

CSE 6643 Homework 2

Karl Hiner

1 Less positive [12.5 pts]

Consider a symmetric positive definite matrix \mathbf{A} separated into subblocks according to

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} \end{pmatrix}.$$

Show that the matrix $\mathbf{A}_{2,2} - \mathbf{A}_{2,1}\mathbf{A}_{1,1}^{-1}\mathbf{A}_{1,2}$ is symmetric and positive definite.

First, note that $\mathbf{A}_{2,2} - \mathbf{A}_{2,1}\mathbf{A}_{1,1}^{-1}\mathbf{A}_{1,2}$ is the *Schur complement* of \mathbf{A} , which we will refer to as \mathbf{S} for brevity. For \mathbf{S} to be symmetric, we must show that $\mathbf{S}^T = \mathbf{S}$:

$$\begin{aligned} \mathbf{S}^T &= (\mathbf{A}_{2,2} - \mathbf{A}_{2,1}\mathbf{A}_{1,1}^{-1}\mathbf{A}_{1,2})^T \\ &= \mathbf{A}_{2,2}^T - \mathbf{A}_{1,2}^T(\mathbf{A}_{1,1}^{-1})^T\mathbf{A}_{2,1}^T \\ &= \mathbf{A}_{2,2} - \mathbf{A}_{2,1}\mathbf{A}_{1,1}^{-1}\mathbf{A}_{1,2} = \mathbf{S} \end{aligned}$$

Since \mathbf{A} is symmetric, $\mathbf{A}_{1,2} = \mathbf{A}_{2,1}^T$, $\mathbf{A}_{2,1} = \mathbf{A}_{1,2}^T$, and $\mathbf{A}_{2,2} = \mathbf{A}_{2,2}^T$. Since \mathbf{A} is positive definite, it is invertible, and since it is symmetric, then its inverse is also symmetric, and $(\mathbf{A}_{1,1}^{-1})^T = \mathbf{A}_{1,1}^{-1}$. Thus, the above equality is satisfied and \mathbf{S} is symmetric.

To show that \mathbf{S} is positive definite, we need to show that for any non-zero vector \mathbf{x} , $\mathbf{x}^T\mathbf{S}\mathbf{x} > 0$.

First, keeping in mind that $\mathbf{A}_{2,1} = (\mathbf{A}_{1,2})^T$, let's define \mathbf{A} more compactly as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{D} \end{pmatrix}.$$

Then the Schur complement is $\mathbf{S} = \mathbf{D} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}$, and we can express \mathbf{A} as

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}^T\mathbf{B}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{B}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \mathbf{N} \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \mathbf{N}^T.$$

\mathbf{N} is invertible, and its inverse can be found like an elementary real matrix, by negating its single off-diagonal element:

$$N = \begin{pmatrix} I & 0 \\ C^T B^{-1} & I \end{pmatrix}$$

$$N^{-1} = \begin{pmatrix} I & 0 \\ -C^T B^{-1} & I \end{pmatrix}$$

Let $M \equiv \begin{pmatrix} B & 0 \\ 0 & S \end{pmatrix}$.

Then,

$$M = N^{-1} A (N^{-1})^T$$

$$\mathbf{u}^T M \mathbf{u} = \mathbf{u}^T N^{-1} A (N^{-1})^T \mathbf{u} \quad (\text{for any vector } \mathbf{u} \neq \mathbf{0})$$

$$= \mathbf{v}^T A \mathbf{v} \quad (\mathbf{v} \equiv (N^{-1})^T \mathbf{u})$$

$$> 0 \quad (\text{since } A \text{ is positive definite})$$

Thus, M is positive-definite.

Since A is positive definite, then B must be positive definite, since it is a principle submatrix of A (and of M , which we just showed to also be positive definite!).

Thus, $S = D - C^T B^{-1} C = A_{2,2} - A_{2,1} A_{1,1}^{-1} A_{1,2}$ is also positive definite, since it is the other principle submatrix of the positive definite M .

2 Just one [12.5 pts]

Consider a matrix $A \in \mathbb{R}^{m \times m}$, with (unpivoted) LU-factorization given by $A = LU$. Propose an algorithm for computing the (i, j) -th entry of A^{-1} using only $\mathcal{O}((m+1-j)^2 + (m+1-i)^2)$ floating point operations.

