Prep: bring ID, water, glasses, jacket, pen, watch and the handwritten version of this set of notes
You got this!

Theory

Convexity and Convex Sets

Convexity of a set $C \subset \mathbb{R}^n$:

- $\forall x, y \in C, \lambda x + (1-\lambda)y \in C \; \forall \lambda \in [0,1]$

<u>Typical convex sets</u>:

- Cone: $x \in C \Rightarrow \alpha x \in C \; \forall \alpha \geq 0$ (all rays)
- Linear hull: $L(\{x_1, \dots, x_n\}) = \{\sum_{i=1}^n \lambda_i x_i\} = \text{Span}(\{x_1, \dots, x_n\})$
- Affine Hull: $\text{aff}(\{x_1, \dots, x_n\}) = \{\sum_{i=1}^n \lambda_i x_i, \sum_{i=1}^n \lambda_i = 1\}$
  - Smallest affine set containing $\{x_1, \dots, x_n\}$
  - Does not necessarily contain 0
  - $\text{aff}(\text{aff}(S)) = \text{aff}(S)$
  - $\text{aff}(C)$ closed if C finite dimensional
  - $\text{aff}(S+T) = \text{aff}(S) + \text{aff}(T)$
  - $0 \in S \Rightarrow \text{aff}(S) = \text{Span}(S)$
- Convex hull: $\text{Co}(\{x_1, \dots, x_n\}) = \{\sum_{i=1}^n \lambda_i x_i, \sum_{i=1}^n \lambda_i = 1, \lambda_i \geq 0\}$
  - smallest convex set containing $\{x_1, \dots, x_n\}$
- Conic hull: $\text{Conic}(\{x_1, \dots, x_n\}) = \{\sum_{i=1}^n \lambda_i x_i, \lambda_i \geq 0\}$. It is the smallest convex cone.

<u>Convexity of function</u> $f: \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ with $\text{dom}(f) = \{x: |f(x)| < \infty\}$ (equivalence)

- $\forall x, y \in \text{dom}(f), \forall \lambda \in [0,1]$,
  $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$
- $epi(f) = \{(x,t) \in \mathbb{R}^{n+1} : f(x) \leq t\}$ is a convex set in $\mathbb{R}^{n+1}$
- $-f$ is concave
- If f differentiable, convex if and only if lower bounded by first order Taylor approximation
  $f(y) \geq f(x) + \nabla f(x)^T (y-x) \; \forall x, y \in \text{dom}(f)$
  $\langle \nabla f(x), x \rangle - f(x) \geq \langle \nabla f(x), y \rangle - f(y)$
- If f twice-differentiable, convex if and only if every local approximation is convex
  $\nabla^2 f(x) \succcurlyeq 0 \;\; \forall x \in \text{dom}(f)$
- Restriction of f to a line is still convex i.e. $f(x+tz)$ convex in t for $x + tz \in \text{dom}(f)$

<u>Properties of convex functions</u>:

- All norms are convex
- All dual norms are convex
- [Sublevel sets] If f convex, sublevel sets $S_\alpha = \{x | f(x) \leq \alpha\}$ are convex $\forall \alpha$

Convex Optimization

$$\min_x f_0(x)$$
s.t. $f_i(x) \leq 0, i \in \{1, \dots, m\}$; $h_j(x) = 0 \; j = \{1, \dots, p\}$
f_i convex and h_j affine

<u>Properties and Theorems</u>:

- Any locally optimal is globally optimal
- Feasible set convex; optimal set convex
- If objective function is strictly convex, then there is at most one optimal point
- [Supporting Hyperplane] If $C \subset \mathbb{R}^n$ convex, non-empty, then $\forall x_0$ on boundary of C, $\exists$ a supporting hyperplane to C at x_0 (i.e. $\exists a \in \mathbb{R}^n, a \neq 0, a^T(x - x_0) \leq 0 \;\; \forall x \in C$)
- [Projection] For a nonempty, closed convex set C and $x \in \mathbb{R}^n$, $\exists m \in C$ s.t. $\|m - x\| \leq \|c - x\| \;\; \forall c \in C$

<u>Optimality Conditions</u>:

- [Unconstrained] $\nabla f_0(x) = 0$
- [Constrained] If and only if $\forall y$ feasible, $\nabla f_0(x)^T (y - x) \geq 0$

Operations that Preserve Convexity

- [Intersection] $(C_\alpha)_{\alpha \in A} \Rightarrow \bigcap_{\alpha \in A} C_\alpha$ convex
  - Half-space convex $\Rightarrow$ polyhedron convex
  - Convex set is intersection of halfspaces
- [Affine Transformation] $f(x) = Ax + b$, $C \subset \mathbb{R}^n$ convex, then $f(C)$ convex.
  - Projections are affine
- [Supremum of Convex Functions]: $f_1, \dots, f_m$ convex, so is $f(x) = \sup_{1 \leq i \leq m} f_i(x)$
- [Composition with Affine Function]: If f convex, so is $g(x) = f(Ax + b)$
- [Nonnegative Linear Combination]: If f, g convex, so is $\alpha f(x) + \beta g(x)$ for $\alpha, \beta \geq 0$.

Lower Semi-Continuous Functions Theory

Definition: $f: \chi \to \mathbb{R} \cup \{+\infty\}$ is <u>lower semi-continuous</u> if for any convergent sequence $(x_n)_n$ s.t. $\lim_{n \to \infty} x_n = x$ in χ, $\lim_{n \to \infty} \inf f(x_n) \geq f(x)$

<u>Theorems and Claims</u>:

- $f: \chi \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous if and only if $epi(f)$ is a closed set
- [Convexity $\Rightarrow$ Max-affine] If $f: \chi \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and convex,

- [Jensen] $f: \mathbb{R}^n \to \mathbb{R}$ convex, $x_1, \dots, x_k \in \mathrm{dom}(f)$, $\theta_1, \dots, \theta_k \geq 0$ with $\sum_{i=1}^{k} \theta_i = 1$:
$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$$

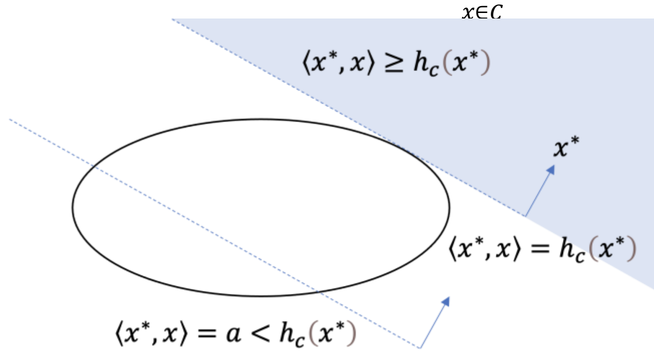
Operations preserving convexity:

