

Robust and Stochastic Optimization Notes

Kevin Kircher, Cornell MAE

These are partial notes from ECE 6990, Robust and Stochastic Optimization, as taught by Prof. [Eilyan Bitar](#) at Cornell University in the fall of 2015. They cover three approaches to convex optimization with uncertain input data:

- robust convex programming, where a solution must be feasible for all possible realizations of the uncertain parameters,
- chance-constrained programming, where a solution must be feasible with high probability under the uncertain parameter distribution, and
- sampled convex programming, where a solution must be feasible for a number of independent samples of the uncertain parameters.

These notes assume knowledge of convex optimization at the level of Boyd and Vandenberghe [1], with which we attempt to maintain consistent notation. Some course material, such as affine policies for multistage stochastic programming, has been omitted. If you find any errors – which are the fault of the scribe, not the lecturer – feel free to let [Kevin](#) know.

Contents

- 1 Three paradigms for optimization under uncertainty 3**
 - 1.1 Robust convex programming 4
 - 1.2 Chance-constrained programming 5
 - 1.3 Sampled convex programming 6

- 2 Sample bounds for sampled convex programming 7**
 - 2.1 A loose, general bound 7
 - 2.2 A tight bound for structured problems 9
 - 2.3 Comparison 10

- 3 Convex, conservative approximations to chance-constrained programs 11**
 - 3.1 Generating functions 11
 - 3.2 Enlarging the inner approximation 13
 - 3.3 Joint chance constraints 15
 - 3.4 Bernstein approximations for affinely perturbed constraints 16

- 4 Tractable robust convex programs 18**
 - 4.1 Robust linear programs 18

- 5 Affine policies for stochastic linear programs 22**
 - 5.1 Primal affine approximation 23
 - 5.2 Dual affine approximation 25
 - 5.3 Example: a two-stage stochastic linear program 27

Chapter 1

Three paradigms for optimization under uncertainty

We consider the uncertain convex program

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && f(x, \delta) \leq 0 \\ & && x \in X, \quad \delta \in \Delta, \end{aligned} \tag{UCP}$$

where $x \in \mathbf{R}^n$ is the optimization variable, $\delta \in \mathbf{R}^d$ is an uncertain parameter, $f : \mathbf{R}^n \times \mathbf{R}^d \rightarrow \mathbf{R}$ is convex, $X \subseteq \mathbf{R}^n$ is closed and convex, and $\Delta \subseteq \mathbf{R}^d$ is the set of all uncertain parameters.

This problem is ill-posed in its current form, since we have no information about δ other than that it belongs to Δ . We also have not specified for which δ the constraint $f(x, \delta) \leq 0$ must hold, *e.g.*, for all $\delta \in \Delta$, for randomly selected δ with some probability, or for particular realizations of δ . Each of these approaches gives rise to a well-posed problem closely related to UCP.

Although it may seem restrictive to require a linear objective function and a single inequality constraint coupling x and δ , in fact UCP is general. To see this, consider the problem

$$\begin{aligned} & \text{minimize} && f_0(y, \delta) \\ & \text{subject to} && f_i(y, \delta) \leq 0, \quad i = 1, \dots, m \\ & && y \in Y, \quad \delta \in \Delta, \end{aligned} \tag{1.1}$$

where $y \in \mathbf{R}^{n-1}$, Y is closed and convex, and f_0, \dots, f_m are convex. The epigraph form of this problem is

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && f_0(y, \delta) \leq t \\ & && f_i(y, \delta) \leq 0, \quad i = 1, \dots, m \\ & && y \in Y, \quad \delta \in \Delta. \end{aligned}$$

This problem fits into the UCP framework with $x = (y, t)$, $X = Y \times \mathbf{R}$, and $f(x, \delta) = \max \{f_0(y, \delta) - t, f_1(y, \delta), \dots, f_m(y, \delta)\}$.

Example (optimal control). A special case of UCP is the open-loop optimal control problem

$$\begin{aligned}
& \text{minimize} && f_0(z_0, \dots, z_T, u_0, \dots, u_{T-1}, w_0, \dots, w_{T-1}) \\
& \text{subject to} && z_{t+1} = A_t z_t + B_t u_t + C_t w_t \\
& && u_t \in U_t \\
& && z_{t+1} \in Z_{t+1} \\
& && (z_0, w_0, \dots, w_{T-1}) \in \Delta,
\end{aligned} \tag{1.2}$$

where the constraints hold for $t = 0, \dots, T-1$ and T is the time horizon. At each time t , $z_t \in \mathbf{R}^{n_z}$ is the system state, $u_t \in \mathbf{R}^{n_u}$ is the control action, and $w_t \in \mathbf{R}^{n_w}$ is the disturbance. The initial state z_0 and disturbances w_0, \dots, w_{T-1} are uncertain. The optimization variables are the controls u_0, \dots, u_{T-1} and states z_1, \dots, z_T .

To see that problem (1.2) is a special case of UCP, we can write each equality constraint $z_{t+1} = A_t z_t + B_t u_t + C_t w_t$ as the two inequality constraints $z_{t+1} - (A_t z_t + B_t u_t + C_t w_t) \leq 0$ and $-z_{t+1} + A_t z_t + B_t u_t + C_t w_t \leq 0$. With this transformation, problem (1.2) is a special case of problem (1.1) with $y = (u_0, \dots, u_{K-1}, z_1, \dots, z_K)$, $Y = U_0 \times \dots \times U_{T-1} \times Z_1 \times \dots \times Z_T$, $\delta = (z_0, w_0, \dots, w_{K-1})$, and appropriately defined constraint functions $f_i, i = 1, \dots, m$. Since problem (1.1) reduces to UCP, so does the optimal control problem (1.2). The resulting UCP instance has $n = T(n_u + n_z) + 1$ optimization variables and $d = n_z + T n_w$ uncertain parameters.

1.1 Robust convex programming

A *robust convex program* requires that the uncertain constraint $f(x, \delta) \leq 0$ in UCP hold for all possible δ :

$$\begin{aligned}
& \text{minimize} && c^T x \\
& \text{subject to} && f(x, \delta) \leq 0 \text{ for all } \delta \in \Delta \\
& && x \in X.
\end{aligned} \tag{RCP}$$

When one exists, we denote a solution of RCP by x_k^* .

RCP is a semi-infinite problem, since the constraint “for all $\delta \in \Delta$ ” is infinite-dimensional, but x is finite-dimensional. The infinite-dimensional constraint makes RCP intractable in general. RCP is also conservative, since low-probability realizations of δ may significantly increase the optimal value. A goal of these notes is to derive convex, finite-dimensional approximations of RCP.

Example (minimax optimization). The minimax problem

$$\begin{aligned}
& \text{minimize} && \sup \{f(x, \delta) \mid \delta \in \Delta\} \\
& \text{subject to} && x \in X,
\end{aligned}$$

where f is convex and X is closed and convex, can be written as an RCP. To see this, consider the epigraph form

$$\begin{aligned}
& \text{minimize} && t \\
& \text{subject to} && \sup \{f(x, \delta) \mid \delta \in \Delta\} \leq t \\
& && x \in X.
\end{aligned}$$

The first constraint is equivalent to $f(x, \delta) \leq t$ for all $\delta \in \Delta$, so we have an RCP.

Example (semidefinite programming). The standard form of a semidefinite program is

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n \preceq 0 \\ & && Ax = b, \end{aligned} \tag{SDP}$$

where $x \in \mathbf{R}^n$, $F_0, \dots, F_n \in \mathbf{S}_+^n$, and $A \in \mathbf{R}^{m \times n}$. The constraint $F(x) \preceq 0$ holds if and only if $\delta^T F(x) \delta \leq 0$ for all δ with $\|\delta\|_2 = 1$. An equivalent problem, therefore, is the RCP

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \delta^T (F_0 + x_1 F_1 + \cdots + x_n F_n) \delta \text{ for all } \delta \in \Delta \\ & && Ax = b, \end{aligned}$$

where $\Delta = \{\delta \mid \|\delta\|_2 = 1\}$. We will see that this problem can be solved approximately using linear programming. This gives a method for handling very large semidefinite programs, for which no efficient solver currently exists.

1.2 Chance-constrained programming

A *chance-constrained program* assumes a distribution on δ and bounds the probability of constraint violation by $\varepsilon \in [0, 1]$:

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \mathbf{P}\{f(x, \delta) \leq 0\} \geq 1 - \varepsilon \\ & && x \in X. \end{aligned} \tag{CCP}_\varepsilon$$

When one exists, we denote a solution of CCP_ε by x_ε^* .

CCP_ε is equivalent to RCP for $\varepsilon = 0$, but is less conservative than RCP even for small $\varepsilon > 0$. For general constraint functions f and distributions on δ , CCP_ε is nonconvex. A goal of these notes is to derive convex, finite-dimensional inner approximations of CCP_ε .

