CONTINUOS OPTIMIZATION

General problem: $\min f(x)$ s.t. $x \in M \subseteq \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$

- Discrete opt – M finite / countable
  P: shortest path, min/max spanning tree, min. max-matching
  NP: integer linear programming, traveling salesman problem, max cut

- Continuous opt. – M uncountably infinite
  ! continuous is often easier than discrete – lin. prog is P
  → linear programming, convex programing, semidefinite programming

The general problem is undecidable
→ let $M :=$ set of inputs s.t. a Turing machine halts
→ Halting problem: does the TM halt for x_0?
→ let $f(x_0) := 0$ and $f(x) := 1$ for $x \neq x_0$
⇒ TM halts for x_0 ⟺ $\min\{f(x) \mid x \in M\} = 0$ ↯

Def: The point $x^* \in M \subseteq \mathbb{R}^n$ is called a

1. (global) minimum ≡ $\forall x \in M: f(x^*) \leq f(x)$
2. strict (global) min ≡ $\forall x \in M: f(x^*) < f(x)$
3. local minimum ≡ $\exists \varepsilon > 0$ s.t. $\forall x \in M \cap \mathcal{U}_\varepsilon(x^*) : f(x^*) \leq f(x)$
4. strict local min ≡ $\exists \varepsilon > 0$ s.t. $\forall x \in M \cap \mathcal{U}_\varepsilon(x^*) : f(x^*) < f(x)$

The minimum doesnt have to exist – note $\min\{e^x \mid x \in \mathbb{R}\}$

Fact: In general it is impossible to reliably find a global minimum

Weierstrass: If M is compact & f continuous, then global min. exists.
↳ in our case (Euclidean space) compact = closed & bounded

Problem classification - typical M

$$\min f(x) \quad s.t. \quad x \in M = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_j(x) \leq 0 , & j = 1, \dots J \\ h_\ell(x) = 0 , & \ell = 1, \dots L \end{array} \right\}$$

Where $f, g_j, h_\ell : \mathbb{R}^n \to \mathbb{R}$.

Notation: $\tilde{g} : \mathbb{R}^n \to \mathbb{R}^J$, $\tilde{h} : \mathbb{R}^n \to \mathbb{R}^L$ $\Rightarrow M = \{x \in \mathbb{R}^n \mid \tilde{g}(x) \leq 0, \tilde{h}(x) = 0\}$

① Linear programming: f, g_j, h_ℓ are linear

② Unconstrained opt: $M = \mathbb{R}^n$

③ Convex optimization: f, g_j are convex & h_ℓ are linear

Transformations of opt. problems

$$M = \{x \in \mathbb{R}^n \mid \tilde{g}(x) \leq 0, \tilde{h}(x) = 0\}$$

① Finding maximum $\to \max\{f(x) \mid x \in M\} = -\min\{-f(x) \mid x \in M\}$

② Transformation of functions

The optimization problem

$$\min f(x) \quad s.t. \quad \tilde{g}(x) \leq 0 , \quad \tilde{h}(x) = 0$$

can be transformed to

$$\min \varphi(f(x)) \quad s.t. \quad \tilde{\psi}(\tilde{g}(x)) \leq 0 , \quad \tilde{\eta}(\tilde{h}(x)) = 0$$

provided:

- $\varphi(z)$ is increasing on its domain: $f(x) < f(y) \Rightarrow \varphi(f(x)) < \varphi(f(y))$
  ↳ z^2, $z^{1/2}$, $\log(z)$
- $\psi_j(z)$ preserves nonnegativity: $g_j(x) \leq 0 \iff \psi_j(g_j(x)) \leq 0$
  ↳ z^3, $\log(z+1)$, $\log(z+k) - \log(k)$
- $\eta_\ell(z)$ preserves roots: $h_\ell(x) = 0 \iff \eta_\ell(h_\ell(x)) = 0$
  ↳ z^2, absolute value

both opt. problems have the same minima points

the optimal values are different, but they can be computed from the optimal solutions

Lemma: The condition $g(x) \geq \alpha$ can be transformed to
$f(g(x)) \geq f(\alpha)$ for any increasing f

Proof: Condition $g(x) \geq \alpha \rightsquigarrow g(x) - \alpha \geq 0$

define $\Psi(z) := f(z+\alpha) - f(\alpha)$ ... this preserves nonnegativity

$\Psi(g(x) - \alpha) \geq 0 \iff f(g(x)) - f(\alpha) \geq 0 \iff f(g(x)) \geq f(\alpha)$ ∎

Example: Convert to linear program

$\min x^3 y^2$ s.t. $2x^2 y^{-1} \leq 1$, $3x^{-1} y^3 \leq 2$, $x > 0$, $y > 0$

log: $x^3 y^2 \rightsquigarrow 3 \log x + 2 \log y$
$2x^2 y^{-1} \leq 1 \rightsquigarrow \log 2 + 2 \log x - \log y \leq \log 1 = 0$
$3x^{-1} y^3 \leq 2 \rightsquigarrow \log 3 - \log x + 3 \log y \leq \log 2$

$x > 0 \iff \log x \in \mathbb{R}$
$y > 0 \iff \log y \in \mathbb{R}$

define $X := \log x$, $Y := \log Y$

$\min 3X + 2Y$ s.t. $2X - Y \leq -\log 2$, $-X + 3Y \leq \log 2 - \log 3$, $X \in \mathbb{R}$, $Y \in \mathbb{R}$

③ Move objective function to constraints

$\min f(x)$ s.t. $x \in M$ $\rightsquigarrow$ $\min z$ s.t. $x \in M$, $z = f(x)$
$\rightsquigarrow$ $\min z$ s.t. $x \in M$, $f(x) \leq z$

same optimum

Example: Transform to linear obj. function

a) $\min x^2 + y^2$ s.t. $x + y \leq 1$, $x \geq 0$, $y \geq 0$

$\rightsquigarrow \min z$ s.t. $x + y \leq 1$, $x \geq 0$, $y \geq 0$, $z = x^2 + y^2$

b) $\min \max\{x^2, y^2, xy\}$ s.t. $x + y \leq 1$

$\rightsquigarrow \min z$ s.t. $x + y \leq 1$, $z \geq \max\{x^2, y^2, xy\}$
$\Rightarrow \min z$ s.t. $x + y \leq 1$, $z \geq x^2$, $z \geq y^2$, $z \geq xy$

④ Elimination of equations and variables

→ some equations or variables of the program might be redundant
⇒ removing them reduces complexity

Example: Simplify the following problem

$$\min\ x^2+y^2+z^2-8x \quad \text{s.t.} \quad x+y+z=1, \quad x-y=0$$

- $x-y=0 \Rightarrow y=x$
- $x+y+z=1 \Rightarrow z=1-2x$

$$\left.\right\} \quad x^2+y^2+z^2-8x = x^2+x^2+(1-2x)^2-8x$$
$$= 2x^2+1-4x+4x^2-8x$$
$$= 6x^2-12x+1$$

$\Rightarrow \min\ 6x^2-12x+1 \quad \text{s.t.} \quad x\in\mathbb{R}$ has the same opt. value

↳ $\frac{d}{dx} = 12x-12 = 0 \Rightarrow x^*=1$, opt. value $= 6-12+1 = \underline{\underline{-5}}$

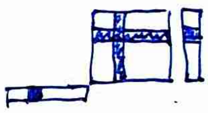
$\Rightarrow$ the original problem has opt. solution $\underline{\underline{x=1,\ y=1,\ z=-1}}$

Linear Algebra Recap

Def: A symmetric matrix $A\in\mathbb{R}^{n\times n}$ is called

- positive definite (PD) $\equiv \forall x\in\mathbb{R}^n,\ x\neq 0:\ x^TAx>0$
- positive semi-definite (PSD) $\equiv \forall x\in\mathbb{R}^n:\ x^TAx\geq 0$

PD $\Rightarrow$ PSD



A is PSD $\Rightarrow \forall i: A_{ii}\geq 0$

↳ $x := e^i = (0,\dots 0, \overset{i}{1}, 0,\dots 0)$, then $x^TAx = A_{ii}\geq 0$

Lemma: $A\in\mathbb{R}^{n\times n}$ symmetric $\Rightarrow$ all eigenvalues of A are real

Theorem (Spectral decomposition): $A\in\mathbb{R}^{n\times n}$ symmetric $\Rightarrow A = Q\Lambda Q^T$ where

- Λ is a diagonal matrix containing the (real) eigenvalues of A
- Q is orthogonal $\equiv Q^T = Q^{-1} \equiv$ columns are orthonormal
- columns of Q are the normalized eigenvectors of A