1. Compute the j -th column of A^{-1} by solving $Ly = e_j$ for y , where e_j is the j -th standard basis vector. This step takes $\mathcal{O}((m+1-j)^2)$ operations.
2. Compute the (i, j) -th entry of A^{-1} by solving $Ux = y$ for x . Then the i -th entry of x is the (i, j) -th entry of A^{-1} . This step takes $\mathcal{O}((m+1-i)^2)$ operations.

The total number of operations is $\mathcal{O}((m+1-j)^2 + (m+1-i)^2)$.

3 SVD [25 pts]

In class, we have reminded ourselves of the SVD of square matrices. Use the recommended literature from the class syllabus to catch up on the SVD applied to non-square matrices.

We denote as σ_{\min} and σ_{\max} the smallest and largest singular value of a matrix and by λ_i its i -th eigenvalue. We consider general matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times l}$. Show that the following results hold

(a) [5 pts]

$$\sigma_{\max}(A) \|\mathbf{x}\|_2 \geq \|A\mathbf{x}\|_2, \forall \mathbf{x} \in \mathbb{R}^n$$

$$\begin{aligned}\|\mathbf{Ax}\|_2 &\leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2 && \text{(consistency of } \|\cdot\|_2 \text{ operator norm with induced vector norm)} \\ &= \sigma_{\max}(\mathbf{A}) \|\mathbf{x}\|_2 && \text{(Thm. 5.3 in Trefethen \& Bau: } \|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A}))\end{aligned}$$

(b) [5 pts]

$$m \geq n \Rightarrow \|\mathbf{Ax}\|_2 \geq \sigma_{\min}(\mathbf{A}) \|\mathbf{x}\|_2, \forall \mathbf{x} \in \mathbb{R}^n$$

$$\begin{aligned}\frac{\mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2} &\geq \inf_{\mathbf{y} \in \mathbb{R}^m \setminus \{0\}} \frac{\mathbf{y}^* \mathbf{A}^* \mathbf{A} \mathbf{y}}{\|\mathbf{y}\|_2^2} && (\forall \mathbf{x} \in \mathbb{R}^n) \\ \mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x} &\geq \left(\inf_{\mathbf{y} \in \mathbb{R}^m \setminus \{0\}} \frac{\mathbf{y}^* \mathbf{A}^* \mathbf{A} \mathbf{y}}{\|\mathbf{y}\|_2^2} \right) \|\mathbf{x}\|_2^2 && \text{(multiply both sides by } \|\mathbf{x}\|_2^2) \\ \mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x} &\geq \lambda_{\min}(\mathbf{A}^* \mathbf{A}) \|\mathbf{x}\|_2^2 && \left(\lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{y} \in \mathbb{R}^m \setminus \{0\}} \frac{\mathbf{y}^* \mathbf{A} \mathbf{y}}{\|\mathbf{y}\|_2^2} \right) \\ \|\mathbf{Ax}\|_2^2 &\geq \lambda_{\min}(\mathbf{A}^* \mathbf{A}) \|\mathbf{x}\|_2^2 && (\|\mathbf{v}\|_2 = \sqrt{\mathbf{v}^* \mathbf{v}}) \\ \|\mathbf{Ax}\|_2 &\geq \sqrt{\lambda_{\min}(\mathbf{A}^* \mathbf{A})} \|\mathbf{x}\|_2 && \text{(take square root of both sides)} \\ &= \sigma_{\min}(\mathbf{A}) \|\mathbf{x}\|_2 && \text{(def. of singular values)}\end{aligned}$$

(c) [5 pts]

$$m = n \Rightarrow \sigma_{\min}(\mathbf{A}) \leq |\lambda_i(\mathbf{A})| \leq \sigma_{\max}(\mathbf{A}), \forall i$$

Let λ_i be an eigenvalue of \mathbf{A} , and let \mathbf{v}_i be a corresponding unit eigenvector such that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Taking the 2-norm of both sides:

$$\begin{aligned}\|\mathbf{A}\mathbf{v}_i\|_2 &= \|\lambda_i \mathbf{v}_i\|_2 \\ &= |\lambda_i| \|\mathbf{v}_i\|_2 \\ &= |\lambda_i| && \text{(since } \|\mathbf{v}_i\|_2 = 1)\end{aligned}$$

Also, note that since \mathbf{A} is square when $m = n$, $\mathbf{A}^* \mathbf{A}$ is positive semidefinite, and can thus be decomposed into $\mathbf{U} \mathbf{\Lambda} \mathbf{U}^*$, where $\mathbf{\Lambda}$ is an $m \times m$ diagonal matrix whose entries are the eigenvalues of $\mathbf{A}^* \mathbf{A}$.

