Optimal Control

MPC Basics: Numerical Optimization Sequential Quadratic Programming

Erchao Rong

Sun Yat-sun University

2024-04-19

1. Reference

Reference

Numerical Optimization in Robotics by Wang et al.

Convex Optimization by Boyd et al.

Convex Optimization by Ryan et al.

https://www.stat.cmu.edu/~ryantibs/convexopt/

2. Convex optimization

Problem formulation

Convex optimization

convex optimization problem:

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

- variable $x \in \mathbf{R}^n$
- equality constraints are linear
- $f_0, \ldots, f_m$ are **convex**: for $\theta \in [0, 1]$,

$$f_i(\theta x + (1 - \theta)y) \leq \theta f_i(x) + (1 - \theta) f_i(y)$$

i.e., f_i have nonnegative (upward) curvature

Convex set

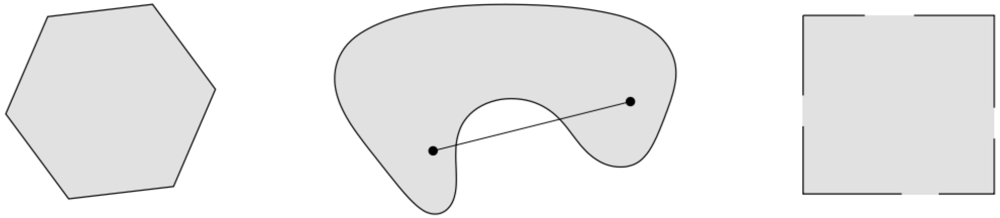
Convex set

line segment between x_1 and x_2: all points of form $x = \theta x_1 + (1-\theta)x_2$, with $0 \le \theta \le 1$

convex set: contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \le \theta \le 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C$$

examples (one convex, two nonconvex sets)



Convex Optimization Boyd and Vandenberghe 2.3

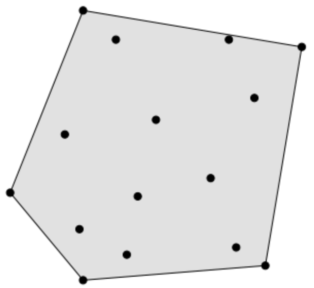
Convex combination and convex hull

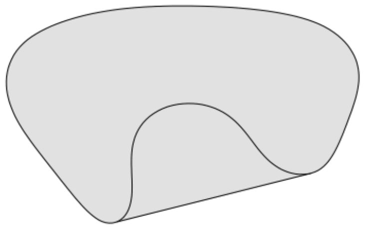
convex combination of $x_1, \ldots, x_k$: any point x of the form

