Convex Optimization

Prof. Nati Srebro

Lecture 10:

Generalized Inequalities

Multiple Objectives

Reading: Boyd and Vandenberghe Sections 2.4,2.6,3.6,4.6—4.7,5.9

Matrix Inequalities (Semi Definite Constraints)

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq 0 \\ & h_i(x) = 0 \end{array}}$$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ convex

$f_i : \mathbb{R}^n \to S^{k_i} \subset \mathbb{R}^{k_i \times k_i}$ **matrix-convex**:

$$f_i(\theta x + (1-\theta)x') \preccurlyeq \theta f_i(x) + (1-\theta) f_i(x')$$

$h_i : \mathbb{R}^n \to \mathbb{R}$ linear (or $\to S^k$, doesn't matter much)

Sometimes convenient to represent $x \in S^n$: doesn't matter much, its just a vector space of dimensionality $n(n+1)/2$

Semi-Definite Programming (SDP): f_0, f_i **linear**

Duality with Matrix Inequalities

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq 0 \quad \lambda_i \in S^{k_i} \\ & h_i(x) = 0 \quad \nu_i \in \mathbb{R} \end{array}$$

$$\begin{array}{ll} \max\limits_{\lambda_i \in S^{k_i}, \nu \in \mathbb{R}^p} & g(\lambda, \nu) \\ s.t. & \lambda_i \succcurlyeq 0 \end{array}$$

$f_0 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex, $f_i : \mathbb{R}^n \to S^{k_i}$ matrix convex, $h_j : \mathbb{R}^n \to \mathbb{R}$ linear

$$L\big(x, (\lambda, \nu)\big) = f_0(x) + \sum_{i=1}^{m} \langle \lambda_i, f_i\,(x) \rangle + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$p^* = \inf_x \sup_{\lambda_i \succcurlyeq 0, \nu} L\big(x, (\lambda, \nu)\big) \geq \sup_{\lambda_i \succcurlyeq 0, \nu} \inf_x L\big(x, (\lambda, \nu)\big) = d^*$$

$$g(\lambda, \nu) = \inf_x L\big(x, (\lambda, \nu)\big)$$

Slater ($\exists$ feasible x s.t. $f_i(x) \prec 0$) ➔ Strong duality

Optimality Condition (assuming Strong Duality)

x^* optimal for (P)

λ^*, ν^* optimal for (D)

KKT Conditions

$f_i(x^*) \preccurlyeq 0 \; \forall_{i=1\ldots m}$

$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$

$\lambda_i^* \succcurlyeq 0 \; \forall_{i=1..m}$

$\nabla_x L\big(x^*, (\lambda^*, \nu^*)\big) = 0$

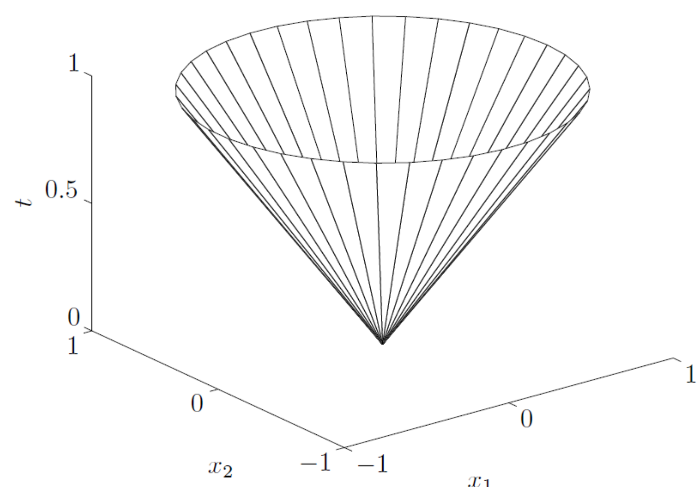
or $0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$

$\langle \lambda_i^*, f_i\,(x^*)\rangle = 0 \; \forall_{i=1..m}$

$\boxed{x^* \text{ optimal for (P)} \Leftrightarrow \exists_{\lambda^*,\nu^*} \text{ s.t. } x^*, \lambda^*, \nu^* \text{ satisfy KKT}}$

Generalized Inequalities

- More generally: $f_i: \mathbb{R}^n \to \mathbb{R}^{k_i}$ with the constraint $-f_i(x) \in K_i$ where $K_i \subset \mathbb{R}^{k_i}$ is a **closed pointed convex cone** with non-empty interior (“proper cone”)
  - Cone: $x \in K, \theta > 0 \to \theta x \in K$
  - Pointed: $x \in K, x \neq 0 \to -x \notin K$
- Examples:
  - $K = \mathbb{R}_+$: scalar non-negativity constraints
  - $K = \mathbb{R}_+^k$ (positive orthant): elementwise non-negativity
  - $K = S_+^k = \{X \in S^k | X \succcurlyeq 0\}$: semi-definite constraint
  - $K = \{\ [x\ t], x \in \mathbb{R}^{k-1}, t \in \mathbb{R}\ |\ \|x\| \leq t\}$ (norm cone)

