Pr. 1. (sol/hs049)

- Approach 1. No SVD, just using properties of **pseudo-inverse**:

$$\begin{aligned} \boldsymbol{x}'(\boldsymbol{I}-\boldsymbol{A}^+\boldsymbol{A})'(\boldsymbol{A}^+\boldsymbol{b}) &= \boldsymbol{x}'(\boldsymbol{I}-(\boldsymbol{A}^+\boldsymbol{A})')(\boldsymbol{A}^+\boldsymbol{b}) = \boldsymbol{x}'(\boldsymbol{I}-(\boldsymbol{A}^+\boldsymbol{A}))(\boldsymbol{A}^+\boldsymbol{b}) \\ &= \boldsymbol{x}'(\boldsymbol{A}^+ - \boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+)\boldsymbol{b}) = \boldsymbol{x}'(\boldsymbol{A}^+ - \boldsymbol{A}^+)\boldsymbol{b}) = \boldsymbol{0} \end{aligned}$$

- Approach 2. full SVD:

$$\begin{aligned} (\boldsymbol{A}^\dagger\boldsymbol{b})'(\boldsymbol{I}-\boldsymbol{A}^\dagger\boldsymbol{A})\boldsymbol{x} &= \boldsymbol{b}'(\boldsymbol{A}^\dagger)'(\boldsymbol{I}-\boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}'\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}')\boldsymbol{x} = \boldsymbol{b}'\boldsymbol{U}(\boldsymbol{\Sigma}^\dagger)'\boldsymbol{V}'(\boldsymbol{V}\boldsymbol{V}'-\boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma}\boldsymbol{V}')\boldsymbol{x} \\ &= \boldsymbol{b}'\boldsymbol{U}((\boldsymbol{\Sigma}^\dagger)' - (\boldsymbol{\Sigma}^\dagger)'\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma})\boldsymbol{V}'\boldsymbol{x} = \boldsymbol{0}, \end{aligned}$$

  because direct multiplication verifies that $(\boldsymbol{\Sigma}^\dagger)' = (\boldsymbol{\Sigma}^\dagger)'\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma}$.

- Approach 3. **compact SVD**: $\boldsymbol{A} = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r' \implies \boldsymbol{A}^\dagger = \boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{U}_r'$ so:

$$(\boldsymbol{A}^\dagger\boldsymbol{b})'(\boldsymbol{I}-\boldsymbol{A}^\dagger\boldsymbol{A})\boldsymbol{x} = \boldsymbol{b}'\boldsymbol{U}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{V}_r'(\boldsymbol{I}-\boldsymbol{V}_r\boldsymbol{V}_r')\boldsymbol{x} = \boldsymbol{0}.$$

Pr. 2. (sol/hs051)

(a) Every point $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ on the plane satisfies the equation $\begin{bmatrix} a & b & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$. Thus every point on the plane must satisfy $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathcal{N}([a \quad b \quad c])$. Clearly: $\begin{bmatrix} a & b & c \end{bmatrix} = \underbrace{1}_{=\boldsymbol{u}_1}\underbrace{\sqrt{a^2+b^2+c^2}}_{=\sigma_1}\underbrace{\frac{\begin{bmatrix} a & b & c \end{bmatrix}}{\sqrt{a^2+b^2+c^2}}}_{=\boldsymbol{v}_1^T}$. Performing a full SVD of $\begin{bmatrix} a & b & c \end{bmatrix}$ will provide a 3×3 matrix $\boldsymbol{V}$ where $\mathcal{N}(\begin{bmatrix} a & b & c \end{bmatrix}) = \text{span}(\{\boldsymbol{v}_2, \boldsymbol{v}_3\})$ because we are considering a rank–1 matrix. Thus $\{\boldsymbol{v}_2, \boldsymbol{v}_3\}$ form an orthonormal basis for the plane. Two basis vectors are required to express every point on the plane.

