Lecture 1

Historical Background, Optimization problem, Classification of optimization problems, and Mathematical background

Historical Background

* Initial work can be traced back to Newton, Lagrange and Cauchy
* Calculus of variation and minimization of functions were developed by Bernouli, Euler, Lagrange and Weirstrass
* Digital Computers motivated research in numerical optimization techiques.

Statement of An optimization problem

$$\min_{\underline{x}} \; f(\underline{x})$$

Subject to $\quad g_j(\underline{x}) \leq 0, \; j=1,2,\dots,m$

$\quad \ell_j(\underline{x}) = 0, \; j=1,2,\dots,p$

$f(\underline{x})$: objective function

$\underline{x} = [x_1\; x_2 \cdots x_n]^T \in R^n$: vector of parameters

$g_j(\underline{x})$: jth inequality Constraint

$\ell_j(\underline{x})$: jth equality Constraint

Example

minimize $f(x_1, x_2, x_3) = x_1^4 - 2x_1^2 x_2 + x_1^2 + x_1 x_2^2 - 2x_1 + 4$

subject to $x_1^2 + x_2^2 - 2 = 0$

$$0.25\,x_1^2 + 0.75\,x_2^2 - 1 + x_3 = 0$$

$$0 \leq x_1 \leq 5$$

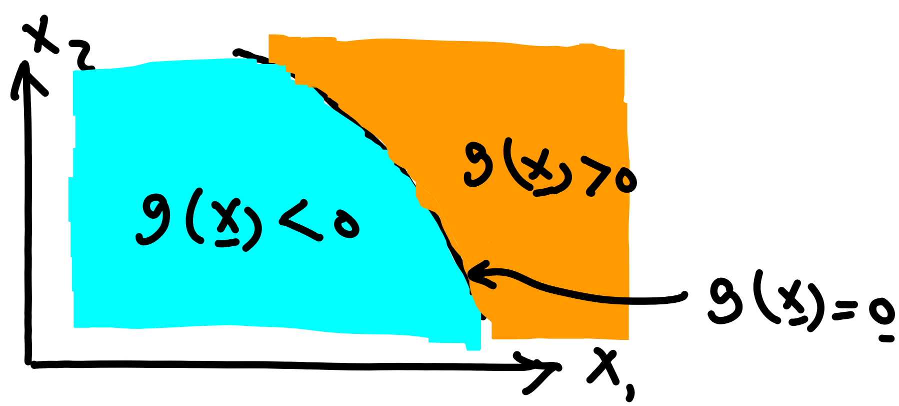
$$0 \leq x_2 \leq 5$$

$$x_3 \geq 0$$

Put in the standard form!

General Notations

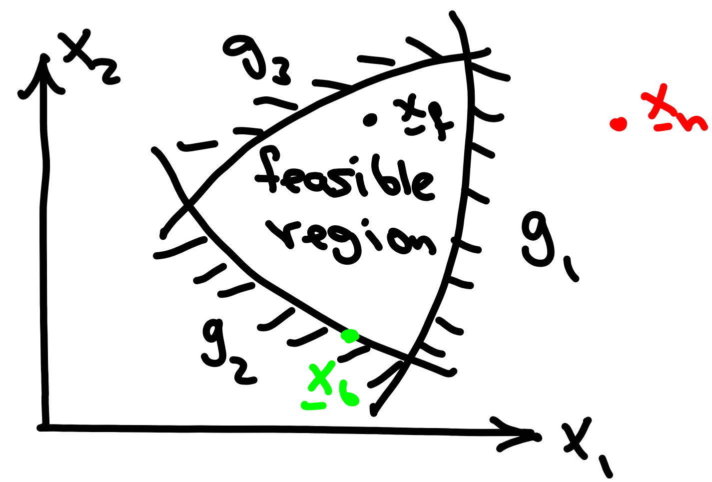
* A constraint divides the parameter space into 3 subsets $g(\underline{x}) > 0$, $g(\underline{x}) = 0$, and $g(\underline{x}) < 0$



