Submitted to *Mathematics of Operations Research*
manuscript (Please, provide the mansucript number!)

Optimal Information Blending with Measurements in the L2 Sphere

Boris Defourny
Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544,
defourny@princeton.edu, http://www.princeton.edu/~defourny

Ilya O. Ryzhov
Department of Decision, Operations and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742,
iryzhov@rhsmith.umd.edu, http://www.rhsmith.umd.edu/faculty/iryzhov

Warren B. Powell
Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544,
powell@princeton.edu, http://castlelab.princeton.edu/

We consider a sequential information collection problem where a risk-averse decision-maker updates a Bayesian belief about the unknown objective function of a linear program. The information is collected in the form of a linear combination of the objective coefficients, subject to random noise. We have the ability to choose the weights in the linear combination, creating a new, nonconvex continuous optimization problem, which we refer to as information blending. We develop two optimal blending strategies: an active learning method that maximizes uncertainty reduction, and an economic approach that maximizes an expected improvement criterion. Semidefinite programming relaxations are used to create efficient convex approximations to the nonconvex blending problem.

Key words: Stochastic programming; semidefinite programming; value of information; Markov decision process; risk aversion
MSC2000 subject classification: Primary: 90C22; secondary: 90C40
OR/MS subject classification: Primary: Programming, stochastic
History:

1. Introduction. Consider planning problems that can be reformulated as linear programs (LPs) in standard form:

$$\text{maximize } c^\top x \quad \text{subject to } Ax = b,\ x \succeq 0\ . \tag{1}$$

Suppose, however, that the vector of objective coefficients is unknown, and is modeled as a random vector following some multivariate probability distribution. Problems where c is random are well studied in the field of stochastic optimization, covering applications such as production problems with unknown profit margins, or logistics and network problems with uncertain costs. The model with random coefficients can also be applied to characterize the optimal policy solving a finite-state Markov decision process (MDP) [34], where randomness in c corresponds to the situation of a one-period reward function that is not perfectly known.

There are several standard techniques for converting (1) with random c into a well-defined and tractable optimization problem. For instance, the problem max $\mathbb{E}\{c^\top x\}$ over $x \in \mathcal{X}$, that is,

$$\max_{x\in\mathcal{X}} \bar{c}^\top x\,, \quad \text{where } \mathcal{X} = \{x \in \mathbb{R}^n : Ax = b, x \succeq 0\}\ , \quad \bar{c} = \mathbb{E}\{c\}\ , \tag{2}$$

optimizes the original objective function $c^\top x$ in expectation. This approach may be reasonable when the decision-maker is risk-neutral with respect to performance. However, in many applications, the decision-maker is likely to be risk-averse, preferring to sacrifice some performance on average while hedging against worst-case scenarios. In such situations, robust optimization [4] offers a way to obtain computationally tractable, conservative decisions. Typically, one would infer a bounded uncertainty set $\mathcal{C}$ with good geometric properties from the distribution of c, and then optimize the worst-case bilinear objective $\max_{x\in\mathcal{X}} \min_{c\in\mathcal{C}} c^\top x$. For instance, suppose that c follows a multivariate normal distribution with covariance matrix Σ, an assumption that will hold throughout the paper:

$$c \sim \mathcal{N}(\bar{c}, \Sigma)\ . \tag{3}$$

As we show in this paper, under some assumptions on the choice of $\mathcal{C}$, we can reformulate the worst-case maximization as the second-order cone program (SOCP) [1]

$$\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}, \tag{4}$$

for some $\alpha > 0$. This problem is polynomially solvable to a fixed precision by interior-point methods. Note that, if $\alpha = 0$, the robust formulation (4) reduces to the risk-neutral formulation (2).

It is possible to reinterpret the distribution in (3) as a Bayesian prior representing the decision-maker's subjective beliefs about c. To emphasize this interpretation, we use the notation c^{true} to represent the unknown vector of objective coefficients, indicating that the decision-maker wishes to estimate some fixed but unknown "true" value. The multivariate normal prior $\mathcal{N}(\bar{c}, \Sigma)$ is a convenient way to incorporate correlations in the decision-maker's beliefs about the unknown c^{true}. Correlations reflect a belief about the similarity of the unknown coefficients. For example, the unknown profit margins for two similar products can reasonably be assumed to be correlated.

With this interpretation, we consider situations where the decision-maker has the ability to collect additional information about c^{true} before implementing a solution $x \in \mathcal{X}$ in production. A single piece of information about c^{true} will change the parameters of the belief distribution (3), thus changing the optimal solution of (4). Simply put, the uncertainty set from which we draw the worst-case scenario is now itself subject to change. Moreover, if we have multiple opportunities to collect information, we face a new problem of optimal multi-stage information collection. In this problem, the goal is to guide the evolution of the uncertainty set in a way that improves the performance of the robust solution to

A single piece of information about c^{true} will change the parameters of the belief distribution (3), thus changing the optimal solution of (4). Simply put, the uncertainty set from which we draw the worst-case scenario is now itself subject to change. Moreover, if we have multiple opportunities to collect information, we face a new problem of optimal multi-stage information collection. In this problem, the goal is to guide the evolution of the uncertainty set in a way that improves the performance of the robust solution to

$$v_\alpha(\bar{c}, \Sigma) = \max_{x:\, Ax=b,\, x\succeq 0} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}. \tag{5}$$

We seek to develop sequential, adaptive information collection policies with the ability to learn from the outcomes of previous observations.

The work by [41] studies the information collection problem in the context of the risk-neutral decision made in (2), under the assumption that the decision-maker collects information in the form of scalar noisy measurements of individual objective coefficients c_j^{true}. The present study adds the dimension of risk-averse decision-making, and makes the following major generalization: we study systems where information on the full vector $c^{\text{true}} \in \mathbb{R}^n$ can be acquired by observing

$$y = u^\top c^{\text{true}} + w, \tag{6}$$

where $u \in \mathbb{R}^n$ is a *measurement vector* chosen in the ball $\mathbb{B} = \{u \in \mathbb{R}^n : ||u||_2 \leq 1\}$, $w \sim \mathcal{N}(0, \sigma_w^2)$ is an independent Gaussian noise of variance $\sigma_w^2 > 0$, and y is the noisy observation that depends on the measurement vector. Given the observation y, by Bayesian updating, c follows the posterior distribution $\mathcal{N}(\bar{c}', \Sigma')$ given by

$$\bar{c}' = \bar{c} + \frac{\Sigma u}{u^\top \Sigma u + \sigma_w^2}(y - \bar{c}^\top u), \tag{7}$$

$$\Sigma' = \Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top \Sigma u + \sigma_w^2} \quad , \tag{8}$$

and the optimal value $v_\alpha(\bar{c}, \Sigma)$ in (5) is updated to $v_\alpha(\bar{c}', \Sigma')$. Instead of requiring measurements of individual objective coefficients, which can be done by restricting $u = e_j$ where e_j is the j-th unit vector in $\mathbb{R}^n$, we allow a "blended" observation that provides information about multiple unknown values simultaneously. To motivate this approach, suppose that u is a feasible solution in $\mathcal{X}$; that is, the decision-maker can only collect information about c^{true} by implementing a particular feasible solution of the LP and observing the outcome.

In this paper, we develop policies for choosing u that are optimal with respect to various criteria. First, we analytically derive a policy that achieves the optimal rate of uncertainty reduction in our beliefs about c^{true}. We show that this policy chooses u to be a dominant eigenvector of the posterior covariance matrix of c at each time step. Second, we develop a different policy that trades uncertainty reduction against the performance of the robust solution in (5) using the expected improvement criterion:

$$\mathbb{K}_\alpha(u, \bar{c}, \Sigma) = \mathbb{E}_y\{v_\alpha(\bar{c}', \Sigma') \mid u, \bar{c}, \Sigma\} - v_\alpha(\bar{c}, \Sigma), \tag{9}$$

$$u^* \in \arg\max_{u \in \mathbb{B}} \mathbb{K}_\alpha(u, \bar{c}, \Sigma). \tag{10}$$

The problem (10) takes the nonlinear update equations (7,8) into account. Inside the expectation, there is a change of optimal x for each outcome y, so as to obtain $v_\alpha(\bar{c}', \Sigma')$.

Although (10) defines a nonconvex optimization problem, we develop computationally tractable convex relaxations that reformulate (10) as a semidefinite program. We then show numerically that the SDP relaxation enables us to find directions u that achieve higher expected improvement than the unit-vector policy of [41]. The key theoretical insight of these results is that the information content of a scalar observation can be significantly improved by optimally blending information, instead of observing individual problem parameters.

The paper is organized as follows. Section 2 discusses related work. Section 3 derives the robust objective (5) from the definition of uncertainty sets for c. Section 4 applies the framework to Markov decision processes. Section 5 establishes properties of optimal solutions for the measurement selection problem (10). Section 6 studies the measurement policy that maximizes the rate of uncertainty reduction. Section 7 presents the main results of the paper on the optimization of (10). Section 8 presents numerical work, and Section 9 concludes.

2. Context and related work. The present paper relates primarily to two different streams of literature. The first is the literature on statistical learning and sequential information collection, usually known in different communities by the name of a particular problem. Examples include ranking and selection in simulation [27], multi-armed bandits in applied probability [20] and computer science [2], and global optimization [26]. This paper is closest to the simulation perspective, in which the information collection process ("ranking") is usually separated from the final implementation decision ("selection"). We also assume that the decision-maker first collects a number of observations of the form (6), before committing to a solution of the problem (2) or (4). The implementation decision in ranking and selection typically consists of choosing the largest value in a finite set; by contrast, our model is closer to [40, 41], where the ranking and selection framework is generalized to allow implementation decisions that optimize a mathematical program with unknown parameters. We also adopt the Bayesian framework for information collection; see [7] or [33] for a survey of Bayesian learning methods in simulation optimization.

The second major stream of literature is the work on robust optimization [4, 6]. Robust solutions to linear programs with uncertainty have been extensively studied [5], and the theory of robust optimization has also been developed for Markov decision processes [30, 25, 36]. See also [38] for recent work connecting the robust solution and the uncertainty set to a risk measure chosen by the decision-maker. Particularly relevant to the present paper is [11], which derived an expression of the form (5) applied specifically to MDPs. However, the notion that sequential information collection may change the uncertainty set over time, thus also changing the robust solution, has received much less attention. To give an example, equation (9) for measurement selection in robust MDPs was previously stated in [10] for $u \in \{e_1, \ldots, e_n\}$; however, the computational approach in this study was based on an approximation that did not take into account the change of the optimal solution from $\arg\max_x v_\alpha(\bar{c}, \Sigma)$ to $\arg\max_x v_\alpha(\bar{c}', \Sigma')$. In Section 4, we discuss this approach in more detail, in the perspective of contrasting it with our new results, which use SDP relaxations to approximate (9) more closely, while also allowing information blending.

Previous work in sequential learning has generally assumed that we always collect scalar observations of *individual* unknown parameters, even when such observations can be used to learn about a set of parameters [17, 35]. In [29], the unknown values of a finite set of alternatives are expressed as a linear combination of parameters via a linear regression model, producing the same Bayesian update as in (7-8). However, in this case, the vector u is pre-specified by the regression features. To our knowledge, the continuous optimization problem of choosing an *optimal* u is completely new. We use the term "information blending" to describe this new type of decision.

We choose the information blend u in two ways. Our first policy maximizes the rate of uncertainty reduction achieved by each measurement. This approach is along the lines of active learning in statistics [9], where the objective is to minimize uncertainty (i.e. improve the accuracy of a statistical model), with no regard for the economic value of a set of estimates. Conversely, our second policy is based on the expected improvement criterion, previously developed by [26] for global optimization and [24] for ranking and selection. This approach, also known by the names "value of information" [7] or "knowledge gradient" [18], provides an economic valuation of information in terms of the average improvement contributed by a single measurement to the optimal value of (2) or (4). This computation balances the expected value of the current solution to (2) or (4) against the decision-maker's uncertainty about that solution (and therefore the potential to improve it).

In the simulation literature, the decision-maker is almost always assumed to be risk-neutral [8], and the expected improvement criterion is defined in terms of the risk-neutral problem (2). Recently, however, there has been some interest in integrating concepts of risk-aversion and robust optimization into simulation optimization [43, 12]. To our knowledge, the work by [39] is the first to formally link ranking and selection with robust optimization. This work provides a theoretical justification for using the expected improvement criterion to learn about the risk-averse problem (4).

Essentially, a brief experiment on a possible solution is less expensive than a final implementation of that solution; thus, the decision-maker is assumed to be risk-neutral with respect to the measurement decision, but risk-averse with respect to the implementation decision. The present paper also adopts this approach, and the formulation in (5) covers both the risk-neutral and risk-averse cases.

To summarize, our work contributes to the literature on sequential learning as well as robust optimization. We show how two types of optimal information blends can be computed via an SDP reformulation, which also enables us to improve on a heuristic previously developed for robust Markov decision processes.

