Nonlinear Optimization: Interesting Problems

Jonathan Mostovoy - 1002142665
University of Toronto

January 23, 2017

Contents

1 Definition to Be Used in The Proceeding Problems

Definition. 1.1: Metric Space

A metric space is an ordered pair (M, d) where M is a set and d is a function $d : M \times M \to \mathbb{R}$ such that the following three conditions hold:

1. $d(x, y) = 0 \iff x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$

Definition. 1.2: Open Ball

The open ball inside the metric space (M, d) denoted $B_r(x)$ is the set:

$$B_r(x) := \{y \in (M, d) : d(y, x) < r\}$$

Definition. 1.3: Open

A set $U \subset (M, d)$ is called open if $\forall x \in U$, $\exists r > 0, r \in \mathbb{R}$ such that:

$$B_r(x) \subset U$$

Definition. 1.4: Closed

A set $U \subset (M, d)$ is called closed if $U^c = \{y \in (M, d) : y \notin U\}$ (the compliment of U) is open.

Definition. 1.5: Bounded

A set $U \subset (M, d)$ is called bounded if $\exists\ r < \infty, r \in \mathbb{R}$ s.t. $U \subset B_r(0)$.

Definition. 1.6: Convex Function

We say that a function, $f : \mathbb{R}^n \to \mathbb{R}$, is *Convex* if and only if epi(f) is a convex set, where

$$\text{epi}(f) := \left\{(x, \mu) \,\middle|\, x \in \mathbb{R}^n,\ \mu \in \mathbb{R},\ \mu \geq f(x)\right\} \subseteq \mathbb{R}^{n+1}$$

2 Applications of Real Analysis, Multivariate Calculus & Linear Algebra

2.1 The Existence of a Solution for a Minimization Problem

Show that if $f(x)$ is continuous on $\mathbb{R}^n$ and $\lim_{||x||\to\infty} f(x) = \infty$ then there exists $\hat{x}$ with $f(\hat{x}) \leq f(x)$ for all $x \in \mathbb{R}^n$ (i.e. the unconstrained minimization problem for $f(x)$ has a solution).

Proof. Let us define our metric as the standard Euclidean metric in $\mathbb{R}^n$ $\left(d(x,y) := \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\right)$ (and hence $d(x,0) = ||x|| = \sqrt{\sum_{i=1}^{n}(x_i)^2}$) so that we are working in a the metric space: $(\mathbb{R}^n, d)$. Let us now choose an $r \in \mathbb{R}$ which satisfies the following:

$$\text{If } m = \inf_{||\mathbf{x}||=r} f(\mathbf{x}), \text{ then } \forall\ ||y|| \geq r, f(y) \geq m$$

And we know by the continuity and the fact that $\lim_{||x||\to\infty} f(x) = \infty$, there must exist (many) rs. We can show this explicitly from the definition of a limit. We say $\lim_{||x||\to\infty} f(x) = \infty$ if $\forall \epsilon > 0\ \exists \delta$ s.t. if $|||x|| - \infty| < \delta \implies ||f(||x||) - \infty|| < \epsilon$. This is intuitively equivalent to saying: $\forall M >> 0$, $\exists r$ s.t. if $||x - 0|| \geq r \implies ||f(||x||)|| > M$. Next, we recall the following theorem:

Theorem. 2.1: The Extreme Value Theorem

If f is continuous on a compact set Ω, then f attains an absolute maximum and an absolute minimum in Ω.

Let us now recall that $S \subset R^n$ is compact $\iff$ S is closed and bounded (this is not true in general for any (M,d), but is for $\mathbb{R}^n$). As such, $\overline{B_r(0)}$ is trivially bounded (by r), and is closed by the definition of closure. Therefore, by Theorem 2.1, we know $\exists\ \hat{x}$ s.t. $\forall x \in \overline{B_r(0)}$, $f(\hat{x}) \leq f(x)$, and since we have the property that $\forall y \in [B_r(0)]^c, f(y) \geq m \implies f(y) \geq f(\hat{x})$. Thus, we have now proven what we wanted to show ($\exists\ \hat{x}$ which is a global minimum).

□

2.2 Intersection of Closed Sets is Closed

Show that the intersection of any number of closed sets is a closed set.

Let us recall a fundamental Theorem of Set Theory:

Theorem. 2.2: De Morgan's Law

If $A_i \subset S$ is a set $\forall i \in I$, where I is an indexing set which may be finite or infinite, then:

$$\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} (A_i)^c.$$

Proof. We prove De Morgan's Law by induction (for the case that $|I| \leq \aleph_0$):

For $|I| = 1$ the result is trivial. When $|I| = 2$, we have $(A_1 \cap A_2)^c = A_1^c \cup A_2^c$. Assume $x \in (A_1 \cap A_2)^c \iff x \notin (A_1 \cap A_2) \iff x \notin A_1$ or $\notin A_2 \iff x \in A_1^c$ or $x \in A_2^c \iff x \in A_1^c \cup A_2^c$, which $\implies (A_1 \cap A_2)^c \subseteq A_1^c \cup A_2^c$ and by assuming the last step of our process ($x \in A_1^c \cup A_2^c$), we find $(A_1 \cap A_2)^c \supseteq A_1^c \cup A_2^c$ which $\implies (A_1 \cap A_2)^c = A_1^c \cup A_2^c$.

We now assume De Morgan's Law holds for $|I| = n - 1$, then for $|I| = n$,

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \left[\left(\bigcap_{i=1}^{n-1} A_i\right) \bigcap A_n\right]^c \overset{*}{=} \left(\bigcap_{i=1}^{n-1} A_i\right)^c \bigcup A_n^c \overset{**}{=} \left(\bigcup_{i=1}^{n-1} A_i^c\right) \bigcup A_n^c = \bigcup_{i=1}^{n} A_i^c$$

Where the $\overset{*}{=}$ step uses De Morgan's Law for $|I| = 2$, and $\overset{**}{=}$ uses our inductive hypothesis. □

We next consider another Theorem to be used to answer this problem:

Theorem. 2.3: Union of Open Sets is Open

If $A_i \subseteq (M, d)$ is open $\forall i \in I$, (I an indexing set), then $\bigcup_{i\in I} A_i$ is open.

Proof. If $x \in \bigcup_{i\in I}$, then $\exists j$ s.t. $x \in A_j$, since A_j is open, $\exists r > 0$ s.t. $B_r(x) \subset A_j$ and therefore $\forall x \in \bigcup_{i\in I} A_i$, $B_r(x) \subset A_j \subset \bigcup_{i\in I} A_i$ which is our definition of openness. □

Now, our question may be considered a Corollary of De Morgan's Law (and Theorem 2.3):

Corollary. 2.1: Intersection of Closed Sets is Closed (I.e. the Question)

The intersection of any number of closed sets is a closed set.