A constraint is active at all points satisfying $g(\underline{x}) = 0$

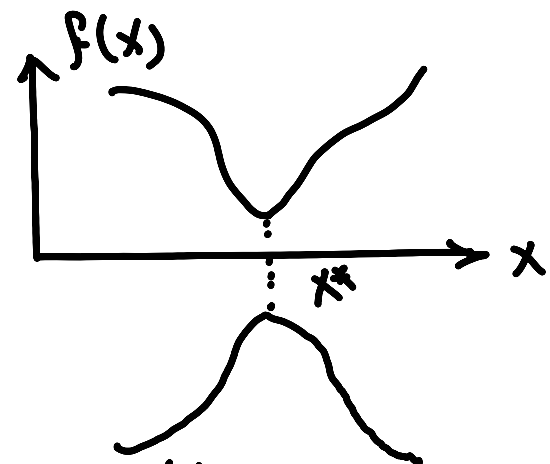
General Notation (Cont'd)

* region at which all constraints are satisfied is called the feasible region
* $\underline{x}_f$: a feasible point
* $\underline{x}_n$: a non feasible point
* $\underline{x}_b$: a boundary point
* We want to minimize $f(\underline{x})$ over feasible region



General Notation (Cont'd)

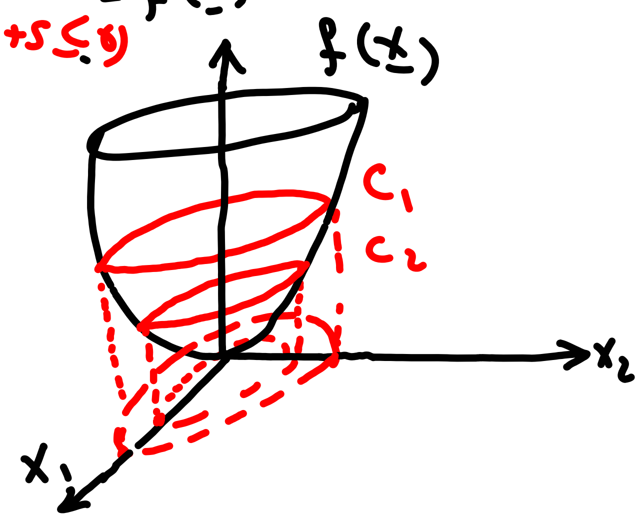
* $\min\limits_{\underline{x}} f(\underline{x}) \equiv \max\limits_{\underline{x}} (-f(\underline{x}))$



* $g(\underline{x}) \geq 0 \Rightarrow (-g(\underline{x})) \leq 0$

$(x_1 - x_2 + x_3 - 5 \geq 0 \Rightarrow -x_1 + x_2 - x_3 + 5 \leq 0)$

* $f(\underline{x}) = c$ is an objective function surface. For 2D they give objective function contours.



General Notation (Cont'd)

* The optimal solution $\underline{x}^*$ is given by

$$\underline{x}^* = \arg\min_{\underline{x}} f(\underline{x}) \text{ subject to}$$

$$g_j(\underline{x}) \leq 0, \quad j = 1, 2, \ldots, m$$

$$l_j(\underline{x}) = 0, \quad j = 1, 2, \ldots, p$$

* An inequality constraint can be transformed to equality constraint using slack variables

$$g(\underline{x}) \leq 0 \Rightarrow g(\underline{x}) + x_s = 0, \quad x_s \geq 0$$

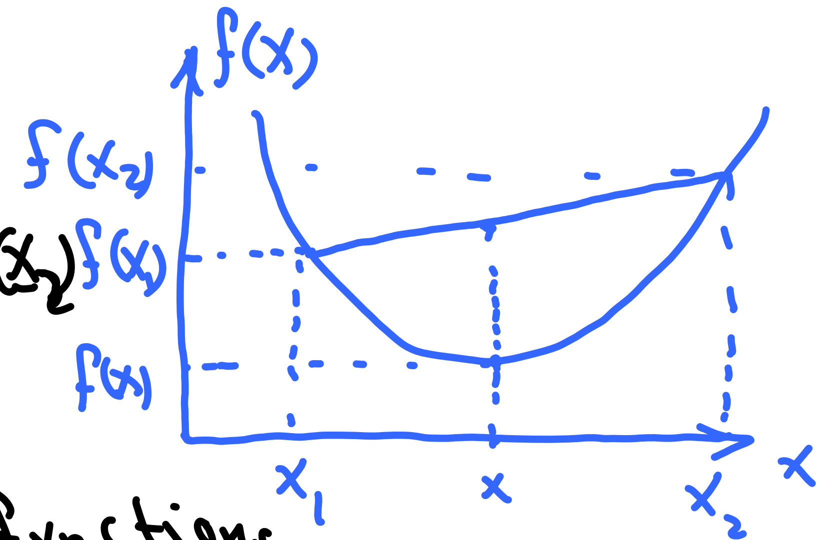
* Multiobjectives: $f(\underline{x}) = \alpha_1 f_1(\underline{x}) + \alpha_2 f_2(\underline{x})$, $\alpha_1, \alpha_2 > 0$

