Asymptotic Properties of Statistical Estimators using Multivariate Chi-squared Measurements

Damián Marelli[a,b], Minyue Fu[c,a]

[a] *School of Automation, Guandong University of Technology, Guangzhou, China.*
[b] *French Argentine International Center for Information and Systems Sciences, National Scientific and Technical Research Council, Rosario, Argentina.*
[c] *School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW, Australia.*

Abstract

This paper studies the problem of estimating a parameter vector from measurements having a multivariate chi-squared distribution. Maximum likelihood estimation in this setting is unfeasible because the multivariate chi-squared distribution has no closed form expression. The typical approach to go around this consists in considering a sub-optimal solution by replacing the chi-squared distribution with a normal one. We investigate the theoretical properties of this approximation as the number of measurements approach infinity. More precisely, we show that this approximation is strongly consistency, asymptotically normal and asymptotically efficient. We consider a source localization problem as a case study.

Keywords: Asymptotic statistical properties, multivariate chi-squared distribution, parameter estimation, maximum likelihood estimation, Cramér-Rao lower bound.

1. Introduction

Statistical estimation [1] finds abundant applications in signal processing, communications and control [2, 3]. It consists in providing an estimate of a given set of parameters, based on a set of measurements whose probability distribution depends on those parameters. In this work, we are concerned with estimation problems where measurements have a multivariate chi-square distribution. This is the distribution of the vector of component-wise square sums of i.i.d. normal random vectors with zero means and arbitrary covariance matrix. A number of estimation problem in engineering meet this assumption. Generally speaking, these are problems where measurements are in the form of a vector of received powers. Examples of this are source localization [4, 5, 6, 7, 8, 9], network localization [10, 11, 12], target tracking [13, 14, 15], etc.

A preferred statistical estimation method is called maximum likelihood (ML). This is because ML estimates enjoy a number of asymptotic statistical properties. More precisely, if certain regularity conditions are satisfied, as the number of available samples used to build the measurements tends to infinity, the sequence of ML estimates is: *strongly consistent*, i.e., it converges with probability one (w.p.1) to the true value, *asymptotically normal*, i.e., the distribution

☆This work was partially supported by the Argentinean Agency for Scientific and Technological Promotion (PICT- 201-0985) and the National Natural Science Foundation of China (Grant Nos. 61633014, 61803101 and U1701264).

Email addresses: Damian.Marelli@newcastle.edu.au (Damián Marelli), Minyue.Fu@newcastle.edu.au (Minyue Fu)

of the estimation error converges to a normal one, and *asymptotically efficient*, i.e., the asymptotic covariance of the estimation error equals the inverse of the asymptotic Fisher information matrix (AFIM), and therefore it attains the Cramér-Rao lower bound (CRLB) [16, 17].

For estimation problems with measurements having a multivariate chi-square distribution, we face the problem that the probability density function (PDF) has no closed-form expression [18]. To go around this difficulty, practical estimation methods typically use the central limit theorem to approximate the multivariate chi-square distribution by a multivariate normal one [5, 8, 6, 9, 7]. This leads to a computationally feasible but sub-optimal estimate which we call *quasi-ML*. The goal of this work is to provide a mathematical backing for quasi-ML estimation. More precisely, we give conditions under which quasi-ML estimates enjoy the same aforementioned asymptotic properties enjoyed by ML estimates. A consequence of our result is that, if many measurements are available, the computational advantages of the quasi-ML estimate come with negligible performance loss.

The rest of the paper is organized as follows: In Section 2 we describe the research problem. In Section 3 we state our main results, namely the conditions required for strong consistency, asymptotic normality and asymptotic efficiency of the quasi-ML estimate. After presenting some preliminary results in Section 4.1, we provide the proofs of the three main theorems in Sections 4.2, 4.3 and 4.4, respectively. Finally, in Section 5 we show how our results permit asserting the asymptotic properties of the quasi-ML estimator in a case study, namely, a source localization problem.

Notation 1. For a vector x we use $[x]_i$ or x_i to denote its i-th entry, $\|x\|$ to denote its 2-norm, and $\operatorname{diag}(x)$ to denote the diagonal matrix whose diagonal entries equal the entries of x. For a matrix X, we use $[X]_{i,j}$ or $X_{i,j}$ to denote its (i,j)-th entry, $\|X\|$ to denote its operator norm, $\operatorname{diag}(X)$ to denote the vector whose entries equal the diagonal entries of X, $\operatorname{Tr}\{X\}$ to denote its trace, and $\overrightarrow{X}$ to denote the vector obtained by stacking the columns of X. The superscript $\cdot^\top$ denotes transposition. We use $\mathbf{I}$ to denote the identity matrix and $\mathcal{P}^I(\mathbb{R}) \subset \mathbb{R}^{I\times I}$ to denote the set of real positive definite $I \times I$ matrices. For a random vector x we use $\mathcal{E}\{x\}$ to denote its expected value, and $\mathcal{C}\{x\}$ to denote its covariance matrix. We use $\mathcal{N}(\mu, \Sigma)$ to denote the multivariate normal distribution with mean μ and covariance Σ, and $\mathcal{N}(x; \mu, \Sigma)$ to denote the PDF of that distribution evaluated at x.

2. Problem description

2.1. Estimation problem

We start by introducing the multivariate chi-square distribution [18].

Definition 1. An I-dimensional, real, random vector s is said to have a multivariate chi-squared distribution, with $N \in \mathbb{N}$ degrees of freedom and parameter matrix $\Sigma \in \mathcal{P}^I(\mathbb{R})$, denoted $\chi^2(N, \Sigma)$, if its PDF p is given by

$$p(s) = \frac{1}{(2\pi)^I} \int \exp\left(-it^\top s\right) \phi(t) dt, \tag{1}$$

where $\phi(t) = \det(I - 2i\Sigma T)^{-N/2}$ denotes the characteristic function of p, with $t = [t_1, \ldots, t_I]^\top$ and $T = \operatorname{diag}(t)$.

We assume that we have a sequence of random vector samples $x_n = [x_{n,1}, \cdots, x_{n,I}]^\top \in \mathbb{R}^I$, $n = 1, \cdots, N$, independently drawn from the distribution

$$x_n \sim \mathcal{N}(0, \Sigma(\theta_\star)), \tag{2}$$

where $\Sigma : \mathcal{D} \to \mathcal{P}^I(\mathbb{R})^1$, $\mathcal{D} \subseteq \mathbb{R}^D$ and $\theta_\star \in \mathcal{D}$ is the unknown true value of the parameter $\theta = [\theta_1, \cdots, \theta_D]^\top \in \mathcal{D}$. We use these samples to obtain an estimate of the variance of each entry of x_n. To this end, we build the vector measurement s_N, whose i-th entry $s_{N,i}$ is given by

$$s_{N,i} = \frac{1}{N} \sum_{n=1}^{N} x_{n,i}^2. \tag{3}$$

As it is known, $s_N \sim \chi^2\left(N, \frac{1}{N}\Sigma(\theta_\star)\right)$ [19]. The estimation problem is to estimate the unknown $\theta_\star$ using s_N.

2.2. Motivation

The estimation problem described in Section 2.1 appears in a number of applications where certain parameter $\theta_\star$ (e.g., the position of a target) needs to be estimated based on the knowledge of a measurement of the strength (power) of signals received at a number of sensors. More precisely, we assume that we have I sensors. At time N, Sensor i measures the signal $x_{N,i}$, and updates an estimate $s_{N,i}$ of its power computing (3). Let $x_n^\top = [x_{n,1}, \cdots, x_{n,I}]$ and $s_N^\top = [s_{N,1}, \cdots, s_{N,I}]$. We assume that, at a given sample time n, measurements from different sensors are possibly statistically dependent, but measurements taken at different sample times are independent. To model this, we assume that vector samples $x_n \in \mathbb{R}^I$, $n = 1, \cdots, N$, are independently drawn from (2). Then, the sequence $s_N \sim \chi^2\left(N, \frac{1}{N}\Sigma(\theta_\star)\right)$, $N \in \mathbb{N}$, satisfies the conditions described in Section 2.1.

2.3. Maximum likelihood estimation

A preferred approach to solve the above consists in using the ML criterion, i.e.,

$$\hat{\theta}_N^{\mathrm{ML}} = \arg\max_{\theta \in \mathcal{D}} L_N(\theta), \tag{4}$$

where

$$L_N(\theta) = \frac{1}{N} \log p_\theta(s_N), \tag{5}$$

and $p_\theta(s_N)$ denotes the PDF of s_N, parameterized by θ. This is because ML estimates typically enjoy certain asymptotic statistical properties, namely, *strong consistency*, *asymptotic normality* and *asymptotic efficiency*, which are properly defined below.

As mentioned in Section 2.1, the PDF $p_\theta(s_N)$ follows a multivariate chi-squared law. A stumbling block for solving (4) in practice is that (1) does not have a closed-form expression [18]. A popular approach [5, 8, 6, 9, 7] to go around this difficulty consists in replacing the ML estimation criterion by an approximate one, which we call quasi-ML estimation criterion, and describe below.

2.4. Quasi-maximum likelihood estimation

The essential idea is to make use of the central limit theorem [20, Th. 15.56] to approximate the PDF $p_\theta(s_N)$ using a multivariate normal distribution. We assume that the distribution of s_N is, instead of (1), given by the following one

$$s_N \sim \mathcal{N}\left(\mu(\theta_\star), \frac{1}{N} C(\theta_\star)\right), \tag{6}$$

[1]Notice that the definition of Σ guarantees that $\Sigma(\theta)$ is positive definite for all $\theta \in \mathcal{D}$.

with

$$\mu(\theta) = \mathcal{E}_\theta\{s_N\} \qquad \text{and} \qquad C(\theta) = N\mathcal{C}_\theta\{s_N\},$$

where $\mathcal{E}_\theta$ and $\mathcal{C}_\theta$ denote expectation and covariance, respectively, with respect to p_θ. Then, following (4), the ML estimate (4) is replaced by the following one, which we call the quasi-ML estimate

$$\hat{\theta}_N = \arg\max_{\theta \in \mathcal{D}} \tilde{L}_N(\theta), \tag{7}$$

where

$$\tilde{L}_N(\theta) = \frac{1}{N} \log \tilde{p}_\theta(s_N), \tag{8}$$

with

$$\tilde{p}_\theta(s_N) = \mathcal{N}\left(s_N; \mu(\theta), \frac{1}{N} C(\theta)\right). \tag{9}$$

The quasi-ML estimate (7)-(9) can be interpreted as an approximation to the ML estimate. This is because the former is derived under the simplifying assumption that the measurements s_N have the multivariate normal distribution $\tilde{p}_\theta$, while the actual distribution p_θ follows a multivariate chi-squared law. The advantage of the quasi-ML estimate, over the ML one, is that $\tilde{p}_\theta$ has a closed-form. This permits computing the quasi-ML estimate using standard numerical optimization techniques.

