COT 5615 Math for Intelligent Systems, Fall 2021

Homework 9

Due Dec. 7

1. *Solving regularized least-squares.* Consider the least-squares problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda\|\boldsymbol{x}\|^2.$$

   The solution is

$$\widehat{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{A} \\ \sqrt{\lambda}\boldsymbol{I} \end{bmatrix}^{\dagger} \begin{bmatrix} \boldsymbol{b} \\ 0 \end{bmatrix} = (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^\top \boldsymbol{b}.$$

   This may suggest that the complexity is at least $\mathcal{O}(mn^2+n^3)$. Suppose $\boldsymbol{A}$ is wide, i.e., $m < n$. Explain how to solve it using $\mathcal{O}(m^2 n + m^3)$ flops. *Hint.* Use the "kernel trick" in §15.5.2 of the textbook.

2. *Companion matrix.* The matrix $\boldsymbol{C}$ below is called a companion matrix. Consider the companion matrix $\boldsymbol{C}$ (left) and the Vandermonde matrix $\boldsymbol{V}$ (right):

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & 0 & -c_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix} \qquad \boldsymbol{V} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdots & \lambda_3^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix}.$$

   (a) What is the characteristic polynomial of $\boldsymbol{C}$?

   (b) Assume that $\boldsymbol{C}$ has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Show that $\boldsymbol{V}$ is the matrix of left eigenvectors of $\boldsymbol{C}$, i.e., that we have $\boldsymbol{V}\boldsymbol{C} = \boldsymbol{\Lambda}\boldsymbol{V}$ where $\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$.

   (Recall that the eigenvalues are the roots of the characteristic polynomial. Some of you may know that polynomials of degree 5 or higher does not have analytical expressions of their roots. This is how their roots are found numerically in practice.)

3. *Circulant matrix.* The matrix $\boldsymbol{C}$ below is called a circulant matrix. The matrix $\boldsymbol{F}$ below is called the DFT matrix (discrete Fourier transform).

$$\boldsymbol{C} = \begin{bmatrix} c_1 & c_n & \cdots & c_3 & c_2 \\ c_2 & c_1 & \cdots & c_4 & c_3 \\ c_3 & c_2 & \cdots & c_5 & c_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_n & c_{n-1} & \cdots & c_2 & c_1 \end{bmatrix} \qquad \boldsymbol{F} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix},$$

   where $\omega = e^{-2\pi i/n}$ and $i = \sqrt{-1}$. Denote the first column of $\boldsymbol{C}$ as $\boldsymbol{c}$.

(a) Verify that $\boldsymbol{C} = \boldsymbol{F}^* \boldsymbol{\Lambda} \boldsymbol{F}$ where $\boldsymbol{\Lambda} = \mathrm{Diag}(\boldsymbol{F}\boldsymbol{c})$. The operation $\boldsymbol{F}\boldsymbol{c}$ is called the discrete Fourier transform (DFT) of $\boldsymbol{c}$. Since $\boldsymbol{F}^{-1} = \boldsymbol{F}^*$, this shows the eigenvectors of a circulant matrix are columns of the DFT matrix.

(b) Suppose you want to solve the linear equation $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{b}$ where $\boldsymbol{C}$ is a circulant matrix. How would you do it and what is the complexity? *Hint.* For the $n \times n$ DFT matrix $\boldsymbol{F}$, the matrix-vector multiplication can be done in $n \log(n)$ flops (rather than n^2 flops for general $n \times n$ matrices) using the fast Fourier transform (FFT) algorithm, and similarly for its conjugate transpose using the inverse FFT.

4. *General least-squares.* Consider the least-squares problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2.$$

If the columns of $\boldsymbol{A}$ are linearly independent, we have shown that the solution is unique and is

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^\dagger \boldsymbol{b} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{b}.$$

Now suppose that the columns may not be linearly independent. If we are given the 'thin' SVD of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, then we can show that a solution is

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^\dagger \boldsymbol{b} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^\top \boldsymbol{b}.$$

In this question we try to verify this result.

(a) Setting the gradient equal to zero, we obtain the equation $\boldsymbol{A}^\top \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{b}$. Verify that the solution obtained by the SVD satisfy this equation.

(b) Follow the steps on page 230 of the textbook, verify that it indeed minimizes the objective function.

5. *QRSVD.* For a $m \times n$ matrix $\boldsymbol{A}$ (assume $m \geq n$) the QR SVD algorithm is as follows

```
A = Q_0 R_0                         ▷ QR factorization of A
for k = 0, 1, 2, ... do
    R_k^T = Q_{k+1} R_{k+1}         ▷ QR factorization of R_k
end for
```

In the QR factorization it is assumed that diagonal entries of $\boldsymbol{R}$ are positive. If $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the 'thin' SVD of $\boldsymbol{A}$, then it is observed that the product $\boldsymbol{Q}_1\boldsymbol{Q}_3 \cdots \boldsymbol{Q}_{2k+1} \cdots$ converges to $\boldsymbol{V}$, the product $\boldsymbol{Q}_0\boldsymbol{Q}_2 \cdots \boldsymbol{Q}_{2k} \cdots$ converges to $\boldsymbol{U}$, and $\boldsymbol{R}_{2k}$ converges to $\boldsymbol{\Sigma}$.

(a) Try the algorithm on a random 6×3 and compare it with `svd(A)`.

(b) Show that $\boldsymbol{Q}_1(\boldsymbol{R}_1\boldsymbol{R}_0)$ is the QR decomposition of $\boldsymbol{R}_0^\top \boldsymbol{R}_0$ ($\boldsymbol{Q}_1$ is the orthonormal factor and $\boldsymbol{R}_1\boldsymbol{R}_0$ is the upper triangular factor).

(c) Show that $(\boldsymbol{R}_1\boldsymbol{R}_0)\boldsymbol{Q}_1 = \boldsymbol{R}_1\boldsymbol{R}_1^\top$ (Hint: focus first on the middle matrix $\boldsymbol{R}_0$ on the left hand side) and then that $\boldsymbol{R}_1\boldsymbol{R}_1^\top = \boldsymbol{R}_2^\top \boldsymbol{R}_2$. Show that $\boldsymbol{B}_0 = \boldsymbol{R}_0^\top \boldsymbol{R}_0$ and $\boldsymbol{B}_1 = \boldsymbol{R}_2^\top \boldsymbol{R}_2$ have the same set of eigenvalues.

(d) More generally consider the matrices $\boldsymbol{B}_k = \boldsymbol{R}_{2k}^\top \boldsymbol{R}_{2k}$ for $k = 1, 2, 3, \ldots$. Show that these are the QR iterations applied to the matrix $\boldsymbol{B}_0 = \boldsymbol{R}_0^\top \boldsymbol{R}_0$.