Def: $K\subseteq\mathbb{R}^n$ is a cone $\equiv \forall\alpha\geq 0:\ x\in K\Rightarrow \alpha x\in K$

↳ K is a pointed cone $\equiv$ it doesn't contain any line

$\mathbb{R}^n$ is a cone but not pointed

Def: For vectors $v_1,\dots v_k$, matrices $A_1,\dots A_k$ define cones

- $\text{cone}(\{v_1,\dots v_k\}) := \{x \mid x=\sum \alpha_i v_i,\ \alpha_i\geq 0\}$
- $\text{cone}(\{A_1,\dots A_k\}) := \{A \mid A=\sum \alpha_i A_i,\ \alpha_i\geq 0\}$

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be symmetric. The following conditions are equivalent

① A is PSD PD

② all eigenvalues of A are non-negative positive

③ $A \in \operatorname{cone}\{xx^T \mid x \in \mathbb{R}^n\}$ $A \in \text{Interior}(\operatorname{cone}\{\dots\})$

Proof: ① $\Rightarrow$ ② $\Rightarrow$ ③ $\Rightarrow$ ①

Not difficult, uses observations bellow and spectral theorem

- $x \in \mathbb{R}^n \Rightarrow xx^T \in \mathbb{R}^{n\times n}$ is PSD $\because$ $\alpha^T xx^T \alpha = \underbrace{(\alpha^T x)}_{\in\mathbb{R}}\underbrace{(x^T\alpha)}_{\in\mathbb{R}} = (\alpha^T x)^2 \geq 0$
- A PSD, $\alpha > 0 \Rightarrow \alpha A$ PSD $\because$ $x^T(\alpha A)x = \alpha x^T A x \geq 0$
- A, B PSD $\Rightarrow A+B$ PSD $\because$ $x^T(A+B)x = x^T A x + x^T B x \geq 0$

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be PSD. Then $\operatorname{rank}(A) = \#$ of positive eigenvalues of A

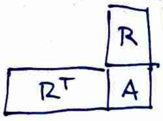
Pf: Spectral theorem: $A = Q\Lambda Q^T$, $\operatorname{rank}(A) = \operatorname{rank}(\Lambda)$

Theorem: $A \in \mathbb{R}^{n\times n}$ is PSD $\iff \exists R \in \mathbb{R}^{m\times n}$, $m \geq n$ s.t. $A = R^T R$

Moreover, $\operatorname{rank}(R) = \operatorname{rank}(A) = \#$ positive eigenvalues of A

Corollary: $A \in \mathbb{R}^{n\times n}$ is PD $\iff \exists R \in \mathbb{R}^{m\times n}$, $m \geq n$, full-rank s.t. $A = R^T R$

$\hookrightarrow \operatorname{rank}(R) = n$



Mathematical Analysis Recap

Def: $f: \mathbb{R}^n \to \mathbb{R}$ define

- gradient $\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n}\right)^T$ $\nabla f: \mathbb{R}^n \to \mathbb{R}^n$

- hessian $\nabla^2 f = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & & \vdots \\ \vdots & & \ddots & \\ f_{x_nx_1} & \dots & & f_{x_nx_n} \end{bmatrix}$ $f_{x_ix_j} := \frac{\partial^2 f}{\partial x_i \partial x_j}$

$\nabla^2 f: \mathbb{R}^n \to \mathbb{R}^{n\times n}$

Note: If all second partial derivatives of f are continuous ($f \in C^2$), then

$f_{x_ix_j} = f_{x_jx_i} \Rightarrow \nabla^2 f$ is symmetric

Intuition: $\nabla f \sim$ first derivative

$\nabla^2 f \sim$ second derivative

Proposition: Let $A \in \mathbb{R}^{n\times n}$, $a \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, $f: \mathbb{R}^n \to \mathbb{R}$

function f	∇f	$\nabla^2 f$
α	0_n	$0_{n\times n}$
$a^T x = x^T a$	a	$0_{n\times n}$
$x^T x$	$2x$	$2I_{n\times n}$
$x^T A x$	$(A+A^T)x$	$A+A^T$
Symmetric A $x^T A x$	$2Ax$	$2A$

Examples

a) $f(x) = x^T A x + a^T x + \alpha$ $\quad \nabla f = (A+A^T)x + a$, $\quad \nabla^2 f = A + A^T$

b) $f(x) = (Ax+a)^T x = x^T(Ax+a) = x^T A x + x^T a = x^T A x + a^T x$

$= x^T A^T x + a^T x \in \mathbb{R} \Rightarrow \alpha^T = \alpha$

c) $f(x) = (Ax+a)^T(Bx+b)$

Recall: $(A+B)^T C = A^T C + B^T C$

$$f(x) = x^T A^T B x + x^T A^T b + a^T B x + a^T b$$
$$= x^T A B x + x^T A b + x^T B^T a + a^T b$$

Assume A, B symmetric

! $\not\Rightarrow$ $A \cdot B$ symmetric !

$$\nabla f(x) = (AB + B^T A^T)x + Ab + B^T a$$
$$= (AB + BA)x + Ab + Ba$$
$$\nabla^2 f(x) = AB + BA$$

Theorem (Taylor with Lagrange remainder): Let $f: \mathbb{R}^n \to \mathbb{R}$, $\tilde{x}^0 \in \mathbb{R}^n$.
Then for $\forall \tilde{h} \in \mathbb{R}^n$ exists $\theta \in (0,1)$ s.t.

$$f(\tilde{x}^0 + \tilde{h}) = f(\tilde{x}^0) + \nabla f(\tilde{x}^0)^T \tilde{h} + \frac{1}{2}\tilde{h}^T \nabla^2 f(\tilde{x}^0 + \theta\tilde{h})\tilde{h}$$

for $f: \mathbb{R} \to \mathbb{R}$: $f(x^0 + h) = f(x^0) + f'(x^0)\cdot h + \frac{1}{2} f''(x^0 + \theta h)\cdot h^2$

Note: We will often use this for $\tilde{h} = \lambda\tilde{y}$, which results in

$$f(\tilde{x}^0 + \lambda\tilde{y}) = f(\tilde{x}^0) + \lambda \cdot \nabla f(\tilde{x}^0)^T \tilde{y} + \frac{1}{2}\lambda^2 \tilde{y}^T \nabla^2 f(\tilde{x}^0 + \theta\lambda\tilde{y})\tilde{y}$$

Unconstrained optimization

$\min f(x) \quad$ s.t. $\quad x \in \mathbb{R}^n \quad , \quad f: \mathbb{R}^n \to \mathbb{R}$ differentiable

Theorem (FO necessary): $X^* \in \mathbb{R}^n$ local extrema $\Rightarrow \nabla f(X^*) = \sigma$.

Proof: WLOG assume X^* is a local minimum $\Rightarrow f(x^*)$ ∵ minima

$$\nabla_i f(X^*) = \frac{\partial f(X^*)}{\partial X_i} = \lim_{h\to 0} \frac{f(X_1^*, \dots X_{i-1}^*, X_i^* + h, X_{i+1}^*, \dots X_n^*) - f(X^*)}{h}$$

$$\left.\begin{aligned} &= \lim_{h\to 0^+} \frac{\geq 0}{0^+} \geq 0 \\ &= \lim_{h \to 0^-} \frac{\geq 0}{0^-} \leq 0 \end{aligned}\right\} = 0$$

∎

Theorem (SO necessary): $f \in C^2$, X^* local minimum. Then $\nabla^2 f(X^*)$ is PSD.

Proof: $\nabla^2 f$ is symmetric ∵ $f \in C^2$

For $\forall \tilde{y} \in \mathbb{R}^n$, $\forall \lambda \in \mathbb{R}$, $\exists \theta \in (0,1)$ s.t.

$$\underbrace{f(X^* + \lambda\tilde{y})}_{\geq f(X^*)} = \underbrace{f(X^*)}_{f(X^*)} + \underbrace{\lambda \cdot \nabla f(X^*)^T \tilde{y}}_{0 \,∵\, \nabla f(X^*) = \sigma} + \underbrace{\tfrac{1}{2}\lambda^2 y^T \nabla^2 f(X^* + \theta\lambda\tilde{y})\tilde{y}}_{(*) \text{ must be } \geq 0}$$

$\Rightarrow \lambda^2 \tilde{y}^T \nabla^2 f(X^* + \theta\lambda\tilde{y})\tilde{y} \geq 0 \iff \tilde{y}^T \nabla^2 f(X^* + \theta\lambda\tilde{y})\tilde{y} \geq 0$

$\Rightarrow$ for $\forall y \in \mathbb{R}^n$ do $\lambda \to 0 \Rightarrow y^T \cdot \nabla^2 f(X^*) y \geq 0$ ∎

Theorem (SO sufficient): $f \in C^2$, $X^* \in \mathbb{R}^n$. If $\nabla f(X^*) = \sigma$ & $\nabla^2 f(X^*)$ is PD, then X^* is a strict local minimum.