- [Composition] $f \circ g$ is convex if f is convex, nondecreasing and g convex
- [Composition with affine] f convex $\Rightarrow$ $g(x) = f(Ax + b)$ convex, $A \in \mathbb{R}^{m \times n}$
- [Pointwise supremum] $f(x) = \max_{\alpha \in A} f_\alpha(x)$
- [Nonnegative linear combination] $f_1, \dots, f_n$ convex $\Rightarrow \lambda_1 f_1 + \cdots + \lambda_n f_n$ convex, $\lambda_i \geq 0$
- [Partial minimum] f convex in $x = (y, z) \Rightarrow g(y) = \min_z f(y, z)$ convex

then f equals supremum of all affine minorants i.e. $f(x) = \sup_{a \leq f, a: \chi \to \mathbb{R}} a(x) \ \forall x \in \chi$

Support Function Theory

Definition: For a set C, $h_C(x^*) = \sup_{x \in C} \langle x^*, x \rangle$



Properties:

- $h_C(x^*) \equiv I_C^*(x)$, where $I_C = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$
- Always convex regardless of C

Theorem:
Every closed convex set $C \subset X$ is an intersection of (possibly uncountably infinite) halfspaces defined by support functions
$$C = \bigcap_{x^* \in \chi^*} \{x : \langle x^*, x \rangle \leq h_c(x^*)\}$$

Conjugate (Fenchel) Duality

For $f: \chi \to \mathbb{R} \cup \{+\infty\}$ with $\mathrm{dom}\, f \neq \phi$, define convex conjugate $f^*: \chi^* \to \mathbb{R} \cup \{+\infty\}$
$$f^*(x^*) = \sup_{x \in \chi} \{\langle x^*, x \rangle - f(x)\}$$

Properties of f^*

- Pointwise maximum of affine function in x^*
- Convex and lower semi-continuous

Properties:

- [Fenchel's inequality]
$$\langle x^*, x \rangle \leq f(x) + f^*(x^*) \ \ \forall x \in \chi, x^* \in \chi^*$$
- [Order reversal]
$$f \leq g \Rightarrow g^* \leq f^* \Rightarrow f^{**} \leq g^{**}$$
- [Biconjugation]
$$f^{**}(x) = (f^*)^*(x) \coloneqq \sup_{x^* \in \chi^*} \{\langle x^*, x \rangle - f^*(x^*)\}$$
- [Weak Duality for Biconjugates] $f^{**} \leq f$
- [Fenchel-Moreau] Let $f: \chi \to \mathbb{R} \cup \{+\infty\}$, then f is convex and lower-semi-continuous $\Leftrightarrow f^{**} = f$
- [Convex lower-envelope] f^{**} is the pointwise largest convex lower semi-continuous function that lies below f

Applications on convex, differentiable functions:

Conjugate Table

$f(x)$	$f^*(x^*)$
$I_K(x) = \begin{cases} 0, & x \in K \\ \infty, & x \notin K \end{cases}$	$I_K^*(x^*) = h_K(x^*)$
$a(x) = \langle x_a^*, x \rangle + b$	$a^*(x^*) = h_\chi(x^* - x_a^*) - b = \begin{cases} \infty, & x^* \neq x_a^* \\ -b, & x^* = x_a^* \end{cases}$
$a^*(x^*) = \begin{cases} \infty, & x^* \neq x_a^* \\ -b, & x^* = x_a^* \end{cases}$	$a^{**}(x) = a(x) = \langle x_a^*, x \rangle + b$
$f(x) = \lVert x \rVert_2$	$f^*(x^*) = \begin{cases} 0, & \lVert x^* \rVert \leq 1 \\ \infty, & \lVert x^* \rVert > 1 \end{cases}$
$f(x) = \dfrac{\lvert x \rvert^p}{p}$	$f^*(x^*) = \dfrac{\lvert x^* \rvert^q}{q}$
$f(x) = \lvert x \rvert$	$f^*(x^*) = \begin{cases} 0, & \lvert x \rvert \leq 1 \\ \infty, & \lvert x \rvert > 1 \end{cases}$
$f(x) = \lVert x \rVert$	$f^*(x^*) = \begin{cases} 0, & \lVert x^* \rVert_* \leq 1 \\ \infty, & \text{otherwise} \end{cases}$
$f(x) = \sum_{i=1}^{n} x_i \log x_i$	$f^*(x^*) = \sum_{i=1}^{n} e^{x_i^* - 1}$
$f(X) = \log \det X^{-1}$, $\mathrm{dom}\, f = S_{++}^n$	$f^*(X^*) = \log \det(-X^*)^{-1} - n$, $\mathrm{dom}\, f = -S_{++}^n$
$f(x) = \log\left(\sum_{i=1}^{m} e^{x_i}\right)$	$f^*(x^*) = \begin{cases} \sum_{i=1}^{m} x_i^* \log x_i^*, & x^* \succcurlyeq 0,\ 1^T x^* = 1 \\ \infty, & \text{otherwise} \end{cases}$

- $f^*(\nabla f(x)) + f(x) = \langle \nabla f(x), x\rangle$
- If f strictly convex, twice differentiable, then $\nabla f^*(\nabla f(x)) = x$ i.e. $\nabla f^*: X^* \to X$ is the inverse of $\nabla f: X \to X^*$

$f(x) = \frac{1}{2}\|x\|^2$	$f^*(x^*) = \frac{1}{2}\|x^*\|_*^2$

General Duality Theory

Primal problem (P):

$$f: \chi \to \mathbb{R} \cup \{+\infty\} \quad \inf_{x\in\chi} f(x)$$

Define perturbation function $F: \chi \times \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ which satisfies $F(x,0) = f(x)$, and F^*

$$F^*: (\chi \times \mathcal{Y})^* = \chi^* \times \mathcal{Y}^* \to \mathbb{R} \cup \{+\infty\}$$

$$F^*(x^*, y^*) = \sup_{x\in\chi, y\in\mathcal{Y}} \{\langle x^*, x\rangle + \langle y^*, y\rangle - F(x,y)\}$$

