Pr. 1. (sol/hs010)
If λ is an eigenvalue of $\boldsymbol{A}$, then there exists a nonzero vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, *i.e.*, $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$.
Thus $\boldsymbol{B}\boldsymbol{x} = (\boldsymbol{A} - 10\boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} - 10\boldsymbol{x} = \lambda\boldsymbol{x} - 10\boldsymbol{x} = (\lambda - 10)\boldsymbol{x}$.
Thus $\lambda - 10$ is an eigenvalue of $\boldsymbol{B}$ corresponding to the same eigenvector $\boldsymbol{x}$.
Conversely, if z is an eigenvalue of $\boldsymbol{B}$ with eigenvector $\boldsymbol{y}$, then using $\boldsymbol{A} = \boldsymbol{B} + 10\boldsymbol{I}$ the same logic shows that $\boldsymbol{y}$ is also an eigenvector of $\boldsymbol{A}$ with eigenvalue $z + 10$.
In summary, $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvectors and $\lambda_B = \lambda_A - 10$.

What follows here is an answer that applies *only* when $\boldsymbol{A}$ is diagonalizable. Student answers that match the following earn partial but not full credit, because the problem did not state that $\boldsymbol{A}$ is diagonalizable.
Suppose $\boldsymbol{A}$ has eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$. Then because $\boldsymbol{I} = \boldsymbol{V}\boldsymbol{V}^{-1}$ we may write

$$\boldsymbol{B} = \boldsymbol{A} - 10\boldsymbol{I} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} - 10\boldsymbol{V}\boldsymbol{V}^{-1} = \boldsymbol{V}(\boldsymbol{\Lambda} - 10\boldsymbol{I})\boldsymbol{V}^{-1}. \tag{1}$$

Because $\boldsymbol{\Lambda} - 10\boldsymbol{I}$ is a diagonal matrix, the right hand side of (1) is an eigendecomposition of $\boldsymbol{B}$. So we conclude that the eigenvectors of $\boldsymbol{B}$ are identical to the eigenvectors of $\boldsymbol{A}$. Further, if λ_B is an eigenvalue of $\boldsymbol{B}$, and λ_A is an eigenvalue of $\boldsymbol{A}$, then $\lambda_B = \lambda_A - 10$.

Pr. 2. (sol/hs013)
Only if:
We are given that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$ are orthonormal and that $\boldsymbol{A}\boldsymbol{v}_i, i = 1, \ldots, N$ are orthonormal. We want to show $\boldsymbol{A}$ is an orthogonal matrix. By assumption, the matrix $\boldsymbol{A}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_N\end{bmatrix} \triangleq \boldsymbol{A}\boldsymbol{V}$ is orthogonal. Thus

$$\begin{aligned} \boldsymbol{I} &= (\boldsymbol{A}\boldsymbol{V})^T(\boldsymbol{A}\boldsymbol{V}) = \boldsymbol{V}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{V} \\ \Rightarrow \boldsymbol{V}\boldsymbol{I}\boldsymbol{V}' &= \boldsymbol{V}\boldsymbol{V}'\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{V}\boldsymbol{V}' \Rightarrow \boldsymbol{I} = \boldsymbol{A}^T\boldsymbol{A}, \end{aligned}$$

using the orthogonality of $\boldsymbol{V}$. Using similar steps, we can show that $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$, and conclude that $\boldsymbol{A}$ is orthogonal.

If:
We are given that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$ are orthonormal and that $\boldsymbol{A}$ is orthogonal. Then

$$(\boldsymbol{A}\boldsymbol{v}_i)^T(\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_i^T \underbrace{\boldsymbol{A}^T\boldsymbol{A}}_{=\boldsymbol{I}} \boldsymbol{v}_j = \boldsymbol{v}_i^T\boldsymbol{v}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \qquad \text{(since } \boldsymbol{v}_i \text{ are orthonormal).}$$

Thus $\{\boldsymbol{A}\boldsymbol{v}_i\}$ are also orthonormal vectors.

