Week 5, Lecture 1

16 February 2024 16:44

It is of interest to know not only when a particular Line-Search (LS) method converges, but also, how fast it converges. We look at that question in this lecture. Unfortunately our results will not be as general as the previous ones which relied on Zoutendijk's Theorem: we will instead limit ourselves to proving a key result about the convergence rate of the Steepest Descent method for the model quadratic problem. The optimization problems we are interested in all resemble the model problem sufficiently close to the minimizer, so this restriction does not involve much loss of generality.

Convergence Rate — S.D. (§ 6.2)

Quadratic cost function:

$$f(\underline{x}) = c + \langle \underline{a}, \underline{x} \rangle + \frac{1}{2} \langle \underline{x}, B\underline{x} \rangle$$

where $\underline{a}$ is a constant vector and B is a symmetric, <u>positive-definite</u> matrix.

Descent direction:

$$\underline{p} = -\nabla f = -B\underline{x}$$

Update step:

$$\underline{x}_{k+1} = \underline{x}_k - \alpha_k \underline{\nabla} f(\underline{x}_k).$$

$$= \underline{x}_k - \alpha_k \underline{\nabla} f_k .$$

Choose α_k:

$$\alpha_k = \arg\min_{\alpha > 0} f(\underline{x}_k - \alpha \nabla f_k)$$

Choose α_k:

$$\alpha_k = \arg\min_{\alpha > 0} f(\underline{x}_k - \alpha \nabla f_k)$$

Model Problem:

$$\alpha_k = \frac{\langle \nabla f_k, \nabla f_k \rangle}{\langle \nabla f_k, B \nabla f_k \rangle}$$

Multiply across by B:

$$B(\underline{x}_{k+1} - \underline{x}_*) = B(\underline{x}_k - \underline{x}_*) - \alpha_k B \nabla f_k$$

Take the dot product / inner product of both sides with $\underline{x}_{k+1} - \underline{x}_*$:

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2 = \langle \underline{x}_{k+1} - \underline{x}_*, B(\underline{x}_{k+1} - \underline{x}_*) \rangle$$

$$= \langle (\underline{x}_{k+1} - \underline{x}_*), B(\underline{x}_k - \underline{x}_*) - \alpha_k B \nabla f_k \rangle$$

Expand out:

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2$$

$$= \langle (\underline{x}_k - \underline{x}_*) - \alpha_k \nabla f_k, B(\underline{x}_k - \underline{x}_*) - \alpha_k B \nabla f_k \rangle$$

Hence:

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2$$

$$= \langle (\underline{x}_k - \underline{x}_*), B(\underline{x}_k - \underline{x}_*) \rangle \quad \| \underline{x}_k - \underline{x}_* \|_B^2$$

$$- \alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle$$

$$- \alpha_k \langle \nabla f_k, B(\underline{x}_k - \underline{x}_*) \rangle$$

$$+ \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

Or:

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2 = \langle (\underline{x}_k - \underline{x}_*), B(\underline{x}_k - \underline{x}_*) \rangle - \alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle - \alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle + \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

Hence:

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2 = \underline{\underline{\langle (\underline{x}_k - \underline{x}_*), B(\underline{x}_k - \underline{x}_*) \rangle}} - 2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle + \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

Identify a weighted norm:

$$\| \underline{v} \|_B^2 = \langle \underline{v}, B \underline{v} \rangle$$

[Norm: $\| \underline{v} \|_B \geq 0$; If $\| \underline{v} \|_B = 0$, then $\underline{v} = 0$

Since B is pos. definite]

Hence,

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2 = \| \underline{x}_k - \underline{x}_* \|_B^2 - 2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle + \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

$\Longrightarrow$

$$\| \underline{x}_{k+1} - \underline{x}_* \|_B^2 - \| \underline{x}_k - \underline{x}_* \|_B^2 = -2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle + \alpha_k^2 \ldots$$

Multiply across by -1 :

$$\overbrace{\| \underline{x}_k - \underline{x}_* \|_B^2 - \| \underline{x}_{k+1} - \underline{x}_* \|_B^2}^{\Delta} = 2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle - \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

Hence,

$$\Delta = 2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B \nabla f_k \rangle - \alpha_k^2 \langle \nabla f_k, B \nabla f_k \rangle$$

Model problem: $B x_* = -a$

$$\nabla f_k = \underline{a} + B \underline{x}_k = -B \underline{x}_* + B \underline{x}_k$$

$$\Longrightarrow \quad \nabla f_k = B (\underline{x}_k - \underline{x}_*) .$$

Back to Δ:

$$\Delta = 2\alpha_k \langle (\underline{x}_k - \underline{x}_*), B\nabla f_k \rangle - \alpha_k^2 \langle \nabla f_k, B\nabla f_k \rangle .$$

$$= 2\alpha_k \langle B(\underline{x}_k - \underline{x}_*), \nabla f_k \rangle - \alpha_k^2 \langle \nabla f_k, B\nabla f_k \rangle .$$

$$\Rightarrow \Delta = 2\alpha_k \langle \nabla f_k, \nabla f_k \rangle - \alpha_k^2 \langle \nabla f_k, B\nabla f_k \rangle$$

Fill in for α_k:

$$\Delta = 2 \overbrace{\frac{\langle \nabla f_k, \nabla f_k \rangle}{\langle \nabla f_k, B\nabla f_k \rangle}}^{\alpha_k} \langle \nabla f_k, \nabla f_k \rangle - \overbrace{\frac{\langle \nabla f_k, \nabla f_k \rangle}{\langle \nabla f_k, B\nabla f_k \rangle} \frac{\langle \nabla f_k, \nabla f_k \rangle}{\cancel{\langle \nabla f_k, B\nabla f_k \rangle}}}^{\alpha_k^2} \cancel{\langle \nabla f_k, B\nabla f_k \rangle}$$

