LINEAR MATRIX INEQUALITIES WITH STOCHASTICALLY DEPENDENT PERTURBATIONS AND APPLICATIONS TO CHANCE–CONSTRAINED SEMIDEFINITE OPTIMIZATION*

SIN–SHUEN CHEUNG†, ANTHONY MAN–CHO SO‡, AND KUNCHENG WANG§

Abstract. The wide applicability of chance–constrained programming, together with advances in convex optimization and probability theory, has created a surge of interest in finding efficient methods for processing chance constraints in recent years. One of the successes is the development of so–called safe tractable approximations of chance–constrained programs, where a chance constraint is replaced by a deterministic and efficiently computable inner approximation. Currently, such approach applies mainly to chance–constrained linear inequalities, in which the data perturbations are either independent or define a known covariance matrix. However, its applicability to chance–constrained conic inequalities with dependent perturbations—which arises in finance, control and signal processing applications—remains largely unexplored. In this paper, we develop safe tractable approximations of chance–constrained affinely perturbed linear matrix inequalities, in which the perturbations are not necessarily independent, and the only information available about the dependence structure is a list of independence relations. To achieve this, we establish new large deviation bounds for sums of dependent matrix–valued random variables, which are of independent interest. A nice feature of our approximations is that they can be expressed as systems of linear matrix inequalities, thus allowing them to be solved easily and efficiently by off–the–shelf solvers. We also provide a numerical illustration of our constructions through a problem in control theory.

Key words. linear matrix inequalities, large deviation bounds, chance–constrained programming, stochastic programming, semidefinite programming

AMS subject classifications. 15A52, 60F10, 90C15, 90C22, 90C59

1. Introduction.

It has long been recognized that traditional optimization models, in which data are assumed to be precisely known, can be inadequate in the presence of data uncertainties. For instance, the notion of a feasible solution may no longer be well defined, as a solution that is optimal with respect to one particular realization of the uncertain data can be sub–optimal or even infeasible with respect to another. Therefore, much effort has been made to develop models that can incorporate data uncertainties in the optimization process. One idea that was first proposed by Charnes et al. [19, 18] is to treat the uncertain data as a random vector $\xi \in \mathbb{R}^m$ with known probability distribution and find a solution that is feasible with respect to *most* realizations of ξ. In the context of conic optimization, such idea leads to so–called *probabilistic* or *chance constraints* of the form

$$\inf_{\mathbb{P}\in\mathscr{P}} \Pr_{\xi\sim\mathbb{P}} (F(x,\xi) \in K) \geq 1-\epsilon, \tag{1.1}$$

where $\mathscr{P}$ is a set of distributions that are consistent with our a priori knowledge of the uncertain data vector ξ, $\xi \sim \mathbb{P}$ means that ξ is distributed according to $\mathbb{P}$, $x \in \mathbb{R}^n$

*This research was supported in part by Project #MMT–p2–09 of the Shun Hing Institute of Advanced Engineering at The Chinese University of Hong Kong, and in part by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) Project CUHK 416012.

†School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853, USA. E–mail: sc2392@cornell.edu.

‡Department of Systems Engineering and Engineering Management, and, by courtesy, the Department of Computer Science and Engineering and the CUHK–BGI Innovation Institute of Trans–omics, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E–mail: manchoso@se.cuhk.edu.hk.

§Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E–mail: kcwang@se.cuhk.edu.hk.

is the decision vector, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$ is a random vector–valued function, $K \subset \mathbb{R}^l$ is a closed convex pointed cone, and $\epsilon \in [0, 1)$ is a tolerance parameter specified by the modeler. Note that by including the set $\mathscr{P}$ in (1.1), it is possible to model the situation where the distribution of ξ is only *partially* known. In particular, the constraint (1.1) offers certain degree of robustness against errors in the specification of the distribution of ξ.

Although the use of chance constraints is perfectly natural when dealing with data uncertainties, it also creates significant computational difficulties. Indeed, even for simple classes of distributions, the set of solutions satisfying (1.1) can be non–convex. Moreover, the probability on the left–hand side of (1.1) is often difficult to compute accurately. Thus, a fundamental problem is to derive efficiently computable (approximate) descriptions of the feasible set defined by (1.1). In this paper, we study the said problem under the assumption that F is affine in both x and ξ, and K is the positive semidefinite cone. In other words, we assume that F is of the form

$$F(x, \xi) = A_0(x) + \sum_{i=1}^{m} \xi_i A_i(x), \tag{1.2}$$

where $A_0, A_1, \ldots, A_m : \mathbb{R}^n \to \mathcal{S}^d$ are affine functions that take values in the space $\mathcal{S}^d$ of $d \times d$ real symmetric matrices. The functional form (1.2) models the situation where the nominal value $A_0(x)$ is randomly perturbed along the directions $A_1(x), \ldots, A_m(x)$, and the resulting constraint (1.1) encapsulates chance–constrained linear, second–order cone and semidefinite programming problems. Note that when $\epsilon = 0$, (1.1) essentially reduces to a robust feasibility problem. In this case, only the support of the distribution (known as the *uncertainty set*) is relevant, and the efficient representability of (1.1) is known for various classes of uncertainty sets; see the book [3] and the references therein. In the sequel, we shall consider $\epsilon \in (0, 1)$, so that it is possible to take advantage of other properties of the distributions in $\mathscr{P}$. We are particularly interested in the case where there is some dependence among the random variables $\xi_1, \ldots, \xi_m$, but the only information available about the dependence structure is a list of independence relations, i.e., a list specifying which subsets of random variables are mutually independent. In particular, we do not assume precise knowledge of the covariance matrix. Such a setting is motivated by applications in finance, control and signal processing, and, to the best of our knowledge, has not been previously addressed—even for the case where $d = 1$. Before we state our results and give an overview of our techniques, let us review some related work in the literature.

1.1. Related Work. For $d = 1$, it was known very early on that if ξ is Gaussian with given mean vector and covariance matrix and $\epsilon \le 1/2$, then (1.1) can be reformulated as a conic quadratic inequality; see, e.g., [35, 60]. In fact, the same result holds when ξ has a radial distribution, of which the Gaussian distribution is a special case [15]. On the other hand, if ξ has a symmetric log–concave distribution and $\epsilon \le 1/2$, then one can show that the feasible set defined by (1.1) is convex [36, 37]. However, whether this convex set has an efficiently computable representation will depend on specific properties of the given distribution.

Although the aforementioned results provide exact convex reformulations of (1.1), they are mainly concerned with the case where $\mathscr{P}$ is a singleton (i.e., the distribution is completely known). Recently, El Ghaoui et al. [24] and Calafiore and El Ghaoui [15] considered the case where $\mathscr{P}$ is the set of distributions that have the same given mean vector and covariance matrix. They showed that with this choice of $\mathscr{P}$, the chance constraint (1.1) can be reformulated as a conic quadratic inequality. Unfortunately, such exact and efficient reformulations are often not possible for other classes of distributions. To circumvent this problem, one can construct a so–called *safe tractable approximation* of (1.1), i.e., a system of efficiently computable constraints whose feasible solutions can be efficiently converted into feasible solutions to (1.1). In general, there are many ways to construct such approximation. For instance, one can derive an analytic upper bound on the violation probability $\Pr_{\xi\sim\mathbb{P}}(F(x,\xi) \prec 0)$. This was first pursued by Pintér [48], who proposed to bound the violation probability using Chernoff–Hoeffding–type inequalities. As later observed by various researchers [4, 8] (see also [3, Chapter 4]), the safe tractable approximations obtained from Pintér's approach are just robust counterparts of the affinely perturbed linear constraint with suitably defined uncertainty sets. Such connection allows one to utilize powerful results in robust optimization to construct other safe tractable approximations of (1.1); see, e.g., [8, 9, 22].

Alternatively, one can use a generating function of the random variable $F(x,\xi)$ to bound the violation probability. A natural bound that results from this approach is the conditional value–at–risk (CVaR) functional applied to $F(x,\xi)$, which can be shown to give the tightest convex conservative approximation of the violation probability; see [26, Remark 4.51 and Theorem 4.61] and [45]. Furthermore, there is a close connection between the CVaR functional and uncertainty sets in robust optimization [7, 42]. However, it is generally difficult to evaluate the CVaR functional accurately. Thus, many efficiently computable bounds on the CVaR functional have been developed [45, 20, 21], and each of them yields a safe tractable approximation of (1.1).

For $d > 1$, most previous work focused on the case where the matrices $A_0(x), A_1(x), \ldots, A_m(x)$ are diagonal for all $x \in \mathbb{R}^n$, which corresponds to what is commonly known as *joint chance constraints*. In an early paper by Miller and Wagner [41], joint chance constraints with *random right–hand side*[1] and independent random perturbations $\xi_1, \ldots, \xi_m$ were considered, and a mathematical and algorithmic treatment was provided. Subsequently, Prékopa initiated a systematic study of joint chance constraints that involve random right–hand sides and more general types of random perturbations. By developing a theory of multivariate log–concave measures, Prékopa showed that when ξ has a log–concave distribution, a large class of joint chance constraints with random right–hand side can be reformulated as deterministic convex constraints [49, 50] (see also [52, 55] and the references therein for related and recent results). This opens up the possibility of using modern convex optimization techniques to efficiently process those chance constraints. Later, similar convexity results were obtained for certain classes of joint chance constraints, in which the randomness is not necessarily on the right–hand side; see, e.g., [51, 32, 53]. It should be noted, however, that exact convex reformulations of joint chance constraints may not be possible in general. Recently, there has been some effort to construct safe tractable approximations of general joint chance constraints using various bounds on the CVaR functional [21, 65]. An upshot of this approach is that it can handle the case where the distribution of ξ is only partially specified, say, e.g., by the mean vector and covariance matrix. By contrast, all the aforementioned exact reformulations assume that complete knowledge of the distribution is available.

As we move beyond the case of joint chance constraints and consider the case

[1]That is, the matrices $A_1(x), \ldots, A_m(x)$ do not depend on x and take the form $A_i(x) = -\mathrm{diag}(e_i)$, where e_i is the i–th basis vector for $i = 1, \ldots, d$.

where the symmetric matrices $A_0(x), A_1(x), \ldots, A_m(x)$ are arbitrary, results are much scarcer. Indeed, the only results that we are aware of are those by Nemirovski [43, 44], Bertsimas and Sim [9], Ben–Tal and Nemirovski [6], and So [56]. These authors showed that certain linear matrix inequality can serve as a safe tractable approximation of (1.1) when the random variables $\xi_1, \ldots, \xi_m$ are independent and have light tails.

It should be pointed out that in all the aforementioned work, guarantees on the violation probability of a solution are established either by utilizing precise distributional information (such as the density function or the covariance matrix of ξ), or by assuming independence of the random variables $\xi_1, \ldots, \xi_m$. Thus, they are not directly applicable to our setting (i.e., when the dependence structure of $\xi_1, \ldots, \xi_m$ is revealed only through a list of independence relations). Although techniques such as the generating function method [3, Chapter 4.5] and moment uncertainty sets method [23, 56] have been developed to tackle dependent perturbations with limited distributional information, they are still not sufficient for our purposes. First, they apply only to the case where $d = 1$. Secondly, the tractability of the safe approximations derived using the generating function method depends on our ability to evaluate certain convex functions accurately and efficiently, while the moment uncertainty sets proposed in [23, 56] can be difficult to define when only a list of independence relations is available. Of course, one can also use Monte Carlo sampling to tackle the general chance constraint (1.1). However, in order to assert the feasibility of the solution obtained by this method with high confidence, the number of samples required is on the order of $1/\epsilon$ [13, 14, 25, 39, 17, 16], which can render the computation prohibitively expensive. The above issues thus motivate us to explore other approaches for constructing safe tractable approximations of chance constraints with dependent perturbations.

1.2. Our Contributions. In this paper, we establish upper bounds on the violation probability

$$\Pr_{\xi}\left(A_0(x) + \sum_{i=1}^{m} \xi_i A_i(x) \not\preceq \mathbf{0} \right) \tag{1.3}$$

by utilizing a list of independence relations of and some additional information (such as support or tail behavior) about the collection of real–valued mean–zero random variables $\xi = (\xi_1, \ldots, \xi_m)$. We then show that those upper bounds can be expressed as systems of linear matrix inequalities in the variable $x \in \mathbb{R}^n$ and hence are efficiently computable. As an immediate corollary, we obtain safe tractable approximations of the chance constraint (1.1) for the setting described above. Our results generalize those in [44, 6, 56], which only deal with the case where $\xi_1, \ldots, \xi_m$ are independent.

The main idea of our approach is to first split the sum $\sum_{i=1}^{m} \xi_i A_i(x)$ into its independent parts using the given list of independence relations, i.e., we write

$$\sum_{i=1}^{m} \xi_i A_i(x) = \sum_{j} w_j \sum_{i \in \mathcal{A}_j} \xi_i A_i(x) \tag{1.4}$$

for some appropriate sets $\{\mathcal{A}_j\}_j$ and positive weights $\{w_j\}_j$, so that for each j, the random variables in $\{\xi_i : i \in \mathcal{A}_j\}$ are mutually independent. The upshot of (1.4) is that for each j, the term $\sum_{i \in \mathcal{A}_j} \xi_i A_i(x)$ is a sum of independent random variables, which makes it more amenable to analysis. Such an idea was previously used by Janson [34] to establish large deviation bounds for sums of dependent, *real–valued* and bounded random variables. These in turn yield upper bounds on the violation probability (1.3) for the case where $d = 1$. In this paper, we extend Janson's techniques to obtain large deviation bounds for sums of dependent *matrix–valued* random variables. Our proof relies on various properties of the matrix exponential, as well as some recently developed tools for handling matrix–valued random variables. We believe that our extension of Janson's result to the matrix case is of independent interest.

To demonstrate the power of our approach, we use it to construct safe tractable approximations of chance–constrained *quadratically perturbed* linear matrix inequalities, i.e., chance constraints of the form

$$\Pr_{\zeta}\left(A_0(x)+\sum_{i=1}^{m}\zeta_i A_i(x)+\sum_{1\le j\le k\le m}\zeta_j\zeta_k B_{jk}(x)\preceq \mathbf{0}\right)\ge 1-\epsilon, \tag{1.5}$$

where $A_i, B_{jk} : \mathbb{R}^n \to \mathcal{S}^d$ are affine functions for $0 \le i \le m$ and $1 \le j \le k \le m$, and $\zeta_1, \ldots, \zeta_m$ are i.i.d. real–valued mean–zero random variables with various tail behavior. Such a chance constraint arises in many areas, such as finance [64], control [57, 54] and signal processing [38, 61, 62]. However, it has not been investigated systematically in the literature. Indeed, to the best of our knowledge, the only results concerning (1.5) are those by Ben–Tal et al. [3, Chapter 4.5] and Zymler et al. [64], which apply only to the case where $d = 1$. The former requires evaluation of certain convex functions, which could be computationally expensive; while the latter requires precise knowledge of the covariance matrix. By contrast, our approach only requires a list of independence relations, and the resulting safe tractable approximations can be formulated as systems of linear matrix inequalities. As such, they can be efficiently solved by standard packages. Moreover, by specializing our results to the case where $A_i(x)$ and $B_{jk}(x)$ are diagonal for all $x \in \mathbb{R}^n$, $0 \le i \le m$ and $1 \le j \le k \le m$, we obtain safe tractable approximations of joint quadratically perturbed scalar chance constraints.

1.3. Outline of the Paper. The paper is organized as follows. In Section 2, we introduce some terminologies and give an overview of our approach. Then, we prove large deviation bounds for sums of dependent matrix–valued random variables in Section 3. In Section 4, we show how those bounds can be used to construct safe tractable approximations of chance–constrained linear matrix inequalities with dependent perturbations. In Section 5, we discuss the conservatism of the proposed safe tractable approximations and introduce a numerical procedure to iteratively relax them while retaining their safety and tractability. In Section 6, we report some numerical results obtained when applying our constructions to a problem in control theory. Finally, we end with some closing remarks in Section 7.

2. Preliminaries. Let $\xi_1, \ldots, \xi_m$ be real–valued mean–zero random variables, and let $A_0, A_1, \ldots, A_m : \mathbb{R}^n \to \mathcal{S}^d$ be deterministic affine functions taking values in the space $\mathcal{S}^d$ of $d \times d$ real symmetric matrices. As mentioned in the Introduction, a key step in constructing safe tractable approximations of the chance constraint (1.3) is to understand the behavior of the matrix–valued random variable $S(x) = \sum_{i=1}^{m} \xi_i A_i(x)$. When the random variables $\xi_1, \ldots, \xi_m$ are mutually independent, the behavior of $S(x)$ is relatively well understood; see, e.g., [56, 59] and the references therein. However, not much is known when there is some dependence among $\xi_1, \ldots, \xi_m$. To handle this case, one idea is to decompose $S \equiv S(x)$ using the notion of exact proper fractional cover. Specifically, let $\mathcal{A} = \{1, \ldots, m\}$. We say that a collection of pairs $\{(\mathcal{A}_j, w_j)\}_j$, where $\mathcal{A}_j \subset \mathcal{A}$ and $w_j > 0$ for all j, forms an *exact proper fractional cover* of $\mathcal{A}$ if

1. for each $i \in \mathcal{A}$, $\sum_{j:i\in\mathcal{A}_j} w_j = 1$,

2. for each j, the random variables in $\{\xi_i : i \in \mathcal{A}_j\}$ are mutually independent.

