A Robust Optimization Perspective of Stochastic Programming

Xin Chen[*], Melvyn Sim[†] and Peng Sun[‡]

December 2004

Revised: June 2005

Abstract

In this paper, we introduce an approach for constructing uncertainty sets for robust optimization using new deviation measures for bounded random variables known as the forward and backward deviations. These deviation measures capture distributional asymmetry and lead to better approximations of chance constraints. We also propose a tractable robust optimization approach for obtaining robust solutions to a class of stochastic linear optimization problems where the risk of infeasibility can be tolerated as a tradeoff to improve upon the objective value. An attractive feature of the framework is the computational scalability to multiperiod models. We show an application of the framework for solving a project management problem with uncertain activity completion time.

[*]Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign. Email: xinchen@uiuc.edu

[†]NUS Business School, National University of Singapore. Email: dscsimm@nus.edu.sg. The research of the author is supported by NUS academic research grant R-314-000-066-122.

[‡]Fuqua School of Business, Duke University Box 90120, Durham, NC 27708, USA. Email: psun@duke.edu

1 Introduction

In recent years, robust optimization has gained substantial popularity as a competing methodology for solving several types of stochastic optimization models. Robust optimization has been successful in immunizing uncertain mathematical optimization. The first step in this direction is taken by Soyster [24] who proposes a worst case model to linear optimization. Subsequently, more elaborate uncertainty sets and computationally attractive robust optimization methodologies are proposed by Ben-Tal and Nemirovski [2, 3, 4], El-Ghaoui et al. [16, 17], Iyangar and Goldfarb [18], Bertsimas and Sim [9, 10, 11, 12] and Atamtürk [1]. To address the issue of over-conservatism in robust linear optimization, these papers propose less conservative models by considering uncertainty sets in the form of ellipsoidal and more complex intersection of ellipsoidal sets. The *robust counterparts* of the nominal problems generally are in the form of conic quadratic problems (see Ben-Tal and Nemirovski [4]) and even linear optimization problems of slightly larger size (see Bertsimas and Sim [10]). Examples of using ellipsoidal random parameter domains also appear in stochastic optimization that focuses on exact and approximate solutions to chance constraints for specific distributions (see, for instance, Kibzun and Kan [21]).

The methodology of robust optimization has also been applied to dynamic settings involving multi-period optimization in which future decisions (recourse variables) depend on the realization of present data. Such models are in general intractable. Ben-Tal et al. [5] propose a tractable approach for solving fixed recourse instances under affine restrictions on the recourse variables with respect to the uncertain data. Some applications of robust optimization on dynamic environment include inventory management (Bertsimas and Thiele [13], Ben-Tal et al. [5]) and supply contracts (Ben-Tal et al. [6]).

There are two important characteristics of robust linear optimization that are practically appealing.

(a) Robust linear optimization models are polynomial in size and in the form of linear programming or second order cone programming (SOCP). We could leverage on the state-of-the-art LP and SOCP solvers, which are increasingly becoming more powerful, efficient and robust.

(b) Robust optimization requires mild assumptions on distributions, such as known mean and bounded support. This relieves users from having to know the probabilistic distributions of the underlying stochastic parameters, which are often unavailable.

Despite its tractability, one of the main criticisms of robust optimization has been its inability to incorporate any available distributional information to achieve better performance. In linear optimization, Bertsimas and Sim [10] and Ben-Tal and Nemirovski [4] obtain probability bounds against constraint violation by assuming that the coefficients are independent and symmetrically bounded and neglecting any deviational information of the random variables. For instance, in cases where the variances of the random variables are small while the support of the distributions are wide, the robust solutions obtained via this approach can be rather conservative. The assumption of symmetric distribution is also limiting in many applications such as financial modeling in which distributions are often known to be asymmetric.

Our goal of this paper is two-folded. First, we refine the framework for robust linear optimization by introducing a new uncertainty set that captures the asymmetry of the underlying random variables. For this purpose, we introduce new deviation measures associated with a random variable, namely the forward and backward deviations and apply to the design of uncertainty set. Hence, this enables us to capture the asymmetry of random variables in order to obtain better solutions that satisfy probabilistic or chance constraints. Our robust linear optimization framework generalizes previous works of Bertsimas and Sim [10] and Ben-Tal and Nemirovski [4]. Second, we propose a tractable robust optimization approach for solving a class of stochastic linear optimization problems with chance constraints. Again, applying the forward and backward deviations of the underlying distributions, our robust optimization approach provides feasible solutions to the stochastic linear optimization. The optimal solution from our model is an upper bound to the minimum objective value for all underlying distributions that satisfy the parameters of deviations. An attractive feature of this framework is the computational scalability to multiperiod models. We emphasize that literatures on multiperiod stochastic programs with chance constraints are rather limited, which could be due to the lack of tractable methodologies.

In Section 2, we introduce a new uncertainty set and formulate the robust counterpart. In Section 3, we present new deviation measures that capture distributional asymmetry. Following which, Section 4 shows how we can integrate the new uncertainty set with the new deviation measures to obtain solutions to chance constrained problems. We present in Section 5 a robust optimization approach for obtaining less conservative solutions for stochastic programming with chance constraints. We also show how our robust optimization framework is applied to a project management problem with uncertain completion time in Section 6. Finally Section 7 concludes this paper.

Notations We denote a random variable, $\tilde{x}$, with the tilde sign. Bold face lower case letters such as $\boldsymbol{x}$ represent vectors and the corresponding upper case letters such as $\boldsymbol{A}$ denote matrices.

2 Robust Formulation of a Stochastic Linear Constraint

Consider a stochastic linear constraint,

$$\tilde{\boldsymbol{a}}'\boldsymbol{x} \leq \tilde{b}, \tag{1}$$

in which the input parameters $(\tilde{\boldsymbol{a}}, \tilde{b})$ are random. We assume that the uncertain data, $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{a}}, \tilde{b})$ has the following underlying perturbations.

Affine Data Perturbation:

We represent uncertainties on the data $\tilde{\boldsymbol{D}}$ as affinely dependent on a set of independent random variables, $\{\tilde{z}_j\}_{j=1:N}$ as follows,

$$\tilde{\boldsymbol{D}} = \boldsymbol{D}^0 + \sum_{j=1}^{N} \boldsymbol{\Delta D}^j \tilde{z}_j,$$

where $\boldsymbol{D}^0$ is the nominal value of the data, $\boldsymbol{\Delta D}^j$, $j \in N$ is a direction of data perturbation. We call $\tilde{z}_j$ the primitive uncertainties which has mean zero and support in $[-\underline{z}_j, \bar{z}_j]$, $\underline{z}_j, \bar{z}_j > 0$. If N is small, we model situations involving a small collection of primitive independent uncertainties, or large, potentially as large as the number of entries in the data. In the former case, the elements of $\tilde{\boldsymbol{D}}$ are strongly dependent, while in the latter case the elements of $\tilde{\boldsymbol{D}}$ are weakly dependent or even independent (when the number of entries in the data equals N).

We desire a set of solutions $X(\epsilon)$ such that $\boldsymbol{x} \in X(\epsilon)$ is feasible to the linear constraint (1) with probability at least $1 - \epsilon$. Formally, we can describe the set $X(\epsilon)$ aptly using the following chance constraint representation (see Charnes and Cooper [15]),

$$X(\epsilon) = \left\{ \boldsymbol{x} \ : \ \mathrm{P}(\tilde{\boldsymbol{a}}'\boldsymbol{x} \leq \tilde{b}) \geq 1 - \epsilon \right\}. \tag{2}$$

The parameter ϵ in the set $X(\epsilon)$ varies the conservatism of the solution. Unfortunately for $\epsilon > 0$, the set $X(\epsilon)$ is often non-convex and computationally intractable (see Birge and Louveaux [14]). Furthermore, the evaluation of probability requires full knowledge of data distributions which is often an unrealistic assumption. In view of the difficulties, robust optimization presents a different approach to handling data uncertainty. Specifically, in addressing the uncertain linear constraint of (1), we represent the set of robust feasible solution

$$X_r(\Omega) = \{\boldsymbol{x} \ : \ \boldsymbol{a}'\boldsymbol{x} \leq b \ \forall (\boldsymbol{a}, b) \in \mathcal{U}_\Omega\}, \tag{3}$$

in which the uncertain set, $\mathcal{U}_\Omega$ is compact and the parameter Ω, referred to as the budget of uncertainty, varies the size of the uncertainty set radially from the central point $\mathcal{U}_{\Omega=0} = (\boldsymbol{a}^0, b^0)$, such that $\mathcal{U}_\Omega \subseteq \mathcal{U}_{\Omega'} \subseteq \mathcal{W}$ for all $\Omega_{\max} \geq \Omega' \geq \Omega \geq 0$. Here the worst case uncertainty set $\mathcal{W}$ is the convex support of the uncertain data, which is the smallest closed convex set satisfying $\mathrm{P}((\tilde{\boldsymbol{a}}, \tilde{b}) \in \mathcal{W}) = 1$, and $\Omega_{\max}$ is the worst case budget of uncertainty, i.e., the minimum parameter Ω such that $\mathcal{U}_\Omega = \mathcal{W}$. For the stochastic linear constraint, $\boldsymbol{D}^0 = (\boldsymbol{a}^0, b^0)$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta a}^j, \Delta b^j)$, the convex support of the uncertain parameter is as follows,

$$\mathcal{W} = \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^N (\boldsymbol{\Delta a}^j, \Delta b^j) z_j, -\underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \bar{\boldsymbol{z}} \right\}. \tag{4}$$

Therefore, under affine data perturbation, the worst case uncertainty set is a parallelotope in which the feasible solution is characterized by Soyster [24], which, of course, is a very conservative approximation to $X(\epsilon)$. To derive a less conservative approximation, we need to choose the budget of uncertainty, Ω, appropriately.

In designing such uncertainty set, we want to preserve the computational tractability both theoretically and most importantly practically of the nominal problem. Furthermore, we want to find a guarantee on the probability such that the robust solution is feasible without being over conservative. In other words, for a reasonable choice of ϵ such as 0.001, there exists a parameter Ω such that $X_r(\Omega) \subseteq X(\epsilon)$. Furthermore, the budget of uncertainty Ω should be substantially smaller than the worst case budget $\Omega_{\max}$, and so that the solution is potentially less conservative than the worst case solution.

For symmetric bounded distributions, we can assume without loss of generality that the primitive uncertainties $\tilde{z}_j$ are distributed in $[-1, 1]$, that is $\underline{\boldsymbol{z}} = \bar{\boldsymbol{z}} = \boldsymbol{1}$. The natural uncertainty set to consider is the intersection of a norm uncertainty set, $\mathcal{V}_\Omega$ and the worst case support set, $\mathcal{W}$ as follows.

$$\begin{aligned} \mathcal{S}_\Omega &= \underbrace{\left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^N (\boldsymbol{\Delta a}^j, \Delta b^j) z_j, \|\boldsymbol{z}\| \leq \Omega \right\}}_{=\mathcal{V}_\Omega} \cap \mathcal{W} \\ &= \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^N (\boldsymbol{\Delta a}^j, \Delta b^j) z_j, \|\boldsymbol{z}\| \leq \Omega, \|\boldsymbol{z}\|_\infty \leq 1 \right\}. \end{aligned} \tag{5}$$

As the budget of uncertainty Ω increases, the norm uncertainty set, $\mathcal{V}_\Omega$ expands radially from the point $(\boldsymbol{a}^0, b^0)$ until it engulfs the set $\mathcal{W}$. In which case, the uncertainty set $\mathcal{S}_\Omega = \mathcal{W}$. Hence, for any choice of Ω, the uncertainty set $\mathcal{S}_\Omega$ is always less conservative than the worst case uncertainty set $\mathcal{W}$. Various choices of norms, $\|\cdot\|$ are considered in robust optimization. Under the l_2 or Ellipsoidal norm proposed by Ben-Tal and Nemirovski [4], the feasible solutions to the robust counterpart of (3) in which $\mathcal{U}_\Omega = \mathcal{S}_\Omega$ is guaranteed feasible to the linear constraint with probability at least $1 - \exp(-\Omega^2/2)$. The robust counterpart is also equivalent to a formulation with second order cones constraints. In Bertsimas and

Sim [10], they consider a $l_1 \cap l_\infty$ norm of the form $\|\boldsymbol{z}\|_{l_1 \cap l_\infty} = \max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\}$, and show that the feasibility guarantee is also $1 - \exp\left(-\Omega^2/2\right)$. The resultant robust counterpart under consideration remains a linear optimization problem of about the same size which is practically suited for optimization over integers. However, in the worst case, this approach can be more conservative than the use of Ellipsoidal norm. In both approaches, the value of Ω is relatively small. For example, for feasibility guarantee of 99.9%, we only need to choose $\Omega = 3.72$. To compare with the worst case uncertainty set, $\mathcal{W}$, we note that for Ω greater than $\sqrt{N}$, the constraints $\|\boldsymbol{z}\|_2 \leq \Omega$ and $\max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\} \leq \Omega$ are the consequence of $\boldsymbol{z}$ satisfying, $\|\boldsymbol{z}\|_\infty \leq 1$. Hence, it is apparent that for both approaches, the budget of uncertainty Ω is substantially smaller than the worst case budget in which $\Omega_{\max} = \sqrt{N}$.