Example (affine constraint, Gaussian parameter). In the special case of affine f , Gaussian δ , and $\varepsilon \leq 1/2$, CCP_ε is convex. To see this, let $f(x, \delta) = \delta^T x + b$ and $\delta \sim \mathcal{N}(\mu, \Sigma)$. Then $\delta^T x + b \sim \mathcal{N}(\mu^T x + b, x^T \Sigma x)$, and

$$\mathbf{P}\{f(x, \delta) \leq 0\} \geq 1 - \varepsilon \iff 0 \geq \mu^T x + b + \Phi^{-1}(1 - \varepsilon) \|Cx\|_2,$$

where $C^T C = \Sigma$ and Φ is the cumulative distribution function of a standard normal random variable. The constant $\Phi^{-1}(1 - \varepsilon)$ is nonnegative for $\varepsilon \leq 1/2$, so this is a (convex) second-order cone constraint. This example demonstrates that chance constraints can promote a deterministic problem to a more general – and likely harder to solve – problem class. For instance, if X is a polyhedron, then the chance constraint promotes a linear program to a second-order cone program.

1.3 Sampled convex programming

A *sampled* (or *scenario* or *random*) *convex program* requires that the constraints be met for each of the independent, identically distributed samples $\delta_1, \dots, \delta_N$ from the distribution of δ :

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && f(x, \delta_i) \leq 0, \quad i = 1, \dots, N \\ & && x \in X. \end{aligned} \tag{SCP}_N$$

When one exists, we denote a solution of SCP_N by x_N^* .

SCP_N is a convex program of the same problem class as the original version. Unlike CCP_ε , SCP_N requires no assumptions on the distribution of δ , only the ability to independently sample from it. SCP_N is also less conservative than RCP, and can accommodate unbounded uncertainty sets. A downside of SCP_N is that in general, its solutions are feasible for some $\delta \in \Delta$, but not all. Only as $N \rightarrow \infty$ are the SCP_N solutions feasible for RCP.

Chapter 2

Sample bounds for sampled convex programming

How many samples must we take to guarantee that the SCP_N solutions are feasible for CCP_ε with high probability? More precisely, for what value of N can we certify that

$$\mathbf{P} \{ \mathbf{P} \{ f(x_N^*, \delta) \leq 0 \} \geq 1 - \varepsilon \} \geq 1 - \beta$$

for $\varepsilon, \beta \in [0, 1]$? This chapter provides two lower bounds on the number of samples. The first bound is general but loose. The second bound is tight but requires some assumptions on the underlying UCP.

It is worth noting that the results in this chapter only concern feasibility. The optimality gap between the SCP_N and CCP_ε solutions is an active area of current research.

2.1 A loose, general bound

We begin by defining the *violation probability* of $x \in X$ by

$$V(x) = \mathbf{P} \{ f(x, \delta) > 0 \}.$$

A point x is feasible for CCP_ε if $V(x) < \varepsilon$, so we seek a lower bound on N that guarantees $\mathbf{P} \{ V(x_N^*) > \varepsilon \} < \beta$.

One bound follows immediately from Markov's inequality,

$$\mathbf{P} \{ V(x_N^*) > \varepsilon \} \leq \frac{1}{\varepsilon} \mathbf{E} V(x_N^*).$$

In 2005, Califiori and Campi [2] showed that

$$\mathbf{E} V(x_N^*) \leq \frac{n}{N+1}. \tag{2.1}$$

Plugging this into the Markov bound gives our main result:

$$\boxed{N \geq \frac{n}{\varepsilon\beta} - 1 \quad \Longrightarrow \quad \mathbf{P}\{V(x_N^*) > \varepsilon\} \leq \beta.} \quad (2.2)$$

We now show that inequality (2.1) holds. This requires a result from convex analysis involving the problem

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && x \in X_i, \quad i = 1, \dots, m, \end{aligned} \quad (2.3)$$

where $x \in \mathbf{R}^n$ and the X_i are closed and convex. Let x^* denote the solution to problem (2.3), and x_k^* denote the solution to the same problem with the k^{th} constraint deleted. We call the constraint $x \in X_k$ a *support constraint* if $c^T x_k^* < c^T x^*$. The result, which we state without proof, is that problem (2.3) has at most n support constraints.

The expectation in (2.1) is taken over the samples $\delta^N = \{\delta_1, \dots, \delta_N\}$. By definition of V , we have

$$\begin{aligned} \mathbf{E}_{\delta^N} V(x_N^*) &= \mathbf{E}_{\delta^N} \mathbf{P}_{\delta|\delta^N} \{f(x_N^*, \delta) > 0 \mid \delta^N\} \\ &= \mathbf{E}_{\delta^N} \mathbf{E}_{\delta|\delta^N} [I_{++}(f(x_N^*, \delta)) \mid \delta^N], \end{aligned}$$

where I_{++} is the indicator function of \mathbf{R}_{++} ,

$$I_{++}(z) = \begin{cases} 0 & z \notin \mathbf{R}_{++} \\ 1 & z \in \mathbf{R}_{++}. \end{cases}$$

Letting $\delta^{-k} = \{\delta_1, \dots, \delta_{k-1}, \delta_{k+1}, \dots, \delta_N\}$, denoting by x_N^k the solution of the SCP_N with the constraint $f(x, \delta_k) \leq 0$ deleted, and applying iterated expectation, we have

$$\begin{aligned} \mathbf{E}_{\delta^N} V(x_N^*) &= \mathbf{E}_{\delta^{-k}} \mathbf{E}_{\delta_k|\delta^{-k}} [I_{++}(f(x_N^*, \delta)) \mid \delta^{-k}] \\ &= \mathbf{E}_{\delta^{-k}} \mathbf{E}_{\delta_k|\delta^{-k}} [v_k \mid \delta^{-k}] \\ &= \mathbf{E}_{\delta^{N+1}} v_k, \end{aligned}$$

where $\delta^{N+1} = \delta^N \cup \delta_{N+1}$, and $v_k = I_{++}(f(x_N^*, \delta))$ indicates whether the k^{th} constraint is a support constraint.

From our support constraint result,

$$\sum_{k=1}^{N+1} v_k \leq n$$

with probability one. Taking the sample mean of both sides of the equation $\mathbf{E}_{\delta^N} V(x_N^*) = \mathbf{E}_{\delta^{N+1}} v_k$, we have

$$\frac{1}{N+1} \sum_{k=1}^{N+1} \mathbf{E}_{\delta^N} V(x_N^*) = \frac{1}{N+1} \sum_{k=1}^{N+1} \mathbf{E}_{\delta^{N+1}} v_k$$

Table 2.1: samples required for $1 - \varepsilon$ violation with $n = 101$, $\beta = 10^{-4}$.

ε	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
§2.1 bound	10^7	2×10^7	4×10^7	10^8	2×10^8	4×10^8	10^9
§2.2 bound	2204	4408	8817	22042	44084	88168	220421

But $\mathbf{E}_{\delta^N} V(x_N^*)$ is a constant and expectation is a linear operator, so we have

$$\begin{aligned} \mathbf{E}_{\delta^N} V(x_N^*) &= \frac{1}{N+1} \mathbf{E}_{\delta^{N+1}} \sum_{k=1}^{N+1} v_k \\ &\leq \frac{n}{N+1}. \end{aligned}$$

2.2 A tight bound for structured problems

In this section, we present a tight bound on the required number of samples for a special class of sampled convex programs. This bound is highly desirable, since the required number of samples $n/\varepsilon\beta - 1$ from §2.1 is huge even for small problems and moderate risk parameters, as the following example shows.

Example (optimal control). We consider an instance of the optimal control problem (1.2) with a single-input, single-output system ($n_u = n_z = 1$) and time horizon of $T = 50$ time steps. As discussed in §1, this problem can be written as an uncertain convex program with $n = T(n_u + n_z) + 1 = 101$ optimization variables. If the corresponding CCP_ε instance is intractable, we can generate an approximate solution x_N^* by solving the corresponding SCP_N instance. If we require that x_N^* be feasible for CCP_ε with 99.99% probability ($\beta = 10^{-4}$), the bound from §2.1 suggests taking $n/\varepsilon \times 10^4 - 1$ samples. The first row of Table 2.1 shows this number for several values of ε . The second row shows the number of required samples with the tightened bound developed in this section.

In 2008, Campi and Garatti [3] showed that if for each N , the SCP_N derived from UCP has

- (i) a unique solution, and
- (ii) a feasible set with nonempty interior,

then

$$\mathbf{P} \{V(x_N^*) > \varepsilon\} \leq \sum_{i=1}^{n-1} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i}. \quad (2.4)$$

Moreover, if UCP is fully supported – meaning that for all $N \geq n$, SCP_N has exactly n support constraints – then the bound holds with equality. A proof of this result can be found in [3].