Note that such a cover always exists, as we can take $\mathcal{A}_j = \{j\}$ and $w_j = 1$ for $j = 1, \ldots, m$. Moreover, if the random variables $\xi_1, \ldots, \xi_m$ are mutually independent, then $\{(\mathcal{A}, 1)\}$ is an exact proper fractional cover of $\mathcal{A}$. Now, suppose that $\{(\mathcal{A}_j, w_j)\}_j$ is an exact proper fractional cover of $\mathcal{A}$. Then, we have

$$S = \sum_{i=1}^{m} \xi_i A_i = \sum_{i=1}^{m} \sum_{j: i \in \mathcal{A}_j} w_j \xi_i A_i = \sum_j w_j \sum_{i \in \mathcal{A}_j} \xi_i A_i. \tag{2.1}$$

In other words, every exact proper fractional cover of $\mathcal{A}$ induces a decomposition of S into its independent parts. The upshot of the decomposition (2.1) is that it can be used to deduce the behavior of S. For the case where $d = 1$ (i.e., $A_1, \ldots, A_m$ are scalars), this has already been observed by Janson [34]. To fix ideas and motivate our results, let us briefly review Janson's argument. The goal is to provide an upper bound on the probability $\Pr(S \geq t)$ for any $t > 0$. Towards that end, consider a collection $\{p_j\}_j$ of positive numbers, each corresponds to a pair in the exact proper fractional cover of $\mathcal{A}$, such that $\sum_j p_j = 1$. For any $u \in \mathbb{R}$, we compute

$$\begin{aligned}
\mathbb{E}\left[\exp(uS)\right] &= \mathbb{E}\left[\exp\left(\sum_j p_j \frac{uw_j}{p_j} \sum_{i \in \mathcal{A}_j} \xi_i A_i\right)\right] \\
&\leq \sum_j p_j \cdot \mathbb{E}\left[\exp\left(\frac{uw_j}{p_j} \sum_{i \in \mathcal{A}_j} \xi_i A_i\right)\right] && (2.2) \\
&= \sum_j p_j \cdot \prod_{i \in \mathcal{A}_j} \mathbb{E}\left[\exp\left(\frac{uw_j}{p_j} \xi_i A_i\right)\right], && (2.3)
\end{aligned}$$

where (2.2) follows from Jensen's inequality, and (2.3) follows from the independence of the random variables in $\{\xi_i : i \in \mathcal{A}_j\}$. Suppose now that the moment generating functions of the random variables $\xi_1, \ldots, \xi_m$ have *subgaussian–type growth*, i.e., there exist constants $\{v_i\}_i$ satisfying

$$\mathbb{E}\left[\exp(\theta \xi_i)\right] \leq \exp\left(\theta^2 v_i^2\right) \tag{2.4}$$

for all $\theta \in \mathbb{R}$ and $i = 1, \ldots, m$. Then, we deduce from (2.3) that

$$\mathbb{E}\left[\exp(uS)\right] \leq \sum_j p_j \cdot \prod_{i \in \mathcal{A}_j} \mathbb{E}\left[\exp\left(\frac{uw_j}{p_j} \xi_i A_i\right)\right] \leq \sum_j p_j \cdot \prod_{i \in \mathcal{A}_j} \exp\left(\frac{u^2 w_j^2}{p_j^2} v_i^2 A_i^2\right). \tag{2.5}$$

Upon setting

$$c_j = \sum_{i \in \mathcal{A}_j} v_i^2 A_i^2, \quad T = \sum_j w_j c_j^{1/2}, \quad p_j = \frac{w_j c_j^{1/2}}{T}$$

and using (2.5), we obtain

$$\mathbb{E}\left[\exp(uS)\right] \leq \sum_j p_j \exp\left(\frac{u^2 w_j^2 c_j}{p_j^2}\right) = \sum_j p_j \exp\left(u^2 T^2\right) = \exp\left(u^2 T^2\right).$$

The desired upper bound

$$\Pr(S \geq t) \leq \inf_{u>0} \left\{ \exp\left(-ut + u^2T^2\right) \right\} = \exp\left(-\frac{t^2}{4T^2}\right) \quad \text{for } t > 0 \tag{2.6}$$

then follows by an application of Markov's inequality.

Since our ultimate goal is to construct safe tractable approximations of chance–constrained linear matrix inequalities, we need to extend the above result to the case where $d > 1$. Before we proceed, however, some remarks on the above derivation are in order. Observe that the quality of the upper bound (2.6) depends on the tightness of the moment generating function bounds (2.4), as well as on the effectiveness of the exact proper fractional cover we use. While the former depends on the class of random variables under consideration, the latter depends on the choice of weights $\{w_j\}_j$, which suggests that some optimization is possible. Indeed, since the bound (2.6) is tighter when T is smaller, it seems reasonable to consider the following optimization problem:

$$T^* = \min \left\{ \sum_j w_j c_j^{1/2} \right\}, \tag{2.7}$$

where the minimization is taken over all exact proper fractional covers $\{(\mathcal{A}_j, w_j)\}_j$ of $\mathcal{A}$. Unfortunately, there are several obstacles that make Problem (2.7) difficult to solve in general. First, the objective function in (2.7) is nonlinear, as both w_j and c_j depend on the choice of the exact proper fractional cover. Secondly, in our applications, the quantities $\{c_j\}_j$ are functions of the decision vector $x \in \mathbb{R}^n$. Thus, the optimal exact proper fractional cover will depend on x in general. Thirdly, given a list of independence relations, it is often possible to derive additional independence relations from it. However, determining whether a particular independence relation follows from a given list of independence relations is far from trivial; see, e.g., the discussion in [63, Section 13.5]. In view of the above obstacles, we shall consider upper bounds on T^* instead. One way of obtaining such bounds is to find a collection of weighted independent sets in certain dependence graph. Specifically, consider a graph G whose vertex set is $\mathcal{A}$, and that the random variables in $\{\xi_i : i \in \mathcal{A}'\}$ are independent whenever $\mathcal{A}' \subset \mathcal{A}$ is an independent set[2] in G. We call G a *dependence graph* of the random variables $\xi_1, \ldots, \xi_m$. Now, let $\{\mathcal{I}_j\}_j$ be the collection of all possible independent sets in G, w_j be a non–negative weight associated with the independent set $\mathcal{I}_j$, and $\mathcal{F}$ be the polyhedron

$$\mathcal{F} = \left\{ w \geq \mathbf{0} : \sum_{j:i\in\mathcal{I}_j} w_j = 1 \text{ for } i \in \mathcal{A} \right\}.$$

Observe that each vector $\bar{w} = (\bar{w}_j)_j \in \mathcal{F}$ corresponds to an exact proper fractional cover of $\mathcal{A}$, viz. the collection $\{(\mathcal{I}_j, \bar{w}_j) : \bar{w}_j > 0\}$. Moreover, we have $T^* \leq \sum_j \bar{w}_j \left(\sum_{i\in\mathcal{I}_j} v_i^2 A_i^2 \right)^{1/2}$ by definition. Thus, any vector in $\mathcal{F}$ yields an upper bound on T^*. If the dependence graph G is given, then a vector in $\mathcal{F}$ can be found in polynomial time by greedy–coloring the vertices of G; see, e.g., [11, Chapter V.1] for the algorithm.

[2] Recall that an independent set in a graph is a set of pairwise non–adjacent vertices.

Alternatively, one can bound T^* directly by exploiting properties of the dependence graph G. For instance, given any exact proper fractional cover $\{(\mathcal{A}_j, w_j)\}_j$ of $\mathcal{A}$, we can apply the Cauchy–Schwarz inequality to obtain

$$\left(\sum_j w_j c_j^{1/2}\right)^2 \le \left(\sum_j w_j\right)\left(\sum_j w_j c_j\right) = \left(\sum_j w_j\right)\left(\sum_{i=1}^m v_i^2 A_i^2\right),$$

where the last equality follows from the fact that

$$\sum_j w_j c_j = \sum_j w_j \sum_{i\in\mathcal{A}_j} v_i^2 A_i^2 = \sum_{i=1}^m v_i^2 A_i^2 \sum_{j:i\in\mathcal{A}_j} w_j = \sum_{i=1}^m v_i^2 A_i^2.$$

Hence, the optimal value $T_{\rm fc}$ of the linear program

$$\min\left\{\sum_j w_j : w \in \mathcal{F}\right\} \tag{2.8}$$

serves as an upper bound on $(T^*)^2/\sum_{i=1}^m v_i^2 A_i^2$. Although the number $T_{\rm fc}$, which is known as the minimum fractional chromatic number of G, is generally $\mathcal{NP}$–hard to even approximate [40], it can be upper bounded by other easily computable quantities (for instance, it is well known that $T_{\rm fc} \le \Delta + 1$, where Δ is the maximum degree of G, and that $T_{\rm fc} \le \Delta$ if G is a simple connected graph that is not a complete graph or an odd cycle [11, Chapter V]). Such upper bounds on $T_{\rm fc}$ can then be used to bound T^*.

3. Large Deviations of Sums of Dependent Random Matrices. In this section, we prove a large deviation bound similar to (2.6) for the case where $A_1, \ldots, A_m$ are $d \times d$ real symmetric matrices. A natural idea is to extend Janson's argument in the previous section and study the matrix moment generating function $\mathbb{E}[\exp(uS)]$. However, since many properties of the scalar exponential function do not carry over to the matrix exponential function, several difficulties arise. Fortunately, as we shall soon see, those difficulties can be overcome by utilizing some classical results in matrix analysis.

3.1. The Matrix Exponential Function and Its Properties. To begin, let A be an arbitrary $d \times d$ real symmetric matrix. An object that plays a central role in our investigation is the *matrix exponential of* A, which is denoted by $\exp(A)$ and defined via

$$\exp(A) = I + \sum_{i=1}^\infty \frac{A^i}{i!}. \tag{3.1}$$

It is easy to verify that if $\lambda \in \mathbb{R}$ is an eigenvalue of A, then $\exp(\lambda)$ is an eigenvalue of $\exp(A)$. In particular, we see that for any $A \in \mathcal{S}^d$, $\exp(A)$ is positive definite and $\|\exp(A)\| = \exp(\|A\|)$, where $\|A\|$ denotes the spectral norm of A.

In contrast with the scalar case, the identity $\exp(A+B) = \exp(A)\exp(B)$ does not hold for general $A, B \in \mathcal{S}^d$ and is valid only when A and B commute. Moreover, the function $A \mapsto \exp(A)$ is not matrix convex[3]. However, we have the following properties, which are sufficient for our purpose:

[3]We say that a function $f : \mathcal{S}^d \to \mathcal{S}^d$ is matrix convex if $f(\lambda A + (1-\lambda)B) \preceq \lambda f(A) + (1-\lambda) f(B)$ for any $A, B \in \mathcal{S}^d$ and $\lambda \in (0,1)$.

FACT 3.1. (GOLDEN–THOMPSON INEQUALITY [28, 58]) *Let $A, B \in \mathcal{S}^d$ be arbitrary. Then, we have*

$$\mathrm{tr}(\exp(A+B)) \le \mathrm{tr}(\exp(A)\exp(B)),$$

where $\mathrm{tr}(A)$ *denotes the trace of A.*

FACT 3.2. (CONVEXITY OF TRACE EXPONENTIAL) *The function $A \mapsto \mathrm{tr}(\exp(A))$ is convex on $\mathcal{S}^d$.*

The proofs of these properties can be found in [47]. We remark that Fact 3.2 is a special case of Jensen's trace inequality; see, e.g., [30] for further details.

3.2. Main Theorem. To prove large deviation bounds for the sum $S = \sum_{i=1}^m \xi_i A_i$, we need some control on the behavior of the random variables $\xi_1, \ldots, \xi_m$. Towards that end, let us introduce the following definition:

DEFINITION 3.1. *A real–valued mean–zero random variable ξ is said to satisfy* moment growth condition (M) with parameters $(\bar{\theta}, v)$ *if*

$$\mathbb{E}\left[\exp(\theta\xi Q)\right] \preceq \exp\left(\theta^2 v^2 Q^2\right)$$

for all $\theta \in (0, \bar{\theta})$ and $Q \in \mathcal{S}^d$ with $\|Q\| = 1$.

REMARK. Using the power series expansion (3.1), it can be shown that the moment growth condition (M) is satisfied by a wide range of random variables. For instance, if ξ is a standard Gaussian or a Bernoulli random variable, then $\mathbb{E}\left[\exp(\theta\xi Q)\right] \preceq \exp\left(\theta^2 Q^2/2\right)$ for all $\theta \in \mathbb{R}$ and $Q \in \mathcal{S}^d$, i.e., ξ satisfies (M) with parameters $(+\infty, 1/\sqrt{2})$ [46, 59].

We are now ready to state our first main result.

THEOREM 3.2. *Let $\xi_1, \ldots, \xi_m$ be real–valued mean–zero random variables satisfying moment growth condition (M) with parameters $(\bar{\theta}_1, v_1), \ldots, (\bar{\theta}_m, v_m)$, respectively. Suppose that an exact proper fractional cover $\{(\mathcal{A}_j, w_j)\}_j$ of $\mathcal{A} = \{1, \ldots, m\}$ is given. Then, for any $A_1, \ldots, A_m \in \mathcal{S}^d$, we have*

$$\Pr\left(\sum_{i=1}^m \xi_i A_i \not\preceq tI\right) \le \begin{cases} d \cdot \exp\left(-\dfrac{t^2}{4T^2}\right) & \text{for } 0 < t < 2\Gamma T, \\[2ex] d \cdot \exp\left(-\dfrac{\Gamma t}{T} + \Gamma^2\right) & \text{for } t \ge 2\Gamma T, \end{cases}$$

where

$$T = \sum_j w_j c_j^{1/2}, \quad c_j = \left\| \sum_{i \in \mathcal{A}_j} v_i^2 A_i^2 \right\|, \quad \Gamma = \min_{1 \le i \le m} \left\{\bar{\theta}_i v_i\right\}. \tag{3.2}$$

The proof of Theorem 3.2 relies on Facts 3.1 and 3.2, as well as the following two results. The first can be viewed as an extension of the so–called exponential Markov inequality to matrix–valued random variables. The second is a variant of a result by Oliveira [46].

FACT 3.3. (AHLSWEDE–WINTER INEQUALITY [1]) *Let Y be a random $d \times d$ real symmetric matrix. Then, for any $B \in \mathcal{S}^d$ and $U \in \mathcal{S}_+^d$ such that $U^T U \succ \mathbf{0}$, we have*

$$\Pr(Y \not\preceq B) \le \mathrm{tr}\left(\mathbb{E}\left[\exp\left(U(Y-B)U^T\right)\right]\right).$$

The proof of Fact 3.3 can be found in [1, Lemma 17].

PROPOSITION 3.3. *Let $\zeta_1,\ldots,\zeta_l$ be independent real–valued mean–zero random variables satisfying moment growth condition (M) with parameters $(\bar{\theta}_1,v_1),\ldots,(\bar{\theta}_l,v_l)$, respectively. Then, for any $A_1,\ldots,A_l\in\mathcal{S}^d$, we have*

$$\operatorname{tr}\left(\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^{l}\zeta_iA_i\right)\right]\right)\le d\cdot\exp\left(\theta^2\left\|\sum_{i=1}^{l}v_i^2A_i^2\right\|\right)$$

for all $\theta\in(0,\bar{\theta})$, where $\bar{\theta}=\min_{1\le i\le l}\left\{\bar{\theta}_i/\|A_i\|\right\}$.

Proof. The argument is similar to that in [46]. For any $\theta>0$, define

$$D_0=\sum_{i=1}^{l}\theta^2v_i^2A_i^2,\quad D_j=D_0+\sum_{i=1}^{j}\left(\theta\zeta_iA_i-\theta^2v_i^2A_i^2\right)\quad\text{for } j=1,\ldots,l.$$

Since $\operatorname{tr}(\cdot)$ and $\mathbb{E}[\cdot]$ commute, for $j=1,\ldots,l$, we have

$$\begin{aligned}\operatorname{tr}\left(\mathbb{E}\left[\exp(D_j)\right]\right)&=\mathbb{E}\left[\operatorname{tr}\left(\exp\left(D_{j-1}+\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right)\right]\\&\le\mathbb{E}\left[\operatorname{tr}\left(\exp(D_{j-1})\exp\left(\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right)\right]&&(3.3)\\&=\operatorname{tr}\left(\mathbb{E}\left[\exp(D_{j-1})\exp\left(\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right]\right)\\&=\operatorname{tr}\left(\mathbb{E}\left[\exp(D_{j-1})\right]\mathbb{E}\left[\exp\left(\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right]\right),&&(3.4)\end{aligned}$$

where (3.3) follows from the Golden–Thompson inequality (Fact 3.1), and (3.4) follows from the independence of D_{j-1} and $\theta\zeta_jA_j-\theta^2v_j^2A_j^2$.

We claim that

$$\mathbb{E}\left[\exp\left(\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right]\preceq I\tag{3.5}$$

whenever $\theta\in\left(0,\bar{\theta}_j/\|A_j\|\right)$. Indeed, since $\theta\zeta_jA_j$ and $\theta^2v_j^2A_j^2$ commute, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\theta\zeta_jA_j-\theta^2v_j^2A_j^2\right)\right]&=\mathbb{E}\left[\exp(\theta\zeta_jA_j)\exp\left(-\theta^2v_j^2A_j^2\right)\right]\\&=\mathbb{E}\left[\exp(\theta\zeta_jA_j)\right]\exp\left(-\theta^2v_j^2A_j^2\right).\end{aligned}$$

Now, let

$$P_j=\mathbb{E}\left[\exp(\theta\zeta_jA_j)\right]\quad\text{and}\quad Q_j=\exp\left(-\theta^2v_j^2A_j^2\right).$$

By assumption, for any $\theta\in\left(0,\bar{\theta}_j/\|A_j\|\right)$, we have $P_j\preceq\exp\left(\theta^2v_j^2A_j^2\right)=Q_j^{-1}$, which implies that $Q_j^{1/2}P_jQ_j^{1/2}\preceq I$. Since the matrices P_jQ_j and $Q_j^{1/2}P_jQ_j^{1/2}$ are similar, we conclude that $P_jQ_j\preceq I$, as desired.