3. Robust optimization criterion. In statistics, confidence intervals can describe uncertain scalar parameters. The intervals are often mean-centered, although nonsymmetric choices are possible. The width of the interval is chosen to achieve a given confidence level $1-\epsilon$. For $c \sim \mathcal{N}(\bar{c}, \Sigma)$ with Σ positive definite ($\Sigma \succ 0$), we consider for some $\alpha > 0$ the confidence ellipsoid

$$\mathcal{C} = \{c \in \mathbb{R}^n : (c-\bar{c})^\top \Sigma^{-1} (c-\bar{c}) \leq \alpha^2\}. \tag{11}$$

Choosing $\alpha^2 = F^{-1}_{\chi^2_n}(1-\epsilon)$, where $F^{-1}_{\chi^2_n}(\cdot)$ is the inverse cumulative distribution function (cdf) of the chi-square distribution with n degrees of freedom, ensures that $c \in \mathcal{C}$ with probability $1-\epsilon$.

By selecting $\mathcal{C}$ as the uncertainty set for c, tractable robust optimization programs can be obtained.

LEMMA 1. *With* $\mathcal{X} = \{x \in \mathbb{R}^n : Ax = b, x \succeq 0\}$ *and* $\mathcal{C}$ *given by* (11)*, the problem* $\max_{x\in\mathcal{X}} \min_{\tilde{c}\in\mathcal{C}} \tilde{c}^\top x$ *is equivalent to* $\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$.

Proof. If $\alpha = 0$, $\mathcal{C} = \{\bar{c}\}$ and the result is trivially verified. If $\alpha > 0$, for any fixed x, the inner minimum $\min_{\tilde{c}\in\mathcal{C}} \tilde{c}^\top x$ is computed by applying the change of variable $z = \Sigma^{-1/2}(\tilde{c} - \bar{c})$, which yields $\min_{z:z^\top z \leq \alpha^2} (\Sigma^{1/2} z + \bar{c})^\top x$ where $\bar{c}^\top x$ is fixed. The minimum is attained at $z^* = -\beta \Sigma^{1/2} x$ for β such that $||z^*||_2^2 = \alpha^2$, that is, $\beta = \alpha/\sqrt{x^\top \Sigma x}$. In terms of $\tilde{c}$ the optimal solution is $\tilde{c} = \bar{c} - \alpha \Sigma x/\sqrt{x^\top \Sigma x}$, hence the value for the inner minimum, $\bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$. □

If Σ is only positive semidefinite ($\Sigma \succeq 0$ but $\Sigma \not\succ 0$), we consider the confidence ellipsoid

$$\begin{aligned} \widetilde{\mathcal{C}} &= \{c = Q_0 Q_0^\top \bar{c} + Q_+ c_+ \in \mathbb{R}^n : c_+ \in \mathcal{C}_+\} \\ \mathcal{C}_+ &= \{c_+ \in \mathbb{R}^p : (c_+ - Q_+^\top \bar{c})^\top \Sigma_+^{-1} (c_+ - Q_+^\top \bar{c}) \leq \alpha^2\}, \end{aligned} \tag{12}$$

where $Q_+ \in \mathbb{R}^{n\times p}$ and $Q_0 \in \mathbb{R}^{n\times(n-p)}$ come from the singular value decomposition (svd)

$$\Sigma = QSQ^\top = [Q_+ \; Q_0] \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} [Q_+ \; Q_0]^\top, \tag{13}$$

Σ_+ being the diagonal matrix containing the p positive singular values of Σ.

LEMMA 2. *With* $\mathcal{X} = \{x \in \mathbb{R}^n : Ax = b, x \succeq 0\}$ *and* $\widetilde{\mathcal{C}}$ *given by* (12)*, the problem* $\max_{x\in\mathcal{X}} \min_{c\in\widetilde{\mathcal{C}}} c^\top x$ *is equivalent to* $\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$.

Proof. Using (13), c can be reexpressed as

$$c = Q_0 Q_0^\top \bar{c} + Q_+ c_+, \qquad\qquad c_+ \sim \mathcal{N}(Q_+^\top \bar{c}, \Sigma_+).$$

Then,

$$\max_{x\in\mathcal{X}} \min_{c\in\mathcal{C}} c^\top x = \max_{x\in\mathcal{X}} \min_{c_+\in\mathcal{C}_+} (Q_0 Q_0^\top \bar{c} + Q_+ c_+)^\top x$$

$$
\begin{aligned}
&= \max_{x\in\mathcal{X}} \{\bar{c}^\top Q_0 Q_0^\top x + \min_{c_+\in\mathcal{C}_+} c_+^\top (Q_+^\top x)\} \\
&= \max_{x\in\mathcal{X}} \bar{c}^\top Q_0 Q_0^\top x + (Q_+^\top \bar{c})^\top (Q_+^\top x) - \alpha\sqrt{(Q_+^\top x)^\top \Sigma_+ Q_+^\top x} \\
&= \max_{x\in\mathcal{X}} \bar{c}^\top (Q_0 Q_0^\top + Q_+ Q_+^\top) x - \alpha\sqrt{x^\top Q_+ \Sigma_+ Q_+^\top x},
\end{aligned}
$$

which reduces to $\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$ using (13) and $Q_0 Q_0^\top + Q_+ Q_+^\top = [Q_0\ Q_+][Q_0\ Q_+]^\top = Q^\top Q = I$. □

Under the Bayesian modeling assumptions on c^{true}, the random variable $x^\top c^{\text{true}}$ follows $\mathcal{N}(x^\top \bar{c}, x^\top \Sigma x)$. In particular, if $\bar{x}$ attains $v_\alpha(\bar{c}, \Sigma)$, we have

$$
\mathbb{P}\{\bar{x}^\top c^{\text{true}} \geq v_\alpha(\bar{c}, \Sigma)\} = 1 - \Phi\left(\frac{v_\alpha(\bar{c}, \Sigma) - \bar{x}^\top c^{\text{true}}}{\sqrt{\bar{x}^\top \Sigma \bar{x}}}\right) = \Phi(\alpha) \ ,
$$

where Φ is the cumulative distribution function (cdf) of $\mathcal{N}(0,1)$. Thus if we want to ensure with confidence $1-\epsilon$ that $\bar{x}^\top c^{\text{true}} \geq v_\alpha(\bar{c}, \Sigma)$, we can choose $\alpha = \Phi^{-1}(1-\epsilon)$, which is less conservative than the choice $\alpha^2 = F^{-1}_{\chi^2_n}(1-\epsilon)$.

Finally, let us mention that (5) can be solved by commercial solvers as a quadratic program with quadratic constraints (QCQP):

LEMMA 3. *If $\Sigma \succ 0$, a dual formulation to (5) is*

$$
v_\alpha(\bar{c}, \Sigma) = \min_{c,z} b^\top z \ \text{ subject to } c \in \mathcal{C},\ A^\top z \succeq c \ ,
$$

using $\mathcal{C}$ given by (11). Otherwise, using $\widetilde{\mathcal{C}}$ given by (12),

$$
v_\alpha(\bar{c}, \Sigma) = \min_{c_+,z} b^\top z \ \text{ subject to } c_+ \in \mathcal{C}_+,\ \ A^\top z \succeq Q_0 Q_0^\top \bar{c} + Q_+ c_+ \ .
$$

Proof. A dual problem to $\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$ or equivalently $\max_{x\in\mathcal{X}} \min_{c\in\mathcal{C}} c^\top x$ is $\min_{c\in\mathcal{C}} \max_{x\in\mathcal{X}} c^\top x$. The dual to $\max_{x\in\mathcal{X}} c^\top x$ is $\min_{z\in\mathcal{Z}} b^\top z$ for $\mathcal{Z} = \{\mathbb{R}^m : A^\top z \succeq c\}$, hence the overall problem. The version with $\widetilde{\mathcal{C}}$ can be established similarly. □

4. Application to Markov decision processes. Let the tuple (S, A, P, R) define a Markov decision process [34] where S is a finite state space with $|S|$ states, A is a finite action space with $|A|$ actions, $P : S \times A \times S \mapsto [0,1]$ with values $p(s'|s,a)$ is a transition probability function, and $R : S \times A \mapsto \mathbb{R}$ is a reward function with bounded values $r(s,a)$. Let $0 < \gamma < 1$ be a discount factor, and let $b(j) = \mathbb{P}\{s_0 = j\}$ specify an initial state distribution, states being labeled from 1 to $|S|$. The maximization of the expected discounted cumulated reward

$$
v_\gamma^\pi = \mathbb{E}^\pi\{\textstyle\sum_{t=0}^\infty \gamma^t r(s_t, a_t)\} \tag{14}
$$

by the choice of a stochastic policy $\pi : S \times A \mapsto [0,1]$ with values $\pi(s,a) = \mathbb{P}\{a_t = a | s_t = s\}$ admits a dual linear programming formulation [13]

$$
\begin{aligned}
\text{maximize} \quad & \sum_{s\in S}\sum_{a\in A} r(s,a) x(s,a) && && (15)\\
\text{subject to} \quad & \sum_{a\in A} x(j,a) - \sum_{s\in S}\sum_{a\in A} \gamma\, p(j|s,a) x(s,a) = b(j) && \text{for } j \in S, \\
& x(s,a) \geq 0 && \text{for } s \in S, a \in A,
\end{aligned}
$$

which is of the form (1). Given an optimal $x^* \in \mathbb{R}^{|S|\times|A|}$,

$$\pi^*(s,a) = x^*(s,a) / \textstyle\sum_{a'\in A} x^*(s,a') \tag{16}$$

is an optimal stochastic policy. The optimization variables $x(s,a)$ (occupation measures) represent the total discounted probability of being in state s and choosing action a, when the system starts from state j with probability $b(j)$. The optimal policy (16) will be independent of the initial distribution.

4.1. MDP with Bayesian prior. In our framework, we assume that the rewards $r(s,a)$ are unknown but endowed with a prior $\mathcal{N}(\bar{r},\Sigma)$, where $\bar{r}$ collects the means $\bar{r}(s,a)$ and Σ is the covariance matrix collecting elements $\Sigma(s,a;s',a')$. The framework is less general than (Bayesian, model-based) reinforcement learning (RL), where transition probabilities would also be endowed with a prior. Nonetheless, the framework is already a valuable step for studying model ambiguity in Markov decision processes from a Bayesian standpoint.

Under the risk-neutral approach ($\alpha = 0$), the rewards $r(s,a)$ in (15) are set to their Bayesian mean $\bar{r}(s,a)$. The optimization problem has still the structure of a MDP, implying the existence of an optimal deterministic policy. To see that from (15), note that the simplex algorithm returns a vertex solution x^* defined by $|S|\cdot|A|$ linear equations, $|S|$ coming from the equality constraints and $|S|\cdot|A| - |S|$ coming from active inequalities $x(s,a) = 0$. Hence x^* has at most $|S|$ nonzero coordinates. The definition of a proper policy requires one nonzero coordinate being assigned to each state, implying that the policy (16) is in fact deterministic.

When the robust optimization approach is used ($\alpha > 0$), the program for finding an optimal policy becomes

$$\begin{aligned}
\text{maximize} \quad & \sum_{s\in S}\sum_{a\in A} \bar{r}(s,a)x(s,a) - \alpha\sqrt{\sum_{s\in S}\sum_{a\in A}\sum_{s'\in S}\sum_{a'\in A} x(s,a)\Sigma(s,a;s',a')x(s',a')} \qquad (17)\\
\text{subject to} \quad & \sum_{a\in A} x(j,a) - \sum_{s\in S}\sum_{a\in A} \gamma\, p(j|s,a)x(s,a) = b(j) \qquad \text{for } j\in S,\\
& x(s,a) \geq 0 \qquad \text{for } s\in S,\ a\in A.
\end{aligned}$$

Generically, optimal solutions to SOCPs are not vertex solutions. Thus more elements $x^*(s,a)$ will be nonzero, and the resulting stochastic policy (16) does not necessarily degenerate into a deterministic one.

The program (17) is a tractable robust MDP obtained by applying generic robust linear programming techniques. The covariance matrix $\Sigma(s,a;s'a')$ allows one to model worst-case reward dependencies among state-action pairs.

4.2. Optimal measurements with fixed decisions. Consider now the measurement selection problem based on the maximization over u of $\mathbb{K}_\alpha(u,\bar{c},\Sigma)$ as defined by (9). An approximation proposed in [10] for robust MDPs with the measurement u valued in $\{e_1,\dots,e_n\}$ assumes that, inside the expectation in (9), for each outcome y, the optimal solution x' attaining $v_\alpha(\bar{c}',\Sigma')$ is replaced by the solution $\bar{x}$ attaining $v_\alpha(\bar{c},\Sigma)$. By doing so, $\mathbb{K}_\alpha(u,\bar{c},\Sigma)$ is approximated by

$$\widetilde{\mathbb{K}}_\alpha(u,\bar{c},\Sigma) = \mathbb{E}_y\{[\bar{c}'^\top\bar{x} - \alpha\sqrt{\bar{x}^\top\Sigma'\bar{x}}] - [\bar{c}^\top\bar{x} - \alpha\sqrt{\bar{x}^\top\Sigma\bar{x}}] \mid u,\bar{c},\Sigma\} = \alpha(\sqrt{\bar{x}^\top\Sigma\bar{x}} - \sqrt{\bar{x}^\top\Sigma'\bar{x}}) \ , \tag{18}$$

where $\mathbb{E}_y\{\bar{c}'\} = \bar{c}$ has been used.

