Optimal Joint Detection and Estimation in Linear Models

Jianshu Chen, Yue Zhao, Andrea Goldsmith, and H. Vincent Poor

Abstract— The problem of optimal joint detection and estimation in linear models with Gaussian noise is studied. A simple closed-form expression for the joint posterior distribution of the (multiple) hypotheses and the states is derived. The expression crystalizes the dependence of the optimal detector on the state estimates. The joint posterior distribution characterizes the beliefs ("soft information") about the hypotheses and the values of the states. Furthermore, it is a *sufficient statistic* for jointly detecting multiple hypotheses and estimating the states. The developed expressions give us a unified framework for joint detection and estimation under all performance criteria.

I. INTRODUCTION

Detection and estimation problems appear simultaneously and are naturally coupled in many engineering systems. Several prominent examples are as follows. To achieve situational awareness in power grids, it is essential to have timely detection of outages as well as estimation of system states [1], [2]. Radar systems detect the existence of targets and also estimate their positions and velocities [3]. Wireless communication systems often need to decode messages and estimate channel states at the same time [4]. In different engineering systems, the problem settings of joint detection and estimation can vary greatly, and many application-specific solutions have been developed in practice.

A classic approach that addresses the detection problem in the presence of unknown states/parameters is composite hypothesis testing [3]. Accordingly, a straightforward approach for joint hypothesis testing and state/parameter estimation is to perform composite hypothesis testing first, followed by state/parameter estimation based on the *hard decision* made from hypothesis testing. However, such an approach cannot provide optimality guarantees under general performance criteria that *depend jointly on detection and estimation results*. In the literature, several studies have addressed such joint performance criteria. The *structure* of the jointly optimal Bayes detector and estimator with discrete-time data was developed in [5] and [6], and was extended to the continuous-time data case in [7]. There, the detector structure was expressed in terms of some generalized forms of likelihood ratios. The structure of the optimal Bayesian estimator under any given constraints on false alarm probability and probability of missed detection have also been developed for the binary hypothesis case [8].

In this paper, we study the problem of optimal joint detection and estimation for a general class of observation models, namely, *linear models with Gaussian noise*. Linear models appear in a wide range of engineering applications including power systems [1], channel estimation [9], [10], adaptive array processing [11]–[13], and spectrum estimation [14]. In these applications, not only is state estimation of primary interest, but also the observation matrix can often change over time, and it is essential to detect which observation matrix among many possibilities is currently effective. We formulate these problems as joint multiple hypothesis testing and state estimation problems. Instead of focusing on a particular form of performance criterion and developing the corresponding optimal joint detector and estimator, *we develop a unified Bayesian approach that can be applied to any given criterion*. Specifically, employing a conjugate prior, we provide *closed form expressions for the joint posterior* of the hypotheses and the system states given all measurement samples. The developed expressions reveal the exact dependence of the optimal detectors on the state estimates. Because the joint posterior is a *sufficient statistic* for joint hypothesis testing and state estimation, the derived explicit forms of such *soft information* (as opposed to hard decisions) can be applied to all performance criteria with optimality guarantees.

The remainder of the paper is organized as follows. In Section II, we describe the system model and formulate the joint detection and estimation problem. In Section III, we provide a factorization of the likelihood function and derive a simple closed form expression for the joint posterior distribution. Finally, we conclude our paper and remark on future directions in Section IV.

Notation. We use boldface letters to denote random quantities and use regular letters to denote realizations or deterministic quantities.

J. Chen is with the Dept. of Electrical Engineering, University of California, Los Angeles, CA 90095 USA. (e-mail: jshchen@ee.ucla.edu)

Y. Zhao is with the Dept. of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA, and with the Dept. of Electrical Engineering, Stanford University, Stanford, CA, 94305 USA (e-mail: yuez@princeton.edu).

A. Goldsmith is with the Dept. of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@stanford.edu)

H. V. Poor is with the Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu)

II. PROBLEM FORMULATION

We consider the following observation model which entails a joint detection and estimation problem. Given each of the $K + 1$ hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_K$, the $M \times 1$ sensor measurement vector $\boldsymbol{x}_t$ at time t is obtained according to the following linear model:

$$\mathcal{H}_k : \boldsymbol{x}_t = H_k \boldsymbol{\theta} + \boldsymbol{v}_t, \qquad k = 0, 1, \ldots, K, \qquad (1)$$