In this paper, we restrict the vector norm, $\|.\|$ we consider in an uncertainty set as follows,

$$\|\boldsymbol{u}\| = \||\boldsymbol{u}|\|, \tag{6}$$

where $|\boldsymbol{u}|$ is the vector with the j component equal to $|u_j|\ \forall j \in \{1, \ldots, N\}$ and

$$\|\boldsymbol{u}\| \leq \|\boldsymbol{u}\|_2, \forall \boldsymbol{u}. \tag{7}$$

We call this a regular norm. It is easy to see that the Ellipsoidal norm and the $l_1 \cap l_\infty$ norm we mentioned satisfy these properties. The dual norm $\|.\|^*$ is defined as

$$\|\boldsymbol{u}\|^* = \max_{\|\boldsymbol{x}\| \leq 1} \boldsymbol{u}'\boldsymbol{x}.$$

We next show some basic properties of regular norms which we will subsequently use in our development.

Proposition 1 *If the norm $\|\cdot\|$ satisfies Eq. (6) and Eq. (7), then we have*

(a) $\|\boldsymbol{w}\|^* = \||\boldsymbol{w}|\|^*$.

(b) *For all $\boldsymbol{v}, \boldsymbol{w}$ such that $|\boldsymbol{v}| \leq |\boldsymbol{w}|$, $\|\boldsymbol{v}\|^* \leq \|\boldsymbol{w}\|^*$.*

(c) *For all $\boldsymbol{v}, \boldsymbol{w}$ such that $|\boldsymbol{v}| \leq |\boldsymbol{w}|$, $\|\boldsymbol{v}\| \leq \|\boldsymbol{w}\|$.*

(d) $\|\boldsymbol{t}\|^* \geq \|\boldsymbol{t}\|_2, \forall \boldsymbol{t}$.

Proof : The proofs of **(a)**, **(b)** and **(c)** are shown in Bertsimas and Sim [11].

(d) It is well known that the dual norm of Euclidian norm is also the Euclidian norm, that is, self dual. For all $\boldsymbol{t}$ observe that

$$\|\boldsymbol{t}\|^* = \max_{\|\boldsymbol{z}\| \leq 1} \boldsymbol{t}'\boldsymbol{z} \geq \max_{\|\boldsymbol{z}\|_2 \leq 1} \boldsymbol{t}'\boldsymbol{z} = \|\boldsymbol{t}\|_2^* = \|\boldsymbol{t}\|_2.$$

■

To build a generalization of the uncertainty set that takes into account the primitive uncertainties being asymmetrically distributed, we first ignore the worst case support set, $\mathcal{W}$ and define the asymmetric norm uncertainty set as follows,

$$\mathcal{A}_\Omega = \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, (\mathrm{a}, b) = (\mathrm{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j)(v_j - w_j), \right. \\ \left. \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega, \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0} \right\}, \quad (8)$$

where $\boldsymbol{P} = \mathrm{diag}(p_1, \ldots, p_N)$ and likewise, $\boldsymbol{Q} = \mathrm{diag}(q_1, \ldots, q_N)$ with $p_j, q_j > 0$, $j \in \{1, \ldots, N\}$. In the subsequent section, it will be clear how $\boldsymbol{P}$ and $\boldsymbol{Q}$ relate to the forward and backward deviations of the underlying primitive uncertainties. The following proposition shows the connection of the set $\mathcal{A}_\Omega$ with the uncertainty set described by norm, $\mathcal{V}_\Omega$ defined in (5).

Proposition 2 *When $p_j = q_j = 1$ for all $j \in \{1, \ldots, N\}$, the uncertainty sets, $\mathcal{A}_\Omega$ and $\mathcal{V}_\Omega$ are equivalent.*

The proof is shown in Appendix A.

Speaking intuitively, to capture distributional asymmetries, we decompose the primitive data uncertainty, $\tilde{z}$ into two random variables, $\tilde{v} = \max\{\tilde{z}, 0\}$ and $\tilde{w} = \max\{-\tilde{z}, 0\}$ such that $\tilde{z} = \tilde{v} - \tilde{w}$. The multipliers $1/p_j$ and $1/q_j$ normalize the effective perturbation contributed by both $\tilde{v}$ and $\tilde{w}$ such that the norm of the aggregated values falls within the budget of uncertainty.

Since $p_j, q_j > 0$, for $\Omega > 0$, the point $(\boldsymbol{a}^0, b^0)$ lies in the interior of the uncertainty set $\mathcal{A}_\Omega$. Hence, we can easily evoke strong duality to obtain an equivalent formulation of the robust counterpart of (3) that is computationally attractive, such as in the form of linear or second order cone optimization problems. To facilitate our expositions, we need the following proposition.

Proposition 3 *Let*

$$\begin{aligned} z^* = \quad & \max \quad \boldsymbol{a}'\boldsymbol{v} + \boldsymbol{b}'\boldsymbol{w} \\ & \text{s.t.} \quad \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega \\ & \qquad \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0}, \end{aligned} \quad (9)$$

then $\Omega\|\boldsymbol{t}\|^* = z^*$, where $t_j = \max\{a_j, b_j, 0\}$, $j \in \{1, \ldots, N\}$.

We present the proof in Appendix B.

Theorem 1 *The robust counterpart of* (3) *in which* $\mathcal{U}_\Omega = \mathcal{A}_\Omega$ *is equivalent to*

$$\left\{ \boldsymbol{x} : \begin{array}{l} \exists \boldsymbol{u} \in \Re^N, h \in \Re \\ \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega h \leq b^0 \\ \|\boldsymbol{u}\|^* \leq h \\ u_j \geq p_j(\boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j), \ \forall j \in \{1,\ldots,N\} \\ u_j \geq -q_j(\boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j), \ \forall j \in \{1,\ldots,N\}. \end{array} \right\} \tag{10}$$

Proof : We first express the robust counterpart of (3) in which $\mathcal{U}_\Omega = \mathcal{A}_\Omega$ as follows,

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \sum_{j=1}^{N} \underbrace{\left(\boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j\right)}_{=y_j}(v_j - w_j) \leq b^0 \quad \forall \boldsymbol{v}, \boldsymbol{w} \in \Re^N, \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega, \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0}$$

$$\Updownarrow$$

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \max_{\substack{\{\boldsymbol{v},\boldsymbol{w} \,:\, \|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega \\ \boldsymbol{v},\boldsymbol{w} \geq \boldsymbol{0}\}}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \leq b^0$$

Observe that

$$\begin{aligned} & \max_{\substack{\{\boldsymbol{v},\boldsymbol{w} \,:\, \|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega \\ \boldsymbol{v},\boldsymbol{w} \geq \boldsymbol{0}\}}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \\ = & \max_{\substack{\{\boldsymbol{v},\boldsymbol{w} \,:\, \|\boldsymbol{v}+\boldsymbol{w}\| \leq \Omega \\ \boldsymbol{v},\boldsymbol{w} \geq \boldsymbol{0}\}}} (\boldsymbol{P}\boldsymbol{y})'\boldsymbol{v} - (\boldsymbol{Q}\boldsymbol{y})'\boldsymbol{w} \\ = & \ \Omega\|\boldsymbol{t}\|^* \end{aligned} \tag{11}$$

where $t_j = \max\{p_j y_j, -q_j y_j, 0\} = \max\{p_j y_j, -q_j y_j\}$, since $p_j, q_j > 0$ for all $j \in \{1,\ldots,N\}$. Furthermore, the equality (11) follows from direct transformation of vectors $\boldsymbol{v}, \boldsymbol{w}$ to respectively $\boldsymbol{Pv}, \boldsymbol{Qw}$. The last equality follows directly from Proposition 3. Hence, the equivalent formulation of the robust counterpart is

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{t}\|^* \leq b^0. \tag{12}$$

Finally, suppose $\boldsymbol{x}$ is feasible in the robust counterpart of (3) in which $\mathcal{U}_\Omega = \mathcal{A}_\Omega$, we let $\boldsymbol{u} = \boldsymbol{t}$, $h = \|\boldsymbol{t}\|^*$ and following from Eq. (12), the constraint (10) is also feasible. Conversely, if $\boldsymbol{x}$ is feasible in (10), then $\boldsymbol{u} \geq \boldsymbol{t}$. Following Proposition 1(b), we have

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{t}\|^* \leq \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{u}\|^* \leq \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega h \leq b^0.$$

■

The complete formulation and complexity class of the robust counterpart depends on the representation of the dual norm constraint, $\|\boldsymbol{u}\|^* \leq y$. In Appendix C, we tabulate the common choices of regular norms, the representation of their dual norms and the corresponding references. In terms of keeping the model linear and moderately increase in size, the $l_1 \cap l_\infty$ norm is an attractive choice.

Incorporating Worst Case Support Set, $\mathcal{W}$

We now incorporate the worst case support set $\mathcal{W}$ as follows

$$\mathcal{G}_\Omega = \mathcal{A}_\Omega \cap \mathcal{W}.$$

Since we represent the support set of $\mathcal{W}$ equivalently as

$$\mathcal{W} = \left\{ (\boldsymbol{a}, b) \; : \; \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j)(v_j - w_j), -\underline{\boldsymbol{z}} \le \boldsymbol{v} - \boldsymbol{w} \le \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \ge \boldsymbol{0} \right\}, \tag{13}$$

it follows trivially that

$$\begin{aligned} \mathcal{G}_\Omega = \Bigg\{ (\boldsymbol{a}, b) \; : \; \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, (\mathrm{a}, b) = (\mathrm{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j)(v_j - w_j), \\ \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, -\underline{\boldsymbol{z}} \le \boldsymbol{v} - \boldsymbol{w} \le \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \ge \boldsymbol{0} \Bigg\}. \end{aligned} \tag{14}$$

We will show an equivalent formulation of the corresponding robust counterpart under the generalized uncertainty set, $\mathcal{G}_\Omega$.

Theorem 2 *The robust counterpart of* (3) *in which* $\mathcal{U}_\Omega = \mathcal{G}_\Omega$ *is equivalent to*

$$\left\{ \boldsymbol{x} \; : \; \begin{array}{l} \exists \boldsymbol{u}, \boldsymbol{r}, \boldsymbol{s} \in \Re^N, h \in \Re \\ \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega h + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \le b^0 \\ \|\boldsymbol{u}\|^* \le h \\ u_j \ge p_j(\boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j - r_j + s_j) \;\; \forall j = \{1, \ldots, N\}, \\ u_j \ge -q_j(\boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j - r_j + s_j) \;\; \forall j = \{1, \ldots, N\}, \\ \boldsymbol{u}, \boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}. \end{array} \right\} \tag{15}$$

Proof : Similar to the exposition of Theorem 1, the robust counterpart of (3) in which $\mathcal{U}_\Omega = \mathcal{G}_\Omega$ is as follows,

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \max_{(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{C}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \le b^0$$

where

$$\mathcal{C} = \left\{ (\boldsymbol{v}, \boldsymbol{w}) \; : \; \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, -\underline{\boldsymbol{z}} \le \boldsymbol{v} - \boldsymbol{w} \le \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \ge \boldsymbol{0} \right\}$$

and $y_j = \mathbf{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j$. Since $\mathcal{C}$ is a compact convex set with nonempty interior, we can use strong duality to obtain the equivalent representation. Observe that

$$
\begin{aligned}
& \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}\,:\,\|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\|\le\Omega,\\ -\underline{\boldsymbol{z}}\le\boldsymbol{v}-\boldsymbol{w}\le\bar{\boldsymbol{z}},\boldsymbol{w},\boldsymbol{v}\ge\mathbf{0}\}}} (\boldsymbol{v}-\boldsymbol{w})'\boldsymbol{y} \\
= & \min_{\boldsymbol{r},\boldsymbol{s}\ge\mathbf{0}} \left\{ \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}:\|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\|\le\Omega,\\ \boldsymbol{v},\boldsymbol{w}\ge\mathbf{0}\}}} (\boldsymbol{v}-\boldsymbol{w})'\boldsymbol{y} + \boldsymbol{r}'(\bar{\boldsymbol{z}}-\boldsymbol{v}+\boldsymbol{w}) + \boldsymbol{s}'(\underline{\boldsymbol{z}}+\boldsymbol{v}-\boldsymbol{w}) \right\} \\
= & \min_{\boldsymbol{r},\boldsymbol{s}\ge\mathbf{0}} \left\{ \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}:\|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\|\le\Omega,\\ \boldsymbol{v},\boldsymbol{w}\ge\mathbf{0}\}}} (\boldsymbol{y}-\boldsymbol{r}+\boldsymbol{s})'\boldsymbol{v} - (\boldsymbol{y}-\boldsymbol{r}+\boldsymbol{s})'\boldsymbol{w} + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\} \\
= & \min_{\boldsymbol{r},\boldsymbol{s}\ge\mathbf{0}} \left\{ \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}:\|\boldsymbol{v}+\boldsymbol{w}\|\le\Omega,\\ \boldsymbol{v},\boldsymbol{w}\ge\mathbf{0}\}}} (\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{r}+\boldsymbol{s}))'\boldsymbol{v} - (\boldsymbol{Q}(\boldsymbol{y}-\boldsymbol{r}+\boldsymbol{s}))'\boldsymbol{w} + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\} \\
= & \min_{\boldsymbol{r},\boldsymbol{s}\ge\mathbf{0}} \left\{ \Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\},
\end{aligned}
$$

where the first equality is due to strong Lagrangian duality (see for instance Bertsekas [7]) and the last inequality follows from Proposition 3 in which

$$
\begin{aligned}
\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s}) &= \begin{bmatrix} \max(p_1(y_1-r_1+s_1), -q_j(y_1-r_1+s_1), 0) \\ \vdots \\ \max(p_N(y_N-r_N+s_N), -q_j(y_N-r_N+s_N), 0) \end{bmatrix} \\
&= \begin{bmatrix} \max(p_1(y_1-r_1+s_1), -q_j(y_1-r_1+s_1)) \\ \vdots \\ \max(p_N(y_N-r_N+s_N), -q_j(y_N-r_N+s_N)) \end{bmatrix}.
\end{aligned}
$$

Hence the robust counterpart is the same as

$$
\boldsymbol{a}^{0\prime}\boldsymbol{x} + \min_{\boldsymbol{r},\boldsymbol{s}\ge\mathbf{0}} \left\{ \Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\} \le b^0. \tag{16}
$$

Using similar arguments as in Theorem 1, we can easily show that the feasible solution of (16) is equivalent to (15). ■

3 Deviation Measures for Bounded Distributions

When incorporated in optimization models, operations on random variables in which the distributions are given are often cumbersome and computationally intractable. Moreover, in many practical problems, we do not know the precise data distributions and hence solutions based on assumed distributions may be unjustified. Instead of using full distributional information, our aim is to identify some salient characteristics of data distribution that we could use to exploit in robust optimization so as to obtain nontrivial probability bounds against constraint violation.

