Optimization and Numerical Linear Algebra Past Quals Solutions

Trevor Loe

January 2025

This document contains solutions that I came up with for past ONLA qual problems. This is in no way a complete solution guide to all the past quals, but I have tried to cover as many past problems as possible.

Thank you for Zerrin Vural,

Past ONLA qualifying exams can be found at `https://ww3.math.ucla.edu/past-qualifying-exams/`

Fall 2021

1a

Find the critical points and local extremizers of

$$f(x_1, x_2, x_3) = x_1^2 + 3x_2^2 + x_3$$

subject to

$$x_1^2 + x_2^2 + x_3^2 = 4$$

To solve this, we use Lagrange's theorem, stating that if x^* is a local minizer of f subject to $h(x) = 0$ then (as long as $\nabla h_1, ..., \nabla h_m$ are linearly independent, there exists some λ such that

$$Df(x^*) + \lambda^T Dh(x^*) = 0$$

We have

$$Df(x) = \begin{pmatrix} 2x_1 \\ 6x_2 \\ 1 \end{pmatrix}$$

and as there is only one h function, we just have

$$Dh(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{pmatrix}$$

so we want a λ and x_1, x_2, x_3 solving

$$\begin{pmatrix} 2x_1 \\ 6x_2 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{pmatrix} = 0 \quad x_1^2 + x_2^2 + x_3^2 = 4$$

From the third equation we have $x_3 = -1/(2\lambda)$. From the first equation we have

$$x_1(1 + \lambda) = 0$$

So either $x_1 = 0$ or $\lambda = -1$. First consider when $x_1 = 0$. We then have

$$x_2^2 + x_3^2 = 4$$

and

$$x_2(3 + \lambda) = 0$$

So $x_2 = 0$ or $\lambda = -3$. First consider when $x_2 = 0$. Then $x_3 = \pm 2$ and $\lambda = \pm(-1/4)$. So two solutions are

$$(x_1, x_2, x_3, \lambda) = (0, 0, 2, -1/4), (0, 0, -2, 1/4)$$

Next we consider when $\lambda = -3$. Then $x_3 = 1/6$. Thus

$$x_2^2 + 1/36 = 4$$

Meaning $x_2 = \pm\sqrt{\frac{143}{36}}$. So two more solutions are

$$(0, \sqrt{\frac{143}{36}}, 1/6, -3), (0, -\sqrt{\frac{143}{36}}, 1/6, -3)$$

Now we consider the case where $x_1 \neq 0$, so $\lambda = -1$. For this case we have $x_3 = 1/2$. Also $6x_2 - 2x_2 = 0$ so $x_2(6-2) = 0$ so $x_2 = 0$. So we get

$$x_1^2 + 1/4 = 4$$

meaning $x_1 = \pm\sqrt{\frac{15}{16}}$. So our final two solutions are

$$(\sqrt{\frac{15}{16}}, 0, 1/2, -1), (-\sqrt{\frac{15}{16}}, 0, 1/2, -1)$$

We can summarize our results as follows To figure out if these points are minimizers for the function, we need

x_1	x_2	x_3	λ
0	0	2	$-\frac{1}{4}$
0	0	-2	$\frac{1}{4}$
0	$\sqrt{\frac{143}{36}}$	$\frac{1}{6}$	-3
0	$-\sqrt{\frac{143}{36}}$	$\frac{1}{6}$	-3
$\sqrt{\frac{15}{16}}$	0	$\frac{1}{2}$	-1
$-\sqrt{\frac{15}{16}}$	0	$\frac{1}{2}$	-1

Table 1: Stationary points to function

the second order necessary/sufficient conditions, which would require that

$$L(x, \lambda) = F(x) + \lambda H(x)$$

is positive semi-definite or positive definite (but only for points in the tangent space). Similar, for negative semi-definiteness, we would get a local max rather than a min.
We can compute L as follows

$$L(x, \lambda) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

If $\lambda = -1/4$ we get

$$L = \begin{pmatrix} 3/2 & 0 & 0 \\ 0 & 11/2 & 0 \\ 0 & 0 & -1/2 \end{pmatrix}$$

