Mathematical Notations and Preliminaries

Yinyu Ye

Department of Management Science and Engineering

Stanford University

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

Real n-Space; Euclidean Space

- $\mathcal{R}$: real numbers
- $\mathcal{R}^n$: n-dimensional Euclidean space
- $\mathbf{x} \geq \mathbf{y}$ means $x_j \geq y_j$ for $j = 1, 2, ..., n$
- $\mathbf{0}$: vector of all zeros; $\mathbf{e}$: vector of all ones
- Inner-product of two vectors:

$$\mathbf{x} \bullet \mathbf{y} := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^{n} x_j y_j$$

- Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$,

  Infinity-norm: $\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\}$,

  p-norm: $\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$

- Column vector:

$$\mathbf{x} = (x_1; x_2; \ldots; x_n)$$

and row vector:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- Transpose operation: A^T
- A set of vectors $\mathbf{a}_1, ..., \mathbf{a}_m$ is said to be linearly dependent if there are scalars $\lambda_1, ..., \lambda_m$, not all zero, such that the linear combination

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}$$

- A linearly independent set of vectors that span R^n is a basis.

Hyper Planes

The most important type of convex set is a hyperplane. Hyperplanes dominate the entire theory of optimization. Let a be a nonzero n-dimensional vector, and let b be a real number. The set

$$H = \{x \in \mathcal{R}^n : a^T x = b\}$$

is a hyperplane in $\mathcal{R}^n$. Relating to hyperplane, positive and negative closed half spaces are given by

$$H_+ = \{x : a^T x \geq b\}$$

$$H_- = \{x : a^T x \leq b\}.$$

System of Linear Equations

Solve for $\mathbf{x} \in \mathcal{R}^n$ from:

$$\left.\begin{array}{rcl} \mathbf{a}_1\mathbf{x} & = & b_1 \\ \mathbf{a}_2\mathbf{x} & = & b_2 \\ \ldots & . & . \\ \mathbf{a}_m\mathbf{x} & = & b_m \end{array}\right. \quad \Rightarrow A\mathbf{x} = \mathbf{b}$$

Basic solution: select m columns from A to form a square matrix A_B such that

$$A_B\mathbf{x}_B = \mathbf{b}, \quad \text{the rest of} \quad \mathbf{x}_N = \mathbf{0}$$

where B is the index set of selected m coolumns.

