Optimization Methods
Lecture 6

Solmaz S. Kia
Mechanical and Aerospace Engineering Dept.
University of California Irvine
solmaz@uci.edu

Reading assignment: Ch 9 of Ref[2]

Numerical solvers for unconstrained optimization

Unconstrained optimization:

$$x^\star = \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x)$$

Iterative solution method $x_{k+1} = x_k + \alpha_k \, d_k$

Observations:

- Steepest descent algorithm can be very slow with lots of zig-zaging
- Newton method is faster bit numerically is expensive due to information equipment associated with the evaluation, storage and inversion of Hessian.

%pause Q: Is it possible to accelerate convergence with low numerical cost?

Conjugate direction method

- Conjugate direction methods proposed for

$$x^\star = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} x^\top Q x - b^\top x, \qquad Q > 0$$

- Conjugate direction methodsconverge in n steps
- Can be extended to solve other nonlinear optimization problems

Defintion:

- Two vectors $d_1 \in \mathbb{R}^n$ and $d_2 \in \mathbb{R}^n$ are orthogonal if and only if $d_1^\top d_2 = 0$.
- Given a symmetric matrix $Q \in \mathbb{R}^{n \times n}$, two vectors d_1 and d_2 are said to be Q-orthogonal or conjugate with respect to Q if $d_1^\top Q d_2 = 0$.
- A finite set of vectors $\{d_1, d_2, \cdots, d_k\}$ is said to be Q-orthogonal or conjugate with respect to Q, if

$$d_i^\top Q d_j = 0, \quad \forall i, j, \; i \neq j.$$

Preliminaries

- Linearly independent vectors: A set of vectors $\{d_1, d_2, \cdots, d_k\}$ is linearly independent if the relation

$$\alpha_1 d_1 + \alpha_2 d_2 + \cdots + \alpha_k d_k = 0$$

  implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

- Lemma If $\{d_1, d_2, \cdots, d_k\}$ is linearly independt set of vectors and $x \in \text{span}\{d_1, d_2, \cdots, d_k\}$, the relation

$$x = \alpha_1 d_1 + \alpha_2 d_2 + \cdots + \alpha_k d_k$$

  is unique.

- Lemma If $\{d_1, d_2, \cdots, d_n\}$ is a set of n linearly independent vectors in $\mathbb{R}^n$, then every $x \in \mathbb{R}^n$ can be written as linear combination of $\{d_1, d_2, \cdots, d_n\}$, i.e.,

$$x = \alpha_1 d_1 + \alpha_2 d_2 + \cdots + \alpha_n d_n$$

Properties of conjugate vectors

Proposition If $Q > 0$, and the set of non-zero vectors $\{d_0, \cdots, d_k\}$ are Q-orthogonal then these vectors are linearly independent.

Proof Suppose $\exists \alpha_i$, $i = 0, 1, \cdots, k$ such that

$$\alpha_0 d_0 + \alpha_1 d_1 + \cdots + \alpha_k d_k = 0$$

Multiply by Q and taking the scalar product with d_i:

$$\alpha_i d_i^\top Q d_i = 0$$

Since $d_i \neq 0$ and $Q > 0$, we obtain $\alpha_i = 0$. Therefore, $\{d_0, \cdots, d_k\}$ are linearly independent.

Conjugate gradient algorithm (preliminary version)

Given x_0
set $g_0 \leftarrow Qx_0 - b, \quad d_0 \leftarrow -g_0, \quad k \leftarrow 0$
while $g_k \neq 0$

$$
\begin{aligned}
\alpha_k &\leftarrow -\frac{g_k^\top d_k}{d_k^\top Q d_k}, \\
x_{k+1} &\leftarrow x_k + \alpha_k\, d_k, \\
g_{k+1} &\leftarrow Q\, x_{k+1} - b, \\
\beta_{k+1} &\leftarrow \frac{g_{k+1}^\top Q\, d_k}{d_k^\top Q d_k}, \\
d_{k+1} &\leftarrow -g_{k+1} + \beta_{k+1}\, d_k, \\
k &\leftarrow k + 1;
\end{aligned}
$$

end(while)

Conjugate gradient algorithm

Given x_0
set $g_0 \leftarrow Qx_0 - b, \quad d_0 \leftarrow -g_0, \quad k \leftarrow 0$
while $g_k \neq 0$

$$
\begin{aligned}
\alpha_k &\leftarrow \frac{g_k^\top g_k}{d_k^\top Q d_k}, \\
x_{k+1} &\leftarrow x_k + \alpha_k d_k, \\
g_{k+1} &\leftarrow g_k + \alpha_k Q d_k, \\
\beta_{k+1} &\leftarrow \frac{g_{k+1}^\top g_{k+1}}{g_k^\top g_k}, \\
d_{k+1} &\leftarrow -g_{k+1} + \beta_{k+1} d_k, \\
k &\leftarrow k+1;
\end{aligned}
$$

end(while)

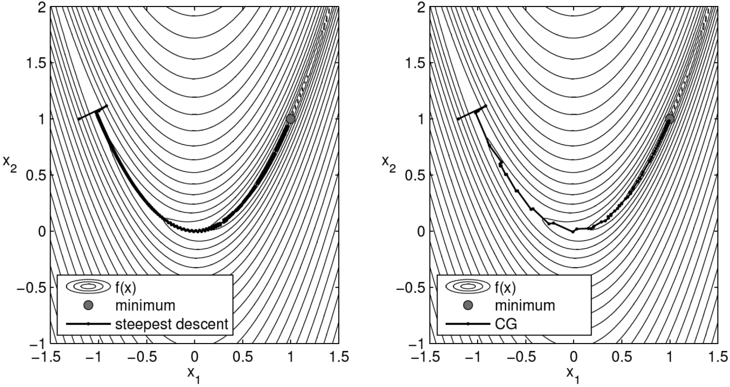
Note:
$g_{k+1} = Q\,x_{k+1} - b = Q(x_k + \alpha_k d_k) - b = Qx_k - b + Q\alpha_k d_k = g_k + \alpha_k Q\, d_k.$

Numerical example

Minimize Rosenbrock's function,

$$f(x) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2,$$

starting from $x_0 = (-1.2, 1.0)^T$.



Solution path of the steepest descent and conjugate gradient methods