Optimization

Marcin Kuta

Stationary points

Stationary point (critical point, equilibrium point)

$$\nabla F(x) = 0 \tag{1}$$

Theorem (First-order necessary conditions for minimum)

If x^* is a local minimizer and f is continuously differentiable in an open neighborhood of x^*, then $\nabla F(x^*) = 0$

Theorem (Second-order necessary conditions for minimum)

If x^* is a local minimizer and $\nabla^2 F$ is continuous in an open neighborhood of x^*, then $\nabla F(x^*) = 0$ and $\nabla^2 F(x^*)$ is positive semidefinite

Stationary points

Theorem (Second-order sufficient conditions)

Suppose that $\nabla^2 F$ is continuous in an open neighborhood of x^* and that $\nabla F(x^*) = 0$.
If $\nabla^2 F(x^*)$ is

- *positive definite, then x^* is a strict local minimizer of f.*
- *negative definite, then x^* is a strict local maximizer of f.*
- *indefinite, then x^* is a saddle point.*
- *singular, then various pathological situations can occur.*

Stationary points

$$\begin{vmatrix} a_{11} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

If sequence of signs of determinants is

- all positive $\Rightarrow$ A positive definite $\Rightarrow$ minimum
- alternates, starting from negative $\Rightarrow$ A negative definite $\Rightarrow$ maximum
- any sign is wrong $\Rightarrow$ A indefinite $\Rightarrow$ saddle point.
- no sign is wrong, but one or more terms is 0 $\Rightarrow$ A is positive semidefinite or negative semidefinite $\Rightarrow$ more delicate work is needed

Methods

Solving equations	Optimization
bisection method	golden section search
secant method	successive parabolic interpolation
Newton's method	Newton's method
Conjugate Gradient method	Conjugate Gradient method

Convergence of optimization methods

Convergence

- linear

$$\lim_{n\to\infty} \frac{||x_{n+1} - x^*||}{||x_n - x^*||} = C, \quad 0 < C < 1 \tag{2}$$

- superlinear

$$\lim_{n\to\infty} \frac{||x_{n+1} - x^*||}{||x_n - x^*||} = 0, \quad 1 < p \tag{3}$$

- quadratic

$$\lim_{n\to\infty} \frac{||x_{n+1} - x^*||}{||x_n - x^*||^2} = C \tag{4}$$

- convergence of order p at rate μ

$$\lim_{k\to\infty} \frac{||x_{n+1} - x^*||}{||x_n - x^*||^p} = C \tag{5}$$

Convergence of optimization methods

Linear order of convergence (p = 1):

- $||x_n - x^*|| = O(\lambda^n)$

Quadratic order of convergence (p = 2):

- the number of correct digits approximately doubles at each iteration.

Convergence of optimization methods

Method	Convergence
coordinate descent method	no convergence
golden section search	linear, $r = 1$, $C \approx 0.618$
successive parabolic interpolation	superlinear, $r \approx 1.324$
steepest descent	linear, $r = 1$
Newton's method	quadratic, $r = 2$
Quasi-Netwon methods	superlinear
–BFGS	superlinear
Conjugate Gradient method	linear

Steepest descent with line search

- $f : \mathbb{R}^n \to \mathbb{R}$
- $x_k \in \mathbb{R}^n$
- $\alpha_k \in \mathbb{R}$

Steepest descent

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{6}$$

Line search

$$\min_{\alpha_k} f(x_k - \alpha_k \nabla f(x_k)) \tag{7}$$

Newton's methods

$$H_f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1 +} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{8}$$

$$\begin{gathered} H_f(x_k) s_k = -\nabla f(x_k) \\ x_{k+1} = x_k + s_k \end{gathered} \tag{9}$$

References

[1] http://heath.cs.illinois.edu/scicomp/notes/cs450_chapt06.pdf

[2] Jorge Nocedal, Stephen J. Wright,
Numerical Optimization, 2006,
https://doi.org/10.1007/978-0-387-40065-5

[3] http://www.benfrederickson.com/numerical-optimization