MEE5114 Advanced Control for Robotics

Lecture 8: Basics of Optimization

Prof. Wei Zhang

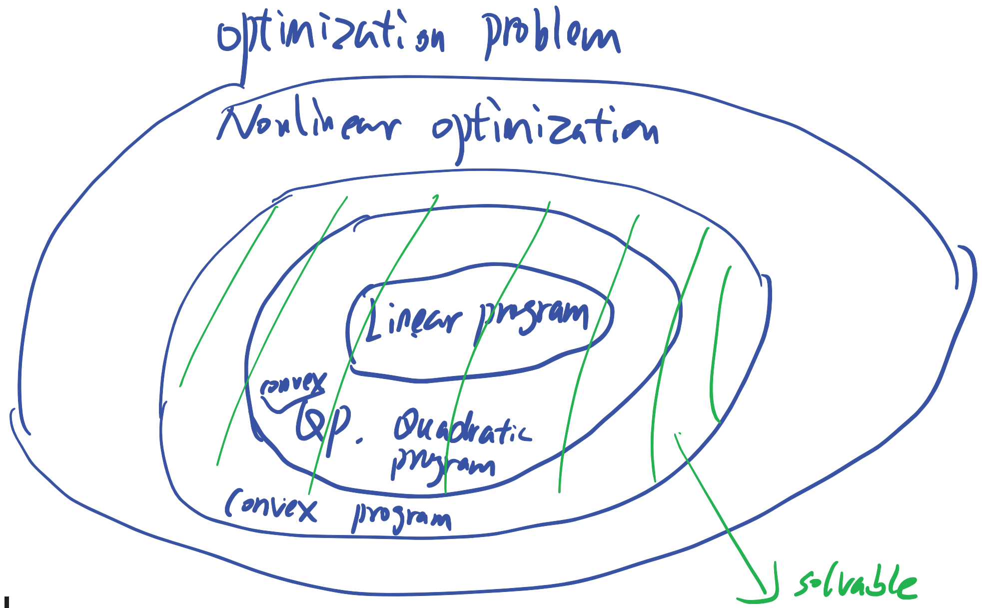
SUSTech Insitute of Robotics
Department of Mechanical and Energy Engineering
Southern University of Science and Technology, Shenzhen, China

Outline

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

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
- Some Multivariable Calculus
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program
- Some Examples

Real Symmetric Matrices

- $\mathcal{S}^n$: set of real symmetric matrices $\subseteq \mathbb{R}^{n\times n}$

  $A \in S^n$ means A symmetric $\Leftrightarrow A = A^T$

- All eigenvalues are real (diagonalizable)

  $A = T\Lambda T^{-1}$ for some nonsingular T

  Λ: diagonal matrix with eigs on the diag.

- There exists a full set of orthogonal eigenvectors

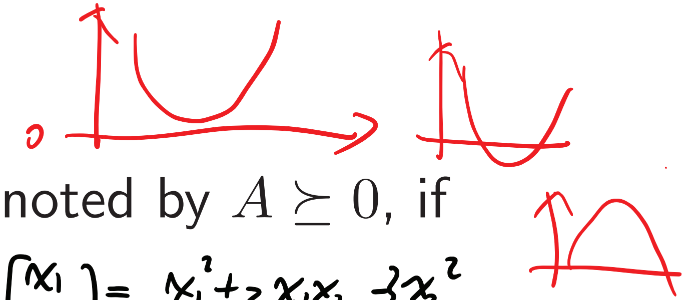
  $\Downarrow$

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

  Q is unitary if $Q^TQ = I$

  $Q = [q_1, q_2, \cdots q_n] \Rightarrow q_i^T q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

Positive Semidefinite Matrices (1/3)



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

  → quadratic form

  e.g. $x^T A x = [x_1 \ x_2] A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 2x_1x_2 - 3x_2^2$

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

  $V(x) = [x_1 \ x_2] \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 > 0$

  $\Downarrow$

  $\Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- We assume p.s.d. and p.d. are symmetric (unless otherwise noted)

- Notation: $\overset{\text{partial order}}{A \succeq B}$ (resp. $A \succ B$) means $A - B \in \mathcal{S}^n_+$ (resp. $A - B \in \mathcal{S}^n_{++}$)

  $5 \geq 3$

  $A \succeq B$, $B \succeq C \Rightarrow A \succeq C$

Positive Semidefinite Matrices (2/3)

- 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 (3/3)

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

    not necessarily unitary

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

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

    $a, b \in \mathbb{R}^n$

    $\langle a, b\rangle = a^T b$

    $\forall A, B \in \mathbb{R}^{m\times n}, \quad tr(A^TB) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A^T_{ij} B_{ji}\right)$

  - For $A, B \in S^n_+$, $tr(AB) \geq 0$

    $\langle A, B\rangle = tr(A^TB) = tr(AB) \geq 0$

    property

    Angle between A, B

    $\cos\theta = \dfrac{\langle A, B\rangle}{\sqrt{\langle A, A\rangle}\sqrt{\langle B, B\rangle}}$