Hence

$$\Delta = \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B\nabla f_k \rangle} .$$

Fill in for Δ:

$$\|\underline{x}_k - \underline{x}_*\|_B^2 - \|\underline{x}_{k+1} - \underline{x}_*\|_B^2 = \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B\nabla f_k \rangle}$$

Re-arrange:

$$\|\underline{x}_k - \underline{x}_*\|_B^2 - \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B\nabla f_k \rangle} = \|\underline{x}_{k+1} - \underline{x}_*\|_B^2 .$$

Re-write:

$$\|\underline{x}_{k+1} - \underline{x}_*\|_B^2 = \|\underline{x}_k - \underline{x}_*\|_B^2 - \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B \nabla f_k \rangle}.$$

$$= \|\underline{x}_k - \underline{x}_*\|_B^2 - \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B \nabla f_k \rangle} \frac{\|\underline{x}_k - \underline{x}_*\|_B^2}{\|\underline{x}_k - \underline{x}_*\|_B^2}$$

$$= \|\underline{x}_k - \underline{x}_*\|_B^2 \left[1 - \frac{\langle \nabla f_k, \nabla f_k \rangle^2}{\langle \nabla f_k, B \nabla f_k \rangle} \frac{1}{\|\underline{x}_k - \underline{x}_*\|_B^2} \right]$$

Back to: $\nabla f_k = B(\underline{x}_k - \underline{x}_*)$.

$$\Rightarrow \quad B^{-1} \nabla f_k = \underline{x}_k - \underline{x}_*.$$

$$\|\underline{x}_k - \underline{x}_*\|_B^2 = \langle (\underline{x}_k - \underline{x}_*), B(\underline{x}_k - \underline{x}_*) \rangle$$

$$= \langle B^{-1} \nabla f_k, \cancel{B}\,\cancel{B^{-1}} \nabla f_k \rangle$$

$$= \langle \nabla f_k, B^{-1} \nabla f_k \rangle$$

Back to:

$$\|\underline{x}_{k+1} - \underline{x}_*\|_B^2 = \|\underline{x}_k - \underline{x}_*\|_B^2 \left[1 - \frac{\langle \nabla f_k, \nabla f_k \rangle \langle \nabla f_k, \nabla f_k \rangle}{\langle \nabla f_k, B \nabla f_k \rangle \langle \nabla f_k, B^{-1} \nabla f_k \rangle} \right]$$

Look at :

$$\frac{\langle \nabla f_k, B \nabla f_k \rangle}{\langle \nabla f_k, \nabla f_k \rangle} \frac{\langle \nabla f_k, B^{-1} \nabla f_k \rangle}{\langle \nabla f_k, \nabla f_k \rangle} \leq \|B\|_2 \, \|B^{-1}\|_2$$

$$= \kappa(B)$$

$$= \frac{\lambda_{max}}{\lambda_{min}}$$

$$\frac{1}{(\cdots)} \geq \frac{1}{\kappa(B)}$$

$$- \frac{1}{(\cdot \cdot)} \leq - \frac{1}{\kappa(B)}$$

$$\Rightarrow \quad 1 - \frac{1}{(\cdots)} \leq 1 - \frac{1}{\kappa(B)}$$

Hence,

$$\|\underline{x}_{k+1} - \underline{x}_* \|_B^2 = \|\underline{x}_k - \underline{x}_* \|_B^2 \left[1 - \frac{1}{(\cdots)} \right]$$

$$\leq \|\underline{x}_k - \underline{x}_* \|_B^2 \left(1 - \frac{1}{\kappa(B)} \right)$$

$$= \|\underline{x}_k - \underline{x}_* \|_B^2 \left(1 - \frac{\lambda_{min}}{\lambda_{max}} \right)$$

Final result:

$$\boxed{\|\underline{x}_{k+1} - \underline{x}_*\|_B \le \|\underline{x}_k - \underline{x}_*\|_B \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max}}\right)^{1/2}.}$$

Where things can go wrong with S.D.

If $\lambda_{max} \gg \lambda_{min}$ $(\kappa(B) \gg 1)$,

then $\dfrac{\lambda_{max} - \lambda_{min}}{\lambda_{max}} \lesssim 1$.

and

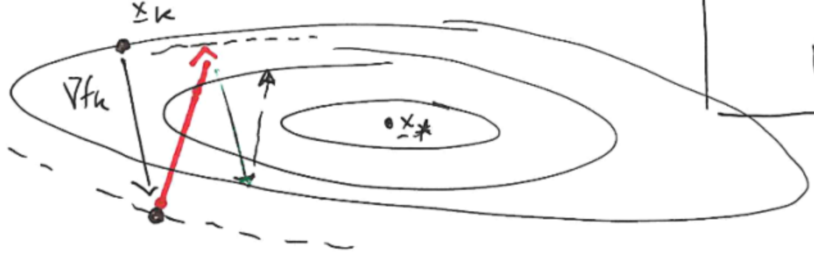
$$\|\underline{x}_{k+1} - \underline{x}_*\|_B \simeq \|\underline{x}_k - \underline{x}_*\|_B.$$

No shrinking!

Hence, SD does not perform well for ill-conditioned problems.

Another way of looking at this:

If $\lambda_{max} >> \lambda_{min}$, then the level sets of the cost function look like an elongated ellipse:



ZIGZAG Pattern