rightarrow \mathbb{P}(\mathbf{T}(\tilde{\boldsymbol{\xi}}) \mathbf{x} \leq \mathbf{u}(\tilde{\boldsymbol{\xi}}) + \delta \mathbf{e})$, see [89]. Since $\lambda_0^*(\delta)$ is also non-decreasing in δ , we conclude that $\lambda_0^*(\delta)$ is right-continuous, that is, $\lambda_0^*(\delta)$ indeed converges to $\lambda_0^*(0)$, the worst-case probability of the chance constraint in (3.2), as δ approaches zero.

As for the second step, we note that the feasibility of $(\{\lambda_i^*\}_{i \in \mathcal{I}_0(\mathbf{x})}, \{\chi_i^*\}_{i \in \mathcal{I}_0(\mathbf{x})})$ in (3.6) implies

$$\mathbf{W}\chi_i^* \leq \lambda_i^* \mathbf{w} \iff \mathbf{W}\chi_i^*/\lambda_i^* \leq \mathbf{w} \quad \forall i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* > 0$$

as well as

$$\begin{aligned} \mathbf{W}\chi_i^* \leq \lambda_i^* \mathbf{w} \quad \forall i \in \mathcal{I}_0(\mathbf{x}) &\implies \mathbf{W}(\chi_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \chi_i^*) \leq (\lambda_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \lambda_i^*) \mathbf{w} \\ &\iff \mathbf{W}(\chi_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \chi_i^*)/\lambda_0^* \leq \mathbf{w}, \end{aligned}$$

where the second equivalence holds because $\lambda_0^* > 0$ (see previous paragraph). Thus, the probability distribution $\mathbb{P}_\delta^*$ is indeed supported on Ξ . We can similarly verify that

$$\mathbb{E}_{\mathbb{P}_\delta^*}[\tilde{\boldsymbol{\xi}}] = \lambda_0^* \left(\chi_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \chi_i^* \right) / \lambda_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* > 0} \lambda_i^* (\chi_i^* / \lambda_i^*) = \sum_{i \in \mathcal{I}_0(\mathbf{x})} \chi_i^* = \boldsymbol{\mu},$$

where the first identity is due to the definition of $\mathbb{P}_\delta^*$ in (3.7), and the second identity holds due to the second constraint in (3.6). Moreover, for any $s \in \mathcal{S}$, we have that

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_\delta^*} \left[\max_{t \in \mathcal{T}} \{ \mathbf{g}_{st}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) + h_{st} \} \right] \\ &= \lambda_0^* \max_{t \in \mathcal{T}} \left\{ \mathbf{g}_{st}^\top \left[(\chi_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \chi_i^*) / \lambda_0^* - \boldsymbol{\mu} \right] + h_{st} \right\} + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* > 0} \lambda_i^* \max_{t \in \mathcal{T}} \left\{ \mathbf{g}_{st}^\top (\chi_i^* / \lambda_i^* - \boldsymbol{\mu}) + h_{st} \right\} \\ &= \max_{t \in \mathcal{T}} \left\{ \mathbf{g}_{st}^\top \left(\chi_0^* + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* = 0} \chi_i^* - \lambda_0^* \boldsymbol{\mu} \right) + \lambda_0^* h_{st} \right\} + \sum_{i \in \mathcal{I}(\mathbf{x}) : \lambda_i^* > 0} \max_{t \in \mathcal{T}} \left\{ \mathbf{g}_{st}^\top (\chi_i^* - \lambda_i^* \boldsymbol{\mu}) + \lambda_i^* h_{st} \right\} \\ &\leq \sum_{i \in \mathcal{I}(\mathbf{x}) \cup \{0\}} \max_{t \in \mathcal{T}} \left\{ \mathbf{g}_{st}^\top (\chi_i^* - \lambda_i^* \boldsymbol{\mu}) + \lambda_i^* h_{st} \right\} \leq d_s, \end{aligned}$$

where the first identity follows from the definition of $\mathbb{P}_\delta^*$ in (3.7), the first inequality is due to the subadditivity property of the maximum operator, and the second inequality holds because of the third constraint in (3.6). We thus conclude that $\mathbb{P}_\delta^*$ is indeed contained in $\mathcal{P}$.

To prove the last step, we observe that the realisation $\tilde{\boldsymbol{\xi}} = \chi_i^* / \lambda_i^*$, $i \in \mathcal{I}(\mathbf{x})$ with $\lambda_i^* > 0$, violates the i -th row of the chance constraint in (3.2) since

$$\mathbf{t}_i(\chi_i^* / \lambda_i^*)^\top \mathbf{x} = (\chi_i^* / \lambda_i^*)^\top \mathbf{T}_i \mathbf{x} + \mathbf{s}_i^\top \mathbf{x} \geq \mathbf{u}_i^\top (\chi_i^* / \lambda_i^*) + (v_i + \delta) > \mathbf{u}_i^\top (\chi_i^* / \lambda_i^*) + v_i.$$

Here, the identity follows from the definition of the left-hand side coefficient matrix $\mathbf{T}$, and the two inequalities are due to the penultimate constraint in (3.6) and the fact that $\delta > 0$, respectively. We conclude that the satisfaction probability of the chance constraint is bounded above by λ_0^* . $\square$

Proof of Lemma 3.1. For any fixed $\mathbf{x} \in [0, 1]^n$, we define the sets $\underline{\mathcal{I}}(\mathbf{x}) = \{i \in \mathcal{I} : x_i = 0\}$, $\bar{\mathcal{I}}(\mathbf{x}) = \{i \in \mathcal{I} : x_i = 1\}$ and $\mathcal{I}^o(\mathbf{x}) = \{i \in \mathcal{I} : x_i \in (0, 1)\}$. Thus, we may re-express the worst-case probability in problem (3.9) as

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\tilde{\xi}_i \geq 0 \ \forall i \in \bar{\mathcal{I}}(\mathbf{x}), \ \tilde{\xi}_{n+i} \geq 0 \ \forall i \in \underline{\mathcal{I}}(\mathbf{x}), \ \tilde{\xi}_i \geq 0, \ \tilde{\xi}_{n+i} \geq 0 \ \forall i \in \mathcal{I}^o(\mathbf{x}) \right).$$

As it constitutes a special case of the worst-case probability problem in (3.2), we can use Theorem 3.1 to reformulate the above moment problem as the following tractable LP.

$$\begin{aligned} \max \quad & \alpha + \mathbf{e}^\top \boldsymbol{\beta} - \frac{2\varepsilon}{n} \mathbf{e}^\top \boldsymbol{\gamma} \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^{2n}, \ \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}_0, \boldsymbol{\eta}_0, \boldsymbol{\theta}_i, \boldsymbol{\eta}_i \in \mathbb{R}_+^{2n} & \forall i \in \{1, \dots, 2n\} \\ & \alpha + \mathbf{e}^\top (\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0) \leq 1, \ \boldsymbol{\beta} = \boldsymbol{\theta}_0 - \boldsymbol{\eta}_0, \ \boldsymbol{\gamma} = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_0 \\ & \alpha + \mathbf{e}^\top (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i) \leq 0, \ \boldsymbol{\beta} - \tau_i \mathbf{e}_i = \boldsymbol{\theta}_i - \boldsymbol{\eta}_i, \ \boldsymbol{\gamma} = \boldsymbol{\theta}_i + \boldsymbol{\eta}_i & \forall i \in \bar{\mathcal{I}}(\mathbf{x}) \\ & \alpha + \mathbf{e}^\top (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i) \leq 0, \ \boldsymbol{\beta} - \tau_i \mathbf{e}_{n+i} = \boldsymbol{\theta}_i - \boldsymbol{\eta}_i, \ \boldsymbol{\gamma} = \boldsymbol{\theta}_i + \boldsymbol{\eta}_i & \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\ & \alpha + \mathbf{e}^\top (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i) \leq 0, \ \boldsymbol{\beta} - \tau_i \mathbf{e}_i = \boldsymbol{\theta}_i - \boldsymbol{\eta}_i, \ \boldsymbol{\gamma} = \boldsymbol{\theta}_i + \boldsymbol{\eta}_i & \forall i \in \mathcal{I}^o(\mathbf{x}) \\ & \alpha + \mathbf{e}^\top (\boldsymbol{\theta}_{n+i} - \boldsymbol{\eta}_{n+i}) \leq 0, \ \boldsymbol{\beta} - \tau_{n+i} \mathbf{e}_{n+i} = \boldsymbol{\theta}_{n+i} - \boldsymbol{\eta}_{n+i}, \ \boldsymbol{\gamma} = \boldsymbol{\theta}_{n+i} + \boldsymbol{\eta}_{n+i} & \forall i \in \mathcal{I}^o(\mathbf{x}). \end{aligned} \tag{3.45}$$

Note that this LP is both feasible and bounded because its optimal value represents a probability. In the rest of the proof we investigate two mutually exclusive scenarios: (i) the first inequality in (3.45) is *not binding* at optimality and (ii) the inequality is *binding* at optimality. We argue that the optimal value of (3.45) must be 0 in scenario (i). Indeed, the objective function of the dual of (3.45) consists solely of the dual variable corresponding to the inequality $\alpha + \mathbf{e}^\top (\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0) \leq 1$ because no other constraint has a non-zero right-hand side constant. Under the assumption of scenario (i), complementary slackness implies that this dual variable attains the value 0 at optimality, that is, the optimal value of (3.45) is 0 as well. In scenario (ii), on the other hand, we can replace the first inequality with the equality

$\alpha + \mathbf{e}^\top(\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0) = 1$. By using the other equality constraints to eliminate the variables $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ as well as $\boldsymbol{\theta}_i$ and $\boldsymbol{\eta}_i$ for $i \in \{1, \dots, 2n\}$, (3.45) reduces to

$$\begin{aligned}
\max \quad & \alpha + \mathbf{e}^\top \boldsymbol{\beta} - \frac{2\varepsilon}{n} \mathbf{e}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^{2n}, \quad \boldsymbol{\gamma}, \boldsymbol{\tau} \in \mathbb{R}_+^{2n} \\
& \alpha + \mathbf{e}^\top \boldsymbol{\beta} = 1, \quad \boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq -\boldsymbol{\gamma} \\
& \alpha + \mathbf{e}^\top \boldsymbol{\beta} \leq \tau_i, \quad \boldsymbol{\gamma} \geq \boldsymbol{\beta} - \tau_i \mathbf{e}_i \geq -\boldsymbol{\gamma} \quad \forall i \in \overline{\mathcal{I}}(\mathbf{x}) \\
& \alpha + \mathbf{e}^\top \boldsymbol{\beta} \leq \tau_i, \quad \boldsymbol{\gamma} \geq \boldsymbol{\beta} - \tau_i \mathbf{e}_{n+i} \geq -\boldsymbol{\gamma} \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& \alpha + \mathbf{e}^\top \boldsymbol{\beta} \leq \tau_i, \quad \boldsymbol{\gamma} \geq \boldsymbol{\beta} - \tau_i \mathbf{e}_i \geq -\boldsymbol{\gamma} \quad \forall i \in \mathcal{I}^o(\mathbf{x}) \\
& \alpha + \mathbf{e}^\top \boldsymbol{\beta} \leq \tau_{n+i}, \quad \boldsymbol{\gamma} \geq \boldsymbol{\beta} - \tau_{n+i} \mathbf{e}_{n+i} \geq -\boldsymbol{\gamma} \quad \forall i \in \mathcal{I}^o(\mathbf{x}).
\end{aligned}$$

Next, the first equality of the resulting equivalent LP allows us to replace α with $1 - \mathbf{e}^\top \boldsymbol{\beta}$. Noting that $\boldsymbol{\tau}^* = \mathbf{e}$ at optimality, (3.45) can be further simplified to

$$\begin{aligned}
\max \quad & 1 - \frac{2\varepsilon}{n} \mathbf{e}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \boldsymbol{\beta} \in \mathbb{R}^{2n}, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{2n} \\
& \boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq -\boldsymbol{\gamma} \\
& \beta_i + \gamma_i \geq 1 \quad \forall i \in \overline{\mathcal{I}}(\mathbf{x}) \\
& \beta_{n+i} + \gamma_{n+i} \geq 1 \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& \beta_i + \gamma_i \geq 1, \quad \beta_{n+i} + \gamma_{n+i} \geq 1 \quad \forall i \in \mathcal{I}^o(\mathbf{x}).
\end{aligned}$$

Due to the signs of the objective function coefficients, it is clear that all components of $\boldsymbol{\gamma}$ should be as small as possible at optimality. It is therefore easy to verify that the optimal solution is $\gamma_i^* = \beta_i^* = \frac{1}{2}$ and $\gamma_{n+i}^* = \beta_{n+i}^* = 0$ for $i \in \overline{\mathcal{I}}(\mathbf{x})$, $\gamma_i^* = \beta_i^* = 0$ and $\gamma_{n+i}^* = \beta_{n+i}^* = \frac{1}{2}$ for $i \in \underline{\mathcal{I}}(\mathbf{x})$ and $\gamma_i^* = \beta_i^* = \frac{1}{2}$ and $\gamma_{n+i}^* = \beta_{n+i}^* = \frac{1}{2}$ for $i \in \mathcal{I}^o(\mathbf{x})$. In scenario (ii), the optimal value of (3.45) is therefore given by

$$1 - \frac{2\varepsilon}{n} \left(\sum_{i \in \overline{\mathcal{I}}(\mathbf{x})} \frac{1}{2} + \sum_{i \in \underline{\mathcal{I}}(\mathbf{x})} \frac{1}{2} + \sum_{i \in \mathcal{I}^o(\mathbf{x})} 1 \right) = 1 - \varepsilon \left(2 - \frac{|\{i \in \mathcal{I} : x_i \in \{0, 1\}\}|}{n} \right). \quad (3.46)$$

In summary, the optimal value of (3.45) must be given by the maximum of 0 and (3.46). If $0 \geq (3.46)$, then the assumption of scenario (i) must hold true. Otherwise, scenario (ii) must be correct. $\square$

Proof of Lemma 3.2. The random parameters $\tilde{\xi}$ only appear with non-negative coefficients on the right-hand sides of the constraints inside the probabilistic expression in (3.27). Thus, the worst-case probability in (3.27) does not change if we replace the ambiguity set $\mathcal{P}$ in (3.28) with

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] \geq 1 + c \left(1 - \frac{\varepsilon}{n}\right) \mathbf{e} \right\},$$

where we have relaxed the expectation constraint to an inequality. For the new ambiguity set $\mathcal{P}'$, the worst-case probability in problem (3.27) affords the following semi-infinite dual formulation.

$$\begin{aligned} \max \quad & \alpha + \left(1 + c \left[1 - \frac{\varepsilon}{n}\right]\right) \mathbf{e}^\top \boldsymbol{\beta} \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\ & \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq \mathbb{I}_{[x_i \leq \xi_i, -x_i \leq \xi_{n+i} \quad \forall i \in \mathcal{I}]} \quad \forall \boldsymbol{\xi} \in \Xi \end{aligned} \tag{3.47}$$

Strong duality holds as the Dirac distribution that places all probability mass on $(1 + c[1 - \varepsilon/2n])\mathbf{e}$ satisfies the support and expectation constraints of the primal problem strictly. By construction, the semi-infinite constraint in (3.47) is satisfied if and only if

$$\begin{aligned} \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} &\leq 1 & \forall \boldsymbol{\xi} \in \Xi \\ \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} &\leq 0 & \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \xi_i < x_i \\ \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} &\leq 0 & \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \xi_{n+i} < -x_i. \end{aligned}$$

Note that the second constraint vanishes whenever $x_i = -1$, and a continuity argument allows us to replace the strict inequality with a weak one whenever $x_i > -1$. Likewise, the third constraint vanishes whenever $x_i = 1$, and we can replace the strict inequality with a weak one whenever $x_i < 1$.

Problem (3.47) is thus equivalent to

$$\begin{aligned}
\max \quad & \alpha + \left(1 + c \left[1 - \frac{\varepsilon}{n}\right]\right) \mathbf{e}^\top \boldsymbol{\beta} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 1 \quad \forall \boldsymbol{\xi} \in \Xi \\
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}), \forall \boldsymbol{\xi} \in \Xi : \xi_i \leq x_i \\
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 \quad \forall i \in \overline{\mathcal{I}}(\mathbf{x}), \forall \boldsymbol{\xi} \in \Xi : \xi_{n+i} \leq -x_i,
\end{aligned}$$

where we have introduced the shorthand notations $\underline{\mathcal{I}}(\mathbf{x}) = \{i \in \mathcal{I} : x_i > -1\}$ and $\overline{\mathcal{I}}(\mathbf{x}) = \{i \in \mathcal{I} : x_i < 1\}$ for $\mathbf{x} \in [-1, 1]^n$. Since $\boldsymbol{\beta} \geq \mathbf{0}$, we can replace each semi-infinite constraint with a single constraint where $\boldsymbol{\xi}$ attains its component-wise largest value:

$$\begin{aligned}
\max \quad & \alpha + \left(1 + c \left[1 - \frac{\varepsilon}{n}\right]\right) \mathbf{e}^\top \boldsymbol{\beta} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\
& \alpha + (1 + c) \mathbf{e}^\top \boldsymbol{\beta} \leq 1 \tag{3.48} \\
& \alpha + (1 + c) \mathbf{e}^\top \boldsymbol{\beta} - (1 + c - x_i) \beta_i \leq 0 \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& \alpha + (1 + c) \mathbf{e}^\top \boldsymbol{\beta} - (1 + c + x_i) \beta_{n+i} \leq 0 \quad \forall i \in \overline{\mathcal{I}}(\mathbf{x})
\end{aligned}$$

As in the proof of Lemma 3.1 we can employ a complementary slackness argument to conclude that either the optimal value of (3.48) is 0 or the first constraint in (3.48) is binding. In the second case, we can replace α with $1 - (1 + c) \mathbf{e}^\top \boldsymbol{\beta}$ to obtain the equivalent reformulation

$$\begin{aligned}
\max \quad & 1 - \frac{c\varepsilon}{n} \cdot \mathbf{e}^\top \boldsymbol{\beta} \\
\text{s.t.} \quad & \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\
& 1/(1 + c - x_i) \leq \beta_i \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& 1/(1 + c + x_i) \leq \beta_{n+i} \quad \forall i \in \overline{\mathcal{I}}(\mathbf{x}).
\end{aligned}$$

Since the objective function is strictly monotonously decreasing in $\boldsymbol{\beta}$, this problem has the optimal solution $\beta_i^* := 1/(1 + c - x_i)$ if $i \in \underline{\mathcal{I}}(\mathbf{x})$; $:= 1/(1 + c + x_i)$ if $i \in \overline{\mathcal{I}}(\mathbf{x})$; $:= 0$ otherwise. In summary, the

optimal value of (3.48) is given by the maximum of 0 and

$$1 - \frac{c\varepsilon}{n} \left(\sum_{i \in \underline{\mathcal{I}}(\mathbf{x})} \frac{1}{1 + c - x_i} + \sum_{i \in \bar{\mathcal{I}}(\mathbf{x})} \frac{1}{1 + c + x_i} \right),$$

which concludes the proof. □

Chapter 4

Worst-Case Ambiguous Chance Constraints: Cloud Computing

4.1 Introduction

In this chapter we study the classical profit maximisation problem

$$\begin{aligned} &\text{maximise} && p d(p) - c(d(p)) \\ &\text{subject to} && p \in P, \end{aligned} \tag{4.1}$$

where a seller has control over the price p and wishes to maximise profit. Profit is the difference between total revenue and total cost. We assume there is a demand function, $d(p)$, which indicates how many people are willing to buy at any price p . Thus, $d(p)$ is the quantity we should be producing. The demand function should obey the “law of demand” [78, §4], i.e. $d(p)$ should be non-increasing in p . The total cost in the objective function of (4.1) is a function of the quantity produced. By construction, it is non-increasing in its argument. The set P contains all admissible prices. A non-negativity constraint on p can be explicitly included in P or, as we will see, is implicitly satisfied for most reasonable choices of d . Problems of the type (4.1) have been studied for a long time, and are now a staple of any beginner’s course in microeconomics.

From the beginning, we take the decision variable to be the price, since this is what is set by sellers and observed by customers. This slightly differs from how the topic is presented in the microeconomic literature, see e.g. [78], where the decision variable is the production quantity. However, if $d(p)$ is invertible then both formulations are equivalent.

We can distil useful managerial insights from the above formulation. For example, the first-order necessary conditions for optimality translate to the fact that any optimal price will equate marginal revenue and marginal cost. Our aim here is to derive such insights in more general settings. We first formulate the above problem to include demand uncertainty. We will assume that there is a known underlying relationship between price and demand, which we then perturb with randomness.

After taking a more realistic view on demand uncertainty, we begin to look at both single- and multi-product monopolists. To illustrate our results, we present a novel application from computer science which we now introduce. However, the structure of models we consider will directly apply to many other revenue management problems.

All numerical experiments are performed on a desktop machine with 8GB memory and quad-core 3.0GHz processor, using MOSEK 7.1 [82] solver accessed through the YALMIP [76] interface in MATLAB.

4.2 Cloud Computing

In this section we give some background on cloud computing and show that important pricing problems in cloud computing give rise to optimisation problems of the form (4.1). Cloud computing is an increasingly popular paradigm for obtaining computing resources. Traditionally, computing resources are procured by building a datacentre and equipping it with thousands of servers. There is also a need for networking, power and cooling equipment and for support staff that will setup, maintain and troubleshoot the system. Building a datacentre is complicated and requires a large initial investment, e.g. see [77]. Deciding exactly how much computing capacity this datacentre should provide is difficult, and once the datacentre is built, its capacity becomes essentially fixed. If it is built to specifically handle periods of high load, for example when a new product is launched, then there will be a lot of unused resources most of the time.

To prevent the buildup of underutilised computing resources, one could hire the necessary resources from a cloud computing provider. A cloud computing provider such as Amazon EC2 [2] or Google

Compute Engine [54] offers computing resources directly in an elastic, “pay-as you go”, and on-demand basis. For example, suppose you run a company that launches a new product and expects a much higher than average number of visitors to its website over the next week. Just before you announce this new product, you can ask the provider to provision (on-demand) 10 servers, to which you replicate your website. However, during the launch, you notice that 10 servers are not enough, so you just request 5 more (elasticity/flexibility). After launch, when you see that the load has reverted back to its mean, you can release the servers back to the cloud provider. Throughout this transaction, you paid a flat rate per server per hour, without any up-front costs or contracts. Neither did you physically have to buy and maintain any computing resources. In fact, nowadays it is very common for startup companies to hire all of their computing power from cloud providers, commonly known as infrastructure-as-a-service (IaaS).

It is the cloud provider’s duty to build and manage the physical datacentres. Most cloud computing providers begin by providing this computational power for their primary business, e.g. Amazon’s Marketplace, or Google’s Search Engine. Over time, they start to sell off unused capacity to others on the basis described above. Nowadays there are even companies whose core business model is to provide cloud computing services. We now introduce the parameters and structure of a basic stochastic pricing model for cloud computing services, and show how it can be represented in the form (4.1).

4.3 Stochastic Model

We consider a fictional cloud computing provider whose goal is to select a price $p \in \mathbb{R}_+$ for a cloud computing service, as well as the capacity $c \in \mathbb{R}_+$ to be installed in order to maximise expected profit.

The provider first buys and installs hardware (servers, networking equipment, etc.) into datacentres. This provides c units of computational capacity, which are procured at a total cost of $k(c)$ per unit, and can then be rented out. This provider would typically reserve some capacity for its own use, perhaps for primary business. We expect this to be more or less constant and therefore ignore it in the model. Of course, the provider could purchase extra capacity from itself just like one of its own customers, so we have no loss of generality.

For a price p , a user would receive a unit of capacity for a prespecified amount of time (e.g. one hour). The cloud computing provider’s planning horizon is divided into time intervals of length Δt , and

at each time point t a realisation of demand for services is measured and a decision is made on how much of this demand can actually be served, based on available capacity c .

The cloud computing provider expects that demand is dependent on the price p set. For example, setting a very high price would mean nobody would be willing to buy. The provider also expects uncertainty about future demand as it is unknown. They cannot know how successful and/or computationally demanding the enterprises of its clients will be. We model this demand as $d(p)\tilde{\xi}_t$, where $d(p)$ captures the relationship between price and demand (satisfying the law of demand) and $\{\tilde{\xi}_t\}_{t \in \mathbb{N}}$ is a non-negative stochastic process (with known distribution $\mathbb{Q}$ and $\tilde{\xi}_t \geq 0$ $\mathbb{Q}$ -almost surely for all $t \in \mathbb{N}$) which reflects demand uncertainty. We chose a multiplicative concatenation of underlying demand function and randomness to ensure that all realisations of demand are non-negative, since both multiplied quantities are themselves non-negative. Consistent with law of demand, we assume that $d(p)$ is non-increasing in p , that $d(0) > 0$, and that there is a maximum price $\bar{p}$ such that $d(\bar{p}) = 0$. We can interpret this choke price $\bar{p}$ as the price for which anyone can purchase the equivalent physical capacity themselves, without relying on any cloud provider. Since the cloud computing provider must also choose a positive price, the set of admissible prices is given by the constraint $0 \leq p \leq \bar{p}$.

At time t the cloud computing provider observes a realisation of the stochastic demand ξ_t and receives $p \min\{c, d(p)\xi_t\}$ in revenue, i.e., it only charges for demand which it is able to satisfy with its capacity. If $d(p)\xi_t > c$, then there are some customers who are not served.

The provider will select an on-demand price p that maximises the expected net present value of their revenue stream. Without much loss of generality, an infinite planning horizon may be adopted since the cloud computing provider hopes to be in this business for an indefinite amount of time. By λ_t we denote a discount factor applied to cash flows arising at time t . These factors transform future cashflows to their present values. To ensure that the infinite series is convergent (i.e. the NPV is finite) we require that $\sum_{t=1}^{\infty} \lambda_t$ is finite. We shall see that doing this will allow us to consider (equivalently) a time-averaged version of the model.

As with any technology service where it is impossible to guarantee a deterministic response behaviour, a Service Level Agreement (SLA) [113] is present. It is a contract between the provider and user about the quality of service that *each individual user* can expect. For example, such agreements can include

guarantees on mean time to complete a given task, mean time taken to recover after a service outage, and uptime. In the cloud computing context, the main clause is about uptime — the availability of capacity. In this work, we assume that an SLA requires that over the course of an agreed time horizon (e.g. 1 year), the percentage of time periods (e.g. of 5 minute intervals) when the demand of a specific user is successfully serviced, should exceed an agreed percentage, $(1 - \varepsilon)\%$. For example, if an SLA guarantees that the percentage of minutes in a month that the service is available will be at least 99.95%, then this allows just over 20 minutes of downtime per month for maintenance and/or unexpected service outages.