Using (3.4) and (3.5), we see that when $\theta\in(0,\bar{\theta})$, where $\bar{\theta}=\min_{1\le j\le l}\left\{\bar{\theta}_j/\|A_j\|\right\}$, we have $\operatorname{tr}(\mathbb{E}\left[\exp(D_j)\right])\le\operatorname{tr}(\mathbb{E}\left[\exp(D_{j-1})\right])$ for $j=1,\ldots,l$. This implies that

$$\begin{aligned}\operatorname{tr}\left(\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^{l}\zeta_iA_i\right)\right]\right)&=\operatorname{tr}(\mathbb{E}\left[\exp(D_l)\right])\le\operatorname{tr}(\mathbb{E}\left[\exp(D_0)\right])\\&=\operatorname{tr}\left(\exp\left(\theta^2\sum_{i=1}^{l}v_i^2A_i^2\right)\right).\end{aligned}$$

Upon observing that

$$\mathrm{tr}\left(\exp\left(\theta^2\sum_{i=1}^{l}v_i^2A_i^2\right)\right) \le d\left\|\exp\left(\theta^2\sum_{i=1}^{l}v_i^2A_i^2\right)\right\| = d\cdot\exp\left(\theta^2\left\|\sum_{i=1}^{l}v_i^2A_i^2\right\|\right),$$

the proof is completed. □

Proof of Theorem 3.2. Since $\{(\mathcal{A}_j, w_j)\}_j$ is an exact proper fractional cover of $\mathcal{A}$, we can decompose $S = \sum_{i=1}^{m}\xi_iA_i$ as in (2.1). Let $u, t > 0$ be arbitrary. By taking $Y = S$, $B = tI$ and $U = \sqrt{u}I$ in Fact 3.3, we have

$$\begin{aligned}\Pr(S \not\preceq tI) &\le \exp(-ut)\cdot\mathrm{tr}(\mathbb{E}\left[\exp(uS)\right]) \\ &= \exp(-ut)\cdot\mathbb{E}\left[\mathrm{tr}\left(\exp\left(u\sum_j w_j\sum_{i\in\mathcal{A}_j}\xi_iA_i\right)\right)\right].\end{aligned}$$

Let $S_j = \sum_{i\in\mathcal{A}_j}\xi_iA_i$. Consider a collection $\{p_j\}_j$ of positive numbers, each corresponds to a pair in the exact proper fractional cover of $\mathcal{A}$, such that $\sum_j p_j = 1$. By Fact 3.2, we have

$$\begin{aligned}\mathrm{tr}\left(\exp\left(u\sum_j w_jS_j\right)\right) &= \mathrm{tr}\left(\exp\left(\sum_j p_j\frac{uw_j}{p_j}S_j\right)\right) \\ &\le \sum_j p_j\cdot\mathrm{tr}\left(\exp\left(\frac{uw_j}{p_j}S_j\right)\right).\end{aligned}$$

Moreover, by definition of an exact proper fractional cover, S_j is a sum of independent random matrices. Hence, it follows from moment growth condition (M) and Proposition 3.3 that

$$\begin{aligned}\Pr(S \not\preceq tI) &\le \exp(-ut)\cdot\left[\sum_j p_j\cdot\mathrm{tr}\left(\mathbb{E}\left[\exp\left(\frac{uw_j}{p_j}S_j\right)\right]\right)\right] \\ &\le d\cdot\exp(-ut)\cdot\left[\sum_j p_j\cdot\exp\left(\frac{u^2w_j^2}{p_j^2}\left\|\sum_{i\in\mathcal{A}_j}v_i^2A_i^2\right\|\right)\right]\end{aligned}$$

whenever $uw_j/p_j < \Theta_j$ for all j, where $\Theta_j = \min_{i\in\mathcal{A}_j}\left\{\bar{\theta}_i/\|A_i\|\right\}$. In particular, if we set $p_j = w_jc_j^{1/2}/T$, then

$$\Pr(S \not\preceq tI) \le d\cdot\exp\left(-ut + u^2T^2\right) \tag{3.6}$$

whenever $uT < \Theta_jc_j^{1/2}$ for all j. Now, note that the right–hand side of (3.6) is minimized at $u^* = t/(2T^2)$, and that $\Theta_jc_j^{1/2} \ge \min_{i\in\mathcal{A}_j}\left\{\bar{\theta}_iv_i\right\} \ge \Gamma$. Thus, if $t < 2\Gamma T$, then $u^*T < \Theta_jc_j^{1/2}$ for all j, which implies that

$$\Pr(S \not\preceq tI) \le d\cdot\exp\left(-\frac{t^2}{4T^2}\right).$$

On the other hand, if $t \ge 2\Gamma T$, then for any $u' \in (0, \Gamma/T)$, we have $u'T < \Theta_j c_j^{1/2}$ for all j. This implies that

$$\Pr(S \not\preceq tI) \le d \cdot \inf_{\delta \in (0,1)} \left\{ \exp\left(-\frac{(1-\delta)\Gamma t}{T} + ((1-\delta)\Gamma)^2 \right) \right\} = d \cdot \exp\left(-\frac{\Gamma t}{T} + \Gamma^2 \right),$$

and the proof of Theorem 3.2 is completed. $\square$

Note that in order to apply Theorem 3.2, we need to have an exact proper fractional cover of $\mathcal{A}$. However, such a cover may not be easy to find. Moreover, in the context of computation, some exact proper fractional covers may not admit efficient representations (e.g., when the weight vector $w = (w_j)_j$ has exponentially many non–zero entries). To circumvent these problems, we may follow the idea in Section 2 and provide an upper bound on the quantity $T^* = \min\left\{ \sum_j w_j c_j^{1/2} \right\}$, where c_j is now given by (3.2). For instance, we can bound

$$\left(\sum_j w_j c_j^{1/2} \right)^2 = \left(\sum_j w_j \left\| \sum_{i \in \mathcal{A}_j} v_i^2 A_i^2 \right\|^{1/2} \right)^2 \le \left(\sum_j w_j \right)^2 \left\| \sum_{i=1}^m v_i^2 A_i^2 \right\|,$$

since $\left\| \sum_{i \in \mathcal{A}_j} v_i^2 A_i^2 \right\| \le \left\| \sum_{i=1}^m v_i^2 A_i^2 \right\|$ for all j. Then, we have $(T^*)^2 \le T_{\mathrm{fc}}^2 \left\| \sum_{i=1}^m v_i^2 A_i^2 \right\|$, where T_{fc} is the minimum fractional chromatic number of a dependence graph of the random variables $\xi_1, \ldots, \xi_m$; see (2.8). This yields the following corollary of Theorem 3.2, which does not require knowing any exact proper fractional cover of $\mathcal{A}$:

Corollary 3.4. *Let $\xi_1, \ldots, \xi_m$ be real–valued mean–zero random variables satisfying moment growth condition (M) with parameters $(\bar{\theta}_1, v_1), \ldots, (\bar{\theta}_m, v_m)$, respectively. Let G be a dependence graph of $\xi_1, \ldots, \xi_m$. Then, for any $A_1, \ldots, A_m \in \mathcal{S}^d$, we have*

$$\Pr\left(\sum_{i=1}^m \xi_i A_i \not\preceq tI \right) \le \begin{cases} d \cdot \exp\left(-\dfrac{t^2}{4T^2} \right) & \text{for } 0 < t < 2\Gamma T, \\ d \cdot \exp\left(-\dfrac{\Gamma t}{T} + \Gamma^2 \right) & \text{for } t \ge 2\Gamma T, \end{cases}$$

where

$$T = T_{\mathrm{fc}} \left\| \sum_{i=1}^m v_i^2 A_i^2 \right\|^{1/2}, \quad \Gamma = \min_{1 \le i \le m} \left\{ \bar{\theta}_i v_i \right\},$$

and T_{fc} is the minimum fractional chromatic number of G.

4. Chance–Constrained Linear Matrix Inequalities with Dependent Perturbations: From Large Deviations to Safe Tractable Approximations.

4.1. General Results. Armed with the results in the previous section, we are now ready to address the central question of this paper, namely, to develop *safe tractable approximations* of the chance–constrained linear matrix inequality

$$\Pr_{\xi}\left(A_0(x) + \sum_{i=1}^m \xi_i A_i(x) \preceq \mathbf{0} \right) \ge 1 - \epsilon, \tag{4.1}$$

where $\xi_1, \ldots, \xi_m$ are real–valued mean–zero random variables with a given list of independence relations, $A_0, A_1, \ldots, A_m : \mathbb{R}^n \to \mathcal{S}^d$ are affine functions of the decision vector $x \in \mathbb{R}^n$, and $\epsilon \in (0,1)$ is a tolerance parameter. In other words, we are interested in finding a set $\mathcal{H}$ of deterministic constraints with the following properties:

1. (Tractability) The constraints in $\mathcal{H}$ are efficiently computable.
2. (Safe Approximation) Every feasible solution to $\mathcal{H}$ can be efficiently converted into a feasible solution to (4.1).

As in [6, 56], we shall restrict our attention to those $x \in \mathbb{R}^n$ that satisfy $A_0(x) \prec \mathbf{0}$. Note that such a restriction is almost essential if we want the chance constraint (4.1) to capture sufficiently general settings. Indeed, it is not hard to verify that when $\epsilon \in (0, 1/2)$ and $\xi_1, \ldots, \xi_m$ are mutually independent and symmetric, a necessary condition for (4.1) to hold is that the decision vector $x \in \mathbb{R}^n$ satisfies $A_0(x) \preceq \mathbf{0}$. Now, given $Q_0, Q_1, \ldots, Q_l \in \mathcal{S}^d$, define the $d(l+1) \times d(l+1)$ symmetric matrix $\mathrm{Arrow}(Q_0, Q_1, \ldots, Q_l)$ by

$$\mathrm{Arrow}(Q_0, Q_1, \ldots, Q_l) = \begin{bmatrix} Q_0 & Q_1 & \cdots & Q_l \\ Q_1 & Q_0 & & \\ \vdots & & \ddots & \\ Q_l & & & Q_0 \end{bmatrix}.$$

Then, we have the following theorem:

Theorem 4.1. *Let $\xi_1, \ldots, \xi_m$ be real–valued mean–zero random variables satisfying moment growth condition (M) with parameters $(\bar\theta_1, v_1), \ldots, (\bar\theta_m, v_m)$, respectively. Suppose that an exact proper fractional cover $\{(\mathcal{A}_j, w_j)\}_{j=1}^s$ of $\mathcal{A} = \{1, \ldots, m\}$ is given. Let*

$$\mathcal{A}_j = \left\{ i_1^j, \ldots, i_{s_j}^j \right\} \quad \text{for } j = 1, \ldots, s,$$

$\Gamma = \min_{1 \le i \le m} \left\{ \bar\theta_i v_i \right\}$, and set

$$\tau(\epsilon) = \begin{cases} 2\sqrt{\ln(d/\epsilon)} & \text{if } \Gamma > \sqrt{\ln(d/\epsilon)}, \\ \Gamma + \dfrac{\ln(d/\epsilon)}{\Gamma} & \text{otherwise.} \end{cases}$$

Then, for any given $\epsilon \in (0,1)$, the following system of linear matrix inequalities is a safe tractable approximation of the chance constraint (4.1):

$$\begin{array}{lll} \text{find} & x \in \mathbb{R}^n, y \in \mathbb{R}^s & \\ \text{such that} & A_0(x) \preceq -\left(\tau(\epsilon) \displaystyle\sum_{j=1}^s w_j y_j \right) I, & (a) \\ & \mathrm{Arrow}\left(y_j I, v_{i_1^j} A_{i_1^j}(x), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}(x) \right) \succeq \mathbf{0} \quad \text{for } j = 1, \ldots, s. & (b) \end{array} \tag{4.2}$$

Specifically, if $(x, y) \in \mathbb{R}^n \times \mathbb{R}^s$ is a feasible solution to (4.2), then $x \in \mathbb{R}^n$ is a feasible solution to (4.1).

Remarks.

1. The size of the above linear matrix inequality system depends on s, the size of the exact proper fractional cover. Thus, if s is polynomial in the input parameters, then so is the size of the above system.

2. In general, the quality of the safe tractable approximation (4.2) will depend on the choice of the parameters $(\bar{\theta}_1, v_1), \ldots, (\bar{\theta}_m, v_m)$. Indeed, if one chooses those parameters so that Γ is made larger (but finite), then we have $\tau(\epsilon) = 2\sqrt{\ln(d/\epsilon)}$ for a wider range of ϵ, thus making the constraint (4.2a) easier to satisfy for those ϵ. However, if this is achieved by making some of the v_i's larger, then constraint (4.2b) will be harder to satisfy. It remains an interesting question to determine how the choice of the parameters $(\bar{\theta}_1, v_1), \ldots, (\bar{\theta}_m, v_m)$ affects the feasible region defined by the linear matrix inequalities in (4.2).
3. If $A_0(x), A_1(x), \ldots, A_m(x)$ are diagonal for each $x \in \mathbb{R}^n$ (e.g., in the case of a joint scalar chance constraint), then the linear matrix inequalities in (4.2) reduce to conic quadratic inequalities, which can be solved more efficiently.

Proof. Consider an $x \in \mathbb{R}^n$ that satisfies $A_0(x) \prec \mathbf{0}$. Let $t > 0$ be such that $A_0(x) \preceq -tI$. By Theorem 3.2, we compute

$$\Pr_{\xi}\left(A_0(x) + \sum_{i=1}^{m} \xi_i A_i(x) \not\preceq \mathbf{0} \right) \le \Pr_{\xi}\left(\sum_{i=1}^{m} \xi_i A_i(x) \not\preceq tI \right)$$

$$\le \begin{cases} d \cdot \exp\left(-\dfrac{t^2}{4T^2} \right) & \text{for } 0 < t < 2\Gamma T, \\ d \cdot \exp\left(-\dfrac{\Gamma t}{T} + \Gamma^2 \right) & \text{for } t \ge 2\Gamma T, \end{cases} \tag{4.3}$$

where

$$T = \sum_j w_j \left\| \sum_{i \in \mathcal{A}_j} v_i^2 A_i^2(x) \right\|^{1/2}.$$

In particular, the chance constraint (4.1) will be satisfied if

$$2\sqrt{\ln(d/\epsilon)} \cdot T \le t < 2\Gamma T \tag{4.4}$$

$$\text{or} \quad t \ge \max\left\{ \left(\Gamma + \frac{\ln(d/\epsilon)}{\Gamma} \right) T, 2\Gamma T \right\}. \tag{4.5}$$

Suppose that $\Gamma > \sqrt{\ln(d/\epsilon)}$. Then, condition (4.4) is non–vacuous. We claim that in this case, if $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ is a feasible solution to the system

$$A_0(x) \preceq -tI, \quad 2\sqrt{\ln(d/\epsilon)} \cdot T \le t, \tag{4.6}$$

then $x \in \mathbb{R}^n$ is a feasible solution to (4.1). Indeed, suppose that $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ is feasible for (4.6). If t satisfies (4.4), then x is feasible for (4.1). Otherwise, we have $t \ge 2\Gamma T$, which together with (4.3) yields

$$\Pr_{\xi}\left(A_0(x) + \sum_{i=1}^{m} \xi_i A_i(x) \not\preceq \mathbf{0} \right) \le d \cdot \exp\left(-\frac{\Gamma t}{T} + \Gamma^2 \right) \le d \cdot \exp\left(-\Gamma^2 \right) < \epsilon.$$

This again implies that x is feasible for (4.1), and the claim is established. Now, using the Schur complement, we can reformulate (4.6) as the system of linear matrix inequalities (4.2). This proves the theorem for the case where $\Gamma > \sqrt{\ln(d/\epsilon)}$.

On the other hand, if $\Gamma \leq \sqrt{\ln(d/\epsilon)}$, then only condition (4.5) is non–vacuous. Using the above argument, one can verify that in this case, if $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ is a feasible solution to the system

$$A_0(x) \preceq -tI, \quad \left(\Gamma + \frac{\ln(d/\epsilon)}{\Gamma}\right) T \leq t, \tag{4.7}$$

then $x \in \mathbb{R}^n$ is a feasible solution to (4.1). Moreover, the constraints in (4.7) can be reformulated as the system of linear matrix inequalities (4.2). This completes the proof of Theorem 4.1. □

Remarks.

1. Recall that if the random variables $\xi_1, \ldots, \xi_m$ are mutually independent, then $\{(\mathcal{A}, 1)\}$ is an exact proper fractional cover of $\mathcal{A}$. In this case, we can simplify the arguments in the proof of Theorem 4.1 and obtain the following safe tractable approximation of (4.1):

$$\begin{array}{ll} \text{find} & x \in \mathbb{R}^n \\ \text{such that} & \text{Arrow}\left(-\dfrac{1}{\tau(\epsilon)}A_0(x), v_1A_1(x), \ldots, v_mA_m(x)\right) \succeq \mathbf{0}. \end{array} \tag{4.8}$$

   The safe tractable approximation (4.8) has a similar form as those developed in [6, 56]. However, it is worth noting that even for the case where $\xi_1, \ldots, \xi_m$ are mutually independent, our result extends those in [6, 56], as it does not only apply to Gaussian or bounded–support random variables but also to those that satisfy moment growth condition (M).
2. When an exact proper fractional cover of $\mathcal{A}$ is not readily available, one can still construct a safe tractable approximation of (4.1) by using the minimum fractional chromatic number of a dependence graph of the random variables $\xi_1, \ldots, \xi_m$ and applying Corollary 3.4. Since the derivation largely follows that of Theorem 4.1, we shall not repeat it here.

4.2. Application to Chance–Constrained Quadratically Perturbed Linear Matrix Inequalities. The results in the preceding sections show that the problem of constructing a safe tractable approximation of the chance–constrained linear matrix inequality (4.1) can be essentially reduced to two tasks: (i) find an exact proper fractional cover to decompose the sum $\sum_{i=1}^m \xi_i A_i(x)$ into its independent parts, and (ii) show that the random variables $\xi_1, \ldots, \xi_m$ satisfy moment growth condition (M) and determine the parameters. In this section, we will illustrate the construction by studying chance–constrained quadratically perturbed linear matrix inequalities, i.e., chance constraints of the form shown in (1.5). It is worth noting that quadratically perturbed chance constraints have found applications in many areas, such as finance [64], control [57, 54] and signal processing [38, 61, 62], and the results developed in this section will allow us to tackle those constraints efficiently.