We commonly measure the variability of a random variable using variance or second moments which is, however, insensitive to distributional asymmetry. In this section, we introduce new deviation measures for bounded random variables that will capture distributional asymmetries and when applied to our proposed robust methodology, we can achieve the desired probabilistic guarantee against constraint violations.

Based on these deviation measures, we have a method that adapts to our knowledge of data distribution. Specifically, if only the support and the mean are known, we can use suitable parameters to achieve solutions that are reasonably less conservative compared to the worst case solution. Likewise, whenever the data distributions are available, we can incorporate these information to yield better solutions.

In the following, we present a specific pair of deviation measures suitable for bounded random variables. A general framework of deviation measures, which is useful for broader settings, can also be defined. However, in order not to interrupt the flow of the context, we present the general framework in Appendix D.

Forward and Backward Deviations Let $\tilde{z}$ be a bounded random variable and $M_{\tilde{z}}(s) = \mathrm{E}(\exp(s\tilde{z}))$ be its moment generating function. We define the set of values associated with forward deviations of $\tilde{z}$ as follows,

$$\mathcal{P}(\tilde{z}) = \left\{ \alpha \ : \ \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(\frac{\phi}{\alpha}\right) \leq \exp\left(\frac{\phi^2}{2}\right) \quad \forall \phi \geq 0 \right\}. \tag{17}$$

Likewise, for backward deviations, we define the following set,

$$\mathcal{Q}(\tilde{z}) = \left\{ \alpha \ : \ \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(-\frac{\phi}{\alpha}\right) \leq \exp\left(\frac{\phi^2}{2}\right) \quad \forall \phi \geq 0 \right\}. \tag{18}$$

For completeness, we also define $\mathcal{P}(c) = \mathcal{Q}(c) = \Re_+$ for any constant c. Observe that $\mathcal{P}(\tilde{z}) = \mathcal{Q}(\tilde{z})$ if $\tilde{z}$ is symmetrically distributed around its expectation. For known distributions, we define the forward deviation of $\tilde{z}$ as $p^*_{\tilde{z}} = \inf \mathcal{P}(\tilde{z})$ and the backward deviation as $q^*_{\tilde{z}} = \inf \mathcal{Q}(\tilde{z})$.

We note that the deviation measures are defined for some distributions with unbounded support such as normal distributions. However, some distributions do not have finite deviation measures, for example, exponentials and gamma distributions. Later in this section we will discuss the practicality of considering bounded support, where the existence of finite deviation measures is guaranteed.

The following results summarize the key properties of the deviation measure after we perform linear operations on independent random variables.

Theorem 3 *Let $\tilde{x}$ and $\tilde{y}$ be two independent random variables with zero means such that $p_{\tilde{x}} \in \mathcal{P}(\tilde{x})$, $q_{\tilde{x}} \in \mathcal{Q}(\tilde{x})$, $p_{\tilde{y}} \in \mathcal{P}(\tilde{y})$ and $q_{\tilde{y}} \in \mathcal{Q}(\tilde{y})$.*

(a) *If $\tilde{z} = a\tilde{x}$, then*

$$(p_{\tilde{z}}, q_{\tilde{z}}) = \begin{cases} (ap_{\tilde{x}}, aq_{\tilde{x}}) & \text{if } a \geq 0 \\ (-aq_{\tilde{x}}, -ap_{\tilde{x}}) & \text{otherwsie} \end{cases}$$

satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$. In other words, $p_{\tilde{z}} = \max\{ap_{\tilde{x}}, -aq_x\}$ and $q_{\tilde{z}} = \max\{aq_{\tilde{x}}, -ap_x\}$.

(b) *If $\tilde{z} = \tilde{x} + \tilde{y}$, then $(p_{\tilde{z}}, q_{\tilde{z}}) = \left(\sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2}, \sqrt{q_{\tilde{x}}^2 + q_{\tilde{y}}^2}\right)$ satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$.*

(c) *For all $p \geq p_{\tilde{x}}$ and $q \geq q_{\tilde{x}}$, we have $p \in \mathcal{P}(\tilde{x})$ and $q \in \mathcal{Q}(\tilde{x})$.*

(d)

$$\mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) \leq \exp\left(-\frac{\Omega^2}{2}\right)$$

and

$$\mathrm{P}\left(\tilde{x} < -\Omega q_{\tilde{x}}\right) \leq \exp\left(-\frac{\Omega^2}{2}\right).$$

Proof : (a) We can examine this condition easily from the definitions of $\mathcal{P}(\tilde{z})$ and $\mathcal{Q}(\tilde{z})$.

(b) To prove part (b), let $p_{\tilde{z}} = \sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2}$. We have that for any $\phi \geq 0$,

$$\begin{aligned}
& \mathrm{E}\left(\exp\left(\phi \tfrac{\tilde{x}+\tilde{y}}{p_{\tilde{z}}}\right)\right) \\
= \; & \mathrm{E}\left(\exp\left(\phi \tfrac{\tilde{x}}{p_{\tilde{z}}}\right) \exp\left(\phi \tfrac{\tilde{y}}{p_{\tilde{z}}}\right)\right) && [\text{since } \tilde{x} \text{ and } \tilde{y} \text{ are independent }] \\
= \; & \mathrm{E}\left(\exp\left(\phi \tfrac{p_{\tilde{x}}}{p_{\tilde{z}}} \tfrac{\tilde{x}}{p_{\tilde{x}}}\right)\right) \mathrm{E}\left(\exp\left(\phi \tfrac{p_{\tilde{y}}}{p_{\tilde{z}}} \tfrac{\tilde{y}}{p_{\tilde{y}}}\right)\right) \\
\leq \; & \exp\left(\tfrac{\phi^2}{2} \tfrac{p_{\tilde{x}}^2}{p_{\tilde{z}}^2}\right) \exp\left(\tfrac{\phi^2}{2} \tfrac{p_{\tilde{y}}^2}{p_{\tilde{z}}^2}\right) \\
= \; & \exp\left(\tfrac{\phi^2}{2}\right).
\end{aligned}$$

Thus, $p_{\tilde{z}} = \sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2} \in \mathcal{P}(\tilde{z})$. Similarly, we can show that $\sqrt{q_{\tilde{x}}^2 + q_{\tilde{y}}^2} \in \mathcal{Q}(\tilde{z})$

(c) Observe that

$$\mathrm{E}\left(\exp\left(\phi \frac{\tilde{x}}{p}\right)\right) = \mathrm{E}\left(\exp\left(\phi \frac{p_{\tilde{x}}}{p} \frac{\tilde{x}}{p_{\tilde{x}}}\right)\right) \leq \exp\left(\frac{\phi^2}{2} \frac{p_{\tilde{x}}^2}{p^2}\right) \leq \exp\left(\frac{\phi^2}{2}\right).$$

The proof relating to backward deviation is similar.

(d) Note that

$$\mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) = \mathrm{P}\left(\frac{\Omega \tilde{x}}{p_{\tilde{x}}} > \Omega^2\right) \leq \frac{\mathrm{E}\left(\exp\left(\frac{\Omega \tilde{x}}{p_{\tilde{x}}}\right)\right)}{\exp(\Omega^2)} \leq \exp\left(-\frac{\Omega^2}{2}\right),$$

where the first inequality follows from Chebyshev's inequality. The proof relating to backward deviation is the same. ■

For some distributions, we can find bounds on the deviations p^* and q^* or even close form expressions. In particular, for general distribution, we can show that these values are no less than the standard deviation. Interestingly, under normal distribution, these values coincide with standard deviation.

Proposition 4 *If the random variable $\tilde{z}$ has mean zero and standard deviation σ, then $p^*_{\tilde{z}} \geq \sigma$ and $q^*_{\tilde{z}} \geq \sigma$. If in addition, $\tilde{z}$ is normally distributed, then $p^*_{\tilde{z}} = q^*_{\tilde{z}} = \sigma$.*

Proof : Notice that for any $p \in \mathcal{P}(\tilde{z})$, we have

$$\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{p}\right)\right) = 1 + \frac{1}{2}\phi^2\frac{\sigma^2}{p^2} + \sum_{k=3}^{\infty}\frac{\phi^k \mathrm{E}[\tilde{z}^k]}{p^k k!},$$

and

$$\exp\left(\frac{\phi^2}{2}\right) = 1 + \frac{\phi^2}{2} + \sum_{k=2}^{\infty}\frac{\phi^{2k}}{2^k k!}.$$

According to the definition of $\mathcal{P}(\tilde{z})$, we have $\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{p}\right)\right) \leq \exp\left(\frac{\phi^2}{2}\right)$ for any $\phi \geq 0$. In particular, this inequality is true for ϕ close to zero, which implies that

$$\frac{1}{2}\phi^2\frac{\sigma^2}{p^2} \leq \frac{\phi^2}{2}.$$

Thus, $p \geq \sigma$. Similarly, for any $q \in \mathcal{Q}(\tilde{z})$, $q \geq \sigma$.

For the normal distribution, the proof follows trivially from the fact that

$$\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{\alpha}\right)\right) = \mathrm{E}\left(\exp\left(\phi\frac{\sigma}{\alpha}\frac{\tilde{z}}{\sigma}\right)\right) = \exp\left(\frac{\phi^2\sigma^2}{2\alpha^2}\right).$$

■

For most distributions, we are unable to obtain close form solutions of p^* and q^*. Nevertheless, we can still determine their values numerically. For instance, if $\tilde{z}$ is uniformly distributed over $[-1, 1]$, we can determine numerically that $p^* = q^* = 0.58$, which is close to the standard deviation 0.5774. In Table 1 we compare the values of p^*, q^* and standard deviation σ in which $\tilde{z}$ has discrete distributions as follows

$$\mathrm{P}(\tilde{z} = k) = \begin{cases} \beta & \text{if } k = 1 \\ 1-\beta & \text{if } k = -\frac{\beta}{1-\beta} \end{cases}. \tag{19}$$

In this example, the standard deviation is close to q^* but underestimates the value of p^*. Hence, it is apparent that if the distribution is asymmetric, the forward and backward deviations can differ from the standard deviation.

It will be clear in the subsequent section that we can use the values of $p^* = \inf\{\mathcal{P}(\tilde{z})\}$ and $q^* = \inf\{\mathcal{Q}(\tilde{z})\}$ in our uncertainty set to obtain the desired probability bound against constraint violation.

β	p^*	q^*	σ	$\bar{p}$	$\bar{q}$
0.5	1	1	1	1	1
0.4	0.83	0.82	0.82	0.83	0.82
0.3	0.69	0.65	0.65	0.69	0.65
0.2	0.58	0.50	0.50	0.58	0.50
0.1	0.47	0.33	0.33	0.47	0.33
0.01	0.33	0.10	0.10	0.33	0.10

Table 1: Numerical comparisons of different deviation measures for centered Bernoulli distributions.

Unfortunately, if the distribution of $\tilde{z}$ is not precisely known, we would not be able to determine values of p^* and q^*. Under such circumstances, as long as we can determine (p, q) such that $p \in \mathcal{P}(\tilde{z})$ and $q \in \mathcal{Q}(\tilde{z})$, we can still construct the uncertainty set that achieves the probabilistic guarantees, albeit more conservatively. In the following, we identify such (p, q) for a random variable $\tilde{z}$ assuming we only know its mean and support.

Theorem 4 *If $\tilde{z}$ has zero mean and distributed in $[-\underline{z}, \bar{z}]$, $\underline{z}, \bar{z} > 0$, then*

$$\bar{p} = \frac{\underline{z} + \bar{z}}{2} \sqrt{g\left(\frac{\underline{z} - \bar{z}}{\underline{z} + \bar{z}}\right)} \in \mathcal{P}(\tilde{z})$$

and

$$\bar{q} = \frac{\underline{z} + \bar{z}}{2} \sqrt{g\left(\frac{\bar{z} - \underline{z}}{\underline{z} + \bar{z}}\right)} \in \mathcal{Q}(\tilde{z}),$$

where

$$g(\mu) = 2 \max_{s>0} \frac{\phi_\mu(s) - \mu}{s^2},$$

and

$$\phi_\mu(s) = \ln\left(\frac{e^s + e^{-s}}{2} + \frac{e^s - e^{-s}}{2}\mu\right).$$

Proof : It is clear that through scaling and shifting,

$$\tilde{x} = \frac{\tilde{z} - (\bar{z} - \underline{z})/2}{(\underline{z} + \bar{z})/2} \in [-1, 1].$$

Thus, it suffices to show that

$$\sqrt{g(\mu)} \in \mathcal{P}(\tilde{x}),$$

where

$$\mu = \mathrm{E}[\tilde{x}] = \frac{\underline{z} - \bar{z}}{\underline{z} + \bar{z}} \in (-1, 1).$$

The proof related to backward deviation is similar.

