Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals

Petter Wirfält*, Magnus Jansson*

*ACCESS Linnaeus Center, KTH Royal Institute of Technology, Signal Processing, Osqudas v 10, 100 44 Stockholm, Sweden.

Abstract

In this article, we investigate the performance of the recently proposed Direction-Of-Arrival (DOA) estimator POWDER (Prior Orthogonally Weighted Direction Estimator). The method is exploiting a specific form of prior information, namely that some DOAs are known, as well as that the correlation state between some of the source signals are known. In such scenarios, it is desirable to exploit the prior information already in the estimator design such that the knowledge can benefit the estimation of the DOAs of the unknown sources.

Through an asymptotical statistical analysis, we find closed form expressions for the accuracy of the method. We also derive the relevant Cramér-Rao Bound, and we show the algorithm to be efficient under mild assumptions. The realizable performance in the finite sample-case is studied through numerical Monte-Carlo simulations, from which one can conclude that the theoretically predicted accuracies are attained for modest sample sizes and comparatively low SNR. This has the implication that the algorithm is significantly more accurate than other, state-of-art, methods, in a wide range of scenarios.

Keywords: Accuracy, Arrays, Covariance matrix, Direction of arrival estimation, Signal processing algorithms

The research leading to these results has received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement #228044 and the Swedish Research Council under contract 621-2011-5847.

*Corresponding author; tel. +46-8-790-6612; fax +46-8-790-8484

Email addresses: wirfalt@kth.se (Petter Wirfält), janssonm@kth.se (Magnus Jansson)

Preprint submitted to Signal Processing May 4, 2014
1. Introduction

Direction-of-Arrival (DOA) estimation is a classical topic in signal processing and much work has been done in the area in the past decades. A passive array of sensors is receiving signals from a number of distinct sources, and the objective is to estimate the directions these signal are impinging from. In the seminal works [1] and [2], statistically efficient DOA estimation methods are presented. The underlying assumptions for the mentioned methods to be efficient were quite mild, notably, i.i.d. spatially white sensor noise. Thus, the mild assumptions give estimators that are applicable to a wide range of scenarios.

However, many different scenarios exist in practice, and in some of these more restrictive assumptions can be made; for example, some of the source directions might be known a-priori. Exploiting such information in the design of the estimator can be expected to produce methods that are more accurate than [1] and [2], and that this is indeed possible has been shown numerous times, e.g. [3], [4], [5]. Another example of a more restrictive assumption is that it might be known that the source signals are spatially uncorrelated, e.g. [6], [7], and the combination of these two types of prior information was in [8] shown to be beneficial.

Recently, a new DOA algorithm denoted Prior Orthogonally Weighted Direction Estimator (POWDER), was proposed [9], were it was assumed that the known and the unknown DOAs are uncorrelated, but no assumptions were made on the correlation between the signals in the sets of known and unknown signals, respectively. Thus, the method of [9] is applicable in a wider set of scenarios than the one in [8], and it was also shown that for scenarios where both methods are applicable, the former possess better small-sample performance than the latter. Additionally, in other scenarios, POWDER significantly outperformed state-of-the-art methods [1], [5]. One example of a scenario where a corresponding data model is applicable is in wireless communications, where a transmitter at a known location is transmitting a signal which is uncorrelated to the signals emanating from the emitters at the unknown locations. The known location could correspond to, e.g., an interfering base station or wireless access point.

In this article we extend the work in [9] in the following ways: we derive closed form expressions for the asymptotic variances of POWDER; we derive the Cramér-Rao
Bound (CRB) under the particular assumptions studied; we show that under benign conditions the POWDER-method attain the CRB; we conclude the article by, through numerical simulations, investigating the finite-sample, finite-SNR performance of the studied method and the relation to the theoretically derived variance. The POWDER method is theoretically investigated for arbitrary array geometries, and all the results hold for general, unambiguous, arrays. Due to the particularly appealing form of the estimator implementation when a ULA is employed, we however use such a receiver in the numerical simulations.

The article is structured as follows. In Section 2 we revisit the problem formulation given in [9], and in Section 3 we look at the theoretical derivation of POWDER. We derive an expression for the CRB in Section 4, and we also show how to simplify that expression for high SNR. In Section 5 we make a performance analysis of the POWDER-algorithm and we see that, asymptotically in \( N \) and for large SNR, the algorithm obtain the CRB. We conclude the article in Section 6 by numerical Monte-Carlo (MC) experiments, in which we study the performance of the investigated algorithm as well as the applicability of the theoretically derived expression for its accuracy.

Note that we assume, as is common in subspace-based methods, that the number of impinging signals are known, as well as the dimension of the subspace those signals span. See, e.g., [10] for a treatment of the case when such information is unavailable.

We use the following notation: Boldface lowercase (uppercase) letters denote vectors (matrices). The operator \( \otimes \) denotes the Kronecker product, and the operator \( \text{vec}(\mathbf{X}) \) stacks the columns of the matrix \( \mathbf{X} \) into a vector. It can be verified that \( \text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}) \) for matrices of matching dimensions. The superscripts \( T, \quad \mathbf{c}, \quad \mathbf{^*} \) denote transpose, conjugate, and conjugate transpose, respectively. The notation \( \text{Tr}(\mathbf{X}) \) denotes the trace of the matrix \( \mathbf{X} \), i.e. the sum of the diagonal elements in \( \mathbf{X} \). By the symbol \( \triangleq \) we indicate a definition. If a quantity \( X = O(x) \), then \( X/x \) is bounded as \( x \to 0 \), and if \( X = o(x) \), then \( X/x \to 0 \) as \( x \to 0 \). We also use \( O_p(x) \) and \( o_p(x) \), which are the respective in-probability versions [11]. Note that if \( X_n = O_p(a_n) \), \( Y_n = o_p(b_n) \), then \( X_n Y_n = o_p(a_n b_n) \). We denote the Frobenius norm of the matrix \( \mathbf{X} \) by \( \| \mathbf{X} \| \). We define the projection matrix \( \mathbf{\Pi}_\mathbf{X} \triangleq \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \) for any full-rank matrix \( \mathbf{X} \), and the orthogonal projector \( \mathbf{\Pi}_\mathbf{X}^\perp \triangleq \mathbf{I} - \mathbf{\Pi}_\mathbf{X} \).
2. Problem description

Consider the narrow-band signal model (see e.g. [2])

\[ y(t) = A(\hat{\theta})x(t) + n(t), \quad t = 0, 1, \ldots, N - 1. \]  

Here, the vector \( y(t) \in \mathbb{C}^{m \times 1} \) represents the sensor array output, and \( x(t) \in \mathbb{C}^{d \times 1} \) the signal samples, at time \( t \). The matrix \( A(\hat{\theta}) \in \mathbb{C}^{m \times d} \) is the array steering matrix, which is uniquely determined by the array geometry and the (assumed distinct) DOAs \( \hat{\theta} \) of the impinging signals (we reserve \( \theta \) for the unknown DOAs, see below). The dimensions \( m \) and \( d \) correspond to the number of sensors and source signals, respectively. Finally, \( n(t) \in \mathbb{C}^{m \times 1} \) represents the sensor noise. We model both the signal and the noise vectors as zero mean, temporally i.i.d. circularly symmetric complex Gaussian random processes with spatial covariance matrices given by \( \text{cov}(x(t)) = P \) and \( \text{cov}(n(t)) = \sigma^2 I \), respectively.

Using (1) and the definitions above, the sensor output covariance matrix is

\[ R \overset{\Delta}{=} E[y(t)y^*(t)] = APA^* + \sigma^2 I. \]  

The first key assumption in the current article, which delimits it from some well-known state of the art results in this field (e.g. [2]), is that we assume some of the signal directions to be known a-priori; hence we are only interested in estimating \( d_u = d - d_k \) of the DOAs, where the subscripts \( u \) and \( k \) henceforth denote unknown and known, respectively. With that fact in mind we can, without loss of generality, write

\[ \hat{\theta}_0 = \left[ \theta_0^T, \vartheta^T \right]^T; \]  

\[ A(\hat{\theta}_0) = \begin{bmatrix} A(\theta_0) & A(\vartheta) \end{bmatrix} \overset{\Delta}{=} \begin{bmatrix} A_u & A_k \end{bmatrix}; \]  

\[ P = \begin{bmatrix} P_u & P_{uk} \\ P_{uk}^* & P_k \end{bmatrix} \]  

where henceforth \( \theta_0 \) and \( \vartheta \) denote the unknown and the known DOAs, respectively. We further distinguish between \( \theta_0 \), representing the true values of the unknown DOAs, and \( \hat{\theta}_0 \), which is the parametrization of the unknown DOAs. In (5), the subscripts denote the correlation states between the signals emanating from the unknown and known DOAs.