4.2.1. Finding An Exact Proper Fractional Cover. We begin by constructing an exact proper fractional cover for the sum

$$S(x) = \sum_{i=1}^m \zeta_i A_i(x) + \sum_{1 \leq j \leq k \leq m} \zeta_j \zeta_k B_{jk}(x).$$

Towards that end, let $\mathcal{A}_0 = \{1, \ldots, m\}$ and define the sets $\mathcal{A}_1, \ldots, \mathcal{A}_m$ as in Table 4.1.

	1	2	$\cdots$	$m-1$	m
1	$\mathcal{A}_1$	$\mathcal{A}_2$	$\cdots$	$\mathcal{A}_{m-1}$	$\mathcal{A}_m$
2	$\mathcal{A}_2$	$\mathcal{A}_3$	$\cdots$	$\mathcal{A}_m$	$\mathcal{A}_1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$m-1$	$\mathcal{A}_{m-1}$	$\mathcal{A}_m$	$\cdots$	$\mathcal{A}_{m-3}$	$\mathcal{A}_{m-2}$
m	$\mathcal{A}_m$	$\mathcal{A}_1$	$\cdots$	$\mathcal{A}_{m-2}$	$\mathcal{A}_{m-1}$

TABLE 4.1
Construction of the exact proper fractional cover.

In other words, if the (j,k)–th entry of the table is labeled $\mathcal{A}_l$, then $(j,k) \in \mathcal{A}_l$. We do not distinguish the pairs (j,k) and (k,j), and we assume that only the pair (j,k) with $j \le k$ appears in $\mathcal{A}_l$. Our interest in the sets $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_m$ lies in the following result:

PROPOSITION 4.2. *Let $\mathcal{A}_0 = \{1,\dots,m\}$ and $\mathcal{A}_1,\dots,\mathcal{A}_m$ be given by Table 4.1. Then, the following hold:*

(a) The random variables in $\{\zeta_i : i \in \mathcal{A}_0\}$ are independent. Moreover, for each $l = 1,\dots,m$, the random variables in $\{\zeta_j\zeta_k : (j,k) \in \mathcal{A}_l\}$ are independent.

(b) $\{(\mathcal{A}_i, 1)\}_{i=0}^m$ is an exact proper fractional cover of $\mathcal{A} = \{1,\dots,m\} \cup \{(j,k) : 1 \le j \le k \le m\}$.

Proof.

(a) Since $\zeta_1,\dots,\zeta_m$ are independent random variables, the first statement is clear. To prove the second statement, it suffices to show that for each $l = 1,\dots,m$, if $(j,k),(j',k') \in \mathcal{A}_l$, then $\{j,k\} \cap \{j',k'\} = \emptyset$. Without loss of generality, we may assume that $j \le j'$. By considering the j–th row of Table 4.1, we must have $j < j'$, for otherwise $(j,k),(j',k')$ cannot both belong to $\mathcal{A}_l$. Since $j' \le k'$ by construction, we have $j < k'$ as well.
Now, by considering the k–th column of Table 4.1, we have $k \ne k'$. Thus, it remains to show that $k \ne j'$. Suppose to the contrary that $k = j'$. Then, by considering the k–th row of Table 4.1 and using the fact that the (j,k)–th entry has the same label as the (k,j)–th entry, we conclude that $(j,k),(j',k')$ cannot both belong to $\mathcal{A}_l$, which is a contradiction.

(b) By definition of an exact proper fractional cover and the result in (a), it suffices to show that for each element $u \in \mathcal{A}$, there exists a unique l such that $u \in \mathcal{A}_l$. However, this is clear from the construction of the sets $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_m$. $\square$

By Proposition 4.2, we may write

$$\begin{aligned} & A_0(x) + \sum_{i=1}^m \zeta_i A_i(x) + \sum_{1 \le j \le k \le m} \zeta_j\zeta_k B_{jk}(x) \\ = \; & A_0(x) + \sigma^2 \sum_{j=1}^m B_{jj}(x) + \sum_{i \in \mathcal{A}_0} \zeta_i A_i(x) \\ & + \sum_{l=1}^m \left[\left(\sum_{(j,j) \in \mathcal{A}_l} \left(\zeta_j^2 - \sigma^2\right) B_{jj}(x) \right) + \left(\sum_{(j,k) \in \mathcal{A}_l : j<k} \zeta_j\zeta_k B_{jk}(x) \right) \right], \end{aligned} \tag{4.9}$$

where $\sigma^2 = \mathbb{E}\left[\zeta_1^2\right]$. In particular, once we show that the mean–zero random variables ζ_1, $\zeta_1\zeta_2$ and $\zeta_1^2 - \sigma^2$ satisfy moment growth condition (M), we can apply Theorem 4.1 and obtain a safe tractable approximation of the chance constraint (1.5).

4.2.2. Bounding the Matrix Moment Generating Functions. Now, let us study the behavior of $S(x)$ under various moment assumptions on the i.i.d. real–valued mean–zero random variables $\zeta_1, \ldots, \zeta_m$ and develop the corresponding safe tractable approximations of the chance constraint (1.5).

A. Bounded Perturbations.

Suppose that ζ_i is supported on $[-1,1]$ with $\sigma^2 = \mathbb{E}\left[\zeta_i^2\right] \in [\underline{\sigma}^2, \overline{\sigma}^2]$ for $i = 1, \ldots, m$. To prove that the mean–zero random variables ζ_1, $\zeta_1\zeta_2$ and $\zeta_1^2 - \sigma^2$ satisfy moment growth condition (M), we need the following result:

Proposition 4.3. *Let X be a real–valued mean–zero random variable supported on $[a,b]$, where $a, b \in \mathbb{R}$. Then, for any $\theta > 0$ and $Q \in \mathcal{S}^d$, we have*

$$\mathbb{E}\left[\exp(\theta X Q)\right] \preceq \exp\left(\frac{1}{8}\theta^2 (b-a)^2 Q^2\right).$$

The proof of Proposition 4.3 relies on the following fact (see (4.16) of [33]):

Fact 4.1. (Hoeffding Inequality [33]) *Let X be a real–valued mean–zero random variable supported on $[a,b]$, where $a, b \in \mathbb{R}$. Then, for any $\theta > 0$, we have*

$$\mathbb{E}\left[\exp(\theta X)\right] \le \exp\left(\frac{1}{8}\theta^2 (b-a)^2\right).$$

Proof of Proposition 4.3. Let $\theta > 0$ be arbitrary, and let $Q = U\Lambda U^T$ be the spectral decomposition of Q, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix consisting of the eigenvalues of Q. By definition of the matrix exponential (see (3.1)), we have

$$\begin{aligned}
\mathbb{E}\left[\exp(\theta X Q)\right] &= \mathbb{E}\left[U\left(I + \sum_{i=1}^{\infty} \frac{\theta^i \Lambda^i}{i!} X^i\right) U^T\right] \\
&= U \mathbb{E}\left[\exp(\theta X \Lambda)\right] U^T \\
&= U \mathrm{diag}(\mathbb{E}\left[\exp(\theta \lambda_1 X)\right], \ldots, \mathbb{E}\left[\exp(\theta \lambda_d X)\right]) U^T.
\end{aligned}$$

Moreover, by Fact 4.1, we have $\mathbb{E}\left[\exp(\theta \lambda_i X)\right] \le \exp\left(\theta^2 \lambda_i^2 (b-a)^2/8\right)$ for $i = 1, \ldots, d$. It follows that

$$\begin{aligned}
\mathbb{E}\left[\exp(\theta X Q)\right] &\preceq U \mathrm{diag}\left(\exp\left(\frac{1}{8}\theta^2 \lambda_1^2 (b-a)^2\right), \ldots, \exp\left(\frac{1}{8}\theta^2 \lambda_d^2 (b-a)^2\right)\right) U^T \\
&= U \exp\left(\mathrm{diag}\left(\frac{1}{8}\theta^2 \lambda_1^2 (b-a)^2, \ldots, \frac{1}{8}\theta^2 \lambda_d^2 (b-a)^2\right)\right) U^T \\
&= U \exp\left(\frac{1}{8}\theta^2 (b-a)^2 \Lambda^2\right) U^T \\
&= \exp\left(\frac{1}{8}\theta^2 (b-a)^2 U \Lambda^2 U^T\right) \\
&= \exp\left(\frac{1}{8}\theta^2 (b-a)^2 Q^2\right),
\end{aligned}$$

as desired. $\square$

Using Proposition 4.3, it is straightforward to obtain a safe tractable approximation of (1.5) for the case where the random variables $\zeta_1, \ldots, \zeta_m$ are bounded and have bounded second moments.

THEOREM 4.4. *Suppose that $\zeta_1, \ldots, \zeta_m$ are i.i.d. real–valued mean–zero random variables supported on $[-1, 1]$ with $\sigma^2 = \mathbb{E}\left[\zeta_1^2\right] \in [\underline{\sigma}^2, \overline{\sigma}^2]$. Then, the following system of linear matrix inequalities is a safe tractable approximation of the chance constraint (1.5):*

$$
\begin{array}{lll}
\text{find} & x \in \mathbb{R}^n, y \in \mathbb{R}^{m+1} & \\
\text{such that} & A_0(x) + \underline{\sigma}^2 \displaystyle\sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l \right) I, & \\
& A_0(x) + \overline{\sigma}^2 \displaystyle\sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l \right) I, & \\
& \text{Arrow}\left(y_0 I, \dfrac{1}{\sqrt{2}} A_1(x), \ldots, \dfrac{1}{\sqrt{2}} A_m(x) \right) \succeq \mathbf{0}, & \\
& \text{Arrow}\left(y_l I, \left(v_{jk} B_{jk}(x)\right)_{(j,k) \in \mathcal{A}_l} \right) \succeq \mathbf{0} & \textit{for } l = 1, \ldots, m,
\end{array}
$$

where $v_{jj} = 1/\sqrt{8}$ and $v_{jk} = 1/\sqrt{2}$ if $j < k$, for $1 \le j \le k \le m$.

Proof. Since ζ_i is supported on $[-1, 1]$ for $i = 1, \ldots, m$, we see that $\zeta_1\zeta_2$ is also supported on $[-1, 1]$, and that $\zeta_1^2 - \sigma^2$ is supported on $[-\sigma^2, 1 - \sigma^2]$. Hence, by Proposition 4.3, we have

$$
\mathbb{E}\left[\exp(\theta\zeta_1 Q)\right] \preceq \exp\left(\frac{1}{2}\theta^2 Q^2 \right),
$$

$$
\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2 Q)\right] \preceq \exp\left(\frac{1}{2}\theta^2 Q^2 \right),
$$

$$
\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right) Q\right)\right] \preceq \exp\left(\frac{1}{8}\theta^2 Q^2 \right)
$$

for all $\theta > 0$, i.e., ζ_1, $\zeta_1\zeta_2$ and $\zeta_1^2 - \sigma^2$ satisfy moment growth condition (M) with parameters $(+\infty, 1/\sqrt{2})$, $(+\infty, 1/\sqrt{2})$ and $(+\infty, 1/\sqrt{8})$, respectively. Hence, we conclude from (4.9) and Theorem 4.1 that the following system of linear matrix inequalities is a safe tractable approximation of the chance constraint (1.5):

$$
\begin{array}{lll}
\text{find} & x \in \mathbb{R}^n, y \in \mathbb{R}^{m+1} & \\
\text{such that} & A_0(x) + \sigma^2 \displaystyle\sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l \right) I, & (\dagger) \\
& \text{Arrow}\left(y_0 I, \dfrac{1}{\sqrt{2}} A_1(x), \ldots, \dfrac{1}{\sqrt{2}} A_m(x) \right) \succeq \mathbf{0}, & \\
& \text{Arrow}\left(y_l I, \left(v_{jk} B_{jk}(x)\right)_{(j,k) \in \mathcal{A}_l} \right) \succeq \mathbf{0} & \text{for } l = 1, \ldots, m.
\end{array}
$$

Here, $v_{jj} = 1/\sqrt{8}$ and $v_{jk} = 1/\sqrt{2}$ if $j < k$, for $1 \le j \le k \le m$. If we do not know σ^2 exactly but know that $\sigma^2 \in [\underline{\sigma}^2, \overline{\sigma}^2]$, then we can replace (†) with the following robust constraint:

$$A_0(x) + \sigma^2 \sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l\right) I \quad \text{for all } \sigma^2 \in [\underline{\sigma}^2, \overline{\sigma}^2]. \tag{4.10}$$

As can be easily verified, the following system of linear matrix inequalities is equivalent to (4.10):

$$A_0(x) + \underline{\sigma}^2 \sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l\right) I,$$

$$A_0(x) + \overline{\sigma}^2 \sum_{j=1}^m B_{jj}(x) \preceq -\left(2\sqrt{\ln(d/\epsilon)} \cdot \sum_{l=0}^m y_l\right) I.$$

This completes the proof of Theorem 4.4. $\square$

B. Gaussian Perturbations.

Suppose that $\zeta_1, \ldots, \zeta_m$ are i.i.d. standard Gaussian random variables. We then have the following matrix moment generating function bounds:

PROPOSITION 4.5. *Let ζ_1, ζ_2 be independent standard Gaussian random variables. Then, we have*

$$\mathbb{E}\left[\exp(\theta\zeta_1 Q)\right] = \exp\left(\frac{1}{2}\theta^2 Q^2\right) \qquad \textit{for any } \theta > 0, Q \in \mathcal{S}^d, \tag{4.11}$$

$$\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2 Q)\right] \preceq \exp\left(\theta^2 Q^2\right) \qquad \textit{for any } \theta \in (0, 0.89) \textit{ and } Q \in \mathcal{S}^d \textit{ with } \|Q\| = 1, \tag{4.12}$$

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - 1\right)Q\right)\right] \preceq \exp\left(4\theta^2 Q^2\right) \qquad \textit{for any } \theta \in (0, 0.465) \textit{ and } Q \in \mathcal{S}^d \textit{ with } \|Q\| = 1. \tag{4.13}$$

In other words, $\zeta_1, \zeta_1\zeta_2$ and $\zeta_1^2 - 1$ satisfy moment growth condition (M) with parameters $(+\infty, 1/\sqrt{2})$, $(0.89, 1)$ and $(0.465, 2)$, respectively.

REMARK. The constants above are chosen out of convenience and are by no means the only possible choice. However, the quality of the safe tractable approximation will in general depend on those constants; see Remark 2 after Theorem 4.1.

Proof. The proof of (4.11) can be found in [59]. To prove (4.12) and (4.13), recall that

$$\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2)\right] = \frac{1}{\sqrt{1-\theta^2}} \qquad \text{for any } \theta \in (-1, 1),$$

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - 1\right)\right)\right] = \frac{\exp(-\theta)}{\sqrt{1-2\theta}} \qquad \text{for any } \theta < \frac{1}{2}.$$

In particular, it can be verified that

$$\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2)\right] \le \exp\left(\theta^2\right) \qquad \text{for any } \theta \in (-0.89, 0.89),$$

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - 1\right)\right)\right] \le \exp\left(4\theta^2\right) \qquad \text{for any } \theta < 0.465.$$

Hence, using the argument in the proof of Proposition 4.3, we obtain

$$\begin{aligned} \mathbb{E}\left[\exp(\theta\zeta_1\zeta_2 Q)\right] &\preceq \exp\left(\theta^2 Q^2\right) && \text{for any } \theta \in (0, 0.89) \\ & && \text{and } Q \in \mathcal{S}^d \text{ with } \|Q\| = 1, \\ \mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - 1\right) Q\right)\right] &\preceq \exp\left(4\theta^2 Q^2\right) && \text{for any } \theta \in (0, 0.465) \\ & && \text{and } Q \in \mathcal{S}^d \text{ with } \|Q\| = 1, \end{aligned}$$

as desired. ◻

Upon combining Theorem 4.1 and Proposition 4.5 and noting that $\Gamma = \min\{0.89, 0.465 \times 2\} = 0.89$ in Theorem 4.1, we have the following theorem:

THEOREM 4.6. *Suppose that $\zeta_1, \ldots, \zeta_m$ are i.i.d. standard Gaussian random variables. Then, the following system of linear matrix inequalities is a safe tractable approximation of the chance constraint (1.5):*

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n, y \in \mathbb{R}^{m+1} \\ \text{such that} \quad & A_0(x) + \sum_{j=1}^m B_{jj}(x) \preceq -\left(\tau(\epsilon)\sum_{l=0}^m y_l\right) I, \\ & \operatorname{Arrow}\left(y_0 I, \frac{1}{\sqrt{2}}A_1(x), \ldots, \frac{1}{\sqrt{2}}A_m(x)\right) \succeq \mathbf{0}, \\ & \operatorname{Arrow}\left(y_l I, \left(v_{jk}B_{jk}(x)\right)_{(j,k)\in\mathcal{A}_l}\right) \succeq \mathbf{0} \qquad \textit{for } l = 1, \ldots, m. \end{aligned}$$

Here,

$$\tau(\epsilon) = \begin{cases} 2\sqrt{\ln(d/\epsilon)} & \textit{if } \sqrt{\ln(d/\epsilon)} < 0.89, \\ 0.89 + \dfrac{\ln(d/\epsilon)}{0.89} & \textit{otherwise}, \end{cases}$$

$v_{jj} = 2$ and $v_{jk} = 1$ if $j < k$, for $1 \le j \le k \le m$.