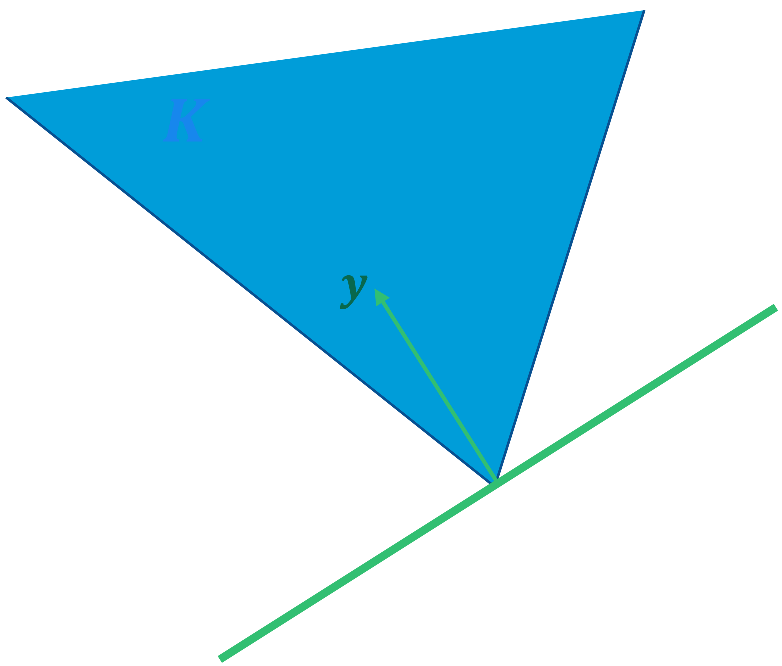


Generalized Inequalities

- More generally: $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ with the constraint $-f_i(x) \in K_i$ where $K_i \subset \mathbb{R}^{k_i}$ is a **closed pointed convex cone** with non-empty interior ("proper cone")
  - Cone: $x \in K, \theta > 0 \to \theta x \in K$
  - Pointed: $x \in K, x \neq 0 \to -x \notin K$
- Examples:
  - $K = \mathbb{R}_+$: scalar non-negativity constraints
  - $K = \mathbb{R}_+^k$ (positive orthant): elementwise non-negativity
  - $K = S_+^k = \{X \in S^k | X \succcurlyeq 0\}$: semi-definite constraint
  - $K = \{\, [x\ t], x \in \mathbb{R}^{k-1}, t \in \mathbb{R} \mid \|x\| \leq t\}$ (norm cone)
- Notation: $x \preccurlyeq_K y$ means $y - x \in K$
  - Constraint: $f_i(x) \preccurlyeq_K 0$
- K-convexity: $f(\theta x + (1-\theta)x') \preccurlyeq_K \theta f(x) + (1-\theta) f(x')$

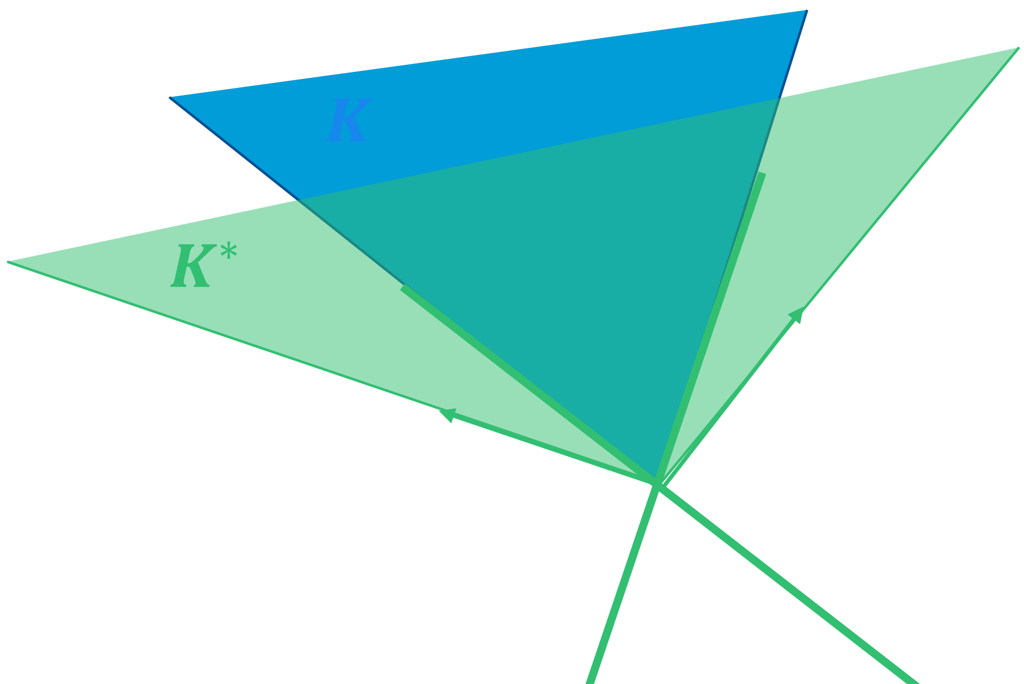
Dual Cone

$$K^* = \left\{ y \in \left(\mathrm{R}^k\right)^* \mid \forall_{(x \in K)} \langle y, x \rangle \geq 0 \right\}$$



Dual Cone

$$K^* = \left\{ y \in \left(\mathrm{R}^k\right)^* \mid \forall_{(x \in K)} \langle y, x \rangle \geq 0 \right\}$$



Dual Cone

$$K^* = \left\{ y \in \left(\mathrm{R}^k \right)^* \mid \forall_{(x \in K)} \langle y, x \rangle \geq 0 \right\}$$

