On Filter Consistency of Discrete-time Nonlinear Systems with Partial-state Measurements

Guoquan P. Huang and Stergios I. Roumeliotis

***Abstract*— Linearized filters, such as the extended Kalman filter (EKF), often become inconsistent, when applied to observable nonlinear systems with partial-state measurements (where the full state cannot be reconstructed from the measurements at each time step). Relying on a novel decomposition of the observability matrix with respect to different measurement sources (sensors), we show that the standard EKF acquires spurious information from the measurements of each source, which erroneously reduces the uncertainty of the state estimates and hence causes inconsistency. With this key insight, we propose two EKF algorithms which compute the Jacobians in such as way so as to ensure that all decompositions of the observability matrix have nullspace of correct dimension. In the first, the linearization points are selected so as to minimize linearization errors under the constraints that the decompositions of the observability matrix have the appropriate nullspace. In the second, we project the most accurate measurement Jacobian (i.e., computed using the latest, and thus the best, state estimates as in the standard EKF) onto the actual information-available directions. We test the proposed algorithms in a two-radar target tracking example and show that significant performance improvement over the standard EKF is attained.**

I. Introduction

An issue of concern with most nonlinear filtering problems (e.g., target tracking [1]) is that of inconsistency; that is, no provably consistent estimator can be constructed for a nonlinear system, and the consistency of every filter has to be evaluated experimentally. As defined in [2], an estimator is consistent if the estimation errors are zero-mean and have covariance smaller or equal to the one calculated by the filter. Consistency is one of the primary criteria for evaluating the performance of any filter; if a filter is inconsistent, then its estimation accuracy is unknown, which in turn makes the filter unreliable. To date, the problem of estimator inconsistency has been studied primarily for the case of unobservable systems (e.g., robot localization [3]–[8]), while very little is known about the causes of estimator inconsistency in observable systems.

In this paper, we revisit the problem of filter inconsistency for a broad class of discrete-time nonlinear systems, by examining the directions of the state space along which information is available from the measurements of each source (sensor). Based on this analysis, we propose a novel methodology to improve consistency by ensuring that the filter acquires information from each source's measurements along the correct directions of the state space.

In particular, the Fisher information matrix (FIM) [2] for given measurements encapsulates all the available information about the entire state of a stochastic system. By marginalizing all but the initial state, we obtain the corresponding FIM that contains all information available in the measurements for determining the initial state. Studying the FIM's structure reveals the directions along which information is (un)available from the measurements. These can be exploited in the design of estimation algorithms, i.e., by enforcing estimators to gain information from measurements only along the "correct" directions (where information is actually available). Moreover, we show that the FIM of the initial state can be factorized in terms of the observability matrix of the corresponding deterministic system. Based on this key finding, and in order to improve consistency, we impose the constraint of acquiring information along the correct directions on the resulting (with respect to different measurement sources) decompositions of the observability matrix, instead of the FIM. To this end, we introduce two different EKF algorithms that compute the appropriate filter Jacobians, either directly (i.e., by projecting the most accurate Jacobians onto the information-available subspace) or indirectly (i.e., by first finding optimal linearization points for computing the Jacobians). As a result, only information actually available from each source's measurements is gained, which substantially improves the estimation consistency and accuracy, as opposed to the standard EKF.

II. Problem Formulation and Solution

We consider a general discrete-time nonlinear system of the following form:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k \tag{1}$$

$$\mathbf{z}_{i,k} = \mathbf{h}(\mathbf{x}_k, \mathbf{s}_{i,k}) + \mathbf{v}_{i,k} \ , \quad i \in \{1, \ldots, s\} \tag{2}$$

where $\mathbf{x}_k \in \mathcal{R}^n$ denotes the state of the system, $\mathbf{u}_k \in \mathcal{R}^\eta$ is the control input, and $\mathbf{w}_k \in \mathcal{R}^n$ is zero-mean white Gaussian process noise, i.e., $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$. $\mathbf{z}_{i,k} \in \mathcal{R}^m$ is the measurement taken from the i-th ($i \in \{1, \ldots, s\}$) measurement source (e.g., sensor), and is generally (although not necessarily) of *lower* dimension than the state vector, i.e., $m < n$, which is the case of partial-state measurements we consider in this work. The parameter $\mathbf{s}_{i,k}$ denotes the known parameters of the i-th measurement source, such as the sensor's location or a binary indicator of the availability of the i-th measurement. The random variable $\mathbf{v}_{i,k} \in \mathcal{R}^m$ is

This work was supported by the University of Minnesota through the Digital Technology Center (DTC) and AFOSR MURI FA9550-10-1-0567.

G. P. Huang is with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: ghuang@csail.mit.edu. S. I. Roumeliotis is with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Email: stergios@cs.umn.edu.

zero-mean white Gaussian measurement noise, i.e., $\mathbf{v}_{i,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{i,k})$.

We employ the EKF to recursively compute the state estimate and error covariance. Specifically, we linearize the nonlinear system at the linearization points, $\mathbf{x}^{\star}_{k|k-1}$ and $\mathbf{x}^{\star}_{k|k}$ (i.e., the linearization points before and after the update at time step k) [see (1) and (2)], and obtain the following linearized error-state system:[1]

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{\Phi}_k \tilde{\mathbf{x}}_{k|k} + \mathbf{w}_k \tag{3}$$

$$\tilde{\mathbf{z}}_{i,k|k-1} = \mathbf{H}_{i,k} \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_{i,k} \ , \quad i \in \{1, \ldots, s\} \tag{4}$$

where

$$\mathbf{\Phi}_k = \nabla_{\mathbf{x}_k} \mathbf{f} \Big|_{\{\mathbf{x}^{\star}_{k|k}, \mathbf{x}^{\star}_{k+1|k}\}} \ , \quad \mathbf{H}_{i,k} = \nabla_{\mathbf{x}_k} \mathbf{h} \Big|_{\{\mathbf{x}^{\star}_{k|k-1}\}} \tag{5}$$

The standard choice of linearization point is the latest state estimate, which, however, we will show is not necessarily the best choice. Once the propagation and measurement Jacobians are computed, we propagate and update the state estimate and covariance, respectively, as follows [2]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k) \tag{6}$$

$$\mathbf{P}_{k+1|k} = \mathbf{\Phi}_k \mathbf{P}_{k|k} \mathbf{\Phi}_k^T + \mathbf{Q}_k \tag{7}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{r}_k \tag{8}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \tag{9}$$

where $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_{i,k}^T \mathbf{S}_k^{-1}$ is the Kalman gain, $\mathbf{r}_k = \mathbf{z}_{i,k} - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{s}_{i,k})$ is the measurement residual, and $\mathbf{S}_k = \mathbf{H}_{i,k} \mathbf{P}_{k|k-1} \mathbf{H}_{i,k}^T + \mathbf{R}_{i,k}$ is the corresponding residual covariance.

A. Observability and Fisher Information

Since the EKF is constructed based on the linearized system [see (3) and (4)], it is important to study the observability properties of the corresponding deterministic system (i.e., noise free). Observability examines whether the information provided by the available measurements is sufficient for estimating the initial state without ambiguity. In particular, the observability matrix for the linearized system (3)-(4) during the time interval $[0, k]$ is defined by [9], [10]:

$$\mathbf{M} = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{\Phi}_0 \\ \vdots \\ \mathbf{H}_k \mathbf{\Phi}_{k-1} \cdots \mathbf{\Phi}_0 \end{bmatrix} \tag{10}$$

If the system is observable, then the corresponding observability matrix $\mathbf{M}$ is full-rank.

The FIM [2] is closely related to the system observability and precisely describes the information available in the measurements. Thus, by studying its properties, we can also gain insight about the directions in the state space along which information is actually available. To this end, we examine the structure of the Hessian (information) matrix of the corresponding batch maximum a posteriori (MAP) estimator over the time interval $[0, k]$, which is known to be optimal [11]. In what follows, we show that the FIM of the initial state $\mathbf{x}_0$ (obtained by marginalization) has the same rank properties as the observability matrix, which motivates us to instead examine the observability matrix in our analysis.

