# Lecture 8 - Convex Optimization

- A convex optimization problem (or just a convex problem) is a problem consisting of minimizing a convex function over a convex set:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in C, \end{array} \tag{1}$$

Given a convex fcn g(x) and a scalar a, {x: g(x)<=a} is convex.

  - $C$ - convex set.
  - $f$ - convex function over $C$.
- A functional form of a convex problem can written as

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, & i = 1, 2, \ldots, m \\ & h_j(\mathbf{x}) = 0, & j = 1, 2, \ldots, p, \end{array}$$

  $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $h_1, h_2, \ldots, h_p : \mathbb{R}^m \to \mathbb{R}$ are affine functions.
- Note that the functional form does fit into the general formulation (1).

# “Convex Problems are Easy” - Local Minima are Global Minima

**Theorem.** Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}^* \in C$ be a local minimum of $f$ over $C$. Then $\mathbf{x}^*$ is a global minimum of $f$ over $C$.

**Proof.**

- $\mathbf{x}^*$ is a local minimum of $f$ over $C \Rightarrow \exists r > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in C \cap B[\mathbf{x}^*, r]$.
- Let $\mathbf{x}^* \neq \mathbf{y} \in C$. We will show that $f(\mathbf{y}) \geq f(\mathbf{x}^*)$.
- Let $\lambda \in (0,1)$ be such that $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in B[\mathbf{x}^*, r]$.
- Since $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in B[\mathbf{x}^*, r]$, it follows that $f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*))$ and hence by Jensen's inequality:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*)) \leq (1 - \lambda) f(\mathbf{x}^*) + \lambda f(\mathbf{y}).$$

- Thus, the desired inequality $f(\mathbf{x}^*) \leq f(\mathbf{y})$ follows.

## More Results

A small variation of the proof of the last theorem yields the following.

**Theorem.** Let $f : C \to \mathbb{R}$ be a strictly convex function defined on the convex set $C$. Let $\mathbf{x}^* \in C$ be a local minimum of $f$ over $C$. Then $\mathbf{x}^*$ is a strict global minimum of $f$ over $C$.

Another important and easily deduced property of convex problems is that set of optimal solutions is also convex.

**Theorem.** Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Then the set of optimal solutions of the problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

is convex. If, in addition, $f$ is *strictly* convex over $C$, then there exists at most one optimal solution of the problem.

**Proof.** In class

E.g. f(x) = e^x is strictly convex on R, but it has no minimizer. (Note: R is not compact, and f is not coercive.)

# Example

- A Convex Problem:

$$\begin{array}{ll} \min & -2x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 \le 3, \end{array}$$

- A Nonconvex Problem:

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 3 \end{array}$$

# Linear Programming

$$
\text{(LP):} \quad \begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b}, \\ & \mathbf{B}\mathbf{x} = \mathbf{g}. \end{array}
$$

- A convex optimization problem (constraints and objective function are linear/affine and hence convex).
- It is *also* equivalent to a problem of maximizing a convex (linear) function subject to a convex constraints set. Hence, if the feasible set is compact ans nonempty, then there exists at least one optimal solution which is an extreme point=basic feasible solution.
- A more general result drops the compactness assumption and is often called the fundamental theorem of linear programming.

# Convex Quadratic Problems

- Convex quadratic problems are problems consisting of minimizing a convex quadratic function subject to affine constraints.
- The general form is

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{c}, \end{array}$$

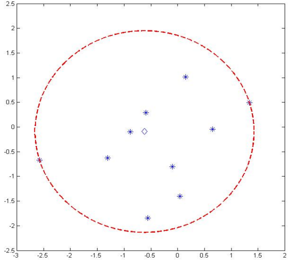
  $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{c} \in \mathbb{R}^m$.

# Chebyshev Center of a Set of Points

**Chebyshev Center Problem.** Given $m$ points $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ in $\mathbb{R}^n$. The objective is to find the center of the minimum radius closed ball containing all the points.