Proof. Let A_i be a closed set for all $i \in I$, where like before I is an indexing set. Then, by De Morgran's Law:

$$\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} (A_i)^c.$$

We now recall that a set, S, is open if and only if its compliment, S^c, is closed. Since $\forall i$, A_i is closed $\implies A_i^c$ is open $\forall i \overset{*}{\implies} \bigcup_{i\in I} A_i$ is open ($\overset{*}{\implies}$ by Theorem 2.3). We now know $(\cup_{i\in I} A_i)^c$ is open $\implies \cup_{i\in I} A_i$ is closed.

□

2.3 The Maximum of a Set of Convex Functions' is Convex

Show that if $f_1, f_2, \dots f_m$ *are convex functions on* $\mathbb{R}^n$*, then the function* $g(x) = \max\big(f_1(x), \dots, f_m(x)\big)$ *is also convex.*

Proof. Let us recall the following definition of a Convex function:

Definition. 2.1: Convex Function

We say that a function, $f : \mathbb{R}^n \to \mathbb{R}$, is *Convex* if and only if epi(f) is a convex set, where

$$\text{epi}(f) := \left\{(x, \mu) \,\middle|\, x \in \mathbb{R}^n,\ \mu \in \mathbb{R},\ \mu \geq f(x)\right\} \subseteq \mathbb{R}^{n+1}$$

And hence, if we let $I = \{1, \dots, m\}$;

$$\begin{aligned}\text{epi}\Bigg(\max\Big(f_1(x), \dots, f_m(x)\Big)\Bigg) &= \left\{(x,\mu)\middle|x \in \mathbb{R}^n, \mu \in \mathbb{R}, \mu \geq \max_{i\in I}(f_i)\right\} \\ &= \bigcap_{i \in I}\left\{(x,\mu)\middle|x \in \mathbb{R}^n, \mu \in \mathbb{R}, \mu \geq f_i\right\} \\ &= \bigcap_{i\in I}\text{epi}\big(f_i(x)\big)\end{aligned}$$

And since the intersection of convex sets is convex, we can conclude that $g(x)$ is convex.

□

We quickly prove that the intersection of convex sets is convex: For $|I| = 2$, we have that if $x_1, x_2 \in S_1 \cap S_2 \implies x_1, x_2 \in S_1$ and S_2, then $\forall y \in S_1$ and S_2, we have $y = \alpha x_1 + (1-\alpha)x_2, \alpha \in [0,1], \implies y \in S_1 \cap S_2$. Since this is true $\forall x_1, x_2 \in S_1 \cap S_2$ and $\alpha \in [0,1]$, we can conclude conclude that $S_1 \cap S_2$ is convex. Naturally, the proof here extends very easily by inductions for all $\mathbb{N}$ (for $|I| = n-1$, we know that $S = \cap_{i=1}^{n-1} S_i$ is convex, and hence $S \cap S_n$ is convex). Hence $\forall |I| \in \mathbb{N}$, we have that if S_i is convex $\forall i \in |I|$, then $\cap_{i\in I} S_i$ is convex.

2.4 Approximating an Arbitrary Function

To approximate the function g *over the interval* $[0,1]$ *by a polynomial* p *of degree* $\leq n$*, we minimize the following criterion:*

$$f(\mathbf{a}) = \int_0^1 \big(g(x) - p(x)\big)^2 dx$$

where $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$*. Find the equations satisfied by the optimal conditions* $\mathbf{a} = (a_0, \dots, a_n)$*.*

Proof. We recall that if $\mathbf{a}$ is optimal $\implies \nabla f(\mathbf{a}) = 0$ (or does not exist). We thus compute:

$$\begin{aligned}\frac{\partial f}{\partial a_i} &= \frac{\partial}{\partial a_i}\left(\int_0^1 (g-p)^2 dx\right) \\ &= \int_0^1 \frac{\partial}{\partial a_i}(g-p)^2 dx \\ &= -2\int_0^1 x^i (g-p) dx\end{aligned}$$

And hence by requiring $\nabla f(\mathbf{a}) = 0 \equiv \frac{\partial f}{\partial a_i} = 0\ \forall i$:

$$\begin{aligned}\int_0^1 x^i g dx &= \int_0^1 x^i p dx \\ &= \int_0^1 \left(a_n x^{n+i} + a_{n-1}x^{n+i-1} + \cdots + a_1 x^{i+1} + a_0 x^i\right) dx \\ &= \sum_{j=0}^{n} a_j \left(\int_0^1 x^{j+i} dx\right) \\ &= \frac{1}{1+i+j}\sum_{j=0}^{n} a_j\end{aligned}$$

We can thus write the set of our equations in the form of $H\mathbf{a}^T = B$:

$$\begin{pmatrix} 1 & 1/2 & \dots & 1/(n+1) \\ 1/2 & 1/3 & \dots & 1/(n+2) \\ \vdots & \vdots & \ddots & \vdots \\ 1/(n+1) & 1/(n+2) & \dots & 1/(2n+1) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \int_0^1 g dx \\ \int_0^1 x g dx \\ \vdots \\ \int_0^1 x^n g dx \end{pmatrix}$$

Now, by noticing that H is what is referred to as a Hilbert Matrix, and by referencing the literature, we can write down H^{-1} as:

$$(H^{-1})_{i,j} = (-1)^{i+j}(i+j-1)\binom{n+i}{n-j+1}\binom{n+j}{n-i+1}\binom{i+j-2}{i-1}^2$$

Therefore, $\mathbf{a} = (H^{-1}B)^T$.

□

2.5 An Explicit Minimization Problem

A) *Using the first-order necessary conditions, find a minimum point of the function:*

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9$$

B) *Verify that the point is a relative minimum point by verifying that the second-order sufficiency conditions hold.*

C) *Prove that the point is a global minimum point.*

A) *Proof.* Let us start by simplifying our function a little:

$$\begin{aligned}f(x, y, z) &= 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9 \\ &= x^2 - xy + (x+y)^2 - y^2 + (y+z)^2 - yz - 6x - 7y - 8z + 9 \\ &= (x-3)^2 + (x+y)^2 + (y+z)^2 - y(y+z+x+7) - 8z\end{aligned}$$

We now take the partial derivatives of each variable:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2(x-3) + 2(x+y) - y = 4x + y - 6 \\ \frac{\partial f}{\partial y} &= 2(x+y) + 2(y+z) - (y+z+x+7) - y = x + 2y + z - 7 \\ \frac{\partial f}{\partial z} &= 2(y+z) - y - 8 = y + 2z - 8\end{aligned}$$

As such, we want to solve when the above partial derivatives are equal to 0. As such, we manipulate the following matrix:

$$\left[\begin{array}{ccc|c} 4 & 1 & 0 & 6 \\ 1 & 2 & 1 & 7 \\ 0 & 1 & 2 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1/4 & 0 & 3/2 \\ 0 & 7/4 & 1 & 11/2 \\ 0 & 1 & 2 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1/4 & 0 & 3/2 \\ 0 & 1 & 4/7 & 22/7 \\ 0 & 0 & 10/7 & 34/7 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c} 1 & 1/4 & 0 & 3/2 \\ 0 & 1 & 4/7 & 22/7 \\ 0 & 0 & 1 & 17/5 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1/4 & 0 & 3/2 \\ 0 & 1 & 0 & 6/5 \\ 0 & 0 & 1 & 17/5 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 6/5 \\ 0 & 1 & 0 & 6/5 \\ 0 & 0 & 1 & 17/5 \end{array}\right]$$