where H_k is the $M \times N$ observation matrix under hypothesis $\mathcal{H}_k$, $\boldsymbol{\theta}$ is the $N \times 1$ unknown state vector[1] to estimate, and $\boldsymbol{v}_t \sim \mathcal{N}(0, R_v)$ is the $M \times 1$ measurement noise that is independent and identically distributed (i.i.d.) over time. From the measurement data $\{\boldsymbol{x}_t\}$, we want to jointly infer a) the true underlying linear model H_k, and b) the true underlying states $\boldsymbol{\theta}$. Note that neither of them is known beforehand, and we need to solve a problem of jointly detecting $\mathcal{H}_k$ and estimating $\boldsymbol{\theta}$. Such problems arise in many applications. We provide in the following an example that arises commonly in power grid monitoring. An outage in a power grid will change the grid topology, and the system operator wants to detect which outage among a candidate set $\{\mathcal{H}_1, \ldots, \mathcal{H}_K\}$ occurs, or no outage occurs ($\mathcal{H}_0$). With a given set of sensors in the grid, the kth outage scenario gives rises to a unique observation matrix H_k, and the sensors measure the states of the grid $\boldsymbol{\theta}$ via (1). Consequently, state estimation depends on knowledge of the true outage, and outage detection depends on knowledge of the true states [2]. Clearly, solving a joint detection and estimation problem is essential for monitoring the health of the power grid in real time.

For these purposes, this paper provides the joint posterior distribution $p(\theta, \mathcal{H}_k | x^i)$ (see (12)–(13) below), which gives us the beliefs about both $\boldsymbol{\theta}$ and $\mathcal{H}_k$. Actually, it is also a sufficient statistic for $\boldsymbol{\theta}$ and $\mathcal{H}_k$ given data x^i, which provides *full information* from the measured data x^i about the hypothesis $\mathcal{H}_k$ and the state vector θ. Therefore, instead of being the optimal functions of x^i, the optimal decision rule and the estimator need only be the optimal functionals of $p(\theta, \mathcal{H}_k | x^i)$. Deriving the expressions for the joint posterior distribution will give us a unified framework for joint detection and estimation under all performance criteria (e.g., minimum risk/minimum probability of error/maximum a-posteriori probability (MAP) detection, and MAP/minimum-mean-square-error (MMSE) estimate).

III. JOINT POSTERIOR OF HYPOTHESES AND STATES

We now derive the joint posterior distribution of the hypothesis $\mathcal{H}_k$ and the unknown states $\boldsymbol{\theta}$. Specifically, we will use $p(\theta, \mathcal{H}_k | x^i)$ as a hybrid probability measure to denote the joint posterior distribution of $\boldsymbol{\theta}$ and $\mathcal{H}_k$:

$$p(\theta, \mathcal{H}_k | x^i) = p(\mathcal{H}_k | x^i) p(\theta | \mathcal{H}_k, x^i) \qquad (2)$$

where $p(\mathcal{H}_k | x^i)$ denotes the posterior probability mass function (PMF) of $\mathcal{H}_k$ and $p(\theta | \mathcal{H}_k, x^i)$ denotes the posterior probability density function (PDF) of $\boldsymbol{\theta}$ given $\mathcal{H}_k$.

A. The Likelihood Function

We begin with a factorization of the likelihood function $p(x^i | \theta, \mathcal{H}_k)$, which will be useful in finding sufficient statistics for jointly detecting $\mathcal{H}_k$ and estimating $\boldsymbol{\theta}$, and in computing the joint posterior distribution.

[1] In addition to states, θ can also include parameters in some applications [12], [13]. For the sake of brevity, we refer to θ as states from now on.

Lemma 1 (Factorization): According to the linear model in (1), we can express the conditional distribution (the likelihood function) $p(x^i | \theta, \mathcal{H}_k)$ in the following form:

$$p(x^i | \theta, \mathcal{H}_k) = p(x^i | \hat{\theta}_{k,\mathrm{ML}}, \mathcal{H}_k)$$
$$\times \exp\left\{ -\frac{1}{2} \left\| \theta - \hat{\theta}_{k,\mathrm{ML}} \right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})} \right\} \qquad (3)$$

where the notation $\|x\|^2_{\Sigma}$ denotes $x^T \Sigma x$ for some positive definite weighting matrix Σ, $\hat{\theta}_{k,\mathrm{ML}}$ is the maximum likelihood estimate of θ given that hypothesis $\mathcal{H}_k$ is true, and $I(\hat{\theta}_{k,\mathrm{ML}})$ is the corresponding Fisher information matrix:

$$\hat{\theta}_{k,\mathrm{ML}} = (H_k^T R_v^{-1} H_k)^{-1} H_k^T R_v^{-1} \overline{x}_i \qquad (4)$$

$$I(\hat{\theta}_{k,\mathrm{ML}}) = i \cdot (H_k^T R_v^{-1} H_k) \qquad (5)$$

$$\overline{x}_i = \frac{1}{i} \sum_{t=1}^{i} x_t. \qquad (6)$$

Proof: See Appendix I. ∎

In [15], an asymptotic expression similar to (3) was derived for general likelihood functions satisfying certain regularity conditions for i large. In comparison, our expression (3) holds for *all* $i \geq 1$ due to the properties of the linear model with Gaussian noise that we have assumed. Furthermore, the linear model (1) also allows us to evaluate the expression for $p(x^i | \hat{\theta}_{k,\mathrm{ML}}, \mathcal{H}_k)$ given by the following lemma.

