A Appendix

Proof of Lemma 2.1

We have

$$\begin{aligned}\boldsymbol{M}\boldsymbol{M} &= \boldsymbol{I}_{m+p} - 2\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\ &\qquad + \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\ &= \boldsymbol{I}_{m+p} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\ &= \boldsymbol{M},\end{aligned}$$

so $\boldsymbol{M}$ is a projection matrix. For any $\boldsymbol{z}$ such that $\boldsymbol{M}\boldsymbol{z} = \boldsymbol{y}$ for some $\boldsymbol{y}$, we have

$$\boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}\boldsymbol{M}\boldsymbol{z} = \boldsymbol{C}\boldsymbol{z} - \boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{z} = \boldsymbol{0}.$$

Thus, $\boldsymbol{M}$ projects any vector onto the space where the constraint $\boldsymbol{C}\boldsymbol{y} = 0$ is satisfied.

Proof of Corollary 2.1

Items 1 and 2 are trivial application of Lemma 2.1. To prove 3, we have

$$\mathrm{E}(\tilde{\boldsymbol{z}}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}|\mathscr{I}_t) = \boldsymbol{M}\,\mathrm{E}(\hat{\boldsymbol{z}}_{t+h}|\mathscr{I}_t) = \boldsymbol{M}\,\mathrm{E}(\boldsymbol{z}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{M}\boldsymbol{z}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{z}_{t+h}|\mathscr{I}_t).$$

Proof of Lemma 2.2

$$\mathrm{Var}(\boldsymbol{z}_{t+h} - \tilde{\boldsymbol{z}}_{t+h}) = \mathrm{Var}(\boldsymbol{M}\boldsymbol{z}_{t+h} - \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}) = \boldsymbol{M}\,\mathrm{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h})\boldsymbol{M}' = \boldsymbol{M}\boldsymbol{W}_h\boldsymbol{M}'.$$

If we simplify it further, we have

$$\begin{aligned}\boldsymbol{M}\boldsymbol{W}_h\boldsymbol{M}' &= (\boldsymbol{I} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C})\boldsymbol{W}_h(\boldsymbol{I} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C})' \\ &= \boldsymbol{W}_h - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h \\ &\qquad + \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h \\ &= \boldsymbol{W}_h - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h \\ &= \boldsymbol{M}\boldsymbol{W}_h.\end{aligned}$$

To get $\text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$, we just need to recognise that it is the first $m \times m$ leading principal submatrix of $\text{Var}(\tilde{\boldsymbol{z}}_{t+h} - \boldsymbol{z}_{t+h})$.

Proof of Theorem 2.1

Trivially, $\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h$ and $\boldsymbol{J}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{J}'$ are positive semi-definite. Note that $\text{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \text{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ is the leading principal submatrix of $\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h$, and the leading principal submatrix of a positive semi-definite matrix is positive semi-definite.

Proof of Theorem 2.2

Suppose now that we want to include q more components $\boldsymbol{c}_t^* = \boldsymbol{\Phi}^*\boldsymbol{y}_t$ in the projection. We define $\boldsymbol{z}_t^* = \begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{c}_t^* \end{bmatrix}$, the constraint matrix

$$\boldsymbol{C}^* = \begin{bmatrix} & \boldsymbol{C} & \underset{p\times q}{\boldsymbol{0}} \\ -\underset{q\times m}{\boldsymbol{\Phi}^*} & \underset{q\times p}{\boldsymbol{0}} & \boldsymbol{I}_q \end{bmatrix} = \begin{bmatrix} -\underset{p\times m}{\boldsymbol{\Phi}} & \boldsymbol{I}_p & \underset{p\times q}{\boldsymbol{0}} \\ -\underset{q\times m}{\boldsymbol{\Phi}^*} & \underset{q\times p}{\boldsymbol{0}} & \boldsymbol{I}_q, \end{bmatrix} = \begin{bmatrix} \overline{\boldsymbol{C}} \\ \underline{\boldsymbol{C}} \end{bmatrix} \tag{12}$$