Pr. 3. (sol/hs022)

$$\text{Here } \boldsymbol{A} = \begin{bmatrix} a_{11} & \ldots & a_{1N} \\ & \ddots & \\ a_{M1} & \ldots & a_{MN} \end{bmatrix}, \text{ so } \boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} \sum_{i=1}^M |a_{i1}|^2 & & \text{off diagonal elements} \\ & \ddots & \\ \text{off diagonal elements} & & \sum_{i=1}^M |a_{iN}|^2 \end{bmatrix}.$$

$$\Rightarrow \operatorname{Tr}(\boldsymbol{A}'\boldsymbol{A}) = \sum_{j=1}^N \sum_{i=1}^M |a_{ij}|^2 = \sum_{(i,j)} |a_{ij}|^2 = \|\boldsymbol{A}\|_{\mathrm{F}}^2.$$

Thus

$$\begin{aligned} \|\boldsymbol{A}\|_{\mathrm{F}} &= \sqrt{\operatorname{Tr}(\boldsymbol{A}'\boldsymbol{A})} = \sqrt{\operatorname{Tr}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\boldsymbol{U}')} && \text{(using the SVD } \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\text{)} \\ &= \sqrt{\operatorname{Tr}(\boldsymbol{U}'\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}')} && \text{(because } \operatorname{Tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{Tr}(\boldsymbol{B}\boldsymbol{A}) \text{ as shown in HW 1)} \\ &= \sqrt{\operatorname{Tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')} = \sqrt{\sum_{i=1}^r \sigma_i^2} = \sqrt{\sum_{i=1}^{\min(M,N)} \sigma_i^2}, \end{aligned}$$

where r is the rank of the matrix $\boldsymbol{A}$. (Either final summation expression is fine.)

Pr. 4. (sol/hs028)
In a previous problem we showed that

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \geq \sigma_1,$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$ are the singular values of $\boldsymbol{A}$, and r is the rank of $\boldsymbol{A}$. We also know that

$$\begin{aligned} r = & \text{ row rank of } \boldsymbol{A} = \# \text{ linearly independent rows of } \boldsymbol{A} \leq M \\ = & \text{ column rank of } \boldsymbol{A} = \# \text{ independent rows of } \boldsymbol{A} \leq N \\ \leq & \min(M, N). \end{aligned}$$

Finally

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \leq \sqrt{\sum_{i=1}^{r} \sigma_1^2} \leq \sqrt{\sum_{i=1}^{\min(M,N)} \sigma_1^2} \leq \sqrt{\min(M,N)}\, \sigma_1.$$

Optional: The upper bound is tight; consider $\boldsymbol{A} = \boldsymbol{I}_N$ for which $\sigma_1 = 1$ and $\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{N}$.

Pr. 5. (sol/hs095)

$$\operatorname{vec}(\boldsymbol{x}\boldsymbol{y}^T) = \operatorname{vec}\left(\begin{bmatrix} x_1y_1 & x_1y_2 & \ldots & x_1y_N \\ x_2y_1 & x_2y_2 & \ldots & x_2y_N \\ & & \ddots & \\ x_My_1 & x_My_2 & \ldots & x_My_N \end{bmatrix}\right) = \begin{bmatrix} x_1y_1 \\ x_2y_1 \\ \vdots \\ x_My_1 \\ --- \\ x_1y_2 \\ x_2y_2 \\ \vdots \\ x_My_2 \\ --- \\ \vdots \\ --- \\ x_1y_N \\ x_2y_N \\ \vdots \\ x_My_N \end{bmatrix} = \begin{bmatrix} y_1\boldsymbol{x} \\ y_2\boldsymbol{x} \\ \vdots \\ y_N\boldsymbol{x} \end{bmatrix} = \boldsymbol{y} \otimes \boldsymbol{x}.$$

Pr. 6. (sol/hs017)
For $\boldsymbol{A} \in \mathbb{F}^{N \times N}$, a SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ always exists. When the rank of $\boldsymbol{A}$ is N, $\boldsymbol{\Sigma}$ is an $N \times N$ diagonal matrix with strictly positive entries; otherwise it is an $N \times N$ diagonal matrix with $r < N$ strictly positive entries and is hence positive semidefinite. Since $\boldsymbol{V}$ is a unitary matrix, $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$ and hence we can express $\boldsymbol{A}$ as:

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \boldsymbol{U}\boldsymbol{I}\boldsymbol{\Sigma}\boldsymbol{V}' \quad \text{Because } \boldsymbol{A} \text{ is square, } \boldsymbol{U}, \boldsymbol{V} \text{ and } \boldsymbol{I} \text{ have the same dimensions.} \\ &= \underbrace{\boldsymbol{U}\boldsymbol{V}'}_{\triangleq \boldsymbol{Q}} \underbrace{\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}'}_{\triangleq \boldsymbol{S}}. \quad \text{This is called a } \textbf{polar factorization}. \end{aligned}$$

Recall from HW1 that $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}'$ is an unitary matrix because $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{V}' = \boldsymbol{V}\boldsymbol{I}\boldsymbol{V}' = \boldsymbol{I}$. The matrix $\boldsymbol{S} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}'$ is symmetric because $\boldsymbol{S}' = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}' = \boldsymbol{S}$. It is positive definite when its eigenvalues, which are exactly equal to the N strictly positive singular values of $\boldsymbol{A}$, are strictly positive, and it is positive semidefinite when the rank of $\boldsymbol{A}$ is less than N. Note $\boldsymbol{S} = \boldsymbol{S}'$ and all the eigenvalues of $\boldsymbol{S}$ being non-negative (or positive) is a sufficient and necessary test for positive semidefiniteness (resp. positive definiteness) of a matrix.

