Lecture 1: Introduction, Necessary and Sufficient Conditions for Minima & Convex Analysis

Prof. Krishna R. Pattipati
Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

ECE 6437

Computational Methods for Optimization

Fall 2009

September 1, 2009

Introduction

- Contact Information
  - Room number: ITE 350
  - Tel/Fax: (860) 486-2890/5585
  - E-mail: krishna@engr.uconn.edu

- Office Hours: Tuesday – Thursday: 11:00-12:00 Noon

- Mission or goal
  - Provide systems analysis with central concepts of widely used optimization techniques
  - Requires skills from both Mathematics and CS
  - Need a strong background in multivariable calculus and linear algebra

Outline of Lecture 1

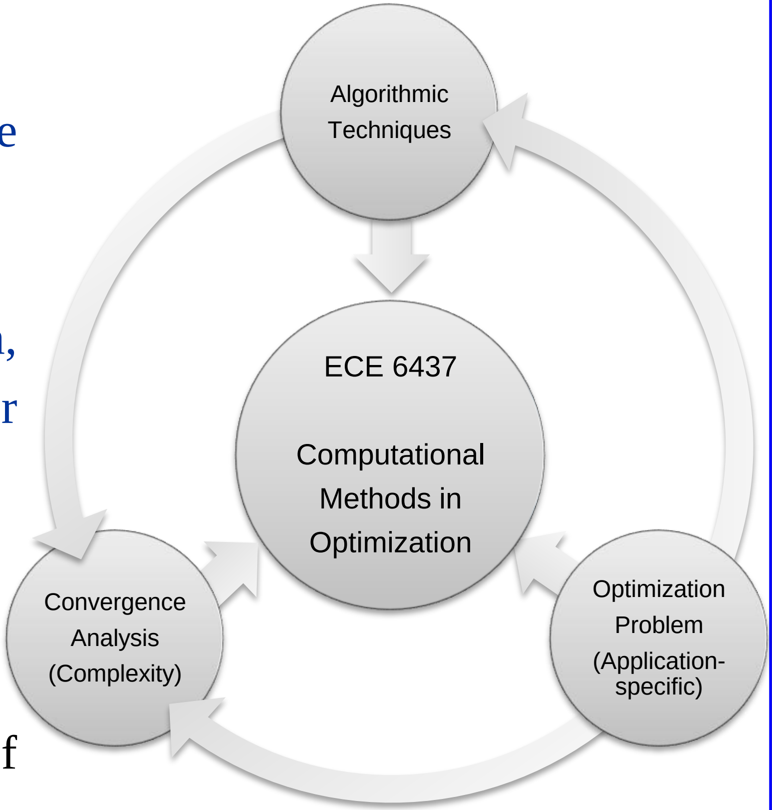
- Three Recurrent Themes
    - Problem, Algorithms, Convergence Analysis
- Optimization Applications
- What is an Optimization Problem?
- Classification of Optimization Problems
- Three Basic Questions of Optimization
    - Optimality conditions, algorithm, convergence
- Optimality Conditions for single variable and Multi-variable Functions
- Elementary Convexity Theory

Three Recurrent Themes

- Need to mathematically understand the optimization problem to be solved

- Design an algorithm to solve the problem, that is, a step-by-step procedure for solving the problem

- Convergence Analysis
  - How fast does the algorithm converge?
  - What is the relationship between rate of convergence and the size of the problem?



Applications of Optimization

- Sample Applications
  - Scheduling in Manufacturing systems
  - Scheduling of police patrol officers in a city
  - Reducing fuel costs in Electric power industry (unit commitment)
  - Gasoline blending in TEXACO
  - Scheduling trucks at North American Van Lines
  - Advertisement to meet certain % of each income group
  - Investment portfolio to maximize expected returns, subject to constrains on risk
- Technical Areas
  - Operations Research, Systems theory (Optimal Control), Statistics (Design of Experiments), Computer Science, Chemical and Civil Engineering, Economics, Medicine, Physics, Math,….

What is an Optimization Problem?

- Three Attributes:

1. A set of independent variables or parameters $(x_1, x_2, \ldots, x_n)$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \; n \text{ vector} \qquad \begin{array}{l} \underline{x} \in R^n \;\; \text{continuous} \\ \underline{x} \in Z^n \;\; (\ldots,-2,-1,0,1,2,\ldots) \;\; \text{integers} \\ \{\underline{x} \mid x_i = 0,1\} \;\; \text{binary} \end{array}$$

2. Conditions or restrictions on the acceptable values of the variables $\Rightarrow$ constraints of the optimization problem, Ω (e.g.,$\underline{x} \geq 0$)

3. A single measure of goodness, termed the objective (utility) function or cost function or goal, which depends on $x_1, x_2, \ldots, x_n$, $f(x_1, x_2, \ldots, x_n)$ or $f(\underline{x})$

$$f : R^n \to R \;\; ; \;\; \text{if } f \in Z \quad \begin{array}{l} R^n \to Z \\ Z^n \to Z \\ Z^n \to (0,1) \end{array}$$

Abstract Formulation

- Abstract Formulation: "Minimize $f(\underline{x})$ subject to $\underline{x} \in \Omega$"

