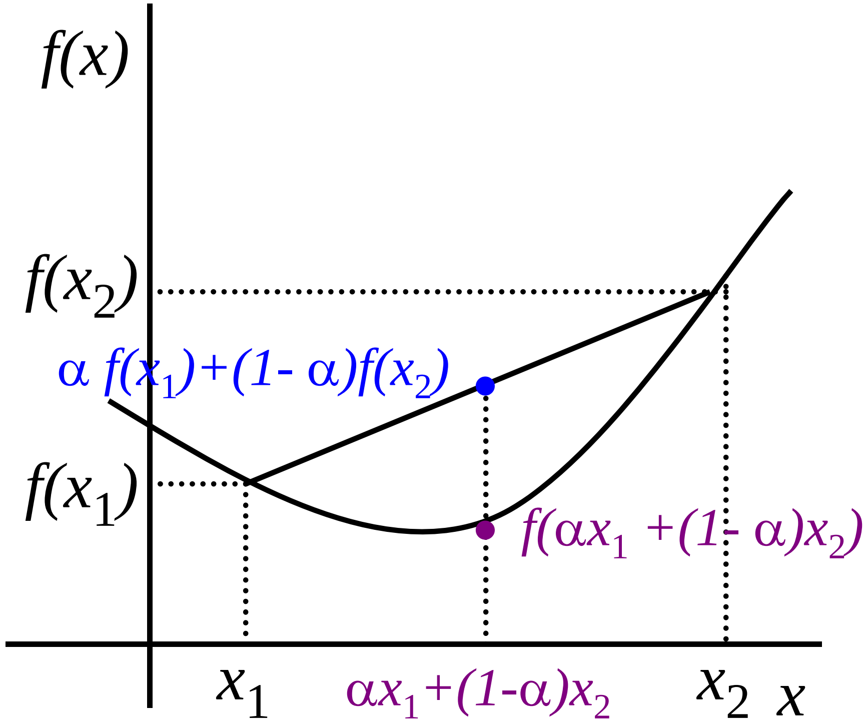
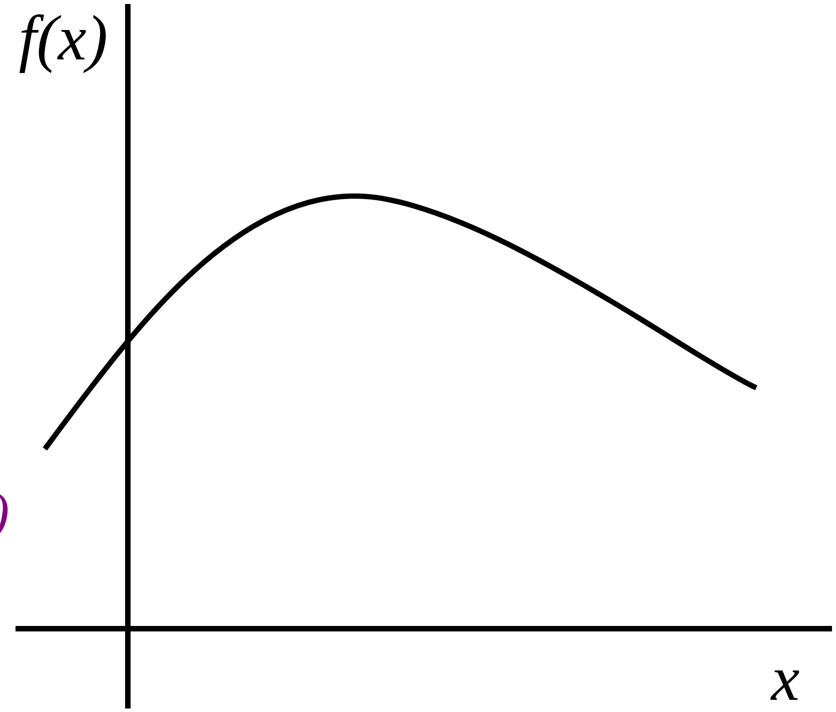
MS&E311 Optimization Review

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/

Math: 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 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).

