CME307/MS&E311 Optimization Model/Theory Summary

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Optimization Problems

- A set of decision variables, *x*, in vector or matrix form with dimension n or nxn
- A continuous and sometime differentiable objective function *f(x)*
- A feasible region where *x* can be in

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \boldsymbol{X} \end{array}$$

- One can smooth them by reformulation as constrained optimization:

$$\max \quad \min_i \{ f_i(x), i=1,\dots,n \} \rightarrow$$

$$\max \quad \alpha \qquad \text{s.t. } \alpha - f_i(x) \le 0, \text{ for } i=1,\dots,n$$

Function, Gradient Vector and Hessian Matrix

- A function f of x in R^n
- The <u>Gradient Vector</u> of f at x

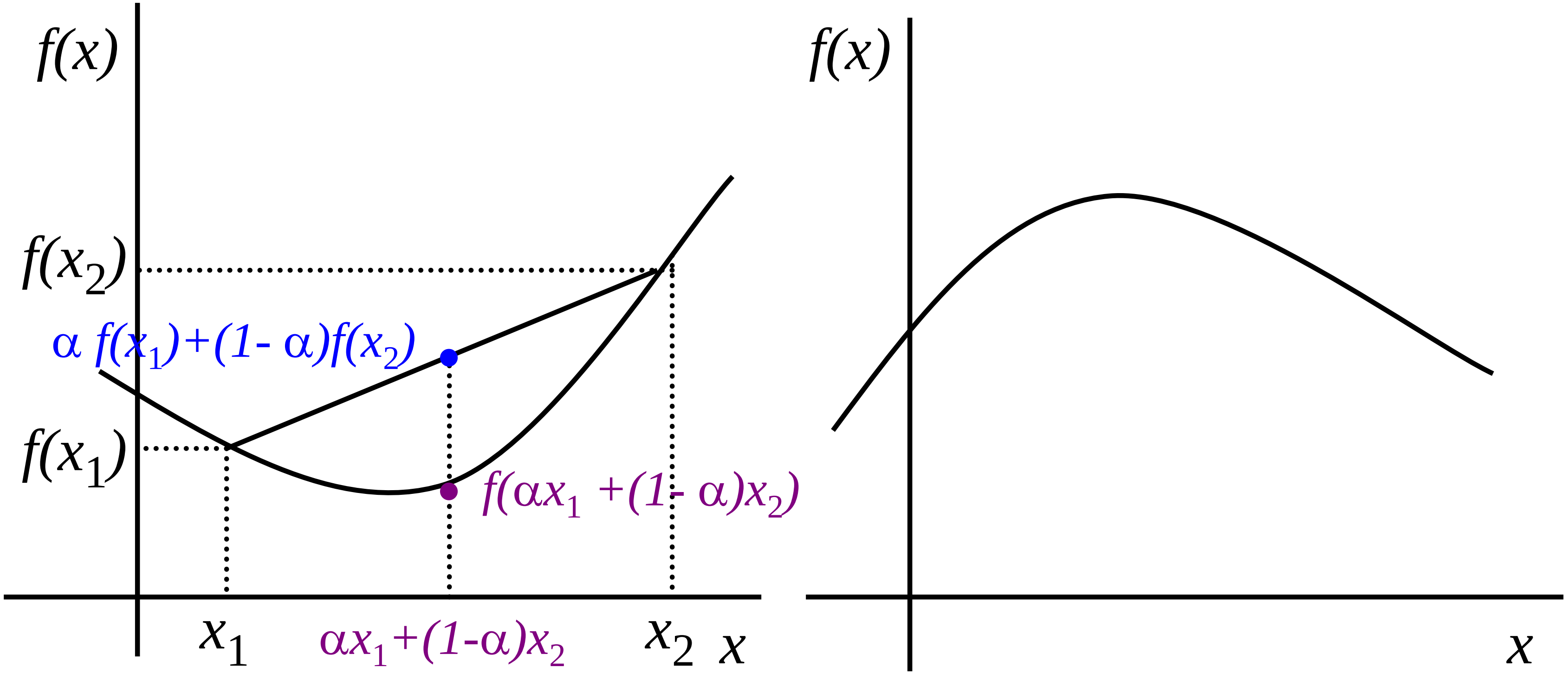
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \cdots\cdots\cdots \frac{\partial f}{\partial x_n} \right)$$

- The <u>Hessian Matrix</u> of f at x

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial f^2}{\partial x_1 \partial x_1} & \cdots\cdots\cdots & \frac{\partial f^2}{\partial x_1 \partial x_n} \\ & \cdots\cdots\cdots\cdots & \\ \frac{\partial f^2}{\partial x_n \partial x_1} & \cdots\cdots\cdots\cdots & \frac{\partial f^2}{\partial x_n \partial x_n} \end{pmatrix}$$

- **Taylor's Expansion Theorem**

Convex and Concave Functions



f(x) is a <u>convex function</u> if and only if for any given two points x_1 and x_2 in the function domain and for any constant $0 \le \alpha \le 1$

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$$

Strongly convex if $x_1 \neq x_2$, $f(0.5x_1 + 0.5x_2) < 0.5f(x_1) + 0.5f(x_2)$

More on Convex Functions

f(x) is a (stronly) convex function if and only if its Hessian matrix is (positive definite PD) positive semi-definite (PSD) in the domain of the function.

A symmetric matrix *Q* is PSD (or PD) if and only if $x^T Q x \geq$ *(or >) 0* for all $x \neq 0$.

