MEE5114 Advanced Control for Robotics

Lecture 10: Basics of Optimization

Prof. Wei Zhang

SUSTech Insitute of Robotics
Department of Mechanical and Energy Engineering
Southern University of Science and Technology, Shenzhen, China

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Motivation

- Optimization is arguably the most important tool for modern engineering
- Robotics
  - Differential Inverse Kinematics
  - Dynamics
  - Motion planning
  - Whole-body control: formulated as a quadratic program
  - SLAM:
  - Perception
- Machine Learning
  - Linear regression
  - Support vector machine:
  - Deep learning
- other domains
  - Check system stability: SDP
  - Compressive sensing
  - Fourier transform: least square problem
- Roughly speaking, most engineering problems (finding a better design, ensure certain properties of the solution, develop an algorithm), can be formulated as optimization/optimal control problems.

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Real Symmetric Matrices

- $\mathcal{S}^n$: set of real symmetric matrices

- All eigenvalues are real

- There exists a full set of orthogonal eigenvectors

- *Spectral decomposition*: If $A \in \mathcal{S}^n$, then $A = Q\Lambda Q^T$, where Λ diagonal and Q is unitary.

Positive Semidefinite Matrices (1/4)

- $A \in \mathcal{S}^n$ is called *positive semidefinite (p.s.d.)*, denoted by $A \succeq 0$, if $x^T A x \geq 0$, $\forall x \in \mathbb{R}^n$
- $A \in \mathcal{S}^n$ is called *positive definite (p.d.)*, denoted by $A \succ 0$, if $x^T A x > 0$ for all nonzero $x \in \mathbb{R}^n$
- $\mathcal{S}^n_+$: set of all p.s.d. (symmetric) matrices
- $\mathcal{S}^n_{++}$: set of all p.d. (symmetric) matrices
- p.s.d. or p.d. matrices can also be defined for non-symmetric matrices.
  e.g.: $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

- We assume p.s.d. and p.d. are symmetric (unless otherwise noted)
- Notation: $A \succeq B$ (resp. $A \succ B$) means $A - B \in \mathcal{S}^n_+$ (resp. $A - B \in \mathcal{S}^n_{++}$)

Positive Semidefinite Matrices (2/4)

- Other equivalent definitions for symmetric p.s.d. matrices:
  - All $2^n - 1$ principal minors of A are nonnegative
  - All eigs of A are nonnegative
  - There exists a factorization $A = B^T B$
- Other equivalent definitions for p.d. matrices:
  - All n leading principal minors of A are positive
  - All eigs of A are strictly positive
  - There exists a factorization $A = B^T B$ with B square and nonsingular.

Positive Semidefinite Matrices (3/4)

- Useful facts:
  - If T nonsingular, $A \succ 0 \Leftrightarrow T^T AT \succ 0$; and $A \succeq 0 \Leftrightarrow T^T AT \succeq 0$
  - Inner product on $\mathbb{R}^{m \times n}$: $< A, B > \triangleq tr(A^T B) \triangleq A \bullet B$.
  - For $A, B \in \mathcal{S}^n_+$, $tr(AB) \geq 0$

Positive Semidefinite Matrices (4/4)

- For any symmetric $A \in \mathcal{S}^n$,

$$\lambda_{\min}(A) \geq \mu \Leftrightarrow A \succeq \mu I \quad \text{and} \quad \lambda_{\max}(A) \leq \beta \Leftrightarrow A \preceq \beta I$$

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Affine Sets and Functions (1/3)

- Linear mapping: $f(x+y) = f(x) + f(y)$ and $f(\alpha x) = \alpha x$, for any x, y in some vector space, and $\alpha \in \mathbb{R}$

  - $f(x) = Ax$, $x \in \mathbb{R}^3$, $A \in SO(3)$

  - $f[x] = \int x(\tau) d\tau$, for all integrable function $x(\cdot)$

  - $E(x)$ expection of a random variable/vector x

  - $f(x) = \mathrm{tr}(x)$, $x \in R^{n \times n}$

Affine Sets and Functions (2/3)