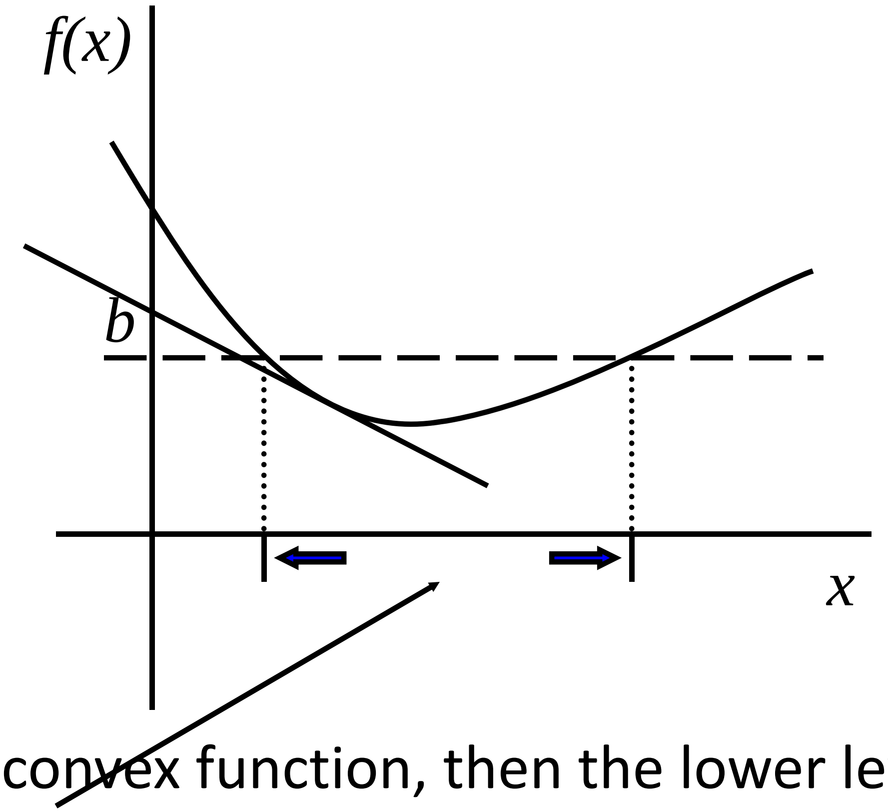
Q is PSD if and only if $x^TQx \geq 0$ for all x.

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 convex if every line segment connecting any two points in the set is contained entirely within the set
  - Ex - polyhedron
  - Ex - ball
- An extreme point 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 tangent line (planes).
The Hessian matrix of a convex function is positive semi-definite.

Optimization problem classes

- Unconstrained Optimization
  - Convex or Nonconvex
- Constrained Optimization
  - Conic Linear Optimization (CLO)
  - Convex Constrained Nonlinear Optimization (CCNO)
    - Feasible region convex; objective convex or non-convex
  - General Nonlinear Optimization (GNO)
- An optimal solution always exists if the intersection of the objective function level set and the feasible set is compact (bounded and closed).

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

Common assumptions

- We generally assume $f(x)$ is continuous and differentiable over the feasible region.
- Sometimes we can smooth them by reformulation as constrained optimization

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

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & \lambda - f_i(x) \le 0, \text{ for } i=1,\ldots,n \end{aligned}$$

Gradient vector and Hessian matrix

- The Gradient Vector of f at x

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- The Hessian Matrix of f at x

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

Optimization Problem Forms

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

Conic Linear Optimization (CLO)

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

This is convex optimization

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

General Nonlinear Optimization (GNO)

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

This is CCO if c_i are all convex, and h_i are all affine functions

Why do we care so much about convex optimization?

- Minimize a convex function over a convex feasible region (as long as it is convex in the feasible region).
- It guarantees that every local optimizer is a global optimizer
- It guarantees that every KKT (or stationary) point is a global optimizer
- This is significant because all of our basic optimization algorithms search for a KKT point
- Sometime the problem can be "convexfied":

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

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

Optimization **Theory**: Mathematical Foundations

Taylor's Expansion and Theorem

Implicit Function Theorem

Separating Hyperplane Theorem
Supporting Hyperplane Theorem

Weierstrass Theorem

Caratheodory's Theorem

Theory: feasibility conditions

- Feasibility Conditions or Farkas' Lemmas are developed to characterize and certify feasibility or infeasibility of a feasible region
- Alternative Systems A and B: A has a feasible solution if and only if B has no feasible solution
  - A and B cannot both have feasible solution
  - Exactly one of them has a feasible solution
- They are special cases of primal and dual pairs