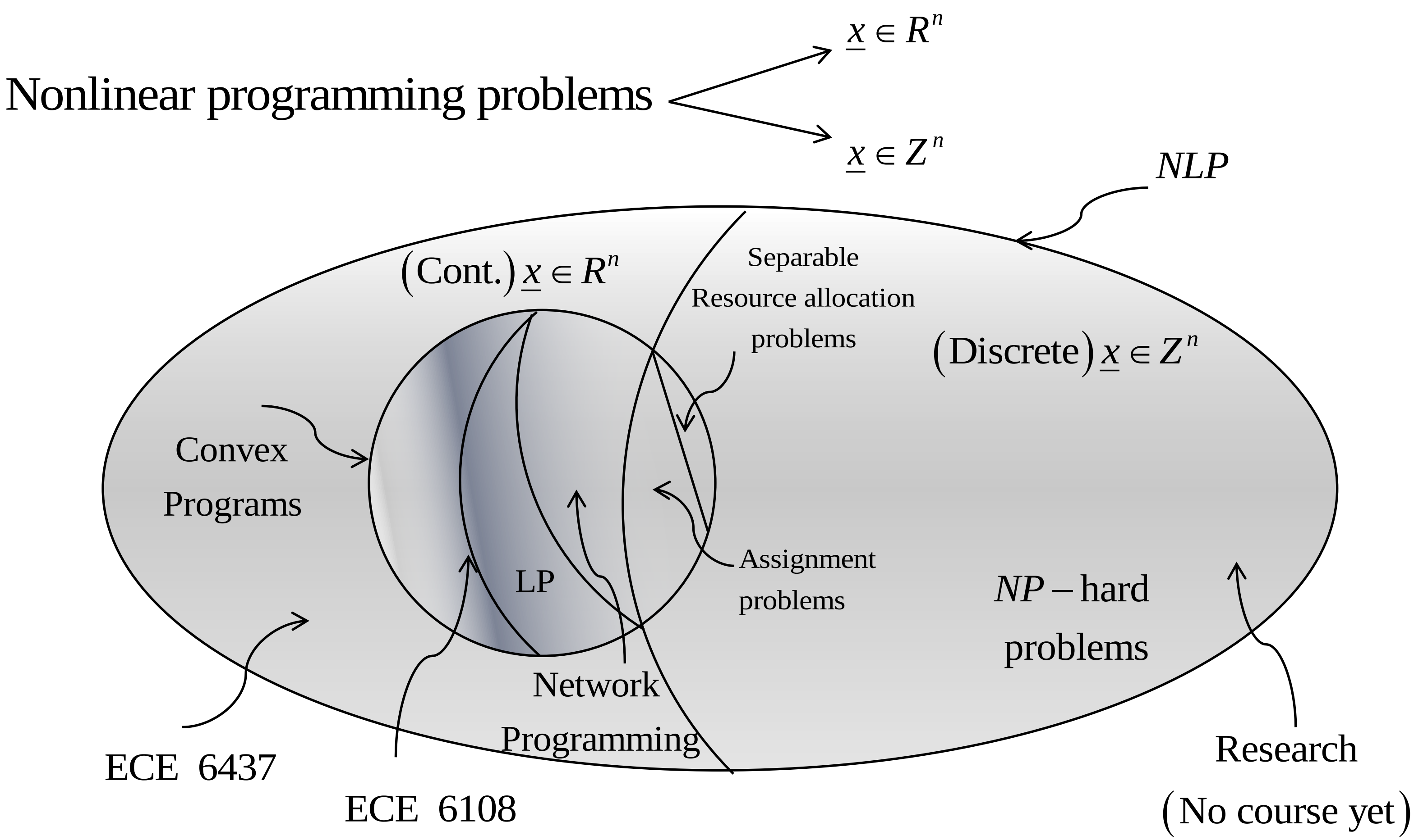
$$\Omega = \text{Feasible set, closed and bounded}$$

- Such problems have been investigated at least since 825 A.D. Persian author Abu Ja'far Muhammad ibn Musa Al-Khwarizmi who wrote the first book on Mathematics

- Since 1950's, a hierarchy of optimization problems have emerged under the general heading "Mathematical Programming". The solution approach is algorithmic in nature, i.e., construct a sequence

$$\underline{x}_0 \rightarrow \underline{x}_1 \rightarrow \ldots \underline{x}^*, \text{ where } \underline{x}^* \text{ minimizes } f(\underline{x})$$

A classification of Mathematical programming problems



- Unconstrained NLP: $\Omega = R^n \Rightarrow$ no constraints on $\underline{x}$
  - Steepest descent (gradient) method
  - Conjugate gradient method
  - Newton, Gauss-Newton methods & variations
  - Quasi-Newton (or) variable metric methods

- Constrained NLP: Ω defined by

$$h_i(\underline{x}) = 0,\ i = 1, 2, \ldots, m < n \quad \text{Equality constraints}$$

$$g_i(\underline{x}) \geq 0,\ i = 1, 2, \ldots, p \quad \text{Inequality constraints}$$

$$x_i^{LB} \leq x_i \leq x_i^{UB},\ i = 1, 2, \ldots, \text{n} \quad \text{Simple bound constraints}$$

  - Penalty methods
  - Multiplier or Augmented Lagrangian methods
  - Reduced gradient method
  - Recursive quadratic programming

Computational Methods in Optimization: ECE 6437 (Cont'd)

Special Case 1: Convex programming problem (CPP)

- Convex cost function with convex constraints
- $f(\underline{x})$ is convex (defined later).
- $g_i(\underline{x})$ is concave (or) $-g_i(\underline{x})$ is convex.
- $h_i(\underline{x})$ linear $\Rightarrow A\,\underline{x} = \underline{b} \Rightarrow \sum_{i=1}^{n} \underline{a}_i x_i = \underline{b} \in R^m$

Local minimum ≡ Global minimum

Linear Programming and Network Flows - ECE 6108

❑ Special Case1.1: Linear Programming (LP) Problem

- $f(\underline{x})$ is linear $\Rightarrow f(\underline{x}) = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = \underline{c}^T \underline{x}$
- $g_i(\underline{x})$ linear $\Rightarrow \underline{a}_i^T \underline{x} \geq b_i;\ i = 1, 2, \ldots, p$
- $x_i \geq 0;\ i = 1, 2, .., n;\ A\underline{x} = \underline{b};\ A$ is m by n matrix
- A striking feature of this problem is that the number of feasible solutions is finite: $$N = \binom{n+p}{m+p}$$
- Efficient algorithms exist for this problem
  - Revised simplex
  - Interior Point algorithms (application of specialized NLP to LP)
- One of the most widely used models in production planning.

❑ Special cases 1.1.x :

- Network Flows (LP on networks, i.e., graphs with weights)
- Shortest paths
- Maximum flow problem
- Transportation problem
- Assignment problem

Integer Programming

- Integer Programming (combinatorial optimization) has hard intractable problems with exponential computational complexity
  - Traveling salesperson problem
  - VLSI routing
  - Testing
  - Multi-processor scheduling to minimize make span
  - bin-packing
  - Knapsack problem
  - …..
- In ECE 6437, our focus will be on the following problems:
  - Unconstrained NLP
  - Constrained NLP
  - Convex Programming

Three Basic Questions of Optimization

1. Static Question: How can one determine whether a given point $\underline{x}^*$ is a minimum → Provides theory, stopping criteria, etc.

2. Dynamic Question: If a given point $\underline{x}$ is not a minimum, then how does one go about finding a solution that is a minimum? → Algorithm

$$\underline{x}_0 \to \underline{x}_1 \to \underline{x}_2 \to \ldots \to \underline{x}^*$$

3. Convergence Analysis:
   - Does the algorithm in 2 converge?
   - If so, how fast?

   How does $\left\| \underline{x}_k - \underline{x}^* \right\|$ or $\left\| f(\underline{x}_k) - f(\underline{x}^*) \right\|$ behave?

   Let us consider the third question first.