Before concluding this section we derive the expressions of the parameters $\mu(\theta)$ and $C(\theta)$ of the normal approximation (6). To this end we use the following lemma.

Lemma 1. *If $[x, y]^\top \sim \mathcal{N}(0, \Sigma)$, then*

$$\mathcal{E}\left\{x^2 y^2\right\} = 2\mathcal{E}\{xy\}^2 + \mathcal{E}\left\{x^2\right\}\mathcal{E}\left\{y^2\right\}.$$

Proof. Let $[u, v]^\top = \Sigma^{-1/2}[x, y]^\top$, and $\Sigma^{1/2} = \begin{bmatrix} a_{x,u} & a_{x,v} \\ a_{y,u} & a_{y,v} \end{bmatrix}$. We can then write $x = a_{x,u}u + a_{x,v}v$ and $y = a_{y,u}u + a_{y,v}v$, with $[u, v]^\top \sim \mathcal{N}(0, \mathbf{I})$. The result then follows since

$$\begin{aligned} &\mathcal{E}\left\{x^2\right\} = a_{x,u}^2 + a_{x,v}^2, \quad \mathcal{E}\left\{y^2\right\} = a_{y,u}^2 + a_{y,v}^2, \quad \mathcal{E}\{xy\} = a_{x,u}a_{y,u} + a_{x,v}a_{y,v}, \\ &\mathcal{E}\left\{x^2 y^2\right\} = 2\left(a_{x,u}a_{y,u} + a_{x,v}a_{y,v}\right)^2 + \left(a_{x,u}^2 + a_{x,v}^2\right)\left(a_{y,u}^2 + a_{y,v}^2\right). \end{aligned}$$

□

The desired expressions are then stated in the following proposition.

Proposition 1. *For all $i, j = 1, \cdots, I$,*

$$\mu_i(\theta) = \Sigma_{i,i}(\theta) \qquad \text{and} \qquad C_{i,j}(\theta) = 2\Sigma_{i,j}^2(\theta). \tag{10}$$

Moreover, $C(\theta) \in \mathcal{P}^I(\mathbb{R})$, for all $\theta \in \mathcal{D}$, i.e., $\Sigma : \mathcal{D} \to \mathcal{P}^I(\mathbb{R})$.

Proof. We have

$$\mu_i(\theta) = [\mathcal{E}\{s_N\}]_i = \mathcal{E}\left\{x_{n,i}^2\right\} = \Sigma_{i,i}(\theta_\star).$$

Also

$$\begin{aligned} [\mathcal{C}\{s_N\}]_{i,j} &= \frac{1}{N^2} \sum_{n,m=1}^{N} \left[\mathcal{E}\left\{x_{n,i}^2 x_{m,j}^2\right\} - \Sigma_{i,i}(\theta_\star)\Sigma_{j,j}(\theta_\star)\right] \\ &= \frac{1}{N^2} \sum_{n=1}^{N} \left[\mathcal{E}\left\{x_{n,i}^2 x_{n,j}^2\right\} - \Sigma_{i,i}(\theta_\star)\Sigma_{j,j}(\theta_\star)\right] = \frac{1}{N}\left[\mathcal{E}\left\{x_{1,i}^2 x_{1,j}^2\right\} - \Sigma_{i,i}(\theta_\star)\Sigma_{j,j}(\theta_\star)\right]. \end{aligned} \tag{11}$$

Now, from Lemma 1 $\mathcal{E}\left\{x_{1,i}^2 x_{1,j}^2\right\} = 2\Sigma_{i,j}^2\left(\theta_\star\right) + \Sigma_{i,i}\left(\theta_\star\right)\Sigma_{j,j}\left(\theta_\star\right)$, and (10) follows by putting this in (11). The last part follows from the Schur product theorem since $\Sigma(\theta) \in \mathcal{P}^I(\mathbb{R})$, for all $\theta \in \mathcal{D}$. □

2.5. Research problem

In view of the above, the question arises as to whether the practically feasible quasi-ML estimate enjoys the following asymptotic statistical properties, which are often enjoyed by ML estimates.

Strong consistency: whether the estimate $\hat{\theta}_N$ converges with probability one (w.p.1) to the true value $\theta_\star$, i.e.,

$$\lim_{N\to\infty} \hat{\theta}_N \overset{\text{w.p.1}}{=} \theta_\star.$$

Asymptotic normality: whether the normalized estimation error $\sqrt{N}\left(\hat{\theta}_{\mathrm{N}} - \theta_\star\right)$ converges in distribution to a normal law, i.e.,

$$\lim_{N\to\infty} \sqrt{N}\left(\hat{\theta}_{\mathrm{N}} - \theta_\star\right) \overset{\text{in dist.}}{=} \mathcal{N}\left(0, \Psi\right),$$

for some matrix $\Psi > 0$ called the *asymptotic covariance.*

Asymptotic efficiency: whether the asymptotic covariance Ψ equals the CRLB, i.e.,

$$\Psi = \mathcal{I}^{-1},$$

where

$$\mathcal{I} = \lim_{N\to\infty} N\mathcal{E}\left\{\nabla L_N\left(\theta_\star\right)\nabla^\top L_N\left(\theta_\star\right)\right\},$$

denotes the AFIM and ∇ denotes the gradient operator.

Remark 1. The three asymptotic properties described above are of particular interest in practice. A number of theoretical results are available giving conditions for these properties to hold [21, Section 24], [22, Sections 3 and 5]. Strong consistency means that the sequence of estimations $\hat{\theta}_N$ always converges to the true value $\theta_\star$, as the number N of samples tends to infinity. Asymptotic normality means that, for large N, the estimation error $\hat{\theta}_{\mathrm{N}} - \theta_\star$ can be approximated by a multivariate normal vector. This facilitates the analysis of systems involving this error. Finally, asymptotic efficiency asserts that, for large N, no unbiased estimator can be better, in the sense of having a smaller covariance.

Our goal is to provide conditions to guarantee that the quasi-ML estimate $\hat{\theta}_{\mathrm{N}}$ satisfies the three asymptotic statistical properties described above.

3. Main results

In this section we state our main results. Before doing so, we introduce an example to illustrate the fact that conditions are indeed needed for the quasi-ML estimate to enjoy the desired properties. To this end, we introduce an example in which the quasi-ML fails to be strongly consistent.

Example 1. Let $x_n \sim \mathcal{N}\left(0, \Sigma\left(\theta_\star\right)\right)$, $n \in \mathbb{N}$, be a sequence of real, scalar variables, with $\Sigma : (0,2) \to (0,2)$ being a discontinuous function of θ given by

$$\Sigma\left(\theta\right) = \begin{cases} \theta, & 0 < \theta \le 1, \\ 3-\theta, & 1 < \theta < 2, \end{cases} \tag{12}$$

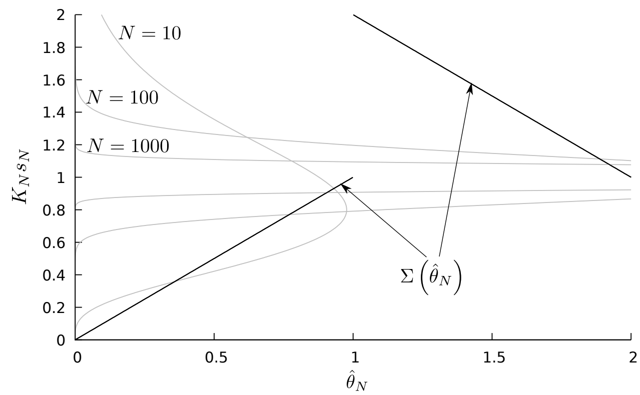

Figure 1: Convergence of the PDF of $K_N s_N$ and the map $\Sigma^{-1} : K_N s_N \mapsto \hat{\theta}_N$.

and $\theta_\star = 1$. Let s_N be defined as in (3). We will show that the estimate $\hat{\theta}_N$ produced by the quasi-ML estimation method fails to converge to $\theta_\star$, with probability one, as $N \to \infty$. To this end we will first show that

$$\hat{\theta}_N = \Sigma^{-1}\left(K_N s_N\right), \tag{13}$$

for some constant K_N. We will then show that the PDF of the random variable $K_N s_N$ converges in a way that half its probability mass falls below 1, and the other half sits above. This is depicted in Figure 1. Since Σ is discontinuous at 1, this leads to

$$\lim_{N\to\infty} p_{\theta_\star}\left(\hat{\theta}_N < 1\right) = \frac{1}{2} \qquad \text{and} \qquad \lim_{N\to\infty} p_{\theta_\star}\left(\frac{3}{2} < \hat{\theta}_N < 2\right) = \frac{1}{2}. \tag{14}$$

Then, for every $0 < \epsilon < \frac{3}{2}$,

$$\begin{aligned}\lim_{N\to\infty} p_{\theta_\star}\left(\left|\hat{\theta}_N - \theta_\star\right| < \epsilon\right) &= 1 - \lim_{N\to\infty} p_{\theta_\star}\left(\left|\hat{\theta}_N - \theta_\star\right| \geq \epsilon\right) \\ &\leq 1 - \lim_{N\to\infty} p_{\theta_\star}\left(\frac{3}{2} < \hat{\theta}_N < 2\right) = \frac{1}{2}.\end{aligned}$$

Hence, $\hat{\theta}_N$ does not converge to $\theta_\star$ in probability. It then follows from [23, Theorem 4.1.2] that $\hat{\theta}_N$ cannot converge to $\theta_\star$ with probability 1.