General Notation (Cont'd)

* A Convex function satisfies

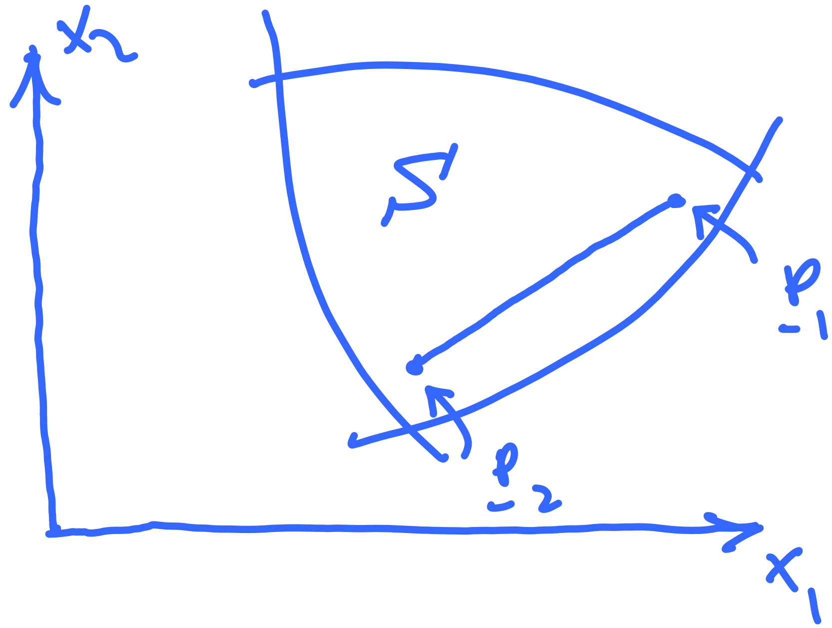
$$f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) \leq \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2)$$

$\alpha \in [0, 1]$



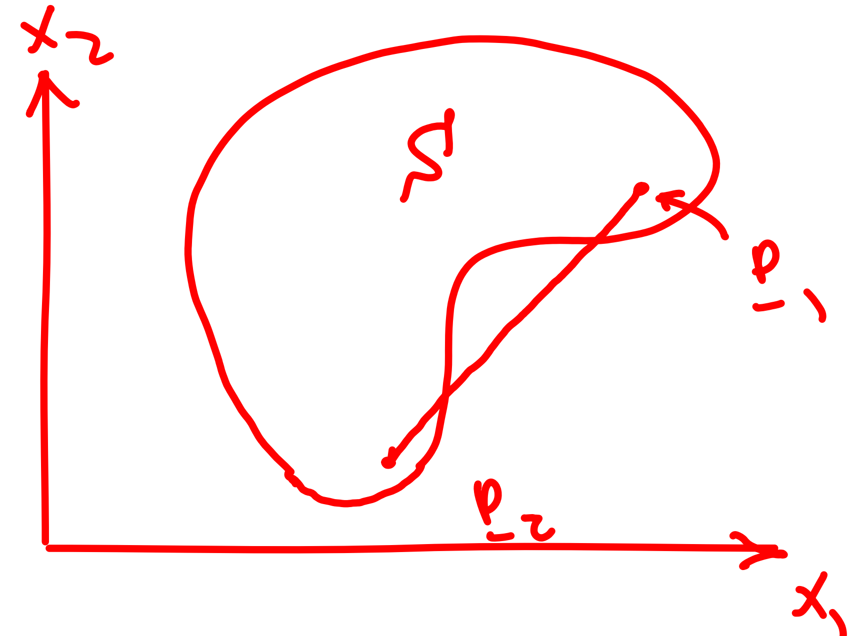
* Only ONE minimum of such functions
* A Convex region satisfies that if $\underline{x}_1$ is in the region and $\underline{x}_2$ is in the region then the line connecting them is also in the region.