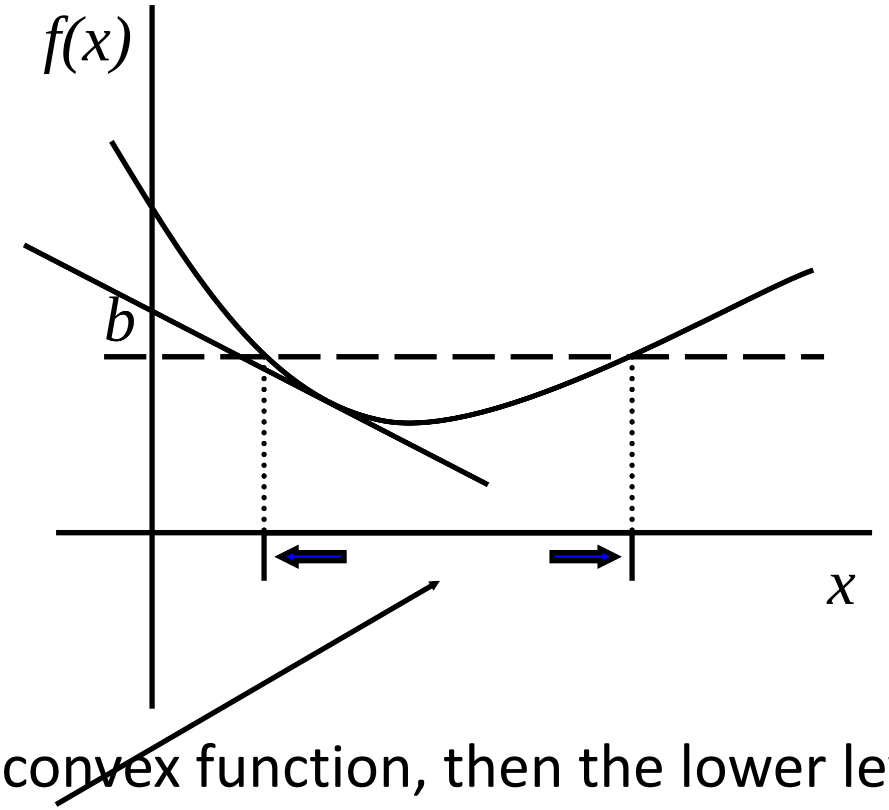
A 2x2 matrix is PSD (or PD) if and only if two diagonal entries and the determinant are nonnegative (or positive).

f(x) is a (strongly) concave function if -f(x) is a (strongly) convex function

Convex Sets

- A set is <u>convex</u> if every line segment connecting any two points in the set is contained entirely within the set
  - Ex - polyhedron
  - Ex - ball
- An <u>extreme point</u> of a convex set is any point that is not on any line segment connecting any other two distinct points of the set
- The intersection of convex sets is a convex set
- A set is closed if the limit of any convergent sequence of the set belongs to the set
- A set is compact if it is bounded and closed.

Convexity of Function and Level Set



If *f(x)* is a convex function, then the lower level set $\{x: f(x) \leq b\}$ is a convex set for any constant *b*.

The graph of a convex function lies above its tangent line (planes).
The Hessian matrix of a convex function is positive semi-definite.

Optimization Problem Classes

- <u>Unconstrained Optimization</u>
  - Convex or Nonconvex
- <u>Constrained Optimization</u>
  - Conic Linear Optimization/Programming (CLO/CLP)
  - Convex Constrained Optimization (CCO)
    - Feasible region/set is convex; objective general
  - Generally Constrained Optimization (GCO)
  - Convex Optimization (CO)
    - Minimize a convex function over a convex feasible set
    - Maximize a concave function over a convex feasible set
    - Changing variable/constraint representation may result CO

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \boldsymbol{X} \end{array}$$

Optimization Problem Forms

$$\begin{aligned} \min \quad & \boldsymbol{c} \bullet \boldsymbol{x} \\ \text{s.t.} \quad & A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}, \\ & \boldsymbol{x} \in \boldsymbol{K} \end{aligned}$$

Conic Linear Optimization (CLO)

A: an m x n matrix
***c*: objective coefficient**
***K*: a closed convex cone**

This is convex optimization

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0, \ \text{i}=1,\dots,\text{m} \\ & c_i(\boldsymbol{x}) \geq 0, \ \text{i}=1,\dots,\text{p} \end{aligned}$$

Generally Constrained Optimization (GCO)

Each function can be continuous, continuously differentiable (C^1), or twice continuously differentiable (C^2)

It is CCO if c_i are all concave, and h_i are all linear/affine functions. In addition, if *f* is convex, it is CO.

Why do we care about convex optimization?