Pr. 7. (sol/hsj13)
Graders: Correct Yes answer earns 2/2 points.
Correct No answer earns 1/2 points; anything similar to the explanations below earns the other 1 point.
There may be other correct reasons for the "No" answers; if you see any please email me.

(a) The set of numbers that are irrational or zero, *i.e.*, the set $(\mathbb{R} - \mathbb{Q}) \cup \{0\}$. No.

- It does not include the "1" element.
- It is not closed under multiplication: $\sqrt{2}$ and $2/\sqrt{2}$ are both irrational, but their product 2 is not in the set.

(b) The set of $N \times N$ diagonal matrices (where the "1" element is $\boldsymbol{I}_N$, and the "0" element is $\mathbf{0}_{N \times N}$.
No, because $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is in the set but does not have an inverse.

(c) The set of $N \times N$ diagonal matrices whose diagonal elements are either all zero or all non-zero.
No, the set is not closed under addition. Consider $\mathrm{Diag}(1, 2) + \mathrm{Diag}(1, -2)$

(d) The set of **rational functions**, *i.e.*, functions of the form $P(x)/Q(x)$ where P and Q are both polynomials and Q is not zero.
Yes! All the conditions hold.

(e) The set of $N \times N$ invertible matrices along with the $N \times N$ zero matrix. No.

- Not closed under addition: $\mathrm{Diag}(1, 2) + \mathrm{Diag}(1, -2)$ is not invertible.
- Matrices in this set do not commute in general.

Pr. 8. (sol/hs012)

(a) For SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, one way to find the eigenvalues of $\boldsymbol{B}$ is to look at

$$\det(\lambda \boldsymbol{I} - \boldsymbol{B}) = \det\left(\begin{bmatrix} \lambda \boldsymbol{I} & -\boldsymbol{A} \\ -\boldsymbol{A}' & \lambda \boldsymbol{I} \end{bmatrix}\right) = \det(\lambda \boldsymbol{I}) \det(\lambda \boldsymbol{I} - \boldsymbol{A}'(\lambda \boldsymbol{I})^{-1}\boldsymbol{A}) = \det(\lambda^2 \boldsymbol{I} - \boldsymbol{A}'\boldsymbol{A}) = \det(\lambda^2 \boldsymbol{I} - \boldsymbol{\Sigma}'\boldsymbol{\Sigma}).$$

So clearly each eigenvalue of $\boldsymbol{B}$ is a zero of $\lambda^2 - \sigma_i^2$, where σ_i denotes the ith singular value of $\boldsymbol{A}$. Because $\boldsymbol{B}$ is symmetric, it has real eigenvalues, so the eigenvalues of $\boldsymbol{B}$ are $\pm\sigma_i$.

To find the eigenvectors of $\boldsymbol{B}$, note that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}' & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}' & \boldsymbol{0} \end{bmatrix} \quad (\boldsymbol{\Sigma}' = \boldsymbol{\Sigma} \text{ because } \boldsymbol{A} \text{ is square}).$$

Hence (using "guess and check" approach):

$$\boldsymbol{B} \begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{v}_i \\ \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}'\boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} \sigma_i \boldsymbol{u}_i \\ \sigma_i \boldsymbol{v}_i \end{bmatrix} = \sigma_i \begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix},$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the left and right singular vectors of $\boldsymbol{A}$ associated with the singular value σ_i. Thus $\boldsymbol{x}_i \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix}$ is an eigenvector of $\boldsymbol{B}$ corresponding to the eigenvalue σ_i. We include the $1/\sqrt{2}$ factor here because in the context of eigendecompositions, we normalize the eigenvectors to have unit norm. (Check that $\boldsymbol{x}_i$ is unit norm!)

