

Scientific Computing HW 3

Ryan Chen

September 18, 2024

1. (a) Compute

$$\begin{aligned} P^T &= I^T - 2(uu^T)^T \\ &= I - 2(u^T)^T u^T \\ &= I - 2uu^T \\ &= P \end{aligned}$$

Using $u^T u = 1$,

$$\begin{aligned} P^2 &= (I - 2uu^T)(I - 2uu^T) \\ &= I^2 - 2Iuu^T - 2uu^T I + 4uu^T uu^T \\ &= I - 4uu^T + 4uu^T \\ &= I \end{aligned}$$

- (b) Using $\text{sign}(a)^2 = 1$ and $\text{sign}(a)a = |a|$, compute

$$\begin{aligned} \tilde{u}^T x &= \|x\|^2 + \text{sign}(x_1)x_1\|x\| = \|x\|^2 + |x_1|\|x\| \\ \tilde{u}^T \tilde{u} &= \|x\|^2 + \text{sign}(x_1)^2\|x\|^2 + 2\text{sign}(x_1)x_1\|x\| = 2\|x\|^2 + 2|x_1|\|x\| \end{aligned}$$

Then compute

$$\begin{aligned} Px &= x - 2 \frac{\tilde{u}^T x}{\tilde{u}^T \tilde{u}} \tilde{u} \\ &= x - \tilde{u} \\ &= -\text{sign}(x_1)\|x\|e_1 \end{aligned}$$

- (c) For convenience, write matrix entries as superscripts, reserving subscripts for the iteration described in the problem. We will prove using finite induction that for $1 \leq j \leq n$, the first j diagonal entries of A_j are

$$-\text{sign}(A_0^{11})\|A_0(1:m, 1)\|, \dots, -\text{sign}(A_{j-1}^{jj})\|A_{j-1}(j:m, j)\|$$

and that the entries below the first j diagonal entries are zero.

The case $j = 1$ is proven by $\text{House}(A_0(1:m, 1))A_0(1:m, 1) = -\text{sign}(A_0^{11})\|A_0(1:m, 1)\|e_1$, which is the first column of $A_1 = P_1 A_0$.

Assume the claim is true for j . Then (use \sim to denote possibly nonzero entries)

$$A_j = \left[\begin{array}{ccc|c} -\text{sign}(A_0^{11})\|A_0(1:m, 1)\| & & \sim & \\ & \ddots & & \\ & & -\text{sign}(A_{j-1}^{jj})\|A_{j-1}(j:m, j)\| & \sim \\ \hline & 0_{(m-j) \times j} & & \sim \end{array} \right]$$

From the algorithm,

$$P_{j+1} = \left[\begin{array}{c|c} I_{j \times j} & 0_{j \times (m-j)} \\ \hline 0_{(m-j) \times j} & \text{House}(A_j(j+1 : m, j+1)) \end{array} \right]$$

Then, by the fact $\text{House}(A_j(j+1 : m, j+1))A_j(j+1 : m, j+1) = -\text{sign}(A_j^{j+1, j+1})\|A_j(j+1 : m, j+1)\|e_1$,

$$A_{j+1} = P_{j+1}A_j = \left[\begin{array}{ccc|c} -\text{sign}(A_0^{11})\|A_0(1 : m, 1)\| & & & \sim \\ & \ddots & & \\ & & -\text{sign}(A_j^{j+1, j+1})\|A_j(j+1 : m, j+1)\| & \sim \\ \hline & & 0_{(m-j-1) \times (j+1)} & \sim \end{array} \right]$$

This closes the induction. Having proven the claim, setting $j = n$ gives

$$A_n = \left[\begin{array}{ccc|c} -\text{sign}(A_0^{11})\|A_0(1 : m, 1)\| & & & \sim \\ & \ddots & & \\ & & -\text{sign}(A_{n-1}^{nn})\|A_{n-1}(n : m, n)\| & \\ \hline & & 0_{(m-n) \times n} & \end{array} \right]$$

To guarantee that A_n has positive diagonal entries, note that

$$-\text{sign}(A_{j-1}^{jj})\|A_{j-1}(j : m, j)\| > 0 \iff -\text{sign}(A_{j-1}^{jj}) > 0 \iff A_{j-1}^{jj} < 0$$

If $A_{j-1}^{jj} > 0$, change its sign before constructing P_j .