So then,

$$\begin{aligned}|\lambda_{\min}| &= \|\mathbf{A}\mathbf{v}_{\min}\|_2 && \text{(let } \mathbf{v}_{\min} \text{ be a unit e.v. of } \mathbf{A} \rightarrow \lambda_{\min}) \\ |\lambda_{\min}|^2 &= \|\mathbf{A}\mathbf{v}_{\min}\|_2^2 && \text{(squaring both sides)} \\ &= \mathbf{v}_{\min}^* \mathbf{A}^* \mathbf{A} \mathbf{v}_{\min} && \text{(def. of vector 2-norm)} \\ &\geq \inf_{\|\mathbf{x}\|=1} \mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x} && \text{(since } \mathbf{v}_{\min} \text{ is min. eigenvalue)} \\ &= \inf_{\|\mathbf{x}\|=1} \mathbf{x}^* \mathbf{U} \mathbf{\Lambda} \mathbf{U}^* \mathbf{x} && (\mathbf{\Lambda}_{i,i} = \lambda_i(\mathbf{A}^* \mathbf{A})) \\ &= \inf_{\|\mathbf{y}\|=1} \mathbf{y}^* \mathbf{\Lambda} \mathbf{y} && \text{(let } \mathbf{y} \equiv \mathbf{U}^* \mathbf{x}) \\ &= \lambda_{\min}(\mathbf{A}^* \mathbf{A}) && \text{(def. of } \lambda_{\min} \text{ - see (4) below)} \\ &= \sigma_{\min}(\mathbf{A})^2 && \text{(def. of singular value)}\end{aligned}$$

Thus, $|\lambda_{\min}(\mathbf{A})|^2 \geq \sigma_{\min}(\mathbf{A})^2 \implies |\lambda_{\min}(\mathbf{A})| \geq \sigma_{\min}(\mathbf{A})$.

The same argument can be applied to show that $|\lambda_{\max}(\mathbf{A})| \leq \sigma_{\max}(\mathbf{A})$, which shows that

$$m = n \implies \sigma_{\min}(\mathbf{A}) \leq |\lambda_i(\mathbf{A})| \leq \sigma_{\max}(\mathbf{A}), \forall i$$

(d) [5 pts]

$$\sigma_{\max}(\mathbf{A}) = 1/\sigma_{\min}(\mathbf{A}^{-1}), \text{ if } \mathbf{A}^{-1} \text{ exists}$$

We can decompose any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ into components $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where \mathbf{U} and \mathbf{V}^T are orthogonal, and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal matrix whose entries $\Sigma_{i,i}$ contain all of the singular values of \mathbf{A} . Since \mathbf{A} is invertible, all of its singular values will be nonzero, and so $\Sigma_{i,i} \neq 0, \forall i$.

If we invert both sides of this equality, we get $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$, which holds since we are given that \mathbf{A}^{-1} exists. $\mathbf{\Sigma}^{-1}$ also exists, since it is a diagonal matrix with no zeros on its diagonal. Its (also nonzero) diagonal entries are the singular values of \mathbf{A}^{-1} , and its inverse can be found by inverting each of its diagonal elements (a property not proven here):

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1(\mathbf{A}) & & \\ & \ddots & \\ & & \sigma_n(\mathbf{A}) \end{bmatrix}$$

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1(\mathbf{A})} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n(\mathbf{A})} \end{bmatrix}.$$

Thus, for any $1 \leq i \leq n$, $\sigma_i(\mathbf{A}) = \Sigma_{i,i} = \frac{1}{(\mathbf{\Sigma}^{-1})_{i,i}} = \frac{1}{\sigma_i(\mathbf{A}^{-1})}$.