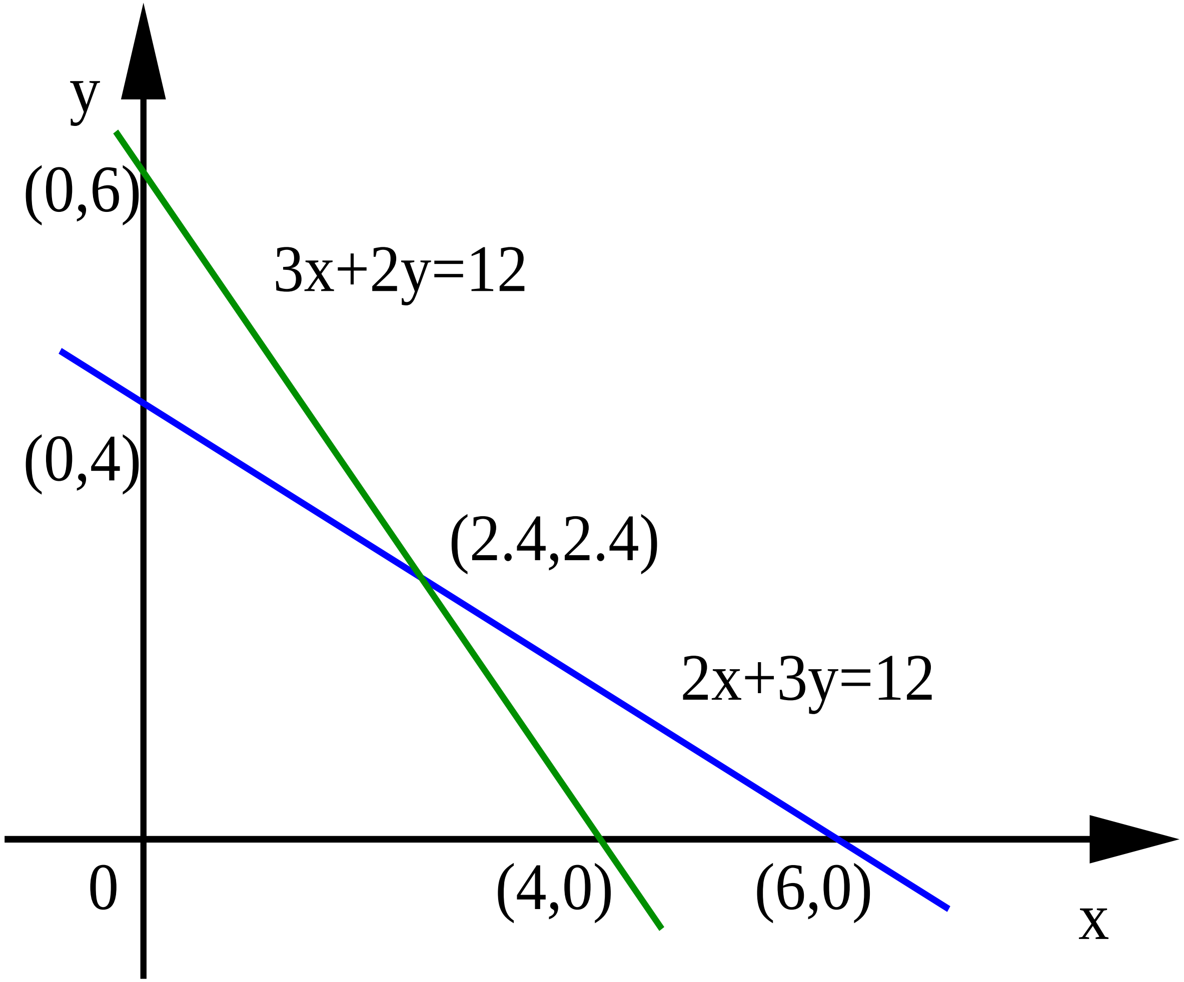

Figure 1: System of Linear Equations

Fundamental theorem of linear equations

Theorem 1 *Let $A \in \mathcal{R}^{m\times n}$ and $\mathbf{b} \in \mathcal{R}^m$. The system $\{\mathbf{x} : \; A\mathbf{x} = \mathbf{b}\}$ has a solution if and only if that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{b}^T\mathbf{y} \neq 0$ has no solution.*

A vector $\mathbf{y}$, with $A^T\mathbf{y} = 0$ and $\mathbf{b}^T\mathbf{y} \neq 0$, is called an infeasibility certificate for the system.

Example Let $A = (1; -1)$ and $\mathbf{b} = (1; 1)$. Then, $\mathbf{y} = (1/2; 1/2)$ is an infeasibility certificate.

Gaussian elimination method

$$\begin{pmatrix} a_{11} & A_{1.} \\ 0 & A' \end{pmatrix} \begin{pmatrix} x_1 \\ x' \end{pmatrix} = \begin{pmatrix} b_1 \\ b' \end{pmatrix}.$$

$$A = L \begin{pmatrix} U & C \\ 0 & 0 \end{pmatrix}$$

Matrices

- Matrix: $\mathcal{R}^{m\times n}$, ith row: $a_{i.}$, jth column: $a_{.j}$, ijth element: a_{ij}
- A_I denotes the submatrix of A whose rows belong to index set I, A_J denotes the submatrix whose columns belong to index set J, A_{IJ} denotes the submatrix whose rows belong to index set I and columns belong to index set J.
- All-zero matrix: $\mathbf{0}$, and identity matrix: I
- Diagonal matrix: $X = \text{diag}(\mathbf{x})$
- Symmetric matrix: $Q = Q^T$
- Positive Definite: $Q \succ 0$ iff $\mathbf{x}^T Q\mathbf{x} > 0$, for all $\mathbf{x} \neq \mathbf{0}$
- Positive Semidefinite: $Q \succeq 0$ iff $\mathbf{x}^T Q\mathbf{x} \geq 0$, for all $\mathbf{x}$

Line and Convex Combination

When $\mathbf{x}$ and $\mathbf{y}$ are two distinct points in R^n and α runs over R ,

$$\{\mathbf{z} : \mathbf{z} = \alpha\mathbf{x} + (1 - \alpha)\mathbf{y}\}$$

is the line determined by $\mathbf{x}$ and $\mathbf{y}$.

When $0 \leq \alpha \leq 1$, it is called the convex combination of $\mathbf{x}$ and $\mathbf{y}$ and it is the line segment between $\mathbf{x}$ and $\mathbf{y}$.