(d) From the iteration,

$$A_n = P_n P_{n-1} \dots P_1 A$$

From $P_j^2 = I$, we have $P_j^{-1} = P_j$, hence

$$(P_n P_{n-1} \dots P_1)^{-1} = P_1^{-1} P_2^{-1} \dots P_n^{-1} = P_1 P_2 \dots P_n$$

Which in turn gives a QR decomposition of A .

$$A = \underbrace{P_1 \dots P_n}_{=:Q} \underbrace{A_n}_{=:R}$$

2. (a) From the eigendecomposition $A = U\Lambda U^T$, we see U is orthogonal, so its columns u_1, \dots, u_n form an orthonormal basis of \mathbb{R}^n , and A has eigenpairs (λ_i, u_i) . Define

$$\Sigma := \text{diag}(\sigma_1, \dots, \sigma_n), \quad \sigma_i := |\lambda_i|$$

$$V := [v_1 \ \dots \ v_n], \quad v_i := s_i u_i, \quad s_i := \text{sign}(\lambda_i)$$

Then we have

$$V = U \text{diag}(s_1, \dots, s_n), \quad V^T = \text{diag}(s_1, \dots, s_n) U^T, \quad \text{diag}(s_1, \dots, s_n)^2 = I_{n \times n}$$

The last equation comes from the fact that $\text{sign}(x)^2 = 1$. Hence

$$\begin{aligned} V^T V &= \text{diag}(s_1, \dots, s_n) U^T U \text{diag}(s_1, \dots, s_n) \\ &= \text{diag}(s_1, \dots, s_n)^2 & U \text{ is orthogonal} \\ &= I_{n \times n} & \text{diag}(s_1, \dots, s_n)^2 = I_{n \times n} \end{aligned}$$

and

$$\begin{aligned} V V^T &= U \text{diag}(s_1, \dots, s_n)^2 U^T \\ &= U U^T & \text{diag}(s_1, \dots, s_n)^2 = I_{n \times n} \\ &= I_{n \times n} & U \text{ is orthogonal} \end{aligned}$$

Thus V is orthogonal.

Lastly, we check that $A = U\Sigma V^T$ by showing that the action of both sides agrees on the basis u_1, \dots, u_n of \mathbb{R}^n , i.e. $Au_i = U\Sigma V^T u_i$ for $1 \leq i \leq n$.

$$\begin{aligned} U\Sigma V^T u_i &= U\Sigma(s_i e_i) & \text{the } u_i \text{'s are orthonormal} \\ &= U(s_i |\lambda_i| e_i) \\ &= U(\lambda_i e_i) & \text{sign}(x)|x| = x \\ &= \lambda_i u_i \\ &= Au_i & (\lambda_i, u_i) \text{ is an eigenpair of } A \end{aligned}$$

- (b) Using the SVD $A = U\Sigma V^T$,

$$\begin{aligned} A^T A &= V\Sigma U^T U\Sigma V^T \\ &= V\Sigma^2 V^T & U^T U = I_{n \times n} \\ &= V \text{diag}(\sigma_1^2, \dots, \sigma_n^2) V^T \end{aligned}$$

This is an eigendecomposition of $A^T A$ with eigenpairs (σ_i^2, v_i) .

- (c) Via the “basis extension theorem” and the Gram–Schmidt process, we can extend the linearly independent orthonormal set u_1, \dots, u_n to an orthonormal basis $u_1, \dots, u_n, u_{n+1}, \dots, u_m$ of \mathbb{R}^m . Since AA^T is $m \times m$, it requires an $m \times m$ eigendecomposition.

$$\begin{aligned} AA^T &= U\Sigma V^T V\Sigma U^T \\ &= U\Sigma^2 U^T \\ &= [U \mid u_{n+1} \mid \dots \mid u_m] \left[\begin{array}{c|c} \Sigma^2 & 0_{n \times (m-n)} \\ \hline 0_{(m-n) \times n} & 0_{(m-n) \times (m-n)} \end{array} \right] \begin{bmatrix} U^T \\ \hline \frac{u_{n+1}^T}{u_m^T} \\ \vdots \\ \hline \end{bmatrix} \end{aligned}$$

This is an eigendecomposition of AA^T with eigenpairs (σ_i^2, u_i) for $1 \leq i \leq n$ and $(0, u_i)$ for $n+1 \leq i \leq m$.

