Assignment 2 (30 pts)

Due March 31, 2021

Q1 (**2pts**) Derive expressions for the gradient vector and Hessian matrix of the 2-D Rosenbrock function $f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$. Write a MATLAB/Python code to compute the first five iterates of the Newton method starting from the initial guess $[-1, -2]^T$. Plot the contours of the function on the domain [-1.5;2;-3;3] and show the iterates of the Newton method on the same plot. (Do not use symbolic computation toolboxes to compute the gradients and Hessian).

Q2 (**10pts**) Implement a quasi-Newton algorithm with the BFGS Hessian approximation scheme. Use back-tracking line search based on Armijo's sufficient decrease condition in your implementation. Evaluate the performance of your algorithm on the Rosenbrock test function (with $n = 2$ and $n = 5$) using 20 randomly generated initial guesses in the interval $[-5, +5]^n$. Present convergence trends of the objective function and the ℓ_2 norm of the gradient and compare the results obtained to the steepest descent algorithm with back-tracking line search.

Q3 (**7pts**) In 1988, Barzilai and Borwein (BB) published a very intriguing gradient-based optimization algorithm.[1] Similar to the steepest descent method, the iterates are computed as $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$. However, the step length is computed so that $\alpha_k \nabla f(\mathbf{x}_k) \approx [\nabla^2 f(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k)$. The BB algorithm is therefore inspired by the Newton method but it does not use the Hessian (similar to quasi-Newton algorithms).

Using the secant equation to estimate the step-length of the BB algorithm, we have

$$\alpha_k = \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{y}_k},$$

where $\mathbf{s}_k = \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{y}_k = \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1})$.

Implement the BB algorithm for minimization of the convex quadratic $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} - \mathbf{x}^T \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is SPD and $\mathbf{b} \in \mathbb{R}^n$. The matrix $\mathbf{A}$ and vector $\mathbf{b}$ for the test case can be found in ConvexQuadratic.mat. Note that $\mathbf{A}$ is a sparse 341×341 matrix for this test case.

Study the convergence behaviour of the BB algorithm for this test case. What observations can you draw from the convergence trends of f and $||\nabla f||_2$?

[1] J. Barzilai and J. Borwein, "Two-point step size gradient method," *IMA Journal of Numerical Analysis*, Vol. 8, 1988, pp. 141–148

Q4 (**3pts**) Formulate and implement the steepest descent method[2] to minimize convex quadratics of the form $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is SPD and $\mathbf{b} \in \mathbb{R}^n$. Compare the performance of the steepest descent method to the Barzilai-Borwein algorithm for the test problem described in Q3.

Q5 (**8pts**) Consider a gradient-based optimization algorithm for solving the unconstrained minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}),$$

wherein the search direction is given by $\mathbf{p}_k = -\mathbf{g}_k = -\nabla f(\mathbf{x}_k)$ when the iteration counter $k = 0$. When the iteration counter $k = 1$, the search direction is set to $\mathbf{p}_k = -\mathbf{g}_k + \mathbf{p}_{k-1}(\mathbf{g}_k^T\mathbf{g}_k)/(\mathbf{g}_{k-1}^T\mathbf{g}_{k-1})$. For subsequent iterations with $k > 1$, the search direction is computed as:

$$\mathbf{p}_k = -\alpha_k\mathbf{g}_k + \beta_k\mathbf{p}_{k-1} + \gamma_k\mathbf{p}_{k-2}, \quad k > 1,$$

where $\alpha_k, \beta_k, \gamma_k \in \mathbb{R}$.

- (**4pts**) Derive an approach to estimate the parameters α_k, β_k and γ_k such that the search direction is close (with respect to some appropriate norm that you are free to choose) to the Newton search direction. Your approach must not assume that any of the parameters are zero and the computational cost must be $\mathcal{O}(n^2)$. In addition, your approach must provide a valid search direction even when the Hessian is not positive definite. Clearly show all working steps in your derivation.
- (**4pts**) Derive an approach to estimate the parameters α_k, β_k and γ_k in the search direction using only the first-order derivatives of f. Your approach must not use finite difference or complex step approximations to the second-order derivatives of f. In addition, your approach must not assume that any of the parameters are zero. Clearly show all working steps in your derivation and explain if your method will always provide a valid direction of descent.

Submission guidelines:

- Your assignment must be submitted in pdf format to Quercus. Reports submitted in other formats such as Microsoft Word are not acceptable.
- You are permitted to use Python for this assignment. Please mention the Python version you are using in your report.
- Submit MATLAB/Python codes archived in tar, zip or gzip format along with input files used to generate results. Do NOT use proprietary data compression formats such as rar. Please verify the integrity of your tar/zip/gzip file before uploading.

[2] Recall that when the objective function is quadratic, line search can be carried out exactly.

- Include a README file that describes how the codes must be run.
- Clearly explain all working steps in your answers.
- All submissions must strictly comply with the AER1415 Collaboration Policy posted on Quercus.