The second defining assumption of this work is

\[ P_{uk} = 0; \]
hence we assume, and exploit, that there is no correlation between the signals from the known and unknown directions. There is no other work known to the authors which have studied this particular instance of the DOA problem; in the DOA scenarios studied in [6] and [12], it was assumed that the source signals were perfectly uncorrelated, i.e. diagonal $P$ — that assumption is a strict subset of the current assumption $P_{uk} = 0$. Accordingly, we do not make any assumptions on $P_k$ or $P_u$; except, we need to know their respective rank and hence introduce the parameters $d'_u = \text{rank}(P_u)$, $d'_k = \text{rank}(P_k)$, and $d' = \text{rank}(P) = d'_u + d'_k$.

We also require that the number of sensors $m$ satisfy

$$m > 2d_u - d'_u$$

$$m \geq d$$  \hspace{1cm} (7)

in order to ensure identifiability of $\theta$; we motivate (7) in Section 5.1.

3. POWDER

3.1. Review of method

Below we study the theoretical properties of the POWDER-method [9]. First we summarize the steps of the estimator. Start by creating the noise-free sensor-output covariance matrix,

$$R - \sigma^2I = \text{APA}^* = \begin{bmatrix} A_u & A_k \\ P_u & P_{uk} & P_{uk}^* & P_k & P_k^* \end{bmatrix} \begin{bmatrix} A_u^* \\ P_{uk}^* \\ P_k^* \end{bmatrix}.$$  \hspace{1cm} (8)

If then (some of) the contribution from the known directions is removed by multiplying with the orthogonal projector $\Pi_{A_k}^\perp$ (such that $A_k^* \Pi_{A_k}^\perp = 0$) from the left, we get

$$(R - \sigma^2I) \Pi_{A_k}^\perp = (A_u P_u + A_k P_{uk}^*) A_k^* \Pi_{A_k}^\perp = A_u P_u A_k^* \Pi_{A_k}^\perp = U_s \Sigma_s V_s^*,$$  \hspace{1cm} (9)

where (9) follows from that $P_{uk} = 0$, and $U_s$, $\Sigma_s$, and $V_s$ follow from the singular value decomposition of (9) (note that only $d'_u$ singular values of $\Sigma_s$ are non-zero). The subscript $s$ denotes the signal subspace (corresponding to the unknown sources). In
In (13), the singular vectors corresponding to the \(d\) largest eigenvalues of \(R\) are collected in \(\hat{V}\) and \(\hat{U}\), and the remaining eigenvectors (and spans the noise subspace) and \(\hat{A}\) are the associated eigenvalues. Now, we can find \(\hat{\sigma}^2 = \frac{1}{m-d'} \text{Tr}(\hat{A}_n)\), which is the ML-estimate of \(\sigma^2\) given the data model in Section 2, [13]. In practice the realization of (10) thus becomes:

\[
\hat{R} = \hat{E}_s \hat{A}_s \hat{E}_s^* + \hat{E}_n \hat{A}_n \hat{E}_n^*,
\]

where \(\hat{E}_s\) contains the eigenvectors associated with the \(d'\) largest eigenvalues of \(\hat{R}\) and \(\hat{A}_s\) is the matrix containing these eigenvalues on the diagonal. Similarly, \(\hat{E}_n\) contains the remaining eigenvectors (and spans the noise subspace) and \(\hat{A}_n\) the associated eigenvalues.

Using (13) and the fact that \(\hat{V}_s\) and \(\hat{V}_n\) form an orthonormal basis,

\[
\hat{U}_s = \left(\hat{R} - \hat{\sigma}^2 I\right) \Pi_{\hat{A}_s} \hat{V}_s \Sigma_s^{-1}.
\]

We now introduce the matrix \(B(\theta) \in \mathbb{C}^{m \times (m-d_n)}\), which (for \(\theta = \theta_0\)) spans the null space of \(A^*_n\) (i.e. \(B^*(\theta_0)A_n = 0\)), such that

\[
B^*(\theta_0)\hat{U}_s = B^*(\theta_0) \left(\hat{R} - \hat{\sigma}^2 I\right) \Pi_{\hat{A}_s} \hat{V}_s \Sigma_s^{-1} = B^*(\theta_0) \left(\hat{R} - \sigma^2 I\right) \Pi_{\hat{A}_s} \hat{V}_s \Sigma_s^{-1} + o_p(1/\sqrt{N}),
\]

since \(\hat{R} \triangleq \hat{R} - R = O_p(1/\sqrt{N})\), \(\sigma^2 \triangleq \sigma^2 - \sigma^2 = O_p(1/\sqrt{N})\), \(B^*(\theta_0)(R - \sigma^2 I)\Pi_{\hat{A}_s} = 0\), \(\hat{V}_s = V_s + o_p(1)\), and \(\hat{\Sigma}_s = \Sigma_s + o_p(1)\). Further, \(\text{Span}(V_s) \subseteq \text{Span}(\Pi_{\hat{A}_s})\) and hence \(\Pi_{\hat{A}_s} V_s = V_s\); then we can re-write (15) as

\[
B^*(\theta_0)\hat{U}_s = B^*(\theta_0) \left(\hat{R} - \hat{\sigma}^2 I\right) \Pi_{\hat{A}_s} \hat{V}_s \Sigma_s^{-1} + o_p(1/\sqrt{N}).
\]

By forming the residual error vector

\[
e(\theta) \triangleq \text{vec} \left(\hat{U}_s^* B(\theta)\right),
\]

we find

\[
\hat{R} = \frac{1}{N} \sum_{t=1}^{N} y(t)y^*(t).
\]
a cost function can be created:

\[ V(\theta) = e^*(\theta) W e(\theta), \]

(18)

where \( W > 0 \) is a positive definite (p.d.) weighting matrix. In [9], estimates of \( \theta_0 \) were found as

\[ \hat{\theta}_{\text{POW}} = \arg \min_{\theta} V(\theta). \]

(19)

It is in (18) implicitly assumed that \( E[\epsilon \epsilon^T] = 0 \), i.e. that \( \epsilon \) is circularly symmetric; this is not necessarily the case, implying that the target function (18) might not capture the full statistics of the problem. We show in Section 3.3 that \( \epsilon \) is circularly symmetric for high SNR, and that the choice (18) thus is asymptotically correct. By creating a cost function based on the extended residual \( \bar{\epsilon} = [\epsilon^T \hspace{1cm} \epsilon^*]^T \), it is possible to form a criterion that takes the non-circularly symmetric aspects of \( \epsilon \) into account. However, numerical simulations have shown that there is practically no benefit even in low-SNR scenarios from exploiting the extended residual as compared to the one given by (17). Hence, the focus will be on (19) in this paper.

For general array models, (19) is typically solved by a Newton-like algorithm. Note that both the gradient and the Hessian of the cost function (18), as needed in such an implementation, are given in Appendix D. For a ULA, a particularly attractive parametrization of \( \theta \) in (19) exists which is discussed in the next section.

3.2. Implementation for uniform linear array

When the receiving array is a uniform linear array (ULA) it is possible to rewrite the minimization problem in a particularly appealing form. This is due to that \( A_u \) in this case has a Vandermonde-structure; we can write its \( l \)th column (for half wave-length intra sensor spacing) as

\[ a(\theta_l) = \begin{bmatrix} 1 \\ \exp(-j\pi \sin(\theta_l)) \\ \vdots \\ \exp(-j(m-1)\pi \sin(\theta_l)) \end{bmatrix}, \]

(20)
in which \( j^2 = -1 \) and \( l = 1, \ldots, d_u \). Then, the matrix \( B \) can be written as [14]

\[
B^\top = \begin{bmatrix}
  b_0 & b_1 & \cdots & b_{d_u} & 0 \\
   \cdots & \cdots & \cdots & \cdots & \cdots \\
  0 & b_0 & b_1 & \cdots & b_{d_u}
\end{bmatrix},
\]

(21)

where \( b_i \) are given as the coefficients of the polynomial

\[
b_0 \prod_{l=1}^{d_u} (z - e^{-j\pi \sin(\theta_l)}) \triangleq b_0 z^{d_u} + b_1 z^{d_u-1} + \ldots + b_{d_u}. \tag{22}
\]

We collect the polynomial coefficients of (22) according to \( b^\top = \begin{bmatrix} b_0 & b_1 & \cdots & b_{d_u} \end{bmatrix}^\top \).