- $\left(\mathbb{R}_+^k\right)^* = \mathbb{R}_+^k$
- $\left(S_+^k\right)^* = S_+^k$
- $\{\ [x\ t], x \in \mathbb{R}^{k-1}, t \in \mathbb{R} \mid \|x\| \leq t\}^*$
  $= \{\ [y\ s], y \in \mathbb{R}^{k-1}, s \in \mathbb{R} \mid \|y\|_* \leq s\}$

Generalized Inequality Constraints

$$\boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

$f_0: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex

$f_i: \mathbb{R}^n \to \mathbb{R}^{k_i}$ K_i-convex

$h_j: \mathbb{R}^n \to \mathbb{R}$ linear

Can always combined to single conic constraint:

$$\tilde{f}(x) = [f_i(x); f_2(x); \dots; f_m(x)] \preccurlyeq_K 0$$

$$\tilde{f}(x): R^n \to R^k \quad k = \sum_i k_i$$

$$K = K_1 \times K_2 \times \cdots \times K_m$$

Cone Programs: Generalized Linear Inequalities

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \preccurlyeq_K h \\ & Ax = b \end{aligned}$$

Inequality form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & \tilde{G}x \preccurlyeq_{\tilde{K}} \tilde{h} \end{aligned}$$

$$\begin{aligned} Gx &\preccurlyeq_K h \\ Ax &\preccurlyeq_K b \\ (-A)x &\preccurlyeq_K -b \end{aligned}$$

Standard form:

$$\begin{aligned} \min_{z \in \mathbb{R}^{\tilde{n}}} \quad & c^\top z \\ s.t. \quad & z \succcurlyeq_{\tilde{K}} 0 \\ & \tilde{A}z = \tilde{b} \end{aligned}$$

$$\begin{aligned} z &= [x_+; x_-;\ t] \\ t &= h - G(x_+ - x_-) \\ A(x_+ - x_-) &= b \end{aligned}$$

Linear Programming (LP)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \le h \\ & Ax = b \end{aligned}$$

Quadratic Programming (QP)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & x^\top P x + q^\top x \\ s.t. \quad & Gx \le h \\ & Ax = b \end{aligned}$$

Second Order Cone Programming (SOCP)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & \|G_i x - h_i\|_2 \le q_i^\top x + r_i \\ & \qquad\qquad\qquad i = 1..m \\ & Ax = b \end{aligned}$$

Quadratically Constrained Quadratic Programming (QCQP)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & x^\top P_0 x + q_0^\top x \\ s.t. \quad & x^\top P_i x + q_i^\top x \le r_i \\ & \qquad\qquad\qquad i = 1..m \\ & Ax = b \end{aligned}$$

Semidefinite Programming (SDP)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ s.t. \quad & Gx \preccurlyeq h \\ & Ax = b \end{aligned}$$

Dual Cones and Generalized Inequalities

$$K^* = \left\{y \in \left(\mathrm{R}^k\right)^* \mid \forall_{x \in K} \langle y, x \rangle \geq 0\right\}$$

- For a proper cone (pointed cone with non-empty interior): $K^{**} = K$, i.e.

$$K^{**} = \left\{x \in R^k \mid \forall_{y \in K^*} \langle y, x \rangle \geq 0\right\} = K$$

- Claim:

$$f_i(x) \preccurlyeq_K 0 \Leftrightarrow \overbrace{\sup_{\lambda \succcurlyeq_{K^*} 0} \langle \lambda, f_i(x) \rangle}^{\text{must be 0 or } \infty} < \infty$$

- Proof of $\Rightarrow$: $\langle \lambda, f_i(x) \rangle = -\langle \overset{K^* \ni}{\lambda}, \overset{K \ni}{-f_i(x)} \rangle \leq 0$
- Proof of $\Leftarrow$: $\forall_{\lambda \in K^*} \langle \lambda, -f_i(x) \rangle \geq 0 \rightarrow -f_i(x) \in K^{**} = K$

Duality with Generalized Inequalities

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_0(x) \\ s.t. \quad & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{aligned}$$

$$\begin{aligned} \max_{\lambda_i\in\mathbb{R}^{k_i},\nu\in\mathbb{R}^p} \quad & g(\lambda,\nu) \\ s.t. \quad & \lambda_i \succcurlyeq_{K_i^*} 0 \end{aligned}$$

$f_0: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex, $f_i: \mathbb{R}^n \to \mathbb{R}^{k_i}$ K_i-convex, $h_j: \mathbb{R}^n \to \mathbb{R}$ linear

- Can write primal as:

$$p^* = \inf_x \sup_{\lambda_i \succcurlyeq_{K_i^*} 0,\nu} f_0(x) + \sum_{i=1}^{m} \langle \lambda_i, f_i(x) \rangle + \sum_{j=1}^{p} \langle \nu_j, h_j(x) \rangle$$

- Dual objective (as usual):

$$g(\lambda,\nu) = \inf_x L\big(x, (\lambda,\nu)\big)$$

Duality with Generalized Inequalities

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_0(x) \\ s.t. \quad & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{aligned}$$

$$\begin{aligned} \max_{\lambda_i\in\mathbb{R}^{k_i},\nu\in\mathbb{R}^p} \quad & g(\lambda,\nu) \\ s.t. \quad & \lambda_i \succcurlyeq_{K_i^*} 0 \end{aligned}$$

$f_0 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex, $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ K_i-convex, $h_j : \mathbb{R}^n \to \mathbb{R}$ linear

x^* optimal for (P)

λ^*, ν^* optimal for (D)

