

Supporting material for the paper:

On the efficiency of misspecified Gaussian inference in nonlinear regression: application to time-delay and Doppler estimation

Stefano Fortunati, Lorenzo Ortega

The aim of this supporting material is to provide a sketch of the proof of the Theorem 2 in the main paper. The interested readers can find all the measure-theoretic aspects on α -mixing processes and the related technical regularity conditions in [1]. For the aim of clarity, let us start by recalling here the Assumption 1 and Theorem 2 reported in the main paper.

Assumption 1 Let $\{n_k : \forall k\}$ be a zero-mean, WSS discrete and circular complex-valued process [2] such that the joint pdf of N samples follows an unspecified pdf $\mathbf{n} \sim p_{\mathbf{n}}, \forall N$. Then, we assume that its autocorrelation function exists and satisfies $|r_n[j]| \triangleq |E_{p_{\mathbf{n}}}[n_{k+j}^* n_k]| = O(|j|^{-\gamma})$, $m \in \mathbb{Z}$, $\gamma > \varrho/(\varrho - 1)$, $\varrho > 1$.¹ Note that the circularity of $\{n_k : \forall k\}$ implies that $E_{p_{\mathbf{n}}}[n_{k+j} n_k] = 0$, $\forall k, j$.

Theorem 2 Let $\{x_k\}_{k=N_1}^{N_2}$ be a sequence of $N = |N_2 - N_1 + 1|$ scalar, complex-valued, random variables s.t.:

$$\mathbb{C} \ni x_k = f_k(\bar{\boldsymbol{\theta}}) + n_k, \quad N_1 \leq k \leq N_2, \quad (75)$$

where $\bar{\boldsymbol{\theta}} \in \Theta \subset \mathbb{R}^p$ indicates the real-valued, true parameter vector and Θ is a compact subset of \mathbb{R}^p . The functions $f_k : \Theta \rightarrow \mathbb{C} \forall k$, are known, continuous and differentiable functions on Θ . Under Assumption 1 and other technical regularity conditions (see A1-A9 in [1]), the estimator

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \left\{ N^{-1} \sum_{k=N_1}^{N_2} |x_k - f_k(\boldsymbol{\theta})|^2 \right\}, \quad (76)$$

satisfies the following properties:

1) Consistency wrt the true parameter vector:

$$\hat{\boldsymbol{\theta}}_N \xrightarrow{a.s.} \bar{\boldsymbol{\theta}}, \quad (77)$$

$\xrightarrow{a.s.}$ indicates the almost sure convergence.

2) Asymptotic normality: Let us indicate as $\underset{N \rightarrow \infty}{\sim}$ the convergence in distribution, we have:

$$\sqrt{N} [\mathbf{P}(\bar{\boldsymbol{\theta}})]^{-1/2} \mathbf{K}(\bar{\boldsymbol{\theta}}) (\hat{\boldsymbol{\theta}}_N - \bar{\boldsymbol{\theta}}) \underset{N \rightarrow \infty}{\sim} \mathcal{N}(0, \mathbf{I}), \quad (78)$$

where the matrices $\mathbf{P}(\bar{\boldsymbol{\theta}})$ and $\mathbf{K}(\bar{\boldsymbol{\theta}})$ are defined in eqs. (27f) and (29) of the main paper.

Proof: To show the consistency property in (77), let us start by introducing the function:

$$Q_N(\boldsymbol{\theta}) \triangleq N^{-1} \sum_{k=N_1}^{N_2} |x_k - f_k(\boldsymbol{\theta})|^2. \quad (79)$$

¹Given a real-valued function $f(x)$ and a positive real-valued function $g(x)$, $f(x) = O(g(x))$ if and only if there exists a positive real number a and a real number x_0 such that $|f(x)| \leq ag(x)$, $\forall x \geq x_0$.

Moreover, under [1, A2], we can define the expected value of $Q_N(\boldsymbol{\theta})$ as:

$$\bar{Q}_N(\boldsymbol{\theta}) \triangleq N^{-1} \sum_{k=N_1}^{N_2} E \{|x_k - f_k(\boldsymbol{\theta})|^2\}. \quad (80)$$

By definition, $\hat{\boldsymbol{\theta}}_N$ in (76) minimizes $\bar{Q}_N(\boldsymbol{\theta})$.