❑ Rate of Convergence Concepts:

Suppose have an algorithm that generates a sequence $\{\underline{x}_k\}$ with a stationary limit point $\underline{x}^*$. Define a scalar error function: $e_k = \left\| \underline{x}_k - \underline{x}^* \right\|$ $e: \mathrm{R}^n \to R$

Rate of Convergence - 1

Rate of Convergence:

Here $\|\underline{x}\|$ is defined as any Holder p-norm defined by:

$$\|\underline{x}\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p}$$

Typically, $\|\underline{x}\|_1 = \sum_{i=1}^{n} |x_i|; \|\underline{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_n^2\right)^{1/2}$; $\|\underline{x}\|_\infty = \max_i |x_i|$

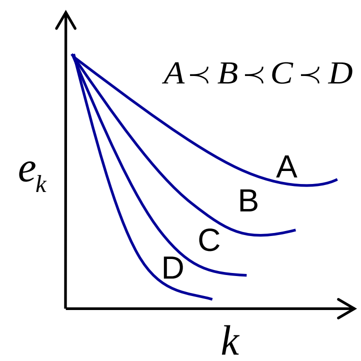
You may also define $e_k = |f(\underline{x}_k) - f(\underline{x}^*)|$

The behavior of e_k as a function of k is directly related to computational efficiency

Time complexity: cost per step * number of iterations

In order to investigate the behavior of e_k, we compare it to "standard" sequences. One standard form is to look for

$$e_{k+1} \cong \beta e_k^r \text{ as } k \to \infty$$



$r \triangleq$ order of convergence (or) asymptotic rate of convergence

$\beta \triangleq$ convergence ratio or asymptotic error constant

Rate of Convergence - 2

$$r \begin{cases} \nearrow 1 \text{ linear convergence (Geometric). Converges if } \beta < 1 \\ \rightarrow 2 \text{ quadratic (fast) convergence} \\ \searrow 3 \text{ cubic (superfast) convergence} \end{cases}$$

$$\text{If } \beta < 1,\ r = 1 \text{ linear } \Rightarrow \lim_{k\to\infty} \frac{e_{k+1}}{e_k} = \beta < 1$$

$$\beta = 1,\ r = 1 \text{ sublinear } \Rightarrow \lim_{k\to\infty} \frac{e_{k+1}}{e_k} = \beta = 1$$

$$\begin{cases} \beta = 0,\ r = 1 \text{ superlinear } \Rightarrow \lim_{k\to\infty} \frac{e_{k+1}}{e_k} = \beta = 0 \\ r > 1 \text{ superlinear } \Rightarrow 0 \le \lim_{k\to\infty} \frac{e_{k+1}}{e_k} < \infty \end{cases}$$

- Examples:

1) $e_{k+1} = \beta e_k\ ,\ \beta < 1 \Rightarrow$ binary search, golden section search, gradient method, regula falsi

2) $e_k = \frac{1}{k} \Rightarrow \frac{e_{k+1}}{e_k} = \frac{k}{k+1} \Rightarrow \beta = 1 \Rightarrow$ sublinear

3) $e_k = \left(\frac{1}{k}\right)^k \Rightarrow e_{k+1} = \left(\frac{1}{k+1}\right)^{k+1} = \frac{1}{k}\left(\frac{k}{k+1}\right)^{k+1} e_k \Rightarrow r = 1;\ \beta \to 0 \text{ as } k \to \infty \Rightarrow$ superlinear

Rate of Convergence - 3

- Examples:

4) $e_k = q\beta^k,\ \ q\left(\beta+\frac{1}{k}\right)^k,\ q\left(\beta-\frac{1}{k}\right)^k,\ q\beta^{k+\frac{1}{k}} \Rightarrow$ linear

5) $e_{2k} = \beta_1^k\beta_2^k;\quad e_{2k+1} = \beta_1^{k+1}\beta_2^k \quad e_{2k+2} = \beta_1^{k+1}\beta_2^{k+1}$

$$\lim_{k\to\infty}\frac{e_{2k+2}}{e_{2k}} = \beta_1\beta_2 \quad\Rightarrow\quad r=1 \text{ and } \beta=\sqrt{\beta_1\beta_2}$$

6) $e_k = a^{2^k} \Rightarrow e_{k+1} = e_k^2 \Rightarrow r=2 \Rightarrow$ quadratic (Newton's Method)

7) $e_{k+1} = Me_k^{\tau};\quad \tau = 1.618$ Golden section number

$r>1 \Rightarrow$ superlinear convergence rate

Examples: secant method, quadratic fit $(\tau = 1.3)$

8) $e_k = a^{2^{-k}}-1; a>0 \Rightarrow$ linear and $\beta = \frac{1}{2}$

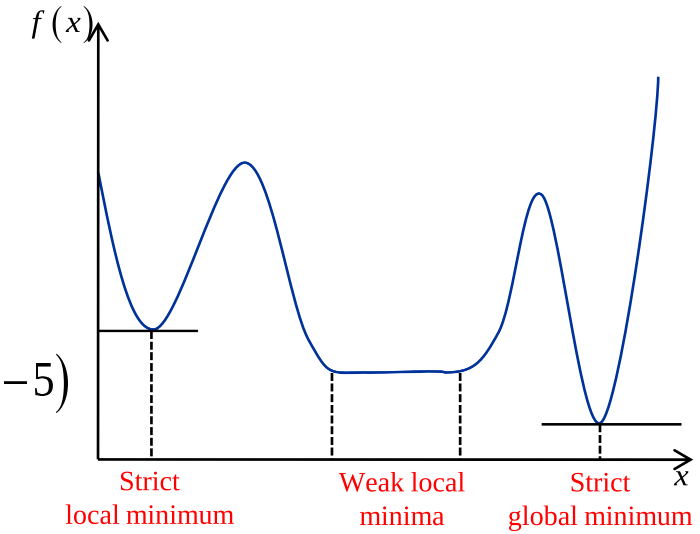
since $\left(a^{2^{-(k+1)}}-1\right)\left(a^{2^{-(k+1)}}+1\right) = \left(a^{2^{-k}}-1\right)$

$$\lim_{k\to\infty}\frac{e_{k+1}}{e_k} = \lim_{k\to\infty}\frac{1}{1+a^{2^{-(k+1)}}} = \frac{1}{2}$$

Most of the methods that we discuss will have $1\le r\le 2$

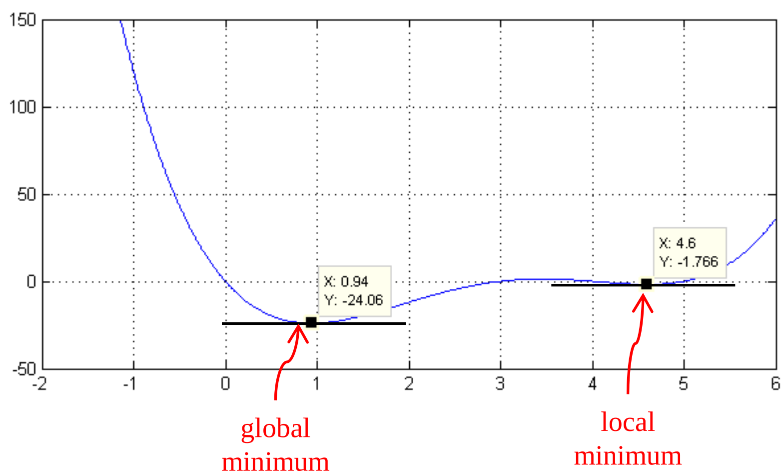
Static Question: Necessary and Sufficient Conditions for Minimum-1