(b) The nearest point on the plane is given by

$$P_{\mathcal{R}\{\boldsymbol{v}_2,\boldsymbol{v}_3\}}\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = (\boldsymbol{v}_2\boldsymbol{v}_2^T + \boldsymbol{v}_3\boldsymbol{v}_3^T)\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = (\boldsymbol{I} - \boldsymbol{v}_1\boldsymbol{v}_1^T)\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}.$$

(c) Code: `v1 = [1, 2, 3] / sqrt(14);` `(I - v1*v1') * [4, 5, 6]`
yields $(-1.714, 0.426, -0.857)$

Pr. 3. (sol/hs069)

Define: $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix}$, and $\boldsymbol{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

(a) The **linear least squares estimate** that minimizes $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ is given by $\widehat{\boldsymbol{x}} = \boldsymbol{A}^\dagger\boldsymbol{b}$. An SVD of $\boldsymbol{A}$ is simply:

$$\boldsymbol{A} = \sum_{i=1}^{2} 1\boldsymbol{q}_i\boldsymbol{e}_i^T,$$

where $\boldsymbol{e}_i$ denotes the ith unit vector Thus:

$$\boldsymbol{A}^\dagger = \sum_{i=1}^{2} 1\boldsymbol{e}_i\boldsymbol{q}_i^T,$$

and hence

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^\dagger\boldsymbol{b} = \sum_{i=1}^{2} 1\boldsymbol{e}_i\boldsymbol{q}_i^T\boldsymbol{b} = \boldsymbol{A}^T\boldsymbol{b}.$$

This solution is unsurprising, because what we get is precisely the first two “coordinates” of the vector $\boldsymbol{b}$ relative to the basis whose first two basis vectors correspond to the columns of $\boldsymbol{A}$.

(b) Here we have that the residual (or error) vector:

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T)\boldsymbol{b} = \sum_{i=3}^{n} \boldsymbol{q}_i \boldsymbol{q}_i^T \boldsymbol{b},$$

where $\boldsymbol{q}_i$ for $i = 3, \ldots, n$ are the $n-2$ (unit norm) basis vectors, orthogonal to $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ so that $\operatorname{span}(\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}) = \mathbb{R}^n$. Hence $\boldsymbol{q}_1^T \boldsymbol{r} = \sum_{i=3}^{n} \boldsymbol{q}_1^T \boldsymbol{q}_i \boldsymbol{b} = 0$ and $\boldsymbol{q}_2^T \boldsymbol{r} = \sum_{i=3}^{n} \boldsymbol{q}_2^T \boldsymbol{q}_i \boldsymbol{b} = 0$. This property is related to the **projection theorem**.

Pr. 4. (sol/hs072)
Here

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{=\boldsymbol{U}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}}_{=\boldsymbol{\Sigma}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^T}_{=\boldsymbol{V}^T},$$

since $\boldsymbol{A}$ is rank-1 and can be written as an outerproduct $\boldsymbol{A} = \boldsymbol{z}\boldsymbol{z}^T$ where $\boldsymbol{z} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and we know the eigen-decomposition of such rank-1 matrices from previous homeworks. Recall that since $\boldsymbol{A}$ is symmetric, positive-semidefinite, its eigen-decomposition is the same as a singular value decomposition.
Consider the minimum norm solution given by

$$\begin{aligned} \widehat{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}\boldsymbol{b} &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T \boldsymbol{b} \\ &= \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.75 \end{bmatrix}. \end{aligned}$$

Here $\operatorname{rank}(\boldsymbol{A}) = 1 < 2$, so the set of all solutions is $\hat{\boldsymbol{x}} + \mathcal{N}(\boldsymbol{A}) = \hat{\boldsymbol{x}} + \operatorname{span}(\boldsymbol{V}_0)$. Moreover

$$\ker(A) = \operatorname{span}(\begin{bmatrix} 1 \\ -1 \end{bmatrix}),$$

so that for all $\alpha \in \mathbb{R}$, vectors of the following form yield the same (minimum) squared error:

$$\widetilde{\boldsymbol{x}} = \widehat{\boldsymbol{x}} + \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Pr. 5. (sol/hs114)

(a) We must show $\forall \alpha \in [0, 1]$ that: $\|\boldsymbol{A}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} - \boldsymbol{b}\|_2 \leq \alpha \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2 + (1-\alpha)\|\boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}\|_2$.
For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$:

$$\begin{aligned} f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) &= \|\boldsymbol{A}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} - \boldsymbol{b}\|_2 \\ &= \|\boldsymbol{A}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} - \boldsymbol{b}(\alpha + 1 - \alpha)\|_2 \\ &= \|(\alpha\boldsymbol{A}\boldsymbol{x} - \alpha\boldsymbol{b}) + ((1-\alpha)\boldsymbol{A}\boldsymbol{y} - (1-\alpha)\boldsymbol{b})\|_2 \\ &\leq \|\alpha(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\|_2 + \|(1-\alpha)(\boldsymbol{A}\boldsymbol{y} - \boldsymbol{b})\|_2, \qquad \text{(via the triangle inequality)} \\ &= \alpha\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2 + (1-\alpha)\|\boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}\|_2, \qquad \text{(because } \alpha \in [0,1]) \\ &= \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}). \end{aligned}$$