The optimal batch-MAP estimator utilizes all available information to estimate the *entire* state trajectory that is formed by stacking all states in the time interval $[0, k]$:

$$\mathbf{x}_{0:k} = \begin{bmatrix} \mathbf{x}_0^T & \mathbf{x}_1^T & \cdots & \mathbf{x}_k^T \end{bmatrix}^T \tag{11}$$

Specifically, the batch-MAP estimator seeks to determine the entire state-space trajectory estimate $\hat{\mathbf{x}}_{0:k|k}$ by maximizing the following posterior pdf (assuming *no* prior is available):

$$p(\mathbf{x}_{0:k} | \mathbf{z}_{0:k}) \propto \prod_{\kappa=0}^{k-1} p(\mathbf{x}_{\kappa+1} | \mathbf{x}_\kappa) \prod_{\kappa=0}^{k} p(\mathbf{z}_{i,\kappa} | \mathbf{x}_\kappa) \tag{12}$$

where $\mathbf{z}_{0:k}$ denotes all the sensor measurements in the time interval $[0, k]$. In the above expression, we have employed the assumption of independent state and measurement noise and the Markovian property of the system dynamics [see (1), and (2), respectively]. Moreover, using the assumption of Gaussian noise, the posterior pdf (12) can be written as:

$$\begin{aligned} & p(\mathbf{x}_{0:k} | \mathbf{z}_{0:k}) \propto \\ & \prod_{\kappa=0}^{k-1} \frac{1}{\sqrt{|2\pi \mathbf{Q}_\kappa|}} \exp\left(-\frac{1}{2} ||\mathbf{x}_{\kappa+1} - \mathbf{f}(\mathbf{x}_\kappa, \mathbf{u}_\kappa)||^2_{\mathbf{Q}_\kappa} \right) \times \\ & \prod_{\kappa=0}^{k} \frac{1}{\sqrt{|2\pi \mathbf{R}_{i,\kappa}|}} \exp\left(-\frac{1}{2} ||\mathbf{z}_{i,\kappa} - \mathbf{h}(\mathbf{x}_\kappa, \mathbf{s}_{i,\kappa})||^2_{\mathbf{R}_{i,\kappa}} \right) \end{aligned} \tag{13}$$

where we have employed the notation, $||\mathbf{a}||^2_{\mathbf{\Lambda}} \triangleq \mathbf{a}^T \mathbf{\Lambda}^{-1} \mathbf{a}$. Due to the monotonicity of the negative logarithm, maximizing (13) is equivalent to minimizing the following cost function under mild assumptions:

$$\begin{aligned} c(\mathbf{x}_{0:k}) = & \sum_{\kappa=0}^{k-1} \frac{1}{2} ||\mathbf{x}_{\kappa+1} - \mathbf{f}(\mathbf{x}_\kappa, \mathbf{u}_\kappa)||^2_{\mathbf{Q}_\kappa} \\ & + \sum_{\kappa=0}^{k} \frac{1}{2} ||\mathbf{z}_{i,\kappa} - \mathbf{h}(\mathbf{x}_\kappa, \mathbf{s}_{i,\kappa})||^2_{\mathbf{R}_{i,\kappa}} \end{aligned} \tag{14}$$

The information (Hessian) matrix can be computed as:

$$\mathbf{A} = \sum_{\kappa=0}^{k-1} \mathcal{F}_\kappa^T \mathbf{Q}_\kappa^{-1} \mathcal{F}_\kappa + \sum_{\kappa=0}^{k} \mathcal{H}_\kappa^T \mathbf{R}_\kappa^{-1} \mathcal{H}_\kappa \tag{15}$$

with

$$\mathcal{F}_\kappa = \begin{bmatrix} \mathbf{0} & \cdots & -\mathbf{\Phi}_\kappa & \mathbf{I}_n & \cdots & \mathbf{0} \end{bmatrix} \tag{16}$$

$$\mathcal{H}_\kappa = \begin{bmatrix} \mathbf{0} & \cdots & -\mathbf{H}_\kappa & \cdots & \mathbf{0} \end{bmatrix} \tag{17}$$

where $\mathbf{H}_\kappa = \mathbf{H}_{i,\kappa}$ and $\mathbf{R}_\kappa = \mathbf{R}_{i,\kappa}$, if the i-th source provides the measurement at time step κ, i.e., $\mathbf{z}_\kappa = \mathbf{z}_{i,\kappa}$ (note that hereafter we will use these notations interchangeably). Note also that due to the sparse structure of $\mathcal{F}_\kappa$ and $\mathcal{H}_\kappa$ [see (16) and (17)], the FIM (15) has banded structure (18).

We now show that the Schur complement of the full FIM with respect to the initial state $\mathbf{x}_0$ (i.e., the information matrix of $\mathbf{x}_0$, denoted by $\mathbf{A}_0$), has the following relation to the observability matrix $\mathbf{M}$:

[1] Throughout this paper, the subscript $\ell|j$ refers to the estimate of a quantity at time step ℓ, after all measurements up to time step j have been processed. $\hat{x}$ is used to denote the estimate of a random variable x, while $\tilde{x} = x - \hat{x}$ is the error in this estimate. $\mathbf{0}_{m \times n}$ denotes $m \times n$ matrices of zeros, and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

$$\mathbf{A} = \begin{bmatrix} \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\mathbf{\Phi}_0 + \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0 & -\mathbf{\Phi}_0^T\mathbf{Q}_0^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{Q}_0^{-1}\mathbf{\Phi}_0 & \mathbf{Q}_0^{-1} + \mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\mathbf{\Phi}_1 + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 & -\mathbf{\Phi}_1^T\mathbf{Q}_1^{-1} & \cdots & \mathbf{0} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0} & \cdots & -\mathbf{Q}_{k-2}^{-1}\mathbf{\Phi}_{k-2} & \mathbf{Q}_{k-2}^{-1} + \mathbf{\Phi}_{k-1}^T\mathbf{Q}_{k-1}^{-1}\mathbf{\Phi}_{k-1} + \mathbf{H}_{k-1}^T\mathbf{R}_{k-1}^{-1}\mathbf{H}_{k-1} & -\mathbf{\Phi}_{k-1}^T\mathbf{Q}_{k-1}^{-1} \\ \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{Q}_{k-1}^{-1}\mathbf{\Phi}_{k-1} & \mathbf{Q}_{k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k \end{bmatrix} \tag{18}$$

Lemma 2.1: The FIM of the initial state $\mathbf{x}_0$, i.e., the corresponding Schur complement of the full FIM, can be factorized as:

$$\mathbf{A}_0 = \mathbf{M}^T\mathbf{\Sigma}\mathbf{M} \tag{19}$$

where $\mathbf{M}$ is the observability matrix and $\mathbf{\Sigma}$ is a nonsingular (full-rank) real symmetric block-diagonal matrix.

Proof: See Appendix I. ■

From this lemma, it is evident that the FIM of the initial state and the observability matrix are closely related. In particular, the FIM can be seen as the "observability gramian" for the corresponding stochastic system. Note also that the stochastic system (1)-(2) is observable if and only if the corresponding deterministic system is observable. In the following, we will exploit this result and decompose the observability matrix in a novel way (i.e., based on the measurement sources), which inspires the proposed filters for improving consistency.

B. Algorithms

Due to the additive property of the measurement information, we decompose the FIM according to the measurements that originated from each of the s sources [see (15)]:

$$\begin{aligned} \mathbf{A} &= \sum_{\kappa=0}^{k-1} \mathcal{F}_\kappa^T\mathbf{Q}_\kappa^{-1}\mathcal{F}_\kappa + \sum_{i=1}^{s}\sum_{\kappa=0}^{k} \mathcal{H}_{i,\kappa}^T\mathbf{R}_{i,\kappa}^{-1}\mathcal{H}_{i,\kappa} \\ &= \sum_{i=1}^{s} \underbrace{\left(\sum_{\kappa=0}^{k-1} \mathcal{F}_\kappa^T\mathbf{Q}_\kappa'^{\,-1}\mathcal{F}_\kappa + \sum_{\kappa=0}^{k} \mathcal{H}_{i,\kappa}^T\mathbf{R}_{i,\kappa}^{-1}\mathcal{H}_{i,\kappa} \right)}_{\mathbf{A}_i} \end{aligned} \tag{20}$$

where $\mathbf{Q}_\kappa' \triangleq s\mathbf{Q}_\kappa$ denotes the inflated state-noise covariance for the i-th source used in order to compensate for the decomposition. Hence, $\mathbf{A}_i$ is considered as the full FIM constructed using measurements *only* from the i-th source. Based on Lemma 2.1, the corresponding FIM of the initial state $\mathbf{A}_{0_i}$ can be written as:

$$\mathbf{A}_{0_i} = \mathbf{M}_i^T\mathbf{\Sigma}\mathbf{M}_i \tag{21}$$

It is important to note that $\mathbf{M}_i$ in (21) is the "observability matrix" which is constructed using the measurements only from the i-th source, *but* padded with zeros in the places corresponding to the measurements from the other sources, in order to match the dimension of the full observability matrix $\mathbf{M}$ [e.g., see (23)]. This immediately results in $\mathbf{M}_i^T\mathbf{M}_j = \mathbf{0}$ for $i \neq j$. Note also that we directly use $\mathbf{\Sigma}$ in (21), since zeros in $\mathbf{M}_i$ will cancel out the corresponding submatrices in $\mathbf{\Sigma}$ to the measurements from the j-th source ($j \neq i$) [see (60)]. Therefore, this result (21) leads to the following decomposition of the observability matrix:

