EE 150 - Applications of Convex Optimization in Signal Processing and Communications
Dr. Andre Tkacenko, JPL
Third Term 2011-2012

Homework Set #2 - Solutions

1. Recall that $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger$ and $\mathbf{B} = \mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger$.

1) $(\mathbf{A}\mathbf{A}^\#\mathbf{A} = \mathbf{A})$
Note that it is trivial to show that $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$. Using this property, we have

$$\mathbf{ABA} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)\left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right) = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma}\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger = \mathbf{A}\,.$$

2) $(\mathbf{A}^\#\mathbf{A}\mathbf{A}^\# = \mathbf{A}^\#)$
It is easy to show that $\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\# = \boldsymbol{\Sigma}^\#$. Using this, we get

$$\mathbf{BAB} = \left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)\left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right) = \mathbf{V}\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger = \mathbf{B}\,.$$

3) $(\left(\mathbf{A}\mathbf{A}^\#\right)^\dagger = \mathbf{A}\mathbf{A}^\#)$
First note that we have

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\# = \underbrace{\begin{bmatrix} \mathbf{I}_\rho & \mathbf{0}_{\rho\times(m-\rho)} \\ \mathbf{0}_{(m-\rho)\times\rho} & \mathbf{0}_{(m-\rho)\times(m-\rho)} \end{bmatrix}}_{\mathbf{C}},$$

where $\mathbf{C}$ is an $m \times m$ Hermitian matrix, i.e., $\mathbf{C} = \mathbf{C}^\dagger$. Using this result, we get

$$(\mathbf{AB})^\dagger = \left(\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)\left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)\right)^\dagger = \left(\mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)^\dagger = \left(\mathbf{U}\mathbf{C}\mathbf{U}^\dagger\right)^\dagger = \mathbf{U}\mathbf{C}^\dagger\mathbf{U}^\dagger = \mathbf{U}\mathbf{C}\mathbf{U}^\dagger\,. \tag{1}$$

Similarly, we have

$$\mathbf{AB} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)\left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right) = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger = \mathbf{U}\mathbf{C}\mathbf{U}^\dagger\,. \tag{2}$$

Comparing (1) with (2), we clearly have $(\mathbf{AB})^\dagger = \mathbf{AB}$.

4) As before, note first that we have

$$\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma} = \underbrace{\begin{bmatrix} \mathbf{I}_\rho & \mathbf{0}_{\rho\times(n-\rho)} \\ \mathbf{0}_{(n-\rho)\times\rho} & \mathbf{0}_{(n-\rho)\times(n-\rho)} \end{bmatrix}}_{\mathbf{D}},$$

where $\mathbf{D}$ is an $n \times n$ Hermitian matrix, i.e., $\mathbf{D} = \mathbf{D}^\dagger$. Using this result, we get

$$(\mathbf{BA})^\dagger = \left(\left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)\right)^\dagger = \left(\mathbf{V}\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma}\mathbf{V}^\dagger\right)^\dagger = \left(\mathbf{V}\mathbf{D}\mathbf{V}^\dagger\right)^\dagger = \mathbf{V}\mathbf{D}^\dagger\mathbf{V}^\dagger = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger\,. \tag{3}$$

Similarly, we have

$$\mathbf{BA} = \left(\mathbf{V}\boldsymbol{\Sigma}^\#\mathbf{U}^\dagger\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\right) = \mathbf{V}\boldsymbol{\Sigma}^\#\boldsymbol{\Sigma}\mathbf{V}^\dagger = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger\,. \tag{4}$$

Comparing (3) with (4), we clearly have $(\mathbf{BA})^\dagger = \mathbf{BA}$.

