Foundations of Machine Learning
Convex Optimization

Mehryar Mohri
Courant Institute and Google Research
mohri@cims.nyu.edu
Convex Optimization
Convexity

Definition: $X \subseteq \mathbb{R}^N$ is said to be convex if for any two points $x, y \in X$ the segment $[x, y]$ lies in $X$:

$$\{\alpha x + (1 - \alpha)y, 0 \leq \alpha \leq 1\} \subseteq X.$$ 

Definition: let $X$ be a convex set. A function $f : X \rightarrow \mathbb{R}$ is said to be convex if for all $x, y \in X$ and $\alpha \in [0, 1]$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

With a strict inequality, $f$ is said to be strictly convex. $f$ is said to be concave when $-f$ is convex.
Properties of Convex Functions

- **Theorem:** let $f$ be a differentiable function. Then, $f$ is convex iff $\text{dom}(f)$ is convex and
  \[
  \forall x, y \in \text{dom}(f), \ f(y) - f(x) \geq \nabla f(x) \cdot (y - x).
  \]

- **Theorem:** let $f$ be a twice differentiable function. Then, $f$ is convex iff its Hessian is positive semi-definite:
  \[
  \forall x \in \text{dom}(f), \ \nabla^2 f(x) \succeq 0.
  \]
Constrained Optimization Problem

Problem: Let $X \subseteq \mathbb{R}^N$ and $f, g_i : X \to \mathbb{R}, i \in [1, m]$. A constrained optimization problem has the form:

$$\min_{x \in X} f(x)$$

subject to: $g_i(x) \leq 0, i \in [1, m]$.

Definition: The Lagrange function or Lagrangian associated to this problem is the function defined by:

$$\forall x \in X, \forall \alpha \geq 0, L(x, \alpha) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x).$$

$\alpha_i$s are called Lagrange or dual variables.
Sufficient Condition

(Lagrange, 1797)

Theorem: Let P be a constrained optimization problem over $X = \mathbb{R}^N$. If $(x^*, \alpha^*)$ is a saddle point, that is $\forall x \in \mathbb{R}^N, \forall \alpha \geq 0$, $L(x^*, \alpha) \leq L(x^*, \alpha^*) \leq L(x, \alpha^*)$, then it is a solution of P.

Proof: By the first inequality,

$$\forall \alpha \geq 0, L(x^*, \alpha) \leq L(x^*, \alpha^*) \Rightarrow \forall \alpha \geq 0, \alpha \cdot g(x^*) \leq \alpha^* \cdot g(x^*)$$

(use $\alpha \to +\infty$ then $\alpha \to 0$) \Rightarrow $g(x^*) \leq 0 \land \alpha^* \cdot g(x^*) = 0$.

- In view of that, the second inequality gives

$$\forall x, L(x^*, \alpha^*) \leq L(x, \alpha^*) \Rightarrow \forall x, f(x^*) \leq f(x) + \alpha^* \cdot g(x).$$

Thus, for all $x$ such that $g(x) \leq 0$, $f(x^*) \leq f(x)$. 
Constraint Qualification

- **Definition:** Assume that \( \text{int} X \neq \emptyset \). Then, the following is the strong constraint qualification or Slater’s condition:

  \[
  \exists \overline{x} \in \text{int} X: g(\overline{x}) < 0.
  \]

- **Definition:** Assume that \( \text{int} X \neq \emptyset \). Then, the following is the weak constraint qualification or Slater’s condition:

  \[
  \exists \overline{x} \in \text{int} X: \forall i \in [1, m], \ (g_i(\overline{x}) < 0) \lor (g_i(\overline{x}) = 0 \land g_i \text{ affine}).
  \]
Necessary Conditions

Theorem: Assume that $f$ and $g_i, i \in [1, m]$, are convex functions and that Slater’s condition holds. If $x$ is a solution of the constrained optimization problem, then there exists $\alpha \geq 0$ such that $(x, \alpha)$ is a saddle point of the Lagrangian.

Theorem: Assume that $f$ and $g_i, i \in [1, m]$, are convex differentiable functions and that the weak Slater’s condition holds. If $x$ is a solution of the constrained optimization problem, then there exists $\alpha \geq 0$ such that $(x, \alpha)$ is a saddle point of the Lagrangian.
Kuhn-Tucker’s Theorem

(Karush 1939; Kuhn-Tucker, 1951)

Theorem: Assume that \( f, g_i : X \to \mathbb{R}, i \in [1, m] \) are convex and differentiable and that the constraints are qualified. Then \( \bar{x} \) is a solution of the constrained program iff there exist \( \bar{\alpha} \geq 0 \) such that:

\[
\nabla_x L(\bar{x}, \bar{\alpha}) = \nabla_x f(\bar{x}) + \bar{\alpha} \cdot \nabla_x g(\bar{x}) = 0
\]

\[
\nabla_\alpha L(\bar{x}, \bar{\alpha}) = g(\bar{x}) \leq 0
\]

\[
\bar{\alpha} \cdot g(\bar{x}) = \sum_{i=1}^{m} \bar{\alpha}_i g_i(\bar{x}) = 0.
\]

Note: Last two conditions equivalent to

\[
(g(\bar{x}) \leq 0) \land (\forall i \in [1, m], \bar{\alpha}_i g_i(\bar{x}) = 0).
\]

complementary conditions
• Since the constraints are qualified, if $\overline{x}$ is solution, then there exists $\overline{\alpha}$ such that $(\overline{x}, \overline{\alpha})$ is a saddle point. In that case, the three conditions are verified (for the 3rd condition see proof of sufficient condition slide).

• Conversely, assume that the conditions are verified. Then, for any $x$ such that $g(x) < 0$,

\[
f(x) - f(\overline{x}) \geq \nabla_x f(\overline{x}) \cdot (x - \overline{x})
\]

(convexity of $f$)

\[
= - \sum_{i=1}^{m} \overline{\alpha}_i \nabla_x g_i(\overline{x}) \cdot (x - \overline{x})
\]

(first condition)

\[
\geq - \sum_{i=1}^{m} \overline{\alpha}_i [g_i(x) - g_i(\overline{x})]
\]

(convexity of $g_i$s)

\[
= - \sum_{i=1}^{m} \overline{\alpha}_i g_i(x) \geq 0,
\]

(third condition)
Primal and Dual Problems

Primal problem:

$$\min_{x \in X} f(x)$$

subject to: $g(x) \leq 0$.

Dual problem:

$$\max \inf_{\alpha} \inf_{x \in X} L(x, \alpha)$$

subject to: $\alpha \geq 0$.

Equivalent problems when constraints qualified.