COT 5615 Math for Intelligent Systems, Fall 2021

Homework 9

1. Equation (15.10) states that

$$(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^\top = \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I})^{-1}.$$

Therefore the solution equals to

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

This implies that we can compute $\widehat{\boldsymbol{x}}$ as follows:

1: Form $\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}$ ($m^2 n$ flops)
2: Cholesky factorization $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}$ ($(1/3)m^3$ flops)
3: back substitution $\boldsymbol{L}^{-1}\boldsymbol{b}$ (n^2 flops)
4: forward substitution $\boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}\boldsymbol{b}$ (n^2 flops)
5: matrix-vector multiplication $\boldsymbol{A}^\top\boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}\boldsymbol{b}$ ($2mn$ flops)

The first two steps dominate, resulting in $\mathcal{O}(m^2 n + m^3)$ complexity.

Alternatively, notice that

$$\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I} = \begin{bmatrix} \boldsymbol{A}^\top \\ \sqrt{\lambda}\boldsymbol{I} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{A}^\top \\ \sqrt{\lambda}\boldsymbol{I} \end{bmatrix}.$$

If we take the QR factorization

$$\begin{bmatrix} \boldsymbol{A}^\top \\ \sqrt{\lambda}\boldsymbol{I} \end{bmatrix} = \boldsymbol{Q}\boldsymbol{R},$$

then

$$\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I} = \boldsymbol{R}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{R}^\top\boldsymbol{R}.$$

This suggests the following algorithm:

1: QR factorization $\begin{bmatrix} \boldsymbol{A}^\top \\ \sqrt{\lambda}\boldsymbol{I} \end{bmatrix} = \boldsymbol{Q}\boldsymbol{R}$ ($2(m+n)m^2$ flops)
2: back substitution $\boldsymbol{R}^{-\top}\boldsymbol{b}$ (n^2 flops)
3: forward substitution $\boldsymbol{R}^{-1}\boldsymbol{R}^{-\top}\boldsymbol{b}$ (n^2 flops)
4: matrix-vector multiplication $\boldsymbol{A}^\top\boldsymbol{R}^{-1}\boldsymbol{R}^{-\top}\boldsymbol{b}$ ($2mn$ flops)

The first step dominates, and the resulting complexity is again $\mathcal{O}(m^2 n + m^3)$.

Remark. Remember the useful rule-of-thumb: smart linear algebra (linear algebra done right) usually leads to the complexity of big times small squared.

2. (a) The characteristic polynomial of $\boldsymbol{C}$ is $p(x) = \det(x\boldsymbol{I} - \boldsymbol{C})$. We use the recursive definition of determinant

$$\det \boldsymbol{A} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \boldsymbol{A}_{ij},$$

where $\boldsymbol{A}_{ij}$ is the matrix obtained by removing the ith row and jth column of $\boldsymbol{A}$. We also use the properties that the determinant of a triangular matrix equals to the product of the diagonal entries and the determinant of a block-diagonal matrix equals to the product of the diagonal-block determinant. Let $\boldsymbol{A} = x\boldsymbol{I} - \boldsymbol{C}$ and apply the recursive definition to the last column of $\boldsymbol{A}$

$$\det(x\boldsymbol{I} - \boldsymbol{C}) = \sum_{i=1}^{n-1} (-1)^{i+n} c_{i-1} \det \boldsymbol{A}_{in} + (c_{n-1} + x) \det \boldsymbol{A}_{nn},$$

where $\boldsymbol{A}_{in}$ looks like

$$\boldsymbol{A}_{in} = \begin{bmatrix} x & 0 & 0 & \cdots & 0 & & & & & \\ -1 & x & 0 & & & & & & & \\ 0 & -1 & x & & \vdots & & & & & \\ \vdots & & \ddots & \ddots & & & & & & \\ 0 & \cdots & 0 & -1 & x & & & & & \\ & & & & & -1 & x & 0 & \cdots & 0 \\ & & & & & 0 & -1 & x & & \vdots \\ & & & & & \vdots & & \ddots & \ddots & \\ & & & & & & & & -1 & x \\ & & & & & 0 & \cdots & & 0 & -1 \end{bmatrix}.$$