- Affine mapping: $f(x)$ is an affine mapping of x if $g(x) \triangleq f(x) - f(x_0)$ is a linear mapping for some fixed x_0

- Finite-dimension representation of affine function: $f(x) = Ax + b$

- Homogeneous representation in $\mathbb{R}^n$:

$$f(x) = Ax + b \quad \Leftrightarrow \quad \tilde{f}(\tilde{x}) = \tilde{A}\tilde{x},$$

$$\text{with } \tilde{A} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}, \tilde{x} = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

- Linear and affine are often used interchangeably

Affine Sets and Functions (3/3)

- Linear/affine sets: $\{x : f(x) \leq 0\}$ for affine mapping f
  - Line/hyperplane: $a^T x = b$
  - Half space: $a^T x \leq b$
  - Polyhedron: $Hx \leq h$
  - For matrix variable $X \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(AX) \leq 0$ for given constant matrix $A \in \mathbb{R}^{n \times n}$ is a halfspace in $\mathbb{R}^{n \times n}$

Quadratic Sets and Functions

- Quadratic functions in $\mathbb{R}^n$: $f(x) = x^T A x + b^T x + c$

- Quadratic functions (homogeneous form): $f(x) = x^T A x$

  - $A \in \mathcal{S}_+ \Leftrightarrow f(x) \geq 0, \forall x \in \mathbb{R}^n$

- Quadratic sets: $\{x :\in \mathbb{R}^n : f(x) \leq 0\}$ for some quadratic function f
  - e.g.: Ball:

  - e.g.: Ellipsoid:

Convex Set

- *Convex Set*: A set S is convex if

$$x_1, x_2 \in S \quad \Rightarrow \quad \alpha x_1 + (1-\alpha)x_2 \in S, \forall \alpha \in [0,1]$$

- *Convex combination* of $x_1, \ldots, x_k$:

$$\left\{ \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k : \alpha_i \geq 0, \text{ and } \sum_i \alpha_i = 1 \right\}$$

- *Convex hull*: $\overline{co}\{S\}$ set of all convex combinations of points in S

Cone

- A set S is called a *cone* if $\lambda > 0$, $x \in S \Rightarrow \lambda x \in S$.

- Conic combination of x_1 and x_2:
  $x = \alpha_1 x_1 + \alpha_2 x_2$ with $\alpha_1, \alpha_2 \geq 0$

- *Convex cone:*
  1. a cone that is convex
  2. equivalently, a set that contains all the conic combinations of points in the set

Positive Semidefinite Cone

- The set of positive semidefinite matrices (i.e. $\mathcal{S}_+^n$) is a convex cone and is referred to as the *positive semidefinite (PSD) cone*

- Recall that if $A, B \in \mathcal{S}_+^n$, then $tr(AB) \geq 0$. This indicates that the cone $\mathcal{S}_+^n$ is acute.

Operations that Preserve Convexity (1/1)

- Intersection of possibly infinite number of convex sets:
  - e.g.: polyhedron:
  - e.g.: PSD cone:
- Affine mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ (i.e. $f(x) = Ax + b$)
  - $f(X) = \{f(x) : x \in X\}$ is convex whenever $X \subseteq \mathbb{R}^n$ is convex
    e.g.: Ellipsoid: $E_1 = \{x \in \mathbb{R}^n : (x - x_c)^T P(x - x_c) \leq 1\}$ or equivalently
    $E_2 = \{x_c + Au : \|u\|_2 \leq 1\}$
  - $f^{-1}(Y) = \{x \in \mathbb{R}^n : f(x) \in Y\}$ is convex whenever $Y \subseteq \mathbb{R}^m$ is convex
    e.g.: $\{Ax \leq b\} = f^{-1}(\mathbb{R}^n_+)$, where $\mathbb{R}^n_+$ is nonnegative orthant

Convex Function

Consider a finite dimensional vector space $\mathcal{X}$. Let $\mathcal{D} \subset \mathcal{X}$ be convex.

