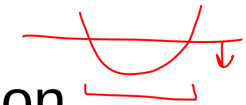
Convex optimization



minimize $f(x)$ subject to $Ax = b$

$g_i(x) \le 0 \quad \forall i \in I$

($x \in \mathbb{R}^n$; f convex; $A \in \mathbb{R}^{m \times n}$; $b \in \mathbb{R}^m$; g_i convex)

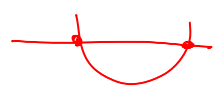
e.g., min $x^2 + y^2$ s.t. $1 - x - y \le 0$ ($g_i(x,y) = 1 - x - y$)

Linear inequalities: OK

Positivity: OK

} just include in g_i's

Terminology



- Feasible $\exists x.\ g_i(x) \le 0\ \forall i \ \wedge\ Ax = b$
- Infeasible not feasible
- Unbounded $\forall v \in \mathbb{R}\ \exists$ feasible $x.\ f(x) \le v$

- Two convex programs are equivalent if:

change of vars, slacks on linear ineq., max/min or $\le$ / $\ge$ by negating

Transformations of CPs

- Optimizing some variables
- min $f(x,y) = y^2 + (x - 2y)^2 / 2$

x y

$\partial f / \partial x = x - 2y = 0 \Rightarrow x = 2y$

min $y^2 + (2y - 2y)^2 / 2$

Transformations of CPs

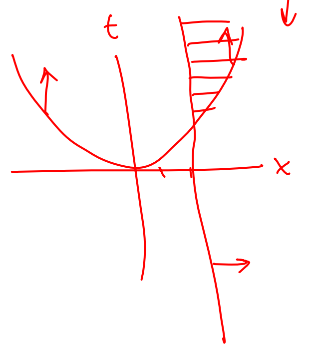
- Implicit constraints
- min x^2 s.t. $x \geq 1$
- min $x^2 + I_{x \geq 1}(x)$

  $I_{x\geq 1}(x) = \begin{cases} 0 & x \geq 1 \\ \infty & o/w \end{cases}$
- For any set S, $I_S(x) = \begin{cases} 0 & x \in S \\ \infty & x \notin S \end{cases}$

Transformations of CPs

- Epigraph form:
- $\min_x \; x^2$ s.t. $x \geq 2$
- $\min_{x,t} \; t$ s.t. $x \geq 2 \quad t \geq x^2$

objective



Optimality

- Optimal value:

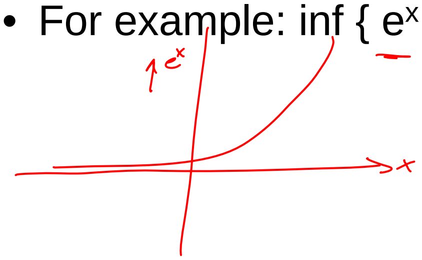
  $v^* = \inf \{ f(x) \mid g_i(x) \leq 0 \ (\forall i), Ax = b \}$

Also, $v^* \leq x \quad \forall x \in S$

- Definition of $v^* = \inf S$: $\subseteq \mathbb{R}$ — infimum

  $\iff \forall \epsilon > 0 \; \exists x_\epsilon \in S. \; v^* + \epsilon \geq x_\epsilon$

- For example: $\inf \{ e^x \mid x \text{ real} \} =$



sup (supremum)

tightest upper bound

Optimal point

- x* is optimal iff:
- Is there always an optimal point?
  - ex. 1:
  - ex. 2:

Local optima

- min f(x) s.t. $g(x) \le 0$
- x_0 is a local optimum iff
- In a convex program,

An optimality criterion

- Suppose f(x) is differentiable, convex
- Then x* = arg min f(x) iff
- What about x* = arg min f(x) s.t. $x \in C$?

Proof of criterion

Example: logistic regression

- Data $x_i \in R$, $y_i \in \{0,1\}$
- $P(y_i = 1 \mid x_i, w, b) = s(w^T x_i + b)$

  $s(z) = 1/(1+\exp(-z))$

- $\max_{wb} P(w, b)\, P(y \mid x, w, b) =$

  $\max_{wb} P(w, b) \prod_i P(y_i \mid x_i, w, b) =$

Ex: minimize top eigenvalue

- Suppose $A_i \in S^{n*n}$ for i = {1, 2, …}
- $\min_{\alpha,A} \;\; \lambda_1(A)$ s.t.

  $A = \alpha_1 A_1 + \alpha_2 A_2 + \ldots$

Ex: minimum volume ellipsoid

- Given points $x_1, x_2, \ldots, x_k$
- $\min_{A, x_C}$ vol(A) s.t.
  $(x_i - x_C)^T A (x_i - x_C) \leq 1 \quad i = 1, \ldots, k$
  $A \in S^{n*n}$
  $A \succcurlyeq 0$

Schur complement

- Symmetric block matrix M =
- Schur complement is S =
- $M \succcurlyeq 0$ iff

Back to min-volume ellipsoid

- $\max_{A,x_C} \log |A|$ s.t.
  $(x_i - x_C)^T A (x_i - x_C) \leq 1 \quad i = 1, \ldots, k$
  $A = A^T, A \succcurlyeq 0$
- $\max_{A,B,x_C,z} \log |A|$ s.t.

More examples

- SDP: learning a distance matrix (multidimensional scaling)
- learning a kernel matrix
- semidefinite embedding
- maximum variance unfolding