2. For simplicity, let us first prove that we always have $\mathbf{AB} = \mathbf{AC}$. To that end, we get

$$\mathbf{AB} = (\mathbf{AB})^\dagger = \mathbf{B}^\dagger\mathbf{A}^\dagger = \mathbf{B}^\dagger(\mathbf{ACA})^\dagger = \mathbf{B}^\dagger\mathbf{A}^\dagger\mathbf{C}^\dagger\mathbf{A}^\dagger = (\mathbf{AB})^\dagger(\mathbf{AC})^\dagger = \mathbf{ABAC} = \mathbf{AC},$$

where the first equality follows from property 3), the third equality follows from property 1), the sixth equality follows from property 3), and the seventh equality follows from property 1). Similarly, we can prove that we always have $\mathbf{BA} = \mathbf{CA}$. For this, we get

$$\mathbf{BA} = (\mathbf{BA})^\dagger = \mathbf{A}^\dagger\mathbf{B}^\dagger = (\mathbf{ACA})^\dagger\mathbf{B}^\dagger = \mathbf{A}^\dagger\mathbf{C}^\dagger\mathbf{A}^\dagger\mathbf{B}^\dagger = (\mathbf{CA})^\dagger(\mathbf{BA})^\dagger = \mathbf{CABA} = \mathbf{CA},$$

where the first equality follows from property 4), the third equality follows from property 1), the sixth equality follows from property 4), and the seventh equality follows from property 1). Now, by exploiting the facts that $\mathbf{AB} = \mathbf{AC}$ and $\mathbf{BA} = \mathbf{CA}$, we have the following:

$$\mathbf{B} = \mathbf{BAB} = \mathbf{BAC} = \mathbf{CAC} = \mathbf{C},$$

where the first equality follows from property 2), the second equality follows from the fact that $\mathbf{AB} = \mathbf{AC}$, the third equality follows from the fact that $\mathbf{BA} = \mathbf{CA}$, and the fourth equality follows from property 2).

3. (a) Note that we have

$$\begin{aligned}
\log(f(\boldsymbol{x}|\boldsymbol{\Sigma})) &= -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}), \\
&= -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\mathrm{tr}\Big((\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big), \\
&= -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\Big),
\end{aligned}$$

Thus, we get

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\Sigma}) &= -\frac{mn}{2} - \frac{n}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\left(\sum_{k=1}^{n}(\boldsymbol{x}_k-\boldsymbol{\mu})(\boldsymbol{x}_k-\boldsymbol{\mu})^T\right)\right), \\
&= -\frac{mn}{2} - \frac{n}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}), \qquad (5)
\end{aligned}$$

where we have defined $\mathbf{C}$ as follows:

$$\mathbf{C} \triangleq \sum_{k=1}^{n}(\boldsymbol{x}_k-\boldsymbol{\mu})(\boldsymbol{x}_k-\boldsymbol{\mu})^T. \qquad (6)$$

Note that $\mathbf{C}$ is symmetric, i.e., $\mathbf{C} = \mathbf{C}^T$. Now, from the hint, we have, for *symmetric* $\mathbf{X}$:

$$\begin{aligned}
\frac{d}{d\mathbf{X}}\log(\det(\mathbf{X})) &= \mathbf{X}^{-1} + (\mathbf{X}^{-1})^T - \mathrm{diag}(\mathbf{X}^{-1}) = 2\mathbf{X}^{-1} - \mathrm{diag}(\mathbf{X}^{-1}), \\
\frac{d}{d\mathbf{X}}\mathrm{tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) &= -\mathbf{X}^{-1}\mathbf{BA}\mathbf{X}^{-1} - (\mathbf{X}^{-1})^T\mathbf{A}^T\mathbf{B}^T(\mathbf{X}^{-1})^T - \mathrm{diag}(-\mathbf{X}^{-1}\mathbf{BA}\mathbf{X}^{-1}), \\
&= -\mathbf{X}^{-1}\mathbf{BA}\mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-1} + \mathrm{diag}(\mathbf{X}^{-1}\mathbf{BA}\mathbf{X}^{-1})
\end{aligned}$$

Hence, using $\mathbf{A} = \mathbf{I}_m$ and $\mathbf{B} = \mathbf{C}$, we get, from (5):

$$\begin{aligned}
\nabla \mathcal{L}(\boldsymbol{\Sigma}) &= \frac{d}{d\boldsymbol{\Sigma}^T} \mathcal{L}(\boldsymbol{\Sigma}) = \frac{d}{d\boldsymbol{\Sigma}} \mathcal{L}(\boldsymbol{\Sigma}) \;, \\
&= -\frac{n}{2} \left[2\boldsymbol{\Sigma}^{-1} - \mathrm{diag}\big(\boldsymbol{\Sigma}^{-1}\big) \right] \\
&\quad - \frac{1}{2} \left[-\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{C}^T\boldsymbol{\Sigma}^{-1} + \mathrm{diag}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1}\big) \right] \;, \\
&= -n\boldsymbol{\Sigma}^{-1} + \frac{n}{2}\mathrm{diag}\big(\boldsymbol{\Sigma}^{-1}\big) + \boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - \frac{1}{2}\mathrm{diag}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1}\big) \;, \\
&= \boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} - \frac{1}{2}\mathrm{diag}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1}\big) \;. \qquad (7)
\end{aligned}$$