Define value function as $V: \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ s.t. $V(y) = \inf_{x\in\chi} F(x,y)$ (i.e. most optimal value of (P) given perturbation by y)

Properties of V:

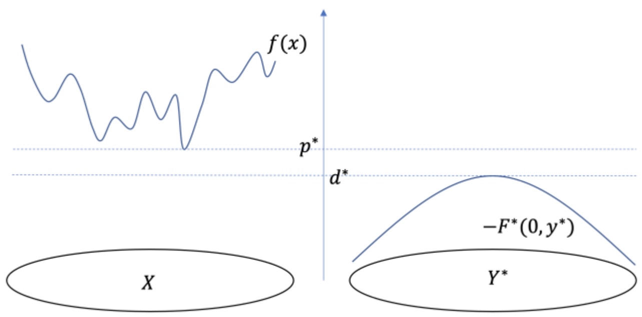
- $V(0) = \inf_{x\in\chi} F(x,0) = \inf_{x\in\chi} f(x) = p^*$
- $V^*(y^*) = F^*(0, y^*) = \sup_{x\in\chi, y\in\mathcal{Y}} \{\langle y^*, y\rangle - F(x,y)\}$
- $V^{**}(y) = \sup_{y^*\in\mathcal{Y}^*} \{\langle y^*, y\rangle - F^*(0, y^*)\}$
- [Weak Duality]

$$p^* = \inf_{x\in\chi} f(x) = V(0) \geq V^{**}(0) = \sup_{y^*\in\mathcal{Y}^*} \{-F^*(0,y^*)\} = d^*$$

- [Dual Problem (D)] Always concave in y^*

$$d^* = \sup_{y^*\in\mathcal{Y}^*} \{-F^*(0,y^*)\}$$

- [Dual Variable] y^*
- [Certificate] x_0, y_0^* s.t. $f(x_0) = -F^*(0, y_0^*)$, then they are optimal for (P) and (D)



Theorem:

- If $\exists x_0 \in \chi$ s.t. $f(x_0) < \infty$ and F convex lower semicontinuous, then strong duality holds by Fenchel-Moreau, i.e. $p^* = d^*$

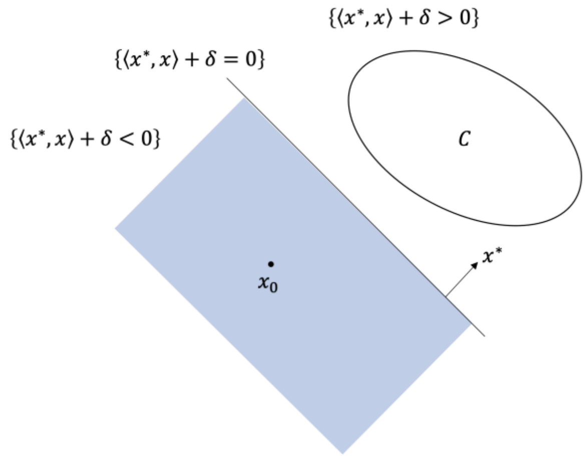
Examples of $F(x,y)$ (perturbation function):

$f_0(x)$	$F(x,y)$
$f_0(x) = \|x - x_0\|_2 + I_C(x)$	$F(x,y) = \|x - x_0\|_2 + I_C(x+y)$ $F^*(0,y) = \begin{cases} +\infty, & \|y^*\| > 1 \\ -\langle y^*, x_0\rangle + h_C(y^*), & \|y^*\| \leq 1 \end{cases}$

Geometrical Duality

[Basic Duality Theorem] Let $C \subset \chi$ be closed convex and $x_0 \in \chi\backslash C$. Then, $\exists$ nonzero $x^* \in \chi^*$ and $\delta > 0$ s.t. $\langle x^*, x_0\rangle + \delta < \langle x^*, x\rangle\ \forall x \in C$ i.e. the hyperplane $\{x: \langle x^*, x\rangle + \delta = 0\}$ separates x from the convex set C.

- Equivalently, $\exists$ nonzero $x^* \in \chi^*$ and $\delta > 0$ s.t. $\langle x^*, x_0\rangle + \delta < \inf_{x\in C}\langle x^*, x\rangle$
- Equivalently, $\exists \delta > 0$ and $x^* \in \chi^*$ s.t. $\sup_{x\in C}\langle x^*, x\rangle + \delta < \langle x^*, x_0\rangle$



[Corollary] Let C, D be closed convex sets and C compact. Then $\exists x^* \in \chi^*$ and $\delta > 0$ s.t. $\langle x^*, c\rangle \geq \langle x^*, d\rangle + \delta\ \forall c \in C, \forall d \in D$.

[Geometric Duality] Convex set C and $x_0 \in \mathbb{R}^m$,

$$\min_{x\in C}\|x - x_0\|_2 = \max_{H:\,H \text{ separates } x_0 \text{ from } C} d(x_0, H)$$

[Farkas' Lemma] Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then, **exactly** one of the following is true

- $\exists x \in \mathbb{R}^n \geq 0$ satisfying $Ax = b$
- $\exists y \in \mathbb{R}^m$ s.t. $A^T y \geq 0$ and $b^T y < 0$.
- [Certificate] If $\exists y$ s.t. $A^T y \geq 0$ and $b^T y < 0$, then $\nexists\ x > 0$ s.t. $Ax = b$.

[Separation theorem] If $C, D \subset \mathbb{R}^n$ convex, $C \cap D = \phi$, then $\exists$ hyperplane separating them, i.e. $\exists a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$ s.t. $a^T x \leq b$ for every $x \in C$ and $a^T x \geq b$ for every $x \in D$.

[Depth]: C closed convex, depth convex in x_0

$$\text{depth}(x_0, C) = \sup_{x^*\in X^*: \|x^*\|_2 = 1} \{\langle x^*, x_0\rangle - h_C(x^*)\}$$

[Projection Theorem]: Let $C \subset \chi$ convex, closed. $\forall x_0 \in \chi$, $\exists$ unique $\Pi_C(x_0) \in C$ i.e. $\|x_0 - \Pi_C(x_0)\|_2 \leq \|x - x_0\|_2 \; \forall x \in C$.

