CPSC 406

Computational Optimization

Dept of Computer Science

University of British Columbia

Scaled Descent

Scaled Gradient Method

$$(P) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \qquad f : \mathbb{R}^n \to \mathbb{R}$$

- Make a linear change of variables with

  S nonsingular $n \times n$, $\quad x = Sy \quad$ ie $\quad y := S^{-1}x$

$$(P_{\text{scaled}}) \quad \underset{y \in \mathbb{R}^n}{\text{minimize}} \quad g(y) := f(Sy)$$

- Apply gradient method to scaled problem:

$$y_{k+1} = y_k - \alpha_k \nabla g(y_k) \quad \text{with} \quad \nabla g(y) = S^T \nabla f(Sy)$$

- Multiply on left by S:

$$x_{k+1} = x_k - \alpha_k SS^T \nabla f(x_k)$$

- Scaled gradient method : with $D = SS^T$,

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

SCALED DESCENT

The scaled gradient $-D\nabla f(x)$ is a descent direction:

$$f'(x; -D\nabla f(x)) = -\nabla f(x)^T D \nabla f(x) < 0$$

because $D = SS^T \succ 0$ (S nonsingular)

Scaled Gradient Method

for $k = 0, 1, 2, \ldots$

- choose scaling matrix D_k
- compute scaled gradient $d_k = D_k \nabla f(x_k)$
- compute steplength α_k by linesearch on the func'n

$$\phi(\alpha) = f(x_k - \alpha d_k)$$

- $x_{k+1} = x_k - \alpha_k d_k$
- STOP if $\|\nabla f(x_{k+1})\| \leq tol$

CHOOSING THE SCALING MATRIX

- Scaled gradient method is just the gradient method applied to g:

$$g(y) = f(D^{\frac{1}{2}}y) \quad = f(x) \qquad [D = SS^T = D^{\frac{1}{2}}D^{\frac{1}{2}}]$$

$$\nabla g(y) = D^{\frac{1}{2}}\nabla f(D^{\frac{1}{2}}y) \quad = D^{\frac{1}{2}}\nabla f(x)$$

$$\nabla^2 g(y) = D^{\frac{1}{2}}\nabla^2 f(D^{\frac{1}{2}}y)D^{\frac{1}{2}} \quad = D^{\frac{1}{2}}\nabla^2 f(x)D^{\frac{1}{2}}$$

- Choose D_k so to make $D_k^{\frac{1}{2}}\nabla^2 f_k D_k^{\frac{1}{2}}$ as well conditioned as possible. Take $H_k \equiv \nabla^2 f(x_k)$:

$$D_k = \begin{cases} H_k^{-1} \succ 0 & [\text{Newton}] \quad D_k^{\frac{1}{2}}H_k D_k^{\frac{1}{2}} = I \quad \text{cond}(I) = 1 \\[2em] (H_k + \lambda I)^{-1} & \left[\begin{array}{l}\text{Damped} \\ \text{Newton}\end{array}\right] \quad D_k^{\frac{1}{2}}H_k D_k^{\frac{1}{2}} \to I \text{ as } \lambda \to 0 \\[2em] \text{diag}\left(\dfrac{\partial^2 f(x_k)}{\partial x_i^2}\right)^{-1} & \left[\begin{array}{l}\text{Diagonal} \\ \text{scaling}\end{array}\right] \\ \succ 0 & \end{cases}$$

GAUSS-NEWTON METHOD

$$\text{(NLS)} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) = \tfrac{1}{2}\|r(x)\|^2 \qquad r_i : \mathbb{R}^n \to \mathbb{R}$$

cont. diff'l

$i = 1, \ldots, m$

Gauss-Newton Method:

$$x_{k+1} = \underset{x}{\text{argmin}} \ \tfrac{1}{2}\|\underbrace{r_k + A_k(x - x_k)}_{\text{[linearization of } r \text{ at } x_k]}\|^2$$

$r_k = r(x_k)$

$$A_k = \begin{bmatrix} \nabla r_1(x_k)^T \\ \vdots \\ \nabla r_m(x_k)^T \end{bmatrix}$$

$$= (A_k^T A_k)^{-1} A_k^T (A_k x_k - r_k)$$

$$= x_k - (A_k^T A_k)^{-1} A_k^T r_k$$

$$= x_k - (A_k^T A_k)^{-1} \nabla f(x_k) \qquad \nabla f(x_k) = A_k^T r_k$$

Thus, we see that the Gauss-Newton method is a scaled gradient method with the scaling matrix

$$D_k = (A_k^T A_k)^{-1}$$

NEWTON'S METHOD

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \qquad f: \mathbb{R}^n \to \mathbb{R} \text{ twice cont. diff'l}$$

