IDENTIFICATION OF MULTI-OBJECT DYNAMICAL SYSTEMS: CONSISTENCY AND FISHER INFORMATION

JEREMIE HOUSSINEAU*, SUMEETPAL S. SINGH†, AND AJAY JASRA‡

Abstract. Learning the model parameters of a multi-object dynamical system from partial and perturbed observations is a challenging task. Despite recent numerical advancements in learning these parameters, theoretical guarantees are extremely scarce. In this article we aim to help fill this gap and study the identifiability of the model parameters and the consistency of the corresponding maximum likelihood estimate (MLE) under assumptions on the different components of the underlying multi-object system. In order to understand the impact of the various sources of observation noise on the ability to learn the model parameters, we study the asymptotic variance of the MLE through the associated Fisher information matrix. For example, we show that specific aspects of the multi-target tracking (MTT) problem such as detection failures and unknown data association lead to a loss of information which is quantified in special cases of interest. To the best of the authors' knowledge, these are new theoretically-backed insights on the subtleties of MTT parameter learning.

Key words. Identifiability, Consistency, Fisher Information

AMS subject classifications. 62F12, 62B10

1. Introduction.

A multi-object dynamical system is comprised of an unknown and randomly varying number of objects, each of which is a partially observed Markov process. Multi-target tracking refers to the problem of estimating the state of each of these objects from noisy observations that are also corrupted by detection failures and false detections (a.k.a. false alarms). This type of problem arises in many different fields such as Systems Biology [2], Robotics [15], Computer Vision [18] or Surveillance [20]. Different formulations of multi-target tracking exist, including extensions of the single-target approach to multiple targets [1] as well as formulations based on simple point processes [13].

One of the main challenges in multi-target tracking is the uncertainty in the *data association*, which refers to the problem of finding the right pairing between targets and recorded observations over time, a task further confounded by the corruption of these observations with false positives and detection failures. Inferentially, multi-target tracking is notoriously difficult to solve as it involves an exponentially growing numbers of possible configurations for the data association. Over the past decade there has been significant advancements towards more practical solutions to this inference problem. Some of these include solutions based on sequential Monte Carlo (SMC) [26], hierarchical SMC [19] or Gaussian mixtures [25].

In this article, both the MTT observation model and the motion model of the constituent individual targets are assumed unknown and are instead parameterised and to be inferred from the data. Although multi-target tracking has been an active research field for decades, questions concerning the identifiability and the consistency of the corresponding model parameter estimates have not received the appropriate attention. In this paper we aim to address this gap and shed some light on this issue. Building on results from the literature on Markov processes (e.g. see [12, 6]), we prove both identifiability and the consistency of the MLE of the MTT model parameters

*Department of Statistics, University of Warwick. Email: jeremie.houssineau@warwick.ac.uk

†Department of Engineering, University of Cambridge and The Alan Turing Institute. Email: sss40@cam.ac.uk

‡Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, KSA. Email: ajay.jasra@kaust.edu.sa

in Theorem 3. Specifically, as each constituent target of the MTT model is a partially observed Markov process, in Theorem 2 we show that identifiability transfers from single to multiple targets under appropriate assumptions. The practical implications of results regarding identifiability include the understanding of the behaviour of Markov chain Monte Carlo (MCMC) techniques in multi-target tracking [17, 10], which is conditioned by the likelihood ratio between the correct parameter value and all the other possible values. The consistency of the maximum likelihood estimator raises the question of its asymptotic normality and the corresponding variance, which in turns motivates the study of the Fisher information matrix (FIM) for this class of problems. It is demonstrated in Theorem 4 that there is a strict loss of information in the presence of data association uncertainty or detection failures. We characterise the Fisher information more precisely in specific illustrated cases, e.g. we show that when increasing the number of targets there is no gain in the Fisher information for the model parameters which are common to all targets if large uncertainties on the origin of the corresponding observations persist (see subsection 5.3). The FIM is useful in applications such as sensor management [7] which aims at optimising the position of the sensor or at finding the best ratio between probability of false alarm and probability of detection.

The MLE and the FIM have been used in different ways in the multi-target tracking literature. For instance, [8] suggests different expectation-maximization algorithms based on the Fisher information for estimating the states of the targets in problems with a known number of targets. Also, the analysis of the Cramér-Rao lower bound (CRLB) proposed in [9], which is an extension of the approach proposed in [24] for multiple targets, brings insight on the evolution of the information on the target states in time under various assumptions on the observation process. A crucial difference between [8, 9, 24] and our paper is that the Fisher information is taken with respect to the targets' state in [8, 9, 24] whereas in this paper the FIM pertains to the estimation of the multi-target model parameters. More recently, MLE has become one of the main techniques for calibrating hidden Markov models, as presented in [11, 21]. These works show that recursive state estimation and maximum likelihood estimation of the model parameters can be performed simultaneously using particle filtering with remarkable accuracy. The application of these ideas to MTT was pioneered in the articles [22] and [27], which provide one of the main motivation for seeking some theoretical justifications for this type of approach in the context of MTT model estimation.

