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

control: linear control / LQR / Adaptive control / Intelligent control
Fuzzy system / robust control / nonlinear control / Hybrid system
/ sliding mode control (variable structure) (OSU)

Outline

MPC → optimal control

distributed control / decentralized / multiagent system

- Motivation

nonlinear stability ← optimization based stability

- Some Linear Algebra

optimization $\dot{x}_i : \dot{x}$

- Sets and Functions

optimal control ← Dynamical system

- Short Introduction to Optimization

optimization

- Linear Program
- Quadratic Program

Motivation

- Optimization is arguably the most important tool for modern engineering
- Robotics
  - Differential Inverse Kinematics ← "QP"
  - Dynamics : RNEA. (ABA ⟷ LQR)
  - 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

optimization

theoretical aspect ⟹ Algorithmic aspect

⇑ this lecture

⇓ Implement coding

code/packages.

Our goal:
- Basic knowledge/key concepts of optimization theory.
- Formulate/reformulate optimization problems.
- Educated users of packages.

convex optimization

boyd

Real Symmetric Matrices

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

$\mathcal{S}^n \subseteq \mathbb{R}^{n\times n}$ $\quad A \in S^n \iff A^T = A$

- All eigenvalues are real (diagonalizable)

eig(B) $\quad B \in \mathbb{R}^{n\times n}$, eig(B) has n eigs

- There exists a full set of orthogonal eigenvectors

$A \in S^n \Rightarrow A = T\Lambda T^{-1}$

Λ: diagonal

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

Q is unitary: $\underbrace{Q^TQ = I \,;\, QQ^T = I}$

$Q = [q_1 \,\vdots\, q_2 \cdots \vdots\, q_n]$

$\Downarrow$

$$q_i^T q_j = \begin{cases} 1 & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

Positive Semidefinite Matrices (1/4)

$x_1^2 + x_1x_2 + 3x_2^2 = \begin{bmatrix}x_1\\x_2\end{bmatrix}^T \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$

- $A \in \mathcal{S}^n$ is called *positive semidefinite (p.s.d.)*, denoted by $A \succeq 0$, if $x^T A x \geq 0$, $\forall x \in \mathbb{R}^n$

  scalar

  x^TAx is a quadratic fun of x.

- $A \in \mathcal{S}^n$ 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.: $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

  $x^TAx = \begin{bmatrix}x_1\\x_2\end{bmatrix}^T \begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = x_1^2 + x_2^2$

$B \Longrightarrow B^{sym} = \frac{1}{2}(B + B^T)$

symmetrization $= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

- 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. $A - B \in \mathcal{S}^n_{++}$)

partial order:

$\Longleftrightarrow A - B \in S_+^n$; $A \succeq 0$

Note: $\exists$ matrices: $C \not\succeq D$ & $D \not\succeq C$

Positive Semidefinite Matrices (2/4)

$A \geq B$ (element-wise)
$\Leftrightarrow A_{ij} \geq B_{ij}$

- 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 eig(A), $\lambda_1, \lambda_2 \cdots \lambda_n \geq 0$
  - There exists a factorization $A = B^T B$
- Other equivalent definitions for $\boxed{\text{p.d.}}$ matrices:
  - All n leading principal minors of A are positive
  - All eigs of A are strictly positive $\lambda_1 \cdots \lambda_n > 0$
  - There exists a factorization $A = B^T B$ with B square and nonsingular.

$\Lambda = \begin{bmatrix} 2 & \\ & 4 \end{bmatrix}$

$\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{2} & \\ & 2 \end{bmatrix}$

· If $A \in S_+^n$, $A = Q \Lambda Q^T = Q \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} Q^T$ $\cancel{= Q \Lambda^{\frac{1}{2}} Q^T \, Q \Lambda^{\frac{1}{2}} Q^T}$

$= B^T B$, $B = \Lambda^{\frac{1}{2}} Q^T$

I $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

Positive Semidefinite Matrices (3/4)

- Useful facts:
  - If T nonsingular, $A \succ 0 \Leftrightarrow T^T A T \succ 0$; and $A \succeq 0 \Leftrightarrow T^T A T \succeq 0$

does not need to be unitary

$\begin{cases} TAT^{-1} : \text{similarity transformation} \\ T^TAT : \text{congruent transformation} \end{cases}$

$\begin{cases} S_+^n \\ S_{++}^n \end{cases}$ invariant under congruent transformation

  - Inner product on $\mathbb{R}^{m \times n}$: $< A, B > \triangleq tr(A^T B) \triangleq A \bullet B$.