where $\overline{\boldsymbol{C}}$ contains the first p rows of $\boldsymbol{C}^*$ and $\underline{\boldsymbol{C}}$ contains the remaining q rows of $\boldsymbol{C}^*$, the forecast error variance matrix

$$\text{Var}(\boldsymbol{z}_{t+h}^* - \hat{\boldsymbol{z}}_{t+h}^*) = \boldsymbol{W}_h^* = \begin{bmatrix} \boldsymbol{W}_h & \boldsymbol{W}_{yc,h}^* \\ \boldsymbol{W}_{cy,h}^* & \boldsymbol{W}_{c,h}^* \end{bmatrix}.$$

where $\hat{\boldsymbol{z}}_{t+h}^*$ is the h-step-ahead base forecasts of $\boldsymbol{z}_t^*$:

$$\hat{\boldsymbol{z}}_{t+h}^* = \begin{bmatrix} \hat{\boldsymbol{z}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h}^* \end{bmatrix},$$

and the corresponding

$$\boldsymbol{M}^* = \boldsymbol{I} - \boldsymbol{W}_h^*\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}_h^*\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^*.$$

Proving Theorem 2.2 requires proving the following two items.

1. Including additional components in the mapping without including corresponding component constraints is equivalent to not including these additional components at all.
2. For a fixed set of components to be included in the mapping, adding constraints will reduce

forecast error variance.

We start by proving the first statement. Consider the case where we include the additional series $\boldsymbol{c}_t^*$ without using the additional constraint $\boldsymbol{\Phi}^*$. Defining $\boldsymbol{M}^+$ only with $\overline{\boldsymbol{C}}$:

$$\boldsymbol{M}^+ = \boldsymbol{I}_{m+p+q} - \boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}, \tag{13}$$

we have $\tilde{\boldsymbol{z}}_{t+h}^+ = \boldsymbol{M}^+\hat{\boldsymbol{z}}_{t+h}^*$. Further, we obtain

$$\boldsymbol{W}_h^*\overline{\boldsymbol{C}}' = \begin{bmatrix} \boldsymbol{W}_h & \boldsymbol{W}_{yc,h}^* \\ \boldsymbol{W}_{cy,h}^* & \boldsymbol{W}_{c,h}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{C}' \\ \underset{q\times p}{\boldsymbol{O}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W}_h\boldsymbol{C}' \\ \boldsymbol{W}_{cy,h}^*\boldsymbol{C}' \end{bmatrix}$$

and

$$\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}' = \begin{bmatrix} \boldsymbol{C} & \underset{p\times q}{\boldsymbol{O}} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_h\boldsymbol{C}' \\ \boldsymbol{W}_{cy,h}^*\boldsymbol{C}' \end{bmatrix} = \boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}',$$

which gives

$$\begin{aligned}
\boldsymbol{M}^+ &= \boldsymbol{I}_{m+p+q} - \begin{bmatrix} \boldsymbol{W}_h\boldsymbol{C}' \\ \boldsymbol{W}_{cy,h}^*\boldsymbol{C}' \end{bmatrix} (\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1} \begin{bmatrix} \boldsymbol{C} & \underset{p\times q}{\boldsymbol{O}} \end{bmatrix} \\
&= \boldsymbol{I}_{m+p+q} - \begin{bmatrix} \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} & \boldsymbol{O} \\ \boldsymbol{W}_{cy,h}^*\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} & \boldsymbol{O} \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\boldsymbol{z}}_{t+h}^+ &= \boldsymbol{M}^+\hat{\boldsymbol{z}}_{t+h}^* \\
&= \left(\boldsymbol{I}_{m+p+q} - \begin{bmatrix} \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{W}_{cy,h}^*\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} & \boldsymbol{0} \end{bmatrix}\right) \begin{bmatrix} \hat{\boldsymbol{z}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h}^* \end{bmatrix} \\
&= \begin{bmatrix} (\boldsymbol{I}_{m+p} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C})\hat{\boldsymbol{z}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h}^* - \boldsymbol{W}_{cy,h}^*\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\hat{\boldsymbol{z}}_{t+h} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h}^* - \boldsymbol{W}_{cy,h}^*\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\hat{\boldsymbol{z}}_{t+h} \end{bmatrix} \\
&= \begin{bmatrix} \tilde{\boldsymbol{z}}_{t+h} \\ \hat{\boldsymbol{c}}_{t+h}^* - \boldsymbol{W}_{cy,h}^*\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\hat{\boldsymbol{z}}_{t+h} \end{bmatrix}.
\end{aligned}$$