First, observe that $p \in \mathcal{P}(\tilde{x})$ if and only if

$$\ln\left(\mathrm{E}\left[\exp\left(s\tilde{x}\right)\right]\right) \le (\tilde{x})s + \frac{p^2}{2}s^2, \forall s \ge 0. \tag{20}$$

We want to find a p such that the inequality (20) holds for for all possible random variables $\tilde{x}$ distributed in $[-1, 1]$ with mean μ. For this purpose, we formulate an infinite dimensional linear program as follows:

$$\begin{array}{ll} \max & \int_{-1}^{1} \exp(sx) f(x) dx \\ \text{s.t.} & \int_{-1}^{1} f(x) dx = 1 \\ & \int_{-1}^{1} x f(x) dx = \mu \\ & f(x) \ge 0. \end{array} \tag{21}$$

The dual of the above infinite dimensional linear program is

$$\begin{array}{ll} \min & u + v\mu \\ \text{s.t.} & u + vx \ge \exp(sx), \forall x \in [-1, 1]. \end{array}$$

Since $\exp(sx) - vx$ is convex in x, the dual is equivalent to a linear program with two decision variables.

$$\begin{array}{ll} \min & u + v\mu \\ \text{s.t.} & u + v \ge \exp(s) \\ & u - v \ge \exp(-s). \end{array} \tag{22}$$

It is easy to check that $(u^*, v^*) = \left(\frac{e^s + e^{-s}}{2}, \frac{e^s - e^{-s}}{2}\right)$ is the unique extreme point of the feasible set of problem (22) and $\mu \in (-1, 1)$. Hence problem (22) is bounded and in particular, the unique extreme point (u^*, v^*) is optimal. Therefore, $\frac{e^s + e^{-s}}{2} + \frac{e^s - e^{-s}}{2}\mu$ is the optimal objective value and by weak duality, it is an upper bound of the infinite dimensional linear program (21).

Notice that $\phi_\mu(0) = 0$ and $\phi'_\mu(0) = \mu$. Therefore, for any random variable $\tilde{x} \in [-1, 1]$ with mean μ, we have

$$\ln\left(\mathrm{E}\left[\exp\left(s\tilde{x}\right)\right]\right) \le \phi_\mu(s) = \phi_\mu(0) + \phi'_\mu(0)s + \frac{1}{2}s^2 \frac{\phi_\mu(s) - \mu s}{\frac{1}{2}s^2} \le \mu s + \frac{1}{2}s^2 g(\mu).$$

Hence, $\sqrt{g(\mu)} \in \mathcal{P}(\tilde{x})$. ■

Remark 1: This theorem implies that all probability distributions with bounded support have finite forward and backward deviations. It also enables us to find valid deviation measures from the support of distributions. In Table 1, we show the values of $\bar{p}$ and $\bar{q}$, which coincide with the deviation measures. Indeed, it is not difficult to see that $\sqrt{g(\mu)} = \inf\{\mathcal{P}(\tilde{x})\}$ for the two point random variable $\tilde{x}$ which takes value 1 with probability $(1+\mu)/2$ and -1 with probability $(1-\mu)/2$.

Remark 2: The function $g(\mu)$ defined in the theorem appears hard to analyze. Fortunately, the formulation can be simplified to $g(\mu) = 1-\mu^2$ for $\mu \in [0,1)$. Notice that

$$\frac{\phi_\mu(s) - \mu s}{\frac{1}{2}s^2} = 2\int_0^1 \phi''_\mu(s\xi)(1-\xi)d\xi,$$

and

$$\phi''_\mu(s) = 1 - \left(\frac{\alpha(s)+\mu}{1+\alpha(s)\mu}\right)^2,$$

where $\alpha(s) = (e^s - e^{-s})/(e^s + e^{-s}) \in [0,1)$ for $s \geq 0$. Since for $\mu \in (-1,1)$, $\inf_{0\leq\alpha<1} \frac{\alpha+\mu}{1+\alpha\mu} = \mu$, we have that for $\mu \in [0,1)$,

$$\phi''_\mu(s) \leq \phi''_\mu(0) = 1-\mu^2, \forall s \geq 0,$$

which implies that $g(\mu) = 1-\mu^2$ for $\mu \in [0,1)$.

Unfortunately, for $\mu \in (-1,0)$, we do not have a close form expression for $g(\mu)$. However, we can obtain some upper and lower bounds for the function $g(\mu)$. First notice that when $\mu \in (-1,0)$, we have $\phi''_\mu(s) \geq \phi''_\mu(0) = 1-\mu^2$ for s close to 0 and hence $1-\mu^2$ is a lower bound for $g(\mu)$. Further, numerically, we observe from Figure 1 that $g(\mu) \leq 1-0.3\mu^2$. On the other hand, when μ is close to -1, we can have a tighter lower bound for $g(\mu)$ as follows

$$\underline{p}^2(\mu) = \frac{(1-\mu)^2}{-2\ln\left((1+\mu)/2\right)}.$$

Indeed, since any distribution $\tilde{x}$ in $[-1,1]$ with mean μ satisfies

$$\mathrm{P}\left(\tilde{x} - \mu > \Omega\sqrt{g(\mu)}\right) \leq \exp(-\Omega^2/2),$$

we have that

$$\sqrt{g(\mu)} \geq \underline{p} = \inf\{p : \mathrm{P}(\tilde{x} - \mu > \Omega p) < \exp(-\Omega^2/2)\}.$$

In particular, when $\Omega = \sqrt{-2\ln((1+\mu)/2)}$, for the two point distribution $\tilde{x}$ which takes value 1 with probability $(1+\mu)/2$ and -1 with probability $(1-\mu)/2$, we obtain $\underline{p}^2 = \underline{p}^2(\mu) = \frac{(1-\mu)^2}{-2\ln((1+\mu)/2)}$. From Figure 1, we observe $\underline{p}^2(\mu)$ and $g(\mu)$ converge to 0 at the same rate as μ approaches -1.

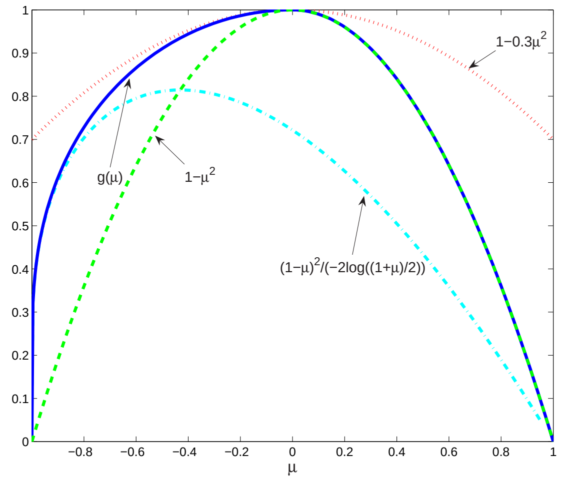

Figure 1: Function $g(\mu)$ and related bounds

3.1 Practicality of using Forward and Backward Deviations

Bounded support

In real world models, measurements are finite and hence, the assumption of bounded support is not practically restrictive. Although there are many interesting unbounded distributions with concise mathematical representations, beautiful properties (example memoryless property), elegant applications to stochastic processes leading to new insights, they are at best an asymptotic approximation of practical stochastic uncertainties. Therefore, to obtain valid deviation measures for unbounded random variables, we have to truncate the distributions so as to reflect upon their practical constraints. For instance, instead of assuming exponential distributions for inter-arrival time, it is reasonable to truncate the distribution to within say four to ten times its mean arrival. As an illustration, we consider a truncated exponential random variable, $\tilde{x}$ in $[0, \bar{x}]$ with distribution

$$f_{\tilde{x}}(x) = \frac{\exp(-x)}{1 - \exp(-\bar{x})}$$

and compare the deviation measures in Table 2. Although the forward deviation of a pure exponential distributed random variable is infinite, the truncated exponential distribution has reasonably small forward deviation compared to the support $\bar{x}$. Even when $\bar{x} = 10$, which is already a conservative level of truncation, the forward deviation is only slightly more than twice its standard deviation.

For random variables with unknown domain, we can use the support set to obtain upper limits of

$\bar{x}$	4	5	6	7	8	9	10	100
σ	0.834	0.911	0.954	0.977	0.989	0.995	0.998	1.000
p^*	1.037	1.239	1.419	1.583	1.733	1.871	2.000	7.000
q^*	0.834	0.911	0.954	0.977	0.989	0.995	0.998	1.000

Table 2: Deviation measures for truncated exponential variable with support $[0, \bar{x}]$.

the deviation measures. It is rather common for modelers to assume support set even in the absence of any data to construct the empirical distributions. For instance, in engineering models, the support sets are often related to known physical limits. In fact, it is a common notion in engineering that a practical random variable never exceeds three times its standard deviation.

Asymptotic behavior of the deviation measure estimators

Given a set of M independent samples, we can estimate the forward and backward deviations by using sample estimates of the moment generating functions. We conjecture that the standard deviations of these estimators decrease at the rate of $1/\sqrt{M}$. To demonstrate this, we perform empirical studies on the convergence of the forward deviation and backward deviation estimators (respectively $\hat{p}^*_M$ and $\hat{q}^*_M$) for standard normal distribution. Indeed, the forward and backward deviation estimators approach one as sample size increases. We repeat the experiment 1000 times to estimate the standard deviations of forward deviation and backward deviation estimators (respectively $\hat{\sigma}\,(\hat{p}^*_M)$ and $\hat{\sigma}\,(\hat{q}^*_M)$) and present the results in Table 3. The convergence rate supports the conjecture. Hence, besides relying on valid support sets, we can also use historical data to estimate the forward and backward deviations. Therefore, as much as we could assume unknown domains with known mean and second moments, we can also assume unknown domains with known mean, forward deviation and backward deviation.

Comparison with probability bounds derived from standard deviation

It is interesting to compare the approach with probability bounds based on standard deviation. We believe that the forward and backward deviations should provide a better bound for a wide variety of practical bounded distributions compared to standard deviation. Here we explain why. For any random variable $\tilde{z}$ with mean zero and standard deviation σ, forward deviation, p^*, and backward deviation, q^* we have from Chebyshev Inequality,

$$\mathrm{P}(\tilde{z} > \Lambda\sigma) \leq 1/\Lambda^2, \tag{23}$$

M	$\hat{\sigma}(\hat{p}^*_M)$	$\hat{\sigma}(\hat{q}^*_M)$	$1/\sqrt{M}$
10	0.2191	0.2182	0.3162
20	0.1697	0.1644	0.2236
50	0.1135	0.1176	0.1414
100	0.0818	0.0821	0.1000
200	0.0645	0.0695	0.0707

Table 3: Convergence rate of deviation estimators.

while the bound provided by the forward deviation is

$$\mathrm{P}(\tilde{z} > \Omega p^*) \leq \exp\left(-\Omega^2/2\right). \tag{24}$$

For the same violation probability, ϵ, bound (23) suggests $\Lambda = 1/\sqrt{\epsilon}$ while bound (24) requires $\Omega = \sqrt{-2\ln(\epsilon)}$. Since the probability bounds are asymptotically tight for some distributions[1], to compare both bounds, we can examine the magnitudes of $\Lambda\sigma$ and Ωp^* for various distributions when ϵ approaches zero. For any distribution with the forward deviation close to the standard deviation (such as normal distribution), we expect the bound (23) to perform poorly comparing to (24). Furthermore, as p^* is finite for bounded distributions, the magnitude of $\Lambda\sigma$ will exceed Ωp^* as ϵ approaches zero. For example, in the case of the centered Bernoulli distribution of (19) with $\beta = 0.01$, we have $\sigma = 0.1$ and $p^* = 0.33$. Hence, $\Lambda\sigma > \Omega p^*$ for $\epsilon < 0.0099$. It is often necessary in robust optimization to protect against low probability "disruptive events" that may result in large deviations. Otherwise, such rare events should be totally ignored in the model. Therefore, since the event $\tilde{z} = 1$ is a "destructive" one, it is reasonable to choose $\epsilon < 0.0099 \approx P(\tilde{z} = 1) = 0.01$, in which case, it would be better to use the bound (24). Another disadvantage of using standard deviation for bounding probabilities is the inability to capture distributional skewness. As evident from the two point distribution of (19) when β is small, the same value of $\Lambda\sigma$ to ensure that $\mathrm{P}(\tilde{z} < -\Lambda\sigma) < \epsilon$ can be large compared to Ωq^*.

[1]The bound (23) is asymptotically tight for the centered Bernoulli distribution of (19) in which $\beta = \epsilon$ and ϵ approaches zero. Indeed, to safeguard against the low probability event of $\tilde{z} = 1$, we require Λ to be at least $1/\sigma = 1/\sqrt{\beta + \beta^2/(1-\beta)} \approx 1/\sqrt{\epsilon}$, so that $\mathrm{P}(\tilde{z} > \Lambda\sigma) < \epsilon$. For the same two point distribution, we verify numerically that Ωp^* converges to one, as $\beta = \epsilon$ approaches to zero, suggesting that the bound (24) is also asymptotically tight.