Which means we have a local minimum at:

$$(x, y, z) = \left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$$

□

B) *Proof.* We begin by finding all the second order derivatives:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial x}\left(4x + y - 6\right) = 4$$
$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial f}{\partial y}\left(x + 2y + z - 7\right) = 2$$
$$\frac{\partial^2 f}{\partial z^2} = \frac{\partial f}{\partial z}\left(y + 2z - 8\right) = 2$$
$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial f}{\partial y}\left(4x + y - 6\right) = 1$$
$$\frac{\partial^2 f}{\partial z \partial x} = \frac{\partial f}{\partial z}\left(4x + y - 6\right) = 0$$
$$\frac{\partial^2 f}{\partial z \partial y} = \frac{\partial f}{\partial z}\left(x + 2y + z - 7\right) = 1$$

And by recalling the following theorem:

Theorem. 2.4: Clairuit's Theorem

If $f : U \subseteq R^n \to R$ is of type C^k then:

$$\frac{\partial^m f}{\partial x_1^{m_1} \cdots \partial x_n^{m_n}} = \frac{\partial^m f}{\partial x_i^{m_i} \cdots \partial x_j^{m_j}}$$

Where $\sum_{k=1}^n m_k = m$ and $(m_i, \ldots, m_k)$ is a rearrangement of $(m_1, \ldots, m_n)$.

As such, we compute the Hessian of our function:

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial z \partial x} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial z \partial y} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix} = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

Next, we recall the following theorem which classifies critical point as local maximums, minimums, saddles points, or inconclusive to be classified as follows:

Theorem. 2.5: Critical Point Classifications

Let $\mathbf{a}$ be a critical point ($\nabla f(\mathbf{a}) = \mathbf{0}$, or $\nabla f(\mathbf{a})$ does not exist), then:

(a) If $H(f(\mathbf{a}))$ is positive definite, then f attains a local minimum at $\mathbf{a}$.

(b) If $H(f(\mathbf{a}))$ is negative definite, then f attains a local maximum at $\mathbf{a}$.

(c) If $H(f(\mathbf{a}))$ has both positive and negative eigenvalues, then $\mathbf{a}$ is a saddle point for $\mathbf{f}$.

And if none of these conditions hold, our test is inconclusive.

We next need a way to check for positive definitiveness, we use the following theorem:

Theorem. 2.6: Sylvester's criterion

If $A \in \mathbb{C}^{n\times n} (\supset \mathbb{R}^{n\times n})$ is a Hermitian matrix, it is positive-definite if and only if all upper-left sub matrices $\in \mathbb{C}^{k\times k}$ $\forall k$ where $k \leq n$ have positive determinants.

As such, we compute:

(a)

$$\det \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} = 4 \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \det \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} = 10 > 0$$

(b)

$$\det \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} = 7 > 0$$

(c)

$$\det(4) = 4 > 0$$

And hence now, we are finally able to conclude by the Critical Point Classifications and Sylvester's criterion, we may conclude that $\mathbf{a} = \left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$ is a relative minimum point. ☐

C) *Proof.* To answer this question, we prove the following theorem:

Theorem. 2.7: One Global Min for Strictly Convex Functions

A strictly convex function will have at most one global minimum.

Proof. Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ has a local minimums at $\mathbf{a}_1$ and $\mathbf{a}_2$, where: $f(\mathbf{a}_1) \leq f(\mathbf{a}_2)$, $\mathbf{a}_1 \neq \mathbf{a}_2$. Since f is a convex function, we have $\forall \mathbf{b}_i \in U$

$$f\big(\beta \mathbf{b}_1 + (1-\beta)\mathbf{a}_2\big) < \beta f(\mathbf{b}_1) + (1-\beta) f(\mathbf{b}_2), \qquad 0 < \beta < 1$$

Now, let $\alpha \in (0, 1)$, then:

$$f(\mathbf{a}_1) \leq f(\mathbf{a}_2) \implies \alpha f(\mathbf{a}_1) \leq \alpha f(\mathbf{a}_2)$$

which implies:

$$\alpha f(\mathbf{a}_1) + (1-\alpha) f(\mathbf{a}_2) \leq \alpha f(\mathbf{a}_2) + (1-\alpha) f(\mathbf{a}_2) = f(\mathbf{a}_2)$$

And due to our function being strictly convex, we have:

$$f\Big(\alpha\mathbf{a}_1 + (1-\alpha)\mathbf{a}_2\Big) < f(\mathbf{a}_2) \tag{1}$$

Since $\mathbf{a}_2$ is local minimum, $\exists\ r > 0, r \in \mathbb{R}$ s.t. $\forall\ \mathbf{x} \in B_r(\mathbf{a}_2)\backslash\{\mathbf{a}_2\}$, we have must have $f(\mathbf{a}_2) \leq f(\mathbf{x})$. We can now choose α small enough s.t. $\alpha\mathbf{a}_1 + (1-\alpha)\mathbf{a}_2 \in B_r(\mathbf{a}_2)$. Which implies

$$f(\alpha\mathbf{a}_1 + (1-\alpha)\mathbf{a}_2) > f(\mathbf{a}_2) \tag{2}$$

And hence we have arrived at a contradiction since it is impossible to satisfy the following equations simultaneously:

$$f\Big(\alpha\mathbf{a}_1 + (1-\alpha)\mathbf{a}_2\Big) \leq f(\mathbf{a}_2) \quad \forall\alpha \in (0,1) \tag{1}$$
$$f(\alpha\mathbf{a}_1 + (1-\alpha)\mathbf{a}_2) > f(\mathbf{a}_2) \quad \alpha \in (0,1) \text{ and small enough} \tag{2}$$

And hence it must be that $\mathbf{a}_1 = \mathbf{a}_2$, and hence if f is strictly convex and has a local minimum, that local minimum is unique and actually f's global minimum. □

And since the positive definiteness of the Hessian is a sufficient condition for strict convexity, which we showed to be so in Part B, we may conclude that $\mathbf{a} = \left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$ is a global minimum.