Lemma 2 (Expression for $p(x^i | \hat{\theta}_{k,\mathrm{ML}}, \mathcal{H}_k)$): The conditional probability $p(x^i | \hat{\theta}_{k,\mathrm{ML}}, \mathcal{H}_k)$ can be expressed as

$$p(x^i | \hat{\theta}_{k,\mathrm{ML}}, \mathcal{H}_k) = \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}} \right]^i$$
$$\cdot \exp\left\{ -\frac{1}{2} \sum_{t=1}^{i} \|x_t\|^2_{R_v^{-1}} \right\}$$
$$\cdot \exp\left\{ \frac{1}{2} \left\| \hat{\theta}_{k,\mathrm{ML}} \right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})} \right\} \qquad (7)$$

where $\hat{\theta}_{k,\mathrm{ML}}$ and $I(\hat{\theta}_{k,\mathrm{ML}})$ are given by (4)–(6).

Proof: See Appendix II. ∎

Substituting (7) into (3), we obtain the following factorization of the likelihood function:

$$p(x^i | \theta, \mathcal{H}_k) = \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}} \right]^i$$
$$\cdot \exp\left\{ -\frac{1}{2} \sum_{t=1}^{i} \|x_t\|^2_{R_v^{-1}} \right\}$$
$$\cdot \exp\left\{ \frac{1}{2} \left\| \hat{\theta}_{k,\mathrm{ML}} \right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})} \right\}$$
$$\cdot \exp\left\{ -\frac{1}{2} \left\| \theta - \hat{\theta}_{k,\mathrm{ML}} \right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})} \right\}. \qquad (8)$$

Note that the first two terms in (8) are independent of the hypothesis index k and the state vector θ, while the other two terms depend on $\{\hat{\theta}_{k,\mathrm{ML}}\}$, which, by (4), is further determined by $\overline{x}_i$. Therefore, by the Neyman-Fisher factorization theorem [3], [15], [16], $\overline{x}_i$ is a sufficient statistic

for jointly detecting $\mathcal{H}_k$ and estimating θ. This fact will also be reflected further ahead in the joint posterior expressions (12)–(16), where $\overline{x}_i$ is the only statistic we need to track over time via, e.g.,

$$\overline{x}_i = \overline{x}_{i-1} + \frac{1}{i}(x_i - \overline{x}_{i-1}). \tag{9}$$

B. Conjugate Prior

For a certain likelihood function, if a prior distribution produces a posterior distribution of the same family, then such a prior distribution is called a conjugate prior. With a conjugate prior, we need only to maintain the recursions for the parameters that describe the distribution family of the prior and the posterior. We will use this kind of prior in our joint detection and estimation problem.

At the beginning (before any measurement data are available), we assume that the prior distribution of θ and $\mathcal{H}_k$ are given by

$$p(\theta, \mathcal{H}_k) = p(\mathcal{H}_k)p(\theta|\mathcal{H}_k) \tag{10}$$

where $p(\mathcal{H}_k)$ is the prior PMF of the hypothesis $\mathcal{H}_k$ and $p(\theta|\mathcal{H}_k)$ is the prior PDF of the state vector θ given hypothesis $\mathcal{H}_k$. Throughout the paper, we assume that given $\mathcal{H}_k$, θ has a Gaussian prior:

$$p(\theta|\mathcal{H}_k) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(C_{k,0})^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \|\theta - \theta_{k,0}\|^2_{C_{k,0}^{-1}} \right\} \tag{11}$$

where $\theta_{k,0}$ and $C_{k,0}$ are the corresponding prior mean and covariance matrix given hypothesis $\mathcal{H}_k$, respectively. We will show in the next section that this prior is indeed a conjugate prior. Furthermore, we will also show that even with an "uninformative prior" about θ (i.e., $C_{k,0} \rightarrow \infty$ in some way), the resulting posterior will take the same form as the joint prior distribution in (11). Therefore, alternatively, we can think of the conjugate prior in (11) as the intermediate knowledge that we learned from earlier data.