We also have that the tangent space at the first point is

$$T(x^*) = \{y \in \mathbb{R}^3 : \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} y = 0\} = \{\begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in \mathbb{R}\}$$

Note that, for any one of these vectors we get that $y^T Ly > 0$. Meaning that this point satisfies the second order sufficient condition and is a minimizer.
Next, note that the second point has

$$\begin{pmatrix} 5/2 & 0 & 0 \\ 0 & 13/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

which is pos. definite for any $y \in \mathbb{R}^3$, so in particular, for any $y \in T(x^*)$ we get $y^T Ly > 0$. So this point is also a minimizer.
Next, when $\lambda = -3$ we get

$$L = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \end{pmatrix}$$

and for this one we have

$$T(x^*) = \{y \in \mathbb{R}^3 : \begin{pmatrix} 0 \\ \sqrt{143}/3 \\ 1/3 \end{pmatrix} y = 0\}$$

This point will satisfy the necessary condition and the sufficient condition, because the only vector that wouldn't be negative from L is $(0, a, 0)^T$. So this point is local maximizer. Similarly, for the third point we get an identical tangent space except with $-\sqrt{143}$. So that will also be a local maximizer.
Now for the fifth and sixth point, we have $\lambda = -1$ so

$$L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

For the fifith point we get a tangent space like

$$y = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

where $a\sqrt{\frac{15}{16}} + c/2 = 0$. As b can be anything, we have $(0, 1, 0) \in T(x^*)$ which would give us a positive number from the quadratic form, but $(a, 0, c)$ would give us a negative. So the matrix is indefinite, meaning the necessary conditions are not satisfied.

1b

Find the solutions to

$$\max x^T \begin{pmatrix} 3 & 5 \\ 0 & 3 \end{pmatrix} x$$

subject to

$$||x|| = 1$$

Similarly, we will use the Lagrange conditions to find the stationary points. We have

$$Df(x) = A^T x + Ax = (A^T + A)x = \begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix} x = \begin{pmatrix} 6x_1 + 5x_2 \\ 5x_1 + 6x_2 \end{pmatrix}$$

Similarly,

$$Dh(x) = D(\sqrt{x_1^2 + x_2^2}) = \begin{pmatrix} \frac{x_1}{||x||} \\ \frac{x_2}{||x||} \end{pmatrix}$$

So the Lagrange conditions are

$$\begin{pmatrix} 6x_1 + 5x_2 \\ 5x_1 + 6x_2 \end{pmatrix} + \lambda \begin{pmatrix} \frac{x_1}{||x||} \\ \frac{x_2}{||x||} \end{pmatrix} = 0$$

and

$$||x|| = 1$$

Note that the second equation implies that the first two equations must satisfy

$$6x_1 + 5x_2 + \lambda x_1 = 0 \qquad 5x_1 + 6x_2 + \lambda x_2 = 0$$

and the last equation implies that

$$x_1^2 + x_2^2 = 1$$

We then can solve

$$x_1(6 + \lambda) = -5x_2$$

and then

$$x_2^2 \left(\left(\frac{5}{6 + \lambda} \right)^2 + 1 \right) = 1$$

Also subbing into the second equation we get

$$5 \frac{-5}{6 + \lambda} x_2 + (6 + \lambda)x_2 = 0$$

So either $x_2 = 0$ or

$$6 + \lambda - \frac{25}{6 + \lambda} = 0$$

But if $x_2 = 0$ then $x_1 = 0$ which is clearly not a solution. So we have

$$(6 + \lambda)^2 - 25 = 0$$

and thus

$$36 + 12\lambda + \lambda^2 - 25 = \lambda^2 + 12\lambda + 11 = 0$$

which means $\lambda = -11$ or $\lambda = -1$. If $\lambda = -11$, then

$$x_2^2((5/5)^2 + 1) = 1$$

so $x_2 = \sqrt{1/2}$. Consequently $x_1 = x_2 = \sqrt{1/2}$. So our first solution is