□

3 Applications of Matrix Calculus

3.1 Minimization of the Modulus of a Linear Equations

Consider the problem $\min_x \big(f(x)\big)$ *for* $f(x) = |Ax - b|^2$, *where* A *is an* $m \times n$ *matrix with zero null space,* b *is an* m *dimensional vector, and the solution* x *is an* n *dimensional vector.*

1. *What is the first order necessary condition for optimality?*
2. *Compute the Hessian of* f *and show that it is positive definite.*
3. *Conclude from (a) and (b) that* f *has a unique global minimum. Indicate what theorems you are using. Then, give a closed form expression for the global minimizer* $\hat{x}$.
4. *Give explicit answers to the questions above when:*

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \textit{and} \quad b = \begin{pmatrix} 2 \\ 6 \\ 2 \\ 0 \end{pmatrix}$$

3.1.1 Part 1

Answer:

Lemma. 3.1: Matrix Calculus Lemmas

If $\mathbf{x}$ is an $(n \times 1)$ variable vector, and A an $(m \times n)$ and b an $(m \times 1)$ constant matrix and vector respectively. Then we have the following general results:

1. $$\frac{\partial b^T A\mathbf{x}}{\partial \mathbf{x}} = A^T b$$

2. $$\frac{\partial \mathbf{x}^T A^T b}{\partial \mathbf{x}} = A^T b$$

3. $$\frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial \mathbf{x}} = (A + A^T)\mathbf{x}$$

Note: we use the "Denominator layout", for both this and Question #2 i.e. by $\mathbf{y}^T$ and $\mathbf{x}$.

We omit the proofs for the above as their are pretty standard in most 2nd year calculus classes.

Second, we have the following lemma:

Lemma. 3.2: Matrix Modulus Lemma

If $\mathbf{x}$ is an $(n \times 1)$ variable vector, and A an $(m \times n)$ and b an $(m \times 1)$ constant matrix and vector respectively. Then we have the following general results:

$$|A\mathbf{x} - b|^2 = (A\mathbf{x} - b)^T (A\mathbf{x} - b)$$

Proof. Computing directly, we have:

$$|A\mathbf{x} - b|^2 = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ji} x_j - b_j \right)^2$$

Furthermore,

$$(Ax - b) = \begin{pmatrix} \sum_{i=1}^{n} a_{1i} x_i - b_1 \\ \vdots \\ \sum_{i=1}^{n} a_{mi} x_i - b_m \end{pmatrix}$$

And hence:

$$\begin{aligned} (Ax - b)^T (Ax - b) &= \left(\sum_{i=1}^{n} a_{1i} x_i - b_1, \ldots, \sum_{i=1}^{n} a_{mi} x_i - b_m \right) \begin{pmatrix} \sum_{i=1}^{n} a_{1i} x_i - b_1 \\ \vdots \\ \sum_{i=1}^{n} a_{mi} x_i - b_m \end{pmatrix} \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ji} x_j - b_j \right)^2 \\ &= |A\mathbf{x} - b|^2 \end{aligned}$$

□

Now, back to the question at hand:

From Lemma 1.2 we know $|A\mathbf{x} - b|^2 = (A\mathbf{x} - b)^T(A\mathbf{x} - b)$, thus, we first multiply out:

$$\begin{aligned} |A\mathbf{x} - b|^2 &= (A\mathbf{x} - b)^T(A\mathbf{x} - b) \\ &= \mathbf{x}^T A^T A\mathbf{x} - b^T A\mathbf{x} - \mathbf{x}^T A^T b + b^T b \\ &\qquad\qquad\qquad \text{since } (AB)^T = B^T A^T \text{ and } (A + B)^T = A^T + B^T \end{aligned}$$

Therefore, we can now compute the gradient:

$$\begin{aligned} \nabla|A\mathbf{x} - b|^2 &= \frac{\partial}{\partial \mathbf{x}}\Big(\mathbf{x}^T A^T A\mathbf{x} - b^T A\mathbf{x} - \mathbf{x}^T A^T b + b^T b\Big) \\ &= (2A^T A)\mathbf{x} - A^T b - A^T b + \mathbf{0} \\ &= 2A^T A\mathbf{x} - 2A^T b \end{aligned}$$

Next, for optimality, we recall that if we are considering a domain $\Omega \subseteq \mathbb{R}^n$, we must have $\nabla f \cdot \mathbf{d} \geq 0 \ \forall \mathbf{d}$, where $\mathbf{d}$ is a feasible direction vector which stays in Ω given $\mathbf{x}$. Thus, for optimality, we must have:

$$2(A^T A\mathbf{x} - A^T b) \cdot \mathbf{d} \geq 0$$

and for the interior of Ω, $\nabla f \cdot \mathbf{d} \geq 0 \iff \nabla f = 0$, we have (and since A has a zero null space we know $AA^T = A^T A$ must have an inverse:

$$\hat{\mathbf{x}} = (A^T A)^{-1}(A^T b)$$

$$\mathcal{H}(f) = 2 \cdot \begin{pmatrix} \sum_{j=1}^m a_{j1}a_{j1} & \dots & \sum_{j=1}^m a_{j1}a_{jn} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^m a_{jn}a_{j1} & \dots & \sum_{j=1}^m a_{jn}a_{jn} \end{pmatrix}$$

3.1.2 Part 2

To compute the Hessian, we recall $\mathcal{H}(f) = \nabla^2 f$, and hence:

$$\mathcal{H}(f) = \nabla^2 f = \nabla(\nabla f) = \nabla(2A^T A\mathbf{x} - 2A^T b) = 2(A^T A)^T = 2A^T A$$

Furthermore, we know this is positive-semi-definite since it is the Gram matrix of linearly independent vectors., i.e., $\frac{1}{2}\mathcal{H}(f) = \langle a_{jk_1}, a_{jk_2}\rangle$.

3.1.3 Part 3

Firstly, we recall that $\mathcal{H}(f)$ is positive definite, which by Prop 5, Ch 7.4 in the textbook implies that f is convex. Next, we recall that since we are considering $\Omega = \mathbb{R}^n$, which is a convex set, we have that f is a convex function on a convex set and hence we can set $\nabla f = 0$ (as we showed how to do explicity above) to find the global minimum.

3.1.4 Part 4

To give explicit answers to the equations above, we need $A^T A$, $A^T b$, and $(A^T A)^{-1}$. We thus compute:

$$A^T A = \begin{pmatrix} 2 & 0 & 0 & 1 \\ -1 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 0 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & -2 \\ -2 & 6 \end{pmatrix}$$

$$A^T b = \begin{pmatrix} 2 & 0 & 0 & 1 \\ -1 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 6 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 12 \end{pmatrix}$$

$$(A^T A)^{-1} = \frac{1}{\det(A^T A)} \begin{pmatrix} 6 & 2 \\ 2 & 5 \end{pmatrix} = \frac{1}{13} \begin{pmatrix} 3 & 1 \\ 1 & 5/2 \end{pmatrix}$$

Therefore, we see that for optimality ($\mathbf{d}$ defined as above), we must have:

$$\nabla f = 2 \left[\begin{pmatrix} 5 & -2 \\ -2 & 6 \end{pmatrix} \mathbf{x} - \begin{pmatrix} 4 \\ 12 \end{pmatrix} \right]$$

For optimality we have:

$$\nabla f = 2 \left[\begin{pmatrix} 5 & -2 \\ -2 & 6 \end{pmatrix} \mathbf{x} - \begin{pmatrix} 4 \\ 12 \end{pmatrix} \right] \cdot \mathbf{d} \geq 0$$

And in the interior (which in this case is all of our domain, $\mathbb{R}^n$) we have:

$$\hat{\mathbf{x}} = \frac{1}{13} \begin{pmatrix} 3 & 1 \\ 1 & 5/2 \end{pmatrix} \begin{pmatrix} 4 \\ 12 \end{pmatrix} = \frac{1}{13} \begin{pmatrix} 24 \\ 34 \end{pmatrix}$$

And for the Hessian, we have:

$$\mathcal{H}(f) = \begin{pmatrix} 10 & -4 \\ -4 & 12 \end{pmatrix}$$

3.2 Newton's Method Applied to the Minimization of the Modulus' Cubic

Consider Newton's method applied to the minimization of the function $f(x) = |x|^3$, where $x \in \mathbb{R}^n$.