- $\langle \Pi_C(x_0) - x_0, \Pi_C(x_0) - x \rangle \leq 0 \; \forall x \in C$

Lagrangian Duality

Primal (P):

$$\min_x f_0(x) \text{ s.t. } f_i(x) \leq 0 \text{ for } i = 1, \dots, m$$

Lagrange Dual (D): $\lambda \in \mathbb{R}^m$ is the dual variable:

$\mathcal{L}(x,\lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$	
$g(\lambda) = \inf_{x \in X} \mathcal{L}(x,\lambda)$	$d^* = \sup_{\lambda \geq 0} g(\lambda)$

- Symmetric form of Primal $\min_x \max_{\lambda \geq 0} \mathcal{L}(x,\lambda)$
- $\max_x f_0(x)$ s.t. $f_i(x) \leq 0$ for $i = 1, \dots, m$

$$d^* = \inf_{\lambda \leq 0} \sup_{x \in X} \mathcal{L}(x,\lambda)$$

Properties:

- No longer any constraints on x
- g concave, upper-semi-continuous
- [Lower bound property] $g(\lambda) \leq p^* \; \forall \lambda \geq 0$

Theorem:

- [Weak Duality]

$$p^* = \inf_{x \in \chi: f_i(x) \leq 0 \forall i} f_0(x) \geq \sup_{\lambda \geq 0} \inf_{x \in \chi} \mathcal{L}(x,\lambda) = d^*$$

Karush-Kuhn-Tucker (KKT) Conditions

[Necessity] If all functions are differentiable, $x^*, (\lambda^*, \mu^*)$ primal, dual optimal and strong duality holds, then KKT conditions are satisfied:

- [Stationarity]

$$\nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_j \mu_j^* \nabla f_j(x^*) = 0$$

- [Feasibility] x^* primal feasible; (λ^*, μ^*) dual feasible
- [Complementary Slackness] $\lambda_i^* f_i(x_i^*) = 0$
  - If $\lambda_i^* > 0$, then $f_i(x_i^*) = 0$
  - If $f_i(x_i^*) < 0$, then $\lambda_i^* = 0$
  - $\lambda_i^* = 0$ unless f_i active at optimum

[Sufficiency] If f_i convex and differentiable and $x, (\lambda, \mu)$ satisfy KKT conditions, then:

- $x^*, (\lambda^*, \mu^*) = x, (\lambda, \mu)$ primal, dual optimal
- Strong duality holds

Perturbation and Sensitivity Analysis

Constraint Qualification

- [Convex LSC, Primal Feasibility] If f_i convex, lower semi-continuous and $\exists x$ feasible

- [Slater's Condition] Let $D = \bigcap_{i=0}^m \operatorname{dom} f_i$ i.e. $x \in D \Rightarrow f_i(x) < \infty$. If f_i convex and $\exists$ a **strictly** feasible $x_0 \in D$

- [Slater's Condition Weakened]
  - h_j affine: if $\exists$ strictly feasible point $\in \operatorname{relint}(D)$ i.e. $h_j(x) = 0, f_i(x) < 0$
  - Affine inequality constraints need not hold with strict inequality

- [KKT Sufficiency] f_i convex, differentiable and KKT conditions hold for some $(x, (\lambda, \mu))$

Lagrangian Duality Linear Programming (LP)

Primal (P)	Dual (D)
$\inf_{x \in \mathbb{R}^n} c^T x$ s.t. $Ax \leq b$	$\sup_{\lambda \geq 0} -b^T \lambda$ s.t. $A^T \lambda = -c$

Theorems:

- [Strong Duality for LP] If either (P) or (D) feasible, then strong duality holds.
- [HW9] If primal feasible and dual is not, then strong duality holds $p^* = d^* = -\infty$

Sion's Minimax Theorem

X compact, convex, Y convex. If $f: X \times Y \to \mathbb{R}$ with $f(x,\cdot)$ USC, concave on Y and $f(\cdot,y)$ LSC, convex on X, then:

$$\min_{x \in X} \max_{y \in Y} f(x,y) = \max_{y \in Y} \min_{x \in X} f(x,y)$$

[Lagrangian Min-Max]

$$p^* = \min_x \max_{\lambda \geq 0} L(x,\lambda)$$

$$p^* = \min_x \max_{\lambda \geq 0} L(x,\lambda) \geq \max_{\lambda \geq 0} \min_x L(x,\lambda) = d^*$$

- All constraints get shifted along with the interchanging of max and min

Fenchel-Rockafellar Duality Theorem

Perturbed problem: $\min_x f_0(x)$
subject to $f_i(x) \le u_i$, $h_j(x) = v_j$.
If strong duality holds and dual optimum (λ^*, μ^*) is achieved, then:

$$p^*(u,v) \ge p^*(0,0) - {\lambda^*}^T u - {\mu^*}^T v$$

- $\lambda_i^* \gg 1, u_i < 0 \Rightarrow p^*(u,v)$ increases greatly
- $|\mu_i^*| \gg 1, \operatorname{sign}(v_i) \neq \operatorname{sign}(\mu_i^*) \Rightarrow p^*(u,v)$ increases greatly
- $\lambda_i^* \ll 1, u_i > 0 \Rightarrow p^*(u,v)$ will not decrease too much
- $|\mu_i^*| \ll 1, \operatorname{sign}(v_i) = \operatorname{sign}(\mu_i^*) \Rightarrow p^*(u,v)$ will not decrease too much

λ^* gives a measure of sensitivity of (P) w.r.t. constraints. λ_i can be interpreted as how much you are willing to pay to relax f_i

Local Sensitivity Analysis:
Assume $p^*(u,v)$ differentiable at $u = 0, v = 0$.
If strong duality holds, symmetric relation:

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i},\ \mu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

- $f_i(x^*) < 0 \Rightarrow$ constraint inactive i.e. can be tightened or loosened with no effect on $p^* \Rightarrow \lambda_i^* = 0$
- $f_i(x^*) = 0 \Rightarrow$ constraint active i.e. sensitive to perturbation (no slackness)

Toolkit

- [Young] $ab \le \frac{a^p}{p} + \frac{b^q}{q}$ where p, q are Holder's conjugate
- [Jensen] $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$
- [Hölder] $\sum_{i=1}^n |a_i b_i| \le (\sum_{i=1}^n |a_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |b_i|^q)^{\frac{1}{q}}$
- [Hölder] Equality when $|b_i| = c|a_i|^{p-1}$
- [Hölder] $\sum_{i=1}^n |x_i|^\theta |y_i|^{1-\theta} \le (\sum_{i=1}^n x_i)^\theta (\sum_{i=1}^n y_i)^{1-\theta}$
- [Taylor] $f(x+\delta) = f(x) + \left(\nabla f(x)\right)^T \delta$
- [Taylor] $f(x+\delta) = f(x) + \left(\nabla f(x)\right)^T \delta + \frac{1}{2}\delta^T (\nabla^2 f(x)) \delta$