Definition 1 (Convex Function).

A function $f : \mathcal{D} \to \mathbb{R}$ is called convex if

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2), \forall x_1, x_2 \in \mathcal{D}, \forall \alpha \in [0,1]$$

- $f : \mathcal{D} \to \mathbb{R}$ is called strictly convex if $f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2), \forall x_1 \neq x_2 \in \mathcal{D}, \forall \alpha \in [0,1]$

- $f : \mathcal{D} \to \mathbb{R}$ is called concave if $-f$ is convex

How to Check a Function is Convex?

- Directly use definition

- First-order condition: If f is differentiable over an open set that contains $\mathcal{D}$, then f is convex over $\mathcal{D}$ iff

$$f(z) \geq f(x) + \nabla f(x)^T (z - x), \forall x, z \in \mathcal{D}$$

- Second-order condition: Suppose f is twicely differentiable over an open set that contains $\mathcal{D}$, then f is convex over $\mathcal{D}$ iff

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathcal{D}$$

- Many other conditions, tricks,...

Examples of Convex Functions

- In general, affine functions are both convex and concave
  - e.g.: $f(x) = a^T x + b$, for $x \in \mathbb{R}^n$
  - e.g.: $f(X) = tr(A^T X) + c = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + c$, for $X \in \mathbb{R}^{m \times n}$
- Quadratic functions: $f(x) = x^T Q x + b^T x + c$ is convex iff $Q \succeq 0$
- All norms are convex
  - e.g. in $\mathbb{R}^n$: $f(x) = \|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$; $f(x) = \|x\|_\infty = \max_k |x_k|$
  - e.g. in $\mathbb{R}^{m \times n}$: $f(X) = \|X\|_2 = \sigma_{\max}(X)$

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Nonlinear Optimization Problems

Nonlinear Optimization:

$$\begin{cases} \text{minimize:} & f_0(x) \\ \text{subject to:} & f_i(x) \leq 0, i = 1, \ldots m \\ & h_i(x) = 0, i = 1, \ldots, q \end{cases}$$

- decision variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, referred to as *primal problem*
- optimal value p^*
- is called a convex optimization problem if $f_0, \ldots, f_m$ are convex and $h_1, \ldots, h_q$ are affine
- typically convex optimization can be solved efficiently

Nonlinear Optimization Problems

Lagrangian

Associated **Lagrangian**: $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}$

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

- weighted sum of objective and constraints functions
- λ_i: Lagrangian multiplier associated with $f_i(x) \leq 0$
- ν_i: Lagrangian multiplier associated with $h_i(x) = 0$

Lagrange Dual Problems (1/2)

Lagrange dual function: $g : \mathbb{R}^m \times \mathbb{R}^q :\to \mathbb{R}$

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

- g is concave, can be $-\infty$ for some λ, ν
- **Lower bound property**: If $\lambda \succeq 0$ (elementwise), then $g(\lambda, \nu) \leq p^*$

Lagrange Dual Problems (2/2)

Lagrange Dual Problem:

$$\begin{cases} \text{maximize :} & g(\lambda, \nu) \\ \text{subject to:} & \lambda \succeq 0 \end{cases}$$

- Find the best lower bound on p^* using the Lagrange dual function
- a convex optimization problem even when the primal is nonconvex
- optimal value denoted d^*
- (λ, ν) is called **dual feasible** if $\lambda \succeq 0$ and $(\lambda, \nu) \in \textbf{dom}(g)$
- Often simplified by making the implicit constraint $(\lambda, \nu) \in \textbf{dom}(g)$ explicit

Duality Theorems

- **Weak Duality:** $d^* \leq p^*$
  - always hold (for convex and nonconvex problems)
  - can be used to find nontrivial lower bounds for difficult problems
- **Strong Duality:** $d^* = p^*$
  - not true in general, but typically holds for convex problems
  - conditions that guarantee strong duality in convex problems are called *constraint qualifications*
  - Slater's constraint qualification: Primal is strictly feasible

General Optimality Conditions (1/3)