$$(x_1, x_2, \lambda) = (1/\sqrt{2}, 1/\sqrt{2}, -11)$$

Now consider when $\lambda = -1$. In this case,

$$x_2^2(2) = 1$$

so $x_2 = 1/\sqrt{2}$. Also $x_1(5) = -5x_2$. So $x_1 = -x_2 = -1/\sqrt{2}$. Thus, our other solution is

$$(-1/\sqrt{2}, 1/\sqrt{2}, -1)$$

We can now check the second order conditions by considering first that $||x|| = 1$ if and only if $||x||^2 = 1$. So $h(x) = x_1^2 + x_2^2$ and we get

$$F(x) = \begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix}$$

$$H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

So

$$L(x, \lambda) = \begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix} + \lambda \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

If $\lambda = -11$ we get

$$L(x, \lambda) = \begin{pmatrix} -16 & 5 \\ 5 & -16 \end{pmatrix}$$

This matrix has eigenvalues -21 and -11, so it is negative definite. Thus the first point is a maximizer. We would expect that, being a quadratic form, the function has a unique max. But just for fun, we can compute L for $\lambda = -1$ to be

$$L(x, \lambda) = \begin{pmatrix} 4 & 5 \\ 5 & 4 \end{pmatrix}$$

Which has eigenvalues 9 and -1. The tangent space for this point will be vectors of the form $(1, 1)$, which just so happens to be the eigenvector for the 9 eigenvalue. So this second point is a minimizer.

2

Let

$$B = \begin{pmatrix} I & A \\ A^* & I \end{pmatrix}$$

With A square the $||A||_2 \leq 1$. Prove that the condition number of B satisfies

$$\kappa(B) = \frac{1 + ||A||_2}{1 - ||A||_2}$$

For this problem we first consider the SVD of A given by

$$A = U\Sigma V^*$$

Then note we can write B as

$$B = \begin{pmatrix} I & U\Sigma V^* \\ V\Sigma U^* & I \end{pmatrix}$$

Now consider the vector $\begin{pmatrix} u_1 \\ v_1 \end{pmatrix}$, where $Av_1 = \sigma_{max}u_1$, the largest singular value for A. We then get

$$B\begin{pmatrix} u_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} I & U\Sigma V^* \\ V\Sigma U^* & I \end{pmatrix}\begin{pmatrix} u_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} u_1 + \sigma_{max}u_1 \\ v_1\sigma_{max} + v_1 \end{pmatrix} = (1 + \sigma_{max})\begin{pmatrix} u_1 \\ v_1 \end{pmatrix}$$

Note that for any vector of that form, the upper vector cannot grow in magnitude by more than $1 + \sigma_{max}$ as it will be the identity matrix plug a vector scaled by A.
Similarly, note that if we multiply by $\begin{pmatrix} u_1 \\ -v_1 \end{pmatrix}$ we get

$$B\begin{pmatrix} u_1 \\ -v_1 \end{pmatrix} = (1 - \sigma_{max})\begin{pmatrix} u_1 \\ -v_1 \end{pmatrix}$$

Observe that this is smallest scaling that can be applied to the vector for the same reason as above. Thus, we have found the maximum scaling that B can achieve, which is in fact an eigenvalue for B, and the smallest. Thus we get

$$\overline{\sigma_{max}} = 1 + \sigma_{max}$$

where $\overline{\sigma_{max}}$ denotes the largest singular value of B. Similarly,

$$\overline{\sigma_{min}} = 1 - \sigma_{max}$$

and σ_{max} denotes the largest singular value for A. Thus, the condition number for B, being the ratio of the largest to the smallest singular value is

$$\kappa(B) = \frac{1 + \sigma_{max}}{1 - \sigma_{max}}$$

3

The Jacobi iteration to solve $Ax = b$ is given by

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$$

where

$$M = D \quad N = -(L + U)$$

And the Jacobi Over-Relaxation is

$$M = \frac{1}{\omega}D \quad N = -\left(\left(1 - \frac{1}{\omega}\right)D + L + U\right)$$

