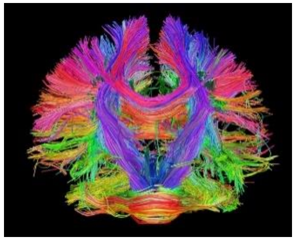



CS-570
Statistical Signal Processing

Lecture 3: Review of Convex Optimization

Spring Semester 2019

Grigorios Tsagkatakis

Today's Objectives

- Review of convex optimization

Disclaimer: Material used:

- Convex Optimization – S. Boyd and L. Vandenberghe http://web.stanford.edu/~boyd/cvxbook/
- Matrix Calculus - Po-Chen Wu

Matrix calculus

- The derivative of $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, by x is written as:

$$\frac{\partial \mathbf{y}}{\partial \mathrm{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

- The derivative of y by $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is written as:

$$\frac{\partial \mathrm{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

Matrix calculus

- The derivative of $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ with respect to $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & & \frac{\partial y_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

  - Also known as the Jacobian matrix

Example

- Given $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, and $y_1 = x_1^2 - 2x_2$, $y_2 = x_3^2 - 4x_2$,

  the Jacobian matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & -2 & 0 \\ 0 & -4 & 2x_3 \end{bmatrix}$$

- The derivative of a matrix function $\mathbf{Y}$ by a scalar x is known as the tangent matrix and is given by

$$\frac{\partial \mathbf{Y}}{\partial x} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial Y_{11}}{\partial x} & \frac{\partial Y_{12}}{\partial x} & \cdots & \frac{\partial Y_{1n}}{\partial x} \\ \frac{\partial Y_{21}}{\partial x} & \frac{\partial Y_{22}}{\partial x} & & \frac{\partial Y_{2n}}{\partial x} \\ \vdots & & \ddots & \vdots \\ \frac{\partial Y_{m1}}{\partial x} & \frac{\partial Y_{m2}}{\partial x} & \cdots & \frac{\partial Y_{mn}}{\partial x} \end{bmatrix}$$

- The derivative of a scalar y function by a matrix $\mathbf{X}$ is known as the gradient matrix and is given by

$$\frac{\partial y}{\partial \mathbf{X}} \overset{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial X_{11}} & \frac{\partial y}{\partial X_{21}} & \cdots & \frac{\partial y}{\partial X_{m1}} \\ \frac{\partial y}{\partial X_{12}} & \frac{\partial y}{\partial X_{22}} & & \frac{\partial y}{\partial X_{m2}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y}{\partial X_{1n}} & \frac{\partial y}{\partial X_{2n}} & \cdots & \frac{\partial y}{\partial X_{mn}} \end{bmatrix}$$

List of Differentiation

	Scalar y		Vector $\mathbf{y}$ (size m)		Matrix $\mathbf{Y}$ (size $m \times n$)	
	Notation	Type	Notation	Type	Notation	Type
Scalar x	$\frac{\partial y}{\partial x}$	scalar	$\frac{\partial \mathbf{y}}{\partial x}$	size-m column vector	$\frac{\partial \mathbf{Y}}{\partial x}$	$m \times n$ matrix
Vector $\mathbf{x}$ (size n)	$\frac{\partial y}{\partial \mathbf{x}}$	size-n row vector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	$m \times n$ matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	–
Matrix $\mathbf{X}$ (size $p \times q$)	$\frac{\partial y}{\partial \mathbf{X}}$	$q \times p$ matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$	–	$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	–

Derivative Formulas

$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{A}\mathbf{x}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T$

Hint: Derive $\mathbf{x}$

- If you have to differentiate $\mathbf{x}^T$, transpose the rest.
- If you have two $\mathbf{x}$-terms, differentiate them separately in turn and then sum up the two derivatives.

Chain rule

- Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{bmatrix}$, where $\mathbf{z}$ is a function of $\mathbf{y}$, which is in turn a function of $\mathbf{x}$. Then

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & & \frac{\partial z_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial z_r}{\partial x_1} & \frac{\partial z_r}{\partial x_2} & \cdots & \frac{\partial z_r}{\partial x_n} \end{bmatrix},$$

$$\text{where } \frac{\partial z_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial x_j} \quad \begin{cases} i = 1,2,\cdots,r \\ j = 1,2,\cdots,n \end{cases}$$

Chain rule

- $$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_2} & \cdots & \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_n} \\ \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_2} & & \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_n} \\ & \vdots & \ddots & \vdots \\ \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_2} & \cdots & \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_m} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & & \frac{\partial z_2}{\partial y_m} \\ \vdots & & \ddots & \vdots \\ \frac{\partial z_r}{\partial y_1} & \frac{\partial z_r}{\partial y_2} & \cdots & \frac{\partial z_r}{\partial y_m} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & & \frac{\partial y_2}{\partial x_n} \\ & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} = \frac{\partial \mathbf{z}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

