Notes from *Covariance Matching Estimation Techniques for Array Signal Processing Applications*

Daniel Eriksson

June 13, 2011

1 COMET

These notes are based on Sections 1 - 4 and Appendix A in [1] which describes the COMET (COvariance Matching Estimation Techniques) method for estimating the parameters of a signal model. The derivation of the method is based on EXIP (Extended invariance principle).

1.1 Data models

Here the model and the parameters that is going to be estimated are introduced. Some notations that will be used later are defined and some required assumtions are specified. Note that the parameters defining the noise covariances enters linearly and are separated from the signal parameters which enters non-linearly.

The data model that is considered is

$$y(t) = Ax(t) + e(t) \tag{1}$$

where $y(t) \in \mathbb{C}^{m\times 1}$ and $A = A(\theta) \in \mathbb{C}^{m\times n_\theta}$.
It is assumed that:

- Emitter signals $x(t)$ are random.
- Observation vectors $\{y(t)\}_{t=1,2,...}$ are i.i.d. circular Gaussian random variables with zero mean (, i.e. $e^{i\phi}Z$ has the same probability distribution as Z for all real ϕ, see for example [3]).
- The emitter signal $x(t)$ and the noise $e(t)$ are uncorrelated which gives that

$$R(\theta,\mu,\sigma) = E\left[y(t)y^*(t)\right] = \underbrace{R_s(\theta,\mu)}_{=E[Ax(t)x^*(t)A^*]} + \underbrace{Q(\sigma)}_{=E[e(t)e^*(t)]} . \tag{2}$$

- The matrices $R_s(\theta,\mu)$ and $Q(\sigma)$ are linearly parameterized by μ and σ, i.e.,

$$R_s(\theta,\mu) = \Psi(\theta)\mu$$
$$Q(\sigma) = \Sigma\sigma$$

  where Σ is known, $\Psi(\theta)$ is a given function, μ are referred to emitter signal covariance parameters and σ as noise covariance parameters.

The operator vec(C) of a matrix $C \in \mathbb{R}^{n\times n}$ returns a vector where all the columns in C are stacked on top of each other as

$$\text{vec}(C) = \text{vec}\left(\begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \dots & c_{nn} \end{pmatrix}\right) = \begin{pmatrix} c_{11} \\ \vdots \\ c_{n1} \\ c_{12} \\ \vdots \\ c_{nn} \end{pmatrix}.$$

Then (2) can be written as

$$r(\theta, \mu, \sigma) = \text{vec}\left(R(\theta, \mu, \sigma)\right) = \Psi(\theta)\mu + \Sigma\sigma = \begin{pmatrix}\Psi(\theta) & \Sigma\end{pmatrix}\begin{pmatrix}\mu \\ \sigma\end{pmatrix} \triangleq \Phi(\theta)\alpha \tag{3}$$

where α are the linear parameter vectors parameterizing the covariances.
To ensure that the used parametrization is "uniquely identifiable" the following restrictions are necessary:

- One parameter per emitter signal, i.e. $n_\theta = \dim(x(t)) = \dim(\theta)$, where n_θ is known.
- $\Psi(\theta') = \Psi(\theta'') \implies \theta' = \theta''$ (unique solution)
- $\Phi(\theta)$ is full rank at true values θ.
- Number of unknowns $\leq$ number of estimating equations available, i.e., $n_\theta + n_\mu + n_\sigma \leq m^2$.

These restrictions above are necessary but not sufficient. Sufficient conditions for identifiability are application specific since they depend on the structure of parametrization.

1.2 Covariance Matching Estimation Technique

When the model and necessary assumptions are set, a way to solve the problem is formed. Note in the solution, that the complex valued entries in R (and $\hat{R}$) are represented as real values by an invertible transformation.

Let

$$\hat{r} = \text{vec}(\hat{R}) \tag{4}$$

be a vector based on the sample covariance matrix

$$\hat{R} = \frac{1}{N}\sum_{t=1}^{N} y(t)y^*(t) \tag{5}$$

and $r = \text{vec}(R)$. The vector $\hat{r}$ has several related entries because $\hat{R}$ is Hermitian where off-diagonal entries are complex valued. Form a vector $\hat{\gamma}$ where all entries are real valued which can be described by a linear transformation

$$\hat{\gamma} = J\text{vec}(\hat{R}) = J\hat{r} \tag{6}$$

where $J \in \mathbb{C}^{m^2 \times m^2}$, whose entries are $0, 1, \frac{1}{2}$ or $\pm \frac{j}{2}$, is invertible.