The right-hand side of inequality (2.4) is equal to $\mathbf{P}\{X \leq n - 1\}$, where $X \sim \text{Bin}(N, \varepsilon)$. The smallest N for which the right-hand side is no larger than β is upper bounded by

$$\inf_{N \in \mathbf{N}} \left\{ \sum_{i=1}^{n-1} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i} \leq \beta \right\} \leq \frac{2}{\varepsilon} \left(\ln \left(\frac{1}{\beta} \right) + n \right).$$

This gives our second result: if for all N , the SCP_N derived from UCP has a unique solution and a feasible set with nonempty interior, then

$$\boxed{N \geq \frac{2}{\varepsilon} \left(\ln \left(\frac{1}{\beta} \right) + n \right) \implies \mathbf{P}\{V(x_N^*) > \varepsilon\} \leq \beta.} \quad (2.5)$$

2.3 Comparison

The bounds (2.2) and (2.5) both grow linearly with n and $1/\varepsilon$, but depend on β as $\mathcal{O}(1/\beta)$ and $\mathcal{O}(\ln(1/\beta))$, respectively. This difference is substantial: for $\beta = 10^{-4}$, for example, $1/\beta = 10^4$, but $\ln(1/\beta) = 9.2$. The logarithmic dependence of (2.5) on $1/\beta$ means that under mild assumptions and for reasonably small N , we can guarantee that the SCP_N solution is feasible for CCP_ε with high probability. By imposing further structure on f and the distribution of δ , the ε -dependence can also be tightened to $\mathcal{O}(\ln(1/\varepsilon))$.

Chapter 3

Convex, conservative approximations to chance-constrained programs

Recall that the general form of the chance-constrained program is

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \mathbf{P} \{f(x, \delta) \leq 0\} \geq 1 - \varepsilon \\ & && x \in X. \end{aligned} \tag{CCP}_\varepsilon$$

The chance constraint makes CCP_ε nonconvex in general. In this chapter, following §2 of Nemirovski and Shapiro's 2006 paper [4], we develop methods to construct convex subsets of the CCP_ε feasible region that lead to tractable optimization problems. This technique is *conservative*: solutions to the approximate problems are feasible but generally suboptimal for CCP_ε .

3.1 Generating functions

We begin by considering the set of variables that are feasible for the chance constraint,

$$X_\varepsilon = \{x \mid \mathbf{P} \{f(x, \delta) \leq 0\} \geq 1 - \varepsilon\}.$$

The set X_ε is convex only in special cases, *e.g.*, when $f(x, \delta) = a^T x - \delta$ and the distribution of δ is logconcave, or when $f(x, \delta) = \delta^T x - b$ and δ is normal.

In the general, nonconvex case, we seek a function $g : \mathbf{R}^n \rightarrow \mathbf{R}$ such that $\{x \mid g(x) \leq \varepsilon\}$ is a convex subset of X_ε . This holds if

1. g is convex, and
2. for all x , $\mathbf{P} \{f(x, \delta) \geq 0\} \leq g(x)$.

An equivalent expression of condition 2 is

$$\mathbf{E} I_+(f(x, \delta)) \leq g(x),$$

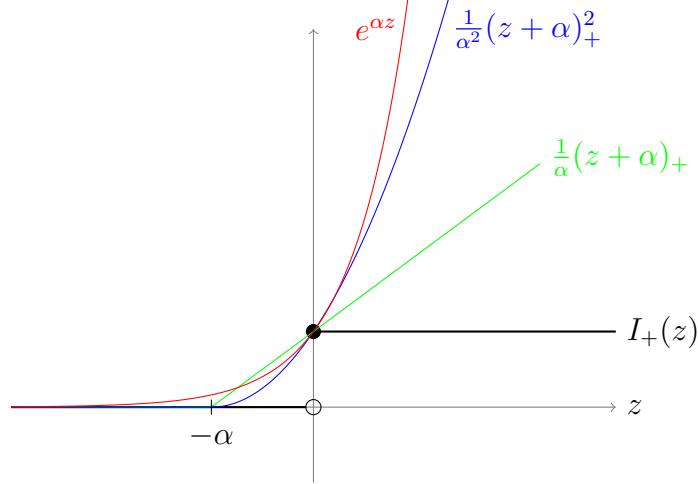


Figure 3.1: the three generating functions $e^{\alpha z}$, $(1/\alpha^2)(z + \alpha)_+^2$, and $(1/\alpha)(z + \alpha)_+$ are all convex, nondecreasing, and upper bound the indicator function $I_+(z)$ on \mathbf{R} .

where I_+ is the indicator function of \mathbf{R}_+ ,

$$I_+(z) = \begin{cases} 0 & z \notin \mathbf{R}_+ \\ 1 & z \in \mathbf{R}_+. \end{cases}$$

We can guarantee that $\mathbf{E} I_+(f(x, \delta)) \leq g(x)$ by ensuring that for all x and for all $\delta \in \Delta$,

$$I_+(f(x, \delta)) \leq g(x). \quad (3.1)$$

To see this, consider the case where δ is a discrete random variable that takes value $\delta_i \in \mathbf{R}^d$ with probability p_i for $i = 1, \dots, r$. In this case, if inequality (3.1) holds, then

$$\mathbf{E} I_+(f(x, \delta)) = \sum_{i=1}^r p_i I_+(f(x, \delta_i)) \leq \sum_{i=1}^r p_i g(x) = g(x),$$

since $p \succeq 0$ and $\mathbf{1}^T p = 1$. A similar argument can be made when δ is a continuous or mixed random variable.

To summarize, we seek a convex function $g(x)$ that upper bounds $I_+(f(x, \delta))$ for all x and $\delta \in \Delta$. This motivates defining a family of *generating functions* $\psi : \mathbf{R} \rightarrow \mathbf{R}$, which

- (i) are convex, nonnegative, and nondecreasing, and
- (ii) satisfy $\psi(z) > \psi(0) = 1$ for all $z > 0$.

Because a convex nondecreasing function of a convex function is convex, the function $x \mapsto \psi(f(x, \delta))$ is convex in x for any generating function ψ . Because ψ is nonnegative and

satisfies condition (ii), we also have $\psi(z) \geq I_+(z)$ for all $z \in \mathbf{R}$, with equality at $z = 0$ (see figure 3.1). For any generating function ψ , therefore, $\{x \mid \mathbf{E} \psi(f(x, \delta)) \leq \varepsilon\}$ is a convex subset of X_ε , and $X \cap \{x \mid \mathbf{E} \psi(f(x, \delta)) \leq \varepsilon\}$ is a convex subset of the CCP_ε feasible region. We can generate an approximate solution to CCP_ε by solving the problem

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \mathbf{E} \psi(f(x, \delta)) \leq \varepsilon \\ & && x \in X. \end{aligned}$$

3.2 Enlarging the inner approximation

In §3.1 we developed a conservative approximation of CCP_ε . So far, we have said nothing about how good or bad this approximation may be. In this section, we improve the approximation by (loosely speaking) expanding our convex subset of the CCP_ε feasible region.

To do this, we introduce a parameter $t > 0$ and replace $f(x, \delta)$ by $f(x, \delta)/t$. The argument in §3.1 carries through exactly under this replacement because an equivalent description of the set of variables that are feasible for the chance constraint is

$$X_\varepsilon = \left\{ x \mid \mathbf{P} \left\{ \frac{1}{t} f(x, \delta) \leq 0 \right\} \geq 1 - \varepsilon \right\}.$$

For any $t > 0$, therefore, $\{x \mid \mathbf{E} \psi(f(x, \delta)/t) \leq \varepsilon\}$ is a convex subset of X_ε .

In the discussion following result (2.4) of [4], Nemirovski and Shapiro strengthen this statement to

$$\left\{ x \mid \inf_{t>0} \left\{ -t\varepsilon + t \mathbf{E} \psi \left(\frac{1}{t} f(x, \delta) \right) \right\} \leq 0 \right\} \subseteq X_\varepsilon.$$

Using this fact, a conservative approximation to CCP_ε is

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \inf_{t>0} \left\{ -t\varepsilon + t \mathbf{E} \psi \left(f(x, \delta)/t \right) \right\} \leq 0 \\ & && x \in X. \end{aligned} \tag{3.2}$$

For problem (3.2) to be convex, the function

$$x \mapsto \inf_{t>0} \left\{ -t\varepsilon + t \mathbf{E} \psi \left(\frac{1}{t} f(x, \delta) \right) \right\}$$

must be convex. Because partial minimization preserves convexity and the sum of a linear and a convex function is convex, this is true as long as $(x, t) \mapsto t \mathbf{E} \psi(f(x, \delta)/t)$ is a convex function. But $(f(x, \delta), t) \mapsto t \mathbf{E} \psi(f(x, \delta)/t)$ is the perspective of the convex function $f(x, \delta) \mapsto \mathbf{E} \psi(f(x, \delta))$, and the perspective of a convex function is convex (see §3.2.6 of [1]). Therefore, problem (3.2) is convex.

While convex, problem (3.2) may still be intractable: the expectation integral may be costly or impossible to compute for some distributions on δ , and the infimum may have

a closed-form solution only for certain choices of ψ , as demonstrated in the following two examples.