Alternative Systems

$$\boxed{\begin{array}{c} A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}, \\ \boldsymbol{x} \in \boldsymbol{K} \end{array}}$$

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

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

System B

K* **is the dual cone**

$$\boxed{\begin{array}{c} A^T\boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} = \boldsymbol{0}, \\ \boldsymbol{s} \in \boldsymbol{K} \end{array}}$$

$$\boxed{\begin{array}{c} \boldsymbol{c}^T\boldsymbol{x}=-1(<0) \\ A\boldsymbol{x} = \boldsymbol{0}, \\ \boldsymbol{x} \in \boldsymbol{K^*} \end{array}}$$

Feasibility Test Machine

Is system **A** feasible?

Yes

Is system **B** feasible?

"No" under any circumstances

No

Is system **B** feasible?

"Yes" only under certain conditions of *K* and *A*:

a) *K* would be a polyhedron cone, or

b) Part of B has an interior feasible solution

Farkas' Lemma for CLO

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

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

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

Dual Problem: System B

K* **is the dual cone**

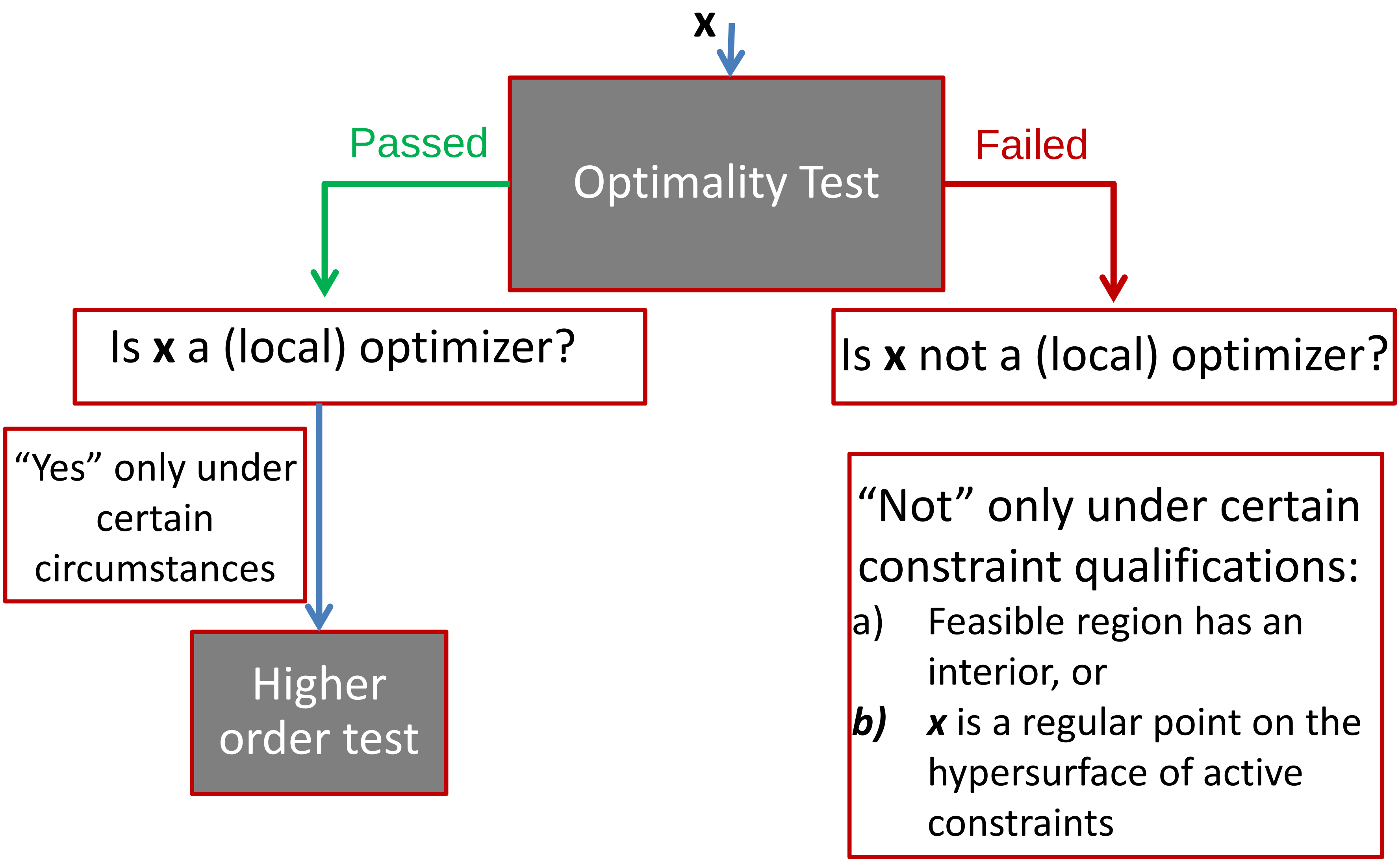
$$\begin{aligned} p^*=\max \quad & \boldsymbol{0}^T\boldsymbol{y} \\ \text{s.t.} \quad & A^T\boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} = \boldsymbol{0}, \\ & \boldsymbol{s} \in \boldsymbol{K} \end{aligned}$$

