Pr. 1.

(a) Prove the "**vec trick**," *i.e.*, that for arbitrary (possibly complex-valued) matrices of compatible sizes:

$$\text{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^T) = (\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{X}), \tag{1}$$

where **vec(.)** is the operator that stacks the columns of the input matrix into a vector, and $\otimes$ denotes the **Kronecker product**. Note that here we really mean transpose $\boldsymbol{B}^T$ even if $\boldsymbol{B}$ is complex valued!
For your solution, use: $\boldsymbol{A}$ is $p \times m$, $\boldsymbol{X}$ is $m \times n$, and $\boldsymbol{B}$ is $q \times n$.

(b) Suppose $\boldsymbol{A}$, $\boldsymbol{X}$, and $\boldsymbol{B}$ are all $N \times N$ dense matrices. Determine how many scalar multiplications are needed for the LHS $\text{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^T)$ and compare to the number for the RHS $(\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{X})$.

Which version uses fewer multiplications?

One application of (1) is computing a 2D **discrete Fourier transform** (DFT), which may be explored in a later HW problem.

Pr. 2.
Let $\boldsymbol{A}$ be a diagonal matrix with real entries that are distinct and non-zero. Let $\boldsymbol{x}$ be a vector with all non-zero entries.

(a) Determine how the **eigenvalues** of the **rank-one 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 look at 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. 3.
Let $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ denote two orthonormal vectors and $\boldsymbol{b}$ some fixed vector, all in $\mathbb{R}^n$.

(a) Find the optimal linear combination $\alpha\boldsymbol{q}_1 + \beta\boldsymbol{q}_2$ that is **closest** to $\boldsymbol{b}$ (in the **2-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. 4.
For $\boldsymbol{A} \in \mathbb{F}^{M \times N}$, $\boldsymbol{b} \in \mathbb{F}^M$ and $\boldsymbol{x} \in \mathbb{F}^N$, show that $(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}$ and $\boldsymbol{A}^+\boldsymbol{b}$ are **orthogonal** vectors, where $\boldsymbol{A}^+$ denotes the **Moore-Penrose pseudo-inverse** of $\boldsymbol{A}$. Hint: use a **compact SVD**.

Pr. 5.
A function $f : \mathcal{V} \mapsto \mathbb{R}$ is called a **convex function** on vector space $\mathcal{V}$ iff for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{V}$, and every $\alpha \in [0, 1]$,

$$f(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \leq \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2).$$

Show that:

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

(b) 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 $\sigma_1(\boldsymbol{X}) = \max_{\|\boldsymbol{u}\|_2 = 1} \|\boldsymbol{X}\boldsymbol{u}\|_2$.

Pr. 6.

Linear regression has a myriad of uses, including investigation of **social justice** issues. In the Canvas HW folder is a file named `hpj4s.csv` that has data collected by the College Board–the organization that runs the **SAT exam** for high-school students. 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.

First, load the data and 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 let's just set it to be 120 (K\$). To get you started, here is `Julia` code for loading and plotting.

```julia
using CSV # you might need to add this Package
using Plots; default(markerstrokecolor=:auto)

data = CSV.File("hpj4s.csv", normalizenames=true, transpose=true)
math = data.Math # math scores from the file (should be from 457 to 564)
income = [5:10:75; 90; 120]
scatter(income, math, label="Data", ylim=[425,575], ytick=425:50:575)
plot!(xlabel = "Family Annual Income (1000\$)")
plot!(ylabel = "SAT Average Math Score", legend = :bottomright)
```

From the scatter plot, it looks like 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 least-squares 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. 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 = LinRange(0, 2, 16)`

Use `Julia` to generate 16 exact measurements: $b_i = f(t_i), \ i = 1, \ldots, 16, \quad t_i \in \boldsymbol{T}$.

Now determine the coefficients of the least square error polynomial approximation of the data $\boldsymbol{b}$ for

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

(b) Now suppose the measurements are affected by some noise. Generate the measurements using $y_i = f(t_i) + e_i, \ i = 1, \ldots, 16, \quad t_i \in \boldsymbol{T}$, where you generate the vector of noise values as follows:

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

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

(a) a polynomial of degree 15: $p_{15}(t)$;
(b) 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 1.0 and 1.3.

(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.
Spherical manifold optimization problems

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

- $\boldsymbol{x}_{\text{opt}} = \arg\max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{Ax}\|_2 = \,?$
- $\boldsymbol{y}_{\text{opt}} = \arg\max_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{A}'\boldsymbol{y}\|_2 = \,?$
- When are $\boldsymbol{x}_{\text{opt}}$ and $\boldsymbol{y}_{\text{opt}}$ unique (to within a sign ambiguity)?
- Do answers change if you replace $\boldsymbol{A}$ with $-\boldsymbol{A}$ in the optimization problems? Explain why or why not?
- 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 $\boldsymbol{A} = \boldsymbol{V\Lambda V}' = \sum_{k=1}^N \lambda_k \boldsymbol{v}_k \boldsymbol{v}_k'$, with descending eigenvalue ordering $\lambda_1 \geq \ldots \geq \lambda_N$ and $\lambda_k \in \mathbb{R}$. Define:

$$\boldsymbol{x}_{\text{opt}} = \arg\max_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}'\boldsymbol{Ax}.$$

- Prove that $\boldsymbol{x}_{\text{opt}} = z\boldsymbol{v}_1$ where $|z| = 1$.
- What is $\boldsymbol{x}_{\text{opt}}'\boldsymbol{A}\boldsymbol{x}_{\text{opt}}$?
- When is $\boldsymbol{x}_{\text{opt}}$ unique (aside from the sign ambiguity)?

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

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

- Prove that $\boldsymbol{x}_{\text{opt}} = z\boldsymbol{v}_K$ where $|z| = 1$, via an equivalent formulation involving projections.
- What is $\boldsymbol{x}_{\text{opt}}'\boldsymbol{A}\boldsymbol{x}_{\text{opt}}$?
- When is $\boldsymbol{x}_{\text{opt}}$ unique (aside from the sign ambiguity)?

Pr. 9.
Suppose $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ with $M \geq N$ has **full column rank**, *i.e.*, $\text{rank}(\boldsymbol{A}) = N$.
Show, using an SVD of $\boldsymbol{A}$, that $\boldsymbol{A}^+ = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. When $\boldsymbol{A}$ is known to be full column rank, one could use this direct formula to compute the **pseudo-inverse** instead of employing an SVD.

Pr. 10.
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 (a).