We have that

$$\mathcal{E}\left\{s_N\right\} = \Sigma\left(\theta_\star\right) \qquad \text{and} \qquad \mathcal{C}\left\{s_N\right\} = \frac{2\Sigma\left(\theta_\star\right)^2}{N}.$$

Hence, from (6) we obtain

$$\tilde{p}_\theta\left(s_N\right) = \mathcal{N}\left(s_N; \Sigma\left(\theta_\star\right), \frac{2}{N}\Sigma\left(\theta_\star\right)^2\right).$$

The quasi-ML estimate is then given by

$$\hat{\theta}_N = \arg\max_{\theta\in\mathcal{D}} \tilde{L}_N(\theta) = \arg\max_{\theta\in\mathcal{D}} \frac{1}{N}\left[-\frac{1}{2}\log\frac{4\pi}{N} - \log\Sigma\left(\theta\right) - \frac{N}{4}\left(\frac{s_N}{\Sigma\left(\theta\right)} - 1\right)^2\right]$$

Tanking the derivative with respect to $\Sigma(\theta)$, we obtain

$$\frac{\partial \tilde{L}_N\left(\hat{\theta}_N\right)}{\partial \Sigma\left(\theta\right)} = -\left[\frac{2}{N}\Sigma^2\left(\theta\right) + s_N\Sigma\left(\theta\right) - s_N^2\right]\frac{1}{2\Sigma\left(\theta\right)^3} = 0.$$

The above is solved when

$$\Sigma\left(\hat{\theta}_N\right) = K_N s_N \qquad \text{with} \qquad K_N = \frac{N}{4}\left(\sqrt{1+\frac{8}{N}} - 1\right). \tag{15}$$

Since the map Σ is invertible, equation (13) follows.

Let $r_N = N s_N$. Clearly, $r_N \sim \chi^2\left(N\right)$. Let

$$F_N\left(x\right) = p_{\theta_\star}\left(r_N < x\right) = \frac{1}{\Gamma\left(\frac{N}{2}\right)}\gamma\left(\frac{N}{2}, \frac{x}{2}\right)$$

denote the CDF of r_N, with Γ being the gamma function and γ being the lower incomplete gamma function. Equation (14) then follows from (13) since

$$p_{\theta_\star}\left(\hat{\theta}_N < 1\right) = p_{\theta_\star}\left(K_N s_N < 1\right) = p_{\theta_\star}\left(r_N < \frac{N}{K_N}\right) = F_N\left(\frac{N}{K_N}\right) \to \frac{1}{2},$$
$$p_{\theta_\star}\left(\frac{3}{2} < \hat{\theta}_N < 2\right) = p_{\theta_\star}\left(1 < K_N s_N < \frac{3}{2}\right) = F_N\left(\frac{3}{2}\frac{N}{K_N}\right) - F_N\left(\frac{N}{K_N}\right) \to \frac{1}{2}.$$

We now state our main results. Their proofs are delayed to later sections. Our first result states conditions under which the quasi-ML estimate $\hat{\theta}_N$ enjoys strong consistency. Its proof appears in Section 4.2.

Theorem 1 (Strong consistency)**.** *If:*

1. The set $\mathcal{D}$ is compact;
2. The map $\Sigma : \mathcal{D} \to \mathcal{P}^I\left(\mathbb{R}\right)$ is continuous;
3. For all $\theta \in \mathcal{D}$, $\operatorname{diag}\left(\Sigma\left(\theta\right)\right) = \operatorname{diag}\left(\Sigma\left(\theta_\star\right)\right)$ if and only if $\theta = \theta_\star$;
   then

$$\lim_{N\to\infty} \hat{\theta}_N \overset{\text{w.p.1}}{=} \theta_\star. \tag{16}$$

Our next result states conditions to guarantee the asymptotic normality of the quasi-ML estimate. Its proof appears in Section 4.3.

Theorem 2 (Asymptotic normality)**.** *If, in addition to the conditions of Theorem 1:*

1. The true vector of parameters lies in the interior $\operatorname{int}(\mathcal{D})$ of $\mathcal{D}$, i.e., $\theta_\star \in \operatorname{int}(\mathcal{D})$;
2. The map $\Sigma : \mathcal{D} \to \mathcal{P}^I\left(\mathbb{R}\right)$ is twice continuously differentiable on $\operatorname{int}(\mathcal{D})$;
3. The Jacobian $\mathbf{J}_\mu\left(\theta_\star\right)$ of the map $\mu : \theta \mapsto \operatorname{diag}\left(\Sigma\left(\theta\right)\right)$ evaluated at $\theta_\star$ has full column rank;
   then

$$\lim_{N\to\infty} \sqrt{N}\left(\hat{\theta}_{\mathrm{N}} - \theta_\star\right) \overset{\text{in dist.}}{=} \mathcal{N}\left(0, \tilde{\mathcal{I}}^{-1}\right), \tag{17}$$

   where

$$\tilde{\mathcal{I}} = \lim_{N\to\infty} \frac{1}{N}\mathcal{E}\left\{\left.\nabla_\theta \log \tilde{p}_\theta\left(s_N\right)\right|_{\theta=\theta_\star} \left.\nabla_\theta^\top \log \tilde{p}_\theta\left(s_N\right)\right|_{\theta=\theta_\star}\right\}. \tag{18}$$

The above results assert the asymptotic consistency and normality of the quasi-ML estimate. In particular, Theorem 2 states that the asymptotic covariance $\tilde{\mathcal{I}}^{-1}$ of this estimate is given by (18). The next theorem states that $\tilde{\mathcal{I}}$ equals the AFIM $\mathcal{I}$, from where it follows that the quasi-ML estimate is asymptotically efficient. The proof of this result appears in Section 4.4.

Theorem 3 (Asymptotic efficiency)**.** *If the conditions of Theorem 2 hold,* $\tilde{\mathcal{I}} = \mathcal{I}$.

We finish by giving the expression of the AFIM.

Proposition 2. *If the conditions of Theorem 2 hold,*

$$\mathcal{I} = \mathbf{J}_{\mu}\left(\theta_{\star}\right)^{\top} C^{-1}\left(\theta_{\star}\right) \mathbf{J}_{\mu}\left(\theta_{\star}\right),$$

where $\mathbf{J}_{\mu}(\theta)$ *denotes the Jacobian of* $\mu(\theta)$*, i.e.,*

$$\left[\mathbf{J}_{\mu}(\theta)\right]_{i,d} = \frac{\partial \mu_i(\theta)}{\partial \theta_d}.$$

The quasi-ML estimate $\hat{\theta}_{\mathrm{N}}$ introduced in Example 1 is not strongly consistent. Notice that Assumptions 1 and 2 of Theorem 1 are not satisfied in that example. In particular, Assumption 2 fails. In Example 2 below, we modify Example 1 to satisfy the assumptions of Theorem 1, and show how this leads to a strongly consistent quasi-ML estimate.

Example 2. Suppose that the definition of Σ in (12) is replaced by any invertible map $\Sigma : [0,2] \to [0,\infty)$ with continuous inverse. From (15) we have $\lim_{N\to\infty} K_N = 1$. Also, from the Kolmogorov's strong law of large numbers [23, Th. 5.4.2], $\lim_{N\to\infty} s_N \overset{\text{w.p.1}}{=} 1$. It then follows from (13) and the continuity of Σ^{-1}, that

$$\lim_{N\to\infty} \hat{\theta}_N = \lim_{N\to\infty} \Sigma^{-1}\left(K_N s_N\right) \overset{\text{w.p.1}}{=} \Sigma^{-1}(1),$$

or equivalently, that $\hat{\theta}_N$ is strongly consistent.

4. Proofs of the main results

4.1. Preliminary results

4.1.1. Log-likelihood function of the normal approximation and its derivatives

We start by giving the expressions of $\tilde{L}_N(\theta)$, its gradient and its Hessian.

Lemma 2. *Let* θ_d*,* $d = 1, \cdots, D$*, denote the* D*-th entry of* θ*. Then*

$$\tilde{L}_N(\theta) = -\frac{1}{2}\left[\frac{1}{N}\log\left|\frac{2\pi}{N}C(\theta)\right| + \operatorname{Tr}\left\{C(\theta)^{-1}\Delta_N(\theta)\right\}\right], \tag{19}$$

$$\begin{aligned}\left[\nabla \tilde{L}_N(\theta)\right]_d =& \frac{\partial \mu(\theta)}{\partial \theta_d}^{\top} C^{-1}(\theta)\,\delta_N(\theta) \\ &+ \frac{1}{2}\operatorname{Tr}\left\{C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_d}\left(C(\theta)^{-1}\Delta_N(\theta) - \frac{1}{N}I\right)\right\},\end{aligned} \tag{20}$$

$$\begin{aligned}\left[\nabla^2 \tilde{L}_N(\theta)\right]_{d,e} =& -\frac{\partial \mu(\theta)}{\partial \theta_d}^{\top} C^{-1}(\theta)\frac{\partial \mu(\theta)}{\partial \theta_e} \\ &- \frac{1}{2}\operatorname{Tr}\left\{C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_d}C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_e}\left[2C^{-1}(\theta)\Delta_N(\theta) - \frac{1}{N}I\right]\right\} \\ &+ \frac{1}{2}\operatorname{Tr}\left\{C^{-1}(\theta)\,\delta_N(\theta)\left(\frac{\partial \mu(\theta)}{\partial \theta_d}^{\top} C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_e} + \frac{\partial^2 \mu(\theta)}{\partial \theta_d \partial \theta_e}^{\top}\right)\right\}.\end{aligned} \tag{21}$$

where

$$\delta_N(\theta) = s_N - \mu(\theta), \tag{22}$$
$$\Delta_N(\theta) = \delta_N(\theta)\,\delta_N(\theta)^\top. \tag{23}$$