Example (square-positive generating function). For the generating function $\psi(z) = (1/\alpha^2)(z + \alpha)_+^2$ (the square of the positive part), we can derive a sufficient condition for the constraint

$$\inf_{t>0} \left\{ -t\varepsilon + t \mathbf{E} \psi \left(\frac{1}{t} f(x, \delta) \right) \right\} \leq 0 \quad (3.3)$$

in problem (3.2) to hold. We have

$$\begin{aligned} t \mathbf{E} \psi \left(\frac{1}{t} f(x, \delta) \right) &= \frac{t}{\alpha^2} \mathbf{E} \left[\left(\frac{1}{t} f(x, \delta) + \alpha \right)_+^2 \right] \\ &= \frac{1}{\alpha^2 t} \mathbf{E} \left[(f(x, \delta) + \alpha t)_+^2 \right] \\ &\leq \frac{1}{\alpha^2 t} \mathbf{E} \left[(f(x, \delta) + \alpha t)^2 \right], \end{aligned}$$

where the second line follows from the fact that $t > 0$, and the third from the fact that $z_+^2 \leq z^2$ for all z . Expanding the quadratic term and simplifying, constraint (3.3) holds whenever

$$\inf_{t>0} \left\{ (1 - \varepsilon)t + \frac{1}{\alpha^2 t} \mathbf{E} [f(x, \delta)^2] + \frac{2}{\alpha} \mathbf{E} f(x, \delta) \right\} \leq 0. \quad (3.4)$$

Applying the first-order optimality conditions to the term inside the infimum (a convex function of t), we have

$$t^* = \sqrt{\frac{\mathbf{E} [f(x, \delta)^2]}{\alpha^2(1 - \varepsilon)}}.$$

Substituting t^* and simplifying, constraint (3.4) is equivalent to

$$\left(1 + \alpha^{3/2}\right) \sqrt{(1 - \varepsilon) \mathbf{E} [f(x, \delta)^2]} + 2 \mathbf{E} f(x, \delta) \leq 0. \quad (3.5)$$

Importantly, this result depends only on the first two moments of $f(x, \delta)$. For example, suppose $f(x, \delta) = \delta^T x$, $\mathbf{E} \delta = \mu$, and $\mathbf{E} \delta \delta^T = \Sigma \in \mathbf{S}_{++}^n$. Then $\mathbf{E} f(x, \delta) = \mu^T x$ and $\mathbf{E} [f(x, \delta)^2] = x^T \Sigma x$, so inequality (3.5) reduces to the second-order cone constraint

$$\left(1 + \alpha^{3/2}\right) \sqrt{(1 - \varepsilon)} \left\| \Sigma^{1/2} x \right\|_2 + 2\mu^T x \leq 0.$$

This gives the following conservative approximation of CCP_ε :

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \left(1 + \alpha^{3/2}\right) \sqrt{(1 - \varepsilon)} \left\| \Sigma^{1/2} x \right\|_2 + 2\mu^T x \leq 0 \\ &&& x \in X. \end{aligned}$$

Assuming X is described by a system of linear equalities and linear, convex quadratic, or second-order cone inequalities, this is a (tractable) second-order cone program. The only required distributional information is the first two moments of δ .

Example (hinge generating function). With the generating function $\psi(z) = (1/\alpha)(z + \alpha)_+$, it is less clear how to ensure satisfaction of the constraint (3.3) in problem (3.2). We have

$$\begin{aligned} t \mathbf{E} \psi \left(\frac{1}{t} f(x, \delta) \right) &= \frac{t}{\alpha} \mathbf{E} \left[\left(\frac{1}{t} f(x, \delta) + \alpha \right)_+ \right] \\ &= \frac{1}{\alpha} \mathbf{E} [(f(x, \delta) + \alpha t)_+] \\ &\leq \frac{1}{\alpha} \mathbf{E} |f(x, \delta) + \alpha t|. \end{aligned}$$

Thus, constraint (3.3) holds whenever

$$\inf_{t > 0} \left\{ -t\varepsilon + \frac{1}{\alpha} \mathbf{E} |f(x, \delta) + \alpha t| \right\} \leq 0.$$

While convex, the term inside the infimum is not differentiable, so we cannot apply the first-order optimality condition.

3.3 Joint chance constraints

We now extend the results in §3.2 to the case of multiple chance constraints. While this case can technically be reduced to the case of a single chance constraint by taking the pointwise maximum of the constraint functions, this may destroy useful structure.

We consider the problem

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \mathbf{P} \{f(x, \delta) \leq 0\} \geq 1 - \varepsilon \\ &&& x \in X, \end{aligned}$$

where $f : \mathbf{R}^n \times \mathbf{R}^d \rightarrow \mathbf{R}^m$ is convex. Writing $f(x, \delta) = (f_1(x, \delta), \dots, f_m(x, \delta))$, a point x is feasible for the chance constraint if $\mathbf{P} \cap_{i=1}^m \{f_i(x, \delta) \leq 0\} \geq 1 - \varepsilon$ or, equivalently, if $\mathbf{P} \cup_{i=1}^m \{f_i(x, \delta) > 0\} \leq \varepsilon$. But

$$\mathbf{P} \bigcup_{i=1}^m \{f_i(x, \delta) > 0\} \leq \sum_{i=1}^m \mathbf{P} \{f_i(x, \delta) > 0\},$$

so we can construct a conservative approximation of the chance constraint by bounding the right-hand probability.

One way to accomplish this is to require $\mathbf{P} \{f_i(x, \delta) > 0\} \leq \varepsilon_i$ for $i = 1, \dots, m$, for some $\varepsilon_1, \dots, \varepsilon_m \geq 0$ satisfying $\sum_{i=1}^m \varepsilon_i \leq \varepsilon$. This gives the following conservative approximation of our jointly chance-constrained problem:

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \mathbf{P} \{f_i(x, \delta) \leq 0\} \geq 1 - \varepsilon_i, \quad i = 1, \dots, m \\ &&& x \in X. \end{aligned}$$

Applying the results in §3.1-3.2 to each chance constraint gives the following (again conservative) approximation:

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \inf_{t>0} \{-t\varepsilon_i + t \mathbf{E} \psi_i(f_i(x, \delta)/t)\} \leq 0, \quad i = 1, \dots, m \\ & && x \in X. \end{aligned}$$

Remark. Specifying the generating functions ψ_1, \dots, ψ_m and bounds $\varepsilon_1, \dots, \varepsilon_m$ are important modeling steps that can strongly influence the tractability and quality of this approximation scheme.

3.4 Bernstein approximations for affinely perturbed constraints

We now consider a special case of CCP_ε with $f(x, \delta)$ affine in δ :

$$f(x, \delta) = f_0(x) + \sum_{i=1}^d \delta_i f_i(x).$$

Here the known, deterministic functions $f_0, \dots, f_d : \mathbf{R}^n \rightarrow \mathbf{R}$ are convex on X . For $i = 1, \dots, d$, we assume

- A1. the component δ_i of δ is independent of δ_j for all $j \neq i$,
- A2. the support Δ_i of δ_i is a compact set,
- A3. the moment generating function $M_i(t) = \mathbf{E} e^{t\delta_i}$ of δ_i is finite-valued for all $t \in \mathbf{R}$, and
- A4. if Δ_i contains a negative number, then f_i is affine on X .

Under these assumptions, the problem

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \inf_{t>0} \left\{ -t \ln \varepsilon + f_0(x) + \sum_{i=1}^d t \ln M_i(f_i(x)/t) \right\} \leq 0 \\ & && x \in X \end{aligned}$$

is a convex, conservative approximation to CCP_ε . One way to show this is to set $\psi(z) = e^z$ and apply the results in §3.2. It can also be shown more directly as follows. For any $t > 0$, the chance constraint $\mathbf{P}\{f(x, \delta) \leq 0\} \geq 1 - \varepsilon$ can be written as

$$\mathbf{P} \left\{ \frac{1}{t} \left(f_0(x) + \sum_{i=1}^d \delta_i f_i(x) \right) > 0 \right\} \leq \varepsilon$$

or, equivalently, as

$$\mathbf{P} \left\{ \exp \left(\frac{1}{t} \left(f_0(x) + \sum_{i=1}^d \delta_i f_i(x) \right) \right) > 1 \right\} \leq \varepsilon.$$

By Markov's inequality,

$$\begin{aligned} \mathbf{P} \left\{ \exp \left(\frac{1}{t} \left(f_0(x) + \sum_{i=1}^d \delta_i f_i(x) \right) \right) > 1 \right\} &\leq \mathbf{E} \exp \left(\frac{1}{t} \left(f_0(x) + \sum_{i=1}^d \delta_i f_i(x) \right) \right) \\ &= e^{f_0(x)/t} \mathbf{E} \prod_{i=1}^d e^{\delta_i f_i(x)/t} \\ &= e^{f_0(x)/t} \prod_{i=1}^d M_i(\delta_i f_i(x)/t), \end{aligned}$$

where the last line follows from the mutual independence of $\delta_1, \dots, \delta_d$ and the definition of M_i . Therefore, the chance constraint holds if there exists a $t > 0$ such that

$$e^{f_0(x)/t} \prod_{i=1}^d M_i(\delta_i f_i(x)/t) \leq \varepsilon,$$

or, equivalently,

$$f_0(x) + t \sum_{i=1}^d \ln M_i \left(\frac{1}{t} \delta_i f_i(x) \right) \leq t \ln \varepsilon.$$

By a similar argument to the one mentioned in §3.2, we can sharpen this result to say that the chance constraint holds if

$$\inf_{t>0} \left\{ f_0(x) + t \sum_{i=1}^d \ln M_i \left(\frac{1}{t} \delta_i f_i(x) \right) - t \ln \varepsilon \right\} \leq 0.$$

The convexity of this constraint can be established using an argument similar to that in §3.2. It involves the perspective function and the convexity and monotonicity of logarithmic moment generating functions.