Now introduce the selection matrix \( \Psi_1 \) such that \( \text{vec}(B) = \Psi_1 b \). Then, we can write

\[
\epsilon = \text{vec} \left( \hat{U}_s^* B \right) = \left( I \otimes \hat{U}_s^* \right) \text{vec}(B) = \hat{\Psi} \epsilon,
\]

(23)

where \( \hat{\Psi} \triangleq \left( I \otimes \hat{U}_s^* \right) \Psi_1 \). With (23), (18) can be re-written as a function of \( b \) according to

\[
V(\theta) = b^*(\theta) \Psi^* W \Psi b(\theta).
\]

(24)

The minimization of \( V(\theta) \) with respect to \( b(\theta) \) can be performed as an eigenvalue problem for \( b(\theta) \). The DOAs \( \theta \) are then estimated by rooting (22), with estimated coefficients from (24).

For some technical details regarding the minimization of (24) with respect to \( b \), please see [15].

3.3. Statistical properties of the residual

We now explore the statistical properties of the residual (17), and derive the optimal choice of the weighting matrix \( W \) in (18).

**Theorem 1.** Given the data model and assumptions in Section 2, the following expressions hold:

\[
C_\epsilon \triangleq \lim_{N \to \infty} N \mathbb{E} \left[ \epsilon \epsilon^\top \right] = MF (R^\top \otimes R) F^* M^*,
\]

(25)

\[
C_\zeta \triangleq \lim_{N \to \infty} N \mathbb{E} \left[ \epsilon^\top \epsilon \right] = MF (R^\top \otimes R) L_{mm} F^T M^T,
\]

(26)
where
\[ M \triangleq (B^T \otimes \Sigma_s^{-1} V_s^+), \] (27)
\[ F \triangleq \left( I_m - \frac{1}{m-d^p} \text{vec}(I_m) \pi^* \right), \] (28)
\[ \pi \triangleq \text{vec}(\Pi_{E_s}), \] (29)
and the permutation matrix \( L_{mm} \) is defined such that
\[ L_{mm} \text{vec} p \hat{m} q = \text{vec} p X_T \hat{m} q. \]

Further,
\[ C_\epsilon = \tilde{C}_\epsilon + O(\sigma^4), \] (30)
\[ C_{\tilde{\epsilon}} = O(\sigma^4), \] (31)
in which
\[ \tilde{C}_\epsilon \triangleq (B^T R^T B \otimes \Sigma_s^{-1} V_s^+ \Sigma_s^{-1}) > 0. \] (32)

Proof. See Appendix A. \( \square \)

It is well known (see e.g. [16]) that the criterion (19) gives minimum variance estimates (for the given residual) if the weighting matrix \( W \) is chosen as (a scaled version of) the inverse covariance of the residual; accordingly, we choose
\[ W_{\text{OPT}} = \sigma^2 C_{\epsilon}, \] (33)
where the scaling \( \sigma^2 \) is introduced in order to keep the weighting matrices finite for small \( \sigma^2 \).

We will explore the small sample, low SNR, properties of the optimally weighted estimator in Section 6.

From Theorem 1, it can be seen that \( W_{\text{OPT}} \), through \( M \), is a function of \( B \). Based on the consistency analysis in Section 5.1, it can however be seen that estimates of \( \theta_0 \) are consistent for any p.d. choice of \( W \); hence initial estimates of \( \theta_0 \) can be used to construct (consistent) estimates of \( W_{\text{OPT}} \).

Further, \( N \) is in practice finite; thus \( C_\epsilon \) must be estimated from sample data. From an analysis similar to the one in (15), it can be deduced that \( \tilde{W}_{\text{OPT}} \triangleq \sigma^2 \tilde{C}_{\epsilon}^{-1} = W_{\text{OPT}} + o_p(1) \); hence, \( \tilde{W}_{\text{OPT}} \) is a consistent estimate of \( W_{\text{OPT}} \) and we can use \( \tilde{W}_{\text{OPT}} \) in (19) without impairing the asymptotic properties of the estimator.
4. Cramér-Rao bound

Since POWDER is designed to exploit a block-diagonal structure of the source covariance matrix $P$, it is of interest to compare its accuracy with the theoretically best achievable performance under such assumptions. Let

$$\alpha \triangleq [\theta^T \ \theta^T \ \sigma^2]^T$$

(34)

denote the vector of unknown parameters in the model; we thus have $\varrho = [\varrho_k^T \ \varrho_u^T]^T$ with $\varrho_k$ being the vector made from $\{[P_k]_{ii}\}$ and $\{\text{Re} \ ([P_k]_{ij}); \text{Im} \ ([P_k]_{ij}); \ i > j\}$, and $\varrho_u$ is made from $P_u$ in a similar fashion. The primary interest herein is an accuracy bound for the DOAs $\theta$. Hence, a bound on the relevant subset of $\alpha$ is found in the following theorem.

**Theorem 2.** Given the assumptions in Section 2,

$$\text{CRB}_{\theta} = \frac{1}{N} \left( \mathbf{G}^* \Pi_{\Delta} \mathbf{G} \right)^{-1},$$

(35)

where we by $\theta$ signify that the bound is only on the DOAs, but the entire $\alpha$ is unknown.

In (35), we have

$$\mathbf{G} \triangleq \mathbf{C}_r^{1/2} \frac{\partial \mathbf{r}}{\partial \theta^T},$$

(36)

where $\mathbf{r} = \text{vec}(\mathbf{R})$, $\mathbf{C}_r^{1/2} \triangleq \mathbf{R}^{-T/2} \otimes \mathbf{R}^{-1/2}$, and

$$\Delta \triangleq \mathbf{C}_r^{1/2} \left[ (A_k^c \otimes A_k) \ (A_u^c \otimes A_u) \ \text{vec}(I) \right].$$

(37)

**Proof.** See Appendix B

It is possible to simplify (35) under more restrictive assumptions.

**Theorem 3.** Given the assumptions in Section 2 and the additional condition $\lambda_{\min}^{-1}(P_k) \to 0$, we have that

$$\text{CRB}_{\theta} = \text{CRB}_{\theta} + O(\|P_k^{-2}\|),$$

(38)

where

$$\overline{\text{CRB}}_{\theta} \triangleq \frac{1}{2N} \left[ \text{Re} \left( \left( \mathbf{D}^* \mathbf{R}^{-1/2} \Pi_{\Delta_{uu}} \mathbf{R}^{-1/2} \mathbf{D} \right) \otimes \left( \mathbf{P}_u^T \mathbf{A}_{uu}^c \mathbf{A}_{uu}^c \mathbf{P}_u^c \right) \right) \right]^{-1},$$

(39)
D is given in (C.23), \( \mathbf{A}_{uw} = \mathbf{R}^{-1/2} \mathbf{A_u} \), and \( \odot \) is the Hadamard (or Schur) product, i.e. element-wise multiplication.

**Proof.** See Appendix C. \( \square \)

As will be seen in Section 6, the additional assumption required in Theorem 3 is in practice not so restrictive; for example, the assumption of \( \lambda_{\min}^{-1} (\mathbf{P}_k) \to 0 \) may be valid already at moderate SNRs of the sources in the known directions.

## 5. Performance analysis

In this section we analyze the asymptotic properties of the estimator given by (19).

We now establish the consistency of the estimator and also find the asymptotic distribution of the errors in the estimates.

### 5.1. Consistency of estimates

The cost function (18) converges, asymptotically in \( N \) and uniformly in \( \theta \), with probability 1 to

\[
V_x(\theta) = \lim_{N \to \infty} V(\theta) = \text{vec}^* (\mathbf{U}_s^* \mathbf{B}(\theta)) \mathbf{W} \text{vec} (\mathbf{U}_s^* \mathbf{B}(\theta)).
\]

Since \( \mathbf{W} \) in (40) is p.d. (by definition), \( V_x \geq 0 \) with equality if and only if \( \mathbf{U}_s^* \mathbf{B}(\theta) = 0 \). Now we establish that

\[
V_x(\theta) = 0 \implies \theta = \theta_0
\]

under the conditions in Section 2.