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

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

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

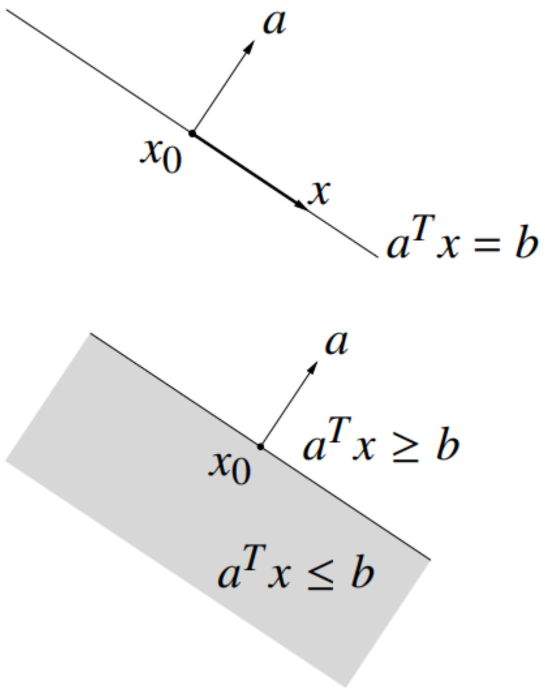




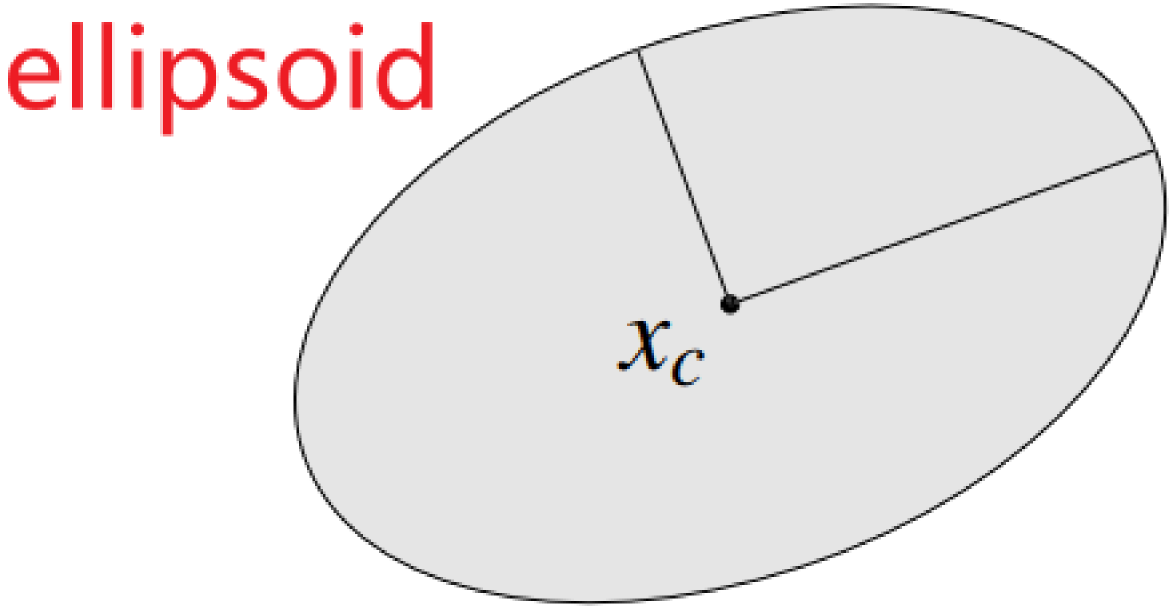
Convex Optimization Boyd and Vandenberghe

Convex set

Hyperplanes&halfspaces



Convex set



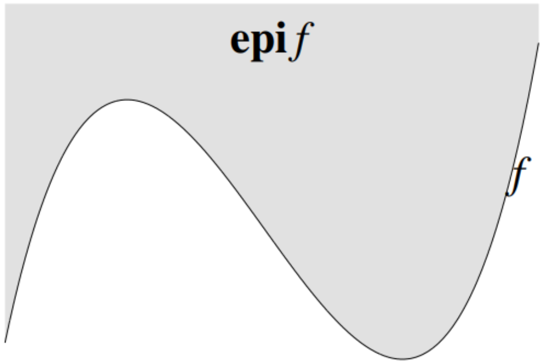
Operations perserves convexity

1. 😀 **Intersection**
   the intersection of (any number of) convex sets is convex
2. 😀 **Affine mappings**
   if $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine: $f(x) = Ax + b$, $A \in \mathbb{R}^{\{m \times n\}}$ and $b \in \mathbb{R}^{\{m\}}$
3. 😀 **scaling, translation**
4. 😀 **projection onto some coordinates**: $\{x | (x, y) \in S\}$
5. 😀 **solution set of linear matrix inequality**:
   $\{x | x_1 A_1 + \ldots + x_n A_n + B < 0\}$ with $A_i \in S^p$ and $B \in S^p$
6. 😀 **persepctive function** $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $P(x, t) = \frac{x}{t}$, $t > 0$
7. 😀 **linear-fractional function** $f : \mathbb{R}^n \to \mathbb{R}^m$:
   $f(x) = \frac{Ax+b}{c^T x+d}$, $c^T x + d > 0$

Convex function

Epigraph and sublevel set

- α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$ is $C_\alpha = \{x \in \mathbf{dom} f \mid f(x) \leq \alpha\}$
- sublevel sets of convex functions are convex sets (but converse is false)
- **epigraph** of $f : \mathbf{R}^n \to \mathbf{R}$ is $\mathbf{epi} f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom} f,\ f(x) \leq t\}$



- f is convex if and only if $\mathbf{epi} f$ is a convex set

Operations perserves convexity

Showing a function is convex

methods for establishing convexity of a function f

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
   - recommended only for **very simple** functions
3. show that f is obtained from simple convex functions by operations that preserve convexity
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective

you'll mostly use methods 2 and 3

Convex Optimization Boyd and Vandenberghe 3.18

Proper cones

a convex cone $K \subseteq \mathbf{R}^n$ is a **proper cone** if

- K is closed (contains its boundary)
- K is solid (has nonempty interior)
- K is pointed (contains no line)

examples

- nonnegative orthant $K = \mathbf{R}^n_+ = \{x \in \mathbf{R}^n \mid x_i \geq 0, \ i = 1, \ldots, n\}$
- positive semidefinite cone $K = \mathbf{S}^n_+$
- nonnegative polynomials on $[0, 1]$:

$$K = \{x \in \mathbf{R}^n \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0, 1]\}$$