If we only consider the forecast performance relevant to $\boldsymbol{y}_{t+h}$, and define $\boldsymbol{J}^* = \boldsymbol{J}_{m,p+q} = \begin{bmatrix}\boldsymbol{I}_m & \boldsymbol{0}_{m\times(p+q)}\end{bmatrix}$, we have

$$\tilde{\boldsymbol{y}}^+_{t+h} = \boldsymbol{J}^*\tilde{\boldsymbol{z}}^+_{t+h} = \boldsymbol{J}\tilde{\boldsymbol{z}}_{t+h} = \tilde{\boldsymbol{y}}_{t+h}.$$

This means adding additional components without imposing the corresponding constraints will yield the same projected forecasts as if these additional components are not added, which implies that the forecast error variance stays the same:

$$\mathrm{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}^+_{t+h}) = \mathrm{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}) = \boldsymbol{J}\boldsymbol{M}\boldsymbol{W}_h\boldsymbol{J}'. \tag{14}$$

This finishes the proof of the first statement. Now we move on to proving the second statement. We have the forecast error variance matrices

$$\mathrm{Var}(\boldsymbol{z}^*_{t+h} - \tilde{\boldsymbol{z}}^+_{t+h}) = \boldsymbol{M}^+\boldsymbol{W}^*_h = (\boldsymbol{I}_{m+p+q} - \boldsymbol{W}^*_h\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}^*_h\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}})\boldsymbol{W}^*_h$$

and $$\mathrm{Var}(\boldsymbol{z}^*_{t+h} - \tilde{\boldsymbol{z}}^*_{t+h}) = \boldsymbol{M}^*\boldsymbol{W}^*_h = (\boldsymbol{I}_{m+p+q} - \boldsymbol{W}^*_h\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}^*_h\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^*)\boldsymbol{W}^*_h.$$

Taking the difference, we have

$$\begin{aligned}\mathrm{Var}(\boldsymbol{z}^*_{t+h} - \tilde{\boldsymbol{z}}^+_{t+h}) - \mathrm{Var}(\boldsymbol{z}^*_{t+h} - \tilde{\boldsymbol{z}}^*_{t+h}) &= (\boldsymbol{W}^*_h\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}^*_h\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^* - \boldsymbol{W}^*_h\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}^*_h\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}})\boldsymbol{W}^*_h \\ &= \boldsymbol{W}^*_h(\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}^*_h\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^* - \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}^*_h\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}})\boldsymbol{W}^*_h.\end{aligned}$$