Proof. Equation (19) follows straightforwardly from the definition. For (20) and (21), we use the following two identities

$$\frac{\partial X(\alpha)^{-1}}{\partial \alpha} = -X(\alpha)^{-1}\frac{\partial X(\alpha)}{\partial \alpha}X(\alpha)^{-1}, \tag{24}$$
$$\frac{\partial \log|X(\alpha)|}{\partial \alpha} = \operatorname{Tr}\left\{X(\alpha)^{-1}\frac{\partial X(\alpha)}{\partial \alpha}\right\}. \tag{25}$$

The result then follows after some routine algebraic steps. □

4.1.2. Strong convergence

Definition 2. A sequence x_n, $n \in \mathbb{N}$, of random variables is said to be strongly convergent (SC) if $\lim_{n\to\infty} \mathcal{E}\{x_n\}$ exists and

$$\lim_{n\to\infty} x_n \overset{\text{w.p.1}}{=} \lim_{n\to\infty} \mathcal{E}\{x_n\}. \tag{26}$$

Lemma 3. *The sequences $\delta_N(\theta)$ and $\Delta_N(\theta)$, $N \in \mathbb{N}$, are SC, for all $\theta \in \mathcal{D}$.*

Proof. Recall the definition of δ_N from (22). We have

$$\delta_N(\theta) = \frac{1}{N}\sum_{n=0}^{N-1} \xi_n(\theta), \tag{27}$$

where $\xi_n(\theta) = [\xi_{n,1}(\theta), \ldots, \xi_{n,I}(\theta)]^\top$, with $\xi_i(\theta) = x_{n,i}^2 - \mu_i(\theta)$. It is easy to see that, for all $\theta \in \mathcal{D}$, $\sup_n \mathcal{E}\left\{\xi_{n,i}^2(\theta)\right\} < \infty$. Then, since, $\xi_n(\theta)$ and $\xi_m(\theta)$ are independent, whenever $n \neq m$, from Rajchman's strong law of large numbers [23, Th. 5.1.2], we have

$$\lim_{N\to\infty} \delta_N(\theta) - \mathcal{E}\{\delta_N(\theta)\} \overset{\text{w.p.1}}{=} 0. \tag{28}$$

□

Now

$$\lim_{N\to\infty} \mathcal{E}\{\delta_N(\theta)\} = \mu(\theta_\star) - \mu(\theta).$$

Hence, from (28),

$$\lim_{N\to\infty} \delta_N(\theta)\delta_N(\theta)^\top \overset{\text{w.p.1}}{=} (\mu(\theta_\star) - \mu(\theta))(\mu(\theta_\star) - \mu(\theta))^\top. \tag{29}$$

We also have

$$\mathcal{E}\left\{s_N s_N^\top\right\} = \frac{1}{N}C(\theta_\star) + \mu(\theta_\star)\,\mu(\theta_\star)^\top.$$

Then,

$$\begin{aligned}\mathcal{E}\left\{\delta_N(\theta)\delta_N(\theta)^\top\right\} =& \mathcal{E}\left\{\left(s_N-\mu(\theta)\right)\left(s_N-\mu(\theta)\right)^\top\right\}\\ =&\mathcal{E}\left\{s_N s_N^\top\right\}-\mu\left(\theta_\star\right)\mu\left(\theta\right)^\top-\mu\left(\theta\right)\mu\left(\theta_\star\right)^\top+\mu(\theta)\mu(\theta)^\top\\ =&\frac{1}{N}C\left(\theta_\star\right)+\left(\mu\left(\theta_\star\right)-\mu(\theta)\right)\left(\mu\left(\theta_\star\right)-\mu(\theta)\right)^\top.\end{aligned}\tag{30}$$

From (30) and (29), we obtain

$$\lim_{N\to\infty}\delta_N(\theta)\delta_N(\theta)^\top-\mathcal{E}\left\{\delta_N(\theta)\delta_N(\theta)^\top\right\}\overset{\text{w.p.1}}{=}0,\tag{31}$$

and the result follows from (28) and (31).

Lemma 4. *Under Assumption 2 of Theorem 2, the sequences $\tilde{L}_N(\theta)$ and $\nabla\tilde{L}_N(\theta)$ are SC, for all $\theta\in\mathcal{D}$.*

Proof. The result follows immediately by using Lemma 3 in (19) and (20). □

4.1.3. Strong uniform convergence

Definition 3. Let $\mathcal{D}\subseteq\mathbb{R}^d$. A sequence of stochastic functions $f_N:\mathcal{D}\to\mathbb{R}^q$, $N\in\mathbb{N}$, is said to be *continuous and strongly uniformly convergent* (CSUC) in $\mathcal{D}$ if $f(x)=\lim_{N\to\infty}\mathcal{E}\left\{f_N(x)\right\}$ exists and, w.p.1, every f_N is a continuous function on $\mathcal{D}$, and

$$\lim_{N\to\infty}\sup_{x\in\mathcal{D}}\left\|f_N(x)-f(x)\right\|\overset{\text{w.p.1}}{=}0.$$

Lemma 5. *Under Assumptions 1 of Theorem 1 and 2 of Theorem 2, $\nabla^2\tilde{L}_N(\theta)$ is CSUC on $\mathcal{D}$.*

Proof. We have

$$\left\|\delta_N(\theta)-\mathcal{E}\left\{\delta_N(\theta)\right\}\right\|=\left\|s_N-\mathcal{E}\left\{s_N\right\}\right\|\tag{32}$$

and

$$\begin{aligned}&\left\|\delta_N(\theta)\delta_N(\theta)^\top-\mathcal{E}\left\{\delta_N(\theta)\delta_N(\theta)^\top\right\}\right\|\\ &\quad=\left\|s_N s_N^\top-\mathcal{E}\left\{s_N s_N^\top\right\}-\left(s_N-\mathcal{E}\left\{s_N\right\}\right)\mu(\theta)^\top-\mu(\theta)\left(s_N-\mathcal{E}\left\{s_N\right\}\right)^\top\right\|\\ &\qquad\le\left\|s_N s_N^\top-\mathcal{E}\left\{s_N s_N^\top\right\}\right\|+2\left\|\mu(\theta)\right\|_2\left\|\left(s_N-\mathcal{E}\left\{s_N\right\}\right)\right\|.\end{aligned}\tag{33}$$

The compactness of $\mathcal{D}$ and continuity of $\Sigma(\theta)$ imply that there exists $k_1>0$ such that $\left\|\mu(\theta)\right\|_2\le k_1$, for all $\theta\in\mathcal{D}$. Hence, from Lemma 3, the convergences of (32) and (33) are strong and uniform in θ. □

Since $\Sigma(\theta)$ is twice continuously differentiable, there exists $k_2>0$ such that

$$\left\|C(\theta)\right\|,\left\|\frac{\partial C}{\partial\theta_i}(\theta)\right\|,\left\|\frac{\partial^2 C}{\partial\theta_i\partial\theta_j}(\theta)\right\|\le k_2,\ \text{for all }\theta\in\mathcal{D}.\tag{34}$$

Also, from the same assumption, there exists $\epsilon>0$ such that

$$\left\|C(\theta)^{-1}\right\|^{-1}\ge\epsilon,\ \text{for all }\theta\in\mathcal{D}.\tag{35}$$

Now, in view of (21), equations (34) and (35) imply that, w.p.1, $\nabla^2 L_N(\theta)$ is continuous. Also, (32) and (33) imply that, w.p.1, $\nabla^2 L_N(\theta)$ converges uniformly to $\lim_{N\to\infty}\mathcal{E}\left\{\nabla^2 L_N(\theta)\right\}$. The result then follows.

4.2. Proof of Theorem 1

Lemma 6. *Let* $\bar{\tilde{L}}(\theta) = \lim_{N\to\infty} \mathcal{E}\left\{\tilde{L}_N(\theta)\right\}$*. Under Assumption 3 of theorem 1*

$$\arg\max_{\theta\in\mathcal{D}} \bar{\tilde{L}}(\theta) = \{\theta_\star\}, \tag{36}$$

i.e., $\theta_\star$ *is the unique maximizer for* $\bar{\tilde{L}}(\theta)$.

Proof. From (19), we have

$$-2\bar{\tilde{L}}(\theta) = \mathrm{Tr}\left\{C(\theta)^{-1} \lim_{N\to\infty} \mathcal{E}\left\{\delta_N(\theta)\,\delta_N(\theta)^\top\right\}\right\}.$$

Then, from (30)

$$\begin{aligned}-2\bar{\tilde{L}}(\theta) &= \mathrm{Tr}\left\{C(\theta)^{-1}\left(\mu(\theta_\star)-\mu(\theta)\right)\left(\mu(\theta_\star)-\mu(\theta)\right)^\top\right\}\\ &= \left(\mu(\theta_\star)-\mu(\theta)\right)^\top C(\theta)^{-1}\left(\mu(\theta_\star)-\mu(\theta)\right).\end{aligned}$$

Also, from (35), there exists $\epsilon > 0$ such that $C(\theta)^{-1} > \epsilon I$, for all $\theta \in \mathcal{D}$. Hence, $\bar{\tilde{L}}(\theta)$ is maximized if and only if $\mu(\theta_\star) = \mu(\theta)$. The result then follows from Assumption 3 of Theorem 1. □

Proof of Theorem 1. From Lemmas 4 and 5, two applications of [24, Lemma 8] give that $\tilde{L}_N(\theta)$ is CSUC on $\mathcal{D}$. This, together with Assumption 1 and Lemma 6, assert that the conditions of [21, Property 24.2] are satisfied, and the result follows. □

4.3. Proof of Theorem 2

Lemma 7. *Let* $\overline{\nabla^2\tilde{L}}(\theta) = \lim_{N\to\infty}\mathcal{E}\left\{\nabla^2\tilde{L}_N(\theta)\right\}$*. Then*

$$\overline{\nabla^2\tilde{L}}(\theta_\star) = -\mathbf{J}_\mu(\theta_\star)^\top C^{-1}(\theta_\star)\,\mathbf{J}_\mu(\theta_\star).$$