A naive way to model this as a constraint is to initially allocate a share of total capacity to each user. While conceptually simple, this is unrealistic and too restrictive, as the price p will be set too high and there will be a lot of unused capacity. For example, consider two users, one idle and the other consuming a lot of resources. It makes sense to give the unused capacity to the latter user to prevent breaking the SLA with them. Instead, the more economic solution is to share the whole capacity between all users and enforce the following aggregate constraint

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{d(p)\tilde{\xi}_t > c} \leq \varepsilon \quad \mathbb{Q}\text{-a.s.},$$

where, for a logical expression $\mathcal{E}$, the value of $\mathbb{1}_{\mathcal{E}}$ is 1 if $\mathcal{E}$ is true and 0 otherwise. For a fixed T , the indicator function counts the number of time periods where demand exceeds capacity. This is summed over the whole time period and gives the percentage of time periods where demand exceeds capacity, meaning that at least one customer was not served. This is a conservative approximation for the SLA. Indeed, if this aggregate SLA is violated, then the SLA for at least one user is violated. If the aggregated SLA is enforced, we cannot say precisely what ε each user will actually experience. To do this would mean having to model individual users, and is outside the scope of this work. We note that the limit in the above expression exists since it is bounded above by 1.

If the demand exceeds capacity we should expect to make some form of compensation to the user. Likewise, if we have unused capacity, we incur an opportunity cost per unit since computing capacity cannot be accumulated over time and there are maintenance, electricity and cooling costs which are due even when a server is idle. Also, we should not forget some overall semi-variable cost $k(c)$ depending on the initial capacity procured. Although important, we disregard these features from the exposition so

that it does not become cluttered with notation. However, we note that all analysis still applies, and can easily be adapted.

To summarise, a monopolistic cloud computing service provider would want to solve the following optimisation problem

$$\begin{aligned}
& \text{maximise} && \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{\infty} \lambda_t \left(p \min \left\{ c, d(p) \tilde{\xi}_t \right\} \right) \right] \\
& \text{subject to} && p \in \mathbb{R} \\
& && \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{d(p) \tilde{\xi}_t > c} \leq \varepsilon \quad \mathbb{Q}\text{-a.s.} \\
& && p \geq 0 \quad d(p) \geq 0.
\end{aligned} \tag{4.2}$$

Model (4.2) involves an infinite stream of random cashflows. We now show how (4.2) can be collapsed to an equivalent single-stage stochastic program by a process that can be viewed as time randomisation. At least for now, we also assume that the distribution of the stochastic process $\mathbb{Q}$ is known precisely.

Theorem 4.1. *Problem (4.2) where the stochastic process $\{\tilde{\xi}_t\}_{t \in \mathbb{N}}$ is stationary, ergodic and integrable, is equivalent to the single-stage stochastic program*

$$\begin{aligned}
& \text{maximise} && \Lambda \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ c, d(p) \tilde{\xi} \right\} \right] \\
& \text{subject to} && p \in \mathbb{R} \\
& && \mathbb{Q}(d(p) \tilde{\xi}_1 > c) \leq \varepsilon \quad \mathbb{Q}\text{-a.s.} \\
& && p \geq 0 \quad d(p) \geq 0,
\end{aligned} \tag{4.3}$$

where $\Lambda = \sum_{t=1}^{\infty} \lambda_t$ and $\tilde{\xi}$ is a non-negative random variable with distribution $\mathbb{P}$ defined through

$$\mathbb{P}(\tilde{\xi} \leq \alpha) = \sum_{t=1}^{\infty} \frac{\lambda_t}{\Lambda} \mathbb{Q}(\tilde{\xi}_t \leq \alpha) \quad \forall \alpha \in \mathbb{R}.$$

Proof. We first reformulate the objective function, using a special case of [48, Lemma 2.1], which we reproduce here for completeness. For discount factors $\lambda_t > 0$ we first define a discrete random variable $\tilde{\theta}$ such that $\mathbb{P}(\tilde{\theta} = t) = \lambda_t / \Lambda$ for all $t = 1, \dots, \infty$. The random variable $\tilde{\theta}$ is assumed to be independent

of the stochastic process $\{\tilde{\xi}_t\}_{t \in \mathbb{N}}$. Then, the objective of (4.2) is

$$\begin{aligned} \Lambda \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{\infty} \frac{\lambda_t}{\Lambda} \left(p \min \left\{ c, d(p) \tilde{\xi}_t \right\} \right) \right] &= \Lambda \sum_{t=1}^{\infty} \frac{\lambda_t}{\Lambda} \mathbb{E}_{\mathbb{Q}} \left[p \min \left\{ c, d(p) \tilde{\xi}_t \right\} \right] \\ &= \Lambda \sum_{t=1}^{\infty} \mathbb{P} \left(\tilde{\theta} = t \right) \mathbb{E} \left[p \min \left\{ c, d(p) \tilde{\xi}_t \right\} \mid \tilde{\theta} = t \right] \\ &= \Lambda \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ c, d(p) \tilde{\xi} \right\} \right], \end{aligned}$$

where $\tilde{\xi}$ is a new random variable with distribution defined in the theorem statement, and can be viewed as a mixture of the distributions of each $\{\tilde{\xi}_t\}_{t \in \mathbb{N}}$ where the mixture weights are λ_t/Λ . The first step in the above derivation is due to linearity of expectation, the second due to the law of total probability, and the third follows from the independence of $\tilde{\theta}$ and $\tilde{\xi}_t$.

Due to the assumption on the stochastic process in the statement of the lemma, we can deal with the SLA constraint by a classical ergodic limit theorem [31, Theorem 13.12, page 200]. Thus, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{d(p) \tilde{\xi}_t > c} = \mathbb{E}_{\mathbb{Q}}[\mathbf{1}_{d(p) \tilde{\xi}_1 > c}] = \mathbb{Q}(d(p) \tilde{\xi}_1 > c) \quad \mathbb{Q}\text{-a.s..}$$

Thus, the claim follows. □

The probabilistic constraint in (4.3) can be equivalently written as an upper bound on the demand $d(p)$ if, for example, the distribution function F of $\tilde{\xi}_1$ is monotonically increasing (and therefore with a unique inverse) since

$$\mathbb{Q}(d(p) \tilde{\xi}_1 > c) \leq \varepsilon \iff \mathbb{Q} \left(\tilde{\xi}_1 \leq c/d(p) \right) \geq 1 - \varepsilon \iff F(c/d(p)) \geq 1 - \varepsilon \iff d(p) \leq c/F^{-1}(1 - \varepsilon).$$

If the demand function is itself monotonically decreasing, then the final constraint would be equivalent to a lower bound on the price p . This confirms our intuition. If we want to make sure that enough demand is available, we should not select a price too low. If the capacity c is increased, then this constraint is relaxed. Henceforth, we assume without loss of generality that $\Lambda = 1$ as it is a constant and does not impact the optimal solution. In the next section we address the solution of (4.3).

4.4 Solving the Stochastic Model

A standard existing approach to solving stochastic optimisation problems is via the Sample Average Approximation (SAA) [72]. In this approach, m samples $\xi_1, \dots, \xi_m$ are generated from distribution $\mathbb{P}$ and the following problem is solved:

$$\begin{aligned}
& \text{maximise} && \frac{p}{m} \sum_{i=1}^m \min \{c, d(p)\xi_i\} && = \pi^2, \\
& \text{subject to} && p \in \mathbb{R} \\
& && 0 \leq d(p) \leq c/F^{-1}(1 - \varepsilon) \\
& && 0 \leq p \leq \bar{p}
\end{aligned} \tag{4.4}$$

where π is defined as

$$\begin{aligned}
\pi = & \text{maximise} && \theta/m \\
& \text{subject to} && p \in \mathbb{R}, \mathbf{z} \in \mathbb{R}_+^m \\
& && \left\| \begin{bmatrix} 2\theta \\ p - \mathbf{z}^\top \mathbf{1} \end{bmatrix} \right\|_2 \leq p + \mathbf{z}^\top \mathbf{1} \\
& && \mathbf{z} \leq c\mathbf{1} \\
& && \mathbf{z} \leq d(p)\boldsymbol{\xi} \\
& && 0 \leq d(p) \leq c/F^{-1}(1 - \varepsilon) \\
& && 0 \leq p \leq \bar{p}.
\end{aligned} \tag{4.5}$$

The equality holds by applying the same reformulations as in the proof of Lemma 4.4. We see that the equivalent reformulation (4.5) of the SAA problem (4.4) scales linearly with the number of samples and is a second-order conic program if $d(p)$ is affine.

Since the decision variable is univariate and bounded, an alternative solution method can be applied to solve (4.3) for an arbitrary demand function. The set of admissible prices is discretised and the objective of (4.3) is evaluated for each fixed p , using the sample average approximation.

We show next that $p d(p)$ is necessarily a concave function, by arguing that the second derivative is non-positive. We use a similar argument as in [108]. Suppose there is some exogenous supply of capacity

which lowers the demand experienced by the monopolist cloud provider. The resulting revenue for the cloud computing provider becomes $p d(p + \gamma)$ where γ is the change in price caused by the exogenous supply. The marginal revenue (first derivative) with respect to p is $p d'(p + \gamma) + d(p + \gamma)$. The rate of change of this marginal revenue with respect to γ (and therefore indirectly, the exogenous supply) is $p d''(p + \gamma) + d'(p + \gamma)$, i.e. a further derivative with respect to γ . It is reasonable to assume this quantity is non-positive for all $\gamma \geq 0$, since we would not expect revenue to increase with exogenous supply which we do not sell. We have

$$p d''(p + \gamma) + d'(p + \gamma) \leq 0 \quad \forall \gamma \geq 0 \quad \implies \quad p d''(p + \gamma) + 2d'(p + \gamma) \leq 0 \quad \forall \gamma \geq 0,$$

where the implication follows since $d'(p + \gamma) < 0 \quad \forall \gamma \geq 0$ as demand is a decreasing function. The consequence of the above implication is precisely that the second derivative of revenue is non-positive with respect to p , showing that the revenue function is concave in p . Alternatively, we show next that $p d(p)$ is a concave function in p if either $d(p)$ is concave, or in the regions where $d(p)$ is convex, its “relative curvature” is not too great. Assuming $d(p)$ is decreasing then $p d(p)$ is concave if and only if $-p \frac{d''(p)}{d'(p)} \leq 2$. This inequality is satisfied trivially if $d(p)$ is concave since the second derivative would be non-positive. By standard convex calculus, if $p d(p)$ is concave, then so is the objective function of (4.3), and so (4.3) may be solved by bisection, requiring only the numerical evaluation of the derivative of its objective function, without any sampling. The derivative of the objective of (4.3) is

$$\begin{aligned} \frac{d}{dp} \mathbb{E} \left[p \min \left\{ c, d(p) \tilde{\xi} \right\} \right] &= \frac{d}{dp} \left(p d(p) \int_0^{\frac{c}{d(p)}} \xi F(d\xi) \right) + \frac{d}{dp} \left(p \int_{\frac{c}{d(p)}}^{\infty} F(d\xi) \right) \\ &= \left(\frac{d}{dp} p d(p) \right) \int_0^{\frac{c}{d(p)}} \xi F(d\xi) + \int_{\frac{c}{d(p)}}^{\infty} F(d\xi) \\ &= \left(\frac{d}{dp} p d(p) \right) \int_0^{\frac{c}{d(p)}} \xi F(d\xi) + \left(1 - F\left(\frac{c}{d(p)}\right) \right), \end{aligned}$$

where F is the distribution function for random variable $\tilde{\xi}$ with distribution $\mathbb{P}$, i.e. for all $\alpha \geq 0$ we have $F(\alpha) = \mathbb{P}(\tilde{\xi} \leq \alpha)$.

Commonly used demand functions in the literature include linear $d(p) = a - bp$, power $d(p) = ap^{-(1+b)}$, and exponential $d(p) = a - \exp(bp)$, where $a > 0$ and $b > 0$. We refer the reader to [64] for

survey of widely used price-dependent demand functions.

4.5 Distributionally Robust Model

In general, historical demand data may be scarce. Because of this, in practice, the distribution functions of the random parameters in (4.3) are unlikely to be known. Even if historical data is available, there may be many, strikingly different, distributions from which such data could have come. It is, however, relatively easy to estimate some moments of these distributions such as mean and variability. We therefore consider the following distributionally robust variant of problem (4.3)

$$\begin{aligned}
& \text{maximise} && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ d(p)\tilde{\xi}, c \right\} \right] \\
& \text{subject to} && p \in \mathbb{R} \\
& && \sup_{\mathbb{P} \in \mathcal{Q}} \mathbb{P} \left(d(p)\tilde{\xi} > c \right) \leq \varepsilon \\
& && p \geq 0, \ d(p) \geq 0,
\end{aligned} \tag{4.6}$$

where we have predefined sets of distributions $\mathcal{P}$ and $\mathcal{Q}$ over which we have a second layer of optimisation. This distributionally robust approach can be thought of as a game against nature. For any fixed choice of p , nature selects distributions from the ambiguity sets that would result in the worst possible outcome, i.e. lowest expected revenue and highest probability of not satisfying the service level agreement. As discussed in Chapter 1, this distributionally robust approach not only mitigates the optimiser's curse where we optimise over the wrong distribution due to estimation error, but also such distributionally robust problems are more likely to be tractable.

We first look at the case where $\mathcal{P}$ and $\mathcal{Q}$ contain all univariate distributions supported on the non-negative orthant, and all share a common mean and second moment. To reduce the amount of extra symbols required, we assume $\mathcal{P} = \mathcal{Q}$ and define the ambiguity sets:

$$\mathcal{P} = \mathcal{Q} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) : \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \right] = \mu \quad \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi}^2 \right] = \sigma^2 + \mu^2 \right\},$$

where $\mathcal{P}_0(\mathcal{V})$ is the set of all distributions supported on set $\mathcal{V}$. In the above ambiguity set μ and σ are the mean and standard deviation of historical demand, usually estimated from historical data.

We first consider the objective function and chance constraint separately, and are able to derive equivalent analytic reformulations. We then bring these results together and consider the whole problem (4.6). The proofs for all lemmas can be found in the appendix of this chapter.

4.5.1 Worst-Case Chance Constraint

We first consider the uncertainty quantification problem, i.e. the evaluation of the worst-case probability. Using standard tools from robust optimisation and a case-by-case analysis of the structure of the worst-case distribution we derive the following result

Lemma 4.1. *For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have*

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+)} \mathbb{P}(d(p)\tilde{\xi} > c) &= \inf_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2} \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t. } \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] &= \mu \\
\mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] &= \sigma^2 + \mu^2 \\
\left\| \begin{bmatrix} \beta - \tau_1 \\ \alpha - \gamma \end{bmatrix} \right\|_2 &\leq \alpha + \gamma \\
\left\| \begin{bmatrix} \beta - d(p)\tau_2 \\ \alpha - 1 + c\tau_2 - \gamma \end{bmatrix} \right\|_2 &\leq \alpha - 1 + c\tau_2 + \gamma
\end{aligned}$$

$$= \begin{cases} 1 & \text{if } d(p)\mu > c \\ \min \left\{ \left(1 + \left(\frac{c/d(p) - \mu}{\sigma} \right)^2 \right)^{-1}, \frac{d(p)\mu}{c} \right\} & \text{if } d(p)\mu \leq c. \end{cases}$$

Proof. See Appendix. □

Lemma 4.1 provides two equivalent reformulations of the worst-case probability problem. The first equality, as we will see, shall be useful when we want to include an upper bound on the worst-case probability in an optimisation problem. Indeed, for fixed $d(p)$ evaluating the worst-case probability is equivalent to solving a second-order cone program (SOCP). The second equality, requires no optimisation and therefore gives insight into how each input parameter influences the worst-case probability. For example, the second equality shows that the worst-case probability is quasi-concave in (μ, σ) . We illustrate this relationship graphically in Figure 4.1 where both $d(p)$ and c have been set to 1. As expected, we see pictorially that each super-level set is convex.

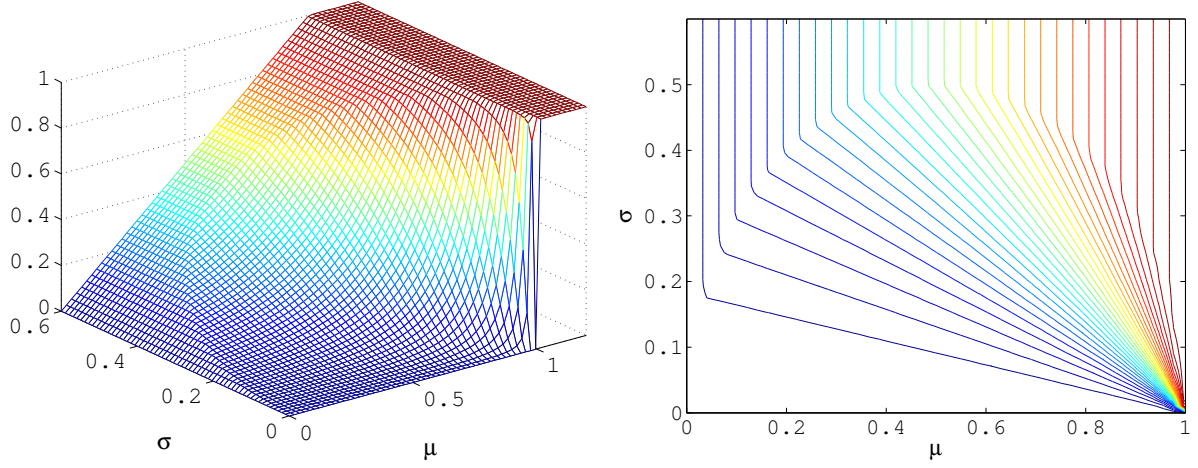


Figure 4.1: The worst-case probability derived in Lemma 4.1 with support constraints as a function of mean μ and standard deviation σ . The right figure visualises the level sets of the worst-case probability.

It is constructive to consider how the results in Lemma 4.1 change if the condition of non-negative support is relaxed. We have shown previously that the inclusion of support constraints in the ambiguity set may render a joint chance constrained problem intractable. In the special case at hand, we are able to analytically calculate the worst-case probability with and without support constraints and may this assess the extend to which support constraints can reduce the model's conservativeness. Using the same techniques from robust optimisation we show the following result

Lemma 4.2. *For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have*

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}_0(\mathbb{R})} \mathbb{P}(d(p)\tilde{\xi} > c) &= \inf_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+} \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t. } \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] &= \mu \\
\mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] &= \sigma^2 + \mu^2 \\
\left\| \begin{bmatrix} \beta \\ \alpha - \gamma \end{bmatrix} \right\|_2 &\leq \alpha + \gamma \\
\left\| \begin{bmatrix} \beta - d(p)\tau \\ \alpha - 1 + c\tau - \gamma \end{bmatrix} \right\|_2 &\leq \alpha - 1 + c\tau + \gamma
\end{aligned}$$

$$= \begin{cases} 1 & \text{if } d(p)\mu > c \\ \left(1 + \left(\frac{c/d(p) - \mu}{\sigma}\right)^2\right)^{-1} & \text{if } d(p)\mu \leq c. \end{cases}$$

Proof. See Appendix. □

The worst-case probability problem without support constraints provides an upper bound on the worst-case probability with support constraints (since we add the support constraint in the latter, thus reducing the feasible set of probability distributions). This is also illustrated in the fact that the dual problem of the latter is a relaxation of the former. In terms of the analytical solution, we have effectively removed case 3 in the proof of Lemma 4.7.

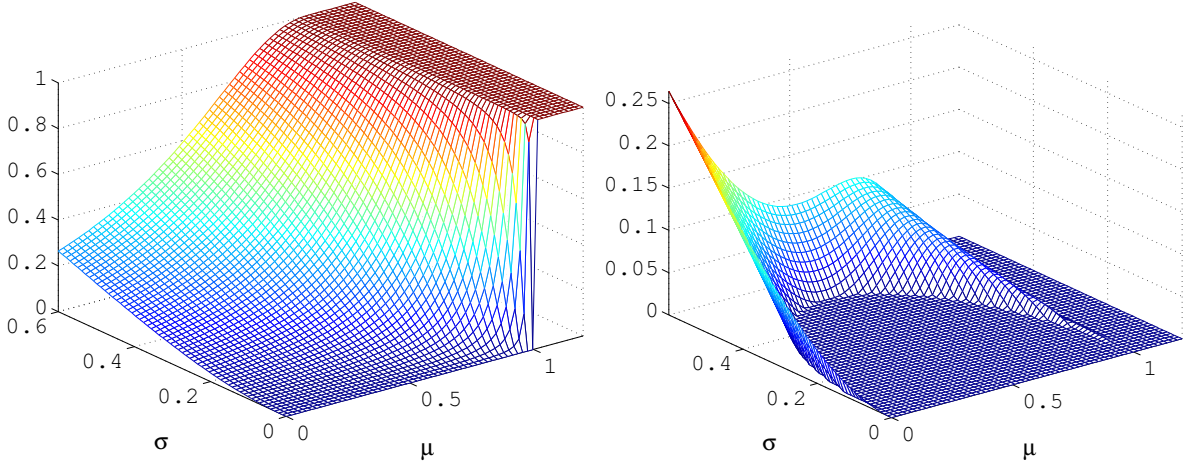


Figure 4.2: The worst-case probability derived in Lemma 4.2 without support constraints as a function of mean μ and standard deviation σ . The right figure shows the difference of the worst-case probabilities corresponding to the models with and without support.

We again graphically illustrate the worst-case probability without support as a function of μ and σ in Figure 4.2.

The comparison of results in Lemmas 4.2 and 4.1 show that it is indeed important to include the support of non-negative random parameters, otherwise the worst-case probability will be over-estimated. We will see later (when we show the optimal decision also), that this leads to an overly conservative decision.

4.5.2 Worst-Case Expected Revenue

Next, we now turn our attention to the worst-case expected profit problem in the objective. Using the robust optimisation techniques developed in Chapters 2 and 3 of this thesis, as well as a seminal result of [102], we arrive at the following result

Lemma 4.3. For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\inf \quad & \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ d(p)\tilde{\xi}, c \right\} \right] = \sup \quad p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) \quad \text{s.t.} \quad \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\
& \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \right] = \mu \quad \left\| \begin{bmatrix} \beta - d(p) + \tau \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\
& \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi}^2 \right] = \sigma^2 + \mu^2 \quad \left\| \begin{bmatrix} \beta \\ \alpha + c - \gamma \end{bmatrix} \right\|_2 \leq \alpha + c + \gamma \\
& = \begin{cases} p \frac{c}{1 + (\sigma/\mu)^2} & \text{if } c \leq \frac{1}{2}d(p) \frac{\mu^2 + \sigma^2}{\mu} \\ \frac{1}{2}p \left(c + d(p)\mu - \sqrt{(c - d(p)\mu)^2 + d(p)^2\sigma^2} \right) & \text{if } c > \frac{1}{2}d(p) \frac{\mu^2 + \sigma^2}{\mu}. \end{cases}
\end{aligned}$$

Furthermore, the worst-case expected revenue is, respectively, bounded above and below by

$$pd(p)\mu \quad \text{and} \quad \begin{cases} pd(p) \left(\mu - \sqrt{\sigma^2 + \mu^2} \right) + pc & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} \geq c \\ pd(p) \left(\mu - \frac{1}{4}d(p)(\sigma^2 + \mu^2)/c \right) & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} < c. \end{cases}$$

Proof. See Appendix. □

Lemma 4.3 shows that for fixed p evaluating the worst-case expectation is equivalent to solving an SOCP. The second, analytical, equality in the statement of Lemma 4.3 again gives insight into the behaviour of the worst-case expectation function as the input parameters are varied. The non-linear function $\frac{1}{2}p \left(c + d(p)\mu - \sqrt{(c - d(p)\mu)^2 + d(p)^2\sigma^2} \right)$ given there is concave only on the region for which it is valid, i.e. whenever $c > \frac{1}{2}d(p) \frac{\mu^2 + \sigma^2}{\mu}$. Overall however, the worst-case expected revenue function is concave since this non-concavity is eliminated with the linear piece $p \frac{c}{1 + (\sigma/\mu)^2}$. The results of Lemma 4.3 can be illustrated graphically in Figure 4.5.

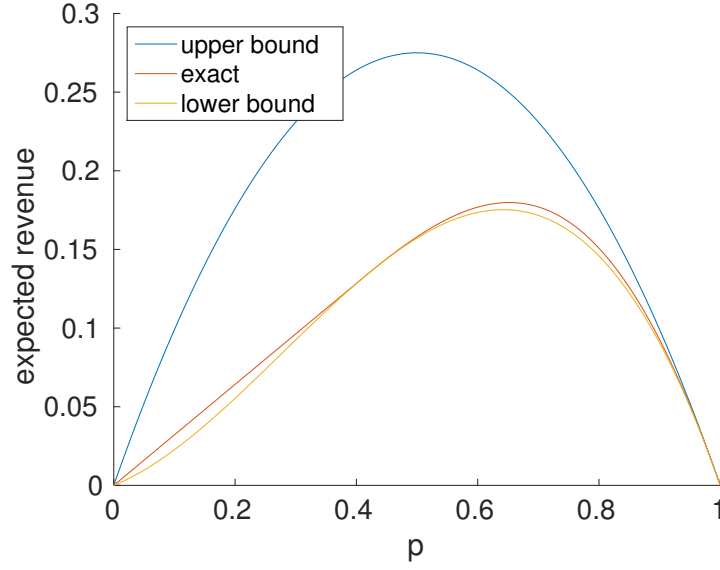


Figure 4.3: The worst-case expected revenue function (non-negative support) from Lemma 4.3 as a function of price p , with $d(p) = 1 - p$, $\mu = 1.1$, $c = 1$, and $\sigma = 1.6$.

Figure 4.3 shows how varying the price affects the worst-case expected revenue, as well as the two bounds in Lemma 4.3. In this case, it is useful to have upper and lower bounds, even though the exact formula is available, since the bounds are simpler in structure, and therefore easier to distil insights from. The upper bound is derived by replacing the random variable $\tilde{\xi}$ with its mean value μ . Doing this leads to a significant over-estimation of the expected revenues, a classical example of the flaw of averages. The lower bound is very close to the true value since the loss of exactness is introduced after the ambiguity set with support constraints has been taken into account, see proof of Lemma 4.3 for details.

4.5.3 Single Product Monopolistic Model

We are now ready to combine the results of the previous two sections to reformulate the complete model (4.6) as an explicit convex optimisation problem.

Lemma 4.4. *For $\varepsilon \in (0, 1)$ we have*

$$\begin{aligned}
 \text{maximise} \quad & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ d(p)\tilde{\xi}, c \right\} \right] = \text{maximise} \quad \frac{1}{2}p \left(c + d(p)\mu - \sqrt{(c - d(p)\mu)^2 + d(p)^2\sigma^2} \right) = \pi^2, \\
 \text{subject to} \quad & p \in \mathbb{R} \quad \quad \quad \text{subject to} \quad p \in \mathbb{R}_+ \\
 & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(d(p)\tilde{\xi} > c \right) \leq \varepsilon \quad \quad \quad 0 \leq d(p) \leq c \max \left\{ \frac{1}{\mu + \sigma\kappa(\varepsilon)}, \frac{\varepsilon}{\mu} \right\} \\
 & p \geq 0, \quad d(p) \geq 0
 \end{aligned}$$

where $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$ and π is defined as

$$\pi = \text{maximise } \theta$$

$$\text{subject to } p \in \mathbb{R}_+, \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+, u \in \mathbb{R}_+, \theta \in \mathbb{R}_+$$

$$\begin{aligned} u &= -\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2) \\ \left\| \begin{bmatrix} 2\theta \\ p - u \end{bmatrix} \right\|_2 &\leq p + u \\ \left\| \begin{bmatrix} \beta - d(p) + \tau \\ \alpha - \gamma \end{bmatrix} \right\|_2 &\leq \alpha + \gamma \\ \left\| \begin{bmatrix} \beta \\ \alpha + c - \gamma \end{bmatrix} \right\|_2 &\leq \alpha + c + \gamma \\ 0 \leq d(p) &\leq c \max \left\{ \frac{1}{\mu + \sigma\kappa(\varepsilon)}, \frac{\varepsilon}{\mu} \right\}. \end{aligned}$$

Proof. See Appendix. □

The first equality in Lemma 4.4 simply combines both analytic results of the worst-case expectation and worst-case chance constraint. However, optimising over p is not possible directly with off-the-shelf tools since the objective function is non-convex over $p \in \mathbb{R}$ and not in a standard form supported by these tools. We see from the second equality in Lemma 4.4 that maximising the square root of the worst-case expectation, subject to the worst-case chance constraint is equivalent to solving an SOCP, if the demand function $d(p)$ is concave.