Proper cone and Generalized inequality

Generalized inequality

- (nonstrict and strict) **generalized inequality** defined by a proper cone K:

$$x \preceq_K y \quad \Longleftrightarrow \quad y - x \in K, \qquad x \prec_K y \quad \Longleftrightarrow \quad y - x \in \textbf{int}\, K$$

- **examples**
  - componentwise inequality ($K = \mathbf{R}^n_+$): $x \preceq_{\mathbf{R}^n_+} y \iff x_i \le y_i, \quad i = 1, \dots, n$
  - matrix inequality ($K = \mathbf{S}^n_+$): $X \preceq_{\mathbf{S}^n_+} Y \iff Y - X$ positive semidefinite

  these two types are so common that we drop the subscript in $\preceq_K$

- many properties of $\preceq_K$ are similar to $\le$ on $\mathbf{R}$, *e.g.*,

$$x \preceq_K y, \quad u \preceq_K v \quad \Longrightarrow \quad x + u \preceq_K y + v$$

3. Unconstrained Optimization

Moving strategy

😀 **Given** current point x_k and a direction p_k , how we move to the new iterate x_{k+1} .

1. 😀 **Line Search**
   - Exact
   - Inexact
     - Backtracking line serach (Armijo rule)

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k p_k^T \nabla f(x_k)$$

     - Wolfe conditions (Curvature condition) $[0 < c_1 < c_2 < 1]$
       1. Weak wolfe condition: $-p_k^T \nabla f(x_k + \alpha_k p_k) \leq -c_2 p_k^T \nabla f(x_k)$
       2. Strong wolfe condition:
          $\left|p_k^T \nabla f(x_k + \alpha_k p_k)\right| \leq c_2 \left|p_k^T \nabla f(x_k)\right|$

Moving strategy

2. 😁 **Trust Region**

the information about f is used to construct a model m_k whose behavior near the current point x_k is similar to that of f.

$\min_p m_k(x_k + p)$ where $x_k + p$ lies inside the trust region. Typically, $m_k(x_k + p) = f_k p^T \nabla f_k + \frac{1}{2} p^T B_k p$.

First Order Method

1. 😁 **Gradient descent:**
   $x^k = x^{k-1} - \alpha \nabla f\left(x^{k-1}\right)$
2. 😁 **Proximal Gradient descent:**
   $x^+ = \arg\min_y f(x) + \nabla f(x)^T (y - x) + \frac{1}{2t} \|y - x\|_2^2$
3. Others

First Order Method

https://towardsdatascience.com/a-visual-explanation-of-gradient-descent-methods-momentum-adagrad-rmsprop-adam-f898b102325c