Convex Set

- Ω is said to be a convex set if for every $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$ and every real number $\alpha \in [0, 1]$, the point $\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in \Omega$.
- The convex hull of a set Ω is the intersection of all convex sets containing Ω
- Intersection of convex sets is convex

Proof of convex set

- All solutions to the system of linear equations, $\{\mathbf{x} : \; A\mathbf{x} = \mathbf{b}\}$, form a convex set.
- All solutions to the system of linear inequalities,

$$\{\mathbf{x} : \; A\mathbf{x} \leq \mathbf{b}\}$$

  , form a convex set.
- All solutions to the system of linear equations and inequalities, $\{\mathbf{x} : \; A\mathbf{x} = \mathbf{b}, \; \mathbf{x} \geq \mathbf{0}\}$, form a convex set.

Homework 3: Prove $\{\mathbf{b} : \;$ there exists $\mathbf{x} \geq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{b}\}$ is a convex set.

Convex Cones

- A set C is a cone if $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C$ for all $\alpha > 0$
- A convex cone is cone plus convex-set.
- Dual cone:

$$C^* := \{\mathbf{y} : \ \mathbf{y} \bullet \mathbf{x} \geq 0 \quad \text{for all} \quad \mathbf{x} \in C\}$$

$-C^*$ is also called the polar of C.

Cone Examples

- Example 2.1: The n-dimensional nonnegative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \ \mathbf{x} \geq 0\}$, is a convex cone; and it's self dual.
- Example 2.2: The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : \ t \geq \|\mathbf{x}\|_p\}$ is a convex cone in $\mathcal{R}^{n+1}$, called the p-order cone.
- Example 2.3: The set of all positive semi-definite matrices in $\mathcal{M}^n$, $\mathcal{M}^n_+$, is a convex cone, called the positive semi-definite cone; and it's self-dual

Homework 4: Find the dual of the p-order cone.

Polyhedral Convex Cones

- A cone C is (convex) polyhedral if C can be represented by

$$C = \{\mathbf{x} : A\mathbf{x} \leq 0\}$$

for some matrix A.