C. Main Results

Theorem 1 (Optimal joint inference): Suppose the prior distribution is given by (10)–(11). Then, the posterior distribution $p(\theta, \mathcal{H}_k|x^i) = p(\mathcal{H}_k|x^i)p(\theta|x^i, \mathcal{H}_k)$ is given by

$$p(\mathcal{H}_k|x^i) = \frac{1}{f(x^i)} p(\mathcal{H}_k) \sqrt{\frac{\det(C_{k,\mathrm{MMSE}})}{\det(C_{k,0})}}$$
$$\cdot \frac{\exp\left\{ \frac{1}{2} \|\hat{\theta}_{k,\mathrm{MMSE}}\|^2_{C_{k,\mathrm{MMSE}}^{-1}} \right\}}{\exp\left\{ \frac{1}{2} \|\theta_{k,0}\|^2_{C_{k,0}^{-1}} \right\}}, \tag{12}$$

$$p(\theta|\mathcal{H}_k, x^i) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(C_{k,\mathrm{MMSE}})^{\frac{1}{2}}}$$
$$\cdot \exp\left\{ -\frac{1}{2} \|\theta - \hat{\theta}_{k,\mathrm{MMSE}}\|^2_{C_{k,\mathrm{MMSE}}^{-1}} \right\}, \tag{13}$$

where

$$f(x^i) \triangleq \sum_{q=0}^{K} p(\mathcal{H}_q) \sqrt{\frac{\det(C_{q,\mathrm{MMSE}})}{\det(C_{q,0})}}$$

$$\cdot \frac{\exp\left\{ \frac{1}{2} \|\hat{\theta}_{q,\mathrm{MMSE}}\|^2_{C_{q,\mathrm{MMSE}}^{-1}} \right\}}{\exp\left\{ \frac{1}{2} \|\theta_{q,0}\|^2_{C_{q,0}^{-1}} \right\}}, \tag{14}$$

$$\hat{\theta}_{k,\mathrm{MMSE}} \triangleq \left(C_{k,0}^{-1} + I(\hat{\theta}_{k,\mathrm{ML}}) \right)^{-1}$$
$$\times \left(I(\hat{\theta}_{k,\mathrm{ML}}) \, \hat{\theta}_{k,\mathrm{ML}} + C_{k,0}^{-1}\theta_{k,0} \right), \tag{15}$$

and $C_{k,\mathrm{MMSE}} \triangleq \left(C_{k,0}^{-1} + I(\hat{\theta}_{k,\mathrm{ML}}) \right)^{-1}$. (16)

Proof: See Appendix III. ∎

Note that $\hat{\theta}_{k,\mathrm{MMSE}}$ is the classical MMSE estimate of θ given $\mathcal{H}_k$ is true, and $C_{k,\mathrm{MMSE}}$ is the corresponding error covariance matrix. In the posterior expression (12),

- $f(x^i)$ is a normalization factor.
- $p(\mathcal{H}_k)$ captures the prior PMF.
- The intuition of $\sqrt{\frac{\det(C_{k,\mathrm{MMSE}})}{\det(C_{k,0})}}$ is to penalize the model complexity of $\mathcal{H}_k$. This term reduces to $\sqrt{\det(I(\hat{\theta}_{k,\mathrm{ML}})^{-1})}$ in the case of an uninformative prior (see (18) below), for which a discussion of its meaning can be found in [15].
- The last term in (12) characterizes the similarity between the data (adjusted by the prior PDF) and the hypothesis $\mathcal{H}_k$.

Moreover, the exact dependence of the optimal detector on the state estimator can be seen from expression (12).

Accordingly, the posterior marginal distribution $p(\theta|x^i)$ for the state vector θ can be expressed as

$$p(\theta|x^i) = \sum_{k=0}^{K} p(\mathcal{H}_k|x^i) \cdot \frac{1}{(2\pi)^{\frac{N}{2}} \det(C_{k,\mathrm{MMSE}})^{\frac{1}{2}}}$$
$$\cdot \exp\left\{ -\frac{1}{2} \|\theta - \hat{\theta}_{k,\mathrm{MMSE}}\|^2_{C_{k,\mathrm{MMSE}}^{-1}} \right\} \tag{17}$$

which is a Gaussian mixture density with $K+1$ components.

We observe from (12)–(13) that the posterior distribution is in the same family as the prior distribution (10)–(11). Therefore, the prior we have chosen is a conjugate prior. As a result, we need only to maintain recursions for the parameters of the posterior distribution. Specifically, we need only to maintain recursions for $C_{k,\mathrm{MMSE}}$ and $\hat{\theta}_{k,\mathrm{MMSE}}$, where $\hat{\theta}_{k,\mathrm{MMSE}}$ is the only term that depends on data via $\hat{\theta}_{k,\mathrm{ML}}$. By (4), $\hat{\theta}_{k,\mathrm{ML}}$ is a linear function of $\overline{x}_i$, which, as we pointed out in Section III-A, is a sufficient statistic and can be updated recursively by (9). Therefore, as new data stream in, the optimal inference over the joint posterior $p(\theta, \mathcal{H}_k|x^i)$ can be implemented recursively using finite memory. This fact is also reflected in the marginal distribution $p(\theta|x^i)$, where the number of Gaussian mixture components remains at $K+1$ over time.