Prove that if Jacobi converges, then the over-relaxation also converges, when $\omega \in (0, 1]$.
To prove this, suppose that Jacobi converges. Suppose that x^* is the solution we seek so

$$Ax^* = b$$

Note that

$$x^{(k+1)} - x^* = M^{-1}Nx^{(k)} + M^{-1}b - x^* = (I - M^{-1}A)x^{(k)} + M^{-1}Ax^* - x^*$$

and we can factor out to the form

$$x^{(k+1)} - x^* = (I - M^{-1}A)(x^{(k)} - x^*)$$

Let $G = I - M^{-1}A$. Then we have that the method will converg if and only if $||G|| < 1$ where the norm here denotes the spectral radius, the largest eigenvalue. We suppose that regular Jacobi converges, meaning that for G define in Jacobi, namely $||\tilde{G}|| < 1$. Also denote $\tilde{M} = D$ and $\tilde{N} = -(L + U)$. Then we get for the over-relation case

$$G = \left(\frac{1}{\omega}D\right)^{-1} \left(-\left(\left(1 - \frac{1}{\omega}\right)D + L + U\right)\right)$$

Note that, as D is diagonal, its inverse is just the matrix with the inverse of its elements. Namely,

$$\left(\frac{1}{\omega}D\right)^{-1} = \omega D^{-1}$$

So

$$G = -\omega D^{-1}\left(\left(1 - \frac{1}{\omega}\right)D + L + U\right) = (1 - \omega)I - \omega D^{-1}(L + U)$$

Now observe that $-D^{-1}(L + U) = \tilde{G}$. Now let u be an arbitrary unit vector. By the triangle inequality

$$||Gu|| \leq (1 - \omega)||u|| + \omega||\tilde{G}u|| < (1 + \omega) + \omega = 1$$

As the spectral radius is bounded by the operator norm we get that the spectral radius must then be less than 1. So the over-relaxation method will converge.

4

a

The Lanczos iteration will tridiagonalize a hermitian matrix A to the form

$$T_n = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & 0 & ... & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & ... & 0 \\ 0 & \beta_2 & \alpha_3 & \beta_3 & ... & 0 \\ . & . & . & & & \\ 0 & 0 & ... & ... & ... & \beta_{n-1} \\ 0 & ... & ... & 0 & \beta_{n-1} & \alpha_n \end{pmatrix}$$

Show that if a symmetric, real matrix A has a multiple eigenvalue, then the algorithm will terminate prematurely. For this problem, we will show that if A has a multiple eigenvalue, then the Krylov subspace $\mathcal{K}_n$ will be dimension strictly less than n. To see this, note that, if A has a multiple eigenvalue, its minimal polynomial, μ_A will be degree strictly less than n (where A is $n \times n$). This means $\deg(\mu_A) \leq n - 1$. Consider the Krylov subspace $\mathcal{K}_n(b)$. Note that, because μ_A has highest power $n - 1$ at most, we have

$$\mu_A(A)b \in \mathcal{K}_n(b)$$

But $\mu_A(A) = 0$, so we have found a vector which is a linear combination of $A^i b$ for $0 \leq i \leq n - 1$ which is 0. Thus the vectors $b, Ab, ..., A^{n-1}b$ are not linearly independent, so $\mathcal{K}_n(b)$ must have dimension less than n.
Because the Krylov subspace is dimension-deficient, we know that at some point, the Gram-Schmit process for finding an orthonormal basis for $\mathcal{K}_n$ will fail for an iteration before n. This would mean that $\beta_i = 0$ for some $i < n$ and the Lanczos algorithm will terminate, prematurely.

b

Premature termination of the algorithm does not necessarily mean we have a multiple eigenvalue. If our initial starting vector b is chosen badly enough, i.e. if it does not have a component in every eigenspace for A, then repeated iterations of A will not hit every eigenspace for A and $\mathcal{K}_n$ will be dimension deficient.
However, if we consider the case where the algorithm terminates for almost every choice of b (as being in a lower dimensional subspace is a measure-0 condition), then we do get the implication that A has a multiple eigenvalue. This would be because pre-mature termination for almost all b means that there is a polynomial of degree $< n$ that zeros out A. Thus, the minimal polynomial of A is degree less than n, so there must be repeated factors in the characteristic polynomial.

