CME307/MS&E311 Optimization Theory Summary

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

http://www.stanford.edu/~yyye
http://www.stanford.edu/class/msande311/

Optimization Problems

- A set of decision variables, x, in vector or matrix form with dimension n
- 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:

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

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

Function, Gradient Vector and Hessian Matrix

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

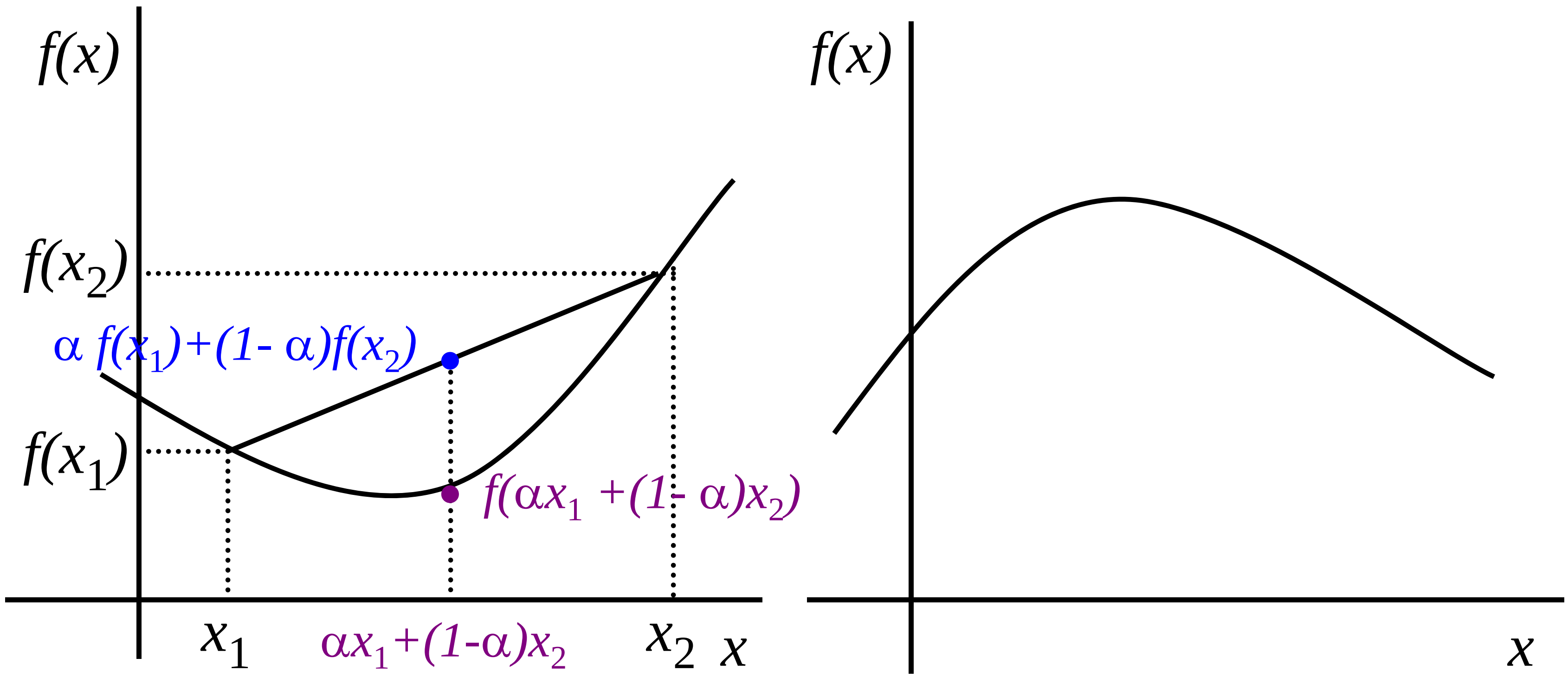
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \cdots\cdots\cdots \frac{\partial f}{\partial x_n} \right)$$

- 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 $0 \leq \alpha \leq 1$

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

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

Convex Quadratic Functions

$f(x)=x^TQx+c^Tx$ is a convex function if and only if Hessian matrix Q is positive semi-definite (PSD).

$f(x)=x^TQx+c^Tx$ is a strictly convex function if and only if Q is positive definite (PD).

