Mid-term Exam: Answers

Course: CS674
Instructor: Wes Cowan
Department of Computer Science, Rutgers University
110 Frelinghuysen Rd., Piscataway, NJ 08854

October 29, 2016

1 Grab Bag

1.1 Lagrange Multipliers

Let Q be an $n \times n$ symmetric, positive definite matrix. Prove that for any $\underline{u}, \underline{v} \in \mathbb{R}^n$, we have

$$\underline{u}^T Q \underline{v} \leqslant \sqrt{(\underline{u}^T Q \underline{u})(\underline{v}^T Q \underline{v})}, \tag{1}$$

by considering the problem

$$\begin{aligned} max_{\underline{u},\underline{v}} \quad & \underline{u}^T Q \underline{v} \\ (s.t.) \quad & \underline{u}^T Q \underline{u} = n_u \\ & \underline{v}^T Q \underline{v} = n_v, \end{aligned} \tag{2}$$

where $n_u, n_v > 0$ are fixed (Q being positive definite means that if either n_u, n_v are 0, then $\underline{u}$, $\underline{v}$ must be 0). When will Ineq. (1) *be satisfied with equality? Hint: As a real symmetric matrix, Q has an orthonormal basis of eigenvectors.*

Answer: Let $\underline{w}_1, \ldots, \underline{w}_n$ be an orthonormal basis of eigenvectors of Q, with $\lambda_i > 0$ the eigenvalue corresponding to eigenvector $\underline{w}_i$. Consider expressing $\underline{u}, \underline{v}$ with respect to this basis: $\underline{u} = \sum_i \alpha_i \underline{w}_i$ and $\underline{v} = \sum_i \beta_i \underline{w}_i$.

The objective function may therefore be expressed as $f(\underline{\alpha}, \underline{\beta}) = \sum_{i=1}^n \lambda_i \alpha_i \beta_i$, and the equality constraint functions as $h_u(\underline{\alpha}) = \sum_{i=1}^n \lambda_i \alpha_i^2$ and $h_v(\underline{\beta}) = \sum_{i=1}^n \lambda_i \beta_i^2$.

The problem may therefore be re-expressed as:

$$\begin{aligned} \max_{\underline{\alpha},\underline{\beta}} \quad & f(\underline{\alpha}, \underline{\beta}) \\ (\text{s.t.}) \quad & h_u(\underline{\alpha}) = n_u \\ & h_v(\underline{\beta}) = n_v. \end{aligned} \tag{3}$$

The Lagrange multiplier condition can be expressed (relative to two multipliers a, b) as:

$$\begin{aligned} \nabla_\alpha f + a\nabla_\alpha h_u + b\nabla_\alpha h_v = 0 \\ \nabla_\beta f + a\nabla_\beta h_u + b\nabla_\beta h_v = 0 \end{aligned} \tag{4}$$

or

$$\begin{aligned} \lambda_i\beta_i + 2a\lambda_i\alpha_i = 0 \text{ for } i = 1,\ldots,n \\ \lambda_i\alpha_i + 2b\lambda_i\beta_i = 0 \text{ for } i = 1,\ldots,n. \end{aligned} \tag{5}$$

Noting that Q is positive definite, the λ_i can be removed from each of the above, yielding the following systems, $\{\beta_i = -2a\alpha_i\}$ and $\{\alpha_i = -2b\beta_i\}$, or $\underline{\beta}^* = -2a\underline{\alpha}^*$ and $\underline{\alpha}^* = -2b\underline{\beta}^*$. At this point, a solution can proceed in one of two ways: it can attempt to solve for the multipliers and the optimal $\underline{\alpha}^*, \underline{\beta}^*$ in full and evaluate the function at those points; alternately, an interesting conclusion to draw here is that any optimal $\underline{\alpha}^*, \underline{\beta}^*$ must be scalar multiples of each other. As a result, any optimal $\underline{u}^*, \underline{v}^*$ of the original problem must be scalar multiples of each other. Take $\underline{v}^* = \delta\underline{u}^*$, and restate the original problem:

$$\begin{aligned} \max_{\underline{u},\delta} \quad & \delta\underline{u}^{\mathrm{T}}Q\underline{u} \\ \text{(s.t.) } & \underline{u}^{\mathrm{T}}Q\underline{u} = n_u \\ & \delta^2\underline{u}^{\mathrm{T}}Q\underline{u} = n_v, \end{aligned} \tag{6}$$

or

$$\begin{aligned} \max_{\delta} \quad & \delta n_u \\ \text{(s.t.) } & \delta^2 n_u = n_v. \end{aligned} \tag{7}$$

The constraint gives $\delta = \pm\sqrt{n_v/n_u}$, which gives a maximum of the above problem at $\sqrt{n_u n_v}$. This completes the problem (with, perhaps, some observations on the regularity condition of the Lagrange multiplier theorem, in particular noting that the (joint) gradients of h_u, h_v will be linearly independent (as $\nabla_\alpha h_v = 0, \nabla_\beta h_u = 0$), and hence all local optimal will be regular.)

1.2 Convex Analysis

*Let X be a convex set. We say that a point $\underline{x} \in X$ is in the **interior** of X if there exist two points $\underline{y}, \underline{z} \in X$ with $\underline{y} \neq \underline{z}$ such that*

$$\underline{x} = \alpha\underline{y} + (1-\alpha)\underline{z} \tag{8}$$

for some α with $0 < \alpha < 1$. That is, $\underline{x}$ lies somewhere on the line segment connecting $\underline{y}$ and $\underline{z}$. Any point in X that is not in the interior is said to be on the boundary.

- *If f is **strictly convex** over X, show that any local maximum $\underline{x} \in X$ of f over X must be located on the boundary of X.*

  Answer: Proceed by contradiction - assume that $\underline{x}$ is a local maximum of f that is in the interior of X. In this case, there are some $\underline{y}, \underline{z} \in X$ such that for some $0 < \alpha < 1$, $\underline{x} = \alpha\underline{y} + (1-\alpha)\underline{z}$. But by the strict convexity of f, we have

$$\begin{aligned} f(\underline{x}) &< \alpha f(\underline{y}) + (1-\alpha)f(\underline{z}) \\ &\leqslant \alpha\max(f(\underline{y}), f(\underline{z})) + (1-\alpha)\max(f(\underline{y}), f(\underline{z})) = \max(f(\underline{y}), f(\underline{z})). \end{aligned} \tag{9}$$

Note, we may also conclude from this that given any line segment that includes $\underline{x}$ (away from the end points), $f(\underline{x})$ is strictly less than the maximum value of f at the end points. We will use the line segment between $\underline{y},\underline{z}$ to construct alternative $\underline{y}',\underline{z}' \in X$ that are arbitrarily close to $\underline{x}$, but (at least one will) have superior values of f.

In particular, for $0 < \delta < 1$, take

$$\begin{aligned} \underline{y}' &= \underline{x} + \delta(\underline{y} - \underline{x}) \\ \underline{z}' &= \underline{x} + \delta(\underline{z} - \underline{x}). \end{aligned} \tag{10}$$

By the convexity of X, we have that $\underline{y}',\underline{z}' \in X$, and note that $||\underline{y}' - \underline{x}|| \leqslant \delta||\underline{y} - \underline{x}||$, and likewise for $\underline{z}'$. Hence, for arbitrarily small values of δ, $\underline{y}',\underline{z}'$ are arbitrarily close to $\underline{x}$.

Note additionally,

$$\begin{aligned} \alpha\underline{y}' + (1-\alpha)\underline{z}' &= \alpha(\underline{x} + \delta(\underline{y} - \underline{x})) + (1-\alpha)(\underline{x} + \delta(\underline{z} - \underline{x})) \\ &= \underline{x} + \delta\left(\alpha(\underline{y} - \underline{x}) + (1-\alpha)(\underline{z} - \underline{x})\right) \\ &= \underline{x} + \delta\left(\alpha\underline{y} + (1-\alpha)\underline{z} - \underline{x}\right) \\ &= \underline{x} + \delta(0) = \underline{x} \end{aligned} \tag{11}$$

Hence, by the previous observation, we have $f(\underline{x}) < \max(f(\underline{y}'), f(\underline{z}'))$. But if we may take $\underline{y}',\underline{z}'$ arbitrarily close to $\underline{x}$, this implies that $\underline{x}$ *cannot* be a local maximum of f.

