Pr. 1.

When $\boldsymbol{D}$ is an $M \times N$ (rectangular) diagonal matrix, its **pseudoinverse** $\boldsymbol{D}^+$ is an $N \times M$ (rectangular) diagonal matrix whose nonzero entries are the reciprocals $1/d_k$ of the nonzero diagonal entries of $\boldsymbol{D}$. A matrix $\boldsymbol{A}$ having SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ has $\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}'$. Working with just this equality, determine by hand (experimenting / checking with code is OK) the pseudoinverse of

(a) $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{y}'$, where neither $\boldsymbol{x} \in \mathbb{F}^M$ nor $\boldsymbol{y} \in \mathbb{F}^N$ is $\mathbf{0}$, (your answer should depend only on $\boldsymbol{x}$ and $\boldsymbol{y}$),

(b) $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{x}'$, where $\boldsymbol{x} \neq \mathbf{0}$.

Pr. 2.

(Projection onto orthogonal complement of a 1D subspace)

(a) Determine an orthonormal basis for the orthogonal complement of the span of the vector $\boldsymbol{v} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$.

(b) Determine an orthonormal basis for the orthogonal complement of the span of the vector $\boldsymbol{z} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$.

(c) Determine (by hand, no `Julia`) the projection of the vector $\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ onto the orthogonal complement of $\mathrm{span}(\{\boldsymbol{z}\})$ *without* using the orthonormal basis you found in the previous part.

Hint. You may want to derive the general mathematical expression needed in the next part first, and then use that expression to solve this problem by hand.

(d) Write a function called `orthcomp1` that projects an input vector $\boldsymbol{y}$ onto the orthogonal complement of $\mathrm{span}(\{\boldsymbol{x}\})$ for an (nonzero) input vector $\boldsymbol{x}$ of the same length as $\boldsymbol{y}$.

For full credit, your final version of the code must be computationally efficient, and it should be able to handle input vectors of length 10 million without running out of memory. Your code must *not* call `svd` or `eig` or `I` and the like. This problem can be solved in one line with elementary vector operations.

In `Julia`, your file should be named `orthcomp1.jl` and should contain the following function:

```
"""
    z = orthcomp1(y, x)

Project `y` onto the orthogonal complement of `Span({x})`

# In:
* `y` vector
* `x` nonzero vector of same length, both possibly very long

# Out:
* `z` vector of same length

For full credit, your solution should be computationally efficient.
"""
function orthcomp1(y, x)
```

Email your solution as an attachment to `eecs551@autograder.eecs.umich.edu`.

Test your code yourself using the example above (and others) *before* submitting to the autograder. Be sure to test it for very long input vectors.

(e) Submit your code (a screen capture is fine) to gradescope so that the grader can verify that your code is computationally *efficient*. (The autograder checks only correctness, not efficiency.)

Pr. 3.

Let $\boldsymbol{A}$ be a diagonal matrix with real entries that are distinct and nonzero. Let $\boldsymbol{x}$ be a vector with all nonzero entries.

(a) Determine how the **eigenvalues** of the **rank-1 update** $\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{x}\boldsymbol{x}'$ are related to the eigenvalues of $\boldsymbol{A}$ and the vector $\boldsymbol{x}$. Your final expression must not have any matrices in it.

Hint. They will be implicitly related (not in a closed form expression) as the solution of an equation involving the eigenvalues of $\boldsymbol{A}$ and the elements of $\boldsymbol{x}$. Hint: $\boldsymbol{G} + \boldsymbol{H} = \boldsymbol{G}(\boldsymbol{I} + \boldsymbol{G}^{-1}\boldsymbol{H})$ if $\boldsymbol{G}$ is invertible, and see HW1.

(b) Use your solution to the previous part to determine the eigenvalues of $\boldsymbol{B}$ when $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ and $\boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

Hint: `Julia`'s `Polynomials.jl` package can be useful here.

Pr. 4.