General Notation (Cont'd)



$\alpha \underline{p}_1 + (1-\alpha) \underline{p}_2 \in S$

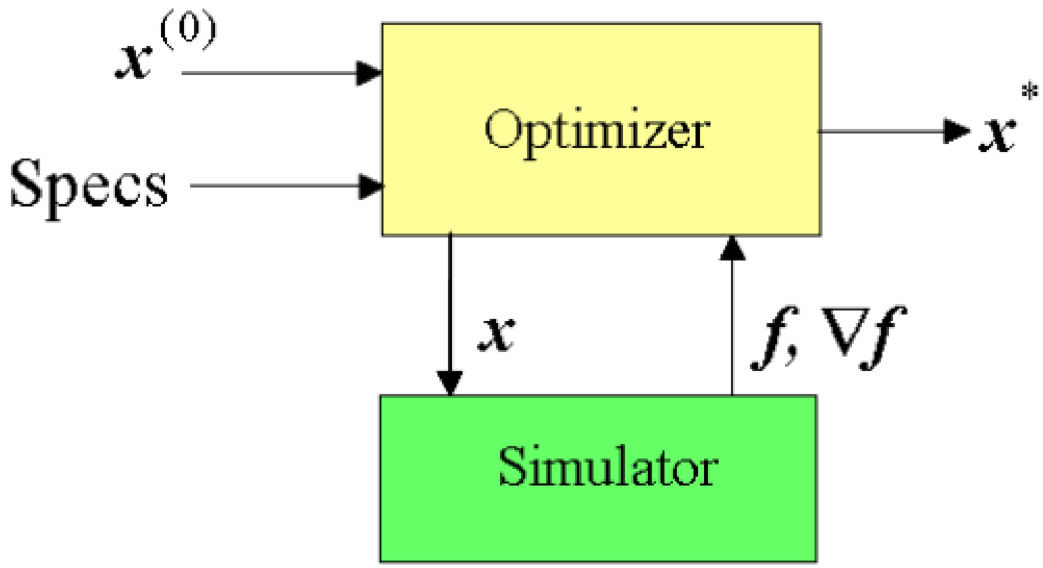
$\alpha \in [0,1]$



$\alpha \underline{p}_1 + (1-\alpha) \underline{p}_2 \notin S$

for some $\alpha \in [0,1]$

General Notation (Cont'd)



Parameters MUST have same order of magnitude!

General Notation (Cont'd)

* Scaling MUST be used when parameters have different order of magnitude

example: $f(\underline{x}) = f(x_1, \underline{x}_2) = 10^6 x_1^2 + (x_2-4)^2$ (ill-conditioned)

Define $\acute{x}_1 = 10^3 x_1$, $\acute{x}_2 = x_2 \Rightarrow \begin{bmatrix} \acute{x}_1 \\ \acute{\underline{x}}_2 \end{bmatrix} = \begin{bmatrix} 10^3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$f(\acute{x}_1, \acute{x}_2) = \acute{x}_1^2 + (\acute{x}_2 - 4)^2$ (well-conditioned)

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 10^{-3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \acute{x}_1^* \\ \acute{x}_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