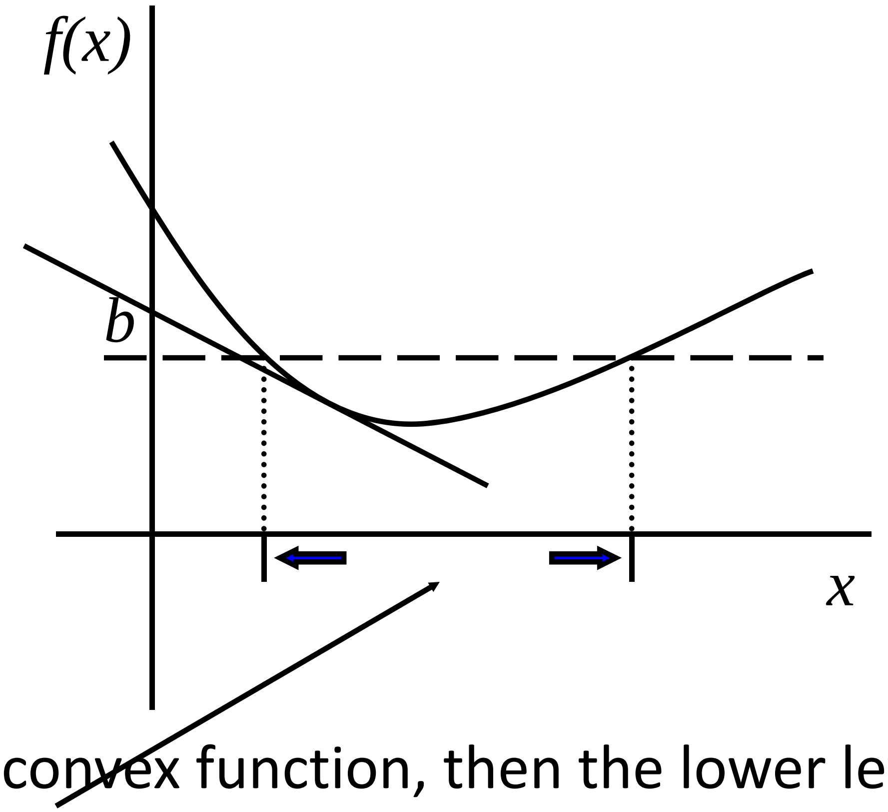
Q is PSD if and only if $x^TQx \geq 0$ for all x.

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

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

Properties of Convex Function



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 <u>tangent line (planes)</u>.
The Hessian matrix of a convex function is <u>positive semi-definite</u>.

Optimization Problem Classes

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

- <u>Unconstrained Optimization</u>
  - Convex or Nonconvex
- <u>Constrained Optimization</u>
  - Conic Linear Optimization/Programming (CLO/CLP)
  - Convex Constrained Optimization (CCO)
    - Feasible region/set 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

Optimization Problem Forms

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

Conic Linear Optimization (CLO)

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

This is convex optimization

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_i(\boldsymbol{x}) = 0, \ \mathrm{i}=1,\ldots,\mathrm{m} \\ & c_i(\boldsymbol{x}) \geq 0, \ \mathrm{i}=1,\ldots,\mathrm{p} \end{array}$$

Generally Contrained 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?

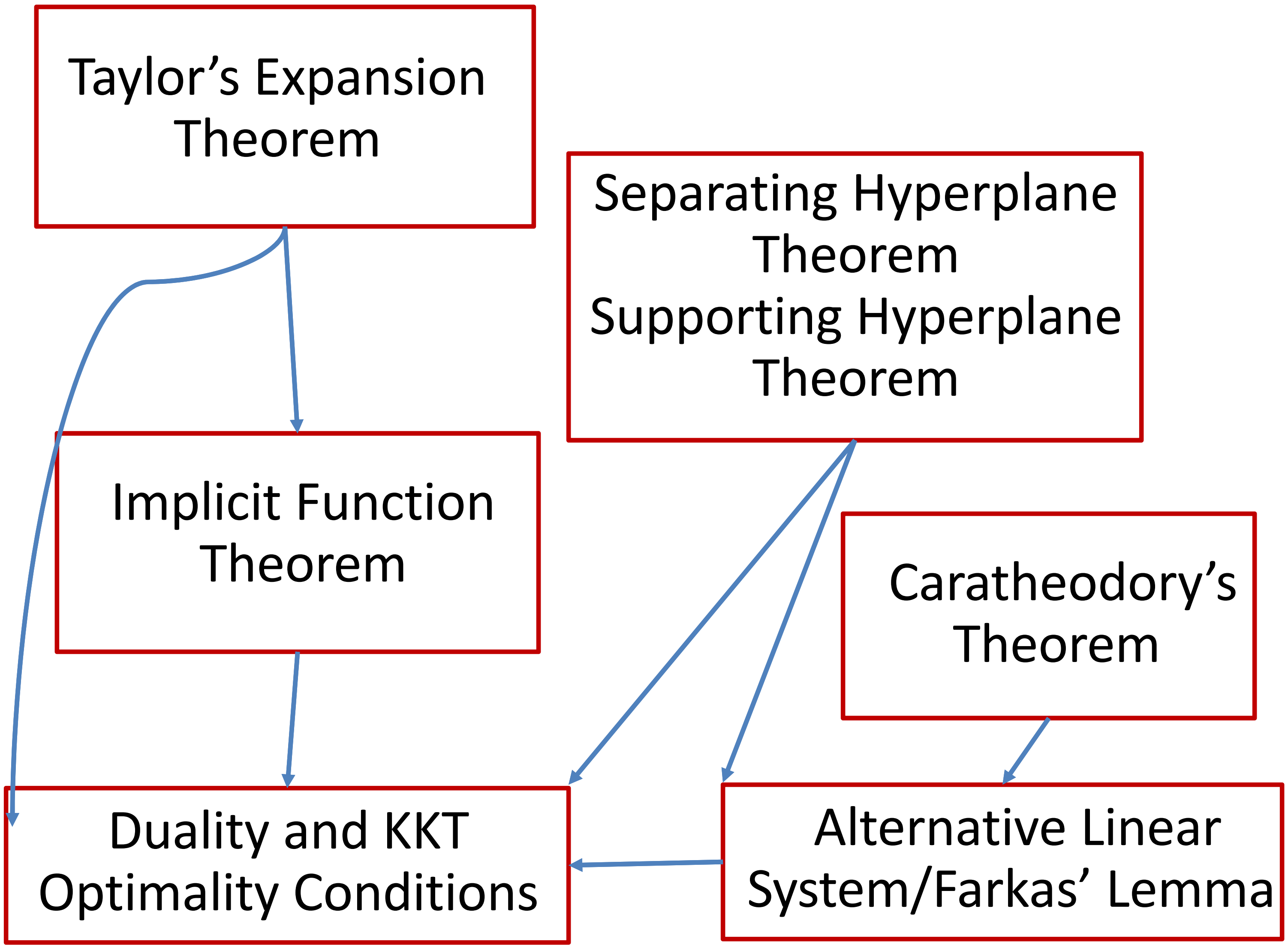
- It guarantees that every local optimizer is a global optimizer
- It guarantees that every (first-order) KKT (or stationary) point/solution is a global optimizer
- This is significant because all of our numerical optimization algorithms search/generate a KKT point/solution
- Sometime the problem can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