$\nabla_X(-\log\det X) = -(X^{-1})^T$	$\nabla_X(a^T X b) = ab^T$
$\nabla_X\left(\operatorname{tr}(AX)\right) = A^T$	$\nabla_X\left(\operatorname{tr}(AX^T)\right) = A$
$\nabla_X\left(\operatorname{tr}(B^T X^T A^T A X B)\right) = A^T A X B B^T$	
$\nabla_X \log\det X = X^{-1}$	$-\log\det X$ convex
$f(x) = -\log\det x + I_{S_{++}^n}(x)$	$f^*(x^*) = \begin{cases} -n - \log\det(-x^*), & x^* \in S_{--}^n \\ \infty, & \text{else} \end{cases}$

- [Dual Norm] $\|z\|_* = \sup_{x:\|x\|\le 1} z^T x = \sup_{x:\|x\|\le 1} |z^T x|$
- [Dual Norm] $\langle z, x\rangle \le \|x\| \|z\|_*$
- $\lambda_{\max}(X) \le t$ is equivalent to $tI - X \in S_+^n$

Perturbation: $F(x,y) = f(x) + g(Ax - y)$

Theorem: Let $f: X \to \mathbb{R} \cup \{+\infty\}$ and $g: Y \to \mathbb{R} \cup \{+\infty\}$ and $A: X \to Y$ be a linear map. Then:

$$\inf_{x\in X}\{f(x) + g(Ax)\} \ge \sup_{y^*\in Y^*}\{-f^*(A^T y^*) - g^*(-y^*)\}$$

If f, g convex and $\exists x_0 \in \operatorname{dom} f \cap \operatorname{dom}(g \circ A)$ s.t. g continuous at Ax_0, then equality holds and the supremum is attained by some $y^* \in Y^*$

Last Resorts and Final Checks

- Instinct (i.e. just set derivative to 0)
- Write down the Lagrangian in proper form
- Conjugate method if inequalities affine
- Component-wise analysis: isolate terms
- Get **ALL** KKT conditions for structure
- Case by case consideration
- Matrix Form

Remember:
- KKT conditions need $*$ (i.e. x^*, λ^*, μ^*)
- Don't be scared of taking derivatives of matrices
- Focus: which variable can vary?
- Did you forget any constraints like $x \ge 0$?
- Did you forget to leave in standard form?
- Try Sion's minimax form; leave $w \ge 0$ in the conditions of minimax.

Problem Solving Techniques:
1. Introduce slack variables
2. Introduce new variables and equality constraints (if affine, use conjugate)
3. Transforming the objective
4. Implicit constraints

Mind Game:
- Clean piece of paper, calm down…
- Slowly, trust the process. At every step, check for correctness. Derive analytic solution if need be.
- Declutter your variables, group similar terms, introduce $\tilde{x}$ if need be.
- Where there is a will, there is a way
- Focus!

Algorithms

Phase I: Finding Feasible Point

$\min_{x,s} s$ subject to $f_i(x) \le s$	$x_0 \in \bigcap_{i=1}^m \operatorname{dom} f_i$ $s_0 = 1 + \max_i f_i(x_0)$

- (x_0, s_0) strictly feasible
- With (x_0, s_0) as star point, obtain (x^*, s^*).
- If initial problem strictly feasible, $s^* < 0 \Rightarrow x^*$ strictly feasible for initial problem

Interior Point Method

Assumptions: f_0, f_i convex and strict feasibility

Heuristic: Unconstrained convex optimization problem $\mathrm{P}(t), t > 0$ with log-barrier $\phi(z)$

$$\min_x f_0(x) + t\sum_{i=1}^m \phi(-f_i(x))$$

$$\phi(z) = \begin{cases} +\infty, & z \le 0 \\ \log\frac{1}{z}, & z > 0 \end{cases}$$

- Let x_t^* be the solution to $\mathrm{P}(t)$.
- Strong duality holds by Slater's Condition
- First order KKT conditions give:

$$\nabla f_0(x_t^*) + \sum_{i=1}^m \frac{t}{-f_i(x_t^*)}\nabla f_i(x_t^*) = 0$$

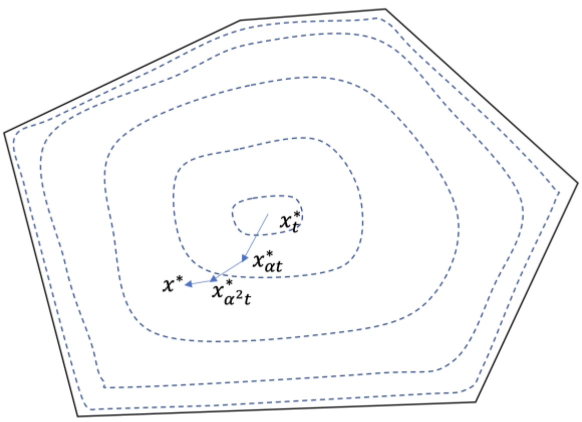
- $\lambda(t)_i \coloneqq \frac{t}{-f_i(x_t^*)} > 0$; hence λ dual feasible

$$p^* = d^* = \sup_{\lambda \ge 0} g(\lambda) \ge g(\lambda(t)) = L(x_t^*, \lambda(t)) = f_0(x_t^*) - mt$$

$$f_0(x_t^*) \le p^* + mt;\ \text{duality gap} \le f_0(x_t^*) - g(\lambda(t)) = mt$$

- Pick t s.t. $mt < \epsilon$ (m = # of conditions)
- Returns feasible solution to initial problem within tolerance mt: $f_0(x_t^*)$ mt suboptimal

Upshot: Given initial feasible point, can get arbitrarily close to optimal point by controlling t



Pseudocode:

- Given strictly feasible x_0, $t = t_0$, $\alpha < 1$

Phase II: Unconstrained Optimization

Problem #1: Step Size s

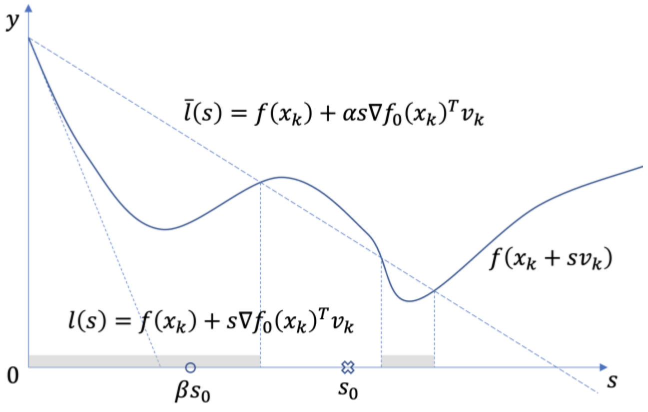
1. Constant step size $s = s_0$
2. Bisection $O\left(\log\frac{1}{\epsilon}\right)$ i.e. exploit monotone f'
   - Assume $x^* \in [L, U]$. Else, double size of interval till $f'(L) < 0, f'(U) > 0$
   - Set $x = \frac{1}{2}(L + U)$
   - If $f'(x) > 0$, $U \leftarrow x$. Else, $L \leftarrow x$.
   - Repeat until $|f'(x)|(U - L) \le \epsilon$

$$p^* = f(x^*) \ge f(x) + f'(x)(x - x^*) \ge f(x) - \epsilon$$

3. Bisection in $\mathbb{R}^n$ (common subroutine)
   - Start at x_0. Choose v.
   - Reduces to 1D optimization on slice

$$\alpha^* = \arg\min_{\alpha \ge 0} f_0(x_0 + \alpha v_0)$$
$$x_{k+1} = x_k + \alpha^* v_k$$

4. Backtracking Line Search
   - Key idea: No need to go to exact minimum along each 1D slice; only move if there is enough decrease, else lower expectation
   - Parameters $\alpha, \beta \in (0,1)$, $x_k, v_k, s_0 = 1$ s.t. $\delta = \nabla f_0(x_k)^T v_k \le 0$ (i.e. direction of ↓)
   - If $f_0(x_k + s v_k) \le f_0(x_k) + s\alpha \nabla f_0(x_k)^T v_k$, then $x_{k+1} \leftarrow x_k + s v_k$, $s \leftarrow s_{\text{init}}$
   - Else, decrease $s \leftarrow \beta s$. Repeat



Problem #2: Direction v

5. Gradient Descent $v_k = -\nabla f_0(x_k)$

$$x_{k+1} = x_k - \alpha^* \nabla f_0(x_k)$$

6. Stochastic Gradient Descent
   - Key idea: high cost of evaluating entire gradient; take a sample $|S| < m$ instead

$\min_w \frac{1}{m}\sum_{i=1}^m L(x_i^T w)$	$\nabla f_0(w) \approx \frac{1}{\lvert S\rvert}\sum_{i \in S} L'(x_i^T w) x_i$

- Solve $\mathrm{P}(t)$ to get $\left(x_t^*, \lambda(t)\right)$
- Update $x_0 \leftarrow x_t^*$, $t \leftarrow \alpha t$
- Repeat until $mt < \epsilon$ (intended accuracy)

Remark: Interior point still works without convex assumption, but not guaranteed 0 duality gap

Simplex Algorithm (Specific to LP)

- Starts at a vertex v
- Greedily chooses a feasible neighboring vertex with more optimal value
- If unable to choose, terminate and declare solved.

$$v_k = -\frac{1}{|S|}\sum_{i\in S} L'(x_i^T w)x_i \approx \nabla f_0(w)$$

7. Coordinate Descent
8. Newton's Method $O\left(\log\log\frac{1}{\epsilon}\right)$
   - Key idea: approximate convex function as a quadratic function locally; travel to the minimizer of quadratic function

$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$	$\left(\nabla^2 f_0(x_k)\right)^{-1}\nabla f(x_k)$ Newton step
$x_{k+1} = x_k - \left(\nabla^2 f_0(x_k)\right)^{-1}\nabla f(x_k)$	

9. Damped Newton's Method $O\left(\log\log\frac{1}{\epsilon}\right)$

$$x_{k+1} = x_k - s_k\left(\nabla^2 f_0(x_k)\right)^{-1}\nabla f_0(x_k)$$

where s_k is chosen by another method like backtracking line search

Applications

Entropy Maximization

Goal: Maximize entropy $\mathbb{H}[p] = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$ subject to constraints

$$\min_x \sum_{i=1}^{n} x_i \log x_i$$

subject to $\mathbb{1}^T x = 1, Ax \leq b$
Note: $x \geq 0$ included in $Ax \leq b$

Dual Problem (using conjugate method):

$$\max_{\lambda \geq 0, \mu} g(\lambda, \mu) = \max_{\lambda \geq 0, \mu} = -b^T\lambda - \mu - e^{-\mu-1} \sum_{i=1}^{n} e^{-a_i^T \lambda}$$

$$\mu = \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right) - 1$$

$$\max_{\lambda \geq 0} -b^T\lambda - \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right)$$

Risk Parity Portfolio

Goal: Find $x \in \mathbb{R}_+^n$, where x_i is amount of money invested in asset i, s.t. risk is distributed equally among all assets $x_i(Cx)_i = \frac{1}{n} x^T C x$. Note: $C = C^T > 0$ is covariance of the assets, measures risk.

- $x_i(Cx)_i$ is the contribution to risk by holding asset i.

Consider the following different convex optimization problem:

$$\min_x f_0(x) + \frac{1}{2} x^T C x$$

$$f_0(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i, & x_i > 0 \ \forall i \\ +\infty, & \text{otherwise} \end{cases}$$

Solutions (KKT)

$$-\frac{1}{x_i^*} + (Cx^*)_i - \lambda_i^* = 0$$
$$\lambda_i^*(-x_i^*) = 0$$
$$x_i^* > 0, \ \lambda_i^* \geq 0$$

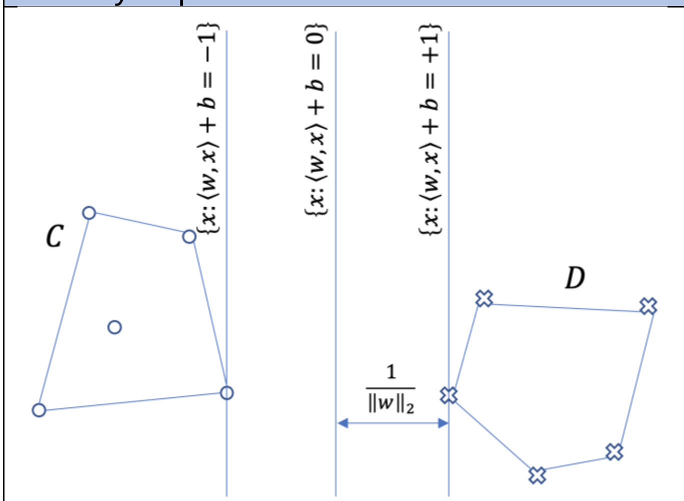
$$\lambda_i^* = 0$$
$$Cx^* = \begin{bmatrix} \frac{1}{x_1^*} & \cdots & \frac{1}{x_n^*} \end{bmatrix}^T$$

Since $x_i^*(Cx^*)_i = 1$, solution is the risk parity portfolio that we are looking for.

