# Iterative Minimization

We are interested in minimizing $\phi : \mathbb{R}^n \to \mathbb{R}$ using an iterative algorithm:

$$x_{k+1} = x_k + \alpha_k d_k . \qquad (1)$$

Question: (i) what should the sequence $d_k$ be?

(ii) For a given choice of $d_k$ what should $\alpha_k$ be?

For (ii) the best choice would be

$$\alpha_k^* = \arg\min_{\alpha} \phi(x_k + \alpha d_k)$$

to be determined by searching in the line defined by $\{x_k + \alpha d_k : \alpha \in \mathbb{R}\}$. Exact one dimensional line search may be costly / impractical and often one uses suboptimal rules (e.g. Armijo's rule)

$d_k$ is called a descent direction if at $x_k$, there exists $\delta > 0$ such that

$$\phi(x_k + \alpha d_k) < \phi(x_k)$$

for all $\alpha \in (0, \delta)$. ~~In particular,~~

Let $\nabla\phi(x)$ be defined as the unique vector field (gradient) such that

$$\langle \nabla\phi(x), h \rangle = D\phi(x) \cdot h$$

$\forall\, h \in \mathbb{R}^n$ and $\langle \cdot , \cdot \rangle$ denotes a positive definite inner product on $\mathbb{R}^n$

Then, $d_k$ is a descent direction if

$$\langle \nabla\phi(x_k), d_k \rangle = D\phi(x_k)\cdot d_k$$

$$= \lim_{\alpha\to 0} \frac{\phi(x_k+\alpha d_k)-\phi(x_k)}{\alpha}$$

$$= \lim_{\alpha\to 0_+} \frac{\phi(x_k+\alpha d_k)-\phi(x_k)}{\alpha}$$

$$< 0 .$$

Consider $\bar{d}_k = -\dfrac{\nabla\phi(x_k)}{\|\nabla\phi(x_k)\|}$

where $\|y\| = \langle y, y\rangle^{1/2}$. It is a descent direction if $\nabla\phi(x_k) \neq 0$.

For any $d_k$, with $\|d_k\| = 1$

$$\langle \nabla\phi(x_k), d_k \rangle \geq -|\langle \nabla\phi(x_k), d_k\rangle|$$

$$\geq -\|\nabla\phi(x_k)\|\cdot\|d_k\|$$

(Cauchy-Schwartz)

$$= -\|\nabla\phi(x_k)\|$$

$$= \langle \nabla\phi(x_k), \bar{d}_k\rangle$$

Hence we call $\bar{d}_k$ the direction of steepest descent at $x_k$ and (1) is a steepest descent alg. if $d_k = \bar{d}_k$

__Theorem__ Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is $C^1$ on the set $S = \{x \in \mathbb{R}^n : \phi(x) \le \phi(x_1)\}$. Suppose that $S$ is a closed and bounded set and $\{x_k\}$ is generated by

$$x_{k+1} = x_k + \alpha_k \left( - \frac{\nabla \phi(x_k)}{\|\nabla \phi(x_k)\|} \right)$$

where $\alpha_k$ is chosen by line search.

Then every point $\bar{x}$ that is a limit point of the sequence $\{x_k\}$, satisfies $\nabla \phi(\bar{x}) = 0$.

__Proof__ (by contradiction).

By Weierstrass, at least one limit point of the sequence $\{x_k\}$ must exist, say $\bar{x}$. Without loss of generality assume that $\lim_{k \to \infty} x_k = \bar{x}$.

Suppose $\nabla \phi(\bar{x}) \neq 0$.

Then, there is $\bar{\alpha} > 0$ such that

$$\delta = \phi(\bar{x}) - \phi(\bar{x} + \bar{\alpha} \bar{d}) > 0$$

where $\bar{d} = -\nabla \phi(\bar{x}) / \|\nabla \phi(\bar{x})\|$. Also $\bar{x} + \bar{\alpha} \bar{d} \in$ the interior of $S$.

Let $d_k$ be the sequence of descent directions, $d_k = -\nabla \phi(x_k) / \|\nabla \phi(x_k)\|$

Since $\phi$ is $C'$, $\lim_{k \to \infty} d_k = \bar{d}$

Since $\bar{x} + \bar{\alpha}\bar{d} \in$ interior $(S)$

and $x_k + \bar{\alpha} d_k \longrightarrow \bar{x} + \bar{\alpha}\bar{d}$ for $k$ large

we have,

$$\phi(x_k + \bar{\alpha} d_k) \leq \phi(\bar{x} + \bar{\alpha}\bar{d}) + \frac{\delta}{2}$$

$$= \phi(\bar{x}) - \delta + \frac{\delta}{2}$$

$$= \phi(\bar{x}) - \frac{\delta}{2}$$

But

$$\phi(\bar{x}) \leq \phi(x_k + \alpha_k d_k)$$

$$\leq \phi(x_k + \bar{\alpha} d_k)$$

$$\leq \phi(\bar{x}) - \frac{\delta}{2}$$

a CONTRADICTION.

Thus $\bar{d} = \nabla \phi(\bar{x}) = 0$ $\square$

In the next page we state a theorem that generalizes the result here to infinite dimensions

Theorem  Let $\phi : X \to \mathbb{R}$ be a functional on a Hilbert space $X$, which is bounded below and twice Fréchet differentiable. Let $x_1 \in X$.

Let $S \subset X$ be defined as the closed convex hull of $\{x : \phi(x) < \phi(x_1)\}$.

Suppose $D^2\phi(x)$ is self adjoint and satisfies $0 < m\,\mathbb{1} \le D^2\phi(x) \le M\,\mathbb{1}$ $\forall\, x \in S'$.

If $\{x_n\}$ is the sequence generated by steepest descent applied to $\phi$ starting at $x_1$, then $\phi'(x_n) = (D\phi)(x_n) \to 0$ as $n \to \infty$.

Furthermore, $\exists\; x^* \in S$ such that $x_n \to x^*$ and $x^* = \inf\{\phi(x) : x \in X\}$

EXAMPLE

Steepest Descent in the Quadratic Case.

Let $X$ be a Hilbert space (i.e. a vector space with positive definite inner product $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ and associated $\| \cdot \|$ defined by $\|x\| = \sqrt{\langle x, x \rangle}$ satisfying $(X, \| \cdot \|)$ is a Banach space).

Let $Q : X \to X$ be a bounded linear operator satisfying

(i) $Q = Q^*$ (self-adjointness)

(ii)
$$m = \inf_{x \neq 0} \frac{\langle x, Qx \rangle}{\langle x, x \rangle} > 0$$

$$M = \sup_{x \neq 0} \frac{\langle x, Qx \rangle}{\langle x, x \rangle} > 0$$

One can then show that $Q$ is invertible.

For each $b \in X$, $Qx = b$ has a unique solution which is simply the minimizer of the function $\phi : X \to \mathbb{R}$

$$x \mapsto \phi(x) = \langle x, Qx \rangle - 2 \langle b, x \rangle .$$

A descent algorithm for $\phi$ is of the form

$$x_{n+1} = x_n + \alpha_n d_n$$

$$= x_n - \frac{1}{2}\alpha_n (D\phi)(x_n)$$

$$= x_n + \alpha_n (b - Qx_n)$$

<u>Steepest Descent</u>

$$\alpha_n = \frac{\langle b - Qx_n, \; b - Qx_n \rangle}{\langle b - Qx_n, \; Q(b - Qx_n) \rangle}$$

Carry out the details of this example.