    Set of positive semi-definite set is acute

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

$$\underbrace{\lambda_{\min}(A)}_{\text{smallest eig}} \geq \mu \Leftrightarrow \underbrace{A \succeq \mu I}_{(A-\mu I)\in S^n_+} \quad \text{and} \quad \lambda_{\max}(A) \leq \beta \Leftrightarrow A \preceq \beta I$$

let $A = Q\Lambda Q^T$

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

$= Q(\Lambda - \mu I)Q^T \Rightarrow A - \mu I \succeq 0 \Leftrightarrow \Lambda - \mu I \succeq 0 \Leftrightarrow \lambda_{\min}(A) \geq \mu$

$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

Outline

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

Gradient and Hessian (1/2)

- Consider a multivariate scalar-valued function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ that takes a matrix (or vector) X to a scalar value in $\mathbb{R}$

- **Gradient** of f wrt $X \in \mathbb{R}^{m\times n}$ is defined as

$$\nabla_X f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial X_{11}} & \frac{\partial f(X)}{\partial X_{12}} & \cdots & \frac{\partial f(X)}{\partial X_{1n}} \\ \frac{\partial f(X)}{\partial X_{21}} & \frac{\partial f(X)}{\partial X_{22}} & \cdots & \frac{\partial f(X)}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(X)}{\partial X_{m1}} & \frac{\partial f(X)}{\partial X_{m2}} & \cdots & \frac{\partial f(X)}{\partial X_{mn}} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

- $\nabla_X f(X)$ is an $m \times n$ matrix, its size is the same as X

Gradient and Hessian (2/2)

$$f(x) = f(\hat{x}) + \underbrace{\nabla_x f(\hat{x})^T (x - \hat{x})} + \cdots$$

$$\underbrace{\frac{\partial f}{\partial x_1}(x_1 - \hat{x}_1) + \frac{\partial f}{\partial x_2}(x_2 - \hat{x}_2) + \cdots}$$

$$\Downarrow$$

$$\langle \nabla f(x), \Delta x \rangle$$

- For vector case, $x \in \mathbb{R}^n$, then $\nabla_x f(x)$ is also a vector:

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

- Let $x \in \mathbb{R}^n$, the Hessian matrix of f wrt x is defined as

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Hessian is always symmetric, and we typically do not consider Hessian wrt matrix variables

Examples (1/2)

- $f(x) = b^T x$

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

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

$= b_1 x_1 + b_2 x_2 + \cdots b_n x_n$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} = b$$

- $f(x) = x^T A x$ (Suppose A is symmetric)

$$f(x) = \left(\sum_i \sum_j A_{ij} x_i x_j \right)$$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_k} = \underbrace{\sum_i A_{ik} x_i + \sum_j A_{kj} x_j}$$

$$\Downarrow$$

$$\nabla_x f(x) = Ax + A^T x$$

If A symmetric $\quad \nabla_x f(x) = 2Ax$

Examples (2/2)

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

$\nabla_X f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots \\ & & \\ \cdot & \cdot & \cdot \end{bmatrix}$

Consider matrix variable $X \in \mathbb{R}^{m \times n}$

$f(X) = f(\hat{X}) + \langle \underbrace{\nabla_X f(X), \Delta X}_{\mathrm{tr}(\ )} \rangle$ $\mathrm{tr}(\ )$

- $f(X) = a^T X b$ $a \in \mathbb{R}^m$ $b \in \mathbb{R}^n$

$\nabla f(X) = ab^T \in \mathbb{R}^{m \times n}$

special case

- $f(X) = \mathrm{tr}(AXB) \Leftrightarrow$ by definition after derivation

$$\boxed{\frac{\partial(\mathrm{tr}(AXB))}{\partial X} = A^T B^T}$$

- $f(X) = \det(X)$ $\Rightarrow \frac{\partial f(X)}{\partial X} = \det(X)(X^{-1})^T$ $\quad \frac{\partial |X|}{\partial X} = |X| \cdot (X^{-1})^T$ $X \in \mathbb{R}^{n \times n}$

- $\det(X) = \underbrace{\sum_{i=1}^{n} (-1)^{i+j} X_{ij} |X_{\setminus i \setminus j}|}$, for all j

- $\frac{\partial \det(X)}{\partial X_{kl}} = \frac{\partial(\ )}{\partial X_{kl}} = (-1)^{k+l} |X_{\setminus k \setminus l}| = (\mathrm{adj}(X))_{lk}$

Outline

$$X^{-1} = \frac{1}{|X|} \operatorname{adj}(X)$$

$$\Rightarrow \frac{\partial \det(X)}{\partial X} = \operatorname{adj}(X)^T = |X| (X^{-1})^T$$

- Motivation
- Some Linear Algebra
- Some Multivariable Calculus
- **Sets and Functions**
- Short Introduction to Optimization
- Linear Program
- Quadratic Program
- Some Examples

Affine Sets and Functions (1/3)

