Descent - L3 - 6/10 - Optimization

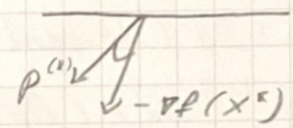
$f(x) \to \min$, $x \in D \subseteq \mathbb{R}^n$, $n \geq 1$, f is smooth

↳ descent dirrection, step size, stopping criterion

$x^{(k+1)} = x^{(k)} + a^{(k)} p^{(k)}$ → descent direction

dyadic/binary search }
fibonacci search } only for unimodal functions

Wolfe conditions: sufficient decrease]
curvatirre condition]



Ex [$f(x) = x^2$ on $[-2, 2]$, one-dimensional gradient descent method
[starting at $x_1 = 2$ in the direction $p_k = -\text{sign}(x_k)$, step $a_k = 2 + 3(2^{-k-1})$

Trust region method

line search methos - find a descent direction → find the next point in this direction

find a region "of posible good step" → find a point in this region

Ussually we approximate the objective function f with a simpler objective $\tilde{f}$

Potential promblem - solution $\tilde{x}^*$ of $\min \tilde{f}(x)$ lies in the region where $\tilde{f}$ poorly approximates f

Solution - restrict the optimization of $\tilde{f}$ to the region where we trust $\tilde{f}$ is similar near a point a we do the quadratic approximation

$$f(x) \approx \tilde{f}(x) = f(d) + \nabla f(d)^T \cdot (x-d) + \frac{1}{2}(x-d)^T \nabla^2 f(d)(x-d)$$

at d, f and $\tilde{f}$ match: $f(0) = \tilde{f}(0)$

the further we go from a the worse the approximation

Generally: given δ, x_1 and $k=0$, repeat

(1) $k \to k+1$

(2) find a solution x_k^* of the minimization problem $\tilde{f} \to$ min subject to $\|x - x^*_{k-1}\| \leq \delta$

(3) if $\tilde{f}(x_k^*) \approx f(x_k^*)$, then increase δ, otherwise decrease δ

- until the required precision is reached →

Rosebrock function: $f(x) = (a - x_1)^2 + b(x_2 - x_1)^2$

global minimum at $x^* = (a, a^2)$

Descent Direction: p_k is a descent direction if $p_k^T \nabla f(x_k) < 0$

more generally (in line search methods)

we consider $p_k = -B_k^{-1} \nabla f(x_k)$ with B_k positive definite

$B_k = Id$ (descent method)

$B_k = Hf(x_k)$ (Newton method)

$B_k \approx Hf(x_k)$ (Quasi Newton method)

$f(x_{k+1}) \cong f(x_k) + p_k^T \nabla f(x_k) + \frac{1}{2} p_k^T \nabla^2 f(x_k) p_k$

↳ up to quadratic error

$p_k = -B_k^{-1} \nabla f(x_k)$

notation: $H = \nabla^2 f$

$p_k = -B_k^{-1} \nabla f(x_k)$

$p_k^T \nabla f(x_k) = -(\nabla f(x_k))^T \cdot B_k \cdot \nabla f(x_k) < 0$

steepest descent method: descent along inverse gradient

assume $f(x) = \frac{1}{2} x^T Q x - b^T x$ where Q is symmetric and positive definite

gradient is given by $\nabla f(x) = Qx - b$ and the minimizer (x^*) is the (unique) solution of $Qx = b$

exact line search writes as $x_{k+1} = x_k - \hat{\alpha}_k \nabla f(x_k)$

to study the rate of convergence we introduce a weighted norm of a vector $x \in \mathbb{R}^n$, $\|x\|_Q^2 = x^T Q x$

Lemma: $\frac{1}{2} \|x - x^*\|_Q^2 = f(x) - f(x^*)$

Theorem: When the steepest descent method with exact line searches $(\hat{\alpha}_k)$ is applied to the strongly convex quadratic function

$$\|x_{k+1} - x^*\|_Q^2 \leq \left[\frac{\lambda_n - \lambda_1}{\lambda_n - \lambda_1}\right]^2 \|x_k - x^*\|_Q^2$$

where $0 \leq \lambda_1 \leq \dots \leq \lambda_n$ are the eigenvalues of Q

convergence of the steepest descent method under the best conditions is linear

Definition: let f be twice differentiable the newton's method is the line search method defined by $p_k = -(Hf(x_k))^{-1} \nabla f(x_k)$ since $(Hf(x_k))^{-1}$ might not always be positive definite, then Newton's method does not always define a descent method. However, near the solutions (minimizers), the convergence is quadratic.

Newton's method:

Assume f is regular class (3 is enough in a neighborhood of a solution x^* (minimum of f) where the sufficient optimality conditions hold.

$x_{k+1} = x_k + p_k$ where p_k is the Newton direction

(a) $x_k \to x^*$ if x_0 is close enough to x^*

(b) the rate of convergence of $\{x_k\}_{k\geq 0}$ is quadratic

(c) $\|\nabla f(x_k)\| \to 0$ quadratically.