REMARK. For the scalar case (i.e., when $d = 1$), it is possible to derive a more compact safe tractable approximation of (1.5) than that offered by Theorem 4.6. To see this, let $x \in \mathbb{R}^n$ be fixed and write

$$A_0(x) + \sum_{i=1}^m \zeta_i A_i(x) + \sum_{1\le j\le k\le m} \zeta_j\zeta_k B_{jk}(x) = A_0(x) + \zeta^T A(x) + \zeta^T \bar{B}(x)\zeta,$$

where $\zeta = (\zeta_1, \ldots, \zeta_m) \in \mathbb{R}^m$, $A(x) = (A_1(x), \ldots, A_m(x)) \in \mathbb{R}^m$ and $\bar{B}(x) \in \mathcal{S}^m$ with

$$\bar{B}_{jk}(x) = \begin{cases} B_{jj}(x) & \text{if } j = k, \\ \dfrac{1}{2}B_{jk}(x) & \text{if } j < k, \end{cases}$$

for $1 \le j \le k \le m$. Let $\bar{B}(x) = U\Lambda U^T$ be the spectral decomposition of $\bar{B}(x)$ (for notational simplicity, we suppress the dependence of U and Λ on x). Then, we have

$$A_0(x) + \zeta^T A(x) + \zeta^T \bar{B}(x)\zeta = A_0(x) + \zeta^T A(x) + \bar{\zeta}^T \Lambda \bar{\zeta},$$

where $\bar{\zeta} = U^T\zeta$. Since ζ is a standard Gaussian random vector and U^T is orthogonal, $\bar{\zeta}$ is also a standard Gaussian random vector. Moreover, by defining $\mathcal{A}_0 = \{1, \ldots, m\}$ and $\mathcal{A}_1 = \{(1,1), \ldots, (m,m)\}$, we see that the random variables in $\{\zeta_i : i \in \mathcal{A}_0\}$ and $\{\bar{\zeta}_{jj}^2 : (j,j) \in \mathcal{A}_1\}$ are mutually independent. Thus, using (4.11), (4.13) and following the proof of Theorem 4.1, we obtain the following conservative approximation of the chance constraint (1.5):

$$A_0(x) + \sum_{j=1}^m \Lambda_{jj} \le -t, \quad \tau(\epsilon)\left[\frac{1}{\sqrt{2}}\left(\sum_{i=1}^m A_i^2(x)\right)^{1/2} + 2\left(\sum_{j=1}^m \Lambda_{jj}^2\right)^{1/2}\right] \le t.$$

Here,

$$\tau(\epsilon) = \begin{cases} 2\sqrt{\ln(1/\epsilon)} & \text{if } \sqrt{\ln(1/\epsilon)} < 0.93, \\ 0.93 + \dfrac{\ln(1/\epsilon)}{0.93} & \text{otherwise.} \end{cases}$$

Since $\sum_{j=1}^m \Lambda_{jj} = \mathrm{tr}(\bar{B}(x))$ and $\sum_{j=1}^m \Lambda_{jj}^2 = \mathrm{tr}\left(\bar{B}^2(x)\right)$, the above constraints can be reformulated as the following system of conic quadratic inequalities:

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n, (y_0, y_1) \in \mathbb{R}^2 \\ \text{such that} \quad & \tau(\epsilon)(y_0 + y_1) \le -A_0(x) - \mathrm{tr}\left(\bar{B}(x)\right), \\ & \frac{1}{\sqrt{2}}\|A(x)\|_2 \le y_0, \\ & 2\left(\sum_{j,k=1}^m \bar{B}_{jk}^2(x)\right)^{1/2} \le y_1. \end{aligned} \tag{4.14}$$

Curiously, for the case where $d = 1$, an alternative safe tractable approximation of (1.5) with Gaussian perturbations can be obtained from a large deviation inequality due to Bechar [2]. In [2] it is shown that if ζ is a standard Gaussian random vector, $Q \in \mathcal{S}^m$ and $c \in \mathbb{R}^m$, then

$$\Pr\left(\zeta^T Q\zeta + c^T\zeta > \mathrm{tr}(Q) + 2\sqrt{\ln\frac{1}{\epsilon}} \cdot \sqrt{\sum_{j,k=1}^m Q_{jk}^2 + \frac{1}{2}\sum_{i=1}^m c_i^2} + 2s^+ \ln\frac{1}{\epsilon}\right) \le \epsilon,$$

where $s^+ = \max\{\lambda_{\max}(Q), 0\}$ and $\lambda_{\max}(Q)$ is the largest eigenvalue of Q. By specializing Bechar's result to our setting, we obtain the following safe tractable approximation of (1.5):

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n, y \in \mathbb{R} \\ \text{such that} \quad & 2\sqrt{\ln\frac{1}{\epsilon}} \cdot \sqrt{\sum_{j,k=1}^m \bar{B}_{jk}^2(x) + \frac{1}{2}\sum_{i=1}^m A_i^2(x)} \le -A_0(x) - \mathrm{tr}\left(\bar{B}(x)\right) - 2y\ln\frac{1}{\epsilon}, \\ & yI \succeq \bar{B}(x), \\ & y \ge 0. \end{aligned} \tag{4.15}$$

One would expect that (4.14) can be solved more efficiently than (4.15), as the former involves only conic quadratic inequalities, while the latter involves both conic quadratic and linear matrix inequalities. This is indeed the case in our numerical experiments with a robust transmit beamforming problem in signal processing. For a comparison of the performance of (4.14) and (4.15) in the context of the transmit beamforming problem, we refer the reader to our recent preprint [62].

C. Subgaussian Perturbations.

Let us now consider a more general class of perturbations, namely the class of subgaussian random variables. As the name suggests, all the random variables in this class possess similar properties (such as tail behavior) as the standard normal random variable. We begin with the definition:

Definition 4.7. *A real–valued random variable X is said to be* subgaussian with exponent $v > 0$ *if* $\mathbb{E}\left[\exp(\theta X)\right] \le \exp\left(\theta^2 v^2\right)$ *for all* $\theta \in \mathbb{R}$.

It is not hard to see that the standard normal and all bounded random variables are subgaussian. Moreover, the notion of forward and backward deviations introduced in [22] is closely related to that of subgaussianity. Specifically, if $\alpha > 0$ belongs to both the forward and backward deviation sets associated with the real–valued mean–zero random variable X, then X is a subgaussian random variable with exponent $v = \alpha/2$.

The following result is an easy consequence of Definition 4.7; see, e.g., [27, Theorem 12.7.1] for a proof.

Fact 4.2. (Basic Properties of Subgaussian Random Variables) *Let X be a subgaussian random variable with exponent $v > 0$. Then, we have $\mathbb{E}[X] = 0$,*

$$\Pr(X > t) \le \exp\left(-\frac{t^2}{4v^2}\right) \quad \textit{and} \quad \Pr(X < -t) \le \exp\left(-\frac{t^2}{4v^2}\right)$$

for all $t > 0$.

To construct a safe tractable approximation of the chance constraint (1.5) with subgaussian perturbations, we need to derive certain matrix moment generating function bounds. This is accomplished in the following proposition:

Proposition 4.8. *Let ζ_1, ζ_2 be i.i.d. subgaussian random variables with exponent $v > 0$, and let $\sigma^2 = \mathbb{E}\left[\zeta_1^2\right]$. Then, we have*

$$\mathbb{E}\left[\exp(\theta\zeta_1 Q)\right] \preceq \exp\left(\theta^2 v^2 Q^2\right) \qquad \textit{for any } \theta > 0, Q \in \mathcal{S}^d,$$

$$\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2 Q)\right] \preceq \exp\left(\frac{64}{(1-\alpha)^3}\theta^2 v^4 Q^2\right) \qquad \textit{for any } \alpha \in (0,1), \theta \in \left(0, \frac{\alpha}{4v^2}\right) \textit{ and } Q \in \mathcal{S}^d \textit{ with } \|Q\| = 1,$$

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)Q\right)\right] \preceq \exp\left(\left(45v^2 - \sigma^2\right)\theta^2 v^2 Q^2\right) \quad \textit{for any } \theta \in \left(0, \frac{1}{8v^2}\right) \textit{ and } Q \in \mathcal{S}^d \textit{ with } \|Q\| = 1.$$

In other words, $\zeta_1, \zeta_1\zeta_2$ and $\zeta_1^2 - 1$ satisfy moment growth condition (M) with parameters $(+\infty, v)$, $\left(\alpha/\left(4v^2\right), 8v^2/(1-\alpha)^{3/2}\right)$ for any $\alpha \in (0,1)$ and $\left(1/\left(8v^2\right), \sqrt{45v^2 - \sigma^2} \cdot v\right)$, respectively.

Remark. It is known that if X is a subgaussian random variable with exponent $v > 0$, then $\mathbb{E}\left[X^2\right] \le 8v^2$; see, e.g., [27, Theorem 12.7.1]. Hence, we always have $45v^2 - \sigma^2 \ge 37v^2 > 0$.

To prove Proposition 4.8, we need the following result:

FACT 4.3. (TAIL BEHAVIOR AND MOMENT GENERATING FUNCTION BOUND) *Let X be a real–valued mean–zero random variable. Suppose that there exist $M > 0$ and $\gamma > 0$ such that $\Pr(|X| > t) \le M \int_t^\infty \exp(-\gamma z)\,dz$ for all $t \ge 0$. Then, for any $\alpha \in (0,1)$, we have*

$$\mathbb{E}\left[\exp(\theta X)\right] \le \exp\left(\frac{2M}{(1-\alpha)^3\gamma^3}\theta^2\right)$$

for all $\theta \in (-\alpha\gamma, \alpha\gamma)$.

For a proof of Fact 4.3, see [31, Lemma 3].

Proof of Proposition 4.8. The proof follows the same line as that of Proposition 4.5. By definition, we have $\mathbb{E}\left[\exp(\theta\zeta_1)\right] \le \exp\left(\theta^2 v^2\right)$ for all $\theta \in \mathbb{R}$, from which it follows that

$$\mathbb{E}\left[\exp(\theta\zeta_1 Q)\right] \preceq \exp\left(\theta^2 v^2 Q^2\right)$$

for all $\theta \in \mathbb{R}$ and $Q \in \mathcal{S}^d$. Next, we compute $\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2)\right]$. Since $\mathbb{E}\left[\zeta_1\zeta_2\right] = 0$ and

$$\begin{aligned}
\Pr(|\zeta_1\zeta_2| > t) &\le \Pr(|\zeta_1| > \sqrt{t}) + \Pr(|\zeta_2| > \sqrt{t}) \\
&\le 2\exp\left(-\frac{t}{4v^2}\right) \\
&= \frac{1}{2v^2}\int_t^\infty \exp\left(-\frac{z}{4v^2}\right)\,dz
\end{aligned}$$

for any $t \ge 0$, by Fact 4.3, we have $\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2)\right] \le \exp\left(64\theta^2 v^4/(1-\alpha)^3\right)$ for all $\alpha \in (0,1)$ and $\theta \in \left(-\alpha/\left(4v^2\right), \alpha/\left(4v^2\right)\right)$. This yields

$$\mathbb{E}\left[\exp(\theta\zeta_1\zeta_2 Q)\right] \preceq \exp\left(\frac{64}{(1-\alpha)^3}\theta^2 v^4 Q^2\right)$$

for all $\alpha \in (0,1)$, $\theta \in \left(0, \alpha/\left(4v^2\right)\right)$ and $Q \in \mathcal{S}^d$ with $\|Q\| = 1$. Finally, let us compute $\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)\right)\right]$. For any $\theta \in \left(0, 1/\left(8v^2\right)\right)$, we have

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)\right)\right] &= \exp\left(-\theta\sigma^2\right)\int_1^\infty \Pr(\exp\left(\theta\zeta_1^2\right) > z)\,dz \\
&= 2\theta \cdot \exp\left(-\theta\sigma^2\right)\int_0^\infty u \cdot \exp\left(\theta u^2\right) \cdot \Pr(|\zeta_1| > u)\,du \\
&\le 4\theta \cdot \exp\left(-\theta\sigma^2\right)\int_0^\infty u \cdot \exp\left(\left(\theta - \frac{1}{4v^2}\right)u^2\right)\,du \\
&= 2\theta\left(\frac{1}{4v^2} - \theta\right)^{-1}\exp\left(-\theta v^2\frac{\sigma^2}{v^2}\right) \\
&\le \exp\left(\left(45v^2 - \sigma^2\right)\theta^2 v^2\right), \qquad (4.16)
\end{aligned}$$

where the last inequality follows from the fact that

$$2\theta\left(\frac{1}{4v^2} - \theta\right)^{-1} \le \exp\left(45\theta^2 v^4\right) \quad \text{and} \quad \theta v^2 > \theta^2 v^4 \qquad \text{for all } \theta \in \left(0, 1/\left(8v^2\right)\right).$$

On the other hand, for any $\theta \in \left(-1/\left(8v^2\right), 0\right)$, we have

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)\right)\right] &= \exp\left(-\theta\sigma^2\right) \cdot \mathbb{E}\left[\exp\left(\theta\zeta_1^2\right)\right] \\
&\le \exp\left(-\theta\sigma^2\right) \cdot \mathbb{E}\left[1 + \theta\zeta_1^2 + \frac{1}{2}\left(\theta\zeta_1^2\right)^2\right] \qquad (4.17) \\
&= \exp\left(-\theta\sigma^2\right) \cdot \left(1 + \theta\sigma^2 + \frac{\theta^2}{2}\mathbb{E}\left[\zeta_1^4\right]\right) \\
&\le \exp\left(-\theta\sigma^2\right) \cdot \left(1 + \theta\sigma^2 + 32\theta^2 v^4\right), \qquad (4.18)
\end{aligned}$$

where (4.17) follows from the inequality $\exp(-x) \le 1 - x + x^2/2$, which is valid for all $x \ge 0$, and (4.18) follows from the fact that $\mathbb{E}\left[\zeta_1^4\right] \le 64v^4$; see [27, Theorem 12.7.1]. Now, observe that $32\theta^2 v^4 \le \left(45v^2 - \sigma^2\right)\theta^2 v^2$ for all $\theta \in \left(-1/\left(8v^2\right), 1/\left(8v^2\right)\right)$. Moreover, using the inequality $\exp(x) \ge 1 + x$, which is valid for all $x \in \mathbb{R}$, we have

$$\begin{aligned}
&\exp\left(-\theta\sigma^2\right) \cdot \left(1 + \theta\sigma^2 + \left(45v^2 - \sigma^2\right)\theta^2 v^2\right) \\
&\le \exp\left(-\theta\sigma^2\right) \cdot \exp\left(\theta\sigma^2 + \left(45v^2 - \sigma^2\right)\theta^2 v^2\right) \\
&= \exp\left(\left(45v^2 - \sigma^2\right)\theta^2 v^2\right)
\end{aligned}$$

for all $\theta \in \left(-1/\left(8v^2\right), 0\right)$. Hence, it follows from (4.18) that

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)\right)\right] \le \exp\left(\left(45v^2 - \sigma^2\right)\theta^2 v^2\right)$$

for all $\theta \in \left(-1/\left(8v^2\right), 0\right)$. This, together with (4.16), implies that

$$\mathbb{E}\left[\exp\left(\theta\left(\zeta_1^2 - \sigma^2\right)Q\right)\right] \preceq \exp\left(\left(45v^2 - \sigma^2\right)\theta^2 v^2 Q^2\right)$$

for all $\theta \in \left(0, 1/\left(8v^2\right)\right)$ and $Q \in \mathcal{S}^d$ with $\|Q\| = 1$, as required. ∎

Now, let $\bar{\alpha} \approx 0.2645$ be the solution to the equation

$$\frac{2\alpha}{(1-\alpha)^{3/2}} = \frac{\sqrt{45}}{8}.$$

Then, we have

$$\Gamma = \min\left\{\frac{2\bar{\alpha}}{(1-\bar{\alpha})^{3/2}}, \frac{\sqrt{45v^2 - \overline{\sigma}^2}}{8v}\right\} = \frac{\sqrt{45v^2 - \overline{\sigma}^2}}{8v}.$$

Upon invoking Theorem 4.1 and Proposition 4.8, we are immediately led to the following theorem:

THEOREM 4.9. *Suppose that $\zeta_1, \ldots, \zeta_m$ are i.i.d. subgaussian random variables with exponent $v > 0$. Let $\sigma^2 = \mathbb{E}\left[\zeta_1^2\right] \in [\underline{\sigma}^2, \overline{\sigma}^2]$ and $\Gamma = \sqrt{45v^2 - \overline{\sigma}^2}/(8v)$. Then, the following system of linear matrix inequalities is a safe tractable approximation of* *the chance constraint (1.5):*

$$
\begin{aligned}
\text{find} \quad & x \in \mathbb{R}^n, y \in \mathbb{R}^{m+1} \\
\text{such that} \quad & A_0(x) + \underline{\sigma}^2 \sum_{j=1}^{m} B_{jj}(x) \preceq -\left(\tau(\epsilon) \sum_{l=0}^{m} y_l \right) I, \\
& A_0(x) + \overline{\sigma}^2 \sum_{j=1}^{m} B_{jj}(x) \preceq -\left(\tau(\epsilon) \sum_{l=0}^{m} y_l \right) I, \\
& \mathrm{Arrow}\,(y_0 I, vA_1(x), \ldots, vA_m(x)) \succeq \mathbf{0}, \\
& \mathrm{Arrow}\left(y_l I, (v_{jk} B_{jk}(x))_{(j,k) \in \mathcal{A}_l} \right) \succeq \mathbf{0} \qquad \text{for } l = 1, \ldots, m.
\end{aligned}
$$

Here,

$$
\tau(\epsilon) = \begin{cases} 2\sqrt{\ln(d/\epsilon)} & \text{if } \Gamma > \sqrt{\ln(d/\epsilon)}, \\ \Gamma + \dfrac{\ln(d/\epsilon)}{\Gamma} & \text{otherwise}, \end{cases}
$$

$v_{jj} = \sqrt{45v^2 - \underline{\sigma}^2} \cdot v$ *and* $v_{jk} = 8v^2/(1-\bar{\alpha})^{3/2}$ *if* $j < k$*, for* $1 \le j \le k \le m$*.*