Similarly:

$$\boldsymbol{B} \begin{bmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{bmatrix} = \begin{bmatrix} -\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{v}_i \\ \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}'\boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} -\sigma_i \boldsymbol{u}_i \\ \sigma_i \boldsymbol{v}_i \end{bmatrix} = -\sigma_i \begin{bmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{bmatrix},$$

so that $\boldsymbol{z}_i \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{bmatrix}$ is a unit-norm eigenvector of $\boldsymbol{B}$ corresponding to the eigenvalue $-\sigma_i$.

(Grader: normalization is not required for part (a), and any nonzero scale factor times the eigenvectors shown above are also valid eigenvectors.)

(b) Optional. Because $\boldsymbol{B}$ is Hermitian, it should have a unitary eigendecomposition, so we check that our unit-norm eigenvectors are orthonormal.

First: $\boldsymbol{x}_i'\boldsymbol{x}_j = \begin{bmatrix} \boldsymbol{u}_i' & \boldsymbol{v}_i' \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{v}_j \end{bmatrix} = \boldsymbol{u}_i\boldsymbol{u}_j + \boldsymbol{v}_i\boldsymbol{v}_j = 0 + 0 = 0$, for $i \neq j$, because $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices in an SVD. Similarly $\boldsymbol{z}_i'\boldsymbol{z}_j = 0$, for $i \neq j$.

On a related note: $\boldsymbol{z}_i'\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{u}_i' & -\boldsymbol{v}_i' \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix} = 1 - 1 = 0$ and $\boldsymbol{z}_i'\boldsymbol{x}_j = \begin{bmatrix} \boldsymbol{u}_i' & -\boldsymbol{v}_i' \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{v}_j \end{bmatrix} = 0 - 0 = 0$ for $i \neq j$, so every $\boldsymbol{x}_i$ is orthogonal to every $\boldsymbol{z}_j$.

Thus the collection of $2N$ eigenvectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$ is an orthonormal set.

Combining into matrix form, a unitary eigendecomposition of $\boldsymbol{B}$ is

$$\boldsymbol{B} = \boldsymbol{Q} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Sigma} \end{bmatrix} \boldsymbol{Q}' \text{ where } \boldsymbol{Q} \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{bmatrix}.$$

Pr. 9. (sol/hsj04)
Thank you for submitting your personal plan.
Grader: given full credit for any submission that looks like a sincere attempt to make a personal plan with more than one goal. There are no "right or wrong" answers here; credit is earned for effort. No answer earns 0 points points. If you see answers that look insincere, then assign them 0 points with a separate rubric and I will take a close look at those.

Pr. 10. (sol/hs074)
We are given that

$$\texttt{FXY} = \begin{bmatrix} f(x_1,y_1) & \dots & f(x_1,y_n) \\ \vdots & \dots & \vdots \\ f(x_m,y_1) & \dots & f(x_m,y_n) \end{bmatrix}$$

$$\texttt{DFDX} = \begin{bmatrix} f(x_2,y_1)-f(x_1,y_1) & \dots & f(x_2,y_n)-f(x_1,y_n) \\ f(x_3,y_1)-f(x_2,y_1) & \dots & f(x_3,y_n)-f(x_2,y_n) \\ \vdots & \dots & \vdots \\ f(x_m,y_1)-f(x_{m-1},y_1) & \dots & f(x_m,y_n)-f(x_{m-1},y_n) \\ f(x_1,y_1)-f(x_m,y_1) & \dots & f(x_1,y_n)-f(x_m,y_n) \end{bmatrix}$$

$$\texttt{DFDY} = \begin{bmatrix} f(x_1,y_2)-f(x_1,y_1) & \dots & f(x_1,y_n)-f(x_1,y_{n-1}) & f(x_1,y_1)-f(x_1,y_n) \\ \vdots & \dots & \vdots & \vdots \\ f(x_m,y_2)-f(x_m,y_1) & \dots & f(x_m,y_n)-f(x_m,y_{n-1}) & f(x_m,y_1)-f(x_m,y_n) \end{bmatrix}.$$

Easy way using the "vec trick" from Ch. 1.
It is clear from the above expressions that $\texttt{DFDX} = \boldsymbol{D}_m \, \texttt{FXY}$. Thus by the vec trick:

$$\texttt{dfdx} = \operatorname{vec}(\texttt{DFDX}) = \operatorname{vec}(\boldsymbol{D}_m \, \texttt{FXY} \, \boldsymbol{I}_n) = (\boldsymbol{I}_n \otimes \boldsymbol{D}_m) \operatorname{vec}(\texttt{FXY}) = (\boldsymbol{I}_n \otimes \boldsymbol{D}_m) \, \texttt{fxy}.$$