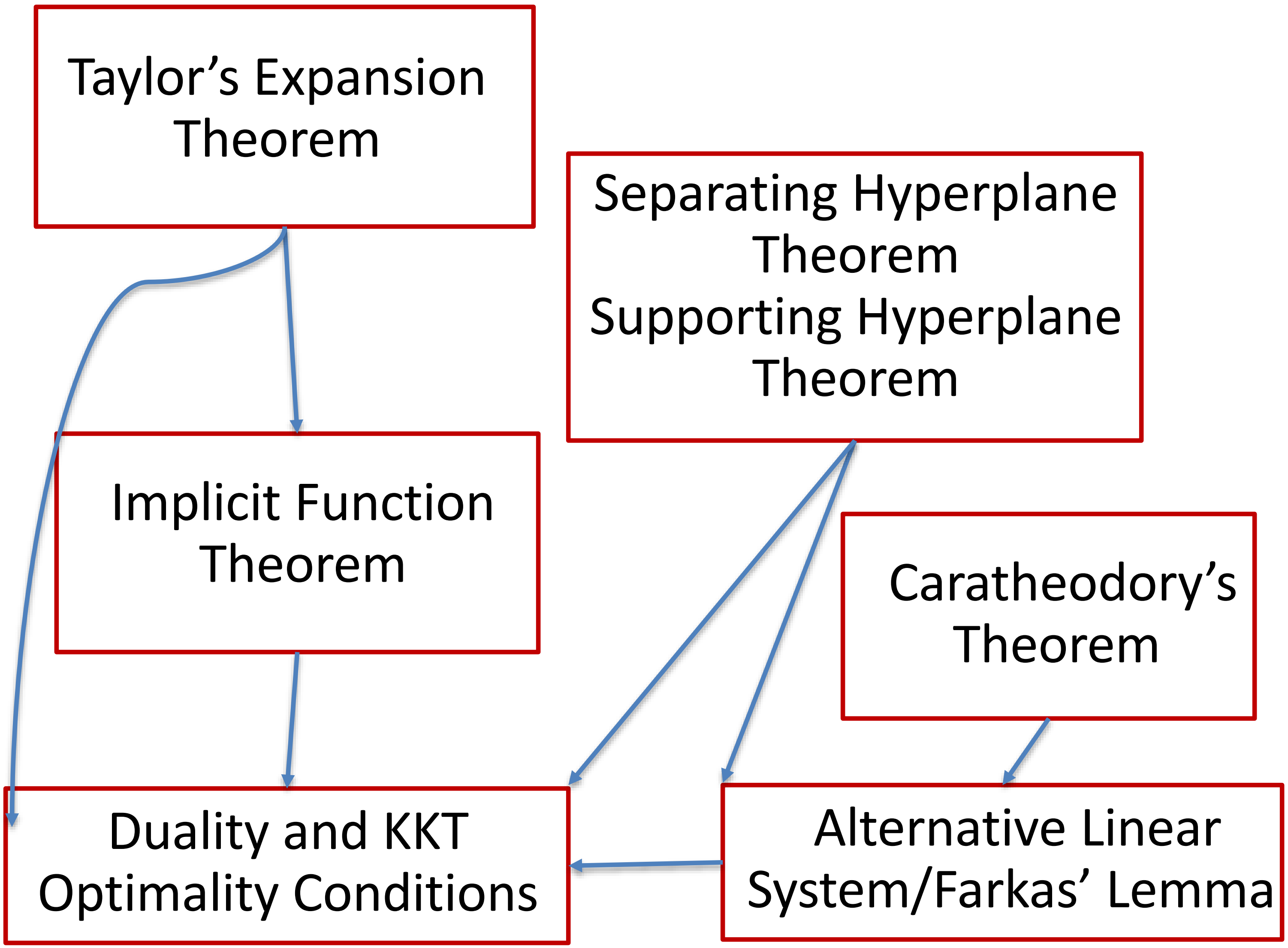
- It guarantees that every local optimizer is a global optimizer
- It guarantees that every (first-order) KKT (or stationary) point/solution is a global optimizer
- This is significant because all of our numerical optimization algorithms search/generate a KKT point/solution
- Sometime the problem can be "convexfied":

$$\min \quad c^T x, \quad \text{s.t.} \quad ||x||^2 = 1$$

$$\Updownarrow$$

$$\min \quad c^T x, \quad \text{s.t.} \quad ||x||^2 \leq 1$$

Optimization **Theory**: Mathematical Foundations



Theory: Feasibility Conditions

- Feasibility Conditions or Farkas' Lemmas are developed to characterize and certify feasibility or infeasibility of a feasible region
- Alternative Systems X and Y: X has a feasible solution if and only if Y has no feasible solution
  - X and Y cannot both have feasible solution
  - Exactly one of them has a feasible solution
- They can be viewed as special cases of Linear Programming primal and dual pairs

Alternative Systems and CLO Pairs I

$$A\mathbf{x} - \mathbf{b} = \mathbf{0},$$

$$\mathbf{x} \in \mathbf{K}$$

$$\mathbf{b}^T\mathbf{y} = 1 (>0)$$

$$A^T\mathbf{y} + \mathbf{s} = \mathbf{0},$$

$$\mathbf{s} \in \mathbf{K}^*$$

System X
A: an m x n matrix
***b*: m-dimension vector**
***K*: a closed convex cone**

System Y

K* **is the dual cone**

$$p^* = \min \quad \mathbf{0}^T\mathbf{x}$$

$$\text{s.t.} \quad A\mathbf{x} - \mathbf{b} = \mathbf{0},$$

$$\mathbf{x} \in \mathbf{K}$$

$$d^* = \max \quad \mathbf{b}^T\mathbf{y}$$

$$\text{s.t.} \quad A^T\mathbf{y} + \mathbf{s} = \mathbf{0},$$

$$\mathbf{s} \in \mathbf{K}^*$$

Alternative Systems and CLO Pairs II

$$\boldsymbol{c}^T\boldsymbol{x}=-1(<0)$$

$$A\boldsymbol{x} = 0,$$

$$\boldsymbol{x} \in K$$

System X

A: an m x n matrix

$\boldsymbol{c}$: n-dimension vector

K: a closed convex cone

$$A^T\boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} = 0,$$

$$\boldsymbol{s} \in K^*$$

System Y

K^* is the dual cone

$$\begin{array}{lll} p^*=\min & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & A\boldsymbol{x} = 0, \\ & \boldsymbol{x} \in K^* \end{array}$$

$$\begin{array}{lll} d^*=\max & \boldsymbol{0}^T\boldsymbol{y} \\ \text{s.t.} & A^T\boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} = 0, \\ & \boldsymbol{s} \in K \end{array}$$

Feasibility Test Machine

Is system **X** feasible?

Yes → Is system **Y** feasible?

“Not” under any circumstances

No → Is system **Y** feasible?