Finally, let $j = \operatorname{argmax}_i \Sigma_{i,i}$. Then $\sigma_j(\mathbf{A}) = \sigma_{\max}(\mathbf{A})$. Since each element in $\mathbf{\Sigma}^{-1}$ is inverted, $j = \operatorname{argmin}_i (\mathbf{\Sigma}^{-1})_{i,i}$, and so $\sigma_j(\mathbf{A}^{-1}) = \sigma_{\min}(\mathbf{A}^{-1})$, and so $\sigma_{\max}(\mathbf{A}) = \frac{1}{\sigma_{\min}(\mathbf{A}^{-1})}$.

(e) [5 pts]

Find the most general conditions on m , n , and l , under which $\sigma_{\min}(\mathbf{AB}) \geq \sigma_{\min}(\mathbf{A})\sigma_{\min}(\mathbf{B})$.

If both \mathbf{A} and \mathbf{B} are positive semi-definite, then the inequality holds.

The product of two positive semi-definite matrices is always positive semi-definite, and so its eigenvalues are non-negative. Thus, the smallest eigenvalue of the product is ≥ 0 , and the product of the smallest eigenvalues of the matrices is also ≥ 0 .

Thus, the smallest eigenvalue of the product is \geq the product of the smallest eigenvalues of the two positive semi-definite matrices.

4 Inverses [25 pts]

We consider $\mathbf{A} \in \mathbb{R}^{m \times m}$ and assume that \mathbf{A}^{-1} exists. You may use without proof that if \mathbf{A} is symmetric, its largest and smallest eigenvalues are characterized as

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$$

and

$$\lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$$

(a) [5 pts]

Show that

$$\|\mathbf{A}^{-1}\|_2^{-1} = \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{A} \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \quad (1)$$

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T && \text{(SVD decomposition of } \mathbf{A} \text{)} \\ \mathbf{A}^{-1} &= \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^T && \text{(invert both sides)} \\ \|\mathbf{A}^{-1}\|_2 &= \|\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^T\|_2 && \text{(take the 2-norm of both sides)} \\ &= \sigma_{\max}(\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^T) && \text{(since } \|\mathbf{B}\|_2 = \sigma_{\max}(\mathbf{B}) \text{ for any matrix } \mathbf{B} \text{)} \\ &= \sigma_{\max}(\mathbf{\Sigma}^{-1}) && \text{(since } \mathbf{V} \text{ and } \mathbf{U} \text{ are orthogonal)} \\ &= \frac{1}{\sigma_{\min}(\mathbf{\Sigma})} && \text{((3d) above)} \\ &= \frac{1}{\sigma_{\min}(\mathbf{A})} && \text{(def. of } \mathbf{\Sigma} \text{)} \\ \|\mathbf{A}^{-1}\|_2^{-1} &= \sigma_{\min}(\mathbf{A}) && \text{(invert both sides)} \\ &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{A} \mathbf{x}\|_2}{\|\mathbf{x}\|_2} && \text{(def. of } \sigma_{\min} \text{)} \end{aligned}$$

(b) [5 pts]

Show that

$$\|\mathbf{A}^{-1}\|_2^{-1} \geq \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{|\mathbf{x}^T \mathbf{A} \mathbf{x}|}{\|\mathbf{x}\|_2^2} \quad (2)$$

$$\begin{aligned} \|\mathbf{A}^{-1}\|_2^{-1} &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{A} \mathbf{x}\|_2}{\|\mathbf{x}\|_2} && \text{(from above)} \\ &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{A} \mathbf{x}\|_2 \|\mathbf{x}\|_2}{\|\mathbf{x}\|_2^2} && \text{(multiply num. and dem. by } \|\mathbf{x}\|_2 \text{)} \\ &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{x}^T \mathbf{A} \mathbf{x}\|_2}{\|\mathbf{x}\|_2^2} && (\|\mathbf{x}^T \mathbf{A} \mathbf{x}\|_2 = \|\mathbf{A} \mathbf{x}\|_2 \|\mathbf{x}\|_2) \\ &\geq \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{|\mathbf{x}^T \mathbf{A} \mathbf{x}|}{\|\mathbf{x}\|_2^2} && (|\mathbf{x}^T \mathbf{A} \mathbf{x}| \leq \|\mathbf{x}^T \mathbf{A} \mathbf{x}\|_2) \end{aligned}$$