Similarly $\texttt{DFDY} = \texttt{FXY} \, \boldsymbol{D}_n'$ so

$$\texttt{dfdy} = \operatorname{vec}(\texttt{DFDY}) = \operatorname{vec}(\boldsymbol{I}_m \, \texttt{FXY} \, \boldsymbol{D}_n') = (\boldsymbol{D}_n \otimes \boldsymbol{I}_m) \operatorname{vec}(\texttt{FXY}) = (\boldsymbol{D}_n \otimes \boldsymbol{I}_m) \, \texttt{fxy}.$$

Stacking up yields $\begin{bmatrix} \texttt{dfdx} \\ \texttt{dfdy} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_m \\ \boldsymbol{D}_n \otimes \boldsymbol{I}_m \end{bmatrix} \texttt{fxy}$.
Harder brute-force way. The $\operatorname{vec}(\cdot)$ versions are:

$$\texttt{fxy} = \left[\begin{array}{c} f(x_1,y_1) \\ \vdots \\ f(x_{m-1},y_1) \\ f(x_m,y_1) \\ \hdashline \vdots \\ \hdashline f(x_1,y_n) \\ \vdots \\ f(x_{m-1},y_n) \\ f(x_m,y_n) \end{array}\right], \quad \texttt{dfdx} = \left[\begin{array}{c} f(x_2,y_1)-f(x_1,y_1) \\ \vdots \\ f(x_m,y_1)-f(x_{m-1},y_1) \\ f(x_1,y_1)-f(x_m,y_1) \\ \hdashline \vdots \\ \hdashline f(x_2,y_n)-f(x_1,y_n) \\ \vdots \\ f(x_m,y_n)-f(x_{m-1},y_n) \\ f(x_1,y_n)-f(x_m,y_n) \end{array}\right], \quad \texttt{dfdy} = \left[\begin{array}{c} f(x_1,y_2)-f(x_1,y_1) \\ \vdots \\ f(x_m,y_2)-f(x_m,y_1) \\ \hdashline \vdots \\ \hdashline f(x_1,y_n)-f(x_1,y_{n-1}) \\ \vdots \\ f(x_m,y_n)-f(x_m,y_{n-1}) \\ \hdashline f(x_1,y_1)-f(x_1,y_n) \\ \vdots \\ f(x_m,y_1)-f(x_m,y_n) \end{array}\right].$$

Thus:

$$
\begin{bmatrix} \texttt{dfdx} \\ \cdots\cdots \\ \texttt{dfdy} \end{bmatrix}
=
\begin{bmatrix}
f(x_2, y_1) - f(x_1, y_1) \\
\vdots \\
f(x_1, y_1) - f(x_m, y_1) \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
\vdots \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
f(x_1, y_n) - f(x_m, y_n) \\
\vdots \\
f(x_1, y_n) - f(x_m, y_n) \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
f(x_1, y_2) - f(x_1, y_1) \\
\vdots \\
f(x_m, y_2) - f(x_m, y_1) \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
\vdots \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
f(x_1, y_1) - f(x_1, y_n) \\
\vdots \\
f(x_m, y_1) - f(x_m, y_n)
\end{bmatrix}
= \boldsymbol{A}
\begin{bmatrix}
f(x_1, y_1) \\
\vdots \\
f(x_m, y_1) \\
\cdots\cdots \\
\vdots \\
\cdots\cdots \\
f(x_1, y_n) \\
\vdots \\
f(x_m, y_n)
\end{bmatrix},
$$

where

$$
\boldsymbol{A} =
\left[\begin{array}{cccccccccccccccc}
-1 & 1 & & & & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & -1 & 1 & & & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & \ddots & \ddots & & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & & -1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
1 & & & & -1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & & & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & & -1 & 1 & & & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & & & \ddots & \ddots & & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & & & & -1 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & & & & -1 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & & & \ddots & \ddots & & & & & \cdots & -1 & 1 & & & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & \cdots & & -1 & 1 & & \\
0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & \cdots & & & \ddots & \ddots & \\
0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & \cdots & & & & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & \cdots & 1 & & & & -1 \\
\hdashline
-1 & & \cdots & & 1 & & & & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & -1 & & \cdots & & 1 & & & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & \ddots & & & & \ddots & & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & & -1 & & \cdots & & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
 & & & & \ddots & & & & & & \ddots & & & & & \\
0 & 0 & 0 & 0 & 0 & -1 & & & & & \cdots & & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & & \ddots & & & & \ddots & & & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & & & -1 & \ddots & & \cdots & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & & & & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 \\
 & & \ddots & & \ddots & & & & & & \ddots & & & & & 0 \\
1 & & & & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & & -1 & 0 & & \\
 & 1 & & & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & & & -1 & & \\
 & & \ddots & & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & & & & \ddots & \\
 & & & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & & & & & -1
\end{array}\right].
$$