Summarizing, a convex, conservative approximation to CCP_ε with $f(x, \delta)$ affine in δ is

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \inf_{t>0} \left\{ f_0(x) + t \sum_{i=1}^d \ln M_i(\delta_i f_i(x)/t) - t \ln \varepsilon \right\} \leq 0 \\ &&& x \in X. \end{aligned} \tag{3.6}$$

The results in this section extend naturally to joint chance constraints.

Chapter 4

Tractable robust convex programs

In this chapter, we consider the following special case of the general robust convex program:

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && Ax \preceq_K b \text{ for all } (A, b) \in \Delta. \end{aligned} \tag{4.1}$$

Here the generalized inequality $Ax \preceq_K b$ means that $b - Ax \in K$, where $K \subset \mathbf{R}^m$ is a proper cone (*i.e.*, K is a closed, convex cone with nonempty interior that contains no line.) See §2.4 of [1] for more on cones and generalized inequalities.

When the proper cone K is specified, problem (4.1) can be reduced to the canonical RCP form of §1.1 by writing the generalized inequality as a system of scalar convex inequalities and constraining the pointwise maximum of their left-hand sides. In order to derive tractable robust convex programs, however, it is useful to retain the conic structure of problem (4.1).

In these notes, we consider only the case of $K = \mathbf{R}_+^m$, so that problem (4.1) is a robust linear program. We enumerate the uncertainty sets Δ that lead to tractable problems. The approach generalizes, with some modifications, to robust second-order cone programs and robust semidefinite programs (see chapters 6 and 8 of [5]).

4.1 Robust linear programs

In this section, we consider the case of $K = \mathbf{R}_+^m$, which gives the robust linear program

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && \delta_i^T x \leq b_i \text{ for all } \delta_i \in \Delta_i, \quad i = 1, \dots, m, \end{aligned} \tag{RLP}$$

where $b_1, \dots, b_m \in \mathbf{R}$ are known, deterministic constants. (The results in this section can be extended to the case of uncertain b_i .) We are interested in forms of the uncertainty sets $\Delta_i \subseteq \mathbf{R}^n$ that generate tractable optimization problems. Loosely speaking, this requires that the maximization problem generated by each robust constraint,

$$\begin{aligned} & \text{maximize} && \delta_i^T x \\ & \text{subject to} && \delta_i \in \Delta_i, \end{aligned}$$

have a tractable dual. This holds in particular if each Δ_i is defined by a set of linear generalized inequalities:

$$\Delta_i = \{\delta_i \mid C_i \delta_i \preceq_{K_i} d_i\},$$

for some proper cone K_i and known, deterministic $C_i \in \mathbf{R}^{l_i \times n}$ and $d_i \in \mathbf{R}^{l_i}$.

4.1.1 Polyhedral uncertainty sets

A simple, solvable case of RLP involves polyhedral uncertainty sets:

$$\Delta_i = \{\delta_i \mid C_i \delta_i \preceq d_i\}.$$

(When applied to vectors, the symbol \preceq without subscript denotes generalized inequality with respect to the nonnegative orthant – in other words, componentwise inequality.) The corresponding robust constraint is

$$\sup \{\delta_i^T x \mid C_i \delta_i \preceq d_i\} \leq b_i.$$

An $x \in \mathbf{R}^n$ is feasible for this constraint if the optimal value of the following linear program in δ_i is no larger than b_i :

$$\begin{aligned} & \text{maximize} && x^T \delta_i \\ & \text{subject to} && C_i \delta_i \preceq d_i. \end{aligned}$$

If Δ_i is nonempty, then this problem is feasible, so strong duality implies that its optimal value equals the optimal value of the dual linear program

$$\begin{aligned} & \text{minimize} && d_i^T \lambda_i \\ & \text{subject to} && C_i^T \lambda_i = x \\ & && \lambda_i \succeq 0, \end{aligned}$$

with variable $\lambda_i \in \mathbf{R}^{l_i}$. Therefore, x is feasible for the i^{th} robust constraint if there exists a $\lambda_i \succeq 0$ satisfying $C_i^T \lambda_i = x$ and $d_i^T \lambda_i \leq b_i$.

We can write RLP with polyhedral uncertainty sets as

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && d_i^T \lambda_i \leq b_i, \quad i = 1, \dots, m \\ & && C_i^T \lambda_i = d_i, \quad i = 1, \dots, m \\ & && \lambda_i \succeq 0, \quad i = 1, \dots, m. \end{aligned} \tag{4.2}$$

This problem remains a linear program, but the variable $(x, \lambda_1, \dots, \lambda_m)$ has dimension $n + l$, where $l = \sum_{i=1}^m l_i$. The robust constraint also introduces m inequality constraints, l equality constraints, and l nonnegativity constraints.

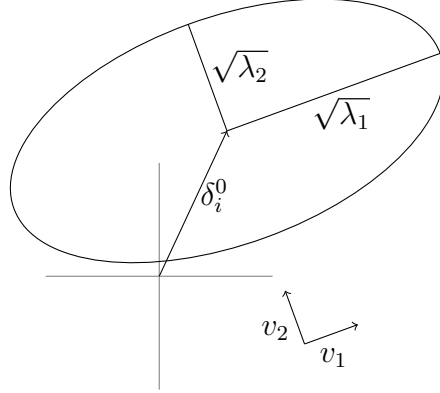


Figure 4.1: the ellipsoid $\{\delta_i^0 + D_i u \mid \|u\|_2 \leq \rho_i\}$. The matrix $(\rho_i D_i)^2$ has eigenvalues λ_1 and λ_2 and eigenvectors v_1 and v_2 .

4.1.2 Ellipsoidal uncertainty sets

Another solvable case of RLP arises when the uncertainty sets are ellipsoids:

$$\Delta_i = \{\delta_i^0 + D_i u \mid \|u\|_2 \leq \rho_i\}.$$

Here the ellipsoid center is δ_i^0 . Its size and shape are determined by the known, deterministic scalar $\rho_i > 0$ and matrix D_i . Without loss of generality, we can assume $D_i \in \mathbf{S}_+^n$. The ellipsoid Δ_i in \mathbf{R}^2 is illustrated in figure 4.1.

The corresponding robust constraint is $\sup\{\delta_i^T x \mid \delta_i \in \Delta_i\} \leq b_i$ or, equivalently,

$$x^T \delta_i^0 + \sup\{x^T D_i u \mid \|u\|_2 \leq \rho\} \leq b_i.$$

An $x \in \mathbf{R}^n$ is feasible for this constraint if the optimal value of the following second-order cone program in u is no larger than $b_i - x^T \delta_i^0$:

$$\begin{aligned} & \text{maximize} && x^T D_i u \\ & \text{subject to} && \|u\|_2 \leq \rho. \end{aligned}$$

This problem maximizes a linear function over the Euclidean ball of radius ρ . The solution is the longest feasible vector in the direction of $D_i^T x$, which is $u^* = (\rho / \|D_i^T x\|_2) D_i^T x$. The optimal value is $\rho \|D_i^T x\|_2$, so x is feasible for the i^{th} robust constraint if $\rho \|D_i^T x\|_2 \leq b_i - x^T \delta_i^0$.

We can therefore write RLP with ellipsoidal uncertainty sets as

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && (\delta_i^0)^T x + \rho \|D_i^T x\|_2 \leq b_i, \quad i = 1, \dots, m. \end{aligned} \tag{4.3}$$

Unlike the polyhedral case, this problem has the same dimension as the original RLP. However, the ellipsoidal uncertainty sets promote the robust linear program to a second-order cone program, which is generally more costly to solve.

4.1.3 Generalization

The argument in §4.1.1-4.1.2 can be generalized to accommodate uncertainty sets of the form

$$\Delta_i = \{\delta_i \mid C_i \delta_i \preceq_{K_i} d_i\},$$

where K_i is a proper cone and $C_i \in \mathbf{R}^{l_i \times n}$ and $d_i \in \mathbf{R}^{l_i}$ are known.

To see this, we note that x is feasible for the constraint

$$\delta_i^T x \leq b_i \text{ for all } \delta_i \in \Delta_i$$

if and only if the optimal value of the problem

$$\begin{aligned} & \text{minimize} && x^T \delta_i \\ & \text{subject to} && C_i \delta_i \preceq_{K_i} d_i \end{aligned}$$

is no larger than b_i . Assuming strong duality holds, an equivalent condition is that the optimal value of the dual

$$\begin{aligned} & \text{maximize} && -d_i^T \lambda_i \\ & \text{subject to} && C_i^T \lambda_i + x = 0 \\ & && \lambda_i \in K_i^* \end{aligned}$$

is no larger than b_i . (Here $K_i^* = \{y \mid x^T y \geq 0 \text{ for all } x \in K\}$ is the dual cone of K .) This holds if and only if there exists a $\lambda_i \in K_i^*$ satisfying $-d_i^T \lambda_i \leq b_i$ and $C_i^T \lambda_i + x = 0$.