4 Probability Bounds of Constraint Violation

In this section, we will show how the new deviation measures relate to the probability bound of constraint violations in the robust framework.

Model of Data Uncertainty, U:

We assume that the primitive uncertainties $\{\tilde{z}_j\}_{j=1:N}$ are independent, zero mean random variables, with support $\tilde{z}_j \in [-\underline{z}_j, \bar{z}_j]$, and deviation measures, (p_j, q_j) satisfying,

$$p_j \in \mathcal{P}(\tilde{z}_j), q_j \in \mathcal{Q}(\tilde{z}_j) \ \forall j = \{1, \ldots, N\}.$$

We consider the generalized uncertainty set $\mathcal{G}_\Omega$, which takes into account the worst case support set, $\mathcal{W}$.

Theorem 5 *Let $\boldsymbol{x}$ be feasible for the robust counterpart of* (3) *in which $\mathcal{U}_\Omega = \mathcal{G}_\Omega$ then*

$$\mathrm{P}\left(\tilde{\boldsymbol{a}}'\boldsymbol{x} > \tilde{b}\right) \leq \exp\left(\frac{-\Omega^2}{2}\right).$$

Proof : Since $\boldsymbol{x}$ is feasible in (15), using the equivalent formulation of inequality (16), it follows that

$$\begin{aligned} & \mathrm{P}\left(\tilde{\boldsymbol{a}}'\boldsymbol{x} > \tilde{b}\right) \\ = \ & \mathrm{P}\left(\boldsymbol{a}^{0\prime}\boldsymbol{x} + \tilde{\boldsymbol{z}}'\boldsymbol{y} > b^0\right) \\ \leq \ & \mathrm{P}\left(\tilde{\boldsymbol{z}}'\boldsymbol{y} > \min_{\boldsymbol{r},\boldsymbol{s}\geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\}\right) \\ \leq \ & \mathrm{P}\left(\tilde{\boldsymbol{z}}'\boldsymbol{y} > \min_{\boldsymbol{r},\boldsymbol{s}\geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|_2 + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\}\right), \end{aligned}$$

where $y_j = \boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j$ and

$$\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s}) = \begin{bmatrix} \max(p_1(y_1 - r_1 + s_1), -q_j(y_1 - r_1 + s_1)) \\ \vdots \\ \max(p_N(y_N - r_N + s_N), -q_j(y_N - r_N + s_N)) \end{bmatrix}.$$

Let

$$(\boldsymbol{r}^*, \boldsymbol{s}^*) = \arg\min_{\boldsymbol{r},\boldsymbol{s}\geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|_2 + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\}$$

and $\boldsymbol{t}^* = \boldsymbol{t}(\boldsymbol{r}^*, \boldsymbol{s}^*)$. Observe that since $-\underline{z}_j \leq \tilde{z}_j \leq \bar{z}_j$, we have $r_j^*\bar{z}_j \geq r_j^*\tilde{z}_j$ and $s_j^*\underline{z}_j \geq -s_j^*\tilde{z}_j$. Therefore,

$$\mathrm{P}\left(\tilde{\boldsymbol{z}}'\boldsymbol{y} > \Omega\|\boldsymbol{t}^*\|_2 + \boldsymbol{r}^{*\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{*\prime}\underline{\boldsymbol{z}}\right) \leq \mathrm{P}\left(\tilde{\boldsymbol{z}}'(\boldsymbol{y} - \boldsymbol{r}^* + \boldsymbol{s}^*) > \Omega\|\boldsymbol{t}^*\|_2\right).$$

From Theorem 3(a), we have $t_j^* \in \mathcal{P}(\tilde{z}_j(y_j - r_j^* + s_j^*))$. Following Theorem 3(b), we have

$$\|\boldsymbol{t}^*\|_2 \in \mathcal{P}\left(\tilde{\boldsymbol{z}}'(\boldsymbol{y} - \boldsymbol{r}^* + \boldsymbol{s}^*)\right).$$

Finally, the desired probability bound follows from Theorem 3(d). ■

We use Euclidian norm as the benchmark to obtain the desired probability bound. Naturally, we can use other norms such as $l_1 \cap l_\infty$-norm $\|\boldsymbol{z}\| = \max\left\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\right\}$ to achieve the same bound. The natural question is whether the approximation is worthwhile. Noting from the inequality (12), the value $\Omega\|\boldsymbol{t}\|^*$, gives the desired "safety distance" against constraint violation. Since, by design $\|\boldsymbol{t}\|^* \geq \|\boldsymbol{t}\|_2$, a way to compare the level of conservativeness between the choice of norms is through the worst case ratio as follows

$$\gamma = \max_{\boldsymbol{t}\neq\boldsymbol{0}} \frac{\|\boldsymbol{t}\|^*}{\|\boldsymbol{t}\|_2}.$$

It turns out that for $l_1 \cap l_\infty$ norm, $\gamma = \sqrt{\lfloor\sqrt{N}\rfloor + (\sqrt{N} - \lfloor\sqrt{N}\rfloor)^2} \approx N^{1/4}$ (see Bertsimas and Sim [10] and Bertsimas et al. [8]). Hence, although the resultant model is linear of manageable size, the choice of the polyhedral norm can lead to more conservative solution compared to the use of Euclidian norm.

Using the forward and backward deviations, the proposed robust counterpart generalizes the results of Ben-Tal and Nemirovski [4] and Bertsimas and Sim [10]. Indeed, if $\tilde{z}_j$ has support in $[-1, 1]$, from Theorem 4, we have $p_j = q_j = 1$ and hence, we will obtain the same robust counterparts. Our result is also stronger, as we do not require distribution symmetry to ensure the same probability bound of $\exp(-\Omega^2/2)$. The worst case budget $\Omega_{\max}$ is at least $\sqrt{N}$ so that $\mathcal{G}_{\Omega_{\max}} = \mathcal{W}$, which can be far conservative when N is large. Therefore, even if little is known about the underlying distribution, except for the mean and support, this approach is potentially less conservative than the worst case solution.

5 Stochastic Programs with Chance Constraints

Consider the following two stage stochastic program,

$$\begin{aligned} Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \mathrm{E}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} \quad & \boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}(\tilde{\boldsymbol{z}}) \leq f_i(\tilde{\boldsymbol{z}}), \text{ a.e.}, \qquad \forall i \in \{1, \ldots, m\}, \\ & \boldsymbol{x} \in \Re^{n_1}, \\ & \boldsymbol{y}(\cdot) \in Y, \end{aligned} \tag{25}$$

where $\boldsymbol{x}$ corresponds to the first stage decision vector, and $\boldsymbol{y}(\tilde{\boldsymbol{z}})$ being the recourse function in a space of measurable functions, Y with domain $\mathcal{W}$ and range $\Re^{n_2}$. Note that to optimize over the space of measurable functions amounts to solving an optimization problem with potentially large or even infinite number of variables. In general, however, finding a first stage solution, $\boldsymbol{x}$, such that there exists feasible recourse for all realization of uncertainty can be a intractable problem (see Ben-Tal et al. [5] and Shapiro and Nemirovski [23]). Nevertheless, in some applications of stochastic optimization, the risk of infeasibility can be tolerated as a tradeoff to improve upon the objective value. Therefore, we consider the following stochastic program with chance constraints,

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \mathrm{E}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}(\tilde{\boldsymbol{z}}) \le f_i(\tilde{\boldsymbol{z}})\right) \ge 1 - \epsilon_i \quad \forall i \in \{1,\ldots,m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}(\cdot) \in Y,
\end{aligned}
\tag{26}
$$

in which $\epsilon_i > 0$. To obtain less conservative solution, we could vary the risks of constraint violations, and hence extend the feasible space for solutions, $\boldsymbol{x}$ and $\boldsymbol{y}(\cdot)$.

Under the *Model of Data Uncertainty, U*, we assume that $\tilde{z}_j \in [-\underline{z}_j, \bar{z}_j]$, $j \in \{1,\ldots,N\}$ are independent random variables with mean zero and deviation parameters (p_j, q_j) satisfying $p_j \in \mathcal{P}(\tilde{z}_j)$ and $q_j \in \mathcal{Q}(\tilde{z}_j)$. For all $i \in \{1,\ldots,m\}$, under the *Affine Data Perturbation*, we have

$$
\boldsymbol{a}_i(\tilde{\boldsymbol{z}}) = \boldsymbol{a}_i^0 + \sum_{j=1}^N \boldsymbol{\Delta a}_i^j \tilde{z}_j,
$$

and

$$
f_i(\tilde{\boldsymbol{z}}) = f_i^0 + \sum_{j=1}^N \Delta f_i^j \tilde{z}_j.
$$

To design a tractable robust optimization approach for solving (26), we restrict the recourse function $\boldsymbol{y}(\cdot)$ to one of *linear decision rule* as follows,

$$
\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum_{j=1}^N \boldsymbol{y}^j z_j. \tag{27}
$$

Apparently, linear decision rules surfaced in early development of stochastic optimization (see Garstka and Wets [19]) and reappeared recently in *affinely adjustable robust counterpart* introduced by Ben-Tal et al. [5]. The linear decision rule enables us to design a tractable robust optimization approach for finding feasible solutions in the model (26) for all distributions satisfying the *Model of Data Uncertainty, U.*

Theorem 6 *The optimal solution to the following robust counterpart,*

$$
\begin{aligned}
Z_r^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{y}^0 \\
\text{s.t.} \quad & \boldsymbol{a}_i^{0\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^0 + \Omega_i h_i + \boldsymbol{r}^{i\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{i\prime}\underline{\boldsymbol{z}} \le f_i^0 \quad \forall i \in \{1, \ldots, m\} \\
& \|\boldsymbol{u}^i\|^* \le h_i \quad \forall i \in \{1, \ldots, m\} \\
& u_j^i \ge p_j(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall i \in \{1, \ldots, m\}, j \in \{1, \ldots, N\} \\
& u_j^i \ge -q_j(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall i \in \{1, \ldots, m\}, j \in \{1, \ldots, N\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}^j \in \Re^k \quad \forall j \in \{0, \ldots, N\} \\
& \boldsymbol{u}^i, \boldsymbol{r}^i, \boldsymbol{s}^i \in \Re_+^N, h_i \in \Re \quad \forall i \in \{1, \ldots, m\},
\end{aligned} \tag{28}
$$

in which $\Omega_i = \sqrt{-2\ln(\epsilon_i)}$ *is feasible in the stochastic optimization model (26) for all distributions that satisfy the Model of Data Uncertainty,* U *and* $Z_r^* \ge Z^*$*.*

Proof : Restricting the space of recourse solutions $\boldsymbol{y}(\boldsymbol{z})$ in the form of Eq. (27), we have

$$\mathrm{E}\left(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})\right) = \boldsymbol{d}'\boldsymbol{y}^0.$$

Hence, the following problem,

$$
\begin{aligned}
Z_1^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{y}^0 \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i^{0\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^0 + \textstyle\sum_{j=1}^N \left(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j\right)\tilde{z}_j \le f_i^0\right) \ge 1 - \epsilon_i \quad \forall i \in \{1, \ldots, m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}^j \in \Re^k \quad \forall j \in \{0, \ldots, N\},
\end{aligned} \tag{29}
$$

gives the upper bound to the model (26). Applying Theorem 5 and using Theorem 2, the feasible solution of the model (28) is also feasible in the model (29) for all distributions that satisfy the *Model of Data Uncertainty, U.* Hence, $Z_r^* \ge Z_1^* \ge Z^*$. ∎

We can easily extend the framework to T stage stochastic program with chance constraints as follows,

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \sum_{t=1}^T \mathrm{E}\left[\boldsymbol{d}_t'\boldsymbol{y}_t(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_t)\right] \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_T)'\boldsymbol{x} + \textstyle\sum_{t=1}^T \boldsymbol{b}_{it}'\boldsymbol{y}_t(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_t) \le f_i(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_T)\right) \ge 1 - \epsilon_i \quad \forall i \in \{1, \ldots, m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}_t(\boldsymbol{z}_1, ..\boldsymbol{z}_t) \in \Re^k \quad \forall t = 1, ..T, \underline{\boldsymbol{z}}_t \le \boldsymbol{z}_t \le \bar{\boldsymbol{z}}_t,
\end{aligned} \tag{30}
$$

In the multi-period model, we assume that the underlying uncertainties, $\tilde{\boldsymbol{z}}_1 \in \Re^{N_1}, \ldots, \tilde{\boldsymbol{z}}_T \in \Re^{N_T}$ unfolds progressively from the first period to the last period. The vector of primitive uncertainties, $\tilde{\boldsymbol{z}}_t$ is only available at the t^{th} period. Hence, under the *Affine Data Perturbation*, we may assume that $\tilde{\boldsymbol{z}}_t$ is statistically independent with other periods. With the above assumptions, we obtain

$$\boldsymbol{a}_i(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_T) = \boldsymbol{a}_i^0 + \sum_{t=1}^{T}\sum_{j=1}^{N_t} \boldsymbol{\Delta a}_{it}^j \tilde{z}_t^j \ ,$$

and

$$f_i(\tilde{\boldsymbol{z}}_1, .., \tilde{\boldsymbol{z}}_T) = f_i^0 + \sum_{t=1}^{T}\sum_{j=1}^{N_t} \Delta f_{it}^j \tilde{z}_t^j.$$

In order to derive the robust formulation of the multi-period model, we similarly use linear decision rules on the recourse function that fulfill the nonanticipativity requirement as follows,

$$\boldsymbol{y}_t(\boldsymbol{z}_1, .., \boldsymbol{z}_t) = \boldsymbol{y}_t^0 + \sum_{\tau=1}^{t}\sum_{j=1}^{N_t} \boldsymbol{y}_\tau^j z_\tau^j \ .$$

Essentially, multiperiod robust model is the same as the two period model we have presented and does not suffer from the "curse of dimensionality".