<u>KKT Conditions</u>

$f_i(x^*) \preccurlyeq_{K_i} 0 \ \forall_{i=1\ldots m}$

$h_j(x^*) = 0 \ \forall_{j=1\ldots p}$

$\lambda_i^* \succcurlyeq_{K_i^*} 0 \ \forall_{i=1..m}$

$\nabla_x L\big(x^*, (\lambda^*, \nu^*)\big) = 0$

or $0 \in \partial_x L\big(x^*, (\lambda^*, \nu^*)\big)$

$\langle \lambda_i^*, f_i(x^*) \rangle = 0 \ \forall_{i=1..m}$

Multi-Dimensional Objective

$$\boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

$$f_0 : \mathbb{R}^n \to \mathbb{R}^k \cup \{\infty\} \; K\text{-convex}$$

$$f_i : \mathbb{R}^n \to \mathbb{R}^{k_i} \; K_i\text{-convex}, \; h_j : \mathbb{R}^n \to \mathbb{R} \text{ linear}$$

- Minimization is w.r.t. cone K: We prefer x over x' if $f_0(x) \preccurlyeq_K f_0(x')$
- Problem: might not be able to compare,
  i.e. possible that $f_0(x) \not\preccurlyeq_K f_0(x')$ but also $f_0(x') \not\preccurlyeq_K f_0(x)$

Multi-Criteria Optimization

$$f_0(x) = [F_1(x) \;\; F_2(x) \;\; \dots F_k(x)] \in \mathbb{R}^k$$

- Optimization w.r.t. positive orthant, i.e. elementwise comparison

$$f_0(x) \preccurlyeq_K f_0(x') \equiv f_0(x) \leq f_0(x') \equiv \forall_i \big(F_i(x) \leq F_i(x')\big)$$

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & F_1(x), F_2(x), \dots, F_k(x) \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

$F_i \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex

$f_i \colon \mathbb{R}^n \to \mathbb{R}^{k_i}$ K_i-convex, $h_j \colon \mathbb{R}^n \to \mathbb{R}$ linear

Examples

- Maximizing expected return and minimizing risk

  n stocks with means $\mu \in \mathbb{R}^n$ and covariance S

$$
\boxed{\begin{array}{ll}
\min\limits_{x \in \mathbb{R}^n} & F_1(x) = -\langle \mu, x \rangle, \quad F_2(x) = x^\top \mathrm{S} x \\
s.t. & x \geq 0 \\
& \sum_{i=1}^n x_i = 1
\end{array}}
$$

- Model fit and model complexity (regularization)

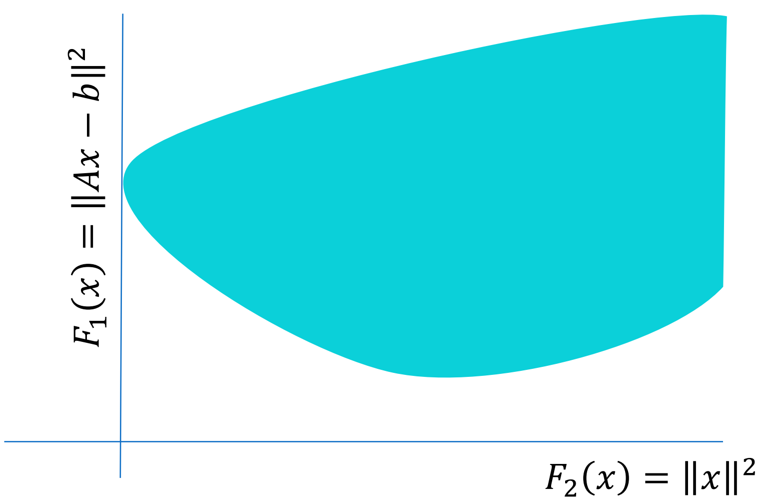
$$
\boxed{\min_{w \in \mathbb{R}^n} \quad F_1(w) = \|Aw - b\|_2^2, \quad F_2(w) = \|w\|_2^2}
$$

What is "Optimal"?

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f_0(x) = [F_1(x), F_2(x), \dots, F_k(x)] \\ s.t. \quad & f_i(x) \leq 0 \\ & h_i(x) = 0 \end{aligned}$$

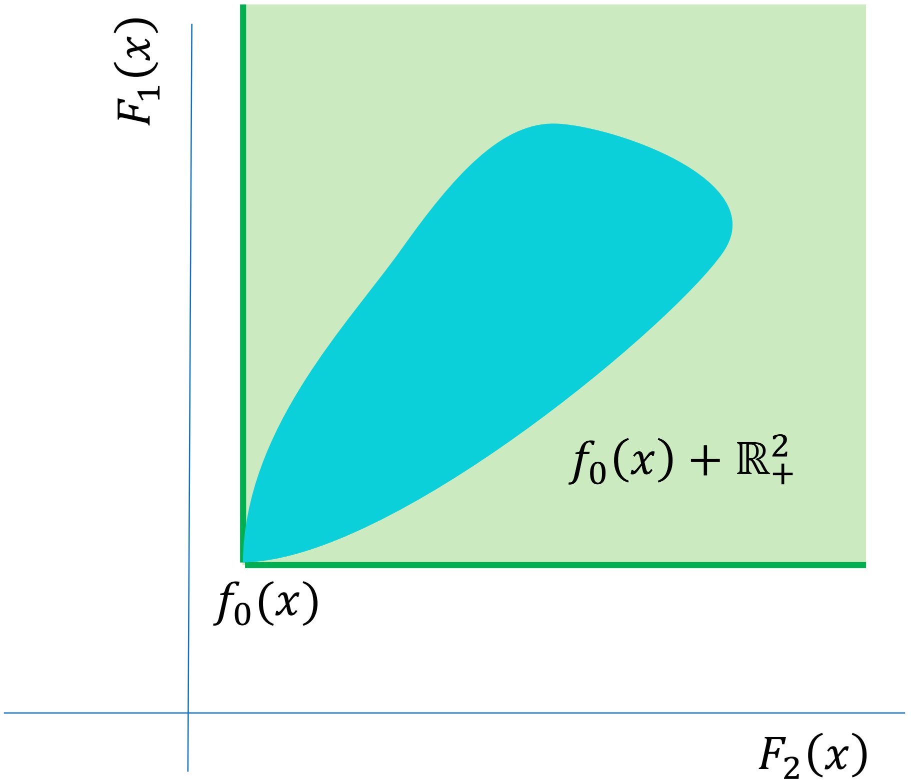
- Definition: A feasible x is **optimal** if $\forall$ feasible y, $f_0(x) \leq f_0(y)$ i.e. if $\forall$ feasible y, $\forall\, i$, $F_i(x) \leq F_i(y)$

