Numerical Optimization

Assignment 3

Xingrong Zong, tnd179

Hold 03, Group 3

February 22th, 2022

Introduction

This assignment investigates a few more of the theoretical properties of the algorithms, such as Goldstein conditions and how they might be suited for Newton's algorithm and Quasi-Newton methods for optimizing arbitrary functions;

Theoretical exercises

Exercise 1

Let f be a convex quadratic, then $f(x) = \frac{1}{2}x^\intercal Qx - b^\intercal x$, where Q is symmetric and positive definite. The gradient is given by $\nabla f(x) = Qx - b$.
Assume p_k is a descent directions and a minimum of one-dimensional function $\phi(\alpha) = f(x_k + \alpha p_k)$, that $\phi'(0) < 0$, so that $\alpha > 0$.
Set α_k as a minimizer and $\phi'(\alpha_k) = \nabla f(x_k + \alpha_k p_k)^\intercal p_k = 0$, then replace

$\nabla f(x_k + \alpha_k p_k)$ with $Q(x_k + \alpha_k p_k) - b$, we get

$$
\begin{aligned}
(Q(x_k + \alpha_k p_k) - b)^\intercal p_k &= 0 \\
(Qx_k + Q\alpha_k p_k - b)^\intercal p_k &= 0 \\
(Qx_k - b)^\intercal p_k + Q^\intercal \alpha_k p_k^\intercal p_k &= 0 \\
Q^\intercal \alpha_k p_k^\intercal p_k &= -(Qx_k - b)^\intercal p_k \\
\alpha_k &= -\frac{(Qx_k - b)^\intercal p_k}{Q^\intercal p_k^\intercal p_k}
\end{aligned}
$$

replace $Qx_k - b$ with $\nabla f(x_k)$, we get $\alpha_k = -\frac{\nabla f(x_k)^\intercal p_k}{Q^\intercal p_k^\intercal p_k}$, and since Q is symmetric and positive definite, which means $Q^\intercal = Q$, so $\alpha_k = -\frac{\nabla f_k^\intercal p_k}{Q p_k^\intercal p_k}$ (3.55).
In conclusion, (3.55) is one-dimensional minimizer along the ray $x_k + \alpha p_k$ on a convex quadratic.

The Goldstein conditions can be stated as a pair of inequalities,

$$
f(x_k) + (1-c)\alpha_k \nabla f_k^\intercal p_k \le f(x_k + \alpha_k p_k) \le f(x_k) + c\alpha_k \nabla f_k^\intercal p_k, 0 < c < \frac{1}{2}
$$

replace α_k with $-\frac{\nabla f_k^\intercal p_k}{Q p_k^\intercal p_k}$, we get

$$
f(x_k) + (1-c)(-\frac{\nabla f_k^\intercal p_k}{Q p_k^\intercal p_k})\nabla f_k^\intercal p_k = f(x_k) - \frac{(\nabla f_k^\intercal p_k)^2}{Q p_k^\intercal p_k} + c\frac{(\nabla f_k^\intercal p_k)^2}{Q p_k^\intercal p_k}
$$

and

$$
f(x_k) + c(-\frac{\nabla f_k^\intercal p_k}{Q p_k^\intercal p_k})\nabla f_k^\intercal p_k = f(x_k) - c\frac{(\nabla f_k^\intercal p_k)^2}{Q p_k^\intercal p_k}
$$

From previous proof, we have $\nabla f(x) = Qx - b$, so we can get $\nabla f_k^\intercal = Qx_k^\intercal - b^\intercal \implies \nabla f_k^\intercal p_k = Qx_k p_k^\intercal - b^\intercal p_k$.