Note that $\widetilde{\mathbb{K}}_\alpha(u,\bar{c},\Sigma) = 0$ for all u if $\alpha = 0$, suggesting that this approximation is uninformative in the risk-neutral case. Despite this undesirable behavior, we can still investigate the problem of maximizing $\widetilde{\mathbb{K}}_\alpha(u,\bar{c},\Sigma)$.

PROPOSITION 1. *Let* $\bar{x} \in \arg\max_{x\in\mathcal{X}} \bar{c}^\top x - \alpha\sqrt{x^\top \Sigma x}$, *and let* $\widetilde{\mathbb{K}}_\alpha(\cdot, \bar{c}, \Sigma)$ *be the approximation relative to* $\bar{x}$. *Then either* $\Sigma\bar{x} = 0$ *and any* $u \in \mathbb{B}$ *is optimal for* $\max_{u\in\mathbb{B}} \widetilde{\mathbb{K}}_\alpha(\cdot, \bar{c}, \Sigma)$, *or* $\Sigma x \neq 0$ *and the maximum of* $\widetilde{\mathbb{K}}_\alpha(\cdot, \bar{c}, \Sigma)$ *over* $\mathbb{B}$ *is attained by selecting*

$$\bar{u} \in \left\{ \pm \frac{(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma\bar{x}}{||(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma\bar{x}||} \right\} , \tag{19}$$

where I_n *is the identity matrix in* $\mathbb{R}^{n\times n}$.

Proof. The proof relies on techniques used in Section 5. It can be found in the appendix. □

From (19), we can better understand the effect of the fixed-decision approximation. If we assume momentarily that σ_w^2 is small with respect to the eigenvalues of Σ, then $(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma$ is close to I_n, so $\bar{u}$ is close to $\bar{x}/||\bar{x}||$. Therefore, $\bar{u}$ tends to measure the coordinates of c^{true} according to the magnitude of their believed contribution to the objective value given the current optimal solution $\bar{x}$. For any value of σ_w^2, if Σ is diagonal, the coordinates c_j for $j \in \{i : \bar{x}_i = 0\}$ are not measured.

This analysis suggests that using the approximation (18) would lead to a measurement policy that is not asymptotically consistent, in the sense that wrong beliefs would not necessarily be corrected by an infinite sequence of measurements.

5. Structural properties for optimal measurements. Convex functions have their supremum on the boundary of their effective domain [37]. A similar result holds for the nonconvex function $\mathbb{K}_\alpha(\cdot, \bar{c}, \Sigma)$.

THEOREM 1. *Let* $\mathcal{U}$ *be an arbitrary nonempty closed convex bounded set. Let* $\partial\mathcal{U}$ *denote the boundary of* $\mathcal{U}$. *We have*

$$\max_{u\in\mathcal{U}} \mathbb{K}_\alpha(u, \bar{c}, \Sigma) = \max_{u\in\partial\mathcal{U}} \mathbb{K}_\alpha(u, \bar{c}, \Sigma).$$

Proof. Fix u in the interior of $\mathcal{U}$. Define u_+ by extending u to $\partial\mathcal{U}$ as follows: define $t^* = \max\{t \geq 0 : tu/||u|| \in \mathcal{U}\}$, $\tau = t^*/||u||$, $u_+ = \tau u \in \partial\mathcal{U}$. Necessarily, $\tau \geq 1$. Essentially, we show that the measurement based on u_+ dominates the measurement based on u, so that optimal measurements are on $\partial\mathcal{U}$.

Define

$$\beta = \frac{u^\top\Sigma u + \sigma_w^2}{u^\top\Sigma u + (\sigma_w/\tau)^2}, \quad \Lambda = \frac{\Sigma u u^\top \Sigma}{u^\top\Sigma u + \sigma_w^2}.$$

Note that $1 \leq \beta \leq \tau^2$. From the update of Σ after measurements $y_u = c^\top u + w$ or $y_{u_+} = c^\top u_+ + w$, we deduce the ordering of the two updated covariance matrices in the cone of the positive semidefinite matrices:

$$\Sigma'_{u^+} = \Sigma - \frac{\Sigma u_+ u_+^\top \Sigma}{u_+^\top\Sigma u_+ + \sigma_w^2} = \Sigma - \frac{\tau^2\Sigma u u^\top\Sigma}{\tau^2[u^\top\Sigma u + (\sigma_w/\tau)^2]} = \Sigma - \beta\Lambda \preceq \Sigma - \Lambda = \Sigma'_u,$$

meaning (informally) that the residual uncertainty is "smaller" with u_+. From the update of $\bar{c}$ after the observations y_u or y_{u_+},

$$\bar{c}'_u = \bar{c} + \frac{\Sigma u}{u^\top\Sigma u + \sigma_w^2}(y_u - \bar{c}^\top u), \qquad \bar{c}'_{u+} = \bar{c} + \frac{\Sigma u_+}{u_+^\top\Sigma u_+ + \sigma_w^2}(y_{u_+} - \bar{c}^\top u),$$

and from the distribution of the observations,

$$y_u \sim \mathcal{N}(u^\top\bar{c}, u^\top\Sigma u + \sigma_w^2), \qquad y_{u_+} \sim \mathcal{N}(\tau u^\top\bar{c}, \tau^2 u^\top\Sigma u + \sigma_w^2),$$

we deduce the distribution of the updated means,

$$\bar{c}'_u \sim \mathcal{N}(\bar{c}, \Lambda), \qquad\qquad \bar{c}'_{u_+} \sim \mathcal{N}(\bar{c}, \beta\Lambda).$$

Using the zero-mean random vector $z \sim \mathcal{N}(0, \Lambda)$, we have

$$\mathbb{E}\{v_\alpha(\bar{c}'_{u_+}, \Sigma'_{u_+})\} = \mathbb{E}\{v_\alpha(\bar{c} + \sqrt{\beta}z, \Sigma'_{u_+})\} \geq \mathbb{E}\{v_\alpha(\bar{c} + z, \Sigma'_{u_+})\} = \mathbb{E}\{v_\alpha(\bar{c}'_u, \Sigma'_{u_+})\},$$

where the inequality is justified by an extension of Jensen's inequality, that states that a function $g(t) = \mathbb{E}\{f(x_0 + tz)\}$ defined for $t \geq 0$ is monotone increasing if f is convex and $\mathbb{E}\{z\} = 0$. Since $\Sigma'_{u_+} \preceq \Sigma'_u$, we have

$$\bar{c}'^{\top}_u x - \alpha\sqrt{x^\top \Sigma'_{u_+} x} \geq \bar{c}'^{\top}_u x - \alpha\sqrt{x^\top \Sigma'_u x},$$

implying $v_\alpha(\bar{c}'_u, \Sigma'_{u_+}) \geq v_\alpha(\bar{c}'_u, \Sigma'_u)$ and thus

$$\mathbb{E}\{v_\alpha(\bar{c}'_u, \Sigma'_{u_+})\} \geq \mathbb{E}\{v_\alpha(c'_u, \Sigma'_u)\}.$$

Therefore, $\mathbb{K}(u^+, \bar{c}, \Sigma) \geq \mathbb{K}(u, \bar{c}, \Sigma)$. Since u was arbitrary, the result follows. □

If we now restrict ourselves to the case where $\mathcal{U}$ is the L2-ball $\mathbb{B}$, Theorem 1 indicates that we should seek solutions u on the L2-sphere $\partial\mathbb{B} = \{u \in \mathbb{R}^n : ||u|| = 1\}$.

It will be convenient to rewrite the objective (9) as

$$\mathbb{K}_\alpha(u, \bar{c}, \Sigma) = \mathbb{E}_t\{v_\alpha(\bar{c} + t\,\Sigma d_u, \Sigma') \;\mid\; u, \bar{c}, \Sigma\} - v_\alpha(\bar{c}, \Sigma) \;\;, \tag{20}$$

where $t \sim \mathcal{N}(0,1)$ and where we have introduced the vector

$$d_u = \frac{u}{\sqrt{u^\top \Sigma u + \sigma_w^2}} \;\;. \tag{21}$$

In the special case $||u|| = 1$, we have $u^\top \Sigma u + \sigma_w^2 = u^\top(\Sigma + \sigma_w^2 \mathrm{I}_n)u$. This leads us to define

$$P = \Sigma + \sigma_w^2 \mathrm{I}_n \;\;. \tag{22}$$

The matrix P is positive definite and thus invertible.

In the risk-neutral case ($\alpha = 0$), we can go further in the characterization of optimal solutions.

THEOREM 2. *Assume the risk-neutral case ($\alpha = 0$). Then either any $u \in \mathbb{B}$ is optimal for $\max_{u\in\mathbb{B}} \mathbb{K}_0(u, \bar{c}, \Sigma)$, or the solutions u^* optimal for $\max_{u\in\mathbb{B}} \mathbb{K}_0(u, \bar{c}, \Sigma)$ satisfy*

$$u^* \in \left\{ \pm \frac{P^{-1}\Sigma\,\mathbb{E}\{t\bar{x}(t)\}}{||P^{-1}\Sigma\,\mathbb{E}\{t\bar{x}(t)\}||} \right\} \;\;, \qquad\qquad \bar{x}(t) \in \arg\max_{x\in\mathcal{X}} \left(\bar{c} + \frac{t\,\Sigma u^*}{||P^{1/2}u^*||} \right)^\top x \;\;,$$

where the expectation is taken over $t \sim \mathcal{N}(0,1)$, and where without loss of generality the vector-valued function $\bar{x}(\cdot)$ is piecewise-constant on $\mathbb{R}$ with a finite number of pieces.

Proof. Let Ξ denote the space of all measurable vector-valued functions $x(\cdot) : \mathbb{R} \mapsto \mathbb{R}^n$ with values $x(t) \in \mathcal{X}$, defined for all $t \in \mathbb{R}$. Note first that for any $u \in \mathbb{B}$, there exists for each t a measurable selection $x(t)$ [14] of the optimal solution set $X(t) = \arg\max_{x\in\mathcal{X}} (\bar{c} + t\Sigma d_u)^\top x$ such that $x(\cdot) \in \Xi$ is

a piecewise-constant, vector-valued function with a finite number of pieces [19, 41]. Thus we can actually restrict Ξ to that space of functions. Consider

$$\max_{u\in\mathbb{B}} \mathbb{K}_0(u,\bar{c},\Sigma) = \max_{u\in\mathbb{B}} \mathbb{E}_t\left\{\max_{x(\cdot)\in\Xi}\left(\bar{c}+t\frac{\Sigma u}{\sqrt{u^\top\Sigma u+\sigma_w^2}}\right)^\top x(t)\right\} - v_0(\bar{c},\Sigma)$$
$$= \max_{x(\cdot)\in\Xi}\max_{u\in\mathbb{B}} \mathbb{E}_t\left\{\left(\bar{c}+t\frac{\Sigma u}{\sqrt{u^\top\Sigma u+\sigma_w^2}}\right)^\top x(t)\right\} - v_0(\bar{c},\Sigma) \ ,$$

where the interchange between $\mathbb{E}_t$ and $\max_{x(\cdot)\in\Xi}$ is possible because the optimization problem is written in terms of a function $x(\cdot)$ that does not explicitly depend on u.

One can check that $\max_{u\in\mathbb{B}} \mathbb{K}_0(u,\bar{c},\Sigma)\geq 0$ by plugging in the constant-valued function $x(\cdot)=\bar{x}_0$, where $\bar{x}_0\in\mathcal{X}$ attains $v_0(\bar{c},\Sigma)$: for any u, one obtains $\mathbb{E}_t\{(\bar{c}+t\Sigma d_u)^\top\bar{x}_0\} = \bar{c}^\top\bar{x}_0 + \mathbb{E}\{t\}d_u^\top\Sigma\bar{x}_0 = v_0(\bar{c},\Sigma)$.

Assume that we are given an optimal function $\bar{x}(\cdot)\in\Xi$ for the problem. The set of the vectors $u\in\mathbb{B}$ that attain $\max_{u\in\mathbb{B}}\mathbb{K}_0(u,\bar{c},\Sigma)$ along with $\bar{x}(\cdot)$ can be expressed by

$$\arg\max_{u\in\mathbb{B}} \mathbb{E}\left\{\left(\bar{c}+t\,\frac{\Sigma u}{\sqrt{u^\top\Sigma u+\sigma_w^2}}\right)^\top \bar{x}(t)\right\} = \arg\max_{u\in\mathbb{B}}\ \frac{u^\top}{\sqrt{u^\top\Sigma u+\sigma_w^2}}\Sigma\mathbb{E}\left\{t\bar{x}(t)\right\} \ ,$$

dropping the constant term $\mathbb{E}\{\bar{c}^\top\bar{x}(t)\}$ on the right-hand side.

