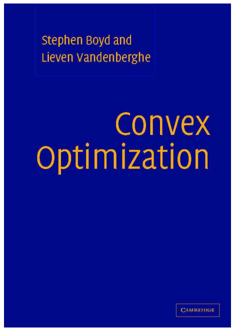
Convex Optimization & Machine Learning

Convex Problems & Duality

mcuturi@i.kyoto-u.ac.jp

Most slides in this lecture are taken from 

Convex optimization problems

Convex optimization problem

standard form convex optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{array}$$

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- often written as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

important property: feasible set of a convex optimization problem is convex

Importance of a good formulation

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1+x_2^2) \le 0 \\ & h_1(x) = (x_1+x_2)^2 = 0 \end{array}$$

- f_0 is convex; feasible set $\{(x_1, x_2) \mid x_1 = -x_2 \le 0\}$ is convex
- not a convex problem (according to our definition):
  - f_1 is not convex,
  - h_1 is not affine
- equivalent (but not identical) to the convex problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \le 0 \\ & x_1 + x_2 = 0 \end{array}$$

Local and global optima

any locally optimal point of a convex problem is (globally) optimal

proof: suppose x is locally optimal and y is optimal with $f_0(y) < f_0(x)$

x locally optimal means there is an $R > 0$ such that

$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \quad \Longrightarrow \quad f_0(z) \geq f_0(x)$$

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
- z is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = R/2$ and

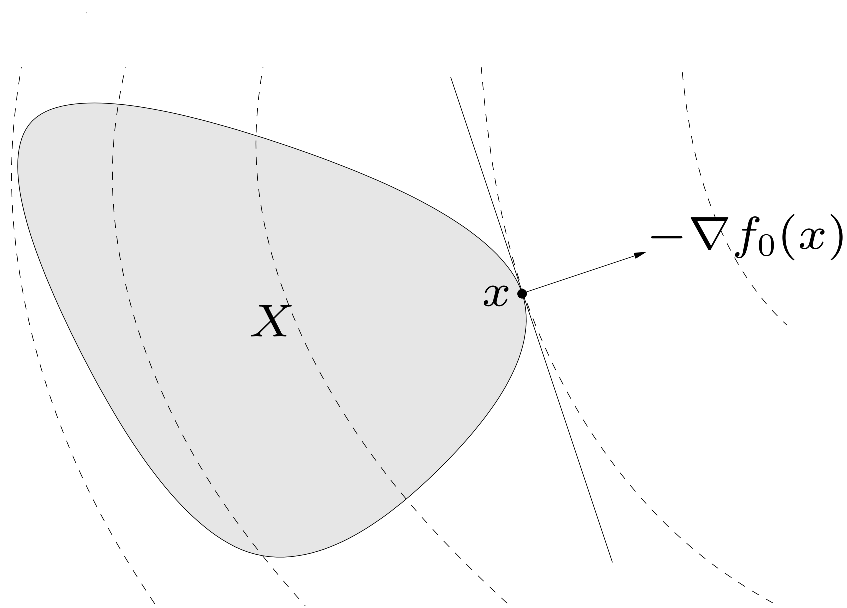
$$f_0(z) \leq \theta f_0(x) + (1 - \theta) f_0(y) < f_0(x)$$

which contradicts our assumption that x is locally optimal

Optimality criterion for differentiable f_0

x is optimal if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \text{for all feasible } y$$



if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

Optimality criterion for differentiable f_0

- **unconstrained problem**: x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad \nabla f_0(x) = 0$$

Optimality criterion for differentiable f_0

- **equality constrained problem**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad Ax = b$$

x is optimal if and only if there exists a ν such that

$$x \in \mathbf{dom}\, f_0, \qquad Ax = b, \qquad \boldsymbol{\nabla f_0(x) + A^T \nu = 0}$$

Optimality criterion for differentiable f_0

- **equality constrained problem**: x optimal iff there exists a ν such that

$$x \in \mathbf{dom}\, f_0, \qquad Ax = b, \qquad \boldsymbol{\nabla f_0(x) + A^T\nu = 0}$$