The proof of identifiability of the MTT model as well as our approach for studying the asymptotic variance of the MLE for the MTT model parameters are original and, to the best of our knowledge, the first of their kind. Consistency of the data association problem in MTT has been studied in [23] in the context of the estimation of multiple splitting and merging targets observed without noise over a fixed time interval during which n observations of the multiple targets are made at discrete times. The result in [23] is limited to the case where the number n of observation tends to infinity which effectively amounts to saying that targets are observed infinitely many times over a fixed interval which is a scenario not typically encountered in practice. In any case, our theoretical results and proof techniques are entirely different as they pertain to the MTT model parameters and not the data association. Point-process-based theoretical studies of MTT have also been conducted in [4, 5] for the stability of specific inference methods.

The structure of the article is as follows: after introducing the required notations and background concepts in section 2 and section 3, the consistency of the maximum likelihood estimator is established along with its asymptotic normality for a large class of multi-object systems in section 4. Finally, in order to better understand the effect of the various parameters on the asymptotic variance, the FIM is computed for important special cases of multi-object systems in section 5. The article concludes in section 6.

2. Notations. All random variables will be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the expectation of a random variable X w.r.t. the probability measure $\mathbb{P}$ is denoted $\mathbb{E}[X]$. Probability densities will be denoted by lower-case letters while probability measures will be denoted by capital letter. Similarly, random variables will be denoted with capital letters whereas their realisations will be in lower-case.

The time is indexed by the set $\mathbb{N}$ of positive integers and for every time $t \in \mathbb{N}$, a finite sequence $\boldsymbol{y}_t$ of $M_t \in \mathbb{N}_0 \doteq \mathbb{N} \cup \{0\}$ observation points in the observation space $\mathbb{Y}$ is made available. This space can be assumed to be a subset of the Euclidean space $\mathbb{R}^d$ with $d > 0$. The sequences of observations of the form $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ will be denoted $\boldsymbol{y}_{1:n}$. In the standard formulation of multi-target tracking, no more than one observation is associated with a given object at a given time step and, conversely, observations are originated from one object only.

Objects' states are modelled as elements of a set $\mathbb{X}$ which is assumed to be a subset of the Euclidean space $\mathbb{R}^{d'}$ with $d' > 0$; usually satisfying $d' \geq d$. They are propagated independently according to a Markov kernel density f_θ from the state space $\mathbb{X}$ to itself, which depends on a parameter θ from a compact set Θ. Densities on $\mathbb{X}$ are defined w.r.t. a reference measure μ. The true value of the parameter θ is denoted θ^*. The random variable X_t describing the state at time t only depends on the state x_{t-1} at time $t-1$, i.e. $X_t \sim f_\theta(\cdot \,|\, x_{t-1})$. This transition does not depend on time so that the associated Markov chain is said to be *homogeneous*. The observation process at time t given the state x_t is modelled by $Y_t \sim g_\theta(\cdot \,|\, x_t)$, where g_θ is a likelihood function from $\mathbb{X}$ to $\mathbb{Y}$, also parametrised by θ, so that the observation Y_t at time t is independent from the states and observations at other times. The process $(X_t, Y_t)_{t \geq 1}$ is usually referred to as a hidden Markov model (HMM). Its law under the parameter $\theta \in \Theta$ is denoted $\bar{P}_\theta$ when initialised with its stationary distribution assuming it exists, and $P_\theta(\cdot \,|\, x_0)$ when initialised at $x_0 \in \mathbb{X}$.

3. Background. The definition of specific properties of Markov chains that will be used in the following sections is given here for completeness. Let $(X_t)_{t \geq 0}$ be a $\mathbb{X}$-valued Markov chain with transition density f and let $P(\cdot \,|\, x)$ be the probability measure on $(\mathbb{X}^{\mathbb{N}_0}, \mathcal{X}^{\otimes \mathbb{N}_0})$, where $\mathcal{X}^{\otimes \mathbb{N}_0}$ is the cylinder σ-algebra on $\mathbb{X}^{\mathbb{N}_0}$, characterising the chain when initialised at point $x \in \mathbb{X}$. Also, let τ_A be the hitting time to a set $A \subseteq \mathbb{X}$ defined as $\tau_A = \inf\{t \geq 1 : X_t \in A\}$.