The Matrix Differential

- For a scalar function $f(\mathbf{x})$, where $\mathbf{x}$ is an n-vector, the ordinary differential of multivariate calculus is defined as

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

- In harmony with this formula, we define the differential of an $m \times n$ matrix $\mathbf{X} = [X_{ij}]$ to be

$$d\mathbf{X} \stackrel{\text{def}}{=} \begin{bmatrix} dX_{11} & dX_{12} & \cdots & dX_{1n} \\ dX_{21} & dX_{22} & & dX_{2n} \\ \vdots & & \ddots & \vdots \\ dX_{m1} & dX_{m2} & \cdots & dX_{mn} \end{bmatrix}$$

The Matrix Differential

- This definition complies with the multiplicative and associative rules

$$d(\alpha\mathbf{X}) = \alpha d\mathbf{X} \qquad\qquad d(\mathbf{X}+\mathbf{Y}) = d\mathbf{X} + d\mathbf{Y}$$

- If $\mathbf{X}$ and $\mathbf{Y}$ are product-conforming matrices, it can be verified that the differential of their product is

$$d(\mathbf{XY}) = (d\mathbf{X})\mathbf{Y} + \mathbf{X}(d\mathbf{Y})$$

Hessian matrix

- Let $f: \mathbb{R}^n \to \mathbb{R}$ and assume all second partial derivative exist in the domain
- The Hessian matrix $\mathbf{H} \in \mathbb{R}^{n \times n}$ is defined as

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Mathematical optimization

(mathematical) optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: constraint functions

solution or **optimal point** $x^\star$ has smallest value of f_0 among all vectors that satisfy the constraints

Solving optimization problems

general optimization problem

- very difficult to solve
- methods involve some compromise, *e.g.*, very long computation time, or not always finding the solution (which may not matter in practice)

exceptions: certain problem classes can be solved efficiently and reliably

- least-squares problems
- linear programming problems
- convex optimization problems

Least squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

solving least-squares problems

- analytical solution: $x^\star = (A^TA)^{-1}A^Tb$
- reliable and efficient algorithms and software
- computation time proportional to n^2k ($A \in \mathbf{R}^{k\times n}$); less if structured
- a mature technology

using least-squares

- least-squares problems are easy to recognize
- a few standard techniques increase flexibility ($e.g.$, including weights, adding regularization terms)

Linear programming

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m \end{array}$$

solving linear programs

- no analytical formula for solution
- reliable and efficient algorithms and software
- computation time proportional to $n^2 m$ if $m \geq n$; less with structure
- a mature technology

using linear programming

- not as easy to recognize as least-squares problems
- a few standard tricks used to convert problems into linear programs ($e.g.$, problems involving ℓ_1- or ℓ_∞-norms, piecewise-linear functions)

Convex optimization problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- objective and constraint functions are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$

- includes least-squares problems and linear programs as special cases

Nonlinear optimization

traditional techniques for general nonconvex problems involve compromises

local optimization methods (nonlinear programming)

- find a point that minimizes f_0 among feasible points near it
- fast, can handle large problems
- require initial guess
- provide no information about distance to (global) optimum

global optimization methods

- find the (global) solution
- worst-case complexity grows exponentially with problem size

these algorithms are often based on solving convex subproblems

History of convex optimization

theory (convex analysis): 1900–1970

algorithms

- 1947: simplex algorithm for linear programming (Dantzig)
- 1970s: ellipsoid method and other subgradient methods
- 1980s & 90s: polynomial-time interior-point methods for convex optimization (Karmarkar 1984, Nesterov & Nemirovski 1994)
- since 2000s: many methods for large-scale convex optimization

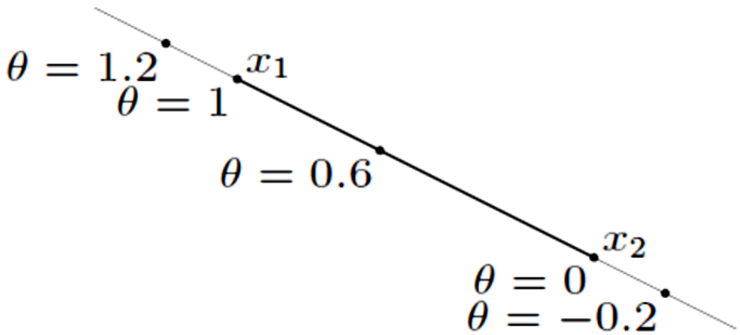
applications

- before 1990: mostly in operations research, a few in engineering
- since 1990: many applications in engineering (control, signal processing, communications, circuit design, . . .)
- since 2000s: machine learning and statistics