(a) Show that $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ is a convex function on $\mathbb{R}^N$ when matrix $\boldsymbol{A}$ has N columns.
Hint. Use the **triangle inequality**: $\|\boldsymbol{x} + \boldsymbol{y}\|_2 \leq \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2$ and the equality $1 = \alpha + (1 - \alpha)$.

(b) Show that the largest singular value of a matrix, *i.e.*, the function $\sigma_1(\boldsymbol{X}) : \mathbb{F}^{M \times N} \mapsto \mathbb{R}$, is a convex function of the elements of the $M \times N$ matrix $\boldsymbol{X}$. Hint. Use the fact that $\sigma_1(\boldsymbol{X}) = \max_{\|\boldsymbol{u}\|_2 = 1} \|\boldsymbol{X}\boldsymbol{u}\|_2$. Also use these two basic properties of maxima / **suprema**: $f(t) \leq g(t) \Rightarrow \max_t f(t) \leq \max_t g(t)$ and $\max_t (f(t) + g(t)) \leq (\max_t f(t)) + (\max_t g(t))$.

(c) (Optional.) Prove or disprove that the second singular value of a matrix, *i.e.*, the function $\sigma_2(\boldsymbol{X}) : \mathbb{F}^{M \times N} \mapsto \mathbb{R}$, is a convex function of the elements of the $M \times N$ matrix $\boldsymbol{X}$ when $\min(M, N) \geq 2$.

Pr. 5.

Linear regression has a myriad of uses, including investigation of **social justice** issues.
Demo 05/sat-regress has data collected by the College Board–the organization that runs the **SAT exam** for high-school students, from Lesser, 2017. This data includes average SAT Math scores for 10 different family annual income brackets. This problem uses this data to explore the relationship between income and SAT scores.
Make a scatter plot of the midpoint of each income bracket versus the SAT Math score. The last income bracket says "100+" which does not have a midpoint, so just set it to be 120 (K\$). Use the template code in the demo.
The scatter plot suggests there is a strong correlation between income and SAT score. You are going to fit four different models to the data (one at a time) and examine the fits and the coefficients:

$$\text{SAT\_Math} \approx \begin{cases} \beta_0, & \text{constant} \\ \beta_1 \cdot \text{ income}, & \text{linear} \\ \beta_0 + \beta_1 \cdot \text{ income}, & \text{affine} \\ \beta_0 + \beta_1 \cdot \text{ income} + \beta_2 \cdot (\text{income})^2, & \text{quadratic.} \end{cases}$$

For each of the four models, use a LLS fit to determine the β coefficient(s).

(a) Make a single plot showing the data and the four fits, for incomes between 0 and 130 (K\$), *i.e.*, `income = 0:130`. (We cannot predict anything for even higher incomes with this data.) Be sure to label your axes and use a legend to explain which points/lines are which.

(b) Make a table to report your β coefficients that looks like this:

Constant Fit	Linear Fit	Affine Fit	Quadratic Fit
β_0	0	β_0	β_0
0	β_1	β_1	β_1
0	0	0	β_2

(c) (Optional) In a statistics class, you would analyze the coefficients β_1 and β_2 to assess whether they are significantly different from 0 and establish evidence of a relationship. Even without formal statistics, you should be able to look at your plot and your table of coefficients and draw some conclusions about how equitable the SAT exam is.

Pr. 6.

Let $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ denote two orthonormal vectors and $\boldsymbol{b}$ some fixed vector, all in $\mathbb{F}^N$.

(a) Find the optimal linear combination $\alpha \boldsymbol{q}_1 + \beta \boldsymbol{q}_2$ that is **closest** to $\boldsymbol{b}$ (in the **Euclidean norm** sense).

(b) Let $\boldsymbol{r} = \boldsymbol{b} - \alpha \boldsymbol{q}_1 - \beta \boldsymbol{q}_2$ denote the "residual error vector." Show that $\boldsymbol{r}$ is **orthogonal** to both $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$.

Pr. 7.

Polynomial regression application

Let $f(t) = 0.5\, e^{0.8t},\ t \in [0, 2]$.