- This ball is called the Chebyshev ball and the corresponding center is the Chebyshev center.
- In mathematical terms, the problem can be written as ($r$ is the radius and $\mathbf{x}$ is the center):

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \mathbf{a}_i \in B[\mathbf{x}, r], \quad i = 1, 2, \ldots, m \end{array}$$




- or:

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \|\mathbf{x} - \mathbf{a}_i\| \leq r, \quad i = 1, 2, \ldots, m. \end{array}$$

# The Portfolio Selection Problem

- We are given $n$ assets numbered as $1, 2, \ldots, n$. Let $Y_j (j = 1, 2, \ldots, n)$ be the RV representing the return from asset $j$.
- We assume that the expected returns are known:

$$\mu_j = E(Y_j), j = 1, 2, \ldots, n,$$

and that the covariances of all the pairs of variables are also known:

$$\sigma_{i,j} = COV(Y_i, Y_j), \quad i, j = 1, 2, \ldots, n.$$

- $x_j (j = 1, 2, \ldots, n)$ - the proportion of budget invested in asset $j$. The decision variables are constrained to satisfy $\mathbf{x} \in \Delta_n$.
- The overall return is the random variable:

$$R = \sum_{j=1}^{n} x_j Y_j,$$

whose expectation and variance are given by:

$$\mathbb{E}(R) = \boldsymbol{\mu}^T \mathbf{x}, \mathbb{V}(R) = \mathbf{x}^T \mathbf{C} \mathbf{x},$$

$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ and $\mathbf{C}$ is the covariance matrix: $C_{i,j} = \sigma_{i,j}$

# The Markowitz Model

- There are several formulations of the portfolio optimization problem, which are all referred to as the "Markowitz model" after Harry Markowitz (1952).
- Minimizing the risk under the constraint that a minimal return level is guaranteed:

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{C}\mathbf{x} \\ \text{s.t} & \boldsymbol{\mu}^T\mathbf{x} \geq \alpha, \\ & \mathbf{e}^T\mathbf{x} = 1, \\ & \mathbf{x} \geq 0, \end{array}$$

- Maximize the expected return subject to a bounded risk constraint:

$$\begin{array}{ll} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t} & \mathbf{x}^T\mathbf{C}\mathbf{x} \leq \beta, \\ & \mathbf{e}^T\mathbf{x} = 1, \\ & \mathbf{x} \geq 0, \end{array}$$

- A penalty approach:

$$\begin{array}{ll} \min & -\boldsymbol{\mu}^T\mathbf{x} + \gamma(\mathbf{x}^T\mathbf{C}\mathbf{x}) \\ \text{s.t} & \mathbf{e}^T\mathbf{x} = 1, \\ & \mathbf{x} \geq 0, \end{array}$$

# QCQP Problems

Quadratically Constrained Quadratic Problems:

$$
\text{(QCQP)} \quad \begin{array}{lll} \min & \mathbf{x}^T \mathbf{A}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 & \\ \text{s.t.} & \mathbf{x}^T \mathbf{A}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i \le 0, & i = 1, 2, \ldots, m, \\ & \mathbf{x}^T \mathbf{A}_j \mathbf{x} + 2\mathbf{b}_j^T \mathbf{x} + c_j = 0, & j = m+1, m+2, \ldots, m+p. \end{array}
$$

$\mathbf{A}_0, \ldots, \mathbf{A}_{m+p}$- $n \times n$ symmetric, $\mathbf{b}_0, \ldots, \mathbf{b}_{m+p} \in \mathbb{R}^n$, $c_0, \ldots, c_{m+p} \in \mathbb{R}$.

- QCQPs are not necessarily convex problems.
- When there are no equality constrainers ($p = 0$) and all the matrices are positive semidefinite: $\mathbf{A}_i \succeq \mathbf{0}, i = 0, 1, \ldots, m$, the problem is convex, and is therefore called a convex QCQP.

?

# The Orthogonal Projection Operator

- Definition. Given a nonempty closed convex set $C$, the orthogonal projection operator $P_C : \mathbb{R}^n \to C$ is defined by

$$P_C(\mathbf{x}) = \operatorname{argmin}\{\|\mathbf{y} - \mathbf{x}\|^2 : \mathbf{y} \in C\}.$$

Nothing orthogonal in this result! And the proof goes through verbatim if ||.|| is any norm with a strictly convex unit ball.

The first important result is that the orthogonal projection exists and is unique.

**The First Projection Theorem.** Let $C \subseteq \mathbb{R}^n$ be a nonempty closed and convex set. Then for any $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection $P_C(\mathbf{x})$ exists and is unique.
?

**Proof.** In class (C does not need to be compact -- it can be unbounded. The existence is guaranteed by the coercivity of the distance function. Uniqueness is guaranteed by strict convexity.)