Thus, from (6), we conclude the following:

$$\begin{aligned}
\nabla \mathcal{L}(\boldsymbol{\Sigma}) &= \boldsymbol{\Sigma}^{-1} \left[\sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^T \right] \boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} \\
&\quad - \frac{1}{2}\mathrm{diag}\left(\boldsymbol{\Sigma}^{-1} \left[\sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^T \right] \boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} \right) \;.
\end{aligned}$$

(b) From (7), setting $\nabla \mathcal{L}(\boldsymbol{\Sigma}) = \mathbf{0}$ yields the following conditions:

$$\begin{cases} \boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} = \mathbf{0}\,, & \text{for off-diagonal entries} \\ \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1}\right) = \mathbf{0}\,, & \text{for on-diagonal entries} \end{cases} .$$

Note that both cases are consistent here and lead to the following for all matrix entries:

$$\boldsymbol{\Sigma}^{-1}\mathbf{C}\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} = \mathbf{0}\,.$$

Pre/post-multiplying both sides by $\boldsymbol{\Sigma}$ yields

$$\mathbf{C} - n\boldsymbol{\Sigma} = \mathbf{0} \Longleftrightarrow \boldsymbol{\Sigma} = \frac{1}{n}\mathbf{C}\,.$$

Thus, from (6), we get

$$\boldsymbol{\Sigma}_{\mathrm{ML}} = \frac{1}{n} \sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^T \;,$$

as desired.

4. Using property 2) and the product rule, we have

$$d\Big(\mathbf{Z}^{\#}\Big) = d\Big(\mathbf{Z}^{\#}\mathbf{Z}\mathbf{Z}^{\#}\Big) = d\Big(\mathbf{Z}^{\#}\left(\mathbf{Z}\mathbf{Z}^{\#}\right)\Big) = \Big(d\Big(\mathbf{Z}^{\#}\Big)\Big)\mathbf{Z}\mathbf{Z}^{\#} + \mathbf{Z}^{\#}\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big) \;. \qquad (8)$$

Now consider $d\big(\mathbf{Z}^{\#}\mathbf{Z}\big)$. Using the product rule, we have

$$d\Big(\mathbf{Z}^{\#}\mathbf{Z}\Big) = \Big(d\Big(\mathbf{Z}^{\#}\Big)\Big)\mathbf{Z} + \mathbf{Z}^{\#}(d\mathbf{Z}) \Longleftrightarrow \Big(d\Big(\mathbf{Z}^{\#}\Big)\Big)\mathbf{Z} = d\Big(\mathbf{Z}^{\#}\mathbf{Z}\Big) - \mathbf{Z}^{\#}(d\mathbf{Z}) \;.$$

Substituting this into (8) yields

$$d\Big(\mathbf{Z}^{\#}\Big) = -\mathbf{Z}^{\#}(d\mathbf{Z})\mathbf{Z}^{\#} + \mathbf{Z}^{\#}\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big) + \Big(d\Big(\mathbf{Z}^{\#}\mathbf{Z}\Big)\Big)\mathbf{Z}^{\#} \;. \qquad (9)$$

As can be seen from (9), it remains to express $d\big(\mathbf{Z}\mathbf{Z}^{\#}\big)$ and $d\big(\mathbf{Z}^{\#}\mathbf{Z}\big)$ in terms of $d\mathbf{Z}$ and $d\mathbf{Z}^{\dagger}$. Focusing first on $d\big(\mathbf{Z}\mathbf{Z}^{\#}\big)$, we have, from the property 1), the product rule, property 3), and the conjugate transpose rule:

$$\begin{aligned} d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) &= d\Big(\mathbf{Z}\mathbf{Z}^{\#}\mathbf{Z}\mathbf{Z}^{\#}\Big) = d\Big(\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big) \,, \\ &= \Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) + \Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big) \,, \\ &= \Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) + \Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)^{\dagger}\Big(d\Big(\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)^{\dagger}\Big)\Big) \,, \\ &= \Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) + \Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)^{\dagger}\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)^{\dagger} \,, \\ &= \Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) + \Big(\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)^{\dagger} \,. \end{aligned} \tag{10}$$