“Yes” under certain conditions of cone *K* and data matrix *A*:

a) *K* is a polyhedron cone, or

b) $A\mathbf{x}$ *or* $A^T\mathbf{y}$ has an interior solution

General Rules to Construct the CLO Dual

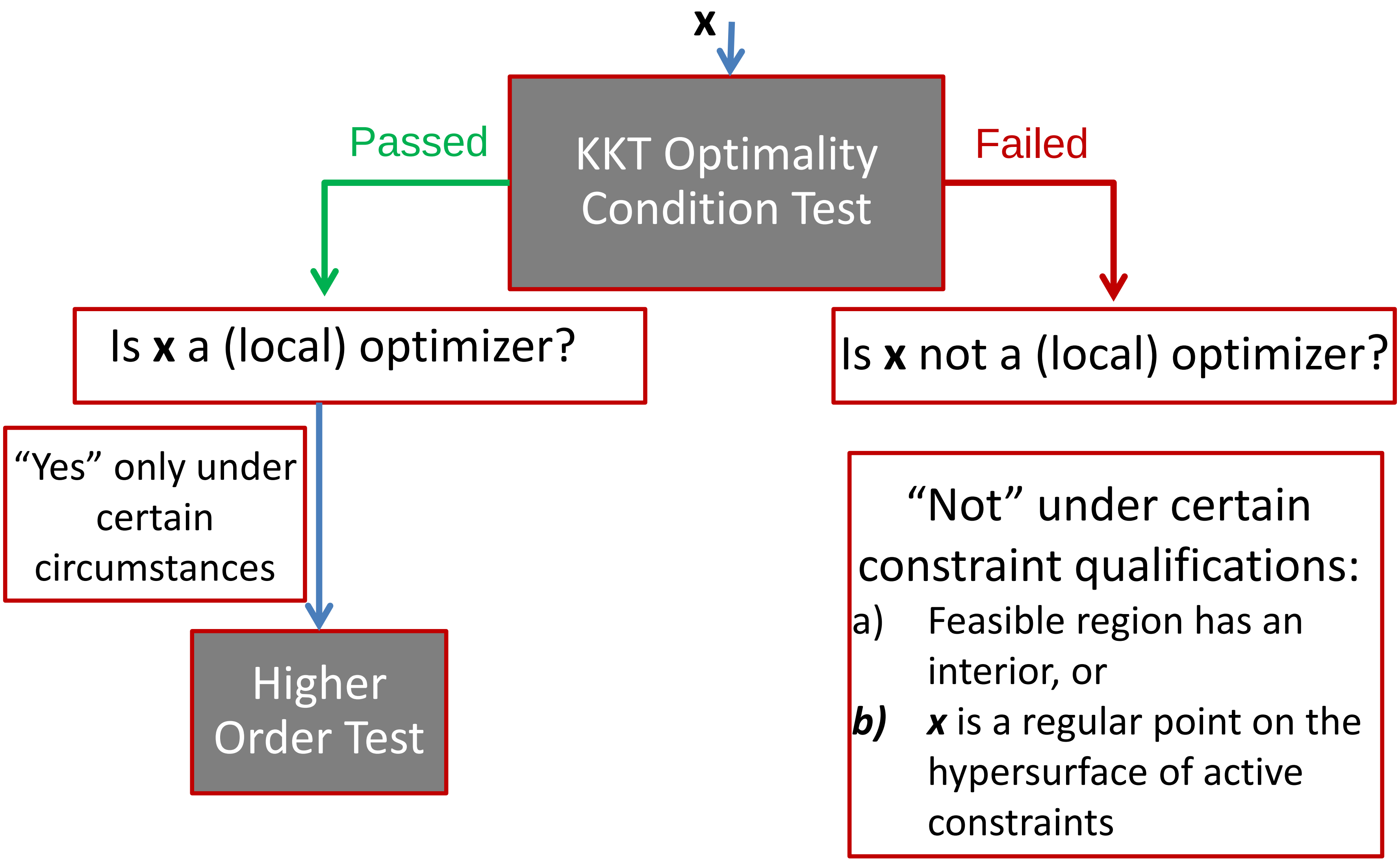
OBJ Vector/Matrix	RHS Vector/Matrix
RHS Vector/Matrix	OBJ Vector/matrix
A	A^T
Max model	Min model
$\boldsymbol{x}_j \geq_K 0$	jth block constraints $\geq_{K^*}$
$\boldsymbol{x}_j \leq_K 0$	jth block constraints $\leq_{K^*}$
$\boldsymbol{x}_j$ free	jth block constraints $=$
ith block constraints $\leq_K$	$\boldsymbol{y}_i \geq_{K^*} 0$
ith block constraints $\geq_K$	$\boldsymbol{y}_i \leq_{K^*} 0$
ith block constraints $=$	$\boldsymbol{y}_i$ free

The dual of the dual is the primal

Theory: Optimality Conditions

- Optimality (KKT) Conditions are developed to characterize and certify possible minimizers
  - Feasibility of original variables
  - Optimality conditions consist of original variables and Lagrange multipliers
  - Zero-order, First-order, Second-order, necessary, sufficient
- They may not lead directly to a very efficient algorithm for solving problems, but they do have a number of benefits:
  - They give insight into what optimal solutions look like
  - They provide a way to set up and solve small problems
  - They provide a method to check solutions to large problems
  - The Lagrange multipliers can be seen as sensitivities of the constraints
- A minimizers may not satisfy optimality conditions unless certain *constraint qualifications* hold.