D. The Case without Prior Knowledge of States

When the inference algorithm has just started, we may not have any prior information about the states $\boldsymbol{\theta}$. In this case, we may assume that we are equally "uninformed" about the states under different hypotheses. In such a setup, we let the covariance matrices of the prior $p(\theta|\mathcal{H}_k)$ be

the same for all $\mathcal{H}_k$, i.e., $C_{k,0} = C_0$ for some invertible matrix, and let $C_0 \to \infty$. Applying this procedure to (15)–(16), the MMSE estimate $\hat{\theta}_{k,\mathrm{MMSE}}$ given $\mathcal{H}_k$ becomes the maximum likelihood estimate $\hat{\theta}_{k,\mathrm{ML}}$, and the MMSE $C_{k,\mathrm{MMSE}}$ given $\mathcal{H}_k$ becomes the inverse Fisher information matrix $I(\hat{\theta}_{k,\mathrm{ML}})^{-1}$. As a result, the optimal joint inference is given by the following corollary.

Corollary 1 ("Uninformative Prior"): Without any prior information about the states, the joint posterior distribution is given by

$$p(\mathcal{H}_k|x^i) = \frac{p(\mathcal{H}_k) \cdot \dfrac{\exp\left\{\frac{1}{2}\left\|\hat{\theta}_k\right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}}{\sqrt{\det\left(I(\hat{\theta}_{k,\mathrm{ML}})\right)}}}{\displaystyle\sum_{q=0}^{K} p(\mathcal{H}_q) \cdot \dfrac{\exp\left\{\frac{1}{2}\left\|\hat{\theta}_{q,\mathrm{ML}}\right\|^2_{I(\hat{\theta}_{q,\mathrm{ML}})}\right\}}{\sqrt{\det\left(I(\hat{\theta}_{q,\mathrm{ML}})\right)}}}, \tag{18}$$

$$p(\theta|\mathcal{H}_k, x^i) = \frac{1}{(2\pi)^{\frac{M}{2}} \det(I(\hat{\theta}_{k,\mathrm{ML}}))^{-\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}\|\theta - \hat{\theta}_k\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}. \tag{19}$$

As a consequence, the posterior marginal distribution $p(\theta|x^i)$ for the state vector θ can be expressed as

$$p(\theta|x^i) = \sum_{k=0}^{K} p(\mathcal{H}_k|x^i) \cdot \frac{1}{(2\pi)^{\frac{M}{2}} \det(I(\hat{\theta}_{k,\mathrm{ML}}))^{-\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}\|\theta - \hat{\theta}_k\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}. \tag{20}$$

∎

Note from (18)–(19) that the joint posterior distribution $p(\theta, \mathcal{H}_k|x^i)$ takes the same form as the joint posterior in (12)–(13) even when we start without any prior knowledge about $\boldsymbol{\theta}$. Therefore, the form of the joint prior distribution in (10)–(11) can also be viewed as the knowledge we learned from earlier data since they take the same form as the one in (18)–(19).

IV. Conclusions and future work

In this paper, we have studied the optimal joint detection and estimation problem in linear models with Gaussian noise. We have proved a factorization lemma of the likelihood function, and showed that the average measurement data vector $\overline{x}_i$ is a sufficient statistic. We then have derived a simple closed form expression of the joint posterior distribution for the hypotheses and the states. This expression reveals the exact dependence of the optimal detector on the state estimates. The joint posterior can then be used to develop optimal joint detector and estimator structures under any given performance criterion.

We have studied the case in which the states are static over time. It is interesting to generalize to the case in which the states evolve according to certain dynamics. In particular, we would like to investigate whether the joint posterior follows similar forms that only depend on the state estimates over

time. For this, it is essential to examine the evolution of the joint posterior from time $i-1$ to i, and we expect that similar techniques will apply. Furthermore, we have focused on the optimal fixed sample size approach. Developing an optimal sequential approach for joint detection and estimation from the joint posterior distribution remains an interesting direction for future work.

APPENDIX I
PROOF OF LEMMA 1

From the definition of Gaussian linear model (1) and because of the i.i.d. assumption on the noise, we can write the joint likelihood function as

$$p(x^i|\theta, \mathcal{H}_k)$$

$$= \prod_{t=1}^{i} p(x_t|\theta, \mathcal{H}_k)$$

$$= \prod_{t=1}^{i} \frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\|x_t - H_k\theta\|^2_{R_v^{-1}}\right\}$$

$$= \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}}\right]^i \exp\left\{-\frac{1}{2}\sum_{t=1}^{i}\|x_t - H_k\theta\|^2_{R_v^{-1}}\right\}$$

$$= \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}}\right]^i$$
$$\cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{i}\left\|x_t - H_k\hat{\theta}_{k,\mathrm{ML}} + H_k(\hat{\theta}_{k,\mathrm{ML}} - \theta)\right\|^2_{R_v^{-1}}\right\}$$