For general optimization problem:

$$\begin{cases} \text{minimize:} & f_0(x) \\ \text{subject to:} & f_i(x) \leq 0, i = 1, \ldots m \\ & h_i(x) = 0, i = 1, \ldots, q \end{cases}$$

General optimality condition:

strong duality and (x^*, λ^*, ν^*) **is primal-dual optimal** $\Leftrightarrow$

- $x^* = \arg\min_x L(x, \lambda^*, \nu^*)$ (Lagrange optimality)
- $\lambda_i^* f_i(x^*) = 0$ for all i (Complementarity)
- $f_i(x^*) \leq 0$ $h_j(x^*) = 0$, for all i, j (primal feasibility)
- $\lambda_i^* \geq 0$ for all i (dual feasibility)

General Optimality Conditions (2/3)

Proof of Necessity

- Assume x^* and (λ^*, ν^*) are primal-dual optimal slns with zero duality gap,

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \min_{x \in \mathcal{D}} \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \nu_j^* h_j(x) \right) \\ &\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \nu_j^* h_j(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

- Therefore, all inequalities are actually equalities

- Replacing the first inequality with equality $\Rightarrow x^* = \text{argmin}_x L(x, \lambda^*, \nu^*)$

- Replacing the second inequality with equality $\Rightarrow$ complementarity condition

General Optimality Conditions (3/3)

Proof of Sufficiency

- Assume (x^*, λ^*, ν^*) satisfies the optimality conditions:

$$\begin{aligned} g(\lambda^*, \nu^*) &= f(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \nu_j^* h_j(x^*) \\ &= f(x^*) \end{aligned}$$

- The first equality is by Lagrange optimality, and the 2nd equality is due to complementarity

- Therefore, the duality gap is zero, and (x^*, λ^*, ν^*) is the primal dual optimal solution

KKT Conditions

For **convex** optimization problem:

$$\begin{cases} \text{minimize:} & f_0(x) \\ \text{subject to:} & f_i(x) \leq 0, i = 1, \ldots m \\ & h_i(x) = 0, i = 1, \ldots, q \end{cases}$$

Suppose duality gap is zero, then (x^*, λ^*, ν^*) is primal-dual optimal if and only if it satisfies the Karush-Kuhn-Tucker (KKT) conditions

- $\frac{\partial L}{\partial x}(x, \lambda^*, \nu^*) = 0$ (Stationarity)
- $\lambda_i^* f_i(x^*) = 0$ for all i (Complementarity)
- $f_i(x^*) \leq 0 \; h_j(x^*) = 0$, for all i, j (primal feasibility)
- $\lambda_i^* \geq 0$ for all i (dual feasibility)

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Linear Program: Primal and Dual Formulations

- **Primal Formulation:** $$\begin{cases} \text{minimize:} & c^T x \\ \text{subject to:} & Ax = b \\ & x \geq 0 \end{cases}$$

- **Its Dual:** $$\begin{cases} \text{maximize:} & -b^T \nu \\ \text{subject to:} & A^T \nu + c \geq 0 \end{cases}$$

Outline

- Motivation
- Some Linear Algebra
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program

Unconstrained Quadratic Program: Least Squares

- minimize: $\quad J(x) = \frac{1}{2}x^T Q x + q^T x + q_0$
- Problem is convex iff $Q \succeq 0$
- When J is convex, it can be written as: $J(x) = \|Q^{\frac{1}{2}}x - y\|^2 + c$
- KKT condition:
- Optimal solution:

Equality Constrained Quadratic Program

- Standard form: $\begin{cases} \min_x & J(x) = x^T Q x + q^T x + q_0 \\ \text{subject to:} & Hx = h \end{cases}$
- The problem is convex if $Q \succeq 0$
- KKT Condition:
- Optimal Solution:

General Quadratic Program

- Standard form: $\begin{cases} \text{minimize:} & J(x) = x^T Q x + q^T x + q_0 \\ \text{subject to:} & Ax \leq b \end{cases}$
- Dual problem:

More Discussions

-

More Discussions

-