$\forall A \in \mathbb{R}^{m\times n}, \ B \in \mathbb{R}^{m \times n}$;

$$\boxed{tr(A^TB) = tr(AB^T) = tr(B^TA) = tr(BA^T) = \sum_i \sum_j A_{ij} B_{ij}}$$

Angle between A, B

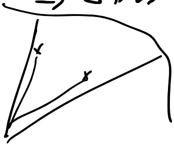
$$\cos\theta = \frac{<A, B>}{\sqrt{<A,A>}\sqrt{<B,B>}}$$

$A \perp B \Leftrightarrow tr(A^TB) = 0$

  - For $A, B \in \mathcal{S}_+^n$, $tr(AB) \geq 0$

A, B square, $\in S_+^n$ P.S.D.

$\Rightarrow <A, B> = tr(A^TB) = tr(AB)$



proof : HW

Positive Semidefinite Matrices (4/4)

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

$$\left(\underbrace{\lambda_{\min}(A) \geq \mu}_{\det(\lambda I - A) = 0}\right) \Leftrightarrow \underbrace{A \succeq \mu I}_{A - \mu I \succeq 0} \quad \text{and} \quad \lambda_{\max}(A) \leq \beta \Leftrightarrow A \preceq \beta I$$

proof: ~~HW~~

$\Rightarrow$ $A = Q \Lambda Q^T$ for unitary Q ($QQ^T = I$)

$A - \mu I = Q(\Lambda - \mu I)Q^T$

congruent transformation does not change psd.

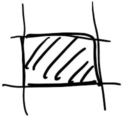
$A - \mu I \succeq 0 \iff \Lambda - \mu I \succeq 0$

$\iff \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} - \begin{bmatrix} \mu & & \\ & \mu & \\ & & \ddots \ \mu \end{bmatrix} \succeq 0$

$\iff \lambda_{\min}(A) \geq \mu$

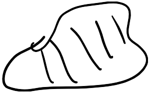
Outline

- Motivation
- Some Linear Algebra
- **Sets and Functions** $\Leftarrow$
- Short Introduction to Optimization
- Linear Program
- Quadratic Program



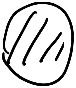
$g(x) \le 0$

optimization



min cost func $\longleftrightarrow f(x)$

subject to constraints $\leftrightarrow x \in \mathcal{X}$.



$\|x\| \le 1$

$\underbrace{(\|x\| - 1)}_{g(x)} \le 0$

Affine Sets and Functions (1/3)

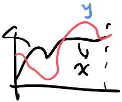
- Linear mapping: $f(x+y) = f(x) + f(y)$ and $f(\underset{\tau}{\alpha} x) = \cancel{\alpha x}\ \alpha f(x)$, for any x, y in some vector space, and $\alpha \in \mathbb{R}$

Examples:

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

  $f(x+y) = A(x+y) = Ax + Ay \cdots$

- $f[x] = \int x(\tau)d\tau$, for all integrable function $x(\cdot)$ $x(\cdot) \in \mathbb{R}$

  f: "traj" $\longrightarrow \mathbb{R}$

  $f[x+y] = \int (x(\tau) + y(\tau))d\tau = \int x(\tau)d\tau + \int y(\tau)d\tau = f[x] + f[y]$

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

  x is random variable $E(x) = \int x f_X(x)dx$

  $E(x+y) = E(x) + E(y)$

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

  $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ $\mathrm{tr}(\alpha A) = \alpha\,\mathrm{tr}(A)$

  $\min_{B} \mathrm{tr}(A_0 B) \in$ L.P.

