Numerical Optimization

Exercise I

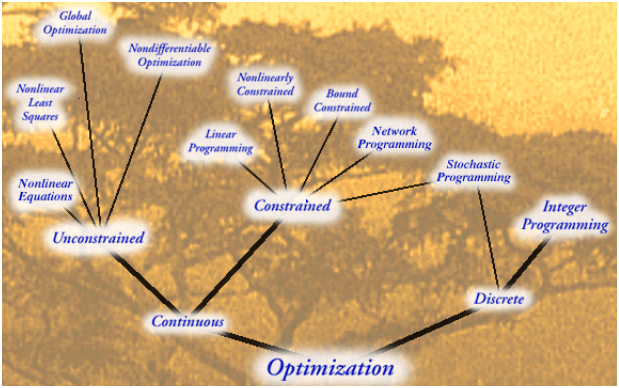
Jonathan Balzer

Geometric Modeling and Scientific Visualization Center
King Abdullah University of Science and Technology

Some terminology

- variables, objective function
- constrained optimization: (in)equality constraints, feasible set
- discrete vs. continuous optimization
- local vs. global optimization
- stochastic vs. deterministic optimization
- convex vs. nonconvex optimization
- strong, weak, isolated optima

The optimization tree



Source: `http://www-fp.mcs.anl.gov/otc/Guide/OptWeb/index.html`

Concepts of local optimization

Theorem (Taylor in 1D)

Given a function $f : \mathbb{R} \mapsto \mathbb{R}$. Suppose that in an open interval containing x_0, f is continuously differentiable $n+1$ times, then for each x in this interval

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_{n+1}(x),$$

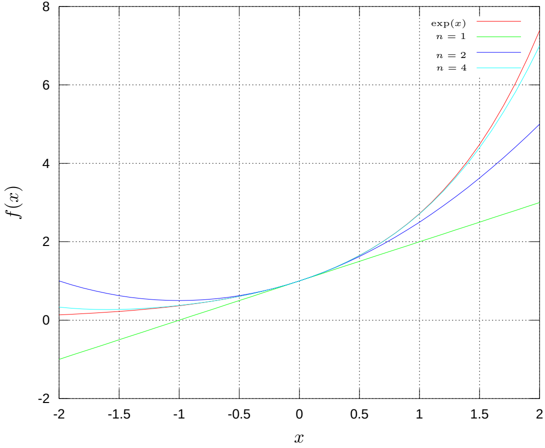
where the error term $R_{n+1}(x)$ satisfies

$$R_{n+1}(x) = \frac{f^{(k)}(\xi)}{k!}(x - x_0)^k$$

for some $\xi \in [x_0, x)$.

Concepts of local optimization

Example:



$$\exp(x) \approx 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{4}x^4 \text{ around } x_0 = 0$$

Concepts of local optimization

Basic idea:

Construct sequence $\{x_k\}$ of points in feasible set such that objective function $f(x_k)$ decreases (monotonically) as $k \to \infty$.

Two strategies:

- Line search
  - descent direction fixed (per iteration): negative gradient, Newton step, Quasi-Newton step, conjugate directions
  - distance variable $\Rightarrow$ one-dimensional subproblem
- Trust region
  - direction variable
  - maximal distance $=$ size of trust region fixed (per iteration)

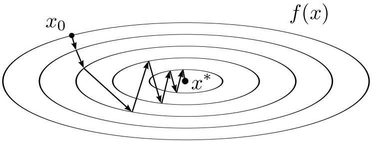
Conditions of optimality

Convention:

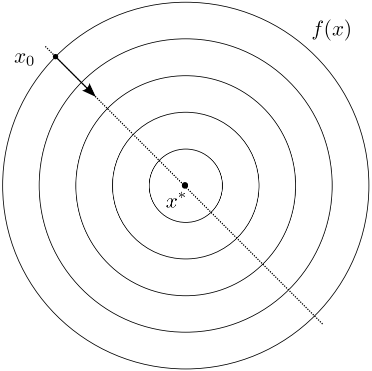
optimization = minimization

- necessary:
  - x^* stationary $\Leftrightarrow$ "$f'(x^*) = 0$"
  - Hessian positive semidefinite $\Leftrightarrow$ "$f''(x^*) \geq 0$"
- sufficient:
  - Hessian positive definite $\Leftrightarrow$ "$f''(x^*) > 0$" (x^* strong)
  - f and feasibility region convex (x^* global)

