Lecture 10: Robust and Stochastic Optimization

goals:

- intro to robust + stochastic optimization
- explictly incorporate uncertainty in problem data
- convex reformulations in several interesting cases

- so far, we've assumed that the problem data defining the objective and constraint functions on our optimization problems are known exactly
- in reality, there is almost always some uncertainty in the problem data, due to
  - parameter measurement or estimation error
  - manufacturing tolerances or implementation errors
  - operation variation, simplifying assumptions, etc.
- subfields of Robust + Stochastic Optimization deal with uncertainties on problem data of optimization problems
  - small uncertainty can have large effects on decision value, including causing infeasibility

Robust + Stochastic Optimization

- objective and inequality constraint functions depend on decision variable **and** a random variable $\omega \in \mathbb{R}^d$
    - objective $f(x, \omega)$
    - constraints $g_i(x, \omega) \leq 0$
    - Note: uncertainty in equality constraints not well posed, not considered here
- ω models problem data uncertainty:
    - do not know its exact value, only some information about its probability distribution

- optimization problem

$$\begin{array}{ll} \text{minimize} & f(x, \omega) \\ \text{subject to} & g_i(x, \omega) \leq 0 \quad i = 1, \dots, m \end{array}$$

Makes no sense! $f(x, \omega)$ and $g(x, \omega)$ are random variables/functions, not just numbers

- need to reformulate the problem
- many ways to do so, depending on assumptions on ω and optimization goals

- goals:
    - constraints satisfied on average, with high probability, or always
    - objective small on average, with high probability, or in the worst case

"Certainty equivalent" Problem

expectation operator

$$\begin{aligned} &\text{minimize} && f(x, Ew) \\ &\text{subject to} && g_i(x, Ew) \le 0 \qquad i=1,\dots,m \end{aligned}$$

- i.e. basically ignore parameter variation
- optimal value, feasibility may change significantly in presence of parameter variations

---

Information about uncertainty w

① Exact probability distribution (e.g. Gaussian)

$$w \sim \mathbb{P}_w$$

② Uncertainty sets (i.e. bounds on support of distribution)

$$w \in W$$

③ Inexact info about distribution

a) moments: e.g. $Ew = \bar{w}$, $Eww^T = \Sigma$

b) a finite dataset (or ability to generate samples of w)

$$\{w_1, w_2, \dots, w_{N_s}\}$$

Robust Optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x,w) \le 0 \quad \forall w \in W \subset \mathbb{R}^d \end{array}$$

- without loss of generality (wlog) can consider uncertainty only in constraints (why?)
- want constraint to hold robustly, i.e. for every possible value of w in the uncertainty set W
- convex problem if g convex in x for each w
  - but it's "semi-infinite" since there are an infinite number of constraints (one for each w)
- equivalent to

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \sup_{w \in W} g(x,w) \le 0 \end{array}$$

- can be reformulated as a problem w/ finitely many convex constraints in several interesting cases
  - already saw one example in Lecture 7:

    Robust LP w/ ellipsoidal uncertainty set $\longrightarrow$ SOCP

$$\begin{array}{ll} \min. & c^T x \\ \text{s.t.} & a_i^T x \leq b_i \quad i=1,\dots,m \\ & \forall a_i \in \underbrace{\{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\}}_{W_i} \end{array} \iff \begin{array}{ll} \min. & c^T x \\ \text{s.t.} & \bar{a}_i^T x + \|P_i x\|_2 \leq b_i \\ & i=1,\dots,m \end{array}$$

Ex

How about if $W_i = \{\bar{a}_i + P_i u \mid \|u\| \leq 1\}$? (general norm)

$$\sup\{a_i^T x \mid a_i \in W_i\} = \bar{a}_i^T x + \sup\{u^T P_i^T x \mid \|u\| \leq 1\}$$

$$= \bar{a}_i^T x + \|P_i^T x\|_*$$

$$\iff \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x + \|P_i^T x\|_* \leq b_i \quad i=1,\dots,m \end{array}$$

- if W_i are generated by 1- or ∞-norm, the robust LP is (equivalent to) an LP!