Classification of Optimization Problems

1- According to Constraints

a) Constrained

$$\underline{x}^* = \arg\min_{\underline{x}} f(\underline{x})$$

subject to

$$g_j(\underline{x}) \leq 0, \quad j = 1, 2, \dots, m$$

$$\ell_j(\underline{x}) = 0, \quad j = 1, 2, \dots, p$$

b) Unconstained

$$\underline{x}^* = \arg\min_{\underline{x}} f(\underline{x})$$

* much earier to solve
* Feasibility Region is R^n

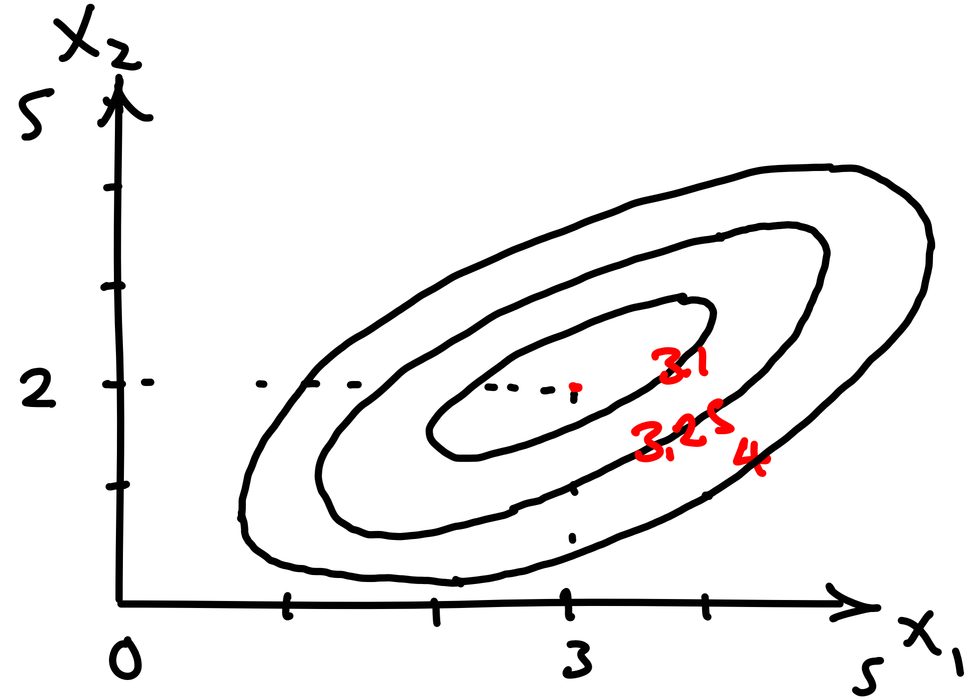
Example

$$\min_{\underline{x}} \; 3 + (x_1 - 1.5x_2)^2 + (x_2 - 2)^2$$

subject to

$$0 \le x_1 \le 5$$

$$0 \le x_2 \le 5$$



Draw Contours using MATLAB!

Classifications (Cont'd)