- Why? Remember $\nabla f_0(x)^T(y-x) \geq 0 \quad$ for all feasible y.
- Yet, for any feasible y, $\exists \nu$ such that $y = x + \nu$ and $A\nu = 0$.
- For any ν such that $A\nu = 0$ (ν in the **null space** $\mathcal{N}(A)$ of A),

$$\nabla f_0(x)^T \nu \geq 0$$

- For $\nabla f_0(x)^T$, linear function, to be negative on a subspace, it must be 0. Hence $\nabla f_0(x) \perp \mathcal{N}(A)$.
- This is equivalent to saying, since $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$, that there exists ν such that $\nabla f_0(x) + A^T\nu = 0$.

Optimality criterion for differentiable f_0

- **minimization over nonnegative orthant**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \succeq 0$$

  x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad x \succeq 0, \qquad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

- Check p.142 of Boyd & Vandenberghe to see why.

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

  is equivalent to

$$\begin{array}{ll} \text{minimize (over } z\text{)} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, \quad i = 1, \ldots, m \end{array}$$

  where F and x_0 are such that

$$Ax = b \quad \Longleftrightarrow \quad x = Fz + x_0 \text{ for some } z$$

- **introducing equality constraints**

$$\begin{array}{ll} \text{minimize} & f_0(A_0x+b_0) \\ \text{subject to} & f_i(A_ix+b_i) \leq 0, \quad i=1,\ldots,m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x,\ y_i) & f_0(y_0) \\ \text{subject to} & f_i(y_i) \leq 0, \quad i=1,\ldots,m \\ & y_i = A_ix+b_i, \quad i=0,1,\ldots,m \end{array}$$

- **introducing slack variables for linear inequalities**

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & a_i^Tx \leq b_i, \quad i=1,\ldots,m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x,\ s) & f_0(x) \\ \text{subject to} & a_i^Tx+s_i = b_i, \quad i=1,\ldots,m \\ & s_i \geq 0, \quad i=1,\ldots m \end{array}$$

- **epigraph form**: standard form convex problem is equivalent to

$$\begin{array}{ll} \text{minimize (over } x, t) & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- **minimizing over some variables**

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

is equivalent to

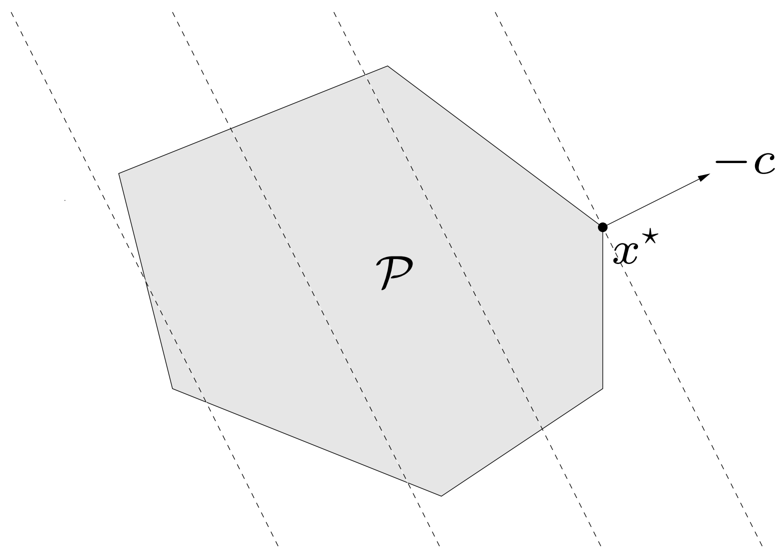
$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$

Linear program (LP)

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \succeq b, \quad x \succeq 0 \end{array}$$

