* Elements of Convex Optimization *
(part II)

(0)

$$\begin{cases} f_0(x) \to \min \\ f_i(x) \le 0, \quad i=1\dots m \\ a_i^T x = b_i \end{cases}$$

General Optimization Problem (*)

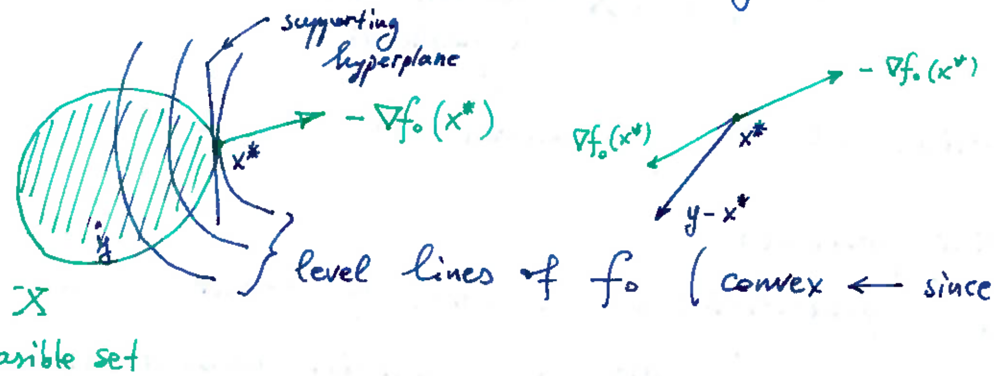
can be also written is a matrix form $Ax=b$

where $f_0, f_1, \dots, f_m$ are convex. Note that equality constraints are affine!

* <u>Note</u>: the feasible set of (*) is convex.

* If f_0 is differentiable, then

$$x^* \text{ optimal} \iff \nabla f_0(x^*)^T (y - x^*) \ge 0 \quad \forall y \in X \qquad (\#)$$



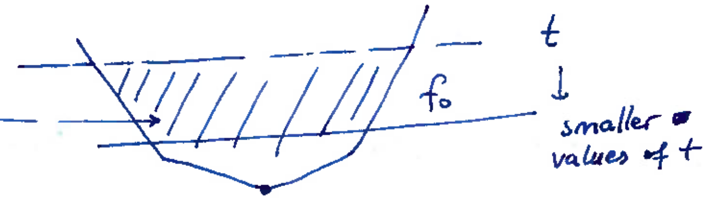
* <u>Note</u>: If the problem is unconstrained, then (#) holds $\forall y \in \mathbb{R}^n$

$$\Rightarrow \nabla f_0(x^*)^T \underbrace{(y - x^*)}_{\text{any vector}} \ge 0 \Rightarrow \nabla f_0(x^*) = 0. \quad \text{(as expected)}$$

In fact, this is if and only if condition.

Standard problem (*) can be re-written is several ways. A useful one is this:

$$\begin{cases} t \to \min \\ f_0(x) - t \le 0 \\ f_i(x) \le 0 \quad i = 1 \ldots m \\ Ax = b \end{cases}$$



This is an optimization problem with <u>linear</u> objective function.

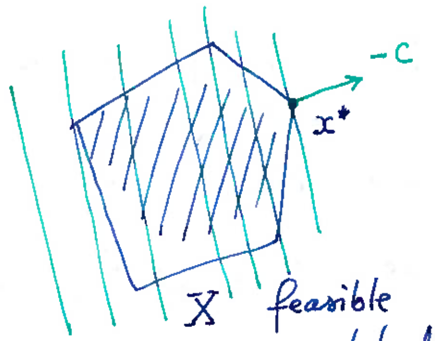
'Linear objective is universal'

① Linear program (LP)

$$\begin{cases} c^T x + b \to \min \\ Gx \le h \\ Ax = b \end{cases}$$

pointwise

Thm: Solution is a vertex



feasible set is a polyhedron

Number of vertices of the polyhedron X can be huge!