- *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ and A be an $n \times m$ matrix, $\underline{b} \in \mathbb{R}^n$. If f is convex, show that the function $g : \mathbb{R}^m \mapsto \mathbb{R}$ is convex, where*

$$g(\underline{x}) = f(A\underline{x} + \underline{b}). \tag{12}$$

Answer: For any $\underline{x},\underline{y} \in \mathbb{R}^n$ and $0 \leqslant \alpha \leqslant 1$

$$\begin{aligned} g(\alpha\underline{x} + (1-\alpha)\underline{y}) &= f(A(\alpha\underline{x} + (1-\alpha)\underline{y}) + \underline{b}) \\ &= f(\alpha(A\underline{x} + \underline{b}) + (1-\alpha)(A\underline{y} + \underline{b})) \\ &\leqslant \alpha f(A\underline{x} + \underline{b}) + (1-\alpha) f(A\underline{y} + \underline{b}) \\ &= \alpha g(\underline{x}) + (1-\alpha) g(\underline{y}). \end{aligned} \tag{13}$$

1.3 Unconstrained Optimization

Identify all local maxima and local minima of

$$f(x,y) = (7x + 9y + 8)e^{-(x^2+y^2)}. \tag{14}$$

Identify the global minima and maxima as well.

Answer: Any local extrema must satisfy $\nabla f(x,y) = 0$. This gives

$$\begin{aligned} e^{-(x^2+y^2)}(7 - 16x - 14x^2 - 18xy) &= 0 \\ e^{-(x^2+y^2)}(9 - 16y - 14xy - 18y^2) &= 0. \end{aligned} \tag{15}$$

The exponentials may be dropped, as they are strictly positive functions for finite x, y. The remaining system, $7-16x-14x^2-18xy=0$ and $9-16y-14xy-18y^2=0$ may be solved by whatever preferred means to yield $(x,y)=(-7/10,-9/10)$ or $(x,y)=(7/26,9/26)$. It remains to characterize these stationary points. One way to approach this is to calculate the Hessian at each of these points, and determine whether it is positive definite or negative definite. This is a somewhat irritating calculation, but immanently doable.

As an alternative, consider the following: note that the exponential term in f dominates as $x^2+y^2\to\infty$, hence $f(x,y)\to 0$ as $(x,y)\to\infty$. Another way to consider this is that outside a circle of sufficiently large radius, the value of the function is arbitrarily close to 0. As the function diminishes in this way, i.e., does not increase to ∞ or decrease to $-\infty$ in any direction, and is additionally continuous everywhere, there must be at least one point that attains a global maximal value that is strictly positive (because there are points (x_p,y_p) for which $f(x_p,y_p)>0$) and at least one point that attains a global minimum value that is strictly negative (because there are points (x_n,y_n) for which $f(x_n,y_n)<0$). Such a global minimum and maximum must satisfy the $\nabla f=0$ condition. There are only two such points, so one of $(-7/10,-9/10)$ and $(7/26,9/26)$ must be the global maximum, the other the global minimum.

We may compute that $f(-7/10,-9/10)\approx -1.36$ and $f(7/26,9/26)\approx 10.73$. By the above reasoning, we see that $(-7/10,-9/10)$ is a global minimum and $(7/26,9/26)$ is a global maximum.

Similarly, identify all minima and maxima (local and global) of

$$f(\underline{x})=(\underline{u}^T\underline{x}-c_0)e^{-\underline{x}^T\underline{x}} \tag{16}$$

for $\underline{u}\in\mathbb{R}^n, c_0\in\mathbb{R}$ *with* $\underline{u}\neq 0, c_0\neq 0$.

Answer: This is simply a generalization of the previous problem. Note that by the product rule we have

$$\begin{aligned}\nabla f(\underline{x}) &= (\nabla[(\underline{u}^{\mathrm{T}}\underline{x}-c_0)])e^{-\underline{x}^{\mathrm{T}}\underline{x}}+(\underline{u}^{\mathrm{T}}\underline{x}-c_0)(\nabla[e^{-\underline{x}^{\mathrm{T}}\underline{x}}])\\ &= \underline{u}e^{-\underline{x}^{\mathrm{T}}\underline{x}}+(\underline{u}^{\mathrm{T}}\underline{x}-c_0)(-2\underline{x}e^{-\underline{x}^{\mathrm{T}}\underline{x}})\\ &= \left(\underline{u}-2\underline{x}(\underline{u}^{\mathrm{T}}\underline{x}-c_0)\right)e^{-\underline{x}^{\mathrm{T}}\underline{x}}.\end{aligned} \tag{17}$$

It follows from the above that any solution to $\nabla f=0$ satisfies $\underline{u}=2\underline{x}(\underline{u}^{\mathrm{T}}\underline{x}-c_0)$. Note that for any $\underline{x}$, $\underline{u}^{\mathrm{T}}\underline{x}-c_0$ is a scalar value; the condition therefore is that $\underline{u}$ is a scalar multiple of $\underline{x}$. As $\underline{u}$ is non-zero by assumption, it follows that said scalar is non-zero and any such optimal $\underline{x}$ is non-zero.

Instead of trying to solve the equation for the individual components of $\underline{x}$, it is convenient to take $\underline{x}=\lambda\underline{u}$, which gives:

$$\begin{aligned}0 &= \underline{u}-2\underline{x}(\underline{u}^{\mathrm{T}}\underline{x}-c_0)\\ &= \underline{u}-2\lambda\underline{u}(\lambda||\underline{u}||^2-c_0)\\ &= \left(1-2\lambda(\lambda||\underline{u}||^2-c_0)\right)\underline{u}.\end{aligned} \tag{18}$$

Solving $1-2\lambda(\lambda||\underline{u}||^2-c_0)=0$ yields

$$\lambda^*_{\pm}=\frac{c_0\pm\sqrt{c_0^2+2||\underline{u}||^2}}{2||\underline{u}||^2}, \tag{19}$$

and solutions of $\underline{x}_+^* = \lambda_+^* \underline{u}$, $\underline{x}_-^* = \lambda_-^* \underline{u}$.

It remains to characterize these potential extrema: minima, maxima, saddle points, etc. One entirely feasible way to do this is to determine the Hessian at each of these points, and calculate whether it is positive definite, negative definite, or neither. This is perhaps not the easiest way to do it, but since enough people asked about it here's how you might do it.

Differentiation of the components of the gradient can be used to derive the following:

$$\nabla^2 f(\underline{x}) = 2e^{-\underline{x}^{\mathrm{T}}\underline{x}} \left[(2\underline{x}\underline{x}^{\mathrm{T}} - \mathbf{I})(\underline{u}^{\mathrm{T}}\underline{x} - c_0) - (\underline{u}\underline{x}^{\mathrm{T}} + \underline{x}\underline{u}^{\mathrm{T}}) \right], \tag{20}$$

and hence

$$\nabla^2 f(\lambda \underline{u}) = 2e^{-\lambda^2 ||\underline{u}||^2} \left[(2\lambda^2 \underline{u}\underline{u}^{\mathrm{T}} - \mathbf{I})(\lambda ||\underline{u}||^2 - c_0) - 2\lambda \underline{u}\underline{u}^{\mathrm{T}} \right]. \tag{21}$$

Note that for the above, we have for $\underline{u}$ and any vector $\underline{v}$ orthogonal to $\underline{u}$:

$$\begin{aligned} \nabla^2 f(\lambda \underline{u})\underline{u} &= 2e^{-\lambda^2 ||\underline{u}||^2} \left[(2\lambda^2 ||\underline{u}||^2 - 1)(\lambda ||\underline{u}||^2 - c_0) - 2\lambda ||\underline{u}||^2 \right] \underline{u} \\ \nabla^2 f(\lambda \underline{u})\underline{v} &= -2e^{-\lambda^2 ||\underline{u}||^2} (\lambda ||\underline{u}||^2 - c_0)\underline{v}. \end{aligned} \tag{22}$$

The above gives that $\underline{u}$ along with a set of vectors $\underline{v}_i$ that are orthogonal to $\underline{u}$ and orthogonal to each other yields an orthogonal basis of eigenvectors. For $\lambda = \lambda_\pm^*$, it can be shown from the above that the corresponding eigenvalues are all positive for $\lambda = \lambda_-^*$ and all negative for $\lambda = \lambda_+^*$. Hence we have that $\underline{x}_-^*$ is a minimum, and $\underline{x}_+^*$ is a maximum.

