Problem 1: Proving what Newton method is up to

Problem Statement

Newton's method minimizes the local quadratic approximation of a function, if its Hessian at the current iterate is positive definite. We showed that Newton's method indeed finds a *local* minimizer of the quadratic approximation. Here, we show that it finds the *global* minimizer through the following steps.

a

If Q is any PD matrix, show that:

$$\frac{1}{2}x^TQx + c^Tx = \frac{1}{2}(x + Q^{-1}c)^TQ(x + Q^{-1}c) - \frac{1}{2}c^TQ^{-1}c$$

b

Using the prior result, argue that the function $\frac{1}{2}x^TQx + c^Tx$ is minimized *globally* at $x^* = -Q^{-1}c$.

c

Recall that for any function $f : \mathbb{R} \to \mathbb{R}$, its local quadratic approximation at x^k is given by:

$$f^q(x) \coloneqq f(x^k) + [\nabla f(x^k)]^T(x - x^k) + \frac{1}{2}(x - x^k)^TH(h^k)(x - x^k)$$

Assume the Hessian $H(x^k)$ is positive definite. Utilize the prior result to conclude that the *global* minimizer of f^q is given by $x^k - [H(x^k)]^{-1}\nabla f(x^k)$. Notice that the minizer is indeed x^{k+1}, as defined by Newton's method.

Solution

a

Simply expand the right side:

$$= \frac{1}{2}(x^TQ(x + Q^{-1}c) + c^T(Q^{-1})^TQ(x + Q^{-1}c)) - \frac{1}{2}c^TQ^{-1}c$$

$$= \frac{1}{2}(x^TQx + x^TQQ^{-1}c + c^T(Q^{-1})^TQx + c^T(Q^{-1})^TQQ^{-1}c) - \frac{1}{2}c^TQ^{-1}c$$

Noting that (per the instructor comment on Piazza) $Q^T = Q$ and that $Q^{-1} \succ 0$:

$$= \frac{1}{2}(x^TQx + x^Tc + c^Tx + c^TQ^{-1}c) - \frac{1}{2}c^TQ^{-1}c$$

Finally, note that $c^Tx = x^Tc \implies c^Tx + x^Tc = 2c^Tx$ and thus:

$$= \frac{1}{2}x^TQx + c^Tx + \frac{1}{2}c^TQ^{-1}c - \frac{1}{2}c^TQ^{-1}c = \frac{1}{2}x^TQx + c^Tx$$

$\square$

b

By definition, if $Q \succ 0$, $x^T Qx > 0 \forall x \neq 0$ and $x^T Qx = 0 \iff x = 0$. As such, $\min_x x^T Qx = 0$ and $\arg\min_x x^T Qx = 0$. Given the prior result, if we substitute in $x := x^* = -Q^{-1}c$ we find:

$$\frac{1}{2}x^T Qx + c^T x = \frac{1}{2}(x + Q^{-1}c)^T Q(x + Q^{-1}c) - \frac{1}{2}c^T Q^{-1}c$$

$$= \frac{1}{2}(-Q^{-1}c + Q^{-1}c)^T Q(-Q^{-1}c + Q^{-1}c) - \frac{1}{2}c^T Q^{-1}c$$

$$= \frac{1}{2}(0)^T Q(0) - \frac{1}{2}c^T Q^{-1}c = -\frac{1}{2}c^T Q^{-1}c$$

which is a constant. Note that the term that was dropped cannot be negative. Thus, this result must be the global minimum of the function. □

c

TODO: rework, made mistake First, note that, for a given x^k, $f(x^k)$ is a constant which we will call a^k. Substituting $x = x^k - [H(x^k)]^{-1}\nabla f(x^k)$ we see that the x^k's cancel leaving us with:

$$f^q(x) = a^k + [\nabla f(x^k)]^T [H(x^k)]^{-1}\nabla f(x^k) + \frac{1}{2}(-H(x^k)^{-1}\nabla f(x^k))^T H(x^k)(-H(x^k)^{-1}\nabla f(x^k))$$

$$= a^k + [\nabla f(x^k)]^T [H(x^k)]^{-1}\nabla f(x^k) + \frac{1}{2}(\nabla f(x^k)^T [H(x^k)^{-1}]^T)H(x^k)(-H(x^k)^{-1}\nabla f(x^k))$$

$$= a^k + [\nabla f(x^k)]^T [H(x^k)]^{-1}\nabla f(x^k) + \frac{1}{2}(\nabla f(x^k)^T (H(x^k)^{-1})\nabla f(x^k))$$

The constant has no effect on the location of the optimimum; i.e. $\arg\min x f(x) = \arg\min f(x) + a$ and can thus be ignore. Then, the resulting equation matches the above form and thus the provided minimizer must be the global minizer.

Problem 2: Newton's method needs a touch-up

Problem Statement

In Newton's method, if the Hessian at the current iterate $H(x^k)$ is not PD, then $-[H(x^k)]^{-1}\nabla f(x^k)$ may not be a descent directon. Then, we modify the Hessian to $H(x^k) + D^k$ where D^k is a diagonal matrix with nonnegative diagonal entries. Let us design D^k to ensure that $H(x^k) + D^k$ is PD. The following corollary of Gershgorin's circle theorem will prove useful.

Theorem 0.1. *If λ is any eigenvalue of an arbitrary matrix $A \in \mathbb{R}^{n\times n}$, then*

$$|\lambda - Aii| \leq \sum_{j\neq i} |A_{ij}|$$

for some $i = 1, \ldots, n$.

a

Using this theorem, find a sufficient condition on the diagonal entries of A s.t. all eigenvalues of A are positive.

b

Using the prior result, find a diagonal matrix D^k s.t. that all eigenvalues of $H(x^k) + D^k$ are nonnegative, and all diagonal entries of D^k are nonnegative.

c

If any eigenvalue of $H(k) + D^k$ is close to zero but positive, then it is close to being singular and its inverse is susceptible to noise. Modify your answer in the prior section to ensure that all the eigenvalues are greater than $\frac{1}{2}$.

d

The file applyNewtonMethod.m implements a basic newton method to the function

$$f(x_1, x_2) := \cos(x_1^2 - 2x_2) + \sin(x_1^2 + x_2^2)$$

starting from $x^0 = (1.2, 0.5)$. The program also draws the contour plot and the surface plot of f.

1. Verify the Hessian at the starting point is NOT PD.
2. Does the Newton method converge? If yes, does it converge to a local minimizer of f?
3. Fill in the missing code in modifyHessian.m that takes $H(x^k)$ as input and gives $H(x^k) + D^k$ as output. Utilize your condition in part (c) to design D^k. Submit your code. Using your code, compute $H(x^0) + D^0$; i.e the modiied Hessian at the starting point.
4. Uncomment the relevant lines in applyNewtonMethod.m to run the modified Newton method. Report if the algorithm converges to a local minimizer of f.

Solution

a