Proof: Similar as last proof, for any $\lambda \neq 0$, $\tilde{y} \neq \sigma$, $\exists \theta \in (0,1)$ s.t.

$$\lambda \cdot \nabla f(X^*)^T \tilde{y} = 0 \quad , \quad \tfrac{1}{2}\lambda^2 y^T \nabla^2 f(X^* + \theta\lambda\tilde{y})\tilde{y} > 0$$

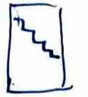
Therefore $f(X^* + \lambda\tilde{y}) > f(X^*)$. ∎

Example: We need $\nabla^2 f$ to be PD, PSD is not enough

$f(x) = -x^4 \quad \dots \quad x^* = 0$ is local maximum

- sufficient not satisfied ... $f'(0) = 0$ but $f''(0) = 0$
- necessary is satisfied ... $f'(0) = 0$ & $f''(0) = 0$

The least squares method

A  full-rank

System of m linear equations $Ax = b$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$
where matrix A has rank $= n$, $m \gg n$ (usually).

- a lot of equations $\Rightarrow$ this system usually doesnt have a solution

$\Rightarrow$ we seek an approximate solution

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \qquad \dots \; \|\cdot\|_2 = \text{Euclidean norm}$$

$\Rightarrow z^2$ is increasing $\Rightarrow$ we can transform this to

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = (Ax-b)^T(Ax-b) = x^T A^T A x - x^T A^T b - b^T A x + b^T b$$

$\Rightarrow$ we need to solve the unconstrained problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad , \quad f(x) = x^T A^T A x - 2 b^T A x + b^T b$$

$\Rightarrow$ solve $\nabla f = 0$ & $\nabla^2 f$ PD ... note: $A^T A$ is always PD
↳ because A is full-rank

$$\nabla f = 2A^T A x - 2A^T b$$

$$\nabla^2 f = 2A^T A \quad \rightarrow \text{PD always}$$

↳ need $\nabla f = 0 \Rightarrow A^T A x = A^T b \Rightarrow \underline{\underline{x = (A^T A)^{-1} A^T b}}$

Convex sets

Def: $M \subseteq \mathbb{R}^n$ is a convex set $\equiv \forall x_1, x_2 \in M,\ \forall \lambda \in [0,1]: \lambda x_1 + (1-\lambda) x_2 \in M$

Theorem: Let $k \geq 2$. Then $M \subseteq \mathbb{R}^n$ is convex $\iff$

$$\forall x_1, \dots x_k \in M,\quad \forall \lambda_1, \dots \lambda_k \geq 0,\quad \sum \lambda_i = 1 \ :\ \sum \lambda_i x_i \in M$$

Proof: $\Leftarrow$ obvious

$\Rightarrow$ induction, base case $k=2$. Now want $\sum_{i=1}^{k} \lambda_i x_i \in M$

$$\sum_{i=1}^{k} \lambda_i x_i = \underbrace{\left(\sum_{i=1}^{k-1} \lambda_i x_i\right)}_{\text{vector } \circledast} + \lambda_k x_k = \underbrace{\sum_{i=1}^{k-1} \lambda_i}_{\lambda} \cdot \underbrace{\frac{\circledast}{\sum_{i=1}^{k-1} \lambda_i}}_{y_1} + \underbrace{\lambda_k}_{1-\lambda} \underbrace{x_k}_{y_2} \in M$$

$\frac{\circledast}{\Sigma} \in M$ by induction for $k-1$. Then use definition of convexity ∎

If $\forall i \in I: M_i$ is convex, then $\bigcap_{i \in I} M_i$ is convex.

Def: The convex hull of a set $M \subseteq \mathbb{R}^n$ is $\operatorname{conv}(M) := \bigcap \left\{ M' \subseteq \mathbb{R}^n \,\middle|\, \begin{matrix} M \subseteq M' \\ M' \text{ convex} \end{matrix} \right\}$

Theorem: $M \subseteq \mathbb{R}^n$ is convex $\iff$ $M = \operatorname{conv}(M)$.

Proof: $\Rightarrow$: because M is convex, it is one of the intersected supersets

$\Leftarrow$: $\operatorname{conv}(M)$ is intersection of convex sets $\Rightarrow$ convex ∎

Def: Two nonempty sets M, N are separable $\equiv \exists a \in \mathbb{R}^n, a \neq 0, \exists b \in \mathbb{R}$ s.t.

$\forall x \in M: a^T x \leq b$ & $\exists x \in M \cup N: a^T x \neq b$

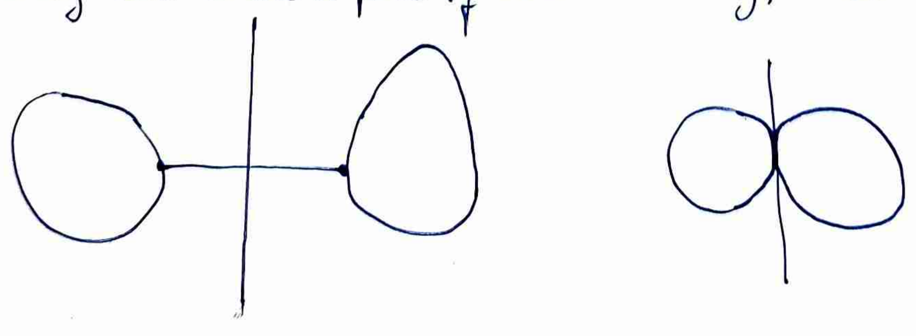
$\forall x \in N: a^T x \geq b$

Intuition: a is the normal vector of the hyperplane $a^T x = 0$

↳ $a^T x = b$ is $a^T x$ shifted from origin in direction of a by b

Theorem: Two nonempty convex sets M, N separable $\iff$ their interiors don't intersect.

↳ they can share a part of the boundary, but not the interior



Def: Let $M \subseteq \mathbb{R}^n$ be closed and convex and let $x^* \in \text{Boundary}(M)$. Then use the separation theorem to separate $\{x^*\}$ from M by a hyperplane $a^T x = b$ (... $x^* \in$ plane $\Rightarrow a^T x^* = b$) s.t. $\forall x \in M: a^T x \leq b$

We call the plane $a^T x = a^T x^*$ a supporting hyperplane of M.

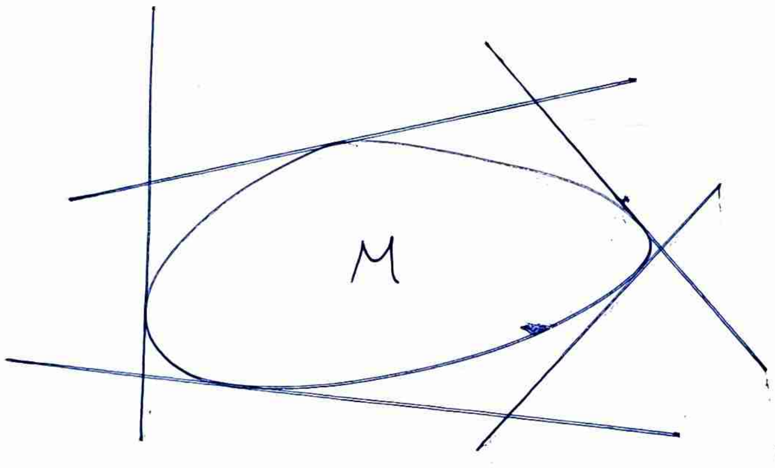
Theorem: $M \subseteq \mathbb{R}^n$ closed and convex. Then M is equal to the intersection of the positive halfspaces defined by all of its supporting hyperplanes

$$M = \bigcap \{ \{x \in \mathbb{R}^n \mid a^T x \leq a^T x^*\} \mid x^* \in M \}$$

Proof:

$M \subseteq \bigcap$: since for $\forall x^*$ we have $\forall x \in M: a^T x \leq b = a^T x^* \Rightarrow M \subseteq \bigcap$

$\bigcap \subseteq M$: for contradiction let $x^0 \in \bigcap$ but $x^0 \notin M$. Then we can separate $\{x^0\}$ from M by a supporting hyperplane $a^T x \leq b$, which can not be included in $\bigcap$, since $a^T x^0 > b$ $\lightning$ ∎



Convex functions

Def: Let $M \subseteq \mathbb{R}^n$ be convex. Then $f: \mathbb{R}^n \to \mathbb{R}$ is