- (d) From A being full rank, the least squares solution to $Ax = b$ is $x^* = (A^T A)^{-1} A^T b$. From the SVD $A = U \Sigma V^T$ and the fact A is full rank (so that $A^T A$ and Σ are nonsingular),

$$A^T A = V \Sigma^2 V^T \implies (A^T A)^{-1} = V (\Sigma^{-1})^2 V^T$$

Hence

$$\begin{aligned} x^* &= (A^T A)^{-1} A^T b \\ &= V (\Sigma^{-1})^2 V^T V \Sigma U^T b \\ &= V (\Sigma^{-1})^2 \Sigma U^T b \\ &= V \Sigma^{-1} U^T b \end{aligned}$$

- (e) We use the fact that $\|A\|_2$ equals the largest singular value of A . From the SVD $A = U \Sigma V^T$ and the fact A is nonsingular (Σ is nonsingular), we have $A^{-1} = V \Sigma^{-1} U^T$, an SVD of A^{-1} . Assume the singular values of A are arranged in decreasing order, $\sigma_1 \geq \dots \geq \sigma_n > 0$. Then Σ^{-1} gives the singular values of A^{-1} , which arranged in decreasing order are $\frac{1}{\sigma_n} \geq \dots \geq \frac{1}{\sigma_1}$. Thus $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$.
- (f) Throughout we use the fact that $Av_i = \sigma_i u_i$ for $1 \leq i \leq n$, and that for vectors w_1, \dots, w_k , $\text{span}(w_1, \dots, w_k)$ is the smallest subspace containing w_1, \dots, w_k .

We first show that $\text{null}(A) = \text{span}(v_{r+1}, \dots, v_n)$.

- Let $x \in \text{null}(A)$. Using the basis v_1, \dots, v_n , write

$$x = \sum_{i=1}^n c_i v_i$$

for some $c_1, \dots, c_n \in \mathbb{R}$. This, along with $Ax = 0$ and $Av_i = \sigma_i u_i$, gives

$$\sum_{i=1}^r c_i \sigma_i u_i = 0$$

Since the u_i 's are linearly independent, $c_i \sigma_i = 0$, hence $c_i = 0$, for $1 \leq i \leq r$. Thus

$$x = \sum_{i=r+1}^n c_i v_i \in \text{span}(v_{r+1}, \dots, v_n)$$

We conclude $\text{null}(A) \subset \text{span}(v_{r+1}, \dots, v_n)$.

- For $r+1 \leq i \leq n$, we have $Av_i = \sigma_i u_i = 0$, so $v_i \in \text{null}(A)$. Then $\text{null}(A)$ is a subspace of \mathbb{R}^n containing v_{r+1}, \dots, v_n , hence $\text{span}(v_{r+1}, \dots, v_n) \subset \text{null}(A)$.

Then we show that $\text{range}(A) = \text{span}(u_1, \dots, u_r)$.

- Let $y \in \text{range}(A)$. Then $y = Ax$ for some $x \in \mathbb{R}^n$. Using the basis v_1, \dots, v_n , write

$$x = \sum_{i=1}^n c_i v_i$$

for some $c_1, \dots, c_n \in \mathbb{R}$. This, along with $Av_i = \sigma_i u_i$, gives

$$y = Ax = \sum_{i=1}^r c_i \sigma_i u_i \in \text{span}(u_1, \dots, u_r)$$

Thus $\text{range}(A) \subset \text{span}(u_1, \dots, u_r)$.

- For $1 \leq i \leq r$, using the fact $Av_i = \sigma_i u_i$, we have $u_i = A \left(\frac{1}{\sigma_i} v_i \right) \in \text{range}(A)$. Then $\text{range}(A)$ is a subspace of \mathbb{R}^m containing u_1, \dots, u_r , hence $\text{span}(u_1, \dots, u_r) \subset \text{range}(A)$.

