CS5489
Lecture 4.1: Duality

Kede Ma

City University of Hong Kong (Dongguan)

Slide template by courtesy of Benjamin M. Marlin
Part of slides are borrowed from Stephen P. Boyd

Outline

Convex Set

- **Line segment** between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$: all points

$$\mathbf{x} = \alpha\mathbf{x}^{(1)} + (1-\alpha)\mathbf{x}^{(2)}$$

  with $0 \leq \alpha \leq 1$

- **Convex set**: contains line segment between any two points in the set

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathcal{X},\ 0 \leq \alpha \leq 1 \implies \alpha\mathbf{x}^{(1)} + (1-\alpha)\mathbf{x}^{(2)} \in \mathcal{X}$$

Convex Function

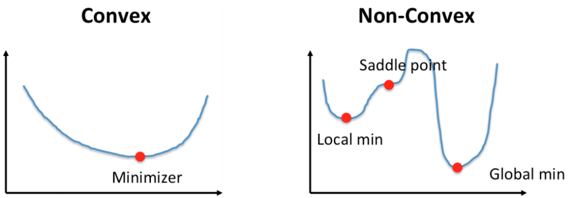
- $f : \mathbb{R}^N \mapsto \mathbb{R}$ is convex if $\mathrm{dom}(f)$ is a convex set and

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$, $0 \leq \alpha \leq 1$

  - This is also called Jensen's inequality

- f is concave if $-f$ is convex



First-Order Condition

- f is **differentiable** if the gradient

$$\nabla f(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_N}\right]^T$$

  exists at each $\mathbf{x} \in \mathrm{dom}(f)$

- **1st-order condition**: differentiable f with convex domain is convex iff

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}), \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$$

- First-order approximation of f is global underestimator

Second-Order Condition

- f is **twice differentiable** if the Hessian $\nabla^2 f(\mathbf{x}) \in \mathbb{S}^N$

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

  exists at each $\mathbf{x} \in \mathrm{dom}(f)$

- **2nd-order condition**: twice differentiable f with convex domain is convex iff

$$\nabla^2 f(\mathbf{x}) \succeq 0$$

  - This means that the Hessian matrix is positive semidefinite

$$\mathbf{A} \succeq 0 \Longleftrightarrow \forall \mathbf{x}, \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$$

Outline

Optimization Problem in Standard Form

- $$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \ i = 1, \ldots, r \\ & h_i(\mathbf{x}) = 0, \ i = 1, \ldots, s \end{aligned}$$
  - $\mathbf{x} \in \mathbb{R}^N$ is the optimization variable
  - $f_0 : \mathbb{R}^N \mapsto \mathbb{R}$ is the objective or cost function
  - $f_i : \mathbb{R}^N \mapsto \mathbb{R}, i = 1, \ldots, r$, are the inequality constraint functions
  - $h_i : \mathbb{R}^N \mapsto \mathbb{R}, i = 1, \ldots, s$ are the equality constraint functions
- **Optimal value**:

$$p^\star = \min\{f_0(\mathbf{x}) | f_i(\mathbf{x}) \leq 0, i = 1, \ldots, r, h_i(\mathbf{x}) = 0, \ i = 1, \ldots, s\}$$

  - $p^\star = \infty$ if problem is infeasible (no $\mathbf{x}$ satisfies the constraints)
  - $p^\star = -\infty$ if the problem is unbounded below

Optimal and Locally Optimal Points

- $\mathbf{x}$ is **feasible** if $\mathbf{x} \in \text{dom}(f_0)$ and it satisfies the constraints
- A feasible $\mathbf{x}$ is optimal if $f_0(\mathbf{x}) = p^\star$
- $\mathbf{x}$ is locally optimal if there is an $R > 0$ such that $\mathbf{x}$ is optimal for

$$\begin{aligned} \underset{\mathbf{z}}{\text{minimize}} \quad & f_0(\mathbf{z}) \\ \text{subject to} \quad & f_i(\mathbf{z}) \leq 0, \ i = 1, \ldots, r \\ & h_i(\mathbf{z}) = 0, \ i = 1, \ldots, s \\ & \|\mathbf{z} - \mathbf{x}\|_2 \leq R \end{aligned}$$