Let us define now as $\boldsymbol{\theta}_0$ the vector that minimizes $\bar{Q}_N(\boldsymbol{\theta})$. Following [1, A4], we assume that $\boldsymbol{\theta}_0$ is unique in a neighborhood of $\bar{\boldsymbol{\theta}}$. It follows directly from the statistical characterization of $\{n_k : \forall k\}$ in Assumption 1 of our main paper that:

$$\begin{aligned} \boldsymbol{\theta}_0 &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \bar{Q}_N(\boldsymbol{\theta}) \\ &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \left\{ \sum_{k=N_1}^{N_2} E \{|f_k(\bar{\boldsymbol{\theta}}) - f_k(\boldsymbol{\theta}) + n_k|^2\} \right\} \\ &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \left\{ \sum_{k=N_1}^{N_2} |(f_k(\bar{\boldsymbol{\theta}}) - f_k(\boldsymbol{\theta}))|^2 + 2\operatorname{Re}\{(f_k(\bar{\boldsymbol{\theta}}) - f_k(\boldsymbol{\theta}))E\{n_k\}\} + E\{|n_k|^2\} \right\} \\ &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \left\{ \sum_{k=N_1}^{N_2} |f_k(\bar{\boldsymbol{\theta}}) - f_k(\boldsymbol{\theta})|^2 \right\} = \bar{\boldsymbol{\theta}}. \end{aligned} \quad (81)$$

This result suggest us that, if $Q_N(\boldsymbol{\theta})$ converges to $\bar{Q}_N(\boldsymbol{\theta})$, one could expect that $\hat{\boldsymbol{\theta}}_N$ (that minimizes $Q_N(\boldsymbol{\theta})$) would converges to the vector $\boldsymbol{\theta}_0$ that minimizes $\bar{Q}_N(\boldsymbol{\theta})$. Then, the consistency property would follows directly from the fact that, as proved in (81), $\boldsymbol{\theta}_0$ equates the true parameter vector $\bar{\boldsymbol{\theta}}$.

This line of reasoning can be formally proved under the Assumption 1. Specifically, by using a generalization of the Strong Law of Large Numbers (SLLN), obtained under Assumption 1 in [1, Theo. 2.3], one can prove that:

$$Q_N(\boldsymbol{\theta}) \xrightarrow{a.s.} \bar{Q}_N(\boldsymbol{\theta}), \quad (82)$$

uniformly in $\boldsymbol{\theta} \in \Theta$. Moreover, Under the Assumption 1, the Theorem 3.1 in [1] assures us that:

$$\hat{\boldsymbol{\theta}}_N \xrightarrow{a.s.} \boldsymbol{\theta}_0, \quad (83)$$

that implies the consistency since $\boldsymbol{\theta}_0 = \bar{\boldsymbol{\theta}}$ as proved in (81).

Let us now move to the asymptotic normality property. By assuming the differentiability (in the Wirtinger sense [3], [4]) of the function $f_k : \mathbb{R}^p \rightarrow \mathbb{C}$, the $\hat{\boldsymbol{\theta}}_N$ in eq. (76) can be rewritten in terms of *estimating equations* as the solution of the following non-linear system:

$$\nabla_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_N} \equiv \nabla_{\boldsymbol{\theta}} Q_N(\hat{\boldsymbol{\theta}}_N) = \mathbf{0}. \quad (84)$$

Using the standard first order Taylor expansion around the true parameter vector $\bar{\boldsymbol{\theta}}$, we have that:²

$$\nabla_{\boldsymbol{\theta}} Q_N(\hat{\boldsymbol{\theta}}_N) - \nabla_{\boldsymbol{\theta}} Q_N(\bar{\boldsymbol{\theta}}) = \boldsymbol{\Omega}_N(\bar{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}_N - \bar{\boldsymbol{\theta}} \right) + o_P(1), \quad (85)$$

²The term $o_P(1)$ characterizes a sequence of random variables converging to 0 in probability. Formally, given a sequence of random variables $\{x_k\}$, the notation $x_k = o_P(1)$ stands for $\lim_{k \rightarrow \infty} \Pr \{|x_k| \geq \varepsilon\} = 0, \forall \varepsilon > 0$.

where $\Omega_N(\tilde{\theta})$ is the Hessian of $Q_N(\theta)$:

$$\Omega_N(\tilde{\theta}) \triangleq \nabla_{\theta} \nabla_{\theta}^{\top} Q_N(\theta) \Big|_{\theta=\tilde{\theta}}, \quad (86)$$

evaluated at $\tilde{\theta}$ lying on the Euclidean path between $\hat{\theta}_N$ and $\bar{\theta}$. Let us now take a closer look at the two terms in the LHS of eq. (85). The first term $\nabla_{\theta} Q_N(\hat{\theta}_N) = 0$ by definition of $\hat{\theta}_N$ given in (84). The second term can be evaluated as follows:

$$\begin{aligned} \nabla_{\theta} Q_N(\bar{\theta}) &= -\frac{1}{N} \sum_{k=N_1}^{N_2} [(x_k - f_k(\bar{\theta}))^* \nabla_{\theta} f_k(\bar{\theta}) + (x_k - f_k(\bar{\theta})) \nabla_{\theta}^* f_k(\bar{\theta})] \\ &= -\frac{2}{N} \sum_{k=N_1}^{N_2} \text{Re} \{ n_k^* \nabla_{\theta} f_k(\bar{\theta}) \}. \end{aligned} \quad (87)$$