Affine set

line through x_1, x_2: all points

$$x = \theta x_1 + (1-\theta)x_2 \qquad (\theta \in \mathbf{R})$$



affine set: contains the line through any two distinct points in the set

example: solution set of linear equations $\{x \mid Ax = b\}$

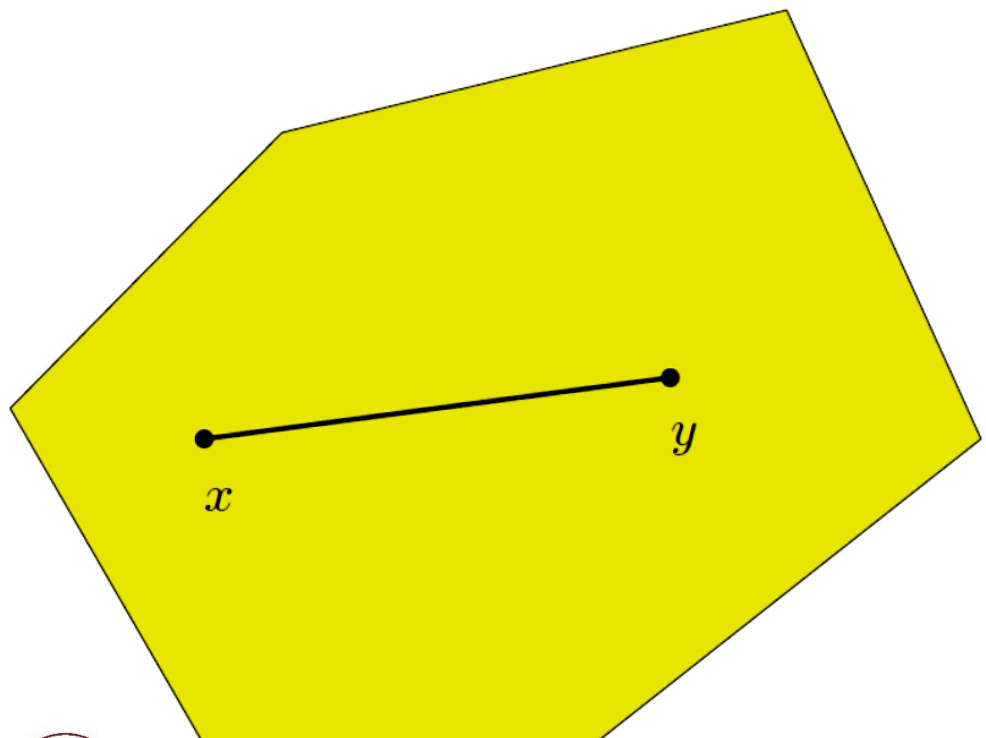
(conversely, every affine set can be expressed as solution set of system of linear equations)

Convex sets

Definition

A set $C \subseteq \mathbb{R}^n$ is *convex* if for $x, y \in C$ and any $\alpha \in [0,1]$,

$$\alpha x + (1-\alpha) y \in C.$$



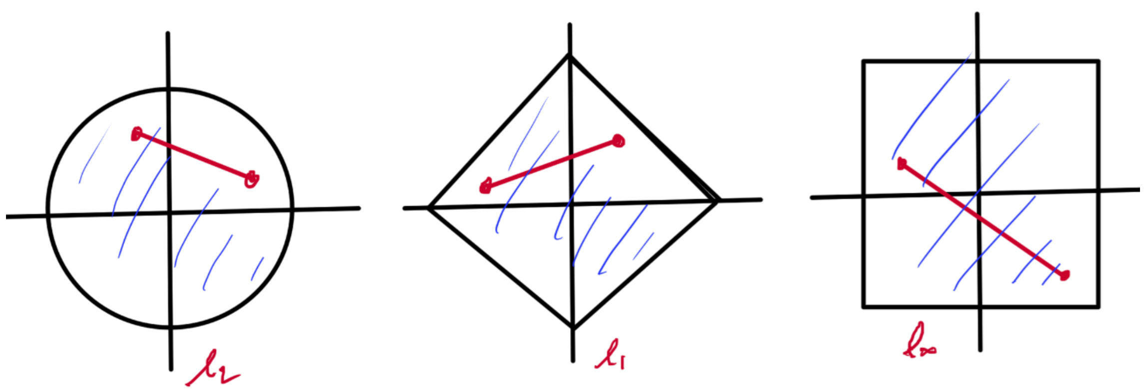
Examples

- All of $\mathbb{R}^n$ (obvious)

- Non-negative orthant, $\mathbb{R}^n_+$: let $x \succeq 0$, $y \succeq 0$, clearly $\alpha x + (1-\alpha) y \succeq 0$.