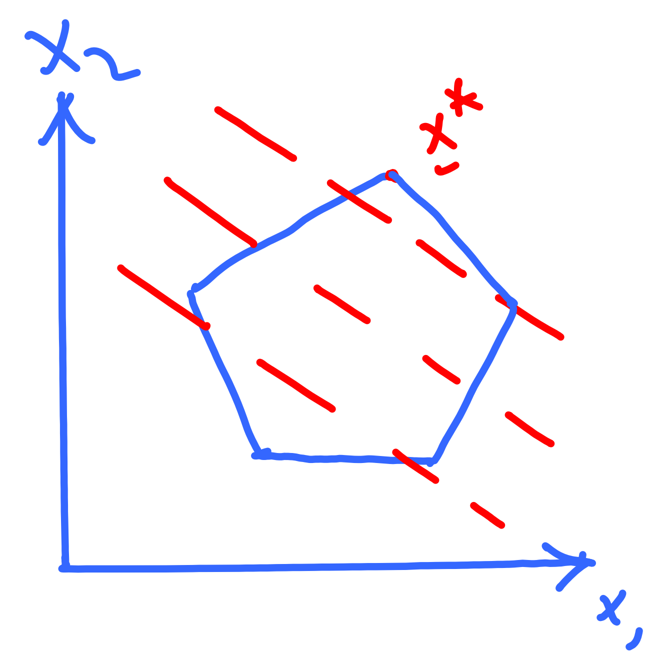
2) According to nature of functions

a) Linear programming

$$\min_{\underline{x}} \ \underline{c}^T \underline{x}$$

subject to $\underline{g}_j^T \underline{x} \leq b_j, \ j = 1, 2, \ldots, m$

$\underline{b}_j^T \underline{x} = 0, \ j = 1, 2, \ldots, p$

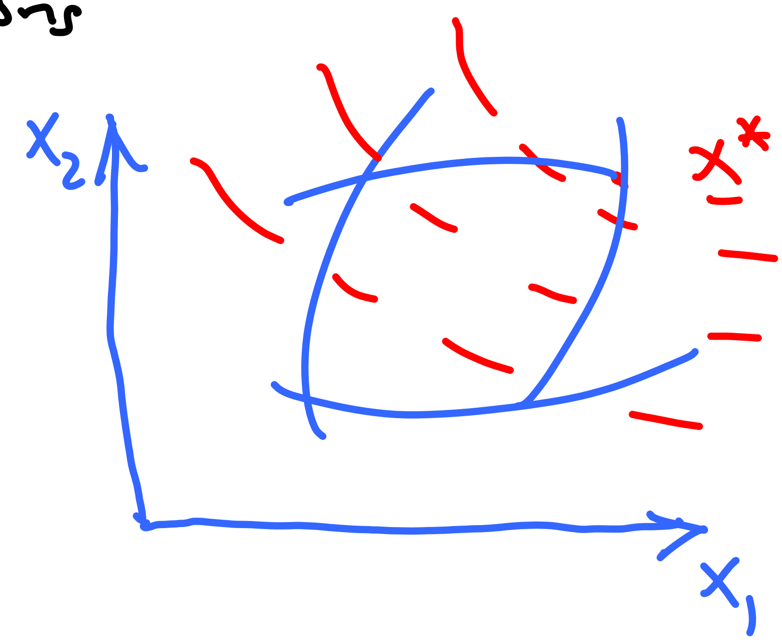


Classifications (Cont'd)

2) According to nature of functions

b) Nonlinear program

Either the objective function or some of the constraints are nonlinear.

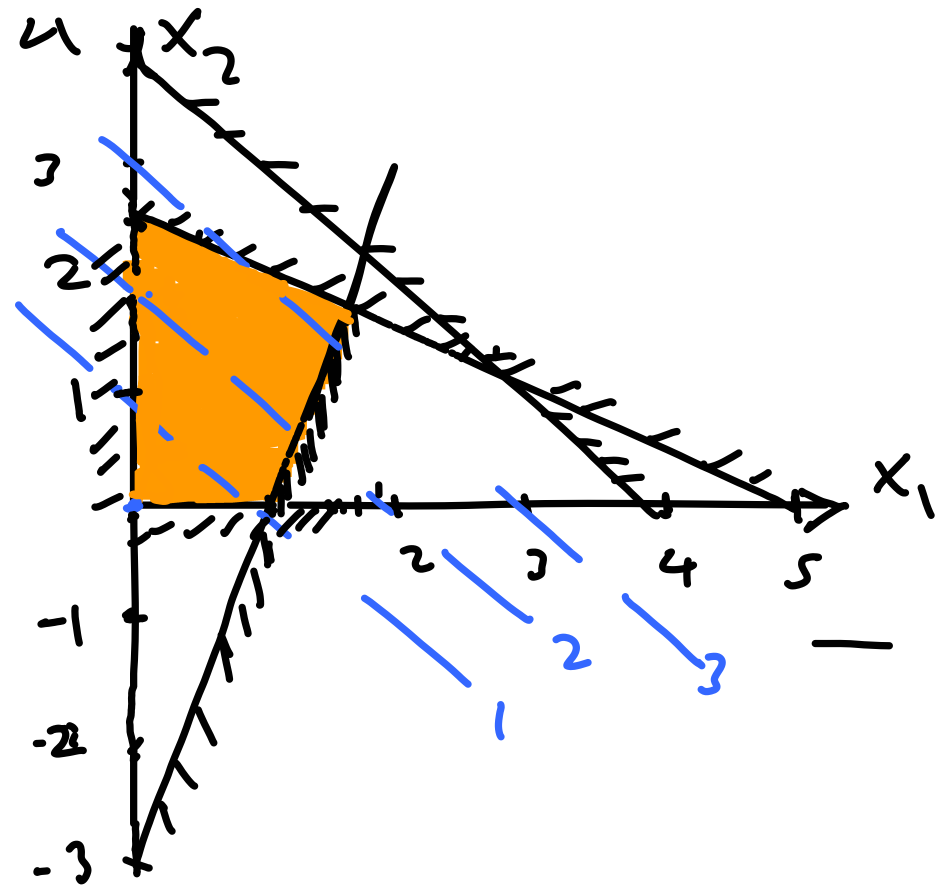


This is the main focus of this course

Example

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & 3x_1 - x_2 \leq 3 \\ & x_1 + 2x_2 \leq 5 \\ & x_1 + x_2 \leq 4 \\ & x_1 \geq 0,\ x_2 \geq 0 \end{aligned}$$