From $\text{range}(A) = \text{span}(u_1, \dots, u_r)$ and the u_i 's being linearly independent, we see that u_1, \dots, u_r form a basis of $\text{range}(A)$, hence $\text{rank}(A) = r$.

3. The claim is that (here we omit the subscript for the 2-norm)

$$\min_{Ax=b} \|x\| = \|x^*\|$$

First we see that x^* solves $Ax = b$.

$$Ax^* = AA^T(AA^T)^{-1}b = b$$

Now let x solve $Ax = b$. Then

$$Ax = Ax^* \implies A(x - x^*) = 0$$

From this,

$$(x - x^*)^T x^* = (x - x^*)^T A^T (AA^T)^{-1} b = (A(x - x^*))^T (AA^T)^{-1} b = 0$$

In turn, we get

$$\|x\|^2 = \|(x - x^*) + x^*\|^2 = \|x - x^*\|^2 + \|x^*\|^2 + 2(x - x^*)^T x^* = \|x - x^*\|^2 + \|x^*\|^2 \geq \|x^*\|^2$$

Hence $Ax^* = b$, with $\|x\| \geq \|x^*\|$ whenever $Ax = b$. This establishes the claim.

4. (a) From the Schur decomposition $A = QTQ^*$, we have $AQ = QT$. If (λ, v) is an eigenpair of T , write $Tv = \lambda v$, so that $AQv = QTv = \lambda Qv$, hence (λ, Qv) is an eigenpair of A .
- (b) Since T is upper triangular, its eigenvalues are the diagonal entries $\lambda_1, \lambda_2, \lambda_3$. Moreover, $Te_1 = \lambda_1 e_1$, confirming (λ_1, v_1) as an eigenpair with $v_1 = e_1$. Seeking an eigenpair of the form (λ_2, v_2) with $v_2 = [a, 1, 0]^T$, compute

$$Tv_2 = \begin{bmatrix} \lambda_1 a + t_{12} \\ \lambda_2 \\ 0 \end{bmatrix}, \quad \lambda_2 v_2 = \begin{bmatrix} \lambda_2 a \\ \lambda_2 \\ 0 \end{bmatrix}$$

Equating the first components,

$$\lambda_1 a + t_{12} = \lambda_2 a \implies (\lambda_2 - \lambda_1) = t_{12} \implies a = \frac{t_{12}}{\lambda_2 - \lambda_1}$$

Seeking an eigenpair of the form (λ_3, v_3) with $v_3 = [b, c, 1]^T$, compute

$$Tv_3 = \begin{bmatrix} \lambda_1 b + t_{12}c + t_{13} \\ \lambda_2 c + t_{23} \\ \lambda_3 \end{bmatrix}, \quad \lambda_3 v_3 = \begin{bmatrix} \lambda_3 b \\ \lambda_3 c \\ \lambda_3 \end{bmatrix}$$

Equating the second components,

$$\lambda_2 c + t_{23} = \lambda_3 c \implies (\lambda_3 - \lambda_2)c = t_{23} \implies c = \frac{t_{23}}{\lambda_3 - \lambda_2}$$

Equating the first components,

$$\lambda_1 b + t_{12}c + t_{13} = \lambda_3 b \implies (\lambda_3 - \lambda_1)b = t_{12}c + t_{13} \implies b = \frac{t_{12}c}{\lambda_3 - \lambda_1} + \frac{t_{13}}{\lambda_3 - \lambda_1}$$

Using the formula for c ,

$$b = \frac{t_{12}t_{23}}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} + \frac{t_{13}}{\lambda_3 - \lambda_1}$$

- (c) By parts (a) and (b), eigenvectors of A are

$$Qv_1 = Qe_1 = q_1$$

$$Qv_2 = aq_1 + q_2$$

$$Qv_3 = bq_1 + cq_2 + q_3$$

where

$$a = \frac{t_{12}}{\lambda_2 - \lambda_1}, \quad b = \frac{t_{12}t_{23}}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} + \frac{t_{13}}{\lambda_3 - \lambda_1}, \quad c = \frac{t_{23}}{\lambda_3 - \lambda_2}$$

5. (a)
(b)