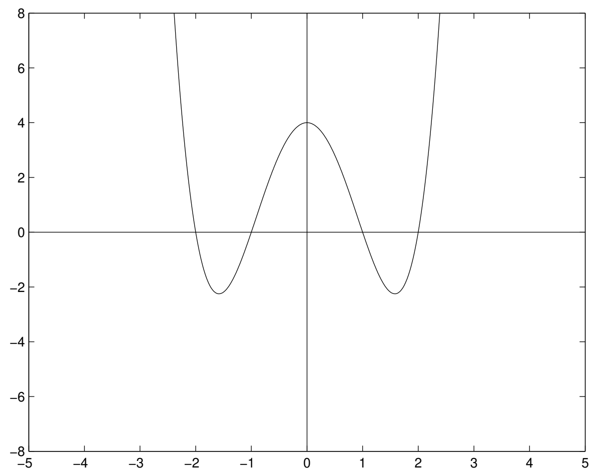
Study Homework Questions 2 Numerical Optimization Fall 2023

Problem 2.1

Let $f(x) : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = x^4 - 5x^2 + 4$$

and consider applying Newton's method for optimization. Here Newton's method refers to the basic form where the step size is 1 and nothing is done to alter the Hessian to guarantee positive definiteness. Note that $f(x)$ is a scalar function of a scalar argument and has the form



(i) What are the values of x that are local minimizers or local maximizers of $f(x)$. Justify your answers.

(ii) Find the value $\beta > 0$ such that $f(x)$ has negative curvature for $-\beta < x < \beta$, and positive curvature outside the interval, i.e., for $x < -\beta$ or $x > \beta$.

(iii) What happens to the Newton step at $x = \beta$?

(iv) Determine $\mu(x) : \mathbb{R} \to \mathbb{R}$ such that the step of Newton's method applied to $f(x)$ can be written as $x_{k+1} = \mu(x_k)x_k$.

(v) Find the value of $\alpha \in \mathbb{R}$ such that $\beta > \alpha > 0$ and Newton's method cycles and does not converge when $x_0 = \alpha$ or $x_0 = -\alpha$. That is, $-\alpha = \mu(\alpha)\alpha$ and $\alpha = -\mu(-\alpha)\alpha$.

(vi) Show that if $-\alpha < x < \alpha$ then

$$|\mu(x)| < 1$$

(vii) Show that if $-\alpha < x_0 < \alpha$ is the initial point for Newton's method then there is a constant $0 < \gamma < 1$ (possibly dependent on x_0 but independent of k) such that

$$|x_{k+1}| < \gamma |x_k|$$

and therefore $x_k \to 0$.

(viii) It can be shown that if $x_0 > \sigma > 0$ where σ is the rightmost local minimizer of $f(x)$ then $x_k \to \sigma$ for Newton's method. Using this fact and those above, show that it is possible to choose $-\beta < x_0 < -\alpha$ so that $x_k \to \sigma$ for Newton's method.

(ix) Implement Newton's method and demonstrate the convergence behavior determined above.

Problem 2.2

Consider solving the unconstrained minimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ using a Quasi-Newton method.

2.2.a

Suppose the Quasi-Newton method guarantees that the symmetric matrix $B_k \in \mathbb{R}^{n \times n}$ used to define the local quadratic model of $f(x)$ is symmetric positive definite.

Show that one can only find a symmetric positive definite matrix B_{k+1} that satisfies the secant condition

$$B_{k+1} s_k = y_k$$

if s_k and y_k satisfy a simple constraint. (**Hint:** Consider the angle between s_k and y_k.)

2.2.b

Consider solving the minimization problem by Newton's method ($\alpha_k = 1$) and BFGS using the H_k update form. Suppose the problem size n is moderate so all linear systems are solved using Cholesky factorization and there is no sparsity to exploit in any matrix operation.

i. Describe the computational complexity of one step of Newton's method.

ii. Describe the computational complexity of one step of BFGS using the H_k update form.

iii. Assuming the search for α_k satisfying Wolfe's conditions is not computationally significant, discuss how much the number of iterations must be reduced to justify using Newton's method.