Compare this result with a standard result in inner-product space: In the special case when C above is a linear subspace of R^n. Then (i) the orthogonal projection operator P_C is *LINEAR* and (ii) x-P_C(x) is orthogonal to C -- all these would *not* hold if the Euclidean norm is replaced by an arbitrary strictly convex norm.

||x|| = sqrt(<x,x>)

# Examples

- $C = \mathbb{R}^n_+$.

$$P_{\mathbb{R}^n_+}(\mathbf{x}) = [\mathbf{x}]_+,$$

where $[\mathbf{v}]_+ = (\max\{v_1, 0\}, \max\{v_2, 0\}, \ldots, \max\{v_n, 0\})^T$.

- A box is a subset of $\mathbb{R}^n$ of the form

$$B = [\ell_1, u_1] \times [\ell_2, u_2] \times \cdots \times [\ell_n, u_n] = \{\mathbf{x} \in \mathbb{R}^n : \ell_i \leq x_i \leq u_i\},$$

where $\ell_i \leq u_i$ for all $i = 1, 2, \ldots, n$.

$$[P_B(\mathbf{x})]_i = \begin{cases} u_i & x_i \geq u_i \\ x_i & \ell_i < x_i < u_i, \\ \ell_i & x_i \leq \ell_i. \end{cases}$$

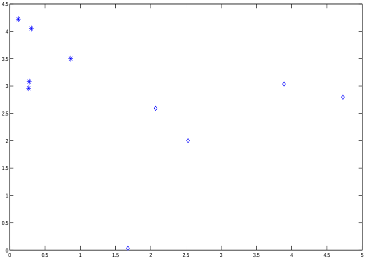
- $C = B[0, r]$.

$$P_{B[\mathbf{0},r]} = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq r, \\ r\frac{\mathbf{x}}{\|\mathbf{x}\|} & \|\mathbf{x}\| > r. \end{cases}$$

What is P_C if C is an affine subspace? See Page 200 of [Beck], basically Lagrange multiplier.

# Linear Classification

- Suppose that we are given two types of points in $\mathbb{R}^n$: type A and type B points.
- $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ - type A.
- $\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \ldots, \mathbf{x}_{m+p} \in \mathbb{R}^n$ - type B.




The objective is to find a linear separator, which is a hyperplane of the form

$$H(\mathbf{w}, \beta) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^T\mathbf{x} + \beta = 0\}$$

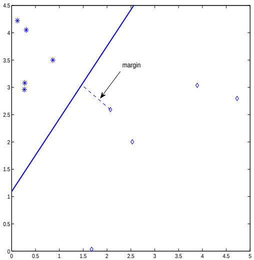
for which the type A and type B points are in its opposite sides:

$$\begin{aligned} \mathbf{w}^T\mathbf{x}_i + \beta &< 0, \quad i = 1, 2, \ldots, m, \\ \mathbf{w}^T\mathbf{x}_i + \beta &> 0, \quad i = m+1, m+2, \ldots, m+p. \end{aligned}$$

**Underlying Assumption:** the two sets of points are linearly separable, meaning that the set of inequalities has a solution.

# Maximizing the Margin

The margin of the separator is the distance of the hyperplane to the closest point.




**The separation problem will thus consist of finding the separator with the largest margin.**

**Lemma.** Let $H(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Let $\mathbf{y} \in \mathbb{R}^n$. Then the distance between $\mathbf{y}$ and the set $H$ is given by

$$d(\mathbf{y}, H(\mathbf{a}, b)) = \frac{|\mathbf{a}^T\mathbf{y} - b|}{\|\mathbf{a}\|}.$$

**Proof.** Later on in lecture 10.

# Mathematical Formulation

- $$\begin{array}{ll} \max & \left\{ \min_{i=1,2,\ldots,m+p} \frac{|\mathbf{w}^T\mathbf{x}_i+\beta|}{\|\mathbf{w}\|} \right\} \\ \text{s.t.} & \mathbf{w}^T\mathbf{x}_i + \beta < 0, \quad i = 1,2,\ldots,m, \\ & \mathbf{w}^T\mathbf{x}_i + \beta > 0, \quad i = m+1, m+2, \ldots, m+p. \end{array}$$

  Nonconvex formulation $\Rightarrow$ difficult to handle.
- the problem has a degree of freedom in the sense that if $(\mathbf{w}, \beta)$ is an optimal solution, then so is any nonzero multiplier of it, that is, $(\alpha\mathbf{w}, \alpha\beta)$ for $\alpha \neq 0$. We can therefore decide that

  $$\min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta| = 1,$$

- Thus, the problem can be written as

  $$\begin{array}{ll} \max & \left\{ \frac{1}{\|\mathbf{w}\|} \right\} \\ \text{s.t.} & \min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta| = 1, \\ & \mathbf{w}^T\mathbf{x}_i + \beta < 0, \quad i = 1,2,\ldots,m, \\ & \mathbf{w}^T\mathbf{x}_i + \beta > 0, \quad i = m+1, 2, \ldots, m+p. \end{array}$$