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

$$\Updownarrow$$

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

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

$$A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0},$$
$$\boldsymbol{x} \in K$$

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

$$\boldsymbol{b}^T\boldsymbol{y}=1(>0)$$
$$A^T\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{0},$$
$$\boldsymbol{s} \in K^*$$

System Y

K* **is the dual cone**

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

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

Alternative Systems and CLO Pairs II

$$c^T x = -1 (<0)$$
$$A\mathbf{x} = \mathbf{0},$$
$$\mathbf{x} \in K$$

$$A^T\mathbf{y} + \mathbf{s} - \mathbf{c} = \mathbf{0},$$
$$\mathbf{s} \in K^*$$

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

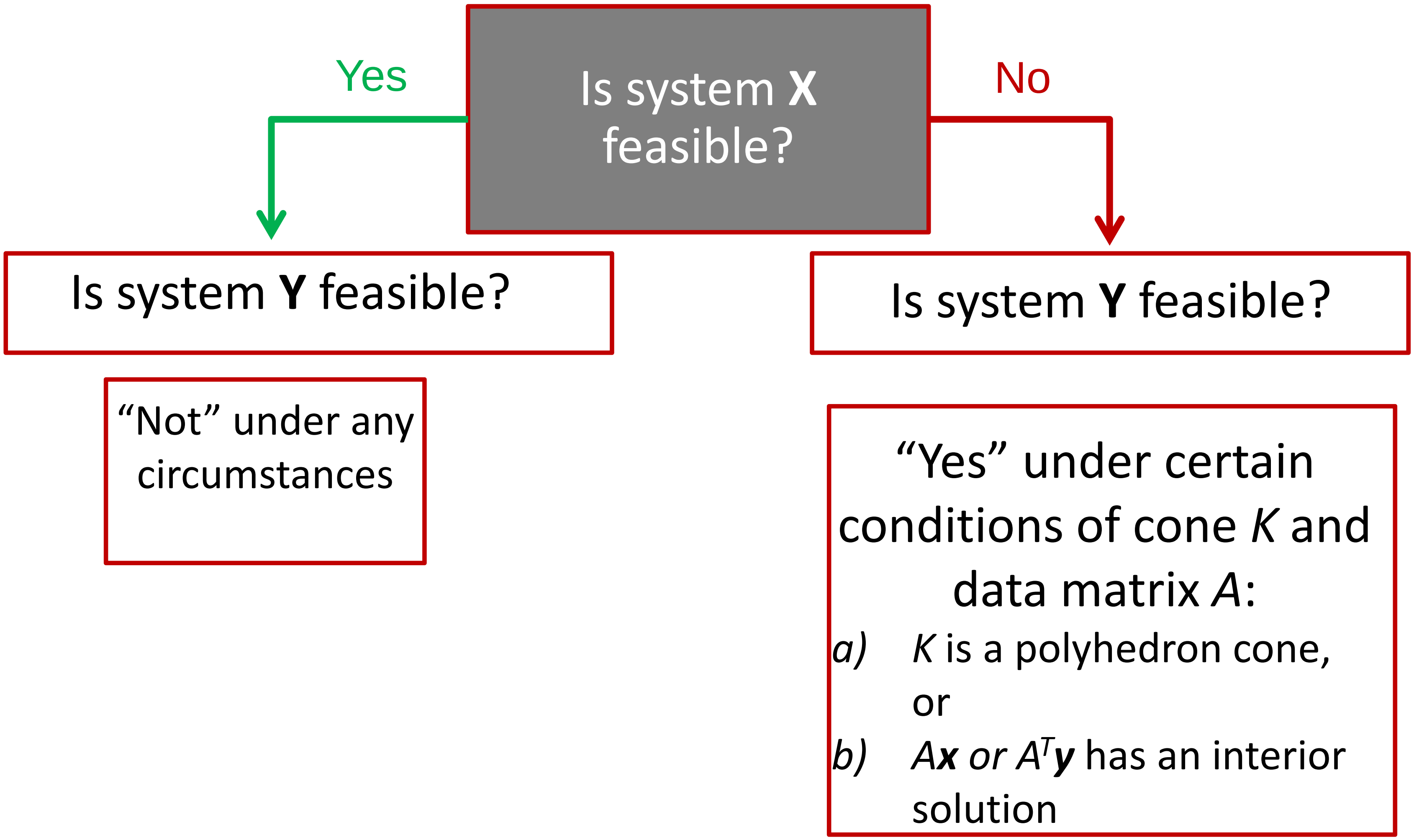
System Y

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

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

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

Feasibility Test Machine



General Rules to Construct the CLO Dual

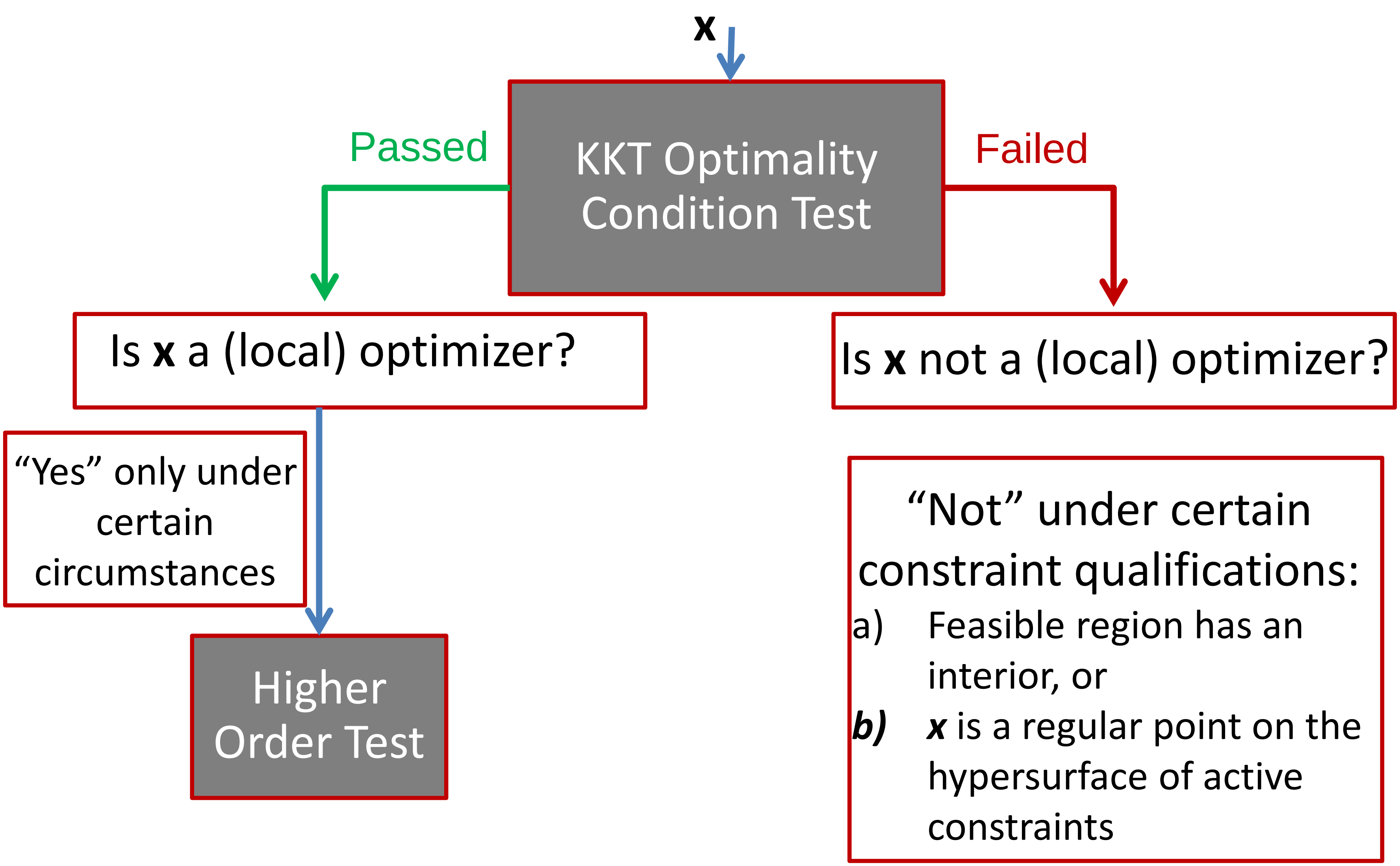
OBJ Vector/Matrix	RHS Vector/Matrix
RHS Vector/Matrix	OBJ Vector/matrix
A	A^T
Max model	Min model
$x_j \geq_K 0$	jth constraint $\geq_{K^*}$
$x_j \leq_K 0$	jth constraint $\leq_{K^*}$
x_j free	jth constraint $=$
ith constraint $\leq_K$	$y_i \geq_{K^*} 0$
ith constraint $\geq_K$	$y_i \leq_{K^*} 0$
ith constraint $=$	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



Duality Theorems for CLO

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

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

Weak Duality Theorem

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

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

The Lagrange Function of GCO

min $\quad f(\boldsymbol{x})$

s.t. $\quad c_i(\boldsymbol{x})\ (\le, =, \ge)\ 0$, i=1,…,m

Restriction on multipliers y_i,

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

The Largrange Function

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

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

y_i free: can be penalized either way
$y_i \ge 0$: can be penalized when $c_i(\boldsymbol{x}) \le 0$
$y_i \le 0$: can be penalized when $c_i(\boldsymbol{x}) \ge 0$

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

Strong Duality Theorem
They must equal?

