Section 5.4 Problems

Ryan Heslin

2023-02-24

1.

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

3.

Say we have bases of V and $V^\perp$. Then the union of the bases forms a basis for R^n, because all vectors in R^n are either linear combinations of V or are orthogonal to, except for 0, which is both.

4.

Yes. The image of A^T is all vectors in R^m obtained by $A^T x$. $\ker(A)^\perp$ is all vectors in R^m for which $Ax \neq 0$.

5.

The solution space V of

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 + 2x_2 + 5x_3 + 4x_4 &= 0 \end{aligned}$$

is the kernel of the matrix. So $V^\perp = \text{im} A^T$:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \\ 1 & 4 \end{bmatrix}$$

6.

If A is $n \times m$, then $im(A) = (AA^T$ is true. $ imA^T$ is the orthogonal complement of the kernel, so it contains all nonzero x for which $Ax \neq 0$ - in other words, all vectors in $(A^\perp)^\perp$

7.

If a matrix is symmetric, the image and kernel are orthogonal complements, because $A^T = A$, so $\ker(A^T) = \ker(A)$. Likewise, the row space and $\ker(A)$ are orthogonal complements.

8.

a. $A^+ = (A^TA)^{-1}A^T$

b.

$$\begin{aligned} A^+ &= (A^TA)^{-1}A^T \\ &= A^{-1}(A^T)^{-1}A^T \\ &= A^{-1} \end{aligned}$$

c. $(A^TA)^{-1}A^TAx = x$

d. $A(A^TA)^{-1}A^Ty = y^{\parallel}$

e.

$$L^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

9.

b.

They are orthogonal complements.

c.

It is a basis for $im(A^T)$.

d.

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

e.

Being the minimal solution, it is the shortest.

10.

a. If x_0 is in $\ker A^{\perp}$, then it lies in the image of the transpose, the orthogonal complement. Then since $x = x_h + x_0$, if we set $x_h = 0$ (the portion in the kernel), then x_0 lies entirely in the image of the transpose. all vectors for which that lead to nonzero b, as well as 0_m

b. 0_m is the only vector shared among $(\ker(A^{\perp}))$ and $\ker(A)$, so x_0 lies entirely in the image of the transpose only if $x_h = 0$.

c. For all linear combinations x_0 of $\ker(A)$, $Ax_0 = 0$ by definition. So combinations of $\ker(A)$ can be freely added to x_h without impacting the solution, since $A(x_0 + x_h) = Ax_0 + Ax_h = 0 + Ax_h$. Since all nonzero vectors have nonzero length, this makes Ax_h the shortest solution.

11.

a. Given the definition of the minimal solution, the minimal least-squares solution is the next best thing: the one solution to $(A^TA)^{-1}A^Tx$ lying in $\ker(A)^{\perp}$). That is the image of the transpose, so this unique solution is purely a linear combination of A^T, without any vectors from $\ker(A)$.

Had these backwards initially.

b. $(A^TA)^{-1}A^TA = I$

c. $A(A^TA)^{-1}A^T$

d. The image is R^n, the kernel is $\ker(A)$

e.The first two elements of y.

12.

The minimal least-squares solution of a system is the shortest solution x^+ that yields an Ax^+ the shortest distance from b. It always lies in $(\ker A)^{\perp}$ because it is the one and only x^+ that lies entirely in A's row space.

13.

a. $L(x) = y$ is linear, so $L(y_1 + y_2)$ is the minimal least-squares solution of $L(x) = y_1 + y_2$, which is the sum of the separate least-squares solutions for y_1 and y_2. For the second property:

$$L(kx) = kL(x)$$
$$L^+(L(kx)) = L^+(k(L(x))) = kL^+(L(x))$$

(I was initially not quite right on these before looking up the answers). b. $L^+(L(X))$ is the minimum least-squares solution of $L(x) = L(X)$ - that is, x. More correctly, the projection of x onto the image of A^T.

c. The projection of y onto the image of the row space.

d. The image and kernel of L^+ are the same as those of A^T.

e. If

$$L(x) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x$$

, then the pseudoinverse is just $\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

15.