Using block matrix inversion, we have

$$\begin{aligned}\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}^*_h\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^* &= \begin{bmatrix}\overline{\boldsymbol{C}}' & \underline{\boldsymbol{C}}'\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{C}}\boldsymbol{W}^*_h\overline{\boldsymbol{C}}' & \overline{\boldsymbol{C}}\boldsymbol{W}^*_h\underline{\boldsymbol{C}}' \\ \underline{\boldsymbol{C}}\boldsymbol{W}^*_h\overline{\boldsymbol{C}}' & \underline{\boldsymbol{C}}\boldsymbol{W}^*_h\underline{\boldsymbol{C}}'\end{bmatrix}^{-1}\begin{bmatrix}\overline{\boldsymbol{C}} \\ \underline{\boldsymbol{C}}\end{bmatrix} \\ &= \begin{bmatrix}\overline{\boldsymbol{C}}' & \underline{\boldsymbol{C}}'\end{bmatrix}\begin{bmatrix}a & b \\ c & d\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{C}} \\ \underline{\boldsymbol{C}}\end{bmatrix} \\ &= \overline{\boldsymbol{C}}'a\overline{\boldsymbol{C}} + \overline{\boldsymbol{C}}'b\underline{\boldsymbol{C}} + \underline{\boldsymbol{C}}'c\overline{\boldsymbol{C}} + \underline{\boldsymbol{C}}'d\underline{\boldsymbol{C}},\end{aligned}$$

where

$$
\begin{aligned}
a &= (\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1} + (\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}' \\
&\qquad (\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}' - \underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1} \\
&= (\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1} + (\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}, \\
b &= -(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}' - \underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1} \\
&= -(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}, \\
c &= -(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}, \\
d &= (\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\boldsymbol{C}^{*\prime}(\boldsymbol{C}^*\boldsymbol{W}_h^*\boldsymbol{C}^{*\prime})^{-1}\boldsymbol{C}^* &= \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} \\
&\quad + \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} \\
&\quad - \overline{\boldsymbol{C}}(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}} \\
&\quad - \underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}'\boldsymbol{W}_h^*\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} \\
&\quad + \underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}} \\
&= \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} \\
&\quad - \overline{\boldsymbol{C}}(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+ \\
&\quad + \underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+ \\
&= \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} + \boldsymbol{M}^{+\prime}\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathrm{Var}(\boldsymbol{z}_{t+h}^* - \tilde{\boldsymbol{z}}_{t+h}^+) - \mathrm{Var}(\boldsymbol{z}_{t+h}^* - \tilde{\boldsymbol{z}}_{t+h}^*) \\
&\qquad = \boldsymbol{W}_h^*(\overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}} + \boldsymbol{M}^{+\prime}\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+ - \overline{\boldsymbol{C}}'(\overline{\boldsymbol{C}}\boldsymbol{W}_h^*\overline{\boldsymbol{C}}')^{-1}\overline{\boldsymbol{C}})\boldsymbol{W}_h^* \\
&\qquad = \boldsymbol{W}_h^*(\boldsymbol{M}^{+\prime}\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}_h^*\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+)\boldsymbol{W}_h^*
\end{aligned}
$$

is positive semi-definite. This concludes the proof of the second statement. Combining the results above, we have

$$
\begin{aligned}
\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}_{t+h})-\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}^*_{t+h}) &= \mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}^+_{t+h})-\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}^*_{t+h}) \\
&= \boldsymbol{J}^*\,\mathrm{Var}(\boldsymbol{z}^*_{t+h}-\tilde{\boldsymbol{z}}^+_{t+h})\boldsymbol{J}^{*\prime}-\boldsymbol{J}^*\,\mathrm{Var}(\boldsymbol{z}^*_{t+h}-\tilde{\boldsymbol{z}}^*_{t+h})\boldsymbol{J}^{*\prime} \qquad (15)\\
&= \boldsymbol{J}^*\boldsymbol{W}^*_h\boldsymbol{M}^{+\prime}\underline{\boldsymbol{C}}'(\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}^*_h\underline{\boldsymbol{C}}')^{-1}\underline{\boldsymbol{C}}\boldsymbol{M}^+\boldsymbol{W}^*_h\boldsymbol{J}^{*\prime}
\end{aligned}
$$

being positive semi-definite. Finally, we have

$$
\begin{aligned}
&(\mathrm{Var}(\boldsymbol{y}_{t+h}-\hat{\boldsymbol{y}}_{t+h})-\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}^*_{t+h}))-(\mathrm{Var}(\boldsymbol{y}_{t+h}-\hat{\boldsymbol{y}}_{t+h})-\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}_{t+h}))\\
&\qquad = \mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}_{t+h})-\mathrm{Var}(\boldsymbol{y}_{t+h}-\tilde{\boldsymbol{y}}^*_{t+h})
\end{aligned}
$$

being a positive semi-definite matrix where the diagonal terms are non-negative, whose trace, therefore, is non-negative. This means using a larger number of components in the mapping achieves lower or equal forecast error variances. In other words, the reduction in forecast error variance of each series is non-decreasing as more components are added, giving Theorem 2.2.