Proof. From (21),

$$\begin{aligned}\left[\overline{\nabla^2\tilde{L}}(\theta)\right]_{d,e} &= \lim_{N\to\infty}\mathcal{E}\left\{\left[\nabla^2\tilde{L}_N(\theta_\star)\right]_{d,e}\right\}\\ &= -\frac{\partial\mu^\top(\theta_\star)}{\partial\theta_d}C^{-1}(\theta)\frac{\partial\mu(\theta_\star)}{\partial\theta_e} - \lim_{N\to\infty}\frac{1}{2N}\mathrm{Tr}\left\{C^{-1}(\theta_\star)\frac{\partial C(\theta_\star)}{\partial\theta_d}C^{-1}(\theta_\star)\frac{\partial C(\theta_\star)}{\partial\theta_e}\right\}\\ &= -\frac{\partial\mu^\top(\theta_\star)}{\partial\theta_d}C^{-1}(\theta)\frac{\partial\mu(\theta_\star)}{\partial\theta_e},\end{aligned}$$

and the result then follows. □

Lemma 8. *Under Assumptions 1 and 2 of Theorem 2,*

$$\sqrt{N}\nabla\tilde{L}_N(\theta_\star) \overset{in\ dist.}{\to} \mathcal{N}\left(0,\tilde{\mathcal{I}}\right).$$

Proof. We proceed in steps: □

Step 1: Using (27), we obtain

$$\sqrt{N}\Delta_N(\theta_\star) = \frac{1}{N^{3/2}}\sum_{n,m=0}^{N-1}\Xi_{n,m},$$

with $\Xi_{n,m} = \xi_n(\theta_\star)\xi_m^\top(\theta_\star)$. Since $\mathcal{E}\{\xi_n(\theta_\star)\} = 0$, it follows that $\mathcal{E}\{\Xi_{n,m}\Xi_{p,q}\} \neq 0$ only if either $(n,m) = (p,q)$ or $(n,m) = (q,p)$ or $(n,p) = (m,q)$. Then

$$\begin{aligned}\mathcal{E}\left\{N\left[\Delta_N(\theta_\star)\right]_{k,l}\right\} =& \frac{1}{N^3}\sum_{(n,m),(p,q)=(0,0)}^{(N-1,N-1)}\mathcal{E}\left\{\left[\Xi_{n,m}\right]_{k,l}\left[\Xi_{p,q}\right]_{k.l}\right\}\\ =& \frac{2}{N^3}\sum_{(n,m)=(0,0)}^{(N-1,N-1)}\mathcal{E}\left\{\left[\xi_n(\theta_\star)\right]_k\left[\xi_n(\theta_\star)\right]_l\left[\xi_m(\theta_\star)\right]_k\left[\xi_m(\theta_\star)\right]_l\right\}\\ \leq& \frac{2}{N}\max_{k,n}\mathcal{E}\left\{\left[\xi_n(\theta_\star)\right]_k^4\right\} < \infty.\end{aligned}$$

Hence, from [23, Th. 4.1.4], $\sqrt{N}\Delta_N(\theta_\star) \to 0$ in probability, and from [23, Th. 4.4.5],

$$\sqrt{N}\Delta_N(\theta_\star) \overset{\text{in dist.}}{\rightarrow} 0. \tag{37}$$

Step 2: Recall (27). Since $\xi_n(\theta_\star)$ and $\xi_m(\theta_\star)$ are independent, whenever $n \neq m$, and $\mathcal{E}\{\xi_n(\theta_\star)\} = 0$, from the central limit theorem [23, Th. 6.4.4],

$$\sqrt{N}\delta_N(\theta_\star) \overset{\text{in dist.}}{\rightarrow} \mathcal{N}(0,U), \tag{38}$$

for some matrix $U \geq 0$.

Step 3: From (20) (37), (38) and [23, Th. 4.4.6], we have that

$$\sqrt{N}\nabla\tilde{L}_N(\theta_\star) \overset{\text{in dist.}}{\rightarrow} \mathcal{N}(\mu,\Sigma),$$

for some vector μ and positive matrix Σ. Now, $\mathcal{E}\left\{\sqrt{N}\nabla\tilde{L}_N(\theta_\star)\right\} = 0$ and

$$\mathcal{C}\left\{\sqrt{N}\nabla\tilde{L}_N(\theta_\star)\right\} = N\mathcal{E}\left\{\nabla\tilde{L}_N(\theta_\star)\nabla^\top\tilde{L}_N(\theta_\star)\right\} \triangleq \tilde{\mathcal{I}}_N.$$

Hence,

$$\tilde{\mathcal{I}}_N^{-1/2}\sqrt{N}\nabla\tilde{L}_N(\theta_\star) \overset{\text{in dist.}}{\rightarrow} \mathcal{N}(0,I),$$

and the result follows from [20, Th. 13.25].

Lemma 9. *The following holds true*

$$\tilde{\mathcal{I}} = \overline{\nabla^2\tilde{L}}(\theta_\star).$$

Proof. From (21), the entries of $\nabla^2\tilde{L}_N(\theta)$ are formed by a deterministic term, plus a linear combination of the elements $\delta_{N,i}$ and $\delta_{N,i}\delta_{N,j}$, for all $i,j = 1,\cdots,I$. Since the first and second moments of s_N are equivalent under the distributions p_θ and $\tilde{p}_\theta$, it follows that

$$\overline{\nabla^2\tilde{L}}(\theta_\star) = \lim_{N\to\infty}\tilde{\mathcal{E}}\left\{\nabla^2\tilde{L}_N(\theta_\star)\right\}. \tag{39}$$

We also have

$$
\begin{aligned}
&N\left[\nabla\tilde{L}_N(\theta)\right]_d\left[\nabla\tilde{L}_N(\theta)\right]_e \\
&= N\frac{\partial\mu(\theta)^\top}{\partial\theta_d}C^{-1}(\theta)\,\delta_N(\theta)\,\frac{\partial\mu(\theta)^\top}{\partial\theta_e}C^{-1}(\theta)\,\delta_N(\theta) \\
&+\frac{N}{2}\frac{\partial\mu(\theta)^\top}{\partial\theta_d}C^{-1}(\theta)\,\delta_N(\theta)\operatorname{Tr}\left\{C^{-1}(\theta)\,\frac{\partial C(\theta)}{\partial\theta_e}\left(C(\theta)^{-1}\Delta_N(\theta)-\frac{1}{N}I\right)\right\} \\
&+\frac{N}{2}\frac{\partial\mu(\theta)^\top}{\partial\theta_e}C^{-1}(\theta)\,\delta_N(\theta)\operatorname{Tr}\left\{C^{-1}(\theta)\,\frac{\partial C(\theta)}{\partial\theta_d}\left(C(\theta)^{-1}\Delta_N(\theta)-\frac{1}{N}I\right)\right\} \\
&+\frac{N}{4}\operatorname{Tr}\left\{C^{-1}(\theta)\,\frac{\partial C(\theta)}{\partial\theta_d}\left(C(\theta)^{-1}\Delta_N(\theta)-\frac{1}{N}I\right)\right\} \\
&\times\operatorname{Tr}\left\{C^{-1}(\theta)\,\frac{\partial C(\theta)}{\partial\theta_e}\left(C(\theta)^{-1}\Delta_N(\theta)-\frac{1}{N}I\right)\right\}.
\end{aligned}
$$

The above expression is the sum of a deterministic term, plus a linear combination of the elements $\delta_{N,i}$, $\delta_{N,i}\delta_{N,j}$, $\delta_{N,i}\delta_{N,j}\delta_{N,k}$ and $\delta_{N,i}\delta_{N,j}\delta_{N,k}\delta_{N,l}$, for all $i,j,k,l=1,\cdots,I$. Now,

$$
\begin{aligned}
\lim_{N\to\infty}\mathcal{E}\left\{\delta_{N,i}\delta_{N,j}\delta_{N,k}\right\} &= \lim_{N\to\infty}\frac{1}{N^3}\sum_{n,m,o=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{m,j}\delta_{o,k}\right\} \\
&= \lim_{N\to\infty}\frac{1}{N^3}\sum_{n=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{n,j}\delta_{n,k}\right\}=0,
\end{aligned}
$$

and

$$
\begin{aligned}
&\lim_{N\to\infty}\mathcal{E}\left\{\delta_{N,i}\delta_{N,j}\delta_{N,k}\delta_{N,l}\right\} = \lim_{N\to\infty}\frac{1}{N^4}\sum_{n,m,o,p=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{m,j}\delta_{o,k}\delta_{p,l}\right\} \\
&\quad=\lim_{N\to\infty}\frac{1}{N^4}\sum_{n=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{n,j}\delta_{n,k}\delta_{n,l}\right\}+\lim_{N\to\infty}\frac{1}{N^4}\sum_{n,o=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{n,j}\delta_{o,k}\delta_{o,l}\right\} \\
&\quad+\lim_{N\to\infty}\frac{1}{N^4}\sum_{n,m=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{m,j}\delta_{n,k}\delta_{m,l}\right\}+\lim_{N\to\infty}\frac{1}{N^4}\sum_{n,m=1}^{N}\mathcal{E}\left\{\delta_{n,i}\delta_{m,j}\delta_{m,k}\delta_{n,l}\right\}=0.
\end{aligned}
$$

The same conclusion can be drawn if we replace $\mathcal{E}$ by $\tilde{\mathcal{E}}$. Then, in view of the aforementioned equivalence of the first and second moments of s_N under p_θ and $\tilde{p}_\theta$, we conclude that

$$
\tilde{\mathcal{I}}=\lim_{N\to\infty}N\tilde{\mathcal{E}}\left\{\nabla\tilde{L}_N(\theta_\star)\,\nabla^\top\tilde{L}_N(\theta_\star)\right\}. \tag{40}
$$