Lemma 2.2: The observability matrix is decomposed as:

$$\mathbf{M} = \sum_{i=1}^{s} \mathbf{M}_i \tag{22}$$

Proof: Using the fact that $\mathbf{M}_i^T\mathbf{M}_j = \mathbf{0}$ for $i \neq j$, we have [see Lemma 2.1, (20) and (21)]:

$$\mathbf{M}^T\mathbf{\Sigma}\mathbf{M} = \mathbf{A}_0 = \sum_{i=1}^{s}\mathbf{A}_{0_i} = \sum_{i=1}^{s}\mathbf{M}_i^T\mathbf{\Sigma}\mathbf{M}_i \;\Rightarrow\; \mathbf{M} = \sum_{i=1}^{s}\mathbf{M}_i$$

■

We thus see that based on the decomposition of the FIM of the initial state according to the measurement sources, the observability matrix can be *accordingly* decomposed. For instance, if the i-th source provides measurement intermittently at even time steps only, then the i-th decomposition of the observability matrix, $\mathbf{M}_i$, assumes the following form (by assuming k is even):

$$\mathbf{M}_i = \begin{bmatrix} \mathbf{H}_{i,0} \\ \mathbf{0} \\ \vdots \\ \mathbf{H}_{i,k-2}\mathbf{\Phi}_{k-3}\cdots\mathbf{\Phi}_0 \\ \mathbf{0} \\ \mathbf{H}_{i,k}\mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_0 \end{bmatrix} \tag{23}$$

It is interesting to note that in many cases (e.g., see Section III) the decomposition of the observability matrix (23), $\mathbf{M}_i$, is *rank-deficient*, although the observability matrix (10), $\mathbf{M}$, is *full-rank*, i.e., the linearized system (3)-(4) is observable. The right nullspace of the matrix, $\mathbf{M}_i$, and thus the decomposition of the FIM (21), $\mathbf{A}_{0_i}$, dictates the directions of the state space along which no information is available from the measurements from the i-th source. If these directions are incorrect, the filter acquires spurious information from the i-th source's measurement, and hence is expected to become inconsistent. Therefore, to ensure consistent estimation, the filter should have $\mathbf{M}_i$, and hence $\mathbf{A}_{0_i}$, of correct nullspace, for $i = 1, \ldots, s$, so that no nonexistent information is gained from the measurements available from each source. To this end, in computing the filter Jacobians at each time step, we explicitly enforce the following constraint on the decompositions of the observability matrix, i.e., each $\mathbf{M}_i$ has correct nullspace denoted by $\mathbf{N}_i$ [see (23)]:

$$\mathbf{M}_i\mathbf{N}_i = \mathbf{0} \;\Leftrightarrow\; \begin{cases} \mathbf{H}_{i,0}\mathbf{N}_i = \mathbf{0} & ,\text{ if } \kappa = 0 \\ \mathbf{H}_{i,\kappa}\mathbf{\Phi}_{\kappa-1}\cdots\mathbf{\Phi}_0\mathbf{N}_i = \mathbf{0} & ,\text{ if } \kappa > 0 \end{cases} \tag{24}$$

In particular, $\mathbf{N}_i$ is a design choice which defines the desired nullspace for the i-th measurement source, and one practical choice will be the nullspace of the first measurement Jacobian $\mathbf{H}_{i,0}$, i.e., $\mathbf{H}_{i,0}\mathbf{N}_i = \mathbf{0}$. Once $\mathbf{N}_i$ has been selected, the next design decision is to compute the filter Jacobians appropriately so that (24) is satisfied.

We first propose to compute the Jacobians *indirectly*, i.e., to find optimal linearization points that minimize the linearization errors of the points $\mathbf{x}^\star_{k|k}$ and $\mathbf{x}^\star_{k+1|k}$ used in computing the filter Jacobians, $\mathbf{\Phi}_k$ and $\mathbf{H}_{k+1}$, at time step $k+1$, subject to the constraint that $\mathbf{M}_i$ has the correct nullspace (24). This can be formulated as the following constrained minimization problem:

$$\min_{\mathbf{x}^\star_{k|k},\mathbf{x}^\star_{k+1|k}} \int ||\mathbf{x}^\star_k - \mathbf{x}_k||^2 p(\mathbf{x}_k|z_{0:k}) d\mathbf{x}_k + \int ||\mathbf{x}^\star_{k+1|k} - \mathbf{x}_{k+1}||^2 p(\mathbf{x}_{k+1}|z_{0:k}) d\mathbf{x}_{k+1} \quad (25)$$

$$\text{subject to} \quad \mathbf{H}_{i,k}\mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_0\mathbf{N}_i = \mathbf{0}\ ,\ \forall i = 1,\ldots,s \quad (26)$$

In general it is intractable to solve this problem analytically. However, when $p(\mathbf{x}_k|z_{0:k})$ and $p(\mathbf{x}_{k+1}|z_{0:k})$ are Gaussian which is the assumption employed in the EKF, this problem can be simplified based on the following lemma, and then solved analytically by using Lagrangian multipliers:

Lemma 2.3: When $p(\mathbf{x}_k|z_{0:k})$ and $p(\mathbf{x}_{k+1}|z_{0:k})$ are Gaussian, the constrained minimization problem (25)-(26) is equivalent to the following problem:

$$\min_{\mathbf{x}^\star_{k|k},\mathbf{x}^\star_{k+1|k}} ||\mathbf{x}^\star_{k|k} - \hat{\mathbf{x}}_{k|k}||^2 + ||\mathbf{x}^\star_{k+1|k} - \hat{\mathbf{x}}_{k+1|k}||^2 \quad (27)$$

$$\text{subject to} \quad \mathbf{H}_{i,k}\mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_0\mathbf{N}_i = \mathbf{0}\ ,\ \forall i = 1,\ldots,s \quad (28)$$

Proof: See Appendix II. ■

Alternatively, we can compute the desired filter Jacobians *directly*. In particular, we compute the propagation Jacobian $\mathbf{\Phi}_\kappa$ ($\kappa = 0,\ldots,k-1$) in the same way as in the standard EKF, while enforcing the information constraint (24) for computing the measurement Jacobian:

$$\min_{\mathbf{H}_{i,k}} ||\mathbf{H}_{i,k} - \mathbf{H}_o||_F^2 \quad (29)$$

$$\text{subject to} \quad \mathbf{H}_{i,k}\mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_0\mathbf{N}_i = \mathbf{0}\ ,\ \forall i = 1,\ldots,s \quad (30)$$

where $||\cdot||_F$ denotes the Frobenius norm. In the above expression, $\mathbf{H}_o$ ideally is the measurement Jacobian computed using the true states, which, however, is not realizable in practice. Hence, we employ the latest, and thus the best, state estimates for computing this Jacobian as in the standard EKF, i.e., $\mathbf{H}_o = \mathbf{H}_o(\hat{\mathbf{x}}_{k|k-1})$. The optimal solution to the above problem (29)-(30) is obtained in *closed form* by application of the following lemma:

Lemma 2.4: The optimal solution to the constrained minimization problem (29)-(30) is:

$$\mathbf{H}_{i,k} = \mathbf{H}_o\left(\mathbf{I}_n - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T\right) \quad (31)$$

where $\mathbf{U}_i = \mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_0\mathbf{N}_i$.

Proof: See Appendix III. ■

Note that $\mathbf{U}_i$ in (31) is the propagated nullspace of the i-th source at time step k, and $\left(\mathbf{I}_n - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T\right)$ is the subspace orthogonal to $\mathbf{U}_i$, i.e., the subspace at time step k where information is available.

III. Application: Two-Radar Target Tracking

In order to verify the preceding analysis and validate the proposed methodology, in this section, we consider a particular application of two radars tracking a target. Consider a target (robot) that moves on a plane and two radars alternatively provide distance measurements to the target (see Fig. 1). Using such intermittent distance measurements as well as the target's motion model, the EKF is employed to estimate the target's pose (position and orientation) in a global frame of reference, denoted by $\mathbf{x}_k = \begin{bmatrix}\mathbf{p}_k^T & \phi_k\end{bmatrix}^T = \begin{bmatrix}x_k & y_k & \phi_k\end{bmatrix}^T$. In what follows, we first describe the motion and measurement models of this system in the context of the standard EKF.

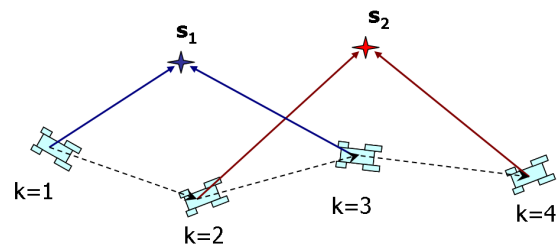


Fig. 1. Illustration of the application of the two-radar target tracking: A target (robot) moves on a plane and two radars, s_1 and s_2, alternate between measuring distance to the target. For example, at time step $k=1$, the first radar, s_1, measures distance to the target; at time step $k=2$, the second radar, s_2, measures distance to the target; at time step $k=3$, s_1 measures distance again; and so on so forth.