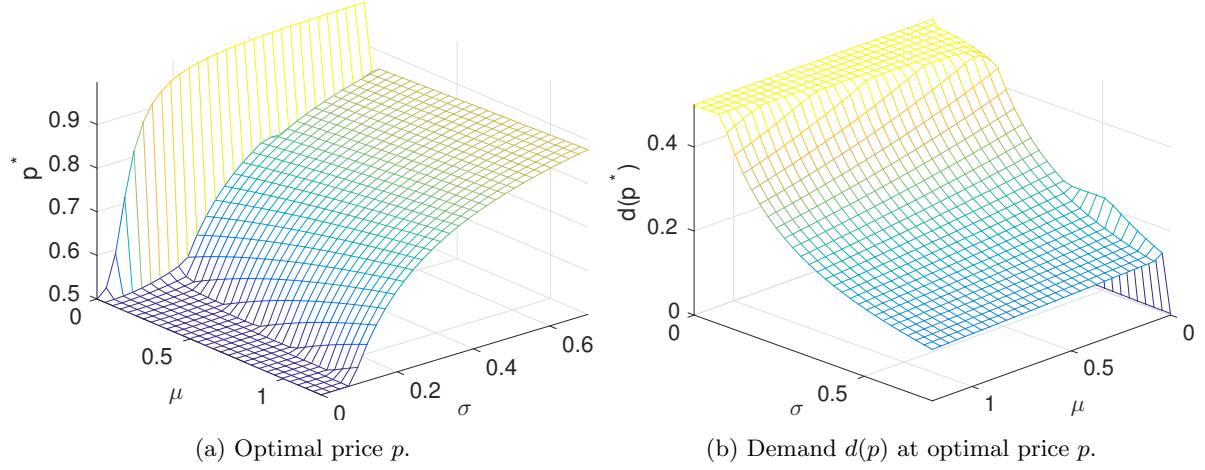


Figure 4.4: Visualisations of the optimal price p and demand as a function of the distribution parameters μ and σ from Lemma 4.4. In both cases $d(p) = 1 - p$, $c = 1$, and $\varepsilon = 1\%$.

Figure 4.4 show the optimal price as a function of the parameters of the ambiguity set μ and σ , with a simple linear demand function with unit elasticity of demand. Figure 4.5 shows how the worst-case expected revenue is affected by μ and σ , with the service level agreement constraint.

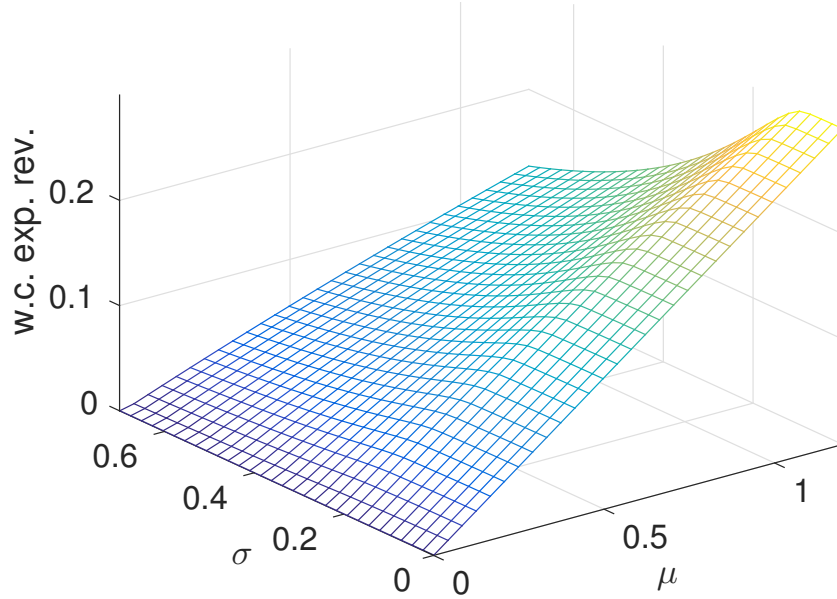


Figure 4.5: The optimal worst-case expected revenue function (non-negative support) from Lemma 4.4 as a function of mean μ and standard deviation σ , with $d(p) = 1 - p$, $c = 1$, and $\varepsilon = 1\%$.

Figure 4.5 confirms our intuition. The worst-case expected revenue is increasing in the mean μ . The higher the mean demand, the more revenue we get in the worst-case. The worst-case expected revenue is decreasing in the standard deviation σ . The less uncertainty there is about the demand, the less conservative the worst-case revenue is.

4.6 Multi-Product Monopolist

In this section we extend the single-product model to a multi-product setting. In the cloud computing setting, products can be memory, processing speed, persistent storage, network bandwidth, etc. Alternatively, single products can be bundled up into composite products such as *general-purpose*, *compute*, or *memory* optimised instance types which consist of different combinations of the base attributes. The *compute* and *memory* optimised instances are available when more processing speed and memory, respectively, are required relative to the *general-purpose* configuration. We still assume that a single monopolist has control over all prices, and can directly extend the single-product model into the following distributionally robust multi-product model

$$\begin{aligned}
& \text{maximise} && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^n p_i \min \left\{ c_i, d_i(\mathbf{p}) \tilde{\xi}_i \right\} \right] \\
& \text{subject to} && \mathbf{p} \in \mathbb{R}^n \\
& && \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(d_i(\mathbf{p}) \tilde{\xi}_i > c_i \right) \leq \varepsilon_i \quad \forall i = 1, \dots, n \\
& && \mathbf{p} \geq \mathbf{0}, \quad \mathbf{d}(\mathbf{p}) \geq \mathbf{0},
\end{aligned} \tag{4.7}$$

where $\mathcal{P}$ is the set of all non-negative joint distributions defined through first- and second-order moments

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+^n) : \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \right] = \boldsymbol{\mu} \quad \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \tilde{\xi}^\top \right] = \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top \right\}.$$

For each product we again have a desired service-level ε_i and capacity c_i , which are elements of the vector of service-level parameters $\boldsymbol{\varepsilon} \in (0, 1)^n$ and vector of capacities $\mathbf{c} \in \mathbb{R}_{++}^n$. As in the single product model, we assume that prices affect demands. For the ambiguity set to be non-empty we require that the second order moment matrix $\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top$ is completely positive [16], i.e. decomposable into the square of a matrix containing non-negative elements. Additionally, the second order moment matrix should be on the interior of the completely positive cone for strong duality to hold. We assume a linear relationship between price and demand

$$\mathbf{d}(\mathbf{p}) = \mathbf{d}_0 + \mathbf{D}\mathbf{p}, \tag{4.8}$$

where $\mathbf{p} \in \mathbb{R}^n$ is the vector of prices, $\mathbf{d}_0 \in \mathbb{R}_+^n$ is the maximum demand when all prices are simultaneously set to zero, and matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ can be interpreted as the cross-price elasticities of demand. The linear demand function is simple in structure, easy to interpret and to estimate the required parameters. If $\mathbf{D}_{ij} < 0$ then products i and j are complements, i.e. an increase in the price of product i leads to a decrease in demand of product j . Alternatively, if $\mathbf{D}_{ij} > 0$, then products i and j are substitutes, i.e. an increase in the price of product i leads to an increase in the demand of product j . The matrix $\mathbf{D}$ need not be symmetric or negative-definite. Since the demand of product i , $d_i(\mathbf{p})$, is expected to follow the law of demand, i.e. is decreasing with respect to its price p_i we have $\mathbf{D}_{ii} \leq 0$. Another reasonable assumption is that the set of all admissible prices $\{\mathbf{p} \in \mathbb{R}^n : \mathbf{p} \geq \mathbf{0} \text{ and } \mathbf{d}(\mathbf{p}) \geq \mathbf{0}\}$ is bounded, otherwise someone would always be willing to buy no matter how high the price is set. This assumption implies that there is an admissible $\mathbf{p}$ such that the demand for *all* products is simultaneously zero, i.e. $\mathbf{d}(\mathbf{p}) = \mathbf{0}$ and that the deterministic revenue function $\mathbf{p}^\top \mathbf{d}(\mathbf{p})$ is bounded above. We make no further assumptions about the matrix $\mathbf{D}$. One could argue that we expect the off-diagonal elements of $\mathbf{D}$ to be much smaller relative to the diagonal elements, small enough for the negative of $\frac{1}{2}(\mathbf{D} + \mathbf{D}^\top)$ to be diagonally dominant, and therefore $\frac{1}{2}(\mathbf{D} + \mathbf{D}^\top)$ to be negative definite. This would imply that the deterministic revenue function is concave in $\mathbf{p}$. We refer the reader to [64] and [116] for further examples of multi-product demand functions and their properties.

Using the same techniques from robust optimisation developed in Chapters 2 and 3, we can derive the results for (4.7). Lemma 4.5 reformulates the worst-case expectation in the objective function of (4.7) and Lemma 4.6 the worst-case chance constraints in (4.7).

Lemma 4.5.

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^n p_i \min \left\{ c_i, d_i(\mathbf{p}) \tilde{\xi}_i \right\} \right] = \sup \quad & \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\mu} - \langle \mathbf{A}, \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top \rangle \\ \text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{S}^n, \mathbf{z} \in \mathbb{R}_+^n, \mathbf{x} \in \mathbb{R}_+^n \\ & \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} \leq \sum_{i=1}^n \min \{ p_i c_i, z_i \xi_i \} \quad \forall \boldsymbol{\xi} \in \mathbb{R}_+^n \\ & x_i^2 \leq \min \{ p_i d_i(\mathbf{p}), z_i \} \quad \forall i = 1, \dots, n \\ & \mathbf{z}^\top \mathbf{e} - \mathbf{x}^\top \mathbf{x} \leq 0. \end{aligned}$$

Proof. Direct application of techniques used in Lemma 4.1. □

Lemma 4.6.

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(d_i(\mathbf{p}) \tilde{\xi}_i > c_i \right) = \inf \quad & \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\mu} + \langle \mathbf{A}, \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top \rangle \\
\text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{S}^n \\
& \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\xi} + \boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}_+^n \\
& \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\xi} + \boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} \geq 1 \quad \forall \boldsymbol{\xi} \in \mathbb{R}_+^n : d_i(\mathbf{p}) \xi_i \geq c_i.
\end{aligned} \tag{4.9}$$

Proof. Direct application of techniques used in Lemma 4.4. □

In the multi-product setting, we do not fare so well with an equivalent reformulation. There are several features of (4.7), each of which independently makes the problem difficult:

- (i) Non-negative support. We have shown earlier that problems with support constraints are, in general, provably difficult. Together with the equality constraint for the second-order moment in the ambiguity set $\mathcal{P}$, this difficulty manifests itself in the dual problems through the semi-infinite constraints. For example, the first constraint in the reformulation in Lemma 4.6 requires the matrix $\mathbf{A}$ to be copositive, i.e. the quadratic form should be non-negative over the non-negative orthant. Although convex, optimisation problems over the copositive cone are in general NP-hard [16]. However, surprisingly, there is an equivalent semidefinite characterisation of the copositive cone for dimension no greater than 4 [36]. A symmetric matrix of dimension no greater than 4 is copositive if and only if it can be decomposed into the sum of a positive semidefinite and non-negative matrix. We can therefore consider a small number of products, or change the second-order moment constraint in the ambiguity set to $\mathbb{E}_{\mathbb{P}} \left[\tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^\top \right] \preceq \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top$ for a conservative approximation. This forces the dual variable $\mathbf{A}$ in both the worst-case expectation and worst-case probability problems to be positive semi-definite. This would allow the application of previously seen convex duality techniques, and the semi-infinite constraints would be equivalent to linear matrix-inequalities.
- (ii) Sum of point-wise minima in the worst-case expectation. In the dual problem, this manifests itself by requiring the bounding of a quadratic function above by a sum of point-wise minima. This is convex, since we may distribute each minimum over two constraints exactly as we did in the

single product case. However in the multi-product setting, this will introduce 2^n constraints. Such problems are, in general, NP-hard [57]. Again, the full expansion would work for a small number of products. We can also conservatively approximate this constraint with quadratic decision rules in the following way, by finding n quadratic functions q_i such that

$$\lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} \leq \sum_{i=1}^n q_i(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}_+^n$$

$$q_i(\boldsymbol{\xi}) \leq \min \{p_i c_i, z_i \xi_i\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}_+^n \quad \forall i = 1, \dots, n,$$

where $q_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \mathbf{A}_i \boldsymbol{\xi} + \boldsymbol{\xi}^\top \mathbf{a}_i + \alpha_i$. This would result in the requirement of $n+1$ quadratic functions being non-negative over the non-negative orthant.

- (iii) Bilinearity between price and demand. This manifests itself in the dual of the worst-case expectation problem (4.9) by the non-convex constraint $\mathbf{z}^\top \mathbf{e} - \mathbf{x}^\top \mathbf{x} \leq 0$. It requires a concave function to be bounded above by zero. Alternating convex optimisation can be applied in this case, meaning we optimise the price of one product in turn, keeping prices of other products fixed. Although there are no guarantees of even a local optimum with this procedure, we have found it performs well in practice. For a small number of products a global optimum can be obtained using spatial branch and bound techniques, since all variables are bounded. Indeed any linearisation of the constraint $\mathbf{z}^\top \mathbf{e} - \mathbf{x}^\top \mathbf{x} \leq 0$ guarantees to give a conservative solution, since the left-hand-side is concave. Additionally, specialised procedures [37, 61] can be employed which guarantee at least a locally optimal solution for difference of convex functions constraints such as this, along with sufficient conditions for global optimality [70].

We proceed in an alternate way by approximating the expected revenue objective function with

$$\mathbb{E}_{\mathbb{P}} \left[\min \left\{ c_i, d_i(\mathbf{p}) \tilde{\xi}_i \right\} \right] \approx \mathbb{E}_{\mathbb{P}} \left[d_i(\mathbf{p}) \tilde{\xi}_i \right]. \quad (4.10)$$

This is justified since we impose constraints which ensure that $\mathbb{P} \left(d_i(\mathbf{p}) \tilde{\xi}_i > c_i \right)$ is small for all i . Since we have closed-form solutions for the single-product service-level constraints we additionally relax the ambiguity sets in the constraints to be consistent only with marginal first- and second-order moments.

We therefore consider the following simplified multi-product model

$$\begin{aligned}
& \text{maximise} && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[d_i(\mathbf{p}) \tilde{\xi}_i \right] \\
& \text{subject to} && \mathbf{p} \in \mathbb{R}^n \\
& && \sup_{\mathbb{P} \in \mathcal{Q}_i} \mathbb{P} \left(d_i(\mathbf{p}) \tilde{\xi}_i > c_i \right) \leq \varepsilon_i \quad \forall i = 1, \dots, n \\
& && \mathbf{p} \geq \mathbf{0}, \quad \mathbf{d}(\mathbf{p}) \geq \mathbf{0},
\end{aligned} \tag{4.11}$$

where $\mathcal{Q}_i$ are ambiguity sets for the marginal distributions of $\tilde{\xi}_i$ consistent with the marginal second-order moment information in $\mathcal{P}$:

$$\mathcal{Q}_i = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) : \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \right] = \mu_i \quad \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi}^2 \right] = \Sigma_{ii}^2 + \mu_i^2 \right\}.$$

Under these assumptions, (4.11) is equivalent to the following optimisation problem.

$$\begin{aligned}
& \text{maximise} && \sum_{i=1}^n p_i d_i(\mathbf{p}) \mu_i \\
& \text{subject to} && \mathbf{p} \in \mathbb{R}^n \\
& && \mathbf{p} \geq \mathbf{0}, \quad \mathbf{0} \leq \mathbf{d}(\mathbf{p}) \leq \bar{\mathbf{d}},
\end{aligned} \tag{4.12}$$

where $\bar{d}_i = c_i \max \left\{ \frac{1}{\mu_i + \sigma_i \kappa(\varepsilon_i)}, \frac{\varepsilon_i}{\mu_i} \right\}$ is the upper bound on the demand for each product such that the chance constraint is satisfied, i.e. the upper bound for the demand given in Lemma 4.1. Even though (4.12) is still a non-convex quadratic maximisation problem subject to a set of linear constraints, there are several options available to solve such problems. QuadProgBB [24] and CPLEX (since version 12.3) support solving these types of problems to global optimality. Additionally one may formulate an equivalent MILP using the KKT conditions of the above problem [17], again guaranteeing global optimality.

For numerical experiments, we consider a two-product setting and show how the optimal price and worst-case expected revenue are affected by the cross price elasticity of demand between the two products.

In Figure 4.6 we show the solution of (4.12) in a two product setting with parameters

$$\mu = \begin{bmatrix} 0.7 \\ 0.8 \end{bmatrix} \quad \sigma = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \quad c = \begin{bmatrix} 0.1 \\ 0.6 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} 10\% \\ 10\% \end{bmatrix} \quad d(p) = \begin{bmatrix} 0.3 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & x \\ x & -0.7 \end{bmatrix},$$

where x , the cross-price elasticity, is varied between -0.1 (complements) and 0.1 (substitutes).

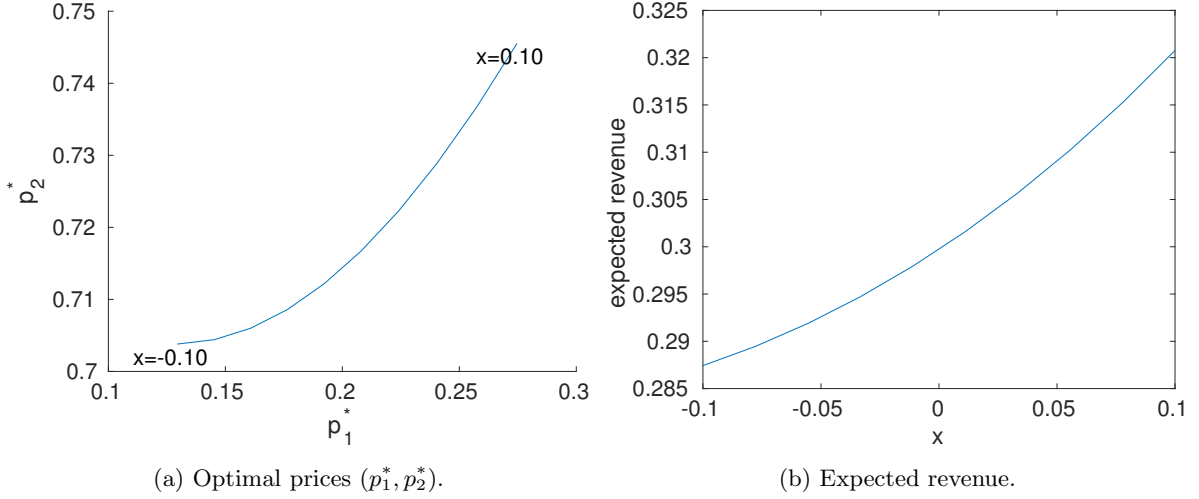


Figure 4.6: Optimal prices (p_1^*, p_2^*) and expected revenue as functions of the cross-price elasticity of demand $-0.1 \leq x \leq 0.1$, for a two-product monopolist.

Figure 4.6 confirms a commonly observed phenomenon in microeconomics. As the cross-price elasticity is varied from $x = -0.1$ (complements) to $x = 0.1$ (substitutes) the expected revenue increases. Thus, it is preferable to make the products offered to be as close substitutes as possible.

4.7 Conclusion

In this chapter, we have considered a novel problem where worst-case ambiguous chance constraints fit naturally into the model. We began with a multi-stage stochastic model which was collapsed to an equivalent single-stage stochastic model using the time-randomisation technique. Next, we motivated why the distributionally robust approach is required to avoid over-estimating the expected revenue.

We then provided both second-order conic and closed form solutions to the resulting single-product pricing problem. We showed that the former gives insight into precisely how the optimal solution will change as input parameters such as capacity, mean, standard of demand are varied. The latter will be

used to numerically solve the pricing problem. We also considered a multi-product model and showed not only the inherent intractabilities, but how this problem can be approximated and still solved effectively.

4.8 Appendix

Lemma (4.1). For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\sup \quad & \mathbb{P} \left(d(p)\tilde{\xi} > c \right) &= \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) & \text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2 \\
& \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi} \right] = \mu & & \left\| \begin{bmatrix} \beta - \tau_1 \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\
& \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi}^2 \right] = \sigma^2 + \mu^2 & & \left\| \begin{bmatrix} \beta - d(p)\tau_2 \\ \alpha - 1 + c\tau_2 - \gamma \end{bmatrix} \right\|_2 \leq \alpha - 1 + c\tau_2 + \gamma \\
& & & = \begin{cases} 1 & \text{if } d(p)\mu > c \\ \min \left\{ \left(1 + \left(\frac{c/d(p) - \mu}{\sigma} \right)^2 \right)^{-1}, \frac{d(p)\mu}{c} \right\} & \text{if } d(p)\mu \leq c. \end{cases}
\end{aligned}$$

Proof. The second equality follows from Lemma 4.7 by applying it with $s \left(\frac{d(p)}{c}\mu, \frac{d(p)}{c}\sigma \right)$ where function s is defined in the statement of the Lemma 4.7. For the first equality, we begin by writing the probability operator in the objective as an expectation. Since

$$\mathbb{P} \left(d(p)\tilde{\xi} > c \right) = \mathbb{E}_{\mathbb{P}} \left[\mathbf{1}_{d(p)\tilde{\xi} > c} \right],$$

the dual of the inf in the above worst-case expectation problem is

$$\begin{aligned}
\inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \\
& \alpha + \beta\xi + \gamma\xi^2 \geq \mathbf{1}_{d(p)\xi > c} \quad \forall \xi \in \mathbb{R}_+.
\end{aligned}$$

We next eliminate the indicator function by decomposing the semi-infinite constraint into two smooth

semi-infinite constraints.

$$\begin{aligned}
& \inf \quad \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \\
& \quad \alpha + \beta\xi + \gamma\xi^2 \geq 0 \quad \forall \xi \in \mathbb{R}_+ \\
& \quad \alpha + \beta\xi + \gamma\xi^2 \geq 1 \quad \forall \xi \in \mathbb{R}_+ : d(p)\xi > c.
\end{aligned}$$

The assumed condition $d(p) \geq 0$ is sufficient to ensure that the second semi-infinite constraint is not vacuously satisfied, i.e. it is always possible to find ξ large enough to make the set $\{\xi \in \mathbb{R} : \xi \geq 0 \wedge d(p)\xi > c\}$ non-empty. Since the quadratic function in ξ is continuous, it is possible to apply a continuity argument to replace the strict inequality with its non-strict variant. It is also necessary that $\gamma \geq 0$. Otherwise we may always select ξ large enough to violate the first semi-infinite constraint. This step ensures that both semi-infinite constraints are (when viewed as minimisations) convex with respect to ξ . This then allows the application of the Farkas Lemma to obtain the following equivalent reformulation

$$\begin{aligned}
& \inf \quad \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2 \\
& \quad \alpha + (\beta - \tau_1)\xi + \gamma\xi^2 \geq 0 \quad \forall \xi \in \mathbb{R} \\
& \quad \alpha + c\tau_2 + (\beta - d(p)\tau_2)\xi + \gamma\xi^2 \geq 1 \quad \forall \xi \in \mathbb{R}.
\end{aligned}$$

Writing the semi-infinite constraints as a quadratic form, we find

$$\begin{aligned}
& \inf \quad \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2 \\
& \quad \begin{bmatrix} \xi \\ 1 \end{bmatrix}^\top \begin{bmatrix} \gamma & \frac{1}{2}(\beta - \tau_1) \\ \frac{1}{2}(\beta - \tau_1) & \alpha \end{bmatrix} \begin{bmatrix} \xi \\ 1 \end{bmatrix} \geq 0 \quad \forall \xi \in \mathbb{R} \\
& \quad \begin{bmatrix} \xi \\ 1 \end{bmatrix}^\top \begin{bmatrix} \gamma & \frac{1}{2}(\beta - d(p)\tau_2) \\ \frac{1}{2}(\beta - d(p)\tau_2) & \alpha + c\tau_2 - 1 \end{bmatrix} \begin{bmatrix} \xi \\ 1 \end{bmatrix} \geq 0 \quad \forall \xi \in \mathbb{R}.
\end{aligned}$$

By a case analysis, it is possible to show that both the above semi-infinite constraints may be homogenised and are therefore equivalent to a semidefinite constraint, i.e.

$$\begin{aligned} \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2 \\ & \begin{bmatrix} \gamma & \frac{1}{2}(\beta - \tau_1) \\ \frac{1}{2}(\beta - \tau_1) & \alpha \end{bmatrix} \succcurlyeq 0 \quad \begin{bmatrix} \gamma & \frac{1}{2}(\beta - d(p)\tau_2) \\ \frac{1}{2}(\beta - d(p)\tau_2) & \alpha + c\tau_2 - 1 \end{bmatrix} \succcurlyeq 0. \end{aligned}$$

We next use the following identity [75]

$$x_0^2 \leq x_1 x_2, \quad x_1 \geq 0, \quad x_2 \geq 0 \iff \begin{bmatrix} x_1 & x_0 \\ x_0 & x_2 \end{bmatrix} \succcurlyeq 0 \iff \left\| \begin{bmatrix} 2x_0 \\ x_1 - x_2 \end{bmatrix} \right\|_2 \leq x_1 + x_2, \quad (4.13)$$

to write the 2×2 matrix semi-definite constraints as second-order cone constraints.

$$\begin{aligned} \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \boldsymbol{\tau} \in \mathbb{R}_+^2 \\ & \left\| \begin{bmatrix} \beta - \tau_1 \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\ & \left\| \begin{bmatrix} \beta - d(p)\tau_2 \\ \alpha - 1 + c\tau_2 - \gamma \end{bmatrix} \right\|_2 \leq \alpha - 1 + c\tau_2 + \gamma. \end{aligned}$$

Similar formulations for the Chebyshev probability bounds can be found in [14, §7.4.1]. $\square$

The next Lemma is an auxiliary result required for the proof of Lemma 4.2.