(c) [5 pts]

Show that for symmetric matrices $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{m \times m}$,

$$\lambda_{\min}(\mathbf{M} + \mathbf{N}) \geq \lambda_{\min}(\mathbf{M}) + \lambda_{\min}(\mathbf{N}) \quad (3)$$

To start, we will substitute the given expression for λ_{\min} , $\lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$, for the left-side term, $\lambda_{\min}(\mathbf{M} + \mathbf{N})$:

$$\begin{aligned} \lambda_{\min}(\mathbf{M} + \mathbf{N}) &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T (\mathbf{M} + \mathbf{N}) \mathbf{x}}{\|\mathbf{x}\|_2^2} \\ &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{M} \mathbf{x} + \mathbf{x}^T \mathbf{N} \mathbf{x}}{\|\mathbf{x}\|_2^2} \\ &= \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{x}^T \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|_2^2} + \frac{\mathbf{x}^T \mathbf{N} \mathbf{x}}{\|\mathbf{x}\|_2^2} \right) \end{aligned}$$

Next, restate the given left- and right-side terms, using an unknown inequality, "?", as a placeholder:

$$\begin{aligned} \lambda_{\min}(\mathbf{M} + \mathbf{N}) & \quad ? \quad \lambda_{\min}(\mathbf{M}) + \lambda_{\min}(\mathbf{N}) \\ \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{x}^T \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|_2^2} + \frac{\mathbf{x}^T \mathbf{N} \mathbf{x}}{\|\mathbf{x}\|_2^2} \right) & \quad ? \quad \inf_{\mathbf{y} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{y}^T \mathbf{M} \mathbf{y}}{\|\mathbf{y}\|_2^2} \right) + \inf_{\mathbf{z} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{z}^T \mathbf{N} \mathbf{z}}{\|\mathbf{z}\|_2^2} \right) \end{aligned}$$

To determine what the unknown inequality, "?", must be, we observe that the left side is the same expression as the right side, except that it is optimizing to find a *single* infimum value for both terms, while the right side is jointly optimizing for *two separate* infimums - one for each term. Another way of thinking about this is that the right side is *less constrained* than the left in its "search" for a minimum (real) value - its "search space" is a superset of that of the left side, since it's allowed to test different vectors \mathbf{y} and \mathbf{z} on the left and right terms, respectively, rather than being constrained to test the same $\mathbf{y} = \mathbf{z}$ as the left side is.

Also, note that this unknown inequality "?", must allow for the case of equivalence, for example when $\mathbf{M} = \mathbf{N}$.

Although this argument is hand wavy, I hope that it is sufficient to convince the reader (grader), that the inequality "?" must be " \geq ".

(d) [5 pts]

In the same setting as (c), show that

$$\lambda_{\max}(\mathbf{M} + \mathbf{N}) \leq \lambda_{\max}(\mathbf{M}) + \lambda_{\max}(\mathbf{N}) \quad (4)$$

Using the same approach as (c), we substitute the given expression for λ_{\max} ,

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2},$$

and determine the unknown inequality, "?":

$$\lambda_{\max}(\mathbf{M} + \mathbf{N}) \quad ? \quad \lambda_{\max}(\mathbf{M}) + \lambda_{\max}(\mathbf{N})$$

$$\sup_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{x}^T \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|_2^2} + \frac{\mathbf{x}^T \mathbf{N} \mathbf{x}}{\|\mathbf{x}\|_2^2} \right) \quad ? \quad \sup_{\mathbf{y} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{y}^T \mathbf{M} \mathbf{y}}{\|\mathbf{y}\|_2^2} \right) + \sup_{\mathbf{z} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \left(\frac{\mathbf{z}^T \mathbf{N} \mathbf{z}}{\|\mathbf{z}\|_2^2} \right)$$

Using the exact same argument as above, but noting that the functions being optimized here are supremums instead of infimums, this inequality "?" must be " \leq ".