5. Iterative Improvement of the Proposed Approximations. From the derivations in Section 4.1, it is clear that there are many ways to construct a safe tractable approximation of the chance constraint (4.1). Thus, it is natural to ask how conservative are our proposed safe tractable approximations. To address this question, it is instructive to revisit the derivation of (4.2). One potential source of conservatism lies in the constraint (4.2a), which is a consequence of the bound $A_0(x) \preceq -tI$ used in the proof of Theorem 4.1. Indeed, such a bound can be quite weak if the eigenvalues of $A_0(x)$ are spread out. To fix this problem, we could try to "precondition" $A_0(x)$ so that its eigenvalues are as equal as possible. Specifically, observe that the chance constraint (4.1) is equivalent to

$$
\Pr_{\xi}\left(D \left(A_0(x) + \sum_{i=1}^{m} \xi_i A_i(x) \right) D \preceq \mathbf{0} \right) \ge 1 - \epsilon, \tag{5.1}
$$

where D is any $d \times d$ invertible symmetric matrix. Under the assumptions of Theorem 4.1, the following is a safe tractable approximation of (5.1):

$$
\begin{aligned}
\textit{find} \quad & x \in \mathbb{R}^n, y \in \mathbb{R}^s \\
\textit{such that} \quad & DA_0(x)D \preceq -\left(\tau(\epsilon) \sum_{j=1}^{s} w_j y_j \right) I, \\
& \mathrm{Arrow}\left(y_j I, v_{i_1^j} DA_{i_1^j}(x)D, \ldots, v_{i_{s_j}^j} DA_{i_{s_j}^j}(x)D \right) \succeq \mathbf{0} \quad \text{for } j = 1, \ldots, s.
\end{aligned}
$$

This can also be written as

$$
\begin{array}{lll}
\textit{find} & x \in \mathbb{R}^n, y \in \mathbb{R}^s & \\
\textit{such that} & A_0(x) \preceq -\left(\tau(\epsilon)\displaystyle\sum_{j=1}^{s} w_j y_j\right) U, & (a) \\
& \mathrm{Arrow}\left(y_j U, v_{i_1^j} A_{i_1^j}(x), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}(x)\right) \succeq \mathbf{0} \quad \text{for } j = 1, \ldots, s & (b)
\end{array}
\tag{5.2}
$$

for some $d \times d$ positive definite matrix U. Now, if we treat U as a decision variable in (5.2), then the constraint (5.2a) can be satisfied as equality. However, the constraint (5.2b) becomes *bilinear* in the decision variables y_j and U, which could cause computational difficulties. To circumvent this problem, we can apply an iterative procedure to tackle the constraints in (5.2). Specifically, consider the following family of optimization problems, parametrized by $U \succeq \mathbf{0}$ and $y \in \mathbb{R}^s$:

$$
g_{\{A_i\}_{i=0}^m}(U, y) = \inf \left\{ f(x) : x \in \Delta_{\{A_i\}_{i=0}^m}(U, y) \right\}, \tag{5.3}
$$

where

$$
\Delta_{\{A_i\}_{i=0}^m}(U, y) = \left\{ x \in X \;\middle|\; \begin{array}{l} A_0(x) \preceq -\left(\tau(\epsilon)\displaystyle\sum_{j=1}^{s} w_j y_j\right) U, \\ \mathrm{Arrow}\left(y_j U, v_{i_1^j} A_{i_1^j}(x), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}(x)\right) \succeq \mathbf{0} \\ \text{for } j = 1, \ldots, s \end{array} \right\}.
$$

To facilitate computation and analysis, we shall make the following assumptions concerning Problem (5.3):

1. (Tractability) Both the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the closed feasible set $X \subset \mathbb{R}^n$ are convex and efficiently computable, and that $A_0, A_1, \ldots, A_m : \mathbb{R}^n \to \mathcal{S}^d$ are affine functions of $x \in \mathbb{R}^n$ with $A_0(x) \prec \mathbf{0}$ for all $x \in X$.
2. (Feasibility) There exist $\bar{U} \succ \mathbf{0}$ and $\bar{y} \in \mathbb{R}^s$ such that $\Delta_{\{A_i\}_{i=0}^m}(\bar{U}, \bar{y}) \neq \emptyset$. Note that without loss of generality, we can take $\bar{U} = I$.

From our earlier discussion, we can view the problem

$$
\tau_0 = \inf_{I \succeq U \succ \mathbf{0},\; y \in \mathbb{R}^s} g_{\{A_i\}_{i=0}^m}(U, y) \tag{5.4}
$$

as an "optimized" safe tractable approximation of the following chance–constrained optimization problem:

$$
\begin{array}{ll}
\inf & f(x) \\
\text{subject to} & \Pr\limits_{\xi}\left(A_0(x) + \displaystyle\sum_{i=1}^{m} \xi_i A_i(x) \preceq \mathbf{0}\right) \geq 1 - \epsilon, \\
& x \in X.
\end{array}
\tag{5.5}
$$

Before proceeding further, let us note that the constraint $I \succeq U \succ \mathbf{0}$ in (5.4) can be replaced by $I \succeq U \succeq \mathbf{0}$ without changing the optimal value. Indeed, let $(\bar{x}, \bar{U}, \bar{y}) \in X \times \mathcal{S}_+^d \times \mathbb{R}^s$ be feasible for (5.4), and suppose that $\bar{U}$ is not positive definite. Since

$A_0(\bar{x}) \prec \mathbf{0}$ by assumption, the positive definite matrix $\bar{U}'$ obtained by replacing the zero eigenvalues of $\bar{U}$ with some small positive number and keeping other eigenvalues of $\bar{U}$ intact will satisfy $I \succeq \bar{U}' \succ \mathbf{0}$ and

$$A_0(\bar{x}) \preceq -\left(\tau(\epsilon)\sum_{j=1}^{s} w_j \bar{y}_j\right)\bar{U}'.$$

In addition, since $\bar{U}' = \bar{U} + (\bar{U}' - \bar{U})$ and $\bar{U}' - \bar{U} \succeq \mathbf{0}$ by construction, it is easy to verify that

$$\text{Arrow}\left(\bar{y}_j\bar{U}', v_{i_1^j}A_{i_1^j}(\bar{x}), \ldots, v_{i_{s_j}^j}A_{i_{s_j}^j}(\bar{x})\right) \succeq \mathbf{0} \quad \text{for } j = 1, \ldots, s.$$

Hence, the solution $(\bar{x}, \bar{U}', \bar{y})$ is also feasible for (5.4).

Now, consider the following iterative procedure for solving (5.4):

<u>Procedure ITER_IMPROVE</u>

1. (Initialization) Let $\eta > 0$ be a given accuracy parameter. Set $A_i^0(x) \longleftarrow A_i(x)$ for $i = 0, 1, \ldots, m$ and $\theta_0 \longleftarrow +\infty$.
2. For $t = 0, 1, \ldots$, do
   (a) Find x^t, y^t such that $x^t \in \Delta_{\{A_i^t\}_{i=0}^m}(I, y^t)$ and
   $$\inf_{y\in\mathbb{R}^s} g_{\{A_i^t\}_{i=0}^m}(I, y) \le f(x^t) \le \min\left\{\theta_t, \inf_{y\in\mathbb{R}^s} g_{\{A_i^t\}_{i=0}^m}(I, y) + \eta\right\}.$$
   (b) Find $\bar{x}^t, \bar{U}^t$ such that $\bar{x}^t \in \Delta_{\{A_i^t\}_{i=0}^m}(\bar{U}^t, y^t)$ and
   $$\inf_{I\succeq U\succeq\mathbf{0}} g_{\{A_i^t\}_{i=0}^m}(U, y^t) \le f(\bar{x}^t) \le \min\left\{f(x^t), \inf_{I\succeq U\succeq\mathbf{0}} g_{\{A_i^t\}_{i=0}^m}(U, y^t) + \eta\right\}.$$
   (c) If some convergence criterion is met, terminate and return $(\bar{x}^t, \bar{U}^t, y^t)$ as the feasible solution to (5.4). Otherwise, set
   $$\begin{aligned} A_i^{t+1}(x) &\longleftarrow (-A_0^t(\bar{x}^t))^{-1/2}A_i^t(x)(-A_0^t(\bar{x}^t))^{-1/2} \quad \text{for } i = 0, 1, \ldots, m, \\ \theta_{t+1} &\longleftarrow f(\bar{x}^t), \\ t &\longleftarrow t + 1. \end{aligned}$$

Intuitively, the above procedure attempts to solve the non–convex optimization problem (5.4) by alternately optimizing over the pair (x, y) (with U fixed) and (x, U) (with y fixed). Specifically, consider a particular iteration $t \ge 0$. In Step 2a, we fix $U = I$ and solve the optimization problem $\inf_{y\in\mathbb{R}^s} g_{\{A_i^t\}_{i=0}^m}(I, y)$. For the moment, let us assume that the optimal value is attained and the optimal solution is given by (x^t, y^t). Then, we can fix $y = y^t$ and consider a new optimization problem, namely, $\inf_{I\succeq U\succeq\mathbf{0}} g_{\{A_i^t\}_{i=0}^m}(U, y^t)$ in Step 2b. Note that (x^t, I) is feasible for this new problem. Suppose again that the optimal value is attained and the optimal solution is given by $(\bar{x}^t, \bar{U}^t)$. Then, we must have $f(\bar{x}^t) \le f(x^t)$. Now, the update rule in Step 2c can be shown to guarantee the existence of some $\bar{y}^t \in \mathbb{R}^s$ such that $(\bar{x}^t, \bar{y}^t)$ is feasible for the optimization problem $\inf_{y\in\mathbb{R}^s} g_{\{A_i^{t+1}\}_{i=0}^m}(I, y)$ in the next iteration. Consequently, if (x^{t+1}, y^{t+1}) is the optimal solution to this problem, then we must have

$f(x^{t+1}) \le f(\bar{x}^t)$. In summary, the above procedure will produce sequences of iterates $\{x^t\}$ and $\{\bar{x}^t\}$ whose objective values are *monotonically decreasing*:

$$f(x^{t+1}) \le f(\bar{x}^t) \le f(x^t) \quad \text{for } t = 0, 1, \ldots.$$

Moreover, since in each iteration of Procedure ITER_IMPROVE we are solving safe tractable approximations of the original chance–constrained optimization problem (5.5), the iterates $\{x^t\}$ and $\{\bar{x}^t\}$ are all *feasible* for (5.5). Hence, by applying Procedure ITER_IMPROVE, we can expect to get better solutions to Problem (5.5) than using the plain safe tractable approximation (4.2).

To turn the above intuition into rigorous statements, we need to make sure that the optimization problems formulated in each iteration are well defined, and to account for the fact that their optimal values may not be attainable. These are achieved in the following proposition.

PROPOSITION 5.1. *Suppose that Problem (5.4) is bounded below, i.e.,* $\tau_0 > -\infty$*. Then, the sequences* $\{x^t\}$*,* $\{y^t\}$*,* $\{\bar{x}^t\}$*,* $\{\bar{U}^t\}$ *and* $\{A_i^t(x)\}$ *specified in Procedure* ITER_IMPROVE *are well defined, and* $A_0^t(x) \prec \mathbf{0}$ *for all* $x \in X$ *and* $t \ge 0$*. In particular, we have*

$$\tau_t = \inf_{I \succeq U \succeq \mathbf{0},\ y \in \mathbb{R}^s} g_{\{A_i^t\}_{i=0}^m}(U, y) > -\infty \quad \text{and} \quad f(x^{t+1}) \le f(\bar{x}^t) \le f(x^t)$$

for all $t \ge 0$*. Moreover, all the steps in Procedure* ITER_IMPROVE *can be done efficiently.*

Proof. Recall from Step 1 that $A_i^0(x) = A_i(x)$ for $i = 0, 1, \ldots, m$, and that $A_0(x) \prec \mathbf{0}$ for all $x \in X$ by assumption. Thus, we shall prove by induction on $t \ge 0$ that in iteration t, the iterates x^t, y^t, $\bar{x}^t$, $\bar{U}^t$, $\{A_i^{t+1}\}_{i=0}^m$ are well defined with $A_0^{t+1}(x) \prec \mathbf{0}$ for all $x \in X$, and $\tau_t > -\infty$. Consider the base case $t = 0$. By assumption, we have $\tau_0 > -\infty$, and there exists some $\bar{y} \in \mathbb{R}^s$ such that $\Delta_{\{A_i^0\}_{i=0}^m}(I, \bar{y}) \ne \emptyset$. Hence, (x^0, y^0) is well defined. Moreover, since $x^0 \in \Delta_{\{A_i^0\}_{i=0}^m}(I, y^0)$, we have

$$\inf_{I \succeq U \succeq \mathbf{0}} g_{\{A_i^0\}_{i=0}^m}(U, y^0) \le f(x^0),$$

which together with $\tau_0 > -\infty$ implies that $(\bar{x}^0, \bar{U}^0)$ is also well defined and $f(\bar{x}^0) \le f(x^0)$. Finally, since $A_0(x) \prec \mathbf{0}$ for all $x \in X$ by assumption, we have $A_0^0(\bar{x}^0) = A_0(\bar{x}^0) \prec \mathbf{0}$, which together with Step 2c implies that $A_i^1(x)$ is well defined for $i = 0, 1, \ldots, m$, and that $A_0^1(x) \prec \mathbf{0}$ for all $x \in X$.

Now, consider a particular iteration $t > 0$. By the inductive hypothesis, the iterates $\bar{x}^{t-1}$, $\{A_i^{t-1}(x)\}_{i=0}^m$, $\{A_i^t(x)\}_{i=0}^m$ are well defined, and $\tau_{t-1} > -\infty$. By Step 2c, for any given $I \succeq U \succeq \mathbf{0}$ and $y \in \mathbb{R}^s$, we have

$$g_{\{A_i^t\}_{i=0}^m}(U, y) = g_{\{A_i^{t-1}\}_{i=0}^m}(\tilde{U}, \tilde{y}),$$

where

$$\tilde{U} = \frac{(-A_0^{t-1}(\bar{x}^{t-1}))^{1/2} U (-A_0^{t-1}(\bar{x}^{t-1}))^{1/2}}{\|(-A_0^{t-1}(\bar{x}^{t-1}))^{1/2}\|^2} \quad \text{and} \quad \tilde{y} = \|(-A_0^{t-1}(\bar{x}^{t-1}))^{1/2}\|^2 \cdot y.$$

In particular, we have $I \succeq \tilde{U} \succeq \mathbf{0}$ and $\tilde{y} \in \mathbb{R}^s$, which together with $\tau_{t-1} > -\infty$ implies that $\tau_t > -\infty$.

Next, we prove that (x^t, y^t) is well defined. By the inductive hypothesis, the iterates $\bar{U}^{t-1}$ and y^{t-1} are well defined. We claim that $\bar{x}^{t-1} \in \Delta_{\{A_i^t\}_{i=0}^m}(I, y^*)$ for some $y^* \in \mathbb{R}^s$. Note that this would imply the well–definedness of x^t and y^t, because we would have

$$\theta_t = f(\bar{x}^{t-1}) \geq \inf_{y \in \mathbb{R}^s} g_{\{A_i^t\}_{i=0}^m}(I, y) \geq \tau_t > -\infty$$

in Step 2a. It would also imply that $f(x^t) \leq f(\bar{x}^{t-1})$. To prove the claim, recall that $\bar{x}^{t-1} \in \Delta_{\{A_i^{t-1}\}_{i=0}^m}(\bar{U}^{t-1}, y^{t-1})$, which is equivalent to

$$A_0^{t-1}(\bar{x}^{t-1}) \preceq -\left(\tau(\epsilon)\sum_{j=1}^s w_j y_j^{t-1}\right)\bar{U}^{t-1}, \tag{5.6}$$

$$\mathrm{Arrow}\left(y_j^{t-1}\bar{U}^{t-1}, v_{i_1^j} A_{i_1^j}^{t-1}(\bar{x}^{t-1}), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}^{t-1}(\bar{x}^{t-1})\right) \succeq \mathbf{0} \quad \text{for } j = 1, \ldots, s. \tag{5.7}$$

In particular, by (5.7), we have $y^{t-1} \geq \mathbf{0}$. If $y^{t-1} = \mathbf{0}$, then $A_1^{t-1}(\bar{x}^{t-1}) = \cdots = A_m^{t-1}(\bar{x}^{t-1}) = \mathbf{0}$, which implies that we can take $y^* = \mathbf{0}$ and have $\bar{x}^{t-1} \in \Delta_{\{A_i^t\}_{i=0}^m}(I, \mathbf{0})$. Hence, we may assume that $y^{t-1} \neq \mathbf{0}$. In this case, define $y^* \in \mathbb{R}^s$ by

$$y_i^* = \frac{y_i^{t-1}}{\tau(\epsilon)\sum_{j=1}^s w_j y_j^{t-1}} \quad \text{for } i = 1, \ldots, s.$$

By definition, we have

$$\begin{aligned} A_0^t(\bar{x}^{t-1}) &= (-A_0^{t-1}(\bar{x}^{t-1}))^{-1/2} A_0^{t-1}(\bar{x}^{t-1})(-A_0^{t-1}(\bar{x}^{t-1}))^{-1/2} \\ &= -I \\ &= -\left(\tau(\epsilon)\sum_{j=1}^s w_j y_j^*\right) I. \end{aligned}$$

Now, consider a fixed $j \in \{1, \ldots, s\}$. Observe that

$$\mathrm{Arrow}\left(y_j^* I, v_{i_1^j} A_{i_1^j}^t(\bar{x}^{t-1}), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}^t(\bar{x}^{t-1})\right) \succeq \mathbf{0}$$

if and only if

$$\mathrm{Arrow}\left(y_j^*(-A_0^{t-1}(\bar{x}^{t-1})), v_{i_1^j} A_{i_1^j}^{t-1}(\bar{x}^{t-1}), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}^{t-1}(\bar{x}^{t-1})\right) \succeq \mathbf{0}. \tag{5.8}$$

If $y_j^{t-1} > 0$, then (5.6) is equivalent to $y_j^*(-A_0^{t-1}(\bar{x}^{t-1})) \succeq y_j^{t-1}\bar{U}^{t-1}$, which together with (5.7) implies that

$$\begin{aligned} &\mathrm{Arrow}\left(y_j^*(-A_0^{t-1}(\bar{x}^{t-1})), v_{i_1^j} A_{i_1^j}^{t-1}(\bar{x}^{t-1}), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}^{t-1}(\bar{x}^{t-1})\right) \\ \succeq\ &\mathrm{Arrow}\left(y_j^{t-1}\bar{U}^{t-1}, v_{i_1^j} A_{i_1^j}^{t-1}(\bar{x}^{t-1}), \ldots, v_{i_{s_j}^j} A_{i_{s_j}^j}^{t-1}(\bar{x}^{t-1})\right) \\ \succeq\ &\mathbf{0}. \end{aligned}$$

On the other hand, if $y_j^{t-1} = 0$, then $y_j^* = 0$. Moreover, by (5.7), we have $A_{i_1^j}^{t-1}(\bar{x}^{t-1}) = \cdots = A_{i_{s_j}^j}^{t-1}(\bar{x}^{t-1}) = \mathbf{0}$. This implies that (5.8) holds in this case as well, and hence $\bar{x}^{t-1} \in \Delta_{\{A_i^t\}_{i=0}^m}(I, y^*)$, as desired. As a corollary, we conclude that (x^t, y^t) is well defined.