But we can also extend the argument from the first section: as $\underline{x} \to \infty$, $f(\underline{x}) \to 0$ (more precisely, outside arbitrarily large spheres, f is arbitrarily close to 0) because of the exponential term dominating. There must be a strictly positive global maximum that satisfies the $\nabla f = 0$ condition, and a strictly positive global minimum. The only options are $\underline{x}_+^*$ and $\underline{x}_-^*$, and it can be shown that $f(\underline{x}_-^*)$ is negative and $f(\underline{x}_+^*)$ is positive.

1.4 Taylor Series

Give the second order Taylor polynomial of

$$-\cos(3x + 4y) \tag{23}$$

around the point $(x, y) = (0, 0)$. What kind of stationary point occurs here, i.e., maxima or minima, strict, local, etc.?

Answer: This really comes down to determining the appropriate first and second derivatives of the function at $(x, y) = (0, 0)$. In particular,

$$\begin{aligned} f_2(x,y) &= -1 + \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 9 & 12 \\ 12 & 16 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= -1 + (1/2)(9x^2 + 12xy + 12xy + 16y^2). \end{aligned} \tag{24}$$

Note that computing the eigenvalues of the Hessian as given above yields eigenvalues of 25 and 0. This is a non-strict local minimum. This can be seen even more clearly noting that the smallest

possible value of $-\cos$ is -1, and that this minimal value will occur at all points along the line $3x+4y=0$ passing through the origin.

Do the same for

$$(x^2+y^2)\cos(3x+4y). \tag{25}$$

Answer: Computing the relevant derivatives again, we have

$$\begin{aligned} f_2(x,y) &= 0+\begin{bmatrix}0 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+\frac{1}{2}\begin{bmatrix}x & y\end{bmatrix}\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} \\ &= x^2+y^2. \end{aligned} \tag{26}$$

The eigenvalues of the Hessian at $(0,0)$ 1 and 1 - this represents a strict local minimum of the function. It is not a global minimum (why?).

2 A Look at Newton's Method

Consider the function $f:\mathbb{R}^n \mapsto \mathbb{R}$ defined by

$$f(\underline{x}) = (||\underline{x}||^2-1)^2 = (\underline{x}^T\underline{x}-1)^2. \tag{27}$$

1. *Compute the gradient $\nabla f(\underline{x})$; show that the gradient at any non-stationary point $\underline{x}$ is in the direction of $\underline{x}$, i.e., $\nabla f(\underline{x}) = \beta_{\underline{x}}\underline{x}$ for some scalar function $\beta_{\underline{x}}$.*

   Answer: Computing the partials,

   $$\partial f/\partial x_i(\underline{x}) = (2x_i)(2)(\underline{x}^{\mathrm{T}}\underline{x}-1) = 4(||\underline{x}||^2-1)x_i. \tag{28}$$

   Summarizing, we have $\nabla f(\underline{x}) = 4(\underline{x}^{\mathrm{T}}\underline{x}-1)\underline{x}$.

2. *Show under Steepest Descent, where the stepsize is chosen to produce the maximum descent (i.e., minimum function value in the next step), the function is minimized in a single step.*

   Answer: Note that the function is strictly non-negative, achieving a minimum value of 0 for any $\underline{x}$ such that $||\underline{x}|| = 1$. If $||\underline{x}|| \neq 1$ and $\underline{x} \neq 0$ ($\underline{x}=0$ is a local maximum, giving a gradient of 0), consider a new point defined by some stepsize $\alpha \geqslant 0$:

   $$\underline{x}' = \underline{x} - \alpha\nabla f(\underline{x}) = (1-4\alpha(\underline{x}^{\mathrm{T}}\underline{x}-1))\underline{x}. \tag{29}$$

   Any stepsize that brings $\underline{x}'$ to $||\underline{x}'|| = 1$ will minimize the function completely (yielding $f(\underline{x}') = 0$). This leads to taking $(1-4\alpha(\underline{x}^{\mathrm{T}}\underline{x}-1)) = 1/||\underline{x}||$, or

   $$\alpha^* = \frac{1}{4||\underline{x}||(1+||\underline{x}||)}. \tag{30}$$

   There is at least one other choice for α^* but all may be found systematically by minimizing $f(\underline{x}')$ with respect to α.

3. *Compute the Hessian $\nabla^2 f(\underline{x})$; show that the Hessian at any point $\underline{x}$ is given by*

$$\nabla^2 f(\underline{x}) = 4(\underline{x}^T\underline{x} - 1)\mathbf{I} + 8\underline{x}\underline{x}^T, \tag{31}$$

where $\underline{x}\underline{x}^T$ is the matrix such that the (i,j)-th entry is $x_i x_j$.

Answer: We have

$$\partial f/\partial x_i(\underline{x}) = 4x_i(\underline{x}^{\mathrm{T}}\underline{x} - 1). \tag{32}$$

The diagonal elements of the Hessian will be given by

$$\begin{aligned}\partial^2 f/\partial x_i^2(\underline{x}) &= 4(\partial/\partial x_i[x_i])(\underline{x}^{\mathrm{T}}\underline{x} - 1) + 4x_i(\partial/\partial x_i[(\underline{x}^{\mathrm{T}}\underline{x} - 1)] \\ &= 4(\underline{x}^{\mathrm{T}}\underline{x} - 1) + 8x_i^2.\end{aligned} \tag{33}$$

The off-diagonal elements of the Hessian will be given by

$$\begin{aligned}\partial^2 f/\partial x_j\partial x_i(\underline{x}) &= 4(\partial/\partial x_j[x_i])(\underline{x}^{\mathrm{T}}\underline{x} - 1) + 4x_i(\partial/\partial x_j[(\underline{x}^{\mathrm{T}}\underline{x} - 1)] \\ &= 8x_i 8x_j.\end{aligned} \tag{34}$$

This is effectively a diagonal matrix with $4(\underline{x}^{\mathrm{T}}\underline{x} - 1)$ along the diagonal added to 8 times a matrix of the products $[x_i x_j]$. This yields the indicated formula.

4. *Show that for $\underline{x} \neq 0$, $\underline{x}$ is an eigenvector of $\nabla^2 f(\underline{x})$. What is the eigenvalue?*

Answer: For any $\underline{x}$, we have

$$\begin{aligned}\nabla^2 f(\underline{x})\underline{x} &= 4(\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{x} + 8\underline{x}\underline{x}^{\mathrm{T}}\underline{x} \\ &= 4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{x}.\end{aligned} \tag{35}$$

Hence, for any non-zero $\underline{x}$, we see that $\underline{x}$ is an eigenvector of $\nabla^2 f(\underline{x})$ with eigenvalue $4(3||\underline{x}||^2 - 1)$.

5. *Under what conditions is the Hessian as above invertible? Hint: For $\underline{x}^T\underline{x} \neq 1$, show that any non-zero $\underline{v}$ that is orthogonal to $\underline{x}$ is an eigenvector of $\nabla^2 f(\underline{x})$ with non-zero eigenvalue. When is $\underline{x}$ an eigenvector with non-zero eigenvalue?*

Answer: For $\underline{x} \neq 0$, let $\underline{v}_1, \ldots, \underline{v}_{n-1}$ be a set of non-zero vectors that are orthogonal to $\underline{x}$, and mutually orthogonal among each other. Note, $\{\underline{x}, \underline{v}_1, \ldots, \underline{v}_{n-1}\}$ forms a basis of $\mathbb{R}^n$.

We have

$$\nabla^2 f(\underline{x})\underline{x} = 4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{x}, \tag{36}$$

and additionally, for any i,

$$\begin{aligned}\nabla^2 f(\underline{x})\underline{v}_i &= 4(\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{v}_i + 8\underline{x}\underline{x}^{\mathrm{T}}\underline{v}_i \\ &= 4(\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{v}_i.\end{aligned} \tag{37}$$

Hence we see that for $\underline{x} \neq 0$, there is an orthogonal system of eigenvectors of $\nabla^2 f(\underline{x})$ with non-zero eigenvalues as long as $||\underline{x}||^2 \neq 1/3$ and $||\underline{x}||^2 \neq 1$.