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

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.

Optimality Test Machine



Conic Duality for CLO

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

Primal Problem

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

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

Dual Problem

K* **is the dual cone**

0-order Condition: $\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{b}^T\boldsymbol{y}$

The Lagrange function of GNO

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

Restriction on multipliers y_i,

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

The Largrange Function

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

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

y_i free: can be penalized either way
$y_i \geq 0$: can be penalized when $c_i(\mathbf{x}) \geq 0$
$y_i \leq 0$: can be penalized when $c_i(\mathbf{x}) \leq 0$

The Lagrangian Duality for GNO

$f^*=\min \quad f(\boldsymbol{x})$

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

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

$\phi^*=\max \quad \phi(\boldsymbol{y})$

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

0-order Condition: $f(\boldsymbol{x}) = \phi(\boldsymbol{y})$

Zero-Order Optimality Test for CLO and GNO

x

0-order Optimality Test: 0-duality gap?

Passed

Is **x** an optimizer?

"Yes" under any circumstances

Higher order test

Failed

Is **x** not a (local) optimizer?

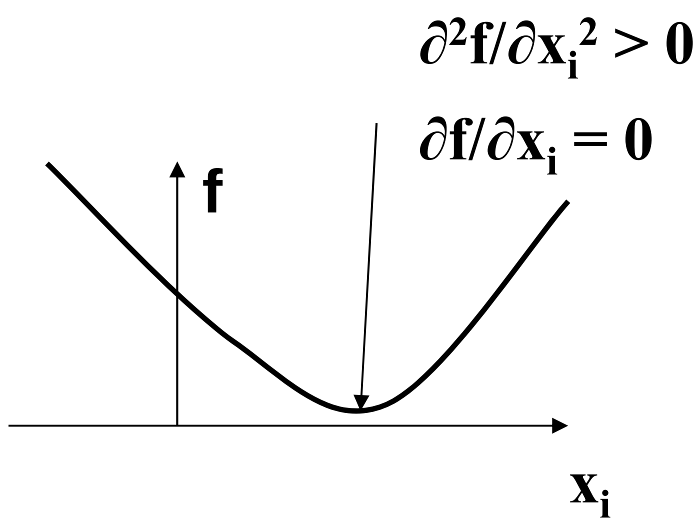
a) "Not" for sure if K is a polyhedral cone in CLO; or
b) "Not" when Feasible region has an interior in CLO; otherwise
c) Inconclusive in GNO.