- Examples:
  - $f_0(x) = -\log x$, $\text{dom}(f_0) = \mathbb{R}_{++}$: $p^\star = -\infty$
  - $f_0(x) = x \log x$, $\text{dom}(f_0) = \mathbb{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
  - $f_0(x) = x^3 - 3x$: $p^\star = -\infty$, locally optimum at $x = 1$

Feasibility Problem

- $$\begin{array}{ll} \text{find} & \mathbf{x} \\ \text{subject to} & f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, r \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, s \end{array}$$

  can be considered a special case of the general problem with $f_0(\mathbf{x}) = 0$:

  $$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & 0 \\ \text{subject to} & f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, r \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, s \end{array}$$

  - $p^\star = 0$ if constraints are feasible; any feasible $\mathbf{x}$ is optimal
  - $p^\star = \infty$ if constraints are infeasible

Convex Optimization Problem in Standard Form

- $$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \ i = 1, \ldots, r \\ & \mathbf{a}_i^T \mathbf{x} = b_i, \ i = 1, \ldots, s \end{aligned}$$
  - $f_0, f_1, \ldots, f_r$ are convex
  - Equality constraints are affine
- Often written as
  $$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \ i = 1, \ldots, r \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$
  - Important property: feasible set of a convex optimization problem is convex

Example

- $$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{subject to} \quad & f_1(\mathbf{x}) = x_1/(1 + x_2^2) \leq 0 \\ & h_1(\mathbf{x}) = (x_1 + x_2)^2 = 0 \end{aligned}$$
  - f_0 is convex
  - Feasible set $\{(x_1, x_2) | x_1 = -x_2 \leq 0\}$ is convex
  - Not a convex problem (in standard form): f_1 is not convex, h_1 is not affine
- Writes it as a convex problem in standard form

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{subject to} \quad & f_1(\mathbf{x}) = x_1 \leq 0 \\ & h_1(\mathbf{x}) = x_1 + x_2 = 0 \end{aligned}$$

Local and Global Optima in Convex Optimization

- Any locally optimal point of is (globally) optimal

Proof. Suppose $\mathbf{x}$ is locally optimal and $\mathbf{y}$ is optimal with $f_0(\mathbf{y}) < f_0(\mathbf{x})$. Locally optimal $\mathbf{x}$ means there is an $R > 0$ such that

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

Consider $\mathbf{z} = \alpha\mathbf{y} + (1 - \alpha)\mathbf{x}$, with $\alpha = R/(2\|\mathbf{y} - \mathbf{x}\|_2)$

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

$$f_0(\mathbf{z}) \leq \alpha f_0(\mathbf{y}) + (1 - \alpha) f_0(\mathbf{x}) < f_0(\mathbf{x})$$

which contradicts our assumption that $\mathbf{x}$ is locally optimal

Outline

Lagrangian

- Standard form problem (not necessarily convex):

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \; i = 1, \ldots, r \\ & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, s \end{aligned}$$

  - Variable $\mathbf{x} \in \mathbb{R}^N$, domain $\mathcal{X}$, optimal value $p^\star$
- **Lagrangian**: $L : \mathbb{R}^N \times \mathbb{R}^r \times \mathbb{R}^s \mapsto \mathbb{R}$, with $\operatorname{dom}(L) = \mathcal{X} \times \mathbb{R}^r \times \mathbb{R}^s$

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{s} \nu_i h_i(\mathbf{x})$$

- Weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(\mathbf{x}) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(\mathbf{x}) = 0$

Lagrange Dual Function

- **Lagrange dual function**: $g : \mathbb{R}^r \times \mathbb{R}^s \mapsto \mathbb{R}$,

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathcal{X}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{s} \nu_i h_i(\mathbf{x}) \right)$$

  g is concave and can be $-\infty$ for some $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$

- **Lower bound property**: if $\boldsymbol{\lambda} \geq 0$, then $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$

  Proof. If $\tilde{\mathbf{x}}$ is feasible and $\boldsymbol{\lambda} \geq 0$, then

$$f_0(\tilde{\mathbf{x}}) \geq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\mathbf{x}\in\mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

  Minimizing over all feasible $\tilde{\mathbf{x}}$ gives $p^\star \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$