- Norm balls: let $\|x\| \leq 1$, $\|y\| \leq 1$, then

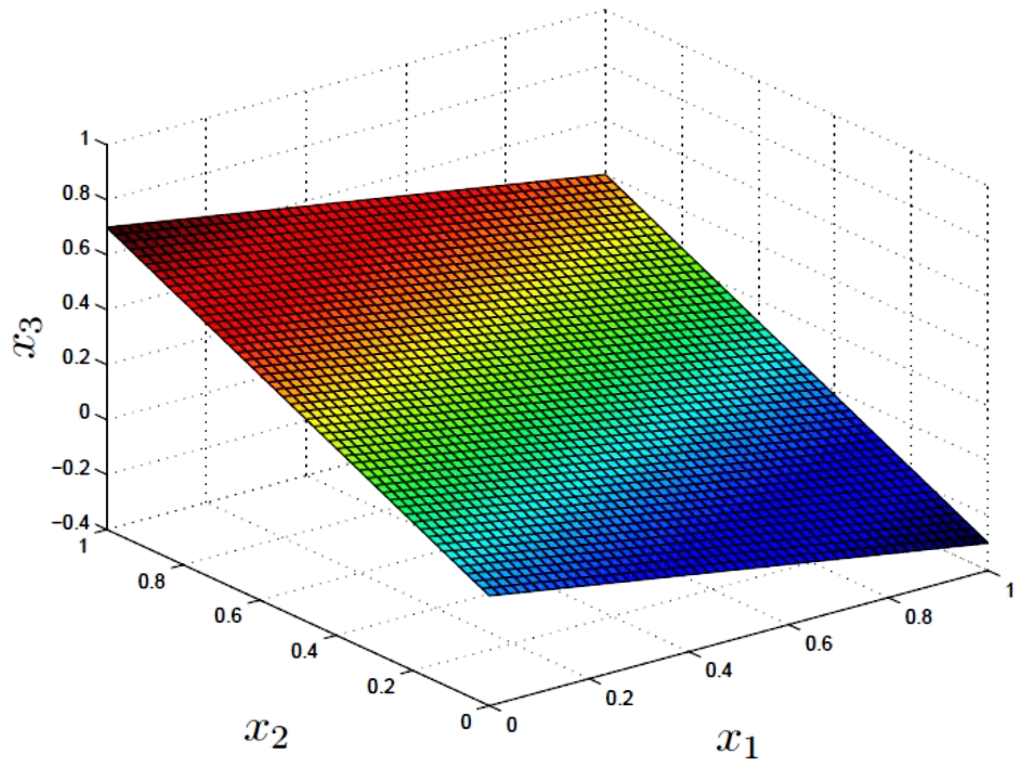
$$\|\alpha x + (1-\alpha) y\| \leq \|\alpha x\| + \|(1-\alpha) y\| = \alpha \|x\| + (1-\alpha) \|y\| \leq 1.$$



Examples

- Affine subspaces: $Ax = b$, $Ay = b$, then

$$A(\alpha x + (1-\alpha)y) = \alpha Ax + (1-\alpha)Ay = \alpha b + (1-\alpha)b = b.$$

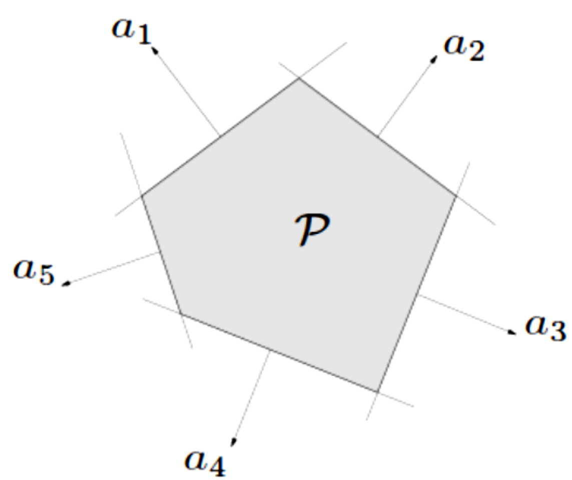


Polyhedra

solution set of finitely many linear inequalities and equalities

$$Ax \preceq b, \qquad Cx = d$$

($A \in \mathbf{R}^{m \times n}$, $C \in \mathbf{R}^{p \times n}$, $\preceq$ is componentwise inequality)