(b) Next we need to show that, given two matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\forall \alpha \in [0, 1]$:

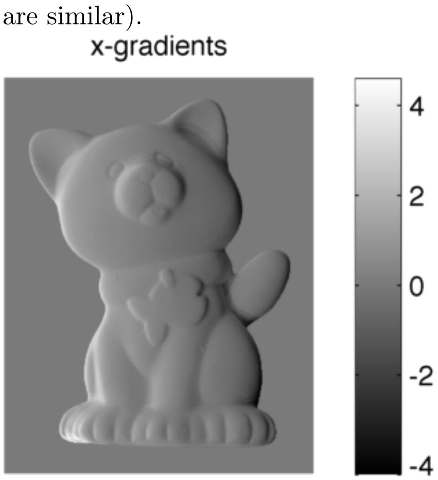
$$\sigma_1(\alpha\boldsymbol{A} + (1-\alpha)\boldsymbol{B}) \leq \alpha\,\sigma_1(\boldsymbol{A}) + (1-\alpha)\,\sigma_1(\boldsymbol{B}).$$

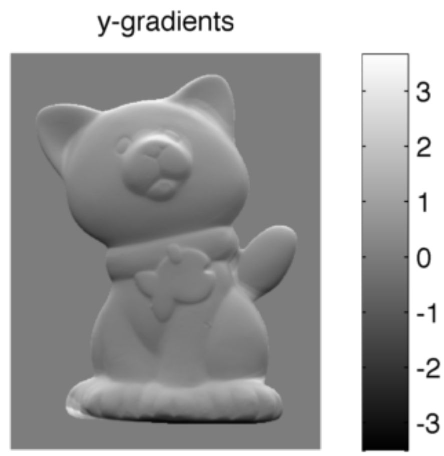
From the hint:

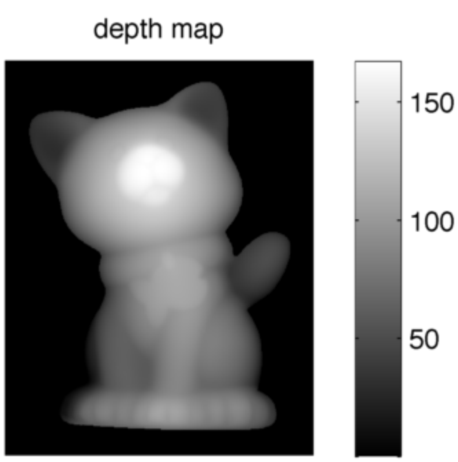
$$
\begin{aligned}
\sigma_1\left(\alpha \boldsymbol{A} + (1-\alpha)\boldsymbol{B}\right) &= \max_{\|\boldsymbol{u}\|_2=1} \|\left(\alpha \boldsymbol{A} + (1-\alpha)\boldsymbol{B}\right)\boldsymbol{u}\|_2 \\
&\le \max_{\|\boldsymbol{u}\|_2=1} \left(\|\alpha \boldsymbol{A}\boldsymbol{u}\|_2 + \|(1-\alpha)\boldsymbol{B}\boldsymbol{u}\|_2\right) \quad \text{(by the triangle inequality)} \\
&= \max_{\|\boldsymbol{u}\|_2=1} \left(\alpha\|\boldsymbol{A}\boldsymbol{u}\|_2 + (1-\alpha)\|\boldsymbol{B}\boldsymbol{u}\|_2\right) \\
&\le \max_{\|\boldsymbol{u}\|_2=1} \alpha\|\boldsymbol{A}\boldsymbol{u}\|_2 + \max_{\|\boldsymbol{u}\|_2=1} (1-\alpha)\|\boldsymbol{B}\boldsymbol{u}\|_2 \\
&= \alpha\,\sigma_1(\boldsymbol{A}) + (1-\alpha)\,\sigma_1(\boldsymbol{B}).
\end{aligned}
$$

