EE 227A: Convex Optimization and Applications **April 24, 2008**

Lecture 24: **Robust Optimization: Chance Constraints**

Lecturer: Laurent El Ghaoui

Reading assignment: Chapter 2 of the book on Robust Optimization[1]

24.1 Stochastic Optimization

Example 1: Stochastic LP

$$\begin{aligned} &\min c^T x \\ &st : A(w_i)x \leq b(w_i), \forall i \end{aligned} \tag{24.1}$$

where $A(w_i), b(w_i)$ are realizations of the random variables A, b.

Example 2: Stochastic portfolio optimization

$$\max_{u_1,u_2} E(y_1u_1 + y_2u_2 - \frac{1}{2}\sigma_1^2 u_1^2 - \frac{1}{2}\sigma_1^2 u_1^2 - c_1|u_1 - u_0| - c_2|u_2 - u_1|) \tag{24.2}$$

where y_1, y_2 are random variables. The problem is maximizing the expectation.

The stochastic recourse problem, however, becomes complicated. (Taking the expectation of the objective function in (24.11), $E(c_2|(\hat{u}_2 + U_2y_1) - u_1|)$ usually is not in a simple form.)

24.2 Chance Optimization

Consider the LP

$$\begin{aligned} &\min c^T x \\ &st : a_i^T x \leq b_i \end{aligned} \tag{24.3}$$

where $a_i \sim N(\hat{\sigma}_i, \Sigma_i)$. We can formulate a chance constrained LP

$$\begin{aligned} &\min c^T x \\ &st : \Pr\{a_i^T x \leq b_i\} \geq 1 - \epsilon \end{aligned} \tag{24.4}$$

If we take into account the correlation between a_i's, the problem is

$$\begin{aligned} &\min c^T x \\ &st : \Pr\{Ax \leq b\} \geq 1 - \epsilon \end{aligned} \tag{24.5}$$

[1] A. Ben Tal, L. El Ghaoui, A. Nemirovski, *Robust Optimization.*

where A is the random variable.

If c is also a random variable, the problem can be formulated as

$$\begin{aligned} &\min t \\ &st : \Pr\{t \geq c^T x\} \geq 1 - \epsilon, \Pr\{Ax \leq b\} \geq 1 - \epsilon \end{aligned} \tag{24.6}$$

24.3 Robust Optimization topics

In the following lectures, we will cover the following topics about Robust Optimization

- Form efficient convex approximation
- Chance constraints relaxation (including the case with partially known distribution)
- Evaluate quality of relaxation
- Recourse programming

24.4 Chance Constraints

Chance constraints are a probabilistic way of handling probabilistic uncertainty. We would like to convert chance constraints into robustness constraints, which are easier to deal with. Consider the LP:

$$\min\ c^T x \tag{24.7}$$
$$\text{s.t. } Ax \leq b \tag{24.8}$$

Our basic problem with chance constraints is:

$$\min\ c^T x \tag{24.9}$$
$$\text{s.t. } \operatorname{Prob}\{Ax \leq b\} \geq 1 - \epsilon \tag{24.10}$$

The technique we use to simplify the the chance constraints is to assume no correlation between rows of the A-matrix, which allows us to write the chance constraint as:

$$\operatorname{Prob}\{a_i^T x \leq b_i\} \geq 1 - \epsilon, \quad i = 1, 2, \ldots m \tag{24.11}$$

This is a less general model, but it is often much easier to solve. The question now becomes: how do we handle such scalar chance constraints?

We will consider a constraint of the form

$$\operatorname{Prob}\{a^T x \leq b\} \geq 1 - \epsilon,$$

where a, b are random, and say that x is ϵ-reliable if it satisfies the above constraint.

24.5 Example: Gaussian Case

In this particular case, this scalar constraint is easy to handle. Suppose $a \sim \mathcal{N}(\hat{a}, \Sigma)$, then for fixed x:

$$\xi = a^T x - b \sim \mathcal{N}(\hat{a}^T x - b, x^T \Sigma x) \tag{24.12}$$

Scaling with $\xi = \hat{\xi} + \sigma u$, where $\hat{\xi} = \hat{a}^T x - b$, $\sigma^2 = x^T \Sigma x$, and $u \sim \mathcal{N}(0,1)$, gives that:

$$\text{Prob}\{\xi \leq 0\} = \text{Prob}\left\{ u \leq -\frac{\hat{\xi}}{\sigma} \right\} = \int_{-\infty}^{-\frac{\hat{\xi}}{\sigma}} \frac{1}{2\pi} \exp(-u^2/2) du \geq 1 - \epsilon \tag{24.13}$$