Note, in the case that $\underline{x} = 0$, we have $\nabla^2 f(\underline{x}) = 4(-1)\mathbf{I} = -4\mathbf{I}$, which is invertible.

Hence, if $\underline{x}$ is not on the surface of the sphere of radius 1 or radius $1/\sqrt{3}$, the Hessian is invertible.

6. *When the Hessian is invertible, use the fact that $\underline{x}$ is an eigenvector of $\nabla^2 f(\underline{x})$ to compute $[\nabla^2 f(\underline{x})]^{-1}\underline{x}$.*

   Answer: We have

$$\nabla^2 f(\underline{x})\underline{x} = 4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)\underline{x}. \tag{38}$$

   Hitting each side with $[\nabla^2 f(\underline{x})]^{-1}$, when it exists:

$$\underline{x} = 4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)[\nabla^2 f(\underline{x})]^{-1}\underline{x}, \tag{39}$$

   or

$$[\nabla^2 f(\underline{x})]^{-1}\underline{x} = \frac{1}{4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)}\underline{x}. \tag{40}$$

7. *Show that applying one step of 'pure' Newton's method to any initial point $\underline{x}$ produces a point in the direction of $\underline{x}$, i.e.,*

$$\underline{x} - [\nabla^2 f(\underline{x})]^{-1}\nabla f(\underline{x}) = \gamma_{\underline{x}}\underline{x}, \tag{41}$$

   for some scalar function $\gamma_{\underline{x}}$, as long as $||\underline{x}||^2 \neq 1/3$ and $||\underline{x}||^2 \neq 1$.

   Answer: The above conditions ensure that the Hessian is invertible and that Newton's method may be applied at all. We then have

$$\begin{aligned} \underline{x} - [\nabla^2 f(\underline{x})]^{-1}\nabla f(\underline{x}) &= \underline{x} - 4(\underline{x}^{\mathrm{T}}\underline{x} - 1)[\nabla^2 f(\underline{x})]^{-1}\underline{x} \\ &= \underline{x} - \frac{4(\underline{x}^{\mathrm{T}}\underline{x} - 1)}{4(3\underline{x}^{\mathrm{T}}\underline{x} - 1)}\underline{x} \\ &= \left(\frac{2\underline{x}^{\mathrm{T}}\underline{x}}{3\underline{x}^{\mathrm{T}}\underline{x} - 1}\right)\underline{x}. \end{aligned} \tag{42}$$

8. *Show that if Newton's method is started with an initial point $\underline{x}_0$ such that $||\underline{x}_0||^2 < 1/3$, the method will converge to $\underline{x}^* = 0$. If it is started with $||\underline{x}_0||^2 > 1$, the method will converge to $\underline{x}^* = \underline{x}/||\underline{x}||$. What happens if $1/3 < ||\underline{x}_0||^2 < 1$? Why does Newton's method 'fail' if $||\underline{x}_0||^2 = 1/3$? Hint: What is the reasoning behind the construction of Newton's method, what is it trying to accomplish?*

   Note: It has been brought to my attention that I completely dropped the ball on this problem. Full credit will be awarded for any attempt.

   Answer: We have established that 'one newton iteration' follows the map:

$$\mathrm{NM}(\underline{x}) = \left(\frac{2\underline{x}^{\mathrm{T}}\underline{x}}{3\underline{x}^{\mathrm{T}}\underline{x} - 1}\right)\underline{x}, \tag{43}$$

   which amounts to a dilation or a contraction of $\underline{x}$, towards or away from the origin. Note, if the scaling factor is negative, it will actually flip the direction from $\underline{x}$ to $-\underline{x}$. Studying the convergence of Newton's method in this case amounts to studying the effect of repeated scaling according to the above map. (Note, interestingly, the map is in fact well defined when $||\underline{x}|| = 1$, even though the Hessian is not invertible (and hence, Newton's method cannot be applied.)

   There are a couple of regions of interest:

- $\mathbf{1} < ||\underline{\mathbf{x}}||^{\mathbf{2}}$: Over this region, the scaling factor is positive, but less than 1. This amounts to a contraction along the ray $\underline{x}$. Additionally, however, it can be shown that the contraction is never enough to actually bring the points inside the ball of radius 1, i.e., $||\mathrm{NM}(\underline{x})|| > 1$ over this region. Repeated iterations of Newton's method contract points towards the surface of the sphere, converging in the limit.
- $\mathbf{1/3} < ||\underline{\mathbf{x}}||^{\mathbf{2}} < \mathbf{1}$: Over this region, the scaling factor is positive, and strictly greater than 1. This is a dilation along the ray $\underline{x}$. In fact, it can be shown that over this region, the scaling factor is sufficient to throw the point outside the sphere, i.e., $||\mathrm{NM}(\underline{x})|| > 1$. As such, after one iteration of Newton's method in this region, the next iterate lies in the previous region, and the previous analysis applies.
- $\mathbf{1/4} < ||\underline{\mathbf{x}}||^{\mathbf{2}} < \mathbf{1/3}$: Over this region, the scaling factor is negative, and strictly less than -1. This is a reflection and a dilation along the ray $\underline{x}$. In fact, it can be shown that over this region again, the scaling factor is sufficient to throw the point outside the sphere, i.e., $||\mathrm{NM}(\underline{x})|| > 1$. Again, after this point, the analysis of the first case applies.
- $\mathbf{1/5} < ||\underline{\mathbf{x}}||^{\mathbf{2}} < \mathbf{1/4}$: Over this region, the scaling factor is negative and strictly less than -1, but is insufficient to throw the point outside the sphere, i.e., $||\underline{x}|| < ||\mathrm{NM}(\underline{x})|| < 1$. It can be shown that over this region, finitely many iterations of Newton's method are sufficient to reach a point in the previous region, which will throw the iterates outside the sphere, at which point the analysis of prior cases applies. **Note:** This region can actually be further subdivided into annuli of increasing complicated radii, based on the number of iterations it takes to be thrown outside the sphere. But for simplicity, we give only the uppermost region.
- $\mathbf{0} < ||\underline{\mathbf{x}}||^{\mathbf{2}} < \mathbf{1/5}$: Over this region, the scaling factor is negative, but greater than -1; this amounts to a reflection and a contraction, so that $||\mathrm{NM}(\underline{x})|| < ||\underline{x}||$. Repeated iterations of Newton's method continues to contract this region, so that any initial guess $\underline{x}_0$ in this region is contracted to $\underline{x}^* = 0$.

Hence, by the above cases, analysis of the actual convergence of Newton's method can be restricted to the first and last cases, $||\underline{x}||^2 > 1$ and $||\underline{x}||^2 < 1/5$. It can further be shown that for $||\underline{x}||$ arbitrarily close to 1 (but greater than 1, we have

$$||\mathrm{NM}(\underline{x})|| \approx 1 + \frac{3}{2}(||\underline{x}|| - 1)^2 + O((||\underline{x}|| - 1)^3). \tag{44}$$

In the limit as $\underline{x}$ approaches the surface of the sphere then, we have

$$\frac{||\mathrm{NM}(\underline{x})|| - 1}{(||\underline{x}|| - 1)^2} \leqslant \frac{3}{2} + o(1). \tag{45}$$

This gives quadratic convergence to the sphere.

In the limit as $\underline{x}$ approaches 0, we have

$$||\mathrm{NM}(\underline{x})|| \approx 2||\underline{x}||^3 + O(||\underline{x}||^5), \tag{46}$$

Hence, in the limit as $\underline{x}$ approaches the origin (a local maximum of f), we have

$$\frac{||\mathrm{NM}(\underline{x})||}{||\underline{x}||^3} \leqslant 2 + o(1). \tag{47}$$

This gives cubic convergence to the origin.

Note: Newton's method fails to work at $||\underline{x}||^2 = 1/3$, because this represents an inflection point of the problem, a point at which the second order variation vanishes. As such, there is no quadratic approximation to f at such points - any such approximation degenerates to a plane. As Newton's method depends on sequentially minimizing quadratic approximations, it cannot be implemented at such a point.