Proof of Theorem 2.3

Denote $\boldsymbol{\psi}_i = \begin{bmatrix}-\boldsymbol{\phi}_i & \mathbf{0}_{1\times(i-1)} & 1\end{bmatrix}$ and $\boldsymbol{W}_h^{(i)}$ to be the base forecast error variance of the original series and the first i components. Starting with the first component, Equation 4 becomes

$$
\mathrm{tr}(\boldsymbol{J}_{m,1}\boldsymbol{W}_h^{(1)}\boldsymbol{\psi}_1'(\boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{\psi}_1')^{-1}\boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{J}_{m,1}') = (\boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{\psi}_1')^{-1}\boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{J}_{m,1}'\boldsymbol{J}_{m,1}\boldsymbol{W}_h^{(1)}\boldsymbol{\psi}_1', \qquad (16)
$$

$$
\text{where} \qquad \boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{J}_{m,1}' = \begin{bmatrix}-\boldsymbol{\phi}_1 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{W}_{y,h} & \boldsymbol{w}_{c_1y,h}' \\ \boldsymbol{w}_{c_1y,h} & W_{c_1,h}\end{bmatrix}\begin{bmatrix}\boldsymbol{I}_m \\ 0\end{bmatrix}
$$

$$
= -\boldsymbol{\phi}_1\boldsymbol{W}_{y,h}+\boldsymbol{w}_{c_1y,h}.
$$

Equation 16 is obviously non-negative. For it to be larger than 0, we need $\boldsymbol{\psi}_1\boldsymbol{W}_h^{(1)}\boldsymbol{J}_{m,1}' \neq 0$, which gives $\boldsymbol{\phi}_1\boldsymbol{W}_{y,h} \neq \boldsymbol{w}_{c_1y,h}$.

When it comes to adding the ith component on top of the first $i-1$ components, we define

$$\overline{\boldsymbol{C}}_i = \begin{bmatrix} \boldsymbol{\psi}_1 & \boldsymbol{0}_{1\times i} \\ \boldsymbol{\psi}_2 & \boldsymbol{0}_{1\times(i-1)} \\ \vdots & \vdots \\ \boldsymbol{\psi}_i & 0 \end{bmatrix}$$

and

$$\boldsymbol{M}_i^+ = \boldsymbol{I}_{m+i} - \boldsymbol{W}_h^{(i)}\overline{\boldsymbol{C}}_{i-1}'(\overline{\boldsymbol{C}}_{i-1}\boldsymbol{W}_h^{(i)}\overline{\boldsymbol{C}}_{i-1}')^{-1}\overline{\boldsymbol{C}}_{i-1}$$

analogously to Equation 12 and Equation 13. Following Equation 15, the additional reduction of forecast error variance when adding the ith component becomes

$$\boldsymbol{J}_{m,i}\boldsymbol{W}_h^{(i)}\boldsymbol{M}_i^{+\prime}\boldsymbol{\psi}_i'(\boldsymbol{\psi}_i\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{\psi}_i')^{-1}\boldsymbol{\psi}_i\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{J}_{m,i}' = (\boldsymbol{\psi}_i\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{\psi}_i')^{-1}\boldsymbol{\psi}_i\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{J}_{m,i}'\boldsymbol{J}_{m,i}\boldsymbol{W}_h^{(i)}\boldsymbol{M}_i^{+\prime}\boldsymbol{\psi}_i'.$$