5

Let A be a positive definite symmetric $n \times n$ matrix and we seek a solution to

$$Ax = b$$

Let $\{z_1, ..., z_n\}$ be a set of A-orthogonal non-zero vectors. Given a starting point x_0, define the conjugate directions

$$w_k = \frac{\langle z_k, b - Ax_{k-1}\rangle}{\langle z_k, Az_k\rangle}, \qquad x_k = x_{k-1} + w_k z_k$$

Prove that $Ax_n = b$.
To prove this, first we show what happens when you take the product of x_k and Az_k.

$$\langle Az_k, x_k\rangle = \langle Az_k, x_{k-1}\rangle + \frac{\langle z_k, b - Ax_{k-1}\rangle}{\langle z_k, Az_k\rangle}\langle Az_k, z_k\rangle$$

canceling out terms and using linearity we get

$$\langle Az_k, x_{k-1}\rangle + \langle z_k, b\rangle - \langle Ax_{k-1}, z_k\rangle$$

As A is symmetric, we have

$$\langle Ax_{k-1}, z_k\rangle = \langle x_{k-1}, Az_k\rangle$$

Thus we are left with

$$\langle Az_k, x_k\rangle = \langle z_k, b\rangle$$

Now if we similarly consider

$$\langle Az_j, x_k\rangle = \langle Az_j, x_{k-1}\rangle + \frac{\langle z_k, b - Ax_{k-1}\rangle}{\langle z_k, Az_k\rangle}\langle Az_j, z_k\rangle$$

for $j \neq k$ we have the second term is 0 so we are left with just $\langle Az_j, x_{k-1}\rangle$. We then get, by repeatedly applying our two formulas to x_n that

$$\langle Az_k, x_n\rangle = \langle z_k, b\rangle$$

This happens because multiplication by Az_k will reduce the index of x from n until it his k, leaving us with $\langle Az_k, b\rangle$.
Now consider that because all the z_k are A orthogonal, they are linearly independent, meaning they form a basis. Similarly, Az_i form a basis, because A is positive-definite. Let

$$b = \beta_1 Az_1 + ... + \beta_n Az_n$$

and

$$x_n = \alpha_1 z_1 + ... + \alpha_n z_n$$

We then have

$$\langle Az_k, x_n\rangle = \langle Az_k, \alpha_1 z_1 + ... + \alpha_n z_n\rangle = \alpha_k\langle Az_k, z_k\rangle$$

and also

$$\langle z_k, b\rangle = \langle z_k, \beta_1 Az_1 + ... + \beta_n Az_n\rangle = \beta_k\langle Az_k, z_k\rangle$$

So by our equality above, we have

$$\alpha_k = \beta_k$$

for all k. Thus,

$$Ax_n = \alpha_1 Az_1 + ... + \alpha_n Az_n = \beta_1 Az_1 + ... + \beta_n Az_n = b$$

6

Let $A = QR$ be a reduced QR factorization for a tall matrix A, which is $N \times n$ (for $N > n$). Prove that if R has m nonzero values on its diagonal, then $\text{rank}(A) \geq m$.
To prove this, we will show that $Ae_1, ..., Ae_m$ are linearly independent, and thus $\text{im}(A)$ has dimension at least m. Without loss of generality, suppose that the matrix R is ordered so the m nonzero entries are first. Then we have

$$Ae_1 = QRe_1 = Qr_{11}e_1 = r_{11}q_1$$

where q_i is the ith column of Q. Similarly, we have

$$Ae_2 = r_{21}q_1 + r_{22}q_2 \qquad Ae_3 = r_{31}q_1 + r_{32}q_2 + r_{33}q_3$$

In general we have $Ae_k = \sum_{i=1}^{k} r_{i,k}q_i$. Now suppose that for some α_is we have

