ECE 490 (Introduction to Optimization) – In-Class Problem Discussions

03/01/2022

Problem 1. Consider the following quadratic function:

$$f(x) = f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 2x_2^2 - 5x_2 + 2$$

(a) If you use the steepest descent method to minimize the above function, how to choose the stepsize? Provide some reasons. What type of convergence behaviors will you get?

(b) If Newton's method is applied, what type of convergence behavior will you get?

(c) Find the minimum and maximum of f over $\mathbb{R}^2$ if they exist.

(d) In general, what are the cons/pros of the steepest descent method when compared with Newton's method?

Solutions

(a) We can choose the stepsize as $\alpha = \frac{1}{\lambda_{\max}}$ or $\alpha = \frac{2}{\lambda_{\max}+\lambda_{\min}}$ where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of the Q matrix which is $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$. The method converges linearly. We can also use line search or Armijo rule.

(b) Newton's method converges in one step for this strongly-convex quadratic function.

(c) The maximum does not exist. The minimum can be solved by setting the gradient to be 0. The minimum is achieved at $(x_1, x_2) = (-5/6, 5/3)$.

(d) For this problem, Newton's method converges in one step. This is faster than the steepest descent method. However, Newton's method requires the matrix inversion step, and this is quite expensive. So the per step cost for Newton's method is higher. In general, Newton's method has faster local convergence but may diverge if initialized from some place far from the optimal point.

Problem 2. True or False. Provide reasons.

(a) If S_1 and S_2 are two convex sets, then $S_1 \cup S_2$ is convex.

(b) If S_1 and S_2 are two convex sets, then $S_1 \cap S_2$ is convex.

(c) If f and g are both convex, then $f(g(x))$ is also convex.

(d) Gradient descent algorithm always converges to a local optimizer for a smooth function.

(e) For a strictly convex function, Newton's method always converges to a minimizer, starting from any point within the domain of this function.

Solutions

(a) False. Consider $S_1 = [0, 1]$, and $S_2 = [2, 3]$. Both are convex. The union is not convex.

(b) True. Take two points x and y which are both in $\in S_1 \cap S_2$. Since x and y are both in S_1 which is convex, the line segment between x and y has to be in S_1. Similarly, the line segment between x and y has to be in S_1, and hence has to be in $S_1 \cap S_2$. Therefore, $S_1 \cap S_2$ is convex.

(c) False. Think $f(x) = -x$ and $g(x) = x^2$. We have $f(g(x)) = -x^2$ which is not convex.

(d) False. With a reasonable stepsize (i.e. $\alpha = \frac{1}{L}$), the gradient descent method can find a stationary point. However, having zero gradient does not mean that the point is a local optimizer. It can be a saddle point.

(e) False. Think about the function $f(x) = \sqrt{1 + x^2}$. This is a strictly convex function. Newton's method iterates as $x_{k+1} = -x_k^3$. For $|x_0| > 1$, Newton's method diverges. In general, Newton's method in pure form ($\alpha = 1$ does not have good global guarantees.

Problem 3. Convergence under PL condition: Suppose f is L-smooth and also satisfies the PL condition:

$$f(x) - f(x^*) \leq \frac{1}{2\mu}||\nabla f(x)||^2,$$

where x^* is the unique global min of f. If we apply the steepest descent method to minimize f, does $f(x_k)$ converge to $f(x^*)$? Try to prove a linear convergence bound in the following form:

$$f(x_k) - f(x^*) \leq \rho^k C$$

where $0 < \rho < 1$ and C are fixed constants. What type of stepsize shall we use?

Solutions We have $x_{k+1} - x_k = -\alpha \nabla f(x_k)$. By smoothness, we have

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \nabla f(x_k)^T (x_{k+1} - x_k) + \frac{L}{2}||x_{k+1} - x_k||^2 \\ &= f(x_k) - \left(\alpha - \frac{L\alpha^2}{2}\right) ||\nabla f(x_k)||^2 \end{aligned}$$

From the PL inequality, we have $-||\nabla f(x)||^2 \leq -2\mu(f(x_k) - f(x^*))$. If $\alpha - \frac{L\alpha^2}{2} > 0$, we have

$$f(x_{k+1}) - f(x^*) \leq f(x_k) - f(x^*) - 2\mu \left(\alpha - \frac{L\alpha^2}{2}\right)(f(x_k) - f(x^*)) = (1 - 2\mu\alpha + \mu L\alpha^2)(f(x_k) - f(x^*))$$

Hence we have the bound $f(x_k) - f(x^*) \leq (1 - 2\mu\alpha + \mu L\alpha^2)^k (f(x_0) - f(x^*))$. We can choose the stepsize as $\alpha = \frac{1}{L}$ and get the linear convergence rate bound $f(x_k) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)^k (f(x_0) - f(x^*))$.