1. *Compute the gradient and the Hessian of f.*
2. *Use the formula $\left(I + uu^T\right)^{-1} = I - \frac{1}{2}uu^T$, where I is the $n \times n$ identity matrix, and $u \in \mathbb{R}^n$ is a unit vector to compute the inverse of the Hessian of f. Use this to give the explicit formula for the iteration step in Newton's method for this function.*

Answer:

3.2.1 Part 1

$$|x|^3 = \left(\sum_{i=1}^{n} (x_i)^2 \right)^{3/2}$$

And hence:

$$\frac{\partial}{\partial x_j}|x|^3 = 3x_j\left(\sum_{i=1}^{n}(x_i)^2\right)^{1/2}$$

$$\implies \nabla|x|^3 = \left(3x_1\left(\sum_{i=1}^{n}(x_i)^2\right)^{1/2}, \dots, 3x_n\left(\sum_{i=1}^{n}(x_i)^2\right)^{1/2}\right) = 3\mathbf{x}^T|\mathbf{x}|$$

And as such,

$$\begin{aligned}
\mathcal{H}(f) = \nabla(3\mathbf{x}^T|\mathbf{x}|) &= 3\left(\frac{\partial|x|}{\partial\mathbf{x}}\cdot\mathbf{x}^T + \frac{\partial\mathbf{x}^T}{\partial\mathbf{x}}\cdot|\mathbf{x}|\right) && \text{By the Matrix Chain Rule} \\
&= 3\left(\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\mathbf{x}^T + \mathbb{1}_{n\times n}\cdot|\mathbf{x}|\right) && \text{Since } \frac{\partial|\mathbf{x}|}{\partial\mathbf{x}} = \frac{\mathbf{x}}{|\mathbf{x}|} \text{ and } \frac{\partial\mathbf{x}^T}{\partial\mathbf{x}} = \mathbb{1}_{n\times n} \\
&= 3|\mathbf{x}|\left(\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\frac{\mathbf{x}^T}{|\mathbf{x}|} + \mathbb{1}_{n\times n}\right)
\end{aligned}$$

We can check that our above computations are indeed correct directly, since:

$$\frac{\partial^2}{\partial x_k x_j}|x|^3 = 3x_j\sum_{i=1}^{n}(x_i)^2 = \begin{cases} 6(x_j)^2 + 3\sum_{i=1}^{n}(x_i)^2 & \text{if } k = j \\ 6x_k x_j & \text{if } k \neq j \end{cases}$$

Therefore,

$$\mathcal{H}(f) = \begin{pmatrix} 6(x_1)^2 + 3\sum_{i=1}^{n}(x_i)^2 & 6x_2x_1 & \dots & 6x_{n-1}x_1 & 6x_nx_1 \\ 6x_1x_2 & 6(x_2)^2 + 3\sum_{i=1}^{n}(x_i)^2 & \dots & 6x_{n-1}x_2 & 6x_nx_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 6x_1x_{n-1} & 6x_2x_{n-1} & \dots & 6(x_{n-1})^2 + 3\sum_{i=1}^{n}(x_i)^2 & 6x_nx_{n-1} \\ 6x_1x_n & 6x_2x_n & \dots & 6x_{n-1}x_n & 6(x_n)^2 + 3\sum_{i=1}^{n}(x_i)^2 \end{pmatrix}$$

$$= 3\left(\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\mathbf{x}^T + \mathbb{1}_{n\times n}\cdot|\mathbf{x}|\right)$$

3.2.2 Part 2

From the identity of $\left(\mathbb{1}_{n\times n} + uu^T\right)^{-1} = \mathbb{1}_{n\times n} - \frac{1}{2}uu^T$, we can see that:

$$\begin{aligned}
\left(\mathcal{H}(f)\right)^{-1} &= \left(3|\mathbf{x}|\left(\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\frac{\mathbf{x}^T}{|\mathbf{x}|} + \mathbb{1}_{n\times n}\right)\right)^{-1} \\
&= \left(\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\frac{\mathbf{x}^T}{|\mathbf{x}|} + \mathbb{1}_{n\times n}\right)^{-1}(3|\mathbf{x}|)^{-1} \\
&= \left(\mathbb{1}_{n\times n} - \frac{1}{2}\frac{\mathbf{x}}{|\mathbf{x}|}\cdot\frac{\mathbf{x}^T}{|\mathbf{x}|}\right)\left(\frac{1}{3|\mathbf{x}|}\right)
\end{aligned}$$

Therefore, since in Newton's Method, have that $\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\mathcal{H}(f)\right)^{-1}(\nabla f)^T$, and hence:

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k - \left(\mathbb{1}_{n\times n} - \frac{1}{2}\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\cdot\frac{\mathbf{x}_k^T}{|\mathbf{x}_k|}\right)\left(\frac{1}{3|\mathbf{x}_k|}\right)\left(3\mathbf{x}_k^T|\mathbf{x}_k|\right)^T \\
&= \mathbf{x}_k - \left(\mathbb{1}_{n\times n} - \frac{1}{2}\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\cdot\frac{\mathbf{x}^T}{|\mathbf{x}_k|}\right)\cdot\mathbf{x}_k \\
&= \mathbb{1}_{n\times n}\mathbf{x}_k - \left(\mathbb{1}_{n\times n} - \frac{1}{2}\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\cdot\frac{\mathbf{x}_k^T}{|\mathbf{x}_k|}\right)\cdot\mathbf{x}_k \\
&= \frac{1}{2}\left(\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\cdot\frac{\mathbf{x}_k^T}{|\mathbf{x}_k|}\right)\cdot\mathbf{x}_k
\end{aligned}$$

3.3 Steepest Decent

Let $f(\mathbf{x}) = \frac{1}{2}\big((x_1)^2 + c(x_2)^2\big)$ *where* $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *and* $c > 1$. *Consider applying the method of of steepest descent to this function, starting at* $\mathbf{x}^0 = \begin{pmatrix} c \\ 1 \end{pmatrix}$. *Find a formula for* $\mathbf{x}^k$.