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

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

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

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

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

    $$E(x) = \int x \, p(x) \, dx$$

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

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

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

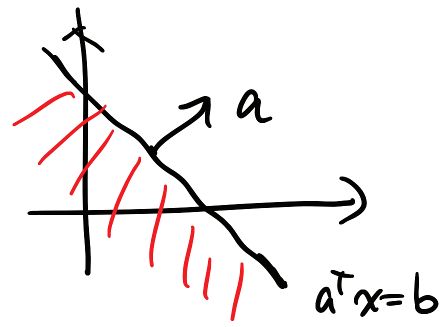
Affine Sets and Functions (3/3)

- Linear/affine sets: $\{x : f(x) \leq 0\}$ for affine mapping f

  $\to$ $f(x)$ zero-sublevel set

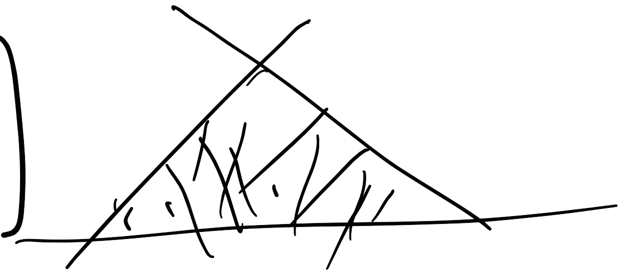
  - Line/hyperplane: $a^T x = b$ $\{x : a^T x = b\}$

  - Half space: $a^T x \leq b$ $\{x : a^T x - b \leq 0\}$



  - Polyhedron: $Hx \leq h$ $H = \begin{bmatrix} H_1^T \\ H_2^T \\ H_3^T \end{bmatrix}$, $h = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$

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



  - 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

$\tilde{x} = \begin{bmatrix} x \\ 1 \end{bmatrix}$

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

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

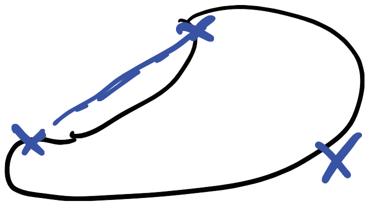
- Quadratic functions (homogeneous form): $f(x) = x^T A x$
  - $A \in \mathcal{S}_+ \Leftrightarrow f(x) \geq 0, \forall x \in \mathbb{R}^n$
- Quadratic sets: $\{x :\in \mathbb{R}^n : f(x) \leq 0\}$ for some quadratic function f
  - e.g.: Ball: $\{x \in \mathbb{R}^n : \|x - x_c\|_2^2 \leq r^2\}$
  - e.g.: Ellipsoid: $\{x \in \mathbb{R}^n : (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$

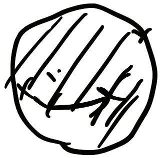
$P \in \mathcal{S}_+^n$

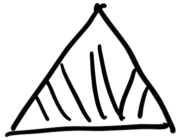
Convex Set

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

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

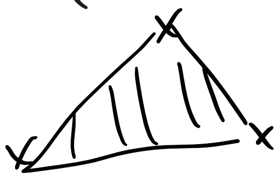




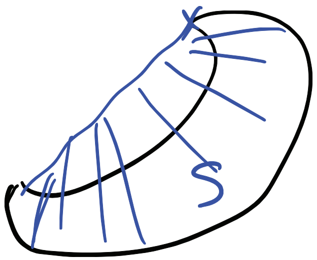


- *Convex combination* of $x_1, \ldots, x_k$:

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

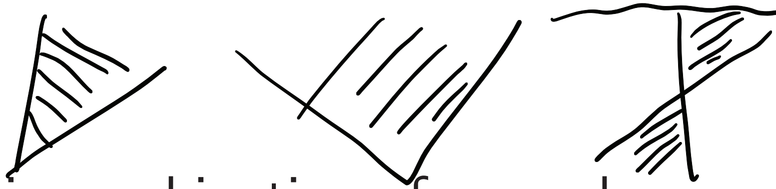


- *Convex hull*: $\overline{co}\{S\}$ set of all convex combinations of points in S

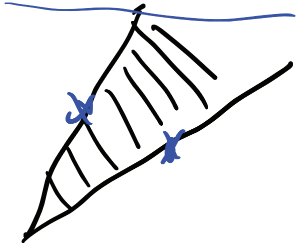


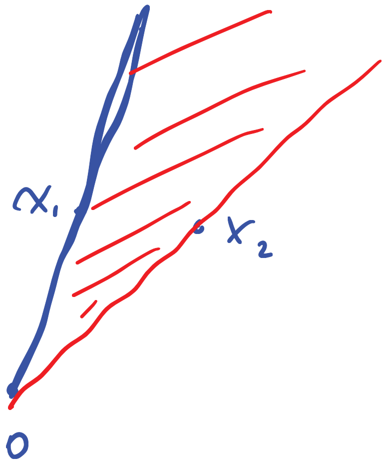
Convex Cone

- 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*

$\forall A, B \in S_+^n$, $(\alpha A + \beta B) \in S_+^n$ for any $\alpha \geq 0$, $\beta \geq 0$

