Continuous Optimization

Unconstrained Optimization (part 1)

Sections covered in the textbook (2nd edition):

- Chapter 2: 1, 2
- Chapter 3: 1, 2, 3, 4
- Chapter 5: 1, 2
- Chapter 6: 1
- Chapter 10: 1, 2, 3

Existence questions

What's the problem with the following functions?

1. $f(x) = -1/|x|$
2. $f(x) = e^{-x^2}$
3. $$f(x) = \begin{cases} 1, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Existence questions for $\min_{x \in \mathbb{R}^n} f(x)$

1. f should be bounded below
2. The minimizer should be obtained for some finite x, for example

$$(1) \qquad \lim_{|x| \to \infty} f(x) = \infty$$

or

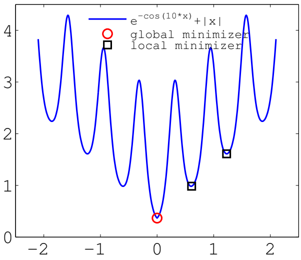
$$(2) \qquad \text{The set } \{x : f(x) < c\} \text{ for some } c \text{ is bounded.}$$

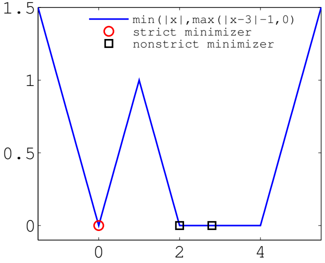
3. f is low-semicontinuous.

$$f(x^*) \le \lim_{x \to x^*} f(x)$$

Different types of minimizers

- Global minimizer vs Local minimizer
- Strict minimizer vs nonstrict minimizer





Usually, we can only find local minimizers. With more information (like convexity) we can show the minimizer we found is global.

If we only need the function value at the minimizer, there is no need to worry about whether it is strict or not.

Optimality conditions of local minimizer x^* for smooth functions f

Theorem (Necessary conditions)

if x^* is optimal, then

- *1st-order necessary condition (NC1):* $\nabla f(x^*) = 0$
- *2nd-order necessary condition (NC2): the Hessian* $\nabla^2 f(x^*)$ *is positive definite*

Theorem (Suffient condition (SC2))

if x^* is such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is posotive definite, then x^* is a local minimum.

Therefore, to find the minimizer of a function, we can just find all the solutions to the system $\mathbf{F}(x) = \nabla f(x) = 0$.

Can you write a minimization problem such whose minimizer x^* is the same as the solution of the system of equations $\mathbf{F}(x) = 0$?

Uniqueness of the minimizer: Convexity

Theorem

If f is convex, then any local minimizer x^* is a global minimizer.

Remarks:

- x^* may not be a strict minimizer: $f(x) = \min(|x| - 1, 0)$
- f does not need to be a smooth function: $f(x) = |x|$
- For convex function f, we can generalize the concept of gradient $\nabla f(x)$ to *subgradient* (a set) $\partial f(x)$ such that

$$f(y) \geq f(x) + (p, y - x) \qquad \text{for all } p \in \partial f(x).$$

Then x^* is a minimizer if and only if $\mathbf{0} \in \partial f(x)$.

What's $\partial f(x)$ for $f(x) = |x|$?

Some Examples of functions with one single variable

Example 1.

$$f(x) = x^4 + 4x^3 + 6x^2 + 4x$$

Example 2.

$$f(0) = 0, f'(x) = (x+1)x^2(x-1)^3$$

Example 3.

$$f(x) = x + \sin(x)$$

Example 4.

$$f(x) = |x - x_1| + |x - x_1| + \cdots + |x - x_m|, \qquad x_1 < x_2 < \cdots < x_m.$$

Example: Data Fitting

For a set of data $(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)$, find the parameter a_0 and a_1 for the linear fit $y = a_0 + a_1 x$.

The objective function

$$f(a_0, a_1) = \sum_{i=1}^{m} (y_i - a_0 - a_1 x_i)^2$$

How about fit with quadratic term $y = a_0 + a_1 x + a_2 x^2$ or higher order?