The first diagonal block is $(i-1) \times (i-1)$. The second diagonal block is $(n-i) \times (n-i)$. Therefore $\det \boldsymbol{A}_{in} = (-1)^{n-i} x^{i-1}$. Plug this back, we get

$$\det(x\boldsymbol{I} - \boldsymbol{C}) = \sum_{i=1}^{n-1} (-1)^{i+n} c_{i-1} (-1)^{n-i} x^{i-1} + (c_{n-1} + x) x^{n-1} = x^n + \sum_{k=0}^{n-1} c_k x^k.$$

If the characteristic polynomial is defined as $\det(\boldsymbol{A} - x\boldsymbol{I})$, then there is a difference of $(-1)^n$ factor, which does not affect the roots.

(b) An eigenvalue of $\boldsymbol{C}$ is a root of its characteristic polynomial, so

$$\lambda_i^n + \sum_{k=0}^{n-1} c_k \lambda_i^k = 0, \quad \Longrightarrow \quad \lambda_i^n = -\sum_{k=0}^{n-1} c_k \lambda_i^k, \quad i = 1, \ldots, n.$$

Consider the ith row of $\boldsymbol{V}$

$$\boldsymbol{v}_i^\top = \left[\, 1 \ \lambda_i \ \lambda_i^2 \ \cdots \ \lambda_i^{n-1} \,\right].$$

We have

$$\boldsymbol{v}_i^\top \boldsymbol{C} = \left[\, \lambda_i \ \cdots \ \lambda_i^{n-1} \ \ -\sum_{k=0}^{n-1} c_k \lambda_i^k \,\right] = \left[\, \lambda_i \ \lambda_i^2 \ \cdots \ \lambda_i^n \,\right] = \lambda_i \boldsymbol{v}_i^\top.$$

Stacking all rows together gives

$$\boldsymbol{V}\boldsymbol{C} = \boldsymbol{\Lambda}\boldsymbol{V}.$$

3. (a) Equivalently, we verify $\boldsymbol{F}\boldsymbol{C} = \boldsymbol{\Lambda}\boldsymbol{F}$, or more specifically their jth row of both sides $\boldsymbol{f}_j^\top \boldsymbol{C} = \lambda_j \boldsymbol{f}_j^\top$. The kth entry of the left-hand-side is

$$\frac{1}{\sqrt{n}} \sum_{d=1}^{k-1} \omega^{(j-1)(d-1)} c_{n-k+d} + \sum_{d=k}^{n} \omega^{(j-1)(d-1)} c_{d-k+1}.$$

Notice that ω^d is periodic: $\omega^d = \omega^{d+n}$, so that sum becomes

$$\frac{1}{\sqrt{n}} \sum_{d=k}^{n+k-1} \omega^{(j-1)(d-1)} c_{d-k+1}.$$

Replace d with $d - k + 1$ and we have

$$\frac{1}{\sqrt{n}} \omega^{k-1} \sum_{d=1}^{n} \omega^{(j-1)(d-1)} c_d.$$

Moving on to the right-hand-side, λ_j is the jth entry of $\boldsymbol{F}\boldsymbol{c}$

$$\lambda_j = \sum_{d=1}^{n} \omega^{(j-1)(d-1)} c_d,$$

and thus the kth entry of $\lambda_j \boldsymbol{f}_j^\top$ is

$$\frac{1}{\sqrt{n}} \lambda_j \omega^{k-1} = \frac{1}{\sqrt{n}} \omega^{k-1} \sum_{d=1}^{n} \omega^{(j-1)(d-1)} c_d.$$

Therefore for all $j, k = 1, \ldots, n$, $[\boldsymbol{F}\boldsymbol{C}]_{jk} = [\boldsymbol{\Lambda}\boldsymbol{F}]_{jk}$, meaning $\boldsymbol{F}\boldsymbol{C} = \boldsymbol{\Lambda}\boldsymbol{F}$. Multiply both sides from the left with $\boldsymbol{F}^{-1} = \boldsymbol{F}^*$, we get $\boldsymbol{C} = \boldsymbol{F}^* \boldsymbol{\Lambda} \boldsymbol{F}$.