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

linear fun: $f(x) = Ax$

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

$g(x) = f(x) - f(0)$

- 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)

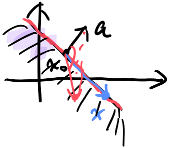
level set : $f(x)=0$

- Linear/affine sets: $\{x : \underbrace{f(x) \le 0}_{\text{sublevel set}}\}$ for affine mapping f

  $f: \mathbb{R}^n \to \mathbb{R}$

  - Line/hyperplane: $a^T x = b$

    $a^T(x-x_0)=0 \Rightarrow a^T x = \underbrace{a^T x_0}_{b}$



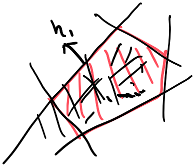
  - Half space: $a^T x \le b$

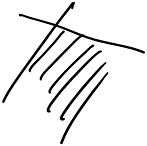
    $a^T(x-x_0) \le 0 \Rightarrow a^T x \le \underbrace{a^T x_0}_{b}$

  - Polyhedron: $Hx \le \cancel{h}$ b

    $\begin{matrix} h_1^T x \le b_1 \\ h_2^T x \le b_2 \\ \vdots \\ h_5^T x \le b_5 \end{matrix} \Rightarrow \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \\ h_5^T \end{bmatrix} x \le \begin{bmatrix} b \end{bmatrix}, \quad \underbrace{H}_{\in \mathbb{R}^{m\times n}} \underbrace{x}_{\in\mathbb{R}^n} \le b$

  - For matrix variable $X \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(AX) \le 0$ for given constant matrix $A \in \mathbb{R}^{n\times n}$ is a halfspace in $\mathbb{R}^{n\times n}$





Quadratic Sets and Functions

$f(x)$ p.s.d iff $f(x) \geq 0 \; \forall x$

$f(x) = x_1^2 + x_2^2 + 3x_1x_2$

$= x^T \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & 1 \end{bmatrix} x$

$x^T A x$

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

  $x \in x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

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

  Let $\tilde{x} = \begin{bmatrix} x \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1}$

  $f(\tilde{x}) = \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}^T}_{\tilde{x}} \underbrace{\begin{bmatrix} A & \frac{b}{2} \\ \frac{b^T}{2} & c \end{bmatrix}}_{\tilde{A}} \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}}_{\tilde{x}}$

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

    positive semidefinite

  $f(x)$ is p.s.d over $D \subseteq \mathbb{R}^n$

  $f(x) \geq 0, \; \forall x \in D$

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

    $\{x \in \mathbb{R}^n : \|x - x_c\|^2 \leq r^2\}$

    $f(x) = \|x - x_c\|^2 - r^2 \leq 0$

    $= (x - x_c)^T (x - x_c) - r^2$

  - e.g.: Ellipsoid:

    $\{x \in \mathbb{R}^n : (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$

    $P \in S_{++}$

    ~~Kalman~~: confidence ellipsoid

    Gaussian

Convex Set

Linear combination: $x_1, x_2 \quad \alpha_1 x_1 + \alpha_2 x_2$

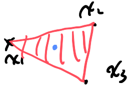
- *Convex Set*: A set S is convex if

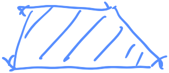
$$x_1, x_2 \in S \quad \Rightarrow \quad \underbrace{\alpha x_1 + (1-\alpha) x_2}_{\text{Convex combination of } x_1, x_2} \in S, \forall \alpha \in [0, 1]$$

$= \alpha_1 x_1 + \alpha_2 x_2, \ \alpha_1, \alpha_2 \geq 0, \ \alpha_1 + \alpha_2 = 1$

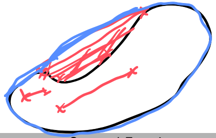
- *Convex combination* of $x_1, \ldots, x_k$:

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





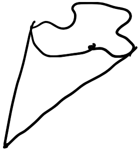
- *Convex hull*: $\overline{co}\,\{S\}$ set of all convex combinations of points in S may not be convex



Cone

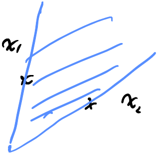
- A set S is called a *cone* if $\lambda > 0$, $x \in S \Rightarrow \lambda x \in S$.