Motivation

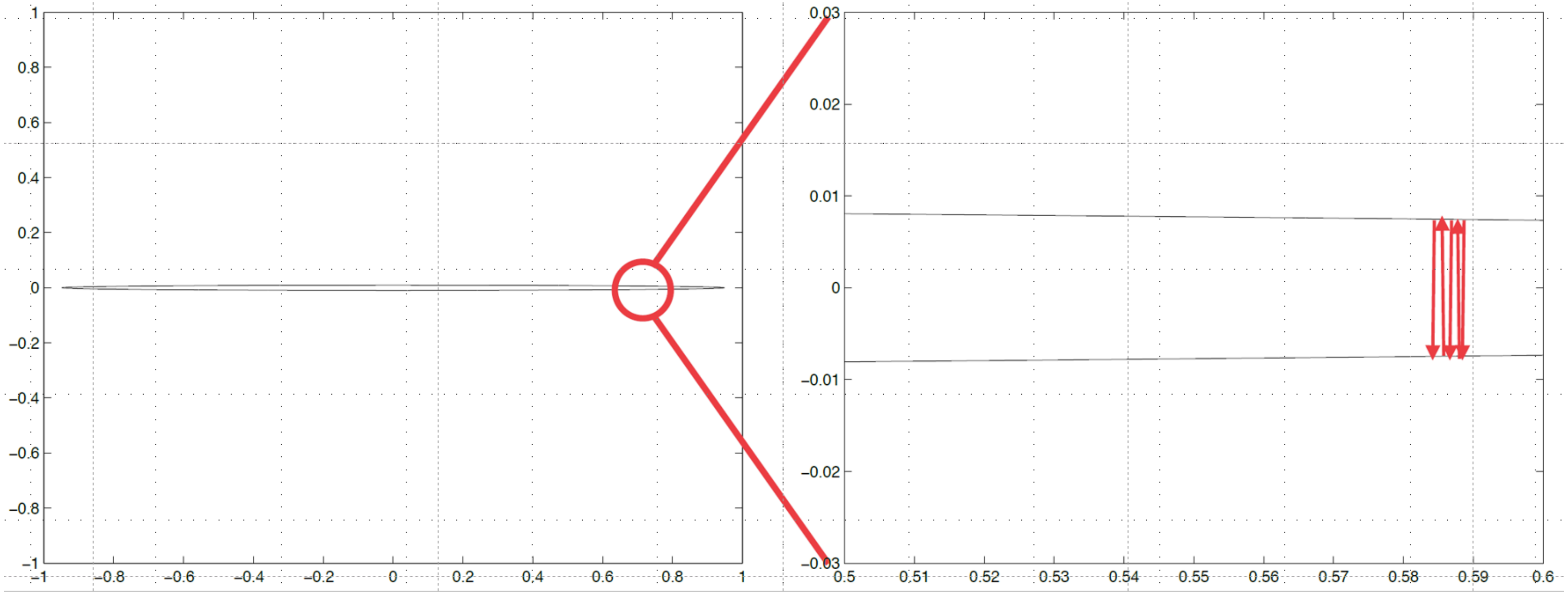
Steepest Gradient Descent

Curvature info is needed!

Drawbacks: Poor conditioning causes performance degeneration

$\kappa = 100$

Contours are (almost) parallel!



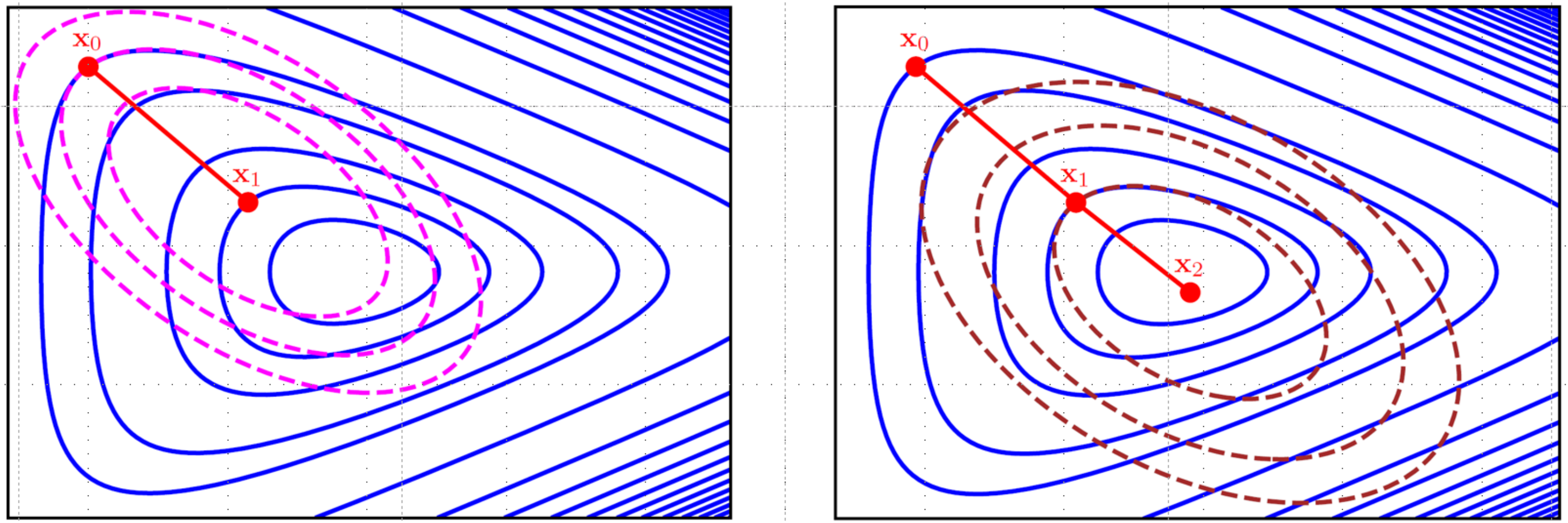
Motivation

Newton's Method

Example

The magenta curves are the level curves of the quadratic approximation at $\boldsymbol{x}_0$

The brown curves are the level curves of the quadratic approximation at $\boldsymbol{x}_1$



Second Order Method

1. 😀 **Newtion**
2. 😀 **BFGS**
3. 😀 **LBFGS**
4. 😀 **Gaussian-Newtion**