Conference Paper

Channel Vector Subspace Estimation from Sample Covariance of Low-Dimensional Projections

Author(s):
Haghighatshoar, Saeid; Caire, Giuseppe

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010646818

Rights / License:
In Copyright - Non-Commercial Use Permitted
Channel Vector Subspace Estimation from Sample Covariance of Low-Dimensional Projections

Saeid Haghighatshoar∗, Giuseppe Caire∗
*Communications and Information Theory Group, Technische Universität Berlin
Email: (saeid.haghighatshoar, caire)@tu-berlin.de

Abstract—In this paper, we propose efficient algorithms for estimating the signal subspace of mobile users in a wireless communication environment with a multi-antenna base-station with \(M\) antennas. When \(M\) is large, because of the high angular resolution of the receiver, any realization of the random channel vector of any given user is approximately contained in a user-specific subspace of dimension \(p \ll M\). Efficient multiuser MIMO schemes can be obtained from such subspace information, which is stable in time and can be accurately estimated even in the presence of fast fading (e.g., for mm-Wave channels). We are interested in the massive MIMO regime of \(M \gg 1\). In order to reduce the RF front-end complexity and overall A/D conversion rate, the \(M\)-antenna base-station transmitter/receiver is split into the product of a baseband linear projection (digital) and an RF reconfigurable beamforming network (analog) with only \(m \ll M\) RF chains. Hence, only \(m\)-dimensional analog observations can be obtained for subspace estimation. We develop efficient algorithms that estimate the dominant signal subspace of the users from sampling only \(m = O(2\sqrt{M})\) specific array elements according to a coprime scheme. For a given target dimension of the signal subspace \(p \leq M\), our algorithms return a \(p\)-dimensional beamformer with a performance comparable with the best \(p\)-dim beamformer designed by knowing the exact covariance matrix of the received signal. We assess the performance of our proposed estimators both analytically and empirically via numerical simulations, and compare it with that of the other state-of-the-art methods in the literature.

1 INTRODUCTION

Consider a multiuser MIMO channel formed by a base-station with \(M\) antennas and \(K\) single-antenna mobile users in a cellular network. We focus here on a flat-fading channel in which the bandwidth of the signal is less than the channel’s coherence bandwidth. Following the current massive MIMO approach [1, 2], we assume that the uplink and the downlink are organized in Time Division Duplexing (TDD), where the base-station estimates the channel vectors of the users from orthogonal pilots that are sent by the users in the uplink in the same channel coherence time [1]. It turns out that for isotropically distributed channel vectors it is optimal to devote half the coherence time to estimate the channel, and to devote the remaining half to serve the users.

In the massive MIMO setup, the number of antennas \(M\) is large, and the receiver antenna at the base-station has a high angular resolution. Consequently, in many relevant scenarios, the channel is far from isotropic. Indeed, as the propagation for a user occurs only through a small set of Angles of Arrivals (AoAs), its channel vectors in consecutive coherence blocks lie on very low-dimensional subspaces. This underlying structure can be exploited to improve the system multiplexing gain via decreasing the training overhead. For example, one approach would be to cluster the users based on the dominant subspace of their channel vectors, and apply the classical channel estimation on a per-group basis, on the low-dimensional projected channels [3]. This requires estimating the dominant signal subspace of each individual user. Although the channel vector changes in every coherence time, in many practical scenarios, the signal subspace remains stationary across many coherence blocks, thus, it can be reliably estimated.

A direct naive approach for estimating the signal subspace is to first estimate the \(M \times M\) covariance matrix of the channel coefficient of each user via sampling the whole array elements, and then identify the signal subspace by applying the singular value decomposition (SVD). This requires sampling the whole array elements which requires \(M\) RF chains. Since in massive MIMO setup \(M \gg K\), this is inefficient and very difficult to implement. Different architectures such as Hybrid Digital Analog (HDA) have been proposed to reduce hardware complexity (notably, the A/D overall bit-rate and the number of RF modulation/demodulation chains). The main idea is to implement the \(M \times K\) beamforming matrix as the product of two matrices: an \(M \times m\) beamforming matrix implemented in the RF analog domain, and an \(m \times K\) precoding matrix implemented in the digital baseband domain, so that only \(m \ll M\) A/D converter and RF chains be used. This implies that exploiting the subspace information is possible only when it can be extracted (estimated) from \(m\)-dimensional sketches \((m \ll M)\) of the received signal.