polyhedron is intersection of finite number of halfspaces and hyperplanes

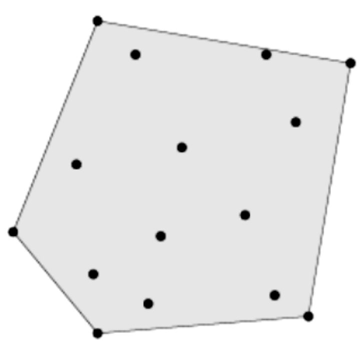
Convex combination and convex hull

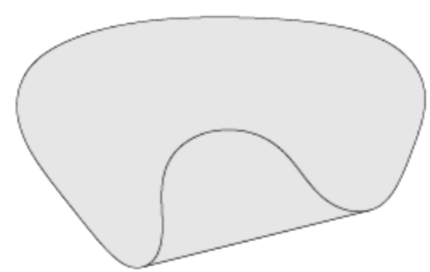
convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \geq 0$

convex hull conv S: set of all convex combinations of points in S



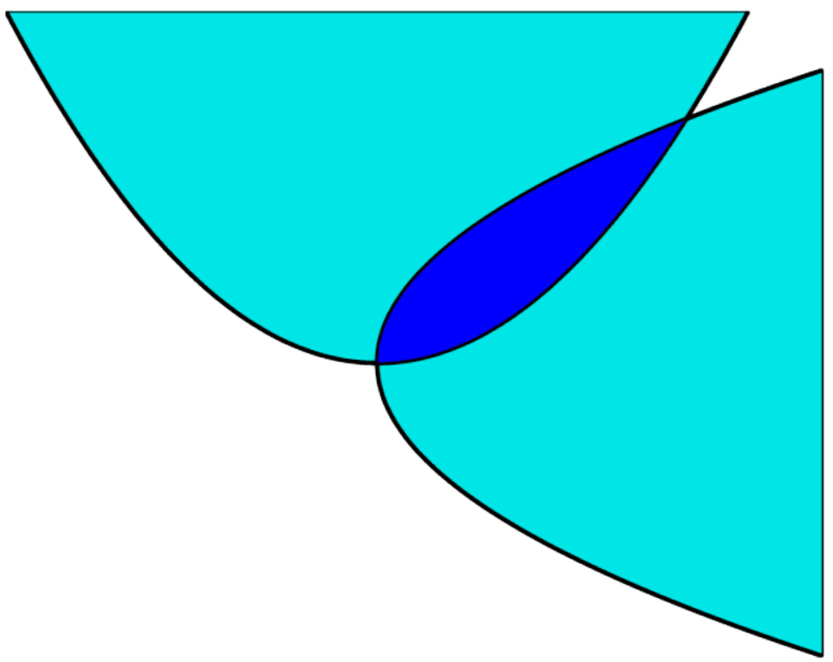


Examples

- Arbitrary intersections of convex sets: let C_i be convex for $i \in \mathcal{I}$, $C = \bigcap_i C_i$, then

$$x \in C, y \in C \quad \Rightarrow \quad \alpha x + (1-\alpha)y \in C_i \ \forall\, i \in \mathcal{I}$$

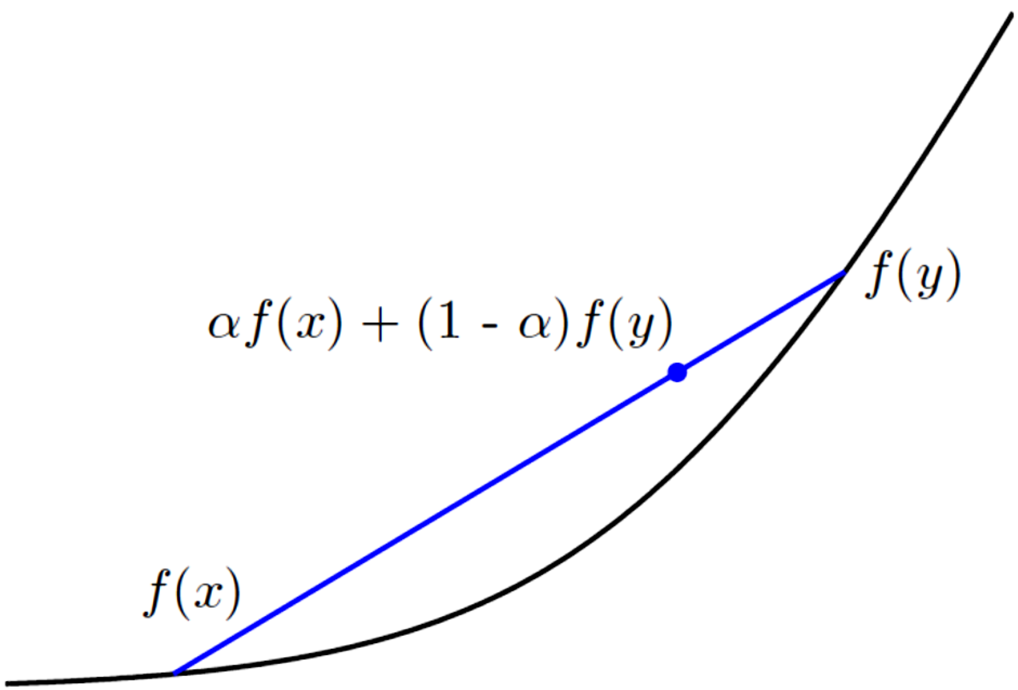
so $\alpha x + (1-\alpha)y \in C$.



Convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if for $x, y \in \operatorname{dom} f$ and any $\alpha \in [0, 1]$,

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



Examples on R^n and $R^{m\times n}$

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$

- affine function $f(x) = a^T x + b$
- norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

examples on $\mathbf{R}^{m\times n}$ ($m \times n$ matrices)

- affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b \;=<\boldsymbol{A},\boldsymbol{X}> +\boldsymbol{b}$$

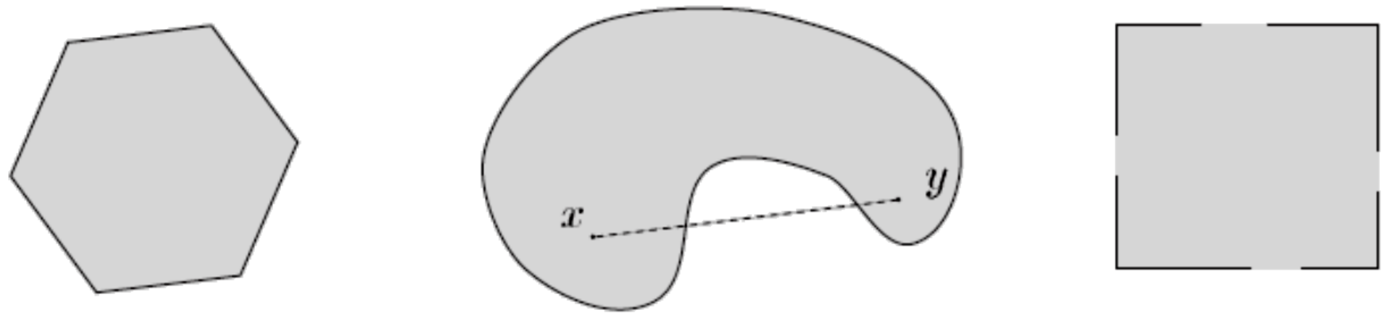
- spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

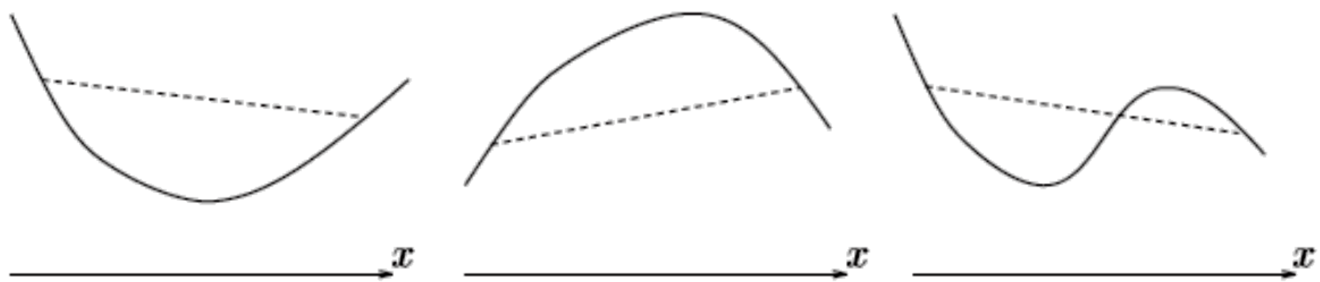
Examples

Sets:

convex — not convex



and functions:



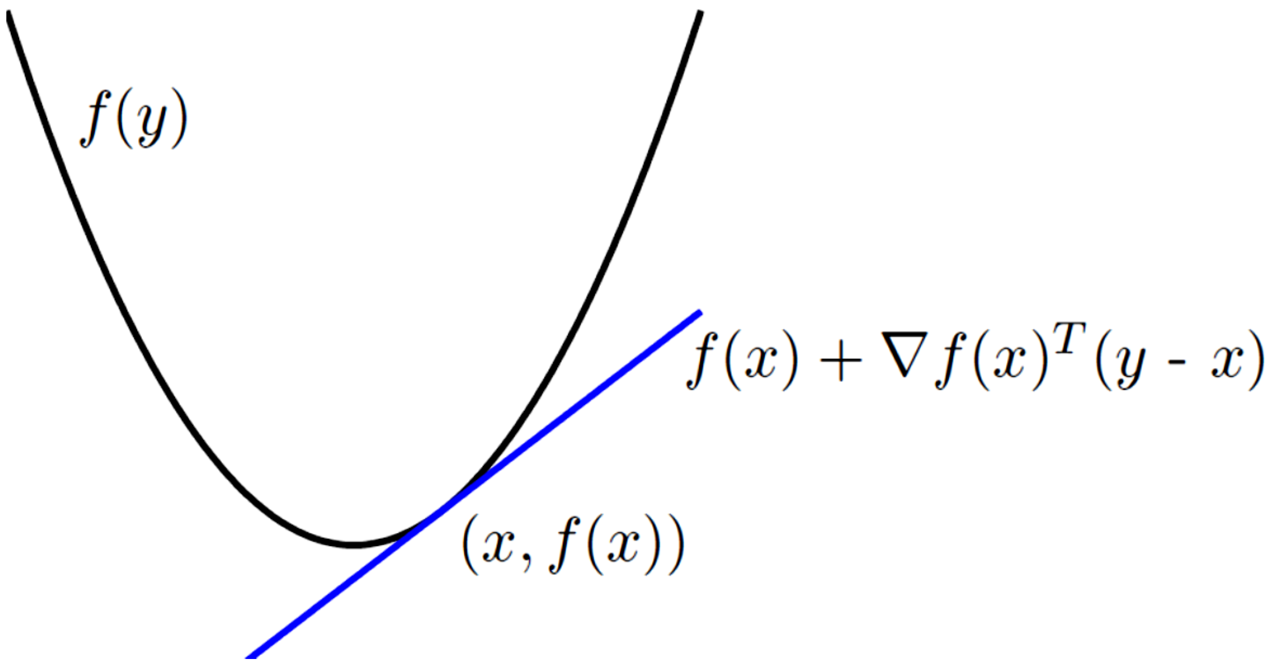
convex — concave — neither

First order convexity conditions

Theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ *is differentiable. Then* f *is convex if and only if for all* $x, y \in \operatorname{dom} f$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$



First order convexity conditions

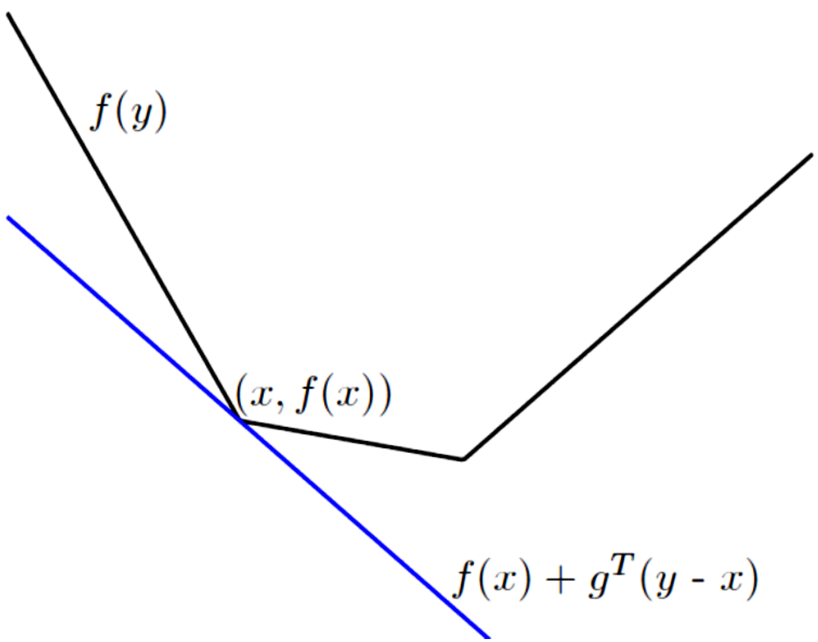
Definition

The *subgradient set*, or subdifferential set, $\partial f(x)$ of f at x is

$$\partial f(x) = \left\{ g : f(y) \geq f(x) + g^T(y - x) \ \text{ for all } y \right\}.$$

Theorem

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if it has non-empty subdifferential set everywhere.