5.1 On Linear Decision Rule

Linear decision rule is the key enabling mechanism that permits scalability to mutistage models. It has appeared in earlier proposals of stochastic optimization, however, due to the perceived limitations, the method was short-lived (see Garstka and Wets [19]). Indeed, *hard* constraints, such as, $y(\tilde{\boldsymbol{z}}) \geq 0$, can nullify any benefit of linear decision rules on the recourse function, $y(\tilde{\boldsymbol{z}})$. As an illustration, we consider a linear decision rule,

$$y(\tilde{\boldsymbol{z}}) = y_0 + \sum_{j=1}^{N} y_j \tilde{z}_j,$$

where the primitive uncertainties, $\tilde{\boldsymbol{z}}$ have unbounded support and finite forward and backward deviations (such as normal distributions). In this extreme example, we investigate the following hard constraints,

$$\begin{array}{c} y(\tilde{\boldsymbol{z}}) \geq 0 \\ y(\tilde{\boldsymbol{z}}) \geq b(\tilde{\boldsymbol{z}}) = b_0 + \sum_{j=1}^{N} b_j \tilde{z}_j, \end{array} \tag{31}$$

where $b_j \neq 0$, on the linear decision rule. Since the support of $\tilde{\boldsymbol{z}}$ is unbounded, the constraints (31) imply $y_j = 0$ and $y_j = b_j$ for all $j \in \{1, \ldots, N\}$, which resulted in infeasibility of the linear decision rule. However, the linear decision rule can survive in *soft* constraints such as

$$\mathrm{P}(y(\tilde{\boldsymbol{z}}) \geq 0) \leq 1 - \epsilon,$$

$$\mathrm{P}(y(\tilde{\boldsymbol{z}}) \geq b(\tilde{\boldsymbol{z}})) \leq 1 - \epsilon$$

even for very high reliability. For instance, suppose $p_j = q_j = 1$, and $\epsilon = 10^{-7}$, the following robust counterpart approximation of the chance constraints becomes

$$y_0 \geq \Omega \| [y_1, \ldots, y_N] \|_2,$$
$$y_0 - b_0 \geq \Omega \| [y_1 - b_1, \ldots, y_N - b_N] \|_2,$$

where $\Omega = 5.68$. Since, $\Omega = \sqrt{-2\ln(\epsilon)}$ is a small number even for high robustness, the solution space for the linear decision rule is not overly constrained. Hence, the linear decision rule may remain viable if risk of infeasibility in the stochastic optimization model can be tolerated. For some applications, the linear decision rule seems to perform reasonably well (see Ben-Tal et al [5, 6]). As a matter of fact, the project management example presented next will further illuminate the benefits of linear decision rules.

6 Application Example: Project Management under Uncertain Activity Time

Project management problems can be represented by a directed graph with m arcs and n nodes. Each node on the graph represents an event marking the completion of a particular subset of activities. We denote the set of directed arcs on the graph as $\mathcal{E}$. Hence, an arc $(i, j) \in \mathcal{E}$ is an activity that connects event i to event j. By convention, we use node 1 as the start event and the last node n as the end event.

We consider a project management example of several activities. Each activity has random completion time, $\tilde{t}_{ij}$. The completion of activities must adhere to precedent constraints. For instance, activity e_1 precedes activity e_2 if activity e_1 must be completed before activity e_2. Analysis on stochastic project management problem such as determining the expected completion time and quantile of completion time is well known to be notoriously difficult (Hagstrom [20]).

In our computational experiment, we assume that the random completion time $\tilde{t}_{ij}$ is independent of the completion times of other activities and its mean is affinely dependent on the additional amount of resource, $x_{ij} \in [0, \bar{x}_{ij}]$ committed on the activity as follows

$$\tilde{t}_{ij} = (1 + \tilde{z}_{ij})b_{ij} - a_{ij}x_{ij}, \tag{32}$$

where $\tilde{z}_{ij} \in [-\underline{z}_{ij}, \bar{z}_{ij}]$, $\underline{z}_{ij} \leq 1$, $(i, j) \in \mathcal{E}$ are independent random variables with zero means and deviation measures, (p_{ij}, q_{ij}) satisfying $p_{ij} \in \mathcal{P}(\tilde{z}_{ij})$ and $q_{ij} \in \mathcal{Q}(\tilde{z}_{ij})$. We also assume that $\tilde{t}_{ij} \geq 0$ for all realization of $\tilde{z}_{ij}$ and valid range of x_{ij}. We note that the assumption of independent activity completion times can be rather strong and difficult to verify in practice. Nevertheless, we have to identify some form of independence in order to have any benefit of risk pooling. Otherwise, the solution would have been a conservative one. We highlight that the model can easily be extended to include linear dependency of activity completion times such as sharing common resources with independent failure probabilities.

Denote c_{ij} to be the cost of using each unit of resource for the activity on the arc (i,j). Our goal is to find a resource allocation to each activity $(i,j) \in \mathcal{E}$ that minimizes the total project cost while ensuring that the probability we can complete the project within T days is at least $1-\epsilon$.

Proposition 5 *For any ϵ_0 and ϵ_{ij}, $\forall (i,j) \in \mathcal{E}$ in $(0,1)$ such that*

$$\epsilon_0 + \sum_{(i,j)\in\mathcal{E}} \epsilon_{ij} \leq \epsilon \ , \tag{33}$$

if there exists measurable functions $y_i(\tilde{z})$ for every node i satisfying

$$\begin{aligned}
&\mathrm{P}\left(y_n(\tilde{\boldsymbol{z}}) \leq T\right) \geq 1-\epsilon_0 \\
&\mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \geq \tilde{t}_{ij}\Big) \geq 1-\epsilon_{ij} \quad \forall (i,j) \in \mathcal{E} \\
&y_1(\tilde{\boldsymbol{z}}) = 0 \ ,
\end{aligned}$$

the probability that the project is finished within T days is at least $1-\epsilon$.

Proof : For any realization $\boldsymbol{z}$ of $\tilde{\boldsymbol{z}}$, the project to can be finished within time period T if and only if there exists y_i for all nodes i such that the following inequalities are satisfied

$$\begin{aligned}
y_n &\leq T \\
y_1 &= 0 \\
y_j &\geq y_i + t_{ij}(\boldsymbol{z}) \qquad \forall (i,j) \in \mathcal{E}.
\end{aligned}$$

Since $y_1(\tilde{\boldsymbol{z}}) = 0$, we have

$$\begin{aligned}
&\mathrm{P}\,(\text{ Project finished within } T\ ) \\
\geq\ & \mathrm{P}\left(\{y_n(\tilde{\boldsymbol{z}}) \leq T\} \bigcap_{(i,j)\in\mathcal{E}} \{y_j(\tilde{\boldsymbol{z}}) \geq y_i(\tilde{\boldsymbol{z}}) + \tilde{t}_{ij}\}\right) \\
=\ & 1 - \mathrm{P}\left(\{y_n(\tilde{\boldsymbol{z}}) > T\} \bigcup_{(i,j)\in\mathcal{E}} \{y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) < \tilde{t}_{ij}\}\right) \\
\geq\ & 1 - \left(\mathrm{P}\Big(y_n(\tilde{\boldsymbol{z}}) > T\Big) + \sum_{(i,j)\in\mathcal{E}} \mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) < \tilde{t}_{ij}\Big)\right) \quad \text{(Bonferroni inequality)} \\
\geq\ & 1 - \left(\epsilon_0 + \sum_{(i,j)\in\mathcal{E}} \epsilon_{ij}\right) \\
\geq\ & 1 - \epsilon \ .
\end{aligned}$$

$\blacksquare$

Therefore, a stochastic model to address the project management problem is as follows:

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \mathrm{P}\left(y_n(\tilde{\boldsymbol{z}}) \le T\right) \ge 1 - \epsilon_0 \\
& \mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \ge (1 + \tilde{z}_{ij})b_{ij} - a_{ij}x_{ij}\Big) \ge 1 - \epsilon_{ij} \quad \forall (i,j) \in \mathcal{E} \\
& y_1(\tilde{\boldsymbol{z}}) = 0 \\
& \boldsymbol{0} \le \boldsymbol{x} \le \bar{\boldsymbol{x}} \\
& \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\
& \boldsymbol{y}(\boldsymbol{z}) \in \Re^n \quad \forall \underline{\boldsymbol{z}} \le \boldsymbol{z} \le \bar{\boldsymbol{z}},
\end{aligned}
\tag{34}
$$

Notice in the deterministic project management formulation, y_i corresponds a feasible completion time at node i.

Applying Theorem 6, we formulate the robust optimization approach as follows

$$
\begin{aligned}
Z_r^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & y_n^0 + \Omega_0 h_0 + \boldsymbol{r}^{0\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{0\prime}\underline{\boldsymbol{z}} \le T \\
& \|\boldsymbol{u}^0\|^* \le h_0 \\
& u_{ij}^0 \ge p_{ij}(y_n^{ij} - r_{ij}^0) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^0 \ge -q_{ij}(y_n^{ij} + s_{ij}^0) \quad \forall (i,j) \in \mathcal{E} \\
& y_j^0 - y_i^0 \ge b_{ij} - a_{ij}x_{ij} + \Omega_{ij}h_{ij} + \boldsymbol{r}^{ij\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{ij\prime}\underline{\boldsymbol{z}} \quad \forall (i,j) \in \mathcal{E} \\
& \|\boldsymbol{u}^{ij}\|^* \le h_{ij} \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{ij} \ge p_{ij}(b_{ij} + y_i^{ij} - y_j^{ij} - r_{ij}^{ij} + s_{ij}^{ij}) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{kl} \ge p_{ij}(y_k^{ij} - y_l^{ij} - r_{ij}^{kl} + s_{ij}^{kl}) \quad \forall (i,j),(k,l) \in \mathcal{E}, (i,j) \ne (k,l) \\
& u_{ij}^{ij} \ge -q_{ij}(b_{ij} + y_i^{ij} - y_j^{ij} - r_{ij}^{ij} + s_{ij}^{ij}) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{kl} \ge -q_{ij}(y_k^{ij} - y_l^{ij} - r_{ij}^{kl} + s_{ij}^{kl}) \quad \forall (i,j),(k,l) \in \mathcal{E}, (i,j) \ne (k,l) \\
& y_1^0 = 0, y_1^{ij} = 0 \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{0} \le \boldsymbol{x} \le \bar{\boldsymbol{x}} \\
& \boldsymbol{u}^0, \boldsymbol{u}^{ij}, \boldsymbol{r}^0, \boldsymbol{r}^{ij}, \boldsymbol{s}^0, \boldsymbol{s}^{ij} \in \Re_+^{|\mathcal{E}|} \quad \forall (i,j) \in \mathcal{E} \\
& h_0, h_{ij} \in \Re \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\
& \boldsymbol{y}^0, \boldsymbol{y}^{ij} \in \Re^n \quad \forall (i,j) \in \mathcal{E}.
\end{aligned}
\tag{35}
$$

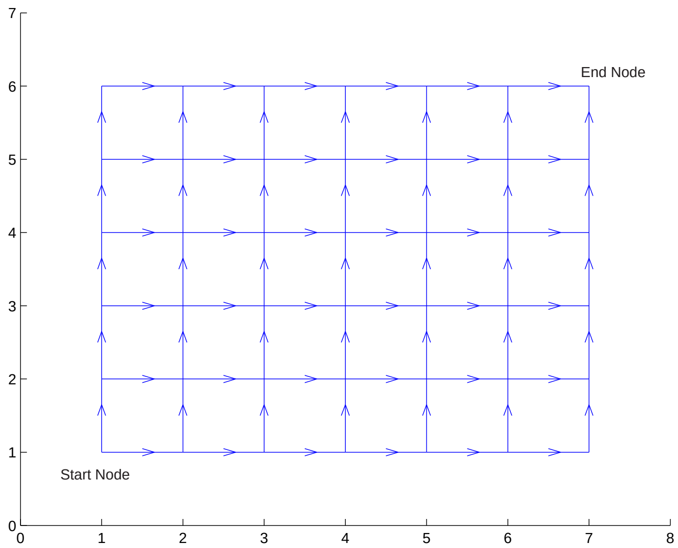

Figure 2: Project management "grid" with height, $H = 6$ and width $W = 7$.

There are multiple ways of selecting ϵ_0 and ϵ_{ij}, $(i,j) \in \mathcal{E}$ so that the project will complete timely with probability at least $1-\epsilon$. As indicated in Proposition 5, a sufficient condition is (33). A reasonable way to select these values, ϵ_0, ϵ_{ij}, $(i,j) \in \mathcal{E}$ is to minimize the total budget of uncertainties for all the constraints as follows

$$\begin{aligned} \min \quad & \Omega_0(\epsilon_0) + \sum_{(i,j)\in\mathcal{E}} \Omega_{ij}(\epsilon_{ij}) \\ \text{s.t.} \quad & \epsilon_0 + \sum_{(i,j)\in\mathcal{E}} \epsilon_{ij} \le \epsilon, \end{aligned}$$

where $\Omega_e(\epsilon_e) = \sqrt{-2\ln(\epsilon_e)}$. Solving the above optimization problem, we obtain $\epsilon_0 = \epsilon_{ij} = \frac{\epsilon}{m+1}$, for all $(i,j) \in \mathcal{E}$. Considering a project with $10{,}000$ jobs, to guarantee that the project will be completed on time with probability 99%, we choose $\Omega = 5.26$, which is essentially a small constant compared to $\Omega_{\max} = \sqrt{m} = 100$.