$$= \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}}\right]^i$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})^T R_v^{-1}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})\right)$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(\hat{\theta}_{k,\mathrm{ML}} - \theta)^T(H_k^T R_v^{-1} H_k)(\hat{\theta}_{k,\mathrm{ML}} - \theta)\right)$$
$$\cdot \exp\left(-\sum_{t=1}^{i}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})^T R_v^{-1} H_k(\hat{\theta}_{k,\mathrm{ML}} - \theta)\right)$$

$$= \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}}\right]^i$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})^T R_v^{-1}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})\right)$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(\hat{\theta}_{k,\mathrm{ML}} - \theta)^T(H_k^T R_v^{-1} H_k)(\hat{\theta}_{k,\mathrm{ML}} - \theta)\right)$$
$$\cdot \exp\left(-i(\overline{x}_i - H_k\hat{\theta}_{k,\mathrm{ML}})^T R_v^{-1} H_k(\hat{\theta}_{k,\mathrm{ML}} - \theta)\right)$$

$$= \left[\frac{1}{(2\pi)^{\frac{M}{2}} \det(R_v)^{\frac{1}{2}}}\right]^i$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})^T R_v^{-1}(x_t - H_k\hat{\theta}_{k,\mathrm{ML}})\right)$$

$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(\hat{\theta}_{k,\mathrm{ML}}-\theta)^{T}(H_{k}^{T}R_{v}^{-1}H_{k})(\hat{\theta}_{k,\mathrm{ML}}-\theta)\right)$$

$$\cdot \exp\Big(-i\Big[\overline{x}_{i}^{T}R_{v}^{-1}H_{k}(\hat{\theta}_{k,\mathrm{ML}}-\theta)$$
$$-\hat{\theta}_{k,\mathrm{ML}}^{T}H_{k}^{T}R_{v}^{-1}H_{k}(\hat{\theta}_{k,\mathrm{ML}}-\theta)\Big]\Big)$$

$$=\left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_{v})^{\frac{1}{2}}}\right]^{i}$$

$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})^{T}R_{v}^{-1}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})\right)$$

$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(\hat{\theta}_{k,\mathrm{ML}}-\theta)^{T}(H_{k}^{T}R_{v}^{-1}H_{k})(\hat{\theta}_{k,\mathrm{ML}}-\theta)\right)$$

$$\cdot \exp\Big(-i\Big[\underbrace{\overline{x}_{i}^{T}R_{v}^{-1}H_{k}(H_{k}^{T}R_{v}^{-1}H_{k})^{-1}}_{\hat{\theta}_{k,\mathrm{ML}}^{T}}$$
$$\times H_{k}^{T}R_{v}^{-1}H_{k}(\hat{\theta}_{k,\mathrm{ML}}-\theta)$$
$$-\hat{\theta}_{k,\mathrm{ML}}^{T}H_{k}^{T}R_{v}^{-1}H_{k}(\hat{\theta}_{k,\mathrm{ML}}-\theta)\Big]\Big)$$

$$=\left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_{v})^{\frac{1}{2}}}\right]^{i}$$

$$\cdot \exp\left(-\frac{1}{2}\sum_{t=1}^{i}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})^{T}R_{v}^{-1}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})\right)$$

$$\cdot \exp\left(-\frac{1}{2}(\hat{\theta}_{k,\mathrm{ML}}-\theta)^{T}(i\cdot H_{k}^{T}R_{v}^{-1}H_{k})(\hat{\theta}_{k,\mathrm{ML}}-\theta)\right)$$

$$=p(x|\hat{\theta}_{k,\mathrm{ML}},\mathcal{H}_{k})$$
$$\cdot \exp\left(-\frac{1}{2}(\hat{\theta}_{k,\mathrm{ML}}-\theta)^{T}I(\hat{\theta}_{k,\mathrm{ML}})(\hat{\theta}_{k,\mathrm{ML}}-\theta)\right). \tag{21}$$

APPENDIX II
PROOF OF LEMMA 2

By (21), the expression for $p(x|\hat{\theta}_{k,\mathrm{ML}},\mathcal{H}_{k})$ is given by

$$p(x^{i}|\hat{\theta}_{k,\mathrm{ML}},\mathcal{H}_{k})$$

$$=\left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_{v})^{\frac{1}{2}}}\right]^{i}$$

$$\cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{i}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})^{T}R_{v}^{-1}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})\right\} \tag{22}$$