(a) Suppose you are given 16 exact measurements of $f(t)$, taken at the times t in the following 1D array:

```
T = range(0, 2, 16)
```

Use `Julia` to generate 16 exact measurements: $b_i = f(t_i),\ i = 1, \ldots, 16$, for $t_i \in$ `T`.
Now determine the coefficients of the **least-squares** polynomial approximation of the data $\boldsymbol{b}$ for

(i) a polynomial of degree 15: $p_{15}(t)$;

(ii) a polynomial of degree 2: $p_2(t)$.

Compare the quality of the two approximations graphically. Use `scatter` to first show b_i vs t_i for $i = 1, \ldots, 16$, then use `plot!` to add plots of $p_{15}(t)$ and $p_2(t)$ and $f(t)$ to see how well they approximate the function on the interval $[0, 2]$. Pick a very fine grid for the interval, *e.g.*, `t = (0:1000)/500`. Make the y-axis range equal $[-1, 4]$ by using the `ylim=(-1,4)` option. As always, include axis labels and a clear legend. Submit your plot and also summarize the results qualitatively in one or two sentences.

For this problem, write your own `Julia` code rather than using `fit` in the `Polynomials` package. (You may check your solutions using that function.)

(b) Now suppose the measurements are affected by some noise. Generate the measurements using $y_i = f(t_i) + e_i,\ i = 1, \ldots, 16,$, as follows. You *must* use the `seed` to get the correct values!

```
using Random:  seed!
seed!(3); e = randn(length(T))
```

Determine the coefficients of the least-squares polynomial approximation of the (noisy) measurements $\boldsymbol{y}$ for

(i) a polynomial of degree 15: $p_{15}(t)$;

(ii) a polynomial of degree 2: $p_2(t)$.

Compare the two approximations as in part (a). Again make the y-axis range equal $[-1, 4]$ by using the `ylim=(-1,4)` option. Submit your plot and also summarize the results qualitatively in a couple of sentences, including comparing the behavior in (a) and (b).

(c) Let $\hat{\boldsymbol{x}}_n(\boldsymbol{b})$ and $\hat{\boldsymbol{x}}_n(\boldsymbol{y})$ denote the LLS polynomial coeffients from noiseless $\boldsymbol{b}$ and noisy $\boldsymbol{y}$, respectively, for a polynomial of degree n. Report the values of the **residual** norms $\|\boldsymbol{A}\hat{\boldsymbol{x}}_n(\boldsymbol{b}) - \boldsymbol{b}\|_2$ and $\|\boldsymbol{A}\hat{\boldsymbol{x}}_n(\boldsymbol{y}) - \boldsymbol{y}\|_2$ for the polynomial fits of degree 2 and degree 15. These residual norms describe how closely the fitted curve fits the *data*.
Also report the fitting errors $\|\boldsymbol{A}\hat{\boldsymbol{x}}_n(\boldsymbol{y}) - \boldsymbol{b}\|_2$ for $n = 2, 15$. Arrange your results in a table as follows:

polynomial degree:	$d = 2$	$d = 15$
Residual norm $\|\boldsymbol{A}\hat{\boldsymbol{x}}(\boldsymbol{b}) - \boldsymbol{b}\|_2$ noiseless (a)	?	?
Residual norm $\|\boldsymbol{A}\hat{\boldsymbol{x}}(\boldsymbol{y}) - \boldsymbol{y}\|_2$ noisy (b)	?	?
Fitting error $\|\boldsymbol{A}\hat{\boldsymbol{x}}(\boldsymbol{y}) - \boldsymbol{b}\|_2$	?	?

Hint: the bottom left entry is between 1.6 and 1.9.

(d) Explain why the residual norm for degree 2 is smaller or larger than that of degree 15.

Non-graded problem(s) below
(Solutions will be provided for self check; do not submit to gradescope.)

Pr. 8.
A matrix $\boldsymbol{T}$ has the property $\boldsymbol{T}^3 = \boldsymbol{I}$. What its possible **eigenvalues**?