From [17], (41) follows for \( m > 2d_u - d_u' \) if \( \mathbf{U}_s \) from (10) can be written

\[
\mathbf{U}_s = \mathbf{A}_u(\theta_0) \mathbf{T},
\]

for some full rank matrix \( \mathbf{T} \in d_u \times d_u' \). Since \( \mathbf{U}_s \) is defined from the singular value decomposition of \( \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \mathbf{\Pi}_{\mathbf{A}_k}^\perp \), it follows that

\[
\text{rank}(\mathbf{T}) = \text{rank}(\mathbf{P}_u \mathbf{A}_u^* \mathbf{\Pi}_{\mathbf{A}_k}^\perp).
\]
Using Frobenius’ rank inequality (e.g., [18]),
\[
\text{rank}(T) \geq \text{rank}(P_uA_u^* + \text{rank}(A_u^*\Pi_{\hat{A}_k}^\perp) - \text{rank}(A_u^*). \tag{44}
\]

Now we note the following: if \(m \geq d\), \(A(\theta_0) = \begin{bmatrix} A_u & A_k \end{bmatrix}\) is full rank and then the Schur-complement of \(A_k\) in \(A^*A\), \(A_u^*\Pi_{\hat{A}_k}^\perp A_u > 0\); hence, \(A_u^*\Pi_{\hat{A}_k}^\perp\) is of full rank \(d_u\). Thus, for \(m \geq d\), we have that
\[
\text{rank}(T) \geq d'_u + d_u - d_u, \tag{45}
\]
or, equivalently, \(\text{rank}(T) = d'_u\). Hence, (41) is established; thus the estimates \(\hat{\theta}\) from (19) converges to \(\theta_0\) with probability 1 as \(N \to \infty\).

### 5.2. Asymptotic distribution

To find the asymptotic variance of the estimates, we perform a Taylor expansion of
\[
\frac{\partial V}{\partial \theta} \bigg|_{\theta = \hat{\theta}} \triangleq V'(\hat{\theta})
\]
around \(\theta_0\):
\[
0 = V'(\hat{\theta}) = V'(\theta_0) + \frac{\partial V'}{\partial \theta} \bigg|_{\theta = \theta_0} (\hat{\theta} - \theta_0) + O_p(|\hat{\theta} - \theta_0|). \tag{46}
\]

If we define
\[
\text{H} \triangleq \lim_{N \to \infty} \frac{\partial V'}{\partial \theta} \bigg|_{\theta = \theta_0}, \tag{47}
\]
and let \(\tilde{\theta} = \hat{\theta} - \theta_0\), we have (for small errors \(\tilde{\theta}\), and asymptotically in \(N\)) from (46) that
\[
\tilde{\theta} = -\text{H}^{-1}V'(\theta_0). \tag{48}
\]

Then we can, by defining
\[
\text{Q} \triangleq \lim_{N \to \infty} \mathbb{E} \left[ V'(\theta_0)V'^*(\theta_0) \right], \tag{49}
\]
formulate the following theorem.

**Theorem 4.** Given the assumptions of Section 2, the POWDER estimates asymptotically in \(N\) satisfy
\[
\sqrt{N}(\hat{\theta}_{\text{POW}} - \theta_0) \sim \mathcal{N}(0, C_{\text{POW}}), \tag{50}
\]
where \(C_{\text{POW}} = \text{H}^{-1}\text{QH}^{-1}\), and \(\text{Q}\) and \(\text{H}\) are given by (D.10) and (D.13), respectively.
Proof. The Gaussianity of (50) follows from the fact that $\sqrt{N} \text{vec}(\hat{\mathbf{R}} - \mathbf{R})$ is asymptotically Gaussian (due to the central limit theorem), together with (48), (D.9), and (A.3). That (50) is zero-mean follows from the consistency proven in Section 5.1. For the derivation of $\mathbf{C}_{\text{POW}}$, see Appendix D.

When using the optimal weighting matrix (33), we can formulate a stronger result given by the following theorem.

**Theorem 5.** Given the assumptions in Section 2 we have that, when using the optimal weighting matrices (33),

$$\mathbf{C}_{\text{POW}} = N \cdot \bar{\mathbf{CRB}}_{\theta} + O(\sigma^4),$$

(51)

and $\bar{\mathbf{CRB}}_{\theta}$ is given by (39).

**Proof.** See Appendix E

Thus we have shown that for large $N$ and high SNR, POWDER is optimal in the sense that a higher accuracy is not obtainable. We will in Section 6 examine the applicability of the respective expressions (50) and (51) as predictors of estimator accuracy for finite $N$ and SNR.

### 6. Numerical examples

In this section we investigate the practical accuracy of the investigated estimators through Monte-Carlo (MC) simulations, as well as the asymptotic accuracy results (50) and (51). In the MC-simulations, we generate pseudo-random Gaussian noise sequences $\mathbf{x}(t)$ and $\mathbf{n}(t)$ with the prescribed statistics. We have two unknown and two known sources: $\boldsymbol{\theta}_0 = \begin{bmatrix} 10^\circ & 15^\circ \end{bmatrix}^T$, and $\boldsymbol{\vartheta} = \begin{bmatrix} 12^\circ & 20^\circ \end{bmatrix}^T$. In order to streamline the presentation, we control the source powers by the two parameters $\text{SNR}_u$ and $\text{SNR}_k$, where the subscripts refer to unknown and known, respectively. Thus, we let the signals from each set of emitters be equipowered, according to

$$P_u = \text{SNR}_u \begin{bmatrix} 1 & \rho_u \\ \rho_u & 1 \end{bmatrix}, \quad P_k = \text{SNR}_k \begin{bmatrix} 1 & \rho_k \\ \rho_k & 1 \end{bmatrix},$$

(52)
where $\rho_k$ and $\rho_u$ are parameters. We satisfy the assumption (6), except in the scenario depicted in Fig. 4. The receiving array is consistently a ULA; its number of sensors $m = 10$ unless otherwise specified.

To evaluate the practical performance of the investigated estimators, we use the Root-Mean-Square Error (RMSE) as performance metric, which we define as

$$\text{RMSE}_i = \sqrt{\frac{1}{L} \sum_{l=1}^{L} \left( \theta^{(l)}_i - \hat{\theta}_{0,i} \right)^2},$$

(53)

where the superscript $(l)$ denotes the $l$th MC-realization, and $L$ is the number of MC-realizations.

6.1. Performance comparison to existing state-of-art

In Fig. 1 we compare simulated performance of POWDER to “MODE”, [1], (or, equivalently, WSF [2]) which is not using prior knowledge on source DOA or correlation, and “PLEDGE” [8], which exploits the knowledge of $\vartheta$ but not the block-diagonal structure of $P$. Two CRBs are displayed: “CRB$^P$”, [8], assuming that $\vartheta$ is known, and “CRB$^{BD}$”, given by (35).

Here, all sources are of the same power, thus $\text{SNR}_u = \text{SNR}_k = \text{SNR}$. As can be seen, in the studied scenario PLEDGE is marginally better than MODE for low SNR; at higher SNR, both methods are equivalent, and efficient. However, it is also seen that POWDER is consistently very much more accurate, and sooner (in terms of SNR) efficient. Hence, a significant decrease in SNR is possible for a given accuracy. It can be concluded that at high SNR, knowing $\vartheta$ (without correlation knowledge) becomes irrelevant as MODE and PLEDGE give the same accuracy. POWDER, however, seems to consistently benefit from its prior knowledge.

The figure is also showing the validity of the derived expression (50).

6.2. Asymptotic variance validity

In Fig. 2 we investigate the significance of the assumptions of Theorem 3, and study the empirical performance in comparison to the expressions (50), (51) and (35). In the figure, “CRB” denotes the exact expression (35), and “CRB$^{HSNR}$” (39).
In Fig. 2a, we have fixed the known sources’ power at $\text{SNR}_k = -10\,\text{dB}$. We can see in the figure that for large $\text{SNR}_u$, the respective bounds converge even though $\text{SNR}_k$ is fixed.

In Fig. 2b we study the opposite scenario; we fix $\text{SNR}_u = 5\,\text{dB}$, and vary $\text{SNR}_k$. We can thus see that while the estimation scenario is more difficult for high $\text{SNR}_k$, the assumptions of Theorem 3 are eventually satisfied and the bounds converge. However, for the higher $\text{SNR}_k$, $\text{SNR}_u$ is too small and the estimators break down.

Additionally, in Fig. 3 we study the scenario when all sources are uncorrelated, and $\text{SNR}_u = \text{SNR}_k = \text{SNR}$; it is then possible to use methods relying on the information that $\mathbf{P}$ is diagonal. Hence in this figure we also include “PLEDGE UC”, [12], which is tailored for the scenario with known sources and diagonal $\mathbf{P}$. Additionally, we include the corresponding bound, “CRB$_{PUC}$”, [8]. We observe several interesting phenomena in Fig. 3: even though there is a strict ordering in the sense that $\text{VAR}_{\text{POW}} > \text{CRB} > \text{CRB}_{\text{PUC}}$ for $\sigma^2 > 0$, the difference is in this scenario practically indistinguishable. Further, POWDER has better low-SNR properties than PLEDGE UC, and outperforms PLEDGE UC in the studied scenario even though POWDER does not exploit the fact that $\mathbf{P}$ is strictly diagonal.