If $\Sigma\mathbb{E}\left\{t\bar{x}(t)\right\}=0$, then any $u\in\mathbb{B}$ is optimal. Otherwise, $\Sigma\mathbb{E}\left\{t\bar{x}(t)\right\}\neq 0$, and by theorem 1,

$$\arg\max_{u\in\mathbb{B}}\mathbb{K}_0(u,\bar{c},\Sigma) = \arg\max_{u:\ ||u||=1}\ \frac{u^\top\Sigma\mathbb{E}\{t\bar{x}(t)\}}{\sqrt{u^\top Pu}}.$$

Moreover, using $v=P^{1/2}u$, we have

$$\max_{u:\ ||u||=1}\ \frac{u^\top\Sigma\mathbb{E}\{t\bar{x}(t)\}}{\sqrt{u^\top Pu}} = \max_{v:\ ||P^{-1/2}v||=1}\ \frac{v^\top P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}}{||v||} \ .$$

Recall that for any z, here taken to be $z=P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}$,

$$||z|| = \max_{y\in\mathbb{B}}\ y^\top z = \max_{y\neq 0}\ y^\top z/||y|| \ .$$

Therefore, an optimal v is given by $v^*=\beta P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}$ with β such that $||P^{-1/2}v^*||=1$. Then it follows that $u^*=P^{-1/2}v=P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}/||P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}||$ is optimal. Moreover, if u^* is optimal, then $-u^*$ is optimal, by the symmetry of the Gaussian distribution and the expression of d_{u^*}. □

COROLLARY 1 **(Norm-maximization reformulation)**. *In the risk-neutral case ($\alpha=0$), we have*

$$\max_{u\in\mathbb{B}}\mathbb{K}_0(u,\bar{c},\Sigma) = \max_{x(\cdot):\,x(t)\in\mathcal{X}}\left\{\mathbb{E}_t\{\bar{c}^\top x(t)\} + ||P^{-1/2}\Sigma\mathbb{E}_t\{tx(t)\}||\right\} - v_0(\bar{c},\Sigma) \ , \tag{23}$$

where u is recovered from an optimal $x^*(\cdot)$ by $u^*=P^{-1}\Sigma\mathbb{E}_t\{tx^*(t)\}/||P^{-1}\Sigma\mathbb{E}_t\{tx^*(t)\}||$.

Proof. Let $f(x(\cdot))=\mathbb{E}_t\{\bar{c}^\top x(t)\}+||P^{-1/2}\Sigma\mathbb{E}_t\{tx(t)\}||\}$. Since f is convex, optimal solutions are attained on the extreme points of the feasibility set. Thus without loss of generality we can assume that $x(t)$ is a vertex of $\mathcal{X}$ for each t. Let $\bar{x}(\cdot)\in\Xi$ be an optimal solution with Ξ defined as in Theorem 2.

First, consider the degenerate case where $\Sigma\mathbb{E}_t\{t\bar{x}(t)\}=0$. Then, $f(\bar{x}(\cdot))=\mathbb{E}_t\{\bar{c}^\top\bar{x}(t)\}$. Since $\bar{x}(t)$ is optimal by assumption, and since any solution $\bar{x}_0$ that attains $v_0(\bar{c},\Sigma)$ is in $\arg\max_{x\in\mathcal{X}}\bar{c}^\top x$, we can assume without loss of generality that $\bar{x}(t)=\bar{x}_0$ almost surely, so that $\mathbb{E}_t\{\bar{c}^\top\bar{x}(t)\}=\bar{c}^\top\bar{x}_0=v_0(\bar{c},\Sigma)$. Hence in that case any measurement is optimal (in fact no new measurement is needed).

Next, consider the nondegenerate case where $\Sigma\mathbb{E}_t\{t\bar{x}(t)\}\neq 0$. The relation $\max_{u\in\mathbb{B}}\mathbb{K}_0(\bar{c},\Sigma)=\max_{x(\cdot)\in\Xi:x(t)\in\mathcal{X}}f(x(\cdot))-v_0(\bar{c},\Sigma)$ can be checked by comparing the two objectives with u set to $P^{-1}\Sigma\mathbb{E}\{tx(t)\}/||P^{-1}\Sigma\mathbb{E}\{tx(t)\}||$: one gets

$$\mathbb{E}\left\{\left(\bar{c}+t\frac{\Sigma\bar{u}}{\sqrt{\bar{u}^\top\Sigma\bar{u}+\sigma_w^2}}\right)^\top\bar{x}(t)\right\}-v_0(\bar{c},\Sigma)=\bar{c}^\top\mathbb{E}\{\bar{x}(t)\}+||P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}||-v_0(\bar{c},\Sigma)\ .$$

At the same time, with $\Sigma\mathbb{E}_t\{t\bar{x}(t)\neq 0$, the subdifferential of $f(x(\cdot))$ at $\bar{x}(\cdot)$ is a singleton corresponding to the gradient of $f(x(\cdot))$ at $\bar{x}(\cdot)$. The gradient of $f(x(\cdot))$ with respect to $x(t')$ for some fixed t' is given by

$$\nabla_{x(t')}f(x(\cdot))=\phi(t')\bar{c}+\phi(t')(t'P^{-1/2}\Sigma)^\top\frac{P^{-1/2}\Sigma\mathbb{E}\{tx(t)\}}{||P^{-1/2}\Sigma\mathbb{E}\{tx(t)\}||}=\phi(t')\left[\bar{c}+t'\frac{\Sigma P^{-1}\Sigma\mathbb{E}\{tx(t)\}}{||P^{-1/2}\Sigma\mathbb{E}\{tx(t)\}||}\right].$$

At $\bar{x}(\cdot)$, we have the implicit definition $\bar{u}=P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}/||P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}||$, so we have

$$\frac{\Sigma\bar{u}}{||P^{1/2}\bar{u}||}=\frac{\Sigma P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}}{||P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}||}\ .$$

Therefore, the gradient with respect to $x(t')$ at $\bar{x}(\cdot)$ can be written as

$$\nabla_{x(t')}f(x(\cdot))|_{\bar{x}}=\phi(t')\left[\bar{c}+t'\frac{\Sigma P^{-1}\Sigma\mathbb{E}\{t\bar{x}(t)\}}{||P^{-1/2}\Sigma\mathbb{E}\{t\bar{x}(t)\}||}\right]=\phi(t')\left[\bar{c}+t'\frac{\Sigma\bar{u}}{||P^{1/2}\bar{u}||}\right]\ .$$

From the basic variational inequality for minimization [14, Thm. 2A.6], a necessary condition for attaining a maximum is $\nabla_{x(t')}f(\bar{x}(\cdot))_{\bar{x}}\in N_{\mathcal{X}}(\bar{x}(t'))$ for almost every t', where $N_{\mathcal{X}}(\bar{x}(t'))$ is the normal cone to $\mathcal{X}$ at $\bar{x}(t')$. Since $\phi(t')>0$, we can invoke the property that $x\in K$ iff $ax\in K$ for a cone K and some positive scalar a, and deduce that $\bar{x}(\cdot)$ must satisfy

$$\bar{c}+\frac{t\Sigma\bar{u}}{||P^{1/2}\bar{u}||}\in N_{\mathcal{X}}(\bar{x}(t))\quad\text{ for almost every } t\ .$$

Now, note that these conditions are necessary and sufficient for ensuring that

$$\bar{x}(t)\in\arg\max_{x\in\mathcal{X}}\left(\bar{c}+\frac{t\Sigma\bar{u}}{||P^{1/2}\bar{u}||}\right)^\top x\quad\text{ for almost every } t\ ,$$

since the latter problem is convex. We have thus verified that (23) fulfills at optimality the necessary conditions of Theorem 2. □

Theorem 2 and its corollary concern the case $\alpha=0$ only. They will not be used in the rest of the paper. However, the structure of the problem (23) makes it easier to establish a complexity result:

PROPOSITION 2 **(NP-completeness)**. *The decision problem associated with (9) with a discretized expectation is NP-complete.*

Proof. For establishing a complexity result, without loss of generality we can set $\alpha = 0$, $\bar{c} = 0$, $\Sigma = \mathrm{I}_n$, and consider $\max_{u\in\mathbb{B}} \mathbb{K}_0(u, 0, \mathrm{I}_n)$. From (23) we obtain $\max_{x(\cdot):\, x(t)\in\mathcal{X}} (1+\sigma_w^2)^{-1/2} ||\mathbb{E}\{tx(t)\}||$, which is equivalent to $\max_{z\in\mathcal{Z}} ||z||$ with $\mathcal{Z} = \{z \in \mathbb{R}^n : z = \mathbb{E}\{tx(t)\}, x(t) \in \mathcal{X}\}$. By discretizing the random variable t into N samples t_i, we obtain a set $\mathcal{Z}^N$ in $\mathbb{R}^n$ which is the projection of a polyhedral set in $\mathbb{R}^{n(N+1)}$ where each $x(t_i)$ can be assumed to be a vertex of $\mathcal{X}$. In that case, $\mathcal{Z}^N$ is polyhedral. The decision problem associated to the maximization of the L2-norm of a vector over a polyhedral set is known to be NP-complete [28]. □

Proposition 2 indicates that we should not expect to develop exact solution algorithms for our problem. Rather it emphasizes the need for good approximations.

6. Optimal uncertainty reduction. Consider the sequential measurement setting, where measurements are taken iteratively. For a given sequence $\{u_k : k \geq 1\}$ of measurements, let $\Sigma_1 = \Sigma \in \{S \in \mathbb{R}^{n\times n} : S = S^\top, S \succeq 0\}$ be the initial covariance matrix, and consider the matrix sequence $\{\Sigma_k : k \geq 1\}$ defined from (8) by

$$\Sigma_{k+1} = \Sigma_k - \Sigma_k u_k u_k^\top \Sigma_k / (u_k^\top \Sigma_k u_k + \sigma_w^2).$$

Independently of the objective (10) based on the expected value of information from the next measurement, a direct approach for reducing the uncertainty is to acquire information on c^{true} by making measurements u_k such that Σ_k provably tends to the zero matrix. By the degeneracy of the posterior distribution of $c_k \sim \mathcal{N}(\bar{c}_k, \Sigma_k)$, Doob's consistency theorem [15] implies that the sequence of updated means $\bar{c}_k$ tends to c^{true}.

This section studies such a method, and shows that it achieves a rate of convergence which is optimal in a certain sense. Namely, we consider u_k taken as a dominant eigenvector of Σ_k:

$$u_k \in E_{\max}(\Sigma_k), \tag{24}$$

using the following notations defined for any symmetric matrix $S \in \mathbb{R}^{n\times n}$:

- $\lambda_{\max}(S) = \max\{\lambda \in \mathbb{R} : Su = \lambda u,\ u^\top u = 1\}$: largest eigenvalue of S;
- $E_{\max}(S) = \{u \in \mathbb{R}^n : Su = \lambda_{\max}(S)u,\ u^\top u = 1\}$: the set of normalized eigenvectors in the eigenspace associated to $\lambda_{\max}(S)$, excluding the zero vector.

For any $\epsilon > 0$, we can ensure that $\operatorname{trace}\Sigma_k < \epsilon$ after a certain number of measurements, as made precise by the following lemma.

LEMMA 4. *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of Σ_1, with repetition according to eigenvalue multiplicity. Fix $\epsilon > 0$. Then the matrix sequence $\{\Sigma_k : k \geq 1\}$ associated with u_k given by (24) satisfies $\operatorname{trace}\Sigma_k < \epsilon$ for any $k > k_0 = \sum_{i=1}^n \log(n/\epsilon)/\log(1/s_i)$, where $s_i = [1 - \lambda_i/(\lambda_i + \sigma_w^2)]$ for $i = 1, \ldots, n$.*

Proof. By the eigenvalue decomposition of $\Sigma_k \succeq 0$, we have $\Sigma_k = \sum_{i=1}^n \lambda_{ik} u_{ik} u_{ik}^\top$, where $\lambda_{1k} \geq \lambda_{2k} \geq \cdots \geq \lambda_{nk} \geq 0$, and where $u_{ik}^\top u_{jk} = 1$ if $i = j$, $u_{ik}^\top u_{jk} = 0$ if $i \neq j$. Taking $u_k = u_{1k}$ in the update equation gives $\Sigma_{k+1} = \Sigma_k - (\lambda_{1k}^2 u_{1k} u_{1k}^\top)/(\lambda_{1k} + \sigma_w^2) = \lambda_{1k}(1 - \lambda_{1k}/(\lambda_{1k} + \sigma_w^2)) u_{1k} u_{1k}^\top + \sum_{i=2}^n \lambda_{ik} u_{ik} u_{ik}^\top$. Therefore, iterations leave the original eigenvectors unchanged.