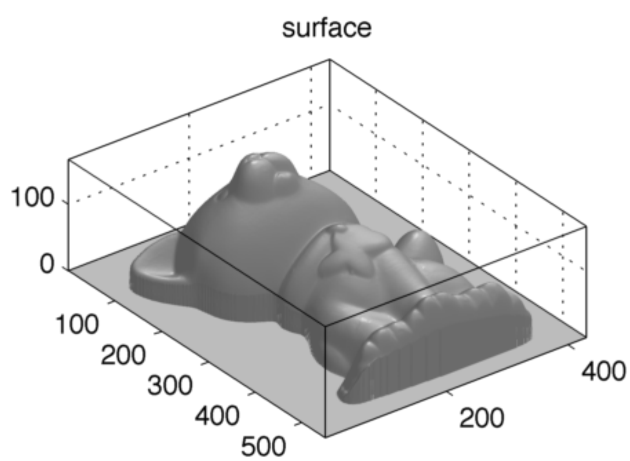
Thus, $\sigma_1(\cdot)$ is a convex function.

Pr. 6. (sol/hs075)
Here is the photometric stereo reconstruction of depth and surface (from an older version of `Julia`, but the results are similar).









Pr. 7. (sol/hs076)

(a) A possible `Julia` implementation is

```
function lsgd(A, b; mu=1/maximum(sum(abs.(A),dims=1))^2, x0=zeros(size(A,2)), nIters=200)
#
# Syntax:       x = lsgd(A, b, mu, x0, nIters)
#
# Inputs:       A is a m x n matrix
#
#               b is a vector of length m
#
#               mu is the step size to use, and must satisfy
#               0 < mu < 2 / sigma_1(A)^2 to guarantee convergence,
#               where sigma_1(A) is the first (largest) singular value.
#               The default value for mu will be explained in Ch.5.
#
#               x0 is the initial starting vector (of length n) to use.
#               Its default value is all zeros for simplicity.
#
#               nIters is the number of iterations to perform (default 200)
#
# Outputs:      x is a vector of length n containing the approximate solution
#
# Description:  Performs gradient descent to solve the least squares problem:
#               \argmin_x \| b - A x \|_2
#

    # Parse inputs
    b = vec(b)
    x0 = vec(x0)

    # Gradient descent
    x = x0
    for _ in 1:nIters
        x -= mu * (A' * (A * x - b))
    end

    return x
end
```

(b) Figure 1 shows $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$ versus iteration k for one realization of the system with step size $\mu = 1/\sigma_1^2(\boldsymbol{A})$, for four values of noise standard deviation σ. Clearly the $\boldsymbol{x}_k$ iterates are converging to $\hat{\boldsymbol{x}} = \boldsymbol{A}^+\boldsymbol{b}$.

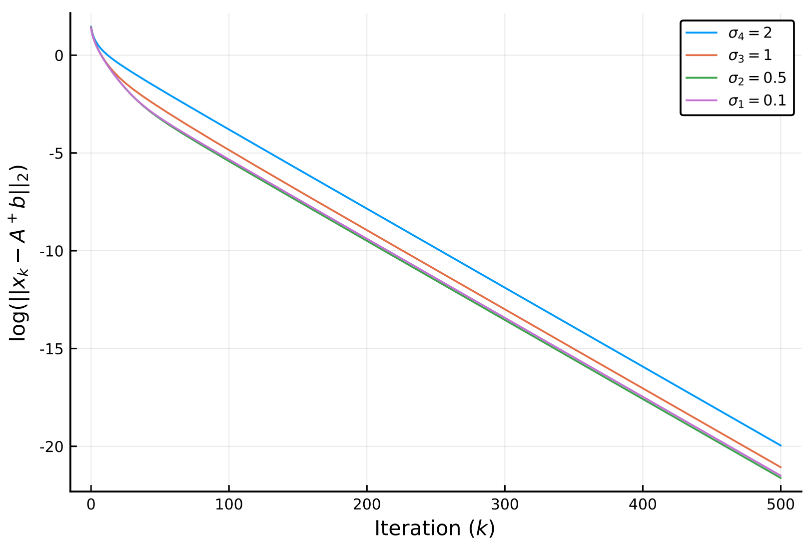


Figure 1: Convergence of gradient descent for least squares problems with different noise levels.