6.3. Robustness to modeling errors

In this simulation scenario we study the effects of violating the assumption that the signals from the known and unknown DOAs are uncorrelated. Thus we let $\mathbf{P}_{uk} = \rho_{uk} \mathbf{1}_2 \mathbf{1}_2^T$, where we vary the parameter $\rho_{uk}$ according to Fig. 4, and $\mathbf{1}_2$ denotes the vector of all ones of length 2. The SNR is fixed at 15dB for both sets of DOAs, $\rho_u = \rho_k = 0.9$, and we plot against the number of obtained data samples $N$, in order to study finite sample effects. In addition to the RMSEs for POWDER for various choices of $\rho_{uk}$, we also show the CRB for the case $\mathbf{P}_{uk} = \mathbf{0}$, i.e. the ideal scenario the method is designed for.

From Fig. 4, and by analyzing the individual MC-estimates, it can be seen that the modeling error introduces a bias in the estimates of the unknown DOAs. As $\rho_{uk}$ grows larger, the method increasingly suffers from not being able to distinguish between the known and unknown DOAs, cf. (10). POWDER appears to be fairly robust to the induced modeling error, in the sense that the estimator does not completely fail. However, as the
error due to finite-samples decrease with increasing $N$, the error due to the induced bias limit the estimation accuracy.

The case of modeling errors in the known angles has been treated in e.g. [19]; a similar analysis of the herein studied estimator can be performed but is out of scope of the present article.

7. Conclusions

In this article we have analyzed the asymptotic performance of the recently proposed method [9] POWDER. We have shown that the method, asymptotically in $N$ and SNR, is statistically efficient. We also explored practical estimation results using Monte Carlo simulations, which showed the theoretically predicted results to be accurate, and we investigated robustness to modeling errors. Our simulations indicate that POWDER obtains its asymptotical (in $N$) accuracy already at moderate sample sizes and low SNR. Additionally, we saw a graceful convergence to the (proven) CRB-equivalent accuracy for moderate SNR.

Appendix A.

First, derive an expression for $C_t = \lim_{N \to \infty} N E \left[ \epsilon \epsilon^* \right]$. To do so, re-write (17) based on (16):

\[ \epsilon = \text{vec} \left( \tilde{U}^* B \right) = \left( B^T \otimes \Sigma_s^{-1} V^*_s \right) \text{vec} \left( \tilde{R} - \tilde{\sigma}^2 I \right) + o_p(1/\sqrt{N}) \]  

(A.1)

From, e.g., [7]:

\[ (m - d')\tilde{\sigma}^2 = \text{Tr}(\tilde{A}_n) - \text{Tr}(A_n) \]
\[ = \text{vec}^*(I_m - E_s E_s^*) \text{vec}(\tilde{R}) + o_p(1/\sqrt{N}) \]
\[ \overset{\Delta}{=} \pi^* \text{vec}(\tilde{R}) + o_p(1/\sqrt{N}), \]  

(A.2)

since $E_s^* E_s = I_{d'}$, and $\pi$ is defined in (29). Using (A.2) in (A.1), with the definitions (27) and (28), we have

\[ \epsilon = MF \text{vec}(\tilde{R}) + o_p(1/\sqrt{N}). \]  

(A.3)
From, e.g., [20], we have that
\[ E(\text{vec}(\bar{R}) \text{vec}^*(\bar{R})) = \frac{1}{N} (R^T \otimes R). \] (A.4)

Using (A.3) together with (A.4) gives
\[ \lim_{N \to \infty} N E \left[ \epsilon \epsilon^* \right] = MF (R^T \otimes R) F^* M^*. \] (A.5)

Expand (A.5) to read
\[ C_\epsilon = \left( M - \frac{1}{m - d'} \text{Mi} \pi^* \right) (R^T \otimes R) \left( M^* - \frac{1}{m - d'} \pi i^* M^* \right), \] (A.6)

where \( i = \text{vec}(I_m) \). Evaluating (A.6) we will have four terms. The first one is
\[ C_1 = M (R^T \otimes R) M^* = (B^T R^T B^* \otimes \Sigma_s^{-1} V_s^* R V_s \Sigma_s^{-1}), \] (A.7)

which is guaranteed to be p.d. for \( \sigma^2 > 0 \) since the constituent matrices of each Kronecker factor then are p.d. The second term in (A.6) is given by
\[ C_2 = -\frac{1}{m - d'} \text{Mi} \pi^* (R^T \otimes R) M^*. \] (A.8)

Note that from (A.8)
\[ \pi^* (R^T \otimes R) M^* = \text{vec}^* (R(I - E_n E_n^*) R) M^* = \text{vec}^* (E_n A_n^2 E_n^*) M^* = \sigma^4 \text{vec}^* (\Sigma_s^{-1} V_s^* E_n E_n^* B), \] (A.9)

where the second equality follows from the EVD of \( R \) and the third from (27). By noting that \( Mi = \text{vec}(\Sigma_s^{-1} V_s^* B) \), we have that (A.8) evaluates to
\[ C_2 = -\frac{\sigma^4}{m - d'} \text{vec}(\Sigma_s^{-1} V_s^* B) \text{vec}^* (\Sigma_s^{-1} V_s^* E_n E_n^* B). \] (A.10)

Obviously, \( C_3 = C_2^* \).

It is easily verified that \( \pi^* (R^T \otimes R) \pi = \sigma^4 \text{Tr}(E_n E_n^*) = \sigma^4 (m - d') \), which gives
\[ C_4 = \frac{1}{(m - d')^2} \text{Mi} \pi^* (R^T \otimes R) \pi i^* M^* = \frac{\sigma^4}{m - d'} \text{vec}(\Sigma_s^{-1} V_s^* B) \text{vec}^* (\Sigma_s^{-1} V_s^* B). \] (A.11)

Hence we can write
\[ C_\epsilon = C_1 + O(\sigma^4), \] (A.12)

which establishes (30).

To continue, we need the following Lemma:
Lemma 1. For a permutation matrix $L_{mn}$ such that

$$L_{mn} \text{vec}(X_{m \times n}) = \text{vec}(X_{m \times n}^T),$$  \hspace{1cm} (A.13)

it holds that

$$L_{mn} (X_{n \times d} \otimes Y_{m \times p}) = (Y_{m \times p} \otimes X_{n \times d}) L_{pd}. \hspace{1cm} (A.14)$$

Proof. Write

$$L_{mn} (X_{n \times d} \otimes Y_{m \times p}) \text{vec}(Z_{p \times d}) = L_{mn} \text{vec}(YZX^T) = \text{vec}(XZ^TY^T)$$

$$= (Y \otimes X) \text{vec}(Z^T)$$

$$= (Y \otimes X) L_{pd} \text{vec}(Z); \hspace{1cm} (A.15)$$

since $Z$ is arbitrary, (A.14) follows. \hfill \Box

We will in the following omit the dimensional marker of these permutation matrices to simplify notation. Thus, if we write $\text{vec}^T(\hat{R}) = \text{vec}^*(\hat{R}^T) = \text{vec}^*(\hat{R}) L$ (since $\hat{R}$ is Hermitian), we have from (A.4) that

$$E(\text{vec}(\hat{R}) \text{vec}^T(\hat{R})) = N^{-1} (R^T \otimes R) L; \hspace{1cm} (A.16)$$

analogously to (A.5), we have

$$C_{\hat{\epsilon}} = M F (R^T \otimes R) L F^T M^T \hspace{1cm} (A.17)$$

$$= M (R^T \otimes R) L M^T + O(\sigma^4), \hspace{1cm} (A.18)$$

where the last equality follows from (28), (A.14), and (A.12). Now,

$$LM^T = L (B \otimes V_s^c \Sigma_s^{-1}) = (V_s^c \Sigma_s^{-1} \otimes B) L, \hspace{1cm} (A.19)$$

again using (A.14). Using (A.19) in (A.18),

$$M (R^T \otimes R) LM^T = (B^T R^T V_s^c \Sigma_s^{-1} \otimes \Sigma_s^{-1} V_s^* R B) L = O(\sigma^4), \hspace{1cm} (A.20)$$

since $V_s^* R B = \sigma^2 V_s^* B$, which is due to $B^* A_u = 0$ and $V_s = \Pi_A \hat{V}_s$. Hence,

$$C_{\hat{\epsilon}} = O(\sigma^4). \hspace{1cm} (A.21)$$
Appendix B.

