Optimization Methods
Lecture 4

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

Consult: pages 29-33, 41-43, 62-67 from Ref[1];
Section 8.5 and 8.6 from Ref[2]

Common choices of the stepsize

$$x_{k+1} = x_k - \alpha_k B_k \nabla f(x_k), \quad B_k > 0$$

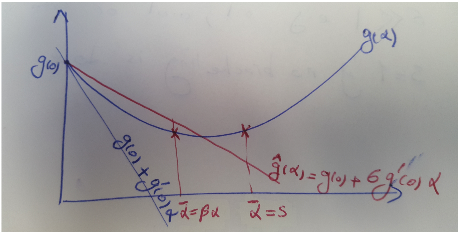
- **Exact line search**: $\alpha_k = \underset{\alpha \geqslant 0}{\operatorname{argmin}} f(x_k + \alpha d_k)$
    - A minimization problem itself, but an easier one (one dimensional).
    - If f convex, the one dimensional minimization problem also convex (why?).
- **Limited minimization**: $\alpha_k = \underset{\alpha \in [0,s]}{\operatorname{argmin}} f(x_k + \alpha d_k)$

  (tries not to stop too far)
- **Constant stepsize**: $\alpha_k = s > 0$ for all k
  (simple rule but may not converge if it is too large or may converge too slow because it is too small)
- **Diminishing step size**: $\alpha_k \to 0$, and $\sum_{k=1}^{\infty} \alpha_k = \infty$. For example $\alpha_k = \frac{1}{k}$
    - Descent not guaranteed at each step; only later when becomes small.
    - $\sum_{k=1}^{\infty} \alpha_k = \infty$ imposed to guarantee progress does not become too slow.
    - Good theoretical guarantees, but unless the right sequence is chosen, can also be a slow method.
- **Successive step size reduction**: well-known examples are Armijo rule (also called Backtracking) and Goldstein rule
  (search but not minimization)

Stepsize selection via successive reduction: Armijo rule

- It is an *inexact line search method*: it does not find the exact minimum but guarantees sufficient decrease
- computationally is cheap
- Armijo parameters: $\beta \in (0, 1)$ and $\sigma \in (0, 1)$

Recall: $g(0) = f(x_k)$, $g'(0) = \nabla f(x_k)^\top d_k < 0$ (d_k is a descent direction)



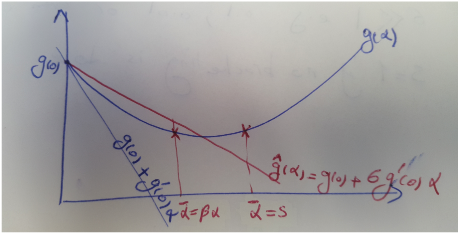
$$\hat{g}(\alpha) = g(0) + \sigma g'(0)\alpha$$

Armijo stepsize should satisfy:

- $g(\bar{\alpha}) \leqslant \hat{g}(\bar{\alpha})$ (ensure sufficient decrease)
- $g(\gamma\bar{\alpha}) \geqslant \hat{g}(\gamma\bar{\alpha})$ (ensure stepsize is not too small)

where $\gamma = \frac{1}{\beta}$

Stepsize selection via successive reduction: Armijo rule



$$\hat{g}(\alpha) = g(0) + \sigma g'(0)\alpha$$

Armijo Line Search Algorithm :

1. Start with $\alpha_k = s$, $0 < \beta < 1$ and $0 < \sigma < 1$
2. If $f(x_k) - f(x_k + \alpha_k d_k) > \sigma\alpha_k \left(-\nabla f(x_k)^\top d_k\right)$
   STOP
   else
   $\alpha_k \leftarrow \beta\alpha_k$ and repeat

In practice the following choices are used

- β: 1/2 to 1/10
- $\sigma \in [10^{-5}, 10^{-1}]$
- if no bracketing is not use $s = 1$