$$\alpha_1 Ae_1 + ... + \alpha_m Ae_m = 0$$

We then consider

$$0 = \langle q_m, \alpha_1 Ae_1 + ... + \alpha_m Ae_m \rangle = r_{mm}\alpha_m$$

because Ae_m is the only vector in the sum that has a component in the q_m direction. As we assumed $r_{ii} \neq 0$ for $i \leq m$ we get $\alpha_m = 0$. We then get, by the same argument

$$0 = \langle q_{m-1}, \alpha_1 Ae_1 + ... + \alpha_m Ae_m \rangle = r_{(m-1),(m-1)}\alpha_{m-1}$$

so $\alpha_{m-1} = 0$. Continuing through all m coefficients we get that α_i must be 0 for all i. Meaning that the vectors Ae_i for $i \leq m$ are linearly independent, by definition. So A will have rank at least m.

7

Consider the GMRES algorithm, which combines the Arnoldi algorithm with a least squares solver. Define $\mathcal{K}_n = \text{span}\{b, Ab, ..., A^{n-1}b\}$, the nth Krylov subspace for b. Suppose that at iteration n we have "arnoldi breakdown" so $h_{n+1,n} = 0$.

a

Show that $A\mathcal{K}_n \subseteq \mathcal{K}_n$.
This problem is asking us to consider the case where Arnoldi breakdown occurs, which happens when the Krylov subspace for $n+1$ is dimension less than $n+1$. This happens because

$$h_{n+1,n} = ||Aq_n - \sum_{j=1}^{n} h_{j,n}q_j|| = 0$$

where $h_{j,n} = \langle q_j, Aq_n \rangle$. This would imply that $Aq_n \in \text{span}\{q_1, ..., q_n\}$. In general, for $j < n$ the Arnoldi iteration gives us

$$Aq_j = \sum_{i=1}^{j+1} h_{i,j}q_i$$

Now consider Au where $u \in \mathcal{K}_n$. We can write

$$u = \alpha_1 q_1 + ... + \alpha_n q_n$$

so

$$Au = \alpha_1 Aq_1 + ... + \alpha_n Aq_n$$

Note that for $j < n$, we have $Aq_j \in \text{span}\{q_1, ..., q_n\} \subseteq \mathcal{K}_n$. Also for Aq_n we showed that $Aq_n \in \text{span}\{q_1, ..., q_n\} \subseteq \mathcal{K}_n$ as well. So $Au \in \mathcal{K}_n$. So $A\mathcal{K}_n \subseteq \mathcal{K}_n$.

b

Show that this guarentees $x \in \mathcal{K}_n$ with x solving $Ax = b$.
In the case described above, we have that the algoritm breaks down at iteration n and no earlier, so $h_{j+1,j} \neq 0$ when $j < n$. This means that for all $j < n$ we get

$$Aq_j = \sum_{i=1}^{j+1} h_{i,j} q_i$$

Because A is invertible, the dimension of $A\mathcal{K}$ cannot be reduced from $\dim(\mathcal{K}_n)$. Thus $A\mathcal{K}_n$ is a $\dim(\mathcal{K}_n)$ dimensional subspace in $\mathcal{K}_n$. Thus $A\mathcal{K} = \mathcal{K}_n$. We then get that

$$b \in \mathcal{K}_n \in A\mathcal{K}_n$$

So, by definition of $A\mathcal{K}_n$, $b = Ax$ for some $x \in \mathcal{K}_n$.

c

Assuming A is diagonalizable and we are given $n < m$. Describe a method for determining b such that this breakdown will occur no later than step n.
We want a vector b such that $A\mathcal{K}_n(b) = \mathcal{K}_n(b)$. Consider the orthonormal eigenbasis for A with eigenvectors $v_1, ..., v_m$. If we choose $b \in \text{span}\{v_1, ..., v_n\}$ we get that

$$A^n b = \lambda_1^n v_1 + \lambda_2^n v_2 + ... + \lambda_n^n v_n \in \text{span}\{v_1, ..., v_n\}$$

So $A^j b \in \text{span}\{v_1, ..., v_n\}$ for all j. So the Krylov subspace must be dimension at most n. If $A\mathcal{K}_n$ was not contained in $\mathcal{K}_n$, then we would have $n + 1$ linearly independent vectors in $\text{span}\{v_1, ..., v_n\}$, which is clearly a contradiction.