This proof follows the methodology of [21] (see also [7]), in which the general scenario without prior knowledge of $\theta$ or $P_{uk} = 0$ was studied. The Fisher Information Matrix (FIM) for the parameter vector $\alpha$ is given by

$$\frac{1}{N} \text{FIM} = \left( \frac{\partial \mathbf{r}}{\partial \alpha^T} \right)^* \left( R^{-T} \otimes R^{-1} \right) \left( \frac{\partial \mathbf{r}}{\partial \alpha^T} \right), \quad (B.1)$$

where

$$\mathbf{r} = \text{vec} \left( A_u P_u A_u^* \right) + \text{vec} \left( A_k P_k A_k^* \right) + \sigma^2 \text{vec} (I), \quad (B.2)$$

due to $P$ being block-diagonal. If we let

$$\left[ \mathbf{G} \mid \Delta \right] \overset{\Delta}{=} C_r^{1/2} \left[ \frac{\partial \mathbf{r}}{\partial \theta^T} \mid \frac{\partial \mathbf{r}}{\partial \theta^T} \frac{\partial \mathbf{r}}{\partial \sigma^2} \right], \quad (B.3)$$

we can, by comparing (B.3) to (B.1), write

$$\frac{1}{N} \text{FIM} = \left[ \mathbf{G}^* \mid \Delta^* \right] \left[ \mathbf{G} \mid \Delta \right]. \quad (B.4)$$

The respective elements of (B.3) are found as (for $i = 1, \ldots, d_u$)

$$\frac{\partial \mathbf{r}}{\partial \theta_i} = \left( A_u^* P_u^c \otimes I \right) d_i + \left( I \otimes A_u P_u \right) L d_i, \quad (B.5)$$

where

$$d_i = \frac{\partial \text{vec}(A_u)}{\partial \theta_i}, \quad (B.6)$$

and $L$ follows from Lemma 1. Collecting $\bar{\mathbf{D}} = \begin{bmatrix} d_1 & \cdots & d_{d_u} \end{bmatrix}$, we have that

$$\frac{\partial \mathbf{r}}{\partial \theta^T} = \left( A_u^* P_u^c \otimes I \right) \bar{\mathbf{D}} + \left( I \otimes A_u P_u \right) L \bar{\mathbf{D}}^c. \quad (B.7)$$

Continuing with the other terms of (B.3),

$$\frac{\partial \mathbf{r}}{\partial \sigma^2} = \left[ \left( A_k^* \otimes A_k \right) J_k \mid \left( A_u^* \otimes A_u \right) J_u \right], \quad (B.8)$$

where $J_k$ is the selection matrix such that $\text{vec}(P_k) = J_k \theta_k$, and $J_u$ is defined from $\text{vec}(P_u) = J_u \theta_u$. Finally, $\partial \mathbf{r} / \partial \sigma^2 = \text{vec}(I)$.
Since the interest is in a lower bound for the DOAs $\theta$ (and not a bound for all the parameters $\alpha$), a standard result on block-matrix inversion on (B.4) gives
\[
\text{CRB}_\theta = \frac{1}{N} \left( G^* \Pi^\dagger_G G \right)^{-1},
\] (B.9)
where we have also exploited that since $J_k$ and $J_u$ are invertible, $\text{Span}(\Delta) = \text{Span}(\bar{\Delta})$ (and $\Delta$ is defined in (37)).

Appendix C.

Start by partitioning $\bar{\Delta} = \begin{bmatrix} \bar{v} & \bar{r} \end{bmatrix}$, where we define
\[
\bar{v} \triangleq \begin{bmatrix} A_k & A_u \end{bmatrix}, \\
\bar{A}_k \triangleq C^{1/2}_\nu (A^*_k \otimes A_k), \\
\bar{A}_u \triangleq C^{1/2}_\nu (A^*_u \otimes A_u), \\
r_i \triangleq \text{vec}(R_i^{-1}).
\]
Then it is possible to decompose the projection operator $\Pi^\dagger_{\bar{\Delta}}$ (see e.g. [22]) according to
\[
\Pi^\dagger_{\bar{\Delta}} = \Pi^\dagger_{\bar{v}} - \frac{\Pi^\dagger_{\bar{r}} r_i^* \Pi^\dagger_{\bar{v}}}{r_i \Pi^\dagger_{\bar{v}} r_i}. 
\] (C.1)
Similarly, we decompose $\Pi^\dagger_{\bar{v}}$ in (C.1) according to
\[
\Pi^\dagger_{\bar{v}} = \Pi^\dagger_{A_u} + \Pi^\dagger_{A_k} A_k \left( A^*_k \Pi^\dagger_{A_u} A_u \right)^{-1} A^*_k \Pi^\dagger_{A_u}. 
\] (C.2)
Introduce
\[
R_u \triangleq A_u P_u A_u^* + \sigma^2 I; 
\] (C.3)
with the assumption that $P_k$ in (5) is p.d. (together with (6)), it is straight-forward to write the inverse of $R$ as
\[
R^{-1} = R_u^{-1} - R_u^{-1} A_k \left( P_k^{-1} + A^*_k R_u^{-1} A_k \right)^{-1} A^*_k R_u^{-1}. 
\] (C.4)
If we then exploit that $\lambda_{\min}^{-1}(P_k) \to 0$ (i.e., the signals from the known directions are strong), we immediately have from (C.4) that
\[
A^*_k R^{-1} = O(||P_k^{-1}||), 
\] (C.5)
due to the expansion

$$(X + Y)^{-1} = X^{-1} - X^{-1} Y X^{-1} + O(\|Y^2\|)$$  \hspace{1cm} (C.6)$$

for small $Y$. The result of (C.5) allows us to write

$$\tilde{A}_k^* \Pi_{A_n}^\perp \tilde{A}_k = \tilde{A}_k^* \tilde{A}_k + O(\|P_k^{-1}\|),$$  \hspace{1cm} (C.7)$$

which in turn, with (C.6), lets us rewrite (C.2) as

$$\Pi_\varphi = \Pi_{A_n} + \Pi_{A_n}^\perp \Pi_{A_n} \Pi_{A_n}^\perp + O(\|P_k^{-2}\|).$$  \hspace{1cm} (C.8)$$

From (B.3),

$$G_{\perp} = \Pi_{A_n}^\perp G = \left( \frac{\partial r}{\partial \theta} \right)^* C_r^{s/2} \Pi_{\varphi} C_r^{1/2} \frac{\partial r}{\partial \theta} T,$$  \hspace{1cm} (C.9)$$

together with (C.1), we have that

$$C_r^{s/2} \Pi_{\varphi} C_r^{1/2} = C_r^{s/2} \Pi_{A_n}^\perp C_r^{1/2} - C_r^{s/2} \frac{\Pi_{\varphi}^\perp r_i r_i^* \Pi_{A_n}^\perp}{r_i^* \Pi_{A_n}^\perp r_i} C_r^{1/2}. \hspace{1cm} (C.10)$$

Due to (C.8) and (C.5), we have that

$$C_r^{s/2} \Pi_{A_n}^\perp C_r^{1/2} = C_r^{s/2} \Pi_{A_n}^\perp C_r^{1/2} + O(\|P_k^{-2}\|).$$  \hspace{1cm} (C.11)$$

By a similar argument, the second term in (C.10) becomes

$$C_r^{s/2} \frac{\Pi_{A_n}^\perp r_i r_i^* \Pi_{A_n}^\perp}{r_i^* \Pi_{A_n}^\perp r_i} C_r^{1/2} = C_r^{s/2} \frac{\Pi_{A_n}^\perp}{r_i^* \Pi_{A_n}^\perp r_i} \frac{\Pi_{A_n}^\perp}{r_i^* \Pi_{A_n}^\perp r_i} C_r^{1/2} + O(\|P_k^{-2}\|). \hspace{1cm} (C.12)$$

Using (C.11) and (C.12) in (C.9), together with (B.7), gives

$$G^* \Pi_{A_n} \Gamma = 2 \text{Re} \left[ \tilde{D}^* \left( P_{\varphi}^T A_{uu}^T \otimes R^{-1/2} \right) \left( \Pi_{A_n}^\perp - \frac{\Pi_{A_n}^\perp r_i r_i^* \Pi_{A_n}^\perp}{r_i^* \Pi_{A_n}^\perp r_i} \right) \right] \times \left[ \left( A_{uu}^c P_{\varphi}^c \otimes R^{-1/2} \right) \tilde{D} + \left( R^{-1/2} \otimes A_{uu} P_u \right) L \tilde{D}^c \right] + O(\|P_k^{-2}\|),$$  \hspace{1cm} (C.13)$$

where we have introduced

$$A_{uu} \overset{\Delta}{=} R^{-1/2} A_n. \hspace{1cm} (C.14)$$
Note that in (C.13),
\[
\left( P_u^T A_u w \otimes R^{-1/2} \right) \left( \Pi_{\tilde{A}_w} r_i \right) \\
= \left( P_u^T A_u w \otimes R^{-1/2} \right) \left( r_i - \left( \Pi_{\tilde{A}_uw} \otimes \Pi_{A_u w} \right) r_i \right) \\
= \text{vec} \left( R^{-1/2} R^{-1} A_u w P_u \right) - \text{vec} \left( R^{-1/2} \Pi_{A_u w} R^{-1} A_u w P_u \right) \\
= \text{vec} \left( R^{-1/2} \Pi_{\tilde{A}_uw} R^{-1} A_u w P_u \right) = O(||P_k^{-1}||). \tag{C.15}
\]