(e) [5 pts]

For \mathbf{A} symmetric positive definite, show that

$$\|\mathbf{A}^{-1}\|_2^{-1} = \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$$

In (4a) above, we showed that

$$\|\mathbf{A}^{-1}\|_2^{-1} = \sigma_{\min}(\mathbf{A}).$$

Since

$$\lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{x} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2},$$

and since, for symmetric positive definite matrices, the minimum singular value is equal to the minimum eigenvalue, we are done.

(The singular values are equal to the square roots of the eigenvalues, and the eigenvalues for a symmetric positive definite matrix are all positive.)

5 Oblivion [25 pts]

(c) [5 pts]

Use section `HW2_driver.jl` to try the performance of the blocked algorithm for different matrix sizes and block sizes. Describe your results, and for the most interesting set of parameters, hand in the resulting plot with your homework. Describe and explain the results.

I found that, for the default settings, I found that the block size mattered a great deal, with times increasing as block sizes increased after about 512-1024.

However, I found that the absolute performance for large matrix sizes was relatively poor, and improved a great deal by making all matrix sizes the same size, and making them powers of two. This is likely because the matrices are then all evenly divided into the power-of-two block sizes,

as well as perhaps allowing for more contiguous memory access if the power-of-two blocks fits well inside the cache.

Shown in the plot below in Fig. 1 is an experiment with all matrix sizes equal to 2048. Here, we can see that performance first improves with increases block size, until block size exceeds 512, and then maxes out when the block size equals the matrix size. This likely corresponds to the blocks fitting inside the cache evenly at these smaller power-of-two block sizes.

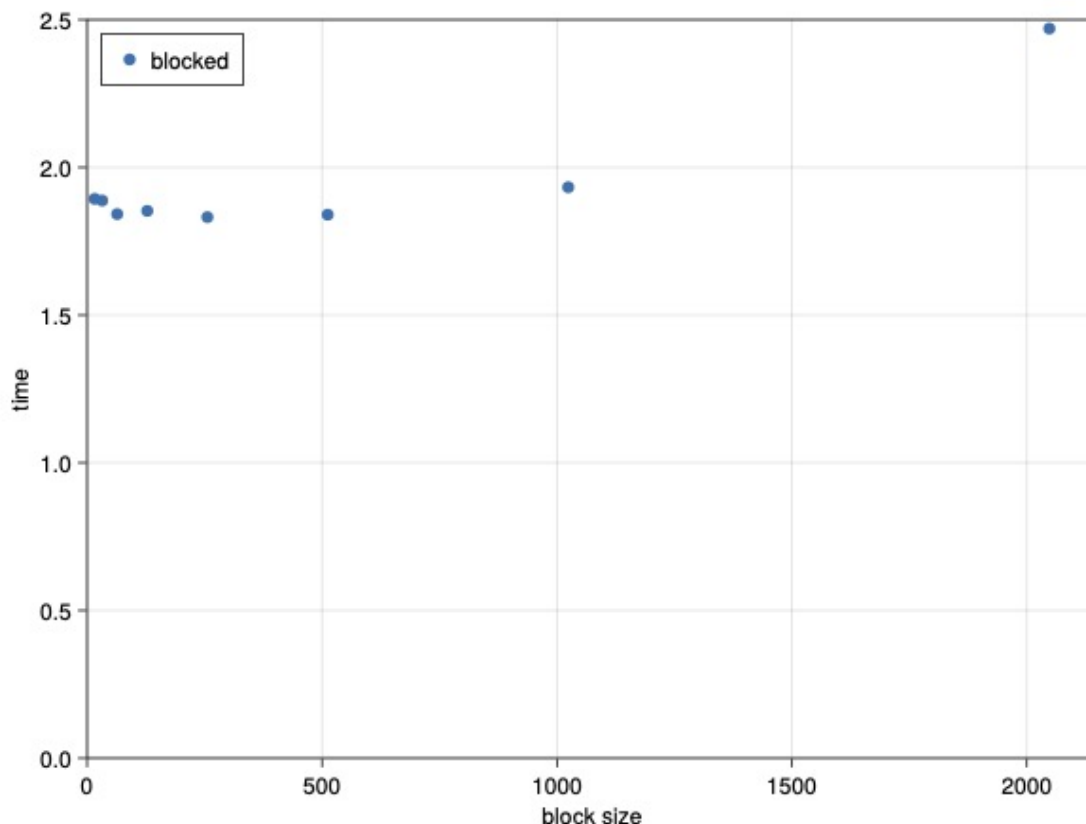


Figure 1 All matrices length 2048

(e) [5 pts]