It is the pseudoinverse, $(A^TA)^{-1}A^T$. We have:

$$(A^TA)^{-1}A^TA = I$$

17.

Yes. If this were not true, then $\dim(\ker(A^T)) \geq \dim(\ker(A))$. This is impossible, because the orthogonal complement of A's image is $\ker(A^T)$ So no x solving $A^T x = 0$ may be produced by a linear combination of A. Therefore, $\ker(A^T A) = \ker(A)$, and both matrices have m columns, so ranks are equal as well.

18.

Yes. We proved above $A^T A$ and A have equal rank. $\ker(A^T)$ is the orthogonal complement of the image of A^T, so any x for which $Ax = 0$ cannot come from a linear combination of A^T (except for 0_m). So the kernel does not expand, and rank remains the same.

19.

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

28.

For an orthonormal basis, the least squares solution is b, since $A(A^T A)^{-1} A^T = I$.

34.

```r
x <- b <- c(0, 0.5, 1, 1.5, 2, 2.5, 3)
X <- cbind(1, sin(x), cos(x), sin(2 * x), cos(2 *
    x))

solve(t(X) %*% X) %*% t(X) %*% b
```

```
            [,1]
[1,]  1.50000000
[2,]  0.10897422
[3,] -1.53669122
[4,]  0.30269197
[5,]  0.04314769
```

36.

I fit a model predicting day length by time of year. The error vector is surprisingly small.

```r
fit <- function(A) {
    solve(t(A) %*% A) %*% t(A)
}
days <- c(32, 77, 121, 152)
b <- c(10, 12, 14, 15)

A <- cbind(rep(1, 4), sin(((2 * pi)/366) * days),
    cos(((2 * pi)/366) * days))
```

```r
betas <- fit(A) %*% b

b - A %*% betas
```

```
        [,1]
```

[1,] -0.01328337 [2,] 0.03559242 [3,] -0.04424210 [4,] 0.02193305

39.

Another exponential fit problem.

a.

```r
A <- cbind(rep(1, 5), log(c(6e+05, 2e+05, 60000,
    10000, 2500)))
z <- c(5, 12, 25, 60, 250)

betas <- fit(A) %*% log(z)
betas
```

```
       [,1]
```

[1,] 10.6031951 [2,] -0.6739273

c.

The exponential base of the fitted function is about 0.5, very close to the theoretical $a = k\sqrt{g}$.

```r
k <- exp(betas[1])
g <- exp(betas[2])
sprintf("k = %.2f, g = %.2f", k, g)
```

```
[1] "k = 40263.28, g = 0.51"
```

41.

Let's predict the US national debt!

```r
A <- cbind(rep(1, 4), seq(0, 30, by = 10))

b <- log(c(533, 1823, 4974, 7933))
betas <- fit(A) %*% b

cbind(A[, 2], exp(b)) %*% betas
```

```
      [,1]
```

[1,] 48.52718 [2,] 230.51271 [3,] 581.93366 [4,] 915.87389

```
# exp(betas[1] +betas[2] * c(A[,2], 40))
```

For 2015, the model predicts a debt of more than $24 trillion.

70.

Let $A_1 = A - \lambda_1 I_n$ and $A_2 = A - \lambda_2 I_n$. By the rank-nullity theorem, $\text{nullity}(A_1 A_2) = \text{nullity} A_1 + \text{nullity} A_2$ if and only if $\ker A_1 \in \text{im} A_2$; otherwise, only vectors mapped by A_2 to 0 are in the kernel. Since both domain and codomain are $\mathbb{R}^n$, this requires the kernels to be linearly independent subspaces. If we set $A_1\prime = A_1 A_2$ and repeat this process for each eigenvalue, the same logic still holds. The matrix is only diagonalizable if $\ker A_1 \dots A_n = \mathbb{R}^n$, which means the combined eigenspaces span all of $\mathbb{R}^n$. But if an eigenvalue of multiplicity m is mapped to an eigenspace of dimension $n < m$, then the dimension of the eigenspaces fails to sum to n and (non-Jordan) diagonalization is impossible.