Unconstrained minimization

Mathématiques Informatique et Statistique Appliquées (MISA)
Université d'Antananarivo

Unconstrained minimization problems

$$\text{minimize } f(x)$$

where $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex and twice differentiable. We assume that the optimal value $p^\star = \inf_x f(x)$ is attained (and finite).

Optimality condition

f is differentiable and convex, a point $x^\star$ is optimal iff $\nabla f(x^\star) = 0$.

Sometimes, we can analytically solve the optimality equation but usually the problem must be solved by an **iterative algorithm**.

Descent methods

We want to produce a minimizing sequence $x^{(k)}$, $k = 1, \ldots$, where $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$ with $f(x^{(k+1)}) < f(x^{(k)})$

- $\Delta x^{(k)}$ is the **search direction**
- $t^{(k)} > 0$ is the **step size**

Descent direction

$\Delta x^{(k)}$ is a **descent direction** if $\nabla f(x^{(k)})^\top \Delta x^{(k)} < 0$ (from convexity).

Algorithm 1: General descent method

Start at a point $x \in \textbf{dom } f$

repeat

Determine a descent direction Δx

Choose a step size $t > 0$

Update $x = x + t\Delta x$

until *stopping criterion is satisfied*;

Choosing the step size - Line search

Exact line search

Choose t to minimize f along the ray $\{x + t\Delta x | t > 0\}$

$$t = \arg\min_{t>0} f(x + t\Delta x)$$

Backtracking line search

An inexact line search approach that chooses t just to reduce f enough. It depends on two constants $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$.

Algorithm 2: Backtracking line search

$t = 1$

repeat

  $t = \beta t$

until $f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^\top \Delta x$;

Gradient descent

It is a descent method with search direction $\Delta x = -\nabla f(x)$.

Algorithm 3: Gradient descent

Start at a point $x \in \textbf{dom } f$

repeat

$\Delta x = -\nabla f(x)$

Choose a step size $t > 0$

Update $x = x + t\Delta x$

until *stopping criterion is satisfied*;

The stopping criterion is usually of the form $||\nabla f(x)||_2 < \epsilon$.

Newton's method

It is a descent method with search direction $\Delta x = -(\nabla^2 f(x))^{-1} \nabla f(x)$.

We can say that $v = \Delta_x$

- minimizes the second order approximation

$$f(x+v) \approx f(x) + \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x) v$$

- or equivalently solves the linearized optimality condition

$$\nabla f(x+v) \approx \nabla f(x) + \nabla^2 f(x) v = 0$$

More

Explore further

- Steepest descent methods (generalizes gradient descent)
- Quasi-Newton methods (like BFGS, L-BFGS)
- Convergence analysis
- Self-concordance
- Gradient-free optimization