Therefore, the robust LP with uncertainty sets defined by general conic inequalities is equivalent to

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && -d_i^T \lambda_i \leq b_i, \quad i = 1, \dots, m \\ & && C_i^T \lambda_i + x = 0, \quad i = 1, \dots, m \\ & && \lambda_i \in K_i^*, \quad i = 1, \dots, m. \end{aligned} \tag{4.4}$$

Here the optimization variable $(x, \lambda_1, \dots, \lambda_m)$ has dimension $n + l$, where $l = \sum_{i=1}^m l_i$. The m robust constraints have been reduced to m inequality constraints, l equality constraints, and l cone constraints.

It is easy to see that problem (4.4) reduces to problem (4.2) when $K_i = \mathbf{R}_+^{l_i}$, $i = 1, \dots, m$. Problem (4.4) also reduces to problem (4.3) when $K_i = \mathbf{R}^{l_i}$, $i = 1, \dots, m$. To see this, recall that

$$\|D_i^T x\|_2 \leq (1/\rho)(-(\delta_i^0)^T x + b_i) \iff \begin{bmatrix} D_i^T x \\ -(\delta_i^0)^T x + b_i \end{bmatrix} \in \mathbf{SO}^{n+1},$$

where \mathbf{SO}^{n+1} is the second-order cone in \mathbf{R}^{n+1} .

Chapter 5

Affine policies for stochastic linear programs

In this chapter, we consider the following situation. Given some matrices A , B , and C , we must decide a function x (also called a *policy* or *decision rule*). At some point in the future, an uncertain parameter δ will be revealed. At that time, we will incur a cost $\delta^T C^T x(\delta)$. Our goal is to minimize the expected value of this cost, subject to the constraint that our policy x must satisfy $Ax(\delta) \preceq B\delta$ almost surely.

Formally, we consider the single-stage stochastic linear program

$$\begin{array}{ll} \text{minimize} & \mathbf{E} \delta^T C^T x(\delta) \\ \text{subject to} & \left. \begin{array}{l} Ax(\delta) + s(\delta) = B\delta \\ s(\delta) \succeq 0 \end{array} \right\} \mathbf{P}\text{-a.s.} \end{array} \quad (\text{SP})$$

The optimization variables are $x \in \mathcal{L}_{k,n}^2$ and $s \in \mathcal{L}_{k,m}^2$, where $\mathcal{L}_{r,s}^2$ denotes the space of Borel-measurable, square-integrable functions from \mathbf{R}^r to \mathbf{R}^s . The notation “ \mathbf{P} -a.s.” means “almost surely under the probability measure \mathbf{P} .” The matrices $A \in \mathbf{R}^{m \times n}$, $B \in \mathbf{R}^{m \times k}$, and $C \in \mathbf{R}^{n \times k}$ are known exactly. The uncertain parameter $\delta \in \mathbf{R}^k$ has support

$$\Delta = \{\delta \mid W\delta \succeq h\},$$

where

$$W = \begin{bmatrix} e_1^T \\ -e_1^T \\ \hat{W} \end{bmatrix} \in \mathbf{R}^{l \times k}, \quad h = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \in \mathbf{R}^l,$$

for some $\hat{W} \in \mathbf{R}^{(l-2) \times k}$. In other words, any $\delta \in \Delta$ satisfies $\delta_1 = 1$ and $\hat{W}\delta \succeq 0$.

Problem SP is intractable in general. In fact, even if δ is uniformly distributed on the unit hypercube, problem SP is known to be $\#\text{P}$ -hard (meaning, loosely speaking, that it is at least as hard as any problem in NP.) Therefore, in §5.1, we will derive a tractable approximation of problem SP by restricting attention to policies that are affine in δ . In §5.2, we will employ similar methods to derive a bound on the suboptimality of our approximate solutions.

To facilitate these approximations, we assume

A1. Δ is nonempty and bounded, and

A2. Δ spans \mathbf{R}^k .

Assumption A1 will allow us to apply linear programming strong duality in §5.1. Assumption A2 guarantees that the moment matrix

$$M = \mathbf{E} \delta \delta^T$$

is positive definite (hence invertible). To see this, first note that M is positive semidefinite: For any v ,

$$v^T M v = \mathbf{E} v^T \delta \delta^T v = \mathbf{E} (v^T \delta)^2 \geq 0.$$

To see that M is also positive definite, suppose that $v \neq 0$. Then $v^T M v = 0$ only if $v^T \delta = 0$ for all $\delta \in \Delta$. But Δ spans \mathbf{R}^k by assumption A2, so this can hold only if $v = 0$, a contradiction. Thus, $v^T M v > 0$ for all $v \neq 0$.

5.1 Primal affine approximation

In this section, we develop a method for solving problem SP approximately. We restrict attention to *linear* policies $x(\delta) = X\delta$ and $s(\delta) = S\delta$, for some $X \in \mathbf{R}^{n \times k}$ and $S \in \mathbf{R}^{m \times k}$. Since $\delta_1 = 1$ with probability one for any $\delta \in \Delta$, a linear policy in δ is actually *affine* in the truly uncertain parameters $\delta_2, \dots, \delta_k$. Hence the title of this section.

With the restriction to linear policies, we can simplify the objective function of problem SP, and derive tractable reformulations of its constraints. The objective function becomes

$$\begin{aligned} \mathbf{E} \delta^T C^T x(\delta) &= \mathbf{E} \delta^T C^T X \delta \\ &= \mathbf{E} \operatorname{tr}(\delta^T C^T X \delta) \\ &= \mathbf{E} \operatorname{tr}(\delta \delta^T C^T X) \\ &= \operatorname{tr}(M C^T X). \end{aligned}$$

Since linear functions are continuous, the equality constraint

$$Ax(\delta) + s(\delta) = B\delta \quad \mathbf{P}\text{-a.s.} \tag{5.1}$$

is equivalent to

$$\begin{aligned} &AX\delta + S\delta = B\delta \text{ for all } \delta \in \Delta \\ \iff &(AX + S - B)\delta = 0 \text{ for all } \delta \in \Delta. \end{aligned}$$

But Δ spans \mathbf{R}^k by assumption A2, so this holds if and only if

$$AX + S = B. \tag{5.2}$$

This is a finite-dimensional reformulation of the original equality constraint (5.1), with $x(\cdot)$ and $s(\cdot)$ restricted to be linear functions.

Similarly, linearity (hence continuity) of s implies that the inequality constraint

$$s(\delta) \succeq 0 \quad \mathbf{P}\text{-a.s.} \quad (5.3)$$

is equivalent to

$$s_i^T \delta \geq 0 \text{ for all } \delta \in \Delta, \quad i = 1, \dots, m,$$

where $S = [s_1 \ \dots \ s_m]^T$. But by definition of Δ , the constraint $s_i^T \delta \geq 0$ for all $\delta \in \Delta$ holds if and only if the optimal value of the linear program

$$\begin{aligned} & \text{minimize} && s_i^T \delta \\ & \text{subject to} && W\delta \succeq h \end{aligned}$$

is nonnegative. The dual of this linear program is

$$\begin{aligned} & \text{maximize} && h^T \lambda_i \\ & \text{subject to} && W^T \lambda_i = s_i \\ & && \lambda_i \succeq 0. \end{aligned}$$

Since Δ is nonempty and bounded by assumption A1, strong duality implies that $s_i^T \delta \geq 0$ for all $\delta \in \Delta$ if and only if there exists a $\lambda_i \in \mathbf{R}_+^l$ such that $W^T \lambda_i = s_i$ and $h^T \lambda_i \geq 0$.

Applying this result for each $i = 1, \dots, m$, the constraint (5.3) holds if and only if there exist $\lambda_1, \dots, \lambda_m \in \mathbf{R}^l$ such that

$$\begin{aligned} \lambda_i^T W &= s_i^T, \quad i = 1, \dots, m \\ \lambda_i^T h &\geq 0, \quad i = 1, \dots, m \\ \lambda_i &\succeq 0, \quad i = 1, \dots, m. \end{aligned}$$

This system can be written more compactly by introducing a matrix $\Lambda = [\lambda_1 \ \dots \ \lambda_m]^T$. With this notation, the constraint (5.3) holds if and only if there exists a $\Lambda \in \mathbf{R}^{m \times l}$ such that

$$\begin{aligned} \Lambda W &= S \\ \Lambda h &\succeq 0 \\ \Lambda_{ij} &\geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, l. \end{aligned}$$

Collecting our objective function and constraint reformulations, we have the following primal linear approximation of problem SP:

$$\begin{aligned} & \text{minimize} && \text{tr}(MC^T X) \\ & \text{subject to} && AX + \Lambda W = B \\ & && \Lambda h \succeq 0 \\ & && \Lambda_{ij} \geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, l. \end{aligned} \quad (\text{SP}^u)$$

This is a finite-dimensional, deterministic linear program with optimization variables $X \in \mathbf{R}^{n \times k}$ and $\Lambda \in \mathbf{R}^{m \times l}$. Because we have reduced the search space from square-integrable functions to linear functions, the optimal value of problem SP^u is an upper bound on the optimal value of problem SP.