Ex Robust LP with polytopic uncertainty

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \\ & \forall a_i \in W_i = \{a_i \mid D_i a_i \leq d_i\} \end{array}$$

where $D_i \in \mathbb{R}^{k_i \times n}$ and $d_i \in \mathbb{R}^{k_i}$ are given

Equivalently,

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \begin{bmatrix} \max_{a_i} & a_i^T x \\ \text{s.t.} & D_i a_i \le d_i \end{bmatrix} \le b_i \quad i=1,\dots,m \end{aligned}$$

Let's take the dual of the inner LP, giving

$$\begin{aligned} &\text{minimize} && \lambda_i^T d_i \\ &\text{subject to} && D_i^T \lambda_i = x \\ &&& \lambda_i \ge 0 \end{aligned}$$

Then due to strong duality of LPs, we get

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && \begin{bmatrix} \min_{\lambda_i} & \lambda_i^T d_i \\ \text{s.t.} & D_i^T \lambda_i = x \\ & \lambda_i \ge 0 \end{bmatrix} \le b_i \\ &&& i=1,\dots,m \end{aligned} \iff \boxed{\begin{aligned} &\underset{x,\lambda_i}{\text{minimize}} && c^T x \\ &\text{subject to} && \lambda_i^T d_i \le b_i \\ &&& D_i^T \lambda_i = x \\ &&& \lambda_i \ge 0 \\ &&& i=1,\dots,m \end{aligned}}$$

Another LP! (in x, λ_i)

- Robust optimization not restricted to LP
- Many robust reformulations of other convex optimization problems with various types of uncertainty sets
  - e.g. Robust SOCP w/ ellipsoidal uncertainty $\longrightarrow$ SDP
- However, not always possible to tractably reformulate all robust convex problems
  - e.g. Robust SDPs almost always NP-hard
- Active area of research: see e.g. 2009 book by Ben-Tal, El Ghaoui, + Nemirovski

Stochastic Optimization

- one basic form using expectation:

$$\begin{array}{ll} \text{minimize} & \mathbf{E} f(x, w) = F(x) \\ \text{subject to} & \underbrace{\mathbf{E} g_i(x, w)}_{G_i(x)} \le 0 \qquad i = 1, \ldots, m \end{array}$$

- minimize cost + satisfy constraints on average
- if f, g_i are convex in x for each w
  - F, G_i are convex (why?)
  - thus stochastic problem is convex
- F, G_i have analytical expressions in only a few cases; otherwise have to approximate

Ex $f(x) = \|Ax - b\|_2^2$ with A a random matrix and b a random vector

$$\begin{aligned} F(x) = \mathbf{E}_{A,b} f(x) &= \mathbf{E}\,(Ax - b)^T (Ax - b) \\ &= \mathbf{E}\left[x^T (A^T A) x - 2 b^T A x + b^T b\right] \\ &= x^T \underbrace{(\mathbf{E} A^T A)}_{\triangleq P} x - 2 \underbrace{\mathbf{E}\, b^T A}_{\triangleq q^T} x + \underbrace{\mathbf{E}\, b^T b}_{\triangleq r} \end{aligned}$$

$\Rightarrow \mathbf{E} f(x) \le 0 \iff x^T P x - 2 q^T x + r \le 0$

- a standard quadratic inequality that depends only on second moments (covariances) of A, b

Chance Constraints

$$\text{Prob}\left(g_i(x,\omega) \le 0\right) \ge 1-\varepsilon$$

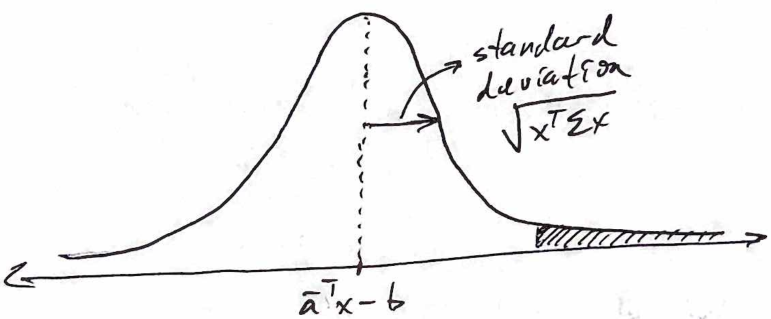
$$\iff \text{Prob}\left(g_i(x,\omega) > 0\right) \le \varepsilon$$

- want constraint satisfied with high probability
  - typically $\varepsilon = 0.1, 0.05, 0.01$
- convex only in a few special cases

Gaussian (normal) distribution

Ex Consider $a^T x \le b$ with $a \sim N(\bar{a}, \Sigma)$

$$\implies a^T x - b \sim N\left(\bar{a}^T x - b,\ x^T \Sigma x\right)$$



$$\text{Prob}(a^T x \le b) = \phi\left(\frac{b - \bar{a}^T x}{\sqrt{x^T \Sigma x}}\right)$$

cdf of $N(0,1)$

so we have

$$\text{Prob}(a^T x \le b) \ge 1-\varepsilon \iff \bar{a}^T x + \underbrace{\phi^{-1}(1-\varepsilon)\,\|\Sigma^{\frac{1}{2}} x\|_2}_{\text{constraint tightening}} \le b$$