The last equality in (C.15) is proven in the following: we have that
\[
R^{-1} A_u w = R^{-1/2} R^{-1} A_u \\
= R^{-1/2} \left( R_u^{-1} A_u - R_u^{-1} A_k \left( A_k^* R_u^{-1} A_k \right)^{-1} \right. \\
\left. \times A_k^* R_u^{-1} A_u \right) + O(||P_k^{-1}||), \tag{C.16}
\]
where the last equality follows from (C.4) and (C.5). Using
\[
R_u^{-1} = \sigma^{-2} \left( I - A_u \left( P_u A_u^* A_u + \sigma^2 I \right)^{-1} P_u A_u^* \right), \tag{C.17}
\]
we can re-write the first term in (C.16):
\[
R^{-1/2} R_u^{-1} A_u = R^{-1/2} A_u \left( P_u A_u^* A_u + \sigma^2 I \right)^{-1} P_u A_u^* \\
= A_u w \left( P_u A_u^* A_u + \sigma^2 I \right)^{-1}. \tag{C.18}
\]

Define $\Xi \triangleq \left( A_k^* R_u^{-1} A_k \right)^{-1} A_k^*$ to shorten notation; then examine the second term in (C.16) and exploit (C.17):
\[
-R^{-1/2} R_u^{-1} A_k \Xi R_u^{-1} A_u \\
= -\sigma^{-2} R^{-1/2} \left( I - A_u \left( P_u A_u^* A_u + \sigma^2 I \right)^{-1} P_u A_u^* \right) \\
\times A_k \Xi R_u^{-1} A_u \\
= -\sigma^{-2} R^{-1/2} A_k \Xi R_u^{-1} A_u + A_u w \Xi_2, \tag{C.19}
\]
where $\Xi_2 \triangleq \sigma^{-2} \left( P_u A_u^* A_u + \sigma^2 I \right)^{-1} P_u A_u^* A_k \Xi R_u^{-1} A_u$. Using (C.18) and (C.19) in (C.16), we have
\[
R^{-1} A_u w = A_u w \Xi_3 - R^{-1/2} A_k \Xi_4, \tag{C.20}
\]
where \( \Xi_3 \triangleq \Xi_2 + (P_u A_u^* A_u + \sigma^2 I)^{-1} \) and \( \Xi_4 \triangleq \sigma^{-2} \Xi R_u^{-1} A_u \). Now compare (C.20) to (C.15): the first term of (C.20) vanishes due to the orthogonal projector in (C.15). The second term inserted into (C.15) gives

\[
\text{vec} \left( R^{-1/2} \Sigma_{\hat{A}_{uu}} R^{-1} A_u P_u \right) \\
= - \text{vec} \left( R^{-1} A_k \Xi_u P_u - R^{-1/2} \Sigma_{\hat{A}_{uu}} R^{-1/2} A_k \Xi_u P_u \right) \\
= O(\|P_k^{-1}\|), \tag{C.21}
\]
due to (C.5).

Hence, we can write (C.13) as

\[
G^* \Sigma \hat{A} G = 2 \Re \left( \tilde{D}^* \left( P_u^T A_u^T \otimes R^{-1/2} \right) \right) \\
\times \Sigma_{\hat{A}_{uu}} \left[ \left( A_u^c P_u \otimes R^{-1/2} \right) \tilde{D} + \left( R^{-T/2} \otimes A_u P_u \right) L \tilde{D} \right] + O(\|P_k^{-2}\|), \tag{C.22}
\]
where the last equality follows due to \( \Sigma_{\hat{A}_{uu}} \). Additionally, if we let

\[
D = \left[ \frac{\partial a(\theta_1)}{\partial \theta_1}, \frac{\partial a(\theta_2)}{\partial \theta_2}, \ldots, \frac{\partial a(\theta_d)}{\partial \theta_d} \right], \tag{C.23}
\]
and take the inverse of (C.22), we can write

\[
\text{CRB}_{\theta} = \frac{1}{2N} \left[ \Re \left( \left( D^* R^{-1/2} \Sigma_{\hat{A}_{uu}} R^{-1/2} D \right) \otimes \left( P_u^T A_u^T A_u^c P_u^c \right) \right) \right]^{-1} \\
+ \frac{1}{N} O(\|P_k^{-2}\|), \tag{C.24}
\]
where \( \otimes \) denotes element-wise multiplication.

**Appendix D.**

\( V' \) in (48) is found by looking at the derivatives of \( V \) with respect to the individual elements \( \theta_i \) of \( \theta \):

\[
V_i' = \frac{\partial V}{\partial \theta_i} = \epsilon_i^* W \epsilon + \epsilon^* W \epsilon_i \\
= \epsilon_i^* W \epsilon + \epsilon^* W^T \epsilon \triangleq \hat{\epsilon}_i^* W \epsilon, \tag{D.1}
\]
where
\[ \epsilon_i = \frac{\partial \epsilon}{\partial \theta_i}, \quad \epsilon_i = \begin{bmatrix} \epsilon_i \\ \epsilon_i^T \end{bmatrix} \quad \text{and} \quad \bar{W} = \begin{bmatrix} W & 0 \\ 0 & W^T \end{bmatrix}. \] (D.2)

From (17),
\[ \frac{\partial \epsilon}{\partial \theta_i} = \text{vec} \left( \hat{U}_s^* \frac{\partial B}{\partial \theta_i} \right) = \text{vec} \left( U_s^* \frac{\partial B}{\partial \theta_i} \right) + o_p(1). \] (D.3)

Note that
\[ U_s^* \frac{\partial B}{\partial \theta_i} = T^* A_u^* \frac{\partial \hat{B}^*}{\partial \theta_i} = -T^* \frac{\partial A_u^*}{\partial \theta_i} B, \] (D.4)

where \( T \) is full rank and defined from
\[ U_s = A_u T, \] (D.5)

and the third equality follows from the fact that \( B^* A_u = 0 \). Using (D.4) together with (B.6), we see
\[ \text{vec} \left( U_s^* \frac{\partial B}{\partial \theta_i} \right) = -(B^T \otimes T^*) L \hat{d}_i^c. \] (D.6)

Then we have
\[ \frac{\partial \epsilon}{\partial \theta^T} = -(B^T \otimes T^*) L \hat{d}_i^c + o_p(1), \] (D.7)

and if we additionally introduce \( K = \begin{bmatrix} K \\ K^c \end{bmatrix} \), where
\[ K \triangleq -(B^T \otimes T^*) L \hat{d}_i^c, \] (D.8)

(and hence \( k_i \), the \( i \)th column of \( K \), is given by (D.6)), we can compactly write
\[ \frac{\partial V}{\partial \theta} = K^* W \epsilon + o_p(1/\sqrt{N}). \] (D.9)

With the definition (49), we use (D.9) to state the expression for \( Q \):
\[ Q = K^* \hat{W} C_\epsilon \hat{W}^* K. \] (D.10)

In (D.10), we have defined
\[ C_\epsilon \triangleq \begin{bmatrix} C_\epsilon & C_\epsilon^c \\ C_\epsilon^c & C_\epsilon^T \end{bmatrix}, \] (D.11)
in which \( C_\epsilon \) is given by (30), and \( \tilde{C}_\epsilon \) by (31).

From (47), and using the expressions (D.6) and (D.9),

\[
H_{ij} = \lim_{N \to \infty} \left[ \frac{\partial k_i^*}{\partial \theta_j} \frac{\partial k_i^T}{\partial \theta_j} \right] \bar{W} \epsilon + \left[ k_i^* \ k_i^T \right] \bar{W} \begin{bmatrix} k_j \\ k_j^c \end{bmatrix}
\]

(D.12)

since \( \lim_{N \to \infty} \epsilon = 0 \), and hence

\[
H = K^* \bar{W} K.
\]

(D.13)

Appendix E.