Lemma 4.7. *Define the function*

$$\begin{aligned} s(\mu, \sigma) = \sup \quad & \mathbb{P}(\tilde{\xi} > 1) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] = \sigma^2 + \mu^2. \end{aligned} \quad (4.14)$$

Then, for $\mu > 0$ and $\sigma > 0$, we have

$$s(\mu, \sigma) = \begin{cases} 1 & \text{if } \mu > 1 \\ \min \left\{ \frac{\sigma^2}{\sigma^2 + (1 - \mu)^2}, \mu \right\} & \text{if } \mu \leq 1. \end{cases}$$

Proof. The techniques we apply are the subject of [102]. For convenience define

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) : \mathbb{E}_{\mathbb{P}} [\tilde{\xi}] = \mu \quad \mathbb{E}_{\mathbb{P}} [\tilde{\xi}^2] = \sigma^2 + \mu^2 \right\}.$$

Rewriting the objective function of (4.14) as an expectation of an indicator function and dualising the corresponding moment problem in the same way as in the proof of Lemma 4.1, we find

$$\begin{aligned} \inf \quad & \lambda_0 + \lambda_1 \mu + \lambda_2 (\mu^2 + \sigma^2) \\ \text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \lambda_1 \in \mathbb{R}, \lambda_2 \in \mathbb{R} \\ & \lambda_0 + \lambda_1 \xi + \lambda_2 \xi^2 \geq \mathbb{1}_{\xi > 1} \quad \forall \xi \geq 0. \end{aligned}$$

For convenience letting $z(\xi) = \lambda_0 + \lambda_1 \xi + \lambda_2 \xi^2$, the dual problem can be written as

$$\begin{aligned} \inf \quad & \mathbb{E}_{\mathbb{P}} [z(\xi)] \\ \text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \lambda_1 \in \mathbb{R}, \lambda_2 \in \mathbb{R} \\ & z(\xi) \geq \mathbb{1}_{\xi \geq 1} \quad \forall \xi \geq 0. \end{aligned} \tag{4.15}$$

Note that the strict inequality is replaced by a non-strict one due to the continuity of polynomials. We proceed by constructing a primal feasible distribution $\mathbb{P}^*$ as well as dual feasible quadratic function $z(\xi)$ such that primal and dual objectives are equal. By weak duality these feasible solutions must be optimal in the respective problems. We will show that $\mathbb{P}^*$ is a discrete distribution. Discreteness of the worst-case distribution also holds in general. Intuitively, the dual problem may be thought of finding the locations of the probability masses in $\mathbb{R}_+$, while the primal problem assigns probability mass to $\mathbb{R}_+$.

We note that the objective values of the primal and dual semi-infinite problems coincide only for any convex combination of point masses situated at the points in $\{\xi \in \mathbb{R}_+ : z(\xi) = \mathbb{1}_{\xi \geq 1}\}$. Therefore, given

a primal feasible distribution supported on $\{\xi \in \mathbb{R}_+ : z(\xi) = \mathbb{1}_{\xi \geq 1}\}$, for some dual feasible quadratic function $z(\xi)$, the objective values of the primal and dual problem would be equated directly, and thus feasible solution would also be optimal.

For the primal distributions, we first parameterise all two-point distributions in the ambiguity set, i.e. we characterise all two-point distributions in $\mathcal{P}$. Due to the first and second moment constraints, the location of one probability mass uniquely determines the location of the other, as well as the probability value at each mass point. Solving a system of non-linear equations, we find that there is a bijection between the set of two-point distributions in $\mathcal{P}$ and the interval $[0, \mu)$, namely, for any $z \in [0, \mu)$ the distributions place a mass at z with probability $q(z) = \frac{\sigma^2}{\sigma^2 + (\mu - z)^2}$ and at $h(z) = \mu + \frac{\sigma^2}{\mu - z}$ with probability $1 - q(z)$.

We now consider the dual feasible solutions. There are several cases depicted in Figure 4.7.

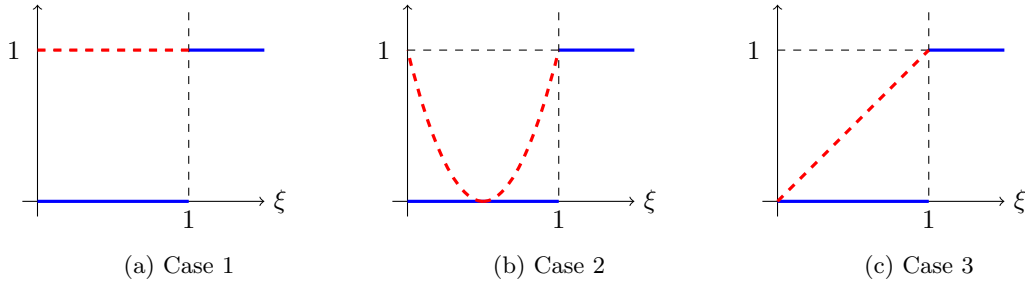


Figure 4.7: The three cases. The red dashed line indicates the dual curve $z(\xi)$, while the solid blue line is the primal objective $\mathbb{1}_{\xi \geq 1}$.

We can assume w.l.o.g. that at optimality, one support point is at 1, otherwise the dual solution would be suboptimal. We now split the analysis into cases depending where the other support point $h(1)$ is. That is, for each possible structure of primal distribution, we will find a dual feasible solution for which both objective values are the same. We note that $h(1) = 1 \iff \sigma^2 = 0$ and $\mu = 1$.

Case 1. (Figure 4.7a) Here $h(1) > 1 \iff \mu > 1$. We select the dual feasible solution $z(\xi) = 1$. The common objective value of both problems is 1.

Case 2. (Figure 4.7b) Here $0 \leq h(1) < 1 \iff \left(\mu - \frac{1}{2}\right)^2 + \sigma^2 \leq \left(\frac{1}{2}\right)^2$. The dual solution $z(\xi)$ must satisfy

$$z(1) = 1 \quad z(h(1)) = 0 \quad z'(h(1)) = 0$$

These three conditions uniquely identify the 3 coefficients for the quadratic function $z(\xi)$. Their derivation is tedious and unnecessary to reproduce here. The objective value of both problems is $q(1) = \frac{\sigma^2}{\sigma^2 + (1 - \mu)^2}$.

Case 3. (Figure 4.7c) Here $h(1) < 0 \iff \left(\mu - \frac{1}{2}\right)^2 + \sigma^2 > \left(\frac{1}{2}\right)^2$. In this case a two-point distribution cannot be feasible — $\mathcal{P}$ only contains distributions supported on $\mathbb{R}_+$. We therefore expand to three-point distributions. Consider a three-point distribution at a_1, a_2, a_3 with masses q_1, q_2 and q_3 respectively, then the moment constraints imply the following relations

$$q_1 + q_2 + q_3 = 1 \quad a_1 q_1 + a_2 q_2 + a_3 q_3 = \mu \quad a_1^2 q_1 + a_2^2 q_2 + a_3^2 q_3 = \mu^2 + \sigma^2.$$

Let us suppose that $a_1 = 0$, $a_2 = 1$ and $a_3 > 1$. By algebraic manipulation, we can express q_1, q_2, q_3 as

$$q_1 = 1 - q_2 - q_3 \quad q_2 = \frac{a_3 \mu - (\mu^2 + \sigma^2)}{a_3 - 1} \quad q_3 = \frac{\left(\mu - \frac{1}{2}\right)^2 + \sigma^2 - \left(\frac{1}{2}\right)^2}{a_3(a_3 - 1)}.$$

We can additionally assume that $\mu < 1$, otherwise we would be in case 1. We have

$$q_2 \geq 0 \iff a_3 \geq \frac{\mu^2 + \sigma^2}{\mu} \quad \text{and} \quad q_2 \leq 1 \iff a_3 \geq \frac{\mu^2 + \sigma^2 - 1}{\mu - 1}.$$

We see that selecting a_3 large enough will make $q_1, q_2, q_3 \geq 0$ and $q_1 + q_2 + q_3 = 1$, i.e. we construct a valid probability distribution. As $a_3 \rightarrow \infty$ it still remains valid. We note that $\lim_{a_3 \rightarrow \infty} q_1 = 1 - \mu$, $\lim_{a_3 \rightarrow \infty} q_2 = \mu$ and $\lim_{a_3 \rightarrow \infty} q_3 = 0$. In this case the optimal solution is shown graphically in Figure 4.7c. The dual solution is $\lambda_0 = 0, \lambda_1 = d(p)$ and $\lambda_2 = 0$. We have constructed a sequence of probability distributions which converge to this two point distribution and therefore our previous duality argument still applies. The

objective value of both problems is μ . To summarise, we have

$$s(\mu, \sigma) = \begin{cases} 1 & \text{if } \mu > 1 \\ \frac{\sigma^2}{\sigma^2 + (1 - \mu)^2} & \text{if } \mu \leq 1 \text{ and } \left(\mu - \frac{1}{2}\right)^2 + \sigma^2 \leq \left(\frac{1}{2}\right)^2 \\ \mu & \text{if } \mu \leq 1 \text{ and } \left(\mu - \frac{1}{2}\right)^2 + \sigma^2 > \left(\frac{1}{2}\right)^2. \end{cases}$$

Next, we note that

$$\frac{\sigma^2}{\sigma^2 + (1 - \mu)^2} = \mu \iff \mu = 1 \text{ or } \left(\mu - \frac{1}{2}\right)^2 + \sigma^2 = \left(\frac{1}{2}\right)^2.$$

It is now sufficient to evaluate both sides on any point in the interior of the circle, i.e.

$$\frac{\sigma^2}{\sigma^2 + (1 - \mu)^2} < \mu \text{ for } (\mu, \sigma) = \frac{1}{2}(1, 1 + \sqrt{2}).$$

The claim now follows directly. $\square$

Lemma (4.2). For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have

$$\begin{aligned} \sup \quad & \mathbb{P}(d(p)\tilde{\xi} > c) &= \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) & \text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu & & \left\| \begin{bmatrix} \beta \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] = \sigma^2 + \mu^2 & & \left\| \begin{bmatrix} \beta - d(p)\tau \\ \alpha - 1 + c\tau - \gamma \end{bmatrix} \right\|_2 \leq \alpha - 1 + c\tau + \gamma \\ & & & = \begin{cases} 1 & \text{if } d(p)\mu > c \\ \left(1 + \left(\frac{c/d(p) - \mu}{\sigma}\right)^2\right)^{-1} & \text{if } d(p)\mu \leq c. \end{cases} \end{aligned}$$

Proof. The second equality is derived in a similar way as in Lemma 4.7. For the first equality, we perform the same initial steps as in the proof of Lemma 4.1 but this time disregard the non-negative worst-case distribution requirement since, in this lemma, we do not have a support constraint. Therefore, the dual

problem becomes

$$\begin{aligned}
\inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) &= \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+ &\text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\
& \alpha + \beta\xi + \gamma\xi^2 \geq 0 \quad \forall \xi \in \mathbb{R} && \alpha + \beta\xi + \gamma\xi^2 \geq 0 \quad \forall \xi \in \mathbb{R} \\
& \alpha + \beta\xi + \gamma\xi^2 \geq 1 \quad \forall \xi \in \mathbb{R} : d(p)\xi \geq c && \alpha - c\tau + (\beta - d(p)\tau)\xi + \gamma\xi^2 \geq 1 \quad \forall \xi \in \mathbb{R},
\end{aligned}$$

where the equality is due to the Farkas Lemma, which enables us to reformulate the second semi-infinite constraint as one over all $\xi \in \mathbb{R}$. Next, we directly perform the minimisation over $\xi \in \mathbb{R}$, which yields

$$\begin{aligned}
\inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) &= \inf \quad & \alpha + \beta\mu + \gamma(\sigma^2 + \mu^2) \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ &\text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\
& \frac{1}{4}\beta^2 \leq \alpha\gamma && \left\| \begin{bmatrix} \beta \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\
& \frac{1}{4}(\beta - d(p)\tau)^2 \leq (\alpha - c\tau - 1)\gamma && \left\| \begin{bmatrix} \beta - d(p)\tau \\ \alpha - 1 + c\tau - \gamma \end{bmatrix} \right\|_2 \leq \alpha - 1 + c\tau + \gamma.
\end{aligned}$$

□

Lemma (4.3). For $d(p) \geq 0$, $c \geq 0$, $\mu > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\inf \quad & \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ d(p)\tilde{\xi}, c \right\} \right] = \sup \quad p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}_+) &\text{s.t.} \quad & \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\
& \mathbb{E}_{\mathbb{P}} [\tilde{\xi}] = \mu && \left\| \begin{bmatrix} \beta - d(p) + \tau \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\
& \mathbb{E}_{\mathbb{P}} [\tilde{\xi}^2] = \sigma^2 + \mu^2 && \left\| \begin{bmatrix} \beta \\ \alpha + c - \gamma \end{bmatrix} \right\|_2 \leq \alpha + c + \gamma \\
& && = \begin{cases} p \frac{c}{1 + (\sigma/\mu)^2} & \text{if } c \leq \frac{1}{2}d(p) \frac{\mu^2 + \sigma^2}{\mu} \\ \frac{1}{2}p \left(c + d(p)\mu - \sqrt{(c - d(p)\mu)^2 + d(p)^2\sigma^2} \right) & \text{if } c > \frac{1}{2}d(p) \frac{\mu^2 + \sigma^2}{\mu}. \end{cases}
\end{aligned}$$

Furthermore, the worst-case expected profit is bounded above and below, respectively, by

$$pd(p)\mu \quad \text{and} \quad \begin{cases} pd(p) \left(\mu - \sqrt{\sigma^2 + \mu^2} \right) + pc & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} \geq c \\ pd(p) \left(\mu - \frac{1}{4}d(p)(\sigma^2 + \mu^2)/c \right) & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} < c. \end{cases}$$

Proof. The dual of the corresponding moment problem of the primal worst-case expectation problem in the statement of the lemma is

$$\begin{aligned} \sup \quad & p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \\ & -\alpha + \beta\xi - \gamma\xi^2 \leq \min \{d(p)\xi, c\} \quad \forall \xi \in \mathbb{R}_+, \end{aligned}$$

where we have intentionally left p in the objective, with the view that we will optimise over it as well later on. We may decompose the min operator in the semi-infinite constraint into two smooth, semi-infinite constraints. Additionally, $\gamma \geq 0$ and $\alpha \geq 0$ since the semi-infinite constraints are infeasible otherwise, and applying the Farkas Lemma only to the first semi-infinite constraint yields

$$\begin{aligned} \sup \quad & p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) & = \sup \quad & p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\ \text{s.t.} \quad & \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ & \text{s.t.} \quad & \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+ \\ & -\alpha + (\beta - d(p) + \tau)\xi - \gamma\xi^2 \leq 0 \quad \forall \xi \in \mathbb{R} & & \frac{1}{4}(\beta - d(p) + \tau)^2 \leq \alpha\gamma \\ & -\alpha + \beta\xi - \gamma\xi^2 \leq c \quad \forall \xi \in \mathbb{R} & & \frac{1}{4}\beta^2 \leq (c + \alpha)\gamma, \end{aligned}$$

where the equality follows from performing the maximisations over ξ . The first equality in the statement of the Lemma follows directly by writing the rotated quadratic cone constraints as explicit second-order cone constraints using the relation (4.13).

The worst-case expectation problem is precisely of the form considered in [102, 44]. We directly apply the seminal result [102, p.206] with $r \leftarrow p$, $y \leftarrow c$, $\mu \leftarrow d(p)\mu$, and $\sigma \leftarrow d(p)\sigma$ to derive the second equality in the statement of the lemma.

To derive the bounds on the worst-case expected profit, we note that without loss of generality τ may take on only two possible values at optimality, namely 0 if $\beta \geq d(p)$ or $d(p) - \beta$ if $0 \leq \beta \leq d(p)$.

Therefore, the optimal value of the above optimisation problem is equivalent to the pointwise maximum of two optimisation problems which we separate into two cases, depending on whether $\beta \geq d(p)$ (case 1) or $0 \leq \beta \leq d(p)$ (case 2).

Case 1. If $\beta \geq d(p)$ then

$$\begin{aligned}
\sup \quad & p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) = \sup \quad p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\
\text{s.t.} \quad & \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+ \quad \quad \quad \text{s.t.} \quad \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+ \\
& \beta - d(p) \geq 0 \quad \quad \quad d(p) \leq \beta \leq \min\{2\alpha^{\frac{1}{2}}\gamma^{\frac{1}{2}} + d(p), 2(c + \alpha)^{\frac{1}{2}}\gamma^{\frac{1}{2}}\} \\
& (\beta - d(p))^2 \leq 4\alpha\gamma \\
& \beta^2 \leq 4(c + \alpha)\gamma \\
& = \sup \quad p \left[-\alpha + \mu \min \left\{ 2\alpha^{\frac{1}{2}}\gamma + d(p), 2(c + \alpha)^{\frac{1}{2}}\gamma \right\} - \gamma^2(\sigma^2 + \mu^2) \right] \\
& \text{s.t.} \quad \alpha \in \mathbb{R}_+, \gamma \in \mathbb{R}_+ \\
& \quad \quad \quad d(p) \leq (c + \alpha)^{\frac{1}{2}}\gamma^{\frac{1}{2}} \\
& = \sup \quad p \left[d(p)\mu - \alpha + \frac{\mu d(p)}{(c/\alpha + 1)^{\frac{1}{2}} - 1} - \frac{d(p)^2(\sigma^2 + \mu^2)}{4((c + \alpha)^{\frac{1}{2}} - \alpha^{\frac{1}{2}})^2} \right] \\
& \text{s.t.} \quad \alpha \in \mathbb{R}_+.
\end{aligned}$$

In the third step of the above chain of equalities, we eliminate γ since $2\alpha^{\frac{1}{2}}\gamma + d(p) = 2(c + \alpha)^{\frac{1}{2}}\gamma$ without loss of generality at optimality, which implies that $\gamma = \frac{1}{2}d(p)/((c + \alpha)^{\frac{1}{2}} - \alpha^{\frac{1}{2}})$. This unconstrained optimal value for γ is both feasible (non-negative) and makes the single constraint redundant.

Case 2. If $0 \leq \beta \leq d(p)$ then

$$\begin{aligned}
& \sup_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+ \\ 0 \leq \beta \leq d(p) \\ \frac{1}{4}\beta^2 \leq (c + \alpha)\gamma}} p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) = \sup_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+ \\ 0 \leq \beta \leq \min\left\{d(p), 2(c + \alpha)^{\frac{1}{2}}\gamma^{\frac{1}{2}}\right\}}} p(-\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2)) \\
& = \sup_{\substack{\alpha \in \mathbb{R}_+, \gamma \in \mathbb{R}_+}} p\left(-\alpha + \mu \min\left\{d(p), 2(c + \alpha)^{\frac{1}{2}}\gamma^{\frac{1}{2}}\right\} - \gamma^2(\sigma^2 + \mu^2)\right) \\
& = \sup_{\alpha \in \mathbb{R}_+} p\left[d(p)\mu - \left(\alpha + \frac{1}{4}d(p)^2(\sigma^2 + \mu^2)(c + \alpha)^{-1}\right)\right] \\
& = \begin{cases} p d(p) \left(\mu - \sqrt{\sigma^2 + \mu^2}\right) + pc & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} \geq c \\ p d(p) \left(\mu - \frac{1}{4}d(p)(\sigma^2 + \mu^2)/c\right) & \text{if } \frac{1}{2}d(p)\sqrt{\sigma^2 + \mu^2} < c. \end{cases}
\end{aligned}$$

In the second step of the above chain of equalities, we make the variable transformation $\gamma \leftarrow \gamma^{\frac{1}{2}}$. In the third step of the above chain of equalities, we eliminate γ . This is possible since the maximum (over γ) of the minimum of the two concave quadratic functions will occur at the intersection point (if one exists) of the two functions. This intersection happens when $\gamma = \frac{1}{2}d(p)(c + \alpha)^{-\frac{1}{2}}$, always exists since $c > 0$, and is non-negative. In the final step, we use the fact that for $a \geq 0$, $b \geq 0$, and $c \geq 0$

$$\begin{aligned}
& \text{minimise } ax + b(x + c)^{-1} \\
& \text{subject to } x \geq 0
\end{aligned} = \begin{cases} 2a^{\frac{1}{2}}b^{\frac{1}{2}} - ac & \text{if } (b/a)^{\frac{1}{2}} \geq c \\ bc^{-1} & \text{if } (b/a)^{\frac{1}{2}} < c. \end{cases}$$

The claim now follows directly. □

Lemma (4.4). For $\varepsilon \in (0, 1)$, we have

$$\begin{aligned}
& \text{maximise} \quad \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[p \min \left\{ d(p)\tilde{\xi}, c \right\} \right] = \text{maximise} \quad \frac{1}{2}p \left(c + d(p)\mu - \sqrt{(c - d(p)\mu)^2 + d(p)^2\sigma^2} \right) = \pi^2, \\
& \text{subject to} \quad p \in \mathbb{R} \quad \quad \quad \text{subject to} \quad p \in \mathbb{R}_+ \\
& \quad \quad \quad \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(d(p)\tilde{\xi} > c \right) \leq \varepsilon \quad \quad \quad 0 \leq d(p) \leq c \max \left\{ \frac{1}{\mu + \sigma\kappa(\varepsilon)}, \frac{\varepsilon}{\mu} \right\} \\
& \quad \quad \quad p \geq 0, d(p) \geq 0
\end{aligned} \tag{4.16}$$

where $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$ and π is defined as

$$\begin{aligned}
& \pi = \text{maximise} \quad \theta \\
& \text{subject to} \quad p \in \mathbb{R}_+, \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \gamma \in \mathbb{R}_+, \tau \in \mathbb{R}_+, u \in \mathbb{R}_+, \theta \in \mathbb{R}_+ \\
& \quad \quad \quad u = -\alpha + \beta\mu - \gamma(\sigma^2 + \mu^2) \\
& \quad \quad \quad \left\| \begin{bmatrix} 2\theta \\ p - u \end{bmatrix} \right\|_2 \leq p + u \\
& \quad \quad \quad \left\| \begin{bmatrix} \beta - d(p) + \tau \\ \alpha - \gamma \end{bmatrix} \right\|_2 \leq \alpha + \gamma \\
& \quad \quad \quad \left\| \begin{bmatrix} \beta \\ \alpha + c - \gamma \end{bmatrix} \right\|_2 \leq \alpha + c + \gamma \\
& \quad \quad \quad 0 \leq d(p) \leq c \max \left\{ \frac{1}{\mu + \sigma\kappa(\varepsilon)}, \frac{\varepsilon}{\mu} \right\}.
\end{aligned}$$

Proof. We begin with the first, analytical, equality. For the objective, we use the analytical equality in Lemma 4.3, noting that the linear piece is increasing, and therefore without loss of generality at optimality we will have $c \geq \frac{1}{2}d(p)\frac{\mu^2 + \sigma^2}{\mu}$. In the chance constraint, since $\varepsilon \in (0, 1)$, we must have $d(p)\mu \leq c$. Lemma 4.2 then implies

$$\min \left\{ \left(1 + \left(\frac{c/d(p) - \mu}{\sigma} \right)^2 \right)^{-1}, \frac{d(p)\mu}{c} \right\} \leq \varepsilon \iff d(p) \leq c \max \left\{ \frac{1}{\mu + \sigma\kappa(\varepsilon)}, \frac{\varepsilon}{\mu} \right\}.$$

It is useful to compare this to what is derived in [50]. In Theorem 1, the authors consider a similar

worst-case probability problem, but in a multivariate setting. The first piece of the upper bound is the same as their analytical solution for the case where no support information is injected into the problem. Indeed, as we have seen, in that case the information about the support is not used. However, when the support constraints impact the optimal solution of (4.16), the authors of [50] can only compute a lower bound on (4.16), whereas we are able to state the exact solution in the one-dimensional case under consideration.

We make use of the SOCP result from Lemma 4.3 to derive the equivalent SOCP formulation here too. In order to remove the bilinearity in the objective (since we now optimise over p too) we apply a hypograph transformation and represent the resulting constraint as a quadratic cone constraint using the relation (4.13) as shown below:

$$\begin{aligned}
& \begin{array}{ll} \text{maximise} & f(\mathbf{x})g(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X} \\ & f(\mathbf{x}) \geq 0 \quad g(\mathbf{x}) \geq 0 \end{array} & = \begin{array}{ll} \text{maximise} & \theta^2 \\ \text{subject to} & \mathbf{x} \in \mathcal{X} \\ & \theta^2 \leq f(\mathbf{x})g(\mathbf{x}) \\ & f(\mathbf{x}) \geq 0 \quad g(\mathbf{x}) \geq 0 \end{array} \\
& & = \begin{array}{ll} \text{maximise} & \theta^2 \\ \text{subject to} & \mathbf{x} \in \mathcal{X} \\ & \left\| \begin{bmatrix} 2\theta \\ f(\mathbf{x}) - g(\mathbf{x}) \end{bmatrix} \right\|_2 \leq f(\mathbf{x}) + g(\mathbf{x}) \\ & f(\mathbf{x}) \geq 0 \quad g(\mathbf{x}) \geq 0. \end{array}
\end{aligned}$$

The statement in the lemma follows from a monotonic transformation of the objective. □

Chapter 5

Best-Case Ambiguous Chance

Constraints: Image Denoising

5.1 Introduction

In this chapter, we show how chance constraints can be used to solve an image denoising problem. We consider both optimistic and pessimistic ambiguous joint chance constraints, and show their equivalent deterministic reformulations.

Surprisingly, we show that if the image denoising problem is formulated with an optimistic joint chance constraint, this problem generalises a popular total variation approach to image-denoising [53]. On the other hand, we argue that pessimistic chance constraints are unsuitable for this application.

All numerical experiments are performed on a desktop machine with 8GB memory and quad-core 3.0GHz processor, using MOSEK 7.1 [82] solver accessed through the C interface. Test images are taken from the USC-SIPI image database as well as the Kodak True Colour Image Suite.

Image denoising (or noise reduction) is a standard and important problem in the image analysis and processing community. An image may be corrupted during the process of acquisition, transmission, or storage, resulting in a degradation of quality. The process of image denoising is to restore the corrupted image as close as possible to its original quality.

The image denoising problem has been (and still is) well studied in the literature. A good survey

on the different techniques which have been applied [67] broadly classifies existing techniques into two categories; spatial and transform domain filtering. Spatial domain filtering techniques work directly with the noisy image and apply various filtering processes on it. In contrast, transform domain filtering typically apply denoising procedures on a transformation of the noisy image.

The focus of this chapter is on one promising spatial domain filtering technique based on total-variation minimisation. This approach, introduced in [101], minimises the total variation of the noisy image, subject to constraints involving statistics of the noise present. One nice property of this technique is that it is edge-preserving, i.e. the process preserves important features such as edges, corners, and other sharp structures. This is vital to the visual appearance of the denoised image. Building on this seminal work, [53] design reformulations and algorithms which allow total variation image denoising techniques to be implemented using off-the-shelf solvers, and thus more computationally accessible than specialised techniques. Although the second-order conic reformulations in [53] work well to denoise an image, the number of variables and constraints in the formulation depends on the number of pixels in the image. To combat this, the authors make use of the structure of the problem to construct efficient interior point solvers.