Using the definition of $\boldsymbol{D}_n$ given in the problem statement and Kronecker product, we write this as

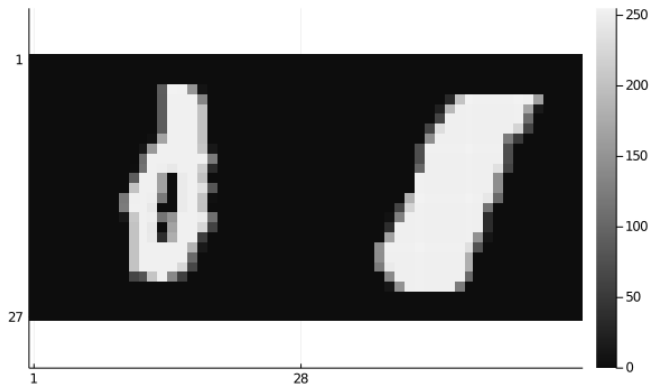
$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_m \\ \boldsymbol{D}_n \otimes \boldsymbol{I}_m \end{bmatrix}.$$

A possible `Julia` implementation is

```
using LinearAlgebra: I
using SparseArrays: sparse, spdiagm

"""
    A = first_diffs_2d_matrix(m, n)

In:
- `m` and `n` are positive integers

Out:
- `A` is a `2mn × mn` sparse matrix such that `A * vec(X)` computes the
first differences down the columns (along x direction)
and across the (along y direction) of the `m × n` matrix `X`.
"""
function first_diffs_2d_matrix(m, n)
    D = [kron(I(n), Dn(m)); # First differences down columns (x)
         kron(Dn(n), I(m))] # First differences across rows (y)
    return D
end


"""
    Dn = D(n)

- In: `n` is a positive integer
- Out: `Dn` is an `n x n` sparse circulant first differences matrix
"""
function Dn(n::Int)
    return spdiagm(0 => -ones(Int, n), 1 => ones(Int, n-1), 1-n => [1])
end
```

Pr. 11. (sol/hsjt1)
(Part 1)
Fraction 0 correct = 0.999
Fraction 1 correct = 0.999
Note that Julia will display 0.9988888888888889, but we should not report all of those digits because the are not statistically meaningful!
Graders: give full credit for the correct fraction (regardless of the number of decimal places given) and any comment acknowledging that this is a (perhaps surprisingly) good classification accuracy.

(Part 2)
Indexes of the misclassified digits: 43 and 673
Display of the misclassified digits:



Comment on why you think these digits were misclassified: The misclassified 0 is quite thin for a zero while the misclassified 1 is very wide. Neither of these are very clear digits!

(Part 3)
Percent 0 correct = 0.999
Percent 1 correct = 0.999
Comment about the classification accuracy: This is still very good! In fact, it is the same as the classification accuracy of the classifier using the angle between vectors. In general, this similarity may not hold, and either classifier may perform better on a given dataset.

Non-graded problem(s) below

Pr. 12. (sol/hs016)
Here $\boldsymbol{T}$ is an invertible (square) matrix and $\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$. Let λ be an eigenvalue of $\boldsymbol{B}$ associated with the eigenvector $\boldsymbol{v}$. By definition:

$$\boldsymbol{B}\boldsymbol{v} = \lambda\boldsymbol{v} \Rightarrow \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}\boldsymbol{v} = \lambda\boldsymbol{v} \Rightarrow \underbrace{\boldsymbol{T}\boldsymbol{T}^{-1}}_{=\boldsymbol{I}}\boldsymbol{A}\boldsymbol{T}\boldsymbol{v} = \lambda\boldsymbol{T}\boldsymbol{v} \Rightarrow \boldsymbol{A}(\boldsymbol{T}\boldsymbol{v}) = \lambda(\boldsymbol{T}\boldsymbol{v}).$$