- Set of attainable solutions: $\mathcal{O} = \{f_0(x) \mid x \text{ is feasible}\}$



When Does an Optimum Exist

- Definition: A feasible x is **optimal** if $\forall$ feasible y, $f_0(x) \leq f_0(y)$
- Claim: A feasible x is optimal iff $\mathcal{O} \subseteq f_0(x) + \mathbb{R}_+^k$

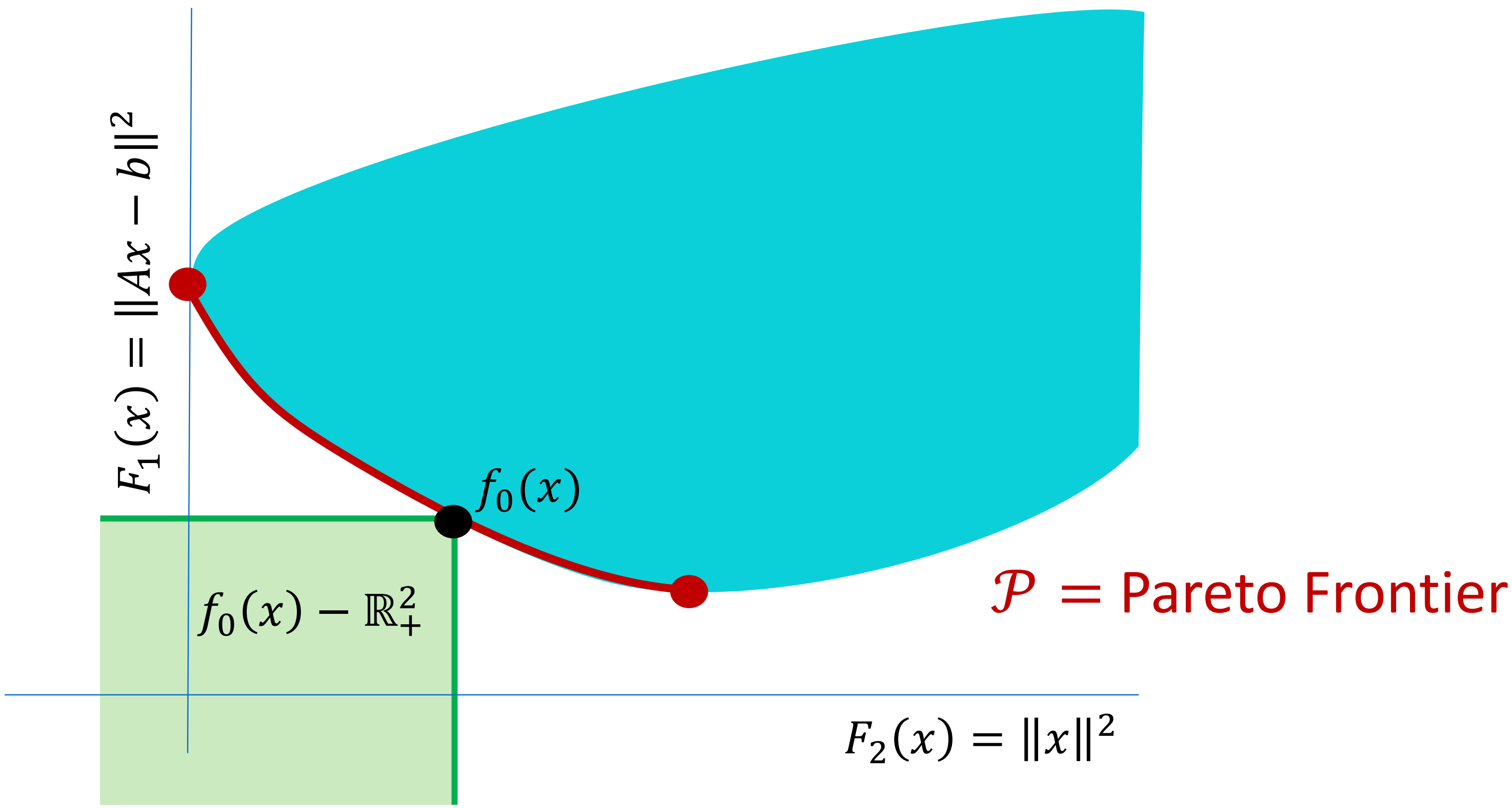


The Pareto Frontier

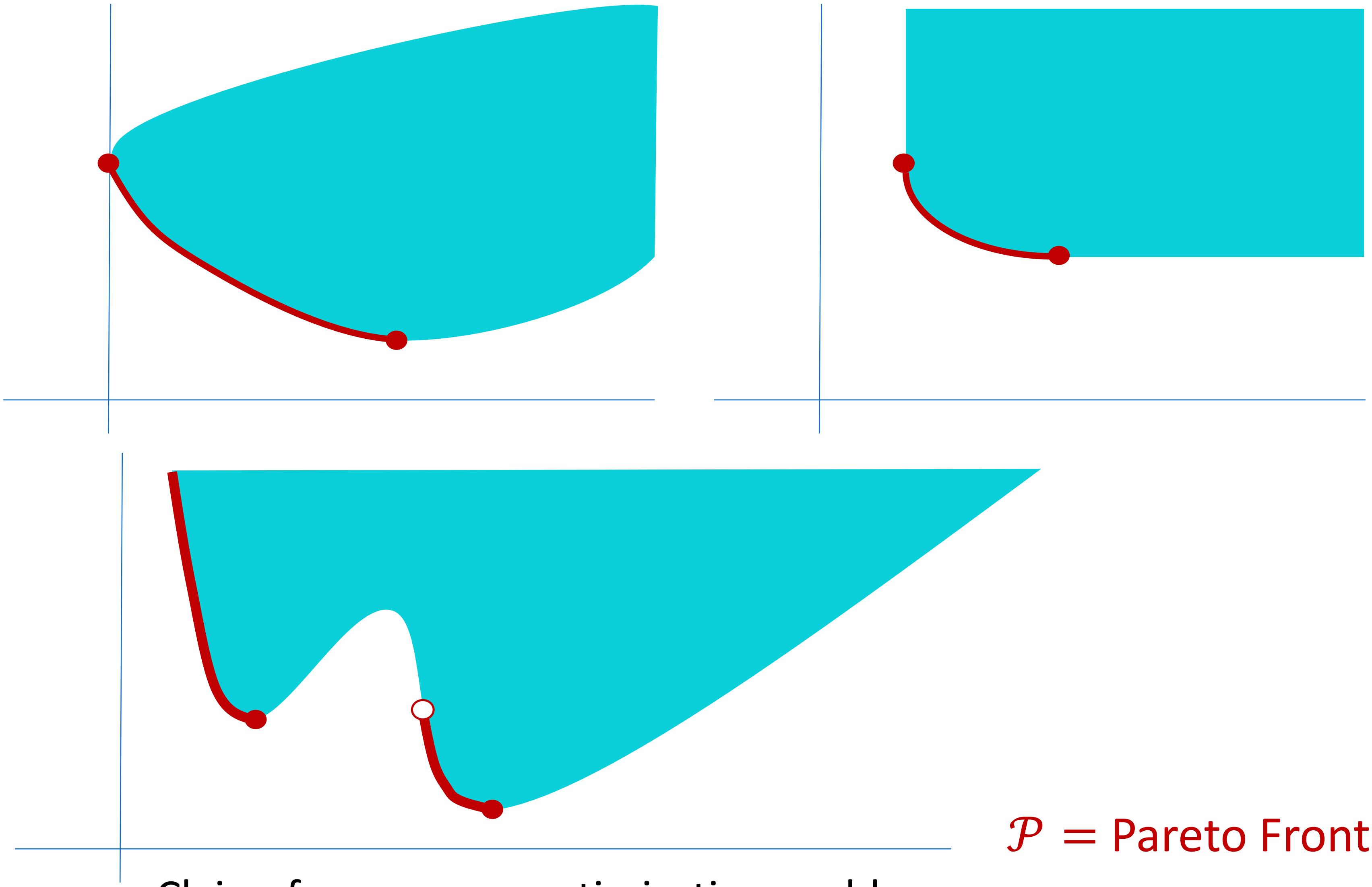
- Definition: A feasible x is **Pareto Optimal** iff

$$\forall \text{ feasible } y,\ f_0(y) \le f_0(x) \to f_0(y) = f_0(x)$$

- Claim: A feasible x is Pareto Optimal iff $\left(f_0(x) - \mathbb{R}_+^k\right) \cap \mathcal{O} = \{f_0(x)\}$



Examples



$\mathcal{P}$ = Pareto Frontier

Claim: for a convex optimization problem,

$\mathcal{O} + \mathbb{R}^n_+$ always convex and has same Pareto Frontier as $\mathcal{O}$

Scalarization

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) = [F_1(x), F_2(x), \dots, F_k(x)] \\ s.t. & f_i(x) \leq 0 \\ & h_i(x) = 0 \end{array}$$

- For $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \geq 0$, introduce the problem:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \langle \lambda, f_0(x) \rangle = \sum_{i=1}^{k} \lambda_i F_i(x) \\ s.t. & f_i(x) \leq 0 \\ & h_i(x) = 0 \end{array}$$

- Claim: for $\lambda > 0$, an optimum x_λ^* of the λ-problem is Pareto optimal

Proof: if $f_0(x) \leq f_0(x_\lambda^*)$ and better on at least one F_i, then $\langle \lambda, f_0(x) \rangle < \langle \lambda, f_0(x_\lambda^*) \rangle$