Similar to before, we would want $\boldsymbol{\psi}_i\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{J}_{m,i}' \neq \boldsymbol{0}$. Note that $\boldsymbol{\psi}_i$ concerns the first m rows and the last row of $\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}$, and $\boldsymbol{J}_{m,i}'$ concerns the first m columns. Combined with the implication from Equation 14 that the $m\times m$ leading principal submatrix in equation $\boldsymbol{J}_{m,i}\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{J}_{m,i}' = \boldsymbol{J}_{m,i-1}\boldsymbol{M}_{i-1}\boldsymbol{W}_h^{(i-1)}\boldsymbol{J}_{m,i-1}'$ is the same, we suppress the straightforward yet tiresome details, and obtain

$$\boldsymbol{\phi}_i\boldsymbol{W}_{\tilde{y},h}^{(i-1)} \neq \begin{bmatrix}\boldsymbol{0}_{1\times m+i-1} & 1\end{bmatrix}\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}\boldsymbol{J}_{m,i}',$$

where $\boldsymbol{W}_{\tilde{y},h}^{(i-1)} = \boldsymbol{J}_{m,i-1}\boldsymbol{M}_{i-1}\boldsymbol{W}_h^{(i-1)}\boldsymbol{J}_{m,i-1}'$ is the projected forecast error variance of the original series using the first $i-1$ components, and the right hand side of the inequality is simply a one-row matrix consisting of the first m elements in the last row of $\boldsymbol{M}_i^+\boldsymbol{W}_h^{(i)}$, which can be denoted as $\boldsymbol{w}_{\tilde{c}_i\tilde{y},h}^{(i-1)}$ and interpreted as the covariance between the projected forecast of the original series using the first $i-1$ components, and the projected forecast of the ith component using the first $i-1$ components.

Proof of Lemma 2.3

If $\boldsymbol{GS} = \boldsymbol{I}$, $\boldsymbol{SG}$ is a projection matrix: $\boldsymbol{SGSG} = \boldsymbol{SG}$.

For any $\boldsymbol{z}$ such that $\boldsymbol{SGz} = \boldsymbol{y}$ for some $\boldsymbol{y}$, we have $\boldsymbol{Cy} = \boldsymbol{CSGz} = \boldsymbol{0}$ because $\boldsymbol{CS} = \begin{bmatrix} -\boldsymbol{\Phi} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{\Phi}' \end{bmatrix}' = \boldsymbol{0}$. Similarly to $\boldsymbol{M}$, $\boldsymbol{SG}$ projects a vector to the same space where $\boldsymbol{C}$ is satisfied.

Proof of Corollary 2.2

Item 1 is an direct application of Lemma 2.3. From Lemma 2.3 and Lemma 2.4 in Rao (1974), we have

$$\boldsymbol{SGz}_{t+h} = \boldsymbol{z}_{t+h} = \boldsymbol{Sy}_{t+h}.$$

Left multiplying by $\boldsymbol{G}$ on both sides, we have $\boldsymbol{Gz}_{t+h} = \boldsymbol{y}_{t+h}$ and item 2 is proven. To prove Item 3, we have

$$\mathrm{E}(\tilde{\boldsymbol{y}}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{G}\hat{\boldsymbol{z}}_{t+h}|\mathscr{I}_t) = \boldsymbol{G}\,\mathrm{E}(\hat{\boldsymbol{z}}_{t+h}|\mathscr{I}_t) = \boldsymbol{G}\,\mathrm{E}(\boldsymbol{z}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{Gz}_{t+h}|\mathscr{I}_t) = \mathrm{E}(\boldsymbol{y}_{t+h}|\mathscr{I}_t).$$