Our contribution in this chapter is to use ambiguous joint chance constraints to formulate the image denoising problem. We minimise a total variation objective as in [53], but instead of having a deterministic constraint, we encode the a priori knowledge of the statistics of the noise via ambiguous chance constraints. Surprisingly, doing this with best-case chance constraints leads to a generalisation of the result presented in [53]. Crucially, we are able to give further insight into this type of problem and how the input parameters should be chosen, based on statistical estimation of the noise.

5.2 Deterministic Image Denoising

We first introduce the existing image denoising problem and formulation as it appears in [53] to keep this chapter self-contained. For simplicity, we consider only grayscale images where each pixel has an intensity in the range $[0, 1]$, 0 being black and 1 being white. A two-dimensional image of width w and height h is represented as a single vector of length $n = wh$, i.e. we stack all the pixels into a single

column vector. The total variation based image denoising problem can then be formulated as follows:

$$\begin{aligned}
h(\sigma) = \text{minimise} \quad & \text{TV}_p(\mathbf{u}) \\
\text{subject to} \quad & \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n \\
& \mathbf{0} \leq \mathbf{u} \leq \mathbf{1} \\
& \mathbf{u} + \mathbf{v} = \mathbf{f} \\
& \|\mathbf{v}\|_q \leq \sigma.
\end{aligned} \tag{5.1}$$

In (5.1) we observe a noisy image $\mathbf{f}$ and try to reproduce it by finding an image $\mathbf{u}$ and noise $\mathbf{v}$ which when additively combined equal $\mathbf{f}$. However, some q -norm function of the noise must not exceed a given value $\sigma \in \mathbb{R}_+$. Additionally, the denoised image $\mathbf{u}$ should have minimum total variation subject to these constraints. These heuristic constraints are based on the assumption that the noise is believed to be small, and the transition of intensities in the original image is primarily smooth. We note that trivial bounds on $\mathbf{v}$ are $-\mathbf{1} \leq \mathbf{v} \leq \mathbf{1}$ as well as $\mathbf{f} - \mathbf{1} \leq \mathbf{v} \leq \mathbf{f}$ due to the equality constraint. The bound constraints on $\mathbf{u}$ are redundant in most cases (based on the value of σ) since they are implied by the equality and norm constraint. Minimising the total variation function of the image ensures that the intensities of neighbouring pixels (row- and column-wise) do not differ too much, i.e. it is a regularisation which smoothens the image. It is convenient to write the total variation function as a sum of norms of a linear transformation of the pixels

$$\text{TV}_p(\mathbf{u}) = \sum_{i=1}^n \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_p,$$

where $\mathbf{x} = \mathbf{T}_x \mathbf{u}$ and $\mathbf{y} = \mathbf{T}_y \mathbf{u}$ for fixed matrices $\mathbf{T}_x$ and $\mathbf{T}_y$ that take the difference in pixel intensities for both vertical and horizontal neighbours of each pixel, i.e. for every pixel i the value of x_i is the intensity of the right neighbour of pixel i minus the intensity of pixel i . Similarly, for every pixel i the value of y_i is the intensity of the top neighbour of pixel i minus the intensity of pixel i . The respective differences are defined to be zero for pixels on top and right boundaries of the image. Doing this allows us to write

down the following equivalent dual of the full image denoising problem

$$\begin{aligned}
h(\sigma) = \text{maximise} \quad & -(\mathbf{b}^\top \mathbf{f} + \sigma \|\mathbf{b}\|_{q*}) \\
\text{subject to} \quad & \mathbf{b} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n \\
& \mathbf{b} = \mathbf{T}_x^\top \mathbf{x} + \mathbf{T}_y^\top \mathbf{y} \\
& \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{p*} \leq 1 \quad \forall i = 1, \dots, n,
\end{aligned}$$

where the dual norm satisfies $\frac{1}{p} + \frac{1}{p*} = 1$. Due to a classical envelope theorem [79] the first derivative, $h'(\sigma)$, satisfies $h'(\sigma) = -\|\mathbf{b}^*\|_{q*}$. From the primal formulation it is easier to see that h is decreasing in σ since the feasible set is expanding. From the dual problem it is easier to see that additionally h is convex in σ , since it is a maximum of an infinite number of linear functions of σ . The dual formulation can be useful to reduce the computation time as the feasible set (once $\mathbf{b}$ is eliminated) only has pairwise couplings between decision variables. This makes it a prime target for decomposition methods, allowing denoising of extremely large images without decomposing the image itself. It is also possible to derive cutting planes for the dual formulation using first-order Taylor series expansion, and equivalently solve a sequence of linearly constrained problems. If the dual problem is solved, the denoised image is precisely the optimal vector of dual variables corresponding to the linear constraint. After solving (5.1) for a given σ , we can measure, a posteriori, the quality of the denoised image $\mathbf{u}$ by defining the metric

$$d(\mathbf{u}) = \frac{\|\bar{\mathbf{u}} - \mathbf{u}\|_2}{\|\bar{\mathbf{u}} - \mathbf{f}\|_2},$$

where $\bar{\mathbf{u}}$ is the original image (without noise), $\mathbf{f}$ is the noisy image, and $\mathbf{u}$ is the denoised image, i.e. the optimal solution of (5.1). We note that $d(\bar{\mathbf{u}}) = 0$ and if $d(\mathbf{u}) < 1$ then an improvement has been made. We should note that this is a naive way to measure image denoising performance, for example $d(\bar{\mathbf{u}} + \varepsilon \mathbf{1})$ could be high but visually the images may be indistinguishable. The problem is to select σ so that $d(\mathbf{u}^*)$ is minimised, where $\mathbf{u}^*$ is optimal solution of (5.1). To do this, we can use the property of weak duality, i.e. if we have found primal feasible and dual feasible solutions for which the primal objective value is equal to the dual objective value, then those feasible solutions must also be optimal. To this end, we

formulate the following a posteriori parameter selection problem

$$\begin{aligned} & \text{minimise} && f(\sigma) \\ & \text{subject to} && \sigma \in \mathbb{R}_+, \end{aligned} \tag{5.2}$$

where

$$\begin{aligned} f(\sigma) = & \text{minimise} && \|\bar{\mathbf{u}} - \mathbf{u}\|_r && = \text{maximise} && -(\boldsymbol{\alpha}^\top \bar{\mathbf{u}} + \boldsymbol{\beta}^\top \mathbf{f} + \sigma \|\boldsymbol{\beta}\|_{q*} + \gamma h(\sigma)) \\ & \text{subject to} && \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n && \text{subject to} && \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n, \boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R} \\ &&& \text{TV}_p(\mathbf{u}) \leq h(\sigma) && && \boldsymbol{\alpha} + \boldsymbol{\beta} = \mathbf{T}_x^\top \mathbf{x} + \mathbf{T}_y^\top \mathbf{y} \\ &&& \mathbf{u} + \mathbf{v} = \mathbf{f} && && \|\boldsymbol{\alpha}\|_{r*} \leq 1 \\ &&& \|\mathbf{v}\|_q \leq \sigma && && \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{p*} \leq \gamma \quad \forall i = 1, \dots, n. \end{aligned}$$

The constraint $\text{TV}_p(\mathbf{u}) \leq h(\sigma)$ ensures that $(\mathbf{u}, \mathbf{v})$ is optimal in (5.1). This optimality is guaranteed by strong duality. The other constraints ensure that $(\mathbf{u}, \mathbf{v})$ is a feasible solutions to the image denoising problem (5.1). The equivalence above follows from convex duality. Since h is convex, the parameter selection problem is not necessarily convex. We note that a different r -norm may be used in the objective of f . By considering a large number of varied images we have found, empirically, that f is quasi-convex. Although we have been unable to prove this rigorously, we believe the quasi-convexity property may be true since f is the maximum of the sum of a decreasing linear and increasing function of σ . Figure 5.1 shows how f and h vary with σ .

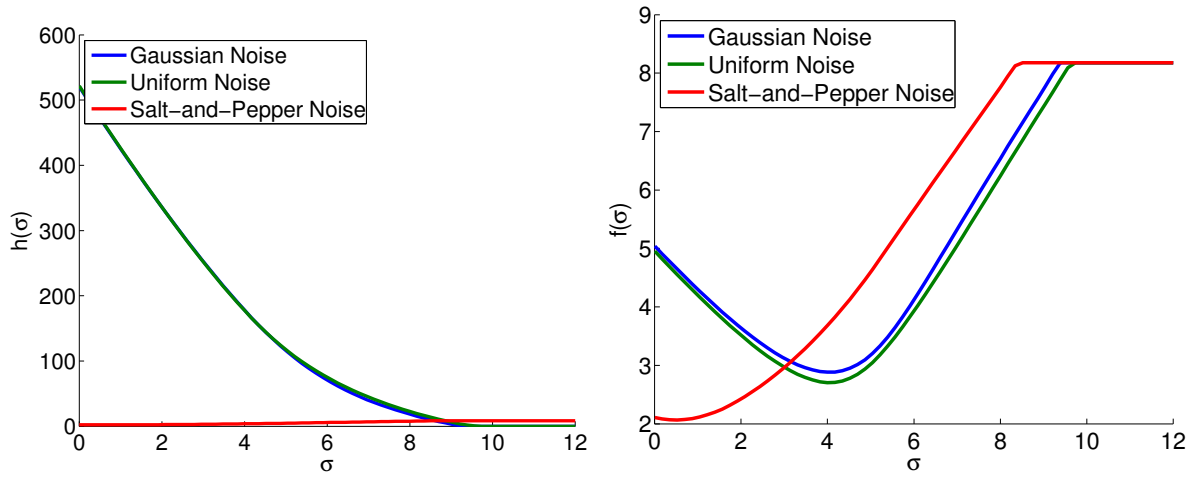


Figure 5.1: Plot of h and f as a function of σ . Image size is 50x50, standard deviation of noise is 0.1. The point where the functions become constant is the upper bound given in Lemma 5.1.

Under this conjecture we can use an envelope theorem to again numerically determine the gradient of f and thus (because σ is one-dimensional) whether we are on the left or the right of the global optimal value of σ in the parameter selection problem. Thus we find $f'(\sigma) = -(\|\mathcal{B}^*\|_{q^*} + \gamma^* h'(\sigma))$. We next give an upper bound on σ which is based on a quantity related to the standard deviation of the pixel intensities of the noisy image $\mathbf{f}$.

Lemma 5.1. *The optimal solution, σ^* , of (5.2) is bounded above and below (for $q = 2$) by*

$$0 \leq \sigma^* \leq \sqrt{\mathbf{f}^\top \mathbf{f} - \frac{(\mathbf{f}^\top \mathbf{1})^2}{n}},$$

where the second inequality is $\sqrt{n}$ multiplied by the standard deviation of the noisy image $\mathbf{f}$.

Proof. $\text{TV}(\mathbf{u}) = 0$ if and only if all pixels of $\mathbf{u}$ have the same intensity. Therefore, it is unlikely that the denoised image will have zero total variation, unless σ is chosen too high. We therefore try to find the smallest σ for which the total variation is zero by solving the following problem

$$\begin{aligned} \text{minimise} \quad & \|\mathbf{v}\|_2 &= \text{minimise} \quad & \sqrt{\sum_{i=1}^n (f_i - \hat{u})^2} \\ \text{subject to} \quad & \hat{u} \in \mathbb{R}, \mathbf{v} \in \mathbb{R}^n & \text{subject to} \quad & \hat{u} \in \mathbb{R} \\ & 0 \leq \hat{u} \leq 1 & & 0 \leq \hat{u} \leq 1. \\ & \hat{u}\mathbf{1} + \mathbf{v} = \mathbf{f} & & \end{aligned}$$

Via the first-order optimality conditions we find that $\hat{u} = \frac{\mathbf{f}^\top \mathbf{1}}{n}$ and the claim follows. $\square$

5.2.1 Computational Experiments

For the numerical experiments, we will contaminate images with three types of noise: Gaussian, uniform, and salt-and-pepper. All three will be constrained to have mean 0. Gaussian noise will have a standard deviation of s . Uniform noise will be supported on $[-x, x]$ for $0 \leq x \leq 1$ and therefore has a standard deviation of $x/\sqrt{3}$. For salt-and-pepper noise, given a threshold value $0 \leq \tau \leq 1$, we draw a value from a uniform distribution on $[-1, 1]$. If the drawn value is above τ , then the noise value is set to 1, below $-\tau$ the noise is set to -1 , and otherwise the noise is set to 0. Therefore the salt-and-pepper noise distribution is discrete taking the values -1, 0, and 1 with probabilities $\frac{1}{2}(1-\tau)$, τ , and $\frac{1}{2}(1-\tau)$, respectively, and the standard deviation is $\sqrt{1-\tau}$. After the noise is applied, the image intensity values are clamped between 0 and 1 (projected onto the interval $[0, 1]$). Figure 5.2 shows an image corrupted by these three types of noise, where the standard deviation was set to 0.1, and the noise applied to each pixel is pairwise independent.



Figure 5.2: Examples of no noise, Gaussian noise, uniform noise, and salt-and-pepper noise (left to right). The standard deviation of the noise of each individual pixel is set to 0.1. The image is the commonly used “Lena” photograph.

We now show how the optimal value of σ is affected by the image size and the standard deviation of the noise. In Figure 5.3a the standard deviation of the noise is varied and the optimal σ (i.e. the one which achieves the best denoised image) is plotted for both Gaussian and uniform noise. Figure 5.3b shows the ratio $\|\bar{\mathbf{u}} - \mathbf{u}^*\|_2 / \|\mathbf{f} - \mathbf{u}^*\|_2$ where $\mathbf{u}^*$ is the denoised image and $\mathbf{f}$ is the noisy image (which varies with the standard deviation). For this experiment the image size was again 50x50.

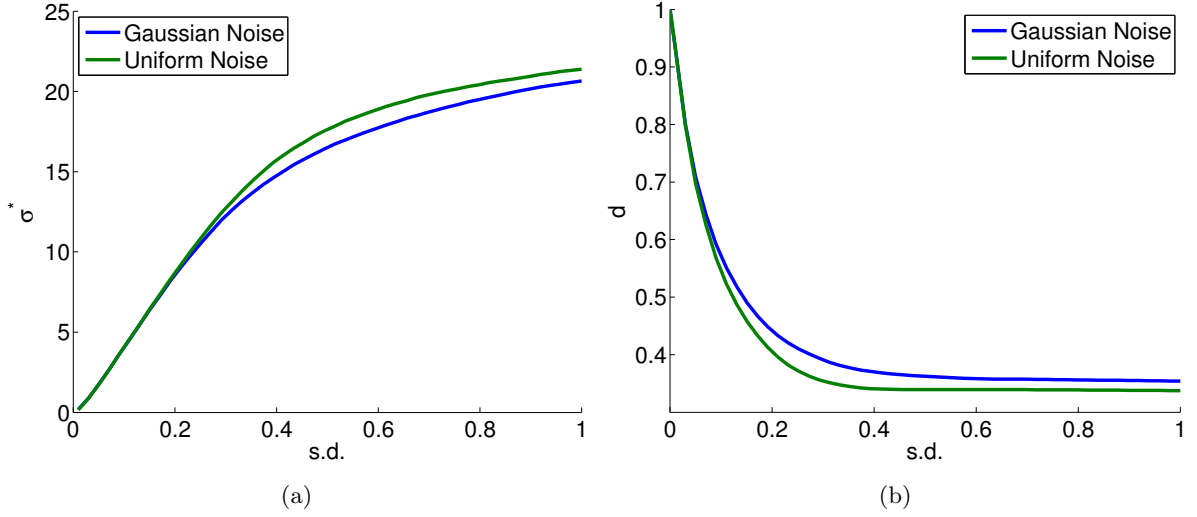


Figure 5.3: Plot showing how the standard deviation of the noise affects the optimal σ and the quality of the denoised image.

In Figure 5.4, we take 30 images and calculate the optimal σ for each image resized to contain a specific number of pixels n . The plot shows a linear relationship between $\sqrt{n}$ and the optimal σ . The red (bottom) line is the average optimal σ over all 30 images for each given n , and the blue (top) line shows a statistical estimate of σ which we will introduce in the next section.

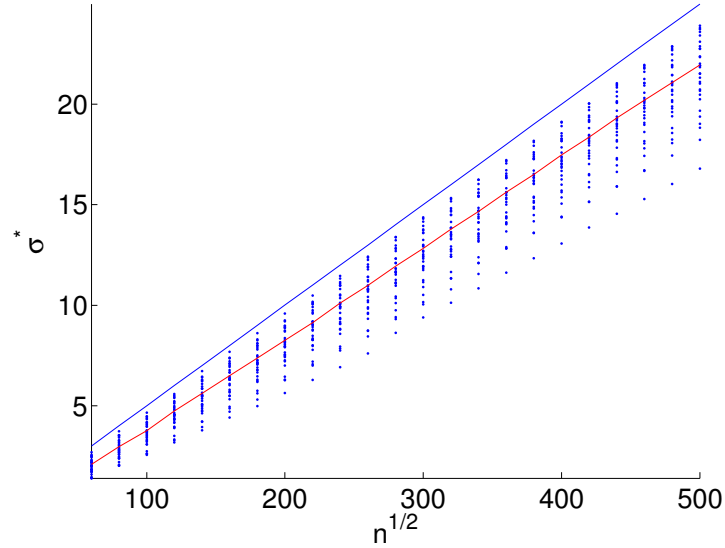


Figure 5.4: Plot showing how the size of the image (square root of the number of pixels) is related to the optimal σ .

Figure 5.5 shows the output of the denoising optimisation problem. In these experiments we take images from the image databases and resize them to size 400x400. Then we add the noise to the image from a prescribed distribution. Finally, we formulate the primal problem (5.1) and solve using the off-

the-shelf second-order cone solver Mosek. The resulting optimisation problem has order of 400^2 variables and constraints and solves in around 40 seconds.



Figure 5.5: The photo on the left is the original image. The middle photo is noisy image with Gaussian noise and a standard deviation 0.05. The right image is the denoised image $\mathbf{u}^*$ with $\sigma = 3.9115$ and $d(\mathbf{u}^*) = 62\%$. The image is the commonly used “Lena” photograph.

5.3 Stochastic Image Denoising

In this next section, we introduce a stochastic variant of (5.1) which makes use of ambiguous chance constraints. We will consider the following image denoising formulation

$$\begin{aligned}
& \text{minimise} && \text{TV}(\mathbf{u}) && \text{where} && g(\mathbf{u}, \mathbf{f}) = \inf / \sup && \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) && (5.3) \\
& \text{subject to} && \mathbf{u} \in \mathbb{R}^n && && \text{s.t.} && \mathbb{P} \in \mathcal{P}_0(\mathcal{V}) \\
& && \mathbf{0} \leq \mathbf{u} \leq \mathbf{1} && && && \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \\
& && g(\mathbf{u}, \mathbf{f}) \geq 1 - \varepsilon && && && \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_p] \leq \sigma,
\end{aligned}$$

and $\mathcal{P}_0(\mathcal{V})$ is the set of all probability measures on $\mathcal{V} \subseteq \mathbb{R}^n$. We have replaced the deterministic noise $\mathbf{v}$ with a random variable $\tilde{\mathbf{v}}$ governed by a distribution $\mathbb{P}$, and relaxed the equality constraint in (5.1) by introducing a margin $\delta > 0$.

5.3.1 Pessimistic Chance Constraints

In this section we consider the case where in (5.3) the function g is an infimum over $\mathcal{P}$. Using standard techniques from robust optimisation introduced in the preceding chapters, we have the following equivalence for the worst-case chance constraint.

Lemma 5.2. For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\inf \quad & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \iff \exists \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) & 1 \leq \tau_j^+ (\delta + (f_j - u_j)) \\
& \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} & 1 \leq \tau_j^- (\delta - (f_j - u_j)) \quad \left. \vphantom{\begin{aligned} \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \end{aligned}} \right\} \forall j = 1, \dots, n \\
& \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_p] \leq \sigma & \min_{\boldsymbol{\beta} \in \mathbb{R}^n} \max_{j=1, \dots, n} \{ \|\boldsymbol{\beta}\|_{p^*}, \|\boldsymbol{\beta} + \tau_j^+ \mathbf{e}_j\|_{p^*}, \|\boldsymbol{\beta} - \tau_j^- \mathbf{e}_j\|_{p^*} \} \leq \frac{\varepsilon}{\sigma},
\end{aligned}$$

where $\mathbf{e}_j$ is the j th standard basis vector and p^* satisfies $\frac{1}{p} + \frac{1}{p^*} = 1$.

Proof. See Appendix. □

Although convex, the last constraint in the reformulated version makes it difficult to understand what the joint chance constraint is enforcing. To get a clearer picture, we consider the case when $p = 1$ and obtain the following result for this special case.

Lemma 5.3. For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\inf \quad & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \iff \|\mathbf{f} - \mathbf{u}\|_{\infty} \leq (\delta(\delta - \sigma/\varepsilon))^{\frac{1}{2}}. \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \\
& \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \\
& \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_1] \leq \sigma
\end{aligned}$$

Proof. See Appendix. □

Lemma 5.3 shows that when $p = 1$, the image denoising constraint is equivalent to a linear constraint where the maximum deviation of $\mathbf{u}$ from $\mathbf{f}$ is bounded above. If we want to relate the ambiguous chance constrained image denoising problem to the one introduced in Section 5.2, then we should set δ to a small and σ to a large value. Additionally, we require ε to be a small, as is usually the case in distributionally robust chance constraints. However, in the reformulation given in Lemma 5.3 it is necessary, for the constraint to be feasible, that

$$(\delta(\delta - \sigma/\varepsilon))^{\frac{1}{2}} \geq 0 \implies \sigma/\varepsilon < \delta.$$

If we select a small value for ε , then either δ should be large or σ should be small, both of which would not make sense in the interpretation of the image denoising problem as a pessimistic ambiguous chance constraint. We observe this numerically for the general case in Lemma 5.2, due to the fact that the feasible set of the chance constraint becomes empty even if ε , σ , and δ are set to “reasonable” values. The image denoising problem, however, also provides the following example which is tractable, but where the deviation bound is not positive homogeneous.

Lemma 5.4. *For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have*

$$\begin{aligned}
\inf \quad & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \quad \Longleftrightarrow \quad \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \quad \quad \quad (\boldsymbol{\beta} + (\mathbf{f} - \mathbf{u}))^\top (\boldsymbol{\beta} + (\mathbf{f} - \mathbf{u})) + \sigma \leq \varepsilon \gamma^2 \\
& \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \quad \quad \quad \|\boldsymbol{\beta}\|_\infty \leq \delta - \gamma. \\
& \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}^\top \tilde{\mathbf{v}}] \leq \sigma
\end{aligned}$$

Proof. See Appendix. □

The chance constraint is feasible if $\delta \geq (\sigma/\varepsilon)^{\frac{1}{2}}$, and again exhibits the same problem of parameter selection discussed previously. If δ is made smaller, γ tends to 0 and $\boldsymbol{\beta}$ also tends to $\mathbf{0}$, and the constraint becomes infeasible. We saw in the Chapter 3 that positive homogeneity of the dispersion bound is a sufficient condition for the existence of a tractable reformulation. The above result shows that this condition is not necessary, thus opening up the potential of expanding the class of tractable ambiguous chance constraints.

5.3.2 Optimistic Chance Constraints

In this section we consider the case where in (5.3) the function g is a supremum over $\mathcal{P}$. Using standard techniques from robust optimisation, we have the following equivalence for the best-case chance constraint.

Lemma 5.5. For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
\sup \quad & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \iff \exists \mathbf{v} \in \mathbb{R}^n : \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathcal{V}) & \mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \mathbf{v} \leq \mathbf{f} + \delta \mathbf{1} \\
& \mathbb{E}[\tilde{\mathbf{v}}] = \mathbf{0} & \|\mathbf{v}\|_p \leq \frac{1}{2}\sigma/(1 - \varepsilon) \\
& \mathbb{E}[\|\tilde{\mathbf{v}}\|_p] \leq \sigma & \max \left\{ \frac{\varepsilon}{1 - \varepsilon} \mathbf{b}_l, -\mathbf{b}_u \right\} \leq \mathbf{v} \leq \min \left\{ -\mathbf{b}_l, \frac{\varepsilon}{1 - \varepsilon} \mathbf{b}_u \right\},
\end{aligned} \tag{5.4}$$

where $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u\}$ and $\mathbf{b}_l < \mathbf{0} < \mathbf{b}_u$.

Proof. See Appendix. □

We can see that if $\mathbf{b}_l = -\mathbf{1}$, $\mathbf{b}_u = \mathbf{1}$, $p = 2$, and $\varepsilon = \frac{1}{2}$, we recover precisely the image denoising constraint in (5.1) as $\delta \downarrow 0$. As before, we can equivalently solve the dual of this image denoising problem if suitable decomposition techniques are applied.

Lemma 5.6. For $\delta > 0$, $\kappa > 0$, and $\kappa_l < \mathbf{0} < \kappa_u$

$$\begin{aligned}
\text{minimise} \quad & \text{TV}_p(\mathbf{u}) & = \text{maximise} \quad & -\mathbf{b}^\top \mathbf{f} - \delta \|\mathbf{b}\|_1 - \kappa \|\mathbf{b} + (\bar{\kappa}_l - \bar{\kappa}_u)\|_{q*} + \bar{\kappa}_l^\top \kappa_l - \bar{\kappa}_u^\top \kappa_u \\
\text{subject to} \quad & \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n & \text{subject to} \quad & \mathbf{b} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n, \bar{\kappa}_l \in \mathbb{R}_+^n, \bar{\kappa}_u \in \mathbb{R}_+^n \\
& \|(\mathbf{f} - \mathbf{u}) + \mathbf{v}\|_\infty \leq \delta & & \mathbf{b} = \mathbf{T}_x^\top \mathbf{x} + \mathbf{T}_y^\top \mathbf{y} \\
& \|\mathbf{v}\|_q \leq \kappa & & \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{p*} \leq 1 \quad \forall i = 1, \dots, n. \\
& \kappa_l \leq \mathbf{v} \leq \kappa_u & &
\end{aligned}$$

Proof. See Appendix. □

We now give some numerical results for the image denoising problem with the constraint (5.4) and parameters $\mathbf{b}_l = -\mathbf{f}$, $\mathbf{b}_u = \mathbf{1} - \mathbf{f}$, and $p = q = 2$. In the following experiment, we choose two values of σ , the best choice in the deterministic setting (i.e. the optimal value of σ for problem (5.2)), as well as a statistical estimate based on the distribution of the noise. To calculate the latter, we remember that in the probabilistic setting, the constraint $\mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_2] \leq \sigma$ is present in the ambiguity set. Given a concrete noise distribution $\mathbb{P}$, our statistical estimate of σ will be $\mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_2]$, which is calculated exactly by evaluating

a multi-dimensional integral if the distribution function permits, or via a sample average approximation otherwise. For this experiment, we present average results over 30 100x100 images. Gaussian noise is applied with a standard deviation of 0.1.

Tables 5.1 and 5.2 show the quality of the denoising as the parameters ε and δ are varied. In Table 5.1 the parameter σ is chosen optimally, i.e. no other value of σ makes the denoised image closer to the uncorrupted image. We find that values of $\delta = 0.04$ and $\varepsilon = 0.2$ provide the best denoising result, on average. However, since in practice the uncorrupted image is not available, Table 5.2 shows the performance of the image denoising scheme when σ is chosen based on statistical estimation of the noise. We find that, on average, the performance is comparable to the case where σ is chosen optimally. In this case we find that the best parameters are $\varepsilon = 0.3$ and $\delta = 0.02$.