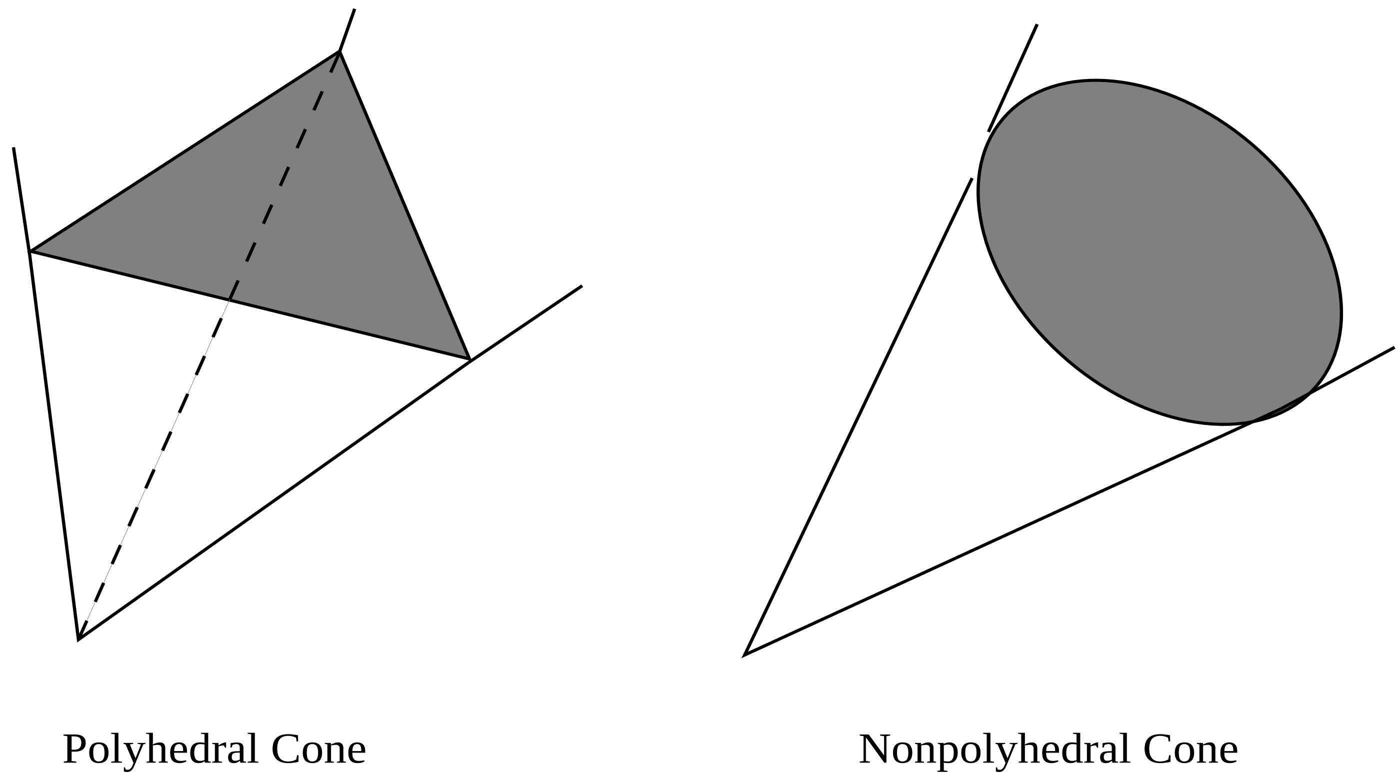

Figure 2: Polyhedral and non-polyhedral cones.

- The nonnegative orthant is a polyhedral cone but the second-order cone is not polyhedral.

Real Functions

- Continuous functions C

- Weierstrass theorem: a continuous function $f(\mathbf{x})$ defined on a compact set (bounded and closed) $\Omega \subset \mathcal{R}^n$ has a minimizer in Ω.

- The least upper bound or supremum of f over Ω

$$\sup\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

  and the greatest lower bound or infimum of f over Ω

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

- The gradient vector C^1:

$$\nabla f(\mathbf{x}) = \{\partial f / \partial x_i\}, \quad \text{for} \quad i = 1, ..., n.$$