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)**
    - **[$x \bullet \nabla^2 f(x) \bullet x \geq 0$ for all x]**
- <u>Second Order Sufficient Condition</u>

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



1-order KKT Condition for CCNO and GNO

Recall the Largrange Function

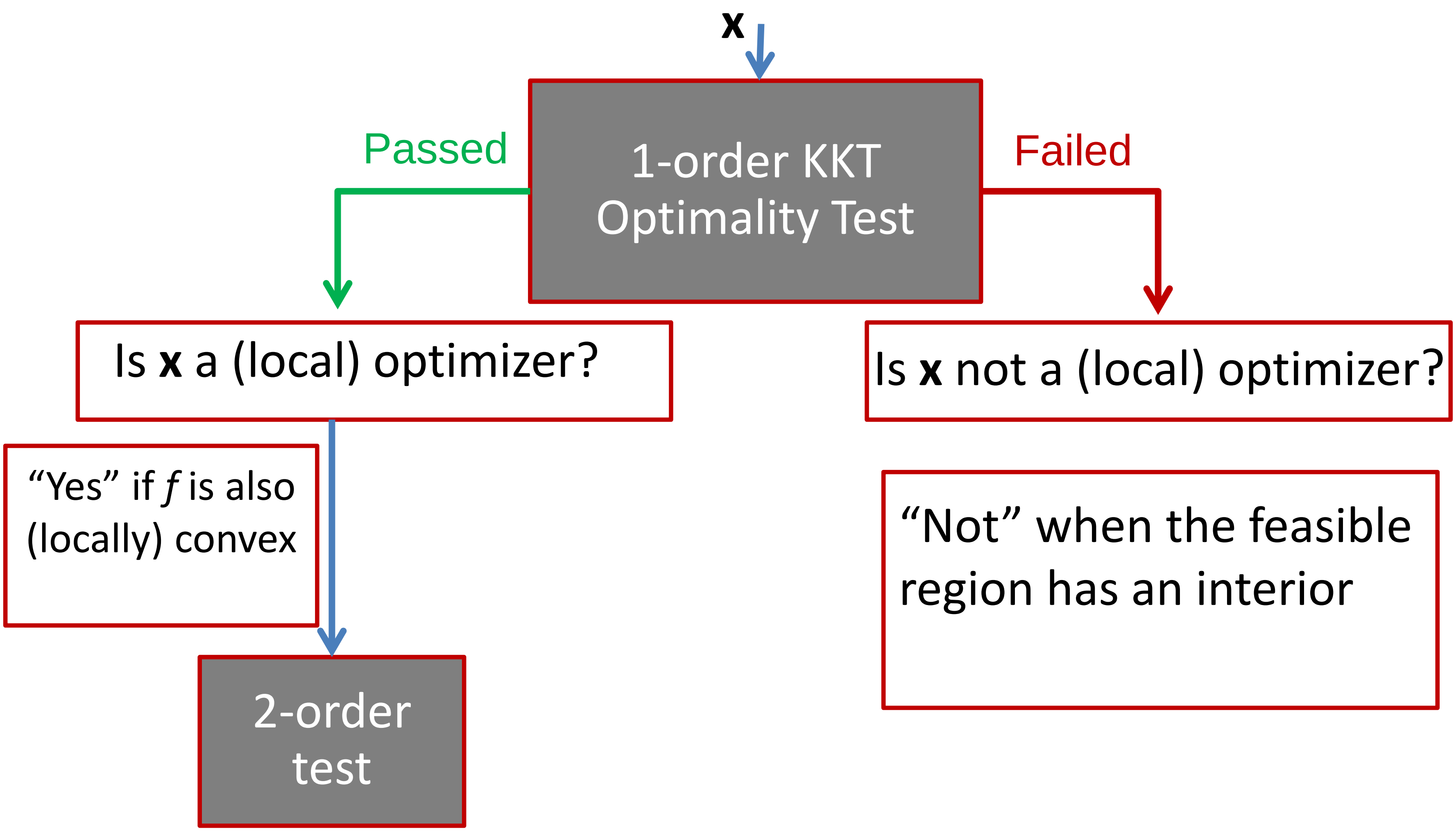
$$L(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) + \textstyle\sum_i c_i(\mathbf{x}) y_i$$

