TMA 4180 Optimeringsteori, Midterm Test, spring 2008

February 26, 2008, 15:15-17:00.

SOLUTION

Problem 1:

Let

$$f(\mathbf{x}) = x^4 - 2x^2 + 3y^2 - 12y - 5$$

(a) Compute the gradient and the Hessian and determine the domain (that is, for which $\mathbf{x}$*) in* $\mathbb{R}^2$ *the function* f *is strictly convex.*

(b) Solve

$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}). \tag{1}$$

Solution:

(a)

$$\begin{aligned}\nabla f(\mathbf{x}) &= \left(4x^3 - 4x, 6y - 12\right), \\ \nabla^2 f &= \begin{bmatrix} 12x^2 - 4 & 0 \\ 0 & 6 \end{bmatrix}.\end{aligned} \tag{2}$$

The function f is strictly convex when both eigenvalues of $\nabla^2 f$ of λ_1 and λ_2 are positive, and since

$$\begin{aligned}\lambda_1 &= 12x^2 - 4, \\ \lambda_2 &= 6,\end{aligned} \tag{3}$$

$\mathcal{D} = \left\{(x, y); |x| > 1/\sqrt{3}\right\}$.

(b) Consider $\nabla f = 0$, that is,

$$\begin{aligned}4x^3 - 4x &= 0, \\ 6y - 12 &= 0.\end{aligned} \tag{4}$$

The solutions are obviously $(0,2)$, $(1,2)$, $(-1,2)$. The last two solutions($(1,2)$ and $(-1,2)$) are in the domain $\mathcal{D}$, are are strict minima. The function values in both points are equal, $f(x^*) = 1 - 2 + 12 - 24 - 5 = -18$, and both points are global minima.

Problem 2:

(a) Explain how x_{k+1} *in the Steepest Descent method is determined from* x_k *and* $g_k = \nabla\phi(x_k)'$ *for the quadratic model problem*

$$\begin{aligned}&\min_{x\in\mathbb{R}^n} \frac{1}{2}x'Ax + b'x, \\ &A = A',\ A > 0.\end{aligned} \tag{5}$$

(b) Why do we often see very slow convergence for the Steepest Descent method even for smooth strictly convex functions close to the solution?

(c) How many iterations are needed to solve the quadratic model problem in $\mathbb{R}^n$ for the Steepest Descent, Conjugate Gradient, and Newton's methods?

Solution:

(a) Set $\phi(x) = \frac{1}{2}x'Ax + b'x$ such that

$$g = \nabla\phi' = Ax + b, \tag{6}$$

Now, the search direction at x_k is $p_k = -g_k = -(Ax_k + b)$, and

$$x_{k+1} = x_k - \alpha_{k+1}g_k, \tag{7}$$

where

$$\alpha_{k+1} = \arg\min_{\alpha>0}(\phi(x_k - \alpha g_k)). \tag{8}$$

At the minimum we have that

$$\nabla\phi(x_{k+1})g_k = [A(x_k - \alpha g_k) + b]'g_k = (g_k - \alpha Ag_k)'g_k = 0, \tag{9}$$

from which it follows that

$$\alpha_{k+1} = \frac{g_k'g_k}{g_k'Ag_k}. \tag{10}$$

(May also be found by expanding $\frac{1}{2}(x_k - \alpha g_k)'A(x_k - \alpha g_k) + b'(x_k - \alpha g_k)$ and determine the minimum w.r.t. α).

(b) When f is a smooth function near the solution x^*, we may write $f(x^* + p) - f(x^*) \approx \frac{1}{2}p'Ap$ where $A = \nabla^2 f(x^*)$. The error $\|x_k - x^*\|_A$ decreases about as

$$\|x_{k+1} - x^*\|_A \leq \frac{1-\kappa}{1+\kappa}\|x_k - x^*\|_A, \tag{11}$$

where $\kappa = \lambda_{\max}/\lambda_{\min}$ is the condition number of A. Large matrices from applications tend to have large condition numbers, and we expect slow convergence.

(c) Unless the first search direction is an eigenvector, SD needs infinitely many iterations. For the Conjugate Gradient method the (theoretical) number of iterations is in general n, or k if the number of different eigenvalues, k, is less than n (As some have remarked, if the first direction is chosen as $-g_0$ and this happens to be an eigenvector for A, also CG converges in one iteration). Newton's Method amounts basically to solve

$$\nabla^2(x_0)(x_1 - x_0) = A(x_1 - x_0) = -\nabla(x_0)' = -(Ax_0 + b), \tag{12}$$

which leads directly to

$$Ax_1 = -b. \tag{13}$$

We thus need only 1 iteration regardless the starting point.

Problem 3

Consider a non-linear least square problem

$$\min_{x\in\mathbb{R}^n} f(x) = \min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|h(x)\|_2^2\right\} \tag{14}$$

where $h(x) \in \mathbb{R}^m$. It is easy to show that

$$\nabla f(x)' = J'(x)h(x), \tag{15}$$

$$\nabla^2 f(x) = J'(x)J(x) + \sum_{i=1}^{m} h_i(x)\nabla^2 h_i(x), \tag{16}$$

where $J = \{\partial h_i/\partial x_j\}$. Explain briefly the ideas behind the Gauss-Newton and Levenberg-Marquardt methods. When do we expect the methods to work well?

Solution:

Both methods use $J'(x)J(x)$ as an approximation for $\nabla^2 f(x)$.