In this paper, we aim to design suitable subspace estimators from low-dimensional sketches of the input signal for a uniform linear array (ULA). The geometry of the array and the scattering channel is shown in Fig. 1. Array elements have a uniform distance \(d = \frac{\lambda}{2\sin(\theta_m)}\), and scan the angular range \([-\theta_m, \theta_m]\) for some \(\theta_m \in (0, \pi/2)\). We use a coprime sampling scheme, introduced in [4], that samples only \(O(2\sqrt{M})\) specific array elements. We propose several algorithms for estimating the signal subspace and cast them as convex optimization problems that can be solved efficiently. We also analyze the performance of our estimators both analytically and empirically via numerical simulations, and compare it with that of the other state-of-the-art methods in the literature.

2 RELATED WORK

Several works in the literature are related to the problem addressed in this paper, which can be summarized in the
following four categories: Subspace tracking, Low-rank matrix recovery, Direction-of-arrival (DoA) estimation, and Multiple Measurement Vectors (MMV) problem in compressed sensing (CS). Let us consider a simple model in which the transmission between a user and the base-station occurs through p scatterers (see Fig. 1). One snapshot of the received signal is given by

$$y = \sum_{\ell=1}^{p} a(\theta_\ell) w_\ell x + n, \quad (1)$$

where x is the transmitted (training) symbol, $w_\ell \sim \mathcal{CN}(0, \sigma^2)$ is the channel gain of the $\ell$-th multipath component, n $\sim \mathcal{CN}(0, I_M)$ is the additive white Gaussian noise of the receiver antenna, and where $a(\theta) \in \mathbb{C}^M$ is the array response at AoA $\theta$, whose $k$-th component is given by

$$[a(\theta)]_k = e^{j k 2 \pi d \sin(\theta)}, \quad (2)$$

According to the WSSUS model, the channel gains for different paths, i.e., \{w_\ell\}_{\ell=1}^p, are uncorrelated, and since they are (jointly) Gaussian, they are statistically independent. Without loss of generality, we suppose $x = 1$ in all training snapshots. Letting $A = [a(\theta_1), a(\theta_2), \ldots, a(\theta_p)]$, we have

$$y(t) = Aw(t) + n(t), \quad t \in [T], \quad (3)$$

where $w(t) = (w_1(t), w_2(t), \ldots, w_p(t))^T$ for different $t \in [T] := \{0, 1, \ldots, T-1\}$ is statistically independent. We assume that the AoAs $\{\theta_\ell\}_{\ell=1}^p$ remain invariant over a long time T. From (3), the covariance of y(t) is given by

$$C_y = A\Sigma A^H + I_M = \sum_{\ell=1}^{p} \sigma_\ell^2 a(\theta_\ell)a(\theta_\ell)^H + I_M. \quad (4)$$

Let $C_y = UAU^H$ be the singular value decomposition (SVD) of $C_y$, where $A = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$ denotes the diagonal matrix of singular values. We always assume that the singular values are sorted in a non-increasing order. If we denote by $U_p$ the $M \times p$ matrix consisting of the first p columns of $U$, it is not difficult to see that the columns of $U_p$ span the signal space. In particular, span($A$) = span($U_p$). We need to identify this subspace from noisy low-dimensional sketches $x(t) = By(t)$, where $B$ is the sampling matrix. This problem for the noiseless case was studied by Chi et al. in [5] where they developed PETREL algorithm to estimate the underlying subspace. Another algorithm named GROUSE was proposed by Balzano et. al. in [6] which uses a low-complexity stochastic gradient update over the Grassmanian manifold. Both algorithms mainly optimize the computational complexity rather than the data size, and principally suit situations in which the dimension is high (very large $M$ and $T$). We empirically compare the performance of our proposed algorithms with PETREL for a fixed data size in Section 5.