Pr. 9.
Determine the **eigenvalues** and **eigenvectors** of $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{x}' + \boldsymbol{y}\boldsymbol{y}'$, where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^N$.
You need only find the eigenvector(s) that correspond to nonzero eigenvalue(s).
Assume that $\boldsymbol{x}'\boldsymbol{y} = \rho \neq 0$ and focus on the case where $\boldsymbol{A}$ has its maximum possible **rank**.
Hint. The desired eigenvector(s) of $\boldsymbol{A}$ must be in the **range** of $\boldsymbol{A}$. Hence, any eigenvector must be linearly related to $\boldsymbol{x}$ and $\boldsymbol{y}$. Also, write $\boldsymbol{A}$ as $\boldsymbol{Z}\boldsymbol{Z}'$ for some simple matrix $\boldsymbol{Z}$. Check your answers numerically.

Pr. 10.
Spherical manifold optimization problems

(a) Suppose $\boldsymbol{A} \in \mathbb{F}^{M\times N}$ has rank $0 < r \leq \min(M, N)$ and SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k'$.

(1) $\boldsymbol{x}_{\mathsf{opt}} = \arg\max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = ?$

(2) $\boldsymbol{y}_{\mathsf{opt}} = \arg\max_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{A}'\boldsymbol{y}\|_2 = ?$

(3) When are $\boldsymbol{x}_{\mathsf{opt}}$ and $\boldsymbol{y}_{\mathsf{opt}}$ unique (to within a sign ambiguity)?

(4) Do answers change if you replace $\boldsymbol{A}$ with $-\boldsymbol{A}$ in the optimization problems? Explain why or why not.

(5) What constraints could you add to above manifold optimization problem (with the same objective function) so you get $\boldsymbol{u}_r$ and $\boldsymbol{v}_r$?

(b) Suppose $\boldsymbol{A} \in \mathbb{F}^{N\times N}$ is Hermitian, and has eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' = \sum_{k=1}^{N} \lambda_k \boldsymbol{v}_k \boldsymbol{v}_k'$, with descending eigenvalue ordering $\lambda_1 \geq \ldots \geq \lambda_N \in \mathbb{R}$. Define: $\boldsymbol{x}_{\mathsf{opt}} = \arg\max_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$.

(1) Prove that $\boldsymbol{x}_{\mathsf{opt}} = z\boldsymbol{v}_1$ where $|z| = 1$.

(2) What is $\boldsymbol{x}_{\mathsf{opt}}'\boldsymbol{A}\boldsymbol{x}_{\mathsf{opt}}$?

(3) When is $\boldsymbol{x}_{\mathsf{opt}}$ unique (aside from the sign ambiguity)?

(c) Same $\boldsymbol{A}$ as in previous part. For some $1 < K \leq r$, define:

$$\boldsymbol{x}_{\mathsf{opt}} = \arg\max_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \text{ subject to } \boldsymbol{x} \perp \boldsymbol{v}_1,\ \boldsymbol{x} \perp \boldsymbol{v}_2,\ \ldots,\ \boldsymbol{x} \perp \boldsymbol{v}_{K-1}.$$

(1) Prove that $\boldsymbol{x}_{\mathsf{opt}} = z\boldsymbol{v}_K$ where $|z| = 1$, via an equivalent formulation involving projections.

(2) What is $\boldsymbol{x}_{\mathsf{opt}}'\boldsymbol{A}\boldsymbol{x}_{\mathsf{opt}}$?

(3) When is $\boldsymbol{x}_{\mathsf{opt}}$ unique (aside from the sign ambiguity)?

Pr. 11.
Consider the following set of three measurements (x_i, y_i): $(1, 2),\ (2, 1),\ (3, 3)$.

(a) Find the line of the form $y = \alpha x + \beta$ that **best fits** (in the 2-norm sense) this data.

(b) Find the line of the form $x = \gamma y + \delta$ that best fits (in the 2-norm sense) this data.
Hint. Re-use your answer from part (a).