- Minima
  - Local or relative
    - – weak $\Rightarrow$ several equivalent minima
    - – strong (strict) $\Rightarrow$ strict local minimum
  - Global
    - – weak
    - – strong (strict)



- Example:

$$f(x) = x^4 - 12x^3 + 47x^2 - 60x = x(x-3)(x-4)(x-5)$$



Extends to multivariable functions readily

Static Question: Necessary and Sufficient Conditions for Minimum-2

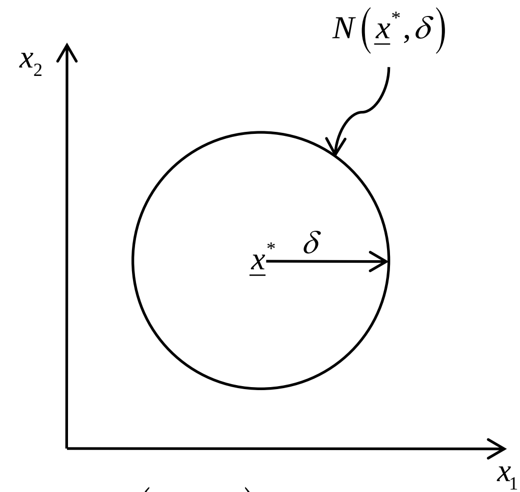
- Definition : $\underline{x}^* \in \Omega$ is a local minimum of $f(\underline{x})$ over Ω if for some δ>0, we have

$$f(\underline{x}^*) \le f(\underline{x}) \forall \underline{x} \in \Omega \text{ and } \|\underline{x} - \underline{x}^*\| \le \delta$$

$$(\text{or})\ f(\underline{x}^*) \le f(\underline{x}) \forall \underline{x} \in \Omega \cap N(\underline{x}^*, \delta)$$

$$N(\underline{x}^*, \delta) = \delta - \text{neighbourhood of } \underline{x}^*$$

$$N(\underline{x}^*, \delta) = \{\underline{x} : \|\underline{x} - \underline{x}^*\| \le \delta\}$$



- Remark: strict local minimum if $f(\underline{x}^*) < f(\underline{x}) \forall \underline{x} \in \Omega \cap N(\underline{x}^*, \delta) \setminus \underline{x}^*$

- Definition : $\underline{x}^* \in \Omega$ is weak (strict) global minimum of $f(\underline{x})$ over $\underline{x} \in \Omega$

$$\text{if } f(\underline{x}^*) \le f(\underline{x}) \left(f(\underline{x}^*) < f(\underline{x}) \right) \forall \underline{x} \in \Omega$$

- Note : strict global minimum $\Rightarrow$ strict local minimum

  strict local minimum $\not\Rightarrow$ strict global minimum except for convex functions

Optimality Conditions of Univariate Functions: Necessary Conditions

- For univariate functions:
  - Tangent is horizontal $\Rightarrow$ slope $\left.\frac{df}{dx}\right|_{x=x^*} = f'(x^*) = 0 \Rightarrow$ 1^{st} order condition
  - Curvature up $\Rightarrow$ second derivative $\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} \geq 0 \Rightarrow$ 2^{nd} order condition

- Proof: Suppose x^* is a local minimum. Let $y = x^* + \delta x$. Then, by the mean value theorem

$$f(y) = f(x^* + \delta x) = f(x^*) + f'(x^*)\delta x + \frac{1}{2} f''(x^* + \alpha \delta x)\delta x^2$$

Suppose $f'(x^*) \neq 0$. Then pick $\delta x = -\varepsilon f'(x^*)$; ε sufficiently small

$$\Rightarrow f(y) - f(x^*) = -\varepsilon\left[f'(x^*)\right]^2 + \frac{1}{2}\varepsilon^2\left[f'(x^*)\right]^2 f''(x^*) < 0$$

a contradiction $\Rightarrow$ need $f'(x^*) = 0$

From the first order condition, we have $f(x^* + \delta x) = f(x^*) + \frac{1}{2} f''(x^* + \alpha \delta x)\delta x^2; 0 < \alpha < 1$

if $f''(x^*) < 0 \Rightarrow f''(x^* + \alpha \delta x) < 0$ for some small α by continuity

$f(x^* + \delta x) < f(x^*) \Rightarrow$ a contradiction $\therefore f''(x^*) > 0$

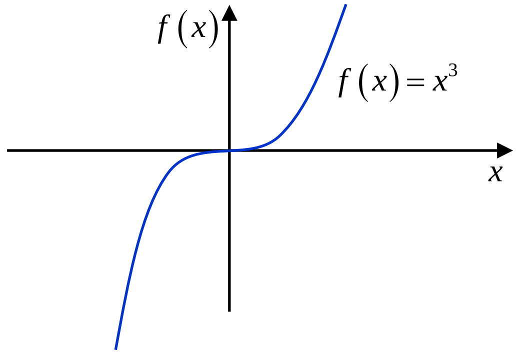
Optimality Conditions of Univariate Functions: Remarks

- For univariate functions:
  1. The proof provides a method of advancing from one x to the next.

     Take a step of $-\varepsilon f'(x)$ s.t. $f\left(x-\varepsilon f'(x)\right) < f(x)$

     Steepest descent or Gradient or Cauchy Method.
  2. These are only necessary conditions. They are not sufficient.

     Example: $f(x)=x^3;\ f'(0)=f''(0)=0$

     Not a local minimum, such point is called a saddle point or point of inflection.
  3. Note that first order condition is satisfied by minima, maxima and saddle point. Such points are refered to as stationary points.

Sufficient Conditions of Optimality for a Univariate Function

- For univariate functions:

$$(i) \quad f'(x^*) = 0$$

$$(ii) \quad f''\left(x^*\right) > 0$$

- (i) was proven earlier. To show (ii), note that $f\left(x^* - \delta x\right) > f\left(x^*\right)$ only if $f''\left(x^* + \alpha\delta x\right) > 0$ wich by continuity implies that $f''\left(x^*\right) > 0$.
- The above results extend directly to multivariable functions, i.e., functions of several variables.
- Assume $f\left(\underline{x}\right) \in C^2 \Rightarrow \frac{\partial f}{\partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exist and are continuous

Univariate		Multivariate
derivative	$\leftrightarrow$	gradient (vector of first order partial derivatives)
second derivative	$\leftrightarrow$	Hessian (Matrix of second order partial derivatives)

Conditions of Optimality for a Multivariate Function-1

- Gradient:

$$\nabla \underline{f}(\underline{x}) = \underline{g}(\underline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix};$$

$$\frac{\partial f}{\partial x_i} = \lim_{\delta \to 0} \frac{f(x_1, x_2, \ldots, x_i + \delta, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{\delta}$$

$$= \lim_{\delta \to 0} \frac{f(\underline{x} + \underline{e}_i \delta) - f(\underline{x})}{\delta}$$