For $p \ll M$ and for a high SNR, the covariance matrix $C_y$ in (4) is nearly low-rank. Recovery of low-rank matrices from a collection of a few possibly noisy samples is of great importance in signal processing and machine learning. Recently, it has been shown that this can be done via nuclear norm minimization, which is a convex optimization and can be efficiently solved [7]:

$$X^* = \arg\min_M \|M\|_*, \text{ subject to } M\Omega = X\Omega, \quad (5)$$

where the nuclear norm $\|M\|_*$ is given by the sum of the eigen-values of M, and reduces to $\text{Tr}(M)$ when M is positive semi-definite (PSD). In practice, we have only a collection of T snapshots of sketches $x(t) = By(t)$, $t \in [T]$, rather than the whole covariance matrix or even the sketches thereof. Let

$$\hat{C}_y = \frac{1}{T} \sum_{t=1}^{T} y(t)y(t)^H, \quad \hat{C}_x = B\hat{C}_y B^H \quad (6)$$

be the sample covariance of the full and subsampled signal. We compare the performance of our algorithms with the following nuclear norm minimization

$$\min_{M} \text{Tr}(M) \text{ subject to } M \in T_+, \|\hat{C}_x - MBM^H\| \leq \epsilon, \quad (7)$$

where $T_+$ is the space of $M \times M$ PSD Toeplitz matrices, and where $\epsilon$ is an estimate of the $\ell_2$-norm of the error.

From (3), it is seen that the received signal $y(t)$ is a noisy superposition of p independent Gaussian sources arriving from $p$ different angles. This is the same model studied for direction-of-arrival (DoA) estimation. There are two main categories of algorithms for DoA estimation: classical algorithms such as MUSIC and ESPRIT that use subspace methods to locate the AoAs, and more recent compressed sensing based algorithms that use the angular sparsity of the signal over a prespecified grid (see [8, 9] and refs. therein). Recently, Candès and Fernandez-Granda [10] developed an off-grid super-resolution (SR) technique using total-variation (TV) minimization. This algorithm was extended by Tan et. al. in [11] to DoA estimation with coprime arrays when the sources are sufficiently separated. In a wireless environment the AoAs are clustered. This implies that the separation requirement for the super-resolution setup may not be met. Since in this paper we aim at estimating the subspace of the signal rather than DoAs, in section 4.2.3 we extend the super-resolution method to develop a new algorithm for estimating the signal subspace.

It is seen from (3) that, neglecting the measurement noise $n(t)$, the signal $y(t)$ has typically a sparse representation over the continuous dictionary $\{a(\theta), \theta \in [-\theta_m, \theta_m]\}$, i.e., only p atoms of the dictionary, i.e., $\{a(\theta_\ell)\}_{\ell=1}^p$, are needed to represent the signal. Thus, $x(t) = By(t)$ can be seen as identifying a sparse vector from a collection of sketches, which coincides with the traditional CS problem. An extension of this problem involves Multiple Measurement Vectors (MMV).
The underlying assumption is that \( y(t) \), for different snapshots \( t \in [T] \), have the same sparsity pattern or support over the underlying dictionary even though they might have different coefficients \( w(t) \) for each \( t \). This problem has been widely studied in the literature (see [9, 12, 13] and refs. therein), where two main approaches have been proposed for estimating the common support of the signals: using a greedy algorithm or convex optimization via a regularizer promoting group sparsity; and using covariance matrix of data and subspace techniques. Once the support is identified, the standard Least-Squares method can be used to find the coefficients. Since the underlying dictionary is continuous, both classes exploit either grid-based or more recently developed off-grid techniques. We will compare the performance of our algorithm with grid-based approach in [13], and the grid-less one in [14–16].

3 Statement of the Problem

In (3), we introduced the channel model given by \( y(t) = AW(t) + n(t), t \in [T] \), where \( A \) contains the array response for the AoAs. Let \( \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) \) be the matrix containing the channel strengths for the AoAs \( \{\theta_p\}_{p=1}^P \). We can prove the following simple result.

**Proposition 3.1:** Let \( C_z = \frac{1}{T} \sum_{t=1}^{T} x(t)x(t)^H \) be the sample covariance of the sketches \( x(t), t \in [T] \). Then \( C_z \) is a sufficient statistics for estimating \( A \) and \( \Sigma \).

For a more practical scenario, we consider the following continuum model

\[
y(t) = \sqrt{\text{snr}} \int_{-1}^{1} \sqrt{f(u)}a(u)z(u,t)du + n(t), \quad t \in [T],
\]  

where \( \text{snr} \) is the SNR and \( z(u,t) \) is a circularly symmetric Gaussian process with \( \mathbb{E}\{z(u,t)z(u',t')^*\} = \delta(u-u')\delta_{t,t'} \). The measure \( f(u) \) models the distribution of the received signal's power over \( u \in [-1,1] \), where \( u = \frac{\sin(\theta)}{\sin(\theta_m)} \) for \( \theta \in [-\theta_m, \theta_m] \). With some abuse of notation, we denote the array vector in the \( u \) domain by \( a(u) \) where \( [a(u)]_k = e^{jku} \).