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

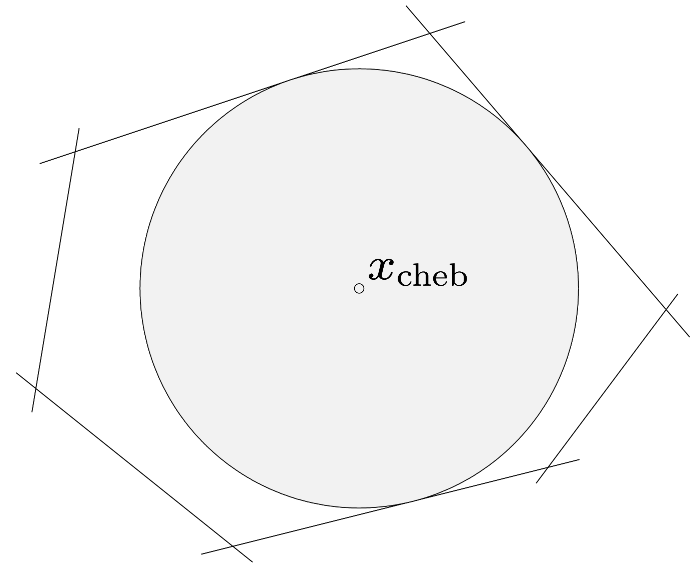
Chebyshev center of a polyhedron

Chebyshev center of

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$$

is center of largest inscribed ball

$$\mathcal{B} = \{x_c + u \mid \|u\|_2 \leq r\}$$



- $a_i^T x \leq b_i$ for all $x \in \mathcal{B}$ if and only if

$$\sup\{a_i^T(x_c + u) \mid \|u\|_2 \leq r\} = a_i^T x_c + r\|a_i\|_2 \leq b_i$$

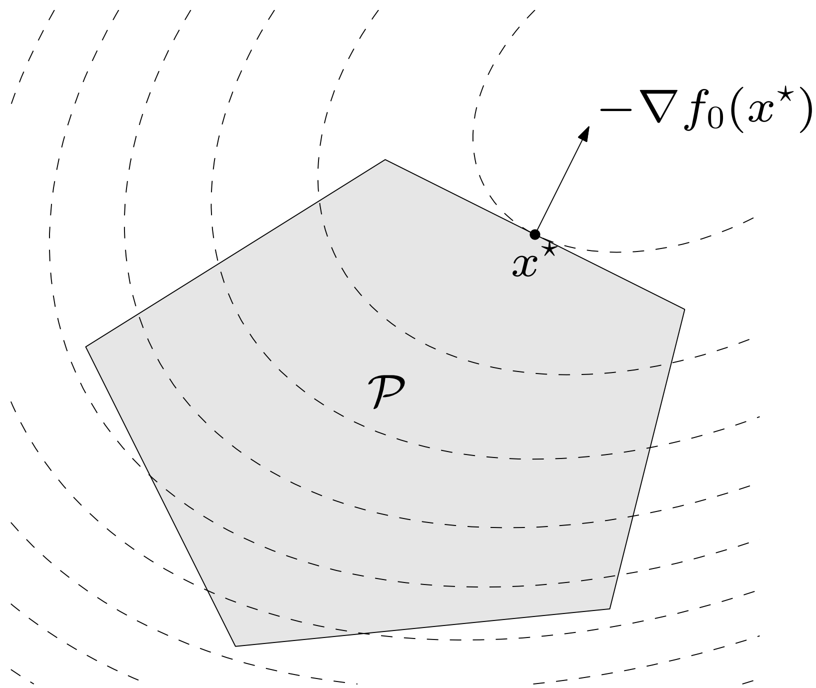
- hence, x_c, r can be determined by solving the LP

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T x_c + r\|a_i\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

Quadratic program (QP)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

- $P \in \mathbf{S}^n_+$, so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron



Examples

least-squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- analytical solution $x^\star = A^\dagger b$ ($A^\dagger$ is pseudo-inverse)
- can add linear constraints, *e.g.*, $l \preceq x \preceq u$

linear program with random cost

$$\begin{array}{ll} \text{minimize} & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma\, \mathbf{var}(c^T x) \\ \text{subject to} & Gx \preceq h, \quad Ax = b \end{array}$$