Additional Commentary: It is additionally interesting to consider the behavior of Newton's method on the boundaries between the indicated regions. For instance, if $||\underline{x}||^2 = 1$, then $\text{NM}(\underline{x}) = \underline{x}$, and the point is stationary (and is already a solution). If $||\underline{x}||^2 = 1/3$, NM is undefined. If $||\underline{x}||^2 = 1/4$, $\text{NM}(\underline{x}) = -2\underline{x}$ and Newton's method converges in one step. If $||\underline{x}||^2 = 1/5$, then $\text{NM}(\underline{x}) = -\underline{x}$, and Newton's method never converges, infinitely flipping back and forth. We may further consider the pre-images of such points, points $\underline{x}'$ such that $||\text{NM}(\underline{x}')||^2 = 1, 1/3, 1/4, 1/5$. Such points demonstrate the indicated behavior after two iterations of Newton's method. However, any solution to $||\text{NM}(\underline{x}')||^2 = 1$ must have $||\underline{x}'||^2 = 1$ or $1/4$, cases that have already been covered. The only feasible solutions to $||\text{NM}(\underline{x}')||^2 = 1/5$ satisfy $||\underline{x}'||^2 = 1/5$. But the solutions to $||\text{NM}(\underline{x}')||^2 = 1/3$ or $1/4$ are non-trivial, and can in turn be used to generate further pre-images - generating an infinite sequence of nested spheres, with interesting finite iteration behavior of Newton's method on the surface of these spheres.

9. *In this case, for this specific function, Optimal Steepest Descent finds the minimum in a single step. Newton's Method, however, converges (when it converges) geometrically, no matter how close to a minimum it is started. Discuss. Hint: How does the Hessian behave at the minima of f?*

   Answer: As the answer to this question depended on understanding the answer to the previous question, again any reasonable attempt at a solution will be given full credit. Some things worth noting: because of the structure of the problem, the negative gradient from any point $\underline{x} \neq 0$ points in the direction of a minimum. The 1-dimensional problem along this direction is a 4-th order polynomial, which can be minimized to yield the absolute minimum of the function. However, Newton's method attempts to treat the problem as a 2-nd order polynomial, and minimize that as an approximation - in general (barring a few special cases) this cannot lead to minimization of the full problem because it is throwing information (higher order variation) away. It is also worth noting that at the minima, $||\underline{x}|| = 1$, the Hessian is given by $\nabla^2 f(\underline{x}) = 8\underline{x}\underline{x}^{\text{T}}$, which is not invertible - the problem becomes singular at these points *when approaching the minimum along a direction not in the direction of the origin.* As the formulation given above *only* approaches the minima along rays from the origin, this problem is effectively avoided, as we are never subject to the behavior of the inverse of the Hessian on the space orthogonal to $\underline{x}$.

3 Solving Linear Systems

Let A be an $m \times n$ real matrix, and $\underline{b} \in \mathbb{R}^m$. Consider the linear system of equations given by

$$A\underline{x} = \underline{b}. \tag{48}$$

*If $n = m$ and A is invertible, we have the unique solution $\underline{x}^* = A^{-1}\underline{b}$. If $m < n$ and the rows of A are linearly independent, there is no unique solution and the system is **underspecified**. If $m > n$ and the columns of A are linearly independent, the system may be **overspecified**, and admit no solutions at all.*

Even in the simplest case of $n = m$ with A invertible, computing the solution may be particularly expensive, especially if A is quite large. So in general, consider computing $\underline{x}^*$ as the minimizer of the following function:

$$f(\underline{x}) = \frac{1}{2}||A\underline{x} - \underline{b}||^2. \tag{49}$$

1. *What is the gradient, $\nabla f(\underline{x})$? Hint: Consider writing the norm of v as $\underline{v}^T\underline{v}$ and expanding.*

   Answer: Consider writing $f(\underline{x}) = (1/2)(A\underline{x} - \underline{b})^{\mathrm{T}}(A\underline{x} - \underline{b})$. We may expand this as

   $$f(\underline{x}) = (1/2)\underline{x}^{\mathrm{T}}A^{\mathrm{T}}A\underline{x} - \underline{b}^{\mathrm{T}}A\underline{x} + (1/2)\underline{b}^{\mathrm{T}}\underline{b}. \tag{50}$$

   This is a quadratic form ($A^{\mathrm{T}}A$ is a square, symmetric matrix), and hence the gradient may be taken easily as

   $$\nabla f(\underline{x}) = A^{\mathrm{T}}A\underline{x} - A^{\mathrm{T}}\underline{b} = A^{\mathrm{T}}(A\underline{x} - \underline{b}). \tag{51}$$

2. *Argue that for any $A, \underline{b}$, the objective function has a unique minimal value (all minimizers realize the same value of f), and no maximal value.*

   Answer: This problem may be approached a number of ways - one would be to argue from the convexity of the function (in conjunction with one of the convexity properties proven in Section 1). This would grant the uniqueness of the minimal values. There is no maximal value as for any $\underline{x}$ that does not satisfy $A\underline{x} = 0$ (and there must be such $\underline{x}$ for non-trivial A), then $f(\lambda\underline{x})$ may be made arbitrarily large by taking large values of λ (see the above expansion if this is not clear).

   Another approach to the uniqueness of the minimal value would be to consider the necessary condition for a minima, i.e., $\nabla f(\underline{x}) = 0$. In this case, we have for any minimizer $\underline{x}^*$:

   $$(A^{\mathrm{T}}A)\underline{x}^* = A^{\mathrm{T}}\underline{b}. \tag{52}$$

   If the matrix $A^{\mathrm{T}}A$ is invertible, then the above grants a unique minimizer $\underline{x}^*$ - if the minimizer is unique, the minimal value is necessarily unique.

   Consider the case of $A^{\mathrm{T}}A$ non-invertible. In this case, note that if $A^{\mathrm{T}}A\underline{v} = 0$, then $\underline{v}^{\mathrm{T}}A^{\mathrm{T}}A\underline{v} = 0$, and hence $(A\underline{v})^{\mathrm{T}}(A\underline{v}) = ||A\underline{v}||^2 = 0$, and $A\underline{v} = 0$. Let $\underline{x}_0^*$ be a given solution to $(A^{\mathrm{T}}A)\underline{x}_0^* = A^{\mathrm{T}}\underline{b}$. *Any* solution $\underline{x}^*$ to $(A^{\mathrm{T}}A)\underline{x}^* = A^{\mathrm{T}}\underline{b}$ may be expressed as $\underline{x}^* = \underline{x}_0^* + \underline{v}$ for some $\underline{v}$ satisfying $A^{\mathrm{T}}A\underline{v} = 0$, or equivalently $A\underline{v} = 0$.

We then have

$$\begin{aligned} f(\underline{x}^*) &= (1/2)(\underline{x}^*)^{\mathrm{T}}\left(A^{\mathrm{T}}A\underline{x}^*\right) - \underline{b}^{\mathrm{T}}A\underline{x}^* + (1/2)\underline{b}^{\mathrm{T}}\underline{b} \\ &= (1/2)(\underline{x}^{\mathrm{T}})^* A^{\mathrm{T}}\underline{b} - \underline{b}^{\mathrm{T}}A\underline{x}^* + (1/2)\underline{b}^{\mathrm{T}}\underline{b} \\ &= -(1/2)\underline{b}^{\mathrm{T}}A\underline{x}^* + (1/2)\underline{b}^{\mathrm{T}}\underline{b} \\ &= -(1/2)\underline{b}^{\mathrm{T}}\left(A\underline{x}^* - \underline{b}\right) \\ &= -(1/2)\underline{b}^{\mathrm{T}}\left(A(\underline{x}_0^* + \underline{v}) - \underline{b}\right) \\ &= -(1/2)\underline{b}^{\mathrm{T}}\left(A\underline{x}_0^* - \underline{b}\right). \end{aligned} \tag{53}$$

As $\underline{x}_0^*$ is fixed, the above shows that any minimizer $\underline{x}^*$ yields the same functional value, $f(\underline{x}_0^*)$. The minimal functional value is unique.