Defining $\text{Erf}(\alpha) = \int_\alpha^\infty \frac{1}{2\pi} \exp(-u^2/2) du$, we can rewrite the inequality above as $\text{Erf}\left(-\frac{\hat{\xi}}{\sigma}\right) \leq \epsilon$. Note that $\text{Erf}(\cdot)$ is a decreasing function and that $\text{Erf}(0) = 1/2$. Further simplifying our expression, we get that:

$$\text{Erf}\left(-\frac{\hat{\xi}}{\sigma}\right) \leq \epsilon \quad \Leftrightarrow -\tfrac{\hat{\xi}}{\sigma} \geq \text{Erfinv}(\epsilon) =: \psi(\epsilon) \tag{24.14}$$

$$\Leftrightarrow \hat{\xi} + \psi(\epsilon)\sigma \leq 0 \tag{24.15}$$

$$\Leftrightarrow \hat{a}^T x + \psi(\epsilon)\sqrt{x^T \Sigma x} \leq b \tag{24.16}$$

Note that $\psi(\epsilon) > 0$ if $\epsilon < 1/2$, and $\psi(\epsilon)$ is like a coefficient that can be precomputed. Consequently, the inequality above is a SOCP constraint. Making ϵ smaller causes $\psi(\epsilon)$ to become larger, which makes the constraint harder to satisfy. Also, we can interpret $\psi(\epsilon)\sqrt{x^T \Sigma x}$ as a risk term and $\hat{a}^T x$ as an average term.

Besides this case, there are very few cases where we can do symbolic calculations. In practice, assuming Gaussian uncertainty is fine for most cases, even when the randomness is not Gaussian. Also, notice that the SOCP constraint derived above is the same as the one obtained using a robust optimization approach with an uncertainty model. Indeed, suppose that all we know is that:

$$a \in \mathcal{U} := \{\hat{a} + \Sigma^{1/2} u \mid \quad \|u\|_2 \leq \psi(\epsilon)\} \tag{24.17}$$

Thenn, we have that $a^T x \leq b, \quad \forall a \in \mathcal{U}$ if and only if $\hat{a}^T x + \psi(\epsilon)\sqrt{x^T \Sigma x}$. This uncertainty model seems very restrictive, but it can be interpreted as a probabilistic approach; it is fairly reasonable.

24.6 Case with Independent, Bounded Random Uncertainties

Suppose that we have a vector uncertainty $\xi \in \mathbf{R}^L$, which is a random variable, such that the coefficients ξ_l's are mutually independent, and $E(\xi_l) = 0$, and $|\xi_l| \leq 1, \forall l$. The last constraint is a deterministic bound, which tells us about the support of the random variable ξ. We consider the chance constraint

$$\mathrm{Prob}\{a(\xi)^T x \leq b(\xi)\} \geq 1 - \epsilon \tag{24.18}$$

where a, b depend on the random variable ξ in an affine fashion, precisely $a(\xi) = a_0 + \Sigma \xi_i a_i$ and $b(\xi) = b_0 + \Sigma \xi_i b_i$, where vector a_i and scalars b_i, $i = 0, \ldots, L$, are given.

In the case that $\xi_i = 1$ with probability $1/2$ and $\xi_i = -1$ with probability $1/2$, computing the probability in (24.29) for fixed x can be shown to be NP-hard, and so it is quite reasonable to use an approximation, precisely, a sufficient condition for the chance constraint to hold.

Our chance constraint can be rewritten as

$$\mathrm{Prob}\{\xi^T z > b_0 - a_o^T x\} \leq \epsilon. \tag{24.19}$$

We will show later the following lemma. Let $z \in \mathbf{R}^L$ be a deterministic vector, and $\xi_1, \ldots, \xi_L$ be independent random variables with zero mean taking values in $[-1, 1]$. Then

$$\mathrm{Prob}\{\xi^T z > \Omega \|z\|_2\} \leq \exp(-\Omega^2/2).$$

Applying the lemma with $z_i = a_i^T x - b_i$, we see that the constraint

$$b_0 - a_o^T x \geq \Omega \sqrt{\sum_{i=1}^{L} (a_i^T x - b_i)^2} \tag{24.20}$$

implies that

$$\mathrm{Prob}\{\xi^T z > b_0 - a_o^T x\} \leq \mathrm{Prob}\{\xi^T z > \Omega \|z\|_2\} \leq \epsilon,$$

provided $\Omega \geq \sqrt{2 \log(1/\epsilon)}$ Thus, the SOC condition (24.31) is a sufficient for our chance constraint (24.30) to hold.