- c is random vector with mean $\bar{c}$ and covariance Σ
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)

Quadratically constrained quadratic program (QCQP)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- $P_i \in \mathbf{S}^n_+$; objective and constraints are convex quadratic
- if $P_1, \ldots, P_m \in \mathbf{S}^n_{++}$, feasible region is intersection of m ellipsoids and an affine set

Second-order cone programming

$$
\begin{array}{ll}
\text{minimize} & f^T x \\
\text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\
& Fx = g
\end{array}
$$

($A_i \in \mathbf{R}^{n_i \times n}$, $F \in \mathbf{R}^{p \times n}$)

- inequalities are called second-order cone (SOC) constraints:

$$
(A_i x + b_i, c_i^T x + d_i) \in \text{second-order cone in } \mathbf{R}^{n_i+1}
$$

- for $n_i = 0$, reduces to an LP; if $c_i = 0$, reduces to a QCQP
- more general than QCQP and LP

Robust linear programming

the parameters in optimization problems are often uncertain, *e.g.*, in an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

there can be uncertainty in c, a_i, b_i

two common approaches to handling uncertainty (in a_i, for simplicity)

- deterministic model: constraints must hold for all $a_i \in \mathcal{E}_i$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \text{ for all } a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m, \end{array}$$

- stochastic model: a_i is random variable; constraints must hold with probability η

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{Prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m \end{array}$$

deterministic approach via SOCP

- choose an ellipsoid as $\mathcal{E}_i$:

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\} \qquad (\bar{a}_i \in \mathbf{R}^n, \quad P_i \in \mathbf{R}^{n \times n})$$

  center is $\bar{a}_i$, semi-axes determined by singular values/vectors of P_i

- robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \dots, m \end{array}$$

  is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \|P_i^T x\|_2 \le b_i, \quad i = 1, \dots, m \end{array}$$

  (follows from $\sup_{\|u\|_2 \le 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2$)

stochastic approach via SOCP

- assume a_i is Gaussian with mean $\bar{a}_i$, covariance Σ_i ($a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$)
- $a_i^T x$ is Gaussian r.v. with mean $\bar{a}_i^T x$, variance $x^T \Sigma_i x$; hence

$$\mathbf{Prob}(a_i^T x \le b_i) = \Phi\left(\frac{b_i - \bar{a}_i^T x}{\|\Sigma_i^{1/2} x\|_2}\right)$$

where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt$ is CDF of $\mathcal{N}(0,1)$

- robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{Prob}(a_i^T x \le b_i) \ge \eta, \quad i = 1, \ldots, m, \end{array}$$

with $\eta \ge 1/2$, is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \le b_i, \quad i = 1, \ldots, m \end{array}$$

Geometric programming

monomial function

$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \qquad \mathbf{dom}\, f = \mathbf{R}_{++}^n$$

with $c > 0$; exponent α_i can be any real number

posynomial function: sum of monomials

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}, \qquad \mathbf{dom}\, f = \mathbf{R}_{++}^n$$

geometric program (GP)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 1, \quad i = 1, \ldots, m \\ & h_i(x) = 1, \quad i = 1, \ldots, p \end{array}$$

with f_i posynomial, h_i monomial

Geometric program in convex form

change variables to $y_i = \log x_i$, and take logarithm of cost, constraints

- monomial $f(x) = cx_1^{a_1} \cdots x_n^{a_n}$ transforms to

$$\log f(e^{y_1}, \ldots, e^{y_n}) = a^T y + b \qquad (b = \log c)$$

- posynomial $f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$ transforms to

$$\log f(e^{y_1}, \ldots, e^{y_n}) = \log \left(\sum_{k=1}^{K} e^{a_k^T y + b_k} \right) \qquad (b_k = \log c_k)$$