- convex on M $\equiv$ $(\forall x_1, x_2 \in M)(\forall \lambda_1, \lambda_2 \geq 0,\ \lambda_1 + \lambda_2 = 1)$:

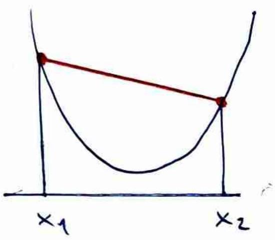
$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

- strictly convex on M $\equiv$ $(\forall x_1, x_2 \in M,\ x_1 \neq x_2)(\forall \lambda_1, \lambda_2 > 0,\ \lambda_1 + \lambda_2 = 1)$:

$$f(\lambda_1 x_1 + \lambda_2 x_2) < \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

- concave on M $\equiv$ $-f(x)$ is convex on M
- strictly concave on M $\equiv$ $-f(x)$ is strictly concave on M

f is linear (affine) on M $\equiv$ it is convex & concave on M



convex: the entire segment is above the graph

strictly convex: the entire segment without its endpoints is strictly above the graph

Example: All vector norms are convex

$$\|\lambda_1 x_1 + \lambda_2 x_2\| \leq \|\lambda_1 x_1\| + \|\lambda_2 x_2\| = \lambda_1 \|x_1\| + \lambda_2 \|x_2\|$$

Theorem (Jensen's inequality): Let $k \geq 2$, $M \subseteq \mathbb{R}^n$ be convex. Then $f: \mathbb{R}^n \to \mathbb{R}$ is convex on M $\Leftrightarrow$

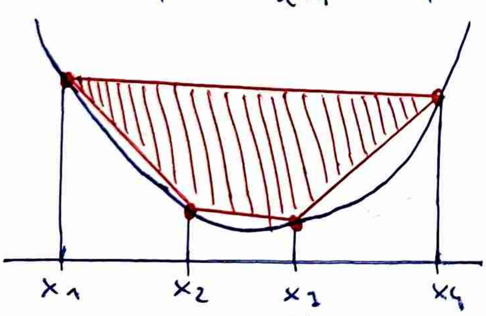
$$(\forall x_1, \dots x_k \in M)(\forall \lambda_1, \dots \lambda_k \geq 0,\ \textstyle\sum_i \lambda_i = 1): f\left(\sum_i \lambda_i x_i\right) \leq \sum_i \lambda_i f(x_i)$$

Proof: Same as for convex sets. Induction ... let $\Delta = \sum_{i=1}^{k-1} \lambda_i$

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) = f\Big(\Delta \cdot \underbrace{\sum_{i=1}^{k-1} \frac{\lambda_i}{\Delta} x_i}_{*} + \lambda_k x_k\Big) \leq \Delta \cdot f(*) + \lambda_k f(x_k) \leq \sum_{i=1}^{k} \lambda_i f(x_i)$$

↳ f is convex ↳ induction step

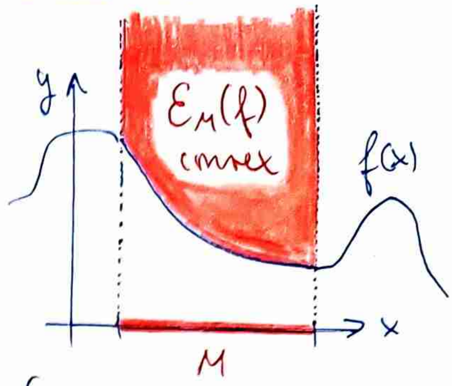
Note: $f(*) = \sum_{i=1}^{k-1} \frac{\lambda_i}{\Delta} f(x_i)$ by induction step for $k-1$ ■



convex: The convex hull of the points $x_1, \dots x_k$ is entirely above the graph of f

Def: The <u>epigraph</u> of a convex function $f:\mathbb{R}^n \to \mathbb{R}$ on $M \subseteq \mathbb{R}^n$ is the set

$$\mathcal{E}_M(f) := \{(x,z) \in \mathbb{R}^{n+1} \mid x \in M,\ f(x) \leq z\}$$



Theorem: Let $M \subseteq \mathbb{R}^n$ be convex, then $f:\mathbb{R}^n \to \mathbb{R}$ is
(Fenchel) <u>convex on M</u> $\iff$ <u>$\mathcal{E}_M(f)$ is a convex set</u>

Proof: $\Rightarrow$: Assume f is convex and let $(x_1, z_1), (x_2, z_2) \in \mathcal{E}$

$$\lambda_1(x_1, z_1) + \lambda_2(x_2, z_2) = (\underbrace{\lambda_1 x_1 + \lambda_2 x_2}_{x}, \underbrace{\lambda_1 z_1 + \lambda_2 z_2}_{z}), \text{ need } f(x) \leq z$$

M convex $\Rightarrow \lambda_1 x_1 + \lambda_2 x_2 \in M$

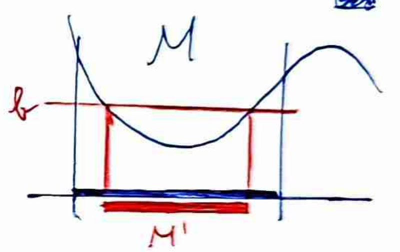
f convex on $M \Rightarrow f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \leq \lambda_1 z_1 + \lambda_2 z_2$

$\Leftarrow$: Assume $\mathcal{E}$ convex and let $x_1, x_2 \in M$. Consider $\lambda_1 x_1 + \lambda_2 x_2 \in M$

$\mathcal{E}$ convex: $\lambda_1(x_1, f(x_1)) + \lambda_2(x_2, f(x_2)) = (\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 f(x_1) + \lambda_2 f(x_2)) \in \mathcal{E}$

$f(x) \leq z$: $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$ ■

Theorem: Let $M \subseteq \mathbb{R}^n$ be convex, $f:\mathbb{R}^n \to \mathbb{R}$ convex on M.
Then for $\forall b \in \mathbb{R}$ is <u>$M' := \{x \in M \mid f(x) \leq b\}$ convex</u>



Note: This is used in optimization, usually the feasible set M is described by a system of inequalities $g_j(x) \leq 0$. If g_j convex $\Rightarrow M$ convex.

Proof: We will show that $M_b := \{x \in \mathbb{R} \mid f(x) \leq b\}$ is convex

then $M' = M \cap M_b$ is convex

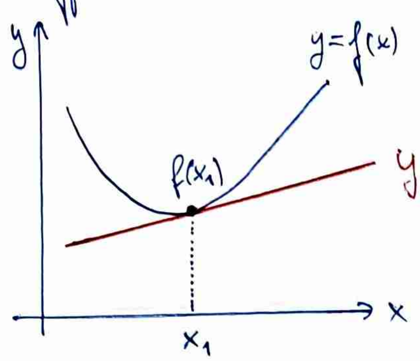
$\Rightarrow$ let $x_1, x_2 \in M_b \Rightarrow f(x_1) \leq b,\ f(x_2) \leq b$, then

$$f(\lambda_1 x_1 + \lambda_2 x_2) \underset{\hookrightarrow f \text{ convex}}{\leq} \lambda_1 f(x_1) + \lambda_2 f(x_2) \leq \lambda_1 b + \lambda_2 b = (\lambda_1 + \lambda_2) b = b$$ ■

Fact: If $f:\mathbb{R}^n \to \mathbb{R}$ is convex on $M \subseteq \mathbb{R}^n$, then f is <u>continuous</u> on Interior(M).