Again, the conic constraint (24.31) can be reinterpreted as a robust condition, against ellipsoidal uncertainties:

$$a^T(\xi) x \leq b(\xi), \quad \forall \xi \in \mathcal{U} \tag{24.21}$$

$$\mathcal{U} = \{\xi \mid \quad \|\xi\|_2 \leq \Omega\} \tag{24.22}$$

Define $B_\Omega = \{\xi | \quad \|\xi\|_2 \leq \Omega\}$ and $\mathrm{Box}_1 = \{\xi | \quad \|\xi\|_\infty \leq 1\}$. We can impose robustness against a Box uncertainty:

$$a^T(\xi) x \leq b(\xi), \forall \xi \in \mathrm{Box}_1 \Leftrightarrow b_0 \geq a_0^T x + \Sigma_{i=1}^L \left| a_i^T x - b_i \right|$$

This condition gives us 100% reliability, while a ball uncertainty, which corresponds to the SOCP condition (24.31) gives a reliability of ϵ. If $\Omega \geq 7.44$, then $\exp(-\Omega^2/2) \leq 10^{-12} = \epsilon$, so we get huge reliability for small Ω. In high dimensions, the ball is much smaller in volume than the box, precisely:

$$\frac{\mathbf{vol}B_\Omega}{\mathbf{vol}\mathrm{Box}_1} = \left(\frac{\Omega\sqrt{e\pi/2}}{\sqrt{m}}\right)^m \rightarrow_{m\rightarrow\infty} 0 \tag{24.23}$$

Hence, in high dimensions, it is better to use B_Ω: although it does not give 100% reliability, it does provide a highly reliable solution. A solution robust against a set that is much smaller in volume than the box is still highly reliable over the entire box.

We can even arrange that the set over which we enforce a robustness condition is actually entirely included in the original box (this is not the case with the previous setting). Indeed, consider $\xi \in \mathcal{U} := \mathrm{Box}_1 \cap B_\Omega$. It turns out that an equivalent representation of the robustness constraint $a(\xi)^T x \leq b(\xi),\ \forall \xi \in \mathcal{U}$ is: there exist z, w such that

$$\|z\|_1 + \Omega\sqrt{\Sigma w_i^2} \leq b_0 - a_0^T,\ \ z_i + w_i = b_i - a_i^T x,\ \ \forall i. \tag{24.24}$$

Now this robustness constraint still guarantees a high reliability: If x is feasible for the above robustness condition, and infeasible for the constraint $a(\xi)^T x \leq b(\xi)$ for some realization ξ, then

$$\begin{aligned}
\Sigma\xi_i(a_i^T x - b_i) > b_0 - a_0^T x \ \ &\Rightarrow -\Sigma z_i\xi_i - \Sigma w_i\xi_i > b_0 - a_0^T x \\
&\Rightarrow \Sigma|z_i| - \Sigma w_i\xi_i > b_0 - a_0^T x \\
&\Rightarrow -\Sigma w_i\xi_i > \Omega\sqrt{\Sigma w_i^2}
\end{aligned}$$

Consequently, we have that $\mathrm{Prob}\{a(\xi)^T x > b(\xi)\} \leq \mathrm{Prob}\{-w^T\xi > \Omega\|w\|_2\} \leq \exp(-\Omega^2/2)$.

When $m > \Omega^2$, ξ takes only ± 1 values, the set $\mathrm{Box}_1 \cap B_\Omega$ does not even contain a single realization of ξ. Still, we can find a highly reliable solution by enforcing the robustness constraint (24.35).

This illustrates that a solution that is robust with respect to an uncertainty set that not only is smaller, but is actually contained in the original box, is actually still highly reliable.

24.7 Example: Investment Problem

We have several assets including cash (or, risk-free asset) and stocks, and denote by r the $n+1$ vector of returns over the investment period, with r_0 being the return of cash (say, $r_0 = 1.05$ for a 5% CD). How should we distribute some sum (say, one dollar) over these assets, where we take the returns to be a random variable?

We invest amounts $y_0, y_1, \ldots, y_n$, such that $\Sigma y_i = 1$ and $y_i \geq 0$. The total return of the portfolio is given by $r_0 y_0 + \Sigma r_i y_i \geq t$. We make the further assumption that the returns are random in the form $r_i = \mu_i + \sigma_i\xi_i$, where μ_i is the mean and σ_i is the standard deviation.

We have two models, the first is the robust case where $\xi \in \mathrm{Box}_1$ and the second is the ball-box case where $\xi \in \mathrm{Box}_1 \cap B_\Omega$. Take the tolerance of $\Omega = 3.255$ or $\epsilon = 0.005$ and use $\mu_i = 1.05 + 0.3(200 - i)/199$ and $\sigma_i = 0.05 + 0.6(200 - i)/199$. Using the first model, we get $y_0 = 1$ and $y_i = 0$, which means that a purely robust approach leads to the very conservative investment of putting everything in a risk-free asset. The second model gives a worst-case return of 10.1% with a 0.5% chance of not getting this return. We thus observe that chance constraints allow to make the robustness condition much less conservative, at the expense of a very slight increase in risk.