Now consider an alternate expression for $d\mathbf{Z}$. From property 1) and the product rule, we get

$$d\mathbf{Z} = d\Big(\mathbf{Z}\mathbf{Z}^{\#}\mathbf{Z}\Big) = d\Big(\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\mathbf{Z}\Big) = \Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\mathbf{Z} + \Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z}) \,.$$

Rearranging terms yields

$$\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big)\mathbf{Z} = d\mathbf{Z} - \Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z}) = \Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z}) \,.$$

Substituting this expression into (10) yields the following upon using the conjugate transpose rule and property 3):

$$\begin{aligned} d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big) &= \Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z})\Big(\mathbf{Z}^{\#}\Big) + \Big(\Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z})\Big(\mathbf{Z}^{\#}\Big)\Big)^{\dagger} \,, \\ &= \Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z})\Big(\mathbf{Z}^{\#}\Big) + \Big(\mathbf{Z}^{\#}\Big)^{\dagger}(d\mathbf{Z})^{\dagger}\Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)^{\dagger} \,, \\ &= \Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big)(d\mathbf{Z})\Big(\mathbf{Z}^{\#}\Big) + \Big(\mathbf{Z}^{\#}\Big)^{\dagger}\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big) \,. \end{aligned} \tag{11}$$

Multiplying the above expression by $\mathbf{Z}^{\#}$ on the left yields the following upon invoking property 2):

$$\mathbf{Z}^{\#}\Big(d\Big(\mathbf{Z}\mathbf{Z}^{\#}\Big)\Big) = \mathbf{Z}^{\#}\Big(\mathbf{Z}^{\#}\Big)^{\dagger}\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big) \,. \tag{12}$$

Using a similar set of arguments (which include using property 4) in this case), we can show that $d\big(\mathbf{Z}^{\#}\mathbf{Z}\big)$ is given by

$$d\Big(\mathbf{Z}^{\#}\mathbf{Z}\Big) = \Big(\mathbf{I}_n - \mathbf{Z}^{\#}\mathbf{Z}\Big)\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{Z}^{\#}\Big)^{\dagger} + \Big(\mathbf{Z}^{\#}\Big)(d\mathbf{Z})\Big(\mathbf{I}_n - \mathbf{Z}^{\#}\mathbf{Z}\Big) \,.$$

Multiplying the above expression by $\mathbf{Z}^{\#}$ on the right yields the following upon invoking property 2):

$$\Big(d\Big(\mathbf{Z}^{\#}\mathbf{Z}\Big)\Big)\mathbf{Z}^{\#} = \Big(\mathbf{I}_n - \mathbf{Z}^{\#}\mathbf{Z}\Big)\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{Z}^{\#}\Big)^{\dagger}\mathbf{Z}^{\#} \,. \tag{13}$$

Substituting (12) and (13) in (9) yields

$$d\Big(\mathbf{Z}^{\#}\Big) = -\mathbf{Z}^{\#}(d\mathbf{Z})\mathbf{Z}^{\#} + \mathbf{Z}^{\#}\Big(\mathbf{Z}^{\#}\Big)^{\dagger}\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\Big) + \Big(\mathbf{I}_n - \mathbf{Z}^{\#}\mathbf{Z}\Big)\Big(d\mathbf{Z}^{\dagger}\Big)\Big(\mathbf{Z}^{\#}\Big)^{\dagger}\mathbf{Z}^{\#} \,,$$

as desired.

