4.5.3 Optimality of the Inverse Noise Covariance as Weighting Matrix

As mentioned at the end of the previous section, the choice of the inverse of the noise covariance as weighting matrix is optimal in the sense that no other choice of weighting matrix delivers a smaller covariance. We will be able to prove an even stronger statement, namely that the optimally weighted least squares estimator is the best among all unbiased linear estimators. In this subsection we drop the subindex $(\cdot)_N$ for notational simplicity.

Theorem 7 (Cramer-Rao Lower Bound for Unbiased Linear Estimators) *Assume measurements $y \in \mathbb{R}^N$ are generated according to a model $y = \Phi\theta_0 + \epsilon$ with $\theta_0 \in \mathbb{R}^d$ the true (but unknown) parameter and $\epsilon \in \mathbb{R}^N$ zero-mean measurement noise with positive definite covariance matrix $\mathrm{cov}(\epsilon) = \Sigma_\epsilon$ and $\mathrm{rank}(\Phi) = d$. Regard any unbiased linear estimator $\hat{\theta}_A := Ay$ with fixed matrix $A \in \mathbb{R}^{d\times N}$. Then*

$$\mathrm{cov}(\hat{\theta}_A) \quad \succcurlyeq \quad (\Phi^\top \Sigma_\epsilon^{-1} \Phi)^{-1}. \tag{4.32}$$

The lower bound is achieved by the optimally weighted least squares estimator with matrix $A^* := (\Phi^\top W^* \Phi)^{-1}\Phi^\top W^*$ using $W^* := \Sigma_\epsilon^{-1}$, i.e., it has the smallest covariance matrix among all unbiased linear estimators.

Proof: First, we note that for any given matrix A, the estimator is unbiased if and only if $A\Phi = \mathbb{I}$, because $\mathbb{E}\{Ay\} = \mathbb{E}\{A\Phi\theta_0 + \epsilon\} = A\Phi\theta_0$. Second, we observe that $\mathrm{cov}(\hat{\theta}_A) = A\Sigma_\epsilon A^\top$. Third, we have shown above in Eq. (4.31) that $\mathrm{cov}(\hat{\theta}_{A^*}) = (\Phi^\top \Sigma_\epsilon^{-1}\Phi)^{-1}$, i.e., that the lower bound is indeed achieved by the optimally weighted least squares estimator. Now, the main statement of the above theorem is equivalent to

$$\forall A \in \mathbb{R}^{d\times N}: \quad \Big(\quad A\Phi = \mathbb{I} \quad \Big) \quad \Rightarrow \quad \Big(\quad A\Sigma_\epsilon A^\top \quad \succcurlyeq \quad (\Phi^\top \Sigma_\epsilon^{-1}\Phi)^{-1} \quad \Big). \tag{4.33}$$

We need to show that the matrix $A\Sigma_\epsilon A^\top - (\Phi^\top \Sigma_\epsilon^{-1}\Phi)^{-1}$ is positive semidefinite if $A\Phi = \mathbb{I}$. In order to show this we introduce the matrix

$$B = \begin{bmatrix} A\,\Sigma_\epsilon^{\frac{1}{2}} \\ \Phi^\top \Sigma_\epsilon^{-\frac{1}{2}} \end{bmatrix} \quad \text{for which we have} \quad BB^\top = \begin{bmatrix} A\Sigma_\epsilon A^\top & A\Phi \\ (A\Phi)^\top & \Phi^\top\Sigma_\epsilon^{-1}\Phi \end{bmatrix}$$

Now there is a famous lemma about the "Schur complement" (cf. A 5.5 in [**?**]) that states that if the lower right block in a symmetric matrix is positive definite, the whole matrix is positive semidefinite if and only if the so called "Schur complement" of the lower right block is positive definite.

In our case, the lower right block is given by $\Phi^\top\Sigma_\epsilon^{-1}\Phi$, which is positive definite. The Schur complement – which must be positive semidefinite due to the fact that the whole matrix $BB^\top$ is positive semidefinite – is in our case given by $A\Sigma_\epsilon A^\top - A\Phi\,(\Phi^\top\Sigma_\epsilon^{-1}\Phi)^{-1}(A\Phi)^\top$. Because $A\Phi = \mathbb{I}$ we deduce that $A\Sigma_\epsilon A^\top - (\Phi^\top\Sigma_\epsilon^{-1}\Phi)^{-1} \succcurlyeq 0$. This is what we wanted to prove.