Let \( C(f) = S(f) + IM \) be the covariance matrix of the received signal, where \( S(f) = \text{snr} \int_{-1}^{1} f(u)a(u)a(u)^Hdu \) is the covariance of the signal of the user with power distribution \( f(u) \). We define the best \( p \)-dim beamformer for \( S(f) \) as \( V_p = \arg \max_{V \in \mathbb{H}(M,p)} \langle S, VV^H \rangle \), where \( \mathbb{H}(M,p) \) is the space of all \( M \times p \) matrices \( U \) with \( U^H U = I_p \). We assess the efficiency of \( V_p \) for capturing the signal's power by

\[
\delta_p = \frac{\langle S, V_p V_p^H \rangle}{\langle Tr(S) \rangle} = \frac{\text{Tr}(V_p^H S V_p)}{\text{Tr}(S)},
\]

where \( \delta_p \approx 1 \) implies that a significant amount of signal's power is concentrated in a \( p \)-dim subspace. Let \( S \) be an estimate of \( S \) and let \( \hat{V}_p \) be its best \( p \)-dim beamformer. We can use \( \hat{V}_p \) as an estimate of the optimal beamformer \( V_p \). We define the following metric for the efficiency of \( \hat{V}_p \)

\[
\Gamma_p = \frac{\langle S, \hat{V}_p \hat{V}_p^H \rangle}{\langle S, V_p V_p^H \rangle} = 1 - \frac{\langle S, V_p V_p^H \rangle - \langle S, \hat{V}_p \hat{V}_p^H \rangle}{\langle S, V_p V_p^H \rangle},
\]

where \( \langle S, V_p V_p^H \rangle - \langle S, \hat{V}_p \hat{V}_p^H \rangle \geq 0 \) is the amount of power lost due to the mismatch between \( V_p \) and the estimate \( \hat{V}_p \).

Note that \( \Gamma_p \in [0,1] \), and the aim is to design an estimator with a \( \Gamma_p \) as close to 1 as possible.

4 Sampling Operator, Algorithms and Results

4.1 Coprime Subsampling Operator

In this section, we introduce our coprime sampling scheme, which samples only \( m \approx 2\sqrt{M} \) carefully selected array elements. Suppose \( q_1, q_2 \) are coprime numbers, i.e., \( \gcd(q_1, q_2) = 1 \), with \( q_1q_2 \approx M \) and \( q_1 \approx q_2 \approx \sqrt{M} \). Let \( D \) be the set of all nonnegative integer combinations of \( q_1 \) and \( q_2 \) less than or equal to \( M-1 \), i.e., \( D = \cup_{i=1,2} \{k \in [M], \mod(k, q_i) = 0\} \), where \( [M] = \{0, 1, \ldots, M-1\} \). Note that \( |D| \approx 2\sqrt{M} \). Since \( q_1 \) and \( q_2 \) are coprime, for sufficiently large \( M \), we have \( D = \{0, \ldots, M-1\} \). Suppose the elements of \( D \) are sorted in an increasing order with \( d_i \in D \) being the \( i \)-th largest element in the list. Also, let \( m = |D| \) be the number of elements in \( D \) and let \( B \) be the \( m \times M \) binary matrix with \( B_{d,i} = 1 \) for \( i \in \{1, \ldots, m\} \) and zero otherwise. We can simply check that \( BB^H = I_m \). We also prove the following result, which shows the efficiency of the coprime matrix \( B \) for sampling Hermitian Toeplitz matrices.

**Proposition 4.1:** Let \( S \) be an \( M \times M \) Hermitian Toeplitz matrix and let \( B \) be the coprime sampling matrix. Then the mapping \( S \to BB^H \) is a bijection.