***5.** (a) First note that optimizing the Frobenius norm of a matrix is equivalent to optimizing the square of the Frobenius norm. So, to that end, consider

$$\begin{aligned}\xi^2 &= ||\mathbf{AX}-\mathbf{B}||_F^2 = \mathrm{tr}\Big((\mathbf{AX}-\mathbf{B})^\dagger(\mathbf{AX}-\mathbf{B})\Big) \,, \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AX} + \mathbf{B}^\dagger\mathbf{B}\Big) \,. \qquad (14)\end{aligned}$$

We will show that ξ^2 can be expressed in the following form, given by completing the square:

$$\xi^2 = \mathrm{tr}\Big(\Big[\mathbf{A}\left(\mathbf{X}-\mathbf{A}^\#\mathbf{B}\right)\Big]^\dagger\Big[\mathbf{A}\left(\mathbf{X}-\mathbf{A}^\#\mathbf{B}\right)\Big]\Big) + \mathrm{tr}\Big(\mathbf{B}^\dagger\left(\mathbf{I}_m - \mathbf{AA}^\#\right)\mathbf{B}\Big) \,. \qquad (15)$$

To do so, let us define the right-hand side of the above quantity to be $\widehat{\xi^2}$. Expanding the terms in (15) yields

$$\begin{aligned}\widehat{\xi^2} &= \mathrm{tr}\Big(\Big(\mathbf{X}^\dagger - \mathbf{B}^\dagger\left(\mathbf{A}^\#\right)^\dagger\Big)\mathbf{A}^\dagger\mathbf{A}\left(\mathbf{X}-\mathbf{A}^\#\mathbf{B}\right)\Big) + \mathrm{tr}\Big(\mathbf{B}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B}\Big) \,, \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AA}^\#\mathbf{B} - \mathbf{B}^\dagger\left(\mathbf{A}^\#\right)^\dagger\mathbf{A}^\dagger\mathbf{AX} + \mathbf{B}^\dagger\left(\mathbf{A}^\#\right)^\dagger\mathbf{A}^\dagger\mathbf{AA}^\#\mathbf{B} \\ &\quad + \mathbf{B}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B}\Big) \,, \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\mathbf{A}^\dagger\left(\mathbf{AA}^\#\right)^\dagger\mathbf{B} - \mathbf{B}^\dagger\left(\mathbf{AA}^\#\right)^\dagger\mathbf{AX} \\ &\quad + \mathbf{B}^\dagger\left(\mathbf{AA}^\#\right)^\dagger\mathbf{AA}^\#\mathbf{B} + \mathbf{B}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B}\Big) \,, \qquad (16) \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\left(\mathbf{AA}^\#\mathbf{A}\right)^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{AX} \\ &\quad + \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{AA}^\#\mathbf{B} + \mathbf{B}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B}\Big) \,, \qquad (17) \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AX} + \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B} + \mathbf{B}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AA}^\#\mathbf{B}\Big) \,, \qquad (18) \\ &= \mathrm{tr}\Big(\mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{AX} - \mathbf{X}^\dagger\mathbf{A}^\dagger\mathbf{B} - \mathbf{B}^\dagger\mathbf{AX} + \mathbf{B}^\dagger\mathbf{B}\Big) \,. \qquad (19)\end{aligned}$$

With reference to the four defining properties of a pseudoinverse, we have that (16) and (17) follow from property 3), while (18) follows from property 1). Comparing (19) with (14), it is clear that $\widehat{\xi^2} = \xi^2$, and so (15) is valid. Prior to deriving a least-squares solution, it is useful to simplify (15) further. For this, note that we have

$$\begin{aligned}\left(\mathbf{I}_m - \mathbf{AA}^\#\right)^\dagger\left(\mathbf{I}_m - \mathbf{AA}^\#\right) &= \left(\mathbf{I}_m - \mathbf{AA}^\#\right)\left(\mathbf{I}_m - \mathbf{AA}^\#\right) \,, \\ &= \mathbf{I}_m - 2\mathbf{AA}^\# + \mathbf{AA}^\#\mathbf{AA}^\# \,, \\ &= \mathbf{I}_m - 2\mathbf{AA}^\# + \mathbf{AA}^\# = \left(\mathbf{I}_m - \mathbf{AA}^\#\right) \,. \qquad (20)\end{aligned}$$

Here, the first equality follows from property 3) and the third equality follows from property 1). Substituting (20) into (15) yields the following:

$$\begin{aligned} \xi^2 &= \operatorname{tr}\left(\left[\mathbf{A}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right)\right]^{\dagger}\left[\mathbf{A}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right)\right]\right) \\ &\quad + \operatorname{tr}\left(\mathbf{B}^{\dagger}\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^{\#}\right)\mathbf{B}\right), \\ &= \left|\left|\mathbf{A}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right)\right|\right|_F^2 + \left|\left|\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^{\#}\right)\mathbf{B}\right|\right|_F^2 \end{aligned} \tag{21}$$