(b) Now we know every circulant matrix $\boldsymbol{C}$ factors into $\boldsymbol{F}^* \boldsymbol{\Lambda} \boldsymbol{F}$ where $\boldsymbol{\Lambda} = \sqrt{n}\,\mathrm{Diag}(\boldsymbol{F}\boldsymbol{c})$. To solve the equation $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{b}$ we know mathematically $\boldsymbol{x} = \boldsymbol{C}^{-1}\boldsymbol{b} = \boldsymbol{F}^* \boldsymbol{\Lambda}^{-1} \boldsymbol{F}\boldsymbol{b}$. This can be computed as follows:

1: Compute $\boldsymbol{F}\boldsymbol{b}$ via FFT ($n \log n$ flops)
2: Compute $\boldsymbol{F}\boldsymbol{c}$ via FFT ($n \log n$ flops)
3: Compute $\boldsymbol{\Lambda}^{-1}\boldsymbol{F}\boldsymbol{b}$ (n flops)
4: Compute $\boldsymbol{F}^* \boldsymbol{\Lambda}^{-1} \boldsymbol{F}\boldsymbol{b}$ via IFFT ($n \log n$ flops)

The overall complexity is $\mathcal{O}(n \log n)$.

If you have taken a course on Fourier analysis, you might resonate the following statements: The Fourier transform of a real signal is conjugate symmetric; the Fourier transform of a (conjugate) symmetric signal is real; (circular) convolution in the primal domain is direct multiplication in the Fourier domain; etc. These are all consistent with the eigen analysis of the corresponding circulant matrix.

4. (a) Replace $\boldsymbol{A}$ with $\boldsymbol{U\Sigma V}^\top$ and $\widehat{\boldsymbol{x}}$ with $\boldsymbol{V\Sigma}^{-1}\boldsymbol{U}^\top\boldsymbol{b}$ into the equation $\boldsymbol{A}^\top\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^\top\boldsymbol{b}$, we get

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

The two sides are indeed equal.

(b) For a particular vector $\widehat{\boldsymbol{x}}$ that satisfies $\boldsymbol{A}^\top\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^\top\boldsymbol{b}$, we have

$$\begin{aligned}
\|\boldsymbol{Ax}-\boldsymbol{b}\|^2 &= \|\boldsymbol{Ax}-\boldsymbol{A}\widehat{\boldsymbol{x}}+\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{b}\|^2 \\
&= \|\boldsymbol{Ax}-\boldsymbol{A}\widehat{\boldsymbol{x}}\|^2 + \|\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{b}\|^2 - 2(\boldsymbol{Ax}-\boldsymbol{A}\widehat{\boldsymbol{x}})^\top(\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{b}) \\
&= \underbrace{\|\boldsymbol{Ax}-\boldsymbol{A}\widehat{\boldsymbol{x}}\|^2}_{\geq 0} + \|\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{b}\|^2 - 2(\boldsymbol{x}-\widehat{\boldsymbol{x}})^\top\underbrace{(\boldsymbol{A}^\top\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{A}^\top\boldsymbol{b})}_{=0} \\
&\geq \|\boldsymbol{A}\widehat{\boldsymbol{x}}-\boldsymbol{b}\|^2
\end{aligned}$$

So any $\widehat{\boldsymbol{x}}$ that satisfies $\boldsymbol{A}^\top\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^\top\boldsymbol{b}$ is an optimal solution that minimizes $\|\boldsymbol{Ax}-\boldsymbol{b}\|^2$. It may not be unique, and one particular choice can be obtained via the SVD. If $\boldsymbol{A}^\top\boldsymbol{A}$ is invertible, then the solution is unique and given by $(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{b}$, and is a special case of the SVD solution.

5. (a) The purpose is to show you that this simple algorithm actually works.