For our computational experiment, we create a fictitious project with the activity network in the form of H by W grid (see Fig. 2). There are a total of $H \times W$ nodes, with the first node at the left bottom corner and the the last node at the right upper corner. Each arc on the graph proceeds either towards the right node or the upper node. We assume that every activity on arc has independent and

H	W	m	n	Ω	Z_r^*	Z_w^*	$\frac{Z_r^*}{Z_w^*}$
4	4	24	16	3.95	511.05	576	0.89
5	5	40	25	4.07	856.01	960	0.89
6	6	60	36	4.17	1294.54	1440	0.90
3	3	12	9	3.77	269.82	288	0.94
3	4	17	12	3.86	367.06	408	0.90
3	8	37	24	4.05	519.69	888	0.59
3	12	57	36	4.16	587.09	1368	0.43

Table 4: Computation results for project management.

identical completion time. In particular, for all arcs (i, j),

$$\mathrm{P}(\tilde{z}_{ij} = z) = \begin{cases} 0.6 & \text{if } z = -0.06 \\ 0.3 & \text{if } z = 0.04 \\ 0.1 & \text{if } z = 0.24, \end{cases}$$

hence, $\underline{z}_{ij} = 0.06$, $\bar{z}_{ij} = 0.24$ and we can determine numerically that $p_{ij} = 0.1154$, $q_{ij} = 0.0917$. We assume that for all activities, $a_{ij} = c_{ij} = 1$, $\bar{x}_{ij} = 24$ and $b_{ij} = 100$. We also want high confidence (at least 99%) that the completion time of the project is no more than $100(H+W-2)$, which is the average completion time of any path with $x_{ij} = 0$ for all arcs. Therefore, additional resources are needed to meet the desired confidence level of project completion.

In the worst case scenario, every activity will be delayed, that is $\tilde{z}_{ij} = 0.24$ and $(1+\tilde{z}_{ij})b_{ij} = 124$. In this case, each activity on arc must be assigned to the maximum resource at $x_{ij} = 24$, so that $\tilde{t}_{ij} = 100$ and all critical paths (longest paths) can meet the targeted completion time of $100(H + W - 2)$. Since there are a total of $m = H(W - 1) + W(H - 1)$ arcs on the H by W grid, the objective according to the worst case scenario is $Z_w^* = 24m$.

We use Euclidian norm to construct our uncertainty sets. Therefore the resultant model is an SOCP. We solve the optimization model to optimality using SDPT3 [25] and display the results in Table 4.

Interestingly, the relative savings for using the robust model depends on the activity network. We first note that when the activity network is a square, the relative cost savings hover around 10% and somewhat indifferent to the size of the problem. In Fig. 3 we illustrate the optimal solution, the relative amount of resource for each activity, by varying the thickness of the arcs on the grid. We can see that except for the upper-left and lower-right activities, all other activities require almost full resources. As

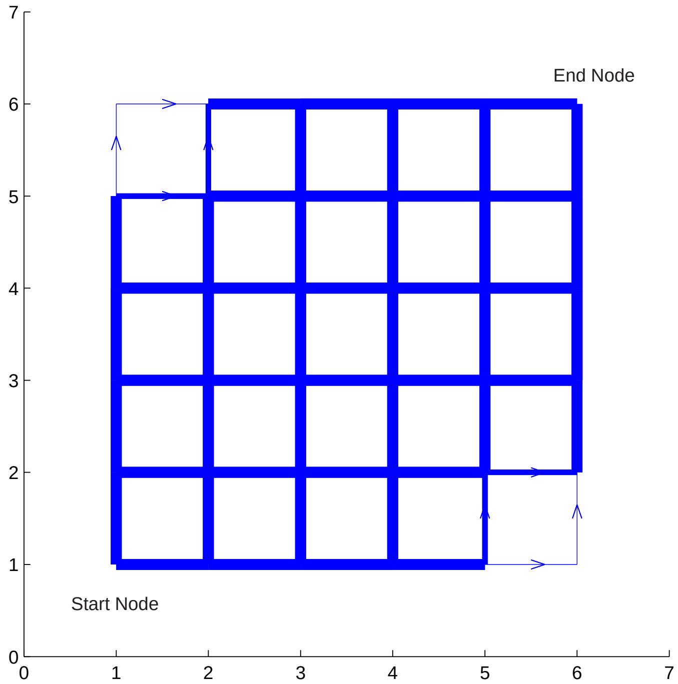

Figure 3: Project management solution for $H = 6$ and $W = 6$.

the activities are tightly connected, any delay in one activity can lead to a substantial delay in the entire project. Hence, there is little leeway for cost savings without having to compromise the completion time.

In Fig. 4 we illustrate the solution on the grid for a rectangular activity graph. When the activity network is more rectangular, we see that we only need to pay for a fraction of the worst case cost in order to guarantee with high confidence that the project will be completed on time. It seems to suggest that if there are more sequential activities and less parallel activities, the robust project management can lead to substantial cost savings while guaranteeing high degree of confidence that the project will complete on time.

7 Conclusions

With the new deviation measures, we are able to refine the descriptions of uncertainty sets by including distributional asymmetry. This in turn enables us to obtain less conservative solutions while achieving better approximation to the chance constraints. We also use linear decision rules to formulate multiperiod stochastic models with chance constraints as a tractable robust counterpart.

We also demonstrate that using Euclidian norm in the uncertainty set gives the least conservative

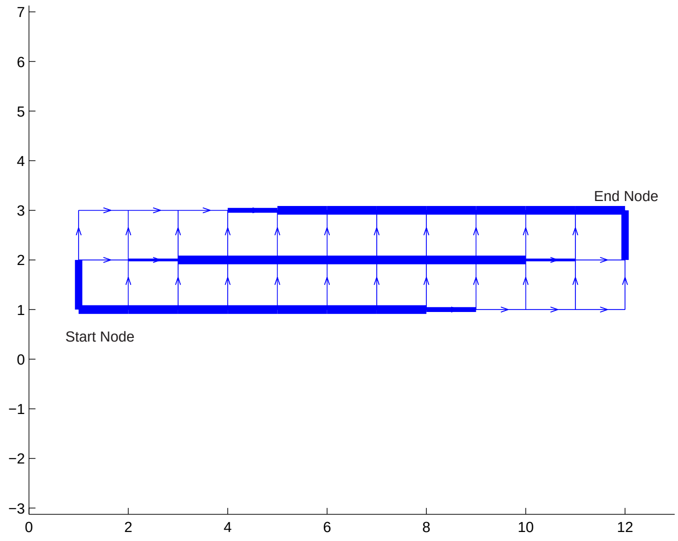

Figure 4: Project management solution for $H = 3$ and $W = 12$.

solution for the same budget of uncertainty. With advances of second order cone solvers, we are able to solve robust models of decent size. When integrality constraints are present, we propose using the $l_1 \cap l_\infty$ norm because the robust counterpart remains linear and the increase in size over the nominal problem is moderate. Hence, we feel that using the robust optimization approach to tackle certain types of stochastic optimization problems can be both practically useful and computationally appealing.

Acknowledgements

We would like to thank the reviewers of the paper for several insightful comments.

A Proof of Proposition 2

Let

$$X = \{\boldsymbol{u} : \|\boldsymbol{u}\| \le \Omega\}$$

and

$$Y = \left\{\boldsymbol{u} : \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, \boldsymbol{u} = \boldsymbol{v} - \boldsymbol{w}, \|\boldsymbol{v} + \boldsymbol{w}\| \le \Omega, \boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0}\right\}.$$

It suffices to show that $X = Y$. For every $\boldsymbol{u} \in X$, let

$$(v_j, w_j) = \begin{cases} (u_j, 0) & \text{if } u_j \ge 0 \\ (0, -u_j) & \text{if } u_j < 0 \end{cases}$$

Clearly, $\boldsymbol{v}, \boldsymbol{w} \geq \mathbf{0}$ and $v_j + w_j = |u_j|$ for all $j = 1, \dots, N$. Since the norm is regular, we have

$$\|\boldsymbol{v} + \boldsymbol{w}\| = \||\boldsymbol{u}|\| = \|\boldsymbol{u}\| \leq \Omega,$$

hence, $X \subseteq Y$. Conversely, suppose $\boldsymbol{u} \in Y$ and hence, $u_j = v_j - w_j$, for all $j = 1, \dots, N$. Clearly

$$|u_j| \leq v_j + w_j.$$

Therefore, since the norm is regular, we have

$$\|\boldsymbol{u}\| = \||\boldsymbol{u}|\| \leq \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega,$$

hence $Y \subseteq X$. ■

B Proof of Proposition 3

Observe that

$$\begin{aligned} \Omega\|\boldsymbol{t}\|^* = \max \quad & \sum_{j=1}^{N} \max\{a_j, b_j, 0\} r_j \\ \text{s.t.} \quad & \|\boldsymbol{r}\| \leq \Omega. \end{aligned} \tag{36}$$

Suppose $\boldsymbol{r}^*$ is an optimal solution to (36). For all $j \in \{1, \dots, N\}$, let

$$\begin{aligned} & v_j = w_j = 0 && \text{if } \max\{a_j, b_j\} \leq 0 \\ & v_j = |r_j^*|, w_j = 0 && \text{if } a_j \geq b_j, a_j > 0 \\ & w_j = |r_j^*|, v_j = 0 && \text{if } b_j > a_j, b_j > 0. \end{aligned}$$

Observe that $a_j v_j + b_j w_j \geq \max\{a_j, b_j, 0\} r_j^*$ and $w_j + v_j \leq |r_j^*|\ \forall j \in \{1, \dots, N\}$. From Proposition 1(c) we have $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{r}^*\| \leq \Omega$, and thus $\boldsymbol{v}, \boldsymbol{w}$ are feasible in Problem (9), leading to

$$z^* \geq \sum_{j=1}^{N} (a_j v_j + b_j w_j) \geq \sum_{j=1}^{N} \max\{a_j, b_j, 0\} r_j^* = \Omega\|\boldsymbol{t}\|^*$$

Conversely, let $\boldsymbol{v}^*, \boldsymbol{w}^*$ be an optimal solution to Problem (9). Let $\boldsymbol{r} = \boldsymbol{v}^* + \boldsymbol{w}^*$. Clearly $\|\boldsymbol{r}\| \leq \Omega$ and observe that

$$r_j \max\{a_j, b_j, 0\} \geq a_j v_j^* + b_j w_j^* \quad \forall j \in \{1, \dots, N\}.$$

Therefore, we have

$$\Omega\|\boldsymbol{t}\|^* \geq \sum_{j=1}^{N} \max\{a_j, b_j, 0\} r_j \geq \sum_{j=1}^{N} (a_j v_j^* + b_j w_j^*) = z^* .$$

■

Regular Norms	$\|\boldsymbol{z}\|$	$\|\boldsymbol{u}\|^* \leq h$	References
l_2	$\|\boldsymbol{z}\|_2$	$\|\boldsymbol{u}\|_2 \leq h$	[4]
Scaled l_1	$\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1$	$\sqrt{N}u_j \leq h, \forall j \in \{1,\ldots,N\}$	[8]
l_∞	$\|\boldsymbol{z}\|_\infty$	$\sum_{j=1}^N u_j \leq h$	[8]
Scaled l_p, $p \geq 1$	$\min\{N^{\frac{1}{2}-\frac{1}{p}}, 1\}\|\boldsymbol{z}\|_p$	$\max\{N^{\frac{1}{p}-\frac{1}{2}}, 1\}\left(\sum_{j=1}^N u_j^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} \leq h$	[8]
$l_1 \cap l_\infty$ norm	$\max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\}$	$\sqrt{N}p + \sum_{j=1}^N s_j \leq h$ $s_j + p \geq u_j, \ \forall j \in \{1,\ldots,N\}$ $p \in \Re_+, \boldsymbol{s} \in \Re_+^N$	[8]

Table 5: Representation of the dual norm for $\boldsymbol{u} \geq \boldsymbol{0}$

C Table of Dual Norm Representation

Table 5 list the common choices of regular norms, the representation of their dual norm inequalities, $\|\boldsymbol{u}\|^* \leq h$ in which $\boldsymbol{u} \geq \boldsymbol{0}$ and the corresponding references.

D Deviation Measures for Asymmetric Distributions

Recently, general deviation measures are introduced and analyzed systematically in Rockafellar et al. [22], in which they define deviation measures on $\mathcal{L}^2(L, \mathcal{F}, Q)$, where L is the sample space, $\mathcal{F}$ is a field of sets in L and Q is a probability measure on $(L, \mathcal{F})$. Among those requirements for a deviation measure, positive homogeneity and subadditivity play a fundamental role. Careful analysis of the forward and backward deviations defined in Section 3 reveals that these properties are also essential for Theorem 3 and Theorem 5, which suggests that it is possible to adapt the deviation measure definition of Rockafellar et al. [22] to our setting.