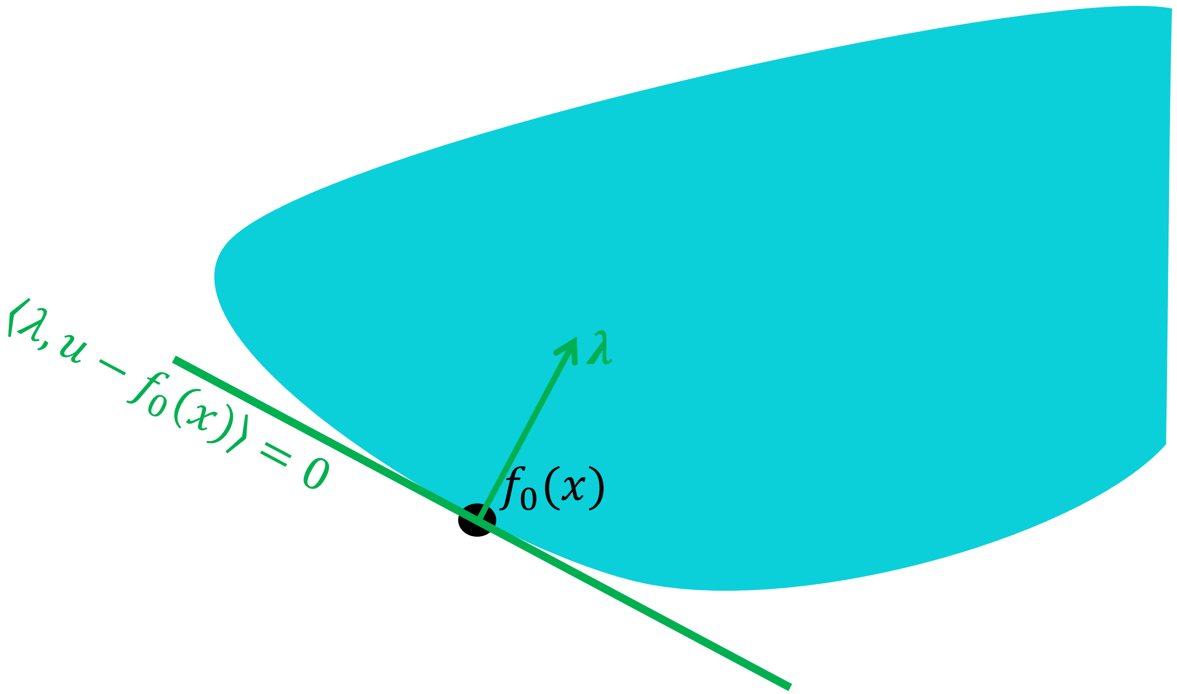
- Geometrically: $\forall_{feasible\ y} \langle \lambda, f_0(x_\lambda^*) \rangle \leq \langle \lambda, f_0(y) \rangle$

  ➔ $\forall_{u \in \mathcal{O}} \langle \lambda, f_0(x_\lambda^*) \rangle \leq \langle \lambda, u \rangle$

  ➔ $\{u | \langle \lambda, u - f_0(x_\lambda^*) \rangle = 0\}$ is a supporting HP of $\mathcal{O}$

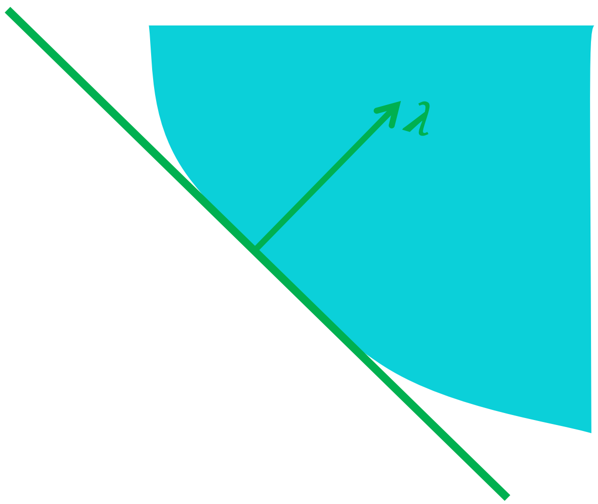
Scalarization

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \langle \lambda, f_0(x) \rangle = \sum_{i=1}^{k} \lambda_i F_i(x) \\ s.t. \quad & f_i(x) \le 0 \\ & h_i(x) = 0 \end{aligned}$$



$$\{\, x_\lambda^* \text{ optimal for } \lambda - \text{problem} \mid \lambda > 0 \} \subseteq \mathcal{P}$$

