# Review: Convex Optimization

Hongwei Zhang

http://www.cs.wayne.edu/~hzhang

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Convexity

- A set $X \subset R^n$ is *convex* if for all x1, x2 $\in$ X and any $\lambda \in [0, 1]$, $\lambda x1 + (1-\lambda) x2 \in X$
- Given a convex set $X \subset R^n$, a function $f: X \rightarrow R$ is *convex* (respectively *concave*) if for all x1, x2 $\in$ X and any $\lambda \in [0, 1]$, $f(\lambda x1 + (1-\lambda)x2) \le$ (respectively $\ge$) $\lambda f(x1) + (1-\lambda)f(x2)$
  - Strict convex/concave if the inequality is strict for distinct x1 and x2
- <u>Theorem C.1</u>: A function $f: R \rightarrow R$ is convex iff. for every a<b<c

$$\frac{f(b)-f(a)}{b-a} \le \frac{f(c)-f(a)}{c-a}$$

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Local and global optima

- Consider the problem of minimizing a function $f:X \rightarrow R$
- Any $x \in X$ is a *feasible* solution
- An element $x^* \in X$ is a *global optimal* solution (or a *solution*) if $f(x^*) \leq f(x)$ for all $x \in X$
- An element $x' \in X$ is *locally optimal* if for some $\varepsilon>0$, $f(x') \leq f(x)$, for all $x \in \{x \in X: ||x-x'||<\varepsilon\}$

- A global optimal solution need not be unique, but must be locally optimal

- Theorem C.2: given a convex set X and a convex function f over X,
  - A local minimum of f over X is also a global minimum;
  - If f is strictly convex, then a local optimum is the unique global optimum.

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Primal problem

$$\min\ f(x)$$

$$s.t.$$

$$g_i(x) \le 0,\ 1 \le i \le m$$

$$x \in R^n$$

- We only consider the case where f and all $g_i$'s are convex and differentiable functions

# Theorem C.3:
# Karush-Kuhn-Tucker (KKT) conditions

$$\text{Given a feasible } x^* \in R^n, \text{ if } \exists \lambda \in R^m, \text{ with } \lambda \geq 0, \text{ s.t.}$$

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0, \text{ and}$$

$$\sum_{i=1}^{m} \lambda_i g_i(x^*) = 0,$$

← Complementary slackness condition

$$\textit{Then}, \ x^* \text{ is a global optimal solution for the primal problem.}$$

- If x* satisfies KKT conditions, it is called a *KKT point*
- λ are called *Lagrange multipliers* or *dual variables*
- In general, KKT conditions are not necessary
- A constraint is *binding or active* at point x* if the constraint is met with equality at the point; otherwise, it is *slack* at the point (i.e., met with strict inequality)