If the noise variance $\sigma_w^2 = 0$, the covariance would become the zero matrix after at most n iterations (exactly n iterations if the matrix is full rank). With $\sigma_w^2 > 0$, we evaluate the number of iterations needed to have $\operatorname{trace}(\Sigma_k) < \epsilon$ as follows. For each i, let $s_i = 1 - \lambda_{i1}/(\lambda_{i1} + \sigma_w^2)$. Define $k_i = \inf\{k \in \mathbb{N} : s_i^k < \epsilon/n\}$, that is, $k_i = \lceil \log(\epsilon/n)/\log(s_i) \rceil$. Since each iteration shrinks the current largest eigenvalue, we are guaranteed to have $\lambda_{ik} < \epsilon/n$ for each i after $k_0 = \sum_{i=1}^n k_i$ iterations. This implies $\operatorname{trace}\Sigma_k = \sum_{i=1}^n \lambda_{ik} < \epsilon$. □

COROLLARY 2. *The matrix sequence $\{\Sigma_k : k \geq 1\}$ associated to $u_k \in E_{max}(\Sigma_k)$ converges to the zero matrix (in the metric space of the Frobenius norm).*

Proof. $||\Sigma_k||_F = (\sum_{i=1}^n \sum_{j=1}^n \Sigma_{k,ij}^2)^{1/2} = (\sum_{i=1}^n \lambda_{ik}^2)^{1/2} \leq \sum_{i=1}^n |\lambda_{ik}| = \sum_{i=1}^n \lambda_{ik} = \text{trace}(\Sigma_k)$, so $\text{trace}(\Sigma_k) < \epsilon$ implies $||\Sigma_k||_F < \epsilon$. □

COROLLARY 3. *To any measurement policy π with values $u_k = \pi(\bar{c}_k, \Sigma_k)$ can be associated a family $\{\pi^\kappa : \kappa = 2, 3, \dots\}$ of asymptotically consistent modified policies with value $u_k = \pi^\kappa(\bar{c}_k, \Sigma_k, k)$ and such that $\text{trace}\,\Sigma_k < \epsilon$ for any $k > \kappa k_0$, where k_0 is given by Lemma 4.*

Proof. By construction: we define $\pi^\kappa(\bar{c}_k, \Sigma_k, k) = \pi(\bar{c}_k, \Sigma_k)$ if $\mod(k, \kappa) \neq \kappa - 1$, $\pi^\kappa(\bar{c}_k, \Sigma_k, k) \in E_{\max}(\Sigma^k)$ if $\mod(k, \kappa) = \kappa - 1$. □

The following result shows that the rate of convergence cannot be improved.

THEOREM 3. *All the measurement sequences defined by $u_k \in E_{max}(\Sigma_k)$ achieve the optimal rate of convergence of $\{\text{trace}(\Sigma_k) : k \geq 1\}$ to 0, among the sequences such that $||u_k|| \leq 1$.*

Proof. The rate of convergence is maximized if we minimize the trace of Σ_{k+1} given Σ_k. Writing Σ' for Σ_{k+1} and Σ for Σ_k, we consider

$$\min_{u:\,||u||=1} \text{trace}\left(\Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top \Sigma u + \sigma_w^2}\right) = \text{trace}(\Sigma) - \max_{u:\,||u||=1} \frac{u^\top \Sigma \Sigma u}{u^\top P u}.$$

The solution to the maximization problem in the second term is obtained by considering the generalized eigenvalue problem $\Sigma^2 u = \lambda P u$ and taking the vector u associated to the dominant generalized eigenvalue λ. Since P is nonsingular, the generalized eigenvalue problem is equivalent to the standard eigenvalue problem $P^{-1}\Sigma^2 u = \lambda u$. Therefore, the sequence defined by $u_k \in E_{\max}((\Sigma_k + \mathrm{I}_n \sigma_w^2)^{-1}\Sigma_k^2)$ maximizes the rate of convergence of $\text{trace}(\Sigma_k)$ to 0.

We will now prove that $E_{\max}(P^{-1}\Sigma^2) = E_{\max}(\Sigma)$, allowing us to conclude that $u_k \in E_{\max}(\Sigma_k)$ is also optimal. To do that, we use the eigenvalue decomposition $\Sigma = QDQ^\top$, where D is diagonal with elements $D_{ii} = \lambda_i$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, and $Q = [q_1 \dots q_n]$ is the matrix of eigenvectors such that $Q^\top Q = \mathrm{I}_n = QQ^\top$. By the Rayleigh quotient representation, $u_k \in E_{\max}(P^{-1}\Sigma^2)$ iff $u_k \in \arg\max_{u:\,||u||=1} u^\top P^{-1}\Sigma^2 u$. Now, we have

$$\begin{aligned}
&\arg\max_{u:\,||u||=1} u^\top (\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma^2 u \\
&= \arg\max_{u:\,||u||=1} u^\top (Q(D + \mathrm{I}_n\sigma_w^2)Q^\top)^{-1} Q D^2 Q^\top u = \arg\max_{u:\,||u||=1} u^\top Q (D + \mathrm{I}_n\sigma_w^2)^{-1} D^2 Q^\top u \\
&= \arg\max_{\theta:\,||\theta||=1} \theta^\top (D + \mathrm{I}_n\sigma_w^2)^{-1} D^2 \theta = \arg\max_{\theta:\,||\theta||=1} \sum_{i=1}^n \frac{\lambda_i^2 \theta_i}{\lambda_i + \sigma_w^2} = \arg\max_{\theta:\,||\theta||=1} \sum_{i=1}^n \nu_i \theta_i,
\end{aligned}$$

where we have used the change of variable $\theta = Q^\top u$ and defined $\nu_i = \lambda_i^2/(\lambda_i + \sigma_w^2)$. We have $\nu_i = \nu_j$ iff $\lambda_i = \lambda_j$. The ordering of the λ_i's implies $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_n \geq 0$. If $\nu_1 > \nu_2$, the optimal solution θ^* is the unit vector e_1, so $u^* = Q\theta^* = Qe_1 = q_1$. If $\nu_1 = \cdots = \nu_k > \nu_{k+1}$, we have $\theta^* \in \{\sum_{i=1}^k w_k e_k : \sum_{i=1}^k w_i = 1,\, w_i \geq 0\}$ and thus $u^* \in \{\sum_{i=1}^k w_k q_k : \sum_{i=1}^k w_i = 1,\, w_i \geq 0\}$, showing that the principal eigenspaces of Σ and $P^{-1}\Sigma^2$ coincide. □

Note that the condition $\Sigma_k \to 0$ is sufficient but not necessary for the convergence of x_k to a maximizer of the true problem (1). To see that, imagine that some coefficient c_j plays no role in the optimization problem, because of a constraint $x_j = 0$. Say that c_j is statistically independent of the other coefficients, and has a prior with an arbitrarily large variance. A sequential measurement algorithm defined by (24) will dedicate many measurements to the reduction of uncertainty on c_j. However, with $\alpha = 0$ we should never measure c_j, since updates of $\bar{c}_j$ never improve the objective.

7. Optimal expected improvement. We now come back to the problem of solving (10) as a stochastic program. A prerequisite is the construction of a finite approximation to the expectation in (9). To do that, consider

- $\phi(t) = (2\pi)^{-1/2}\exp\{-t^2/2\}$: pdf of $\mathcal{N}(0,1)$
- $\Phi(t) = \int_{-\infty}^{t}\phi(t')dt'$: cdf of $\mathcal{N}(0,1)$
- $\{t_i\}_{1\le i\le N}$: sequence defined by $t_0 = -\infty$, $t_{N+1} = +\infty$,

$$\int_{(t_{i-1}+t_i)/2}^{(t_i+t_{i+1})/2}(t-t_i)\phi(t)\mathrm{d}t = 0, \quad 1\le i\le N. \tag{25}$$

The relation (25) expresses a stationary property satisfied by the optimal solution to the quantization problem [21]

$$D_N = \inf_{q\in\mathcal{Q}_N}\mathbb{E}\{||t-q(t)||^2\}, \quad t\sim\mathcal{N}(0,1),$$

where $\mathcal{Q}_N$ denotes the class of measurable functions $q:\mathbb{R}\mapsto\mathbb{R}$ with at most N values $t_1,\dots,t_N$. Because $\mathcal{N}(0,1)$ is one-dimensional and strongly unimodal, the points t_i are uniquely determined by (25) [21, Thm I.5.1]. The points can be computed by methods described in [31].

- $\{p_i\}_{1\le i\le N}$ with $p_i = \Phi\left(\frac{t_i+t_{i+1}}{2}\right) - \Phi\left(\frac{t_{i-1}+t_i}{2}\right)$. For a function f that is Lipschitz continuous modulus L,

$$\left|\mathbb{E}\{f(t)\} - \textstyle\sum_{i=1}^{N} p_i f(t_i)\right| \le L\mathbb{E}\{||t-q(t)||\}.$$

For a convex function f, we have [31]

$$\textstyle\sum_{i=1}^{N} p_i f(t_i) \le \mathbb{E}\{f(t)\}. \tag{26}$$

Using the optimal N-quantization of $\mathcal{N}(0,1)$, we then define

$$\widehat{\mathbb{K}}_\alpha^N(u,\bar{c},\Sigma) = \textstyle\sum_{i=1}^{N} p_i v_\alpha(\bar{c}+t_i\Sigma d_u,\Sigma') - v_\alpha(\bar{c},\Sigma). \tag{27}$$

Lemma 5. *For all N, $\widehat{\mathbb{K}}_\alpha^N(u,\bar{c},\Sigma) \le \mathbb{K}_\alpha(u,\bar{c},\Sigma)$.*

Proof. For each fixed (x,Σ), the function $\bar{c}^\top x - \alpha\sqrt{x^\top\Sigma x}$ is linear in $\bar{c}$ and thus convex in $\bar{c}$. The maximum over an infinite family of convex functions indexed by x is convex, thus $v_\alpha(\bar{c},\Sigma)$ is convex in $\bar{c}$. Since composition with linear functions preserves convexity, $v_\alpha(\bar{c}+t\Sigma d_u,\Sigma')$ is convex in t. The inequality of the lemma follows from (26). □

Finally, noting that to each $v_\alpha(\bar{c}+t_i\Sigma d_u,\Sigma')$, $i=1,\dots,N$, is associated a program with decision vector $x_i\in\mathbb{R}^n$, and using the update formula for the inverse covariance matrix $[\Sigma']^{-1} = \Sigma^{-1} + uu^\top/\sigma_w^2$, we expand (27) as

$$\widehat{\mathbb{K}}_\alpha^N(u,\bar{c},\Sigma) = \max_{x_1\in\mathcal{X},\dots,x_N\in\mathcal{X}} \sum_{i=1}^{N} p_i\left[(\bar{c}+t_i\Sigma d_u)^\top x_i - \alpha\sqrt{x_i^\top\left(\Sigma^{-1}+uu^\top/\sigma_w^2\right)^{-1}x_i}\right] - v_\alpha(\bar{c},\Sigma)\,. \tag{28}$$

In $\max_u \widehat{\mathbb{K}}_\alpha^N(u,\bar{c},\Sigma)$ the term $-v_\alpha(\bar{c},\Sigma)$ is constant with u, so one can omit it.

7.1. The case $N = 1$. We first study the maximization of $\widehat{\mathbb{K}}^N_\alpha(\cdot, \bar{c}, \Sigma)$ with $N = 1$, where $\mathcal{N}(0,1)$ is reduced to a single mass point. In that case, $t_1 = 0$ and $p_1 = 1$ in (28), and we obtain the problem

$$\max_{u:||u||\leq 1, x: Ax=b, x\succeq 0} \bar{c}^\top x - \alpha\sqrt{x^\top(\Sigma^{-1} + uu^\top/\sigma_w^2)^{-1}x} \ . \tag{29}$$

To get some insights on the nature of (29), suppose momentarily that we are given an optimal solution x for (29), say $\bar{x}$. Then a corresponding optimal u is given by

$$\bar{u} \in \arg\max_{u:\ ||u||=1} \bar{c}^\top\bar{x} - \alpha\sqrt{\bar{x}^\top(\Sigma - \frac{\Sigma uu^\top\Sigma}{u^\top\Sigma u + \sigma_w^2})\bar{x}} = \arg\max_{u:\ ||u||=1} \frac{u^\top\Sigma\bar{x}\bar{x}^\top\Sigma u}{u^\top\Sigma u + \sigma_w^2} \ .$$

This is formally equivalent to the problem solved for establishing Proposition 1, so we immediately obtain $\bar{u} = P^{-1}\Sigma\bar{x}/||P^{-1}\Sigma\bar{x}||$. The maximization of $\widehat{\mathbb{K}}^N_\alpha(\cdot, \bar{c}, \Sigma)$ with $N = 1$ is thus closely related to the fixed-decision heuristic, except that the reference solution $x = \bar{x}$ is now optimal for the problem with the current $\bar{c}$ and the *updated* covariance matrix Σ', which depends on u.

PROPOSITION 3. *With $\alpha > 0$, the problem (29) is equivalent to the following program over $x \in \mathbb{R}^n$, $s \in \mathbb{R}$, and the symmetric matrix $W \in \mathbb{R}^{n\times n}$:*

$$\begin{aligned} &\textit{maximize} \quad \bar{c}^\top x - \alpha\, s \\ &\textit{subject to} \quad Ax = b \ , \quad x \succeq 0 \ , \\ &\qquad\qquad \begin{bmatrix} s & x^\top \\ x & s\Sigma^{-1} + W \end{bmatrix} \succeq 0 \ , \quad \text{trace}(W) = s/\sigma_w^2 \ , \quad \text{rank}(W) = 1 \ , \end{aligned}$$

where u corresponds to a normalized dominant eigenvector of W.