Letting $\boldsymbol{u} \triangleq \boldsymbol{T}\boldsymbol{v}$, we have $\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u}$, so $\boldsymbol{u}/\|\boldsymbol{u}\|$ is a unit-norm eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ. Because this is true for *every* eigenvector $\boldsymbol{v}$ of $\boldsymbol{B}$ associated with *every* eigenvalue λ, this implies that the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$ are identical.
Thus for $i = 1, \ldots, n$, the unit-norm eigenvector $\boldsymbol{v}_i$ of $\boldsymbol{B}$ associated with the eigenvalue λ_i is related to the unit-norm eigenvector $\boldsymbol{u}_i$ of $\boldsymbol{A}$ associated with the same λ_i by :

$$\boldsymbol{v}_i = \frac{\boldsymbol{T}^{-1}\boldsymbol{u}_i}{||\boldsymbol{T}^{-1}\boldsymbol{u}_i||_2}.$$

When speaking of the *set* of eigenvectors associated with a particular eigenvalue it is important to normalize the eigenvectors to have unit-norm. If we did not normalize $\boldsymbol{v}_i$ as above then we would have to multiply the ith eigenvalue of $\boldsymbol{B}$ by $||\boldsymbol{T}^{-1}\boldsymbol{u}_i||_2^2$ so that $\boldsymbol{B} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^T$ as needed.

Pr. 13. (sol/hs018)
This is a simple application of SVD properties. Let $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$. Then $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\boldsymbol{V}' = \boldsymbol{0} \Rightarrow \boldsymbol{\Sigma}'\boldsymbol{\Sigma} = \boldsymbol{0}$, because $\boldsymbol{V}$ is unitary and hence invertible. The singular values of $\boldsymbol{X}$ are non-negative numbers, so the matrix $\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$ is a diagonal matrix with non-negative numbers equal to the singular values squared along the diagonal. Thus $\boldsymbol{\Sigma}'\boldsymbol{\Sigma} = \boldsymbol{0} \Rightarrow \boldsymbol{\Sigma} = 0 \Rightarrow \boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \boldsymbol{0}$.

For completeness, we can prove this particular result without using an SVD (but this proof is not accepted for credit for this HW problem). Because $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{0}$, we know that $\boldsymbol{x}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{x} = 0$ for any vector $\boldsymbol{x} \in \mathbb{F}^N$. Let $\boldsymbol{x} = \boldsymbol{e}_n$ where n is the nth unit vector, *i.e.*, $\boldsymbol{x}$ is all zeros except 1 in the nth element, for any $n \in \{1, \ldots, N\}$. Then $\boldsymbol{X}\boldsymbol{x} = \boldsymbol{X}\boldsymbol{e}_n = \boldsymbol{X}_{:,n}$, so $0 = \boldsymbol{x}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{x} = \boldsymbol{X}_{:,n}'\boldsymbol{X}_{:,n} = \sum_{m=1}^{M} |x_{mn}|^2$ and this can hold only if all of the elements of the nth column of $\boldsymbol{X}$ are zero because $|x_{mn}|^2 \geq 0$. By considering all values of n, we see that all of the columns of $\boldsymbol{X}$ must be $\boldsymbol{0}$, so $\boldsymbol{X} = \boldsymbol{0}$.

Pr. 14. (sol/hs117)

(a) If $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ and $\boldsymbol{A}$ has full rank, then $\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}' = \boldsymbol{V}\operatorname{Diag}(1/\sigma_n)\boldsymbol{U}'$, since $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}' = \boldsymbol{I}$.

(b) Technically $\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}'$ is not an SVD of $\boldsymbol{A}^{-1}$ in general, because the singular values of $\boldsymbol{A}^{-1}$ inside $\boldsymbol{\Sigma}^{-1}$ are not in descending order. To define a valid SVD for $\boldsymbol{A}^{-1}$, we need to first define a permutation matrix $\boldsymbol{P}$ like $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ that reverses the order, for which $\boldsymbol{P}^2 = \boldsymbol{I}$. Then we write a valid SVD like:

$$\boldsymbol{A}^{-1} = \underbrace{\boldsymbol{V}\boldsymbol{P}}_{\tilde{\boldsymbol{U}}}\underbrace{\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P}}_{\tilde{\boldsymbol{\Sigma}}}\underbrace{\boldsymbol{P}\boldsymbol{U}'}_{\tilde{\boldsymbol{V}}'}.$$

Pr. 15. (sol/hs132)

(a) (1) $\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$

(2) $(\boldsymbol{A}\boldsymbol{v}_i)'(\boldsymbol{A}\boldsymbol{v}_j) = (\sigma_i \boldsymbol{u}_i)'(\sigma_j \boldsymbol{u}_j) = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j \end{cases}$