<u>Example 1</u>: Diet problem:

- $x_1 \ldots x_n$ quantities of n types of foods
- a unit of food j costs c_j, contains amount a_{ij} of nutrient i
- health diet requires nutrient i in quantity at least b_i

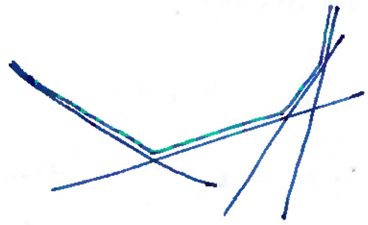
Cheapest health diet?

$$\begin{cases} c^T x \to \min \\ Ax \ge b, \quad x \ge 0 \end{cases}$$

Example 2: Piecewise-linear minimization

$$\max_{i=1\dots m} \left(a_i^T x + b_i\right) \longrightarrow \min$$

Not a linear problem!

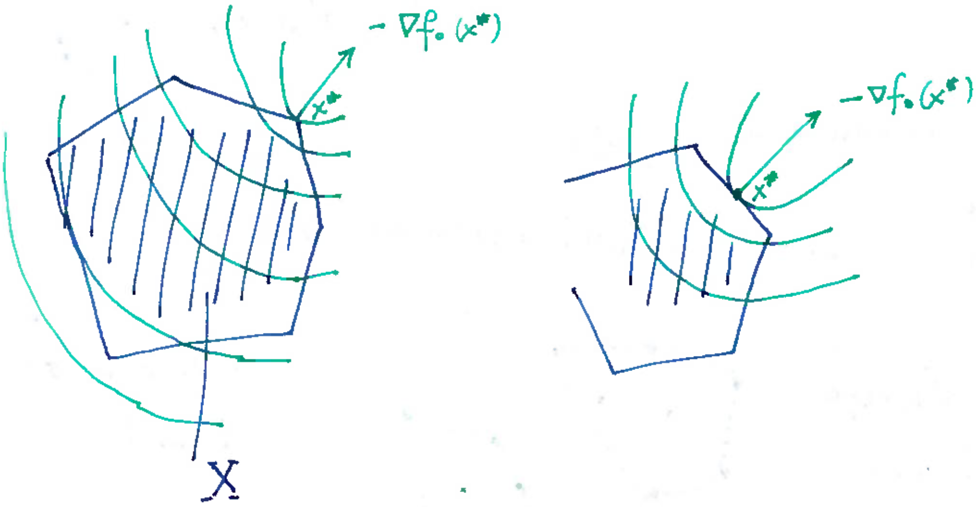


$$\underset{LP}{\Longleftrightarrow} \quad \begin{cases} t \to \min \\ a_i^T x + b_i \leq t, \quad i = 1\dots m \end{cases}$$

? Example 3: Inscribed center?

② Quadratic program (QP)

$$\begin{cases} \underbrace{\tfrac{1}{2} x^T P x + q^T x + c}_{f_0} \longrightarrow \min \\ Gx \leq h \\ Ax = b \end{cases}, \quad P \text{ is positive semi-definite} \quad P \succeq 0$$



Examples 1: $\|Ax - b\|_2^2 \longrightarrow \min$

Analytic solution: $x^* = A^{-1} b$

Can add: $l \leq x \leq u$ range constraint

? Quadratically constrained quadratic program

! Second-order cone programming (SOCP)

③ Robust Linear programming:

* parameters in the ~~LP~~ LP can be coming w/ uncertainty.

$$\begin{cases} c^T x \longrightarrow \min \\ a_i^T x \leq b_i, \quad i = 1 \dots m \end{cases}$$

w/ uncertainty in c, a_i, b_i:

→ deterministic: $\begin{cases} c^T x \longrightarrow \min \\ a_i^T x \leq b_i, \quad a_i \in \mathcal{E}_i \quad i = 1 \dots m \end{cases}$

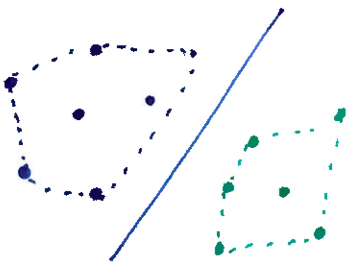
→ probabilistic: a_i is a random variable

$$\begin{cases} c^T x \longrightarrow \min \\ \mathbb{P}(a_i^T x \leq b_i) \geq \eta \end{cases}$$

Same w/ c, b_i.