Rate of change of f along the x_i direction

(or) slope of the tangent line along x_i

(or) direction of increase in f at $\underline{x}$

- Example: $f(\underline{x}) = x_1 x_2^{\ 2} + x_2 \cos(x_1)$

$$\nabla \underline{f}(\underline{x}) = \begin{bmatrix} x_2^{\ 2} - x_2 \sin(x_1) \\ 2x_1 x_2 + \cos(x_1) \end{bmatrix}; \quad \nabla \underline{f}(\underline{x})\Big|_{\substack{x_1 = \pi/2 \\ x_2 = 1}} = \begin{bmatrix} o \\ \pi \end{bmatrix}; \; f(\underline{x}) = \pi/2$$

Conditions of Optimality for a Multivariate Function-2

- **Hessian:**

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] = F(\underline{x}) \text{ Hessian } \Rightarrow n \text{ x } n \text{ matrix}$$

$$\text{since } \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \Rightarrow F(\underline{x}) \text{ is symmetric } \Rightarrow F(\underline{x}) = F^T(\underline{x})$$

$$f_{ij} = f_{ji} \quad \boxed{\text{Need only } \frac{n(n+1)}{2} \text{ elements}}$$

- **Example:**

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -x_2 \cos(x_1) & 2x_2 - \sin(x_1) \\ 2x_2 - \sin(x_1) & 2x_1 \end{bmatrix}\Bigg|_{\substack{x_1 = \pi/2 \\ x_2 = 1}} = \begin{bmatrix} 0 & 1 \\ 1 & \pi \end{bmatrix} < 0$$

- **Example:** A quadratic function

$$f(x) = \frac{1}{2}\underline{x}^T Q \underline{x} + \underline{b}^T \underline{x} + c = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c$$

$$\nabla \underline{f}(\underline{x}) = Q\underline{x} + \underline{b}; \quad \nabla^2 f(x) = Q$$

Summary of Conditions of Optimality for a Multivariate Function-3

Necessary conditions

1. $\nabla \underline{f}\left(\underline{x}^*\right)=\underline{0}$
2. $\nabla^2 f\left(\underline{x}^*\right)\geq 0\ \left(PSD\right)$

1. A symetric matrix A is PSD iff
   $\underline{x}^T A\underline{x}\geq 0\ \forall \underline{x}\in R^n \Rightarrow \lambda_i\left(A\right)\geq 0$
   $$\Downarrow$$
   All principal minors have non-negative determinants
   $$\Downarrow$$
   Matrix A can be factored as $A=LDL^T$
   $d_i\geq 0;\ \ L$ unit lower Δ

Sufficient conditions

1. $\nabla \underline{f}\left(\underline{x}^*\right)=\underline{0}$
2. $\nabla^2 f\left(\underline{x}^*\right)>0\ \left(PD \text{ matrix}\right)$

   A symetric matrix A is PD iff
   $\underline{x}^T A\underline{x}>0\ \forall \underline{x}\in R^n \wedge\ \underline{x}\neq \underline{0}\Rightarrow \lambda_i\left(A\right)>0$
   $$\Downarrow$$
   All principal minors have positive determinants
   $$\Downarrow$$
   $A=LDL^T; d_i>0;\ \ O\left(\frac{n^3}{6}\right)$ Computation

2. For any symmetric matrix A with $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$, we have
   $\lambda_n\underline{x}^T\underline{x}\leq\underline{x}^T A\underline{x}\leq\lambda_1\underline{x}^T\underline{x}$ Rayleigh inequality

- PD: Positive Definite
- PSD: Positive Semi-definite

Proof of Optimality Conditions - 1

- Proof necessity:

From the mean value theorem, we have for any $\underline{x}$ and $\underline{y}$

$$f(\underline{y}) = f(\underline{x}) + \nabla \underline{f}^T(\underline{x})(\underline{y}-\underline{x}) + \frac{1}{2}(\underline{y}-\underline{x})^T \nabla^2 f\left(\underline{x}+\beta(\underline{y}-\underline{x})\right)(\underline{y}-\underline{x}); \beta \in (0,1)$$

Take $\underline{x} = \underline{x}^*$, $\underline{y} = \underline{x} + \alpha \underline{d}$ where $\|\underline{d}\| = 1$ for any norm (usually 1,2,∞)

$$\Rightarrow f(\underline{x}^* + \alpha \underline{d}) = f(\underline{x}^*) + \alpha \nabla \underline{f}^T(\underline{x}^*)\underline{d} + \frac{1}{2}\alpha^2 \underline{d}^T \nabla^2 f(\underline{x}^* + \beta\alpha\underline{d})\underline{d} = g(\alpha)$$

If $\underline{x}^*$ is a minimum, the scalar function $g(\alpha)$ has minimum at $\alpha = 0 \Rightarrow g'(0) = 0$

$$\Rightarrow g'(0) = \nabla \underline{f}^T(\underline{x}^*)\underline{d} = f'(\underline{x}, \underline{d}) \quad \forall\ \underline{d} \in R^n$$

Taking $\underline{d} = \underline{e}_1 \Rightarrow \left.\frac{\partial f}{\partial x_1}\right|_{\underline{x}^*} = 0$ similarly $\left.\frac{\partial f}{\partial x_i}\right|_{\underline{x}^*} = 0$ since $\underline{d}$ is arbitrary

$\Rightarrow \nabla f(\underline{x}^*) = \underline{0} \Rightarrow \|\nabla f(\underline{x}^*)\| = 0$ 1^{st} order condition. so, norm will be small near minimum.

For a local minimum, we also need

$$\frac{d^2 g(0)}{d\alpha^2} \geq 0 \Rightarrow \underline{d}^T \nabla^2 f^T(\underline{x}^*)\underline{d} \geq 0\ \forall\ \underline{d} \in R^n \Rightarrow \nabla^2 f(\underline{x}^*) \text{ is } PSD$$

Proof of Optimality Conditions - 2

- **Sufficiency:**

Suppose $\nabla^2 f(\underline{x}^*) > 0 \Rightarrow$ smallest eigenvalue λ_n of $\nabla^2 f(\underline{x}^*) > 0$

$$f(\underline{x}^* + \alpha \underline{d}) = f(\underline{x}^*) + \frac{1}{2}\alpha^2 \underline{d}^T \nabla^2 f(\underline{x}^* + \beta\alpha\underline{d})\underline{d};\ \beta \in (0,1)$$

For sufficiently small α, $\nabla^2 f(\underline{x}^* + \beta\alpha\underline{d}) > 0$ if $\nabla^2 f(\underline{x}^*) > 0$

Let λ_n be the smallest eigenvalue of $\nabla^2 f(\underline{x}^* + \beta\alpha\underline{d})$. Then

$$f(\underline{x}^* + \alpha \underline{d}) - f(\underline{x}^*) = \frac{1}{2}\alpha^2 \underline{d}^T \nabla^2 f(\underline{x}^* + \beta\alpha\underline{d})\underline{d} \geq \frac{\alpha^2}{2}\lambda_n \|d\|^2$$ ....Recall Rayleigh inequality

$\Rightarrow \underline{x}^*$ is a strict local minimum.