KKT Optimality Condition Test Machine



0-Order Condition: Duality Theorems for CLO

$$p^* = \min \quad \boldsymbol{c}^T\boldsymbol{x}$$
$$\text{s.t.} \quad A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0},$$
$$\boldsymbol{x} \in K$$

Primal Problem
A: an m x n matrix
***c*: objective coefficient**
***K*: a closed convex cone**

Weak Duality Theorem

$$d^* = \max \quad \boldsymbol{b}^T\boldsymbol{y}$$
$$\text{s.t.} \quad A^T\boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} = \boldsymbol{0},$$
$$\boldsymbol{s} \in K^*$$

Dual Problem

***K** is the dual cone**

0-Order Condition: $p^* = d^*$

Sufficient!

Strong Duality Theorem: They must equal?

"Yes" under certain conditions of cone *K* and data matrix ***A,b,c***:

a) *K* is a polyhedron cone, or
b) *either one* has an interior feasible solution (but solution may not be attainable if K is not polyhedron)

The Lagrange Function of GCO

min $f(\boldsymbol{x})$

s.t. $c_i(\boldsymbol{x})\ (\leq, =, \geq)\ 0$, i=1,...,m

Restriction on multipliers y_i,

$y_i\ (\leq, \text{"free"}, \geq)\ 0$, $i=1,\ldots,m$

The Largrange Function

$$L(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) - \textstyle\sum_i y_i c_i(\boldsymbol{x})$$

The Lagrange function can be interpreted as a "penalized" aggregated objective function:

y_i free: can be penalized either way

$y_i \geq 0$ for "≥ 0" constraint: would be penalized only when $c_i(\boldsymbol{x}) \leq 0$

$y_i \leq 0$ for "≤ 0" constraint: would be penalized only when $c_i(\boldsymbol{x}) \geq 0$

$y_i = 0$: no penalty if inequality constraint is strictly satisfied, which leads to complementarity.

The Lagrangian Duality for GCO

$$p^* = \min \quad f(\boldsymbol{x})$$
$$\text{s.t.} \quad c_i(\boldsymbol{x}) \ (\geq, =, \leq) \ 0, \ i=1,\ldots,m$$

Weak Duality Theorem
$P^* \geq d^*$

Let $\phi(\boldsymbol{y}) = \inf_x L(\boldsymbol{x},\boldsymbol{y})$

The Lagrangian dual and CLP dual are equivalent for Conic Linear Optimization

Strong Duality Theorem
They must equal?

$$d^* = \max \quad \phi(\boldsymbol{y})$$
$$\text{s.t.} \quad y_i \ (\leq, \text{"free"}, \geq) \ 0, \ i=1,\ldots,m$$

Not necessarily!

0-Order Condition: $p^* = d^*$

Sufficient!

The Farkas' Lemma for General Constraint System

$$p^* = \min \quad 0^T \boldsymbol{x}$$
$$\text{s.t.} \quad c_i(\boldsymbol{x}) \ (\geq, =, \leq)\ 0,\ i=1,\ldots,m$$

Weak Duality Theorem
$P^* \geq d^*$

Let $\phi(\boldsymbol{y}) = \inf_x L(\boldsymbol{x}, \boldsymbol{y})$

$$\max \quad \phi(\boldsymbol{y})$$
$$\text{s.t.} \quad y_i\ (\leq, \text{"free"}, \geq)\ 0,\ i=1,\ldots,m$$

Not necessarily!

If there exists $\boldsymbol{y}$ such that $\phi(\boldsymbol{y})>0$, then GCS is infeasible

Sufficient!

General Rules to Construct the Dual

$\min f(\mathbf{x})$

$c_i(\mathbf{x})$ $(\geq, =, \leq)$ 0, i=1,…,m (ODC)

← **Primal**

Multiplier Sign Conditions (MSC)

y_i (≥,"free",≤) $0, i=1,\ldots,m$

← **Constraints in the Dual**

Lagrange Derivative Conditions (LDC)

$\partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$, for all $j=1,\ldots,n$.

← If no ***x*** in the equation, set it as an equality constraint in the dual; otherwise, express ***x*** in terms of ***y*** and replace ***x*** in the Lagrange function, which becomes the Dual objective.

