22M174/22C174: Optimization techniques.

Homework 9. Due 05/01/13.

1. Is the convergence of the sequence $x_k := 1/k^2$ for the norm on $\mathbb{R}$ given by the absolute value $|\cdot|$, of Q-order (at least) 1, (at least) Q-linear, (at least) Q-superlinear? Same question for the sequence $x_k := 2^{-k}$.

2. We consider the classical steepest descent (CSD) method with exact line-search on a quadratic $q(x) := b + c^T x + \frac{1}{2}x^T Hx$ with H symmetric and positive definite (Algorithm I.8.1.1).

   (a) First show that by choice of the steplength α_k we have
   $$x_{k+1} = x_k - \left(\frac{\nabla q(x_k)^T \nabla q(x_k)}{\nabla q(x_k)^T H \nabla q(x_k)} \right) \nabla q(x_k).$$

   (b) By using the relation $Hx^* = -c$ prove that
   $$\frac{1}{2}\|x_k - x^*\|_H^2 = q(x_k) - q(x^*).$$

   (c) Use the previous relations to obtain that
   $$\begin{aligned}\|x_k - x^*\|_H^2 - \|x_{k+1} - x^*\|_H^2 &= 2\alpha_k \nabla q(x_k)^T H(x_k - x^*) - \\ &\qquad \alpha_k^2 \nabla q(x_k)^T H \nabla q(x_k).\end{aligned}$$

   (d) Show that
   $$\nabla q(x_k) = H(x_k - x^*).$$

   (e) Use the previous relations to obtain
   $$\|x_k - x^*\|_H^2 - \|x_{k+1} - x^*\|_H^2 = \frac{(\nabla q(x_k)^T \nabla q(x_k))^2}{\nabla q(x_k)^T H \nabla q(x_k)}$$
   and
   $$\|x_k - x^*\|_H^2 = \nabla q(x_k)^T H^{-1} \nabla q(x_k).$$

   (f) Show that
   $$\begin{aligned}&\|x_{k+1} - x^*\|_H^2 = \\ &\qquad \left(1 - \frac{(\nabla q(x_k)^T \nabla q(x_k))^2}{(\nabla q(x_k)^T H \nabla q(x_k))(\nabla q(x_k)^T H^{-1} \nabla q(x_k))}\right) \|x_k - x^*\|_H^2.\end{aligned}$$

(g) For H symmetric and positive definite we have for any vector $y \in \mathbb{R}^n$

$$\frac{(y^T y)^2}{(y^T H y)(y^T H^{-1} y)} \geq 4 \frac{\lambda_{\min}(H)\lambda_{\max}(H)}{(\lambda_{\min}(H) + \lambda_{\max}(H))^2}$$

where $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$ are respectively the smallest and largest eigenvalues of H. This inequality is called the *Kantorovitch inequality* and you are not asked to prove it. From the Kantorovitch inequality and the previous equality, prove finally that

$$\|x_{k+1} - x^*\|_H \leq \left(\frac{\mathrm{cond}_2(H) - 1}{\mathrm{cond}_2(H) + 1}\right) \|x_k - x^*\|_H$$

where $\mathrm{cond}_2(H) = \lambda_{\max}(H)/\lambda_{\min}(H)$.

3. The matrix

$$H_n = S_n \otimes I_n + I_n \otimes S_n$$

where $\otimes$ is the matrix tensor product and

$$S_n := \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$$

appears when discretizing the Laplacian operator Δ on a square domain in $\mathbb{R}^2$ using Cartesian coordinates. It can be shown that the smallest eigenvalue of H_n is $\lambda_{\min} = 8\sin^2(\pi/(2n+2))$ and its largest eigenvalue is $\lambda_{\max} = 8\cos^2(\pi/(2n+2))$. Up to that point you have nothing to do for this exercise, so breathe and relax, the question is coming! For $n = 20, 200, 2000, 20000$ give an approximate number of iterations necessary for the classical steepest descent (CSD) method on quadratics with exact line-search (Algorithm I.8.1.1) to reduce the error of the initial point x_0 by a factor at least 10^{-1} in the norm $\|\cdot\|_{H_n}$ when solving for the Hessian matrix H_n. Same question for the linear conjugate gradient method (Algorithm I.4.3.4). Given the fact that at worst (and this is the most pessimistic figure) the work per iteration of CG is twice the work of CSD, which method is the most efficient on this problem? Hint: use Theorems I.8.1.2 and I.8.2.1 and the approximations for x small $\cos(x) \approx 1 - x^2/2, \ln(1+x) \approx x, \tan(x) \approx x, (1-x)/(1+x) \approx 1 - 2x$.