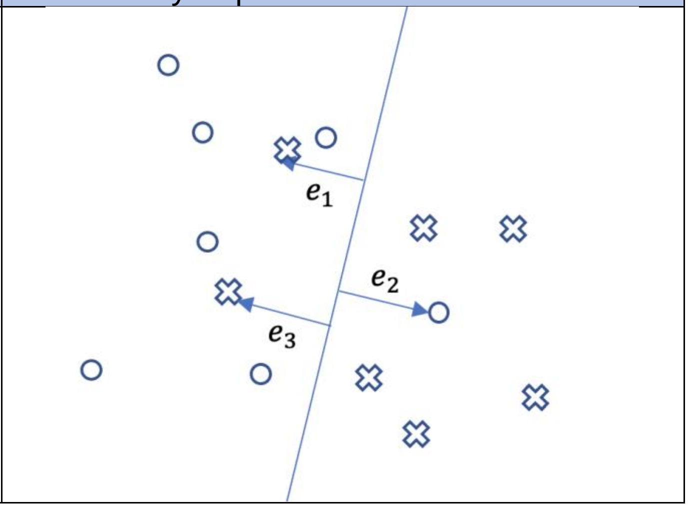
Support Vector Machine (SVM)

Given data points $x_1, \dots, x_m \in \mathbb{R}^n$ and labels $y = (y_1, \dots, y_m) \in \{0,1\}$
<u>Goal</u>: Find hyperplane of maximum margin

Linearly Separable Data



Non-linearly Separable Data



<u>Maximum Margin SVM</u> (Linearly Separable) Note: Quadratic Program

$\min_{w,b} \|w\|_2^2$	subject to $y_i(\langle w_i, x_i\rangle + b) \geq 1$

Non-linearly Separable: Key idea is to introduce slack variables e_i

$\min_{e,w,b} \|w\|_2^2 + \lambda \sum_{i=1}^{m} e_1$	subject to $y_i(\langle w_i, x_i\rangle + b) \geq 1 - e_i$ $e \geq 0$

Analysis:

1. $\lambda \gg 1$: can almost ignore w
    - Will find hyperplane that separates the greatest number of data points perfectly
    - Similar to L_1 norm; encourages sparsity among e_i
2. $0 < \lambda \ll 1$: increases the importance of margin compared to errors
3. λ is a tradeoff between margin (robustness) and classification error.

Variants:

1. Worst Case Loss

$\min_{e,w,b} \max_i e_i$	subject to $e_i \geq \max\big(0, 1 - y_i(\langle w_i, x_i\rangle + b_i)\big)$
$\min_{e,w,b} \max_i \big(0, 1 - y_i(\langle w_i, x_i\rangle + b_i)\big)$	Hinge Loss

- Does not care about margin w at all; just wants a hyperplane that minimizes worst case

2. Robust SVM (SOCP)

- Know $x_i \in B_{r_i}(\hat{x}_i)$ i.e. $\|\hat{x}_i - x_i\|_2 \leq r_i$

$\min_{w,b} \|w\|_2$	subject to $y_i(\langle w, x_i\rangle + b) \geq 1 \;\; \forall x_i \in B_{r_i}(\hat{x}_i)$
$\min_{w,b} \|w\|_2$	subject to $r_i\|w\|_2 + 1 \leq w^T\hat{x}_i + b \;\; (y = +1)$ subject to $r_i\|w\|_2 + 1 \leq -w^T\hat{x}_i - b \;\; (y = -1)$

3. Nonlinear Data

- Reparametrize in terms of new parameters like x_1^2, x_1x_2, x_2^2 (can model circular data)

Supervised Learning

$$\min_{w \in \mathbb{R}^n} L(X^T w, y) + \lambda \cdot p(w)$$

- X: data matrix $X = [x_1 \;\; \cdots \;\; x_m] \in \mathbb{R}^{n \times m}$
- $y = (y_1, \dots, y_m) \in \mathbb{R}^m$ labels
- w: weights; gives prediction rule for new data
- L: loss function; **convex** in first argument
- $\lambda \geq 0$: parameter for regularization
- p: **convex** penalty function; independent of data; reflects prior knowledge

Loss Function	Paradigm
$L(z,y) = \|z - y\|_2$	Linear least-squares regression Assume Gaussian noise
$L(z,y) = \|z - y\|_1$	Disregard outliers
$L(z,y) = \|z - y\|_\infty$	Robust regression
$L(z,y) = \sum_{i=1}^{m} \max(0, 1 - y_i z_i)$	Hinge loss; useful in SVM
$L(z,y) = -\sum_{i=1}^{m} \log(1 + e^{-y_i z_i})$	Logistical loss
$L(z,y) = \|z - y\|_2$ $p(w) = \|w\|_1$	LASSO Encourages sparsity
$p(w) = \|w - x_0\|_2$	Regularization: believes w to be close to x_0
λ	Parameter tuning

Network Economics Problem

R	Set of **routes** (not edges!)
J	Set of resources/edges
S	Set of source/sink pairs
U_s	Utility function of $s \in S$, increasing, strictly concave differentiable (LDMR)

System Problem

$\max_{x,y} \sum U_s(x_s)$	$Hy = x$	Valid flow pattern
	$Ay \leq c$	Capacity constraints
	$x, y \geq 0$	Nonnegative flow

Strong duality holds; primal, dual optimal both attained.