Scaling



poor



better

Rates of Convergence

- Q-convergence
  - sublinear, superlinear
  - quadratic
  - ...
- R-convergence

Exercises

1. Are the following matrices positive definite or positive semidefinite?

(i) $\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & -1 & 1 \end{pmatrix}$

(ii) $\mathbf{B} = \begin{pmatrix} 3 & 5 \\ 12 & 20 \end{pmatrix}$

(iii) $\mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Exercises

2. Compute gradient and Hessian of the following functions.
Identify stationary points and check whether these are local optima.

(i) $f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

(ii) $f(\boldsymbol{x}) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$

Exercises

3. Calculate the n-th order Taylor expansion of the function f around x_0.

(i) $f(x) = \cos x, \; n = 3, \; x_0 = 0$

(ii) $f(x) = \cos\left(\frac{1}{x}\right), \; n = 2$

Matrix Norms I

Definition

Let $\mathbb{K}$ be equal to $\mathbb{C}$ or $\mathbb{R}$. A norm $\|\cdot\|$ on $\mathbb{K}^{n\times n}$, $n\in\mathbb{N}$, is called *matrix norm* if it is submultiplicative, i.e. for two $\mathbf{A},\mathbf{B}\in\mathbb{K}^{n\times n}$, it holds

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|.$$

Definition

A matrix norm $\|\cdot\|$ is *consistent* with a vector norm $\|\cdot\|$ on $\mathbb{K}^n$ if

$$\|\mathbf{A}\| \leq \|\mathbf{A}\|\|\boldsymbol{x}\|$$

for every $\mathbf{A}\in\mathbb{K}^{n\times n}$ and $\boldsymbol{x}\in\mathbb{K}^n$.

Matrix Norms II

Examples

- Frobenius norm:

$$\|\mathbf{A}\|_F = \sqrt{\sum_i^n \sum_j^n |a_{ij}|^2} = \mathrm{tr}(\mathbf{A}^\top \mathbf{A})$$

- Induced norms (consistent!):

$$\|\mathbf{A}\|_p = \sup_{\|\boldsymbol{x}\|_p = 1} \|\mathbf{A}\boldsymbol{x}\|_p$$

$(p = 1, 2, \ldots, \infty)$

Condition of Matrix Inversion

Relative error:

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_p}{\|\boldsymbol{x}\|_p} \leq \kappa \frac{\|\boldsymbol{b} - \tilde{\boldsymbol{b}}\|_p}{\|\boldsymbol{b}\|_p}$$

- $\boldsymbol{b}, \tilde{\boldsymbol{b}}$: true, disturbed right-hand side
- $\boldsymbol{x}, \tilde{\boldsymbol{x}}$: true, erroneous solution
- $\kappa = \|\mathbf{A}\|_p \|\mathbf{A}^{-1}\|_p$: condition number

For $p = 2$:

$$\kappa = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}}$$

Exercises

4. Suppose that a function f of two variables is poorly scaled at the solution $\boldsymbol{x}^*$. Write the Taylor expansion of f around $\boldsymbol{x}^*$ and use it to show that the Hessian $\nabla^2 f$ is ill-conditioned.

Exercises

5. What can you say about the convergence rates of the following sequences?

(i) $x_k = \frac{1}{k}$

(ii) $x_k = 1 + \left(\frac{1}{2}\right)^{2^k}$

(iii) $x_k = \frac{1}{k!}$

(iv) $x_k = \begin{cases} \left(\frac{1}{4}\right)^{2^k}, & k \text{ even}, \\ x_{k-1}/k, & k \text{ odd}. \end{cases}$

Exercises

6. Suppose that f is a convex function. Show that the set of global minimizers of f is a convex set.