C6.2/B2. Continuous Optimization

Problem Sheet 3

Part C and OMMS: Please hand-in for marking Problems 1(i)–(iii), 4 and 5(a)–(c).
MMSC MSc: Please hand-in Problems 1(i)–(iii) and 4 for the fourth intercollegiate class, and Problem 5(a)–(e) for the fifth intercollegiate class.
The other problems are optional for everyone.

1. (i) Show that if $\nabla f(x)$ is the gradient of the quadratic function $f(x) = g^T x + \frac{1}{2}x^T Bx$, then for any s,

$$Bs = \gamma, \text{ where } \gamma = \nabla f(x+s) - \nabla f(x).$$

   (ii) Consider a general nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$, and apply a quasi-Newton generic linesearch method (GLM) for its minimization that generates iterates x^k. Suppose that B^k is a given symmetric matrix (that approximates the Hessian of f on the kth iteration), and that $B^{k+1} = B^k + \beta vv^T$ is a rank-one correction for which the secant condition

$$B^{k+1}\delta^k = \gamma^k, \text{ where } \gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k) \text{ and } \delta^k = x^{k+1} - x^k, \tag{1}$$

   is required to hold. Show that the resulting B^{k+1} is unique, and give its form; state any assumptions you need to make.

   (iii) Now suppose that (in the same set-up as in (ii))

$$B^{k+1} = B^k + \theta\gamma^k(\gamma^k)^T + \beta vv^T \tag{2}$$

   is a rank-two correction to a given symmetric matrix B^k for which (1) holds. Find an expression for θ, β and v in terms of

$$\phi = 1 - \theta(\gamma^k)^T\delta^k,$$

   where $\delta^k = x^{k+1} - x^k$; once again state any assumptions you require. What special cases occur in the formula you have derived when $\phi = 0$ and $\phi = 1$?

   (iv) Show that if B is symmetric positive definite and $\delta^T\gamma > 0$, then the BFGS update

$$B^+ = B + \frac{\gamma\gamma^T}{\gamma^T\delta} - \frac{B\delta\delta^T B}{\delta^T B\delta}$$

   is also positive definite.

2. (i) Consider the Sherman-Morrison-Woodbury formula

$$(B + UV^T)^{-1} = B^{-1} - B^{-1}U(I + V^T B^{-1}U)^{-1}V^T B^{-1},$$

   where B is a $n \times n$ invertible matrix, U and V are $n \times m$ matrices with $m \leq n$, I is the $m \times m$ identity matrix, and where we assume that the matrix $I + V^T B^{-1}U$ is invertible. Show that the formula holds true by multiplying it with $(B + UV^T)$.

   (ii) Can it happen that B and $B + UV^T$ are invertible but $I + V^T B^{-1}U$ is not, so that the formula is not applicable? (Hint: did you ever use the condition $m \leq n$ in part (i)?)

   (iii) Consider applying a quasi-Newton generic linesearch method (GLM) to minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$ and assume that we use the Symmetric rank one or BFGS formula to update the Hessian approximation. Using (i), justify why the cost per iteration of these methods is $\mathcal{O}(n^2)$ (as opposed to the Newton iteration whose cost is $\mathcal{O}(n^3)$). By cost per iteration we mean the number of floating point operations (flops) per iteration which are the number of simple arithmetic operations that are performed (namely, $+$, $-$, $\times$ and division).

3. Let $r : \mathbb{R}^2 \to \mathbb{R}^4$, $x = (x_1\ x_2)^T$, $r_i(x) = e^{x_1 + t_i x_2} - y_i$, $i \in \{1, 2, 3, 4\}$. Suppose that

$$t_1 = -2,\ t_2 = -1,\ t_3 = 0,\ t_4 = 1,\ y_1 = 0.5,\ y_2 = 1,\ y_3 = 2,\ y_4 = 4.$$

Let $f(x) = \frac{1}{2} r(x)^T r(x)$. Note that $f(x^*) = 0$ at $x^* = (\log 2\ \ \log 2)^T$. Calculate one iteration of the Gauss-Newton method (without linesearch) and one iteration of Newton's (applied to f, without linesearch) starting from $x^0 = (1\ 1)^T$. Do the same if the values of y_1 and y_4 are changed to 5 and -4, respectively.