4.2 Proposed Algorithms for Subspace Estimation

4.2.1 Algorithm 1: Approximate Maximum Likelihood (AML) Estimator

Let \( S = \text{snr} \int_{-\pi}^{\pi} f(u)a(u)a(u)^Hdu \) be the covariance matrix of a user with power distribution \( f(u) \). It is easy to check that for the coprime sampling matrix \( B \), we have

\[
p(x(1:T)|S) = \exp \left\{ -T \text{Tr}(C_z(I_m + BB^H)^{-1}) \right\}
\]

where \( C_z = \frac{1}{T} \sum_{t=1}^{T} x(t)x(t)^H \) is the sample covariance of the sketches \( x(t) \). The ML estimator for \( S \) can be written as \( S^* = \arg \min_{S \in T_+} L(S) \), where \( T_+ \) is the set of PSD Toeplitz matrices and where \( L(S) \) is given by

\[
L(S) = \log \det(I_m + BB^H) + \text{Tr}(C_z(I_m + BB^H)^{-1}).
\]

**Proposition 4.2:** Let \( L(S) \) be as in (11). Then, \( L(S) \) is the sum of the concave function \( L_{cap}(S) = \log \det(I_m + BB^H) \) and the convex one \( L_{conv}(S) = \text{Tr}(C_z(I_m + BB^H)^{-1}) \).

As \( L(S) \) is not convex, the ML estimation is generally intractable. However, since the signal covariance matrix \( S \) scales linearly with \( \text{snr} \), it is possible to obtain a convex (indeed, linear) approximation of \( L_{cap}(S) \), which is tight especially for low SNR.

**Proposition 4.3:** Let \( L_{cap}(S) = \log \det(I_m + BB^H) \). Then, \( L_{cap}(S) \leq \text{Tr}(BB^H) \) for all \( S \in T_+ \). Moreover, for low \( \text{SNR}, L_{cap}(S) \approx \text{Tr}(BB^H) + o(\text{snr}) \).

From Proposition 4.3, we define the AML cost function by

\[
L_{aml}(S) = \text{Tr}(BB^H) + \text{Tr}(C_z(I_m + BB^H)^{-1}),
\]

as the best convex upper bound for \( L(S) \), which is tight for very low SNR. In particular, AML can be formulated as a semi-definite program (SDP) that can be solved efficiently.
Proposition 4.4: Let $L_{\text{app}}(S)$ be as in (12). Suppose that $U^H A^H U$ is the SVD of $C_x$ and set $\Delta = C_x^{1/2} = U A^{1/2}$. Then the AML estimate is obtained from the following SDP

$$\begin{align*}
(S^*, W^*) &= \arg \min_{S \in T_A, W \in \mathbb{C}^{m \times m}} \text{Tr}(SB^H B^H) + \text{Tr}(W) \\
\text{subject to} & \begin{bmatrix} S & Z^H \end{bmatrix} \begin{bmatrix} S & W \\ Z & \Delta^H \end{bmatrix} \succeq 0, \ \|X - B Z\| \leq \delta,
\end{align*}$$

where $\delta$ is an estimate of the $\ell_2$-norm of $\bar{N}$. For sufficiently large $m$, the optimal $\delta$ is given by $\delta^* = \sigma \sqrt{m^2 + m\sigma} \approx 2\sigma \sqrt{M}$, where $\sigma^2$ is the noise variance in each array element, which can be estimated during the system’s operation.

4.2.3 Algorithm 3: Super Resolution (SR): Consider a user with a power distribution $f(u)$ and let $S(f)$ be its signal covariance matrix. Note that $S$ is a Toeplitz matrix whose first column is given by $f = (f(a, a)) = f(u) \mid u \in [0, M]$, where $f(u) = \int f(u)e^{j \pi k u} du$ is the $k$-th Fourier coefficient of $f$.

In this section, we assume that $f$ is merely a positive measure and not necessarily a normalized one. Since $S$ is Toeplitz, from Proposition 4.1, it is seen that for the coprime sampling matrix $B$ introduced in Section 4.1, all the elements of $S$, and as a result the vector of Fourier coefficients $f$ can be identified from $S^H$. This implies that for a sufficiently large $T$, we can estimate $f$ accurately using the elements of the sample covariance matrix $\hat{C}_S = B C_x B^H$. Let $\hat{x}_k = \{(i, t') : i \geq t', d_i - d_j = k\}$, where $D$ and $d_i \in D$ are as in Section 4.1. Let $\hat{c}_k = |\hat{x}_k|$, and define the estimator $\hat{f}_k = \sum_{i,j} x_{ij} \hat{C}_{ij}$ for $f_k$. We propose the following TV-minimization to recover the subspace of the signal from the estimates $\hat{f}$

$$\begin{align*}
\min_{f \in \mathbb{T}^n} \|f\|_{TV} \text{ subject to } \|\langle f, a \rangle - \hat{f}\| \leq \epsilon,
\end{align*}$$

where $\epsilon$ is an estimate of the $\ell_2$-norm of the noise in the data. Since $f$ is a positive measure, $\|f\|_{TV}$ is given by $f(\{i, t\}) = \int_0^1 f(u) du = \hat{f}_0$, thus, we obtain the following result.