Use part (e) of `HW2_driver.jl` to benchmark your new algorithm for different values of `bks`. Are you able to obtain an improvement over the blocked algorithm? Hand in the figure produced by part (e) of `HW2_driver.jl` for what you consider the most interesting set of parameters. Algorithms of this type are called “cache-oblivious.” Explain why this is the case and what could be the theoretical advantages of this particular cache-oblivious algorithm.

As shown in Fig. 2 below, I was able to obtain a performance improvement for all but the smallest block sizes.

Cache-oblivious methods take advantage of the hierarchical nature of modern memory systems, which have multiple cache levels. By working with progressively smaller problem sizes, it allows for taking advantage of the different cache sizes within this memory hierarchy. The operating

system can potentially better optimize cache utilization and memory access patterns with multiple problem sizes, rather than working with only a single fixed size.

For very small block sizes, the operating system will have fewer opportunities to take advantage of varying problem sizes, which explains my results.

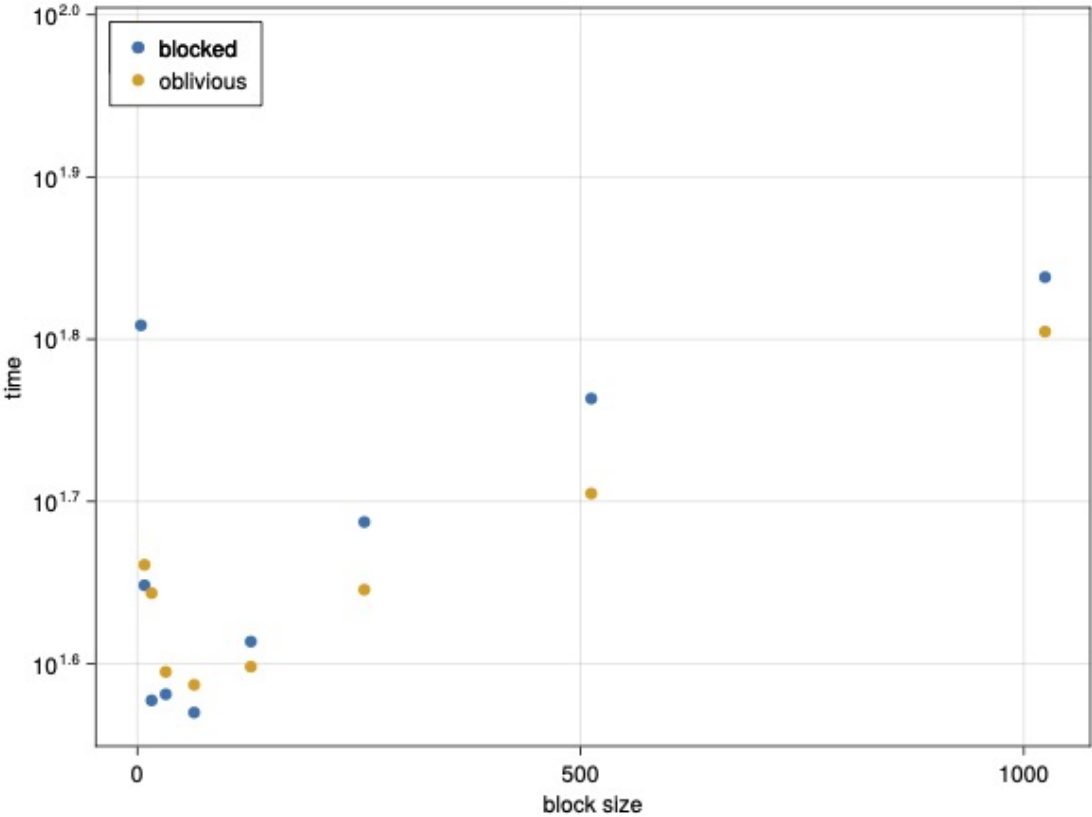


Figure 2 All matrices length 2048