8

a

Consider the algorithm $x^{(k+1)} = x^{(k)} - \alpha_k M_k \nabla f(x^{(k)})$ where $f : \mathbb{R}^2 \to \mathbb{R}$ and $f \in C^1$ and

$$M_k = \begin{pmatrix} 1 & 1 \\ 0 & a \end{pmatrix}$$

and

$$\alpha_k = \text{argmin}_{\alpha \geq 0} f(x^{(k)} - \alpha M_k \nabla f(x^{(k)}))$$

At some iteration we have $\nabla f(x^{(k)}) = (1, -1)^T$. Find the largest range of values for a that guarantees $\alpha_k > 0$.
This is a modified version of the steepest descent algorithm, and we want to find the values for a such that $\alpha_k \neq 0$, meaning there will be a descent direction in $M_k \nabla f(x^{(k)})$. A descent direction will be a direction v in which $\nabla f(x)^T v > 0$. This is essentially asking to find the conditions in which

$$\nabla f(x^{(k)})^T M_k \nabla f(x^{(k)}) > 0$$

We can compute this value as

$$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & a \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ -a \end{pmatrix} = a$$

so for the above quantity to be positive, we must have $a > 0$.

b

Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ with $f(w) \geq c$ for all $w \in \mathbb{R}^d$. Assume there is some $L > 0$ such that

$$f(w') \leq f(w) + \nabla f(w)^T(w' - w) + \frac{L}{2}||w' - w||^2$$

for all $w, w' \in \mathbb{R}^d$. Show that there exists some $\alpha \in \mathbb{R}$ such that if we run gradient descent with fixed step size α, we get

$$\min_{0 \leq t \leq T-1} ||\nabla f(w^{(t)})||^2 \leq \frac{2L}{T}|f(w^{(0)}) - c|$$

If we consider $w' = w^{(t+1)} = w - \alpha \nabla f(w^{(t)})$, we have, from the descent lemma

$$\begin{aligned} f(w') &\leq f(w) - \nabla f(w)^T(\alpha \nabla f(w)) + \frac{L}{2}||\alpha \nabla f(w)||^2 \\ &= f(w) - ||\nabla f(w)||^2(\alpha - \frac{L\alpha^2}{2}) \end{aligned}$$

Rearranging we get

$$f(w') - f(w) \leq -||\nabla f(w)||^2 \alpha(1 - \frac{L\alpha}{2})$$

and

$$f(w) - f(w') \geq ||\nabla f(w)||^2 \alpha(1 - \frac{L\alpha}{2})$$

If we pick $\alpha = 2/L$ we have $\alpha(1 - \frac{L\alpha}{2}) = 1/(2L)$, so we get that

$$f(w) - f(w') \geq ||\nabla f(w)||^2 \frac{1}{2L}$$

Subbing in what w and w' are we have

$$f(x^{(k+1)}) - f(x^{(k)}) \geq ||\nabla f(x^{(k)})||^2 \frac{1}{2L}$$

We sum up the iterates of this inequality for the iterations 1 through T. Giving us

$$f(x^{(T)}) - f(x^{(0)}) \geq \sum_{j=0}^{T-1} ||\nabla f(x^{(j)})||^2 \frac{1}{2L} \geq \min_{0 \leq j \leq T-1} ||\nabla f(x^{(j)})||^2 \frac{T}{2L}$$

Then we use the fact that $f(w) \geq c$ for all w to get

$$c - f(x^{(0)}) \geq \min_{0 \leq j \leq T-1} ||\nabla f(x^{(j)})||^2 \frac{T}{2L}$$

and thus

$$|c - f(x^{(0)})| \frac{2L}{T} \geq \min_{0 \leq j \leq T-1} ||\nabla f(x^{(j)})||^2$$