$\forall x \in \mathbb{R}^n$, $x^T(\alpha A + \beta B)x = \alpha(x^T A x) + \beta(x^T B x) \geq 0$

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

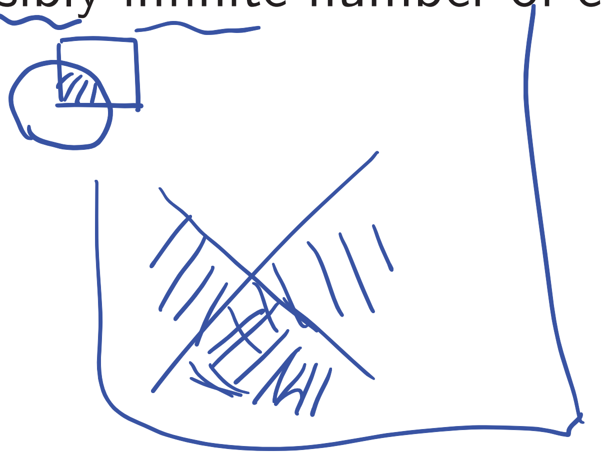
Operations that Preserve Convexity (1/1)

- Intersection of possibly infinite number of convex sets: is also convex

  convex set / like a half space)

  for this particular z

  - e.g.: polyhedron:



  Define $G(z) = \{P \in S^n : z^T P z \geq 0\}$

  $G(z) \supseteq S_+^n, \forall z$

  $S_+^n = \bigcap_{z \in \mathbb{R}^n} G(z)$

  - e.g.: PSD cone:

  $\to S_+^n$

- 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
    $E_2 = \{x_c + Au : \|u\|_2 \leq 1\}$

    $\Downarrow$

    $= \|P^{\frac{1}{2}}(x - x_c)\|^2$

    $E_1 = \{x \in \mathbb{R}^n : \|P^{\frac{1}{2}}(x - x_c)\|^2 \leq 1\}$
  - $f^{-1}(Y) = \{x \in \mathbb{R}^n : f(x) \in Y\}$ 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

    $\text{Ball} = \{x \in \mathbb{R}^n : \|x\|^2 \leq 1\}$

    $f(x) = P^{\frac{1}{2}}(x - x_c)$

    $E_1 = f(\text{Ball})$

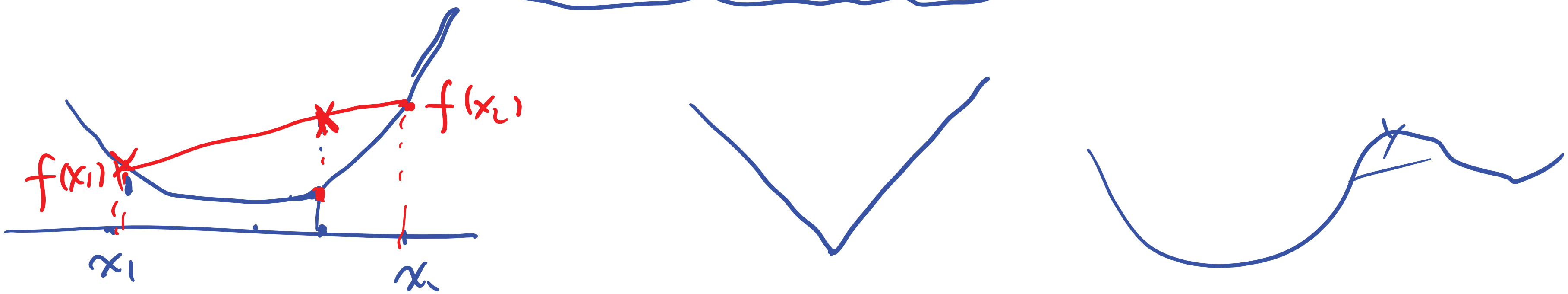
Convex Function

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

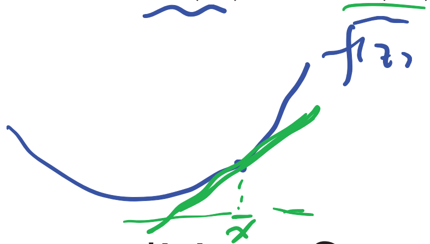
- $f : \mathcal{D} \to \mathbb{R}$ is called concave if $-f$ is convex

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

$$f(z) \geq \overbrace{f(x) + \nabla f(x)^T(z - x)}^{1^{st}\text{ order Taylor around } x}, \forall x, z \in \mathcal{D}$$



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

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

Hessian matrix $\in S_+^n$

- Many other conditions, tricks,... see [BV04].