The matrix J is designed so that the non-diagonal elements in $\hat{R}$ is formed as linear combinations to real valued entries in $\hat{\gamma}$ as

$$\begin{aligned} \text{real}(\hat{R}_{ij}) &= \frac{1}{2}(\hat{R}_{ij} + \hat{R}_{ji}) \\ \text{imag}(\hat{R}_{ij}) &= \frac{1}{2j}(\hat{R}_{ij} - \hat{R}_{ji}). \end{aligned} \tag{7}$$

The model of the covariance matrix (2) can now be expressed as

$$\gamma(\theta, \alpha) = J\text{vec}(R(\theta, \alpha)) = Jr(\theta, \alpha) \tag{8}$$

which is a real vector representing the elements in R. **This representation of $\hat{R}$ as a real valued vector $\hat{\gamma}$ is an important property of the COMET method.**

The COMET solution is that the estimates of θ and α are obtained by fitting the data $\hat{\gamma}$ to the model (8) in a weighted least square sense

$$(\hat{\gamma} - \gamma(\theta, \alpha))^T C^{-1} (\hat{\gamma} - \gamma(\theta, \alpha)) \triangleq \tilde{\gamma}^T(\theta, \alpha) C^{-1} \tilde{\gamma}(\theta, \alpha) \tag{9}$$

where C^{-1} is the inverse of the (asymptotic) covariance matrix of the residuals

$$\tilde{\gamma} = \hat{\gamma} - \gamma \tag{10}$$

An expression of C is required and will be derived in the next section.

1.3 COMET: Detailed Formulas

The COMET estimator first estimates $\hat{\theta}$ and then $\hat{\alpha}$ (as a function of $\hat{\theta}$). The estimated parameters are shown to be real valued and optimal. An important note is that the approach when deriving COMET is the approximation, based on EXIP, by replacing R by $\hat{R}$. Note that the method requires a sufficiently large number of samples N to have positive definite covariance matrices $\hat{R}$ and $\hat{Q}$.

To find an expression for C, the second-order moments of the covariance estimates are determined. The ith $(m \times 1)$ subvector $\hat{r}_i$ of the $(m^2 \times 1)$ vector $\hat{r}$ is given by

$$\hat{r}_i = \frac{1}{N} \sum_{t=1}^{N} y(t) y_i^*(t). \tag{11}$$

Since the signals $y(t)$ are

- Independent from snapshot to snapshot
- circularly Gaussian distributed.

the following relation is established

$$
\begin{aligned}
E[\hat{r}_i\hat{r}_j^*] &= E\left[\frac{1}{N}\sum_{t=1}^{N} y(t)y_i^*(t)\left(\frac{1}{N}\sum_{s=1}^{N} y(s)y_j^*(s)\right)^*\right] = \\
&= \frac{1}{N^2}E\left[\sum_{t=1}^{N} y(t)y_i^*(t)\left(\sum_{s=1}^{N} y(s)y_j^*(s)\right)^*\right] = \frac{1}{N^2}E\left[\sum_{t=1}^{N}\sum_{s=1}^{N} y(t)y_i^*(t)\left(y(s)y_j^*(s)\right)^*\right] = \\
&= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N} E\left[y(t)y_i^*(t)y_j(s)y^*(s)\right] = \\
&= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\left(E\left[y(t)y_i^*(t)y_j(s)y^*(s)\right] - E\left[y(t)y_i^*(t)\right]E\left[y_j(s)y^*(s)\right]\right) + \\
&+ \underbrace{\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N} E\left[y(t)y_i^*(t)\right]E\left[y_j(s)y^*(s)\right]}_{r_i r_j^*} = \\
&= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\left(E\left[y(t)y_i^*(t)y_j(s)y^*(s)\right] - E\left[y(t)y_i^*(t)\right]E\left[y_j(s)y^*(s)\right]\right) + r_i r_j^*.
\end{aligned}
\tag{12}
$$