- Note: Strict local maximum if $\nabla^2 f(\underline{x}^*) < 0$ and saddle point if $\nabla^2 f(\underline{x}^*)$ is indefinite.

- Example: $f(x_1, x_2) = x_1^2 - 6x_1 + x_2^2 + 4x_2 + 5$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 6 \\ 2x_2 + 4 \end{bmatrix} \Rightarrow \begin{matrix} x_1^* = 3 \\ x_2^* = -2 \end{matrix};\ \nabla^2 f(\underline{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0 \Rightarrow$$

Strict local minimum
(It is also global minimum. Why?)

- Example: $f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 14x_1 - 2x_2^2 + 22x_2 - 8$

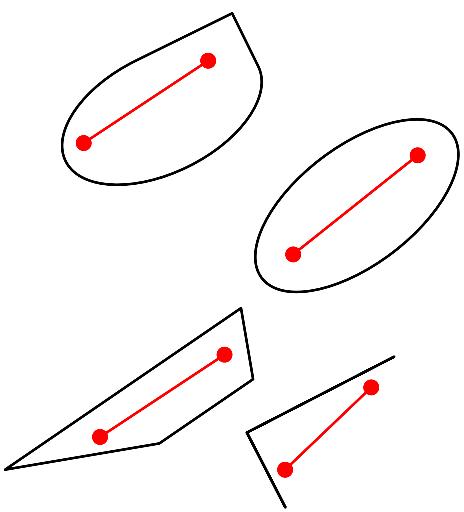
$$\nabla f(x_1, x_2) = \begin{bmatrix} 4x_1 + 2x_2 + 14 \\ 2x_1 - 4x_2 + 22 \end{bmatrix} \Rightarrow \underline{x}^* = \begin{bmatrix} -5 \\ 3 \end{bmatrix};\ \nabla^2 f(\underline{x}^*) = \begin{bmatrix} 4 & 2 \\ 2 & -4 \end{bmatrix}$$

$\lambda_1 = \sqrt{20}, \lambda_2 = -\sqrt{20}$
$\Rightarrow$ Indefinite
$\Rightarrow$ Saddle point

Convex Sets -1

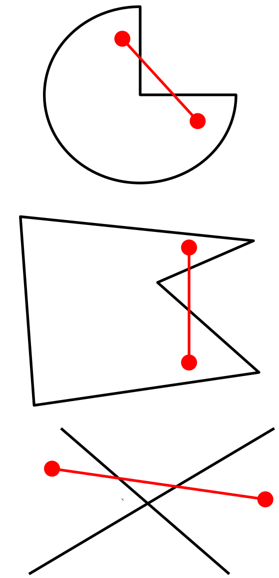
- Important because local optimum $\Leftrightarrow$ global optimum
- Definition : A set $\Omega \in R^n$ convex if for *any* two points $\underline{x}_1, \underline{x}_2 \in \Omega$ and $\forall \alpha \in (0,1)$ we have $\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2 \in \Omega$. In words, Ω is convex if for every two points $\underline{x}_1$ and $\underline{x}_2$, the line segment joining $\underline{x}_1$ and $\underline{x}_2$ is also in Ω.

• Convex



A convex set is one whose boundaries do not bulge inward (or) do not have indentations.

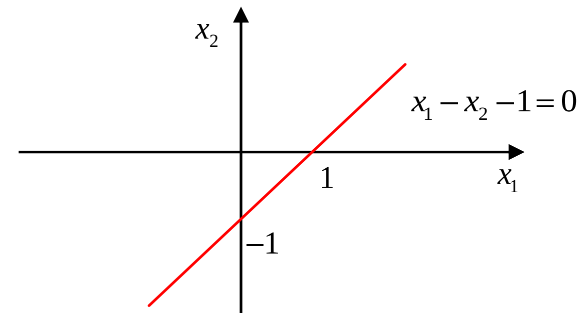
• Nonconvex



Convex Sets -2

❑ Examples:

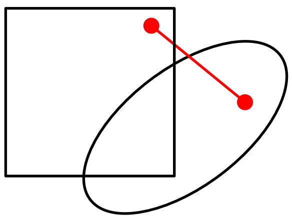
1. A hyperplane $\underline{a}^T \underline{x} = \underline{b}$ is convex.



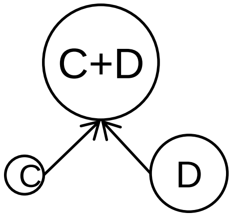
2. Half spaces $H_+ = \{\underline{x} : \underline{a}^T \underline{x} \geq b\}$

   or $H_- = \{\underline{x} : \underline{a}^T \underline{x} \leq b\}$ are convex.

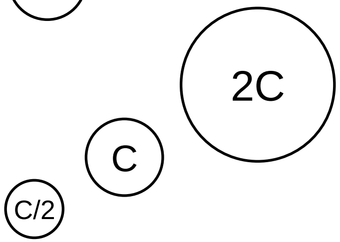
3. $\cap\ c_i$ convex. $\cup\, c_i$ Need not be



4. Sum or difference of convex sets is convex.



5. Expansions or contraction of convex sets is convex.



6. Empty set is convex (by definition).

Convex Functions - 1

Convex Functions

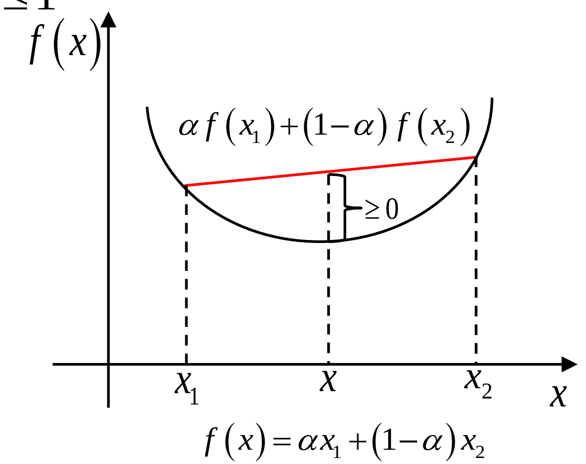
Consider $f(\underline{x}):\Omega \to R;\quad f(\underline{x})$ is a sacalar multivariable function.