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 $Q \succeq 0$

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

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

$$\Leftarrow \begin{cases} \|x\|_1 = |x_1| + |x_2| + \cdots |x_n| \\ \|x\|_\infty = \max_k |x_k| \end{cases}$$

$$\left(\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2} \right)$$

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

$\Downarrow$

$A \in \mathbb{R}$

$$\|A\|_p = \max \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p = 1} \|Ax\|_p$$

Outline

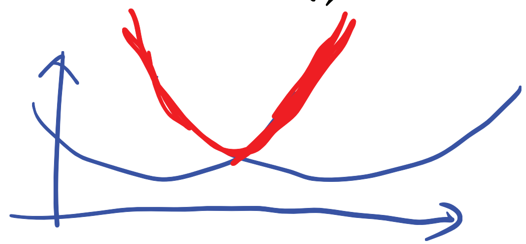
- Motivation
- Some Linear Algebra
- Some Multivariable Calculus
- Sets and Functions
- **Short Introduction to Optimization**
- Linear Program
- Quadratic Program
- Some Examples

– Affine mapping of convex function is still convex

e.g.: suppose $f(x)$ convex. $\Rightarrow$ $\underbrace{Af(x)+b}$ is also convex

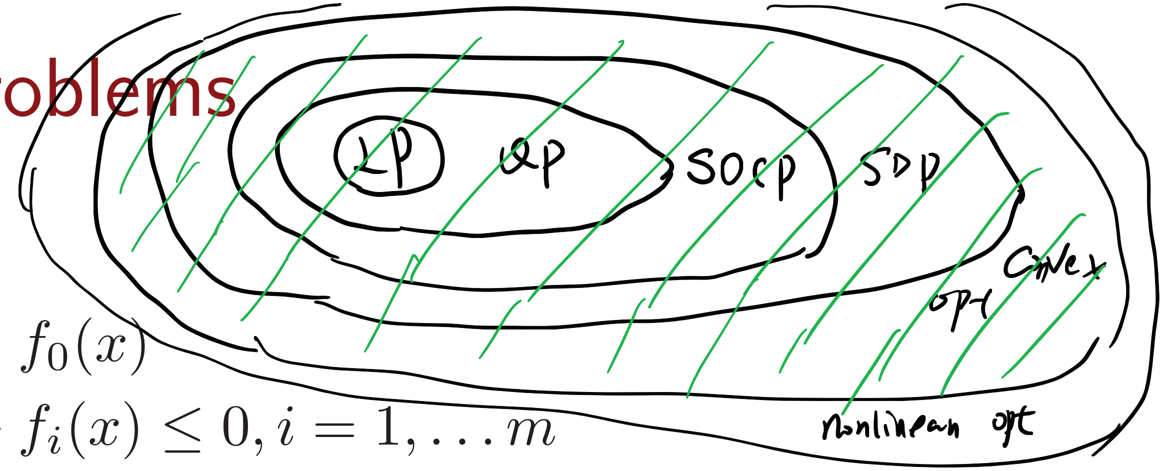
– Pointwise maximum of convex func is convex

$f_1(x)$, $f_2(x)$ convex $\Rightarrow f(x) = \underbrace{\max\{f_1(x), f_2(x)\}}_{\text{convex}}$

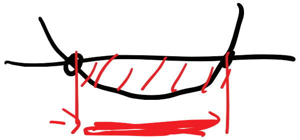


– Pointwise minimum of concave is concave

Nonlinear Optimization Problems



Nonlinear Optimization:



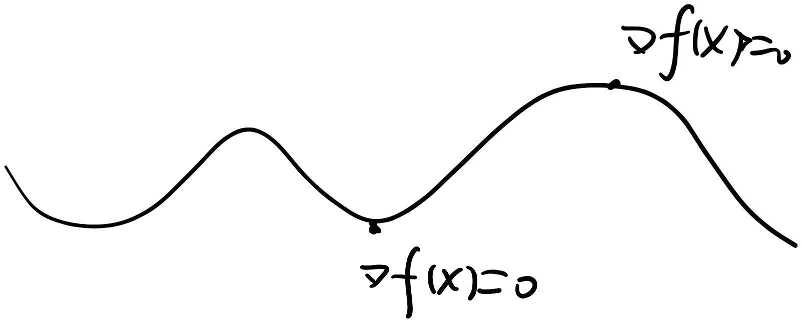
$$\begin{cases} \text{minimize:} & f_0(x) \\ \text{subject to:} & \begin{cases} f_i(x) \leq 0, i = 1, \ldots m \\ h_i(x) = 0, i = 1, \ldots, q \end{cases} \end{cases}$$