Scalarization



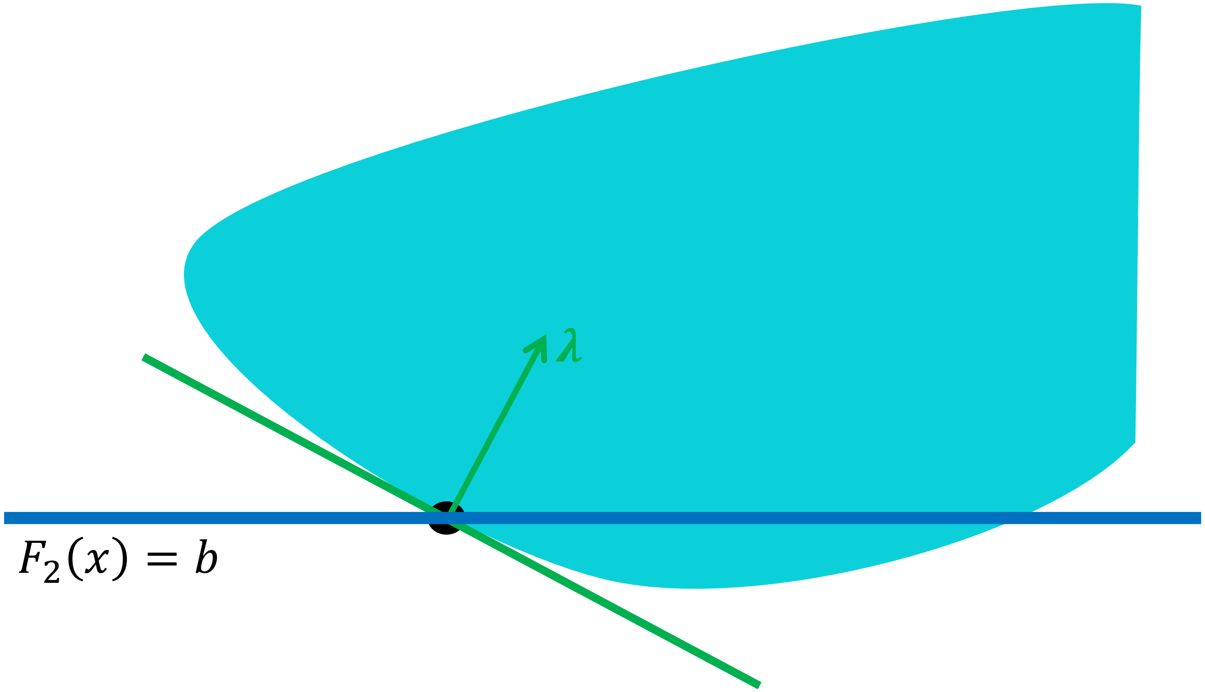
Multiple Pareto optimal x_λ^* for same λ

Not all x_λ^* for $\lambda = (1{,}0)$ and $\lambda = (0{,}1)$ are Pareto optimal

$$\{\, x_\lambda^* \text{ optimal for } \lambda - \text{problem} \mid \lambda > 0\} \subseteq \mathcal{P} \subseteq \{\, x_\lambda^* \text{ optimal for } \lambda - \text{problem} \mid \lambda \geq 0\}$$

Criteria as Constraints

$$\boxed{\min_{x \in \mathbb{R}^n} \quad F_1(x), F_2(x)}$$



$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & F_1(x) \\ s.t. & F_2(x) \leq b \end{array}}$$

$$\boxed{\min_{x \in \mathbb{R}^n} \quad F_1(x) + \lambda F_2(x)}$$

Optimization w.r.t. Cone K

$$\boxed{\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

$$f_0: \mathbb{R}^n \to \mathbb{R}^k \cup \{\infty\}\ K\text{-convex}$$

$$f_i: \mathbb{R}^n \to \mathbb{R}^{k_i}\ K_i\text{-convex},\ h_j: \mathbb{R}^n \to \mathbb{R} \text{ linear}$$

- Definition: A feasible x is **optimal** if $\forall$ feasible y, $f_0(x) \preccurlyeq_K f_0(y)$
- Claim: A feasible x is optimal iff $\mathcal{O} \subseteq f_0(x) + K$

- Definition: A feasible x is **Pareto Optimal** iff
  $\forall$ feasible y, $f_0(y) \preccurlyeq_K f_0(x) \to f_0(y) = f_0(x)$
- Claim: A feasible x is Pareto Optimal iff $(f_0(x) - K) \cap \mathcal{O} = \{f_0(x)\}$

Scalarization for K-Optimization

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

- For $\lambda \succcurlyeq_K 0$, introduce the problem:

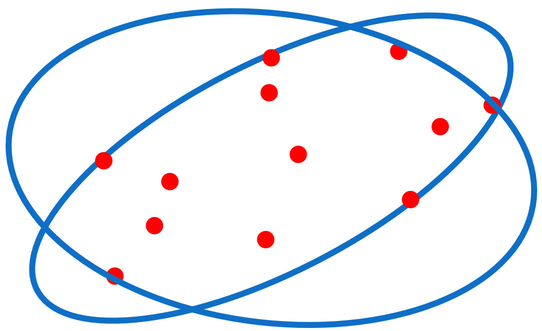
$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \langle \lambda, f_0(x) \rangle \\ s.t. & f_i(x) \preccurlyeq_{K_i} 0 \\ & h_i(x) = 0 \end{array}}$$

- Claim: for $\lambda \succ_K 0$, an optimum x_λ^* of the λ-problem is Pareto optimal

$\lambda \in interior(K)$

$$\{ x_\lambda^* \text{ optimal for } \lambda - \text{problem} \mid \lambda \succ_K 0\}$$
$$\subseteq \mathcal{P} \subseteq$$
$$\{ x_\lambda^* \text{ optimal for } \lambda - \text{problem} \mid \lambda \succcurlyeq_K 0\}$$

Matrix Optimization: Example



- Find ***minimal*** encompassing ellipsoid of $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ (i.e. encompassing ellipsoid that isn't a superset of any other encompassing ellipsoid)

$$\boxed{\begin{array}{ll} \min\limits_{X \in S^n} & (-X) \\ s.t. & a_i^\top X a_i \leq 1 \;\; i = 1..m \end{array}}$$

Optimization is w.r.t. the semi-definite cone (matrix inequality)