(3) $\boldsymbol{A}'\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i$

(4) $(\boldsymbol{A}'\boldsymbol{u}_i)'(\boldsymbol{A}'\boldsymbol{u}_j) = (\sigma_i \boldsymbol{v}_i)'(\sigma_j \boldsymbol{v}_j) = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j \end{cases}$

(5) $\|\boldsymbol{A}\boldsymbol{v}_i\|_2^2 = (\boldsymbol{A}\boldsymbol{v}_i)'(\boldsymbol{A}\boldsymbol{v}_i) = \sigma_i^2$

(b) (1) $\boldsymbol{A}\boldsymbol{v}_i\boldsymbol{v}_i' = \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i'$

(2) $\boldsymbol{A}'\boldsymbol{u}_i\boldsymbol{u}_i' = \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i'$

(3) $\|\boldsymbol{A}\boldsymbol{v}_i\boldsymbol{v}_i'\|_F = \sigma_i$

(4) $\|\boldsymbol{A}'\boldsymbol{u}_i\boldsymbol{u}_i'\|_F = \sigma_i$

(5) $\|\boldsymbol{A}\boldsymbol{v}_i\boldsymbol{u}_j'\|_F = \|\sigma_i \boldsymbol{u}_i \boldsymbol{u}_j'\|_F = \sigma_i$

(6) $\|\boldsymbol{A}\boldsymbol{v}_i\boldsymbol{v}_j'\|_F = \|\sigma_i \boldsymbol{u}_i \boldsymbol{v}_j'\|_F = \sigma_i$

(7) $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\boldsymbol{U}'$

(8) $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\boldsymbol{V}'$

(9) $\|\boldsymbol{A}'\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^r \sigma_i^4}$

(10) $\|\boldsymbol{A}\boldsymbol{A}'\|_F = \sqrt{\sum_{i=1}^r \sigma_i^4}$

(c) (1) $\boldsymbol{U}[:, 1{:}k]'\boldsymbol{A} = \boldsymbol{\Sigma}[1{:}k, 1{:}k]\boldsymbol{V}[:, 1{:}k]'$

(2) $\boldsymbol{A}\boldsymbol{V}[:, 1{:}k] = \boldsymbol{U}[:, 1{:}k]\boldsymbol{\Sigma}[1{:}k, 1{:}k]$

(3) $\boldsymbol{U}[:, 1{:}k]\boldsymbol{U}[:, 1{:}k]'\boldsymbol{A} = \boldsymbol{U}[:, 1{:}k]\boldsymbol{\Sigma}[1{:}k, 1{:}k]\boldsymbol{V}[:, 1{:}k]'$

(4) $\boldsymbol{A}\boldsymbol{V}[:, 1{:}k]\boldsymbol{V}[:, 1{:}k]' = \boldsymbol{U}[:, 1{:}k]\boldsymbol{\Sigma}[1{:}k, 1{:}k]\boldsymbol{V}[:, 1{:}k]'$

(5) $\boldsymbol{U}[:, 1{:}k]'\boldsymbol{A}\boldsymbol{V}[:, 1{:}k] = \boldsymbol{\Sigma}[1{:}k, 1{:}k]$

Pr. 16. (sol/hsj14)

(a) A variation of "version 2" is to use the elements of $\boldsymbol{A}$ and the row partition of $\boldsymbol{B}$:

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} A_{1,1} & \dots & A_{1,K} \\ \vdots & & \\ A_{M,1} & \dots & A_{M,K} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{1,:} \\ \vdots \\ \boldsymbol{B}_{K,:} \end{bmatrix} \Rightarrow \boldsymbol{C}_{m,:} = \begin{bmatrix} A_{m,1} & \dots & A_{m,K} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{1,:} \\ \vdots \\ \boldsymbol{B}_{K,:} \end{bmatrix} = \sum_{k=1}^M A_{m,k} \boldsymbol{B}_{k,:}$$

(b) A variation of "version 3" is to use the row partition of $\boldsymbol{A}$ and the matrix $\boldsymbol{B}$:

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{1,:} \\ \vdots \\ \boldsymbol{A}_{M,:} \end{bmatrix} \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{1,:}\boldsymbol{B} \\ \vdots \\ \boldsymbol{A}_{M,:}\boldsymbol{B} \end{bmatrix} \Rightarrow \boldsymbol{C}_{m,:} = \boldsymbol{A}_{m,:}\boldsymbol{B}$$

These two variations are probably less useful than the versions given in the notes, because `MATLAB` and `Julia` store matrices by column, *i.e.*, with the row index varying fastest. So these two variations would access memory non-sequentially, thus probably running more slowly. But they are mathematically valid.