Consider the first term in the last equality. All expectations where $t \neq s$ are zero since they are independent. For $t = s$, the element (p, k) of $E\left[y(t)y_i^*(t)y_j(t)y^*(t)\right]$ can be written as (where (t) is omitted)

$$
\begin{aligned}
& E\left[y_p y_i^* y_j y_k^*\right] = E\left[y_p y_i^*\right]E\left[y_j y_k^*\right] + E\left[y_p y_j\right]E\left[y_i^* y_k^*\right] + E\left[y_p y_k^*\right]E\left[y_i^* y_j\right] \\
& \implies E\left[y y_i^* y_j y^*\right] = E\left[y y_i^*\right]E\left[y_j y^*\right] + 0 + R_{ji}R
\end{aligned}
\tag{13}
$$

The second term is zero since the observations are circularly Gaussian, i.e. $E[y_p y_j] = E\left[y_i^* y_k^*\right] = 0$. Using the results (13) in (12) gives

$$
\begin{aligned}
& \frac{1}{N^2}\sum_{t=1}^{N}\left(E\left[y y_i^* y_j y^*\right] - E\left[y y_i^*\right]E\left[y_j y^*\right]\right) + r_i r_j^* = \\
& = \frac{1}{N^2}\sum_{t=1}^{N}\left(E\left[y y_i^*\right]E\left[y_j y^*\right] + R_{ji}R - E\left[y y_i^*\right]E\left[y_j y^*\right]\right) + r_i r_j^* = \\
& = \frac{1}{N^2}\sum_{t=1}^{N}\left(R_{ji}R\right) + r_i r_j^* = \frac{1}{N}R_{ji}R + r_i r_j^*
\end{aligned}
\tag{14}
$$

Now writing the covariance matrix of the error of the estimation by using (14) gives

$$
\begin{aligned}
& E\left[(\hat{r}_i - r_i)(\hat{r}_j - r_j)^*\right] = E\left[\hat{r}_i\hat{r}_j^* - \hat{r}_i r_j^* - r_i\hat{r}_j^* - r_i r_j^*\right] = \\
& = E\left[\hat{r}_i\hat{r}_j^*\right] - E\left[\hat{r}_i\right]r_j^* - r_i E\left[\hat{r}_j^*\right] - r_i r_j^* = /E\left[\hat{r}\right] = r/ = \\
& = E\left[\hat{r}_i\hat{r}_j^*\right] - r_i r_j^* - r_i r_j^* + r_i r_j^* = E\left[\hat{r}_i\hat{r}_j^*\right] - r_i r_j^* = \frac{1}{N}R_{ji}R
\end{aligned}
\tag{15}
$$

which gives that

$$
\bar{C} \triangleq E\left[(\hat{r} - r)(\hat{r} - r)^*\right] = \frac{1}{N}(R^T \otimes R).
\tag{16}
$$

This is used to define C as

$$C \triangleq E\left[(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)^*\right] = E\left[J(\hat{r}-r)(\hat{r}-r)^*J^*\right] = \frac{1}{N}J(R^T \otimes R)J^* = J\bar{C}J^*. \tag{17}$$

This shows how a consistent estimate of C can be formed from the data. By normalizing (9) it can be rewritten as

$$\begin{aligned} \frac{1}{N}\tilde{\gamma}^T(\theta,\alpha)C^{-1}\tilde{\gamma}(\theta,\alpha) &= \frac{1}{N}\tilde{r}^*(\theta,\alpha)J^*C^{-1}J\tilde{r}(\theta,\alpha) = \\ = \frac{1}{N}\tilde{r}^*(\theta,\alpha)J^*(J\bar{C}J^*)^{-1}J\tilde{r}(\theta,\alpha) &= \frac{1}{N}\tilde{r}^*(\theta,\alpha)\bar{C}^{-1}\tilde{r}(\theta,\alpha) = \\ &= \tilde{r}^*(\theta,\alpha)(R^T \otimes R)^{-1}\tilde{r}(\theta,\alpha) \end{aligned} \tag{18}$$

The following result is not important for the derivation of COMET. *By using the properties:*

$$\begin{aligned} (B^T \otimes A)\text{vec}(X) &= \text{vec}(AXB) \\ (B \otimes A)^{-1} &= (B^{-1} \otimes A^{-1}) \end{aligned} \tag{19}$$

and that $\tilde{R} = \hat{R} - R$, the cost function can be simplified as

$$\tilde{r}^*(R^T \otimes R)^{-1}\tilde{r} = \tilde{r}^*(R^T \otimes R)^{-1}\text{vec}(\tilde{R}) = \tilde{r}^*(R^{-T} \otimes R^{-1})\text{vec}(\tilde{R}) = \tilde{r}^*\text{vec}(R^{-1}\tilde{R}R^{-1}). \tag{20}$$