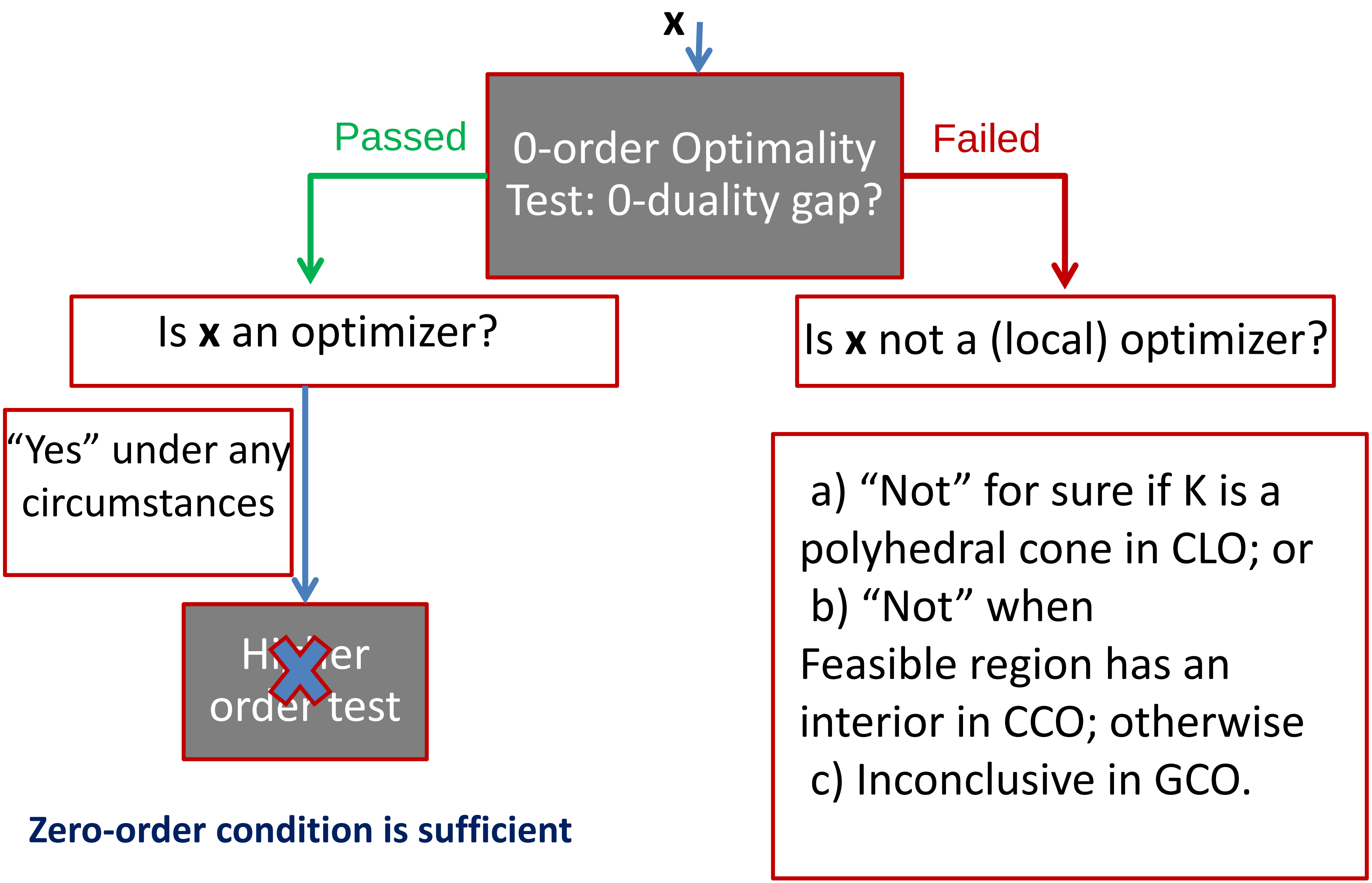
$$d^* = \max \quad \phi(\boldsymbol{y})$$
$$\text{s.t.} \quad y_i \ (\leq, \text{"free"}, \geq)\ 0,\ i=1,\ldots,m$$

Not necessarily!

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

Sufficient!