Proof. The constraint $\text{rank}(W) = 1$ implies that $W = \lambda uu^\top$ for some $\lambda \in \mathbb{R}$, with u corresponding to the unique normalized eigenvector of W. Since $\text{trace}(W) = \lambda$, the condition $\text{trace}(W) = s/\sigma_w^2$ implies $\lambda = s/\sigma_w^2$ and thus $W = suu^\top/\sigma_w^2$. By substitution into the SDP constraint, we have

$$\begin{bmatrix} s & x^\top \\ x & s(\Sigma^{-1} + uu^\top/\sigma_w^2) \end{bmatrix} \succeq 0 \ .$$

By the Schur complement formula, this constraint means that either $s = 0$ (and thus $x = 0$), or $s > 0$ and $s - x^\top(s[\Sigma^{-1} + uu^\top/\sigma_w^2])^{-1}x \geq 0$, that is, $s \geq \sqrt{x^\top(\Sigma^{-1} + uu^\top/\sigma_w^2)^{-1}x}$. The objective with $\alpha > 0$ ensures that s is made small, so at optimality we get $s = \sqrt{x^\top(\Sigma^{-1} + uu^\top/\sigma_w^2)^{-1}x}$. □

Proposition 3 suggests the use of a classical convexification technique where the rank-one constraint is relaxed [42], and then a solution u with $||u|| = 1$ is recovered by extracting the dominant eigenvector of W. When the rank-one constraint is relaxed, we must add the constraint $W \succeq 0$ which is no longer implied by the other constraints. Hence a first approximate solution scheme:

1. Solve the semidefinite program

$$\begin{aligned} &\text{maximize} \quad \bar{c}^\top x - \alpha\, s \\ &\text{subject to} \quad Ax = b \ , \quad x \succeq 0 \ , \quad \begin{bmatrix} s & x^\top \\ x & s\Sigma^{-1} + W \end{bmatrix} \succeq 0 \ , \quad \text{trace}(W) = s/\sigma_w^2 \ , \quad W \succeq 0 \ . \end{aligned} \tag{30}$$

2. Return for u the normalized dominant eigenvector of W.

Step 2 is justified by the fact that the best rank-one approximation to W (in the Frobenius norm metric) is the matrix $X = \lambda_{\max}(W)uu^\top$. If W has rank one, then $\lambda_{\max}(W) = \text{trace}(W) = s/\sigma_w^2$.

Legitimate questions are then to ask whether the relaxation (30) should be tight — can we expect that a rank-one matrix W could be optimal — and then, if the relaxation is tight, can we easily recover the rank-one solution, since the interior-point solver might well return another solution with rank greater than one.

In general, these are difficult questions, but it turns out that one can actually say something on the quality of the relaxation. Let $\nu \in \mathbb{R}^n$ with elements $\nu_1 \geq \cdots \geq \nu_n \geq 0$ denote the vector of sorted eigenvalues of W. We have $\text{trace}(W) = \sum_{i=1}^n \nu_i = \sum_{i=1}^n |\nu_i| = ||\nu||_1$. Since L1-norm regularization induces sparsity in the solution, one can see that the constraint $\text{trace}(W) = s/\sigma_w^2$, combined with the fact that s is minimized in the objective, has a beneficial effect on the formulation: it induces zero eigenvalues in W, and thus rank reduction. Nuclear norm minimization, or trace minimization in the special case of positive semidefinite matrices, is a convex technique for inducing low-rank solutions [16]; in our case the trace minimization effect is a byproduct of the original objective.

This analysis reveals that $\alpha > 0$ plays the additional role of weighting a low-rank regularization term for W. For sufficiently high values of α, we are more likely to obtain tighter relaxations. For any value of $\alpha > 0$, the preference will be given to a low-rank solution for W among all optimal solutions.

7.2. The case $N > 1$, $\alpha = 0$. When $N > 1$, the problem takes into account the update of $\bar{c}$ to $\bar{c}'$, which depends on t and u. The following lemma is instrumental for dealing with the nonlinear dependence of d_u on u, as defined in (21). From Theorem 1, we know we can restrict our attention to measurements u with $||u|| = 1$.

LEMMA 6. *The nonconvex set*

$$D = \left\{ d = \frac{u}{\sqrt{u^\top \Sigma u + \sigma_w^2}} : ||u|| = 1, u \in \mathbb{R}^n \right\} \tag{31}$$

admits the alternative representations

$$D = \{d' = P^{-1/2}u' : ||u'|| = 1, u' \in \mathbb{R}^n\}, \tag{32}$$

$$D = \{d'' \in \mathbb{R}^n : \text{trace}(Pd''d''^\top) = 1\}. \tag{33}$$

Proof. If $d \in D$, there exists $u \in \mathbb{R}^n$ with $u^\top u = 1$ such that

$$d = \frac{u}{\sqrt{u^\top \Sigma u + \sigma_w^2}} = \frac{u}{\sqrt{u^\top P u}} = \frac{P^{-1/2}P^{1/2}u}{||P^{1/2}u||} = P^{-1/2}u'$$

where $u' = P^{1/2}u/||P^{1/2}u||$ satisfies $||u'|| = 1$, showing (31) $\rightarrow$ (32). Conversely, if $d' \in D$, there exists $u' \in \mathbb{R}^n$ with $u'^\top u' = 1$ such that

$$d' = P^{-1/2}u' = ||P^{-1/2}u'||v$$

where we have defined $v = P^{-1/2}u'/||P^{-1/2}u'||$; then noting that $v^\top v = 1$, we evaluate

$$[v^\top \Sigma v + \sigma_w^2]^{-1/2} = [v^\top P v]^{-1/2} = \left[\frac{u'^\top P^{-1/2} P P^{-1/2} u'}{||P^{-1/2}u'||^2}\right]^{-1/2} = ||P^{-1/2}u'||,$$

so that $d' = [v^\top \Sigma v + \sigma_w^2]^{-1/2} v$, showing (32) $\rightarrow$ (31) with $u = v = P^{-1/2}u'/||P^{-1/2}u'||$. This establishes the equivalence between (31) and (32).

The well-know identity $\{Q^{1/2}z : ||z|| = 1, z \in \mathbb{R}^n\} = \{z \in \mathbb{R}^n : z^\top Q^{-1} z = 1\}$ applied to $Q = P^{-1}$, and the relation $z^\top Q^{-1} z = \text{trace}(z^\top Q^{-1} z) = \text{trace}(Q^{-1} z z^\top) = \text{trace}(P z z^\top)$, establish the equivalence between (32) and (33). $\square$

The following lemma, due to [44], will be useful to strengthen the relaxations.

LEMMA 7. *Assume* $\mathcal{X} = \{x \in \mathbb{R}^n : Ax = b,\ x \succeq 0\}$ *is bounded and not reduced to* $\{0\}$. *Fix* $\nu \in \mathbb{R}^n$ *with* $\nu_i > 0$ *for each* i, *and define* $\bar{\gamma}_\nu = \sup_{x\in\mathcal{X}} \nu^\top x$. *Then the following relation holds true for any* $x \in \mathcal{X}$:

$$xx^\top \preceq \bar{\gamma}_\nu \mathrm{Diag}(x)\mathrm{Diag}(\nu)^{-1},$$

where $\mathrm{Diag}(z)$ *denotes the diagonal matrix with elements* z_i.

Proof. Since $x \succeq 0$ and $\mathcal{X} \neq \{0\}$, $\bar{\gamma}_\nu > 0$. Since $\mathcal{X}$ is bounded, $\bar{\gamma}_\nu < \infty$. A lemma established in [44] shows that for any $x \in \mathcal{X}$, $\mathrm{diag}(\nu)xx^\top\mathrm{diag}(\nu) \preceq \bar{\gamma}_\nu \mathrm{diag}(\nu)\mathrm{diag}(x)$. Recall that $S \succeq 0$ iff $PSP^\top \succeq 0$, where P can be any invertible matrix. Applying this rule to the inequality with $P = \mathrm{diag}\{\nu\}^{-1}$ establishes the result. □

We have now the necessary ingredients for proposing a solution scheme to (10), first in the case $\alpha = 0$. As usual, $P = \Sigma + \sigma_w^2 \mathrm{I}_n$.

1. Choose a quantization $\{p_i, t_i\}_{i=1}^N$ of $t \sim \mathcal{N}(0,1)$.

Construct the symmetric matrices

$$C_i = \frac{1}{2}\left[\begin{array}{c|c|c} 0 & \bar{c}^\top & 0^\top \\ \hline \bar{c} & 0 & t_i\Sigma \\ \hline 0 & t_i\Sigma & 0 \end{array}\right] \in \mathbb{R}^{(2n+1)\times(2n+1)}, \qquad 1 \le i \le N.$$

2. Generate a set of vectors $\{\nu_\ell\}_{\ell=1}^M$, $\nu_\ell \succ 0$, and evaluate

$$\bar{\gamma}_\ell = \max_{x\in\mathcal{X}} \nu_\ell^\top x.$$

3. Solve the following SDP over the symmetric optimization matrices $Y \in \mathbb{R}^{n\times n}$ and

$$Z_i = \left[\begin{array}{c|c|c} Z_i^{11} & Z_i^{1x} & Z_i^{1d} \\ \hline Z_i^{x1} & Z_i^{xx} & Z_i^{xd} \\ \hline Z_i^{d1} & Z_i^{dx} & Z_i^{dd} \end{array}\right] = \left[\begin{array}{c|c|c} 1 & x_i^\top & d^\top \\ \hline x_i & Z_i^{xx} & Z_i^{xd} \\ \hline d & Z_i^{dx} & Y \end{array}\right] \in \mathbb{R}^{(2n+1)\times(2n+1)}, \qquad 1 \le i \le N \ :$$

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_{i=1}^N p_i \mathrm{trace}(C_i Z_i) \\ \text{subject to} \quad \forall\, i: \quad & Z_i \succeq 0, \\ & Z_i^{11} = 1, \quad AZ_i^{x1} = b, \quad Z_i^{x1} \succeq 0, \\ & AZ_i^{xx}A^\top = bb^\top,\ [Z_i^{xx}]_{qr} \ge 0 \quad \forall q, r, \\ & Z_i^{xx} \preceq \bar{\gamma}_\ell \mathrm{Diag}(Z_i^{x1})\mathrm{Diag}(\nu_\ell)^{-1} \quad \forall \ell, \\ & Z_i^{dd} = Y, \\ & \mathrm{trace}(PY) = 1. \end{aligned}$$

4. Return for u the eigenvector associated to the largest eigenvalue of Y.

The scheme is based on the relation

$$(\bar{c} + t_i\Sigma d)^\top x_i = \frac{1}{2}\mathrm{trace}\left(\left[\begin{array}{ccc} 0 & \bar{c}^\top & 0^\top \\ \bar{c} & 0 & t_i\Sigma \\ 0 & t_i\Sigma & 0 \end{array}\right]\left[\begin{array}{c} 1 \\ x_i \\ d \end{array}\right]\left[\begin{array}{c} 1 \\ x_i \\ d \end{array}\right]^\top\right) \ ,$$

where we define

$$Z_i = \begin{bmatrix} 1 \\ x_i \\ d \end{bmatrix} \begin{bmatrix} 1 \\ x_i \\ d \end{bmatrix}^\top = \begin{bmatrix} 1 & x_i^\top & \dfrac{u^\top}{\sqrt{u^\top \Sigma u + \sigma_w^2}} \\ x_i & x_i x_i^\top & \dfrac{x_i u^\top}{\sqrt{u^\top \Sigma u + \sigma_w^2}} \\ \dfrac{u}{\sqrt{u^\top \Sigma u + \sigma_w^2}} & \dfrac{u x_i^\top}{\sqrt{u^\top \Sigma u + \sigma_w^2}} & \dfrac{u u^\top}{u^\top \Sigma u + \sigma_w^2} \end{bmatrix},$$

which is semidefinite positive and has rank 1.

The constraints $Ax = b$ and $x \succeq 0$ imply $Ax_i x_i^\top A^\top = bb^\top$ (linear equality between matrices) and $[x_i x_i^\top]_{qr} \geq 0$ for $1 \leq q, r \leq n$ (nonnegativity of the matrix $x_i x_i^\top$). In terms of the matrix Z_i, we write $AZ_i^{xx}A^\top = bb^\top$ and $[Z_i^{xx}]_{qr} \geq 0$. The constraint $\text{trace}(Z_i^{xx} AA^\top) = b^\top b$ would be of no use here because it is implied by $AZ_i^{xx}A^\top = bb^\top$, so we use Lemma 7 to further control Z_i^{xx} by the constraint $Z_i^{xx} \preceq \bar{\gamma}_\ell \text{Diag}(Z_i^{x1})\text{Diag}(\nu_\ell)^{-1}$. A single inequality suffices since we impose $Z_i^{x1} \in \mathcal{X}$, which is bounded by assumption. Introducing additional valid inequalities can strengthen the relaxation but can also increase the rank of the solution Y, since the minimal rank solution is affected by the number of constraints [32, 3].

We introduce the variable $Y = uu^\top/(u^\top \Sigma u + \sigma_w^2)$ to write the constraints $Z_i^{uu} = Y = Z_j^{uu}$, $1 \leq i, j \leq N$. From Theorem 1 we want $||u|| = 1$. From Lemma 6, this is possible by imposing $\text{trace}(PY) = 1$ and $\text{rank}(Y) = 1$. All the rank-one constraints are then relaxed. We obtain our approximation of the optimal u through the normalized eigenvector associated to the largest eigenvalue of Y, since we have $Yu = (u^\top u / u^\top Pu)u = \lambda u$ with $\lambda = u^\top Pu$ when Y follows its rank-one definition.