3. *Consider applying Steepest Descent. What is the optimal stepsize to use at a given point $\underline{x}$?*

   Answer: For a quadratic function of the form $f(\underline{x}) = (1/2)\underline{x}^{\mathrm{T}}Q\underline{x} - \underline{b}^{\mathrm{T}}\underline{x}$, the optimal stepsize for steepest descent is given by

$$\alpha_k^* = \frac{\nabla f(\underline{x}_k)^{\mathrm{T}}\nabla f(\underline{x}_k)}{\nabla f(\underline{x}_k)^{\mathrm{T}}Q\nabla f(\underline{x}_k)}. \tag{54}$$

   Taking $Q = A^{\mathrm{T}}A$, the above may be simplified to

$$\alpha_k^* = \frac{||\nabla f(\underline{x}_k)||^2}{\nabla f(\underline{x}_k)^{\mathrm{T}}A^{\mathrm{T}}A\nabla f(\underline{x}_k)} = \frac{||\nabla f(\underline{x}_k)||^2}{||A\nabla f(\underline{x}_k)||^2} = \frac{||A^{\mathrm{T}}(A\underline{x}_k - \underline{b})||^2}{||AA^{\mathrm{T}}(A\underline{x}_k - \underline{b})||^2}. \tag{55}$$

4. *Apply steepest descent with the above stepsize to the case of:*

$$A = \begin{bmatrix} 5 & -3 & 3 & 2 & 0 \\ 2 & -2 & -5 & 3 & -5 \\ 3 & 3 & -3 & -3 & 2 \\ 0 & 3 & 0 & 5 & 3 \\ -1 & -3 & -4 & 0 & 4 \end{bmatrix}, \ \underline{b} = \begin{bmatrix} -5 \\ -1 \\ 0 \\ 1 \\ -4 \end{bmatrix} \tag{56}$$

   - *Give the first 6 iterates $\{\underline{x}_k\}$ of steepest descent, starting at $\underline{x}_0 = 0$.*
   - *Using this methodology, estimate the solution $\underline{x}^* = A^{-1}\underline{b}$ to five decimal places using this method.*

   Answer: We have the iteration for steepest descent with optimal stepsize:

$$\underline{x}_{k+1} = \underline{x}_k - \frac{||A^{\mathrm{T}}(A\underline{x}_k - \underline{b})||^2}{||AA^{\mathrm{T}}(A\underline{x}_k - \underline{b})||^2}A^{\mathrm{T}}(A\underline{x}_k - \underline{b}). \tag{57}$$

   The first iterates $\underline{x}_0, \ldots, \underline{x}_6$ are given by (the rows, from top to bottom)

$$\begin{pmatrix} 0. & 0. & 0. & 0. & 0. \\ -0.512797 & 0.713457 & 0.133773 & -0.178364 & -0.178364 \\ -0.482427 & 0.723256 & 0.0633021 & -0.0581625 & -0.399535 \\ -0.52539 & 0.819546 & 0.0761291 & -0.0569724 & -0.404608 \\ -0.518851 & 0.822135 & 0.0667826 & -0.0427661 & -0.431146 \\ -0.52246 & 0.834878 & 0.0682455 & -0.0430397 & -0.431454 \\ -0.521283 & 0.835305 & 0.0670368 & -0.0413418 & -0.434787 \end{pmatrix}. \tag{58}$$

Depending on how you are estimating the rate of convergence and how you choose to round, an estimate of $\underline{x}^*$ accurate to five decimal places is given by

$$\underline{x}^* \approx \underline{x}_{12} = (-0.521338, 0.837333, 0.0670587, -0.0412391, -0.435271). \tag{59}$$

5. *Is Steepest Descent the best algorithm to apply to this problem? Discuss.*

   Answer: Note that this is specifically solving for the minimizer of a quadratic form. It can be difficult to know in advance how quickly Steepest Descent is going to converge, even in the optimal case - and we know that the best we can hope for in general is linear or geometric convergence of the error. A better choice of algorithm might be the conjugate direction method, which is guaranteed to solve for the minimizer of a quadratic form in no more than n steps. Newton's method could potentially solve the problem in a single step - however, this would require computation of the inverse of the Hessian, i.e., $[A^{\mathrm{T}}A]^{-1}$. But this is no more difficult than inverting A (when A is invertible), at which point we might simply utilize the formula $\underline{x}^* = A^{-1}\underline{b}$. Steepest descent is not a *bad* algorithm, but in particular the conjugate direction method is worth considering.

6. *Define the matrix A' to be the first two rows of A, and $\underline{b}' \in \mathbb{R}^2$ to be the first two rows of $\underline{b}$. Starting with $\underline{x}'_0 = 0$, give the first 6 iterates of steepest descent on*

$$f(\underline{x}) = \frac{1}{2}||A'\underline{x} - \underline{b}'||^2. \tag{60}$$

   Estimate a solution to the above to five decimal places.

   Answer: Utilizing the steepest descent iteration as before, taking the first two rows of A and $\underline{b}$ as A' and $\underline{b}'$ produces the following iterates $\underline{x}_0, \ldots, \underline{x}_6$:

$$\begin{pmatrix} 0. & 0. & 0. & 0. & 0. \\ -0.51363 & 0.323396 & -0.190233 & -0.247303 & 0.0951166 \\ -0.515274 & 0.312137 & -0.286126 & -0.214285 & 0.0185796 \\ -0.535847 & 0.325091 & -0.293745 & -0.22419 & 0.0223894 \\ -0.535913 & 0.32464 & -0.297586 & -0.222868 & 0.0193238 \\ -0.536737 & 0.325158 & -0.297892 & -0.223264 & 0.0194764 \\ -0.53674 & 0.32514 & -0.298045 & -0.223211 & 0.0193536 \end{pmatrix}. \tag{61}$$

   A sufficient estimator is given by the 10th iterate or so,

$$\underline{x}^* \approx \underline{x}_{10} = (-0.536774, 0.325161, -0.298064, -0.223226, 0.0193548). \tag{62}$$

4 Two-Pronged Descent

In this section, we consider two extensions of the steepest descent algorithm. In each case, we will apply the algorithms to a function of the following form:

$$f(\underline{x}) = \frac{1}{2}\underline{x}^T Q\underline{x} - \underline{b}^T\underline{x}, \tag{63}$$

for symmetric, positive definite matrix Q and vector $\underline{b}$.

We augment the steepest descent step in the following way, by not only considering movement in the direction of $-\nabla f(\underline{x}_k)$, but also movement in a secondary direction $\underline{\Delta}_k$. That is, we consider iterates of the form

$$\underline{x}_{k+1} = \underline{x}_k - \alpha_k \nabla f(\underline{x}_k) + \beta_k \underline{\Delta}_k. \tag{64}$$

We will consider two possible choices for $\underline{\Delta}_k$. But first:

1. *As a function of $\nabla f(\underline{x}_k)$ and $\underline{\Delta}_k$, what are the optimal choices of stepsizes α_k^*, β_k^* to maximize descent, i.e., minimize the value of $f(\underline{x}_{k+1})$? Argue or prove that this optimal choice (α_k^*, β_k^*) indeed maximizes descent, as opposed to representing a saddle point or a minimum of descent. Hint: The property in Ineq.* (1) *may be useful.*

   Answer: We have

$$\begin{aligned} f(\underline{x}_{k+1}) = (1/2)(&\underline{x}_k^{\mathrm{T}} Q \underline{x}_k + \alpha_k^2 \nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k) + \beta_k^2 \underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k) \\ &+ (-\alpha_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{x}_k + \beta_k \underline{\Delta}_k^{\mathrm{T}} Q \underline{x}_k - \alpha_k \beta_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k) \\ &- \underline{b}^{\mathrm{T}} (\underline{x}_k - \alpha_k \nabla f(\underline{x}_k) + \beta_k \underline{\Delta}_k). \end{aligned} \tag{65}$$

   This can be simplified somewhat, to

$$\begin{aligned} f(\underline{x}_{k+1}) = f(\underline{x}_k) &+ (1/2)(\alpha_k^2 \nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k) + \beta_k^2 \underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k) \\ &- \alpha_k \nabla f(\underline{x}_k)^{\mathrm{T}} \nabla f(\underline{x}_k) + \beta_k \underline{\Delta}_k^{\mathrm{T}} \nabla f(\underline{x}_k) \\ &- (\alpha_k \beta_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k). \end{aligned} \tag{66}$$