(b) In the first iteration we have $\boldsymbol{R}_0^\top = \boldsymbol{Q}_1\boldsymbol{R}_1$. Therefore $\boldsymbol{R}_0^\top\boldsymbol{R}_0 = \boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{R}_0$. The matrix $\boldsymbol{Q}_1$ has orthonormal columns. The product of two upper triangular matrices, $\boldsymbol{R}_1\boldsymbol{R}_0$, is also upper triangular. The QR factorization tries to factor the matrix into the product of an orthonormal matrix and an upper triangular matrix, and $\boldsymbol{Q}_1(\boldsymbol{R}_1\boldsymbol{R}_0)$ meets the requirement.

(c) Since $\boldsymbol{R}_0^\top = \boldsymbol{Q}_1\boldsymbol{R}_1$, multiply from the left with $\boldsymbol{Q}_1^\top$ we get $\boldsymbol{Q}_1^\top\boldsymbol{R}_0^\top = \boldsymbol{Q}_1^\top\boldsymbol{Q}_1\boldsymbol{R}_1 = \boldsymbol{R}_1$. Therefore $\boldsymbol{R}_1\boldsymbol{R}_1^\top = \boldsymbol{R}_1\boldsymbol{R}_0\boldsymbol{Q}_1$.
In the second iteration $\boldsymbol{R}_1^\top = \boldsymbol{Q}_2\boldsymbol{R}_2$, so $\boldsymbol{R}_1\boldsymbol{R}_1^\top = \boldsymbol{R}_2^\top\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\boldsymbol{R}_2 = \boldsymbol{R}_2^\top\boldsymbol{R}_2$.
As a result,

$$\boldsymbol{R}_2^\top\boldsymbol{R}_2 = \boldsymbol{R}_1\boldsymbol{R}_1^\top = \boldsymbol{R}_1\boldsymbol{R}_0\boldsymbol{Q}_1 = \boldsymbol{Q}_1^\top\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{R}_0\boldsymbol{Q}_1 = \boldsymbol{Q}_1^\top\boldsymbol{R}_0^\top\boldsymbol{R}_0\boldsymbol{Q}_1.$$

This means if $\boldsymbol{R}_2^\top\boldsymbol{R}_2\boldsymbol{v} = \lambda\boldsymbol{v}$, then $\boldsymbol{R}_0^\top\boldsymbol{R}_0(\boldsymbol{Q}_1\boldsymbol{v}) = \lambda(\boldsymbol{Q}_1\boldsymbol{v})$, for all eigenvalues of $\boldsymbol{R}_2^\top\boldsymbol{R}_2$. In other words, $\boldsymbol{B}_0 = \boldsymbol{R}_0^\top\boldsymbol{R}_0$ and $\boldsymbol{B}_1 = \boldsymbol{R}_2^\top\boldsymbol{R}_2$ have the same set of eigenvalues.

(d) With similar arguments, we see that $\boldsymbol{B}_k$ and $\boldsymbol{B}_{k+1}$ have the same set of eigenvalues. The QR factorization of $\boldsymbol{B}_k$ is $\boldsymbol{Q}_{k+1}(\boldsymbol{R}_{k+1}\boldsymbol{R}_k)$, and $\boldsymbol{B}_{k+1}$ is obtain by $(\boldsymbol{R}_{k+1}\boldsymbol{R}_k)\boldsymbol{Q}_{k+1}$. This indeed shows that $\{\boldsymbol{B}_k\}$ is the sequence obtained by applying the QR iteration to $\boldsymbol{B}_0 = \boldsymbol{R}_0^\top\boldsymbol{R}_0$. Notice that $\boldsymbol{B}_0$ also equals to

$$\boldsymbol{B}_0 = \boldsymbol{R}_0^\top\boldsymbol{R}_0 = \boldsymbol{R}_0^\top\boldsymbol{Q}_0^\top\boldsymbol{Q}_0\boldsymbol{R}_0 = \boldsymbol{A}^\top\boldsymbol{A},$$

and the eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$ are the squared singular values of $\boldsymbol{A}$.