Optional problem(s) below

Pr. 8. (sol/hsj34)
Proof sketch for the case of distinct magnitudes, i.e., $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$ with $|\lambda_1| > \ldots > |\lambda_N|$.
If $\boldsymbol{x}$ is a RSV of $\boldsymbol{A}$ then $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \alpha\boldsymbol{x}$ so $|\boldsymbol{\Lambda}|^2 \boldsymbol{y} = \alpha\boldsymbol{y}$ where $\boldsymbol{y} = \boldsymbol{V}'\boldsymbol{x}$. Because of the distinct magnitudes we must have $\boldsymbol{y} = \gamma\boldsymbol{e}_n$ so $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{y} = \gamma\boldsymbol{v}_n$, an eigenvector of $\boldsymbol{A}$.
Now suppose that $\lambda_1 = \lambda_2$ and all other eigenvalues have distinct magnitudes. Then $|\boldsymbol{\Lambda}|^2 \boldsymbol{y} = \alpha\boldsymbol{y}$ implies that $\boldsymbol{y} = \gamma_1\boldsymbol{e}_1 + \gamma_2\boldsymbol{e}_2$ so $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{y} = \gamma_1\boldsymbol{v}_1 + \gamma_2\boldsymbol{v}_2$. Thus $\boldsymbol{A}\boldsymbol{x} = \gamma_1\boldsymbol{A}\boldsymbol{v}_1 + \gamma_2\boldsymbol{A}\boldsymbol{v}_2 = \gamma_1\lambda_1\boldsymbol{v}_1 + \gamma_2\lambda_2\boldsymbol{v}_2 = \lambda_1(\gamma_1\boldsymbol{v}_1 + \gamma_2\boldsymbol{v}_2) = \lambda_1\boldsymbol{x}$.
The idea generalizes to more repeated eigenvalues having the same value.

Pr. 9. (sol/hs014)

(a) $\tilde{x} = \beta_1/\beta_3$, $\tilde{y} = \beta_2/\beta_3$.

(b) From the first and third elements of vector $\boldsymbol{\beta}$, we get

$$\beta_3\tilde{x} = h_1^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T h_1 \tag{1}$$

$$\beta_3 = h_3^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T h_3 \tag{2}$$

from which we see that

$$\boldsymbol{\alpha}^T h_1 - \tilde{x}\boldsymbol{\alpha}^T h_3 = 0.$$

In matrix-vector form:

$$\begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{0}^T & -\tilde{x}\boldsymbol{\alpha}^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{0}^T & -\tilde{x}\boldsymbol{\alpha}^T \end{bmatrix} \operatorname{vec}(\boldsymbol{H}) = 0.$$

Following the same argument with the second and third elements of β, we get

$$\beta_3\tilde{y} = h_2^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T h_2 \tag{3}$$

$$\beta_3 = h_3^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T h_3$$

from which

$$\begin{bmatrix} \boldsymbol{0}^T & \boldsymbol{\alpha}^T & -\tilde{y}\boldsymbol{\alpha}^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 0$$

so that the required matrix $\boldsymbol{A}$ is

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{0}^T & -\tilde{x}\boldsymbol{\alpha}^T \\ \boldsymbol{0}^T & \boldsymbol{\alpha}^T & -\tilde{y}\boldsymbol{\alpha}^T \end{bmatrix}.$$

Alternatively, one could combine (1), (2) and (3) to see that

$$\boldsymbol{\alpha}^T h_1 + \boldsymbol{\alpha}^T h_2 - \boldsymbol{\alpha}^T(\tilde{x} + \tilde{y})h_3 = \beta_3\tilde{x} + \beta_3\tilde{y} - (\tilde{x} + \tilde{y})\beta_3 = 0,$$

so that

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{\alpha}^T & -(\tilde{x} + \tilde{y})\boldsymbol{\alpha}^T \\ \boldsymbol{\alpha}^T & \boldsymbol{\alpha}^T & -(\tilde{x} + \tilde{y})\boldsymbol{\alpha}^T \end{bmatrix}.$$

(c) Vector $\boldsymbol{h}$ is in the **null space** of $\boldsymbol{A}$, *i.e.*, $\boldsymbol{h} \in \mathcal{N}(\boldsymbol{A})$.

Pr. 10. (sol/hs122)

(a)

$$\begin{aligned}
&(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_y)(\boldsymbol{V}_x \otimes \boldsymbol{V}_y)' \\
=&(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_y)(\boldsymbol{V}_x' \otimes \boldsymbol{V}_y') \\
=&(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)((\boldsymbol{\Sigma}_x \boldsymbol{V}_x') \otimes (\boldsymbol{\Sigma}_y \boldsymbol{V}_y')) \\
=&(\boldsymbol{U}_x \boldsymbol{\Sigma}_x \boldsymbol{V}_x') \otimes (\boldsymbol{U}_y \boldsymbol{\Sigma}_y \boldsymbol{V}_y') = \boldsymbol{X} \otimes \boldsymbol{Y}.
\end{aligned}$$