Answer:
If we assume $\alpha_k = \cdots = \alpha_0$, we can compute that:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} x_1 \\ cx_2 \end{pmatrix}$$

$$\implies \mathbf{x}^{m+1} = \mathbf{x}^m - \alpha\nabla f(\mathbf{x}^m) = \begin{pmatrix} x_1^m \\ x_2^m \end{pmatrix} - \alpha\begin{pmatrix} x_1^m \\ cx_2^m \end{pmatrix} = \begin{pmatrix} (1-\alpha)x_1^m \\ (1-c\alpha)x_2^m \end{pmatrix}$$

Next, since we see that this holds $\forall m > 0$, and hence we naturally will have:

$$\mathbf{x}^k = \begin{pmatrix} (1-\alpha)^k x_1^0 \\ (1-c\alpha)^k x_2^0 \end{pmatrix}$$

And therefore if $\mathbf{x}^0 = (c, 1)^T$, then:

$$\mathbf{x}^k = \begin{pmatrix} c(1-\alpha)^k \\ (1-c\alpha)^k \end{pmatrix}$$

Now, we need to prove that $\alpha_k = \cdots = \alpha_0$. To do this, we note that the Wolfe Conditions imply that $\alpha_k = \frac{2}{1+c}$.

4 Corollaries and Examples of Common Algorithms

4.1 Steepest Descent

Let $x^1, x^2, \ldots, x^k, \ldots$ *be the sequence obtained by applying the method of steepest descent to a continuously differentiable function* f*, starting at* x^0. *Show that for any* $k \geq 0$*, the vector* $(x^{k+2} - x^{k+1})$ *is perpendicular to* $(x^{k+1} - x^k)$. *(This explains the "orthogonal zig-zag pattern" for steepest descent mentioned in class).*

Hint: *recall that* α_k *is chosen to minimize the function:* $\alpha \to f\big(x^k - \alpha(\nabla f(x^k))^T\big)$*; use calculus!*

Proof. We recall that two vectors, u, v in $\mathbb{R}^n$ are orthogonal $\iff$ $\langle u, v\rangle_{\mathbb{R}^n} := u^T v = 0$. Thus, let us first alter the expressions $(x^{k+2} - x^{k+1})$ and $(x^{k+1} - x^k)$ by noting that since we are working within the context of steepest decent, we have that $x^{k+1} = x^k - \alpha_k(\nabla f(x^k))^T \implies$

$$
\begin{aligned}
x^{k+2} - x^{k+1} &= x^{k+1} - \alpha_{k+1}(\nabla f(x^{k+1}))^T - x^k + \alpha_k(\nabla f(x^k))^T \\
&= \Big(x^k - \alpha_k(\nabla f(x^k))^T\Big) - \alpha_{k+1}(\nabla f(x^k - \alpha_k(\nabla f(x^k))^T))^T - \Big(x^k - \alpha_k(\nabla f(x^k))^T\Big) \\
&= -\alpha_{k+1}(\nabla f(x^k - \alpha_k(\nabla f(x^k))^T))^T
\end{aligned}
$$

and similarly for $x^{k+1} - x^k$:

$$
\begin{aligned}
x^{k+1} - x^k &= x^k - \alpha_k(\nabla f(x^k))^T - x^k \\
&= -\alpha_k(\nabla f(x^k))^T
\end{aligned}
$$

And therefore we see that:

$$
\begin{aligned}
&\langle (x^{k+2} - x^{k+1}), (x^{k+1} - x^k)\rangle_{\mathbb{R}^n} = 0 \\
\iff &\Big(- \alpha_{k+1}(\nabla f(x^k - \alpha_k\big(\nabla f(x^k))^T)\big)\Big) \cdot \Big(- \alpha_k(\nabla f(x^k))^T\Big) = 0 \\
\iff &\Big(\nabla f(x^k - \alpha_k\big(\nabla f(x^k))^T\big)\Big) \cdot \Big(\nabla f(x^k))^T\Big) = 0
\end{aligned}
$$

Furthermore, we recall that α_k is chosen s.t. $\alpha_k = \arg\min\big(f\big(x^k - \alpha_k(\nabla f(x^k))^T\big)\big)$, and this condition is satisfied $\iff \frac{\partial}{\partial \alpha_k}\big(f\big(x^k - \alpha_k(\nabla f(x^k))^T\big)\big) = 0$, and since:

$$
\frac{\partial}{\partial \alpha_k}\big(f\big(x^k - \alpha_k(\nabla f(x^k))^T\big)\big) = \Big(\nabla f(x^k - \alpha_k\big(\nabla f(x^k))^T\big)\Big) \cdot \Big(- \nabla f(x^k))^T\Big)
$$

It follows that the condition of $\alpha_k = \arg\min\big(f\big(x^k - \alpha_k(\nabla f(x^k))^T\big)\big)$ is equivalent to the condition of $\langle (x^{k+2} - x^{k+1}), (x^{k+1} - x^k)\rangle_{\mathbb{R}^n} = 0$.

□

4.2 The Rosenbrock Banana Function

Let f be the function defined for all $x = (x_1, x_2)^T$ in $\mathbb{R}^2$ by: $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$. Find the global minimum point $\hat{x}$ of this function, and compute the condition number of the Hessian of f at $\hat{x}$. (This function is known as the Rosenbrock banana function, for which methods such as steepest descent converge very slowly).

Answer: We first compute all first and second partial derivatives:

$$
\begin{aligned}
&\frac{\partial f}{\partial x_1} = -400x_1(x_2 - x_1) - 2(1 - x_1) = 400x_1^2 + 2x_1 - 400x_2 - 2, \qquad \frac{\partial f}{\partial x_2} = 200(x_2 - x_1^2) \\
&\frac{\partial^2 f}{\partial x_1^2} = 800x_1 + 2, \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -400x_1, \qquad \frac{\partial^2 f}{\partial x_2^2} = 200
\end{aligned}
$$

As such, we would like to set $\nabla f((x_1, x_2)^T) = 0$ to find the minimum, doing so will result in a system of linear equations as is evident if we do this explicitly:

$$
\begin{aligned}
&\nabla f((x_1, x_2)^T) = 0 \\
\iff &400x_1^2 + 2x_1 - 400x_2 - 2 = 0 \;\text{ and }\; 200(x_2 - x_1^2) = 0 \\
\iff &400x_2 = 400x_1^2 + 2x_1 - 2 \;\text{ and }\; x_2 = x_1^2 \\
\iff &x_1 = 1 \;\text{ and }\; x_2 = x_1^2 = 1^2
\end{aligned}
$$

Therefore, we see that $\nabla(f(x_1,x_2)^T) = 0 \iff (x_1,x_2) = (1,1)$. We next compute $\mathcal{H}(f(1,1))$.

$$\mathcal{H}(f(1,1)) = \begin{pmatrix} 800x_1 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}\Bigg|_{(x_1,x_2)=(1,1)} = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

And since $\det(\mathcal{H}(f(1,1))) = 400 > 0$ and $\frac{\partial^2 f}{\partial x_1^2}\Big|_{(x_1,x_2)=(1,1)} = 802 > 0$, by Sylvester's criterion, we now know that $\det(\mathcal{H}(f(1,1)))$ is positive-definite. Therefore, $\hat{x} = (1,1)$ is the global minimum of this function (global because $f(1,1) = 0$, and $f(x) = a(x_2 - x_1^2)^2 + b(1-x_1)^2 \geq 0 \forall x \in \mathbb{R}^2$ since $a, b > 0$ghgggy.