Least-Norm Solution of Linear Equations

- $$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{x}^T\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array}$$
- Dual function:
  - Lagrangian is $L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$
  - To minimize L over $\mathbf{x}$, set gradient equal to zero:

    $$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \quad \Longrightarrow \quad \mathbf{x} = -(1/2)\mathbf{A}^T\boldsymbol{\nu}$$
  - Plug in L to obtain g:

    $$g(\boldsymbol{\nu}) = L((-1/2)\mathbf{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T\mathbf{A}\mathbf{A}^T\boldsymbol{\nu} - \mathbf{b}^T\boldsymbol{\nu}$$

    a concave function of $\boldsymbol{\nu}$
- Lower bound property: $p^\star \geq -\frac{1}{4}\boldsymbol{\nu}^T\mathbf{A}\mathbf{A}^T\boldsymbol{\nu} - \mathbf{b}^T\boldsymbol{\nu}$ for all $\boldsymbol{\nu}$

Standard Form Linear Programming

- $$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \succeq 0 \end{aligned}$$
- Dual function:
  - Lagrangian is
  $$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) - \boldsymbol{\lambda}^T\mathbf{x} \\ &= -\mathbf{b}^T\boldsymbol{\nu} + (\mathbf{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{c})^T\mathbf{x} \end{aligned}$$
  - L is affine in $\mathbf{x}$, hence:
  $$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^T\boldsymbol{\nu} & \mathbf{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{c} = 0 \\ -\infty & \text{otherwise} \end{cases}$$
  g is linear on $\{(\boldsymbol{\lambda}, \boldsymbol{\nu}) | \mathbf{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{c} = 0\}$, hence concave
- Lower bound property: $p^\star \geq -\mathbf{b}^T\boldsymbol{\nu}$ if $\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c} \succeq 0$

The Dual Problem

- **Lagrange dual problem**:

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \quad & g(\boldsymbol{\lambda},\boldsymbol{\nu}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned}$$

  - $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq 0$, $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathrm{dom}(g)$
  - Finds best lower bound $p^\star$, obtained from Lagrange dual problem
  - A convex optimization problem; optimal value denoted $d^\star$
- **Weak duality**: $d^\star \leq p^\star$
  - Always holds (for convex and nonconvex problems)
  - Can be used to find nontrivial lower bounds for difficult problems
- **Strong duality**: $d^\star = p^\star$
  - Does not hold in general
  - (Usually) holds for convex problems, *e.g.*, SVM

Complementary Slackness

- Assuming strong duality holds, $\mathbf{x}^\star$ is primal optimal, $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ is dual optimal

$$
\begin{aligned}
f_0(\mathbf{x}^\star) = g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i^\star f_i(\mathbf{x}) + \sum_{i=1}^{s} \nu_i^\star h_i(\mathbf{x}) \right) \\
&\leq f_0(\mathbf{x}^\star) + \sum_{i=1}^{r} \lambda_i^\star f_i(\mathbf{x}^\star) + \sum_{i=1}^{s} \nu_i^\star h_i(\mathbf{x}^\star) \\
&\leq f_0(\mathbf{x}^\star)
\end{aligned}
$$

- Hence, the two inequalities hold with equality
  - $\mathbf{x}^\star$ not only minimizes $f_0(\mathbf{x})$, but also minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$
  - $\lambda_i^\star f_i(\mathbf{x}^\star) = 0$ for $i = 1, \ldots, r$ (complementary slackness):

$$
\lambda_i^\star > 0 \Longrightarrow f_i(\mathbf{x}^\star) = 0, \qquad f_i(\mathbf{x}^\star) < 0 \Longrightarrow \lambda_i^\star = 0
$$

Karush-Kuhn-Tucker (KKT) Conditions

- The following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i)
  1. Primal constraints: $f_i(\mathbf{x}) \leq 0, i = 1, \ldots, r, h_i(\mathbf{x}) = 0, i = 1, \ldots, s$
  2. Dual constraints: $\boldsymbol{\lambda} \succeq 0$
  3. Complementary slackness: $\lambda_i f_i(\mathbf{x}) = 0, i = 1, \ldots, r$
  4. Gradient of Lagrangian with respect to $\mathbf{x}$ vanishes:

$$\nabla f_0(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i \nabla f_i(\mathbf{x}) + \sum_{i=1}^{s} \nu_i \nabla h_i(\mathbf{x}) = 0$$

- If strong duality holds and $\mathbf{x}$, $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions