6.252 NONLINEAR PROGRAMMING

LECTURE 5: RATE OF CONVERGENCE

LECTURE OUTLINE

- Approaches for Rate of Convergence Analysis
- The Local Analysis Method
- Quadratic Model Analysis
- The Role of the Condition Number
- Scaling
- Diagonal Scaling
- Extension to Nonquadratic Problems
- Singular and Difficult Problems

APPROACHES FOR RATE OF CONVERGENCE ANALYSIS

- Computational complexity approach
- Informational complexity approach
- Local analysis
- Why we will focus on the local analysis method

THE LOCAL ANALYSIS APPROACH

- Restrict attention to sequences x^k converging to a local min x^*
- Measure progress in terms of an error function $e(x)$ with $e(x^*) = 0$, such as

$$e(x) = \|x - x^*\|, \qquad e(x) = f(x) - f(x^*)$$

- Compare the tail of the sequence $e(x^k)$ with the tail of standard sequences
- Geometric or linear convergence [if $e(x^k) \leq q\beta^k$ for some $q > 0$ and $\beta \in [0, 1)$, and for all k]. Holds if

$$\limsup_{k \to \infty} \frac{e(x^{k+1})}{e(x^k)} < \beta$$

- Superlinear convergence [if $e(x^k) \leq q \cdot \beta^{p^k}$ for some $q > 0$, $p > 1$ and $\beta \in [0, 1)$, and for all k].
- Sublinear convergence

QUADRATIC MODEL ANALYSIS

- Focus on the quadratic function $f(x) = (1/2)x'Qx$, with $Q > 0$.
- Analysis also applies to nonquadratic problems in the neighborhood of a nonsingular local min
- Consider steepest descent

$$x^{k+1} = x^k - \alpha^k \nabla f(x^k) = (I - \alpha^k Q)x^k$$

$$\begin{aligned} \|x^{k+1}\|^2 &= x^{k'}(I - \alpha^k Q)^2 x^k \\ &\le \big(\max \text{ eig. } (I - \alpha^k Q)^2\big)\|x^k\|^2 \end{aligned}$$

The eigenvalues of $(I - \alpha^k Q)^2$ are equal to $(1 - \alpha^k \lambda_i)^2$, where λ_i are the eigenvalues of Q, so

$$\max \text{ eig of } (I - \alpha^k Q)^2 = \max\big\{(1 - \alpha^k m)^2, (1 - \alpha^k M)^2\big\}$$

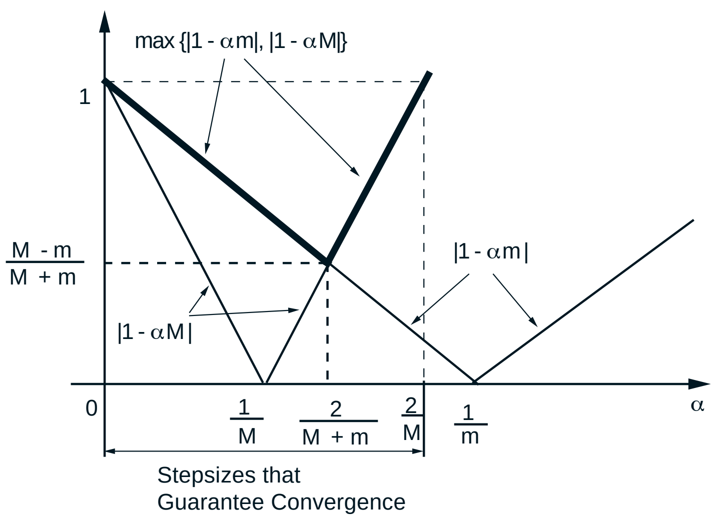
where m, M are the smallest and largest eigenvalues of Q. Thus

$$\frac{\|x^{k+1}\|}{\|x^k\|} \le \max\big\{|1 - \alpha^k m|, |1 - \alpha^k M|\big\}$$

OPTIMAL CONVERGENCE RATE

- The value of α^k that minimizes the bound is $\alpha^* = 2/(M+m)$, in which case

$$\frac{\|x^{k+1}\|}{\|x^k\|} \leq \frac{M-m}{M+m}$$



- Conv. rate for minimization stepsize (see text)

$$\frac{f(x^{k+1})}{f(x^k)} \leq \left(\frac{M-m}{M+m}\right)^2$$

- The ratio M/m is called the *condition number* of Q, and problems with M/m: large are called *ill-conditioned*.

SCALING AND STEEPEST DESCENT

- View the more general method

$$x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$$

as a scaled version of steepest descent.

- Consider a change of variables $x = Sy$ with $S = (D^k)^{1/2}$. In the space of y, the problem is

$$\begin{aligned} &\text{minimize } \ h(y) \equiv f(Sy) \\ &\text{subject to } \ y \in \Re^n \end{aligned}$$

- Apply steepest descent to this problem, multiply with S, and pass back to the space of x, using $\nabla h(y^k) = S\nabla f(x^k)$,

$$y^{k+1} = y^k - \alpha^k \nabla h(y^k)$$

$$Sy^{k+1} = Sy^k - \alpha^k S\nabla h(y^k)$$

$$x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$$

DIAGONAL SCALING

- Apply the results for steepest descent to the scaled iteration $y^{k+1} = y^k - \alpha^k \nabla h(y^k)$:

$$\frac{\|y^{k+1}\|}{\|y^k\|} \leq \max\left\{|1 - \alpha^k m^k|, |1 - \alpha^k M^k|\right\}$$

$$\frac{f(x^{k+1})}{f(x^k)} = \frac{h(y^{k+1})}{h(y^k)} \leq \left(\frac{M^k - m^k}{M^k + m^k}\right)^2$$

where m^k and M^k are the smallest and largest eigenvalues of the Hessian of h, which is

$$\nabla^2 h(y) = S \nabla^2 f(x) S = (D^k)^{1/2} Q (D^k)^{1/2}$$

- It is desirable to choose D^k as close as possible to Q^{-1}. Also if D^k is so chosen, the stepsize $\alpha = 1$ is near the optimal $2/(M^k + m^k)$.
- Using as D^k a diagonal approximation to Q^{-1} is common and often very effective. Corrects for poor choice of units expressing the variables.

NONQUADRATIC PROBLEMS

- Rate of convergence to a nonsingular local minimum of a nonquadratic function is very similar to the quadratic case (linear convergence is typical).

- If $D^k \to \big(\nabla^2 f(x^*)\big)^{-1}$, we asymptotically obtain optimal scaling and superlinear convergence

- More generally, if the direction $d^k = -D^k \nabla f(x^k)$ approaches asymptotically the Newton direction, i.e.,

$$\lim_{k\to\infty} \frac{\|d^k + \big(\nabla^2 f(x^*)\big)^{-1}\nabla f(x^k)\|}{\|\nabla f(x^k)\|} = 0$$

and the Armijo rule is used with initial stepsize equal to one, the rate of convergence is superlinear.

- Convergence rate to a singular local min is typically sublinear (in effect, condition number $= \infty$)