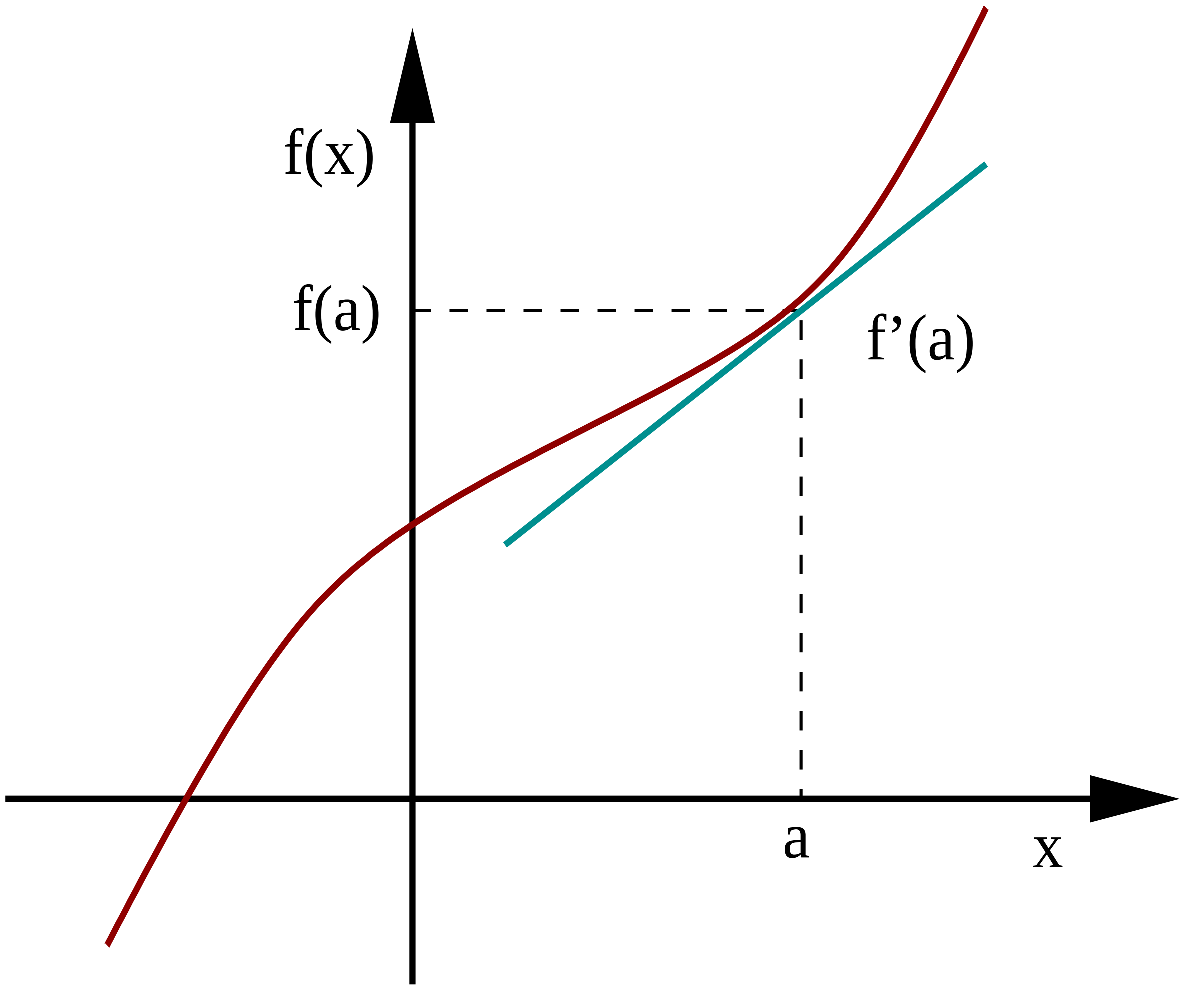

Figure 3: Derivative and slope

- The Hessian matrix C^2:

$$\nabla^2 f(\mathbf{x}) = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} \quad \text{for} \quad i = 1, ..., n; \; j = 1, ..., n.$$

- Vector function: $\mathbf{f} = (f_1; f_2; ...; f_m)$
- The Jacobian matrix of $\mathbf{f}$ is

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ ... \\ \nabla f_m(\mathbf{x}) \end{pmatrix}.$$

Convex Functions

- f convex function iff for $0 \leq \alpha \leq 1$,

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}).$$

- The level set of convex function f

$$L(z) = \{\mathbf{x} : \quad f(\mathbf{x}) \leq z\}$$

is a convex set.