Finally, from [16, Property 3.8], we have

$$
\lim_{N\to\infty}N\tilde{\mathcal{E}}\left\{\nabla\tilde{L}_N(\theta_\star)\,\nabla^\top\tilde{L}_N(\theta_\star)\right\}=\lim_{N\to\infty}\tilde{\mathcal{E}}\left\{\nabla^2\tilde{L}_N(\theta)\right\}. \tag{41}
$$

The result then follows from (39), (40) and (41). □

Proof of Theorem 2. In view of (21), Assumptions 2 implies that $\nabla^2\tilde{L}_N(\theta)$ is continuous. Also, in view of Lemma 7 and Assumption 3, $\overline{\nabla^2\tilde{L}}(\theta_\star)<0$. This, together with Assumption 1, and Lemmas 5 and 8, give the conditions for [21, Property 24.16], from where the result follows. □

4.4. Proof of Theorem 3

Let

$$R_N = \frac{1}{N}\sum_{n=1}^{N} x_n x_n^\top.$$

We know that, for a given $\theta \in \mathcal{D}$, the random matrix NR_N has a Wishart distribution $\mathcal{W}_I(NR_N; \Sigma(\theta), N)$ [25, §7.2]. Hence

$$\begin{aligned} p_\theta(R_N) &= N\mathcal{W}_I(NR_N; \Sigma(\theta), N) \\ &= \frac{N\left|NR_N\right|^{(N-I-1)}}{2^{\frac{NI}{2}}\left|\Sigma(\theta)\right|^{\frac{N}{2}}\Gamma_I\left(\frac{N}{2}\right)} \exp\left(-\frac{N}{2}\mathrm{Tr}\left\{\Sigma(\theta)^{-1} R_N\right\}\right). \end{aligned} \tag{42}$$

Let $\mathcal{J}$ be the AFIM associated to the measurement R_N, i.e.,

$$\mathcal{J} = \lim_{N\to\infty}\frac{1}{N}\mathcal{E}\left\{\nabla \log p_{\theta_\star}(R_N)\,\nabla^T \log p_{\theta_\star}(R_N)\right\}.$$

The following lemma gives an expression for $\mathcal{J}$.

Lemma 10. *The (d,e)-th entry $\mathcal{J}_{d,e}$ of matrix $\mathcal{J}$ is given by*

$$\mathcal{J}_{d,e} = \frac{1}{2}\mathrm{Tr}\left\{\Sigma(\theta_\star)^{-1}\frac{\partial\Sigma(\theta_\star)}{\partial\theta_d}\Sigma(\theta_\star)^{-1}\frac{\partial\Sigma(\theta_\star)}{\partial\theta_e}\right\}.$$

Proof. From (42), we have

$$\log p_\theta(R_N) = \log\frac{N\left|NR_N\right|^{(N-I-1)}}{2^{\frac{NI}{2}}\Gamma_I\left(\frac{N}{2}\right)} - \frac{N}{2}\left[\log\left|\Sigma(\theta)\right| + \mathrm{Tr}\left\{\Sigma(\theta)^{-1}R_N\right\}\right],$$

Then

$$\left[\nabla \log p_\theta(R_N)\right]_d = -\frac{N}{2}\mathrm{Tr}\left\{\Sigma(\theta)^{-1}\frac{\partial\Sigma(\theta)}{\partial\theta_i}\left(I - \Sigma(\theta)^{-1}R_N\right)\right\},$$

and

$$\begin{aligned} &\left[\nabla^2 \log p_\theta(R_N)\right]_{d,e} \\ =&-\frac{N}{2}\mathrm{Tr}\left\{\Sigma(\theta)^{-1}\left[\frac{\partial^2\Sigma(\theta)}{\partial\theta_d\theta_e} - \frac{\partial\Sigma(\theta)}{\partial\theta_e}\Sigma(\theta)^{-1}\frac{\partial\Sigma(\theta)}{\partial\theta_d}\right]\left(I - \Sigma(\theta)^{-1}R_N\right)\right. \\ &\left.+\Sigma(\theta)^{-1}\frac{\partial\Sigma(\theta)}{\partial\theta_d}\Sigma(\theta)^{-1}\frac{\partial\Sigma(\theta)}{\partial\theta_e}\Sigma(\theta)^{-1}R_N\right\} \end{aligned}$$

It then follows from [16, Property 3.8] that

$$\begin{aligned} \mathcal{J}_{d,e} &= -\lim_{N\to\infty}\frac{1}{N}\mathcal{E}\left\{\left[\nabla^2 \log p_{\theta_\star}(R_N)\right]_{d,e}\right\} \\ &= \frac{1}{2}\mathrm{Tr}\left\{\Sigma(\theta_\star)^{-1}\frac{\partial\Sigma(\theta_\star)}{\partial\theta_d}\Sigma(\theta_\star)^{-1}\frac{\partial\Sigma(\theta_\star)}{\partial\theta_e}\right\}. \end{aligned}$$

□

As defined in (9), $\tilde{p}_\theta$ is a distribution of the diagonal entries s_N of R_N. Our next goal is to extend $\tilde{p}_\theta$ to the whole random matrix R_N.

Lemma 11. *The following holds true*

$$\mathcal{E}\left\{\overrightarrow{R_N}\right\} = \overrightarrow{\Sigma\left(\theta_\star\right)}, \tag{43}$$

$$\mathcal{C}\left\{\overrightarrow{R_N}\right\} = \frac{1}{N}\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\Gamma\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right), \tag{44}$$

where $\Gamma = M + \mathbf{I}$ and M is the symmetric permutation matrix described in [26, §1.2].

Proof. Equation (43) follows straightforwardly. For (44), let $X = [x(1), \cdots, x(N)]$ and

$$E = \Sigma^{-1/2}\left(\theta_\star\right) X.$$

Then

$$R_N = \frac{1}{N}\Sigma^{-1/2}\left(\theta_\star\right) E E^\top \Sigma^{-1/2}\left(\theta_\star\right),$$

and therefore

$$\overrightarrow{R_N} = \frac{1}{N}\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\overrightarrow{EE^\top}.$$

We then have

$$\begin{aligned}\mathcal{C}\left\{\overrightarrow{R_N}\right\} &= \mathcal{C}\left\{\frac{1}{N}\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\overrightarrow{EE^\top}\right\}\\ &= \frac{1}{N^2}\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\mathcal{C}\left\{\overrightarrow{EE^\top}\right\}\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right),\end{aligned}$$

and the result follows since, in view of [26, §1.2], $\mathcal{C}\left\{\overrightarrow{EE^\top}\right\} = \Gamma$. □

Since $R_N \in \mathcal{P}^I(\mathbb{R})$, it has $I(I+1)/2$ degrees of freedom. Hence, $M\left(\theta_\star\right)$ has rank $I(I+1)/2$. It can be readily verified from [26, §1.2], that Γ has two eigenvalues, namely, 0 with multiplicity $I(I-1)/2$, and 2 with multiplicity $I(I+1)/2$. It then follows that, for all $R_N \in \mathcal{P}^I(\mathbb{R})$, matrix Γ can be replaced by $2\mathbf{I}$, i.e.,

$$\begin{aligned}&\overrightarrow{R_N}^\top\left[\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\Gamma\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\right]^{-1}\overrightarrow{R_N}\\ &\qquad = \frac{1}{2}\overrightarrow{R_N}^\top M^{-1}\left(\theta_\star\right)\overrightarrow{R_N},\end{aligned}$$

with

$$M\left(\theta_\star\right) = 2\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right)\left(\Sigma^{1/2}\left(\theta_\star\right)\otimes\Sigma^{1/2}\left(\theta_\star\right)\right).$$

We then define, for $R_N \in \mathcal{P}^I(\mathbb{R})$,

$$\tilde{p}_\theta\left(R_N\right) = \mathcal{N}\left(\overrightarrow{R_N}; \overrightarrow{\Sigma\left(\theta\right)}, \frac{1}{N}M\left(\theta\right)\right).$$

Let

$$\tilde{\mathcal{J}} = \lim_{N\to\infty}\frac{1}{N}\mathcal{E}\left\{\nabla\log\tilde{p}_{\theta_\star}\left(R_N\right)\nabla^T\log\tilde{p}_{\theta_\star}\left(R_N\right)\right\}.$$

Our next goal is to provide an expression for $\tilde{\mathcal{J}}$. To this end, we use the following lemma.

Lemma 12 ([27, Eq. (23)])**.** *Let $x \sim \mathcal{N}\left(\mu\left(\theta_\star\right), \Sigma\left(\theta_\star\right)\right)$ and $\mathcal{I}$ denote its associated AFIM. Then*

$$[\mathcal{I}]_{d,e} = \frac{\partial\mu\left(\theta_\star\right)^\top}{\partial\theta_d}\Sigma^{-1}\left(\theta_\star\right)\frac{\partial\mu\left(\theta_\star\right)}{\partial\theta_e} + \frac{1}{2}\mathrm{Tr}\left\{\Sigma^{-1}\left(\theta_\star\right)\frac{\partial\Sigma\left(\theta_\star\right)}{\partial\theta_d}\Sigma^{-1}\left(\theta_\star\right)\frac{\partial\Sigma\left(\theta_\star\right)}{\partial\theta_e}\right\}.$$

Using the above lemma, we obtain:

Lemma 13. *The following holds true* $\tilde{\mathcal{J}} = \mathcal{J}$.