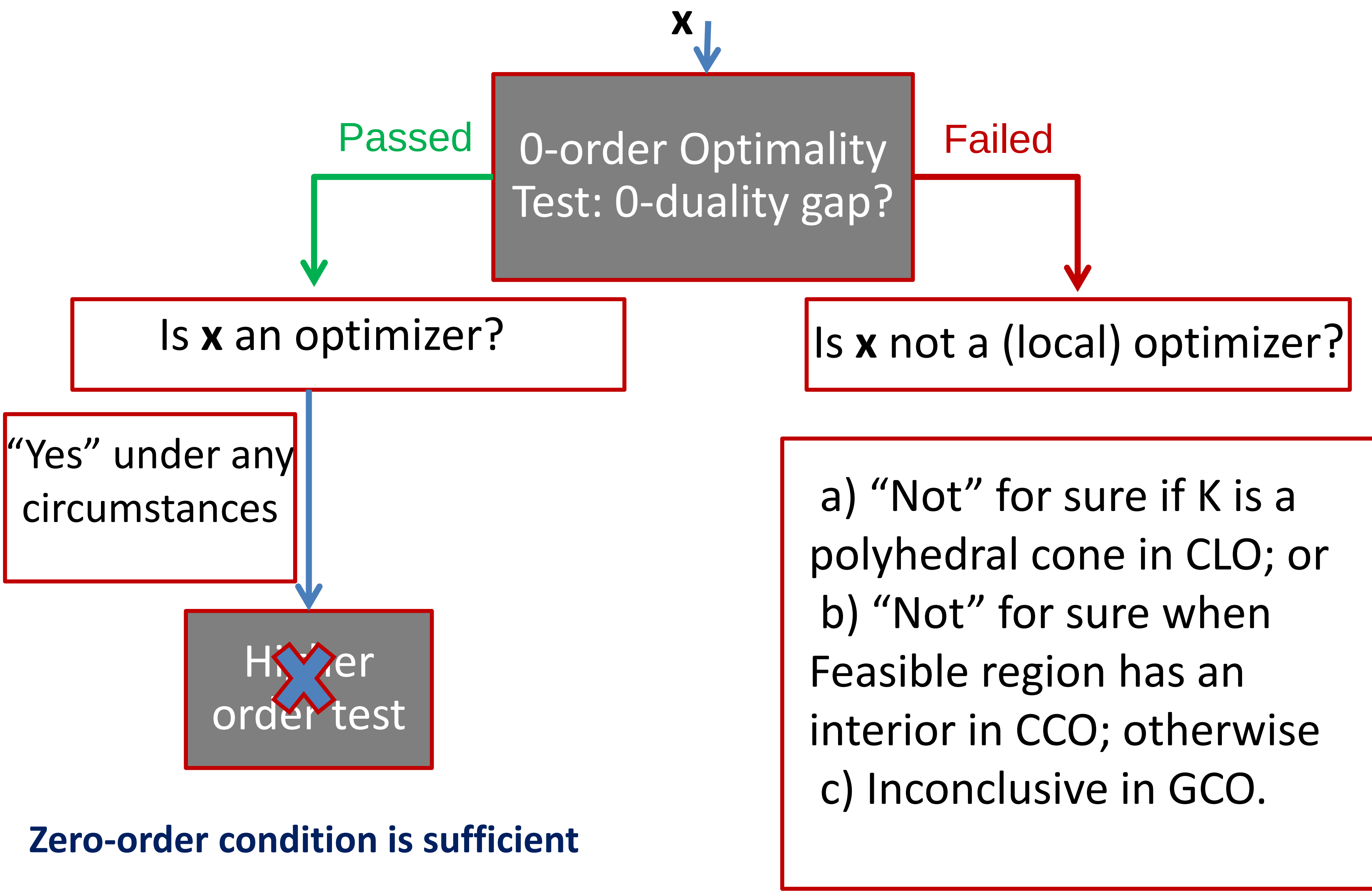
Warning: this may be difficult to do in general!

Complementarity Slackness Condition (CSC)

$y_i c_i(\mathbf{x}) = 0$, for each inequality constraint i. ✘

Not needed for construct Dual

Zero-Order Optimality Test for CLO and GCO

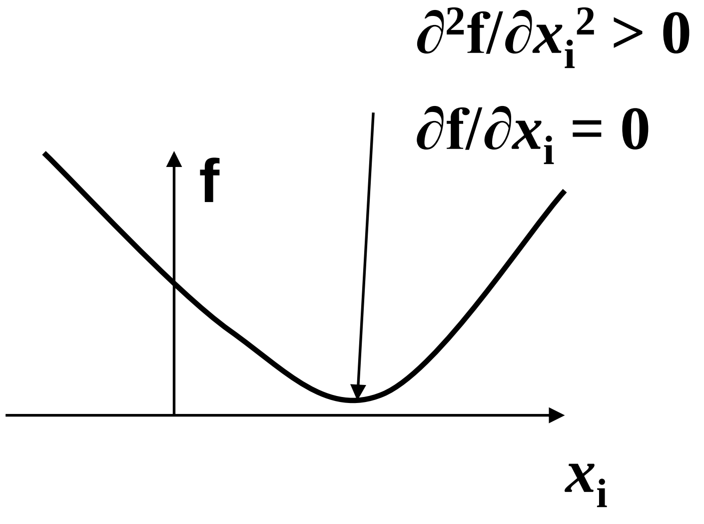


Zero-order condition is sufficient

1 and 2-order Conditions: Unconstrained

- Problem:
  - **Minimize *f(x)*, where *x* is a vector that could have any values, positive or negative**
- First Order Necessary Condition (min or max):
  - **$\nabla f(x) = 0$ ($\partial f/\partial x_i = 0$ for all i) is the first order necessary condition for optimization**
- Second Order Necessary Condition:
  - **$\nabla^2 f(x)$ is positive semidefinite (PSD)**
    - **[$d^T \nabla^2 f(x) d \geq 0$ for all d]**
- Second Order Sufficient Condition

  (Given FONC satisfied)
  - **$\nabla^2 f(x)$ is positive definite (PD)**
    - **[$d^T \nabla^2 f(x) d > 0$ for all $d \neq 0$]**



The First-Order Necessary Conditions for GCO

Original Decision-Var Constraints (ODC)

$c_i(\boldsymbol{x})\ (\geq,=,\leq)\ 0$, i=1,...,m

Multiplier Sign Condition (MSC)

$y_i\ (\geq,\text{"free"},\leq)\ 0,\ i=1,\ldots,m$

Lagrange Derivative Condition (LDC)

$\partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$, for all $j=1,\ldots,n$.

Complementary Slackness Condition (CSC)

$y_i\, c_i(\mathbf{x}) = 0$, for each inequality constraint i.

For maximization, just flip the sign of multipliers, and every condition remains the same.