! continuous $\not\Rightarrow$ differentiable

Differentiation and convex function



$y - f(x_1) = f'(x_1)(x - x_1) \Rightarrow f(x) - f(x_1) \geq f'(x_1)(x - x_1)$

↳ the tangent line must lie bellow the graph

Theorem (FO characterization): Let $\emptyset \neq M \subseteq \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$ differ. on open $N \supseteq M$.

$$f(x) \text{ convex on } M \iff \forall x_1, x_2 \in M: f(x_2) - f(x_1) \geq \nabla f(x_1)^T (x_2 - x_1)$$

Proof: $\Rightarrow$: let $x_1, x_2 \in M$ and $\lambda \in (0,1)$ be arbitrary. Then

f convex: $f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda) f(x_1) + \lambda f(x_2)$

$f((1-\lambda)x_1 + \lambda x_2) - f(x_1) \leq -\lambda f(x_1) + \lambda f(x_2) = \lambda (f(x_2) - f(x_1))$

$\Rightarrow f(x_2) - f(x_1) \geq \dfrac{f((1-\lambda)x_1 + \lambda x_2) - f(x_1)}{\lambda}$

Taking $\lambda \to 0$ we get the derivative of $g(h) = f(x_1 + h(x_2 - x_1))$

$f(x_2) - f(x_1) \geq g'(0) = \nabla f(x_1 + 0) \cdot \frac{d}{dh}(x_1 + h(x_2 - x_1))\Big|_0 = \nabla f(x_1)(x_2 - x_1)$

$\Leftarrow$: let $x_1, x_2 \in M$ and set $x := \lambda_1 x_1 + \lambda_2 x_2$, $\lambda_1 + \lambda_2 = 1$. We know

$(*)$ $f(x_1) - f(x) \geq \nabla f(x)^T (x_1 - x) = \nabla f(x)^T (x_1 - \lambda_1 x_1 - \lambda_2 x_2) = \lambda_2 \nabla f(x)^T (x_1 - x_2)$

$(+)$ $f(x_2) - f(x) \geq \nabla f(x)^T (x_2 - x) = \nabla f(x)^T (x_2 - \lambda_1 x_1 - \lambda_2 x_2) = -\lambda_1 \nabla f(x)^T (x_1 - x_2)$

$\lambda_1 (*) + \lambda_2 (+)$: $\lambda_1 (f(x_1) - f(x)) + \lambda_2 (f(x_2) - f(x)) \geq 0$

$\Rightarrow \lambda_1 f(x_1) + \lambda_2 f(x_2) \geq (\lambda_1 + \lambda_2) f(x) = f(x) = f(\lambda_1 x_1 + \lambda_2 x_2)$ ■

Remark: $f(x)$ strictly convex $\iff \forall x_1 \neq x_2 \in M$: $f(x_2) - f(x_1) > \nabla f(x_1)^T (x_2 - x_1)$

Theorem (SO characterization): Let $\emptyset \neq M \subseteq \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R} \in C^2$.

$$f(x) \text{ convex on } M \iff \forall x \in M: \nabla^2 f(x) \text{ is PSD} \quad \to f''(x) \geq 0$$

Remark:

$f(x)$ strictly convex on $M \Rightarrow \nabla^2 f(x)$ is PD for almost all $x \in M$
otherwise it is PSD →

$\nabla^2 f(x)$ is PD on $M \Rightarrow f(x)$ is strictly convex on M

Proof: Let $x^* \in M$ be arbitrary. Because f is C^2 for $\forall \tilde{y} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ and $x^* + \lambda \tilde{y} \in M$ (we only care about M) there exists $\theta \in (0,1)$ s.t.

$$(*) \quad f(x^* + \lambda \tilde{y}) = f(x^*) + \lambda \cdot \nabla f(x^*)^T \tilde{y} + \tfrac{1}{2}\lambda^2 \tilde{y}^T \nabla^2 f(x^* + \theta\lambda\tilde{y})\,\tilde{y}$$

$\Rightarrow$: let f be convex and use the F.O. characterization for

$$x_2 = x^* + \lambda\tilde{y},\ x_1 = x^* : \quad f(x_2) - f(x_1) \geq \nabla f(x_1)(x_2 - x_1)$$
$$f(x^* + \lambda\tilde{y}) \geq f(x^*) + \lambda \cdot \nabla f(x^*)\,\tilde{y}$$

This combined with $(*)$ implies that

$$\tilde{y}^T \nabla^2 f(x^* + \theta\lambda\tilde{y})\,\tilde{y} \geq 0 \xrightarrow{\lambda \to 0} \tilde{y}^T \nabla^2 f(x^*)\, y \geq 0 \Rightarrow \nabla^2 f(x^*) \text{ PSD}$$

$\Leftarrow$: Because $\nabla^2 f$ is PSD on M, we have in $(*)$

$$\tilde{y}^T \nabla^2 f(x^* + \theta\lambda\tilde{y})\,\tilde{y} \geq 0 \Rightarrow f(x^* + \lambda y) \geq f(x^*) + \lambda \cdot \nabla f(x^*)^T \tilde{y}$$

Which shows convexity of f thanks to the F.O. characterization, $\begin{matrix} x_2 = x^* + \lambda\tilde{y} \\ x_1 = x^* \end{matrix}$ can be made arbitrary using x^*, $\tilde{y}$ and λ. ∎

Example: Consider the quadratic function $f: \mathbb{R}^n \to \mathbb{R}$

$$f(x) = x^T A x + a^T x + \alpha, \quad A \in \mathbb{R}^{n \times n} \text{ symmetric},\ a \in \mathbb{R}^n,\ \alpha \in \mathbb{R}$$

$$\nabla^2 f(x) = 2A \Rightarrow f(x) \text{ convex} \Leftrightarrow A \text{ is PSD}$$

Properties of convex functions

1. f, g convex, $\alpha \geq 0 \Rightarrow \alpha f(x)$, $f(x) + g(x)$ are convex
2. f, g convex, nonnegative, nondecreasing or nonincreasing $\Rightarrow f(x) \cdot g(x)$ is convex
3. f convex, nonnegative, nondecreasing; g concave, positive, nonincreasing $\Rightarrow f(x)/g(x)$ is convex

Now let $\tilde{f}: \mathbb{R}^n \to \mathbb{R}^k$, $\tilde{g}: \mathbb{R}^k \to \mathbb{R}$

4. f_i convex for $i = 1, \dots k$; $\tilde{g}$ convex and nondecreasing in each coordinate $\Rightarrow \tilde{g}(\tilde{f}(x))$ is convex
5. f_i concave for $i = 1, \dots k$; $\tilde{g}$ convex & nonincreasing in each coordinate $\Rightarrow \tilde{g}(\tilde{f}(x))$ is convex

Convex optimization

min $f(x)$ s.t. $x \in M$, where M convex & f convex on M

Typically $M = \{x \in \mathbb{R}^n \mid g_j \leq 0,\ j = 1 \dots J\}$ where g_j are convex.

Examples: The following are convex opt. problems

a) min $x_1 + x_2$ s.t. $x_1^2 + x_2^2 \leq 2$

b) min $x_1^2 + x_2^2 + 2x_2$ s.t. $x_1^2 + x_2^2 \leq 2$

$\hookrightarrow D^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is PSD $\Rightarrow f$ convex

Theorem (Fenchel): For every convex opt. problem we have:

① Each local optimum is a global optimum

② The optimal solution set is convex

③ If $f(x)$ is strictly convex, then the optimum is unique or none

Proof:

① let $x^0 \in M$ be a local opt and suppose $\exists x^* \in M$ s.t. $f(x^*) < f(x^0)$.

$$\Rightarrow f(\lambda x^* + (1-\lambda) x^0) \leq \lambda f(x^*) + (1-\lambda) f(x^0)$$
$$< \lambda f(x^0) + (1-\lambda) f(x^0) = f(x^0)$$

This holds for any point x' on the line x^0 — x^*

$\Rightarrow$ let $\lambda \to 0 \Rightarrow x' \to x^0$ and we have $f(x') < f(x^0)$

$\Rightarrow x^0$ is not a local optimum ↯

② let x_1, x_2 be opt. solutions with $f(x_1) = f(x_2) = z$, then (f convex)

$$\Rightarrow f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) = \lambda_1 z + \lambda_2 z = z$$

But since z is the global optimal value we must have $f(\lambda_1 x_1 + \lambda_2 x_2) = z$

③ suppose $x_1 \neq x_2$ are optimal, $f(x_1) = f(x_2) = z$, then (f strictly convex)

$$\Rightarrow f(\lambda_1 x_1 + \lambda_2 x_2) < \lambda_1 f(x_1) + \lambda_2 f(x_2) = z$$

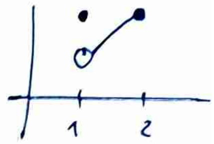
This contradicts the fact that z is the global optimal value ↯ ∎

Example: A convex opt. problem doesn't always have an optimal solution

a) M unbounded: $\min e^x$ s.t. $x \in \mathbb{R}$

b) M compact but f not continuous

$\min f(x)$ s.t. $x \in [1,2]$ where $f(x) := \begin{cases} x, & 1 < x \leq 2 \\ 2, & x = 1 \end{cases}$



Solving convex opt. problems using differentiation

Recall: Unbounded problem $\min_{x \in \mathbb{R}^n} f(x)$ we know that F.O. necessary-cond.

$$x^* \in \mathbb{R}^n \text{ optimal } \Rightarrow \nabla f(x^*) = 0$$

We will show that for convex problems it is sufficient if M is open

Theorem: Let $\emptyset \neq M \subseteq \mathbb{R}^n$ be an open convex set, $f: M \to \mathbb{R}$ convex and differentiable on M. Then consider $\min f(x)$ s.t. $x \in M$

$x^* \in M$ is an optimal solution $\Leftrightarrow$ $\nabla f(x^*) = 0$

Proof: $\Rightarrow$: x^* is optimal sol $\Rightarrow$ local minimum $\Rightarrow \nabla f(x^*) = 0$

$\Leftarrow$: let $\nabla f(x^*) = 0$. By the F.O. characterizations of convex functions:

for $\forall x \in M$: $f(x) - f(x^*) \geq \nabla f(x^*)^T (x - x^*) = 0$

Therefore $f(x) \geq f(x^*)$ and x^* is an optimal solution ∎

Example: We need M to be open!