- a second-order cone constraint!
  (for $\varepsilon \le 0.5$, so that $\phi^{-1}(1-\varepsilon) \ge 0$)

$$z = \frac{a^T x - b - (\bar{a}^T x - b)}{\sqrt{x^T \Sigma x}} \sim N(0,1)$$

$$\begin{aligned}
&\text{Prob}(a^T x \le b) \\
&= \text{Prob}(a^T x - b \le 0) \\
&= \text{Prob}\left(\frac{a^T x - b - (\bar{a}^T x - b) + (\bar{a}^T x - b)}{\sqrt{x^T \Sigma x}} \le 0 \right) \\
&= \text{Prob}\left(z \le \frac{b - \bar{a}^T x}{\sqrt{x^T \Sigma x}} \right) = \phi\left(\frac{b - \bar{a}^T x}{\sqrt{x^T \Sigma x}} \right) \ge 1 - \varepsilon
\end{aligned}$$

cdf of $N(0,1)$

$$\frac{b - \bar{a}^T x}{\sqrt{x^T \Sigma x}} \ge \phi^{-1}(1-\varepsilon)$$

$\| \Sigma^{\frac{1}{2}} x \|_2$

$$b - \bar{a}^T x \ge \phi^{-1}(1-\varepsilon) \| \Sigma^{\frac{1}{2}} x \|_2$$

$$\bar{a}^T x + \phi^{-1}(1-\varepsilon) \| \Sigma^{\frac{1}{2}} x \|_2 \le b$$

Solving stochastic optimization problems

- analytical solution only in very special cases, when expectations/probabilities can be found analytically
  - e.g. f, g_i are linear or quadratic in ω
  - ω is a discrete random variable (i.e. takes on finitely many values)

- in general, must approximate solution via <u>Monte Carlo sampling</u>
  - known as <u>sample average approximation</u> or <u>scenario approach</u>
  - generate N samples (aka scenarios or realizations) $\{\omega_1, \dots, \omega_N\}$
  - form sample average approximations

$$\hat{F}(x) = \frac{1}{N}\sum_{j=1}^{N} f(x, \omega_j), \quad \hat{G}_i(x) = \frac{1}{N}\sum_{j=1}^{N} g_i(x, \omega_j)$$

    - RVs w/ mean $\mathbb{E} f(x,\omega) = F(x)$, $\mathbb{E} g_i(x,\omega) = G_i(x)$

  - Solve

$$\boxed{\begin{array}{ll} \text{minimize} & \hat{F}(x) \\ \text{subject to} & \hat{G}_i(x) \le 0 \quad i = 1, \dots, m \end{array}}$$

    convex if f, g_i convex fn for each ω

    - solution and optimal value are RVs

- Good approximation if N is large enough

- In principle, we can any reformulation of a stochastic optimization problem by allowing transformations of objective and constraint functions and taking expectations:

$$\text{minimize} \quad E\,\varphi_0(f(x,\omega))$$
$$\text{subject to} \quad E\,\varphi_i(g_i(x,\omega)) \le 0 \qquad i=1,\dots,m$$

where φ_i are risk functions that quantify our dissatisfaction with variations in cost and constraint violations

---

Convex bounds on chance constraints

- in general, chance constraints are not convex and arguably not always the right thing to do in practice

- there are several convex reformulations of chance constraints that also have interesting and practical interpretations as alternatives for quantifying risk of constraint violations

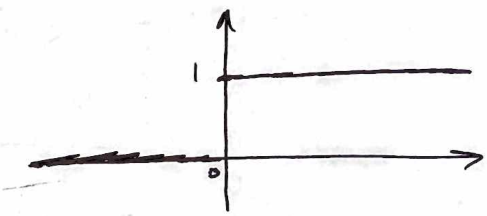
- FACT: For any random vector $z \in \mathbb{R}^d$ and subset $C \subseteq \mathbb{R}^d$

$$\text{Prob}(z \in C) = E\,\mathbb{1}_C(z)$$

where $\mathbb{1}_C(z) = \begin{cases} 1 & z \in C \\ 0 & \text{otherwise} \end{cases}$ indicator function

We can express a chance constraint as

$$E\,\mathbb{1}_{[0,\infty)}(g_i(x, w)) \le \varepsilon$$



risk function

- related to "Value at Risk" (VaR), a common risk metric in finance

- For any scalar RV z and any $\alpha > 0$, we have

$$\text{Prob}(z \ge 0) = \text{Prob}\left(\tfrac{1}{\alpha} z \ge 0\right) = E\,\mathbb{1}_{[0,\infty)}\left(\tfrac{1}{\alpha} z\right).$$