Stepsize selection via successive reduction: Armijo rule

Recall: $g(0) = f(x_k)$, $g'(0) = \nabla f(x_k)^\top d_k$

Armijo: acceptable $\bar{\alpha}$ should satisfy:

$$\begin{cases} g(\bar{\alpha}) \leqslant g(0) + \sigma g'(0)\bar{\alpha} \\ g(\bar{\alpha}\gamma) > g(0) + \sigma g'(0)(\gamma\bar{\alpha}) \end{cases} \Leftrightarrow \begin{cases} f(x_k + \bar{\alpha} d_k) - f(x_k) \leqslant \sigma\bar{\alpha}\nabla f(x_k)^\top d_k \\ f(x_k + \gamma\bar{\alpha} d_k) - f(x_k) > \sigma\gamma\bar{\alpha}\nabla f(x_k)^\top d_k \end{cases}$$

where $\beta \in (0, 1)$ and $\sigma \in (0, 1)$, $\gamma = \frac{1}{\beta}$

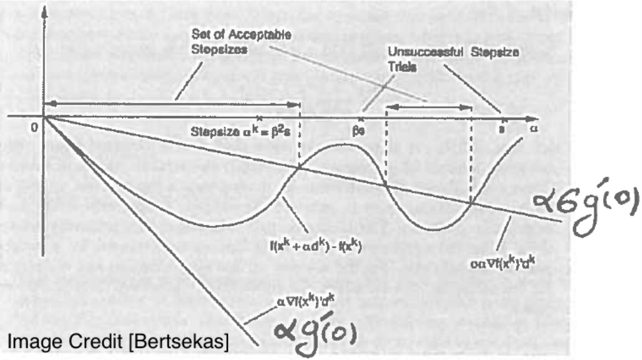
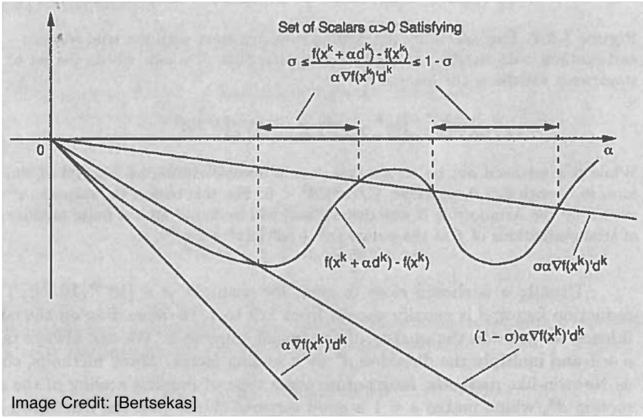


Image Credit [Bertsekas]

Goldenstein



Definition: a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is called L-Lipschitz if and only if

$$\|f(x) - f(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

We denote the class of L-Lipschitz functions by $\mathcal{C}_L$.

— — — — — — — ——

Lemma (Descent Lemma) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Consider any $x, y \in \mathbb{R}^n$. Suppose that

$$\|\nabla f(x + ty) - \nabla f(x)\| \leqslant L\, t\, \|y\|, \quad \forall t \in [0, 1]$$

where L is some scalar. Then.

$$f(x + y) \leqslant f(x) + \nabla f(x)^\top y + \frac{L}{2}\|y\|^2.$$

or

$$f(z) \leqslant f(x) - \nabla f(x)^\top (z - x) + \frac{L}{2}\|z - x\|^2.$$

Admissible fixed stepsize with convergence guarantees for the steepest descent algorithm

Theorem

Consider the steepest descent method $x_{k+1} = x_k - \alpha \nabla f(x_k)$ with fixed stepsize α. Let $\nabla f(x) \in C_L$ and $f^\star = \min f(x) > -\infty$. Then the gradient descent algorithm with fixed stepsize satisfying $0 < \alpha < \frac{2}{L}$ will converge to a stationary point starting from any initial condition.