# Example C.1

$$\min\ (x_1-5)^2+(x_2-5)^2$$

$$s.t.$$

$$x_1^2+x_2^2-5\le 0$$

$$\frac{1}{2}x_1+x_2-2\le 0$$

$$-x_1\le 0$$

$$-x_2\le 0$$

$$x\in R^2$$

$$\nabla f(x)=\begin{pmatrix}2(x_1-5)\\2(x_2-5)\end{pmatrix}$$

$$\nabla g_1(x)=\begin{pmatrix}2x_1\\2x_2\end{pmatrix}$$

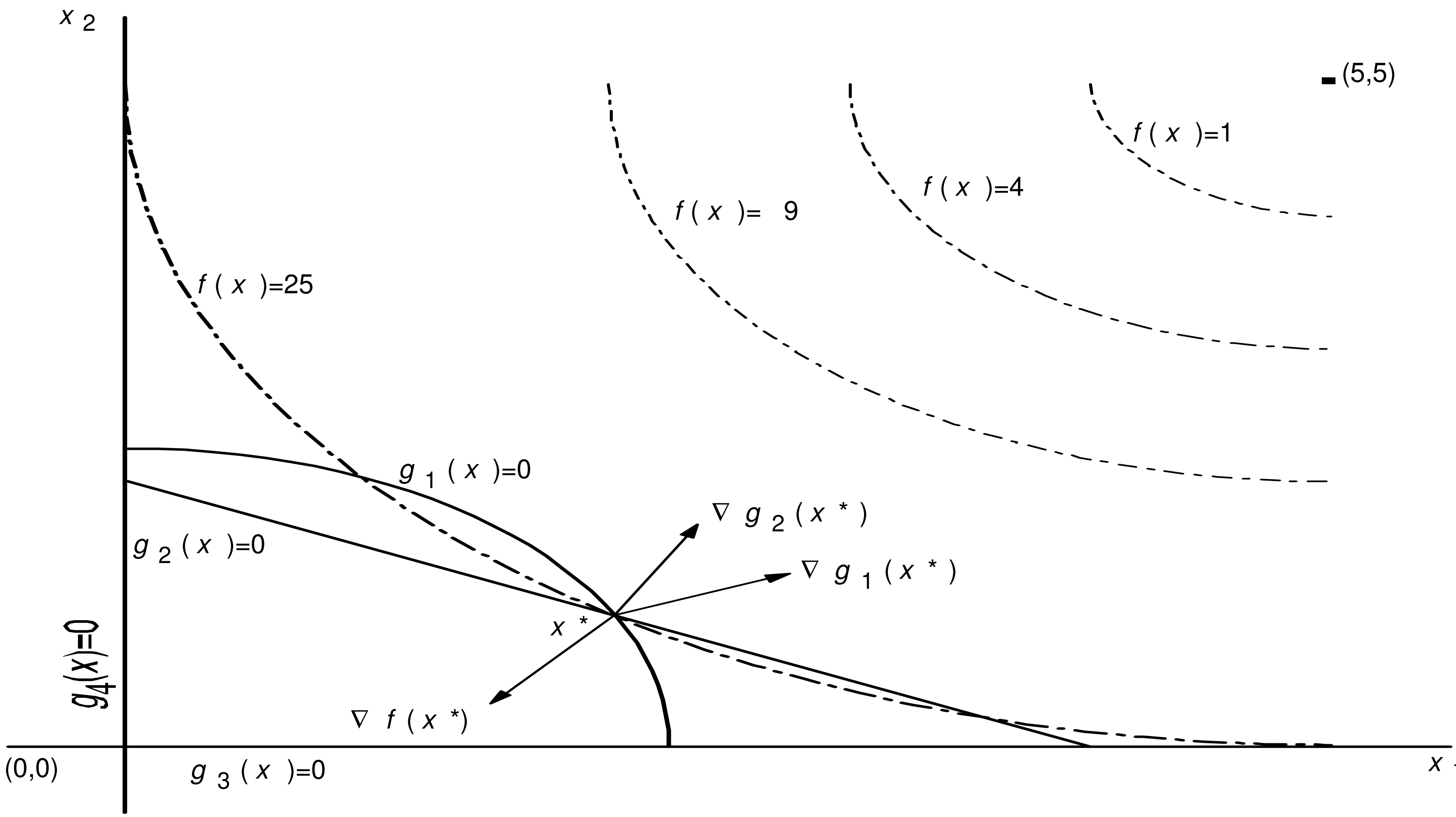
$$\nabla g_2(x)=\begin{pmatrix}1/2\\1\end{pmatrix}$$

$$\nabla g_3(x)=\begin{pmatrix}-1\\0\end{pmatrix}$$

$$\nabla g_4(x)=\begin{pmatrix}0\\-1\end{pmatrix}$$

$$\text{consider } x^*=\begin{pmatrix}2\\1\end{pmatrix}$$

- Geometry of example C.1:



# Theorem C.4: necessity and sufficiency of KKT conditions

- If $x^*$ is s.t. $\exists\ x \in R^n$ with $g_i(x) < 0$ if i-th constraint is active at $x^*$, then $x^*$ is optimal iff. KKT conditions hold at $x^*$

- If constraints are all linear, then $x^*$ is optimal iff. KKT conditions hold at $x^*$

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Linear program (LP)

- LP: objective function and constraints are all linear
- Primal LP:
  - Given $b \in R^n$, A is an m*n matrix with real elements, and $c \in R^m$

$$
\begin{aligned}
&\min b^T x \\
&s.t. \\
&Ax \geq c \\
&x \geq 0
\end{aligned}
$$