Zero-Order Optimality Test for CLO and GCO

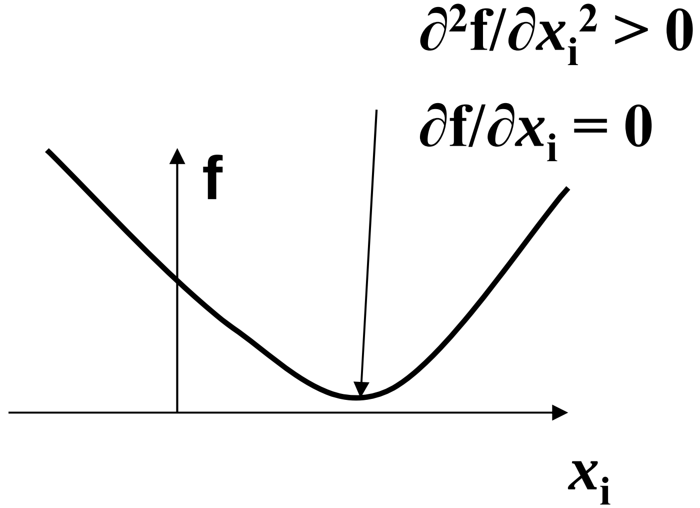


Zero-order condition is sufficient

1 and 2-order Conditions: Unconstrained

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

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



1-Order KKT Condition for GCO

Recall the Largrange Function

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

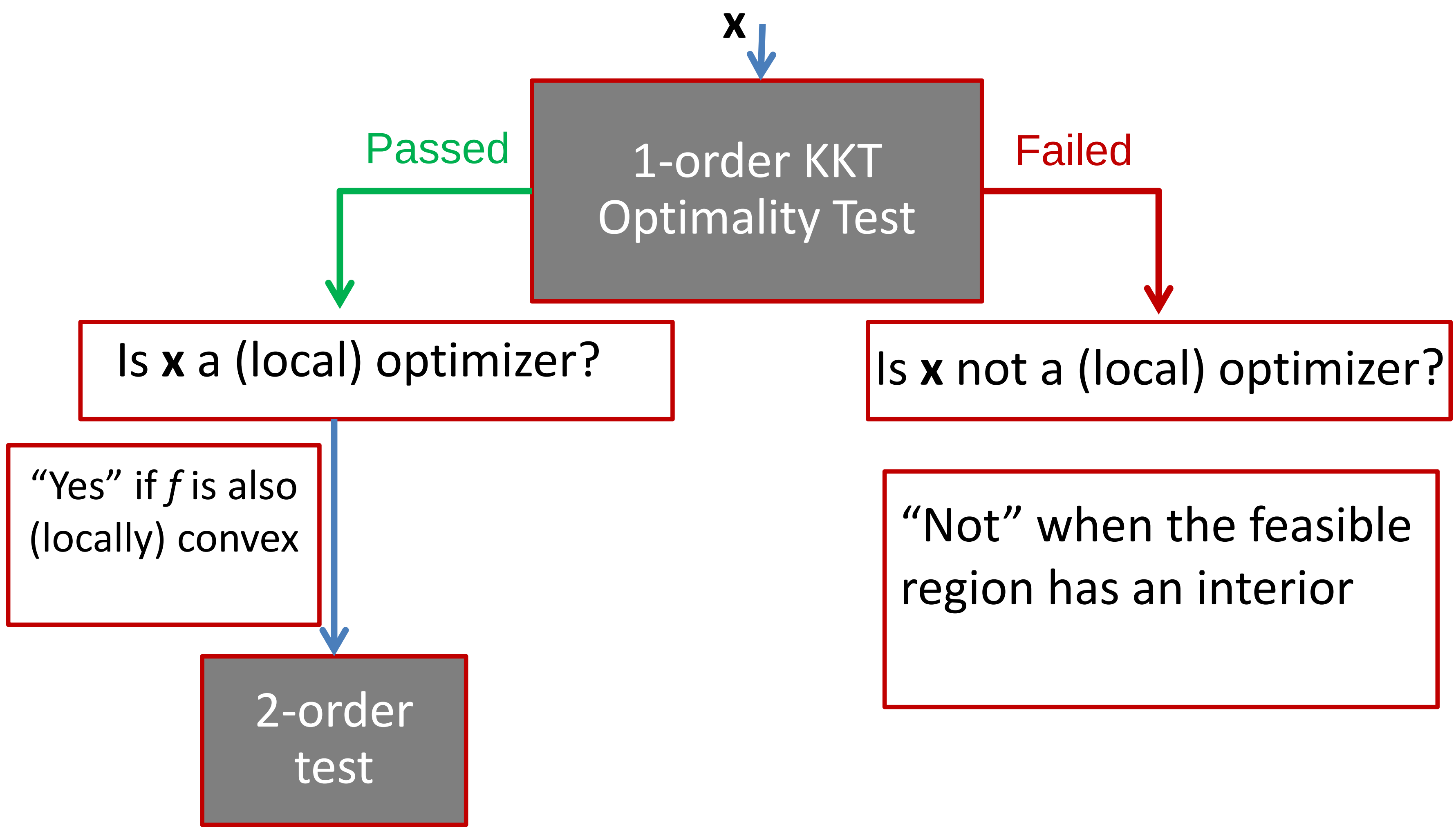
$\nabla_x L(\mathbf{x},\mathbf{y}) = \mathbf{0}$, that is,