$\underline{x}^* = [0 \quad 0]^T$



Classifications (Cont'd)

2) According to nature of functions

C) Quadratic Programming

A general quadratic problem is given by

$$\min_{\underline{x}} \ \tfrac{1}{2}\underline{x}^T \underline{G}\,\underline{x} + \underline{c}^T\underline{x} \qquad (G \text{ is symmetric})$$

$$\text{subject to} \quad \underline{a}_i^T \underline{x} = b_i, \quad i \in E$$

$$\underline{a}_i^T \underline{x} \leq \underline{b}_i, \quad i \in I$$

Example

$$\min \; q(\underline{x}) = 3x_1^2 + 2x_1x_2 + x_1x_3 + 2.5x_2^2 + 2x_2x_3 + 2x_3^2 - 8x_1 - 3x_2 - 3x_3$$

$$\text{s.t.} \quad x_1 + x_3 = 3, \quad x_2 + x_3 = 0$$

$$\min_{\underline{x}} \; \frac{1}{2}\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} -8 & -3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\text{s.t.} \quad x_1 + x_3 = 3, \quad x_2 + x_3 = 0$$

Classifications (Cont'd)

3) According to permissible values of the design parameters

a) Integer programming

All parameters represent integer values (number of workers, etc.)

b) Continuous programming

Mathematical Background

* $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a column vector, $\underline{x}^T = [x_1 \; x_2 \cdots \underline{x}_n]$ is a row vector. $\underline{x} \in \mathbb{R}^n$.

* Standard Inner product for real vectors

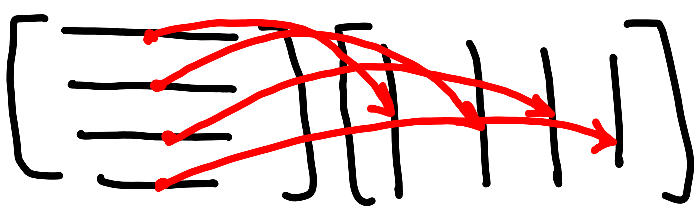
$$\underline{x}^T \underline{y} = [x_1 \; x_2 \cdots x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Mathematical Background

* $\underline{x}^T \underline{y} = 0$ if $\underline{x}$ & $\underline{y}$ are normal vectors
* A matrix $\underline{A}$ is positive semidefinite if $\underline{x}^T \underline{A} \underline{x} \geq 0 \Rightarrow \underline{x}^T (\underline{A}\underline{x}) \geq 0 \Rightarrow$ vector is rotated by less than 90°
* A square matrix $\underline{A} \in R^{n \times n}$ is nonsingular, if for every vector $\underline{b}$ there exists a vector $\underline{x}$ such that $\underline{A}\underline{x} = \underline{b} \Rightarrow (\underline{A}^{-1}\underline{A})\underline{x} = \underline{A}^{-1}\underline{b} \Rightarrow \underline{x} = \underline{A}^{-1}\underline{b}$

Mathematical Background (Cont'd)

* A square matrix $\underset{\sim}{Q}$ is orthogonal if $\underset{\sim}{Q}\,\underset{\sim}{Q}^T = \underset{\sim}{Q}^T\underset{\sim}{Q} = \underset{\sim}{I}$

 $= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow$ Columns are normal & normalized

Example:

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow Q\underset{\sim}{Q}^T = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Mathematical Background (Cont'd)

* A subset $\mathcal{S} \subset \mathbb{R}^n$ is a subspace if for every $\underline{x}, \underline{y} \in \mathcal{S}$, then $\alpha \underline{x} + \beta \underline{y} \in \mathcal{S}$ $\forall \alpha, \beta \in \mathbb{R}$

examples: $\mathbb{R}^n$ is a subspace, Any line through the origin in $\mathbb{R}^2$ is a subspace, the origin is a subspace

* Given any set of vectors $\underline{a}_i \in \mathbb{R}^n$, $i = 1, 2, \dots, m$, the set

$$\mathcal{S} = \{ \underline{w} \in \mathbb{R}^n \mid \underline{a}_i^T \underline{w} = 0,\ i = 1, 2, \dots, m \}$$

is a subspace. Prove it!