This can be further simplified using that $vec(A)^*vec(B) = tr(A^*B)$ which gives

$$\tilde{r}^*\text{vec}(R^{-1}\tilde{R}R^{-1}) = \text{vec}(\tilde{R})^*\text{vec}(R^{-1}\tilde{R}R^{-1}) = \text{tr}(\tilde{R}^*R^{-1}\tilde{R}R^{-1}). \tag{21}$$

$\triangle$

By using that $\hat{R}$ is a consistent estimate of R, according to EXIP, a large-sample ML-estimator of θ and α is obtained by minimizing the cost function (18) where R is replaced by $\hat{R}$ as

$$\tilde{r}^*(\theta,\alpha)(\hat{R}^T \otimes \hat{R})^{-1}\tilde{r}(\theta,\alpha) = \tilde{r}^*(\theta,\alpha)\hat{W}^{-1}\tilde{r}(\theta,\alpha) \tag{22}$$

where $\hat{W} = (\hat{R}^T \otimes \hat{R})$. The computed cost function in (21) and (22) is sometimes referred to as the generalized least square criterion and is a large sample approximation to the ML criterion.

Remember that $r(\theta,\alpha)$ **is a linear function of** α **and that** $\Phi(\theta)$ **have full column rank**. Then θ and α can be uniquely determined from a given R. First $\hat{\alpha}$ can be determined as a function of θ

$$\begin{aligned} \min_{\alpha} \tilde{r}^*(\theta,\alpha)\hat{W}^{-1}\tilde{r}(\theta,\alpha) &= \min_{\alpha} ||\hat{W}^{-1/2}\hat{r} - \hat{W}^{-1/2}\Phi(\theta)\alpha||^2 \\ \Longrightarrow \hat{\alpha} &= \left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1}\hat{r} \end{aligned} \tag{23}$$

where $||\cdot||$ denotes the Euclidian vector norm and $\hat{W}^{-1/2}$ is a Hermitian square root factor of $\hat{W}^{-1}$.

Optimality of the estimated vector $\hat{\alpha}$ can only be ensured if it is real valued which will be shown here. Since J is invertible, (23) can be written as

$$\hat{\alpha} = \left((J\Phi)^*(J\hat{W}J^*)^{-1}J\Phi\right)^{-1}(J\Phi)^*(J\hat{W}J^*)^{-1}J\hat{r} \tag{24}$$

where $J\hat{r} = \hat{\gamma}$ and $J\Phi$ are real (since $Jr = J\Phi\alpha$ is real if α is real-valued implies also that $J\Phi$ is real-valued). Consider a sequence of random numbers $[z(t)]$ that are circularly Gaussian distributed with mean zero and covarance matrix $\hat{R}$ which is fixed. Let $\hat{\rho} = \text{vec}(\sum_{t=1}^{N} z(t)z^*(t)/N)$. Corresponding to $\hat{r}$, the previous calculations, leading to (6) and (18) implies that $J\hat{\rho}$ is real valued and also the covariance of the vector $J(\hat{R}^T \otimes \hat{R})J^*$. This result assures that $J\hat{W}J^*$ in (24) is also real valued and thus $\hat{\alpha}$ is real-valued and optimality is ensured.

Now $\hat{\theta}$ can be computed. Substitute α in

$$\min_{\alpha}(\hat{r} - \Phi(\theta)\alpha)^*\hat{W}^{-1}(\hat{r} - \Phi(\theta)\alpha) \tag{25}$$

with the estimated $\hat{\alpha}$ in (23) which gives

$$\begin{aligned}
&\min_{\alpha}(\hat{r} - \Phi(\theta)\alpha)^*\hat{W}^{-1}(\hat{r} - \Phi(\theta)\alpha) = \\
&= \left(\hat{r} - \Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1}\hat{r}\right)^*\hat{W}^{-1}\left(\hat{r} - \Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1}\hat{r}\right) = \\
&= \left(\hat{W}^{-1/2}(I - \Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1})\hat{r}\right)^* \times \\
&\left(\hat{W}^{-1/2}(I - \Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1})\hat{r}\right) = \\
&= \|\left(I - \hat{W}^{-1/2}\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2}\right)\hat{W}^{-1/2}\hat{r}\|^2 = \\
&= \hat{r}^*\hat{W}^{-1/2}\left(I - \hat{W}^{-1/2}\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2}\right)\hat{W}^{-1/2}\hat{r}
\end{aligned} \tag{26}$$