   We want to minimize with respect to α_k, β_k, which yields the following system of equations from setting the derivatives equal to 0:

$$\begin{aligned} \alpha_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k) - \nabla f(\underline{x}_k)^{\mathrm{T}} \nabla f(\underline{x}_k) - \beta_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k = 0 \\ \beta_k \underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k + \underline{\Delta}_k^{\mathrm{T}} \nabla f(\underline{x}_k) - \alpha_k \nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k = 0. \end{aligned} \tag{67}$$

   If $\nabla f(\underline{x}_k)$ and $\underline{\Delta}_k$ are linearly independent, then the system can be solved for α_k^*, β_k^*:

$$\begin{aligned} \alpha_k^* &= \frac{(\underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k)(\nabla f(\underline{x}_k)^{\mathrm{T}} \nabla f(\underline{x}_k)) - (\nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k)(\underline{\Delta}_k^{\mathrm{T}} \nabla f(\underline{x}_k))}{(\nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k))(\underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k) - (\nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k)^2}, \\ \beta_k^* &= -\frac{-(\nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k)(\nabla f(\underline{x}_k)^{\mathrm{T}} \nabla f(\underline{x}_k)) + (\nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k))(\underline{\Delta}_k^{\mathrm{T}} \nabla f(\underline{x}_k))}{(\nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k))(\underline{\Delta}_k^{\mathrm{T}} Q \underline{\Delta}_k) - (\nabla f(\underline{x}_k)^{\mathrm{T}} Q \underline{\Delta}_k)^2} \end{aligned} \tag{68}$$

   It can be shown that the Hessian corresponding to the second derivatives with respect to α_k, β_k at the above points is positive definite, and hence the point is a local minimum. Because the function is quadratic in terms of these variables, we know that it is a global minimum as well.

   Note, if $\nabla f(\underline{x}_k)$ and $\underline{\Delta}_k$ are linearly dependent, then there is no unique solution for α_k^*, β_k^*; however, in this case the problem of optimal step size effectively reduces steepest descent with optimal stepsize, as the net direction of motion will be parallel to $\nabla f(\underline{x}_k)$ for any α_k, β_k. Hence in this case, we can take

$$\begin{aligned} \alpha_k^* &= \frac{\nabla f(\underline{x}_k)^{\mathrm{T}} \nabla f(\underline{x}_k)}{\nabla f(\underline{x}_k)^{\mathrm{T}} Q \nabla f(\underline{x}_k)}, \\ \beta_k^* &= 0. \end{aligned} \tag{69}$$

For the remaining problems, consider the specific case of $\underline{b} = 0$ and

$$Q = \begin{bmatrix} 71 & 0 & 17 & -10 & -9 \\ 0 & 58 & 7 & 7 & 4 \\ 17 & 7 & 65 & -1 & 2 \\ -10 & 7 & -1 & 48 & -8 \\ -9 & 4 & 2 & -8 & 50 \end{bmatrix}. \tag{70}$$

In this case, $f(\underline{x})$ has a unique minimizer at $\underline{x}^* = 0$, but we are particularly interested in the way in which the iterates of these algorithms approach the minimizer.

2. ***Steepest Descent with Momentum:*** *In this case, we take*

$$\underline{\Delta}_k = \underline{x}_k - \underline{x}_{k-1}. \tag{71}$$

Beginning with $\underline{x}_0 = \underline{x}_1 = (1,1,\ldots,1)$, give the first ten iterates of $\{\underline{x}_k\}$.

Additionally, estimate the least upper bound of

$$\frac{||\underline{x}_{k+1}||}{||\underline{x}_k||} \tag{72}$$

as $k \to \infty$.

Answer: With the above iteration and formulae, this yields as the iterates $\underline{x}_0, \ldots, \underline{x}_{10}$ (rounding to give three non-trivial digits):

$$\begin{pmatrix} 1. & 1. & 1. & 1. & 1. \\ 1. & 1. & 1. & 1. & 1. \\ 0.0742 & -0.0197 & -0.221 & 0.517 & 0.477 \\ 0.165 & -0.138 & -0.0658 & 0.101 & 0.0933 \\ -0.00197 & -0.00365 & 0.00523 & 0.00389 & -0.00158 \\ -0.000513 & -0.000513 & -0.000203 & 0.0000840 & 0.000322 \\ 0.0000635 & -0.0000916 & 0.0000236 & -0.0000166 & 0.000129 \\ 0.0000149 & 1.84*10^{-6} & -9.67*10^{-6} & 0.0000333 & 0.0000266 \\ 4.08*10^{-6} & -3.30*10^{-6} & -3.15*10^{-6} & 3.08*10^{-6} & -2.40*10^{-6} \\ -2.16*10^{-7} & -6.97*10^{-8} & -1.67*10^{-8} & 7.61*10^{-8} & -1.21*10^{-7} \\ -2.08*10^{-8} & -1.67*10^{-8} & 5.65*10^{-8} & -5.03*10^{-9} & -5.66*10^{-8} \end{pmatrix} \tag{73}$$

With the indicated starting point, it can be difficult to estimate the limit of $||\underline{x}_{k+1}||/||\underline{x}_k||$ depending on the precision available to you, but I estimate it (going out to about 200 terms) to be approximately less than 0.364.

3. ***Steepest Descent with Lag Descent:*** *In this case, we take*

$$\underline{\Delta}_k = -\nabla f(\underline{x}_{k-1}). \tag{74}$$

Beginning with $\underline{x}_0 = \underline{x}_1 = (1,1,\ldots,1)$, give the first ten iterates of $\{\underline{x}_k\}$.

Additionally, estimate the least upper bound of

$$\frac{||\underline{x}_{k+1}||}{||\underline{x}_k||} \tag{75}$$

as $k \to \infty$.

Answer: With the above iteration and formulae, this yields as the iterates $\underline{x}_0, \ldots, \underline{x}_{10}$ (rounding to give three non-trivial digits):

$$\begin{pmatrix} 1. & 1. & 1. & 1. & 1. \\ 1. & 1. & 1. & 1. & 1. \\ 0.0742 & -0.0197 & -0.221 & 0.517 & 0.477 \\ 0.165 & -0.138 & -0.0658 & 0.101 & 0.0933 \\ 0.00779 & 0.00525 & 0.0141 & 0.0137 & 0.00827 \\ -0.000286 & -0.00192 & -0.000881 & 0.00635 & 0.00438 \\ 0.00171 & -0.00142 & -0.000711 & 0.000937 & 0.00107 \\ 0.000218 & 0.000245 & -0.0000519 & 0.000134 & 0.000225 \\ 7.55*10^{-6} & 3.61*10^{-6} & -0.0000886 & 0.0000609 & 0.0000672 \\ 0.0000251 & -0.0000225 & -9.26*10^{-6} & 0.0000183 & 0.0000145 \\ 1.84*10^{-6} & 4.81*10^{-7} & 3.23*10^{-6} & 8.29*10^{-6} & 6.57*10^{-6} \end{pmatrix} \tag{76}$$

Estimating the limit of $||\underline{x}_{k+1}||/||\underline{x}_k||$ based on the first 150 terms gives an estimate of less than 0.377.

4. *Compare the previous two algorithms to traditional steepest descent in this case,*

$$\underline{x}_{k+1} = \underline{x}_k - \alpha_k \nabla f(\underline{x}_k). \tag{77}$$

For the optimal stepsize α_k^* in the case of pure steepest descent, and beginning with $\underline{x}_1 = (1,1,\ldots,1)$, give the first ten iterates of $\{\underline{x}_k\}$.

Additionally, estimate the least upper bound of

$$\frac{||\underline{x}_{k+1}||}{||\underline{x}_k||} \tag{78}$$

as $k \to \infty$.