Mathematical Background (Cont'd)

* All vectors satisfying $\{\underset{\sim}{w} \mid \underset{\sim}{A}\underset{\sim}{w} = 0\}$ is a subspace called the null space of $\underset{\sim}{A}$ ($\mathrm{Null}(\underset{\sim}{A})$).
* The range space of a matrix $\underset{\sim}{A}$ is defined by $\mathrm{Range}(\underset{\sim}{A}) = \{\underset{\sim}{w} \mid \underset{\sim}{w} = \underset{\sim}{A}\underset{\sim}{v}\} \Rightarrow \underset{\sim}{w}$ is a linear combination of the columns of $\underset{\sim}{A}$
* The Fundamental Theory of Linear Algebra

$$\mathrm{Null}(\underset{\sim}{A}) \oplus \mathrm{Range}(\underset{\sim}{A}^T) = \mathbb{R}^n \quad (\oplus \text{ is direct sum})$$

Example

$$\underline{A} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 1 & 3 \end{bmatrix}, \quad \underline{w} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

$$\underline{A}\,\underline{w} = \underline{0}, \quad \underline{w} \in \mathrm{Null}(\underline{A})$$

$$\mathrm{Range}(\underline{A}^T) = \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

Notice that $\mathrm{Null}(\underline{A}) \perp \mathrm{Range}(\underline{A}^T)$

Gradient and Hessian

* For a function $f(\underline{x})$, the gradient is given by

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} . \qquad \frac{\partial f}{\partial x_i} = \lim_{\varepsilon \to 0} \frac{f(\underline{x} + \varepsilon \underset{\sim}{e}_i) - f(\underline{x})}{\varepsilon}$$

$$\underset{\sim}{e}_i = [0\ 0 \ldots 0\ \underset{\text{ith component}}{1}\ 0\ 0\ 0\ 0]^T$$

practical calculation

$$\frac{\partial f}{\partial x_i} \simeq \frac{\Delta f}{\Delta x_i} = \frac{f(\underline{x} + \Delta \underset{\sim}{e}_i) - f(\underline{x})}{\Delta}$$

(What about backward & central equations?)

Gradient and Hessian (Cont'd)

* The Hessian of a function $f(\underset{\sim}{x})$ is a square matrix defined by

$$\underset{\sim}{H} = \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Taylor Expansion

$$f(\underset{\sim}{x} + \Delta \underset{\sim}{x}) = f(\underset{\sim}{x}) + (\nabla f)^T \Delta \underset{\sim}{x} + \frac{1}{2} \Delta \underset{\sim}{x}^T \underset{\sim}{H} \Delta \underset{\sim}{x} + O(\|\Delta \underset{\sim}{x}\|^3)$$

Try it!

Convergence Rates

* Optimization Algorithms generate a sequence of points $\{\underline{x}_k\}$ in $\underline{R}^n$ that converge to $\underline{x}^*$, the optimal solution.

* The convergence is linear if it satisfies

$$\|\underline{x}_{k+1} - \underline{x}^*\| \leq r \|\underline{x}_k - \underline{x}^*\| \quad \text{for all } k \text{ large enough}$$

example: $\{1 + (0.5)^k\}$

Slow Convergence!

Convergence Rates (Cont'd)

* Superlinear Convergence

$$\|\underline{x}_{k+1} - \underline{x}^*\| \leq r \|\underline{x}_k - \underline{x}^*\|^p \quad \text{for all } k \text{ large enough}$$

$$1 < p < 2$$

example $(1 + k^{-k})$ (A better rate of Convergence)

* Quadratic Convergence

$$\|\underline{x}_{k+1} - \underline{x}^*\| \leq r \|\underline{x}_k - \underline{x}^*\|^2 \quad \text{for all } k \text{ large enough}$$

example $(1 + (0.5)^{2^k})$ (The best we can get!)