Continuous Optimization

Introduction

January 6

Examples of optimizations

- Financial portfolio

$$\begin{array}{ll} \text{min} & \textit{risk/reward ratio} \\ \text{subject to} & \\ & \text{risk tolerance} \\ & \text{time frame} \end{array}$$

- Nature system: the hanging chain

- Logistics
- Curve fitting

General formulation

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_j(x) = 0 \quad j \in E \\ & c_j(x) \leq 0 \quad j \in I. \end{aligned}$$

Some Notations and conventions

- Variable x, objective function f, constraint c_j.
- "Minimize" is preferred over "Maximize" for some historic reasons.

$$\min_{\mathcal{C}} f(x) = -\max_{\mathcal{C}} -f(x)$$

- Equality constraints can be written as inequality constraints

$$c_j(x) = 0 \qquad \Leftrightarrow \qquad \begin{cases} c_j(x) \leq 0 \\ c_j(x) \geq 0 \end{cases}$$

Course Outline

- Introduction
- Unconstrained optimization:
  - First-order and second-order necessary conditions
  - Line search methods for scalar functions
  - Conjugate gradient methods for quadratic functions
  - Newton (Quasi-Newton) methods
- Constrained optimization:
  - First-order and second-order necessary conditions
  - Lagrange Multiplier for equality constraints
  - KKT condition
  - Penalty, barrier and argumented Lagrangian methods
- Additional Topics:
  - Convex programming
  - Sequential quadratic programming
  - Applications

Review on calculus and linear algebra

- Calculate the gradient $\nabla f(x)$ and Hessian matrix $\nabla^2 f(x)$
- Directional derivative: $x, y \in \mathbb{R}^n$

$$\frac{d}{d\lambda} f(x + \lambda(y - x)) =$$

- Taylor expansion for functions of one variable and multiple variables (up to the quadratic order)
- Matrix (and vector) notations: find the gradient and Hessian matrix of $f(x) = \frac{1}{2} x^t A x - b^t x$.
- Different norms and their convexity:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$$

$$\|(1-\lambda)x + \lambda y\|_p \leq (1-\lambda)\|x\|_p + \lambda \|y\|_p, \ \lambda \in (0,1)$$

Classification of different problems

- Singular variable ($<$) multiple variables
- Linear Problem ($<$) Nonlinear Problem
- Unconstrained ($<$) Constrained
- Convex ($\ll$) Nonconvex

Here $A < B$ means that A is relatively easier to solve (analytically or numerically) than B.

Different algorithms work for different problems. The recognization of a particular class of problems may help use to choose the right algorithm to solve it.

Conversion to simpler problems

These classification are not unique, because some problems can be converted into simpler ones.

- Convert the absolute value $|\cdot|$ into linear ones. If there

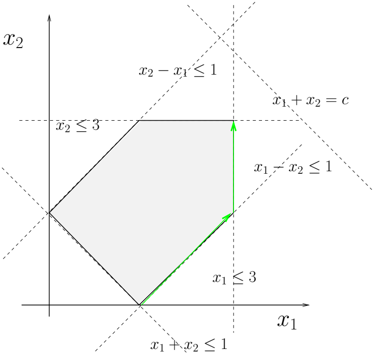
$$\begin{aligned} \min \quad & x_1 \\ \text{subject to} \quad & \\ & |x_1 - 1| + x_2 \le 4 \\ & x_1 - |x_2 - 1| \ge 0 \end{aligned}$$

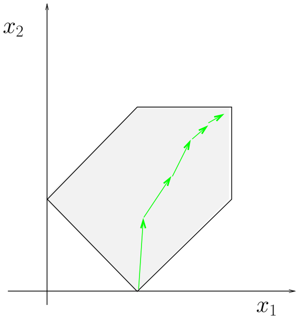
- Convert equality into convex inequality (if the extremer is obtained at that equality)

$$\begin{aligned} \min \quad & x_1 + x_2 + x_3 \\ \text{subject to} \quad & \\ & x_1^2 + x_2^2 + x_3^2 = 1 \end{aligned}$$

The problem can be converted into a convex one with $x_1^2 + x_2^2 + x_3^2 \le 1$.