By substituting the previous results in (85), we have:

$$\sqrt{N} (\hat{\theta}_N - \tilde{\theta}) = \Omega_N^{-1}(\tilde{\theta}) \left[-\frac{2}{\sqrt{N}} \sum_{k=N_1}^{N_2} \text{Re} \{ n_k^* \nabla_{\theta} f_k(\tilde{\theta}) \} \right] + o_P(1), \quad (88)$$

where the matrix $\Omega_N(\tilde{\theta})$ can be evaluated as:

$$\begin{aligned} \Omega_N(\tilde{\theta}) &= \nabla_{\theta} \nabla_{\theta}^{\top} Q_N(\theta) \Big|_{\theta=\tilde{\theta}} \\ &= -\frac{1}{N} \sum_{k=N_1}^{N_2} \left[(x_k - f_k(\tilde{\theta}))^* \nabla_{\theta} \nabla_{\theta}^{\top} f_k(\tilde{\theta}) - \nabla_{\theta} f_k(\tilde{\theta}) \nabla_{\theta}^H f_k(\tilde{\theta}) \right] + \times \\ &\quad \times -\frac{1}{N} \sum_{k=N_1}^{N_2} \left[(x_k - f_k(\tilde{\theta})) \left[\nabla_{\theta} \nabla_{\theta}^{\top} f_k(\tilde{\theta}) \right]^* - \nabla_{\theta}^* f_k(\tilde{\theta}) \nabla_{\theta}^{\top} f_k(\tilde{\theta}) \right] \\ &= -\frac{2}{N} \sum_{k=N_1}^{N_2} \text{Re} \left\{ (x_k - f_k(\tilde{\theta})) \left[\nabla_{\theta} \nabla_{\theta}^{\top} f_k(\tilde{\theta}) \right]^* \right\} + \frac{2}{N} \sum_{k=N_1}^{N_2} \text{Re} \left\{ \nabla_{\theta} f_k(\tilde{\theta}) \nabla_{\theta}^H f_k(\tilde{\theta}) \right\}. \end{aligned} \quad (89)$$

From the consistency result, we have that, since $\hat{\theta}_N \xrightarrow{a.s.} \bar{\theta}$, then $\tilde{\theta} \rightarrow \bar{\theta}$ as $N \rightarrow \infty$. Moreover, using again the generalization of the SLLN [1, Theo. 2.3], we have that:

$$-\frac{2}{N} \sum_{k=N_1}^{N_2} \text{Re} \left\{ (x_k - f_k(\tilde{\theta})) \left[\nabla_{\theta} \nabla_{\theta}^{\top} f_k(\tilde{\theta}) \right]^* \right\} \xrightarrow{a.s.} -\sum_{k=N_1}^{N_2} \text{Re} \left\{ E_{p_n} [x_k - f_k(\bar{\theta})] \left[\nabla_{\theta} \nabla_{\theta}^{\top} f_k(\bar{\theta}) \right]^* \right\} = \mathbf{0}, \quad (90)$$

since $E_{p_n} [x_k - f_k(\bar{\theta})] = E_{p_n} [n_k] = 0 \forall k$ from Assumption 1. As a consequence, we immediately have that:

$$\Omega_N(\tilde{\theta}) \xrightarrow{a.s.} \frac{2}{N} \sum_{k=N_1}^{N_2} \text{Re} \left\{ \nabla_{\theta} f_k(\bar{\theta}) \nabla_{\theta}^H f_k(\bar{\theta}) \right\} = \mathbf{K}(\bar{\theta}), \quad (91)$$

where the matrix $\mathbf{K}(\bar{\theta})$ as been introduced in eq. (29) of our main paper.

Let us now evaluate the covariance matrix of $\sqrt{N} \nabla_{\theta} Q_N(\bar{\theta})$ in (87) as (for ease of notation we will indicate

$\nabla_{\boldsymbol{\theta}} f_k(\bar{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}} \bar{f}_k$:

$$\begin{aligned}
\boldsymbol{\Psi}_N(\bar{\boldsymbol{\theta}}) &\triangleq E_{p_n} \left[\left[\sqrt{N} \nabla_{\boldsymbol{\theta}} Q_N(\bar{\boldsymbol{\theta}}) \right] \left[\sqrt{N} \nabla_{\boldsymbol{\theta}} Q_N(\bar{\boldsymbol{\theta}}) \right]^\top \right] \\
&= E_{p_n} \left[\left[\frac{2}{\sqrt{N}} \sum_{k=N_1}^{N_2} \operatorname{Re} \{ n_k^* \nabla_{\boldsymbol{\theta}} \bar{f}_k \} \right] \left[\frac{2}{\sqrt{N}} \sum_{j=N_1}^{N_2} \operatorname{Re} \{ n_j^* \nabla_{\boldsymbol{\theta}} \bar{f}_j \} \right]^\top \right] \\
&= \frac{4}{N} \sum_{k=N_1}^{N_2} \sum_{j=N_1}^{N_2} E_{p_n} \left[\operatorname{Re} \{ n_k^* \nabla_{\boldsymbol{\theta}} \bar{f}_k \} \operatorname{Re} \{ n_j^* \nabla_{\boldsymbol{\theta}}^\top \bar{f}_j \} \right] \\
&= \frac{4}{N} \sum_{k=N_1}^{N_2} \sum_{j=N_1}^{N_2} E_{p_n} \left[\left(\frac{n_k^* \nabla_{\boldsymbol{\theta}} \bar{f}_k + n_k \nabla_{\boldsymbol{\theta}}^* \bar{f}_k}{2} \right) \left(\frac{n_j^* \nabla_{\boldsymbol{\theta}}^\top \bar{f}_j + n_j \nabla_{\boldsymbol{\theta}}^H \bar{f}_j}{2} \right) \right] \\
&= \frac{2}{N} \sum_{k=N_1}^{N_2} \sum_{j=N_1}^{N_2} \left[\operatorname{Re} \left\{ (E_{p_n} [n_k n_j])^* \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^\top \bar{f}_j \right\} + \operatorname{Re} \left\{ E_{p_n} [n_k^* n_j] \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^H \bar{f}_j \right\} \right] \\
&= \frac{2}{N} \sum_{k=N_1}^{N_2} E_{p_n} [|n_k|^2] \operatorname{Re} \left\{ \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^H \bar{f}_k \right\} + \frac{2}{N} \sum_{k=N_1}^{N_2} \sum_{\substack{j=N_1 \\ j \neq k}}^{N_2} \operatorname{Re} \left\{ E_{p_n} [n_k^* n_j] \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^H \bar{f}_j \right\} \\
&= \frac{2}{N} \sum_{k=N_1}^{N_2} E_{p_n} [|n_k|^2] \operatorname{Re} \left\{ \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^H \bar{f}_k \right\} + \frac{4}{N} \sum_{j=1}^{N_2-N_1} \sum_{k=N_1}^{N_2-j} \operatorname{Re} \left\{ E_{p_n} [n_{k+j}^* n_k] \nabla_{\boldsymbol{\theta}} \bar{f}_{k+j} \nabla_{\boldsymbol{\theta}}^H \bar{f}_k \right\} \\
&= \frac{2}{N} \sum_{k=N_1}^{N_2} r_n[0] \operatorname{Re} \left\{ \nabla_{\boldsymbol{\theta}} \bar{f}_k \nabla_{\boldsymbol{\theta}}^H \bar{f}_k \right\} + \frac{4}{N} \sum_{j=1}^{N_2-N_1} \sum_{k=N_1}^{N_2-j} \operatorname{Re} \left\{ r_n[j] \nabla_{\boldsymbol{\theta}} \bar{f}_{k+j} \nabla_{\boldsymbol{\theta}}^H \bar{f}_k \right\} = \mathbf{P}(\bar{\boldsymbol{\theta}}),
\end{aligned} \tag{92}$$

where the matrix $\mathbf{P}(\bar{\boldsymbol{\theta}})$ as been introduced in eq. (27f) of the main document.

Finally, by recalling that $\boldsymbol{\Omega}_N(\tilde{\boldsymbol{\theta}}) \xrightarrow{a.s.} \mathbf{K}(\bar{\boldsymbol{\theta}})$, from a direct application of the Central Limit Theorem (CLT) for dependent random variables satisfying Assumption 1 [1, Theo. 2.4] to the Taylor expansion in (85), we get:

$$\sqrt{N} [\mathbf{P}(\bar{\boldsymbol{\theta}})]^{-1/2} \mathbf{K}(\bar{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}_N - \bar{\boldsymbol{\theta}} \right) \underset{N \rightarrow \infty}{\sim} \mathcal{N}(0, \mathbf{I}). \tag{93}$$

that implies the asymptotic normality of $\hat{\boldsymbol{\theta}}_N$ in (76),

REFERENCES

- [1] H. White and I. Domowitz, "Nonlinear regression with dependent observations," *Econometrica*, vol. 52, no. 1, pp. 143–161, 1984.
- [2] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.
- [3] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," in *ISI World Statistics Congress 2017 (ISI2017)*, 2017. [Online]. Available: <https://arxiv.org/abs/0906.4835>
- [4] R. Remmert, *Theory of Complex Functions*. New York: Springer, 1991.