Comment: note that you should get that the Gauss-Newton step takes us closer to the solution than the Newton step for the first given choice of y values.

4. Consider the overdetermined system $r : \mathbb{R} \to \mathbb{R}^2$, $r(x) := (r_1(x)\ \ r_2(x))^T = 0$, with $x \in \mathbb{R}$, where

$$r(x) := (x + 1\ \ \lambda x^2 + x - 1)^T, \quad \text{where } \lambda \in \mathbb{R} \text{ is a constant.}$$

Consider the least-squares objective $f(x) = \frac{1}{2} r(x)^T r(x)$; compute its gradient and Hessian, and show that $x^* = 0$ is a stationary point of f for all λ, and it is a minimizer of f if $\lambda < 1$.
Apply Newton's method and Gauss-Newton method, without linesearch (i.e., with stepsize $\alpha^k = 1$, for all k) to $f(x)$: compute the Newton iterate in the form $x^{k+1} = \psi_N(x^k)$ and the Gauss-Newton, $x^{k+1} = \psi_{GN}(x^k)$.
Consider now the functions $\psi_N(x)$ and $\psi_{GN}(x)$. Show that $\psi_N'(0) = 0$ for all λ, and show that this implies that the rate of convergence of Newton is at least quadratic (asymptotically). Show that $\psi_{GN}'(0) = \lambda$ and that this implies that convergence of Gauss-Newton is linear if $|\lambda| < 1$, and does not occur if $|\lambda| > 1$.

Comment: Note that $r(x^*) = r(0) = [1, -1]^T \neq 0$, and so we have a nonzero residual least-squares problem (and hence, the Gauss-Newton approximation to the Hessian is poor, yielding only linear convergence, or even nonconvergence of Gauss-Newton method).

5. Solve the trust-region subproblem

$$\min_{s \in \mathbb{R}^3} s^T g + \frac{1}{2} s^T H s \quad \text{subject to} \quad \|s\| \leq \Delta$$

in the following cases below. [*Hint*: use the characterization of the global minimizer of the trust-region subproblem given in the lectures and solve the resulting nonlinear equation in λ analytically (rather than using Newton's method or the secular equation).]

(a)

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad g = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \Delta = 2$$

(b)

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad g = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \Delta = 5/12$$

[*Hint*: a root of the nonlinear equation $\frac{1}{(1+\lambda)^2} + \frac{1}{(2+\lambda)^2} = \frac{25}{144}$ is $\lambda = 2$.]

(c)

$$H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad g = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \Delta = 5/12$$

(d)

$$H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \Delta = 1/2$$

(e)

$$H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \Delta = \sqrt{2}.$$

6. Consider the trust-region subproblem

$$\min_{s\in\mathbb{R}^n} s^T g + \frac{1}{2} s^T H s \text{ subject to } \|s\| \leq \Delta. \tag{3}$$

Find and sketch the solution of problem (3) with data

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \text{ and } g = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

as a function of the trust-region radius Δ. In which direction does the solution point as Δ shrinks to zero? How does (the Lagrange multiplier) λ for the trust-region constraint depend on the trust-region radius? For what value of the radius does the solution become unconstrained?

7. Instead of the second-order Taylor model, consider using an approximate second-order model in the trust-region subproblem, in each iteration of the GTR algorithm. Thus, the step s^k is assumed to be computed by (approximately or exactly) solving the following trust-region subproblem

$$\min_{s\in\mathbb{R}^n} m_k(s) := f(x^k) + s^T \nabla f(x^k) + \frac{1}{2} s^T B^k s \quad \text{subject to} \quad \|s\| \leq \Delta_k, \tag{4}$$

where B^k is an $n \times n$ symmetric matrix. By following the same steps as in the proof of global convergence of the GTR method given in the lectures, prove the (liminf) global convergence of the GTR method where s^k is computed from the subproblem (4). What conditions do the matrices B^k need to satisfy for this convergence to hold? (*Hint*: note that you can define the Cauchy point in a similar fashion to the lectures.)