Same as above, we have $f(x) = \frac{1}{2}x^\intercal Qx - b^\intercal x$, we get

$$
\begin{aligned}
f(x_k + \alpha_k p_k) - f(x_k) &= (\frac{1}{2}(x_k + \alpha_k p_k)^\intercal Q(x_k + \alpha_k p_k) - b^\intercal(x_k + \alpha_k p_k)) - (\frac{1}{2}x_k^\intercal Q x_k - b^\intercal x_k) \\
&= \frac{1}{2}(Qx_k^\intercal x_k + Qx_k^\intercal \alpha_k p_k + Q\alpha_k p_k^\intercal x_k + Q\alpha_k p_k^\intercal \alpha_k p_k) - b^\intercal x_k - b^\intercal \alpha_k p_k - \frac{1}{2}x_k^\intercal Q x_k + b^\intercal x_k \\
&= \frac{1}{2}(Qx_k^\intercal \alpha_k p_k + Qx_k \alpha_k p_k^\intercal + Q(\alpha_k)^2 p_k p_k^\intercal) - b^\intercal \alpha_k p_k \\
&= \frac{1}{2}(2Q\alpha_k x_k p_k^\intercal) + \frac{1}{2}Q(\alpha_k)^2 p_k p_k^\intercal - b^\intercal \alpha_k p_k \\
&= \alpha_k(Qx_k p_k^\intercal - b^\intercal p_k) + \frac{1}{2}Q(\alpha_k)^2 p_k p_k^\intercal \\
&= -\frac{\nabla f_k^\intercal p_k}{Qp_k^\intercal p_k}\nabla f_k^\intercal p_k + \frac{1}{2}Q(-\frac{\nabla f_k^\intercal p_k}{Qp_k^\intercal p_k})^2 p_k p_k^\intercal \\
&= -\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} + \frac{1}{2}\frac{(\nabla f_k^\intercal p_k)^2 Q p_k p_k^\intercal}{(Qp_k^\intercal p_k)^2} \\
&= -\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} + \frac{1}{2}\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} \\
&= -\frac{1}{2}\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k}
\end{aligned}
$$

Therefore, $f(x_k + \alpha_k p_k) = f(x_k) - \frac{1}{2}\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} \geq f(x_k) - \frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} + c\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k}$, where $-1 < c - 1 < -\frac{1}{2}$, so $f(x_k + \alpha_k p_k) \geq f(x_k) + (1-c)\alpha_k \nabla f_k^\intercal p_k$.
Same method, $f(x_k + \alpha_k p_k) = f(x_k) - \frac{1}{2}\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k} \leq f(x_k) - c\frac{(\nabla f_k^\intercal p_k)^2}{Qp_k^\intercal p_k}$, where $0 < c < \frac{1}{2}$, so $f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f_k^\intercal p_k$.
So it always satisfies the Goldstein conditions (3.11).

The Goldstein conditions can be stated as a pair of inequalities,

$$f(x_k) + (1-c)\alpha_k \nabla f_k^\intercal p_k \leq f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f_k^\intercal p_k, 0 < c < \frac{1}{2}$$

ensure that the step length α achieves sufficient decrease but is not too short. The second inequality is the sufficient decrease condition (3.4), whereas the first inequality is introduced to control the step length from below.

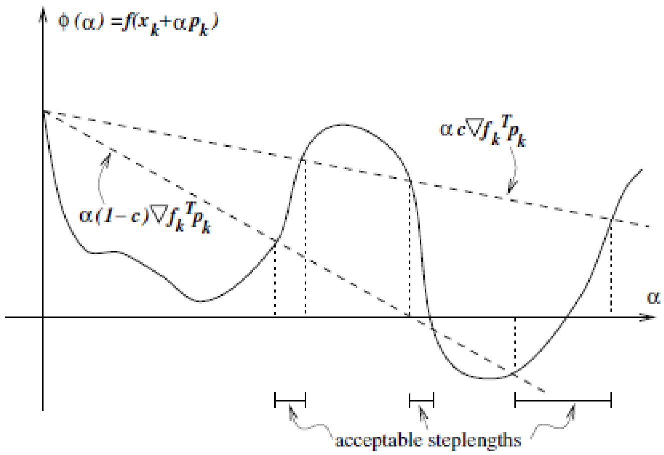

Figure 1: The Goldstein conditions

By using backtracking line search,

$\bar{\alpha} > 0, \rho \in (0,1), c \in (0,1)$; Set $\alpha \leftarrow \bar{\alpha}$;

repeat until $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^{\mathsf{T}} p_k$

$\alpha \leftarrow \rho\alpha$;

end (repeat)

Terminate with $\alpha_k = \alpha$

The initial step length $\bar{\alpha}$ is chosen to be 1 in Newton and quasi-Newton methods, but can have different values in other algorithms such as steepest descent or conjugate gradient. An acceptable step length will be found after a finite number of trials, because α_k will eventually become small enough that the sufficient decrease condition holds. In practice, the contraction factor ρ is often allowed to vary at each iteration of the line search. We need to ensure that at each iteration we have $\rho \in [\rho_{low}, \rho_{high}]$, for some fixed constants $0 < \rho_{low} < \rho_{high} < 1$.
The backtracking line search ensures either that the selected step length α_k is some fixed value (the initial choice $\bar{\alpha}$), or else that it is short enough to satisfy the sufficient decrease condition but not too short. The latter claim holds because

the accepted value α_k is within a factor ρ of the previous trial value, $\frac{\alpha_k}{\rho}$, which was rejected for violating the sufficient decrease condition, that is, for being too long.

This strategy for terminating a ling search is well suited for Newton methods but is less appropriate for quasi-Newton methods that maintain a positive definite Hessian approximation and conjugate gradient methods.

Exercise 2

Suppose that $f(x) = \frac{1}{2}x^\intercal Qx - b^\intercal x$, where Q is symmetric and positive definite. The gradient is given by $\nabla f(x) = Qx - b$ and the minimizer x^* is the unique solution of the linear system $Qx = b$.

Compute the step length α_k that minimizes $f(x_k - \alpha\nabla f_k)$, we get $f(x_k - \alpha\nabla f_k) = \frac{1}{2}(x_k - \alpha\nabla f_k)^\intercal Q(x_k - \alpha\nabla f_k) - b^\intercal(x_k - \alpha\nabla f_k)$, with respect to α, and setting the derivative to 0, we get $\alpha_k = \frac{\nabla f_k^\intercal \nabla f_k}{\nabla f_k^\intercal Q\nabla f_k}$

We introduced weighed norm $||x||_Q^2 = x^\intercal Qx$, we obtain $\frac{1}{2}||x - x^*||_Q^2 = f(x) - f(x^*)$, so $||x_k - x^*||_Q^2 = (x_k - x^*)^\intercal Q(x_k - x^*) = (x_k - x^*)^\intercal QQ^{-1}Q(x_k - x^*) = \nabla f(x_k)^\intercal Q^{-1}\nabla f(x_k)$

$$
\begin{aligned}
||x_{k+1} - x^*||_Q^2 &= (x_{k+1} - x^*)^\intercal Q(x_{k+1} - x^*) \\
&= (x_k - x^* - \alpha_k\nabla f(x_k))^\intercal Q(x_k - x^* - \alpha_k\nabla f(x_k)) \\
&= (x_k - x^*)^\intercal Q(x_k - x^*) - 2\alpha_k\nabla f(x_k)^\intercal Q(x_k - x^*) + \alpha_k^2\nabla f(x_k)^\intercal Q\nabla f(x_k) \\
&= ||x_k - x^*||_Q^2 - 2\alpha_k\nabla f(x_k)^\intercal Q(x_k - x^*) + \alpha_k^2\nabla f(x_k)^\intercal Q\nabla f(x_k) \\
&= ||x_k - x^*||_Q^2 - \frac{(\nabla f(x_k)^\intercal\nabla f(x_k))^2}{\nabla f(x_k)^\intercal Q\nabla f(x_k)} \\
&= \{1 - \frac{(\nabla f_k^\intercal\nabla f_k)^2}{(\nabla f_k^\intercal Q\nabla f_k)(\nabla f_k^\intercal Q^{-1}\nabla f_k)}\}||x_k - x^*||_Q^2
\end{aligned}
$$

Conclusion

From this assignment, I had a deeper view on the convex quadratic function and its one-dimensional minimizer. Also, it always satisfies the Goldstein conditions and how they are well suited for Newton methods but are less appropriate for quasi-Newton methods and conjugate gradient methods.