In the propagation step, the target (robot) is assumed to have an odometer whose measurements are processed to obtain an estimate of the pose change between two consecutive time steps, and then employed in the EKF to propagate the state estimate. The EKF state propagation equations are given by:

$$\hat{\mathbf{p}}_{k+1|k} = \hat{\mathbf{p}}_{k|k} + \mathbf{C}(\hat{\phi}_{k|k})^k\hat{\mathbf{p}}_{k+1} \quad (32)$$

$$\hat{\phi}_{k+1|k} = \hat{\phi}_{k|k} + {}^k\hat{\phi}_{k+1} \quad (33)$$

where $\mathbf{C}(\cdot)$ denotes the 2×2 rotation matrix, and ${}^k\hat{\mathbf{x}}_{k+1} = [{}^k\hat{\mathbf{p}}_{k+1}^T \quad {}^k\hat{\phi}_{k+1}]^T$ is the odometry-based estimate of the target's motion between time steps k and $k+1$. This estimate is corrupted by zero-mean white Gaussian noise $\mathbf{w}_k = {}^k\mathbf{x}_{k+1} - {}^k\hat{\mathbf{x}}_{k+1}$, with covariance matrix $\mathbf{Q}_k$. Clearly this process model is nonlinear, and hence the linearized error-state propagation equation is necessary for the EKF, i.e.,

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{\Phi}_k\tilde{\mathbf{x}}_{k|k} + \mathbf{G}_k\mathbf{w}_k \quad (34)$$

where the state and noise Jacobians are given by:

$$\mathbf{\Phi}_k = \begin{bmatrix}\mathbf{I}_2 & \mathbf{J}\left(\hat{\mathbf{p}}_{k+1|k} - \hat{\mathbf{p}}_{k|k}\right)\\ \mathbf{0}_{1\times 2} & 1\end{bmatrix} \quad (35)$$

$$\mathbf{G}_k = \begin{bmatrix}\mathbf{C}(\hat{\phi}_{k|k}) & \mathbf{0}_{2\times 1}\\ \mathbf{0}_{1\times 2} & 1\end{bmatrix}\ ,\quad \mathbf{J} \triangleq \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}. \quad (36)$$

The distance measurement at time step k is given by:

$$z_{i,k} = ||\mathbf{p}_k - \mathbf{p}_{S_i}|| + v_{i,k} = \sqrt{(x_k - x_{S_i})^2 + (y_k - y_{S_i})^2} + v_{i,k}\ ,\ i = 1,2 \quad (37)$$

where $\mathbf{p}_{S_i} \triangleq [x_{S_i}\ y_{S_i}]^T$ is the known position of the i-th radar expressed in the global frame of reference, and $v_{i,k}$ is zero-mean white Gaussian measurement noise, with variance $\sigma_{i,k}^2$, i.e., $v_{i,k} \sim \mathcal{N}(0,\sigma_{i,k}^2)$. The linearized measurement model is:

$$\tilde{z}_{i,k} \simeq \mathbf{H}_{i,k}\tilde{\mathbf{x}}_{k|k-1} + v_{i,k} \quad (38)$$

with Jacobian:

$$\mathbf{H}_{i,k} = \begin{bmatrix} \frac{(\hat{\mathbf{p}}_{k|k-1}-\mathbf{p}_{S_i})^T}{||\hat{\mathbf{p}}_{k|k-1}-\mathbf{p}_{S_i}||} & 0 \end{bmatrix}$$
$$= \underbrace{\frac{(\hat{\mathbf{p}}_{k|k-1}-\mathbf{p}_{S_i})^T}{||\hat{\mathbf{p}}_{k|k-1}-\mathbf{p}_{S_i}||}}_{\hat{\boldsymbol{\alpha}}_{i,k}} \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{k|k-1}-\mathbf{p}_{S_i}) \end{bmatrix} \quad (39)$$

A. Observability of the EKF Linearized Systems

We now examine the observability matrix (and thus the FIM), and show that in the standard EKF each of the decompositions of the observability matrix, with respect to different radars, has different nullspace than the ideal case, i.e., when the Jacobians are evaluated at the true states and which is expected to have correct observability properties.

1) Ideal EKF: To facilitate the ensuing analysis, we begin with the *ideal* case of a *single* radar providing distance measurements. In this ideal EKF, the filter Jacobians are evaluated using the *true* values of the state variables, i.e., $\mathbf{x}^\star_{k|k} = \mathbf{x}^\star_{k|k-1} = \mathbf{x}_k$, for all k. In the following, all matrices evaluated using the true state values are denoted by the symbol " $\breve{}$ ". In this case, by noting that [see (35) and (39)]

$$\breve{\mathbf{H}}_\kappa \breve{\boldsymbol{\Phi}}_{\kappa-1} \cdots \breve{\boldsymbol{\Phi}}_0 = \breve{\boldsymbol{\alpha}}_\kappa \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_S) \end{bmatrix} \quad (40)$$

the observability matrix is computed by [see (10)]:

$$\breve{\mathbf{M}} = \mathbf{Diag}\left(\breve{\boldsymbol{\alpha}}_0, \breve{\boldsymbol{\alpha}}_1, \cdots, \breve{\boldsymbol{\alpha}}_k\right) \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_S) \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_S) \\ \vdots & \vdots \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_S) \end{bmatrix} \quad (41)$$

Using the theorem of the rank of the matrix product [12], we can show that $\mathrm{rank}(\breve{\mathbf{M}}) = 2$ and $\mathrm{null}(\breve{\mathbf{M}}) = \begin{bmatrix} \mathbf{J}(\mathbf{p}_S - \mathbf{p}_0) \\ 1 \end{bmatrix}$.

Now we extend this analysis to the ideal case where *two* radars alternatively provide distance measurements. Specifically, by proceeding similarly, the observability matrix of the ideal-EKF linearized system can be computed as:

$$\breve{\mathbf{M}} = \underbrace{\mathbf{Diag}\left(\breve{\boldsymbol{\alpha}}_{1,0}, \breve{\boldsymbol{\alpha}}_{2,1}, \cdots, \breve{\boldsymbol{\alpha}}_{1,k-1}, \breve{\boldsymbol{\alpha}}_{2,k}\right)}_{\breve{\mathbf{D}}} \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_{S_1}) \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_1 - \mathbf{p}_{S_2}) \\ \vdots & \vdots \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_{S_1}) \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_1 - \mathbf{p}_{S_2}) \end{bmatrix}$$
$$= \underbrace{\breve{\mathbf{D}} \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_{S_1}) \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \vdots & \vdots \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_0 - \mathbf{p}_{S_1}) \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \end{bmatrix}}_{\breve{\mathbf{M}}_1} + \underbrace{\breve{\mathbf{D}} \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_1 - \mathbf{p}_{S_2}) \\ \vdots & \vdots \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \mathbf{I}_2 & \mathbf{J}(\mathbf{p}_1 - \mathbf{p}_{S_2}) \end{bmatrix}}_{\breve{\mathbf{M}}_2} \quad (42)$$

where $\breve{\mathbf{M}}_1$ and $\breve{\mathbf{M}}_2$ are the decompositions of the observability matrix $\breve{\mathbf{M}}$, with respect to the first and second radar, respectively. It is not difficult to show that $\mathrm{rank}(\breve{\mathbf{M}}_1) = \mathrm{rank}(\breve{\mathbf{M}}_2) = 2$, which agrees with the preceding result in the single-radar case.

2) Standard EKF: The ideal EKF is not realizable in practice since the true states are not available. Therefore, we now consider the standard EKF which computes the Jacobians using the current state estimates, and show that the preceding results do not hold for the standard EKF linearized system. Specifically, the observability matrix for the two-radar scenario under consideration is given by:

$$\mathbf{M} = \underbrace{\mathbf{Diag}\left(\hat{\boldsymbol{\alpha}}_{1,0}, \hat{\boldsymbol{\alpha}}_{2,1}, \cdots, \hat{\boldsymbol{\alpha}}_{1,k-1}, \hat{\boldsymbol{\alpha}}_{2,k}\right)}_{\hat{\mathbf{D}}} \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{0|0} - \mathbf{p}_{S_1}) \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{1|0} - \mathbf{p}_{S_2}) \\ \vdots & \vdots \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{k-1|k-1} + \sum_{\kappa=1}^{k-2} \boldsymbol{\Delta}\hat{\mathbf{p}}_\kappa - \mathbf{p}_{S_1}) \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{k|k} + \sum_{\kappa=1}^{k-1} \boldsymbol{\Delta}\hat{\mathbf{p}}_\kappa - \mathbf{p}_{S_2}) \end{bmatrix}$$
$$= \underbrace{\hat{\mathbf{D}} \begin{bmatrix} \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{0|0} - \mathbf{p}_{S_1}) \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \vdots & \vdots \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{k-1|k-1} + \sum_{\kappa=1}^{k-2} \boldsymbol{\Delta}\hat{\mathbf{p}}_\kappa - \mathbf{p}_{S_1}) \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \end{bmatrix}}_{\mathbf{M}_1} + \underbrace{\hat{\mathbf{D}} \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{1|0} - \mathbf{p}_{S_2}) \\ \vdots & \vdots \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\ \mathbf{I}_2 & \mathbf{J}(\hat{\mathbf{p}}_{k|k} + \sum_{\kappa=1}^{k-1} \boldsymbol{\Delta}\hat{\mathbf{p}}_\kappa - \mathbf{p}_{S_2}) \end{bmatrix}}_{\mathbf{M}_2} \quad (43)$$