Table 5.1: Average denoising quality $d(\mathbf{u}^*)$ for images of size 100x100 for different values of ε (rows) and δ (columns) with the a priori optimally on average selected $\sigma = 8.73$. The smallest distance to the original image is achieved with $\varepsilon = 0.2$ and $\delta = 0.04$.

$\varepsilon \backslash \delta$	0	0.0050	0.0100	0.0200	0.0400	0.0600	0.0800	0.1000
0.0100	0.9578	0.9259	0.8949	0.8364	0.7329	0.6521	0.5952	0.5655
0.0200	0.9166	0.8859	0.8563	0.8005	0.7035	0.6309	0.5831	0.5640
0.0500	0.8026	0.7764	0.7515	0.7059	0.6348	0.5896	0.5738	0.5823
0.1000	0.6682	0.6455	0.6256	0.5955	0.5828	0.6142	0.6457	0.6941
0.2000	0.6207	0.5967	0.5752	0.5400	0.5101	0.5609	0.6918	0.8484
0.3000	0.5820	0.5618	0.5445	0.5200	0.5276	0.6354	0.7905	0.9472
0.4000	0.5434	0.5309	0.5230	0.5245	0.6165	0.7739	0.9351	1.0608
0.5000	0.5397	0.5520	0.5731	0.6359	0.8015	0.9689	1.0345	1.0609

Table 5.2: Average denoising quality $d(\mathbf{u}^*)$ for images of size 100x100 for different values of ε (rows) and δ (columns) with the statistical estimate $\sigma = 9.99$. The smallest distance to the original image is achieved with $\varepsilon = 0.3$ and $\delta = 0.02$.

$\varepsilon \backslash \delta$	0	0.0050	0.0100	0.0200	0.0400	0.0600	0.0800	0.1000
0.0100	0.9578	0.9259	0.8949	0.8364	0.7329	0.6521	0.5952	0.5655
0.0200	0.9166	0.8859	0.8563	0.8005	0.7035	0.6309	0.5831	0.5640
0.0500	0.8026	0.7764	0.7515	0.7059	0.6348	0.5896	0.5738	0.5823
0.1000	0.6586	0.6441	0.6328	0.6238	0.6054	0.6142	0.6457	0.6941
0.2000	0.5813	0.5614	0.5445	0.5217	0.5397	0.6568	0.8135	0.9396
0.3000	0.5475	0.5335	0.5240	0.5210	0.6020	0.7569	0.9171	1.0567
0.4000	0.5313	0.5343	0.5459	0.5935	0.7484	0.9144	1.0345	1.0608
0.5000	0.6260	0.6637	0.7040	0.7901	0.9638	1.0187	1.0345	1.0609

5.4 Conclusion

In this chapter we considered an image denoising problem. We showed that a standard total-variation minimisation objective function together with a best-case ambiguous chance constraint generalises an existing well-known approach [53].

We also considered the same model but with a worst-case (instead of best-case) ambiguous chance constraint and argued that it is unsuitable for the context of the application. It does, however, provide an example of a tractable reformulation where the deviation function in the ambiguity set is not positive homogeneous, showing that the corresponding condition of tractability in Chapter 3 is only sufficient.

5.5 Appendix

The following Lemmas 5.7, 5.8, and 5.9, are auxiliary results which are required in the proof of the main lemmas in the text.

Lemma 5.7. *For $\gamma \in \mathbb{R}_+$, we have*

$$\begin{aligned} \text{maximise } \beta^\top \xi - \gamma \|\xi\|_p &= \text{minimise } \beta^\top \xi + \gamma \|\xi\|_p = \text{maximise } 0 \\ \text{subject to } \xi \in \mathbb{R}^n &\quad \text{subject to } \xi \in \mathbb{R}^n \quad \text{subject to } \|\beta\|_q \leq \gamma. \end{aligned}$$

Proof. Applying an epigraph reformulation to the left-hand side problem yields

$$\begin{aligned} \text{minimise } \beta^\top \xi + \gamma \theta \\ \text{subject to } \xi \in \mathbb{R}^n, \theta \in \mathbb{R} \\ \|\xi\|_p \leq \theta. \end{aligned}$$

Lagrangian (if the norm constraint is thought of as membership in the quadratic cone) is

$$\beta^\top \xi + \gamma \theta - \bar{\xi}^\top \xi - \bar{\theta} \theta.$$

The dual problem is

$$\begin{aligned} \text{maximise } 0 \\ \text{subject to } \bar{\xi} \in \mathbb{R}^n, \bar{\theta} \in \mathbb{R} \\ \beta - \bar{\xi} = 0 \\ \gamma - \bar{\theta} = 0 \\ \|\bar{\xi}\|_q \leq \bar{\theta}, \end{aligned}$$

which concludes the proof. □

Lemma 5.8. For $\beta \in \mathbb{R}^n$, $\gamma \in \mathbb{R}_+$, and $\mathbf{b}_l < \mathbf{0} < \mathbf{b}_u$, we have

$$\begin{aligned} \text{minimise} \quad & \beta^\top \mathbf{v} + \gamma \|\mathbf{v}\|_p = \text{maximise} \quad \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u \\ \text{subject to} \quad & \mathbf{v} \in \mathbb{R}^n \quad \quad \quad \text{subject to} \quad \bar{\mathbf{b}}_l \in \mathbb{R}_+^n, \bar{\mathbf{b}}_u \in \mathbb{R}_+^n \\ & \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u \quad \quad \quad \|\beta + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l)\|_q \leq \gamma, \end{aligned}$$

where $1/p + 1/q = 1$.

Proof. Applying an epigraph transformation, the left-hand side problem becomes

$$\begin{aligned} \text{minimise} \quad & \beta^\top \mathbf{v} + \gamma \theta \\ \text{subject to} \quad & \mathbf{v} \in \mathbb{R}^n, \theta \in \mathbb{R} \\ & \|\mathbf{v}\|_p \leq \theta \\ & \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u. \end{aligned}$$

Introducing dual variables, the Lagrangian of this problem is

$$\beta^\top \mathbf{v} + \gamma \theta - (\bar{\mathbf{v}}^\top \mathbf{v} + \bar{\theta} \theta) + \bar{\mathbf{b}}_l^\top (\mathbf{b}_l - \mathbf{v}) + \bar{\mathbf{b}}_u^\top (\mathbf{v} - \mathbf{b}_u).$$

The dual problem then is

$$\begin{aligned} \text{maximise} \quad & \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u = \text{maximise} \quad \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u \\ \text{subject to} \quad & \bar{\mathbf{v}} \in \mathbb{R}^n, \bar{\theta} \in \mathbb{R}, \bar{\mathbf{b}}_l \in \mathbb{R}_+^n, \bar{\mathbf{b}}_u \in \mathbb{R}_+^n \quad \quad \quad \text{subject to} \quad \bar{\mathbf{b}}_l \in \mathbb{R}_+^n, \bar{\mathbf{b}}_u \in \mathbb{R}_+^n \\ & \bar{\mathbf{v}} = \beta + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l) \quad \quad \quad \|\beta + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l)\|_q \leq \gamma, \\ & \bar{\theta} = \gamma \\ & \|\bar{\mathbf{v}}\|_q \leq \bar{\theta} \end{aligned}$$

where the equality follows from substituting out $\bar{\mathbf{v}}$ and $\bar{\gamma}$. □

Lemma 5.9. For $\beta \in \mathbb{R}^n$, $\gamma \in \mathbb{R}_+$, and $\mathbf{b}_l < \mathbf{0} < \mathbf{b}_u$, we have

$$\begin{aligned}
& \text{minimise} && \beta^\top \mathbf{v} + \gamma \|\mathbf{v}\|_p && = \text{maximise} && \boldsymbol{\lambda}^\top (\mathbf{u} - \mathbf{f}) - \delta \|\boldsymbol{\lambda}\|_1 + \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u \\
& \text{subject to} && \mathbf{v} \in \mathbb{R}^n && \text{subject to} && \boldsymbol{\lambda} \in \mathbb{R}^n \\
& && \|(\mathbf{u} - \mathbf{f}) + \mathbf{v}\|_\infty \leq \delta && && \|\beta + \boldsymbol{\lambda} + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l)\|_q \leq \gamma, \\
& && \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u && &&
\end{aligned}$$

where $1/p + 1/q = 1$.

Proof. After introducing auxiliary variables, the left-hand side problem becomes

$$\begin{aligned}
& \text{minimise} && \beta^\top \mathbf{v} + \gamma \theta \\
& \text{subject to} && \mathbf{v} \in \mathbb{R}^n, \theta \in \mathbb{R}, \mathbf{z} \in \mathbb{R}^n, d \in \mathbb{R} \\
& && \|\mathbf{v}\|_p \leq \theta \\
& && \|\mathbf{z}\|_\infty \leq d \quad \mathbf{z} = (\mathbf{u} - \mathbf{f}) + \mathbf{v} \quad d = \delta \\
& && \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u.
\end{aligned}$$

Introducing dual variables, the Lagrangian of this problem is

$$\beta^\top \mathbf{v} + \gamma \theta - (\bar{\mathbf{v}}^\top \mathbf{v} + \bar{\theta} \theta) - (\bar{\mathbf{z}}^\top \mathbf{z} + \bar{d} d) + \boldsymbol{\lambda}^\top ((\mathbf{u} - \mathbf{f}) + \mathbf{v} - \mathbf{z}) + \lambda_0 (d - \delta) + \bar{\mathbf{b}}_l^\top (\mathbf{b}_l - \mathbf{v}) + \bar{\mathbf{b}}_u^\top (\mathbf{v} - \mathbf{b}_u).$$

The dual problem is

$$\begin{array}{ll}
\text{maximise} & \boldsymbol{\lambda}^\top (\mathbf{u} - \mathbf{f}) - \delta \lambda_0 + \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u \\
\text{subject to} & \bar{\mathbf{z}} \in \mathbb{R}^n, \bar{d} \in \mathbb{R}, \bar{\mathbf{v}} \in \mathbb{R}^n, \bar{\theta} \in \mathbb{R}, \\
& \boldsymbol{\lambda} \in \mathbb{R}^n, \lambda_0 \in \mathbb{R}, \bar{\mathbf{b}}_l \in \mathbb{R}_+^n, \bar{\mathbf{b}}_u \in \mathbb{R}_+^n \\
& \bar{\mathbf{v}} = \boldsymbol{\beta} + \boldsymbol{\lambda} + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l) \\
& \bar{\theta} = \gamma \\
& \bar{\mathbf{z}} = \boldsymbol{\lambda} \\
& \bar{d} = \lambda_0 \\
& \|\bar{\mathbf{v}}\|_q \leq \bar{\theta} \\
& \|\bar{\mathbf{z}}\|_1 \leq \bar{d}
\end{array}
= \text{maximise} \quad \boldsymbol{\lambda}^\top (\mathbf{u} - \mathbf{f}) - \delta \lambda_0 + \bar{\mathbf{b}}_l^\top \mathbf{b}_l - \bar{\mathbf{b}}_u^\top \mathbf{b}_u$$

$$\begin{array}{ll}
\text{subject to} & \boldsymbol{\lambda} \in \mathbb{R}^n, \lambda_0 \in \mathbb{R}, \bar{\mathbf{b}}_l \in \mathbb{R}_+^n, \bar{\mathbf{b}}_u \in \mathbb{R}_+^n \\
& \|\boldsymbol{\beta} + \boldsymbol{\lambda} + (\bar{\mathbf{b}}_u - \bar{\mathbf{b}}_l)\|_q \leq \gamma \\
& \|\boldsymbol{\lambda}\|_1 \leq \lambda_0,
\end{array}$$

where the equality follows from eliminating $\bar{\mathbf{v}}$, $\bar{\theta}$, $\bar{\mathbf{z}}$, and $\bar{d}$. The statement of the lemma now follows directly. $\square$

Lemma (5.2). For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have

$$\begin{array}{ll}
\inf & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \iff \exists \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\
\text{s.t.} & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \\
& \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \\
& \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_p] \leq \sigma
\end{array}
\left\{ \begin{array}{l} 1 \leq \tau_j^+ (\delta + (f_j - u_j)) \\ 1 \leq \tau_j^- (\delta - (f_j - u_j)) \end{array} \right\} \forall j = 1, \dots, n$$

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n} \max_{j=1, \dots, n} \{ \|\boldsymbol{\beta}\|_q, \|\boldsymbol{\beta} + \tau_j^+ \mathbf{e}_j\|_q, \|\boldsymbol{\beta} - \tau_j^- \mathbf{e}_j\|_q \} \leq \frac{\varepsilon}{\sigma},$$

where $\mathbf{e}_j$ is the j th standard basis vector.

Proof. Dualising the left-hand side problem, the chance constraint is equivalent to

$$\begin{array}{ll}
\sup & \alpha - \sigma \gamma \\
\text{s.t.} & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ \\
& \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p \leq \mathbf{1}_{\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \mathbf{v} \leq \mathbf{f} + \delta \mathbf{1}} \quad \forall \mathbf{v} \in \mathbb{R}^n.
\end{array}$$

Separating out the indicator function, we obtain

$$\begin{aligned}
& \sup \quad \alpha - \sigma\gamma && \geq 1 - \varepsilon. \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p \leq 1 \quad \forall \mathbf{v} \in \mathbb{R}^n \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p \leq 0 \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n : v_j > (f_j - u_j) + \delta \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p \leq 0 \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n : v_j < (f_j - u_j) - \delta
\end{aligned}$$

Since $\alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p$ is continuous in $\mathbf{v}$ and the last two semi-infinite constraints cannot be vacuously satisfied (i.e. are not redundant), we can apply a standard continuity argument to replace the strict inequalities with their non-strict variants. We then apply the Farkas Lemma to obtain equivalent semi-infinite constraints which must hold for all $\mathbf{v} \in \mathbb{R}^n$

$$\begin{aligned}
& \sup \quad \alpha - \sigma\gamma && \geq 1 - \varepsilon. \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p \leq 1 \quad \forall \mathbf{v} \in \mathbb{R}^n \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p + \tau_j^+ v_j \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n \\
& \quad \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \|\mathbf{v}\|_p - \tau_j^- v_j \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n
\end{aligned}$$

The semi-infinite constraints can be viewed as unconstrained maximisations over $\mathbf{v} \in \mathbb{R}^n$, which can equivalently be written as a finite number of constraints using Lemma 5.7

$$\begin{aligned}
& \sup \quad \alpha - \sigma\gamma && \geq 1 - \varepsilon, \\
& \text{s.t.} \quad \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\
& \quad \alpha \leq 1 \\
& \quad \left. \begin{aligned} & \|\boldsymbol{\beta}\|_q \leq \gamma \\ & \alpha \leq \tau_j^+ (\delta + (f_j - u_j)) \\ & \alpha \leq \tau_j^- (\delta - (f_j - u_j)) \\ & \|\boldsymbol{\beta} + \tau_j^+ \mathbf{e}_j\|_q \leq \gamma, \quad \|\boldsymbol{\beta} - \tau_j^- \mathbf{e}_j\|_q \leq \gamma \end{aligned} \right\} \forall j = 1, \dots, n
\end{aligned}$$

where q is chosen such that $1/p + 1/q = 1$. Since $\varepsilon < 1$, we can, without loss of generality, assume that $\alpha = 1$ at optimality, otherwise the worst-case probability will be zero and the chance constraint is violated. We can thus reformulate the non-convex constraints into hyperbolic constraints

$$\begin{aligned}
& \inf \quad \sigma\gamma && \leq \varepsilon. \\
& \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\
& \quad \left. \begin{aligned} & \|\boldsymbol{\beta}\|_q \leq \gamma \\ & 1 \leq \tau_j^+ (\delta + (f_j - u_j)) \\ & 1 \leq \tau_j^- (\delta - (f_j - u_j)) \\ & \|\boldsymbol{\beta} + \tau^+ \mathbf{e}_j\|_q \leq \gamma, \quad \|\boldsymbol{\beta} - \tau^- \mathbf{e}_j\|_q \leq \gamma \end{aligned} \right\} \forall j = 1, \dots, n
\end{aligned}$$

The result now follows directly by eliminating γ . □

Lemma (5.3). For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have

$$\begin{aligned}
& \inf \quad \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \quad \Longleftrightarrow \quad \|\mathbf{f} - \mathbf{u}\|_\infty \leq (\delta(\delta - \sigma/\varepsilon))^{\frac{1}{2}}. \\
& \text{s.t.} \quad \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \\
& \quad \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \\
& \quad \mathbb{E}_{\mathbb{P}}[\|\tilde{\mathbf{v}}\|_1] \leq \sigma
\end{aligned}$$

Proof. Directly applying Lemma 5.2 with $p = 1$, we find that the chance constraint is equivalent to

$$\begin{aligned}
& \exists \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\
& \quad \left. \begin{aligned} & 1 \leq \tau_j^+ (\delta + (f_j - u_j)) \\ & 1 \leq \tau_j^- (\delta - (f_j - u_j)) \end{aligned} \right\} \forall j = 1, \dots, n \\
& \quad \min_{\boldsymbol{\beta} \in \mathbb{R}^n} \max \{ \|\boldsymbol{\beta}\|_\infty, \|\boldsymbol{\beta} + \boldsymbol{\tau}^+\|_\infty, \|\boldsymbol{\beta} - \boldsymbol{\tau}^-\|_\infty \} \leq \varepsilon/\sigma,
\end{aligned}$$

where some redundant constraints due to the ∞ -norm have been removed. We next perform the minimisation over $\boldsymbol{\beta}$ analytically using convex duality. First, by definition of ∞ -norm, the left-hand side of

the last constraint is equivalent to

$$\min_{\beta \in \mathbb{R}^n} \max_{i=1, \dots, n} \max \{ |\beta_i|, |\beta_i + \tau_i^+|, |\beta_i - \tau_i^-| \} = \min_{\beta \in \mathbb{R}^n} \max \{ \|\beta + \tau^+\|_\infty, \|\beta - \tau^-\|_\infty \}.$$

Since $\tau^+ \geq \mathbf{0}$ and $\tau^- \geq \mathbf{0}$, it is impossible (shown by a case analysis) that $|\beta_i| > |\beta_i + \tau_i^+|$ and $|\beta_i| > |\beta_i - \tau_i^-|$ simultaneously. Therefore the equality above holds. We next use convex duality to analytically solve the resulting minimisation over β . To this end, we introduce the auxiliary variable $\theta \in \mathbb{R}$ and write down the equivalent epigraph formulation

$$\begin{aligned} \min \quad & \theta & = \min \quad & \theta \\ \text{s.t.} \quad & \theta \in \mathbb{R}, \beta \in \mathbb{R}^n & \text{s.t.} \quad & \theta \in \mathbb{R}, \theta^+ \in \mathbb{R}, \theta^- \in \mathbb{R}, \beta \in \mathbb{R}^n, \mathbf{z}^+ \in \mathbb{R}^n, \mathbf{z}^- \in \mathbb{R}^n \\ & \|\beta + \tau^+\|_\infty \leq \theta & & \|\mathbf{z}^+\|_\infty \leq \theta^+ \quad \mathbf{z}^+ = \beta + \tau^+ \quad \theta = \theta^+ \\ & \|\beta - \tau^-\|_\infty \leq \theta & & \|\mathbf{z}^-\|_\infty \leq \theta^- \quad \mathbf{z}^- = \beta - \tau^- \quad \theta = \theta^- \end{aligned}$$

Here, we have introduced further auxiliary variables to simplify the dualisation. The Lagrangian is

$$\begin{aligned} & \theta - (\bar{\mathbf{z}}^{+\top} \mathbf{z}^+ + \bar{\theta}^+ \theta^+) + \mathbf{x}^{+\top} (\beta + \tau^+ - \mathbf{z}^+) + y^+ (\theta^+ - \theta) \\ & - (\bar{\mathbf{z}}^{-\top} \mathbf{z}^- + \bar{\theta}^- \theta^-) + \mathbf{x}^{-\top} (\beta - \tau^- - \mathbf{z}^-) + y^- (\theta^- - \theta). \end{aligned}$$

The dual problem is

$$\begin{aligned} \max \quad & \mathbf{x}^{+\top} \tau^+ - \mathbf{x}^{-\top} \tau^- \\ \text{s.t.} \quad & \bar{\theta}^+ \in \mathbb{R}, \bar{\theta}^- \in \mathbb{R}, \bar{\mathbf{z}}^+ \in \mathbb{R}^n, \bar{\mathbf{z}}^- \in \mathbb{R}^n, \mathbf{x}^+ \in \mathbb{R}^n, \mathbf{x}^- \in \mathbb{R}^n, y^+ \in \mathbb{R}, y^- \in \mathbb{R} \\ & 1 - y^+ - y^- = 0 \\ & \bar{\theta}^+ = y^+ \quad \bar{\mathbf{z}}^+ = -\mathbf{x}^+ \\ & \bar{\theta}^- = y^- \quad \bar{\mathbf{z}}^- = -\mathbf{x}^- \\ & \mathbf{x}^+ + \mathbf{x}^- = \mathbf{0} \\ & \|\bar{\mathbf{z}}^+\|_1 \leq \bar{\theta}^+ \quad \|\bar{\mathbf{z}}^-\|_1 \leq \bar{\theta}^-. \end{aligned}$$

Eliminating all variables with a bar (using the equality constraints) yields

$$\begin{aligned}
& \max \quad \mathbf{x}^{+\top} \boldsymbol{\tau}^+ - \mathbf{x}^{-\top} \boldsymbol{\tau}^- \\
& \text{s.t.} \quad \mathbf{x}^+ \in \mathbb{R}^n, \mathbf{x}^- \in \mathbb{R}^n, y^+ \in \mathbb{R}, y^- \in \mathbb{R} \\
& \quad 1 - y^+ - y^- = 0 \\
& \quad \mathbf{x}^+ + \mathbf{x}^- = \mathbf{0} \\
& \quad \|\mathbf{x}^+\|_1 \leq y^+ \quad \|\mathbf{x}^-\|_1 \leq y^-.
\end{aligned}$$

Using the equality constraints, we eliminate $\mathbf{x}^-$ and y^-

$$\begin{aligned}
& \max \quad \mathbf{x}^{+\top} (\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-) & = \max \quad \mathbf{x}^\top (\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-) & = \frac{1}{2} \|\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-\|_\infty. \\
& \text{s.t.} \quad \mathbf{x}^+ \in \mathbb{R}^n, y^+ \in \mathbb{R} & \text{s.t.} \quad \mathbf{x} \in \mathbb{R}^n \\
& \quad \|\mathbf{x}^+\|_1 \leq y^+ \quad \|\mathbf{x}^+\|_1 \leq 1 - y^+ & \quad \|\mathbf{x}\|_1 \leq 1/2
\end{aligned}$$

Here, the first equality holds because $\max_{y \in \mathbb{R}} \min\{y, 1 - y\} = 1/2$, and the second equality from the fact that, at optimality, all weight will be placed on the largest component of $\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-$. We return to the chance constraint, which now becomes

$$\begin{aligned}
& \exists \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\
& \left. \begin{aligned} 1 &\leq \tau_j^+ (\delta + (f_j - u_j)) \\ 1 &\leq \tau_j^- (\delta - (f_j - u_j)) \end{aligned} \right\} \forall j = 1, \dots, n \\
& \|\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-\|_\infty \leq 2\varepsilon/\sigma.
\end{aligned}$$

Clearly, $\tau_j^+ > 0$ and $\tau_j^- > 0$ since otherwise the hyperbolic constraints are violated. Similarly, $\delta + (f_j - u_j) > 0$ and $\delta - (f_j - u_j) > 0$. We therefore divide the first and second set of hyperbolic constraints by $(\delta + (f_j - u_j))$ and $(\delta - (f_j - u_j))$ respectively, to isolate τ_j^+ and τ_j^- , and notice that now both τ_j^+ and τ_j^- can be reduced until all these constraints become binding (without losing feasibility)

$$\frac{1}{\delta + (f_j - u_j)} + \frac{1}{\delta - (f_j - u_j)} \leq 2\varepsilon/\sigma \quad \forall j = 1, \dots, n$$

Putting the fractions over a common denominator, we obtain

$$\frac{\delta}{\delta^2 - (f_j - u_j)^2} \leq \varepsilon/\sigma \quad \forall j = 1, \dots, n \quad \Longleftrightarrow \quad \delta \frac{\sigma}{\varepsilon} \leq \delta^2 - (f_j - u_j)^2 \quad \forall j = 1, \dots, n,$$

where the equivalence holds since the denominator is strictly greater than zero. The statement of the lemma now follows directly. $\square$

Lemma (5.4). *For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have*

$$\begin{aligned} \inf \quad & \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon \quad \Longleftrightarrow \quad \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \quad (\boldsymbol{\beta} + (\mathbf{f} - \mathbf{u}))^\top (\boldsymbol{\beta} + (\mathbf{f} - \mathbf{u})) + \sigma \leq \varepsilon \gamma^2 \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}] = \mathbf{0} \quad \|\boldsymbol{\beta}\|_\infty \leq \delta - \gamma. \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{v}}^\top \tilde{\mathbf{v}}] \leq \sigma \end{aligned}$$

Proof. As before, we write down the dual problem and use a continuity argument to obtain the following equivalent problem

$$\begin{aligned} \sup \quad & \alpha - \sigma \gamma \quad \geq 1 - \varepsilon \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} \leq 0 \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n : v_j \geq (f_j - u_j) + \delta \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} \leq 0 \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n : v_j \leq (f_j - u_j) - \delta \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} \leq 1 \quad \forall \mathbf{v} \in \mathbb{R}^n \end{aligned}$$

Applying the Farkas Lemma to first two semi-infinite constraints, we obtain

$$\begin{aligned} \sup \quad & \alpha - \sigma \gamma \quad \geq 1 - \varepsilon. \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} + \tau_j^+ v_j \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} - \tau_j^- v_j \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n, \forall \mathbf{v} \in \mathbb{R}^n \\ & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} \leq 1 \quad \forall \mathbf{v} \in \mathbb{R}^n \end{aligned}$$

Next, we analytically perform the maximisation over $\mathbf{v} \in \mathbb{R}^n$ in both the semi-infinite constraints. To this end we use the following result, which is due to quadratic programming duality. For $\gamma \geq 0$

$$\begin{aligned} \text{maximise} \quad & \alpha + \boldsymbol{\beta}^\top \mathbf{v} - \gamma \mathbf{v}^\top \mathbf{v} = \alpha + \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma}. \\ \text{subject to} \quad & \mathbf{v} \in \mathbb{R}^n \end{aligned}$$

This means that the chance constraint is equivalent to

$$\begin{aligned} \sup \quad & \alpha - \sigma\gamma \quad \geq 1 - \varepsilon. \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\ & \alpha + \frac{1}{4} \frac{(\boldsymbol{\beta} + \tau_j^+ \mathbf{e}_j)^\top (\boldsymbol{\beta} + \tau_j^+ \mathbf{e}_j)}{\gamma} \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ & \alpha + \frac{1}{4} \frac{(\boldsymbol{\beta} - \tau_j^- \mathbf{e}_j)^\top (\boldsymbol{\beta} - \tau_j^- \mathbf{e}_j)}{\gamma} \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n \\ & \alpha + \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma} \leq 1 \end{aligned}$$