where the last equality depends on that $\left(I - \hat{W}^{-1/2}\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2}\right)$ is a projection matrix. This can be summarized as

$$\hat{\theta} = \arg\min_{\theta} \hat{r}^*\hat{W}^{-1/2}\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\hat{r} \tag{27}$$

where

$$\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)} = I - \Pi_{\hat{W}^{-1/2}\Phi(\theta)} = I - \hat{W}^{-1/2}\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2}. \tag{28}$$

This can be further simplified using that

$$\begin{aligned}
\hat{W}^{-1/2}\hat{r} &= (\hat{R}^{-T/2} \otimes \hat{R}^{-1/2})\text{vec}(\hat{R}) = \text{vec}(\hat{R}^{-1/2}\hat{R}\hat{R}^{-1/2}) = \text{vec}(I) \\
\hat{W}^{-1/2}\text{vec}(I) &= (\hat{R}^{-T/2} \otimes \hat{R}^{-1/2})\text{vec}(I) = \text{vec}(\hat{R}^{-1})
\end{aligned} \tag{29}$$

which results in the cost function whose minimizer yields the COMET estimate of the signal parameter vector

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} \hat{r}^*\hat{W}^{-1/2}\Pi_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\hat{r} = \\
&= \arg\max_{\theta} \text{vec}^*(I)\Pi_{\hat{W}^{-1/2}\Phi(\theta)}\text{vec}(I) = \\
&= \arg\max_{\theta} \text{vec}^*(I)\hat{W}^{-1/2}\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2}\text{vec}(I) = \\
&= \arg\max_{\theta} \text{vec}^*(\hat{R}^{-1})\Phi(\theta)\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\text{vec}(\hat{R}^{-1})
\end{aligned} \tag{30}$$

where maximizing $\Pi_{\hat{W}^{-1/2}\Phi(\theta)}$ corresponds to minimizing $\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}$. This is solved by using a Newton-type method. Once $\hat{\theta}$ is obtained, this is used to compute $\hat{\sigma}$ and $\hat{\mu}$ in (24).

1.4 Statistical Analysis and CRB

Focus is on analyzing $\hat{\theta}$. The analysis of the COMET estimation shows that the covariance equals the CRB. Important is that the number of samples N needs to be "sufficiently large" for $\hat{R}_s$ and $\hat{Q}$ to be positive definite. There are several long derivations in this section but in the end it will show that they can be simplifed to a simple formula for the CRB.

The CRB of the parameter $\hat{\theta}$ is derived using the compact form of the COMET cost function (27). First consistency of $\hat{\theta}$ is considered. As $N \to \infty$, $\hat{\theta}$ converges to the gloabal minima of asymptotic function corresponding to (30). Using that

$$\begin{aligned} &\hat{W} \to W \text{ as } N \to \infty, \\ &||Pa||^2 = a^T Pa \text{ if } P \text{ projection matrix} \end{aligned} \tag{31}$$

then (27) can be written as

$$\begin{aligned} &\min_\theta \hat{r}^* \hat{W}^{-1/2} \Pi^\perp_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} \hat{r} = \min_\theta ||\Pi^\perp_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} \hat{r}||^2 = \\ &= ||\Pi^\perp_{W^{-1/2}\Phi(\theta_0)} W^{-1/2} r_0||^2 = ||\Pi^\perp_{W^{-1/2}\Phi(\theta_0)} W^{-1/2} \Phi(\theta_0)\alpha_0||^2 \end{aligned} \tag{32}$$

where (θ_0, α_0) denote the true values. The consistency of $\hat{\theta}$ is given by the unique solution assumption.