where $\boldsymbol{\Delta}\hat{\mathbf{p}}_\kappa \triangleq \hat{\mathbf{p}}_{\kappa|\kappa} - \hat{\mathbf{p}}_{\kappa|\kappa-1}$ is the correction in the target position due to the EKF update at time step κ, and in general does not vanish. As a result, the decompositions of the observability matrix become full-rank, i.e., $\mathrm{rank}(\mathbf{M}_1) = \mathrm{rank}(\mathbf{M}_2) = 3$. This implies that the standard EKF acquires nonexistent information along one direction of the state space from each radar's measurements, which can lead to inconsistency and confirms our preceding analysis. Next, we apply the algorithms presented in Section II-B to this system so as to improve EKF consistency.

B. Application of the Consistency Algorithms

In particular, we choose the desired nullspace of the decompositions of the observability matrix for the two radars as follows (i.e., using the first state estimates when the two sensors provide their first measurements):

$$\mathbf{N}_1 = \begin{bmatrix} \mathbf{J}(\mathbf{p}_{S_1} - \hat{\mathbf{p}}_{0|0}) \\ 1 \end{bmatrix}, \quad \mathbf{N}_2 = \begin{bmatrix} \mathbf{J}(\mathbf{p}_{S_2} - \hat{\mathbf{p}}_{1|0}) \\ 1 \end{bmatrix} \quad (44)$$

We first describe the *indirect* algorithm which first finds the optimal linearization points for computing Jacobians by solving the problem (27)-(28). In this case, the constraint (28) can be simplified as:

$$\mathbf{H}_{i,k}\boldsymbol{\Phi}_{k-1}\cdots\boldsymbol{\Phi}_0\mathbf{N}_i = \mathbf{0}$$
$$\Leftrightarrow \hat{\boldsymbol{\alpha}}_{i,k}\left[\mathbf{I}_2 \quad \mathbf{J}\left(\mathbf{p}^\star_{0|0} + \textstyle\sum_{\kappa=1}^{k}\boldsymbol{\Delta}\mathbf{p}^\star_\kappa - \mathbf{p}_{S_i}\right)\right]\mathbf{N}_i = \mathbf{0}$$
$$\Leftrightarrow \mathbf{p}^\star_{k|k} = \mathbf{p}^\star_{k|k-1} - \sum_{\kappa=1}^{k}\boldsymbol{\Delta}\mathbf{p}^\star_\kappa \quad (45)$$

where we have employed the definition $\boldsymbol{\Delta}\mathbf{p}^\star_\kappa = \mathbf{p}^\star_{\kappa|\kappa} - \mathbf{p}^\star_{\kappa|\kappa-1}$. Using the method of Lagrangian multipliers, we can now analytically solve for the optimal solution to the problem (27) and (45), which is given by:

$$\mathbf{p}^\star_{k|k} = \mathbf{p}^\star_{k|k-1} = \hat{\mathbf{p}}_{k|k-1}\ ,\ \phi^\star_{k|k} = \hat{\phi}_{k|k}\ ,\ \phi^\star_{k|k-1} = \hat{\phi}_{k|k-1} \quad (46)$$

Alternatively, we can use the *direct* algorithm for computing the EKF Jacobians. Specifically, we directly use the optimal solution of the measurement Jacobian (31), i.e., projecting the best-available measurement Jacobian onto the information-available directions (see Lemma 2.4), while computing the propagation Jacobians in the same way as the standard EKF.

It is important to point out that in the both proposed EKF algorithms, once the filter Jacobians are computed, the state estimates and covariance are propagated and updated in the same way as in the standard EKF. Note also that the proposed

EKFs are causal and realizable in practice, since they do not use any information about the future or the true states.

C. Numerical Results

To demonstrate the ability of the proposed algorithms to improve filter consistency, we conducted 100 Monte-Carlo simulations under various conditions. The metrics used to evaluate filter performance are: (i) the root mean square error (RMSE), and (ii) the average normalized (state) estimation error squared (NEES) [2]. Specifically, we compute the average RMSE and average NEES by averaging over all Monte-Carlo runs for each time step. The RMSE provides us with a concise metric of the accuracy of a given estimator. On the other hand, the NEES is a standard criterion for evaluating filter consistency. Specifically, it is known that the NEES of an N-dimensional Gaussian random variable follows a χ^2 distribution with N degrees of freedom. Therefore, if a certain filter is consistent, we expect that the average NEES will be close to 3 for all time steps. The larger the deviations of the NEES from these values, the worse the inconsistency of the filter. By studying both the RMSE and NEES of all the filters considered here, we obtain a comprehensive picture of the estimators' performance.

In the numerical simulation test presented here, a target (robot) with a simple differential drive model moved on a planar surface, at a constant velocity of $v = 0.25$ m/sec. The two-drive wheels were equipped with encoders, which measure their revolutions and provide measurements of velocity (i.e., right and left wheel velocities, v_r and v_l, respectively), with standard deviation equal to $\sigma = 1\% v$ for each wheel. These measurements were used to obtain the linear and rotational velocity measurements for the target, which are given by $v = \frac{v_r+v_l}{2}$ and $\omega = \frac{v_r-v_l}{a}$, where $a = 0.5$ m is the distance between the active wheels. The standard deviations of the linear and rotational velocity measurement noise were thus equal to $\sigma_v = \frac{\sigma}{\sqrt{2}}$ and $\sigma_\omega = \frac{\sqrt{2}\sigma}{a}$, respectively. Two radars with known positions alternatively provide distance measurements to the target. The standard deviation of the distance-measurement noise was equal to 10% of the radar-to-target distance.

Fig. 2 shows the Monte-Carlo results of average NEES and RMSE for the robot (target) pose. As evident, the proposed EKFs (i.e., the indirect and direct EKFs) perform much better than the standard EKF, and very close to the benchmark, the ideal EKF, in terms of both consistency (NEES) and accuracy (RMSE). This is attributed to the fact that the proposed EKFs acquire information only along the correct directions of the state space from *each* radar's measurements.

IV. Conclusions

In this paper, we have studied the problem of filter inconsistency for nonlinear systems where only partial-state measurements are available. We showed that despite the fact that the system is observable, linearized estimators, such as the EKF, can still become inconsistent. To understand the cause of inconsistency, we examined the FIM of the initial state (by marginalizing all but the initial state) and showed that it is closely related to the observability matrix. Moreover, we proposed a novel decomposition of the observability matrix of the standard linearized system with respect to different sources of measurements, and showed that when using the standard EKF each decomposition of the observability matrix has higher rank than that of the ideal case. This implies that the filter gains spurious information from the measurements of each source, which leads to inconsistency. To address this issue, we proposed to compute the filter Jacobians in such a way that ensures that each decomposition of the observability matrix has nullspace of correct dimensions. We applied the proposed algorithms to the problem of two-radar target tracking, and demonstrated the superior performance of the proposed filters over the standard EKF.

Appendix I
Proof of Lemma 2.1

We prove this result by mathematical induction. Specifically, we start by the base case of $k = 0$, in which since no marginalization is involved and the observability matrix is simply the first measurement Jacobian, i.e., $\mathbf{M} = \mathbf{H}_0$ [see (10)], the information matrix can be directly written in the desired form:

$$\mathbf{A} = \mathbf{A}_0 = \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 =: \mathbf{M}^T \mathbf{\Sigma} \mathbf{M} \tag{47}$$

We now consider the case of $k = 1$, in which the state vector is $\mathbf{x}_{0:1} = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix}$ and hence the full information matrix is given by [see (18)]:

$$\begin{aligned} \mathbf{A} &= \mathcal{F}_0^T \mathbf{Q}_0^{-1} \mathcal{F}_0 + \mathcal{H}_0^T \mathbf{R}_0^{-1} \mathcal{H}_0 + \mathcal{H}_1^T \mathbf{R}_1^{-1} \mathcal{H}_1 \\ &= \begin{bmatrix} \mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \mathbf{\Phi}_0 + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 & -\mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \\ -\mathbf{Q}_0^{-1} \mathbf{\Phi}_0 & \mathbf{Q}_0^{-1} + \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1 \end{bmatrix} \end{aligned} \tag{48}$$