- geometric program transforms to convex problem

$$\begin{array}{ll} \text{minimize} & \log \left(\sum_{k=1}^{K} \exp(a_{0k}^T y + b_{0k}) \right) \\ \text{subject to} & \log \left(\sum_{k=1}^{K} \exp(a_{ik}^T y + b_{ik}) \right) \leq 0, \quad i = 1, \ldots, m \\ & Gy + d = 0 \end{array}$$

Semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

with F_i, $G \in \mathbf{S}^k$

- inequality constraint is called linear matrix inequality (LMI)
- includes problems with multiple LMI constraints: for example,

$$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \qquad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$$

  is equivalent to single LMI

$$x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$$

Duality

Duality

- **Duality theory**:
  - Keep this in mind: only a long list of **simple** inequalities. . . .
  - In the end: very powerful results at low technical/numerical cost.
  - A few important, intuitive theorems.

- **In a LP context**:
  - Dual problem provides a different **interpretation** on the same problem.
  - Essentially assigns cost ("displeasure" measure) to constraints.
  - Provides alternative algorithms (dual-simplex).

- **In a more general context**:
  - Very powerful tool to give approximate solutions to intractable problems.

Duality : the general case

Optimization problem

- Consider the following **mathematical program**:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

  where $\mathbf{x} \in \mathcal{D} \subset \mathbf{R}^n$ with optimal value $p^\star$.

- **No particular assumptions** on $\mathcal{D}$ and the functions f and h (nothing about convexity, linearity, continuity, $etc.$)
- Very generic (includes linear programming and many other problems)

Lagrangian

We form the **Lagrangian** of this problem:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda_i} f_i(\mathbf{x}) + \sum_{i=1}^{p} \boldsymbol{\mu_i} h_i(\mathbf{x}).$$

Variables $\lambda \in \mathbf{R}^m$ and $\mu \in \mathbf{R}^p$ are called **Lagrange multipliers**.

- The Lagrangian is a **penalized** version of the original objective
- The Lagrange multipliers λ_i, μ_i control the weight of the penalties.
- The Lagrangian is a smoothed version of the hard problem, we have turned $\mathbf{x} \in C$ into penalties that take into account the constraints that **define** C.

Lagrange dual function

- We originally have

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda_i} f_i(\mathbf{x}) + \sum_{i=1}^{p} \boldsymbol{\mu_i} h_i(\mathbf{x})$$

- The penalized problem is here:

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \inf_{\mathbf{x} \in \mathcal{D}} f_0(\mathbf{x}) + \textstyle\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}) \end{aligned}$$

- The function $g(\lambda, \mu)$ is called the **Lagrange dual function**.
  - Easier to solve than the original one (the constraints are gone)
  - Can often be computed explicitly (more later)

Lower bound

- The function $g(\lambda, \mu)$ produces a lower bound on $p^\star$.
- **Lower bound property**: If $\lambda \geq 0$, then $g(\lambda, \mu) \leq p^\star$
- Why?
  - If $\tilde{\mathbf{x}}$ is feasible,
    - ▷ $f_i(\tilde{\mathbf{x}}) \leq 0$ and thus $\lambda_i f_i(\tilde{x}) \leq 0$
    - ▷ $h_i(\tilde{\mathbf{x}}) = 0$, and thus $\mu_i h_i(\tilde{x}) = 0$
  - thus by construction of L:

$$g(\lambda, \mu) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \mu) \leq L(\tilde{\mathbf{x}}, \lambda, \mu) \leq f_0(\tilde{\mathbf{x}})$$

  - This is true for any feasible $\tilde{\mathbf{x}}$, so it must be true for the optimal one, which means $g(\lambda, \mu) \leq f_0(\mathbf{x}^\star) = p^\star$.

Lower bound

- We have a **systematic** way of producing **lower bounds** on the optimal value $p^\star$ of the original problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

  by computing the value for a given (λ, μ) couple where $\lambda \geq \mathbf{0}$.

- We can look for the best possible one. . .

Dual problem

- We can define the **Lagrange dual** problem:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \mu) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

  in the variables $\lambda \in \mathbf{R}^m$ and $\mu \in \mathbf{R}^p$.