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

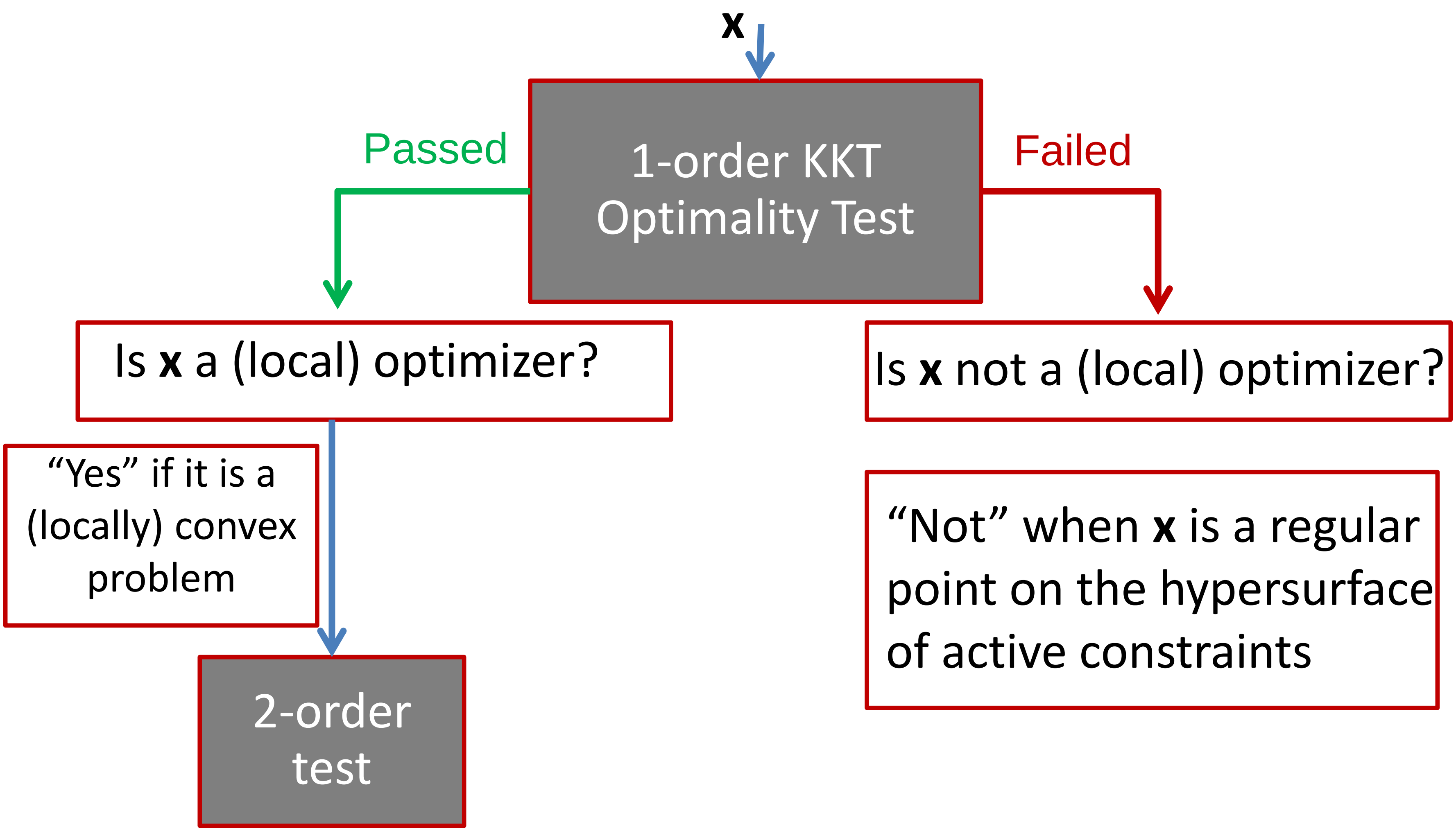
$c_i(\mathbf{x}) y_i = 0$, for all $i=1,\dots,m$

$c_i(\boldsymbol{x})$ ($\leq$, =, $\geq$) 0, y_i ($\leq$,"free", $\geq$) 0, $i=1,\dots,m$

Optimality Test for CCO



Optimality Test for GCO



2-Order KKT Condition for GCO

Tangent Plane:

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

Necessary Condition:

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

Sufficient Condition:

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

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

Example: Optimality Conditions

$$\begin{array}{ll} \min & x_1^{\ 2} + x_2^{\ 2} \\ \text{s.t.} & 1\text{-}0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2 \geq 0 \end{array}$$