- Using Theorems C.3 and C.4, we can derive (p.810 of R0):
  - x* is optimal for the LP iff. $\exists \lambda \in R^m$, $\lambda \geq 0$, such that

$$\lambda^T (Ax^* - c) = 0 \text{ and } (b^T - \lambda^T A)x^* = 0$$

*Complementary slackness* conditions (related to "duality")

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Dual problem

- For the primal problem

$$\min\ f(x)$$

$$s.t.$$

$$g_i(x) \le 0,\ 1 \le i \le m$$

$$x \ge 0$$

$$x \in R^n$$

define for $\lambda \in R^m$, $\lambda \ge 0$,

$$\Theta(\lambda) = \inf\left\{x \ge 0, g_i(x) \le 0, 1 \le i \le m : f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)\right\},$$

$\Theta(\lambda)$ is called the *Lagrangian dual function*

- Then, the dual problem is

$$\max_{\lambda \ge 0} \Theta(\lambda)$$

  - By the definition of the dual function, solution to the dual problem *lower bound* the solution to the primal problem

- <u>Theorem C.5 (Strong Duality)</u>: If $\exists x \ge 0$ s.t. $g_i(x) < 0$, $1 \le i \le m$, then
  - The primal and dual problems have the same optimal values;
  - If the optimal value is finite, and if x* and λ* are solutions to the primal and dual problems, then

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0$$

# Dual LP

| $\underline{\underline{\text{Primal LP}}}:$ | | $\underline{\underline{\text{Dual LP}}}:$ |
|---|---|---|
| $\min\ b^T x$ | | $\max\ \lambda^T c$ |
| $s.t.$ | → | $s.t.$ |
| $Ax \geq c$ | | $\lambda^T A \leq b^T$ |
| $x \geq 0$ | | $\lambda \geq 0$ |

- Recall the *complementary slackness conditions for LP*:

$$\lambda^T (Ax^* - c) = 0 \text{ and } (b^T - \lambda^T A)x^* = 0$$

A slack in the primal constraint => corresponding dual variable must be 0;

A slack in the dual constraint => corresponding primal variable must be 0.

- From "Strong Duality Theorem": if the optimum value of the LP is finite, it is $(\lambda^*)^T c$

# LP with equality constraints

- H is an l*n metrix, and d∈ $R^l$

$$\underline{\underline{\text{Primal LP}}}: \quad \min b^T x \quad s.t. \quad Ax \geq c \quad Hx = d \quad x \geq 0$$

$$\underline{\underline{\text{Dual LP}}}: \quad \max \lambda^T c + \mu^T d \quad s.t. \quad \lambda^T A + \mu^T H \leq b^T \quad \lambda \geq 0 \quad \mu \text{ unrestricted}$$

- Complementary slackness conditions: x* is optimal for the LP iff. ∃λ∈ $R^m$, λ≥0, and μ∈ $R^l$ such that

$$\lambda^T (Ax^* - c) = 0 \text{ and } (b^T - (\lambda^T A + \mu^T H))x^* = 0$$

# Theorem C.6:

Given the primal and dual LPs (with equality constraints),

- Dual LP objective value lower bounds the primal LP objective value
- If the primal LP is infeasible, then the dual LP is unbounded; if the dual LP is infeasible, then the primal LP is unbounded.
- If the primal LP and the dual LP are both feasible, then they both have solutions and have the same optimal objective value; in this case, the complementary slackness conditions hold.

# Outline

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution

# Perturbed primal problem

$$\min f(x)$$

$$s.t.$$

$$g_i(x) \le c_i,\ 1 \le i \le m$$

$$x \ge 0$$

$$x \in R^n$$

- For LP (where Strong Duality holds): if the optimum value of the LP is finite, it is $(\lambda^*)^T c$ => sensitivity is provided by the optima dual variable $\lambda^*$: a small perturbation in $c_i$ leads to a proportional perturbation in the optimal value, and the proportionality factor is $\lambda_i^*$.

# Summary

- Convexity
- Local and global optima
- Karush-Kuhn-Tucker (KKT) conditions
- Linear programming
- Duality
- Sensitivity of the optimal solution