Mihai Anitescu, 03/04/2010

Stat 310, Part II, Optimization. Homework 4.

1. Consider the conjugated gradients method, applied to the linear problem $Ax = b$, with $A \in \mathbb{R}^{n\times n}$ positive definite. We denote by d^i the A-conjugate search directions, by x^i the iterations, and by $r^i = Ax^i - b$ the residuals. Assume that matrix A has exactly p distinct eigenvalues, Prove that
   a. The minimal polynomial of A has at most degree p.
   b. Prove that the linear space (called the Krylov space of A generated by the vector b) $Span\{b, Ab, \ldots, A^k b\}$ has dimension at most p for any k (hint: use Cayley-Hamilton).
   c. Show that, as long as $r^k \neq 0$ $Span\{b, Ab, \ldots, A^k b\} = Span\{r^0, Ar^0, \ldots, A^k r^0\} = Span\{d^0, d^1, \ldots, d^k\}$
   d. Use this observation to show that we must necessarily have $r^p = 0$, and conclude that the conjugated gradient method must end in at most p steps.
2. Use the Sherman-Morrison Lemma (A28, N & W), to compute the Hessian update formula (6.19) from the Hessian inverse update formula (BFGS, 6.17).
3. Implement the damped BFGS method (Procedure 18.2, N & W) to compute a matrix B^k that can be used with the line search approach from Homework 2 *without computing any second-order derivative information* and apply to Fenton's function. In the definition of that procedure, use $s_k = x_{k+1} - x_k; \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ (the notations in 18.13 refer to constrained optimization which we will talk about in the next lecture). **Estimate the rate of convergence**. What seems to be the rationale for modifying the BFGS formula, (6.19)?
4. Implement algorithm 7.2, the CG-Steihaug method. To be on the safe side, use the definition from the book (like discussed in class, the variants of the conjugated gradients methods have to do with the choice of the sign for residual in its definition or in the introduction to the Gramm-Schmidt step).
   a. Convince yourself first that your implementation is correct, by applying it first to Fenton's function.
   b. Apply it to the extended Rosenbrock function.

   $$f(x) = \sum_{i=1}^{\frac{n}{2}} \left[\alpha\left(x_{2i} - x_{2i-1}^2\right)^2 + \left(1 - x_{2i-1}\right)^2 \right]$$

   Pick several values for α, n and start at $(-1, -1, -1, \ldots -1)^T$ and observe the behavior of your program. (For example, report the number of matrix-vector multiplications needed before convergence to some tolerance, for various values of α, n ,). The solution should be $(1,1,1,\ldots 1)$