For the summation in the exponent, we have

$$\sum_{t=1}^{i}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})^{T}R_{v}^{-1}(x_{t}-H_{k}\hat{\theta}_{k,\mathrm{ML}})$$

$$=\sum_{t=1}^{i}\Big[x_{t}^{T}R_{v}^{-1}x_{t}-2x_{t}^{T}R_{v}^{-1}H_{k}\hat{\theta}_{k,\mathrm{ML}}$$
$$+\hat{\theta}_{k,\mathrm{ML}}^{T}H_{k}^{T}R_{v}^{-1}H_{k}\hat{\theta}_{k,\mathrm{ML}}\Big]$$

$$=\sum_{t=1}^{i}x_{t}^{T}R_{v}^{-1}x_{t}-2i\overline{x}_{i}^{T}R_{v}^{-1}H_{k}\hat{\theta}_{k,\mathrm{ML}}$$

$$+i\cdot\hat{\theta}_{k,\mathrm{ML}}^{T}H_{k}^{T}R_{v}^{-1}H_{k}\hat{\theta}_{k,\mathrm{ML}}$$

$$=\sum_{t=1}^{i}x_{t}^{T}R_{v}^{-1}x_{t}-2\hat{\theta}_{k,\mathrm{ML}}^{T}I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}$$

$$+2\hat{\theta}_{k,\mathrm{ML}}^{T}I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}$$

$$=\sum_{t=1}^{i}x_{t}^{T}R_{v}^{-1}x_{t}-\hat{\theta}_{k,\mathrm{ML}}^{T}I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}. \tag{23}$$

Finally, substituting (23) into (22), we establish Lemma 2.

APPENDIX III
PROOF OF THEOREM 1

By Bayes' formula, the joint posterior distribution can be expressed as

$$p(\theta,\mathcal{H}_{k}|x^{i})=\frac{p(\theta,\mathcal{H}_{k})p(x^{i}|\theta,\mathcal{H}_{k})}{p(x^{i})}. \tag{24}$$

To compute the above posterior distribution, we need to have $p(x^{i})$ given by

$$p(x^{i})=\sum_{k=0}^{K}p(\mathcal{H}_{k})\cdot\int_{\theta\in\Theta}p(\theta)p(x^{i}|\theta,\mathcal{H}_{k})d\theta. \tag{25}$$

To proceed, we first introduce the following lemma that gives an integral result that would be useful in deriving both the optimal detection and estimation procedures.

Lemma 3 (A useful integral): Suppose we are given the Gaussian prior distribution (11). Then, the following result holds:

$$\int_{\theta\in\Theta}p(\theta)\exp\left\{-\frac{1}{2}\left\|\theta-\hat{\theta}_{k,\mathrm{ML}}\right\|_{I(\hat{\theta}_{k,\mathrm{ML}})}^{2}\right\}d\theta \tag{26}$$

$$=\exp\left\{\frac{1}{2}\left\|I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right\|_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}}^{2}\right\}$$

$$\times\exp\left\{-\frac{1}{2}\left(\|\theta_{k,0}\|_{C_{k,0}^{-1}}^{2}+\|\hat{\theta}_{k,\mathrm{ML}}\|_{I(\hat{\theta}_{k,\mathrm{ML}})}^{2}\right)\right\}$$

$$\times\frac{1}{\det(C_{k,0})^{\frac{1}{2}}\cdot\det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{\frac{1}{2}}} \tag{27}$$

where $\hat{\theta}_{k,\mathrm{ML}}$ and $I(\hat{\theta}_{k,\mathrm{ML}})$ are defined in (4)–(6).

Proof: See Appendix IV. ∎

Second, we compute the following integral and $p(x^{i})$ using the above lemma:

$$\int_{\theta\in\Theta}p(\theta)p(x^{i}|\theta,\mathcal{H}_{k})d\theta$$

$$=\left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_{v})^{\frac{1}{2}}}\right]^{i}\cdot\exp\left\{-\frac{1}{2}\sum_{t=1}^{i}\|x_{t}\|_{R_{v}^{-1}}^{2}\right\}$$

$$\cdot\exp\left\{\frac{1}{2}\|\hat{\theta}_{k,\mathrm{ML}}\|_{I(\hat{\theta}_{k,\mathrm{ML}})}^{2}\right\}$$

$$\cdot\int_{\theta\in\Theta}p(\theta)\exp\left\{-\frac{1}{2}\|\theta-\hat{\theta}_{k,\mathrm{ML}}\|_{I(\hat{\theta}_{k,\mathrm{ML}})}^{2}\right\}d\theta$$

$$=\left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_{v})^{\frac{1}{2}}}\right]^{i}\cdot\exp\left\{-\frac{1}{2}\sum_{t=1}^{i}\|x_{t}\|_{R_{v}^{-1}}^{2}\right\}$$

$$\cdot \exp\left\{\frac{1}{2}\left\|\hat{\theta}_{k,\mathrm{ML}}\right\|_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}$$

$$\cdot \exp\left\{\frac{1}{2}\left\|I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right\|^2_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}}\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}\left(\|\theta_{k,0}\|^2_{C_{k,0}^{-1}}+\|\hat{\theta}_{k,\mathrm{ML}}\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right)\right\}$$

$$\cdot \det(C_{k,0})^{-\frac{1}{2}}\cdot \det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-\frac{1}{2}} \tag{28}$$