Answer: Using the standard steepest descent iteration, with optimal choice of stepsize, gives iterates $\underline{x}_0, \ldots, \underline{x}_{10}$:

$$\begin{pmatrix} 1. & 1. & 1. & 1. & 1. \\ 1. & 1. & 1. & 1. & 1. \\ 0.0742 & -0.0197 & -0.221 & 0.517 & 0.477 \\ 0.215 & -0.0701 & -0.00186 & 0.155 & 0.148 \\ 0.0183 & -0.0319 & -0.0550 & 0.0975 & 0.0853 \\ 0.0423 & -0.0109 & 0.00280 & 0.0342 & 0.0318 \\ 0.00375 & -0.00698 & -0.0113 & 0.0201 & 0.0175 \\ 0.00868 & -0.00224 & 0.000629 & 0.00708 & 0.00656 \\ 0.000769 & -0.00145 & -0.00232 & 0.00415 & 0.00361 \\ 0.00179 & -0.000463 & 0.000131 & 0.00146 & 0.00135 \\ 0.000159 & -0.000298 & -0.000479 & 0.000856 & 0.000745 \end{pmatrix} \tag{79}$$

Estimating the limit of $||\underline{x}_{k+1}||/||\underline{x}_k||$ gives something less than approximately 0.47.

5. *Which algorithm do your results suggest is best in this case? Can you find starting points for which that algorithm is not the best? When might any of these algorithms be preferred to the others?*

Answer: In the above case, it looks as though steepest descent with momentum is the best choice, though potentially not by much. With a random search, for instance, it is possible to discover starting points where either lag or momentum is better.

5 Initial Feasible Points and Barrier Methods

The **barrier method** optimizes a function over a constrained region of $\mathbb{R}^n$. It reduces a constrained optimization problem to a sequence of unconstrained optimization problems, whose solutions converge to the solution to the constrained problem. However, one drawback of the method is that it relies on being able to *find an initial point in the interior the constraint region*.

If the constraint region is simple, such as a box or a sphere, this can generally be easily done. However if the constraint region is sufficiently complex, defined to be the intersection of a number of complex volumes, identifying an initial interior point can be difficult.

For a given set S, we can consider defining a cost function $C_S : \mathbb{R}^n \mapsto \mathbb{R}$ which will be small over the interior of the set S, and larger for $\underline{x}$ outside of S. For suitably chosen cost functions, minimizing C_S as a **unconstrained** optimization problem will produce a point $\underline{x}^* \in S$. This can be thought of as a variant on the **penalty method**.

Let the constraint set $X \subset \mathbb{R}^n$ be expressed as the intersection of some number of other sets.

$$X = X_1 \cap X_2 \cap \ldots \cap X_m. \tag{80}$$

Note, X_i can be viewed as 'the feasible set for constraint i', for instance

$$X_i = \{\underline{x} : g_i(\underline{x}) \leqslant 0\}. \tag{81}$$

With this view, X as above represents the joint feasible set, where all constraints are satisfied simultaneously.

To find a point in the interior of X, we can consider attempting to minimize the function

$$C_X(\underline{x}) = \sum_{i=1}^{m} C_{X_i}(\underline{x}). \tag{82}$$

In this problem, we will consider X_i as a ball of radius r_i, centered on point $\underline{z}_i$. That is,

$$X_i = \{\underline{x} : ||\underline{x} - \underline{z}_i||^2 - r_i^2 \leqslant 0\}, \tag{83}$$

and define the cost function

$$C_{X_i}(\underline{x}) = e^{\frac{||\underline{x}-\underline{z}_i||^2}{r_i^2}-1}. \tag{84}$$

This cost function is minimized at the center of ball i, and increases exponentially quickly with distance away from the center.

Consider the set of four balls defined by $r_i = 1$ for $i = 1,2,3,4$ and

$$\begin{aligned}
\underline{z}_1 &= (0.226, 0.875, 1.165, 1.302, 0.942) \\
\underline{z}_2 &= (0.747, 0.544, 1.495, 0.602, 1.203) \\
\underline{z}_3 &= (0.853, 0.900, 0.177, 0.517, 1.001) \\
\underline{z}_4 &= (1.218, 0.735, 0.191, 0.690, 1.351).
\end{aligned} \tag{85}$$

1. *Using whatever algorithm you like, starting from* $\underline{x}_0 = (0,0,0,0,0)$*, attempt to minimize*

$$C_X(\underline{x}) = \sum_{i=1}^{m} C_{X_i}(\underline{x}). \tag{86}$$

What is the first iterate produced that lies in the interior of all four balls, i.e., $g_i(\underline{x}_k) < 0$ *for* $i = 1,2,3,4$*?*

Answer: The easiest method is probably a gradient descent. It is straightforward differentiation to show that

$$\nabla C_X(\underline{x}) = \sum_{i=1}^{m} 2(\underline{x} - \underline{z}_i) e^{||\underline{x}-\underline{z}_i||^2 - 1}. \tag{87}$$

Using steepest descent with a stepsize of 0.005, the first iterate I discover that satisfies all the constraints is

$$\underline{x}_{14} = (0.85202, 0.862721, 1.00795, 0.943331, 1.31842). \tag{88}$$

The objective function can be further reduced by pushing to higher iterates, but this is sufficient to yield an interior point.

2. *Consider using the cost function*

$$C_{X_i}(\underline{x}) = \left(\frac{||\underline{x} - \underline{z}_i||^2}{r_i^2} - 1 \right)^2. \tag{89}$$

Are you as successful with this cost function? Why or why not; what difficulties does this cost function introduce? What would an ideal cost function look like? Hint: Look at the graphs of $e^{\delta - 1}$ *vs* $(\delta - 1)^2$ *for* $\delta \geqslant 0$*.*

Answer: In this case, we have

$$\nabla C_X(\underline{x}) = 4 \sum_{i=1}^{m} (\underline{x} - \underline{z}_i) \left(||\underline{x} - \underline{z}_i||^2 - 1 \right). \tag{90}$$

Utilizing steepest descent with a stepsize of 0.005 as in the previous problem failed to produce an iterate in the interior in the first 1000 iterates - additionally, it appears to be converging onto the boundary of one of the spheres, from outside the sphere, i.e., it will never enter the sphere. Part of the problem here, looking at the graph of the respective cost functions, is that the cost function above penalizes points in the interior of the sphere as much as some points in the exterior - it is creating a pressure to converge on the boundary of the spheres rather than in the interior. This may or may not produce an iterate in the interior.

The first cost function is superior in that regard, as any point in the interior of the sphere is penalized strictly less than every point in the exterior of the sphere, and the cost decreases as you approach the center. This creates a pressure towards the interior, away from the boundary. Ideally, all points in the interior would be penalized equally, i.e., the cost would be constant over the interior of the sphere, but increasing for points moving away from the boundary of the sphere. This would create a pressure to drive the iterates into the interior, without biasing them in any particular direction once they are inside. However, it may generally be useful

to include a slight decrease of cost in the immediate neighborhood of the boundary of the sphere from the inside, to keep the iterates away from the boundary and potentially exiting the sphere.

3. *Consider using the cost function*

$$C_{X_i}(\underline{x}) = e^{\frac{||\underline{x}-\underline{z}_i||^4}{r_i^4}-1}. \tag{91}$$

How does this compare?

Answer: In this case, we have:

$$\nabla C_X(\underline{x}) = 4\sum_{i=1}^{m}(\underline{x}-\underline{z}_i)||\underline{x}-\underline{z}_i||^2 e^{||\underline{x}-\underline{z}_i||^4-1}. \tag{92}$$

In this case, we cannot use a constant stepsize. Note that even at the origin (a relatively mild starting point by anyone's standards), we have

$$\nabla C_X(0) = (-1.54095, -1.45318, -3.39147, -1.76299, -2.74522)\times 10^{11}. \tag{93}$$

For any stepsize greater than about 10^{-11}, this is going to result in massive overshoot, numerical instability, and general computational catastrophe. And for stepsizes in that range that might work, convergence will be slow to the point of futility.

However, implementing Armijo's Rule for determining stepsize with an initial stepsize guess of $s_0 = 1$, contraction factor $\beta = 0.5$ and threshold value $\sigma = 0.1$ produces an iterate in the interior of X in 5 steps,

$$\underline{x}_5 = (0.606785, 0.585498, 0.712607, 0.687093, 0.940017). \tag{94}$$

Note: In the above, instead of taking $\underline{d}_k = -\nabla C_X(\underline{x}_k)$, I took $\underline{d}_k$ as the negative normalized gradient, in an attempt to reduce numerical issues - hence the initial stepsize guess of $s_0 = 1$.