Furthermore, we recall that any $\mathcal{H}(f(\hat{x}))$ will be trivially normal, and therefore the condition number of $\mathcal{H}(f(\hat{x}))$, $\kappa(\mathcal{H}(f(\hat{x}))) = \frac{\lambda_{min}}{\lambda_{\min}}$. We thus compute $\mathcal{H}(f(\hat{x}))$'s eigenvalues:

$$\begin{aligned} &\begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = 0 \\ \iff &\lambda^2 - 1002\lambda + 400 = 0 \\ \iff &(\lambda - (501 - \sqrt{250601})(\lambda - (501 + \sqrt{250601}) = 0 \\ \iff &\lambda = 501 \pm \sqrt{250601} \end{aligned}$$

And therefore:

$$\kappa(\mathcal{H}(f(\hat{x}))) = \frac{501 + \sqrt{250601}}{501 - \sqrt{250601}} = \frac{\left(501 + \sqrt{250601}\right)^2}{501602} \approx 2$$

4.3 Quadratic Inequalities

Let f be the function defined on $\mathbb{R}^n$ as $f(x) = \frac{1}{2}x^TQx - b^Tx$, with Q a positive definite symmetric $n \times n$ matrix and b a vector in $\mathbb{R}^n$. Let $\hat{x}$ be the point of global minimum of f. Let $E(x)$ denote the quantity $E(x) = \frac{1}{2}(x-\hat{x})^TQ(x-\hat{x})$. (You can think of it as the Q-norm of $(x-\hat{x})$).

Show that $E(x) = f(x) - f(\hat{x})$. Use this to write the inequality (8.42) from your textbook in terms of values of f. Compare with the weaker inequality (8.47), valid for a more general class of functions.

Proof. We can show the identity of $E(x) = f(x) - f(\hat{x})$ directly. First, we note that $\nabla f(x) = 0 \iff$

$Qx - b = 0 \iff x = Q^{-1}b$. The rest is just algebra:

$$
\begin{aligned}
E(x) &= \frac{1}{2}(x-\hat{x})^T Q(x-\hat{x}) \\
&= \frac{1}{2}(x^T - b^T(Q^{-1})^T)Q(x - Q^{-1}b) \\
&= \frac{1}{2}\left(x^T Qx - b^T(Q^{-1})^T Qx - x^T QQ^{-1}b + b^T(Q^{-1})^T QQ^{-1}b\right) \\
&= \frac{1}{2}\left(x^T Qx - b^T Q^{-1}Qx - x^T QQ^{-1}b + b^T Q^{-1}QQ^{-1}b\right) \\
&= \frac{1}{2}\left(x^T Qx - b^T x - x^T b + b^T Q^{-1}b\right) \\
&= \left(\frac{1}{2}\big(x^T Qx\big) - b^T x\right) - \left(-\frac{1}{2}\big(b^T Q^{-1}b\big)\right) && \text{since } b^T x = x^T b \\
&= f(x) - \left(\frac{1}{2}\big(b^T Q^{-1}b\big) - b^T Q^{-1}b\right) \\
&= f(x) - \left(\frac{1}{2}\Big(\big(Q^{-1}b\big)^T Q\big(Q^{-1}b\big)\Big) - b^T\big(Q^{-1}b\big)\right) \\
&= f(x) - \left(\frac{1}{2}\big(\hat{x}^T Q\hat{x}\big) - b^T\hat{x}\right) \\
&= f(x) - f(\hat{x})
\end{aligned}
$$

Thus, inequality (8.42) may be written as:

$$
E(X_{k+1}) \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 E(x_k) \quad \equiv \quad f(x_{k+1}) - f(\hat{x}) \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 \big(f(x_{k+1}) - f(\hat{x})\big)
$$

And inequality (8.47) we recall is:

$$
f(x_{k+1}) - f(\hat{x}) \leq \left(1 - \frac{\lambda_{\max}}{\lambda_{\min}}\right)\big(f(x_{k+1}) - f(\hat{x})\big)
$$

Thus, since $\lambda_{\max}, \lambda_{\min} > 0$ and $\lambda_{\max} \geq \lambda_{\min}$, then we know that: $0 \leq \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} < 1$ and $0 \leq 1 - \frac{\lambda_{\min}}{\lambda_{\max}} < 1$. Therefore also, $\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} \leq \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}} = 1 - \frac{\lambda_{\min}}{\lambda_{\max}}$. And as such, we also have:

$$
\left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max}} = 1 - \frac{\lambda_{\min}}{\lambda_{\max}}
$$

And hence the Quadratic Case will likely converge faster than a non-Quadratic Case given the same eigenvalues for each. □

4.4 Conjugate Gradient Algorithm Example

Apply the Conjugate Gradient Algorithm to the quadratic function defined for all $x = (x_1, x_2)^T$ *in* $\mathbb{R}^n$ *as* $f(x) = (2x_1 - x_2)^2 + x_2^2 + 2x_2$ *starting at* $x^0 = (5/2, 2)^T$.

Answer:
since degree$(f) = 2$, we may write:

$$
f(x,y) = g(x,y) := \frac{1}{2}\Big((x,y)\mathcal{H}(f)(x,y)^T\Big) - \Big(-K\big(\nabla f(x,y)\big)\Big)^T(x,y) + c
$$

Where $K : \text{Im}(\nabla) \to \mathbb{R}^{1\times 2}$ such that K removes all non-constant terms in $\nabla f(x)$. Here, we have:

$$c = 0, \qquad b = (-K(\nabla(f(x))))^T = \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{and} \quad \mathcal{H}(f(x,y)) = \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} := \mathcal{H}(f)$$

We now show the tremendously laborious calculations below:

$$(1) \qquad r_0 = p_0 = b - \mathcal{H}(f)x_0 = \begin{pmatrix} 0 \\ -2 \end{pmatrix} - \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} 5/2 \\ 2 \end{pmatrix} = \begin{pmatrix} -12 \\ 0 \end{pmatrix}$$

$$(2) \qquad \alpha_0 = \frac{r_0^T r_0}{p_0^T \mathcal{H}(f) p_0} = \frac{(-12, 0)\begin{pmatrix} -12 \\ 0 \end{pmatrix}}{(-12, 0)\begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix}\begin{pmatrix} -12 \\ 0 \end{pmatrix}} = \frac{12^2}{12^2 \cdot 8} = \frac{1}{8}$$

$$(3) \qquad x_1 = x_0 + \alpha_0 p_0 = \begin{pmatrix} 5/2 \\ 2 \end{pmatrix} + \frac{1}{8}\begin{pmatrix} -12 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$(4) \qquad r_1 = r_0 - \alpha_0 \mathcal{H}(f) p_0 = \begin{pmatrix} -12 \\ 0 \end{pmatrix} - \frac{1}{8}\begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix}\begin{pmatrix} -12 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -6 \end{pmatrix}$$