Here follows the derivation of $\text{cov}(\hat{\theta}) = \text{CRB}$. The COMET estimate is given by de minimizing argument of (27)

$$f(\theta) = \hat{r}^* \hat{W}^{-1/2} \Pi^\perp_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} \hat{r}. \tag{33}$$

Then a Taylor expansion around the derivative of the cost function around $\hat{\theta}$ gives

$$\begin{aligned} &0 = \frac{\partial f(\theta)}{\partial \theta} \approx \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\theta_0} + (\lim_{N\to\infty} \frac{\partial^2 f(\theta)}{\partial \theta^2})|_{\theta=\theta_0}(\hat{\theta} - \theta_0) \\ &\hat{\theta} - \theta_0 = -\left(\left(\lim_{N\to\infty} \frac{\partial^2 f(\theta)}{\partial \theta^2}\right)|_{\theta=\theta_0}\right)^{-1} \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\theta_0} = -H^{-1} g \end{aligned} \tag{34}$$

where $\left(\lim_{N\to\infty} \frac{\partial^2 f(\theta)}{\partial \theta^2}\right)|_{\theta=\theta_0} > 0$ is invertible in a neighbourhood of $\theta = \theta_0$.

To derive (33) the following rule is used: Assume that $P_X^\perp$ is projection matrix spanning the null space of X where $X = X(\theta)$. Then the derivative of $P_X^\perp$ can be derived, for example in [2], as

$$
\begin{aligned}
\frac{\partial P_X^{\perp}}{\partial\theta} &= \frac{\partial}{\partial\theta}\left(I - X(X^*X)^{-1}X^*\right) = -\frac{\partial}{\partial\theta}\left(X(X^*X)^{-1}X^*\right) = \\
&= -\left(\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^* + X\frac{\partial}{\partial\theta}\left((X^*X)^{-1}\right)X^* + X(X^*X)^{-1}\frac{\partial X^*}{\partial\theta}\right) = \\
&= \Big/ \frac{\partial}{\partial\theta}\left((X^*X)^{-1}\right) = -X^{-1}\frac{\partial}{\partial\theta}\left(X^*X\right)X^{-1} \Big/ = \\
&= -\left(\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^* - X(X^*X)^{-1}\frac{dX^*X}{\partial\theta}(X^*X)^{-1}X^* + X(X^*X)^{-1}\frac{\partial X^*}{d\theta}\right) = \\
&= -\left(\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^* - X(X^*X)^{-1}X^*\frac{dX}{\partial\theta}(X^*X)^{-1}X^*\right) + \\
&+ \left(X(X^*X)^{-1}\frac{dX^*}{\partial\theta}X(X^*X)^{-1}X^* + X(X^*X)^{-1}\frac{\partial X^*}{\partial\theta}\right) = \\
&= -\left((I - X(X^*X)^{-1}X^*)\frac{dX}{\partial\theta}(X^*X)^{-1}X^* + X(X^*X)^{-1}\frac{\partial X^*}{\partial\theta}(I - X(X^*X)^{-1}X^*)\right) = \\
&= -\left(P_X^{\perp}\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^* + X(X^*X)^{-1}\frac{\partial X^*}{\partial\theta}P_X^{\perp}\right) = \\
&= -\left(P_X^{\perp}\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^* + \left(P_X^{\perp}\frac{\partial X}{\partial\theta}(X^*X)^{-1}X^*\right)^*\right)
\end{aligned} \tag{35}
$$

With this result the derivative of (33) can be derived as

$$
\begin{aligned}
\frac{\partial f(\theta)}{\partial\theta} &= \frac{\partial}{\partial\theta}\left(\hat{r}^*\hat{W}^{-1/2}\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\hat{r}\right) = \hat{r}^*\hat{W}^{-1/2}\frac{\partial\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}}{\partial\theta}\hat{W}^{-1/2}\hat{r} = \\
&= \hat{r}^*\hat{W}^{-1/2}\left(\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta}\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2} + (\ldots)^*\right)\hat{W}^{-1/2}\hat{r}
\end{aligned} \tag{36}
$$

where $(\ldots)^*$ denotes the term equal to the conjugate transpose of the previous term.