Simplex method vs Continuous Optimization

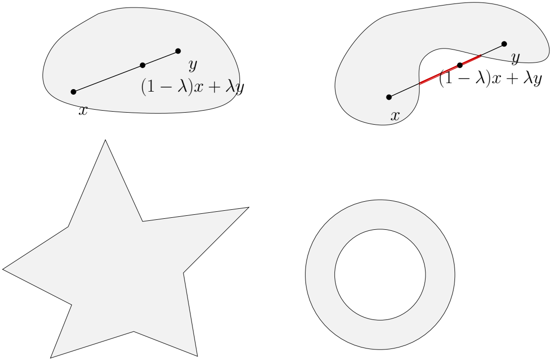




$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{subject to} & x_1 \leq 3,\ x_3 \leq 2,\ x_1 + x_2 \geq 1 \\ & x_1 - x_2 \leq 1, \quad x_2 - x_1 \leq 1. \end{array}$$

Convex set and convex functions

A set Ω is *convex* if for any $x, y \in \Omega$, the line segment $[x, y]$ is in Ω.



A function f is convex if

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

Characterization and properties of convex functions

A smooth function $f(x)$ is convex if and only if the Hessian matrix H is nonnegative definite.

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Properties:

- $f(y) \geq f(x) + (\nabla f(x), y - x)$
- ∇f is monotone, $(\nabla f(y) - \nabla f(x), y - x) \geq 0$

Graph Method: 1D

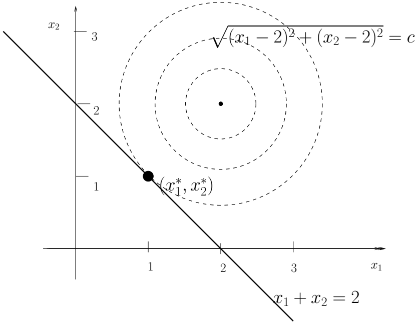
Find the minimizers of the following functions:

$$(1) \qquad f(x) = \max(|x|, |2x-3|)$$

$$(2) \qquad f(x) = |x| + |2x-3|$$

Graph Method: Equality constraint

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x_1, x_2) = \sqrt{(x_1 - 2)^2 + (x_2 - 2)^2} \\ \text{subject to} \quad & x_1 + x_2 = 2 \end{aligned}$$



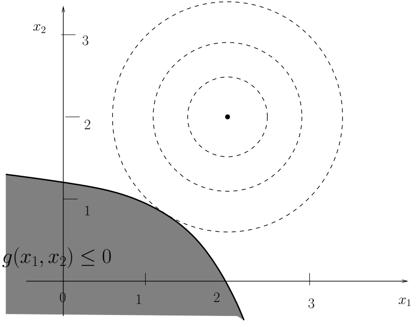
General Procedure:

(a) Plot the feasible region.

(b) Plot the contour lines of the objective function.

Graph Method: Inequality constraint

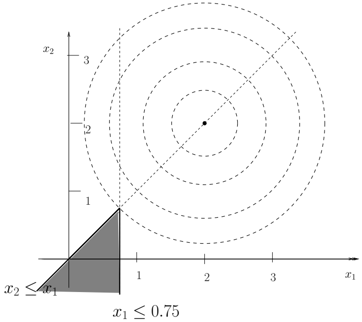
$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x_1, x_2) = \sqrt{(x_1 - 2)^2 + (x_2 - 2)^2} \\ \text{subject to} & g(x_1, x_2) \leq 0 \end{array}$$



What kind of special properties does the minimizer possess?

Graph Method: Optimality condition

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x_1, x_2) = \sqrt{(x_1 - 2)^2 + (x_2 - 2)^2} \\ \text{subject to} \quad & x_1 \le 0.75, \\ & x_2 \le x_1. \end{aligned}$$



Is the feasible region (shaded) “tangent” to the contour lines?

We are going to find these conditions later in this class.

Convergence of algorithms

The minimizer x^* of a problem is usually obtained iteratively, as the limit of $\{x_n\}$. There are some concepts associated with the rate of how fast x_n approaches x^*.

- Global convergence (x_1 can be any initial states) vs Local convergence (x_1 is restricted)
- Convergence rate:
  - Q-linear, Q-superlinear, Q-quadratic: $|x_{n+1} - x^*|/|x_n - x|^*$
  - R-convergence: $|x_n - x^*|^{1/n}$