5.2 Dual affine approximation

In §5.1, we derived a primal approximation to the single-stage stochastic program SP. The derivation involved constructing an inner approximation of the SP feasible region by restricting the search space from square integrable functions to linear ones. The resulting approximation provides an upper bound on the optimal value of SP. Because the primal linear approximation is tractable, it also provides a means of generating solutions to SP that, while suboptimal, might be quite good.

In this section, we develop a means of quantifying how good the primal linear approximate solutions are. This involves constructing a lower bound on the SP optimal value by dualizing its equality constraints and imposing another linearity restriction. Once we have derived a lower bound, we can compare it to the upper bound from the primal linear approximation. The difference between the upper and lower bounds is itself an upper bound on the suboptimality of the primal linear approximate solutions.

We begin by dualizing the equality constraints in problem SP, giving the equivalent reformulation

$$\begin{aligned} & \text{minimize} && \mathbf{E} [\delta^T C^T x(\delta)] + \sup_{y \in \mathcal{L}_{k,m}^2} \{ \mathbf{E} [y(\delta)^T (Ax(\delta) + s(\delta) - B\delta)] \} \\ & \text{subject to} && s(\delta) \succeq 0 \quad \mathbf{P}\text{-a.s.} \end{aligned} \quad (5.4)$$

Here the equivalence follows from the fact that with probability one, the supremum over y yields zero whenever $Ax(\delta) + s(\delta) = B\delta$, and $+\infty$ otherwise.

We seek a tractable problem whose optimal value lower bounds the optimal value of problem (5.4). To that end, we note that

$$\sup_{y \in \mathcal{L}_{k,m}^2} \{ \mathbf{E} [y(\delta)^T (Ax(\delta) + s(\delta) - B\delta)] \} \geq \sup_{Y \in \mathbf{R}^{m \times k}} \{ \mathbf{E} [\delta^T Y^T (Ax(\delta) + s(\delta) - B\delta)] \}. \quad (5.5)$$

In other words, the optimal value of problem (5.4) cannot increase when the dual search space is restricted from square-integrable functions y to linear functions $y(\delta) = Y\delta$.

The right-hand side of inequality (5.5) can be simplified:

$$\begin{aligned} \mathbf{E} [\delta^T Y^T (Ax(\delta) + s(\delta) - B\delta)] &= \mathbf{E} [\text{tr} (\delta^T Y^T (Ax(\delta) + s(\delta) - B\delta))] \\ &= \text{tr} (Y^T \mathbf{E} [(Ax(\delta) + s(\delta) - B\delta) \delta^T]) \end{aligned}$$

Therefore, the optimal value of problem SP cannot be less than the optimal value of the problem

$$\begin{aligned} & \text{minimize} && \mathbf{E} [\delta^T C^T x(\delta)] + \sup_{Y \in \mathbf{R}^{m \times k}} \{ \text{tr} (Y^T \mathbf{E} [(Ax(\delta) + s(\delta) - B\delta) \delta^T]) \} \\ & \text{subject to} && s(\delta) \succeq 0 \quad \mathbf{P}\text{-a.s.} \end{aligned} \quad (5.6)$$

The supremum in problem (5.6) yields zero whenever $\mathbf{E} [(Ax(\delta) + s(\delta) - B\delta) \delta^T] = 0$ and yields $+\infty$ otherwise. Therefore, we can replace the supremum with the constraint

$$\begin{aligned} & \mathbf{E} [(Ax(\delta) + s(\delta) - B\delta) \delta^T] = 0 \\ \iff & A \mathbf{E} x(\delta) \delta^T + \mathbf{E} s(\delta) \delta^T = BM, \end{aligned}$$

where the equivalence follows from the linearity of expectation and the definition of the moment matrix M .

Recall that $x : \mathbf{R}^k \rightarrow \mathbf{R}^n$ and $s : \mathbf{R}^k \rightarrow \mathbf{R}^m$ are general square-integrable functions. We therefore have no clean expressions for the terms $\mathbf{E} x(\delta)\delta^T$ and $\mathbf{E} s(\delta)\delta^T$. However, let us change variables to matrices $X \in \mathbf{R}^{n \times k}$ and $S \in \mathbf{R}^{m \times k}$, defined such that

$$\begin{aligned} XM &= \mathbf{E} x(\delta)\delta^T \\ SM &= \mathbf{E} s(\delta)\delta^T. \end{aligned}$$

We have shown that M is positive definite (hence invertible), so the matrices X and S can be uniquely determined from the functions x and s .

With this change of variables, the constraint $A\mathbf{E} x(\delta)\delta^T + \mathbf{E} s(\delta)\delta^T = BM$ is equivalent to $(AX + S - B)M = 0$. Since M is invertible, this holds if and only if

$$AX + S = B.$$

Given $XM = \mathbf{E} x(\delta)\delta^T$, we also have that

$$\mathbf{E} \delta^T C^T x(\delta) = \text{tr}(C^T \mathbf{E} x(\delta)\delta^T) = \text{tr}(C^T XM).$$

Therefore, an equivalent reformulation of problem (5.6) is

$$\begin{aligned} &\text{minimize} && \text{tr}(MC^T X) \\ &\text{subject to} && AX + S = B \\ &&& \left. \begin{aligned} &\exists x \in \mathcal{L}_{k,n}^2 \text{ s.t. } XM = \mathbf{E} x(\delta)\delta^T \\ &\exists s \in \mathcal{L}_{k,m}^2 \text{ s.t. } SM = \mathbf{E} s(\delta)\delta^T \text{ and } s(\delta) \succeq 0 \end{aligned} \right\} \mathbf{P}\text{-a.s.} \end{aligned} \quad (5.7)$$

Problem (5.7) is still intractable due to the second and third constraints. However, the second constraint is redundant and can be deleted. To see this, note that given any $X \in \mathbf{R}^{n \times k}$, we can produce an $x \in \mathcal{L}_{k,n}^2$ with $XM = \mathbf{E} x(\delta)\delta^T$ by setting $x(\delta) = X\delta$. The third constraint can be replaced by

$$(W - he_1^T)MS^T \succeq 0.$$

A proof of this fact can be found in §2.4 of [6].

Thus, the optimal value of the problem

$$\begin{aligned} &\text{minimize} && \text{tr}(MC^T X) \\ &\text{subject to} && AX + S = B \\ &&& ((W - he_1^T)MS^T)_{ij} \geq 0, \quad i = 1, \dots, l, \quad j = 1, \dots, m. \end{aligned} \quad (5.8)$$

lower bounds the optimal value of problem SP. Like the primal linear approximation problem SP^u , the dual linear approximation problem SP^l is a finite-dimensional linear program with optimization variables $X \in \mathbf{R}^{n \times k}$ and $S \in \mathbf{R}^{m \times k}$.

The primal and dual linear approximations give the following method for solving problem SP approximately and bounding the suboptimality gap.

1. Compute a solution (X^u, S^u) to the primal linear approximation SP^u .
2. Compute a solution (X^l, S^l) to the dual linear approximation SP^l .
3. Observe that

$$\text{tr}(MC^T X^u) - \mathbf{E} \delta^T C^T x^*(\delta) \leq \text{tr}(MC^T (X^u - X^l)),$$

where $x^* \in \mathcal{L}_{k,n}^2$ is the optimal value of problem SP .

5.3 Example: a two-stage stochastic linear program

We consider the problem

$$\begin{aligned} & \text{minimize} && c^T x + \mathbf{E} \delta^T D^T y(\delta) \\ & \text{subject to} && \left. \begin{aligned} A_x x + A_y y(\delta) + s(\delta) &= B\delta \\ s(\delta) &\succeq 0 \end{aligned} \right\} \mathbf{P}\text{-a.s.} \end{aligned} \quad (5.9)$$

Here we must decide both the vector $x \in \mathbf{R}^{n_x}$ and the policy $y \in \mathcal{L}_{k,n_y}^2$. At some point in the future, we will realize δ and incur cost $c^T x + \delta^T D y(\delta)$. Our goal is to choose x and y to minimize this cost, subject to the constraint that $A_x x + A_y y(\delta) \preceq B\delta$ almost surely. The problem data are $A_x \in \mathbf{R}^{m \times n_x}$, $A_y \in \mathbf{R}^{m \times n_y}$, $B \in \mathbf{R}^{m \times k}$, $c \in \mathbf{R}^{n_x}$, and $D \in \mathbf{R}^{n_y \times k}$, as well as $\hat{W} \in \mathbf{R}^{(l-2) \times k}$, which defines the support Δ of δ , and the moment matrix $M \in \mathbf{S}_{++}^k$.