Furthermore, suppose $\psi : \mathbb{R} \to \mathbb{R}$ is a convex, non-decreasing function s.t. $\psi(\tau) \ge \mathbb{1}(\tau)\ \forall \tau$

$$\Longrightarrow E\,\psi\left(\tfrac{1}{\alpha} z\right) \ge E\,\mathbb{1}_{[0,\infty)}\left(\tfrac{1}{\alpha} z\right) = \text{Prob}(z \ge 0)$$

i.e., $E\,\psi(\tfrac{1}{\alpha} z)$ is an upper bound on $\text{Prob}(z \ge 0)$

- This means that

$$E\psi\left(\tfrac{1}{\alpha} g_i(x,\omega)\right) \geq \text{Prob}\left(g_i(x,\omega) > 0\right)$$

and also that

$$\inf_{\alpha > 0} \alpha E\psi\left(\tfrac{1}{\alpha} g_i(x,\omega)\right) \geq \text{Prob}\left(g_i(x,\omega) > 0\right)$$

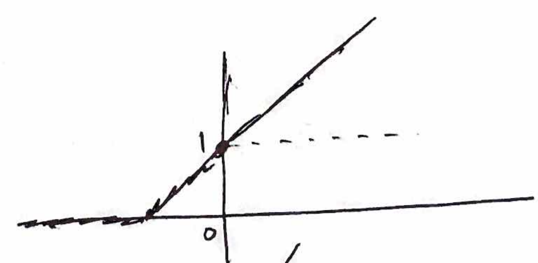
$$\Longrightarrow \boxed{\inf_{\alpha > 0}\left[\alpha E\psi\left(\tfrac{1}{\alpha} g_i(x,\omega)\right) - \alpha\varepsilon\right] \leq 0}$$

this constraint implies the chance constraint and is jointly convex in variables x and α!

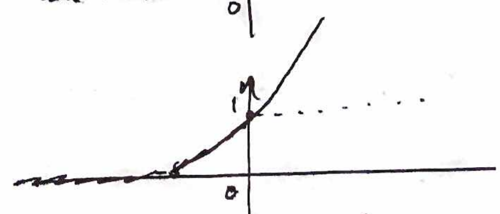
- composition rules with $\psi \circ g$
- perspective function $f(x,\alpha) = \alpha f\left(\tfrac{1}{\alpha} x\right)$

Candidate Risk Functions

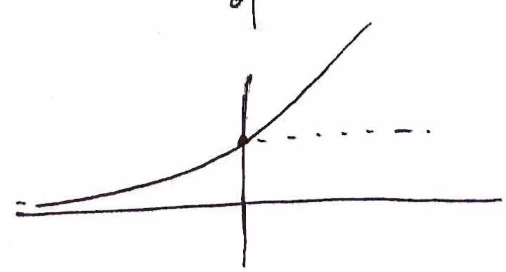
- $\psi(z) = [1+z]_+$ (Markov)



- $\psi(z) = \left([1+z]_+\right)^2$ (Chebyshev)



- $\psi(z) = e^z$ (Bernstein)



- $\psi(z) = (1+z)^2$ ("traditional" Chebyshev)

- Markov:

$$\boxed{\mathbf{E}\left[g_i(x,\omega) + \alpha\right]_+ \le \varepsilon\alpha}$$

with variables x and α (CVaR)

  - closely related to "Conditional Value at Risk", another commonly used risk metric in finance
  - limits both probability/frequency and severity of constraint violations
    - quantifies "how bad is bad"
  - a convex constraint in x and α
  - expectation can't be evaluated analytically, but effectively approximated by sample average approx.

- "Traditional" Chebyshev (assume $g_i^T(x,\omega) = \omega^T x + b$ w/ $\mathbf{E}\omega = \bar{\omega}$, $\mathbf{Var}\,\omega = \Sigma$)
  - w/ a bit of calculus/algebra

$$\boxed{\bar{\omega}^T x + b + \sqrt{\frac{1-\varepsilon}{\varepsilon}}\,\|\Sigma^{\frac{1}{2}} x\|_2 \le 0}$$

  - a second order cone constraint, that only depends on $\bar{\omega}$ and Σ (not necessarily Gaussian)

- equivalent to a "distributionally robust" chance constraint

$$\text{Prob}\left(w^T x + b \leq 0\right) \geq 1-\varepsilon \quad \forall \text{Prob} \in \mathbb{P}(\bar{w}, \Sigma)$$

set of all probability distributions w/ given mean and covariance

- an important emerging area in optimization