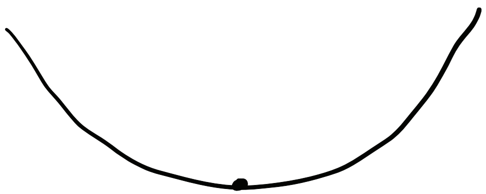
- decision variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, referred to as *primal problem*
- optimal value p^* , optimizer x^* $p^* = f_0(x^*)$
- is called a convex optimization problem if $f_0, \ldots, f_m$ are convex and $h_1, \ldots, h_q$ are affine
- typically convex optimization can be solved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

solve optimization: characterize optimality condition (i.e. condition optimizers need to satisfy) unconstrained optimization: $\min_{x} f(x)$, $f: \mathbb{R}^n \to \mathbb{R}$

optimality condition: $\nabla f(x) = 0$

$$f(x) = f(x^*) + \nabla f(x)^T (x - x^*) + \text{H.O.T.}$$





$$f(x,z) = xz + b$$

$$g(z) = \max_{x \in [1,2]} f(x,z)$$

$g(z)$ is concave

eg: $f_1(z) = z + b$

$f_2(z) = 2z + b$

$f_3(z) = 3z + b$

$\max\{f_1(z), f_2(z), f_3(z)\}$

Lagrangian

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

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

$\lambda_i \geqslant 0$

- 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: $g : \mathbb{R}^m \times \mathbb{R}^q :\to \mathbb{R}$

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

$g(\lambda,\nu)$ is pointwise minimum of $L(x,\lambda,\nu)$, $x = 1, 2$

- g is concave, can be $-\infty$ for some λ, ν $\quad g(\lambda,\nu) = \min\{L(1,\lambda,\nu), L(2,\lambda,\nu)\}$

affine in (λ,ν)

- **Lower bound property**: If $\lambda \succeq 0$ (elementwise), then $g(\lambda, \nu) \leq p^*$

$\lambda_1 \geq 0, \lambda_2 \geq 0, \cdots \lambda_m \geq 0$

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

$\Downarrow$

$f_i(\tilde{x}) \leq 0, \quad h_i(\tilde{x}) = 0$

$$\Rightarrow f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_x L(x, \lambda, \nu) = g(\lambda, \nu)$$

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

Lagrange Dual Problems (2/2)

Lagrange Dual Problem:

optimization problem (Dual) =

$$\begin{cases} \text{maximize :} & g(\lambda, \nu) \\ \text{subject to:} & \lambda \succeq 0 \end{cases} \iff \underbrace{\begin{cases} \min & \overbrace{(-g(\lambda,\nu))}^{\text{convex}} \\ \text{subj to:} & -\lambda \le 0 \end{cases}}_{\text{this is always a convex problem}}$$

- 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^*

optimizers: λ^*, ν^*

$d^* = g(\lambda^*, \nu^*)$

- (λ, ν) is called **dual feasible** if $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}(g)$
- Often simplified by making the implicit constraint $(\lambda, \nu) \in \mathbf{dom}(g)$ 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. $\exists \hat{x}$ such that $f_i(\hat{x}) < 0$, $h_i(\hat{x}) = 0$

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 (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,

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

- Therefore, all inequalities are actually equalities
- Replacing the first inequality with equality $\Rightarrow x^* = \underline{\text{argmin}_x L(x, \lambda^*, \nu^*)}$
- Replacing the second inequality with equality $\Rightarrow$ complementarity condition

General Optimality Conditions (3/3)

$\min_x L(x, \lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$

Proof of Sufficiency

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

$$d^* = g(\lambda^*, \nu^*) = f(x^*) + \underbrace{\sum_i \lambda_i^* f_i(x^*)}_{=0 \text{ (complementarity)}} + \underbrace{\sum_j \nu_j^* h_j(x^*)}_{=0 \text{ primal feasibility}}$$

$$= f(x^*) = p^*$$

- 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

- $\frac{\partial L}{\partial x}(x, \lambda^*, \nu^*)\Big|_{x=x^*} = 0$ (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
- Some Multivariable Calculus
- Sets and Functions
- Short Introduction to Optimization
- Linear Program
- Quadratic Program
- Some Examples

Linear Program: Primal and Dual Formulations

primal:
n-variable
m-equality
n-inequality

- **Primal Formulation:** $\begin{cases} \text{minimize:} & c^T x \\ \text{subject to:} & Ax = b \leftarrow \\ -x \le 0 & x \ge 0 \leftarrow \end{cases}$ $x \in \mathbb{R}^n$ $x_1 \ge 0, \dots x_n \ge 0$