Proof: Using the last lemma from the previous page we can write

$$f(x_k - \alpha\nabla f(x_k)) - f(x_k) - \nabla f(x)^\top (x_k - \alpha \nabla f(x_k) - x_k) \leqslant \frac{L}{2}\|x_k - \alpha\nabla f(x_k) - x_k\|^2$$

$$f(x_k - \alpha\nabla f(x_k)) - f(x_k) \leqslant -\alpha \nabla f(x)^\top \nabla f(x_k) + \frac{L\alpha^2}{2}\|\nabla f(x_k)\|^2$$

$$f(x_{k+1}) - f(x_k) \leqslant -(\alpha - \frac{L\alpha^2}{2})\|\nabla f(x_k)\|^2$$

To achieve reduction we need $(\alpha - \frac{L\alpha^2}{2}) > 0$, therefore $0 < \alpha < \frac{2}{L}$. From the last inequality above we also have

$$f(\bar{x}) - f(x_0) \leqslant -(\frac{2\alpha - L\alpha^2}{2}) \sum_{k=1}^{\infty} \|\nabla f(x_k)\|^2 \Rightarrow$$

$$\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \leqslant \underbrace{\frac{2}{2\alpha - L\alpha^2}(f(x_0) - f(x_\infty))}_{\text{bounded}}$$

Therefore $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$ therefore $x_\infty = x^\star$.

Admissible fixed stepsize with convergence guarantees for the steepest descent algorithm: special case of quadratic costs

Lemma

Consider a quadratic cost function $f(x) = \frac{1}{2}x^\top Qx + b^\top x + c$, with $Q > 0$, and let $x^\star$ be the unique unconstrained minimizer of this cost function. Starting from any initial condition, the following assertions hold:

- **(a)** *For the steepest descent algorithm with exact line search, we have $x_k \to x^\star$ (this is called global convergence).*
- **(b)** *For steepest descent algorithm with fixed stepsize, we have global convergence if and only if the stepsize α satisfies $0 < \alpha < \frac{2}{\lambda_{\max}(Q)}$.*

Note that

$$\nabla f(x) = Qx + b$$

Therefore $\|\nabla f(x) - \nabla f(y)\| = \|Qx + b - (Qy + b)\| = \|Q(x - y)\| \leqslant \|Q\|\|x - y\|$. Since $Q > 0$, its norm is equal to its maximum eigenvalue, i.e., $\|Q\| = \lambda_{\max}(Q)$. Therefore, the proof of assertion (b) follows directly from the results in the previous page.

Proposition: Convergence of a Constant Stepsize

Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$, and assume that $\{d_k\}$ us gradient related. Assume that for some constant $L > 0$, we have

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

and that for all k we have $d_k \neq 0$ and

$$\epsilon \leqslant \alpha_k \leqslant (2 - \epsilon)\bar{\alpha}_k,$$

where

$$\bar{\alpha}_k = \frac{|\nabla f(x_k)^\top d_k|}{L\|d_k\|^2}$$

and ϵ is a fixed positive scalar. Then every limit point of $\{x_k\}$ is a stationary point of f.
For Steepest descent algorithm: $\epsilon \leqslant \alpha_k \leqslant \frac{2-\epsilon}{L}$ (set $\epsilon = 0$ to recover the result we have obtained earlier)

Proposition: **Convergence of a Diminishing Stepsize**

Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$. Assume that for some constant $L > 0$ we have

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

and that there exists positive scalars c_1 and c_2 such that for all k we have

$$c_2\|\nabla f(x_k)\|^2 \leqslant -\nabla f(x_k)^\top d_k, \quad \|d_k\|^2 \leqslant c_2\|\nabla f(x_k)\|^2.$$

Suppose also that

$$\alpha_k \to 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Then either $f(x_k) \to -\infty$ or else $\{f(x_k)\}$ converges to a finite value and $\nabla f(x_k) \to 0$. Furthermore, every limit point of $\{x_k\}$ is a stationary point of f.