Re-writing the quadratic terms yields

$$\begin{aligned} \sup \quad & \alpha - \sigma\gamma \quad \geq 1 - \varepsilon. \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n \\ & \alpha + \frac{1}{4} \frac{\boldsymbol{\beta}_{-j}^\top \boldsymbol{\beta}_{-j}}{\gamma} + \frac{1}{4} \frac{(\beta_j + \tau_j^+)^2}{\gamma} \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ & \alpha + \frac{1}{4} \frac{\boldsymbol{\beta}_{-j}^\top \boldsymbol{\beta}_{-j}}{\gamma} + \frac{1}{4} \frac{(\beta_j - \tau_j^-)^2}{\gamma} \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n \\ & \alpha + \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma} \leq 1 \end{aligned}$$

Without loss of generality we can assume that $\alpha = 1 - \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma}$ at optimality (otherwise worst-case probability is zero) and thus substitute out α

$$\begin{aligned} \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\ & \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma} + \sigma\gamma \leq \varepsilon \\ & 1 - \frac{1}{4} \frac{\beta_j^2}{\gamma} + \frac{1}{4} \frac{(\beta_j + \tau_j^+)^2}{\gamma} \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ & 1 - \frac{1}{4} \frac{\beta_j^2}{\gamma} + \frac{1}{4} \frac{(\beta_j - \tau_j^-)^2}{\gamma} \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n \end{aligned}$$

Next, we can remove the terms involving β_j^2 since they cancel out when the quadratic terms are multiplied out:

$$\begin{aligned} \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\ \frac{1}{4} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\gamma} + \sigma \gamma \leq \varepsilon \\ 1 + \frac{1}{4} \frac{(\tau_j^+)^2 + 2\beta_j \tau_j^+}{\gamma} \leq \tau_j^+ (\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ 1 + \frac{1}{4} \frac{(\tau_j^-)^2 - 2\beta_j \tau_j^-}{\gamma} \leq \tau_j^- (\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n. \end{aligned}$$

We now multiply all constraints by 4γ and divide the second (third) constraint by τ_j^+ (τ_j^-). This is possible since we may assume that $\boldsymbol{\tau}^+ > \mathbf{0}$ and $\boldsymbol{\tau}^- > \mathbf{0}$ at optimality, since otherwise the worst-case probability will again be zero. Thus we obtain

$$\begin{aligned} \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n, \boldsymbol{\tau}^- \in \mathbb{R}_+^n : \\ \boldsymbol{\beta}^\top \boldsymbol{\beta} + 4\sigma\gamma^2 \leq 4\varepsilon\gamma \\ 4\gamma(\tau_j^+)^{-1} + \tau_j^+ + 2\beta_j \leq 4\gamma(\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ 4\gamma(\tau_j^-)^{-1} + \tau_j^- - 2\beta_j \leq 4\gamma(\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n. \end{aligned}$$

Removing $\boldsymbol{\tau}^+$ and $\boldsymbol{\tau}^-$ by minimising the left-hand-side of the second and third constraint (using first-order optimality conditions for the unconstrained problems), we find that $\tau_j^+ = \tau_j^- = 2\gamma^{\frac{1}{2}}$, which is still feasible since $\gamma \geq 0$. This leads to the simplified set of constraints

$$\begin{aligned} \exists \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : \\ \boldsymbol{\beta}^\top \boldsymbol{\beta} + 4\sigma\gamma^2 \leq 4\varepsilon\gamma \\ 4\gamma^{\frac{1}{2}} + 2\beta_j \leq 4\gamma(\delta + (f_j - u_j)) \quad \forall j = 1, \dots, n \\ 4\gamma^{\frac{1}{2}} - 2\beta_j \leq 4\gamma(\delta - (f_j - u_j)) \quad \forall j = 1, \dots, n. \end{aligned}$$

Dividing the second and third constraints by 4γ , the first constraint by $4\gamma^2$, and redefining β , we obtain

$$\begin{array}{ll}
\exists \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : & \iff \exists \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : \\
\beta^\top \beta + \sigma \leq \varepsilon \gamma^{-1} & \beta^\top \beta + \sigma \leq \varepsilon \gamma^2 \\
\gamma^{-\frac{1}{2}} + \beta_j \leq \delta + (f_j - u_j) \quad \forall j = 1, \dots, n & \gamma + \beta_j \leq \delta + (f_j - u_j) \quad \forall j = 1, \dots, n \\
\gamma^{-\frac{1}{2}} - \beta_j \leq \delta - (f_j - u_j) \quad \forall j = 1, \dots, n, & \gamma - \beta_j \leq \delta - (f_j - u_j) \quad \forall j = 1, \dots, n,
\end{array}$$

where the equivalence follows by redefining γ . The last two constraints give the following bounds on β_j

$$\begin{aligned}
& \exists \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ : \\
& \beta^\top \beta + \sigma \leq \varepsilon \gamma^2 \\
& (f_j - u_j) + \gamma - \delta \leq \beta_j \leq (f_j - u_j) + \delta - \gamma \quad \forall j = 1, \dots, n.
\end{aligned}$$

The domain of γ and the last constraint imply that $0 \leq \gamma \leq \delta$. The redefinition $\beta \leftarrow \beta - (\mathbf{f} - \mathbf{u})$ now yields the statement of the lemma. $\square$

Lemma (5.5). *For $0 < \varepsilon < 1$, $\delta > 0$, and $\sigma > 0$, we have*

$$\begin{array}{ll}
\sup \quad \mathbb{P}(\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \tilde{\mathbf{v}} \leq \mathbf{f} + \delta \mathbf{1}) \geq 1 - \varepsilon & \iff \exists \mathbf{v} \in \mathbb{R}^n : \\
\text{s.t.} \quad \mathbb{P} \in \mathcal{P}_0(\mathcal{V}) & \mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \mathbf{v} \leq \mathbf{f} + \delta \mathbf{1} \\
\mathbb{E}[\tilde{\mathbf{v}}] = \mathbf{0} & \|\mathbf{v}\|_p \leq \frac{1}{2}\sigma/(1 - \varepsilon) \\
\mathbb{E}[\|\tilde{\mathbf{v}}\|_p] \leq \sigma & \max \left\{ \frac{\varepsilon}{1 - \varepsilon} \mathbf{b}_l, -\mathbf{b}_u \right\} \leq \mathbf{v} \leq \min \left\{ -\mathbf{b}_l, \frac{\varepsilon}{1 - \varepsilon} \mathbf{b}_u \right\},
\end{array}$$

where $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u\}$ and $\mathbf{b}_l < \mathbf{0} < \mathbf{b}_u$.

Proof. Dualising the left-hand side problem yields

$$\begin{array}{ll}
\inf \quad \alpha + \sigma \gamma & \geq 1 - \varepsilon. \\
\text{s.t.} \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ & \\
\alpha + \beta^\top \mathbf{v} + \gamma \|\mathbf{v}\|_p \geq \mathbf{1}_{\mathbf{f} - \delta \mathbf{1} \leq \mathbf{u} + \mathbf{v} \leq \mathbf{f} + \delta \mathbf{1}} & \forall \mathbf{v} \in \mathbb{R}^n : \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u
\end{array}$$

Separating out the indicator function, we obtain

$$\begin{aligned}
\inf \quad & \alpha + \sigma\gamma && \geq 1 - \varepsilon. \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}_+ \\
& \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} + \gamma \|\boldsymbol{\xi}\|_p \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n : \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u \\
& \alpha + \boldsymbol{\beta}^\top \boldsymbol{\xi} + \gamma \|\boldsymbol{\xi}\|_p \geq 1 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n : \|(\mathbf{u} - \mathbf{f}) + \boldsymbol{\xi}\|_\infty \leq \delta \wedge \mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u
\end{aligned}$$

We now minimise both semi-infinite constraints over $\boldsymbol{\xi}$ using the auxiliary Lemmas 5.8 and 5.9

$$\begin{aligned}
\inf \quad & \alpha + \sigma\gamma && \geq 1 - \varepsilon, \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^n, \mathbf{x}_1 \in \mathbb{R}_+^n, \mathbf{y}_1 \in \mathbb{R}_+^n, \mathbf{x}_2 \in \mathbb{R}_+^n, \mathbf{y}_2 \in \mathbb{R}_+^n \\
& \alpha + \mathbf{x}_1^\top \mathbf{b}_l - \mathbf{y}_1^\top \mathbf{b}_u \geq 0 \\
& \alpha + \mathbf{x}_2^\top \mathbf{b}_l - \mathbf{y}_2^\top \mathbf{b}_u + \boldsymbol{\lambda}^\top (\mathbf{u} - \mathbf{f}) - \delta \|\boldsymbol{\lambda}\|_1 \geq 1 \\
& \|\boldsymbol{\beta} + (\mathbf{x}_1 - \mathbf{y}_1)\|_q \leq \gamma \\
& \|\boldsymbol{\beta} + \boldsymbol{\lambda} + (\mathbf{x}_2 - \mathbf{y}_2)\|_q \leq \gamma
\end{aligned}$$

where $1/p + 1/q = 1$. Introducing auxiliary variables this problem is equivalent to

$$\begin{aligned}
\inf \quad & \alpha + \sigma\gamma && \geq 1 - \varepsilon. \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^n, \mathbf{x}_1 \in \mathbb{R}_+^n, \mathbf{y}_1 \in \mathbb{R}_+^n \\
& \mathbf{x}_2 \in \mathbb{R}_+^n, \mathbf{y}_2 \in \mathbb{R}_+^n, \mathbf{z}_1 \in \mathbb{R}^n, \mathbf{z}_2 \in \mathbb{R}^n, g \in \mathbb{R}, \theta \in \mathbb{R} \\
& \alpha + \mathbf{x}_1^\top \mathbf{b}_l - \mathbf{y}_1^\top \mathbf{b}_u \geq 0 \\
& \alpha + \mathbf{x}_2^\top \mathbf{b}_l - \mathbf{y}_2^\top \mathbf{b}_u + \boldsymbol{\lambda}^\top (\mathbf{u} - \mathbf{f}) - \delta\theta \geq 1 \quad \|\boldsymbol{\lambda}\|_1 \leq \theta \\
& \|\mathbf{z}_1\|_q \leq \gamma \quad \mathbf{z}_1 = \boldsymbol{\beta} + (\mathbf{x}_1 - \mathbf{y}_1) \\
& \|\mathbf{z}_2\|_q \leq g \quad \mathbf{z}_2 = \boldsymbol{\beta} + \boldsymbol{\lambda} + (\mathbf{x}_2 - \mathbf{y}_2) \\
& \gamma = g
\end{aligned}$$

The Lagrangian is

$$\begin{aligned}
& \alpha + \sigma\gamma - \eta_1(\alpha + \mathbf{x}_1^\top \mathbf{b}_l - \mathbf{y}_1^\top \mathbf{b}_u) - \eta_2(\alpha + \mathbf{x}_2^\top \mathbf{b}_l - \mathbf{y}_2^\top \mathbf{b}_u + \boldsymbol{\lambda}^\top (\mathbf{f} - \mathbf{u}) - \delta\theta - 1) \\
& + \eta_3(\gamma - g) + \boldsymbol{\eta}_4^\top (\boldsymbol{\beta} + (\mathbf{x}_1 - \mathbf{y}_1) - \mathbf{z}_1) + \boldsymbol{\eta}_5^\top (\boldsymbol{\beta} + (\mathbf{x}_2 - \mathbf{y}_2) + \boldsymbol{\lambda} - \mathbf{z}_2) \\
& - (\bar{\boldsymbol{\lambda}}^\top \boldsymbol{\lambda} + \bar{\theta}\theta) - (\bar{\mathbf{z}}_1^\top \mathbf{z}_1 + \bar{\gamma}\gamma) - (\bar{\mathbf{z}}_2^\top \mathbf{z}_2 + \bar{g}g) - \bar{\mathbf{x}}_1^\top \mathbf{x}_1 - \bar{\mathbf{x}}_2^\top \mathbf{x}_2 - \bar{\mathbf{y}}_1^\top \mathbf{y}_1 - \bar{\mathbf{y}}_2^\top \mathbf{y}_2.
\end{aligned}$$

The dual problem is

$$\begin{array}{ll}
\max & \eta_2 \\
\text{s.t.} & 1 - \eta_1 - \eta_2 = 0 \\
& \boldsymbol{\eta}_4 + \boldsymbol{\eta}_5 = \mathbf{0} \\
& \bar{\gamma} = \sigma + \eta_3 \\
& \bar{\boldsymbol{\lambda}} = \boldsymbol{\eta}_5 - \eta_2(\mathbf{f} - \mathbf{u}) \\
& \bar{\mathbf{x}}_1 = \boldsymbol{\eta}_4 - \eta_1 \mathbf{b}_l \quad \bar{\mathbf{y}}_1 = \eta_1 \mathbf{b}_u - \boldsymbol{\eta}_4 \\
& \bar{\mathbf{x}}_2 = \boldsymbol{\eta}_5 - \eta_2 \mathbf{b}_l \quad \bar{\mathbf{y}}_2 = \eta_2 \mathbf{b}_u - \boldsymbol{\eta}_5 \\
& \bar{\mathbf{z}}_1 = -\boldsymbol{\eta}_4 \quad \bar{\mathbf{z}}_2 = -\boldsymbol{\eta}_5 \\
& \bar{g} = -\eta_3 \quad \bar{\theta} = \eta_2 \delta \\
& \eta_1 \geq 0 \quad \eta_2 \geq 0 \\
& \bar{\mathbf{x}}_1 \geq \mathbf{0} \quad \bar{\mathbf{y}}_1 \geq \mathbf{0} \\
& \bar{\mathbf{x}}_2 \geq \mathbf{0} \quad \bar{\mathbf{y}}_2 \geq \mathbf{0} \\
& \|\bar{\boldsymbol{\lambda}}\|_\infty \leq \bar{\theta} \\
& \|\bar{\mathbf{z}}_1\|_p \leq \bar{\gamma} \quad \|\bar{\mathbf{z}}_2\|_p \leq \bar{g}
\end{array}
\qquad
\begin{array}{ll}
= \max & \eta_2 \\
\text{s.t.} & 1 - \eta_1 - \eta_2 = 0 \\
& \boldsymbol{\eta}_4 + \boldsymbol{\eta}_5 = \mathbf{0} \\
& \eta_1 \geq 0 \quad \eta_2 \geq 0 \\
& \boldsymbol{\eta}_4 - \eta_1 \mathbf{b}_l \geq \mathbf{0} \quad \eta_1 \mathbf{b}_u - \boldsymbol{\eta}_4 \geq \mathbf{0} \\
& \boldsymbol{\eta}_5 - \eta_2 \mathbf{b}_l \geq \mathbf{0} \quad \eta_2 \mathbf{b}_u - \boldsymbol{\eta}_5 \geq \mathbf{0} \\
& \|\boldsymbol{\eta}_5 - \eta_2(\mathbf{f} - \mathbf{u})\|_\infty \leq \eta_2 \delta \\
& \|\boldsymbol{\eta}_4\|_p \leq \sigma + \eta_3 \quad \|\boldsymbol{\eta}_5\|_p \leq -\eta_3,
\end{array}$$

where the equality holds by eliminating constraints using the equality constraints. We now remove

further variables $\boldsymbol{\eta}_5$ and η_3 in the same way

$$\begin{array}{ll}
\max & \eta_2 \\
\text{s.t.} & 1 - \eta_1 - \eta_2 = 0 \\
& \eta_1 \geq 0 \quad \eta_2 \geq 0 \\
& \boldsymbol{\eta}_4 - \eta_1 \mathbf{b}_l \geq \mathbf{0} \quad \eta_1 \mathbf{b}_u - \boldsymbol{\eta}_4 \geq \mathbf{0} \\
& -\boldsymbol{\eta}_4 - \eta_2 \mathbf{b}_l \geq \mathbf{0} \quad \eta_2 \mathbf{b}_u + \boldsymbol{\eta}_4 \geq \mathbf{0} \\
& \|\boldsymbol{\eta}_4 + \eta_2(\mathbf{f} - \mathbf{u})\|_\infty \leq \eta_2 \delta \\
& \|\boldsymbol{\eta}_4\|_p \leq \sigma + \eta_3 \quad \|\boldsymbol{\eta}_4\|_p \leq -\eta_3
\end{array}
= \max \quad \eta_2
\begin{array}{ll}
\text{s.t.} & 1 - \eta_1 - \eta_2 = 0 \\
& \eta_1 \geq 0 \quad \eta_2 \geq 0 \\
& \eta_1 \mathbf{b}_l \leq \boldsymbol{\eta}_4 \leq \eta_1 \mathbf{b}_u \quad -\eta_2 \mathbf{b}_u \leq \boldsymbol{\eta}_4 \leq -\eta_2 \mathbf{b}_l \\
& \|\boldsymbol{\eta}_4 + \eta_2(\mathbf{f} - \mathbf{u})\|_\infty \leq \eta_2 \delta \\
& \|\boldsymbol{\eta}_4\|_p \leq \sigma + \eta_3 \quad \|\boldsymbol{\eta}_4\|_p \leq -\eta_3.
\end{array}$$

Next, we eliminate η_2 using the equality constraint and remove η_3 since $\max_{\eta_3 \in \mathbb{R}} \min\{\sigma + \eta_3, -\eta_3\} = \frac{1}{2}\sigma$

$$\begin{array}{ll}
\max & 1 - \eta_1 \\
\text{s.t.} & 0 \leq \eta_1 \leq 1 \\
& \eta_1 \mathbf{b}_l \leq \boldsymbol{\eta}_4 \leq \eta_1 \mathbf{b}_u \quad -(1 - \eta_1) \mathbf{b}_u \leq \boldsymbol{\eta}_4 \leq -(1 - \eta_1) \mathbf{b}_l \\
& \|\boldsymbol{\eta}_4 + (1 - \eta_1)(\mathbf{f} - \mathbf{u})\|_\infty \leq (1 - \eta_1) \delta \\
& \|\boldsymbol{\eta}_4\|_p \leq \frac{1}{2}\sigma.
\end{array}$$

Dividing all constraints involving $\boldsymbol{\eta}_4$ by $(1 - \eta_1)$ and conducting the variable transformation $\boldsymbol{\eta}_4 \leftarrow \boldsymbol{\eta}_4 / (1 - \eta_1)$, the chance constraint becomes

$$\begin{array}{ll}
\text{maximise} & 1 - \eta_1 \\
\text{subject to} & 0 \leq \eta_1 \leq 1 \\
& \frac{\eta_1}{1 - \eta_1} \mathbf{b}_l \leq \boldsymbol{\eta}_4 \leq \frac{\eta_1}{1 - \eta_1} \mathbf{b}_u \quad -\mathbf{b}_u \leq \boldsymbol{\eta}_4 \leq -\mathbf{b}_l \\
& \|\boldsymbol{\eta}_4 + (\mathbf{f} - \mathbf{u})\|_\infty \leq \delta \\
& \|\boldsymbol{\eta}_4\|_p \leq \frac{1}{2}\sigma / (1 - \eta_1)
\end{array}
\geq 1 - \varepsilon.$$

Since $\varepsilon < 1$, this problem is equivalent to

$$\begin{aligned}
\exists p \in \mathbb{R}, \mathbf{v} \in \mathbb{R}^n : & \iff \exists \mathbf{v} \in \mathbb{R}^n : \\
0 \leq p \leq \varepsilon & \quad \frac{\varepsilon}{1-\varepsilon} \mathbf{b}_l \leq \mathbf{v} \leq \frac{\varepsilon}{1-\varepsilon} \mathbf{b}_u \quad -\mathbf{b}_u \leq \mathbf{v} \leq -\mathbf{b}_l \\
\frac{p}{1-p} \mathbf{b}_l \leq \mathbf{v} \leq \frac{p}{1-p} \mathbf{b}_u \quad -\mathbf{b}_u \leq \mathbf{v} \leq -\mathbf{b}_l & \quad \|\mathbf{v} + (\mathbf{f} - \mathbf{u})\|_\infty \leq \delta \\
\|\mathbf{v} + (\mathbf{f} - \mathbf{u})\|_\infty \leq \delta & \quad \|\mathbf{v}\|_p \leq \frac{1}{2}\sigma/(1-\varepsilon), \\
\|\mathbf{v}\|_p \leq \frac{1}{2}\sigma/(1-p) &
\end{aligned}$$

where the equivalence holds since both $1/(1-p)$ and $p/(1-p)$ are increasing in p for $0 < p < 1$ and since $\mathbf{b}_l < \mathbf{0} < \mathbf{b}_u$ (for any feasible solution), which implies that we can increase p until $p = \varepsilon$ without losing feasibility. The statement of the lemma now follows directly. $\square$

Lemma (5.6). For $\delta > 0$, $\kappa > 0$, and $\kappa_l < \mathbf{0} < \kappa_u$, we have

$$\begin{aligned}
\text{minimise} \quad & \sum_{i=1}^n \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{p_1} & = \text{maximise} \quad & -\mathbf{b}^\top \mathbf{f} - \delta \|\mathbf{b}\|_1 - \kappa \|\mathbf{b} + (\bar{\kappa}_l - \bar{\kappa}_u)\|_{q_2} + \bar{\kappa}_l^\top \kappa_l - \bar{\kappa}_u^\top \kappa_u \\
\text{subject to} \quad & \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n, & \text{subject to} \quad & \mathbf{b} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n, \bar{\kappa}_l \in \mathbb{R}_+^n, \bar{\kappa}_u \in \mathbb{R}_+^n \\
& \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n & & \mathbf{b} = \mathbf{T}_x^\top \mathbf{x} + \mathbf{T}_y^\top \mathbf{y} \\
& \mathbf{T}_x \mathbf{u} = \mathbf{x} \quad \mathbf{T}_y \mathbf{u} = \mathbf{y} & & \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{q_1} \leq 1 \quad \forall i = 1, \dots, n, \\
& \|(\mathbf{f} - \mathbf{u}) + \mathbf{v}\|_\infty \leq \delta & & \\
& \|\mathbf{v}\|_{p_2} \leq \kappa & & \\
& \kappa_l \leq \mathbf{v} \leq \kappa_u & &
\end{aligned}$$

where $1/p_i + 1/q_i = 1$.

Proof. As before, we introduce auxiliary variables to facilitate the dualisation

$$\begin{aligned}
& \text{minimise} \quad \boldsymbol{\tau}^\top \mathbf{1} \\
& \text{subject to} \quad \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n, \boldsymbol{\tau} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^n, d \in \mathbb{R}, k \in \mathbb{R} \\
& \quad \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_{p_1} \leq \tau_i \quad \forall i = 1, \dots, n \\
& \quad \mathbf{T}_x \mathbf{u} = \mathbf{x} \quad \mathbf{T}_y \mathbf{u} = \mathbf{y} \\
& \quad \|\mathbf{z}\|_\infty \leq d \quad \mathbf{z} = (\mathbf{f} - \mathbf{u}) + \mathbf{v} \quad d = \delta \\
& \quad \|\mathbf{v}\|_{p_2} \leq k \quad k = \kappa \\
& \quad \boldsymbol{\kappa}_l \leq \mathbf{v} \leq \boldsymbol{\kappa}_u.
\end{aligned}$$

The Lagrangian of this problem is

$$\begin{aligned}
& \boldsymbol{\tau}^\top \mathbf{1} - (\bar{\mathbf{x}}^\top \mathbf{x} + \bar{\mathbf{y}}^\top \mathbf{y} + \bar{\boldsymbol{\tau}}^\top \boldsymbol{\tau}) + \mathbf{a}_x^\top (\mathbf{T}_x \mathbf{u} - \mathbf{x}) + \mathbf{a}_y^\top (\mathbf{T}_y \mathbf{u} - \mathbf{y}) \\
& - (\bar{\mathbf{z}}^\top \mathbf{z} + \bar{d}d) - (\bar{\mathbf{v}}^\top \mathbf{v} + \bar{k}k) + \mathbf{b}^\top (\mathbf{z} + (\mathbf{u} - \mathbf{f}) - \mathbf{v}) \\
& + c_1(d - \delta) + c_2(k - \kappa) + \bar{\boldsymbol{\kappa}}_l^\top (\boldsymbol{\kappa}_l - \mathbf{v}) + \bar{\boldsymbol{\kappa}}_u^\top (\mathbf{v} - \boldsymbol{\kappa}_u).
\end{aligned}$$

The dual problem is

$$\begin{aligned}
& \text{maximise} \quad -\mathbf{b}^\top \mathbf{f} - c_1 \delta - c_2 \kappa + \bar{\boldsymbol{\kappa}}_l^\top \boldsymbol{\kappa}_l - \bar{\boldsymbol{\kappa}}_u^\top \boldsymbol{\kappa}_u = \text{maximise} \quad -\mathbf{b}^\top \mathbf{f} - c_1 \delta - c_2 \kappa + \bar{\boldsymbol{\kappa}}_l^\top \boldsymbol{\kappa}_l - \bar{\boldsymbol{\kappa}}_u^\top \boldsymbol{\kappa}_u \\
& \text{subject to} \quad \mathbf{T}_x^\top \mathbf{a}_x + \mathbf{T}_y^\top \mathbf{a}_y + \mathbf{b} = \mathbf{0} \quad \text{subject to} \quad \mathbf{b} \in \mathbb{R}^n, \bar{\mathbf{x}} \in \mathbb{R}^n, \bar{\mathbf{y}} \in \mathbb{R}^n, \\
& \quad \bar{\mathbf{v}} = -\mathbf{b} + (\bar{\boldsymbol{\kappa}}_u - \bar{\boldsymbol{\kappa}}_l) \quad \bar{\boldsymbol{\kappa}}_l \in \mathbb{R}^n, \bar{\boldsymbol{\kappa}}_u \in \mathbb{R}^n, c_1 \in \mathbb{R}, c_2 \in \mathbb{R} \\
& \quad \bar{\mathbf{x}} = -\mathbf{a}_x \quad \bar{\mathbf{y}} = -\mathbf{a}_y \quad \mathbf{b} = \mathbf{T}_x^\top \bar{\mathbf{x}} + \mathbf{T}_y^\top \bar{\mathbf{y}} \\
& \quad \bar{\boldsymbol{\tau}} = \mathbf{1} \quad \bar{\mathbf{z}} = \mathbf{b} \quad \left\| \begin{bmatrix} \bar{x}_i \\ \bar{y}_i \end{bmatrix} \right\|_{q_1} \leq 1 \quad \forall i = 1, \dots, n \\
& \quad \bar{d} = c_1 \quad \bar{k} = c_2 \quad \|\mathbf{b}\|_1 \leq c_1 \quad \|\mathbf{b} + (\bar{\boldsymbol{\kappa}}_l - \bar{\boldsymbol{\kappa}}_u)\|_{q_2} \leq c_2 \\
& \quad \left\| \begin{bmatrix} \bar{x}_i \\ \bar{y}_i \end{bmatrix} \right\|_{q_1} \leq \bar{\tau}_i \quad \forall i = 1, \dots, n \quad \bar{\boldsymbol{\kappa}}_l \geq \mathbf{0} \quad \bar{\boldsymbol{\kappa}}_u \geq \mathbf{0}, \\
& \quad \|\bar{\mathbf{z}}\|_1 \leq \bar{d} \quad \|\bar{\mathbf{v}}\|_{q_2} \leq \bar{k}
\end{aligned}$$

where $1/p_i + 1/q_i = 1$, and the equality follows from eliminating variables using the equality constraints.