For our purpose, we focus on a closed convex cone $\mathcal{S}$ of $\mathcal{L}^2(L, \mathcal{F}, Q)$. A deviation measure on the closed convex cone $\mathcal{S}$ is a functional $\mathcal{D} : \mathcal{S} \to \Re_+$ satisfying the following axioms.

(D1) $\mathcal{D}(\tilde{x} + c) = \mathcal{D}(\tilde{x})$ for any random variable $\tilde{x} \in \mathcal{S}$ and constant c.

(D2) $\mathcal{D}(0) = 0$ and $\mathcal{D}(\lambda\tilde{x}) = \lambda\mathcal{D}(\tilde{x})$ for any $\tilde{x} \in \mathcal{S}$ and $\lambda \geq 0$.

(D3) $\mathcal{D}(\tilde{x} + \tilde{y}) \leq \mathcal{D}(\tilde{x}) + \mathcal{D}(\tilde{y})$ for any $\tilde{x} \in \mathcal{S}$ and $\tilde{y} \in \mathcal{S}$.

(D4) $\mathcal{D}(\tilde{x}) > 0$ for any nonconstant $\tilde{x} \in \mathcal{S}$, whereas $\mathcal{D}(c) = 0$ for any constant c.

Standard deviation, semideviations and conditional value at risk are some examples satisfying the above deviation measure definition. For more details, see Rockafellar et al. [22]. Since in our model, uncertain data are affinely dependent on a set of independent random variables, (D3) can be relaxed for our purpose. In particular, we assume

(D3') $\mathcal{D}(\tilde{x}+\tilde{y}) \leq \mathcal{D}(\tilde{x}) + \mathcal{D}(\tilde{y})$ for any independent $\tilde{x} \in \mathcal{S}$ and $\tilde{y} \in \mathcal{S}$.

In addition, to derive meaningful probability bounds against constraint violation, we prefer a deviation measure with one of the following properties.

(D5) For any random variables $\tilde{x} \in \mathcal{S}$ with mean zero, $P(\tilde{x} > \Omega\mathcal{D}(\tilde{x})) \leq f_{\mathcal{S}}(\Omega)$ for some function $f_{\mathcal{S}}$ depending only on $\mathcal{S}$ such that $f_{\mathcal{S}}(\Omega) \to 0$ as $\Omega \to \infty$.

(D5') For any random variables $\tilde{x} \in \mathcal{S}$ with mean zero, $P(\tilde{x} < -\Omega\mathcal{D}(\tilde{x})) \leq f_{\mathcal{S}}(\Omega)$ for some function $f_{\mathcal{S}}$ depending only on $\mathcal{S}$ such that $f_{\mathcal{S}}(\Omega) \to 0$ as $\Omega \to \infty$.

Assumption (D5) is associated with the upside or forward deviation while Assumption (D5') is associated with the downside or backward deviation. Of particular interest are functions with exponential decay rate, i.e. $f_{\mathcal{S}}(\Omega) = O(e^{-\Omega})$ or $f_{\mathcal{S}}(\Omega) = O(e^{-\Omega^2/2})$.

Notice that from Theorem 3, the forward (backward) deviation defined in Section 3 satisfies Assumptions (D1), (D2), (D3'), (D4) and (D5) (or (D5')) with $f_{\mathcal{S}}(\Omega) = e^{-\Omega^2/2}$. In fact, Theorem 3 (b) is even stronger than (D3'). However, as we already pointed out, the forward and backward deviations may not exist for some unbounded distributions, such as exponential distributions. We now introduce different forward and backward deviation measures, which are defined for more general distributions.

For a given $\kappa > 1$, define

$$\mathcal{P}_\kappa(\tilde{z}) = \left\{\alpha \;:\; \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(\frac{1}{\alpha}\right) \leq \kappa\right\}, \tag{37}$$

and

$$\mathcal{Q}_\kappa(\tilde{z}) = \left\{\alpha \;:\; \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(-\frac{1}{\alpha}\right) \leq \kappa\right\}. \tag{38}$$

Let

$$\hat{\mathcal{S}}_+ = \{\tilde{z} : M_{\tilde{z}}(s) < \infty \text{ for some } s > 0\},$$

and

$$\hat{\mathcal{S}}_- = \{\tilde{z} : M_{\tilde{z}}(s) < \infty \text{ for some } s < 0\},$$

It is clear that $\mathcal{P}_\kappa(\tilde{z})$ is nonempty if and only if $\tilde{z} \in \hat{\mathcal{S}}_+$ and $\mathcal{Q}_\kappa(\tilde{z})$ is nonempty if and only if $\tilde{z} \in \hat{\mathcal{S}}_-$. Note that an exponentially distributed random variable has finite derivation measure p_κ while not p^*. We now show that a result similar to Theorem 3 holds for $\mathcal{P}_\kappa(\tilde{z})$ and $\mathcal{Q}_\kappa(\tilde{z})$.

Theorem 7 *Let $\tilde{x}$ and $\tilde{y}$ be two independent random variables with zero means such that $p_{\tilde{x}} \in \mathcal{P}_\kappa(\tilde{x})$, $q_{\tilde{x}} \in \mathcal{Q}_\kappa(\tilde{x})$, $p_{\tilde{y}} \in \mathcal{P}_\kappa(\tilde{y})$ and $q_{\tilde{y}} \in \mathcal{Q}_\kappa(\tilde{y})$.*
(a) *If $\tilde{z} = a\tilde{x}$, then*

$$(p_{\tilde{z}}, q_{\tilde{z}}) = \begin{cases} (ap_{\tilde{x}}, aq_{\tilde{x}}) & \text{if } a \geq 0 \\ (-aq_{\tilde{x}}, -ap_{\tilde{x}}) & \text{otherwsie} \end{cases}$$

satisfy $p_{\tilde{z}} \in \mathcal{P}_\kappa(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}_\kappa(\tilde{z})$. In other words, $p_{\tilde{z}} = \max\{ap_{\tilde{x}}, -aq_x\}$ and $q_{\tilde{z}} = \max\{aq_{\tilde{x}}, -ap_x\}$.
(b) *If $\tilde{z} = \tilde{x} + \tilde{y}$, then $(p_{\tilde{z}}, q_{\tilde{z}}) = (p_{\tilde{x}} + p_{\tilde{y}}, q_{\tilde{x}} + q_{\tilde{y}})$ satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$.*

Proof : (a) This result directly follows from the definition.
(b) We only show the result for the forward deviation measure. The result for the backward deviation measure follows from a similar argument. Notice that

$$\begin{aligned} M_{\tilde{z}}\left(\frac{1}{p_{\tilde{z}}}\right) &= M_{\tilde{x}}\left(\frac{1}{p_{\tilde{z}}}\right) M_{\tilde{y}}\left(\frac{1}{p_{\tilde{z}}}\right) \text{ [since } \tilde{x} \text{ and } \tilde{y} \text{ are independent]} \\ &= M_{\tilde{x}}\left(\frac{1}{p_{\tilde{x}}}\frac{p_{\tilde{x}}}{p_{\tilde{z}}}\right) M_{\tilde{y}}\left(\frac{1}{p_{\tilde{y}}}\frac{p_{\tilde{y}}}{p_{\tilde{z}}}\right) \\ &\leq \left(M_{\tilde{x}}\left(\frac{1}{p_{\tilde{x}}}\right)\right)^{\frac{p_{\tilde{x}}}{p_{\tilde{z}}}} \left(M_{\tilde{y}}\left(\frac{1}{p_{\tilde{y}}}\right)\right)^{\frac{p_{\tilde{y}}}{p_{\tilde{z}}}} \text{ [Hölder's Inequality]} \\ &\leq \kappa. \end{aligned}$$

Thus, $p_{\tilde{z}} = p_{\tilde{x}} + p_{\tilde{y}} \in \mathcal{P}(\tilde{z})$. ■

Theorem 8 *Let $p_\kappa(\tilde{z}) = \inf \mathcal{P}_\kappa(\tilde{z})$ and $q_\kappa(\tilde{z}) = \inf \mathcal{Q}_\kappa(\tilde{z})$. Then p_κ satisfies (D1), (D2), (D3'), (D4) and (D5) with $f_{\hat{\mathcal{S}}_+}(\Omega) = \kappa \exp(-\Omega)$ and q_κ satisfies (D1), (D2), (D3'), (D4) and (D5) with $f_{\hat{\mathcal{S}}_-}(\Omega) = \kappa \exp(-\Omega)$.*

Proof : We show the claim for p_κ only. The result for q_κ follows from a similar argument. It is clear that (D1), (D2) and (D4) hold for p_κ and (D3') follows from Theorem 7 **(b)**.

It remains to show (D5). Notice that if $\tilde{x} = 0$, (D5) follows trivially for all $\Omega, p_{\tilde{x}} \geq 0$. Otherwise, observe that for all $\Omega \geq 0$,

$$\mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) = \mathrm{P}\left(\frac{\tilde{x}}{p_{\tilde{x}}} > \Omega\right) \leq \frac{\mathrm{E}\left(\exp\left(\frac{\tilde{x}}{p_{\tilde{x}}}\right)\right)}{\exp(\Omega)} \leq \kappa \exp(-\Omega),$$

where the first inequality follows from Chebyshev's inequality and the second inequality follows from the definition of $p_{\tilde{x}}$. ■

Parallel results to Theorem 4 can also be obtained. We would like to point out that all our results hold for the deviation measures p_κ and q_κ with corresponding modifications. As we mentioned, p_κ and q_κ can be applied for certain class of random variables, for instance exponential random variables, in which p^* and q^* are not well defined. Of course, since Theorem 7 part (b) is weaker than Theorem 3 part (b), one can expect that the probability bound derived based on p^* and q^* is stronger than that based on p_κ and q_κ, if both deviation measures are well defined. Thus, for simplicity of presentation, we focus on the deviation measures p^* and q^* in this paper.

References

[1] Atamtürk, A. (2003): Strong formulations of robust mixed 0-1 programming. *Working Paper*, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

[2] Ben-Tal, A., Nemirovski, A. (1998): Robust convex optimization, *Math. Oper. Res.*, 23, 769-805.

[3] Ben-Tal, A., Nemirovski, A. (1999): Robust solutions to uncertain programs, *Oper. Res. Let.*, 25, 1-13.

[4] Ben-Tal, A., Nemirovski, A. (2000): Robust solutions of Linear Programming problems contaminated with uncertain data, *Math. Progr.*, 88, 411-424.

[5] Ben-Tal, A., A. Goryashko, E. Guslitzer and A. Nemirovski. (2002): Adjusting Robust Solutions of Uncertain Linear Programs. *Working Paper*, Technion.

[6] Ben-Tal, A., B. Golany, A Nemirovski., J. Vial. (2003): Supplier-Retailer Flexible Commitments Contracts: A Robust Optimization Approach. *Working Paper*, Technion.

[7] Bertsekas, D. (1999): Nonlinear Programming, *Athena Scientific*, Belmont, MA.

[8] Bertsimas, D., Pachamanova, D., Sim, M. (2003): Robust Linear Optimization under General Norms, to appear in *Oper. Res. Let.*

[9] Bertsimas, D., Sim, M. (2003): Robust Discrete Optimization and Network Flows, *Math. Progr.*, 98, 49-71.

[10] Bertsimas, D., Sim, M. (2004): Price of Robustness, *Oper. Res.*, 52(1), 35-53.

[11] Bertsimas, D., Sim, M. (2004): Robust Conic Opimization, *Working Paper*, MIT.

[12] Bertsimas, D., Sim, M. (2004): Robust Discrete Opimization and Downside Risk Measures, *Working Paper*, National University of Singapore.

[13] Bertsimas, D., A. Thiele. 2003. A Robust Optimization Approach to Supply Chain Management. Working Paper, MIT.

[14] Birge, J. R., Louveaux, F. (1997): Introduction to Stochastic Programming. *Springer*, New York.

[15] Charnes, A. and Cooper, W.W. (1959): Uncertain Convex Programs: Randomize solutions and confidence level, *Management Science*, 6, 73-79.

[16] El-Ghaoui, Lebret, H. (1997): Robust solutions to least-square problems to uncertain data matrices, *SIAM J. Matrix Anal. Appl.*, 18, 1035-1064.

[17] El-Ghaoui, L., Oustry, F., Lebret, H. (1998): Robust solutions to uncertain semidefinite programs, *SIAM J. Optim.*, 9, 33-52.

[18] Goldfarb, D. and Iyengar, G.(2003): Robust quadratically constrained programs, *Math. Prog. Ser. B*, 97(3), 495-515

[19] Garstka S. J. and Wets R.J-B. (1974): On decision rules in stochastic programming, *Math. Progr.*, 117-143.

[20] Hagstrom, J.N. (1988): Computational complexity of PERT problems. *Networks*, 18:139-147, 1988.

[21] Kibzun, A. and Kan, Y. (1996): Stochastic Programming Problems with Probability and Quantile Functions, *Wiley*

[22] Rockafellar, R.T., Uryasev S. and M. Zabarankin. (2004) Generalized Deviations In Risk Analysis. Research Report 2004-4, ISE Dept., University of Florida.

[23] Shapiro, A. and Nemirovski, A. (2004): On Complexity of stochastic programming Problems, *Working Paper*, Georgia Tech, Atlanta.

[24] Soyster, A.L. (1973): Convex programming with set-inclusive constraints and applications to inexact linear programming, *Oper. Res.*, 21, 1154-1157.

[25] Toh, K.C. , Todd, M.J. , and Tütüncü, R.H. (1999): SDPT3- A Matlab Software package Programming, Optimization Methods and Software, 11, pp. 545-581.