Second order convexity conditions

If $f(x)$ is twice continuously differentiable, then

$$f \text{ is convex} \quad \Leftrightarrow \quad \nabla^2 f(x) \geq 0 \text{ for all } x \in \Re^n.$$

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Examples

Non-convex	Non-differentiable convex	Differentiable convex
$1 - \|x\|$	$\|x\|$	$x^T A x,\ A \geq 0$
$\cos(e^T x)$	$\|x\|_1$	$e^{-\|x\|^2}$
$x^T A x,\ A \not\geq 0$	$\max\{\|x\|^2, e^T x\}$	$-\log(t^2 - \|x\|^2)$
$\|x\|^3$	$\|x\|_1^2$	$-\log\det(X)$
$\vdots$	$\vdots$	$\vdots$

Properties of Convex functions

- **Convexity over all lines:**

$$f(x) \text{ is convex} \quad \Leftrightarrow \quad f(x_0 + th) \text{ is convex in } t \text{ for all } x_0 \text{ and } h$$

- **Positive multiple:**

$$f(x) \text{ is convex} \quad \Leftrightarrow \quad \alpha f(x) \text{ is convex, for all } \alpha \geq 0$$

- **Sum of convex functions**:

$$f_1(x),\ f_2(x) \text{ convex} \quad \Rightarrow \quad f_1(x) + f_2(x) \text{ is convex}$$

- **Pointwise maximum:**

$$f_1(x),\ f_2(x) \text{ convex} \quad \Rightarrow \quad \max\{f_1(x), f_2(x)\} \text{ is convex}$$

- **Affine transformation of domain:**

$$f(x) \text{ is convex} \quad \Rightarrow \quad f(Ax + b) \text{ is convex}$$

Some Commonly Used Convex Functions

- **Piecewise-linear functions:** $\max_i\{a_i^T x + b_i\}$ is convex in x
- **Quadratic functions:** $f(x) = x^T Q x + 2q^T x + c$ is convex iff $Q \geq 0$
- **Piecewise-quadratic functions:** $\max_i\{x^T Q_i x + q_i^T x + c_i\}$ is convex in x if $Q_i \geq 0$
- **Norm functions:** $\|x\|_k = \left(\sum_{i=1}^{n} |x_i|^k\right)^{1/k}$, where $k \in [1, \infty)$
- **Convex functions over matrices:** $\mathrm{Tr}(X)$, $\lambda_{\max}(X)$ are convex on $X = X^T$; and $-\log\det(X)$ is convex on the set $\{X \mid X = X^T,\ X \geq 0\}$
- **Logarithmic barrier functions:** $f(x) = \sum_{i=1}^{m} \log(b_i - a_i^T x)^{-1}$ is convex over $\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ 1 \leq i \leq m\}$

Convex Optimization Problems

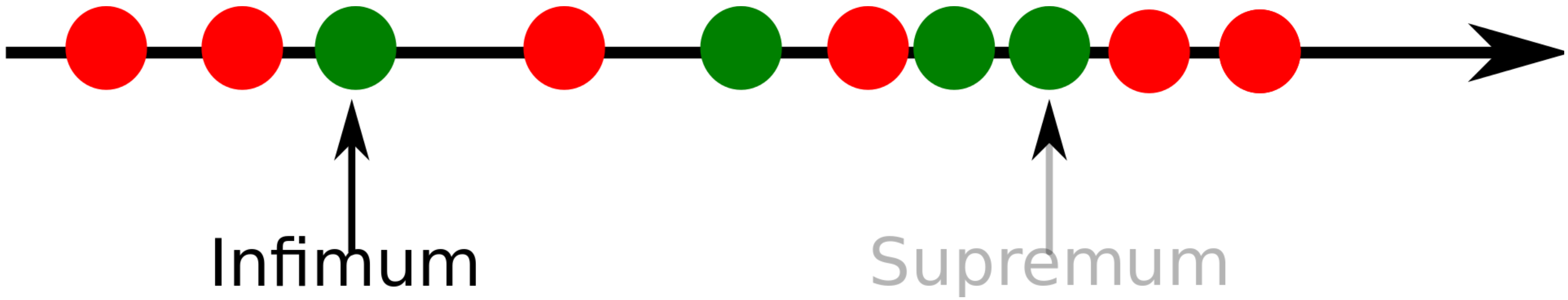
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, are the inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^\star = -\infty$ if problem is unbounded below



A set *T* of real numbers (red and green circles), a subset *S* of *T* (green circles), and the infimum of *S*. Note that for finite, totally ordered sets the infimum and the minimum are equal.

Optimal and locally optimal points

x is **feasible** if $x \in \textbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; $X_{\rm opt}$ is the set of optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \leq 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\ & \|z - x\|_2 \leq R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x \log x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

Feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

can be considered a special case of the general problem with $f_0(x) = 0$:

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $p^\star = 0$ if constraints are feasible; any feasible x is optimal
- $p^\star = \infty$ if constraints are infeasible

Convex optimization problem

standard form convex optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & a_i^T x = b_i, \quad i = 1, \dots, p \end{array}$$

- $f_0, f_1, \dots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \dots, f_m$ convex)

often written as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & Ax = b \end{array}$$

important property: feasible set of a convex optimization problem is convex

Lagrangian

standard form problem (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$