- Finds the best, that is **highest**, possible lower bound $g(\lambda, \mu)$ on the optimal value $p^\star$ of the original (now called **primal**) problem.

- We call its optimal value $d^\star$

Dual problem

- For each given $\mathbf{x}$, the function

$$L(\mathbf{x}, \lambda, \mu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x})$$

is **linear** in the variables λ and μ.

- This means that the function

$$g(\lambda, \mu) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \mu)$$

is a minimum of linear functions of (λ, μ), so it must be **concave** in (λ, μ)

- This means that the dual problem is always a **concave maximization** problem, whatever f, g, h's properties are.

Weak duality

We have shown the following property called **weak duality**:

$$d^\star \leq p^\star$$

the optimal value of the **dual** is
always **less** than the optimal value of the **primal** problem.

- We haven't made any assumptions on the problem... **no mention of convexity**
- Weak duality **always hold**
- Produces lower bounds on the problem at low cost

Are there cases where $d^\star = p^\star$?

Strong duality

When $d^\star = p^\star$ for a class of problems: **strong duality**.

- Because $d^\star$ is a lower bound on the optimal value $p^\star$, if both are equal for some $(\mathbf{x}, \lambda, \mu)$, the current point must be optimal
- For most convex problems, we have **strong duality**. (see next slide)
- The difference $p^\star - d^\star$ is called the **duality gap**
- The **duality gap** measures how optimal the current solution $(\mathbf{x}, \lambda, \mu)$ is.

Slater's conditions

Example of sufficient conditions for **strong duality**:

- **Slater's conditions**. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \le 0, \quad i = 1, \ldots, m \\ & A\mathbf{x} = \mathbf{b}, \quad i = 1, \ldots, p \end{array}$$

  where all the $f_i(\mathbf{x})$ are **convex** and assume that:

$$\text{there exists } \mathbf{x} \in \mathcal{D}: \; f_i(\mathbf{x}) < 0, \; A\mathbf{x} = \mathbf{b}, \quad i = 1, \ldots, m$$

  in other words there is a **strictly feasible point**, then strong duality holds.

- Many other versions exist. . .
- Often easy to check.
- Let's see for linear programs.

Duality: the simple example of linear programming

Duality: linear programming

- Take a **linear program** in standard form:

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T \mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \,(\text{ which is equivalent to } -\mathbf{x} \leq 0) \end{array}$$

- We can form the **Lagrangian**:

$$L(\mathbf{x}, \lambda, \mu) = \mathbf{c}^T \mathbf{x} - \lambda^T \mathbf{x} + \mu^T (A\mathbf{x} - \mathbf{b})$$

- and the **Lagrange dual function**:

$$\begin{aligned} g(\lambda, \mu) &= \inf_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) \\ &= \inf_{\mathbf{x}} \mathbf{c}^T \mathbf{x} - \lambda^T \mathbf{x} + \mu^T (A\mathbf{x} - b) \end{aligned}$$

Duality: linear programming

- For linear programs, the Lagrange dual function can be computed **explicitly**:

$$
\begin{aligned}
g(\lambda, \mu) &= \inf_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \; - \lambda^T\mathbf{x} + \mu^T(A\mathbf{x} - b) \\
&= \inf_{\mathbf{x}} (c - \lambda + A^T\mu)^T\mathbf{x} - \mathbf{b}^T\mu
\end{aligned}
$$

- This is either $-\mathbf{b}^T\mu$ or $-\infty$, so we finally get:

$$
g(\lambda, \mu) = \begin{cases} -\mathbf{b}^T\mu & \text{if } c - \lambda + A^T\mu = 0 \\ -\infty & \text{otherwise} \end{cases}
$$

- If $g(\lambda, \mu) = -\infty$ we say that (λ, μ) are outside the domain of the dual.