Given x_k, define x_{k+1} as the minimizer of the quadratic approx to f at x_k:

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \left\{ f_k + \nabla f_k^T (x - x_k) + \tfrac{1}{2} (x - x_k)^T \nabla^2 f_k (x - x_k) \right\}$$

Minimizer is well-defined if $\nabla^2 f_k \equiv \nabla^2 f(x_k) \succ 0$. In that case,

$$\begin{aligned} x_{k+1} &= x_k - (\nabla^2 f_k)^{-1} \nabla f_k \\ &= x_k + d_k \end{aligned}$$

$(d_k \equiv \text{Newton direction})$

where d_k solves $\nabla^2 f_k d = -\nabla f_k$.

Pure Newton's Method

x_0 given

for $k = 0, 1, 2, \ldots$

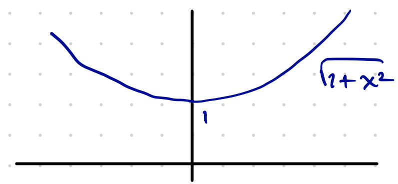
- $g_k \leftarrow \nabla f(x_k)$ compute gradient
- $H_k \leftarrow \nabla^2 f(x_k)$ compute Hessian
- d_k solves $H_k d = -g_k$ compute Newton step
- $x_{k+1} \leftarrow x_k + d_k$
- STOP if $\|\nabla f(x_{k+1})\| \le \text{TOL}$

Convergence of (pure) Newton's Method

- require $\nabla^2 f_k \succ 0$ for all k. Ensures descent:

$$\nabla^2 f_k d = -\nabla f_k \Rightarrow d^T \nabla^2 f_k d = -d^T \nabla f_k \Rightarrow \nabla f_k^T d < 0.$$

- May still diverge even if $\nabla^2 f_k \succ 0$.



Ex $\quad f(x) = \sqrt{1+x^2}$

$$f'(x) = x(1+x^2)^{-\frac{1}{2}} \qquad f''(x) = (1+x^2)^{-3/2}$$

Newton iteration $\quad x_{k+1} = x_k - \dfrac{f'(x_k)}{f''(x_k)} = x_k - x_k(1+x^2) = -x_k^3$

$$x_k \to \begin{cases} 0 & \text{if } |x_0| < 1 \\ \pm 1 & \text{if } |x_0| = 1 \\ +\infty & \text{if } |x_0| > 1 \end{cases}$$

Convergence of Newton's Method (Pure)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice cont. diff'l and

(a) $\nabla^2 f(x) \succcurlyeq \varepsilon I$ for some $\varepsilon > 0$, $\forall x$

(b) $\| \nabla^2 f(x) - \nabla^2 f(y) \| \leq L \| x - y \|$ $\forall x, y$ for some $L > 0$

Then the pure Newton iteration satisfies if x_* unique:

$$\| x_{k+1} - x_* \| \leq \frac{L}{2\varepsilon} \| x_k - x_* \|^2$$

In addition, if $\| x_0 - x_* \| \leq \varepsilon / L$ then

$$\| x_k - x_* \| \leq \left(\frac{2\varepsilon}{L} \right) \left(\frac{1}{4} \right)^{2^k} \quad k = 0, 1, \ldots$$

Rates of Convergence

Measure how fast a sequence $\{x_k\}_{k=1,2,3\cdots}$ converges to its limit (assuming limit exists).

Suppose $x_k \to x_*$, ie, $\lim_{k\to\infty} \|x_k - x_*\| = 0$.

<u>Linear Convergence</u> : there exists a number $\mu \in (0,1)$ st

$$\lim \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = \mu$$

eg $a_k = 2^{-k}$

$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots \to 0$

<u>Sublinear</u>

$$\lim \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 1$$

$a_k = \frac{1}{k+1}$

$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \cdots$

<u>Superlinear</u> (order q)

$$\lim \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^q} \le M \quad (M > 0)$$

$a_k = \left(\frac{1}{2}\right)^{2^k}$

$q = 2 \equiv$ quadratic

$$\frac{2^{-2^{k+1}}}{2^{-2^k}} = 2^{(-2^{k+1} + 2^k)} = 2^{2^k(1-2)} = 2^{-2^k} \to 1$$