$$(5) \qquad \beta_0 = \frac{r_1^T r_1}{r_0^T r_0} = \frac{(0, -6)\begin{pmatrix} 0 \\ -6 \end{pmatrix}}{(-12, 0)\begin{pmatrix} -12 \\ 0 \end{pmatrix}} = \frac{6^2}{12^2} = \frac{1}{4}$$

$$(6) \qquad p_1 = r_1 + \beta_0 p_0 = \begin{pmatrix} 0 \\ -6 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} -12 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ -6 \end{pmatrix}$$

$$(7) \qquad \alpha_1 = \frac{r_1^T r_1}{p_1^T \mathcal{H}(f) p_1} = \frac{(0, -6)\begin{pmatrix} 0 \\ -6 \end{pmatrix}}{(-3, -6)\begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix}\begin{pmatrix} -3 \\ -6 \end{pmatrix}} = \frac{6^2}{2 \cdot 6^2} = \frac{1}{2}$$

$$(8) \qquad x_2 = x_1 + \alpha_1 p_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -3 \\ -6 \end{pmatrix} = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$$

And hence we see that $x_2 = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$ is our final solution to the Conjugate Gradient Algorithm (and is indeed a global minimum).

4.5 Steepest Descent Example

Consider the problem:

$$\text{minimize } 5x^2 + 5y^2 - xy - 11x + 11y + 11$$

(a) *Find a point satisfying the first-order necessary conditions for a solution.*

(b) *Show that this point is a global minimum.*

(c) *What would be the rate of convergence of steepest descent for this problem?*

(d) *Starting at $x = y = 0$, how many steepest descent iterations would it take (at most) to reduce the function value to 10^{-11}?*

Let us first compute all the 1st and 2nd order partial derivatives ($f(x, y) = 5x^2 + 5y^2 - xy - 11x + 11y + 11$):

$$\frac{\partial g}{\partial x} = 10x - y - 11 \qquad \frac{\partial f}{\partial y} = 10y - x + 11$$
$$\frac{\partial^2 f}{\partial x^2} = 10 \quad \frac{\partial^2 f}{\partial x \partial y} = -1 \quad \frac{\partial^2 f}{\partial y^2} = 10$$

Furthermore, since degree$(f) = 2$, we may write:

$$f(x, y) = g(x, y) := \frac{1}{2}\big((x, y)\mathcal{H}(f)(x, y)^T\big) - \big(-K\big(\nabla f(x, y)\big)\big)^T (x, y) + c$$

Where $K : \text{Im}(\nabla) \to \mathbb{R}^{1\times 2}$ such that K removes all non-constant terms in $\nabla f(x)$, i.e., here $K(\nabla f(x, y)) = (-11, 11)$. (Note: This is also why we are easily able to apply the method of steepest descent).

(a) **Answer:** We see that the first-order conditions will be satisfied $\iff \nabla f(x, y) = 0 \iff (10x - y - 11, 10y - x + 11) = (0, 0)$. We can solve this system of linear equations through the following row reductions:
$$\left(\begin{array}{cc|c} 10 & -1 & 11 \\ -1 & -10 & 11 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1/10 & 11/10 \\ 0 & 99/10 & -99/10 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -1 \end{array}\right)$$
And therefore we see that $(\hat{x}, \hat{y}) = (1, -1)$ satisfies our first order conditions.

(b) **Answer:** We can easily see that the hessian is defined as:
$$\mathcal{H}(f) = \begin{pmatrix} 10 & -1 \\ -1 & 10 \end{pmatrix}$$
And since $\det(\mathcal{H}(f)) = 99 > 0$ and $\frac{\partial^2 f}{\partial x^2} = 10 > 0$, by Sylvester's criterion, we now know that $\det(\mathcal{H}(f(1, 1)))$ is positive-definite. Therefore, $(\hat{x}, \hat{y}) = (1, -1)$ is the global minimum of this function.

(c) **Answer:** We can quickly see that the two eigenvalues of the Hessian are $\lambda_1 = 9, \lambda_2 = 11$ since: $\det\big(\mathcal{H}(f(1, 1) - \lambda I_{n\times n}\big) = 0 \iff (10 - \lambda)^2 - 1 = 0 \iff (\lambda - 9)(\lambda - 11) = 0 \iff \lambda = 9, 11$. Therefore, we have the condition number, $\kappa(f) = \frac{\lambda_2}{\lambda_1} = \frac{11}{9}$, and hence we have that the rate of convergence here will be: $\rho = \big(\frac{11/9-1}{11/9+1}\big)^2 = \frac{1}{10^2}$.

(d) **Answer:** We first recall that $E(x) := \frac{1}{2}(x - \hat{x})^T \mathcal{H}(f)(x - \hat{x}) = f(x) - f(\hat{x})$. However, $f(\hat{x}) = 0$, therefore, $E(x_k) = f(x_k)$. We also recall that $E(x_k) \le \rho^k E(x_0)$, and $f(x_0) = 11$. Therefore:
$$\begin{aligned} E(x_k) &\le \rho^k E(x_0) \\ &= \rho^k f(x_0) \\ &= \left(\frac{1}{10^2}\right)^k (11) = 11 \cdot (10)^{-2k} \end{aligned}$$
And hence we see that $11 \cdot (10)^{-2k} \le 10^{-11} \iff \log_{10}(11) \le 2k - 11 \iff k \ge \frac{11+\log_{10}(11)}{2}$ and since this value is ≈ 6.02, we see it will take at most 7 iterations for us to reduce the function value down to 10^{-11}.

4.6 Symmetric Matrices' Eigenvalues are Q-conjugate

Let Q be a symmetric matrix. Show that any two eigenvectors of Q, corresponding to distinct eigenvalues, are Q-conjugate.

Proof. Suppose v_1, v_2 are eigenvectors of Q corresponding to eigenvalues λ_1, λ_2 where $\lambda_1 \neq \lambda_2$ and $Q = Q^T$ (I.e, $Qv_i = \lambda_i v_i,\ i = 1, 2$). We can thus prove this fact quite easily by noticing:

$$\begin{aligned}\lambda_1 \langle v_1, v_2 \rangle_Q &= \lambda_1 v_1^T Q v_2 = (\lambda_1 v_1)^T Q v_2 = (Qv_1)^T Q v_2 = v_1^T Q^T Q v_2 \\ &= v_1^T Q (Q v_2) = v_1^T Q \lambda_2 v_2 = \lambda_2 v_1^T Q v_2 = \lambda_2 \langle v_1, v_2 \rangle_Q\end{aligned}$$

And since $\lambda_1 \neq \lambda_2 \implies \lambda_1 \langle v_1, v_2 \rangle_Q = \lambda_2 \langle v_1, v_2 \rangle_Q \iff \langle v_1, v_2 \rangle_Q = 0$. □