7.3. General case: $N > 1$, $\alpha > 0$. The solution scheme for the general case combines the techniques used in the two preceding cases.

1. Choose a quantization $\{p_i, t_i\}_{i=1}^N$ of $t \sim \mathcal{N}(0,1)$.

Define the symmetric matrices

$$C_i = \frac{1}{2} \begin{bmatrix} 0 & \bar{c}^\top & 0^\top \\ \bar{c} & 0 & t_i \Sigma \\ 0 & t_i \Sigma & 0 \end{bmatrix} \in \mathbb{R}^{(2n+1)\times(2n+1)}, \qquad 1 \leq i \leq N.$$

2. Generate a set of vectors $\{\nu_\ell\}_{\ell=1}^M$, $\nu_\ell \succ 0$, and evaluate $\bar{\gamma}_\ell = \max_{x \in \mathcal{X}} \nu_\ell^\top x$.
3. Solve the following SDP over $u \in \mathbb{R}^n$, $s_i \in \mathbb{R}$ and the symmetric matrices Y, $W_i \in \mathbb{R}^{n \times n}$, and

$$Z_i = \begin{bmatrix} Z_i^{11} & Z_i^{1x} & Z_i^{1d} \\ Z_i^{x1} & Z_i^{xx} & Z_i^{xd} \\ Z_i^{d1} & Z_i^{dx} & Z_i^{dd} \end{bmatrix} = \begin{bmatrix} 1 & x_i^\top & d^\top \\ x_i & Z_i^{xx} & Z_i^{xd} \\ d & Z_i^{dx} & Y \end{bmatrix} \in \mathbb{R}^{(2n+1)\times(2n+1)} \quad , \quad 1 \leq i \leq N \quad :$$

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_{i=1}^N p_i [\text{trace}(C_i Z_i) - \alpha s_i] \\ \text{subject to} \quad & \text{trace}(PY) = 1 \ , \\ \forall i: \quad & Z_i \succeq 0 \ , \\ & \text{trace}(W_i) = s_i / \sigma_w^2 \ , \\ & \begin{bmatrix} s_i & Z_i^{1x} \\ Z_i^{x1} & s_i \Sigma^{-1} + W_i \end{bmatrix} \succeq 0 \ , \end{aligned}$$

$$\begin{bmatrix} W_i & w_i \\ w_i^\top & 1 \end{bmatrix} \succeq 0 \ ,$$
$$\begin{bmatrix} Y & w_i \\ w_i^\top & \mathrm{trace}(PW_i) \end{bmatrix} \succeq 0 \ ,$$
$$Z_i^{11} = 1 \ , \quad AZ_i^{x1} = b \ , \quad Z_i^{x1} \succeq 0 \ ,$$
$$AZ_i^{xx}A^\top = bb^\top \ , \quad [Z_i^{xx}]_{qr} \geq 0 \quad \forall q, r \ ,$$
$$Z_i^{xx} \preceq \bar{\gamma}_\ell \mathrm{Diag}(Z_i^{x1})\mathrm{Diag}(\nu_\ell)^{-1} \quad \forall \ell \ ,$$
$$Z_i^{dd} = Y \ .$$

4. Return for u the eigenvector associated to the largest eigenvalue of Y.

In the SDP, using $||u|| = 1$ we define $Y = dd^\top = uu^\top / u^\top P u = uu^\top / \mathrm{trace}(Puu^\top)$. We have $\mathrm{trace}(PY) = \mathrm{trace}(u^\top P u / u^\top P u) = 1$. For each i, we define $s_i \geq 0$ and $w_i w_i^\top = W_i = s_i uu^\top / \sigma_w^2$. We have $\mathrm{trace}(W_i) = s_i/\sigma_w^2$. Assuming $s_i > 0$, we have $uu^\top = \sigma_w^2 W_i / s_i$, so we can rewrite Y as

$$Y = \frac{\sigma_w^2 W_i / s_i}{\mathrm{trace}(P\sigma_w^2 W_i / s_i)} = \frac{w_i w_i^\top}{\mathrm{trace}(PW_i)} \ .$$

We relax the definitions of W_i and Y to $W_i \succeq w_i w_i^\top$ and $Y \succeq w_i w_i^\top / \mathrm{trace}(PW_i)$, which can be expressed, using a Schur complement logic, by the constraints

$$\begin{bmatrix} W_i & w_i \\ w_i^\top & 1 \end{bmatrix} \succeq 0 \ , \quad \begin{bmatrix} Y & w_i \\ w_i^\top & \mathrm{trace}(PW_i) \end{bmatrix} \succeq 0 \ .$$

The rest of the construction of the program follows the logic of Sections 7.1 and 7.2.

8. Numerical test. We begin by testing the different approximation methods proposed in this paper on a series of random LPs with $n = 20$ optimization variables, n positivity constraints, and $m = 5$ equality constraints. The goal of these experiments is to evaluate the ability of the different methods of maximizing the objective $\mathbb{K}_\alpha(\cdot, \bar{c}, \Sigma)$ stated in (9) for some arbitrary values of $\bar{c}, \Sigma$.

We compare the following optimization strategies:

- RAND: maximum of $\mathbb{K}_\alpha(u, \bar{c}, \Sigma)$ over 1000 random directions $u = u'/||u'||$ with $u' \sim \mathcal{N}(0, \mathrm{I}_n)$.
- EIG: $\mathbb{K}_\alpha(u, \bar{c}, \Sigma)$ with u set to the eigenvector relative to the largest eigenvalue of Σ.
- UNIT: maximum of $\mathbb{K}_\alpha(e_j, \bar{c}, \Sigma)$ over the unit vectors e_j, $1 \leq j \leq n$.
- SDP-1: $\mathbb{K}_\alpha(u, \bar{c}, \Sigma)$ with u set to the output of the one-sample approximation scheme of Section 7.1.
- SDP-2: $\mathbb{K}_\alpha(u, \bar{c}, \Sigma)$ with u set to the output of the general scheme of Section 7.3, using $N = 5$ samples for the approximating the expectation inside the optimization program, and $M = 5$ random positive directions ν_ℓ.

In our simulations, a run is defined as follows:

1. Generation of an initial random problem ($k = 1$):
- $\bar{c}_1 \sim \mathcal{N}(0, \mathrm{I}_n)$,
- $\Sigma_1 = (S + S^\top)(S + S^\top) \in \mathbb{R}^{n \times n}$ with $S_{ij} \sim \mathcal{N}(0, 1)$,
- $A \in \mathbb{R}^{m \times n}$ with A_{ij} drawn uniformly in $[0, 1]$ and rejection of A if $\mathrm{rank}(A) < \min\{m, n\}$,
- $b = A\beta$, where $\beta \in \mathbb{R}^n$ has coordinates $\beta_i = |\beta_i'|$, $\beta' \sim \mathcal{N}(0, 1)$.

2. Optimization of a measurement u by the 5 methods.

3. Selection of u_k from SDP-2, and update of $\bar{c}_k, \Sigma_k$ to $\bar{c}_{k+1}, \Sigma_{k+1}$, pretending that $y = 0$. (Thus $\bar{c}_{k+1} = \bar{c}_k$ in this setting.)

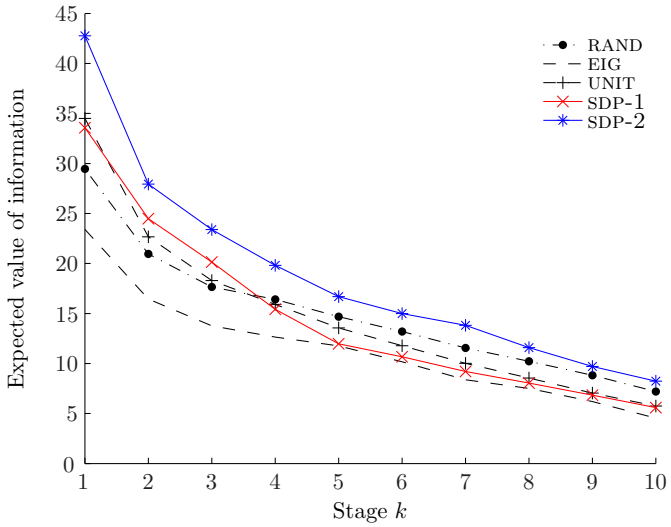

FIGURE 1. Optimization results averaged over 25 random problems. Each problem is updated 10 times, using the previous measurement determined by SDP-2. Higher values indicate better solutions to the maximization of $\mathbb{K}(\cdot, \bar{c}_k, \Sigma_k)$ over the sphere in $\mathbb{R}^{20}$.

4. Return to step 2 with k incremented until 10 stages have been carried out.

In all these experiments we set $\sigma_w^2 = 1$ and $\alpha = 1$.

Experimentally, our random problems are such that the eigenvalues of Σ_k are distributed in some range during the first stages $k = 1, 2, \ldots$ of a run, and then tend to be more concentrated at later stages.

The experiments are implemented in Matlab 7.10. The LPs and SOCPs are solved with Cplex 12.2.0.2. The semidefinite programs are formulated and solved through cvx in Matlab [22, 23]. Averaged results over 25 random problem instances are presented on Figure 1. The values $\mathbb{K}_\alpha(u, \bar{c}^k, \Sigma^k)$ are estimated by using an optimal quadratic quantization on 21 samples.

Recall that $\mathbb{K}_\alpha(u, \bar{c}^k, \Sigma^k)$ is a measure of expected improvement in the quality of the robust solution. This quantity can thus be used as a performance measure. A policy that consistently achieves higher expected improvement than another policy will also achieve (on average) better robust solutions. We see that the values of $\mathbb{K}_\alpha$ exhibit a downward trend over time, reflecting the fact that the marginal value of a measurement tends to decrease as the uncertainty is progressively reduced. From the numerical results, it appears that direct search RAND already begins to break down on these small problems. In fact, on each individual problem, SDP-2 consistently outperforms RAND. We also observe an improvement from SDP-1 to SDP-2. The baseline policy EIG generally gives the worst results, showing that optimal uncertainty reduction does not necessarily lead to solutions with higher economic value. Overall, Figure 1 suggests that SDP-2 consistently finds better solutions, confirming the value of a better approximation of the value of information.

Next, we present results obtained on a randomly generated MDP with $|S| = 10$ states and $|A| = 2$ actions. This time, we compare the algorithms on the basis of the measurement policies they induce over a sequence of 10 measurements. We are interested in the true value of the MDP policy that is obtained after k measurements for $k = 1, \ldots, 10$, that is,

$$f(x_k, c^{\text{true}}) = x_k^\top c^{\text{true}} \quad , \qquad\qquad x_k \in \underset{x:\, Ax=b,\, x \succeq 0}{\arg\max}\; x^\top \bar{c}_k - \alpha\sqrt{x^\top \Sigma_k x} \quad ,$$

where $\bar{c}_k$, Σ_k are the end-result of the method that optimizes the measurement vectors $u_1, \ldots, u_k$, and of the random observations $y_1 = u_1^\top c^{\text{true}} + w_1, \ldots, y_k = u_k^\top c^{\text{true}} + w_k$.

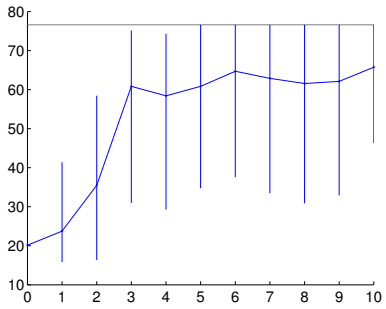

FIGURE 2. Distribution of the true performance with EIG, for a growing number of measurements.

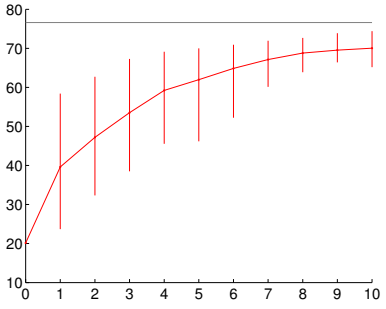

FIGURE 3. Distribution of the true performance with SDP-1, for a growing number of measurements.

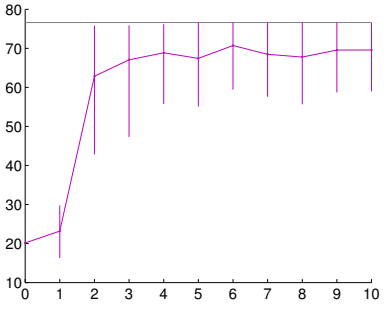

FIGURE 4. Distribution of the true performance with SDP-2, for a growing number of measurements.