Since $x^t \in \Delta_{\{A_i^t\}_{i=0}^m}(I, y^t)$, we have

$$\inf_{I \succeq U \succeq \mathbf{0}} g_{\{A_i^t\}_{i=0}^m}(U, y^t) \leq f(x^t),$$

which together with $\tau_t > -\infty$ implies that $(\bar{x}^t, \bar{U}^t)$ is well defined and $f(\bar{x}^t) \leq f(x^t)$. Moreover, since $A_0^t(x) \prec \mathbf{0}$ for all $x \in X$ by the inductive hypothesis, we see from Step 2c that $\{A_i^{t+1}(x)\}_{i=0}^m$ are well defined with $A_0^{t+1}(x) \prec \mathbf{0}$ for all $x \in X$.

Finally, observe that the optimization problems in Steps 2a and 2b have an efficiently computable objective function f and include only linear matrix inequalities and the efficiently computable set X as constraints. Thus, both steps can be done efficiently. Moreover, it is clear that Steps 1 and 2c can be done efficiently. This completes the proof. □

REMARKS.

1. The conclusion of Proposition 5.1 remains valid if we replace the accuracy parameter $\eta > 0$ in iteration t of Procedure ITER_IMPROVE by $\eta_t > 0$, where $\{\eta_t\}$ is a sequence that tends to zero.
2. Suppose that $\epsilon_0 > 0$ is the given tolerance level. In the discussion above, we simply assume that Procedure ITER_IMPROVE is run with $\epsilon = \epsilon_0$. However, by running Procedure ITER_IMPROVE with different values of ϵ, it is possible to further reduce the conservatism of the proposed safe tractable approximations. Indeed, we can perform a binary search on $[\epsilon_0, 1]$ to find the largest tolerance level ϵ_0' such that when Procedure ITER_IMPROVE is run with $\epsilon = \epsilon_0'$, the solution obtained will still satisfy the original tolerance level ϵ_0 with high confidence. Note that since $\epsilon_0' \geq \epsilon_0$, the objective value of the solution returned by Procedure ITER_IMPROVE when $\epsilon = \epsilon_0'$ will be no worse than that when $\epsilon = \epsilon_0$. For a more thorough treatment of the binary search scheme, we refer the reader to [45].

6. Computational Studies. To illustrate numerically the constructions developed in preceding sections, we apply them to the minimum–volume invariant ellipsoid problem in control theory and compare their performance with some existing methods. Before we present our computational results, let us state the problem and define its chance–constrained counterpart. Consider the following discrete–time controlled dynamical system (cf. [5, Exercise 4.76]):

$$\begin{aligned} x(t+1) &= Ax(t) + bu(t) \qquad\qquad \text{for } t = 0, 1, \ldots, \\ x(0) &= \bar{x}. \end{aligned}$$

Here, $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ are system specifications, $x(t) \in \mathbb{R}^n$ represents the state of the system at time t, $\bar{x} \in \mathbb{R}^n$ is the initial state, and $u(t) \in [-1, 1]$ is the control at time t. Naturally, one is interested in characterizing the trajectory $\{x(t) : t \geq 0\}$ of the dynamical system, so that its influence and stability can be determined. However, an exact characterization is often difficult, as it would depend on the system specifications A and b, as well as the control $u(t)$ at each time $t \geq 0$. Instead, one could find a simple region in $\mathbb{R}^n$ to capture the "stable" part of the trajectory. In general, the shape of such region is chosen so that it is simple enough to have an efficiently computable representation, yet expressive enough to accurately capture the dynamics of the system. One widely accepted choice is the ellipsoid, which has its roots in the notion of quadratic stability; see, e.g., [12, 10]. Specifically, consider an ellipsoid centered at the origin

$$E(Z) = \left\{ x \in \mathbb{R}^n : x^T Z x \le 1 \right\},$$

where $Z \succ \mathbf{0}$ is an $n \times n$ symmetric positive definite matrix. We say that $E(Z)$ is an *invariant ellipsoid* if $Ax \pm b \in E(Z)$ whenever $x \in E(Z)$. It is known that if $E(Z)$ is an invariant ellipsoid and $x(t) \in E(Z)$ for some $t \ge 0$, then $x(t') \in E(Z)$ for all $t' \ge t$ [5, Exercise 4.76]. In other words, once the system is in a state that belongs to the invariant ellipsoid, then the subsequent trajectory of the system will remain inside the invariant ellipsoid, regardless of the control. With this interpretation, it is natural to find an invariant ellipsoid for the given dynamical system that has the smallest volume. Towards that end, let us first recall the following result from [5, Exercise 4.76]:

FACT 6.1. (EXISTENCE AND CHARACTERIZATION OF INVARIANT ELLIPSOIDS) *Suppose that the vectors $b, Ab, \ldots, A^{n-1}b$ are linearly independent. Then, an invariant ellipsoid exists if and only if A is stable (i.e., $\|A\| < 1$). Moreover, the ellipsoid $E(Z)$ is invariant if and only if there exists a $\lambda \ge 0$ such that*

$$\begin{bmatrix} 1 - b^T Z b - \lambda & -b^T Z A \\ -A^T Z b & \lambda Z - A^T Z A \end{bmatrix} \succeq \mathbf{0}. \tag{6.1}$$

Fact 6.1 allows us to formulate the problem of finding the minimum–volume invariant ellipsoid as a bilinear semidefinite programming problem. To see this, recall that the volume of the ellipsoid $E(Z)$ is $\kappa_n (\det Z)^{-1/2}$, where κ_n is the volume of the n–dimensional unit Euclidean ball. Moreover, the function $Z \mapsto (\det Z)^{1/n}$ is concave in $Z \succeq \mathbf{0}$, and the constraints

$$y \le (\det Z)^{1/n}, \quad Z \succeq \mathbf{0}$$

can be expressed as linear matrix inequalities [5, pp. 149–150]. Thus, we have the following bilinear semidefinite programming formulation of the minimum–volume invariant ellipsoid problem (note that constraint (6.1) implies that $\lambda \in [0, 1]$):

$$\text{(MVIE)} \qquad \begin{array}{ll} \text{maximize} & y \\ \text{subject to} & y \le (\det Z)^{1/n}, \\ & \begin{bmatrix} 1 - b^T Z b - \lambda & -b^T Z A \\ -A^T Z b & \lambda Z - A^T Z A \end{bmatrix} \succeq \mathbf{0}, \\ & \lambda \in [0,1], Z \succeq \mathbf{0}. \end{array}$$

Although (MVIE) is difficult to solve in general, it can be approximated by solving a finite collection of semidefinite programming problems $\{(\mathsf{MVIE}(\lambda)) : \lambda \in \mathcal{D}\}$, where

$$
(\mathsf{MVIE}(\lambda)) \qquad \begin{array}{ll} \text{maximize} & y \\ \text{subject to} & y \leq (\det Z)^{1/n}, \\ & \begin{bmatrix} 1 - b^T Z b - \lambda & -b^T Z A \\ -A^T Z b & \lambda Z - A^T Z A \end{bmatrix} \succeq \mathbf{0}, \\ & Z \succeq \mathbf{0} \end{array}
$$

and $\mathcal{D} \subset [0,1]$ is a finite set (e.g., one can take $\mathcal{D} = \{0.00, 0.01, \ldots, 0.99, 1.00\}$). Specifically, we have the following numerical procedure for approximating Problem (MVIE):

<u>Procedure Approx–Nominal–Mvie</u>

1. For each $\lambda \in \mathcal{D}$, let $v_{\mathrm{nom}}(\lambda)$ be the optimal value of and $(y_{\mathrm{nom}}(\lambda), Z_{\mathrm{nom}}(\lambda))$ be the optimal solution to $(\mathsf{MVIE}(\lambda))$.
2. Return $E(Z_{\mathrm{nom}}(\lambda^*))$ as the approximating ellipsoid, where $\lambda^* = \arg\max_{\lambda \in \mathcal{D}} v_{\mathrm{nom}}(\lambda)$.

Note that in the above formulation, the system specifications A and b are assumed to be exactly known. However, it is conceivable that they are corrupted by some random noise. For the sake of simplicity, let us assume that only b is corrupted, and that it is given by

$$
b_i = \bar{b}_i + \rho \zeta_i \quad \text{for } i = 1, \ldots, n,
$$

where $\bar{b} \in \mathbb{R}^n$ is the nominal value of $b \in \mathbb{R}^n$, $\rho \geq 0$ is a fixed constant to control the level of perturbation, and $\zeta_1, \ldots, \zeta_n$ are i.i.d. real–valued mean–zero random variables of one of the following two types:

(B) ζ_i is supported on $[-1,1]$ with $\sigma^2 = 1/3$, for $i = 1, \ldots, n$.
(G) ζ_i is a standard Gaussian random variable, for $i = 1, \ldots, n$.

Under this setting, there is a natural chance–constrained version of the minimum–volume invariant ellipsoid problem, namely, to find a $Z \succ \mathbf{0}$ such that the ellipsoid $E(Z)$ is invariant with probability at least $1 - \epsilon$ and has the smallest volume, where $\epsilon > 0$ is a tolerance parameter. To tackle this problem, let us follow our earlier idea and consider the finite collection of chance–constrained semidefinite programs $\{(\mathsf{CCMVIE}(\lambda)) : \lambda \in \mathcal{D}\}$, where

$$
(\mathsf{CCMVIE}(\lambda)) \qquad \begin{array}{ll} \text{maximize} & y \\ \text{subject to} & y \leq (\det Z)^{1/n}, \\ & \Pr_{\zeta}\underbrace{\left(\begin{bmatrix} -1 + b^T Z b + \lambda & b^T Z A \\ A^T Z b & -\lambda Z + A^T Z A \end{bmatrix} \preceq \mathbf{0} \right)}_{A_0(\lambda,Z) + \sum_{i=1}^n \zeta_i A_i(\lambda,Z) + \sum_{1 \leq j \leq k \leq n} \zeta_j \zeta_k B_{jk}(\lambda,Z) \preceq \mathbf{0}} \geq 1 - \epsilon, \\ & Z \succeq \mathbf{0} \end{array}
$$

and $\mathcal{D} \subset [0,1]$ is a finite set. For each fixed $\lambda \in [0,1]$, we can apply the techniques developed in Section 5 to tackle Problem $(\mathsf{CCMVIE}(\lambda))$. This suggests the following numerical procedure for approximating the chance–constrained minimum–volume invariant ellipsoid problem:

<u>Procedure Approx–Ccmvie</u>

1. For each $\lambda \in \mathcal{D}$, let $v_{\mathrm{sta}}(\lambda)$ be the objective value and $(y_{\mathrm{sta}}(\lambda), Z_{\mathrm{sta}}(\lambda))$ be the corresponding solution obtained by applying Procedure Iter_Improve in Section 5 to $(\mathsf{CCMVIE}(\lambda))$.
2. Return $E(Z_{\mathrm{sta}}(\lambda^*))$ as the approximating ellipsoid, where $\lambda^* = \arg\max_{\lambda\in\mathcal{D}} v_{\mathrm{sta}}(\lambda)$.

Alternatively, one can use Monte Carlo sampling to tackle $(\mathsf{CCMVIE}(\lambda))$. To the best of our knowledge, this is the only other approach in the literature for processing chance–constrained linear matrix inequalities with quadratic perturbations. In this approach, one samples N i.i.d. copies $\zeta^{(1)}, \ldots, \zeta^{(N)}$ of the random vector $\zeta = (\zeta_1, \ldots, \zeta_n)$ and construct the so–called *scenario program*

$$(\mathsf{MCMVIE}_N(\lambda)) \qquad \begin{array}{ll} \text{maximize} & y \\ \text{subject to} & y \le (\det Z)^{1/n}, \\ & A_0(\lambda, Z) + \displaystyle\sum_{i=1}^{n} \zeta_i^{(l)} A_i(\lambda, Z) \\ & \quad + \displaystyle\sum_{1\le j\le k\le n} \zeta_j^{(l)} \zeta_k^{(l)} B_{jk}(\lambda, Z) \preceq \mathbf{0} \quad \text{for } l = 1, \ldots, N, \\ & Z \succeq \mathbf{0}. \end{array}$$

It can be shown that when

$$N \ge \left\lceil \frac{1}{\epsilon}\left(L - 1 + \ln\frac{1}{\delta} + \sqrt{2(L-1)\ln\frac{1}{\delta} + \ln^2\frac{1}{\delta}} \right) \right\rceil, \tag{6.2}$$

where $L = n(n+1)/2$ is the number of decision variables and $\delta \in (0, 1)$ is a confidence parameter, the optimal solution to $(\mathsf{MCMVIE}_N(\lambda))$ will be feasible for $(\mathsf{CCMVIE}(\lambda))$ with probability at least $1 - \delta$; cf. [17]. This yields the following alternative numerical procedure for approximating the chance–constrained minimum–volume invariant ellipsoid problem:

<u>Procedure Approx–Mcmvie</u>

1. Choose $\delta \in (0, 1)$ and N such that (6.2) holds. Generate N i.i.d. copies of ζ.
2. For each $\lambda \in \mathcal{D}$, let $v_{\mathrm{mc}}(\lambda)$ be the optimal value of and $(y_{\mathrm{mc}}(\lambda), Z_{\mathrm{mc}}(\lambda))$ be the optimal solution to the sampled problem $(\mathsf{MCMVIE}_N(\lambda))$.
3. Return $E(Z_{\mathrm{mc}}(\lambda^*))$ as the approximating ellipsoid, where $\lambda^* = \arg\max_{\lambda\in\mathcal{D}} v_{\mathrm{mc}}(\lambda)$.

To compare the above procedures through numerical experiments, we proceed as follows. We set $\mathcal{D} = \{0.00, 0.01, \ldots, 0.99, 1.00\}$ and assume that $\zeta_1, \ldots, \zeta_n$ are i.i.d. real–valued mean–zero random variables of either type (B) or (G). Furthermore, we run the iterative procedure in Section 5 using $|f(\bar{x}^t)/f(\bar{x}^{t-1}) - 1| \le 10^{-4}$ as our convergence criterion. All experiments are run under the Matlab R2011a environment on a Windows® 7 operating system with Intel® Core™2 6700@2.66GHz and 2GB of RAM. The computations are performed using `CVX` version 1.21, a package for specifying and solving convex programs [29].

We first consider the following instance:

$$A = \begin{bmatrix} -0.8147 & -0.4163 \\ 0.8167 & -0.1853 \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} 1 \\ 0.7071 \end{bmatrix}. \tag{6.3}$$

Table 6.1 shows the performance of various procedures when $\epsilon = 0.05$, $\rho = 0.01$ and $\delta = 0.05$, while Figure 6.1 shows the ellipsoids obtained by those procedures.

To compare the sizes of different ellipsoids, we use the *average linear size* measure, which is defined as $\mathrm{ALS}(E(Z)) = (\mathrm{Vol}_n(E(Z)))^{1/n}$; see [5, pp. 268] for the motivation of using such a measure. As can be seen from the table, the average linear sizes of the ellipsoids obtained by the stated procedures are all very close to each other. Moreover, the average runtime (averaged over the $|\mathcal{D}| = 101$ iterations needed to find λ^*) of the safe tractable approximation approach is less than that of the Monte Carlo sampling approach. This demonstrates the advantage of our proposed safe tractable approximations.