$\min x$ s.t. $x \in M = [1,2]$ ... $f(x) = x$ is differentiable, M is convex

↳ but M is not open!

→ optimal solution: $x^* = 1 \Rightarrow f'(x^*) = 1 \neq 0$!

$\Rightarrow$ we need a stronger theorem

Theorem: Let $\emptyset \neq M \subseteq \mathbb{R}^n$ be a convex set, $f: M' \to \mathbb{R}$ convex and differentiable on an open set $M' \supseteq M$. Then consider $\min f(x)$ s.t. $x \in M$

$x^* \in M$ is an optimal solution $\Leftrightarrow$ $\forall y \in M: \nabla f(x^*)^T (y - x^*) \geq 0$.

Corollary: $\nabla f(x^*) = 0 \Rightarrow x^*$ is optimal sol. even if M is not open

↳ $\nabla f(x^*) = 0 \Rightarrow \nabla f(x^*)^T (y - x^*) \geq 0$

Proof: Want: x^* opt $\iff \forall y \in M: \nabla f(x^*)^T(y-x^*) \geq 0$

$\Rightarrow$: suppose $\exists y \in M$ s.t. $\nabla f(x^*)^T(y-x^*) < 0$

Recall: directional derivative of f at $\tilde{x} \in \mathbb{R}^n$ in dir. $\tilde{u} \in \mathbb{R}^n$ is

$$D_u f(\tilde{x}) := \lim_{\lambda \to 0} \frac{f(\tilde{x} + \lambda \tilde{u}) - f(\tilde{x})}{\lambda} = \nabla f(x)^T \tilde{u}$$

↳ if f is differentiable – TD exists

Therefore we have

$$\nabla f(x^*)(y-x^*) = \lim_{\lambda \to 0^+} \frac{f(x^* + \lambda(y-x^*)) - f(x^*)}{\lambda} < 0$$

This means that for sufficiently small λ: $f(\lambda y + (1-\lambda)x^*) < f(x^*)$

Since M is convex: $\lambda y + (1-\lambda)x^* \in M$, but then x^* is not an optimum ↯

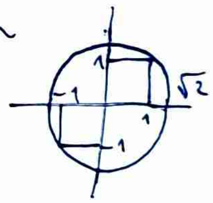
$\Leftarrow$: By F.O. characterization of convex functions we have for $\forall y \in M$

$$f(y) - f(x^*) \geq \nabla f(x^*)(y-x^*) \geq 0$$

Therefore for $\forall y \in M$: $f(x^*) \leq f(y) \Rightarrow x^*$ is optimal sol. ∎

Examples

a) $\min x_1 + x_2$ s.t. $x_1^2 + x_2^2 \leq 2$ ... M not open



👁 optimum is $x^* = (-1,-1)$

verification: $\nabla f(x^*) = (1,1)^T$

need: for $\forall y \in M$: $\nabla f(x^*)^T(y-x^*) = (1,1)\begin{pmatrix} y_1+1 \\ y_2+1 \end{pmatrix} = y_1 + y_2 + 2 \geq 0$

↳ $y_1^2 + y_2^2 \leq 2 \Rightarrow y_1 + y_2 \geq -2$ ... clearly true

b) $\min x_1^2 + x_2^2 + 2x_2$ s.t. $x_1^2 + x_2^2 \leq 2$

$\nabla f = (2x_1, 2x_2 + 2)^T$

$\nabla f(x) = 0 \iff x = (0,-1)^T$

Since $x^* = (0,-1)$ has $\nabla f(x^*) = 0$ and therefore

for $\forall y \in M$: $\nabla f(x^*)(y-x^*) = 0$, it is an optimal sol

$\Rightarrow \underline{\underline{x^* = (0,-1)^T}}$

Quadratic programming

$\min x^T C x + d^T x$ s.t. $x \in M$, where $\begin{cases} C \in \mathbb{R}^{n\times n} \text{ symmetric} \\ d \in \mathbb{R}^n \\ M = \{x \in \mathbb{R} \mid \underbrace{Ax \le b}\} \end{cases}$ $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$

Remark: $f(x)$ convex $\iff \nabla^2 f = C$ is PSD

$\Rightarrow$ convex quadratic program ... special type of convex optimization

Fact
- C is PSD $\Rightarrow$ effectively solvable in polynomial time
- C not PSD $\Rightarrow$ NP-hard, even if only 1 eigenvalue negative

We will show: C is negative definite $\Rightarrow$ NP-hard

equivalently: $\max_{x\in M} x^T C x = -\min_{x \in M} -x^T C x$ ... $-C$ ND $\Rightarrow$ C PD

Theorem: If C is PD then $\max_{x\in M} x^T C x$ is NP-hard.

Proof: We will reduce the NP-complete problem of set-partitioning to a Q. prog.

Problem: Given $\alpha_1, \dots \alpha_n \in \mathbb{N}$, is it possible to split them into two subsets of equal Σ-size?

Equivalently: is there $x \in \{\pm 1\}^n$ s.t. $\sum \alpha_i x_i = 0$?

$\Rightarrow \max \sum_{i=1}^n x_i^2$ s.t. $\sum_{i=1}^n \alpha_i x_i = 0$, $x \in M = [0,1]^n$

optimal value $= n \iff$ SetPartitioning has a solution

This program is quadratic ... $C = I_{n\times n}$ and all constraints linear ■

Example: Investment options (prices) $e_1, \dots e_n > 0$, capital $K > 0$, return of investment i is c_i. Portfolio selection problem

$\max c^T x$ s.t. $e^T x = K$, $x \ge 0$ ← linear program

- We don't know returns c_i, but only expected return values $\tilde{c}_i = \mathbb{E}[c_i]$ and the covariance matrix $\Sigma := \mathrm{cov}(c_1, \dots c_n) = \mathbb{E}[(c-\tilde{c})(c-\tilde{c})^T]$ ... Fact: Σ PSD
- for $x \in \mathbb{R}^n$, the expected obj. function value is $\mathbb{E}[c^T x] = \tilde{c}^T x$ and the variance of the —— " —— is $\mathrm{Var}[c^T x] = x^T \Sigma x$.
- maximizing the expected returns value

$\max \tilde{c}^T x$ s.t. $e^T x = K$, $x \ge 0$ ← linear program
- taking into account the risks of investments, γ = risk aversion coefficient

$\max \tilde{c}^T x - \gamma \cdot x^T \Sigma x$ s.t. $e^T x = K$, $x \ge 0$ ← convex quadratic program

Examples of dual cones

① Nonnegative orthant is self-dual ... $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$

② Lorentz cone is self-dual ... $L^* = L$

③ Cone of PSD matrices is self dual ... $(S^n_+)^* = S^n_+$

↳ ! $K = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is PSD}$

$K^* = \{B \in \mathbb{R}^{n \times n} \mid \forall A \in K : \langle A|B \rangle \geq 0\}$

→ we need to define the inner product

⊛ $\langle A|B \rangle = tr(AB) = \sum_{ij} a_{ij} b_{ij}$ ← sum of element-wise product

Properties of dual cones

① K^* is a closed convex cone ... for arbitrary set $K \subseteq \mathbb{R}^n$

② If K is a closed convex cone, then $(K^*)^* = K$

↳ this gives intuition that dual of the dual is the primal

③ K_1, K_2 cones $\Rightarrow K_1 \times K_2$ is a cone and $(K_1 \times K_2)^* = K_1^* \times K_2^*$

④ $K_1 \subseteq K_2$ cones $\Rightarrow K_1^* \supseteq K_2^*$

Example: Cone combination

$$\begin{aligned} \min\ & c^T x \\ \text{s.t. } & Ax_1 \geq b \\ & Bx_2 \geq_K d \end{aligned} \quad \rightsquigarrow \quad \begin{aligned} \max\ & b^T y + d^T z \\ \text{s.t. } & A^T y + B^T z = c \\ & y \geq 0 \ , \ z \geq_{K^*} 0 \end{aligned}$$

Example: Semidefinite programming

→ variable matrix $Y \in \mathbb{R}^{m \times m}$

PSD order

$$\begin{aligned} \min\ & c^T x \\ \text{s.t. } & \sum_{k=1}^{n} x_k A^{(k)} \succeq B \end{aligned} \quad \rightsquigarrow \quad \begin{aligned} \max\ & tr(BY) \\ \text{s.t. } & tr(A^{(k)} Y) = c_k \ , \ k = 1, \dots n \\ & Y \succeq 0 \end{aligned}$$