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

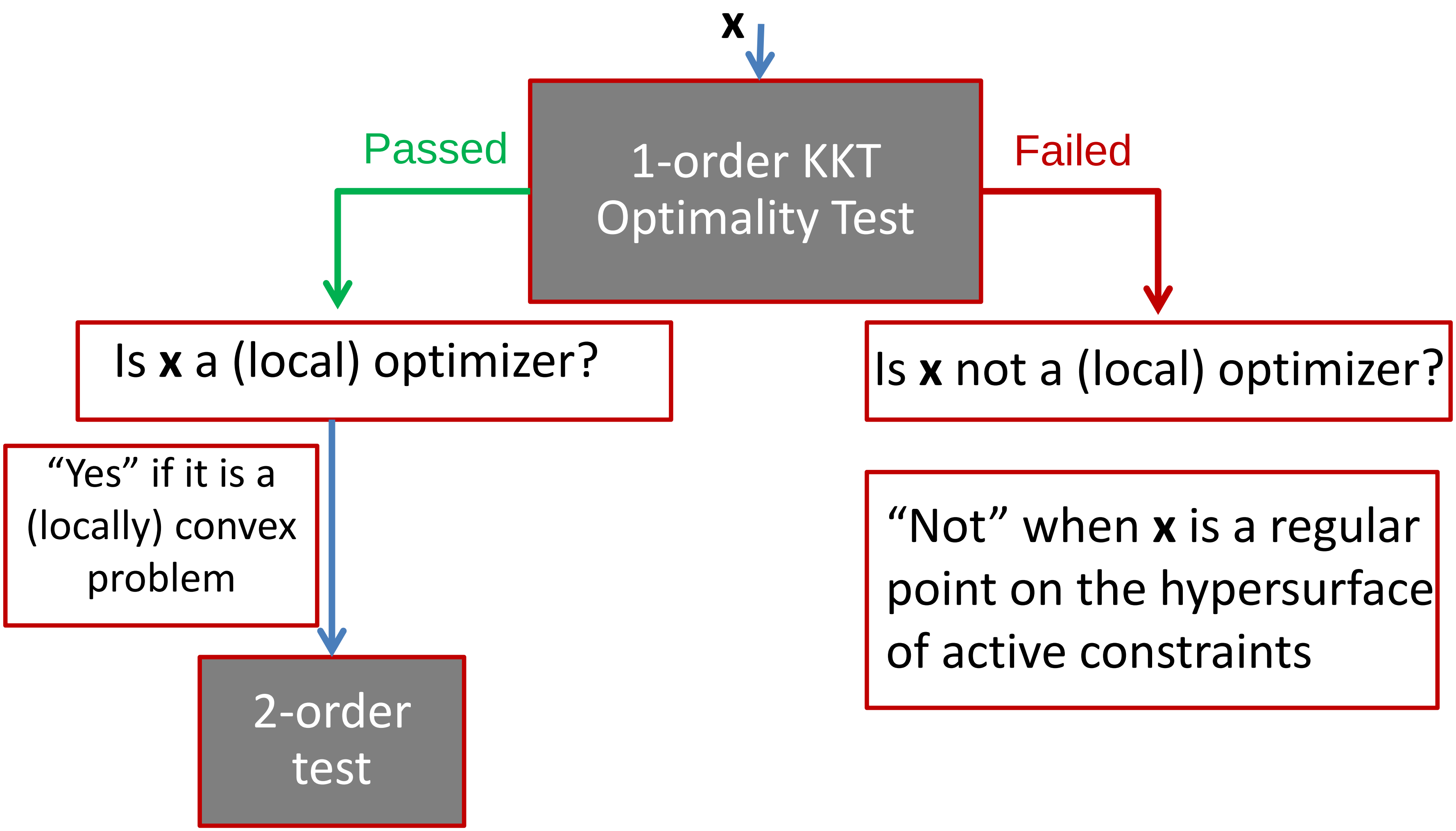
$\partial L(\mathbf{x},\mathbf{y})/\partial x_j = 0$, for all $j=1,\ldots,n$, and

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

Optimality Test for CCNO



Optimality Test for GNO



2-order KKT Condition for CCNO and GNO

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:

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

Sufficient Condition:

$\boldsymbol{z}^{\mathrm{T}}\nabla_x{}^2 L(\mathbf{x},\mathbf{y})\boldsymbol{z} > \mathbf{0}$, 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

$$\min \quad x_1^2 + x_2^2$$

$$\text{s.t.} \quad 1 - 0.25\cdot(x_1-2)^2 - (x_2-2)^2 \geq 0$$

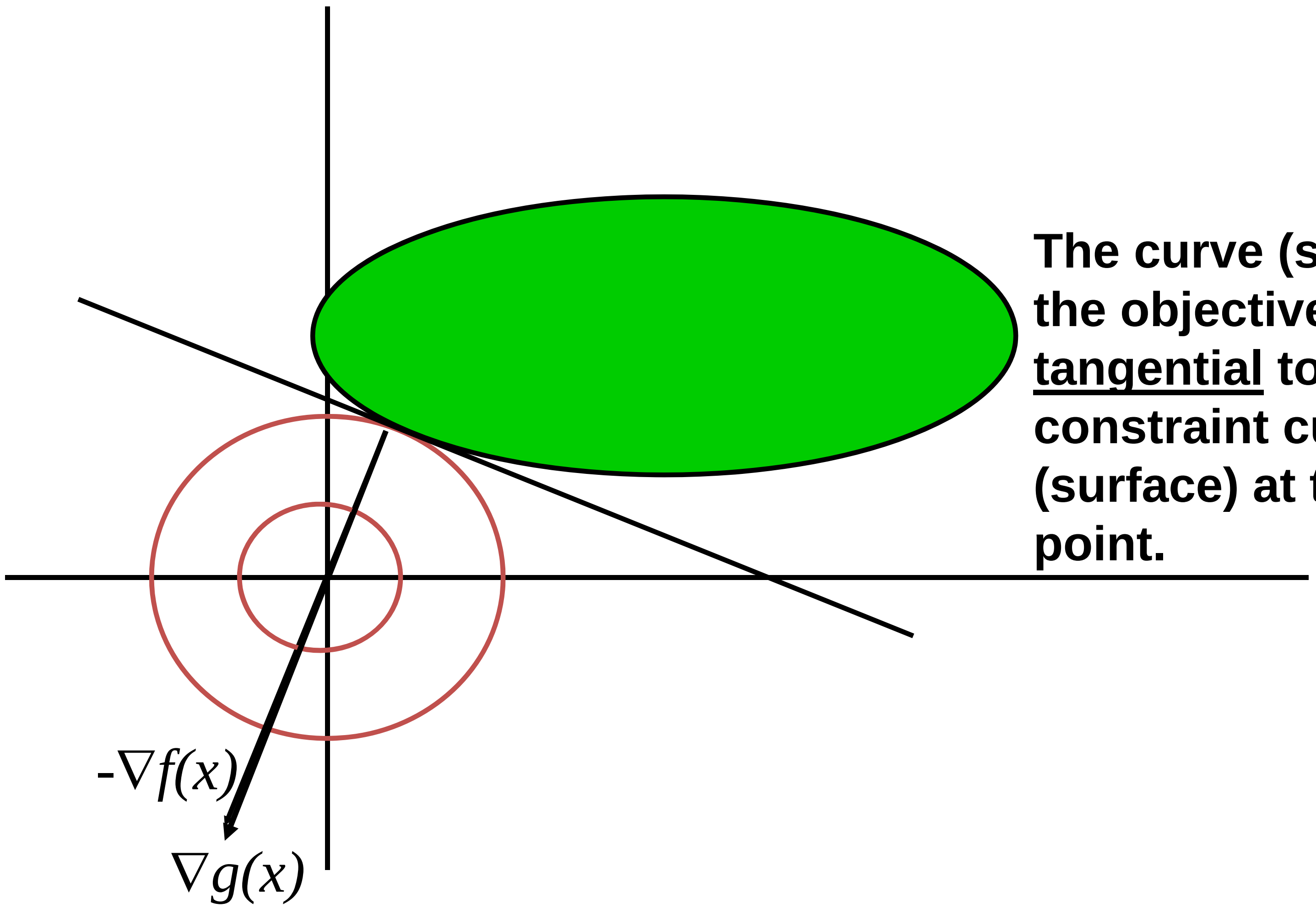
$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 - \lambda\left(1 - 0.25\cdot(x_1-2)^2 - (x_2-2)^2\right)$$

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

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

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

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

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

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

- 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
  - Linear utilities
  - Other utilities?
- Concave regularization
  - L_p norm regulation function for unconstrained or constrained minimization
  - Thresh-holding properties at any KKT point (first or second order)
  - Other concave regulation functions?