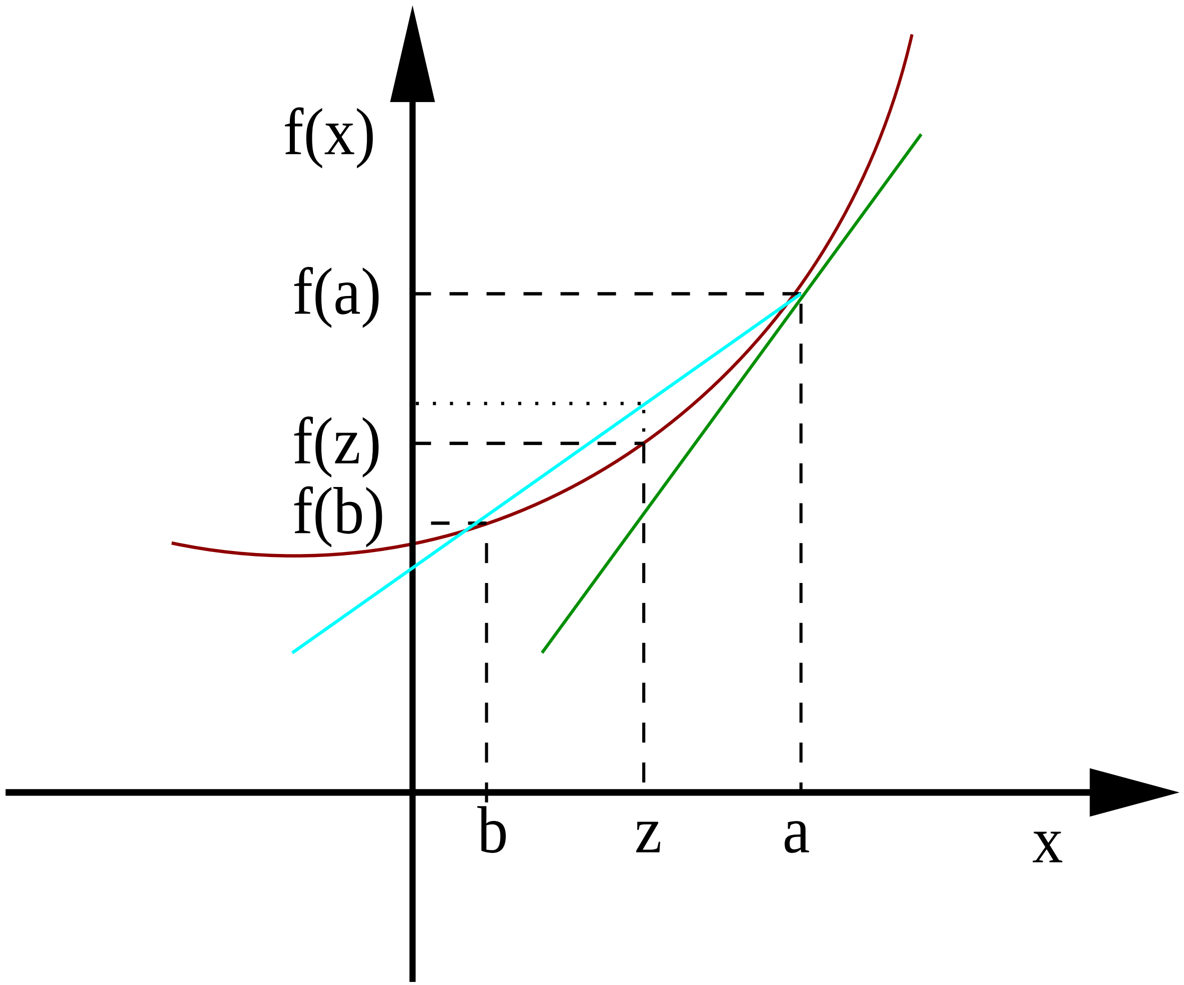

Figure 4: Properties of convex function

Proof of convex function

Consider the minimal-objective function of $\mathbf{b}$ for fixed A and $\mathbf{c}$:

$$\begin{array}{rl} z(\mathbf{b}) := \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Homework 5: Show that $z(\mathbf{b})$ is a convex function in $\mathbf{b}$ for all feasible $\mathbf{b}$.

Theorems on functions

Taylor's theorem or the mean-value theorem:

Theorem 2 *Let* $f \in C^1$ *be in a region containing the line segment* $[\mathbf{x}, \mathbf{y}]$*. Then there is a* α*,* $0 \le \alpha \le 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}).$$

Furthermore, if $f \in C^2$ *then there is a* α*,* $0 \le \alpha \le 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + (1/2)(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y}-\mathbf{x}).$$

Theorem 3 *Let* $f \in C^1$*. Then* f *is convex over a convex set* Ω *if and only if*

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x},\ \mathbf{y} \in \Omega$*.*

Theorem 4 *Let* $f \in C^2$*. Then* f *is convex over a convex set* Ω *if and only if the Hessian matrix of* f *is positive semi-definite throughout* Ω*.*

Known Inequalities

- Cauchy-Schwarz: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\mathbf{x}^T\mathbf{y} \le \|\mathbf{x}\|\|\mathbf{y}\|$.
- Arithmetic-geometric mean: given $\mathbf{x} > \mathbf{0}$,

$$\frac{\sum x_j}{n} \ge \left(\prod x_j\right)^{1/n}.$$

- Harmonic: given $\mathbf{x} > \mathbf{0}$,

$$\left(\sum x_j\right)\left(\sum 1/x_j\right) \ge n^2.$$

Linear least-squares problem

Given $A \in \mathcal{R}^{m \times n}$ and $\mathbf{c} \in \mathcal{R}^n$,

$$
\begin{array}{lll}
(LS) & \text{minimize} & \|A^T\mathbf{y} - \mathbf{c}\|^2 \\
 & \text{subject to} & \mathbf{y} \in \mathcal{R}^m.
\end{array}
$$

$$AA^T\mathbf{y} = A\mathbf{c} \quad \text{or} \quad \mathbf{y} = (AA^T)^{-1}A\mathbf{c}$$

with the projection:

$$A^T\mathbf{y} = A^T(AA^T)^{-1}A\mathbf{c}$$

Projection matrix: $P = A^T(AA^T)^{-1}A$ or $P = I - A^T(AA^T)^{-1}A$

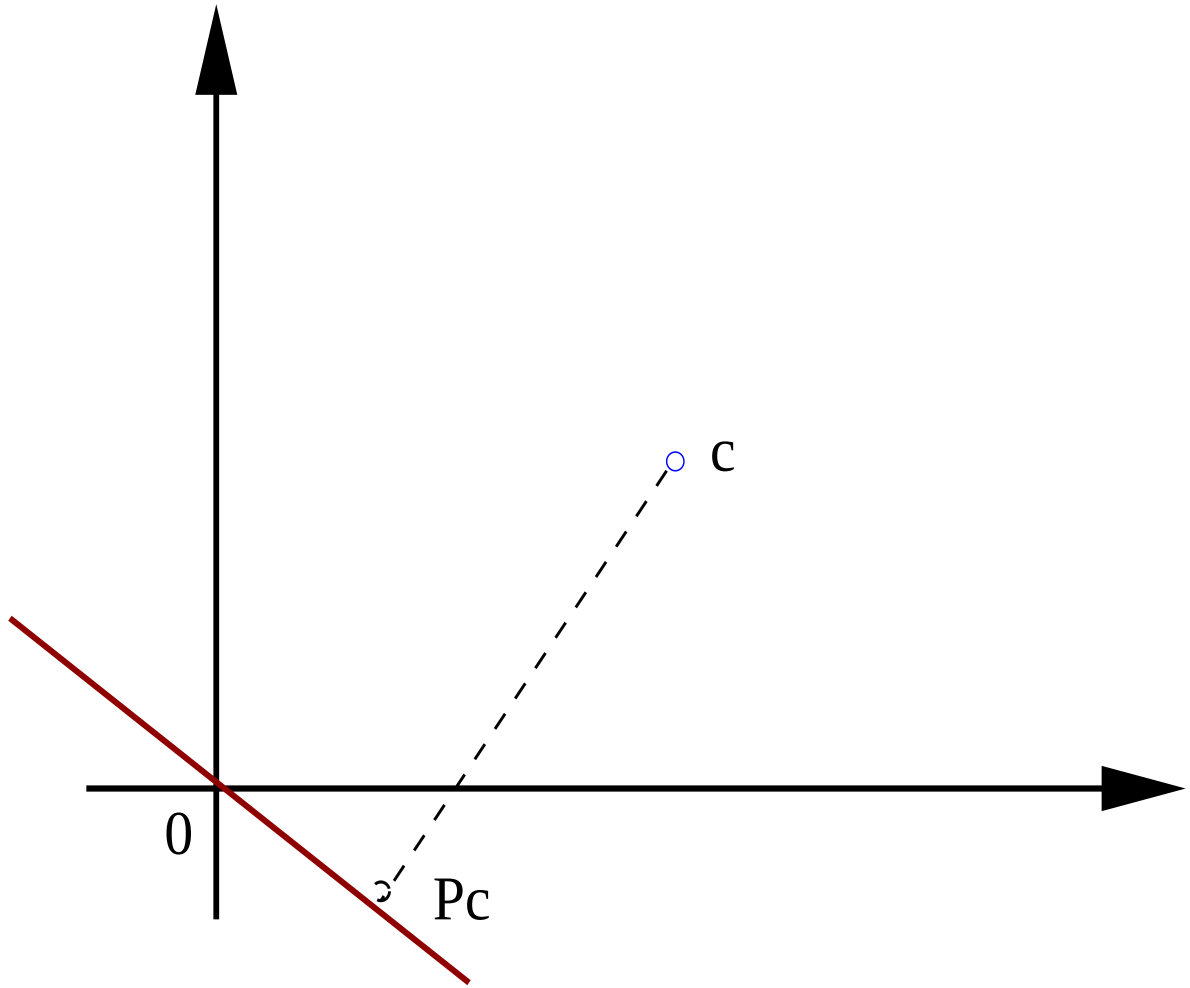

Figure 5: Projection of $\mathbf{c}$ onto a subspace

Newton's method and system of equations

Given $\mathbf{f}(\mathbf{x}) : \mathcal{R}^n \rightarrow \mathcal{R}^n$, the problem is to solve n equations for n unknowns:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}.$$

Given a point $\mathbf{x}^k$, Newton's Method sets

$$f(\mathbf{x}) \simeq f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) = \mathbf{0}.$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\nabla f(\mathbf{x}^k))^{-1} f(\mathbf{x}^k)$$

or solve for direction vector $\mathbf{d}_x$:

$$\nabla f(\mathbf{x}^k)\mathbf{d}_x = -f(\mathbf{x}^k) \quad \text{and} \quad \mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x.$$