The statement of the lemma now follows directly. □

Chapter 6

Conclusion and Future Work

In this thesis we investigated an area of decision making under uncertainty through the use of ambiguous joint chance constraints. We have shown both necessary and sufficient conditions under which these probabilistic constraints are tractable. If any of these tractability conditions are not satisfied, then the problems become NP-hard in general, in which case conservative approximations may be employed. We also show novel and interesting application areas to which we have applied ambiguous joint chance constraints. The applications we show are of “real-world” dimensions, i.e. of a scale that have never been solved exactly with existing techniques up until now.

Our main contributions can be summarised as follows:

- We first considered the problem of uncertainty quantification and chance constrained programming in the distributionally robust setting. The aim here was to certify that a given physical, engineering or economic system satisfies multiple safety conditions with high probability. We assumed that the uncertain parameters of the system are governed by an ambiguous distribution that is only known to belong to an ambiguity set characterised through generalised moment bounds and structural properties such as symmetry, unimodality or independence patterns. We delineated the watershed between tractability and intractability in ambiguity-averse uncertainty quantification and chance constrained programming. Using tools from distributionally robust optimisation, we derived explicit conic reformulations for tractable problem classes and suggested efficiently com-

putable conservative approximations for intractable ones.

- We then looked more closely at joint worst- and best-case ambiguous chance constraints. We again considered a general setting where the ambiguity set contains support, mean, and general conic bounds on the dispersion measure, but this time no structural information. We showed that left- and right-hand side uncertainty is the key separator of computational tractability. In general, problems with left-hand side uncertainty are NP-hard, even with the absence of the support constraints and with arguably the simplest polyhedral dispersion measure. Problems with right-hand side uncertainty are much more computationally amenable and we show, for the first time, equivalent tractable conic representations. Finally, we look at best-case ambiguous chance constraints and show that they may also be handled with the same techniques developed in this thesis.
- We consider an application of worst-case ambiguous chance constraints to a revenue management problem, in the context of a profit maximising cloud computing service provider. The objective is to select a price that maximises expected revenue, while at the same time ensuring an adequate level of quality of service for customers. We are able to obtain some closed-form solutions in this domain and are able to distil managerial insights, such as the direct effect of input parameters on the worst-case quality of service and worst-case revenue.
- We motivated the use of best-case ambiguous chance constraints by considering the image denoising problem. The aim of this application is to restore an image corrupted by noise during acquisition or transfer. We showed that this problem can quite naturally be cast with the help of a best-case chance constraint, and its equivalent conic reformulation surprisingly generalises a well known and powerful existing approach. We show that further insights can be gained from this chance constrained formulation.

We believe there are many interesting open questions and directions for future work. Some of the possibilities include:

- There are many interesting application areas where chance constraints naturally encode a required level of safety of the system. Examples include: finding the optimal vaccination policy for controlling infectious disease epidemics under parameter uncertainty [112]; robust least-squares prob-

lems with uncertain data [49]; surgery planning under uncertainty [4]; operating room scheduling [109]. Although classical chance constraints have been applied in some of these areas, intractability issues highlighted in Chapter 1 limit the size of problems that can be solved. Possibilities for future work include modelling uncertainty via ambiguous chance constraints. This would require adequate definition of ambiguity sets as well as defining the inequalities inside the probability expression in such a way that they capture the model and uncertainty well, as well as satisfy the conditions of tractability from Chapter 3.

- Although the exact tractable reformulations we give in Chapter 3 are effective, from a modelling and implementation perspective they are difficult and error prone to implement. A good future direction would be to define a modelling language (or an extension of an existing modelling system such as AMPL [43], GAMS [18], or YALMIP [76]) for ambiguity averse optimisation. The user describes the inequalities inside the probability expression algebraically and selects what information is available for the ambiguity set. The software would then automatically reformulate the ambiguous chance constraint into its conic equivalent and call the solver. This would greatly accelerate the application and adoption of ambiguous chance constraints to many real-world problems.
- Currently, our ambiguity sets do not depend on the decision variable $\mathbf{x}$. However, it is natural to presume that our decisions should influence the uncertainty we experience over time. The aim would be to investigate to what extent this quality, known as decision dependent uncertainty [52], can be implemented tractably within the ambiguous joint chance constraint framework.
- We have seen in Chapter 3 that if our tractability conditions are not satisfied, then in general the problems become NP-hard. However, this does not mean that the class of tractable ambiguous joint chance constraints cannot be expanded. In Chapters 4 and 5 we have provided examples of ambiguous chance constraints which do not satisfy these tractability conditions but surprisingly *do* admit equivalent tractable reformulations. This motivates us to push the boundaries of tractability further.

Bibliography

- [1] F. Alizadeh and D. Goldfarb. ‘Second-Order Cone Programming’. In: *Mathematical Programming* 95.1 (2003), pp. 3–51.
- [2] Amazon.com. *Amazon EC2*. <http://aws.amazon.com/ec2>.
- [3] H. An and J. W. Eheart. ‘A Screening Technique for Joint Chance-Constrained Programming for Air-Quality Management’. In: *Operations Research* 55.4 (2007), pp. 792–798.
- [4] H. Azimian, M. D. Naish, B. Kiaii and R. V. Patel. ‘A Chance-Constrained Programming Approach to Preoperative Planning of Robotic Cardiac Surgery Under Task-Level Uncertainty’. In: *Biomedical and Health Informatics* 19.2 (2015), pp. 612–622.
- [5] E. M. L. Beale. ‘On Minimizing a Convex Function Subject to Linear Inequalities’. In: *Journal of the Royal Statistical Society, Series B* 17.2 (1955), pp. 173–184.
- [6] A. Ben-Tal, L. El Ghaoui and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
- [7] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg and G. Rennen. ‘Robust Solutions of Optimization Problems Affected by Uncertain Probabilities’. In: *Management Science* 59.2 (2013), pp. 341–357.
- [8] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.
- [9] A. Ben-Tal and A. Nemirovski. ‘On Safe Tractable Approximations of Chance-Constrained Linear Matrix Inequalities’. In: *Mathematics of Operations Research* 34.1 (2009), pp. 1–25.

- [10] A. Ben-Tal and A. Nemirovski. ‘Robust Solutions of Linear Programming Problems Contaminated with Uncertain Data’. In: *Mathematical Programming A* 88.3 (2000), pp. 411–424.
- [11] D. Bertsimas, V. Gupta and N. Kallus. ‘Data-Driven Robust Optimization’. Available on Optimization Online. 2013.
- [12] D. Bertsimas and M. Sim. ‘The Price of Robustness’. In: *Operations Research* 52.1 (2004), pp. 35–53.
- [13] R. F. Bordley and S. M. Pollock. ‘A Decision-Analytic Approach to Reliability-Based Design Optimization’. In: *Operations Research* 57.5 (2009), pp. 1262–1270.
- [14] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [16] S. Burer. ‘Copositive Programming’. In: *Handbook on Semidefinite, Conic and Polynomial Optimization*. Ed. by M. F. Anjos and J. B. Lasserre. Vol. 166. International Series in Operations Research & Management Science. 2012, pp. 201–218.
- [17] S. Burer and D. Vandenbussche. ‘A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations’. In: *Mathematical Programming* 113.2 (2008), pp. 259–282. ISSN: 0025-5610.
- [18] M. R. Bussieck and A. Meeraus. ‘General Algebraic Modeling System (GAMS)’. In: vol. 88. Applied Optimization. 2004, pp. 137–157.
- [19] G. C. Calafiore and L. El Ghaoui. ‘On Distributionally Robust Chance-Constrained Linear Programs’. In: *Journal of Optimization Theory and Applications* 130.1 (2006), pp. 1–22.
- [20] G. Casella and R. L. Berger. *Statistical Inference*. 2nd. Duxbury Thomson Learning, 2002.
- [21] A. Charnes and W. W. Cooper. ‘Chance-Constrained Programming’. In: *Management Science* 6.1 (1959), pp. 73–79.
- [22] A. Charnes, W. W. Cooper and K. Kortanek. ‘Duality in Semi-Infinite Programs and Some Works of Haar and Carathéodory’. In: *Management Science* 9.2 (1963), pp. 209–228.

- [23] A. Charnes, W. W. Cooper and G. H. Symonds. ‘Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil’. In: *Management Science* 4.3 (1958), pp. 235–263.
- [24] J. Chen and S. Burer. ‘Globally solving nonconvex quadratic programming problems via completely positive programming’. In: *Mathematical Programming Computation* 4.1 (2012), pp. 33–52.
- [25] W. Chen, M. Sim, J. Sun and C.-P. Teo. ‘From CVaR to Uncertainty Set: Implications in Joint Chance-Constrained Optimization’. In: *Operations Research* 58.2 (2010), pp. 470–485.
- [26] X. Chen, M. Sim and P. Sun. ‘A Robust Optimization Perspective on Stochastic Programming’. In: *Operations Research* 55.6 (2007), pp. 1058–1071.
- [27] S.-S. Cheung, A. M.-C. So and K. Wang. ‘Linear Matrix Inequalities with Stochastically Dependent Perturbations and Applications to Chance-Constrained Semidefinite Optimization’. In: *SIAM Journal on Optimization* 22.4 (2012), pp. 1394–1430.
- [28] J.-F. Cordeau, G. Laporte, M. W. P. Savelsbergh and D. Vigo. ‘Vehicle Routing’. In: *Transportation*. Ed. by C. Barnhart and G. Laporte. Vol. 14. Handbooks in Operations Research and Management Science. North-Holland, Amsterdam, The Netherlands, 2007. Chap. 6, pp. 367–428.
- [29] D. Bertsimas and I. Popescu. ‘Optimal inequalities in probability theory: A convex optimization approach’. In: *SIAM Journal of Optimization* 15.3 (2004), pp. 780–804.
- [30] G. B. Dantzig. ‘Linear Programming under Uncertainty’. In: *Management Science* 1.3–4 (1955), pp. 197–206.
- [31] J. Davidson. *Stochastic Limit Theory. An Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford University Press, 1994.
- [32] E. Delage and Y. Ye. ‘Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems’. In: *Operations Research* 58.3 (2010), pp. 595–612.
- [33] V. DeMiguel and F. J. Nogales. ‘Portfolio Selection with Robust Estimation’. In: *Operations Research* 57.3 (2009), pp. 560–577.

- [34] C. Dert and B. Oldenkamp. ‘Optimal Guaranteed Return Portfolios and the Casino Effect’. In: *Operations Research* 48.5 (2000), pp. 768–775.
- [35] S. W. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications*. Vol. 27. Probability and Mathematical Statistics. Academic Press, 1988.
- [36] P. H. Diananda. ‘On non-negative forms in real variables some or all of which are non-negative’. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 58 (1 1962), pp. 17–25.
- [37] Q. T. Dinh, S. Gumussoy, W. Michiels and M. Diehl. ‘Combining Convex-Concave Decompositions and Linearization Approaches for Solving BMIs, With Application to Static Output Feedback’. In: *IEEE Transactions on Automatic Control* 57.6 (2012), pp. 1377–1390.
- [38] X. V. Doan, X. Li and K. Natarajan. ‘Robustness to Dependency in Portfolio Optimization Using Overlapping Marginals’. Available on Optimization Online. 2013.
- [39] X. V. Doan and K. Natarajan. ‘On the Complexity of Nonoverlapping Multivariate Marginal Bounds for Probabilistic Combinatorial Optimization Problems’. In: *Operations Research* 60.1 (2012), pp. 138–149.
- [40] M. Dyer and L. Stougie. ‘Computational Complexity of Stochastic Programming Problems’. In: *Mathematical Programming* 106.3 (2006), pp. 423–432.
- [41] L. G. Epstein. ‘A Definition of Uncertainty Aversion’. In: *The Review of Economic Studies* 66.3 (1999), pp. 579–608.
- [42] E. Erdoğan and G. Iyengar. ‘Ambiguous Chance Constrained Problems and Robust Optimization’. In: *Mathematical Programming B* 107.1–2 (2006), pp. 37–61.
- [43] R. Fourer, D. M. Gay and B. Kernighan. ‘Algorithms and Model Formulations in Mathematical Programming’. In: ed. by Stein W. Wallace. 1989. Chap. AMPL: A Mathematical Programming Language, pp. 150–151.
- [44] G. Gallego and I. Moon. ‘The Distribution Free Newsboy Problem: Review and Extensions’. English. In: *The Journal of the Operational Research Society* 44.8 (1993), pp. 825–834.
- [45] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1979.

- [46] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [47] C. F. Gauss. ‘Theoria combinationis observationum erroribus minimis obnoxiae, pars prior’. In: *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores* 33 (1821), pp. 321–327.
- [48] A. Georghiou, W. Wiesemann and D. Kuhn. *Wind of Change: Meeting the UK’s Renewable Energy Targets for 2020*. Working Paper.
- [49] L. El Ghaoui and H. Lebret. ‘Robust Solutions to Least-Squares Problems with Uncertain Data’. In: *SIAM Journal on Matrix Analysis and Applications* 18.4 (1997), pp. 1035–1064.
- [50] L. El Ghaoui, M. Oks and F. Oustry. ‘Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach’. In: *Operations Research* 51.4 (2003), pp. 543–556.
- [51] I. Gilboa and D. Schmeidler. ‘Maxmin Expected Utility with Non-Unique Prior’. In: *Journal of Mathematical Economics* 18.2 (1989), pp. 141–153.
- [52] V. Goel and I. E. Grossmann. ‘A Class of stochastic programs with decision dependent uncertainty’. In: *Mathematical Programming* 108.2–3 (2006), pp. 355–394.
- [53] D. Goldfarb and W. Yin. ‘Second-order cone programming methods for total variation-based image restoration’. In: *SIAM Journal of Scientific Computing* (2004), pp. 622–645.
- [54] Google. *Google Compute Engine*. <https://developers.google.com/compute>.
- [55] I. Gurvich, J. Luedtke and T. Tezcan. ‘Staffing Call Centers with Uncertain Demand Forecasts: A Chance-Constrained Optimization Approach’. In: *Management Science* 56.7 (2010), pp. 1093–1115.
- [56] S. Han, M. Tao, U. Topcu, H. Owhadi and R. M. Murray. ‘Convex Optimal Uncertainty Quantification’. Available on arXiv.org, 2013.
- [57] G. A. Hanasusanto, D. Kuhn, S. W. Wallace and S. Zymler. ‘Distributionally robust multi-item newsvendor problems with multimodal demand distributions’. In: *Mathematical Programming* (2014), pp. 1–32.

- [58] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann. ‘A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming’. In: *Mathematical Programming B* (2015), pp. 1–28.
- [59] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann. ‘A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming’. In: *Technical Report, Imperial College London and École Polytechnique Fédérale de Lausanne* (2015).
- [60] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann. *Ambiguous Joint Chance Constraints with Conic Dispersion Measures*. Working Paper. Imperial College London and École Polytechnique Fédérale de Lausanne, 2015.
- [61] K. L. Hoffman. ‘A method for globally minimizing concave functions over convex sets’. In: *Mathematical Programming* 20.1 (1981), pp. 22–32.
- [62] Z. Hu and L. J. Hong. ‘Kullback-Leibler Divergence Constrained Distributionally Robust Optimization’. Available on Optimization Online. 2012.
- [63] Z. Hu, L. J. Hong and A. M.-C. So. ‘Ambiguous Probabilistic Programs’. Available on Optimization Online. 2013.
- [64] J. Huang, M. Leng and M. Parlar. ‘Demand Functions in Decision Modeling: A Comprehensive Survey and Research Directions’. In: *Decision Sciences* 44.3 (2013), pp. 557–609.
- [65] P. J. Huber. ‘Robust Estimation of a Location Parameter’. In: *Annals of Statistics* 53.1 (1964), pp. 73–101.
- [66] K. Isii. ‘On Sharpness of Tchebycheff-Type Inequalities’. In: *Annals of the Institute of Statistical Mathematics* 14.1 (1962), pp. 185–197.
- [67] P. Jain and V. Tyagi. ‘A survey of edge-preserving image denoising methods’. In: *Information Systems Frontiers* (2014), pp. 1–12.
- [68] B. Jansen, J.J. de Jong, C. Roos and T. Terlaky. ‘Sensitivity analysis in linear programming: just be careful!’ In: *European Journal of Operational Research* 101.1 (1997), pp. 15–28.
- [69] A. Jasour, N. S. Aybat and C. Lagoa. ‘Semidefinite Programming For Chance Optimization Over Semialgebraic Sets’. Available on arXiv.org. 2014.

- [70] V. Jeyakumar and Z. Wu. ‘Conditions for Global Optimality of Quadratic Minimization Problems with LMI Constraints’. In: *Asia-Pacific Journal of Operational Research* 24.02 (2007), pp. 149–160.
- [71] R. Jiang and Y. Guan. ‘Data-driven Chance Constrained Stochastic Program’. Available on Optimization Online. 2012.
- [72] A. J. Kleywegt, A. Shapiro and T. Homem-de-Mello. ‘The Sample Average Approximation Method for Stochastic Discrete Optimization’. In: *SIAM Journal on Optimization* 12.2 (2002), pp. 479–502.
- [73] J. Korsi, F. Pfeuffer and K. Klamroth. ‘Biconvex Sets and Optimization with Biconvex Functions: A Survey and Extensions’. In: *Mathematical Methods of Operations Research* 66.3 (2007), pp. 373–407.
- [74] X. Li, K. Natarajan, C.-P. Teo and Z. Zheng. ‘Distributionally Robust Mixed Integer Linear Programs: Persistency Models with Applications’. In: *European Journal on Operational Research* 233.3 (2014), pp. 459–473.
- [75] M. S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret. ‘Applications of Second-Order Cone Programming’. In: *Linear Algebra and its Applications* 284 (1998), pp. 193–228.
- [76] J. Löfberg. ‘YALMIP : A Toolbox for Modeling and Optimization in MATLAB’. In: *Proceedings of the CACSD Conference*. Taipei, Taiwan, 2004.
- [77] C. Loken et al. ‘SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre’. In: *Journal of Physics: Conference Series* 256.1 (2010), pp. 12–26.
- [78] G. N. Mankiw and M. P. Taylor. *Microeconomics*. Thomson Learning, 2001.
- [79] P. Milgrom and I. Segal. ‘Envelope Theorems for Arbitrary Choice Sets’. In: *Econometrica* 70.2 (2002), pp. 583–601.
- [80] B. L. Miller and H. M. Wagner. ‘Chance Constrained Programming with Joint Constraints’. In: *Operations Research* 13.6 (1965), pp. 930–945.

- [81] P. Mohajerin Esfahani and D. Kuhn. ‘Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations’. In: *Working Paper, École Polytechnique Fédérale de Lausanne* (2015).
- [82] Mosek. *MOSEK Solver*. <https://mosek.com>.
- [83] S. Mukerji. ‘Ambiguity Aversion and Incompleteness of Contractual Form’. In: *The American Economic Review* 88.5 (1998), pp. 1207–1231.
- [84] S. Mukerji and J.-M. Tallon. ‘Ambiguity Aversion and Incompleteness of Financial Markets’. In: *The Review of Economic Studies* 68.4 (2001), pp. 883–904.
- [85] K. Natarajan, D. Pachamanova and M. Sim. ‘Incorporating asymmetric distributional information in robust value-at-risk optimization’. In: *Management Science* 54.3 (2008), pp. 573–585.
- [86] A. Nemirovski. ‘On Safe Tractable Approximations of Chance Constraints’. In: *European Journal of Operational Research* 219.3 (2012), pp. 707–718.
- [87] A. Nemirovski and A. Shapiro. ‘Convex Approximations of Chance Constrained Programs’. In: *SIAM Journal on Optimization* 17.4 (2006), pp. 969–996.
- [88] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns and M. Ortiz. ‘Optimal Uncertainty Quantification’. In: *SIAM Review* 55.2 (2013), pp. 271–345.
- [89] B. K. Pagnoncelli, S. Ahmed and A. Shapiro. ‘Sample average approximation method for chance constrained programming: theory and applications’. In: *Journal of Optimization Theory and Applications* 142.2 (2009), pp. 399–416.
- [90] B. P. G. Van Parys, P. J. Goulart and D. Kuhn. ‘Generalized Gauss Inequalities via Semidefinite Programming’. In: *Available on Optimization Online* (2014).
- [91] B. P. G. Van Parys, D. Kuhn, P. J. Goulart and M. Morari. ‘Distributionally Robust Control of Constrained Stochastic Systems’. Available on Optimization Online. 2013.
- [92] G. Pflug, A. Pichler and D. Wozabal. ‘The $1/N$ Investment Strategy is Optimal Under High Model Ambiguity’. In: *Journal of Banking & Finance* 36.2 (2012), pp. 410–417.
- [93] I. Popescu. ‘An SDP Approach to Optimal Moment Bounds for Convex Classes of Distributions’. In: *Mathematics of Operations Research* 50.3 (2005), pp. 632–657.

- [94] A. Prékopa. ‘On Probabilistic Constrained Programming’. In: *Proceedings of the Princeton Symposium on Mathematical Programming*. Princeton University Press, 1970, pp. 113–138.
- [95] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.
- [96] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. Wiley, 1991.
- [97] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.
- [98] R. T. Rockafellar and S. Uryasev. ‘Conditional Value-at-Risk for General Loss Functions’. In: *Journal of Banking & Finance* 26.7 (2002), pp. 1443–1471.
- [99] R. T. Rockafellar and S. Uryasev. ‘Optimization of Conditional Value-at-Risk’. In: *Journal of Risk* 2.3 (2000), pp. 21–41.
- [100] V. Roitch and D. Kuhn. *Distributionally Robust Pricing of Monopolistic Cloud Computing Services with Service Level Agreements*. Working Paper. Imperial College London and École Polytechnique Fédérale de Lausanne, 2015.
- [101] L. I. Rudin, S. Osher and E. Fatemi. ‘Nonlinear total variation based noise removal algorithms’. In: *Physica D: Nonlinear Phenomena* 60.1–4 (1992), pp. 259–268.
- [102] H. Scarf. ‘Studies in the Mathematical Theory of Inventory and Production’. In: Stanford University Press, 1958. Chap. A min-max solution of an inventory problem.
- [103] A. Shapiro. ‘On Duality Theory of Conic Linear Problems’. In: *Semi-Infinite Programming*. Kluwer Academic Publishers, 2001. Chap. 7, pp. 135–165.
- [104] A. Shapiro. ‘Quantitative Stability in Stochastic Programming’. In: *Mathematical Programming* 67.1–3 (1994), pp. 99–108.
- [105] A. Shapiro, D. Dentcheva and A. Ruszczyński. *Lectures on Stochastic Programming: Modelling and Theory*. SIAM, 2009.
- [106] A. Shapiro and A. Kleywegt. ‘Minimax Analysis of Stochastic Problems’. In: *Optimization Methods and Software* 17.3 (2002), pp. 523–542.
- [107] S. Shen, J. C. Smith and S. Ahmed. ‘Expectation and Chance-Constrained Models and Algorithms for Insuring Critical Paths’. In: *Management Science* 56.10 (2010), pp. 1794–1814.

- [108] H. D. Sherali, A. L. Soyster and F. H. Murphy. ‘Stackelberg-Nash-Cournot Equilibria: Characterizations and Computations’. In: *Operations Research* 31.2 (1983), pp. 253–276.
- [109] O. V. Shylo, O. A. Prokopyev and A. J. Schaefer. ‘Stochastic Operating Room Scheduling for High-Volume Specialties Under Block Booking’. In: *INFORMS Journal on Computing* 25.4 (2013), pp. 682–692.
- [110] M. Sion. ‘On general minimax theorems’. In: *Pacific Journal of Mathematics* 8.1 (1958), pp. 171–176.
- [111] H. Sun and H. Xu. ‘Asymptotic Convergence Analysis for Distributional Robust Optimization and Equilibrium Problems’. Available on Optimization Online. 2013.
- [112] M. W. Tanner, L. Sattenspiel and L. Ntaimo. ‘Finding optimal vaccination strategies under parameter uncertainty using stochastic programming’. In: *Mathematical Biosciences* 215.2 (2008), pp. 144–151.
- [113] J. J. M. Trienekens, J. J. Bouman and M. van der Zwan. ‘Specification of Service Level Agreements: Problems, Principles and Practices’. In: *Software Quality Journal* 12.1 (2004), pp. 43–57.
- [114] L. Vandenberghe and S. Boyd. ‘Semidefinite Programming’. In: *SIAM Rev.* 38.1 (1996), pp. 49–95.
- [115] L. Vandenberghe, S. Boyd and K. Comanor. ‘Generalized Chebyshev Bounds via Semidefinite Programming’. In: *SIAM Review* 49.1 (2007), pp. 52–64.
- [116] G. Zhao W. Soon and J. Zhang. ‘Complementarity demand functions and pricing models for multi-product markets’. In: *European Journal of Applied Mathematics* 20 (5 2009), pp. 399–430.
- [117] J. Wang. ‘The β -Reliable Median on a Network with Discrete Probabilistic Demand Weights’. In: *Operations Research* 55.5 (2007), pp. 966–975.
- [118] J. E. Ward and R. E. Wendell. ‘Approaches to sensitivity analysis in linear programming’. In: *Annals of Operations Research* 27.1 (1990), pp. 3–38.
- [119] R. E. Wendell. ‘The Tolerance Approach to Sensitivity Analysis in Linear Programming’. In: *Management Science* 31.5 (1985), pp. 564–578.

- [120] W. Wiesemann, D. Kuhn and B. Rustem. ‘Robust Markov Decision Processes’. In: *Mathematics of Operations Research* 38.1 (2013), pp. 153–183.
- [121] W. Wiesemann, D. Kuhn and M. Sim. ‘Distributionally Robust Convex Optimization’. In: *Operations Research* 62.6 (2014), pp. 1358–1376.
- [122] H. Xu, C. Caramanis and S. Mannor. ‘Optimization Under Probabilistic Envelope Constraints’. In: *Operations Research* 60.3 (2012), pp. 682–699.
- [123] İ. Yamikoğlu and D. den Hertog. ‘Safe Approximations of Ambiguous Chance Constraints using Historical Data’. In: *INFORMS Journal on Computing* 25.4 (2013), pp. 666–681.
- [124] J. Žáčková. ‘On minimax solutions of stochastic linear programming problems’. In: *Časopis pro pěstování matematiky* 91.4 (1966), pp. 423–430.
- [125] S. Zymmler, D. Kuhn and B. Rustem. ‘Distributionally Robust Joint Chance Constraints with Second-Order Moment Information’. In: *Mathematical Programming A* 137.1–2 (2013), pp. 167–198.
- [126] S. Zymmler, D. Kuhn and B. Rustem. ‘Worst-Case Value-at-Risk of Non-Linear Portfolios’. In: *Management Science* 59.1 (2013), pp. 172–188.