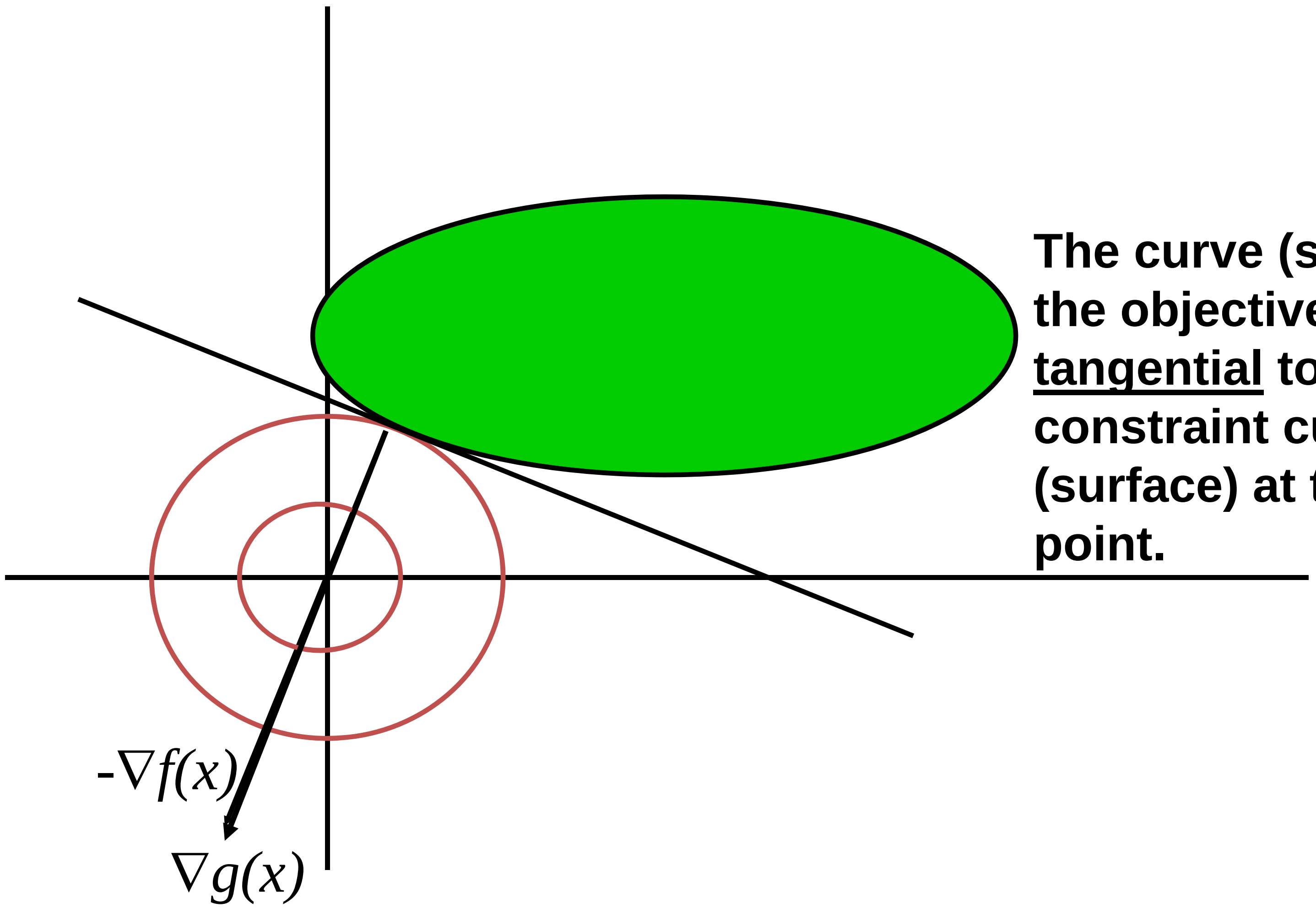
$$L(x_1, x_2, \lambda) = x_1^{\ 2} + x_2^{\ 2} - \lambda\left(1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2\right)$$

$$\begin{pmatrix} \partial L / \partial x_1 \\ \partial L / \partial x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} - \lambda \cdot \begin{pmatrix} 0.5(2 - x_1) \\ 2(2 - x_2) \end{pmatrix} = 0$$

$$1\text{-}0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2 \geq 0, \qquad \lambda \geq 0$$

$$\lambda(1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2) = 0$$

Example: KKT Conditions



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

Example: Computation of a KKT Point

$$\begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} - \lambda \cdot \begin{pmatrix} 0.5(2 - x_1) \\ 2(2 - x_2) \end{pmatrix} = 0$$

$$x_1 = \frac{2\lambda}{4+\lambda}; x_2 = \frac{2\lambda}{1+\lambda}$$

- If $\lambda = 0$, then $x_1 = 0$ and $x_2 = 0$, and thus the constraint would not hold with equality. Therefore, λ must be positive.
- Plugging the two values of $x_1(\lambda)$ and $x_2(\lambda)$ into the constraint with equality gives us $\lambda = 1.8$.
- We can then solve for $x_1 = .61$ and $x_2 = 1.28$.

Applications: Optimality Conditions

- The market equilibrium theory
  - Fisher market, Arrow-Debreu market
  - Duality and optimality lead to equilibrium conditions
- Sensor localization
  - SOCP: KKT conditions explain observations
  - SDP: Duality explains localizability
- Offline and Online LP
  - Learning optimal dual solution helps to make primal decisions online
- Non-convex regularization
  - L_p norm regulation function for unconstrained or constrained minimization
  - KKT conditions establish a desired thresh-holding properties at any KKT solution (first or second order)