C_j	Capacity of $j \in J$
A_{jr}	$= \begin{cases} 1, & r \in R \text{ uses } j \in J \\ 0, & \text{otherwise} \end{cases}$
H_{jr}	$= \begin{cases} 1, & r \in R \text{ serves } s \in S \\ 0, & \text{otherwise} \end{cases}$
y	Assignment of flow in a network along the **routes**
x	Amount of flow from source to sink s

User Problem (maximize utility)

$\text{User}_s(\lambda) = \max_{x_s \geq 0} U_s(x_s) - \lambda_s x_s$	λ_s: cost per flow

Network Problem (maximize profit)

$\text{Network} = \max_{x_s \geq 0} \lambda_s x_s$	$Hy = x$
	$Ay \leq c$
	$x, y \geq 0$

Theorem: There is an equilibrium price vector λ s.t. x^* in both problems are the same and optimal for the system.

$$L(x,y,z) = \sum_{s \in S} U_s(x_s) - \lambda_s x_s + \sum_{r \in R} y_r \left(\lambda_s(r) - \sum \mu_j\right) + \sum_{j \in J} \mu_j (c_j - z_j)$$

λ^* is precisely this magical price vector, justified by KKT.

Network Optimization Problem

- At advertised prices λ, users signal willingness to pay m
- Network solves $\text{Network}(m)$
- Network updates prices $\lambda_s = \frac{m_s}{x_s}$
- Repeating this algorithm converges to equilibrium λ^*

User Problem (maximize utility)

$\text{User}_s(\lambda) = \max_{m_s \geq 0} U_s\left(\frac{m_s}{\lambda_s}\right) - m_s$	m_s: willingness to pay or budget

Network Problem (maximize profit)

$\text{Network}(m) = \max_{x,y} \sum_s m_s \log x_s$	$Hy = x$
	$Ay \leq c$
m: budget of users; can't control	$x, y \geq 0$

Portfolio Optimization

$p^* = \max_{w \geq 0; \mathbb{1}^T w = 1} \hat{r}^T w - \frac{1}{2} w^T D w$	subject to $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$, $\mu > 0$

$$p^* = \max_{w \geq 0} \min_{\mu} \hat{r}^T w - \frac{1}{2} w^T D w + \mu(\mathbb{1}^T w - 1) = \min_{\mu} \max_{w \geq 0} \hat{r}^T w - \frac{1}{2} w^T D w + \mu(\mathbb{1}^T w - 1)$$

Optimization of Norms (Example of Sion's Minimax Application)

$\min_x \lVert Ax - y \rVert_1 + \mu \lVert x \rVert_2$	subject to $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$, $\mu > 0$

Key idea: $\lVert z \rVert_2 = \max_{u: \lVert u \rVert_2 \leq 1} u^T z$, $\lVert z \rVert_1 = \max_{u: \lVert u \rVert_\infty \leq 1} u^T z$

$$p^* = \min_x \max_{\substack{\lVert u \rVert_\infty \leq 1 \\ \lVert v \rVert_2 \leq 1}} u^T(Ax - y) + \mu v^T x = \max_{\substack{\lVert u \rVert_\infty \leq 1 \\ \lVert v \rVert_2 \leq 1}} \min_x u^T(Ax - y) + \mu v^T x = \max_{\substack{\lVert u \rVert_\infty \leq 1 \\ \lVert v \rVert_2 \leq 1 \\ A^T u + \mu v = 0}} -u^T y = d^*$$

Distributed Systems

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n f_i(x_i)$$

subject to $a^T x = b$

f_i: utilities of different users, subject to a resource constraint

$$p^* = \max_{\mu \in \mathbb{R}} g(\mu) = \max_{\mu \in \mathbb{R}} \inf_{x \in \mathbb{R}^n} \sum_{i=1}^n f_i(x_i) - \mu(a^T x - b) = \max_{\mu \in \mathbb{R}} \mu b - \sum_{i=1}^n \max_{x_i} \mu a_i x_i - f_i(x_i)$$

$$= \max_{\mu \in \mathbb{R}} \mu b - \sum_{i=1}^n f_i^*(\mu a_i)$$

Remark: reduces to 1D problem in the dual

Dual of SOCP

$$p^* = \min_{x \in \mathbb{R}^n} c^T x$$

subject to $\lVert Ax + b \rVert_2 \leq c^T x + d$

$$\max_{u, \lambda:\, \lVert u \rVert_2 \leq \lambda} u^T y - t\lambda = \max_{\lambda \geq 0} \lambda(\lVert y \rVert_2 - t)$$

Minimum Volume Covering Ellipsoid

<table>
<tr><td>$\min_X \log\det X^{-1}$
s.t. $a_i^T X a_i \leq 1$ for $i = 1, \dots, m$
$\epsilon_X = \{z \mid z^T X z \leq 1\}, X \in S_{++}^n$</td><td>Min volume of ellipse centered at origin containing $a_1, \dots, a_m$
$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbb{1}^T\lambda + n, & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty, & \text{otherwise} \end{cases}$$
(by conjugate method) Strong duality always obtained</td></tr>
<tr><td colspan="2">Introducing New Variables and Equality Constraints Technique</td></tr>
<tr><td>$\min_x f_0(Ax + b)$
$\min_{x,y} f_0(y)$
subject to $Ax + b = y$</td><td>$g(\mu) = b^T\mu + \inf_y \{f_0(y) - \mu^T y\} = b^T\mu - f_0^*(\mu)$
$\max_\mu b^T\mu - f_0^*(\mu)$
subject to $A^T\mu = 0$</td></tr>
<tr><td colspan="2">Unconstrained Geometric Program</td></tr>
<tr><td>$\min_x \log\left(\sum_{i=1}^m e^{a_i^T x + b_i}\right)$
$\min_{x,y} \log\left(\sum_{i=1}^m e^{y_i}\right)$
subject to $Ax + b = y$</td><td>$\max_\mu b^T\mu - \sum_{i=1}^m \mu_i \log\mu_i$
subject to $1^T\mu = 1$
$A^T\mu = 0$
$\mu \geq 0$
(Entropy Maximization Problem)</td></tr>
<tr><td colspan="2">Norm Approximation Problem</td></tr>
<tr><td>$\min_x \lVert Ax - b \rVert$
$\min_x \lVert y \rVert$
subject to $Ax - b = y$</td><td>$\max_\mu b^T\mu$
subject to $\lVert\mu\rVert_* \leq 1, A^T\mu = 0$</td></tr>
</table>