Solution to Exercise 2.19

⋆2.19 Show that $\boldsymbol{P_X} - \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$, where $\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ is the projection on to the span of $\boldsymbol{M}_1\boldsymbol{X}_2$. This can be done most easily by showing that any vector in $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$ is invariant under the action of $\boldsymbol{P_X} - \boldsymbol{P}_1$, and that any vector orthogonal to this span is annihilated by $\boldsymbol{P_X} - \boldsymbol{P}_1$.

Any vector in $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$ can be expressed as $\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma}$ for some k_2–vector $\boldsymbol{\gamma}$. If we premultiply such a vector by $\boldsymbol{P_X} - \boldsymbol{P}_1$, we find that

$$
\begin{aligned}
(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma} &= \boldsymbol{P_X}\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma} = \boldsymbol{P_X}(\mathbf{I} - \boldsymbol{P}_1)\boldsymbol{X}_2\boldsymbol{\gamma} \\
&= (\boldsymbol{X}_2 - \boldsymbol{P}_1\boldsymbol{X}_2)\boldsymbol{\gamma} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma}.
\end{aligned}
$$

The first equality follows from the fact that, since $\boldsymbol{P}_1$ and $\boldsymbol{M}_1$ are complementary projections, $\boldsymbol{P}_1\boldsymbol{M}_1 = \mathbf{O}$. The second and fourth express the definition of $\boldsymbol{M}_1$. The third uses (2.35) and the fact that $\boldsymbol{P_X}\boldsymbol{X}_2 = \boldsymbol{X}_2$, because each column of $\boldsymbol{X}_2$ is also a column of $\boldsymbol{X}$, and so belongs to $\mathcal{S}(\boldsymbol{X})$, and is thus invariant under $\boldsymbol{P_X}$. This gives the result that any vector in $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$ is invariant under the action of $\boldsymbol{P_X} - \boldsymbol{P}_1$.

Consider now a vector $\boldsymbol{z}$ that is orthogonal to $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, so that $\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{z} = \mathbf{0}$. We wish to show that $(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{z} = \mathbf{0}$, or, equivalently, that

$$\boldsymbol{P_X}\boldsymbol{z} = \boldsymbol{P}_1\boldsymbol{z}. \tag{S2.09}$$

If we premultiply this equation by $\boldsymbol{X}^\top$, we obtain the equation

$$\boldsymbol{X}^\top\boldsymbol{P_X}\boldsymbol{z} = \boldsymbol{X}^\top\boldsymbol{P}_1\boldsymbol{z}. \tag{S2.10}$$

It is clear that (S2.09) implies (S2.10). In the other direction, on premultiplying (S2.10) by $\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ we obtain

$$\boldsymbol{P_X}\boldsymbol{P_X}\boldsymbol{z} = \boldsymbol{P_X}\boldsymbol{P}_1\boldsymbol{z},$$

which is just equation (S2.09). This follows from the facts that $\boldsymbol{P_X}$ is idempotent and that $\boldsymbol{P_X}\boldsymbol{P}_1 = \boldsymbol{P}_1$ by equation (2.35).

In order to show (S2.10), note that

$$\boldsymbol{X}^\top\boldsymbol{M}_1\boldsymbol{z} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \end{bmatrix}\boldsymbol{M}_1\boldsymbol{z} = \begin{bmatrix} \boldsymbol{X}_1^\top\boldsymbol{M}_1\boldsymbol{z} \\ \boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{z} \end{bmatrix} = \mathbf{0}.$$

The upper block vanishes because $\boldsymbol{M}_1$ annihilates $\boldsymbol{X}_1$, and the lower block vanishes because we assumed that $\boldsymbol{z}$ is orthogonal to $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$. Therefore,

$$\mathbf{0} = \boldsymbol{X}^\top\boldsymbol{M}_1\boldsymbol{z} = \boldsymbol{X}^\top(\mathbf{I} - \boldsymbol{P}_1)\boldsymbol{z},$$

which implies (S2.10) because $\boldsymbol{X}^\top \boldsymbol{P_X} = \boldsymbol{X}$. This completes the proof.

The last part of the demonstration could be made simpler by using the result that the trace of an orthogonal projection equals the dimension of its image. The traces we need are

$$\mathrm{Tr}(\boldsymbol{P_X} - \boldsymbol{P}_1) = k - k_1 = k_2, \quad \text{and} \quad \mathrm{Tr}(\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}) = k_2,$$

where the second result follows because $\boldsymbol{M}_1\boldsymbol{X}_2$ has k_2 linearly independent columns. Since we have shown that $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, which is the image of $\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$, is contained in the image of $\boldsymbol{P_X} - \boldsymbol{P}_1$, and that the dimensions of the two images are the same, the two images must coincide.

Alternative proof:

There is another way to prove this result, which is based on a rather different approach from the one suggested in the exercise. Instead of trying to prove directly that $\boldsymbol{P_X} - \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$, we prove the equivalent proposition that $\boldsymbol{P_X} = \boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$. To do this, we must show two things. First, for any vector $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{X})$, $(\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2})\boldsymbol{x} = \boldsymbol{x}$. Second, for any vector $\boldsymbol{W} \in \mathcal{S}^\perp(\boldsymbol{X})$, $(\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2})\boldsymbol{w} = \boldsymbol{0}$. These two results imply that the projection matrices $\boldsymbol{P_X}$ and $\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ are the same.

Since $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{X})$, we can write $\boldsymbol{x} = \boldsymbol{X\gamma} = \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}_2\boldsymbol{\gamma}_2$ for some vector $\boldsymbol{\gamma}$ and subvectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. Thus we have

$$\begin{aligned}(\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2})\boldsymbol{x} &= (\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2})(\boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}_2\boldsymbol{\gamma}_2)\\ &= \boldsymbol{P}_1(\boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}_2\boldsymbol{\gamma}_2) + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}(\boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}_2\boldsymbol{\gamma}_2)\\ &= \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{\gamma}_2 + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma}_2\\ &= \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}_2\boldsymbol{\gamma}_2 = \boldsymbol{x}\end{aligned}$$

The crucial step here is the one from the second line to the third. It follows because of the two relations

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{X}_1 &= \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_1 = \mathbf{O}, \text{ and}\\ \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{X}_2 &= \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{M}_1\boldsymbol{X}_2.\end{aligned}$$

We have now proved the first of the two results.

The second result is easier. We have

$$\begin{aligned}(\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2})\boldsymbol{w} &= \boldsymbol{P}_1\boldsymbol{w} + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{w}\\ &= \boldsymbol{0} + \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{w}\\ &= \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top(\mathbf{I} - \boldsymbol{P}_1)\boldsymbol{w} = \boldsymbol{0}.\end{aligned}$$

Here $\boldsymbol{P}_1\boldsymbol{w} = \boldsymbol{0}$ because $\boldsymbol{w}$, being orthogonal to $\boldsymbol{X}$, is orthogonal to $\boldsymbol{X}_1$ and so is annihilated by $\boldsymbol{P}_1$. Since $\boldsymbol{w}$ is also orthogonal to $\boldsymbol{X}_2$, we see that $\boldsymbol{X}_2^\top\boldsymbol{w} = \boldsymbol{0}$.

We have shown that the projection $\boldsymbol{P}_1 + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$ has exactly the same properties as the projection $\boldsymbol{P_X}$. It annihilates any vector in $\mathcal{S}^\perp(\boldsymbol{X})$, while any vector in $\mathcal{S}(\boldsymbol{X})$ is invariant to it.