$$p(x^i)$$

$$=\sum_{p=0}^{P} p(\mathcal{H}_k)\cdot \int_{\theta\in\Theta} p(\theta)p(x^i|\theta,\mathcal{H}_k)d\theta$$

$$=\sum_{p=0}^{P} p(\mathcal{H}_k)\cdot \left[\frac{1}{(2\pi)^{\frac{M}{2}}\det(R_v)^{\frac{1}{2}}}\right]^i$$

$$\cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{i}\|x_t\|^2_{R_v^{-1}}\right\}\cdot \exp\left\{\frac{1}{2}\|\hat{\theta}_{k,\mathrm{ML}}\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}$$

$$\cdot \exp\left\{\frac{1}{2}\left\|I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right\|^2_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}}\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}\left(\|\theta_{k,0}\|^2_{C_{k,0}^{-1}}+\|\hat{\theta}_{k,\mathrm{ML}}\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right)\right\}$$

$$\times \det(C_{k,0})^{-\frac{1}{2}}\cdot \det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-\frac{1}{2}}. \tag{29}$$

Substituting the above expressions and (8) into (24), we obtain (12)–(16) after some simple algebra.

APPENDIX IV
PROOF OF LEMMA 3

It follows that

$$\int_{\theta\in\Theta} p(\theta)\exp\left\{-\frac{1}{2}\left\|\theta-\hat{\theta}_{k,\mathrm{ML}}\right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}d\theta$$

$$=\frac{1}{(2\pi)^{\frac{N}{2}}\det(C_{k,0})^{\frac{1}{2}}}$$

$$\int_{\theta\in\Theta}\exp\left\{-\frac{1}{2}\|\theta-\theta_{k,0}\|^2_{C_{k,0}^{-1}}\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}\left\|\theta-\hat{\theta}_{k,\mathrm{ML}}\right\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right\}d\theta$$

$$=\exp\left\{\frac{1}{2}\left\|I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right\|^2_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}}\right\}$$

$$\times \exp\left\{-\frac{1}{2}\left(\|\theta_{k,0}\|^2_{C_{k,0}^{-1}}+\|\hat{\theta}_{k,\mathrm{ML}}\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right)\right\}$$

$$\times \frac{1}{\det(C_{k,0})^{\frac{1}{2}}\cdot \det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{\frac{1}{2}}}$$

$$\times \frac{1}{(2\pi)^{\frac{N}{2}}\det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-\frac{1}{2}}}$$

$$\times \int_{\theta\in\Theta}\exp\left\{-\frac{1}{2}\left\|\theta-\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}\right.\right.$$

$$\times \left.\left.\left(I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right)\right\|^2_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)}\right\}d\theta$$

$$=\exp\left\{\frac{1}{2}\left\|I(\hat{\theta}_{k,\mathrm{ML}})\hat{\theta}_{k,\mathrm{ML}}+C_{k,0}^{-1}\theta_{k,0}\right\|^2_{\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{-1}}\right\}$$

$$\times \exp\left\{-\frac{1}{2}\left(\|\theta_{k,0}\|^2_{C_{k,0}^{-1}}+\|\hat{\theta}_{k,\mathrm{ML}}\|^2_{I(\hat{\theta}_{k,\mathrm{ML}})}\right)\right\}$$

$$\times \frac{1}{\det(C_{k,0})^{\frac{1}{2}}\cdot 1\det\left(C_{k,0}^{-1}+I(\hat{\theta}_{k,\mathrm{ML}})\right)^{\frac{1}{2}}} \tag{30}$$

where the notation $\|x\|^2_{\Sigma}=x^T\Sigma x$, and in the last step we used the fact that the integral of a Gaussian distribution over the entire space equals one.

REFERENCES

[1] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*, Marcel Dekker, New York, 2004.

[2] Y. Zhao, A. Goldsmith, and H. V. Poor, "On PMU location selection for line outage detection in wide-area transmission networks," in *Proc. IEEE Power and Energy Society General Meeting*, San Diego, CA, July 2012, pp. 1–8.

[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.

[4] A. Goldsmith, *Wireless Communications*, Cambridge University Press, Cambridge, UK, 2005.

[5] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 434–444, May 1968.

[6] A. Fredriksen, D. Middleton, and V. VandeLinde, "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inf. Theory*, vol. 18, no. 5, pp. 607–614, Sep. 1972.

[7] D. G. Lainiotis, "Joint detection, estimation and system identification," *Information and Control*, vol. 19, no. 1, pp. 75–92, 1971.

[8] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4215–4229, 2012.

[9] O. Edfors, M. Sandell, J.-J. Van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[10] Y. Li, L. J. Cimini Jr, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[11] H. L. Van Trees, *Optimum Array Processing*, John Wiley & Sons, New York, 2002.

[12] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[13] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[14] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[15] S. M Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection theory*, Prentice Hall, Upper Saddle River, NJ, 1998.

[16] S. M Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, 1993.