$f(\underline{x})$ is a convex function on a convex set Ω if for any two points

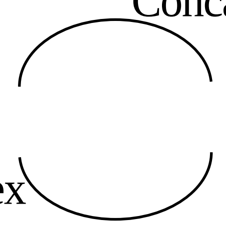
$\underline{x}_1$ and $\underline{x}_2 \in \Omega$

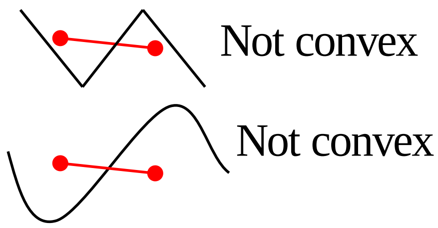
$$f\left(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2\right) \le \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2)\ \forall 0 \le \alpha \le 1$$

- A convex function bends up
- A line segment (chord, secant) between any two points never lies below the graph
- Linear interpolation between any two points x_1 and x_2 overestimates the function



$\Rightarrow$ • Concave if $-f(x)$ is convex.





Convex Functions - 2

- Examples:

1. A linear function is convex $f(\underline{x}) = \underline{c}^T \underline{x}$

$$f(\alpha \underline{x} + (1-\alpha)\underline{x}_2) = \alpha \underline{c}^T \underline{x}_1 + (1-\alpha)\underline{c}^T \underline{x}_2 = \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2)$$

2. A quadratic function $\underline{x}^T Q \underline{x}$ is convex if Q is *PSD*.

$$\text{①}\; f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) = \alpha^2 \underline{x}_1^T Q \underline{x}_1 + \alpha(1-\alpha)\left(\underline{x}_1^T Q \underline{x}_2 + \underline{x}_2^T Q \underline{x}_1\right) + (1-\alpha)^2 \underline{x}_2^T Q \underline{x}_2$$

$$\text{②}\; \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2) = \alpha \underline{x}_1^T Q \underline{x}_1 + (1-\alpha)\underline{x}_2^T Q \underline{x}_2$$

$$\text{①} - \text{②} = -\alpha(1-\alpha)\left[\underline{x}_1^T Q \underline{x}_1 + \underline{x}_2^T Q \underline{x}_2 - \underline{x}_1^T Q \underline{x}_2 - \underline{x}_2^T Q \underline{x}_1\right]$$

$$= -\alpha(1-\alpha)\;(\underline{x}_1 - \underline{x}_2)^T Q (\underline{x}_1 - \underline{x}_2) \le 0 \text{ iff } Q \text{ is } PSD$$

3. In general $f\left(\sum_i \alpha_i \underline{x}_i\right) \le \sum_i \alpha_i f(\underline{x}_i);\quad \sum_i \alpha_i = 1;\ \alpha_i \ge 0$

JENSEN'S INEQUALITY $\quad f[E(x)] \le E\{f(x)\}$

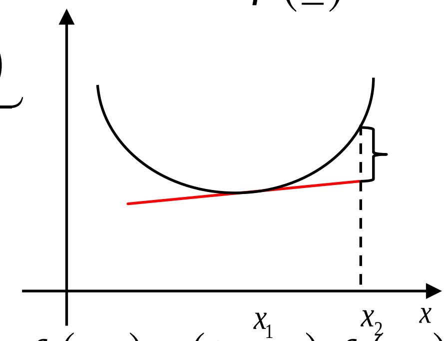
Convex Functions - 3

- Examples (cont'd):

4. The linear extrapolation at a point underestimates a convex function $f(\underline{x})$

$$\text{assume } f(\underline{x}) \in C^2;\; f(\underline{x}_2) \geq f(\underline{x}_1) + \underbrace{\nabla f^T(\underline{x}_1)(\underline{x}_2 - \underline{x}_1)}$$

Defines the tangent plane at $\underline{x}_1$



- Proof:

$$(\text{only if})\; f(\underline{x}) \text{ is convex } \Rightarrow f(\alpha \underline{x}_2 + (1-\alpha)\underline{x}_1) \leq \alpha f(\underline{x}_2) + (1-\alpha) f(\underline{x}_1)$$

$$\lim_{\alpha \to 0} \frac{f(\underline{x}_1 + \alpha(\underline{x}_2 - \underline{x}_1)) - f(\underline{x}_1)}{\alpha} \leq f(\underline{x}_2) - f(\underline{x}_1)$$

$$\nabla \underline{f}^T(\underline{x}_1)(\underline{x}_2 - \underline{x}_1) \leq f(\underline{x}_2) - f(\underline{x}_1)$$

$$(\text{If})\; \text{Assume result is true at } \underline{x}_1 \text{ and } \underline{x}_2 \ni \underline{x}_0 = \alpha \underline{x}_2 + (1-\alpha)\underline{x}_1$$

$$(1-\alpha) f(\underline{x}_1) \geq (1-\alpha)\left[f(\underline{x}_0) + \nabla \underline{f}^T(\underline{x}_0)(\underline{x}_1 - \underline{x}_0) \right]$$

$$\alpha f(\underline{x}_2) \geq \alpha \left[f(\underline{x}_0) + \nabla \underline{f}^T(\underline{x}_0)(\underline{x}_2 - \underline{x}_0) \right]$$

$$\Rightarrow \alpha f(\underline{x}_2) + (1-\alpha) f(\underline{x}_1) \geq f(\underline{x}_0) + \nabla \underline{f}^T(\underline{x}_0)\left[(1-\alpha)\underline{x}_1 + \alpha \underline{x}_2 - \underline{x}_0 \right]$$

Convex Functions - 4

- Examples:

5. $f(\underline{x})$ convex $\in C^2 \Leftrightarrow \nabla^2 f(\underline{x})$ is PSD over $\underline{x} \in \Omega(\Omega = \text{convex})$

   (only if) $f(\underline{x}_2) = f(\underline{x}_1) + \nabla \underline{f}^T(\underline{x}_1)(\underline{x}_2 - \underline{x}_1) + \frac{1}{2}(\underline{x}_2 - \underline{x}_1)^T \nabla^2 f(\underline{x}_1 + \alpha(\underline{x}_2 - \underline{x}_1))(\underline{x}_2 - \underline{x}_1)$

   $\nabla^2 f(\underline{x}_1) \geq 0 \Rightarrow \nabla^2 f(\underline{x}_1 + \alpha(\underline{x}_2 - \underline{x}_1)) \geq 0$ for sufficiently small α

   $\Rightarrow f(\underline{x}_2) \geq f(\underline{x}_1) + \nabla \underline{f}^T(\underline{x}_1)(\underline{x}_2 - \underline{x}_1) \Rightarrow f(\underline{x})$ is convex

   (If): Suppose $\nabla^2 f(\underline{x}_1) < 0 \Rightarrow$ can find $N(\underline{x}^*, \delta) \ni$

   $(\underline{x}_2 - \underline{x}_1)^T \nabla^2 f(\underline{x}_1 + \alpha(\underline{x}_2 - \underline{x}_1))(\underline{x}_2 - \underline{x}_1) < 0$

   $\Rightarrow \quad f(\underline{x}_2) \leq f(\underline{x}_1) + \nabla \underline{f}^T(\underline{x}_1)(\underline{x}_2 - \underline{x}_1)$ a contradiction