$A^{(k)}, B \in \mathbb{R}^{m \times m}$ symmetric

Intuition: $B \sim$ RHS vector, $\dim(B) = m \times m \Rightarrow Y \in \mathbb{R}^{m \times m}$

- dual obj. function $= \langle B|Y \rangle = tr(BY)$ ⊛
- dual constraints

$y^T A = c^T \sim \langle Y | A^{(k)} \rangle = tr(A^{(k)} Y) = c_k$ → only short for some linear equation

Strong duality for convex cone programming

Theorem (strong duality): The primal and dual optimal values are the same if at least one of the following conditions holds

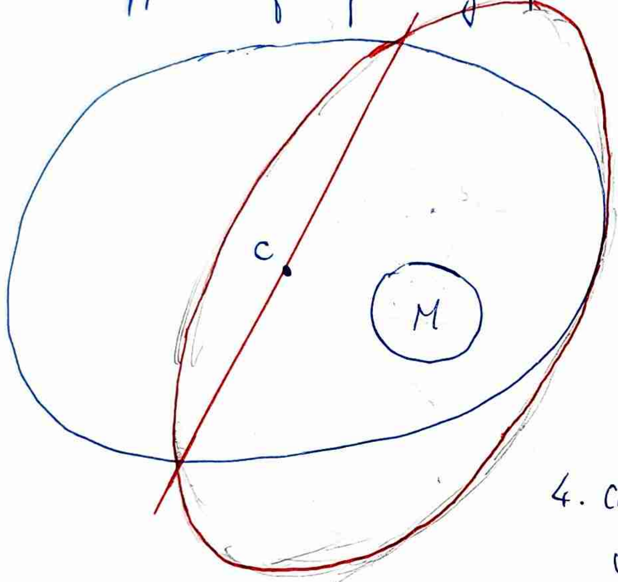
1) P is strictly feasible $\equiv \exists x \in \mathbb{R}^n$ s.t. $Ax >_K b$
2) D is strictly feasible $\equiv \exists y >_{K^*} 0$ s.t. $A^T y = c$

Ellipsoid method for convex optimization

→ unlike quadratic programming, convex optimization is in general usually solvable in polynomial time

→ the ellipsoid method finds a feasible $x \in M$

↳ approach for finding optimal $x^* \in M$ is similar



1. Construct a sufficiently large ellipsoid $\mathcal{E}$ s.t. $M \subseteq \mathcal{E}$
2. if center of $\mathcal{E} = c \in \mathcal{E}$ ... return c
3. else construct a separating hyperplane $a^T x \leq b$ containing c s.t. $M \subseteq \{x \mid a^T x \leq b\}$ ... positive halfspace
4. construct a smaller (min volume) ellipsoid covering $\mathcal{E} \cap \{x \mid a^T x \leq b\}$
5. repeat until $c \in M$ or we prove $M = \emptyset$

Convergence: fact: the volume of $\mathcal{E}$ decreases exponentionaly

Correctness & complexity: for polynomial complexity we need

① M is not too "flat" or too "large"

→ exist reasonably large $r, R > 0$ s.t. $\text{Ball}(r) \subseteq M \subseteq \text{Ball}(R)$

② separation oracle

a) for $\forall x^0 \in \mathbb{R}^n$ we need to be able to check $x^0 \in M$ in polynomial time

b) for $x^0 \notin M$ we need to be able to find a vector $0 \neq a \in \mathbb{R}^n$ s.t.

$$a^T x \geq \sup_{x \in M} a^T x \quad \text{, this gives us separating hyperplane } a^T x = a^T x^0$$

→ implementation: if a constrained $g(x) \leq 0$ was violated $\Rightarrow a := \nabla g(x^0)$

$g(x^0) > 0$ & g convex $\Rightarrow a^T(x - x^0) = \nabla g(x^0)(x - x^0) \leq g(x) - g(x^0) < g(x) \leq 0$, $\forall x \in M$

Karush-Kuhn-Tucker optimality conditions

$$\min f(x) \quad \text{s.t.} \quad x \in M = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_j \leq 0, & j = 1, \dots J \\ h_l = 0, & l = 1, \dots L \end{array} \right\}$$

↳ $f: \mathbb{R}^n \to \mathbb{R}$ differentiable ↳ not necessarily convex

Recall:

- $M = \mathbb{R}^n \Rightarrow$ unconstrained optimization

  x^* optimum $\Rightarrow \nabla f(x^*) = 0$

- M convex & f convex on M $\Rightarrow$ convex optimization

  $\nabla f(x^*) = 0 \Rightarrow x^*$ optimum

Goal: generalize this

Def: $I(x) := \{ j \mid g_j(x) = 0 \} \rightarrow$ active constraints

Theorem (KKT conditions): Let $\{ \nabla h_l(x^*) \mid l = 1, \dots L \} \cup \{ \nabla g_j(x^*) \mid j \in I(x^*) \}$ be linearly independent vectors. Then

x^* local optimum $\Rightarrow (\exists \tilde{\lambda} \in \mathbb{R}^J, \tilde{\lambda} \geq 0)(\exists \tilde{\mu} \in \mathbb{R}^L)$ s.t.

$$\text{KKT conditions} \begin{cases} \nabla f(x^*) + \underbrace{\nabla \tilde{h}(x^*)}_{\text{matrix}} \mu + \nabla \tilde{g}(x^*) \lambda = 0 \\ \lambda^T \tilde{g}(x^*) = 0 \quad \circledast \end{cases}$$

Intuition: $\tilde{\lambda} \geq 0$ together with $\circledast$ says either

$$\left.\begin{array}{ll} \bullet\ g_j(x^*) < 0 & \Rightarrow \lambda = 0 \\ \bullet\ \lambda > 0 & \Rightarrow g_j(x^*) = 0 \end{array}\right\} \text{ because } \lambda_j g_j(x^*) = 0$$

$\Rightarrow$ if $g_j(x^*)$ is not strict, $j \notin I(x^*)$, then the variable λ_j doesn't act in the KKT conditions because $\lambda_j = 0$

Remark (Lagrange multipliers): Let $\{ \nabla h_l(x^*) \mid l = 1, \dots L \}$ be lin. ind.

x^* local optimum $\Rightarrow (\exists \tilde{\mu} \in \mathbb{R}^L)$ s.t. $\nabla f(x^*) + \nabla \tilde{h}(x^*) \cdot \mu = 0$

Intuition: KKT is generalized LM, where we add the active constraints g_j, $j \in I(x^*)$ to the system of equations

Problem: We don't know x^* so checking lin. ind. is difficult.

Solution: Assume something stronger (less general) but easier to check

$\Rightarrow$ Slater's condition: $\exists x^0 \in M$ s.t. $\tilde{g}(x^0) < 0$

Theorem: Consider $\min f(x)$ s.t. $\tilde{g}(x) \leq 0$, $x \in M$, where $f(x)$, $g_j(x)$ are convex functions and M is a convex set. Suppose that Slater's condition is satisfied. Then

x^* optimum $\Rightarrow \exists \tilde{\lambda} \geq 0$ s.t. x^* is also an optimum for

$$\min f(x) + \lambda^T \tilde{g}(x) \quad \text{s.t. } x \in M \qquad \text{and } \lambda^T \tilde{g}(x^*) = 0$$

Corollary: Suppose that Slater's condition is satisfied for the convex optimization problem

$$\min f(x) \quad \text{s.t. } \tilde{g}(x) \leq 0 \qquad \ldots f, g_j \text{ convex}$$

Then

x^* optimum $\Rightarrow \exists \tilde{\lambda} \geq 0$ s.t. the KKT conditions are satisfied $\Rightarrow$

$$\nabla f(x^*) + \nabla g(x^*) \lambda = 0$$
$$\lambda^T g(x^*) = 0$$

Proof:

x^* is optimum for $\min f(x) + \lambda^T \tilde{g}(x)$, $x \in \mathbb{R}^n$ & $\lambda^T g(x^*) = 0$

↳ unconstrained program $\Rightarrow$ gradient $= \nabla f(x^*) + \nabla \tilde{g}(x) \lambda = 0$

Corollary: Suppose the Slater's condition is satisfied for the convex optimization problem

$$\min f(x) \quad \text{s.t. } \tilde{g}(x) \leq 0, \quad Ax = b \qquad \ldots f, g_j \text{ convex}$$