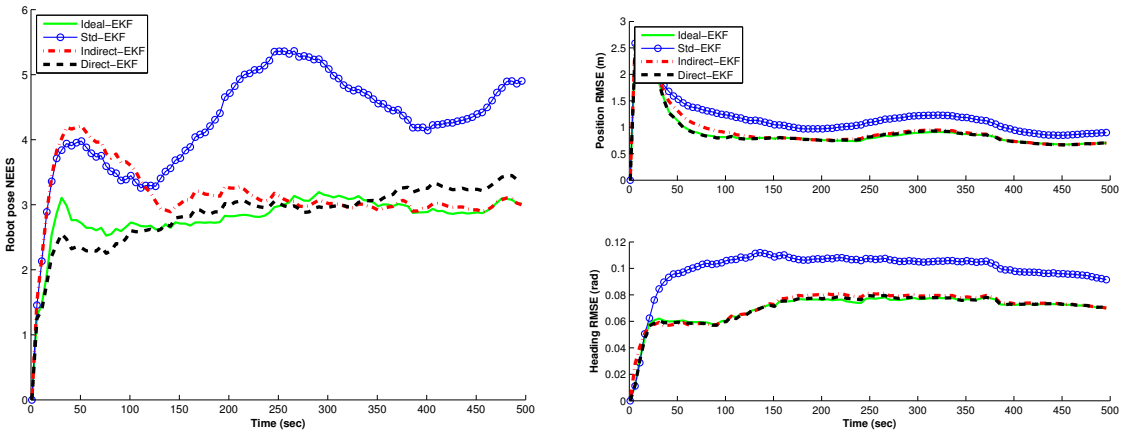
In order to obtain the information matrix of $\mathbf{x}_0$, we marginalize out $\mathbf{x}_1$ by employing the Schur complement:

$$\begin{aligned} \mathbf{A}_0 =& \mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \mathbf{\Phi}_0 + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 - \left(\mathbf{\Phi}_0^T \mathbf{Q}_0^{-1}\right) \underbrace{\left(\mathbf{Q}_0^{-1} + \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1\right)^{-1}}_{\mathbf{\Xi}_1} \left(\mathbf{Q}_0^{-1} \mathbf{\Phi}_0\right) \\ =& \mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \mathbf{\Phi}_0 + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 - \\ & \left(\mathbf{\Phi}_0^T \mathbf{Q}_0^{-1}\right) \left[\mathbf{Q}_0 - \mathbf{Q}_0 \mathbf{H}_1^T (\mathbf{H}_1 \mathbf{Q}_0 \mathbf{H}_1^T + \mathbf{R}_1)^{-1} \mathbf{H}_1 \mathbf{Q}_0\right] \left(\mathbf{Q}_0^{-1} \mathbf{\Phi}_0\right) \\ =& \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 + (\mathbf{H}_1 \mathbf{\Phi}_0)^T (\mathbf{H}_1 \mathbf{Q}_0 \mathbf{H}_1^T + \mathbf{R}_1)^{-1} (\mathbf{H}_1 \mathbf{\Phi}_0) \\ =& \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{\Phi}_0 \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_0^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}_1 \mathbf{Q}_0 \mathbf{H}_1^T + \mathbf{R}_1)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{\Phi}_0 \end{bmatrix} =: \mathbf{M}^T \mathbf{\Sigma} \mathbf{M} \end{aligned} \tag{49}$$

where we have used the Woodbury matrix identity [13] for computing $\mathbf{\Xi}_1$. It becomes clear that in this case the information matrix of $\mathbf{x}_0$ is factorized into the desired form.

To better understand the structure, we also consider the case of $k = 2$, where the full information matrix of the entire state $\mathbf{x}_{0:2}$ is given by [see (18)]:

$$\begin{aligned} \mathbf{A} &= \sum_{\kappa=0}^{1} \mathcal{F}_\kappa^T \mathbf{Q}_\kappa^{-1} \mathcal{F}_\kappa + \sum_{\kappa=0}^{2} \mathcal{H}_\kappa^T \mathbf{R}_\kappa^{-1} \mathcal{H}_\kappa \\ &= \begin{bmatrix} \mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \mathbf{\Phi}_0 + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 & -\mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} & \mathbf{0} \\ -\mathbf{Q}_0^{-1} \mathbf{\Phi}_0 & \mathbf{Q}_0^{-1} + \mathbf{\Phi}_1^T \mathbf{Q}_1^{-1} \mathbf{\Phi}_1 + \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1 & -\mathbf{\Phi}_1^T \mathbf{Q}_1^{-1} \\ \mathbf{0} & -\mathbf{Q}_1^{-1} \mathbf{\Phi}_1 & \mathbf{Q}_1^{-1} + \mathbf{H}_2^T \mathbf{R}_2^{-1} \mathbf{H}_2 \end{bmatrix} \end{aligned} \tag{50}$$



(a) Avg. NEES of target pose

(b) Avg. RMSE of target pose (position and orientation)

Fig. 2. Monte-Carlo results of two-radar tracking. In this simulation, a target moves on a circular trajectory and two radars with known positions alternate between providing distance measurements to the target. Note that in these plots, the RMSE values of the ideal EKF and the two proposed EKFs are very close, which makes the corresponding lines difficult to distinguish.

Similarly, by marginalizing out $\mathbf{x}_1$ and $\mathbf{x}_2$ using the Schur complement, we obtain the information matrix of the initial state $\mathbf{x}_0$:

$$\mathbf{A}_0 = \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\mathbf{\Phi}_0 + \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0 - \begin{bmatrix}\mathbf{Q}_0^{-1}\mathbf{\Phi}_0\\ \mathbf{0}\end{bmatrix}^T \underbrace{\begin{bmatrix}\mathbf{Q}_0^{-1}+\mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\mathbf{\Phi}_1+\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 & -\mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\\ -\mathbf{Q}_1^{-1}\mathbf{\Phi}_1 & \mathbf{Q}_1^{-1}+\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{H}_2\end{bmatrix}^{-1}}_{\mathbf{\Xi}_2} \begin{bmatrix}\mathbf{Q}_0^{-1}\mathbf{\Phi}_0\\ \mathbf{0}\end{bmatrix} \tag{51}$$

It is clear that due to the structure of the above equation, we only need to compute the top-leftmost submatrix $\mathbf{\Xi}_2(1,1)$, obtained by partitioning $\mathbf{\Xi}_2 = \begin{bmatrix}\mathbf{\Xi}_2(1,1) & \mathbf{\Xi}_2(1,2)\\ \mathbf{\Xi}_2(2,1) & \mathbf{\Xi}_2(2,2)\end{bmatrix}$, corresponding to the nonzero block of $\begin{bmatrix}\mathbf{Q}_0^{-1}\mathbf{\Phi}_0\\ \mathbf{0}\end{bmatrix}$. Hence, using the block matrix inversion lemma [13], we have:

$$\begin{aligned}\mathbf{\Xi}_2(1,1) &= \left[\mathbf{Q}_0^{-1}+\mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\mathbf{\Phi}_1+\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1-\mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\left(\mathbf{Q}_1^{-1}+\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{H}_2\right)^{-1}\mathbf{Q}_1^{-1}\mathbf{\Phi}_1^T\right]^{-1}\\ &= \left[\mathbf{Q}_0^{-1}+\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1+\mathbf{\Phi}_1^T\mathbf{H}_2^T(\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T+\mathbf{R}_2)^{-1}\mathbf{H}_2\mathbf{\Phi}_1\right]^{-1}\\ &= \left\{\mathbf{Q}_0^{-1}+\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\mathbf{\Phi}_1\end{bmatrix}^T\begin{bmatrix}\mathbf{R}_1^{-1} & \mathbf{0}\\ \mathbf{0} & (\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T+\mathbf{R}_2)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\mathbf{\Phi}_1\end{bmatrix}\right\}^{-1}\\ &=: \left[\mathbf{Q}_0^{-1}+\mathbf{H}_{1:2}^T\mathbf{R}_{1:2}^{-1}\mathbf{H}_{1:2}\right]^{-1}\\ &= \mathbf{Q}_0-\mathbf{Q}_0\mathbf{H}_{1:2}^T(\mathbf{H}_{1:2}\mathbf{Q}_0\mathbf{H}_{1:2}^T+\mathbf{R}_{1:2})^{-1}\mathbf{H}_{1:2}\mathbf{Q}_0\end{aligned} \tag{52}$$

where we have defined $\mathbf{H}_{1:2} \triangleq \begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\mathbf{\Phi}_1\end{bmatrix}$ and $\mathbf{R}_{1:2} \triangleq \mathrm{Diag}\left(\mathbf{R}_1, \mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T+\mathbf{R}_2\right)$. It is important to notice that, from the first equality of the above equations, $\mathbf{\Xi}_2^{-1}(1,1)$ is the Schur complement of $\mathbf{\Xi}_2^{-1}$ with respect to $\mathbf{x}_1$ (i.e., marginalizing out $\mathbf{x}_2$ from $\mathbf{x}_{1:2}$). In the second equality, we have also employed the Woodbury matrix identity. Substitution of (52) in (51) yields (by noting again that it is not necessary to compute the other submatrices of $\mathbf{\Xi}_2$):