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

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

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

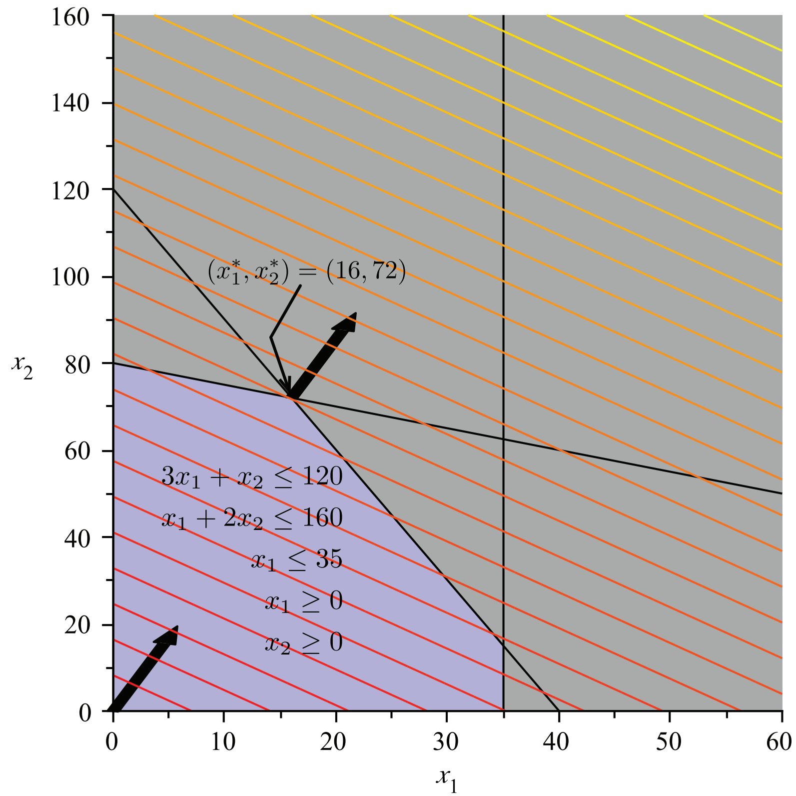
- **Its Dual:** $\begin{cases} \text{maximize:} & -b^T \nu \leftarrow m\text{-variable} \\ \text{subject to:} & A^T \nu + c \ge 0 \end{cases}$ $\max_{\lambda \ge 0} g(\lambda, \nu)$

n-inequality $\Rightarrow$ $\max -b^T \nu$

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

Linear Program: Example

A toy company produces toy planes and toy boats. Price: \$10 per plane and \$8 per boat. Cost: \$3 in raw materials per plane and \$2 per boat. A plane requires 3 hours to make and 1 hour to finish while a boat requires 1 hour to make and 2 hours to finish. The company cannot sell anymore than 35 planes per week. Further, given the number of workers, the company cannot spend anymore than 160 hours per week finishing toys and 120 hours per week making toys. How much of each toy it needs to produce?



Outline

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

Unconstrained Quadratic Program: Least Squares

- minimize: $J(x) = \frac{1}{2}x^T Q x + q^T x + q_0$
- Problem is convex iff $Q \succeq 0$

$Q = Q^T$

- When J is convex, it can be written as: $J(x) = \|Q^{\frac{1}{2}} x - y\|^2 + c$

- KKT condition:

$\nabla J(x) = Qx + q = 0$ (KKT condition)

$\Rightarrow x^* = -Q^{-1} q$

- Optimal solution:

Least square problem: $\min_x \|Hx - y\|^2$

$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$= \min \left(x^T H^T H x - 2 y^T H^T x + y^T y \right)$

$\Rightarrow \hat{x}_{LS} = (H^T H)^{-1} H y$

Equality Constrained Quadratic Program

- Standard form: $\begin{cases} \min_x & J(x) = \frac{1}{2} x^T Q x + q^T x + q_0 \\ \text{subject to:} & Hx = h \end{cases}$
- The problem is convex if $Q \succeq 0$
- KKT Condition: $L(x,\nu) = \frac{1}{2} x^T Q x + q^T x + q_0 + \nu^T (Hx - h)$

$$\frac{\partial L}{\partial x} = 0 \Rightarrow \begin{cases} Qx + q + H^T \nu = 0 \\ Hx = h \end{cases} \Rightarrow \begin{bmatrix} Q & H^T \\ H & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} -q \\ h \end{bmatrix}$$

- Optimal Solution:

$$\Rightarrow \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} Q & H^T \\ H & 0 \end{bmatrix}^{-1} \begin{bmatrix} -q \\ h \end{bmatrix}$$

General Quadratic Program

primal :
n - variables
m - inequality constraints.

- Standard form: $\begin{cases} \text{minimize:} & J(x) = \frac{1}{2} x^T Q x + q^T x + q_0 \\ \text{subject to:} & Ax \leq b \end{cases}$ $A \in \mathbb{R}^{m \times n}$
- Dual problem:

It's dual problem is given by

Dual :
$$\boxed{\max_{\lambda} \left(\frac{1}{2} \lambda^T D \lambda - e^T \lambda \right) - \frac{1}{2} q^T Q^{-1} q \qquad (\lambda \geq 0), \quad \lambda \in \mathbb{R}^m}$$

where : $e = b + A Q^{-1} q$

$D = -A Q^{-1} A$

primal : n- variables

m - inequa constraints

Dual : m- variables

m : nonnegative constraints

Outline

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

Example I: Identification of Robot Dynamics (1/2)

Example I: Identification of Robot Dynamics (2/2)

-

Example I: Identification of Robot Dynamics (3/2)

-
-

Example II: Differential IK (1/2)