Proof of Lemma 2.4

Let the base and projected forecast errors be given as

$$\begin{aligned}
\hat{\boldsymbol{e}}_{y,t+h} &= \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}, \\
\hat{\boldsymbol{e}}_{z,t+h} &= \boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h}, \\
\tilde{\boldsymbol{e}}_{y,t+h} &= \boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}, \\
\text{and} \qquad \tilde{\boldsymbol{e}}_{z,t+h} &= \boldsymbol{z}_{t+h} - \tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{Sy}_{t+h} - \boldsymbol{S}\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{e}}_{y,t+h}. \\
\text{Then we have} \qquad \tilde{\boldsymbol{e}}_{z,t+h} &= \hat{\boldsymbol{e}}_{z,t+h} + \hat{\boldsymbol{z}}_{t+h} - \tilde{\boldsymbol{z}}_{t+h} \\
&= \hat{\boldsymbol{e}}_{z,t+h} + \hat{\boldsymbol{z}}_{t+h} - \boldsymbol{SG}\hat{\boldsymbol{z}}_{t+h} \\
&= \hat{\boldsymbol{e}}_{z,t+h} + (\boldsymbol{I} - \boldsymbol{SG})(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{e}}_{z,t+h}) \\
\text{and} \qquad &= \boldsymbol{SG}\hat{\boldsymbol{e}}_{z,t+h} + (\boldsymbol{I} - \boldsymbol{SG})\boldsymbol{Sy}_{t+h} \\
\boldsymbol{S}\tilde{\boldsymbol{e}}_{y,t+h} &= \boldsymbol{SG}\hat{\boldsymbol{e}}_{z,t+h},
\end{aligned}$$

where the last line comes from $\boldsymbol{GS} = \boldsymbol{I}$. Left multiplying by $\boldsymbol{G}$ on both sides, we have

$$\boldsymbol{GS}\tilde{\boldsymbol{e}}_{y,t+h} = \boldsymbol{GSG}\hat{\boldsymbol{e}}_{z,t+h} \qquad \text{and} \qquad \tilde{\boldsymbol{e}}_{y,t+h} = \boldsymbol{G}\hat{\boldsymbol{e}}_{z,t+h},$$

and therefore

$$\mathrm{Var}(\tilde{\boldsymbol{y}}_{t+h} - \boldsymbol{y}_{t+h}) = \mathrm{Var}(\tilde{\boldsymbol{e}}_{y,t+h}) = \mathrm{Var}(\boldsymbol{G}\hat{\boldsymbol{e}}_{z,t+h}) = \boldsymbol{G}\,\mathrm{Var}(\hat{\boldsymbol{e}}_{z,t+h})\boldsymbol{G}' = \boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}'.$$

Proof of Theorem 2.4

This can be proved in a few different ways. We adopt the approach of Ando and Narita (2024) to obtain the solution to Equation 10, but the procedure from Luenberger (1969,p. 85) can also be used, where the problem is divided to Equation 11 and reconstructed to find the solution to Equation 10.

There exists a Lagrange multiplier $\boldsymbol{\Lambda}$ such that

$$L(\boldsymbol{G}) = \mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}') + \mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{I} - \boldsymbol{G}\boldsymbol{S}))$$

is stationary at an extremum $\boldsymbol{G}$ (Luenberger 1969,p. 243, Theorem 1). We find the numerator of the Gateaux differential (Luenberger 1969,p. 171)

$$\lim_{\alpha\to 0}\frac{L(\boldsymbol{G}+\alpha\boldsymbol{H}) - L(\boldsymbol{G})}{\alpha}$$