Proof. We can readily apply Lemma 9 to obtain

$$\tilde{\mathcal{J}} = -\lim_{N\to\infty} \frac{1}{N} \tilde{\mathcal{E}} \left\{ \nabla^2 \log \tilde{p}_{\theta_\star} \left(R_N \right) \right\}.$$

Now, from Lemma 12 we have

$$\begin{aligned} -\left[\tilde{\mathcal{E}} \left\{ \nabla^2 \log \tilde{p}_{\theta_\star} \left(R_N \right) \right\} \right]_{d,e} =& N \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}^\top}{\partial \theta_d} M^{-1}\left(\theta_\star\right) \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}}{\partial \theta_e} \\ &+ \frac{1}{2} \mathrm{Tr} \left\{ M^{-1}\left(\theta_\star\right) \frac{\partial M\left(\theta_\star\right)}{\partial \theta_d} M^{-1}\left(\theta_\star\right) \frac{\partial M\left(\theta_\star\right)}{\partial \theta_e} \right\}, \end{aligned}$$

hence

$$\left[\tilde{\mathcal{J}} \right]_{d,e} = \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}^\top}{\partial \theta_d} M^{-1}\left(\theta_\star\right) \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}}{\partial \theta_e}.$$

Now

$$\begin{aligned} M^{-1}\left(\theta_\star\right) &= \frac{1}{2} \left(\Sigma^{-1}\left(\theta_\star\right) \otimes \Sigma^{-1}\left(\theta_\star\right) \right) \\ &= \frac{1}{2} \left(\Sigma^{-1/2}\left(\theta_\star\right) \otimes \Sigma^{-1/2}\left(\theta_\star\right) \right) \left(\Sigma^{-1/2}\left(\theta_\star\right) \otimes \Sigma^{-1/2}\left(\theta_\star\right) \right). \end{aligned}$$

Then

$$\begin{aligned} \left[\tilde{\mathcal{J}} \right]_{d,e} &= \frac{1}{2} \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}^\top}{\partial \theta_d} \left(\Sigma^{-1/2}\left(\theta_\star\right) \otimes \Sigma^{-1/2}\left(\theta_\star\right) \right) \left(\Sigma^{-1/2}\left(\theta_\star\right) \otimes \Sigma^{-1/2}\left(\theta_\star\right) \right) \frac{\partial \overrightarrow{\Sigma\left(\theta_\star\right)}}{\partial \theta_e} \\ &= \frac{1}{2} \overrightarrow{\Sigma^{-1/2}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_d} \Sigma^{-1/2}\left(\theta_\star\right)}^\top \overrightarrow{\Sigma^{-1/2}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_e} \Sigma^{-1/2}\left(\theta_\star\right)} \\ &= \frac{1}{2} \mathrm{Tr} \left\{ \left[\Sigma^{-1/2}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_d} \Sigma^{-1/2}\left(\theta_\star\right) \right]^\top \left[\Sigma^{-1/2}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_e} \Sigma^{-1/2}\left(\theta_\star\right) \right] \right\} \\ &= \frac{1}{2} \mathrm{Tr} \left\{ \Sigma^{-1}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_d} \Sigma^{-1}\left(\theta_\star\right) \frac{\partial \Sigma\left(\theta_\star\right)}{\partial \theta_e} \right\}, \end{aligned}$$

and the result follows from Lemma 10. □

We can now show the main result of the section.

Proof of Theorem 3. Let q_N denote the off-diagonal terms of R_N. Since $p_\theta\left(R_N\right) = p_\theta\left(q_N|s_N\right) p_\theta\left(s_N\right)$, it follows from [16, Property 3.8] that

$$\begin{aligned} \mathcal{J} = -\lim_{N\to\infty} \frac{1}{N} \mathcal{E} \left\{ \nabla^2 \log p_{\theta_\star}\left(R_N\right) \right\} \\ = \lim_{N\to\infty} \mathcal{E}\left\{ \mathcal{K}\left(s_N\right) \right\} - \lim_{N\to\infty} \frac{1}{N} \mathcal{E} \left\{ \nabla^2 \log p_{\theta_\star}\left(s_N\right) \right\} = \lim_{N\to\infty} \mathcal{E}\left\{ \mathcal{K}\left(s_N\right) \right\} + \mathcal{I}, \end{aligned}$$

with

$$\begin{aligned} \mathcal{K}\left(s_N\right) &= -\frac{1}{N} \int \nabla^2 \log p_{\theta_\star}\left(q_N|s_N\right) p_{\theta_\star}\left(q_N|s_N\right) dq_N \\ &= -\frac{1}{N} \mathcal{E} \left\{ \nabla^2 \log p_{\theta_\star}\left(q_N|s_N\right) |s_N \right\} = N \mathcal{E} \left\{ \nabla \log p_{\theta_\star}\left(q_N|s_N\right) \nabla^\top \log p_{\theta_\star}\left(q_N|s_N\right) |s_N \right\}. \end{aligned}$$

Following a the same steps, but using Lemma 9 in place of [16, Property 3.8], we obtain

$$\tilde{\mathcal{J}} = \lim_{N\to\infty} \mathcal{E}\left\{\tilde{\mathcal{K}}\left(s_N\right)\right\} + \tilde{\mathcal{I}},$$
$$\tilde{\mathcal{K}}\left(s_N\right) = N\mathcal{E}\left\{\nabla \log \tilde{p}_{\theta_\star}\left(q_N|s_N\right)\nabla^\top \log \tilde{p}_{\theta_\star}\left(q_N|s_N\right)|s_N\right\}.$$

□

From Lemma 13,

$$\mathcal{I} = \tilde{\mathcal{I}} + \lim_{N\to\infty} \mathcal{E}\left\{\tilde{\mathcal{K}}\left(s_N\right)\right\} - \lim_{N\to\infty} \mathcal{E}\left\{\mathcal{K}\left(s_N\right)\right\}.$$

Now, $\mathcal{K}\left(s_N\right)$ denotes the AFIM associated to the estimation of θ, based on the knowledge of q_N, for a given know s_N. Also, $\tilde{\mathcal{K}}\left(s_N\right)^{-1}$ denotes the asymptotic covariance of the same estimate when obtained by maximizing $\tilde{p}_\theta(R_N)$. Hence, from the CRLB

$$\mathcal{K}\left(s_N\right) \geq \tilde{\mathcal{K}}\left(s_N\right).$$

Then

$$\tilde{\mathcal{I}} \geq \mathcal{I},$$

and the result follows from the CRLB [16, Th. 6.1].

4.5. Proof of Proposition 2

It follows from Lemma 9 and Theorem 3 that

$$\mathcal{I} = \overline{\nabla^2 \tilde{L}}\left(\theta_\star\right).$$

The result then follows from Lemma 7.

5. Case study: source localization problem

5.1. Problem formulation

We assume that there are I sensors deployed in a two dimensional surveillance area, with known positions at (a_i, b_i), $i = 1, \ldots, I$. Suppose that a signal source, located at some unknown position $(a_\star, b_\star)$, transmits, at sample time n, a white signal $u_n \sim \mathcal{N}\left(0, z_\star\right)$ with an unknown intensity $z_\star$. I.e., the vector $\theta_\star = \left[a_\star, b_\star, z_\star\right]^T$ represents the true position and transmitting power of the source. According to the log-distance path loss model [28, S 4.9.1], [15], the signal $x_{n,i}$, received at Sensor i, and at sample time n is given by

$$x_{n,i} = \sqrt{\gamma}\frac{u_n}{d_{\star,i}^{\upsilon}} + w_{n,i},$$

where $\gamma > 0$ is the receiver gain,

$$d_{\star,i} = \sqrt{\left(a_\star - a_i\right)^2 + \left(b_\star - b_i\right)^2},$$

represents the signal attenuation due to the path loss, and $w_{n,i} \sim \mathcal{N}\left(0, \sigma^2\right)$ is white, and independent of u_n and $w_{n,j}$, whenever $i \neq j$. The power term υ denotes the path loss factor, describing the intensity decay as the wave propagates. Empirical measurements of υ for different environments are given in [28, Table 4.2]. To simplify the presentation, we assume that σ, γ and υ are the same for all sensors.

At time N, Sensor i computes an estimate $s_{N,i}$ of the power of $x_{n,i}$, by averaging N samples, i.e.,

$$s_{N,i} = \frac{1}{N}\sum_{n=0}^{N-1} x_{n,i}^2. \tag{45}$$

The target localization problem consists in using the measurements

$$s_N = \left[s_{N,1}, \ldots, s_{N,I}\right]^T, \tag{46}$$

from all the sensors, to estimate the unknown target coordinates $(a_\star, b_\star)$. However, since this requires knowledge of the unknown transmission power $z_\star$, the problem turns into that of jointly estimating $\theta_\star = \left[a_\star, b_\star, z_\star\right]^T$.

Clearly, s_N is of the form (3). Hence, the ML estimate (4) cannot be computed, and we have to replace it by the quasi-ML estimate (7). Our goal is to show that, using the results in Theorems 1, 2 and 3, we can guarantee that the quasi-ML estimate is strongly consistent, asymptotically normal and asymptotically efficient.

5.2. Conditions for quasi-ML estimation

Let

$$\Sigma(\theta) = \mathcal{C}_\theta\left\{x_i(n)\right\} = h(\theta) + \sigma^2 \mathbf{1}, \tag{47}$$

with $h(\theta) = \left[h_1(\theta), \ldots, h_I\right]^T$, $h_i(\theta) = \gamma z d_i^{-2\upsilon}$, $i = 1, \cdots, I$, and $\mathbf{1}$ being a column vector of ones.We introduce the following assumption:

Assumption 1. *The set $\mathcal{D} \in \mathbb{R}^3$ is compact, $\theta_\star \in \operatorname{int}(\mathcal{D})$, and does not include the points $\left[a_i, b_i, z'\right]^\top$, for all $i = 1, \ldots, I$ and any $z' \in \mathbb{R}$, and the points $\left[x', y', z'\right]^\top$, for all $z' < 0$ and any $x', y' \in \mathbb{R}$.*

Assumption 1 asserts that Assumptions 1 and 2 of Theorem 1, as well as Assumptions 1 and 2 of Theorem 2 are satisfied. Hence, we need to show that Assumption 3 of Theorem 1 also holds. To this end, we do the following further assumption:

Assumption 2. *The sensors are not arranged on a circle or on a straight line, $I \geq 4$ and $z_\star \neq 0$.*