The primal affine approximation of this problem is

$$\begin{aligned} & \text{minimize} && c^T x + \text{tr}(MD^T Y) \\ & \text{subject to} && \begin{aligned} A_x x e_1^T + A_y Y + \Lambda W &= B \\ \Lambda h &\succeq 0 \\ \Lambda_{ij} &\geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, l, \end{aligned} \end{aligned} \quad (5.10)$$

with variables $x \in \mathbf{R}^{n_x}$, $Y \in \mathbf{R}^{n_y \times k}$, and $\Lambda \in \mathbf{R}^{m \times l}$. The dual affine approximation is

$$\begin{aligned} & \text{minimize} && c^T x + \text{tr}(MD^T Y) \\ & \text{subject to} && \begin{aligned} A_x x e_1^T + A_y Y + S &= B \\ ((W - h e_1^T) M S^T)_{ij} &\geq 0, \quad i = 1, \dots, l, \quad j = 1, \dots, m. \end{aligned} \end{aligned} \quad (5.11)$$

with variables $x \in \mathbf{R}^{n_x}$, $Y \in \mathbf{R}^{n_y \times k}$, and $S \in \mathbf{R}^{m \times k}$.

We will compare the affine policy to a simple approximation method called *sample average approximation*. In this method, we generate the independent samples $\delta_1, \dots, \delta_N$ from the distribution of δ , then solve the problem

$$\begin{aligned} & \text{minimize} && c^T x + (1/N) \sum_{i=1}^N \delta_i^T D^T y_i \\ & \text{subject to} && A_x x + A_y y_i \preceq B \delta_i, \quad i = 1, \dots, N. \end{aligned} \quad (5.12)$$

Here the variables are $x \in \mathbf{R}^{n_x}$ and $y_1, \dots, y_N \in \mathbf{R}^{n_y}$. With probability one, the true realization of δ will not equal any of the samples δ_i , so implementing the sample average

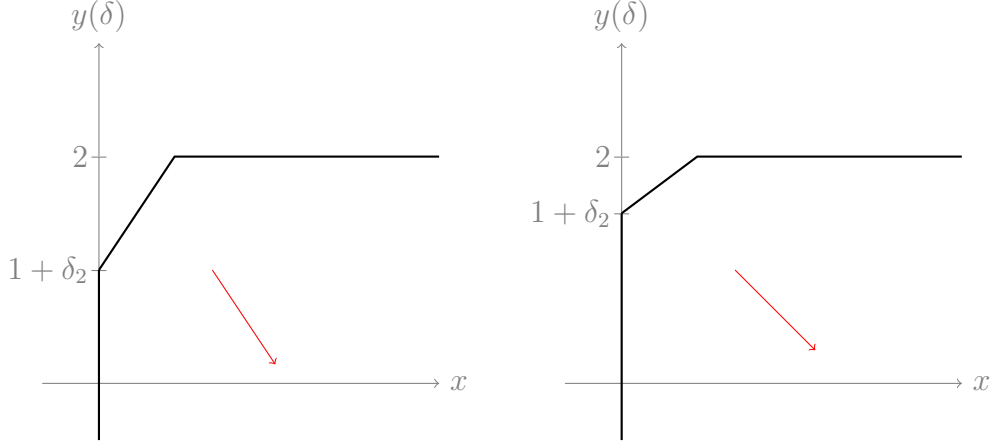


Figure 5.1: feasible region (black) and cost vector $(c, D\delta) = (1, -(1 + \delta_2))$ (red) for $\delta_2 = 0$ (left) and $\delta_2 = 1/2$ (right). Different values of δ_2 give different optimal vertices.

approximation policy requires interpolation. For large N , the optimal value of the sample average approximation approaches the optimal value of the true problem (5.9). The sample average approximation policy is much more costly to compute than the affine policies, however. Problem (5.12) has $n_x + Nn_y$ variables and mN constraints, compared to $n_x + kn_y + ml$ variables and $l + m + lm$ constraints in problem (5.10) and $n_x + kn_y + km$ variables and $lm + m$ constraints in problem (5.11).

5.3.1 Problem instance

We consider a simple problem instance with $k = 2$ and δ uniformly distributed on $\Delta = \{\delta \mid \delta_1 = 1, 0 \leq \delta_2 \leq 1\}$, so that $l = 4$ and

$$\hat{W} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}.$$

The dimensions are $n_x = n_y = 1$, with constraint matrices

$$A_x = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix}, \quad A_y = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 1 & 1 \end{bmatrix},$$

giving the feasible region

$$\{(x, y) \in \mathbf{R} \times \mathcal{L}_{2,1}^2 \mid x \geq 0, y(\delta) \leq \max\{2, \mathbf{1}^T \delta + 2x\} \mathbf{P}\text{-a.s.}\}.$$

We consider costs $c = 1$ and $D = -\mathbf{1}^T$. Figure 5.1 shows the problem geometry for two fixed values of the uncertain parameter δ_2 . Figure 5.2 shows the optimal values of the approximation problems. Figure 5.3 shows a histogram of the costs incurred by each policy, over 100,000 Monte Carlo simulations.

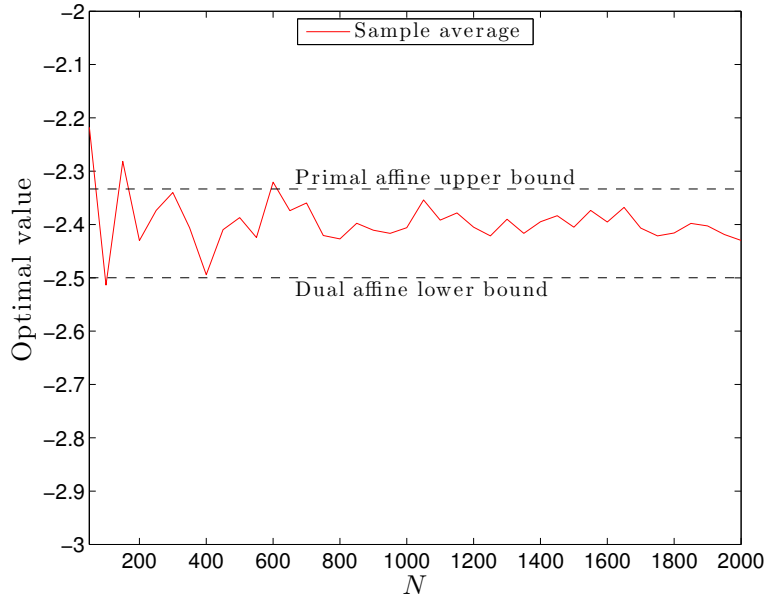


Figure 5.2: for large N , the optimal value of the sample average approximation problem (5.12) (red) approaches the optimal value of the true problem (5.9), which lies between the primal and dual affine approximations.

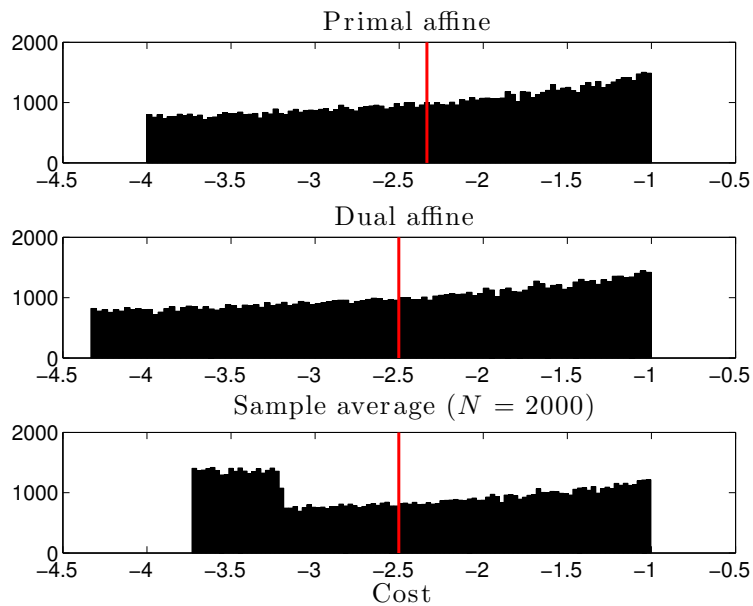


Figure 5.3: cost histogram of the three policies over 100,000 Monte Carlo runs. In the sample average policy, we compute $y(\delta)$ by linearly interpolating the solutions y_1, \dots, y_N corresponding to the samples $\delta_1, \dots, \delta_N$.

Bibliography

- [1] Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [2] Calafiore, Giuseppe, and Marco C. Campi. “Uncertain convex programs: randomized solutions and confidence levels.” *Mathematical Programming* **102.1** (2005): 25-46.
- [3] Campi, Marco C., and Simone Garatti. “The exact feasibility of randomized solutions of uncertain convex programs.” *SIAM Journal on Optimization* **19.3** (2008): 1211-1230.
- [4] Nemirovski, Arkadi, and Alexander Shapiro. “Convex approximations of chance constrained programs.” *SIAM Journal on Optimization* **17.4** (2006): 969-996.
- [5] Ben-Tal, Aharon, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.
- [6] Kuhn, Daniel, Wolfram Wiesemann, and Angelos Georghiou. “Primal and dual linear decision rules in stochastic and robust optimization.” *Mathematical Programming* **130.1** (2011): 177-209.