$$\begin{aligned}\mathbf{A}_0 &= \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\mathbf{\Phi}_0+\mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0-\\ &\quad \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\left[\mathbf{Q}_0-\mathbf{Q}_0\mathbf{H}_{1:2}^T(\mathbf{H}_{1:2}\mathbf{Q}_0\mathbf{H}_{1:2}^T+\mathbf{R}_{1:2})^{-1}\mathbf{H}_{1:2}\mathbf{Q}_0\right]\mathbf{Q}_0^{-1}\mathbf{\Phi}_0\\ &= \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0+\mathbf{\Phi}_0^T\mathbf{H}_{1:2}^T(\mathbf{H}_{1:2}\mathbf{Q}_0\mathbf{H}_{1:2}^T+\mathbf{R}_{1:2})^{-1}\mathbf{H}_{1:2}\mathbf{\Phi}_0\\ &= \begin{bmatrix}\mathbf{H}_0\\ \mathbf{H}_{1:2}\mathbf{\Phi}_0\end{bmatrix}^T\begin{bmatrix}\mathbf{R}_0^{-1} & \mathbf{0}\\ \mathbf{0} & (\mathbf{H}_{1:2}\mathbf{Q}_0\mathbf{H}_{1:2}^T+\mathbf{R}_{1:2})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{H}_0\\ \mathbf{H}_{1:2}\mathbf{\Phi}_0\end{bmatrix} =: \mathbf{M}^T\mathbf{\Sigma}\mathbf{M}\end{aligned} \tag{53}$$

It is clear that in this case we can also factorize the information matrix of $\mathbf{x}_0$ into the desired form.

We now consider the general case of $k=\kappa$. Suppose in the case of $k=\kappa-1$, the information matrix of the initial state $\mathbf{x}_0$ can be factorized into the desired form and in particular $\mathbf{\Xi}_{\kappa-1}(1,1)$ whose inversion is the Schur complement of $\mathbf{\Xi}_{\kappa-1}^{-1}$ with respect to $\mathbf{x}_1$ by marginalizing out $\mathbf{x}_{2:\kappa-1}$ from $\mathbf{x}_{1:\kappa-1}$, assumes the following form [see (52)]:

$$\begin{aligned}\mathbf{\Xi}_{\kappa-1}(1,1) &= \left(\mathbf{Q}_0^{-1}+\mathbf{H}_{1:\kappa-1}^T\mathbf{R}_{1:\kappa-1}^{-1}\mathbf{H}_{1:\kappa-1}\right)^{-1}\\ &= \mathbf{Q}_0-\mathbf{Q}_0\mathbf{H}_{1:\kappa-1}^T\left(\mathbf{H}_{1:\kappa-1}\mathbf{Q}_0\mathbf{H}_{1:\kappa-1}^T+\mathbf{R}_{1:\kappa-1}\right)^{-1}\mathbf{R}_{1:\kappa-1}\mathbf{Q}_0\end{aligned} \tag{54}$$

where $\mathbf{H}_{1:\kappa-1}$ and $\mathbf{R}_{1:\kappa-1}$ are defined similarly as in (52):

$$\mathbf{H}_{1:\kappa-1} \triangleq \begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_{\kappa-1}\mathbf{\Phi}_{\kappa-2}\cdots\mathbf{\Phi}_1\end{bmatrix} \tag{55}$$

$$\mathbf{R}_{1:\kappa-1} \triangleq \mathrm{Diag}\left(\mathbf{R}_1,\cdots,\mathbf{H}_{\kappa-1}\mathbf{Q}_{\kappa-1}\mathbf{H}_{\kappa-1}^T+\mathbf{R}_{\kappa-1}\right) \tag{56}$$

Note that by considering the full state $\mathbf{x}_{2:\kappa}$ (instead of $\mathbf{x}_{1:\kappa-1}$) and marginalizing $\mathbf{x}_{3:\kappa}$ from the full state, we will have a similar matrix as (54) whose time index is shifted by one and which is useful for the ensuing derivations.

$$\begin{aligned}\mathbf{\Xi}'_{\kappa}(1,1) &= \left(\mathbf{Q}_1^{-1}+\mathbf{H}_{2:\kappa}^T\mathbf{R}_{2:\kappa}^{-1}\mathbf{H}_{2:\kappa}\right)^{-1}\\ &= \mathbf{Q}_1-\mathbf{Q}_1\mathbf{H}_{2:\kappa}^T\left(\mathbf{H}_{2:\kappa}\mathbf{Q}_1\mathbf{H}_{2:\kappa}^T+\mathbf{R}_{2:\kappa}\right)^{-1}\mathbf{R}_{2:\kappa}\mathbf{Q}_1\end{aligned} \tag{57}$$

Due to the sparse banded structure of the full information matrix (18), marginalization of $\mathbf{x}_{1:\kappa}$ from the full state $\mathbf{x}_{0:\kappa}$

using the Schur complement yields:

$$\mathbf{A}_0 = \mathbf{\Phi}_0^T \mathbf{Q}_0^{-1} \mathbf{\Phi}_0 + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 - \begin{bmatrix} \mathbf{Q}_0^{-1}\mathbf{\Phi}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}^T \mathbf{\Xi}_\kappa \begin{bmatrix} \mathbf{Q}_0^{-1}\mathbf{\Phi}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (58)$$

As evident, to compute $\mathbf{A}_0$, we only need to calculate the submatrix $\mathbf{\Xi}_\kappa(1,1)$ corresponding to $\mathbf{Q}_0^{-1}\mathbf{\Phi}_0$. Note that analogous to (54), $\mathbf{\Xi}_\kappa^{-1}(1,1)$ is the Schur complement of $\mathbf{\Xi}_\kappa^{-1}$ with respect to $\mathbf{x}_1$ by marginalizing out $\mathbf{x}_{2:\kappa}$ from $\mathbf{x}_{1:\kappa}$. Using (18) and (57), we compute $\mathbf{\Xi}_\kappa(1,1)$:

$$\begin{aligned}
\mathbf{\Xi}_\kappa(1,1) &= \left[\mathbf{Q}_0^{-1} + \mathbf{\Phi}_1^T \mathbf{Q}_1^{-1}\mathbf{\Phi}_1 + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 - \mathbf{\Phi}_1^T\mathbf{Q}_1^{-1}\mathbf{\Xi}'_\kappa(1,1)\mathbf{Q}_1^{-1}\mathbf{\Phi}_1\right]^{-1} \\
&= \left[\mathbf{Q}_0^{-1} + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 + \mathbf{\Phi}_1^T\mathbf{H}_{2:\kappa}^T\left(\mathbf{H}_{2:\kappa}\mathbf{Q}_1\mathbf{H}_{2:\kappa}^T + \mathbf{R}_{2:\kappa}\right)^{-1}\mathbf{H}_{2:\kappa}\mathbf{\Phi}_1\right]^{-1} \\
&= \left\{\mathbf{Q}_0^{-1} + \begin{bmatrix}\mathbf{H}_1 \\ \mathbf{H}_{2:\kappa}\mathbf{\Phi}_1\end{bmatrix}^T \begin{bmatrix}\mathbf{R}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{2:\kappa}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{H}_1 \\ \mathbf{H}_{2:\kappa}\mathbf{\Phi}_1\end{bmatrix}\right\}^{-1} \\
&=: \left(\mathbf{Q}_0^{-1} + \mathbf{H}_{1:\kappa}\mathbf{R}_{1:\kappa}^{-1}\mathbf{H}_{1:\kappa}\right)^{-1} \\
&= \mathbf{Q}_0 - \mathbf{Q}_0\mathbf{H}_{1:\kappa}^T\left(\mathbf{H}_{1:\kappa}\mathbf{Q}_0\mathbf{H}_{1:\kappa}^T + \mathbf{R}_{1:\kappa}\right)^{-1}\mathbf{R}_{1:\kappa}\mathbf{Q}_0 \qquad (59)
\end{aligned}$$

Substituting (59) in (58), we obtain the information matrix of $\mathbf{x}_0$ in the desired form:

$$\begin{aligned}
\mathbf{A}_0 &= \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\mathbf{\Phi}_0 + \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0 - \\
&\quad \mathbf{\Phi}_0^T\mathbf{Q}_0^{-1}\left[\mathbf{Q}_0 - \mathbf{Q}_0\mathbf{H}_{1:\kappa}^T\left(\mathbf{H}_{1:\kappa}\mathbf{Q}_0\mathbf{H}_{1:\kappa}^T + \mathbf{R}_{1:\kappa}\right)^{-1}\mathbf{R}_{1:\kappa}\mathbf{Q}_0\right]\mathbf{Q}_0^{-1}\mathbf{\Phi}_0 \\
&= \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0 + \mathbf{\Phi}_0^T\mathbf{H}_{1:\kappa}^T\left(\mathbf{H}_{1:\kappa}\mathbf{Q}_0\mathbf{H}_{1:\kappa}^T + \mathbf{R}_{1:\kappa}\right)^{-1}\mathbf{H}_{1:\kappa}\mathbf{\Phi}_0 \\
&= \begin{bmatrix}\mathbf{H}_0 \\ \mathbf{H}_{1:\kappa}\mathbf{\Phi}_0\end{bmatrix}^T \begin{bmatrix}\mathbf{R}_0^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{H}_{1:\kappa}\mathbf{Q}_0\mathbf{H}_{1:\kappa}^T + \mathbf{R}_{1:\kappa}\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{H}_0 \\ \mathbf{H}_{1:\kappa}\mathbf{\Phi}_0\end{bmatrix} \\
&=: \mathbf{M}^T\mathbf{\Sigma}\mathbf{M} \qquad (60)
\end{aligned}$$

This completes the proof.

