

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:

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

b

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

c

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

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

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 minimizer is indeed x^{k+1} , as defined by Newton's method.

Solution

a

Simply expand the right side:

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

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

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

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

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

□

b

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

$$\begin{aligned} \frac{1}{2}x^T Q x + c^T x &= \frac{1}{2}(x + Q^{-1}c)^T Q (x + Q^{-1}c) - \frac{1}{2}c^T Q^{-1}c \\ &= \frac{1}{2}(-Q^{-1}c + Q^{-1}c)^T Q (-Q^{-1}c + Q^{-1}c) - \frac{1}{2}c^T Q^{-1}c \\ &= \frac{1}{2}(0)^T Q (0) - \frac{1}{2}c^T Q^{-1}c = -\frac{1}{2}c^T Q^{-1}c \end{aligned}$$

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. \square

c

First, we need to note that $f(x^k)$ is a constant, as x^k is a parameter and will not change as a function of x . Furthermore:

$$\arg \min_x g(x) = \arg \min_x g(x) + k$$

i.e. addition of a constant does not change the location of the optimum.

So: simply let

$$c' := \nabla f, \quad x' = x - x^k, \quad Q' = H$$

We can then see it matches the form:

$$f(x^k) + \frac{1}{2}x'^T Q' x' + c'^T x'$$

Given our expression for x' , we can see

$$x'^* = x^* - x^k = x^k - H^{-1}\nabla f - x^k = -H^{-1}\nabla f = -Q'^{-1}c'$$

which is identical to the result in the prior section.¹ \square

¹Kudos to Will V. for pointing out the expression should be massaged to match the prior expression from the get-go. I had previously spent a few hours chasing my tail trying to make the expressions match *after* I substituted the value for the optimum. This was fruitless.

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 direction. 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*

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

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

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 $H(x^k) + D^k$ 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 modified 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**

Using the geometric interpretation (borrowing heavily from the Wikipedia page²), we can note that the theorem essentially describes a bounded estimate of an eigenvalue, where the center of circle is located at the value of the diagonal element in a row (A_{ii}) and its radius is the sum of the magnitude of all other elements in the row ($\sum_{i \neq j} |A_{ij}|$). In order to bound the eigenvalue to be positive, this circle must exist solely within the RHS of the complex plane. Thus: Given a square matrix $A \in \mathbb{R}^{n \times n}$, a sufficient (albeit not necessary!) condition is:

$$A_{ii} > \sum_{i \neq j} |A_{ij}| \forall i \in 1, \dots, n$$

In plain english: we have to move the bounding circle via a translation of its center (the diagonal element of that row) to be larger than the radius. If we only require non-negative eigenvalues, the relationship could be relaxed to be “greater than or equal to”.

b

$$D_{ii} + A_{ii} > \sum_{i \neq j} |A_{ij}| \implies D_{ii} > \sum_{i \neq j} |A_{ij}| - A_{ii}$$

c

$$D_{ii} > \sum_{i \neq j} |A_{ij}| - A_{ii} + \frac{1}{2}$$

d

Note: I used GNU Octave³ instead of MATLAB because I’m too lazy to get it installed and working on my machine (plus I’m not a fan of closed-source proprietary languages).

1. I checked `all(eig(Hk)>0)` for iterate 0, which reports false. The matrix is not PD.
2. It does after 4 iterations. It is not a local minimum; again, the Hessian is evaluated to be not PSD. The necessary second-order optimality condition does not hold: it is not a local minimum.
3. See code listing at end of the document.

$$\tilde{H}^0 = \begin{bmatrix} 5.3041 & 4.8041 \\ 4.8041 & 5.3041 \end{bmatrix}$$

4. It does. Script output at final iteration:

```
Iteration #17
Current values of (x,y) = [27.4427      -28.5679]
Current norm of gradient =1.2935e-07
Function value at last iterate =-2
Iterate is a local minimum: True
```

²Wikipedia: Gershgorin’s Circle Theorem Example (click me!)

³GNU Octave (click me!)