Using that

$$
AXB + (AXB)^* = 2\Re\left(AXB\right)\left(= 2\Re\left(B^*X^*A^*\right)\right) \tag{37}
$$

and (24), then (36) can be written as

$$
\begin{aligned}
&\hat{r}^*\hat{W}^{-1/2}\left(\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta}\left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta)\right)^{-1}\Phi^*(\theta)\hat{W}^{-1/2} + (\ldots)^*\right)\hat{W}^{-1/2}\hat{r} = \\
&= -2\Re\left(\hat{r}^*\hat{W}^{-1/2}\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta}\hat{\alpha}\right) \\
&= -2\hat{r}^*\hat{W}^{-1/2}\Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)}\hat{W}^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta}\hat{\alpha}
\end{aligned} \tag{38}
$$

It can easily be shown that $\frac{\partial f(\theta)}{\partial \theta}$ is real-valued by writing (38) as

$$\begin{aligned}
\frac{\partial f(\theta)}{\partial \theta} &\approx -2\hat{r}^* \hat{W}^{-1/2} \Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha} = \\
&= -2\hat{r}^* \hat{W}^{-1/2} \left(I - \hat{W}^{-1/2}\Phi(\theta) \left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta) \right)^{-1} \Phi^*(\theta)\hat{W}^{-1/2} \right) \hat{W}^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha} = \\
&= -2\hat{r}^* \hat{W}^{-1} \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha} + 2\hat{r}^* \hat{W}^{-1}\Phi(\theta) \left(\Phi^*(\theta)\hat{W}^{-1}\Phi(\theta) \right)^{-1} \Phi^*(\theta)\hat{W}^{-1} \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha} = \\
&= -2\hat{r}^* J^* (J\hat{W}J^*)^{-1} J \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha} + \\
&+ 2\hat{r}^* J^* (J\hat{W}J^*)^{-1} J\Phi(\theta) \left(\Phi^*(\theta) J^* (J\hat{W}J^*)^{-1} J\Phi(\theta) \right)^{-1} \Phi^*(\theta) J^* (J\hat{W}J^*)^{-1} J \frac{\partial \Phi(\theta)}{\partial \theta} \hat{\alpha}
\end{aligned} \tag{39}$$

where all terms are shown to be real as when $\hat{\alpha}$ was shown to be real. The term $J\frac{\partial \Phi(\theta)}{\partial \theta}$ is real since $J\Phi(\theta)$ is real.

The first steps of **the computation of the Hessian** $\lim_{N\to\infty} \frac{d^2}{\partial \theta_k \theta_p} f(\theta)$ **is similar to** (38). Replace $\hat{W}$ by W, $\hat{\alpha}$ by α, $\hat{r} = r = \Phi\theta_0\alpha$ and $\frac{\partial \Phi(\theta)}{\partial \theta}$ by $\frac{\partial \Phi(\theta)}{\partial \theta_k}$ and then derivate the transpose conjugate of (38), which is equal according to (37), with respect to θ_k

$$\begin{aligned}
&\lim_{N\to\infty} \frac{\partial}{\partial \theta_p} \left(-2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta_k} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} r \right) = \\
&= -2\alpha^* \frac{d^2 \Phi^*(\theta)}{\partial \theta_k \partial \theta_p} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} \hat{r} - 2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta_k} W^{-1/2} \frac{\partial}{\partial \theta_p} \left(\Pi^{\perp}_{W^{-1/2}\Phi(\theta)} \right) W^{-1/2} r = \\
&= -2\alpha^* \frac{d^2 \Phi^*(\theta)}{\partial \theta_k \partial \theta_p} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} \hat{r} + 2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta_k} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta_p} \alpha + \\
&+ 2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta_k} W^{-1/2} \left(\hat{W}^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta} \left(\Phi^*(\theta) W^{-1} \Phi(\theta) \right)^{-1} \Phi^*(\theta) W^{-1/2} \right)^* \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} r
\end{aligned} \tag{40}$$

which, when evaluated at $\theta = \theta_0$, $r = \Phi(\theta_0)\alpha$ gives that the first and last term is zero and thus **the Hessian can be written as**

$$\lim_{N\to\infty} \frac{d^2}{\partial \theta_k \theta_p} f(\theta) = \lim_{N\to\infty} \nabla^2 f(\theta) = 2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta_k} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta_p} \alpha. \tag{41}$$

