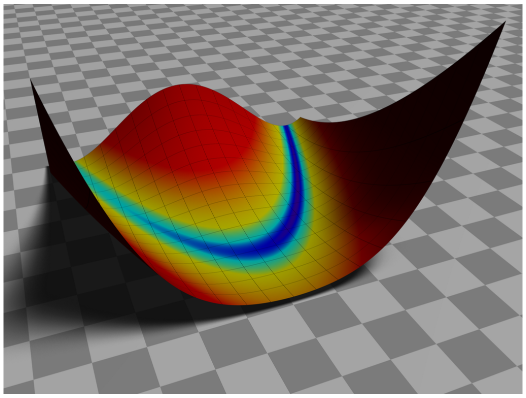
Unconstrained Optimization

Lecture 5



ME EN 575
Andrew Ning
aning@byu.edu

Outline

Conjugate Gradient

Quasi-Newton Methods

gradient descent $\Rightarrow$ conjugate gradient

Newton's method $\Rightarrow$ Quasi-Newton methods

Conjugate Gradient

Only an overview given here, see document on course page for details if interested.

Recall steepest descent:

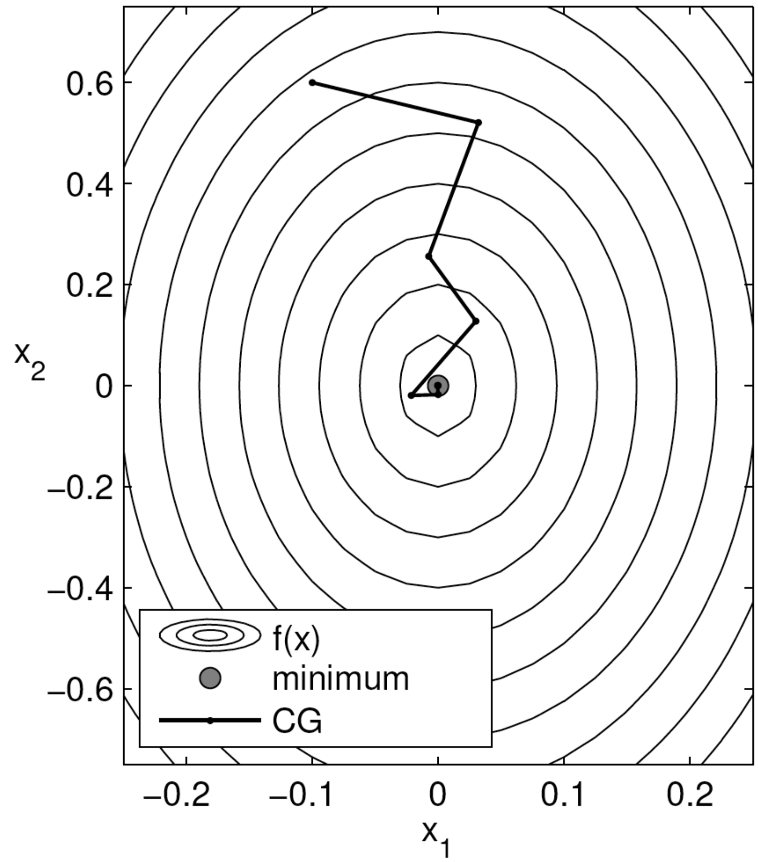
$$p_k = \frac{-g_k}{\|g_k\|}$$

each search direction was orthogonal to the last (for an exact line search)—slow!

Would like to use information from previous iteration, gives us some idea of the skew.

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$$

$$p_k = \frac{-g_k}{\|g_k\|} + \beta_k p_{k-1}$$



Not much harder to implement than steepest descent, but performs much better!

Quasi-Newton Methods

Recall Newton's method:

$$f(x_k + s_k) \approx f_k + g_k^T s_k + \frac{1}{2} s_k^T H_k s_k$$

$$s_k = -H_k^{-1} g_k$$

Recall problem: the Hessian is rarely available.

We can use function values to estimate the gradient. So we can use gradients to estimate the Hessian.

How should we estimate the Hessian?

$$\phi_{k+1}(p) = f_{k+1} + g_{k+1}^T p + \frac{1}{2} p^T H_{k+1} p$$

Choose the new Hessian such that:

- A local quadratic model would yield the same function value at the new point.
- A local quadratic model would yield the same gradient at the new point.
- A local quadratic model would yield the same gradient at the previous point.

$$\phi_{k+1}(p) = f_{k+1} + g_{k+1}^T p + \frac{1}{2} p^T H_{k+1} p$$

First two we get for free, just from the definition of a quadratic model.

$$\begin{aligned} \phi_{k+1}(0) &= f_{k+1} \\ \nabla \phi_{k+1}(0) &= g_{k+1} \end{aligned}$$

The last condition requires

$$
\begin{aligned}
\nabla \phi_{k+1}(-s_k) &= g_k \\
g_{k+1} - H_{k+1} s_k &= g_k
\end{aligned}
$$

Or rearranging:

$$H_{k+1} s_k = y_k$$

where:

- Step size: $s_k = \alpha p_k$ or $s_k = x_{k+1} - x_k$
- Change in gradient: $y_k = g_{k+1} - g_k$

$$H_{k+1} s_k = y_k$$

This is called the secant condition.

Recall other problem: Hessian might not be symmetric and positive definite.

Since we need to estimate the Hessian anyway, we can force the Hessian to be symmetric and positive definite.

Recall that what we really care about is the inverse of the Hessian.

$$p_k = -H_k^{-1} g_k$$

Let's estimate the inverse directly! Let $V_k = H_k^{-1}$ then

$$p_k = -V_k g_k$$

and the secant condition becomes:

$$V_{k+1} y_k = s_k$$