Define \( \bar{W}^{\text{OPT}} \triangleq \sigma^2 \tilde{C}_\epsilon^{-1} \), such that \( \bar{W}^{\text{OPT}} \) exists for small \( \sigma^2 \); then, due to Theorem 1, we can write (33)

\[
W^{\text{OPT}} = \bar{W}^{\text{OPT}} + O(\sigma^2)
\]

(E.1)

since \( \tilde{C}_\epsilon^{-1} = O(\sigma^{-2}) \). Also, let

\[
\bar{W}^{\text{OPT}} = \begin{bmatrix} \bar{W}^{\text{OPT}} & 0 \\ 0 & (\bar{W}^{\text{OPT}})^T \end{bmatrix}.
\]

(E.2)

With these definitions, we can rewrite (D.13) according to

\[
\sigma^2 H^{\text{OPT}} = \bar{H}^{\text{OPT}} + O(\sigma^2),
\]

(E.3)

where

\[
\bar{H}^{\text{OPT}} \triangleq K^* \bar{W}^{\text{OPT}} K.
\]

(E.4)

It follows from (E.4) that

\[
H^{\text{OPT}} = \sigma^2 K^* \tilde{C}_\epsilon^{-1} K + \sigma^2 K^T \tilde{C}_\epsilon^{-c} K^c
\]

\[
= 2\sigma^2 \text{Re} \left( K^* \tilde{C}_\epsilon^{-1} K \right).
\]

(E.5)
Using (D.8), (32), and (A.14), (E.5) can be simplified to read
\[
\mathbf{H}_{\text{OPT}} = 2\sigma^2 \text{Re} \left( \mathbf{D}^* \left[ (\mathbf{T}^* \mathbf{\Sigma}_c^* (\mathbf{V}_s^T \mathbf{R}^* \mathbf{V}_s^c)^{-1} \mathbf{\Sigma}_c^* \mathbf{T}^*) \right] \otimes (\mathbf{B} (\mathbf{B}^* \mathbf{R}^*)^{-1} \mathbf{B}^*) \mathbf{D} \right).
\] 
(E.6)

From (42) and (10), \( \mathbf{T} = \mathbf{A}_u^* \mathbf{U}_s = \mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s \mathbf{\Sigma}_c^{-1} \); using this result together with the observation that \( \mathbf{V}_s^* \mathbf{R} \mathbf{V}_s = \mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s \), we have that the first Kronecker factor in (E.6) can be written as
\[
(T \mathbf{\Sigma}_c (\mathbf{V}_s^* \mathbf{R} \mathbf{V}_s)^{-1} \mathbf{\Sigma}_c \mathbf{T}^*)^c = (\mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s (\mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s)^{-1} \mathbf{V}_s^* \mathbf{P}_u \mathbf{A}_u)^c.
\] 
(E.7)

In (E.7),
\[
(\mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s)^{-1} = (\mathbf{V}_s^* \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s + \sigma^2 \mathbf{I})^{-1}.
\] 
(E.8)

Note that we can decompose \( \mathbf{P}_u = \mathbf{S}_u \mathbf{S}_u^* \), where \( \mathbf{S}_u \in \mathbb{C}^{d_u \times d_u} \) is full rank. From the definition of \( \mathbf{V}_s \), (10), we have \( \text{Span}(\mathbf{V}_s) = \text{Span}(\mathbf{\Pi}_{\mathbf{A}_u}^\perp \mathbf{A}_u \mathbf{S}_u) \). Thus, since \( \mathbf{V}_s^* \mathbf{\Pi}_{\mathbf{A}_u}^\perp \mathbf{A}_u \mathbf{S}_u = \mathbf{V}_s^* \mathbf{A}_u \mathbf{S}_u \) is non-singular. Using these results and (E.8) in (E.7), we have
\[
(T \mathbf{\Sigma}_c (\mathbf{V}_s^* \mathbf{R} \mathbf{V}_s)^{-1} \mathbf{\Sigma}_c \mathbf{T}^*)^c = (\mathbf{S}_u \mathbf{S}_u^*)^c + O(\sigma^2) = \mathbf{P}_u^c + O(\sigma^2).
\] 
(E.9)

By introducing \( \mathbf{B}_w \triangleq \mathbf{R}^{1/2} \mathbf{B} \), we can rewrite the second Kronecker factor of (E.6) according to
\[
\mathbf{B} (\mathbf{B}^* \mathbf{R}^*)^{-1} \mathbf{B}^* = \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{B}_w} \mathbf{R}^{-1/2} = \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_u \mathbf{w}}^\perp \mathbf{R}^{-1/2},
\] 
(E.10)

since \( \mathbf{B}_w \) spans the null-space of \( \mathbf{A}_u^* \).

Now using (E.9) and (E.10) in (E.6), and re-writing using (C.23),
\[
\bar{\mathbf{H}}_{\text{OPT}} = 2\sigma^2 \text{Re} \left( \mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_u \mathbf{w}}^\perp \mathbf{R}^{-1/2} \mathbf{D} \otimes \mathbf{P}_u^c \right) + O(\sigma^2).
\] 
(E.11)
Since \( C_{\text{POW}} = H^{-1}QH^{-1} \), we have from the above derivations and (D.10) that

\[
C_{\text{POW}} = \sigma^2 \left( \sigma^2 H_{\text{OPT}} \right)^{-1} = \sigma^2 \left( \bar{H}_{\text{OPT}} + O(\sigma^2) \right)^{-1}
\]

\[
= \frac{1}{2} \left[ \text{Re} \left( \left( D^* R^{-1/2} \Pi_{A_u}^\dagger R^{-1/2} D \right) \otimes P_u^e \right) \right]^{-1} + O(\sigma^4). \tag{E.12}
\]

We now examine the second factor of the Schur-product in (39), \( P_u A_u^* R^{-1} A_u P_u \).

Let \( P_k = S_k S_k^* \), with \( S_k \in \mathbb{C}^{d_k \times d_k} \) full rank. Then we can write

\[
R = \begin{bmatrix} A_u S_u & A_k S_k \end{bmatrix} \begin{bmatrix} S_u^* A_u^* \\ S_k^* A_k^* \end{bmatrix} + \sigma^2 I. \tag{E.13}
\]

Let \( \bar{A} = \begin{bmatrix} A_u S_u & A_k S_k \end{bmatrix} \); then, as in (F.1), we have that

\[
\sigma^2 R^{-1} = \Pi_{\bar{A}} + \sigma^2 (\bar{A}^\dagger)^* \bar{A}^\dagger + O(\sigma^4). \tag{E.14}
\]

Using (E.14),

\[
P_u A_u^* R^{-1} A_u P_u = \frac{1}{\sigma^2} S_u S_u^* A_u^* \left( \Pi_{\bar{A}}^\dagger + \sigma^2 (\bar{A}^\dagger)^* \bar{A}^\dagger \right) A_u S_u S_u^* + O(\sigma^2)
\]

\[
= S_u \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \Pi_{\bar{A}}^\dagger + \sigma^2 (\bar{A}^\dagger)^* \bar{A}^\dagger \end{bmatrix} S_u^* + O(\sigma^2)
\]

\[
= P_u + O(\sigma^2), \tag{E.15}
\]

which gives

\[
\bar{\text{CRB}}_\theta = \frac{1}{2N} \left[ \text{Re} \left( \left( D^* R^{-1/2} \Pi_{A_u}^\dagger R^{-1/2} D \right) \otimes P_u^e \right) \right]^{-1} + O(\sigma^4). \tag{E.16}
\]

Comparing (E.16) to (E.12), we have that

\[
C_{\text{POW}} = N \cdot \bar{\text{CRB}}_\theta + O(\sigma^4); \tag{E.17}
\]

hence, both POWDER methods are efficient for small \( \sigma^2 \).

**Appendix F.**

For small \( \sigma^2 P^{-1} \) we have that

\[
\sigma^2 R^{-1} = I - A (A^* A + \sigma^2 P^{-1})^{-1} A^*
\]

\[
= \Pi_{\bar{A}}^\dagger + \sigma^2 (\bar{A}^\dagger)^* P^{-1} A^\dagger + O(\sigma^4 \|P^{-2}\|). \tag{F.1}
\]
References


Figure 1: Comparison of POWDER to existing state-of-the-art methods. Normalized RMSE, averaged over 1000 MC realizations, along with the derived theoretical bounds; $\rho_b = \rho_w = 0.9$, and $N = 1000$. 
Figure 2: Scenario as in Fig. 1, except for SNR as detailed in the respective figure; averages of 1000 MC-simulations, $N = 1000$.
Figure 3: Comparison to existing state-of-the-art methods when $\rho_u = \rho_h = 0$. Normalized RMSE, averaged over 500 MC realizations, along with the derived theoretical bounds; $m = 15$ and $N = 1000$. 
Figure 4: Showing the effects of violating the assumption $P_{uk} = 0$, and the influence of finite sample effects. Examining three different values of the parameter $\rho_{uk}$. SNR fixed at 15dB, averages over 500 MC realizations.