Problem 2.3

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ is a cost function for a minimization problem. Suppose the iteration $x_{k+1} = x_k + \alpha_k p_k$ is such that at every step α_k satisfies the Wolfe conditions

$$\begin{gathered} f(x_{k+1}) \leq f(x_k) + \gamma_1 \alpha_k \nabla f_k^T p_k \\ \nabla f(x_{k+1})^T p_k \geq \gamma_2 \nabla f(x_k)^T p_k \\ 0 < \gamma_1 < \gamma_2 < 1. \end{gathered}$$

Consider an affine change of variable $x = Sz + c$ where $S \in \mathbb{R}^{n \times n}$ is a nonsingular matrix and $c \in \mathbb{R}^n$ and let

$$\tilde{f}(z) = f(x) = f(Sz + c)$$

(2.3.a) Determine the relationships between the gradients $\nabla f(x)$ and $\nabla \tilde{f}(z)$ and the Hessians $\nabla^2 f(x)$ and $\nabla^2 \tilde{f}(z)$.

(2.3.b) Show that given the x_k iterates satisfy the Wolfe conditions then so do the iterates $z_{k+1} = z_k + \tilde{\alpha}_k \tilde{p}_k$.

Problem 2.4

Suppose $x_{k+1} = \phi(x_k)$ defines an iteration on $\mathbb{R}^n$. We know that $f(x_{k+1}) < f(x_k)$ is not sufficient for convergence to a local minimizer of a cost function $f(x)$. Demonstrate this by creating an example of a convex cost function $f(x) : \mathbb{R}^2 \to \mathbb{R}$ that has a single stationary point, $x^* = 0 \in \mathbb{R}^2$, that is a global minimum and an iteration $x_{k+1} = \phi(x_k)$ that guarantees $f(x_{k+1}) < f(x_k)$ but that does not converge to the minimum x^*. (In fact, you can generate an example for which $\|x_k - x^*\| > \gamma > 0$.)

Problem 2.5

The potential energy, $E(r)$, of a diatomic molecule as a function of distance, $r \in \mathbb{R}$, is given by

$$E(r) = \frac{B}{r^2} - \frac{zs^2}{r}$$

where $B > 0$, $s > 0$, and $z \geq 1$ are all real parameters. Find the distance r at which the potential energy is a minimum.

Problem 2.6

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the cost function

$$f(x) = \xi_1^2 - 5\xi_1\xi_2 + \xi_2^4 - 25\xi_1 - 8\xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

(2.6.a) Find $\nabla f(x)$ and $\nabla^2 f(x)$.

(2.6.b) Show that $x_*^T = (20, 3)$ is the unique strict, i.e., nondegenerate, local minimizer of $f(x)$.

(2.6.c) Start with $x_0^T = (0, 0)$ and iterate using the Steepest Descent method with fixed stepsize $\alpha_k = \alpha$. Try different α values and determine how many steps are required to get x_* to approximately six digits in each component.

(2.6.d) As mentioned in class, in the limit the behavior of Steepeset Descent converging to a nondegenerate minimizer is governed by reasoning with the local quadratic models in the neighborhood of the minimizer. Consider solving the problem again with the same initial condition but with the change of variables $\xi_2 = 7\zeta$, i.e., $f(\xi_1, \xi_2) = f(\xi_1, \zeta/7))$ allows you to reuse the code already written. Note that you will have to modify $\nabla f(x)$ and $\nabla^2 f(x)$. You should also consider a different value of α for the rescaled problem. Explain why the behavior of the rescaled iteration is better than the original scaling.

Problem 2.7

For each of the following cost functions find all minimizers, if they exist and, if they do not, show why.

1. The cost function $f : \mathbb{R}^3 \to \mathbb{R}$

$$f(x) = 2\xi_1^2 + 3\xi_2^2 + 4\xi_3^2 - 8\xi_1 - 12\xi_2 - 24\xi_3 + 110, \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}$$

2. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = (4\xi_1^2 - \xi_2)^2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

3. The cost function $f : \mathbb{R}^3 \to \mathbb{R}$

$$f(x) = \xi_1^4 - 3\xi_1^2 + \xi_2^2 + 2\xi_2\xi_3 + 2\xi_3^2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}$$

4. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = 2\xi_2^3 - 6\xi_2^2 + 3\xi_1^2\xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

5. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = \xi_1^4 + 2\xi_1^2\xi_2 + \xi_2^2 - 4\xi_1^2 - 8\xi_1 - 8\xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

6. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = (\xi_1 - 2\xi_2)^4 + 64\xi_1\xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

7. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = 2\xi_1^2 + 3\xi_2^2 - 2\xi_1\xi_2 + 2\xi_1 - 3\xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

8. The cost function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = \xi_1^2 + 4\xi_1\xi_2 + \xi_2^2 + \xi_1 - \xi_2 \quad x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$