Gauss-Newton is a *line search method* based on Newton's formula for the search direction,

$$p_{k+1} = -\left(J'(x_k)J(x_k)\right)^{-1} J'(x_k)h(x_k). \tag{17}$$

(This actually solves the linear least square problem $\min_p \|h(x_k) + J(x_k)p\|_2^2$).

The Levenberg-Marquardt method is a *trust region method* where

$$m_k(p) = f(x_k) + h'(x_k)J(x_k)p + \frac{1}{2}p'J'(x_k)J(x_k)p. \tag{18}$$

is used as the quadratic approximation.

The methods work well when the first term on the RHS in Eqn. 16 dominates. This typically occurs when the problem is almost linear ($\nabla^2 h_i$ is small), or h_i is small near at the solution.

Problem 4:

(In the original problem set, the second constraint was written $c_2(\mathbf{x}) = y - x + 1$. The derivation of the solution now ($x^* = (1,0)$) is quite similar and equally hard).

Let

$$f(\mathbf{x}) = x^2 - 4x + y^2 + 2y, \tag{19}$$

$$c_1(\mathbf{x}) = y - (x-1)^2 + 1, \tag{20}$$

$$c_2(\mathbf{x}) = y - x - 1\ ,\ \mathbf{x} = (x, y) \in \mathbb{R}^2, \tag{21}$$

and

$$\Omega = \{x\ ;\ c_1(\mathbf{x}) \geq 0, c_2(\mathbf{x}) \geq 0\}. \tag{22}$$

(a) Show that f and Ω are convex. Is it possible for the problem

$$\min_{\mathbf{x}\in\boldsymbol{\Omega}} f(\mathbf{x}) \tag{23}$$

to have more than one solution? Determine a solution graphically by making a sketch.

(b) State the KKT equations for the problem in (a) and show that the equations are satisfied at the solution. Will an x_0 satisfying the KKT-equations in this particular case necessarily be a minimum?

Solution:

(a) The function f is strictly convex everywhere because $\nabla^2 f = 2I_{2\times 2} > 0$. The constraint $c_2(\mathbf{x}) \geq 0$ defines a half-plane, and $c_1(\mathbf{x}) > 0$ may be identified as the interior of the parabola

$$y = (x-1)^2 - 1. \tag{24}$$

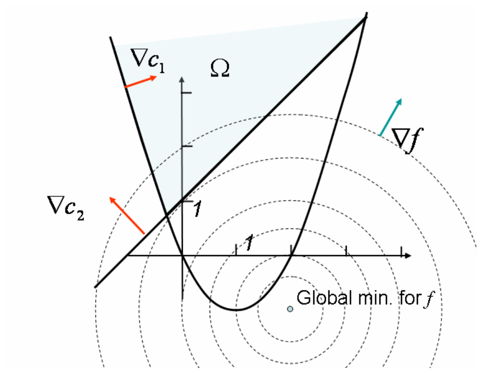

Figure 1: The feasible domain Ω and some contours of $f(\mathbf{x}) = \text{const}$. These are circles centred at $(2, -1)$.

Both these domains are clearly convex, and so is therefore their intersection.

This problem will (at most) have one solution since f is strictly convex and Ω is convex. Also, it clearly *has* a solution since $f(x) \underset{\|x\|\to\infty}{\longrightarrow} \infty$. Figure 1 shows a sketch. The global minimum (outside Ω) for f occurs at (2,-1). The graph indicates that the minimum in Ω is near $\mathbf{x}_0 = (0, 1)$. Noting that the contours are circles and $c_2(\mathbf{x}) = 0$ has a 45° slope, it is obvious that $(0, 1)$ is the solution.

(b) The complete set of KKT equations consists of the gradient of the Lagrangian,

$$\begin{aligned}\nabla\mathcal{L}(\mathbf{x}, \lambda) &= \nabla f(\mathbf{x}) - \lambda_1 \nabla c_1(\mathbf{x}) - \lambda_2 \nabla c_2(\mathbf{x}) \\ &= [2x - 4, 4y] - \lambda_1 [-2(x-2), 1] - \lambda_2 [-1, 1] = 0,\end{aligned} \tag{25}$$

(two equations) and the additional conditions

$$\begin{aligned}\lambda_i c_i(\mathbf{x}) &= 0, \\ \lambda_i &\geq 0, \\ c_i(\mathbf{x}) &\geq 0, \ i = 1, 2.\end{aligned} \tag{26}$$

The conditions provide 2 additional equations,

$$\begin{aligned}\lambda_1 \left(y - (x-1)^2 + 1\right) &= 0, \\ \lambda_2 (y - x + 1\ ) &= 0.\end{aligned} \tag{27}$$

along with the inequalities.

Now, since $\nabla c_2(\mathbf{x}_0) = (-1, 1)$ and $\nabla f(\mathbf{x}_0) = (2 \cdot 0 - 4, 2 \cdot 1 + 2) = (-4, 4)$, we have

$$\nabla f(\mathbf{x}_0) = 4 \nabla c_2(\mathbf{x}_0), \tag{28}$$

which looks promising since we may set $\lambda_1 = 0$ and $\lambda_2 = 4 > 0$:

$$\nabla f(\mathbf{x}_0) = 0 \cdot \nabla c_1(\mathbf{x}_0) + 4 \cdot \nabla c_2(\mathbf{x}_0), \tag{29}$$

We also need the check the constraints and observe that $c_2(\mathbf{x}_0)$ is active, whereas $c_1(\mathbf{x}_0) > 0$, that is, valid, but not active.

Yes, a point satisfying the KKT-equations will in fact be a global minimum because of the "convexity theorem": f as well as $-c_1$ and $-c_2$ (even c_2) are convex.