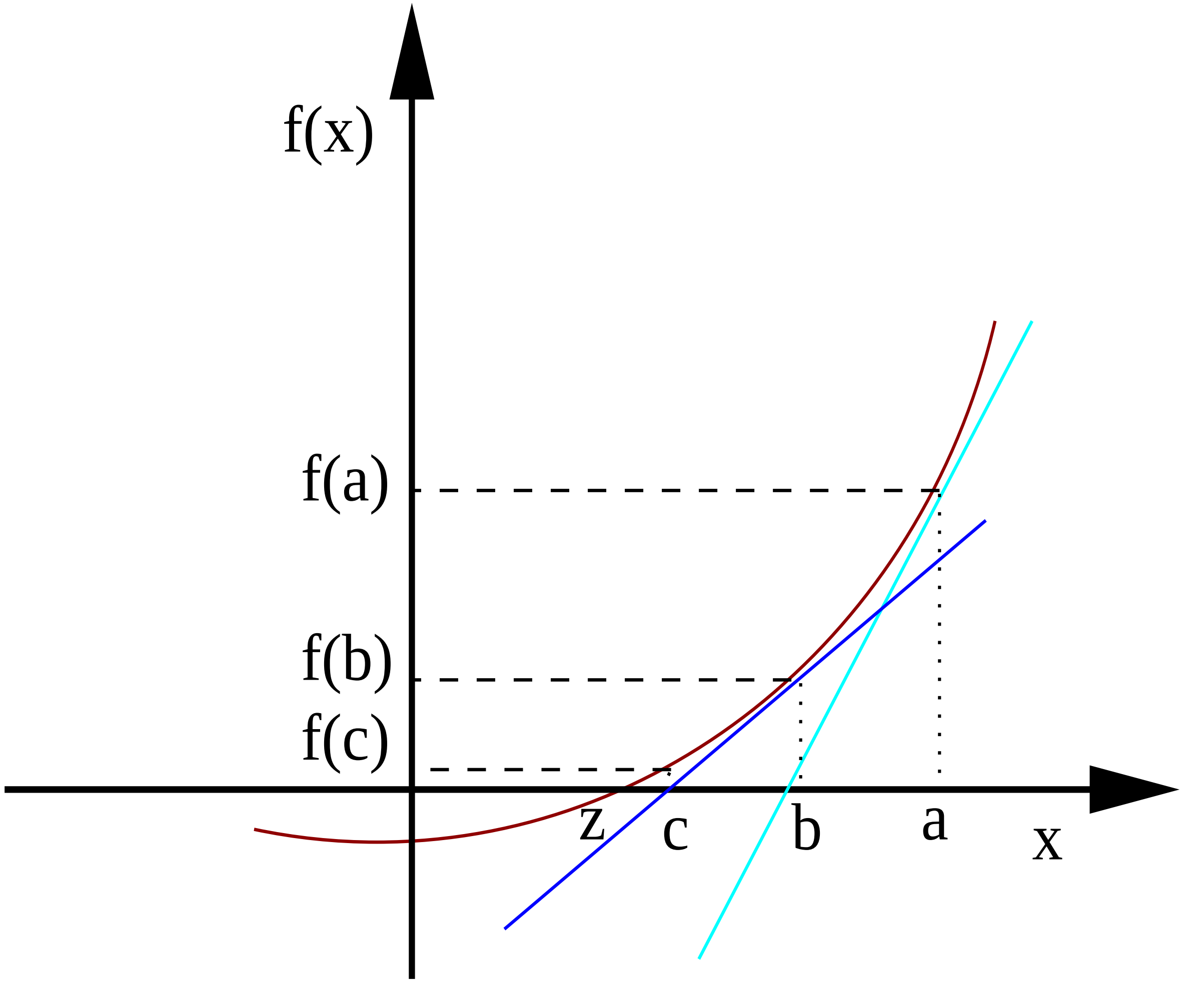

Figure 6: Newton's method for root finding

The quasi Newton method

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha(\nabla f(\mathbf{x}^k))^{-1} f(\mathbf{x}^k)$$

where scalar $\alpha \geq 0$ is called step-size. More generally

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha M^k f(\mathbf{x}^k)$$

where M^k is an $n \times n$ symmetric matrix. In particular, if $M^k = I$, the method is called the gradient method, where f is viewed as the gradient vector of a real function.