6. Sum of convex functions is convex

7. The epigraph or the level set $\Omega_\mu = \{\underline{x} : f(\underline{x}) \leq \mu\}$

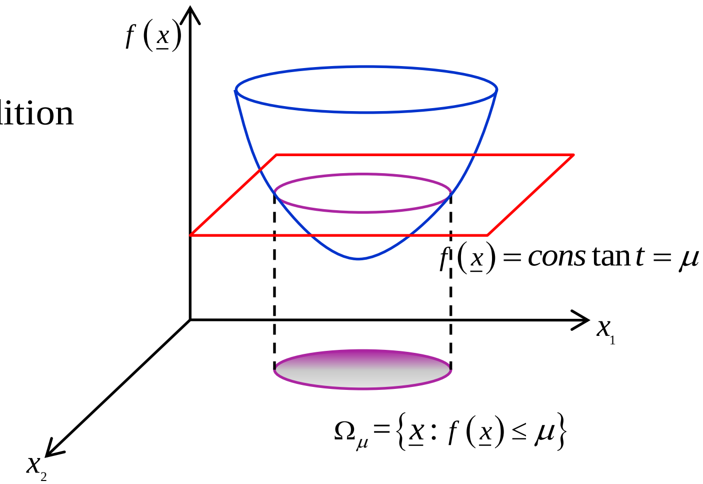
   is convex for all μ if $f(\underline{x})$ is convex.

   Proof: Let $\underline{x}_1$ and $\underline{x}_2 \in \Omega_\mu \Rightarrow f(\underline{x}_1) \leq \mu$

   and $f(\underline{x}_2) \leq \mu$;

   $f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) \leq \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2) \leq \mu$

   $\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2 \in \Omega_\mu$



Convex Functions - 5

- Examples:

8. Convex programming problem

$\min f(\underline{x})$ — $f(\underline{x})$ convex

s.t. $A\underline{x} = \underline{b}$ — $g_i(\underline{x})$ concave $\Rightarrow -g_i(\underline{x})$ convex

$g_i(\underline{x}) \geq 0$

$\Omega_i = \{\underline{x} : -g_i(\underline{x}) \leq 0\} = \{\underline{x} : g_i(\underline{x}) \geq 0\}$ convex; $\Omega_\mu = \{\underline{x} : f(\underline{x}) \leq \mu\}$ convex

$A\underline{x} = \underline{b}$ intersection of hyperplanes $\Rightarrow$ convex set A

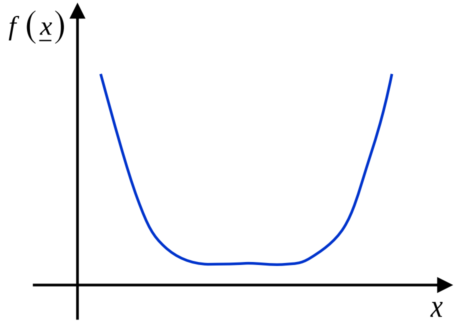
$\Omega = \bigcap_i \Omega_i \cap \Omega_\mu \cap \mathsf{A}$ convex

9. Local optimum $\Leftrightarrow$ global optimum

global $\Rightarrow$ local is always true!!!

To prove local $\Rightarrow$ global, let $\underline{x}^*$ be a local minimum,

but y is a global minimum.



Consider $\underline{x} = \alpha \underline{x}^* + (1-\alpha)\underline{y} \in \Omega$

Convexity $\Rightarrow f\left(\alpha \underline{x}^* + (1-\alpha)\underline{y}\right) \leq \alpha f\left(\underline{x}^*\right) + (1-\alpha) f(y) \leq f\left(\underline{x}^*\right) \forall \alpha$

$\Rightarrow \underline{x}^*$ can not be a local minimum, a contradiction.

As a worst case, local minima must be bunched together as shown.

Convex Functions - 6

- Examples:

10. First order necessary condition is also sufficient

$$f(\underline{x}) \geq f(\underline{x}^*) + \nabla f^T(\underline{x}^*)(\underline{x} - \underline{x}^*) = f(\underline{x}^*) \forall \underline{x} \in R^n$$

11. $f(\underline{x})$ is convex iff the scalar function $g(\alpha) = f(\underline{x} + \alpha \underline{d})$ is convex $\forall \underline{x}$ and $\underline{d}$.

12. Since near $\underline{x}^*$, $\nabla^2 f(\underline{x}^*) \geq 0$, we can apply convex analysis locally.

In addition, from Taylor series, for $\underline{x}$ near $\underline{x}^*$

$$f(\underline{x}^*) \cong f(\underline{x}) + \nabla \underline{f}^T(\underline{x})(\underline{x}^* - \underline{x}) + \frac{1}{2}(\underline{x}^* - \underline{x})^T \nabla^2 f(\underline{x})(\underline{x}^* - \underline{x})$$

$$= \underbrace{f(\underline{x}) - \nabla \underline{f}^T(\underline{x})\underline{x} + \frac{1}{2}\underline{x}^T \nabla^2 f(\underline{x})\underline{x}}_{c} + \underbrace{\left[\nabla f^T(\underline{x}) - \underline{x}^T \nabla^2 f(\underline{x})\right]}_{\underline{b}^T} \underline{x}^* + \frac{1}{2}\underline{x}^{*T} \underbrace{\nabla^2 f(\underline{x})}_{Q} \underline{x}^*$$

$$= c + \underline{b}^T \underline{x}^* + \frac{1}{2}\underline{x}^{*T} Q \underline{x}^*$$ A quadratic approximation near $\underline{x}^*$

Convex Functions - 7

- Example:

$$f(\underline{x}) = -\ln(1 - x_1 - x_2) - \ln x_1 - \ln x_2$$

$$\nabla f(\underline{x}) = \begin{bmatrix} \dfrac{1}{1 - x_1 - x_2} - \dfrac{1}{x_1} \\ \dfrac{1}{1 - x_1 - x_2} - \dfrac{1}{x_2} \end{bmatrix} = \underline{0} \Rightarrow \begin{matrix} 2x_1 + x_2 = 1 \\ x_1 + 2x_2 = 1 \end{matrix} \Rightarrow x_1 = x_2 = \frac{1}{3}$$

$$\nabla^2 f(\underline{x}) = \begin{bmatrix} \dfrac{1}{(1 - x_1 - x_2)^2} + \dfrac{1}{x_1^2} & \dfrac{1}{(1 - x_1 - x_2)^2} \\ \dfrac{1}{(1 - x_1 - x_2)^2} & \dfrac{1}{(1 - x_1 - x_2)^2} + \dfrac{1}{x_2^2} \end{bmatrix} > 0 \forall \Omega = \{\underline{x} : x_1 > 0, x_2 > 0, x_1 + x_2 < 1\}$$

Strictly Convex

Summary

- Abstract Definition of an Optimization Problem
- Classification of Optimization Problems
- Three Basic Questions of Optimization
  - Optimality conditions, algorithm, convergence
- Optimality Conditions for single variable and Multi-variable Functions
- Elementary Convexity Theory