In [1] it is shown that (39) can be simplified since

$$\begin{aligned}
\hat{r}^* \hat{W}^{-1/2} \Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} &= (\hat{r}^* - r) \hat{W}^{-1/2} \Pi^{\perp}_{\hat{W}^{-1/2}\Phi(\theta)} \hat{W}^{-1/2} = \\
&= (\hat{r}^* - r) W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} + o(\frac{1}{\sqrt{N}}) = \mathcal{O}(\frac{1}{\sqrt{N}}).
\end{aligned} \tag{42}$$

which gives that

$$\begin{aligned}
\frac{\partial f(\theta)}{\partial \theta} &\approx -2(\hat{r} - r)^* W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} \frac{\partial \Phi(\theta)}{\partial \theta} \alpha = \\
&= -2\alpha^* \frac{\partial \Phi^*(\theta)}{\partial \theta} W^{-1/2} \Pi^{\perp}_{W^{-1/2}\Phi(\theta)} W^{-1/2} (\hat{r} - r).
\end{aligned} \tag{43}$$

For the following computation of CRB observe that

$$
\begin{aligned}
&\lim_{N\to\infty} E\left[\nabla f(\theta)\nabla f^*(\theta)\right] = \\
&= E\left[4\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}(\hat{r}-r)(\hat{r}-r)^*W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha\right] = \\
&= 4\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}E\left[(\hat{r}-r)(\hat{r}-r)^*\right]W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha = \\
&= 4\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{1}{N}WW^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha = \\
&= 4\frac{1}{N}\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha = \\
&= 4\frac{1}{N}\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha
\end{aligned}
\tag{44}
$$

CRB can be computed as, see [4],

$$
\mathrm{CRB} = \left(\nabla^2 f(\theta_0)\right)^{-1} E\left[\nabla f(\theta_0)\nabla f^*(\theta_0)\right]\left(\nabla^2 f(\theta_0)\right)^{-1} \tag{45}
$$

and

$$
\mathrm{CRB}^{-1} = N\left(\nabla^2 f(\theta_0)\right)\left(NE\left[\nabla f(\theta_0)\nabla f^*(\theta_0)\right]\right)^{-1}\left(\nabla^2 f(\theta_0)\right) \tag{46}
$$

where the element $[\mathrm{CRB}^{-1}]_{kp}$ is written as

$$
\begin{aligned}
&[\mathrm{CRB}^{-1}]_{kp} = N\frac{\partial^2 f(\theta)}{\partial\theta_k\partial\theta_p}\left(NE\left[\frac{\partial f(\theta)}{\partial\theta_k}\frac{\partial f(\theta)}{\partial\theta_p}\right]\right)^{-1}\frac{\partial^2 f(\theta)}{\partial\theta_k\partial\theta_p} = \\
&= N2\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha\left(4\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha\right)^{-1}\times \\
&\times 2\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha = N\alpha^*\frac{\partial\Phi^*(\theta)}{\partial\theta_k}W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha.
\end{aligned}
\tag{47}
$$

When $\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha$ is evaluated at α_0 then

$$
\frac{\partial\Phi(\theta)}{\partial\theta_p}\alpha = \frac{\partial r}{\partial\theta_p} = \frac{\partial\mathrm{vec}(R)}{\partial\theta_p} \tag{48}
$$

which gives the CRB^{-1}

$$
\mathrm{CRB}^{-1} = N\left(\frac{\partial\mathrm{vec}(R)}{\partial\theta}\right)^* W^{-1/2}\Pi^\perp_{W^{-1/2}\Phi(\theta)}W^{-1/2}\frac{\partial\mathrm{vec}(R)}{\partial\theta} \tag{49}
$$

The asymptotical distribution of $\hat{\theta}$ is Gaussian and the covariance matrix equals the CRB. It is worth noting that it requires that N is "sufficiently large" for $\hat{R}$ and $\hat{Q}$ to be positive definite.

References

[1] B. Ottersten and P. Stoica and R. Roy, "Covariance Matching Estimation Techniques for Array Signal Processing Applications", *Digital Signal Processing, 1998.*

[2] *T. V. K. Chaitanya, [PDF] "Notes for Asymptotic Analysis Examples", http://www.commsys.isy.liu.se/ADE/chaitanya-notes.pdf, 2011.*

[3] *R. G. Gallager, [PDF] "Circularly-Symmetric Gaussian random vectors", www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf, 2008*

[4] *E. G. Larsson, [PDF] "Key concepts in asymptotic analysis", http://www.commsys.isy.liu.se/ADE/Notes-asymptotic-analysis.pdf, 2011*