	Nominal	Sta–B	Mc–B	Sta–G	Mc–G
avg. line size	4.0221	4.1464	4.0654	4.1477	4.1405
λ^*	0.71	0.71	0.71	0.71	0.71
avg. runtime (sec)	0.1716	1.7798	4.8665	1.7089	4.8614

Table 6.1

Performance of various procedures when applied to the problem instance (6.3), with $\epsilon = 0.05$, $\rho = 0.01$ and $\delta = 0.05$. (I) Nominal*: Procedure* Approx–Nominal–Mvie*. (II)* Sta–B*: Procedure* Approx–Ccmvie *with* Iter_Improve*, using the safe tractable approximation for bounded perturbations (Theorem 4.4). (III)* Mc–B*: Procedure* Approx–Mcmvie*, assuming that each ζ_i follows the uniform distribution on $[-1, 1]$. (IV)* Sta–G*: Procedure* Approx–Ccmvie *with* Iter_Improve*, using the safe tractable approximation for Gaussian perturbations (Theorem 4.6). (V)* Mc–G*: Procedure* Approx–Mcmvie*, assuming that each ζ_i follows the standard Gaussian distribution.*

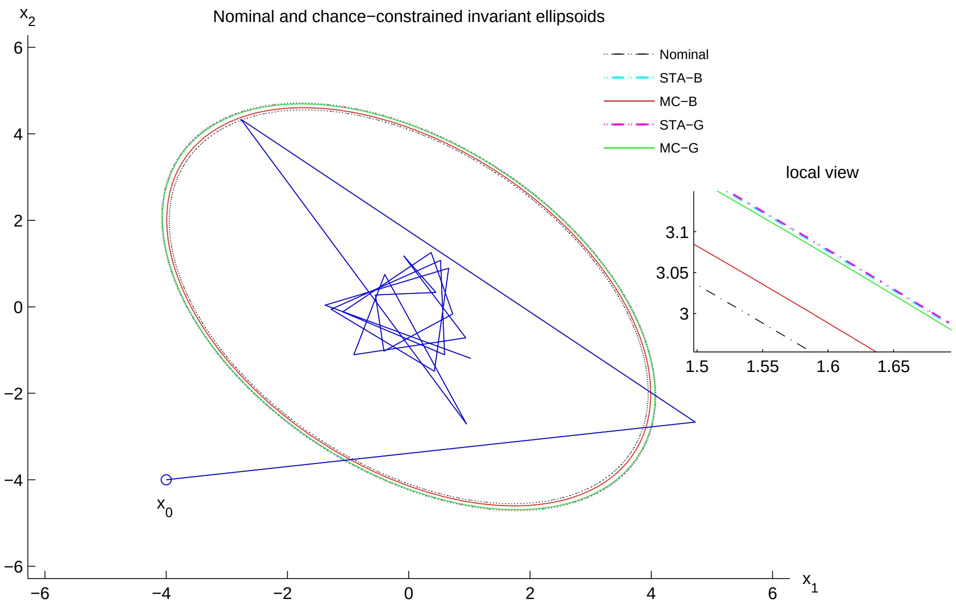

Fig. 6.1. *Nominal and chance–constrained invariant ellipsoids obtained by various procedures for the problem instance (6.3), with $\epsilon = 0.05$, $\rho = 0.01$ and $\delta = 0.05$.*

Figure 6.2 shows the ellipsoids obtained by Sta–B for different values of ϵ, with $\rho = 0.01$. For the case where $\epsilon = 0.001$ or $\epsilon = 0.0001$, our machine ran out of memory when we ran the Monte Carlo sampling approach. By contrast, the complexity of our safe tractable approximation approach does not vary with ϵ. Figure 6.3 shows the ellipsoids obtained by Sta–B for different values of ρ, with $\epsilon = 0.05$.

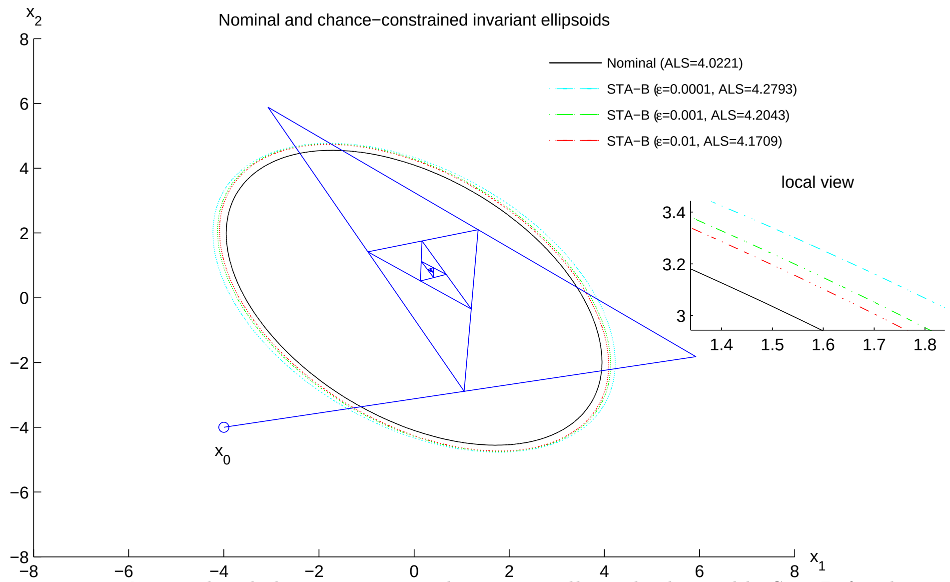

FIG. 6.2. *Nominal and chance–constrained invariant ellipsoids obtained by* STA–B *for the problem instance (6.3), with* $\rho = 0.01$.

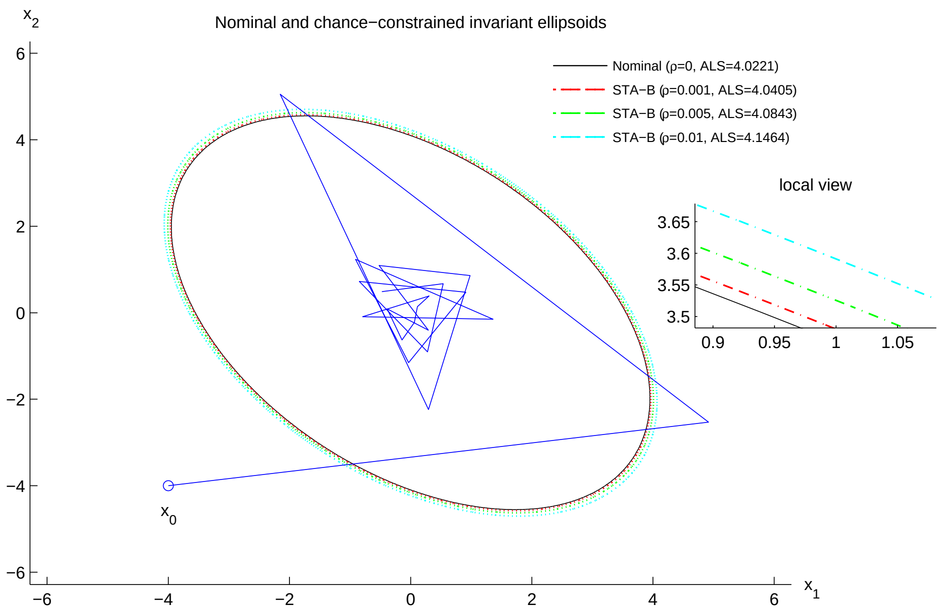

FIG. 6.3. *Nominal and chance–constrained invariant ellipsoids obtained by* STA–B *for the problem instance (6.3), with* $\epsilon = 0.05$.

To test the performance of our proposed procedures on higher dimensional problems, we next consider the following instance:

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0.0028 & 0.0142 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -0.0825 & -0.4126 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} 0 \\ 0.0076 \\ 0 \\ -0.1676 \\ 0 \end{bmatrix}. \tag{6.4}$$

Table 6.2 shows the computational results obtained by our approach and the Monte Carlo approach when $\epsilon = 0.03$, $\rho = 0.01$ and $\delta = 0.05$. We observe that the average linear sizes of the ellipsoids obtained by the two approaches are comparable. However, our approach has a much more favorable average runtime when compared with the Monte Carlo approach. Moreover, it is worth noting that our machine already ran out of memory when the Monte Carlo approach is run for the case where $\epsilon = 0.01$.

	Nominal	Sta–B	Mc–B	Sta–G	Mc–G
avg. line size	0.01038	0.0636	0.0444	0.0955	0.0710
λ^*	0.46	0.42	0.44	0.48	0.45
avg. runtime (sec)	1.1232	11.2503	59.1757	25.9987	53.9058

Table 6.2

Performance of various procedures when applied to the problem instance (6.4), with $\epsilon = 0.03$, $\rho = 0.01$ *and* $\delta = 0.05$. *Please refer to Table 6.1 for descriptions of the listed procedures.*

7. Conclusion. In this paper, we developed safe tractable approximations of chance–constrained linear matrix inequalities with dependent perturbations, where the only information available about the dependence structure is a list of independence relations. An advantage of our safe tractable approximations is that they can be expressed as systems of linear matrix inequalities and hence can be efficiently solved using standard packages. As a crucial initial step of our construction, we proved a large deviation bound for sums of dependent random matrices, which may be of independent interest. Our work is a first attempt to develop a general framework for processing chance–constrained linear matrix inequalities with dependent perturbations. As such, many questions remain. For instance, how conservative are our safe tractable approximations? Is it possible to exploit further the Ahlswede–Winter inequality (see Fact 3.3) to develop better approximations? Another direction is to identify conditions on the random perturbations that could lead to *exact* and efficient reformulations of the chance–constrained linear matrix inequalities considered in this paper. Finally, it would be interesting to find other practical settings to which our results apply.

Acknowledgments. The authors would like to thank the Associate Editor and the referees for their invaluable comments and suggestions. Most of this work was done when the first author was an M. Phil. student in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong.

REFERENCES

[1] R. Ahlswede and A. Winter. Strong Converse for Identification via Quantum Channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[2] I. Bechar. A Bernstein–Type Inequality for Stochastic Processes of Quadratic Forms of Gaussian Variables. Preprint, available at `http://arxiv.org/abs/0909.3595`, 2009.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

[4] A. Ben-Tal and A. Nemirovski. Robust Solutions of Linear Programming Problems Contaminated with Uncertain Data. *Mathematical Programming, Series A*, 88(3):411–424, 2000.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MPS–SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2001.

[6] A. Ben-Tal and A. Nemirovski. On Safe Tractable Approximations of Chance–Constrained Linear Matrix Inequalities. *Mathematics of Operations Research*, 34(1):1–25, 2009.

[7] D. Bertsimas and D. B. Brown. Constructing Uncertainty Sets for Robust Linear Optimization. *Operations Research*, 57(6):1483–1495, 2009.

[8] D. Bertsimas and M. Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004.

[9] D. Bertsimas and M. Sim. Tractable Approximations to Robust Conic Optimization Problems. *Mathematical Programming, Series B*, 107(1–2):5–36, 2006.

[10] F. Blanchini. Set Invariance in Control. *Automatica*, 35(11):1747–1767, 1999.

[11] B. Bollobás. *Modern Graph Theory*, volume 184 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 1998.

[12] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *SIAM Studies in Applied and Numerical Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1994.

[13] G. Calafiore and M. C. Campi. Uncertain Convex Programs: Randomized Solutions and Confidence Levels. *Mathematical Programming, Series A*, 102(1):25–46, 2005.

[14] G. C. Calafiore and M. C. Campi. The Scenario Approach to Robust Control Design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[15] G. C. Calafiore and L. El Ghaoui. On Distributionally Robust Chance–Constrained Linear Programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[16] M. C. Campi and G. C. Calafiore. Notes on the Scenario Design Approach. *IEEE Transactions on Automatic Control*, 54(2):382–385, 2009.

[17] M. C. Campi and S. Garatti. The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[18] A. Charnes and W. W. Cooper. Chance–Constrained Programming. *Management Science*, 6(1):73–79, 1959.

[19] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management Science*, 4(3):253–263, 1958.

[20] W. Chen and M. Sim. Goal–Driven Optimization. *Operations Research*, 57(2):342–357, 2009.

[21] W. Chen, M. Sim, J. Sun, and C.-P. Teo. From CVaR to Uncertainty Set: Implications in Joint Chance–Constrained Optimization. *Operations Research*, 58(2):470–485, 2010.

[22] X. Chen, M. Sim, and P. Sun. A Robust Optimization Perspective on Stochastic Programming. *Operations Research*, 55(6):1058–1071, 2007.

[23] E. Delage and Y. Ye. Distributionally Robust Optimization under Moment Uncertainty with Application to Data–Driven Problems. *Operations Research*, 58(3):595–612, 2010.

[24] L. El Ghaoui, M. Oks, and F. Oustry. Worst–Case Value–at–Risk and Robust Portfolio Optimization: A Conic Programming Approach. *Operations Research*, 51(4):543–556, 2003.

[25] E. Erdoğan and G. Iyengar. Ambiguous Chance Constrained Problems and Robust Optimization. *Mathematical Programming, Series B*, 107(1–2):37–61, 2006.

[26] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter, Berlin, Germany, second revised and extended edition, 2004.

[27] D. J. H. Garling. *Inequalities: A Journey into Linear Analysis*. Cambridge University Press, New York, 2007.

[28] S. Golden. Lower Bounds for the Helmholtz Function. *Physical Review*, 137(4B):B1127–B1128, 1965.

[29] M. Grant and S. Boyd. CVX: Matlab Software for Disciplined Convex Programming. `http://cvxr.com/cvx`, 2011.

[30] F. Hansen and G. K. Pedersen. Jensen's Operator Inequality. *Bulletin of the London Mathematical Society*, 35(4):553–564, 2003.

[31] D. L. Hanson. Some Results Relating Moment Generating Functions and Convergence Rates in the Law of Large Numbers. *The Annals of Mathematical Statistics*, 38(3):742–750, 1967.

[32] R. Henrion and C. Strugarek. Convexity of Chance Constraints with Independent Random Variables. *Computational Optimization and Applications*, 41(2):263–276, 2008.

[33] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the*

American Statistical Association, 58(301):13–30, 1963.

[34] S. Janson. Large Deviations for Sums of Partly Dependent Random Variables. *Random Structures and Algorithms*, 24(3):234–248, 2004.

[35] S. Kataoka. A Stochastic Programming Model. *Econometrica*, 31(1–2):181–196, 1963.

[36] C. M. Lagoa. On the Convexity of Probabilistically Constrained Linear Programs. In *Proceedings of the 38th IEEE Conference on Decision and Control*, pages 516–521, 1999.

[37] C. M. Lagoa, X. Li, and M. Sznaier. Probabilistically Constrained Linear Programs and Risk–Adjusted Controller Design. *SIAM Journal on Optimization*, 15(3):938–951, 2005.

[38] Q. Li, W.-K. Ma, and A. M.-C. So. Safe Convex Approximation to Outage–Based MISO Secrecy Rate Optimization under Imperfect CSI and with Artificial Noise. In *Proceedings of the 45th Annual Asilomar Conference on Signals, Systems, and Computers*, pages 207–211, 2011.

[39] J. Luedtke and S. Ahmed. A Sample Approximation Approach for Optimization with Probabilistic Constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[40] C. Lund and M. Yannakakis. On the Hardness of Approximating Minimization Problems. *Journal of the ACM*, 41(5):960–981, 1994.

[41] B. L. Miller and H. M. Wagner. Chance Constrained Programming with Joint Constraints. *Operations Research*, 13(6):930–945, 1965.

[42] K. Natarajan, D. Pachamanova, and M. Sim. Constructing Risk Measures from Uncertainty Sets. *Operations Research*, 57(5):1129–1141, 2009.

[43] A. Nemirovski. On Tractable Approximations of Randomly Perturbed Convex Constraints. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 2419–2422, 2003.

[44] A. Nemirovski. Sums of Random Symmetric Matrices and Quadratic Optimization under Orthogonality Constraints. *Mathematical Programming, Series B*, 109(2–3):283–317, 2007.

[45] A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[46] R. I. Oliveira. Sums of Random Hermitian Matrices and An Inequality by Rudelson. *Electronic Communications in Probability*, 15:203–212, 2010.

[47] D. Petz. A Survey of Certain Trace Inequalities. In J. Zemánek, editor, *Functional Analysis and Operator Theory*, volume 30 of *Banach Center Publications*, pages 287–298. Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland, 1994.

[48] J. Pintér. Deterministic Approximations of Probability Inequalities. *Mathematical Methods of Operations Research*, 33(4):219–239, 1989.

[49] A. Prékopa. Logarithmic Concave Measures with Application to Stochastic Programming. *Acta Scientiarum Mathematicarum*, 32(3–4):301–316, 1971.

[50] A. Prékopa. Contributions to the Theory of Stochastic Programming. *Mathematical Programming*, 4(1):202–221, 1973.

[51] A. Prékopa. Programming under Probabilistic Constraints with a Random Technology Matrix. *Mathematische Operationsforschung und Statistik, Series Optimization*, 5(2):109–116, 1974.

[52] A. Prékopa. *Stochastic Programming*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[53] A. Prékopa, K. Yoda, and M. M. Subasi. Uniform Quasi–Convexity in Probabilistic Constrained Stochastic Programming. *Operations Research Letters*, 39(3):188–192, 2011.

[54] C. W. Scherer and C. W. J. Hol. Matrix Sum–of–Squares Relaxations for Robust Semi–Definite Programs. *Mathematical Programming, Series B*, 107(1–2):189–211, 2006.

[55] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MPS–SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2009.

[56] A. M.-C. So. Moment Inequalities for Sums of Random Matrices and Their Applications in Optimization. *Mathematical Programming, Series A*, 130(1):125–151, 2011.

[57] R. Tempo, G. Calafiore, and F. Dabbene. *Randomized Algorithms for Analysis and Control of Uncertain Systems*. Communications and Control Engineering. Springer–Verlag, London, 2005.

[58] C. J. Thompson. Inequality with Applications in Statistical Mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813, 1965.

[59] J. A. Tropp. User–Friendly Tail Bounds for Sums of Random Matrices. Accepted for publication in *Foundations of Computational Mathematics*, 2011.

[60] C. van de Panne and W. Popp. Minimum–Cost Cattle Feed under Probabilistic Protein Constraints. *Management Science*, 9(3):405–430, 1963.

[61] K.-Y. Wang, T.-H. Chang, W.-K. Ma, A. M.-C. So, and C.-Y. Chi. Probabilistic SINR Con-

strained Robust Transmit Beamforming: A Bernstein–Type Inequality Based Conservative Approach. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, pages 3080–3083, 2011.

[62] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi. Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks: Tractable Approximations by Conic Optimization. Preprint, available at `http://www.se.cuhk.edu.hk/~manchoso/papers/rbf_rar.pdf`, 2011.

[63] R. W. Yeung. *Information Theory and Network Coding*. Springer Science+Business Media, LLC, New York, 2008.

[64] S. Zymler, D. Kuhn, and B. Rustem. Worst–Case Value–at–Risk of Non–Linear Portfolios. Preprint, available at `http://www.optimization-online.org/DB_HTML/2009/08/2379.html`, 2009.

[65] S. Zymler, D. Kuhn, and B. Rustem. Distributionally Robust Joint Chance Constraints with Second–Order Moment Information. Accepted for publication in *Mathematical Programming*, 2011.