Appendix II
Proof of Lemma 2.3

Under the Gaussianity assumption, $p(\mathbf{x}_k|\mathbf{z}_{0:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$ and $p(\mathbf{x}_{k+1}|\mathbf{z}_{0:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k})$, the first term of the cost function (25) is computed as:

$$\begin{aligned}
&\int \left\|\mathbf{x}_k - \mathbf{x}_{k|k}^\star\right\|^2 p(\mathbf{x}_k|\mathbf{z}_{0:k})d\mathbf{x}_k \\
&= \int \left(\mathbf{x}_k^T\mathbf{x}_k - 2\mathbf{x}_k^T\mathbf{x}_{k|k}^\star + \mathbf{x}_{k|k}^{\star T}\mathbf{x}_{k|k}^\star\right)p(\mathbf{x}_k|\mathbf{z}_{0:k})d\mathbf{x}_k \\
&= \mathbb{E}\left(\mathbf{x}_k^T\mathbf{x}_k\right) - 2\mathbb{E}\left(\mathbf{x}_k^T\right)\mathbf{x}_{k|k}^\star + \mathbf{x}_{k|k}^{\star T}\mathbf{x}_{k|k}^\star \\
&= \mathrm{tr}\left(\mathbf{P}_{k|k} + \hat{\mathbf{x}}_{k|k}\hat{\mathbf{x}}_{k|k}^T\right) - 2\hat{\mathbf{x}}_{k|k}^T\mathbf{x}_{k|k}^\star + \mathbf{x}_{k|k}^{\star T}\mathbf{x}_{k|k}^\star \\
&= \mathrm{tr}\left(\mathbf{P}_{k|k}\right) + \left\|\hat{\mathbf{x}}_{k|k} - \mathbf{x}_{k|k}^\star\right\|^2 \qquad (61)
\end{aligned}$$

where $\mathbb{E}(\cdot)$ denotes expectation and $\mathrm{tr}(\cdot)$ the matrix trace. Proceeding similarly, we can show that the second term of the cost function (25) is equivalent to:

$$\begin{aligned}
&\int ||\mathbf{x}_{k+1} - \mathbf{x}_{k+1|k}^\star||^2 p(\mathbf{x}_{k+1}|z_{0:k})d\mathbf{x}_{k+1} \\
&= \mathrm{tr}\left(\mathbf{P}_{k+1|k}\right) + \left\|\hat{\mathbf{x}}_{k+1|k} - \mathbf{x}_{k+1|k}^\star\right\|^2 \qquad (62)
\end{aligned}$$

Using the fact that the true $\mathbf{P}_{k|k}$ and $\mathbf{P}_{k+1|k}$ are independent of the linearization points, the following is immediate:

$$\begin{aligned}
&\min_{\mathbf{x}_{k|k}^\star, \mathbf{x}_{k+1|k}^\star} \mathrm{tr}\left(\mathbf{P}_{k|k}\right) + \left\|\hat{\mathbf{x}}_{k|k} - \mathbf{x}_{k|k}^\star\right\|^2 + \mathrm{tr}\left(\mathbf{P}_{k+1|k}\right) + \left\|\hat{\mathbf{x}}_{k+1|k} - \mathbf{x}_{k+1|k}^\star\right\|^2 \\
\Leftrightarrow &\min_{\mathbf{x}_{k|k}^\star, \mathbf{x}_{k+1|k}^\star} \left\|\hat{\mathbf{x}}_{k|k} - \mathbf{x}_{k|k}^\star\right\|^2 + \left\|\hat{\mathbf{x}}_{k+1|k} - \mathbf{x}_{k+1|k}^\star\right\|^2
\end{aligned}$$

This completes the proof.

Appendix III
Proof of Lemma 2.4

The constraint (30) states that the rows of $\mathbf{H}_{i,k}$ lie in the left nullspace of the matrix $\mathbf{U}_i$. Therefore, if $\mathbf{L}$ is a matrix whose rows span this nullspace, then $\mathbf{H}_{i,k}$ can be written as:

$$\mathbf{H}_{i,k} = \mathbf{\Theta}\mathbf{L} \qquad (63)$$

where $\mathbf{\Theta}$ is the unknown matrix we seek to find. We note that there are several possible ways of computing an appropriate matrix $\mathbf{L}$, whose rows lie in the nullspace of $\mathbf{U}_i$. For instance, such a matrix is given, in closed form, by the expression:

$$\mathbf{L} = \begin{bmatrix}\mathbf{I}_m & \mathbf{0}_{m\times(n-m)}\end{bmatrix}\left(\mathbf{I}_n - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T\right) =: \mathbf{\Gamma}\mathbf{\Pi} \quad (64)$$

It is not difficult to see that $\mathbf{\Pi} := \mathbf{I}_n - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T$ is an orthogonal projection matrix (i.e., $\mathbf{\Pi}^2 = \mathbf{\Pi}$ and $\mathbf{\Pi}^T = \mathbf{\Pi}$) and hence has the eigenvalues of either 1 or 0, whose reduced SVD is given by $\mathbf{\Pi} = \mathbf{Q}\mathbf{Q}^T$. Using this result, $\mathbf{L}^T$ immediately can be written as $\mathbf{L}^T = \mathbf{Q}\mathbf{Q}^T\mathbf{\Gamma}^T$. By substituting this identity into the cost function, we have:

$$\begin{aligned}
&\min \ ||\mathbf{H}_o - \mathbf{H}_{i,k}||_F^2 = ||\mathbf{Q}^T\mathbf{\Gamma}^T\mathbf{\Theta}^T - \mathbf{Q}^T{\mathbf{H}_o}^T||_F^2 \\
&\Rightarrow \quad \mathbf{\Theta} = \mathbf{H}_o\mathbf{Q}\left(\mathbf{\Gamma}\mathbf{Q}\right)^{-1} \qquad (65)
\end{aligned}$$

Therefore, substitution of the above equation in (63) yields:

$$\begin{aligned}
\mathbf{H}_{i,k} &= \mathbf{H}_o\mathbf{Q}\left(\mathbf{\Gamma}\mathbf{Q}\right)^{-1}\mathbf{\Gamma}\mathbf{Q}\mathbf{Q}^T = \mathbf{H}_o\mathbf{Q}\mathbf{Q}^T = \mathbf{H}_o\mathbf{\Pi} \\
&= \mathbf{H}_o\left(\mathbf{I}_n - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T\right) \qquad (66)
\end{aligned}$$

This completes the proof.

References

[1] S. S. Blackman and R. F. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001.

[3] S. Julier and J. K. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Proc. of the IEEE International Conference on Robotics and Automation*, Seoul, Korea, May 21–26, 2001, pp. 4238–4243.

[4] J. A. Castellanos, J. Neira, and J. Tardos, "Limits to the consistency of EKF-based SLAM," in *Proc. of the 5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisbon, Portugal, Jul. 5–7, 2004, pp. 1244–1249.

[5] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, Oct. 9–15, 2006, pp. 3562–3568.

[6] S. Huang and G. Dissanayake, "Convergence and consistency analysis for extended Kalman filter based SLAM," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 1036–1049, Oct. 2007.

[7] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.

[8] G. P. Huang, N. Trawny, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based consistent EKF estimators for multi-robot cooperative localization," *Autonomous Robots*, vol. 30, no. 1, pp. 99–122, Jan. 2011.

[9] Z. Chen, K. Jiang, and J. Hung, "Local observability matrix and its application to observability analyses," in *Proc. of the 16th Annual Conference of IEEE*, Pacific Grove, CA, Nov. 27–30, 1990, pp. 100–103.

[10] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, ser. Mathematics in Science and Engineering. London: Academic Press, 1979, vol. 141-1.

[11] S. Kay, *Fundamentals of Statistical Signal Processing, Vol. I - Estimation Theory*. Prentice Hall, 1993.

[12] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.

[13] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.