nonconvex cone



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

$\mathrm{cone}(x_1, \cdots x_k) = \left\{ \sum_{i=1}^{k} \alpha_i x_i \ : \ \alpha_i \geq 0 \right\}$



- *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 (PSD cone) S_+^n

- 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*

S_+^n: set of PSD matrices

pick any $A \in S_+^n \Rightarrow \lambda A \succeq 0 \Rightarrow \lambda A \in S_+^n$

$\Rightarrow S_+^n$ is a cone

is S_+^n convex?

pick: $A, B \in S_+^n$

$\alpha A + (1-\alpha) B \overset{?}{\in} S_+^n$

$\alpha (\underbrace{x^T A x}_{\geq 0}) + (1-\alpha) \underbrace{x^T B x}_{\geq 0} \geq 0$

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

$\langle A, B \rangle = tr(A^T B) = tr(AB)$

Operations that Preserve Convexity (1/1)

- Intersection of possibly infinite number of convex sets:
  - e.g.: polyhedron:

    $\left.\begin{array}{l} a_1^T x \le b_1 \\ a_2^T x \le b_2 \end{array}\right\} \Rightarrow \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} x \le \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

  - e.g.: PSD cone:

    X is convex, $f(X) = \{f(x) : x \in X\}$

    $f(x)$ is affine

- 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: $E_1 = \{x \in \mathbb{R}^n : (x - x_c)^T P (x - x_c) \le 1\}$ or equivalently
    $E_2 = \{x_c + Au : \|u\|_2 \le 1\}$

    Ball: $\{x \in \mathbb{R}^n : \|x\|^2 \le 1\}$

    Define: $f(x) = P^{\frac{1}{2}}(x - x_c) \Rightarrow f(\text{Ball}) = E_1$
  - $f^{-1}(Y) = \{x \in \mathbb{R}^n : f(x) \in Y\}$ is convex whenever $Y \subseteq \mathbb{R}^m$ is convex
    e.g.: $\{Ax \le b\} = f^{-1}(\mathbb{R}_+^n)$, where $\mathbb{R}_+^n$ is nonnegative orthant

Convex Function

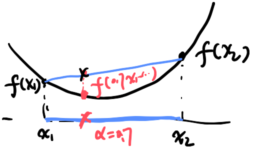
Consider a finite dimensional vector space $\mathcal{X}$. Let $\mathcal{D} \subset \mathcal{X}$ be convex.

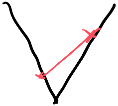
Definition 1 (Convex Function).

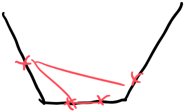
A function $f : \mathcal{D} \to \mathbb{R}$ is called convex if

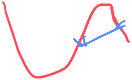
→ "<" strictly convex

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



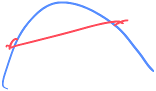






- $f : \mathcal{D} \to \mathbb{R}$ is called strictly convex if $f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2), \forall x_1 \neq x_2 \in \mathcal{D}, \forall \alpha \in [0,1]$

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

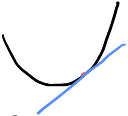


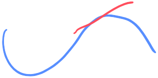
How to Check a Function is Convex?

- 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

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





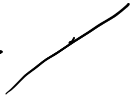
- Second-order condition: Suppose f is twicely differentiable over an open set that contains $\mathcal{D}$, then f is convex over $\mathcal{D}$ iff

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

- 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: $f(x) = \frac{1}{2} x^T Q x + b^T x + c$ is convex iff $Q \succeq 0$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdot \\ & & \\ & & \end{bmatrix} = Q$$

sencond order condition

- All norms are convex
  - e.g. in $\mathbb{R}^n$: $f(x) = \|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$; $f(x) = \|x\|_\infty = \max_k |x_k|$

$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$

  - e.g. in $\mathbb{R}^{m \times n}$: $f(X) = \|X\|_2 = \sigma_{\max}(X)$

$\|x\|_\infty = \max_k |x_k|$

$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$

Outline

- Affine mapping of convex fun is convex

$f(x)$ is convex : $g(x) \triangleq a f(x) + b$ is also convex

$f: \mathbb{R}^n \to \mathbb{R}$

- Motivation

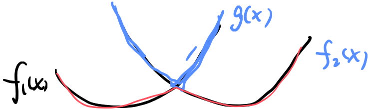
- pointwise maximum of convex func is convex.