To show that $\boldsymbol{U}_x \otimes \boldsymbol{U}_y$ is unitary, observe that

$$\begin{aligned}
&(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)'(\boldsymbol{U}_x \otimes \boldsymbol{U}_y) \\
=&(\boldsymbol{U}_x' \otimes \boldsymbol{U}_y')(\boldsymbol{U}_x \otimes \boldsymbol{U}_y) \\
=&(\boldsymbol{U}_x'\boldsymbol{U}_x) \otimes (\boldsymbol{U}_y'\boldsymbol{U}_y) = \boldsymbol{I}_m \otimes \boldsymbol{I}_p = \boldsymbol{I}_{mp}.
\end{aligned}$$

The above argument can be repeated to show that

- $(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)' = \boldsymbol{I}_{mp}$
- $(\boldsymbol{V}_x \otimes \boldsymbol{V}_y)'(\boldsymbol{V}_x \otimes \boldsymbol{V}_y) = \boldsymbol{I}_{nq}$
- $(\boldsymbol{V}_x \otimes \boldsymbol{V}_y)(\boldsymbol{V}_x \otimes \boldsymbol{V}_y)' = \boldsymbol{I}_{nq}$.

A technicality here is that $\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_y$ is a block rectangular diagonal matrix with rectangular diagonal blocks, which is not our usual arrangement of the "$\boldsymbol{\Sigma}$" for an SVD. To make a "proper" SVD we will need to introduce permutation matrices:

$$\underbrace{(\boldsymbol{U}_x \otimes \boldsymbol{U}_y)\boldsymbol{P}_1}_{\boldsymbol{U}} \underbrace{\boldsymbol{P}_1'(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_y)\boldsymbol{P}_2}_{\boldsymbol{\Sigma}} \underbrace{\boldsymbol{P}_2'(\boldsymbol{V}_x \otimes \boldsymbol{V}_y)'}_{\boldsymbol{V}'}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ permute things to put the singular values in descending order in the proper places.

(b)

$$\begin{aligned}
&(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a \otimes \boldsymbol{\Lambda}_b)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)' \\
=&(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a \otimes \boldsymbol{\Lambda}_b)(\boldsymbol{Q}_a' \otimes \boldsymbol{Q}_b') \\
=&(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a\boldsymbol{Q}_a' \otimes \boldsymbol{\Lambda}_b\boldsymbol{Q}_b') \\
=&\boldsymbol{Q}_a\boldsymbol{\Lambda}_a\boldsymbol{Q}_a' \otimes \boldsymbol{Q}_b\boldsymbol{\Lambda}_b\boldsymbol{Q}_b' = \boldsymbol{A} \otimes \boldsymbol{B}.
\end{aligned}$$

To show that $\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b$ is unitary, observe that

$$\begin{aligned}
&(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)'(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b) \\
=&(\boldsymbol{Q}_a' \otimes \boldsymbol{Q}_b')(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b) \\
=&(\boldsymbol{Q}_a'\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b'\boldsymbol{Q}_b) = \boldsymbol{I}_n \otimes \boldsymbol{I}_m = \boldsymbol{I}_{mn}.
\end{aligned}$$

The above argument can be repeated to show that $(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)' = \boldsymbol{I}_{mn}$.

(c)

$$\begin{aligned}
\boldsymbol{A} \oplus \boldsymbol{B} &= \boldsymbol{A} \otimes \boldsymbol{I}_m + \boldsymbol{I}_n \otimes \boldsymbol{B} \\
&= \boldsymbol{A} \otimes (\boldsymbol{Q}_b\boldsymbol{I}_m\boldsymbol{Q}_b') + (\boldsymbol{Q}_a\boldsymbol{I}_n\boldsymbol{Q}_a') \otimes \boldsymbol{B} \\
\text{by (b)} \rightarrow &= (\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a \otimes \boldsymbol{I}_m)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)' + (\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{I}_n \otimes \boldsymbol{\Lambda}_b)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)' \\
&= (\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a \otimes \boldsymbol{I}_m + \boldsymbol{I}_n \otimes \boldsymbol{\Lambda}_b)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)' \\
&= (\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)(\boldsymbol{\Lambda}_a \oplus \boldsymbol{\Lambda}_b)(\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b)'.
\end{aligned}$$

As in (b), $\boldsymbol{Q}_a \otimes \boldsymbol{Q}_b$ is unitary.