④ Linear discrimination: ~~SVM~~

$a^T x_i + b > 0$
$i = 1 \dots N$



$a^T y_i + b < 0$
$i = 1 \dots M$

Homogeneous in a, b → we can separate

because inequalities are strict

$a^T x_i + b \geq 1$ $\qquad$ $a^T y_i + b \leq -1$

$\mathcal{H}_1 = \{ z \mid a^T z + b = 1 \}$

$\mathcal{H}_2 = \{ z \mid a^T x + b = -1 \}$

$d(\mathcal{H}_1, \mathcal{H}_2) = \dfrac{2}{\|a\|_2} \longrightarrow \max$

$$\rightsquigarrow \begin{cases} \|a\|_2 \longrightarrow \min \iff \|a\|_2^2 \longrightarrow \min \\ a^T x_i + b \geq 1 \quad i = 1 \dots N \\ a^T y_i + b \leq -1 \quad i = 1 \dots M \end{cases}$$

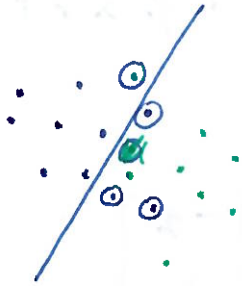
QP in a, b

⑤ Approximate linear separation for non-separable sets.
(spam − or − not)

$$\begin{cases} \mathbb{1}^T u + \mathbb{1}^T v = \underbrace{\sum u_i + \sum v_i}_{\text{slack}} \longrightarrow \min \\ a^T x_i + b \geq 1 - u_i & i = 1 \dots N \\ a^T y_i + b \leq -1 + v_i & i = 1 \dots M \\ u \geq 0, \quad v \geq 0 \end{cases}$$

pointwise

~ minimizing # misclassified pt's



* If u_i, v_i are zero, we get separation.

⟶ LP in a, b, u, v

⑥ Support vector machine

$$\begin{cases} \|a\|_2 + \gamma\left(\mathbb{1}^T u + \mathbb{1}^T v\right) \longrightarrow \min \\ a^T x_i + b \geq 1 - u_i & i = 1 \dots N \\ a^T y_i + b \leq -1 + v_i & i = 1 \dots M \\ u \geq 0, \; v \geq 0 \end{cases}$$

γ is a parameter weight

$\gamma \sim 0.1$

⑦ How to solve LP/QP? Use methods discussed earlier.

* ~~[illegible]~~ : ~~[illegible]~~min
* Unconstrained ~~[illegible]~~

These methods work much better because for convex functions, local optimum is global!

Newton method [illegible] works amazingly well (scalling invariant)

• Convergance analysis → Classical
→ self-concordance

* Equality constrained

$$\begin{cases} f(x) \to \min \\ Ax = b \end{cases}, \quad f \text{ convex (smooth, } A \text{ full rank)}$$

x^* optimal $\underset{\exists \nu}{\iff}$ $\nabla f(x^*) + A^T \nu^* = 0, \quad Ax^* = b$

Newton method

n+p equations in n+p unknowns

Example:

$$\begin{cases} \frac{1}{2} x^T P x + q^T x + r \to \min \\ Ax = b \end{cases} \quad (P \geq 0)$$

$$\underbrace{\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Linear system

KKT matrix.

KKT matrix is non-singular $\iff$ $(Ax = 0, x \neq 0 \Rightarrow x^T P x > 0)$

$\iff$ $P + A^T A > 0$

In general, see page 7.

* Inequality constraints

$$\begin{cases} f_0(x) \to \min \\ f_i(x) \leq 0 \quad i = 1..m \\ Ax = b \end{cases}$$

We can reduce it to

$$\begin{cases} f_0(x) \underbrace{- \frac{1}{t} \sum_{i=1}^{m} \log(-f_i(x))}_{\text{convex}} \to \min \\ Ax = b \end{cases}$$

log-barriers.

* In general, for equality constraint:

Newton step: $$x^{(k+1)} = x^{(k)} - \frac{\nabla f(x^{(k)})}{\nabla^2 f(x^{(k)})}$$

$$\Delta x^{(k)} := x^{(k+1)} - x^{(k)}$$

In the constrained case,

feasible

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

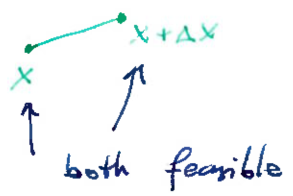
optimality

ensures that we are doing a feasible step

For optimality:

$$\nabla f(x+\Delta x) + A^T w = 0$$

$$A(x+\Delta x) = 0$$



$\nabla f(x+\Delta x) \sim \nabla f(x) + \nabla^2 f(x)\cdot \Delta x$, and this approximation gives the system

~~Method~~