Duality: linear programming

- With $g(\lambda, \mu)$ given by:

$$g(\lambda, \mu) = \begin{cases} -\mathbf{b}^T \mu & \text{if } c - \lambda + A^T \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- we can write the dual program as:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \mu) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

- which is again, writing the domain explicitly:

$$\begin{array}{ll} \text{maximize} & -\mathbf{b}^T \mu \\ \text{subject to} & c - \lambda + A^T \mu = 0 \\ & \lambda \geq 0 \end{array}$$

Duality: linear programming

- After simplification:

$$\begin{cases} c - \lambda + A^T\mu = 0 \\ \lambda \geq 0 \end{cases} \iff c + A^T\mu \geq 0$$

- we conclude that the dual of the linear program:

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \qquad \textbf{(primal)}$$

- is given by:

$$\begin{array}{ll} \text{maximize} & -\mathbf{b}^T\mu \\ \text{subject to} & -A^T\mu \leq c \end{array} \qquad \textbf{(dual)}$$

- equivalently:

$$\begin{array}{ll} \text{maximize} & \mathbf{b}^T\mu \\ \text{subject to} & A^T\mu \leq c \end{array}$$

Dual Linear Program

Up to now, what have we introduced?

- A vector of parameters $\mu \in \mathbf{R}^m$, **one coordinate by constraint**.
- For **any** μ and any feasible $\mathbf{x}$ of the primal = a lower bound on the primal.
- For **some** μ the lower bound is $-\infty$, not useful.
- The **dual problem** computes the **biggest** lower bound.
- We discard values of μ which give $-\infty$ lower bounds.
- This the way **dual constraints** are defined.
- The **dual** is **another linear program** in dimensions $\mathbf{R}^{n \times m}$, that is
  - n constraints,
  - m variables.

From Primal to Dual for general LP's

- Some notations: for $A \in \mathbf{R}^{m \times n}$ we write
  - $\mathbf{a}_j$ for the n column vectors
  - $\mathbf{A}_i$ for the m row vectors of A.
- Following a similar reasoning we can flip from primal to dual changing
  - the constraints linear relationships A,
  - the constraints constants $\mathbf{b}$,
  - the constraints directions $(\leq, \geq, =)$
  - non-negativity conditions,
  - the objective

minimize	$\mathbf{c}^T\mathbf{x}$		maximize	$\mu^T\mathbf{b}$	
subject to	$\mathbf{A}_i^T\mathbf{x} \geq b_i,$	$i \in M_1$	subject to	$\mu_i \geq 0$	$i \in M_1$
	$\mathbf{A}_i^T\mathbf{x} \leq b_i,$	$i \in M_2$		$\mu_i \leq 0$	$i \in M_2$
	$\mathbf{A}_i^T\mathbf{x} = b_i,$	$i \in M_3$		μ_i free	$i \in M_3$
	$x_j \geq 0$	$j \in N_1$		$\mu^T\mathbf{a}_j \leq c_j$	$j \in N_1$
	$x_j \leq 0$	$j \in N_1$		$\mu^T\mathbf{a}_j \geq c_j$	$j \in N_2$
	x_j free	$j \in N_1$		$\mu^T\mathbf{a}_j = c_j$	$j \in N_3$

(1)

Dual Linear Program

- In summary, for any kind of constraint,

primal	minimize	maximize	**dual**
constraints	$\geq b_i$ $\leq b_i$ $= b_i$	≥ 0 ≤ 0 free	variables
variables	≥ 0 ≤ 0 free	$\leq c_j$ $\geq c_j$ $= c_j$	constraints

- For simple cases and in matrix form,

minimize subject to	$\mathbf{c}^T\mathbf{x}$ $A\mathbf{x} = \mathbf{b}$ $\mathbf{x} \geq 0$	$\Rightarrow$	maximize subject to	$\mathbf{b}^T\mu$ $A^T\mu \leq c$
minimize subject to	$\mathbf{c}^T\mathbf{x}$ $A\mathbf{x} \geq \mathbf{b}$	$\Rightarrow$	maximize subject to	$\mathbf{b}^T\mu$ $A^T\mu = c$ $\mu \geq 0$