In the MDP language, x_k encodes the stochastic policy π_k that optimally solves the robust MDP (17) posed on the beliefs after k measurements. The vector c^{true} encodes the true reward function $r^{\text{true}}(\cdot,\cdot)$ of the MDP, and $f(x_k, c^{\text{true}})$ is the expected value of π_k on the true MDP, that we can also write as

$$V^{\pi_k} = \mathbb{E}\Big\{ \sum_{t=0}^{\infty} \gamma^t r^{\text{true}}(s_t, \pi_k(s_t)) \,\Big|\, s_0 \sim q_0 \Big\} \ ,$$

for some initial state distribution q_0 determined by the vector b. Because the sequence $w_1, \ldots, w_k$ of observation noise is random, one should actually look at the distribution of $f(x_k, c^{\text{true}}) = V^{\pi_k}$.

Figures 2 to 4 show the results of 100 simulations run on the same fixed MDP. All simulations start from a same belief distribution $(\bar{c}_0, \Sigma_0)$. There are 3 graphs, corresponding to EIG, SDP-1 and SDP-2 respectively. The same 100 samples of a sequence of Gaussian noises $\{w_k : 1 \leq k \leq 10\}$ for making 10 consecutive measurements are used for comparing the 3 methods. The true maximum is indicated by a horizontal line. We have plotted the curve of the estimated mean of V^{π_k} over the 100 samples as a function of the number $k = 0, \ldots, 10$ of past measurements. We have also plotted vertical bars between the 25-th and the 75-th percentiles of the distribution of V^{π_k}. The support of V^{π_k} cannot cross the horizontal line of the true maximum.

Figure 2 shows that EIG performs well on average, but exhibits a high degree of variation. In other words, there is a higher probability that optimal uncertainty reduction will lead to an MDP policy π_k that performs poorly on the true problem.

By contrast, Figure 3 shows that SDP-1 underperforms EIG on average, but the distribution of performance is more tightly concentrated around the mean (that is, the resulting MDP policy is more robust). Lastly, Figure 4 shows that SDP-2 dominates the other two methods on average, while also achieving smaller variance than EIG.

9. Conclusion. We have posed an optimal learning problem in which a decision-maker improves a robust solution to a stochastic linear program by sequentially collecting information about the unknown objective coefficients. A single piece of information takes the form of a linear combination (a "blend") of the true underlying objective vector, subject to Gaussian noise. Bayesian updating is then used to combine this new information with a multivariate normal prior distribution on the unknown parameters. Previous work has considered weighted sums of unknown parameters where the weights were pre-specified by a linear regression model. To our knowledge, the present paper is the first to pose the continuous optimization problem of choosing the optimal weight vector. Our formulation of this problem allows for both risk-neutral and risk-averse decision-makers.

Within this setting, we have proposed two policies for choosing information blends. The first was shown to optimize uncertainty reduction (analogous to active learning methods in statistics) by

selecting the largest eigenvector of the posterior covariance matrix. The second approximates the optimal solution to an expected improvement criterion (a nonconvex optimization problem) via an SDP reformulation technique. The approach is applicable to robust LP formulations of Markov decision process problems, where risk-averse decision-making policies are desired. We show that our approach generalizes a previous heuristic for such problems. In numerical examples, the SDP approximation consistently outperforms a number of benchmarks. We believe that the present paper contributes to the interface of robust optimization and optimal learning, and that the idea of information blending offers a new way to think about sequential information collection.

Appendix. Proof of Proposition 1.

Assuming $\alpha > 0$, we have, from (18),

$$\arg\max_{u\in\mathbb{B}} \widetilde{\mathbb{K}}_\alpha(u,\bar{c},\Sigma) = \arg\max_{u\in\mathbb{B}} \sqrt{\bar{x}^\top\Sigma\bar{x}} - \sqrt{\bar{x}^\top\left(\Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top\Sigma u + \sigma_w^2}\right)\bar{x}} \ .$$

If $\Sigma\bar{x} = 0$, then any $u \in \mathbb{B}$ is optimal. Otherwise, $\Sigma\bar{x} \neq 0$, and we can justify that any optimal u will satisfy $u^\top u = 1$ by the proof technique used in Theorem 1. Then we have

$$\begin{aligned}
\arg\max_{u\in\mathbb{B}} \widetilde{\mathbb{K}}_\alpha(u,\bar{c},\Sigma) &= \arg\max_{u:\,||u||=1} \sqrt{\bar{x}^\top\Sigma\bar{x}} - \sqrt{\bar{x}^\top\left(\Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top\Sigma u + \sigma_w^2}\right)\bar{x}} \\
&= \arg\min_{u:\,||u||=1} \sqrt{\bar{x}^\top\left(\Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top\Sigma u + \sigma_w^2}\right)\bar{x}} \\
&= \arg\min_{u:\,||u||=1} \bar{x}^\top\left(\Sigma - \frac{\Sigma u u^\top \Sigma}{u^\top\Sigma u + \sigma_w^2}\right)\bar{x} \\
&= \arg\max_{u:\,||u||=1} \frac{\bar{x}^\top\Sigma u u^\top\Sigma\bar{x}}{u^\top\Sigma u + \sigma_w^2} \\
&= \arg\max_{u:\,||u||=1} \frac{u^\top\Sigma\bar{x}\bar{x}^\top\Sigma u}{u^\top(\Sigma + \sigma_w^2 \mathrm{I}_n)u} \ .
\end{aligned}$$

We can then proceed as in the proof of Theorem 2, or observe that an optimal solution $\bar{u}$ can be obtained by considering the generalized eigenvalue problem $(\Sigma\bar{x}\bar{x}^\top\Sigma)u = \lambda(\Sigma + \sigma_w^2\mathrm{I}_n)u$ and taking for $\bar{u}$ a normalized generalized vector associated to the largest generalized eigenvalue λ. Since $(\Sigma + \sigma_w^2\mathrm{I}_n)$ is nonsingular, the generalized eigenvalue problem is equivalent to the standard eigenvalue problem $(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}(\Sigma\bar{x}\bar{x}^\top\Sigma)u = \lambda u$, which is of the form

$$fg^\top u = \lambda u \quad \text{with } f = (\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma\bar{x} \ , \quad g = \Sigma\bar{x}.$$

Therefore, the rank-one matrix $fg^\top$ has a single positive eigenvalue $g^\top f/||f||$ with a normalized eigenvector $f/||f||$ or $-f/||f||$, and $\bar{u} = \pm(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma\bar{x}/||(\Sigma + \mathrm{I}_n\sigma_w^2)^{-1}\Sigma\bar{x}||$.

References

[1] Alizadeh, F., D. Goldfarb. 2003. Second-order cone programming. *Math. Programming* **95** 3–51.

[2] Auer, P., N. Cesa-Bianchi, P. Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **47** 235–256.

[3] Barvinok, A. 2001. A remark on the rank of positive semidefinite matrices subject to affine constraints. *Discrete and Computational Geometry* **25** 23–31.

[4] Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

[5] Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Math. Programming* **99**(2) 351–376.

[6] Bertsimas, D., M. Sim. 2004. The price of robustness. *Oper. Res.* **51**(1) 35–53.

[7] Chick, S. E. 2006. Subjective probability and Bayesian methodology. S. G. Henderson, B. L. Nelson, eds., *Handbooks of Operations Research and Management Science, vol. 13: Simulation*. North-Holland Publishing, Amsterdam, 225–258.

[8] Chick, S. E., N. Gans. 2009. Economic analysis of simulation selection problems. *Management Sci.* **55**(3) 421–437.

[9] Cohn, D.A., Z. Ghahramani, M.I. Jordan. 1996. Active learning with statistical models. *J. Artificial Intelligence Res.* **4** 129–145.

[10] Delage, E., S. Mannor. 2007. Percentile optimization in uncertain Markov decision processes with application to efficient exploration. *Proc. 24th Internat. Conf. Machine Learning (ICML-2007)*. ACM, 225–232.

[11] Delage, E., S. Mannor. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**(1) 203–213.

[12] Dellino, G., J.P.C. Kleijnen, C. Meloni. 2012. Robust optimization in simulation: Taguchi and Krige combined. *Informs J. Comp.* **24**(3) 471–484.

[13] D'Epenoux, F. 1960. Sur un problème de production et de stockage dans l'aléatoire. *Revue Française de Recherche Opérationnelle* **14** 3–16.

[14] Dontchev, A.L., R.T. Rockafellar. 2009. *Implicit Functions and Solution Mappings: A view from Variational Analysis*. Springer, New York.

[15] Doob, J.L. 1949. Application of the theory of martingales. *Le calcul des probabilités et ses applications*. Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13, CNRS, Paris, 23–27.

[16] Fazel, M., H. Hindi, S. Boyd. 2001. A rank minimization heuristic with application to minimum order system approximation. *Proc. 2001 American Control Conference*. Arlington, VA, 4734 – 4739.

[17] Frazier, P. I., W. B. Powell, S. Dayanik. 2009. The knowledge-gradient policy for correlated normal rewards. *Informs J. Comp.* **21**(4) 599–613.

[18] Frazier, P.I., W.B. Powell, S. Dayanik. 2008. A knowledge gradient policy for sequential information collection. *SIAM J. Control Optim.* **47**(5) 2410–2439.

[19] Ghaffari-Hadigheh, A., T. Terlaky. 2006. Sensitivity analysis in linear optimization: Invariant support set intervals. *Eur. J. Oper. Res.* **169** 1158–1175.

[20] Gittins, J. C., K. D. Glazebrook, R. Weber. 2011. *Multi-armed bandit allocation indices*. 2nd ed. John Wiley and Sons, Chichester, UK.

[21] Graf, S., H. Luschgy. 2000. *Foundations of Quantization for Probability Distributions*. Springer-Verlag, Berlin, Germany.

[22] Grant, M., S. Boyd. 2008. Graph implementations for nonsmooth convex programs. V. Blondel, S. Boyd, H. Kimura, eds., *Recent Advances in Learning and Control – A tribute to M. Vidyasagar*. LNCIS, Springer, New York, NY, 95–110.

[23] Grant, M., S. Boyd. 2011. CVX: Matlab software for disciplined convex programming, version 1.21. http://cvxr.com/cvx.

[24] Gupta, S.S., K.J. Miescke. 1996. Bayesian look ahead one-stage sampling allocations for selection of the best population. *J. Statist. Planning and Inference* **54**(2) 229–244.

[25] Iyengar, G. N. 2005. Robust dynamic programming. *Math. Oper. Res.* **30**(2) 257–280.

[26] Jones, D.R., M. Schonlau, W.J. Welch. 1998. Efficient global optimization of expensive black-box functions. *J. Global Optim.* **13**(4) 455–492.

[27] Kim, S.-H., B. L. Nelson. 2006. Selecting the best system. S. G. Henderson, B. L. Nelson, eds., *Handbooks of Operations Research and Management Science, vol. 13: Simulation*. North-Holland Publishing, Amsterdam, 501–534.

[28] Mangasarian, O.L., T.H. Shiau. 1986. A variable-complexity norm maximization problem. *SIAM J. Algebraic Discrete Methods* **7**(3) 455–461.

[29] Negoescu, D. M., P. I. Frazier, W. B. Powell. 2010. The knowledge-gradient algorithm for sequencing experiments in drug discovery. *Informs J. Comp.* **23**(3) 346–363.

[30] Nilim, A., L. El Ghaoui. 2005. Robust control of Markov decision processes with uncertain transition matrices. *Oper. Res.* **53**(5) 780–798.

[31] Pages, G., J. Printems. 2003. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Applications* **9**(2) 135–166.

[32] Pataki, G. 1998. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.* **23**(2) 339–358.

[33] Powell, W. B., I. O. Ryzhov. 2012. *Optimal Learning*. Wiley, Hoboken, NJ.

[34] Puterman, M.L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, NY.

[35] Qu, H., I. O. Ryzhov, M. C. Fu. 2012. Ranking and selection with unknown correlation structures. *Proc. 2012 Winter Simulation Conference*.

[36] Regan, K., C. Boutilier. 2010. Robust policy computation in reward-uncertain MDPs using nondominated policies. *Proc. 24th AAAI Conf. Artificial Intelligence (AAAI-10)*. The AAAI Press, Menlo Park, CA, 1127–1133.

[37] Rockafellar, R.T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

[38] Ruszczyński, A. 2010. Risk-averse dynamic programming for Markov decision processes. *Math. Programming* **125**(2) 235–261.

[39] Ryzhov, I. O., B. Defourny, W. B. Powell. 2012. Ranking and selection meets robust optimization. *Proc. 2012 Winter Simulation Conference*.

[40] Ryzhov, I.O., W.B. Powell. 2011. Information collection on a graph. *Oper. Res.* **59**(1) 188–201.

[41] Ryzhov, I.O., W.B. Powell. 2012. Information collection for linear programs with uncertain objective coefficients. *SIAM Journal on Optimization* To appear.

[42] Shor, N.Z. 1987. Quadratic optimization problems. *Soviet Journal of Circuits and Systems Sciences* **25** 1–11.

[43] Waeber, R., P. I. Frazier, S. G. Henderson. 2010. Performance measures for ranking and selection procedures. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, E. Yücesan, eds., *Proc. 2010 Winter Simulation Conference*. 1235–1245.

[44] Zheng, X.J., X.L. Sun, D. Li. 2011. Convex relaxations for nonconvex quadratically constrained quadratic programming: matrix cone decomposition and polyhedral approximation. *Math. Programming, Ser. B* **129**(2) 301–329.