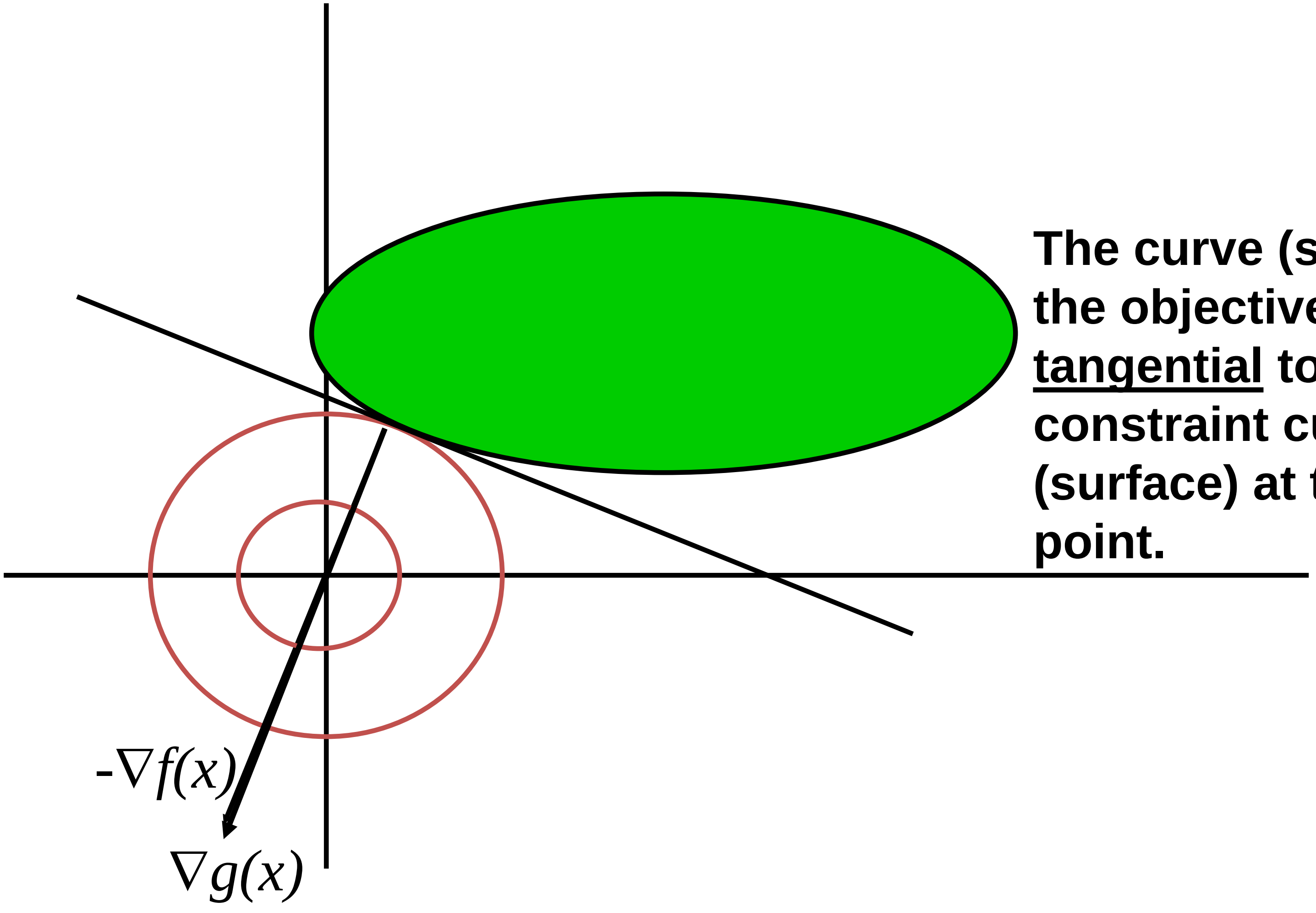
Short cut in dealing

ODC: $x_j \geq 0$

LDC: $\partial L(\mathbf{x},\mathbf{y})/\partial x_j \geq 0$

CSC: $x_j \partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$

Example: KKT Conditions

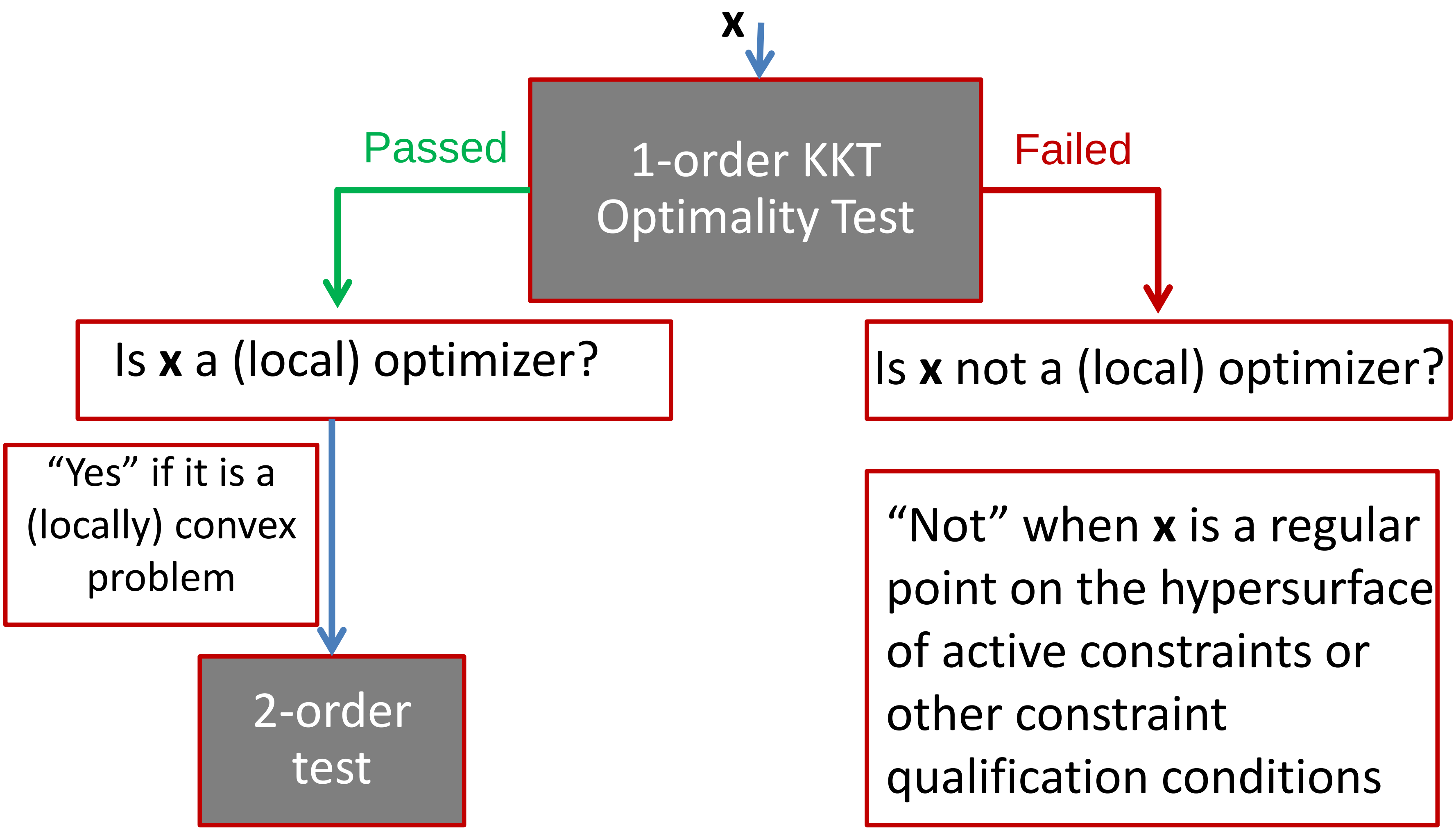


The curve (surface) of the objective function is <u>tangential</u> to the constraint curve (surface) at the optimal point.

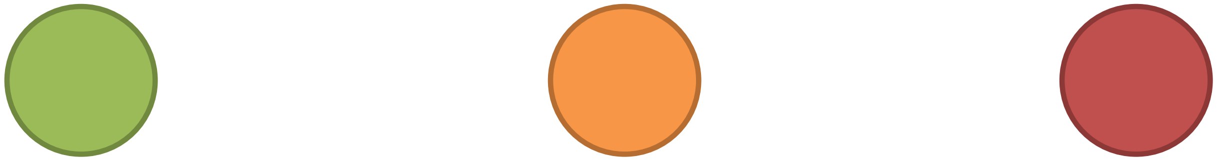
Remarks of First-Order Necessary Conditions

- The conditions only used the first derivatives of the functions involved in the problem, so it is usually called First-Order Necessary conditions (FONC), also named as the KKT conditions.
- Every optimizer must satisfy these conditions (under mild technical assumptions, such as all functions are linear, slate condition, regularity condition, etc.)
- For general optimization, these necessary conditions may not be sufficient; but for convex optimization, they are also sufficient, and the optimal solution is unique if the objective function is stronly convex.
- These conditions are important both theoretically and computationally
- Complementarity Slackness: nonbinding constraint receives zero penalty (multiplier and slack could be both zeros).
- Applications include: the Fisher equilibrium, SVM, etc.

Optimality Test for GCO

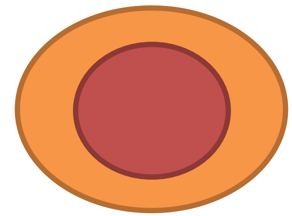