- Some Linear Algebra

suppose $f_1(x)$, $f_2(x)$, $g(x) \triangleq \max\{f_1(x), f_2(x)\}$ is convex

$\downarrow$ convex $\quad \downarrow$ convex

- Sets and Functions



- Short Introduction to Optimization

suppose $f(x;\theta)$ is convex for any $\theta \in [a,b]$

- Linear Program

then $g(x) \triangleq \max_{\theta \in [a,b]} f(x,\theta)$ is convex

- Quadratic Program

e.g. $f(x;\theta) = \theta x + b \Rightarrow g(x) = \left(\max_{\theta \in [1,2]} \theta x + b \right)$

-. X. pointwise minimum of concave is concave.

Nonlinear Optimization Problems

Nonlinear Optimization:

cost func

$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

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

inequalities

equalities

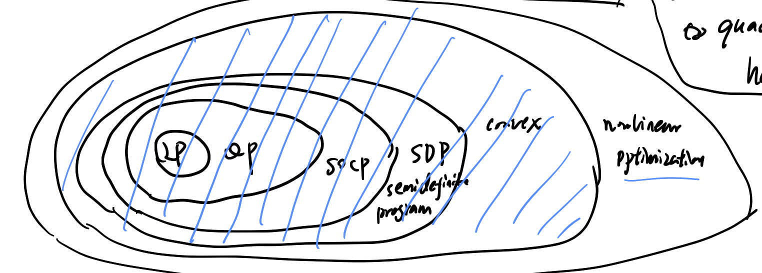
$C = \{x \in \mathbb{R}^n : f_i(x) \le 0, i=1..m;\ h_i(x) = 0, i=1..q\}$

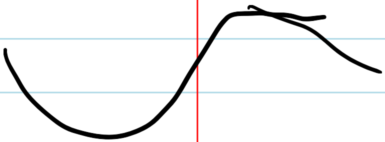
- decision variable $\underline{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, referred to as *primal problem*
- optimal value p^*
- is called a convex optimization problem if $f_0, \dots, f_m$ are convex and $h_1, \dots, h_q$ are affine
- typically convex optimization can be solved efficiently

Nonlinear Optimization Problems

- Categories:

objective func.	constraint set / func.	
- Linear/affine	Linear/affine	⇒ Linear program (LP)
- Quadratic (convex)	Linear/affine	⇒ Quadratic Program (QP)
- Quadratic (convex)	Quadratic	⇒ QCQP Quadratically constrained quadratic program. hard to solve





- How to find optimal sln? $\min_x f(x)$

· optimality condition for unconstrained optimization

· 1^{st} order: x^* is local optimizer ~~for~~

then $\nabla f(x^*) = 0$ ... ① $\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = 0$

x^*

e.g. pick $\frac{\partial f}{\partial x_1} \neq 0 \Rightarrow \begin{cases} \frac{\partial f}{\partial x_1} > 0 \\ \frac{\partial f}{\partial x_1} < 0 \end{cases}$ $x^* + \begin{bmatrix} -\varepsilon \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$

$x^* + \begin{bmatrix} \varepsilon \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$

For convex problem, condition ① guarantees x^* is global minimizer

Question? what about constrained optimization?

Lagrangian

$\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \qquad \nu = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_q \end{bmatrix}$

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

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

$\lambda^T f \qquad \nu^T h \qquad = 0$

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

  require $\lambda_i \geq 0$, $\forall i$
- ν_i: Lagrangian multiplier associated with $h_i(x) = 0$

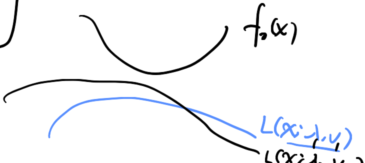
  no requirement for ν_i

If (λ, ν) dual feasible, $\lambda_i \geq 0$,

x is primal feasible

(i.e. $f_i(x) \leq 0$, $h_i(x) = 0$)

then $L(x, \lambda, \nu) \leq f_0(x)$



Lagrange Dual Problems (1/2)