# Mathematical Formulation Contd.

- $$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} & \min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta| = 1, \\ & \mathbf{w}^T\mathbf{x}_i + \beta \le -1, \quad i = 1,2,\ldots,m, \\ & \mathbf{w}^T\mathbf{x}_i + \beta \ge 1, \quad i = m+1, 2, \ldots, m+p, \end{array}$$

- The first constraint can be dropped (why?)

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} & \mathbf{w}^T\mathbf{x}_i + \beta \le -1, \quad i = 1,2,\ldots,m, \\ & \mathbf{w}^T\mathbf{x}_i + \beta \ge 1, \quad i = m+1, m+2, \ldots, m+p. \end{array}$$

Convex Formulation. in fact a QP !

# Hidden Convexity in Trust Region Subproblems

- $$\text{(TRS):} \quad \min\{\mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c : \|\mathbf{x}\|^2 \le 1\}.$$

  where $\mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$ and $\mathbf{A}$ is an $n \times n$ symmetric matrix. **In general, this is a nonconvex problem**
- By the spectral decomposition theorem, there exist an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_n)$ such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, and hence (TRS) can be rewritten as
  $$\min\{\mathbf{x}^T\mathbf{U}\mathbf{D}\mathbf{U}^T\mathbf{x} + 2\mathbf{b}^T\mathbf{U}\mathbf{U}^T\mathbf{x} + c : \|\mathbf{U}^T\mathbf{x}\|^2 \le 1\}.$$
- Making the linear change of variables $\mathbf{y} = \mathbf{U}^T\mathbf{x}$, the problem reduces to
  $$\min\{\mathbf{y}^T\mathbf{D}\mathbf{y} + 2\mathbf{b}^T\mathbf{U}\mathbf{y} + c : \|\mathbf{y}\|^2 \le 1\}.$$
- Denoting $\mathbf{f} = \mathbf{U}^T\mathbf{b}$, we obtain
  $$\begin{array}{ll} \min & \sum_{i=1}^n d_i y_i^2 + 2\sum_{i=1}^n f_i y_i + c \\ \text{s.t.} & \sum_{i=1}^n y_i^2 \le 1. \end{array} \tag{2}$$

# Hidden Convexity in Trust Region Subproblems Contd.

**Lemma.** Let $\mathbf{y}^*$ be an optimal solution of (2). Then $f_i y_i^* \leq 0$ for all $i = 1, 2, \ldots, n$.

**Proof.**

- Denote the objective function of (2) by $g(\mathbf{y}) \equiv \sum_{i=1}^n d_i y_i^2 + 2\sum_{i=1}^n f_i y_i + c$.
- Let $i \in \{1, 2, \ldots, n\}$. Define $\tilde{\mathbf{y}}$ as

$$\tilde{y}_j = \begin{cases} y_j^* & j \neq i, \\ -y_i^* & j = i. \end{cases}$$

- $\tilde{\mathbf{y}}$ is feasible and $g(\mathbf{y}^*) \leq g(\tilde{\mathbf{y}})$.
- $\sum_{i=1}^n d_i (y_i^*)^2 + 2\sum_{i=1}^n f_i y_i^* + c \leq \sum_{i=1}^n d_i (\tilde{y}_i)^2 + 2\sum_{i=1}^n f_i \tilde{y}_i + c$.
- After cancelleation of terms, $2 f_i y_i^* \leq 2 f_i (-y_i^*)$,
- implying the desired inequality $f_i y_i^* \leq 0$.

# Hidden Convexity in Trust Region Subproblems Contd.

Back to the TRS problem –

- Make the change of variable $y_i = -\mathrm{sgn}(f_i)\sqrt{z_i}(z_i \geq 0)$.
- problem (2) becomes

why convex? :)

$$\begin{array}{ll} \min & \sum_{i=1}^n d_i z_i - 2\sum_{i=1}^n |f_i|\sqrt{z_i} + c \\ \text{s.t.} & \sum_{i=1}^n z_i \leq 1, \\ & z_1, z_2, \ldots, z_n \geq 0. \end{array}$$

- convex optimization problem.

neither a LP nor QP

Open question: is there a mathematical principle behind these clever convexification tricks? When can a problem be convexified?