CMO: Minimizing a quadratic function

Eklavya Sharma

In many search algorithms, given the current point x, we choose the next point as $x+\alpha u$, where u is a descent direction (i.e. $\nabla_f(x)^T u \le 0$) and $\alpha > 0$.

The strategy of choosing α as $\operatorname{argmin}_{\alpha>0} f(x+\alpha u)$, is called **exact line search**.

1 Quadratic function

$$f(x) = \frac{1}{2}x^T Q x - d^T x$$

where Q is symmetric and positive definite.

$$\nabla_f(x) = Qx - d$$

$$\mathrm{H}_f(x) = Q$$

Since the hessian is positive definite, f is convex. So a local minimum is also a global minimum.

Define $x^* = Q^{-1}d$ (Q^{-1} exists because Q is positive definite). We find that x^* is a local minimum because it satisfies the sufficient conditions for it.

$$f(x^*) = -\frac{1}{2}{x^*}^T Q x^*$$

Although we have a closed form solution for x^*, this is sometimes not usable, since finding Q^{-1} takes $O(d^3)$ time, which can be too much if Q is large.

We will therefore explore descent-based methods to compute x^*.

2 Descent-based minimization of quadratic function

Let $u = \nabla_f(x) \ne 0$. Therefore, $u = Q(x - x^*)$.

Let $g(\alpha) = f(x - \alpha u)$.

$$g'(\alpha) = -u^T \nabla_f(x - \alpha u) = -u^T Q(x - \alpha u - x^*) = u^T(\alpha Q u - u)$$

Setting $g'(\alpha)$ to 0, we get

$$\alpha^* = \frac{\|u\|^2}{u^T Q u}$$

Since Q is positive definite, $\alpha^* > 0$.

$g''(\alpha) = u^T Q u > 0$, so α^* is a local minimum of g. Since $g''(\alpha) > 0$ for all α, g is convex, so α^* is a global minimum of g.

Apply Taylor series to find $f(x - \alpha^* u)$ around x,

$$
\begin{aligned}
&f(x - \alpha^* u) = f(x) + \nabla_f(x)^T(-\alpha^* u) + \frac{1}{2}(-\alpha^* u)^T \,\mathrm{H}_f(x)(-\alpha^* u) \\
&\implies f(x) - f(x - \alpha^* u) = \alpha^* \,\nabla_f(x)^T u - \frac{(\alpha^*)^2}{2} u^T Q u \\
&= \left(\frac{\|u\|^2}{u^T Q u}\right)\|u\|^2 - \frac{1}{2}\left(\frac{\|u\|^2}{u^T Q u}\right)^2 u^T Q u \\
&= \frac{1}{2}\frac{\|u\|^4}{u^T Q u}
\end{aligned}
$$

Apply Taylor series to find $f(x)$ around x^*,

$$
\begin{aligned}
&f(x) = f(x^*) + \nabla_f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \,\mathrm{H}_f(x - x^*) \\
&\implies f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T Q(x - x^*) = \frac{u^T Q^{-1} u}{2}
\end{aligned}
$$

Before we can analyze the convergence of a descent-based algorithm to minimize f, we must look at an important result – Kantorovich's inequality.

Theorem 1 (Kantorovich's inequality)**.** *Let Q be a symmetric positive definite matrix. Let λ_1 and λ_d be its maximum and minimum eigenvalues respectively. Then*

$$
\frac{\|u\|^4}{(u^T Q u)(u^T Q^{-1} u)} \geq \frac{4\lambda_1\lambda_d}{(\lambda_1 + \lambda_d)^2}
$$

Let $x^{(k+1)} = x^{(k)} - \alpha u$. Let $E(x) = f(x) - f(x^*)$. Then

$$
\begin{aligned}
&\frac{E(x^{(k+1)})}{E(x^{(k)})} \\
&= 1 - \frac{f(x^{(i)}) - f(x^{(i+1)})}{f(x^{(i)}) - f(x^*)} \\
&= 1 - \frac{\|u\|^4}{(u^T Q u)(u^T Q^{-1} u)} \\
&\leq 1 - \frac{4\lambda_1\lambda_d}{(\lambda_1 + \lambda_d)^2} && \text{(by Kantorovich's inequality)} \\
&\leq \left(\frac{\lambda_1 - \lambda_d}{\lambda_1 + \lambda_d}\right)^2
\end{aligned}
$$

Therefore, E linearly converges to 0. We know that linear convergence is very fast, so this is a good descent method.