Lemma 14. *Suppose Assumptions 1 and 2 hold. Then, for all $\theta \in \mathcal{D}$,*

$$\Sigma(\theta) = \Sigma(\theta_\star) \Rightarrow \theta = \theta_\star.$$

Proof. For $\theta = [a, b, z]^\top$, let

$$d_i(\theta) = \sqrt{(a - a_i)^2 + (b - b_i)^2},$$

i.e., $d_i(\theta)$ is the distance from the point (a, b) to Sensor i. It follows from (47) that $\Sigma(\theta) = \Sigma(\theta_\star)$ implies that $h(\theta) = h(\theta_\star)$. Then, for all $i = 1, \ldots, I$,

$$\frac{z}{d_i(\theta)^{2\upsilon}} = \frac{z_\star}{d_{\star i}^{2\upsilon}}.$$

Hence, since $z_\star$, for all $i = 1, \ldots, I$,

$$c \triangleq \left(\frac{z}{z_\star}\right)^{1/\upsilon} = \frac{d_i(\theta)^2}{d_{\star,i}^2}. \tag{48}$$

The above in turn implies that,

$$(a_i - \alpha)^2 + (b_i - \beta)^2 = r^2,$$

with

$$\alpha = \frac{a - ca_\star}{1-c}, \qquad \beta = \frac{b - cb_\star}{1-c} \qquad \text{and} \qquad r^2 = \frac{c}{(1-c)^2}\left[(a - a_\star)^2 + (b - b_\star)^2\right].$$

Now, if $z \neq z_\star$, then $c \neq 1$, and therefore the points (a_i, b_i), $i = 1, \ldots, N$, must lie on a circle. But this contradicts Assumption 2. Hence we conclude that $z = z_\star$. Then, (48) implies that, for all $i = 1, \ldots, I$,

$$d_i(\theta)^2 = d_{\star,i}^2. \tag{49}$$

However, (49) implies that, either the points (a_i, b_i), $i = 1, \ldots, N$, lie on a straight line, or that $(a, b) = (a_\star, b_\star)$. But from Assumption 2, the points (a_i, b_i), $i = 1, \ldots, N$, cannot lie on a straight line. Hence we conclude that $[a, b, z] = (a_\star, b_\star, z_\star)$ completing the proof. $\square$

In view of Lemma 14, Assumption 3 of Theorem 1 holds. Therefore, all the assumptions from Theorems 1, 2 and 3 are satisfied. We then immediately obtain the following corollary:

Corollary 1. *Under Assumptions 1 and 2, the quasi-ML estimate $\hat{\theta}_N$ is strongly consistent, asymptotically normal, and asymptotically efficient.*

5.3. Simulation

For a given N, we use Ψ_N to denote the estimation error covariance. In view of the CRLB, $\Psi_N \geq \mathcal{I}_N^{-1}$, with $\mathcal{I}_N$ denoting the Fisher information matrix, given by

$$\mathcal{I}_N = N^2 \mathcal{E}\left\{\nabla L_N(\theta_\star) \nabla^\top L_N(\theta_\star)\right\}.$$

According to Theorems 2 and 3, we have

$$\lim_{N\to\infty} N\Psi_N - N\mathcal{I}_N^{-1} = 0,$$

i.e., as N increases, Ψ_N converges to its lower bound $\tilde{\mathcal{I}}_N^{-1}$. We present a numerical experiment confirming this claim. To this end, we regularly deploy 25 sensors, over a square region of 100×100 meters, and randomly place the source within the same region. The arrangement of source and sensors is depicted in Figure 2.

For each value of the number N of samples, we estimate Ψ_N using $M = 1000$ Monte Carlo runs. To this end, we solve the quasi-ML problem (7) using the implementation of the BFGS quasi-Newton method available in the GSL C library. To obtain an initialization for this algorithm, we use the one step least-squares method in [4]. We also use $\sigma^2 = 1$, $\upsilon = 1$ and $\gamma = 3000$. The resulting dependence of the traces of Ψ_N and $\mathcal{I}_N^{-1}$ with N are shown in Figure 3, from where we can see their asymptotic equality. Notice that the figure also shows that Ψ_N converges to zero.

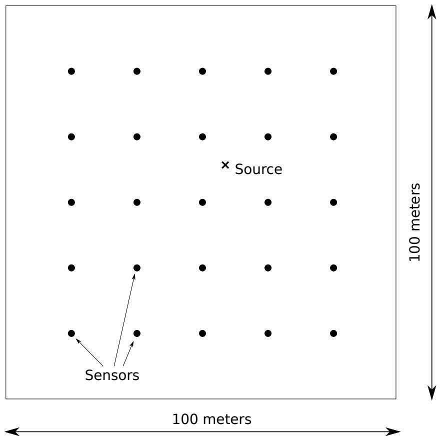

Figure 2: Sensor and source deployment

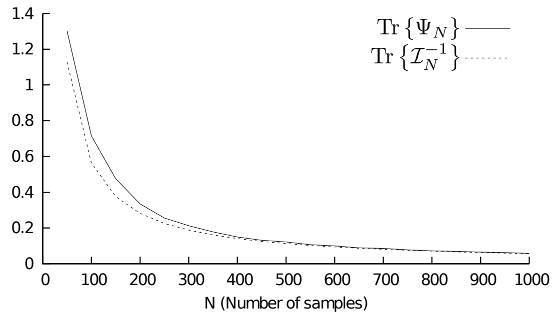

Figure 3: Convergence of the trace of the estimation error covariance $\mathrm{Tr}\left\{\Psi_N\right\}$ to its lower bound $\mathrm{Tr}\left\{\mathcal{I}_N^{-1}\right\}$.

6. Conclusion

We studied statistical properties of the parameter estimation problem in which measurements have multivariate chi-squared distribution, but a normal approximation to the measurement statistics is used to find the estimate. The main technical challenge is given by the fact that the PDF of measurements lacks a closed-form expression. We where able to show strong consistency, asymptotic normality and asymptotic efficiency of the resulting estimate. We considered a source localization problem as a case study, and used our theoretical results to provide conditions to guarantee that the aforementioned three properties hold in this application.

[1] G. Casella, R. L. Berger, Statistical inference, Duxbury Pacific Grove, CA, 2002.

[2] S. M. Kay, Fundamentals of statistical signal processing, Prentice Hall PTR, 1993.

[3] W. Liu, P. Shi, J.-S. Pan, State estimation for discrete-time markov jump linear systems with time-correlated and mode-dependent measurement noise, Automatica 85 (2017) 9–21.

[4] C. Meesookho, U. Mitra, S. Narayanan, On energy-based acoustic source localization for sensor networks, IEEE T Signal Proces 56 (1) (2008) 365–377.

[5] X. Sheng, Y. Hu, Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks, IEEE T Signal Proces 53 (1) (2005) 44–53.

[6] T. Zhao, A. Nehorai, Information-driven distributed maximum likelihood estimation based on gauss-newton method in wireless sensor networks, IEEE T Signal Proces 55 (9) (2007) 4669–4682.

[7] Q. Shi, C. He, A new incremental optimization algorithm for ml-based source localization in sensor networks, Signal Processing Letters, IEEE 15 (2008) 45–48.

[8] M. Rabbat, R. Nowak, Decentralized source localization and tracking, in: IEEE, ICASSP, Vol. 3, 2004, pp. 921–924.

[9] D. Li, Y. H. Hu, Energy-based collaborative source localization using acoustic microsensor array, EURASIP J Appl Si Pr 2003 (4) (2003) 321–337.

[10] N. Patwari, A. Hero, M. Perkins, N. Correal, R. O'Dea, Relative location estimation in wireless sensor networks, IEEE T Signal Proces 51 (8) (2003) 2137–2148.

[11] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, , N. Correal, Locating the nodes: Cooperative localization in wireless sensor networks, IEEE Signal Processing Magazine 22 (4) (2005) 54–69.

[12] L. Cheng, C. Wu, Y. Zhang, H. Wu, M. Li, C. Maple, A survey of localization in wireless sensor network, Int J Distrib Sens N 2012.

[13] X. Wang, M. Fu, H. Zhang, Target tracking in wireless sensor networks based on the combination of KF and MLE using distance measurements, IEEE Transactions on Mobile Computing 11 (4) (2012) 567 –576.

[14] X. Wang, H. Zhang, M. Fu, Collaborative target tracking in wsns using the combination of maximum likelihood estimation and kalman filtering, J Contr Theor Appl 11 (1) (2013) 27–34.

[15] N. Ahmed, M. Rutten, T. Bessell, S. S. Kanhere, N. Gordon, S. Jha, Detection and tracking using particle-filter-based wireless sensor networks, IEEE Transactions on Mobile Computing 9 (9) (2010) 1332 – 1345.

[16] C. Gourieroux, A. Monfort, Statistics and Econometric Models: Volume 1, Cambridge University Press, 1995.

[17] W. K. Newey, D. McFadden, Large sample estimation and hypothesis testing, Handbook of econometrics 4 (1994) 2111–2245.

[18] T. Royen, Expansions for the multivariate chi-square distribution, Journal of Multivariate Analysis 38 (2) (1991) 213–232.

[19] T. Dickhaus, T. Royen, A survey on multivariate chi-square distributions and their applications in testing multiple hypotheses, Statistics 49 (2) (2015) 427–454.

[20] A. Klenke, Probability theory: a comprehensive course, Springer Verlag, 2008.

[21] C. Gourieroux, A. Monfort, Statistics and Econometric Models: Volume 2, Cambridge University Press, 1996.

[22] A. Gallant, H. White, A unified theory of estimation and inference for nonlinear dynamic models, Basil Blackwell New York, 1988.

[23] K. L. Chung, A Course in Probability Theory, 3rd Edition, Academic Press, 2000.

[24] D. Marelli, M. Fu, B. Ninness, Asymptotic optimality of the maximum-likelihood kalman filter for bayesian tracking with multiple nonlinear sensors, IEEE Transactions on Signal Processing 63 (17) (2015) 4502–4515.

[25] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 2003.

[26] S. W. Nydick, The wishart and inverse wishart distributions, J. Stat.

[27] L. Malagò, G. Pistone, Information geometry of the gaussian distribution in view of stochastic optimization, in: Proceedings of ACM FOGA, ACM, 2015, pp. 150–162.

[28] T. S. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall PTR New Jersey, 1996.