

# CSE 6643 Homework 2

Karl Hiner

- There are 2 sections in grade scope: Homework 2 and Homework 2 Programming. Submit your answers as a PDF file to Homework 2 (report the results that you obtain using programming by using plots, tables, and a description of your implementation like you would when writing a paper.) and also submit your code in a zip file to Homework 2 Programming.
- Programming questions are posted in Julia. You are allowed to use basic library functions like sorting, plotting, matrix-vector products etc, but nothing that renders the problem itself trivial. Please use your common sense and ask the instructors if you are unsure. You should never add additional packages to the environment.
- Late homework incurs a penalty of 20% for every 24 hours that it is late. Thus, right after the deadline, it will only be worth 80% credit, and after four days, it will not be worth any credit.
- We recommend the use of LaTeX for typing up your solutions. No credit will be given to unreadable handwriting.
- List explicitly with whom in the class you discussed which problem, if any. Cite all external resources that you were using to complete the homework. For details, consult the collaboration policy in the class syllabus on canvas.

## 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:

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

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

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

Then,

$$\begin{aligned} \mathbf{M} &= \mathbf{N}^{-1} \mathbf{A} (\mathbf{N}^{-1})^T \\ \mathbf{u}^T \mathbf{M} \mathbf{u} &= \mathbf{u}^T \mathbf{N}^{-1} \mathbf{A} (\mathbf{N}^{-1})^T \mathbf{u} && \text{(for any vector } \mathbf{u} \neq \mathbf{0} \text{)} \\ &= \mathbf{v}^T \mathbf{A} \mathbf{v} && (\mathbf{v} \equiv (\mathbf{N}^{-1})^T \mathbf{u}) \\ &> 0 && \text{(since } \mathbf{A} \text{ is positive definite)} \end{aligned}$$

Thus,  $\mathbf{M}$  is positive-definite.

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

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

## 2 Just one [12.5 pts]

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

The algorithm for computing the  $(i, j)$ -th entry of  $\mathbf{A}^{-1}$  is as follows:

1. Compute the  $j$ -th column of  $\mathbf{A}^{-1}$  by solving  $\mathbf{L} \mathbf{y} = \mathbf{e}_j$  for  $\mathbf{y}$ , where  $\mathbf{e}_j$  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  $\mathbf{A}^{-1}$  by solving  $\mathbf{U}\mathbf{x} = \mathbf{y}$  for  $\mathbf{x}$ , where the  $i$ -th entry of  $\mathbf{x}$  is the desired entry. 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  $\sigma_{\min}$  and  $\sigma_{\max}$  the smallest and largest singular value of a matrix and by  $\lambda_i$  its  $i$ -th eigenvalue. We consider general matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{n \times l}$ . Show that the following results hold

**(a) [5 pts]**

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

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

TODO must prove Thm. 5.3

**(b) [5 pts]**

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

$$\begin{aligned} \|\mathbf{Ax}\|_2 &\geq \sigma_{\min}(\mathbf{A})\|\mathbf{x}\|_2 \\ \|\mathbf{Ax}\|_2 &\geq \sqrt{\lambda_{\min}(\mathbf{A}^T \mathbf{A})}\|\mathbf{x}\|_2 && \text{(def. of singular values)} \\ \|\mathbf{Ax}\|_2^2 &\geq \lambda_{\min}(\mathbf{A}^T \mathbf{A})\|\mathbf{x}\|_2^2 && \text{(square both sides)} \\ \lambda_{\max}((\mathbf{Ax})^T(\mathbf{Ax})) &\geq \lambda_{\min}(\mathbf{A}^T \mathbf{A})\|\mathbf{x}\|_2^2 && (\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})}) \\ \lambda_{\max}(\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}) &\geq \lambda_{\min}(\mathbf{A}^T \mathbf{A})(\mathbf{x}^T \mathbf{x}) && \text{(simplify)} \end{aligned}$$

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

Note that the ranks of  $\mathbf{A}^T \mathbf{A}$  and  $\mathbf{A} \mathbf{A}^T$  are different (and thus one of them contains at least one eigenvalue 0 if  $m > n$  or  $m < n$ ).

**(c) [5 pts]**

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

First, we prove  $|\lambda_i(\mathbf{A})| \leq \sigma_{\max}(\mathbf{A})$ .

Let  $\lambda$  be an eigenvalue of  $\mathbf{A}$ , and let  $\mathbf{x}$  be the corresponding unit eigenvector such that  $\mathbf{Ax} = \lambda\mathbf{x}$ . Taking the 2-norm of both sides:

$$\begin{aligned}\|\mathbf{Ax}\|_2 &= \|\lambda\mathbf{x}\|_2 \\ &= |\lambda|\|\mathbf{x}\|_2 \\ &= |\lambda| \quad (\text{since } \|\mathbf{x}\|_2 = 1)\end{aligned}$$

By definition,

$$\sigma_{\max}(\mathbf{A}) = \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2,$$

and so:

$$\sigma_{\max}(\mathbf{A}) \geq \|\mathbf{Ax}\|_2 = |\lambda_i(\mathbf{A})|, \forall i$$

For the other side,  $\sigma_{\min}(\mathbf{A}) \leq |\lambda_i(\mathbf{A})|$ ,

... TODO use (d) below. If  $\mathbf{A}$  is singular, its smallest singular value is zero and the inequality is trivially true.

## 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

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

and

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

### (a) [5 pts]

Show that

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

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

**(b) [5 pts]**

Show that

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

This follows directly from 3c and 4a above, along with the given definition of  $\lambda_{\min}(\mathbf{A})$ :

$$\begin{aligned}
\|\mathbf{A}^{-1}\|_2^{-1} &= \sigma_{\min}(\mathbf{A}) && \text{(From 4a)} \\
&\geq && \text{(from 3c, since } \mathbf{A} \text{ is square } m = n \Rightarrow \sigma_{\min}(\mathbf{A}) \leq |\lambda_i(\mathbf{A})| \text{)}
\end{aligned}$$

**(c) [5 pts]**

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

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

**(d) [5 pts]**

In the same setting as (c), show that

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

**(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

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

For symmetric matrices, the singular values are the absolute values of the eigenvalues.

## 5 Oblivion [25 pts]

### (a) [5 pts]

Have a look at the function `add_to_A_B_times_C!` in `HW2_your_code.jl`. This function adds the product of the input variables `B` and `C` to the input variable `A`. It is written such as to use the optimal loop order, and employs multiple optimizations by means of the `@turbo` macro.

Go to section (a) of the file `HW2_driver.jl` to see how this function performs when compared to the built-in `mul!` function. If necessary, adapt the size of the matrix to your system. Hand in the file `performance_per_size.pdf` produced by the code with your homework.

### (b)locked [5 pts]

Now implement a blocked/tiled variant of `add_to_A_B_times_C!` that takes an integer `bks` as a fourth input. This method should perform the matrix multiplication by calling the original `add_to_A_B_times_C!` on blocks of size (roughly) `bks` times `bks`. Again, your code should be correct and avoid memory allocations as checked by section (b) of `HW2_driver.jl`.

### (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.

### (d) oblivious [5 pts]

We can sometimes obtain better performance by using a hierarchical algorithm. Complete the skeleton for `oblivious_add_to_A_B_times_C!` to obtain an algorithm that equally divides each of its input matrices into four parts and then recurses on the resulting eight subproblems, until one of the problems reaches a size below `bks`. Again, make sure that your algorithm does not allocate memory and is correct using part (d) `HW2_driver.jl`.

### (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.