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

min

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

is func of multipliers (λ,ν)

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

True for all f_i, h_i

inf   min

min x = inf x = 0
$(0,1]$ $(0,1]$

① $L(x,\lambda,\nu)$ is affine in (λ,ν)

② $g(\lambda,\nu)$ pointwise min of $L(x,\lambda,\nu)$

pointwise min of concave is concave

<1> • g is concave, can be $-\infty$ for some λ,ν

<2> • **Lower bound property**: If $\lambda \succeq 0$ (elementwise), then $g(\lambda,\nu) \le p^*$ primal optimal

let $\tilde{x}$ be arbitrary feasible primal variable and $\lambda \geq 0$

$f_i(\tilde{x}) \le 0$, $i=1,\dots,m$ $h_j(\tilde{x}) = 0$, $j=1\dots q$

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

$$\Rightarrow p^* = \min_{\tilde{x}\text{ feasible}} f_0(\tilde{x}) \geq g(\lambda,\nu)$$

Lagrange Dual Problems (2/2)

always convex optimization problem

Lagrange Dual Problem:

$g(\lambda,\nu) \le p^*$

$\Rightarrow d^* \le p^*$

$$\begin{cases} \text{maximize}_{\lambda,\nu} : & g(\lambda,\nu) \\ \text{subject to:} & \lambda \succeq 0 \end{cases} \iff \begin{cases} \min & (-g(\lambda,\nu)) \\ \text{subj} & -\lambda \le 0 \end{cases}$$

- 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^* $\quad d^* \le p^*$
- (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

i.e. $\exists \tilde{x}$ such that $f_i(\tilde{x}) < 0$, $h_i(\tilde{x}) = 0$

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 $\begin{cases} f_i(x^*) < 0, \Rightarrow \lambda_i = 0 \\ f_i(x^*) = 0 \Rightarrow \lambda_i \geq 0 \end{cases}$ (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,

$$p^* = f_0(x^*) \overset{\text{zero duality gap assumption}}{=} g(\lambda^*, \nu^*) = d^*$$

$$\min_x L(x, \lambda^*, \nu^*) = \min_{x \in \mathcal{D}} \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \nu_j^* h_j(x) \right)$$

$$\leq L(x^*, \lambda^*, \nu^*) \qquad \leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \nu_j^* h_j(x^*) = L(x^*, \lambda^*, \nu^*)$$

$$\leq \underbrace{f_0(x^*)}_{=p^*} \quad (\lambda \geq 0, \quad h = 0)$$

- 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:

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

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)
- $f_i(x^*) \le 0$ $h_j(x^*) = 0$, for all i, j (primal feasibility)
- $\lambda_i^* \ge 0$ for all i (dual feasibility)

Outline

Linear Program: Primal and Dual Formulations

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

$x \in \mathbb{R}^n$

q – equality constraint number

n – inequality constraint

$x_1 \geq 0,\ x_2 \geq 0 \ldots \ x_n \geq 0$

$-x_1 \leq 0 \ \ldots, \ -x_n \leq 0$

Lagrangian: $L(x, \lambda, \nu) = c^T x + \lambda^T(-x) + \nu^T(Ax - b)$

$\Rightarrow g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) = \inf_{x \in \mathbb{R}^n} \{ (c^T - \lambda^T + \nu^T A) x - \nu^T b \}$

$= \begin{cases} -\infty & \text{if } c^T - \lambda^T + \nu^T A \neq 0 \\ -b^T \nu & \text{if } \ldots = 0 \end{cases}$

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

$\max_{\lambda, \nu} g(\lambda, \nu)$

subj to: $\begin{cases} \lambda \geq 0 \\ c^T - \lambda^T + \nu^T A = 0 \end{cases}$

- q – variables
- n – inequality constraint

Outline

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

Unconstrained Quadratic Program: Least Squares

$\wedge$

- minimize: $J(x) = \frac{1}{2}x^T Qx + q^T x + q_0$
- 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: $\Rightarrow$ Analytical solution
- 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: $\Rightarrow$ analytical solution
- 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:

Numerical / active set interior point

More Discussions

-

More Discussions

•