To prove that $\mathbf{X}^{\star} \triangleq \mathbf{A}^{\#}\mathbf{B}$ is a least-squares solution, note that from (21) that $\mathbf{X}$ only appears in the first squared Frobenius norm term. Furthermore, this term is greater than or equal to zero, with equality if and only if the argument is identically zero, which occurs if $\mathbf{X} = \mathbf{A}^{\#}\mathbf{B}$. Hence, $\mathbf{X}^{\star} = \mathbf{A}^{\#}\mathbf{B}$ is a least-squares solution. With this choice of $\mathbf{X}$, from (21), the square of the optimal objective function becomes

$$(\xi^{\star})^2 = \left|\left|\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^{\#}\right)\mathbf{B}\right|\right|_F^2 ,$$

from which we conclude

$$\xi^{\star} = \left|\left|\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^{\#}\right)\mathbf{B}\right|\right|_F .$$

(b) As with the previous part, it will be convenient to work with the squared Frobenius norm. We have the following:

$$\begin{aligned} ||\mathbf{X}||_F^2 &= ||(\mathbf{X}-\mathbf{X}^{\star})+\mathbf{X}^{\star}||_F^2 , \\ &= ||\mathbf{X}-\mathbf{X}^{\star}||_F^2 + ||\mathbf{X}^{\star}||_F^2 + \operatorname{tr}\left((\mathbf{X}-\mathbf{X}^{\star})^{\dagger}\mathbf{X}^{\star}\right) + \operatorname{tr}\left((\mathbf{X}^{\star})^{\dagger}(\mathbf{X}-\mathbf{X}^{\star})\right) . \end{aligned}$$

For simplicity, define $\mathbf{Y} \triangleq (\mathbf{X}^{\star})^{\dagger}(\mathbf{X}-\mathbf{X}^{\star})$, so that we have

$$||\mathbf{X}||_F^2 = ||\mathbf{X}-\mathbf{X}^{\star}||_F^2 + ||\mathbf{X}^{\star}||_F^2 + \operatorname{tr}\left(\mathbf{Y}^{\dagger}\right) + \operatorname{tr}(\mathbf{Y}) . \tag{22}$$

Now note that we have the following chain of equalities:

$$\begin{aligned} \mathbf{Y} &= \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right) = \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right), \\ &= \mathbf{B}^{\dagger}\left(\left(\mathbf{A}^{\#}\mathbf{A}\right)\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right) = \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{A}^{\#}\mathbf{A}\right)^{\dagger}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right), \\ &= \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\right)^{\dagger}\mathbf{A}^{\#}\mathbf{A}\left(\mathbf{X}-\mathbf{A}^{\#}\mathbf{B}\right) = \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{A}^{\#}\mathbf{A}\mathbf{X}-\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#}\mathbf{B}\right), \\ &= \mathbf{B}^{\dagger}\left(\mathbf{A}^{\#}\right)^{\dagger}\left(\mathbf{A}^{\#}\mathbf{B}-\mathbf{A}^{\#}\mathbf{B}\right) = \mathbf{0}. \end{aligned}$$

Here, the second equality follows from property 2), the fifth equality follows from property 4), and the seventh equality follows from the fact that $\mathbf{AX} = \mathbf{B}$ as well as property 2). Thus, $\mathbf{Y} = \mathbf{0}$ and so from (22), we get

$$||\mathbf{X}||_F^2 = ||\mathbf{X}-\mathbf{X}^{\star}||_F^2 + ||\mathbf{X}^{\star}||_F^2 .$$

Note that the first squared Frobenius norm term on the right-hand side is greater than or equal to zero with equality if and only if $\mathbf{X} = \mathbf{X}^{\star}$. Thus, we have

$$||\mathbf{X}||_F^2 \geq ||\mathbf{X}^{\star}||_F^2 ,$$

with equality if and only if $\mathbf{X} = \mathbf{X}^\star$. Upon taking square roots, we have, equivalently,

$$||\mathbf{X}||_F \geq ||\mathbf{X}^\star||_F \ ,$$

with equality if and only if $\mathbf{X} = \mathbf{X}^\star$.