- Forward Kinematics: $\underbrace{(R, p)}_{} = T(\theta)$
  - Relate joint angle to end-effector configuration
  - Sometimes we care only about certain part of the configuration

$$x = f(T(\theta))$$

, e.g. $x = R$, ← orientation . $x = (p_x, p_y)$

- Inverse Kinematics: Given desired x_d, find θ_d such that

$$x_d = f(T(\theta_d))$$

← solve for θ_d

- There may be multiple solutions:

- Reliable solver: IKFast in OpenRAVE

Example II: Differential IK (2/3)

- Differential kinematics / Diff Inverse kinematics
-
  - Differential kinematics : $\dot{x} = J_a(\theta)\dot{\theta}$
    - relate joint velocity $\dot{\theta}$ to end-effector velocity $\dot{x}$
    - $J_a(\theta)$: analytical Jacobian
    - $J_a(\theta) = E(\theta)\underbrace{J(\theta)}_{\in \mathbb{R}^{6\times n}}$ geometric Jacobian

      $[\underbrace{J_1(\theta)}_{\begin{bmatrix} w \\ v \end{bmatrix}} \quad J_2(\theta) \quad \cdots]$

  - Differential IK: given $\dot{x}_d$ find $\dot{\theta}_d$ such that

    $$\underbrace{\dot{x}_d}_{\in \mathbb{R}^{n_x}} \approx J_a(\theta)\,\dot{\theta}_d \qquad \dot{\theta}_d \in \mathbb{R}^n \rightarrow \text{\# of joints}$$

- when $J_a(\theta)$ is square and invertable then $\dot{\theta}_d = J_a^{-1}(\theta)\, \dot{x}_a$

- often times $J_a(\theta)$ is not invertable
  - $\boxed{n_x < n}$

$$J_a(\theta) = \underbrace{\begin{bmatrix} | & | & | & | & | & | \end{bmatrix}}_{n\text{-columns}} \Big\} \, n_x \text{ rows}$$

- often times: $J_a(\theta)$ is singular at least at some θ
  - $J_a(\theta)$ is not full rank

    $\text{rank}(J_a(\theta)) < n_x$

e.g.:

$$J_a(\theta) = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}}_{\text{full rank}}$$

$$J_a(\theta) = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 8 & 12 \end{bmatrix}}$$

$$\text{col}(J_a(\theta)) = \text{col}\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}\right)$$

- Singularity:
  - representation singularity: $E(\theta)$ rank deficient
  - kinematic singularity: $J(\theta)$ is rank deficient

- $\dot{\theta}_d^* = \cancel{\arg} \Leftarrow \min_{\dot{\theta}} \| J_a(\theta)\dot{\theta} - \dot{x}_d \|^2$ given

---

- Regardless of $\underbrace{\text{redundancy}}_{n_x < n}$ or singularity of J_a

a reliable way to solve diff IK is through optimization

$$\min_{\dot{\theta}} \| J_a(\theta)\,\dot{\theta} - \dot{x}_d \|^2 , \quad \text{decision variable is } \dot{\theta} \in \mathbb{R}^n$$

$\dot{x}_d \in \mathbb{R}^{n_x}$ $\quad \boxed{J_a(\theta)} \in \mathbb{R}^{n_x \times n}$

- This is a convex QP that can be solved efficiently

- The solution is given by $\dot{\theta}_{d,LS} = J_a^{+}(\theta)\,\dot{x}_d$ pseudo-inverse

If J_a is full row rank

$$J_a^{+}(\theta) = J_a^T (J_a J_a^T)^{-1} \Leftarrow \text{right inverse}$$

Example III: Point Cloud Registration (1/3)

$$J_a(\theta)\cdot \underbrace{J_a^{+}(\theta)} = I \iff J_a \cdot J_a^T (\underbrace{J_a J_a^T})^{-1} = I$$

- If J_a full row rank: $\underset{3\times 6}{J_a}\, \underset{6\times 1}{\dot\theta} - \underset{\in \mathbb{R}^3}{\dot x_d} = 0$ has infinite solutions

$\Downarrow$

$\text{rank}(J_a) = 3$

$\dot\theta_{ls} = J_a^{+}\dot x_d$ is the minimum norm s/n to

If $\text{rank}(J_a) = 2$, $\underbrace{J_a \dot\theta} - \dot x_d = 0$

linear combination of col(J_a)

if $\dot x_d \in$ col(J_a), has solution

$\dot x_d \notin$ col(J_a). no solution

Example III: Point Cloud Registration (2/3)

- – Differential IK with constraints:

$$\min_{\dot{\theta}} \|J_a(\theta)\dot{\theta} - \dot{x}_d\|^2 + \rho \|\theta + \Delta t\dot{\theta} - \theta_d\|^2$$

$$\begin{cases} \dot{\theta}_{min} \leq \dot{\theta} \leq \dot{\theta}_{max} \\ \theta_{min} \leq \theta + \dot{\theta} \cdot \Delta t \leq \theta_{max} \end{cases}$$

given / known at current time

See Russ Tedrake's course on robot manipulation

Example III: Point Cloud Registration (3/1)

-
-

References

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

More Discussions

•

More Discussions

-