Then

x^* optimum $\Rightarrow (\exists \tilde{\lambda} \geq 0)(\exists \tilde{\mu})$ s.t. the KKT conditions are satisfied $\Rightarrow$

$$\nabla f(x^*) + \nabla g(x^*) \lambda + A^T \mu = 0$$
$$\lambda^T g(x^*) = 0$$

Theorem (Sufficient KKT): Let $x^* \in \mathbb{R}^n$ be a feasible solution of

$$\min f(x) \quad \text{s.t.} \quad g(x) \le 0,$$

→ not necessarily all g_j

let $f(x)$ and $g_j(x)$, $j \in I(x^*)$ be convex functions. Then

KKT conditions satisfied for x^* $\Rightarrow$ x^* is optimal solution

Proof: KKT conditions satisfied $\Rightarrow$

$$\exists \lambda \ge 0 \text{ s.t. } \nabla f(x^*) + \nabla \tilde{g}(x^*) \tilde{\lambda} = 0 \quad \& \quad \lambda^T \tilde{g}(x^*) = 0$$

f convex: $f(x) - f(x^*) \ge \nabla f(x^*)(x - x^*)$ ← F.O. characterization

g_j convex: $g_j(x) - g_j(x^*) \ge \nabla g_j(x^*)(x - x^*)$

→ $j \in I(x^*)$

KKT: $\nabla f(x^*) = - \nabla \tilde{g}(x^*) \tilde{\lambda}$

→ Recall: $j \notin I(x^*) \Rightarrow \lambda_j = 0 \Rightarrow$ don't care about g_j

$$f(x) - f(x^*) \ge \nabla f(x^*)(x - x^*) = - \sum_{j \in I(x^*)} \lambda_j \nabla g_j(x^*)(x - x^*)$$

$$\ge - \sum_{j \in I(x^*)} \lambda_j \big(g_j(x) - \overbrace{g_j(x^*)}^{0 \,\to\, j \in I(x^*)}\big)$$

! there is a $-$

$$= - \sum_{j \in I(x^*)} \underbrace{\lambda_j}_{\ge 0} \cdot \underbrace{g_j(x)}_{\le 0} \ge 0$$

Therefore $f(x) \ge f(x^*)$ $\Rightarrow$ x^* is optimum ∎

Corollary: Suppose that Slater's condition is satisfied for

$$\min f(x) \quad \text{s.t.} \quad g(x) \le 0, \quad \text{where } f, g_j \text{ convex}$$

Then

x^* optimal solution $\Leftrightarrow$ KKT conditions are satisfied for x^*

Methods for solving optimization problems

Example: Finding a solution of $Ax = b$ is costly for large sparse A

$\Rightarrow$ consider the unconstrained quadratic program and solve it

$$\min \frac{1}{2} x^T A x - b^T x \quad \text{s.t.} \quad x \in \mathbb{R}^n$$

$$x^* \text{ is optimum} \Rightarrow \nabla f(x^*) = \frac{1}{2}(A + A^T) x^* - b = 0$$
$$\nabla^2 f(x^*) = \frac{1}{2}(A + A^T) \text{ is PSD}$$

$\Rightarrow$ if A is PSD, then we have (symetric isnt enough, need to satisfy)

$$x^* \text{ optimum} \Rightarrow \frac{1}{2}(A + A^T) x^* - b = A x^* - b = 0$$
$$\Rightarrow x^* \text{ is a solution of } A x^* = b$$

Unconstrained problems

$$\min f(x) \quad \text{s.t.} \quad x \in M$$

① Newtons method – iterative, current guess $= x_k$

→ finds $x^* \in \mathbb{R}^n$ s.t. $\nabla f(x^*) = 0$ ... necessary condition for optimum

$$\Rightarrow q(x) := f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

$$q(x_k) = f(x_k), \quad \nabla q(x_k) = \nabla f(x_k), \quad \nabla^2 q(x_k) = \nabla^2 f(x_k)$$

→ approximate $\nabla f = 0$ by solving $\nabla q = 0$

$$\nabla q(x) = 0 \iff \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$$

→ this is a $n \times n$ system of linear equations which we can solve for $\tilde{\delta}$

$$\nabla^2 f(x_k) \cdot \tilde{\delta} = -\nabla f(x_k) \Rightarrow \tilde{x}_{k+1} := \tilde{x}_k + \tilde{\delta}$$

② Gradient methods

→ finds a local minimum, current guess $= x_k$

→ find a direction d_k in which $f(x)$ locally decreases $\Rightarrow \nabla f(x_k)^T d_k < 0$

$$x_{k+1} := x_k + \alpha_k \cdot d_k, \quad \text{for a carefully chosen } \alpha_k \in \mathbb{R}$$

$\Rightarrow \alpha_k :=$ optimal solution of $\min f(x_k + \alpha d_k)$ s.t. $\alpha \in \mathbb{R}$

↳ obj. function is single variable (easy) ... solve e.g. using Newtons m.

Heuristic for direction

• steepest descent : $d_k := -\nabla f(x_k)$

Robust linear programming

→ in practice, there are often uncertainties in the data

⇒ we want solutions which are "robust" - feasible even for data variance

Def: An <u>interval matrix</u> is $[\underline{A}, \overline{A}]$, where

- $\underline{A}$ = lower bounds, $\overline{A}$ = upper bounds, $\overline{A}, \underline{A} \in \mathbb{R}^{m \times n}$

We write $A \in [\underline{A}, \overline{A}] \equiv \forall ij: \underline{A}_{ij} \le A_{ij} \le \overline{A}_{ij}$

<u>Examples</u>

① $\min c^T x$ s.t. $\begin{array}{c} Ax \le b \\ x \ge 0 \end{array}$ where $A \in [\underline{A}, \overline{A}]$, $b \in [\underline{b}, \overline{b}]$

x is <u>robust feasible</u> $\equiv (\forall A \in [\underline{A}, \overline{A}]) (\forall b \in [\underline{b}, \overline{b}]): Ax \le b$

$x \ge 0 \Rightarrow x$ is robust feasible $\Leftrightarrow \overline{A} x \le \underline{b}$

⇒ <u>robust program</u>: $\min c^T x$ s.t. $\overline{A} x \le \underline{b}$, $x \ge 0$

② $\min c^T x$ s.t. $Ax \le b$ where $A \in [\underline{A}, \overline{A}]$, $b \in [\underline{b}, \overline{b}]$

→ we want a nice characterization of robust solutions

→ let $a^T x \le b_i$ be a selected inequality

x is a robust solution of $a^T x \le b_i$

$\Leftrightarrow (\forall a \in [\underline{a}, \overline{a}]) (\forall b_i \in [\underline{b}_i, \overline{b}_i]): a^T x \le b_i$

$\Leftrightarrow \max \{a^T x \mid a \in [\underline{a}, \overline{a}]\} \le \underline{b}_i$

<u>Lemma</u>: Denote

- $a_c := \frac{1}{2}(\underline{a} + \overline{a})$ ... vector of interval midpoints
- $a_\Delta := \frac{1}{2}(\overline{a} - \underline{a})$ ... vector of interval radii

Then $\max\limits_{a \in [\underline{a}, \overline{a}]} a^T x = a_c^T x + a_\Delta^T |x|$, where $|x|$ is entry-wise abs. value

<u>Proof</u>: Focus on the i-th summand of $a^T x$:

$$a_i \cdot x_i = a_{c_i} x_i + (a_i - a_{c_i}) x_i \le a_{c_i} x_i + |a_i - a_{c_i}| \cdot |x_i| \le a_{c_i} x_i + a_{\Delta_i} |x_i|$$ ■

<u>Corollary</u>:

x is a <u>robust solution</u> of $a^T x \le b_i \Leftrightarrow a_c^T x + a_\Delta^T |x| \le \underline{b}_i$

→ we can rewrite this complex constraint into linear constraints:

$$a_c^T x + a_\Delta^T |x| \leq \underline{b_i} \rightsquigarrow a_c^T x + a_\Delta^T \cdot y \leq \underline{b_i}, \quad x \leq y, \quad -x \leq y$$

⇒ robust program:

$$\min \ c^T x \quad \text{s.t.} \quad x^T a^{(i)} \leq b_i \begin{cases} a^{(i)} \in [\underline{a^{(i)}}, \overline{a^{(i)}}] \\ b_i \in [\underline{b_i}, \overline{b_i}] \end{cases}, \quad i = 1, \dots m$$

$$x \in \mathbb{R}^n$$

$$\downarrow$$

$$\min \ c^T x \quad \text{s.t.} \quad x^T a_c^{(i)} + y^T a_\Delta^{(i)} \leq \underline{b_i}, \quad i = 1, \dots m$$

$$x \leq y$$

$$-x \leq y$$