Minimum and KKT Solutions

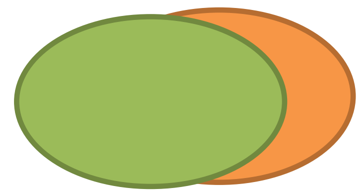


1^{st} order KKT | **Local Opt** | **Global Opt**

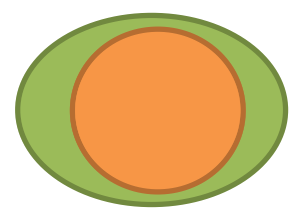
Local vs Global Opt



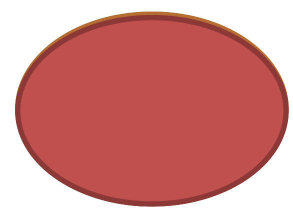
KKT vs Local Opt



KKT vs Local Opt with CQ



KKT vs Global Opt for CO with CQ



2-Order KKT Condition for GCO

Tangent Plane:

$T = \{ \mathbf{z}: \nabla c_i(\mathbf{x})\mathbf{z} = 0$, for all i, such that $c_i(\mathbf{x})=0 \}$

Necessary Condition:

$\mathbf{z}^T\nabla_x^2 L(\mathbf{x},\mathbf{y})\mathbf{z} \geq \mathbf{0}$, for all $\mathbf{z}$ in T

Sufficient Condition:

$\mathbf{z}^T\nabla_x^2 L(\mathbf{x},\mathbf{y})\mathbf{z} > \mathbf{0}$, for all non-zero $\mathbf{z}$ in T

This can be done by checking positive semi-definiteness (or definiteness) of the projected Hessian of the Lagrange function

Applications: Optimality Condition & Duality

- Data Science, Machine Learning, Game/Market Equilibrium Theories
  - LR, SVM, WBC, SNL, MDP, etc.
  - Fisher market, Arrow-Debreu market
  - Duality and optimality lead to equilibrium conditions
- Pricing and learning
  - OLP: online LP by learning prices
  - WBC: distributed computation
  - SDP: Duality explains localizability
- Distributionally robust optimization/learning
  – A model to deal with inaccurate sample-distributions in stochastic optimization and prediction
- etc…