to be

$$\begin{aligned}
L(\boldsymbol{G}+\alpha\boldsymbol{H}) - L(\boldsymbol{G}) =& \,\mathrm{tr}((\boldsymbol{G}+\alpha\boldsymbol{H})\boldsymbol{W}_h(\boldsymbol{G}+\alpha\boldsymbol{H})') + \mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{I} - (\boldsymbol{G}+\alpha\boldsymbol{H})\boldsymbol{S})) - \mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}') - \mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{I} - \boldsymbol{G}\boldsymbol{S}))\\
=& \,\mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}') + \mathrm{tr}(\alpha^2\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{H}') + \mathrm{tr}(\alpha\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{H}') + \mathrm{tr}(\alpha\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{G}') + \mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{I} - \boldsymbol{G}\boldsymbol{S})) - \mathrm{tr}(\alpha\boldsymbol{\Lambda}'\boldsymbol{H}\boldsymbol{S})\\
&- \mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}') - \mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{I} - \boldsymbol{G}\boldsymbol{S}))\\
=& \,\alpha^2\,\mathrm{tr}(\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{H}') + \alpha\,\mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{H}') + \alpha\,\mathrm{tr}(\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{G}') - \alpha\,\mathrm{tr}(\boldsymbol{\Lambda}'\boldsymbol{H}\boldsymbol{S}).
\end{aligned}$$

Thus, the Gateaux differential becomes

$$\begin{aligned}
\lim_{\alpha\to 0}\frac{L(\boldsymbol{G}+\alpha\boldsymbol{H})-L(\boldsymbol{G})}{\alpha} &= \lim_{\alpha\to 0}\alpha\,\mathrm{tr}(\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{H}')+\mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{H}')+\mathrm{tr}(\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{G}')-\mathrm{tr}(\boldsymbol{\Lambda}'\boldsymbol{H}\boldsymbol{S}) \\
&= \mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{H}')+\mathrm{tr}(\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{G}')-\mathrm{tr}(\boldsymbol{\Lambda}'(\boldsymbol{H}\boldsymbol{S})) \\
&= \mathrm{tr}(2\boldsymbol{H}\boldsymbol{W}_h\boldsymbol{G}'-\boldsymbol{\Lambda}'\boldsymbol{H}\boldsymbol{S}) \\
&= \mathrm{tr}(\boldsymbol{H}(2\boldsymbol{W}_h\boldsymbol{G}'-\boldsymbol{S}\boldsymbol{\Lambda}')),
\end{aligned}$$

which we set to zero with a value of $\boldsymbol{G}^*$

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{H}(2\boldsymbol{W}_h\boldsymbol{G}^{*\prime}-\boldsymbol{S}\boldsymbol{\Lambda}')) &= 0 \\
2\boldsymbol{W}_h\boldsymbol{G}^* &= \boldsymbol{S}\boldsymbol{\Lambda}' \\
\boldsymbol{G}^{*\prime} &= \frac{1}{2}\boldsymbol{W}_h^{-1}\boldsymbol{S}\boldsymbol{\Lambda}'.
\end{aligned}$$

Multiplying $\boldsymbol{S}'$ to the left of both sides. we have

$$\boldsymbol{S}'\boldsymbol{G}^{*\prime}=\boldsymbol{I}=\frac{1}{2}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S}\boldsymbol{\Lambda}' \qquad \text{and} \qquad \boldsymbol{\Lambda}'=2(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}$$

because $\boldsymbol{G}^*\boldsymbol{S}=\boldsymbol{I}$. Putting it back in, we have

$$\boldsymbol{G}^{*\prime}=\boldsymbol{W}_h^{-1}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1} \qquad \text{and} \qquad \boldsymbol{G}^*=(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}.$$

To see how Equation 10 can be split into separate problems, recognize

$$\mathrm{tr}(\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}')=\sum_{i=1}^{m}\boldsymbol{g}_i'\boldsymbol{W}_h\boldsymbol{g}_i,$$

where $\boldsymbol{W}_h$ is a positive definite variance covariance matrix, which makes each $\boldsymbol{g}_i'\boldsymbol{W}_h\boldsymbol{g}_i$ positive. Additionally, the element in the ith row and jth column of $\boldsymbol{G}\boldsymbol{S}$ is $\boldsymbol{g}_i'\boldsymbol{s}_j$, and the element in the ith row and jth column of an identity matrix is δ_{ij}. Therefore, the problem in Equation 10 is m separate problems.