What's the difference compared with fitting the model $y = a_0 e^{a_1 x}$?

Example: Maximum Likelihood Estimation

Estimate (μ, σ) in the Gaussian distribution based on x_i.
The objective function is

$$L = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Estimate λ in the Poisson distribution based on t_i.
The objective function is

$$L = \prod_{i=1}^{m} \lambda e^{-\lambda t_i}$$

General features of an algorithm

$$\min_{x \in \mathbb{R}^n} f(x)$$

If we can not find the minimizers analytically, we have to use numerical techniques.

- Successive approximiations $x_1 \to x_2 \to \cdots \to x^*$
- Either requiring $f(x_1) \geq f(x_2) \geq \cdots \geq f(x^*)$ or $\nabla f(x_k) \to 0$.
- Stoping criteria (Is x_k optimal?): (i) $|f(x_{k+1}) - f(x_k)| < \epsilon$ or (ii) $|\nabla f(x_k)| < \epsilon$.
- Different algorithms find different x_{k+1} from x_k:
  - Linear Search: Steepest decent, conjugate gradient, Newton's method
  - Trust Region

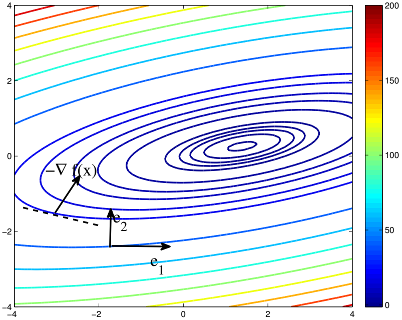
General problems for unconstrained optimization

- Multiple local minimizers
- Large scale (many variables)
- Complicated function evaluations

Line Search Method

$$x_{k+1} = x_k + \alpha_k p_k$$

Basic questions: how do we choose the direction p_k and the length α_k?

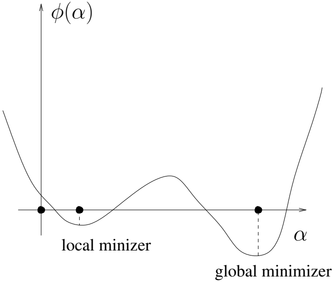


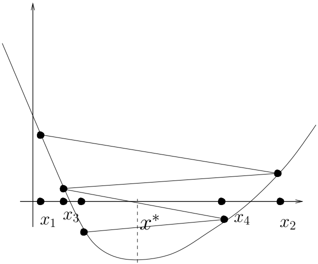
In general, it is NOT a good idea to choose p_k to be aligned with the coordinate axis e_i? We should choose a problem-dependent p_k, like $p_k = -\nabla f(x_k)$.

Step length α_k

Assuming p_k is chosen, then we want to find the (global) minimizer α_k for

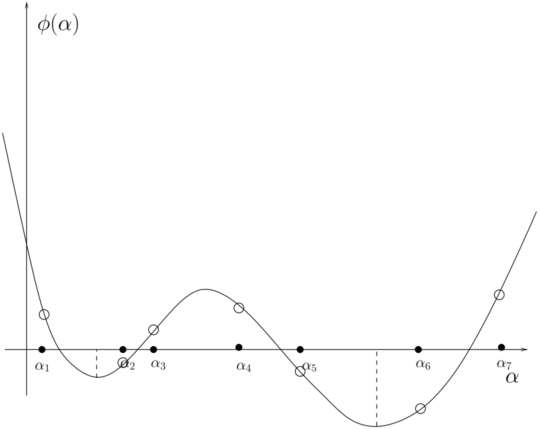
$$\phi(\alpha) = f(x_k + \alpha p_k), \quad \alpha > 0.$$





The search algorithm should allow step length α large enough to find the global minimizer, but also small enough to reduce the oscillations around the minimizer.

Optimal Steplength



- Locate the intervals with local minimizers
- Find the local minimizers and the global one