Proposition 4.5: Consider the TV-minimization in (16).

Then, (16) can be equivalently written as

$$\begin{align*}
S^* &= \arg \min_{T} \text{Tr}(T) \text{ subject to } T \in \mathbb{T}_+,
\end{align*}$$

where $e_1 = (1, 0, \ldots, 0)^T$ is an $M \times 1$ vector, where $[T]_{11}$ is the diagonal element of the Toeplitz matrix $T$ (equivalent to $\hat{f}_0$), and where $\sigma^2$ is an estimate of noise variance.

Algorithm (17) is a convex optimization that can be efficiently solved to recover the signal covariance matrix $S$. In particular, no prior knowledge of SNR is necessary.

4.2.4 Algorithm 4: Covariance Matrix Projection (CMP): Let $B$ be the $m \times M$ subsampling matrix as in Section 4.1, and $C_x$ be the sample covariance of the subsampled signal. In order to recover the dominant $p$-dim subspace of the signal, we first find an estimate of the signal covariance matrix $C_y^*$

$$\begin{align*}
C_y^* &= \arg \min_{R \in \mathbb{T}_+} \|LT(\hat{C}_S) - LT(B R B^H)\|_F,
\end{align*}$$

where $LT$ keeps the lower-diagonal elements of $C_x$. Then, an estimate of signal subspace is obtained from $C_y^*$. The following theorem shows the resulting performance.

Theorem 4.6: Consider the optimization problem (18).

Then, for a given $p$ with $1 \leq p \leq M$, the CMP estimator recovers a $p$-dim subspace of the signal, and has a performance measure $\Gamma_p$ satisfying

$$\begin{align*}
E(\Gamma_p) &\geq \max \left\{ 1 - \frac{2\sqrt{p}}{\sigma \sqrt{p}} \left(1 + \frac{1}{\text{snr}}\right), 0 \right\}, \\
\text{Var}(\Gamma_p) &\leq \frac{4p}{M^2} \left(1 + \frac{1}{\text{snr}}\right)^2,
\end{align*}$$

106
where \( \delta_p \) is defined as in (9), and where \( \text{snr} \) is the received SNR in one snapshot \( t \in [T] \).

5 Simulations

In this section, we assess the performance of our proposed estimators via numerical simulations. We use the CVX package [17] for running all the convex optimizations. We assume that the AoAs are uniformly distributed in \( \Theta = [40, 50] \cup [100, 110] \) with an angular spread of 20 degrees. We use an array of size \( M = 80 \), and a coprime sampling with \( q_1 = 7 \), \( q_2 = 9 \), where we sample only 19 out of 80 array elements that are located at \( D = \{0, 7, \ldots, 77\} \cup \{0, 9, \ldots, 72\} \).

Fig. 2 compares the performance of our proposed algorithms with PETRELS, nuclear norm minimization (NucNorm) in (7), grid-based (GBMMV) in [13], and grid-less MMV (GLMMV) in [14–16].

Fig. 3 compares the scaling of the different estimators with training length \( T \in \{200, 400, 800, 1600\} \).

**Fig. 2:** Comparison of the performance of the estimators versus SNR for the training length \( T = 100 \). It is seen that AML, RMMV, and SR perform comparably with the GLMMV but they have lower computational complexity which does not scale with \( T \). The CMP is as good as GBMMV and better than NucNorm especially for higher SNR but its complexity is much lower than GBMMV since it does not scale with \( T \). PETRELS does not perform very well for the fixed data size, e.g., its performance even for \( T = 800 \) is worse than that of the other algorithms for \( T = 100 \).

**Fig. 3:** Scaling of the performance of different estimators with training length \( T \). As the performance of AML and RMMV is comparable with the GLMMV and better than GBMMV and since for large training length \( T \), these algorithms are really time-consuming to run, we have not included them in this figure.

References