Consider the following concepts: A set $A \subseteq \mathbb{X}$ is said to be *accessible* if $\tau_A < \infty$ has positive probability under $P(\cdot \,|\, x)$ for all $x \in \mathbb{X}$. The Markov chain $(X_t)_{t \geq 0}$ is said to be *phi-irreducible* if there exists a density ϕ on $\mathbb{X}$ such that for any subset $A \subseteq \mathbb{X}$, $\int_A \phi(x) \mathrm{d}x > 0$ implies that A is accessible. A set $A \subseteq \mathbb{X}$ is said to be *Harris recurrent* if the event $\tau_A < \infty$ happens almost surely (a.s.) under $P(\cdot \,|\, x)$ for all $x \in \mathbb{X}$. A phi-irreducible Markov chain is said to be Harris recurrent if any accessible set is Harris recurrent. A density q is called *invariant* if it holds that $q(x) = \int f(x \,|\, x') q(x') \mathrm{d}x'$ for all $x \in \mathbb{X}$. A phi-irreducible Markov chain is called *positive* if it admits an invariant probability density function (p.d.f.). More details about these notions expressed in a measure-theoretic formulation can be found in [14]. These concepts will be useful when considering the long-time behaviour of the Markov chains involved in multi-target tracking problems.

4. Consistency of the maximum likelihood estimator.

4.1. The multi-target tracking model. In order to bring the target number within the scope of parameter estimation, the true number of objects in the considered system will be assumed to be fixed and will be denoted by $K^* \in \mathbb{N}$. We consider a Markov chain $(\boldsymbol{X}_t)_{t\geq 0}$ in $\mathbb{X}^{K^*}$ with components independently evolving via the Markov transition f_θ from $\mathbb{X}$ to $\mathbb{X}$. Observations at time t are gathered into a vector $\boldsymbol{y}_t$ in the space $\mathbb{Y}^\times \doteq \bigcup_{k\geq 0} \mathbb{Y}^k$ where $\mathbb{Y}^0$ is a notation for the set containing the empty sequence only. The observation $\boldsymbol{y}_t$ is a superposition of

1. the independent observation of components of $\boldsymbol{X}_t$ via the likelihood g_θ from $\mathbb{X}$ to $\mathbb{Y}$ followed by a Bernoulli thinning with parameter p_{D} corresponding to detection failure, and

2. false alarms, or *clutter*, generated independently of the object-originated observations and assumed to come from an i.i.d. process whose cardinality at each time is Poisson with parameter λ and common distribution P_ψ which depends on the parameter ψ in a compact set Ψ and which true value is denoted ψ^*.

The number of objects K^* is not assumed to be known so that it will also be considered as a parameter of the model. The parameter for the multi-target model is then defined as $\boldsymbol{\theta} \doteq [\theta, K, p_{\mathrm{D}}, \lambda, \psi]^{\mathrm{t}} \in \boldsymbol{\Theta} \doteq \Theta \times S^{\mathrm{T}} \times (0,1) \times S^{\mathrm{C}} \times \Psi$ where t is the vector transposition and where S^{T} and S^{C} are compact subsets of $\mathbb{N}$ and $(0,\infty)$ respectively, with "T" and "C" standing for target and clutter respectively. The true parameter $\boldsymbol{\theta}^*$ is assumed to be an interior point of $\boldsymbol{\Theta}$. Special parameter sets that are not subsets of $\boldsymbol{\Theta}$ can also be introduced by fixing one or several parameters to special values, for instance $\boldsymbol{\Theta}_{\lambda=0} \doteq \Theta \times S^{\mathrm{T}} \times (0,1)$, $\boldsymbol{\Theta}_{p_{\mathrm{D}}=1} \doteq \Theta \times S^{\mathrm{T}} \times S^{\mathrm{C}} \times \Psi$ or $\boldsymbol{\Theta}_{\lambda=0,p_{\mathrm{D}}=1} \doteq \Theta \times S^{\mathrm{T}}$ correspond respectively to cases where the parameters λ, p_{D} or both have known values that are outside of their domain of definition in $\boldsymbol{\Theta}$. Alternatively, if the value of a parameter is known but inside of its domain of definition, e.g. it is known that $K = 1$, then the corresponding hyperplane will be expressed as $\boldsymbol{\Theta}|_{K=1}$. Although the Poisson distribution is not defined for the parameter $\lambda = 0$, this parameter value is simply assumed to represent the case where there is no false alarm.

The Markov transition $\boldsymbol{f_\theta}$ associated with the K-target process $(\boldsymbol{X}_t)_t$ can simply be expressed as $\boldsymbol{f_\theta}(\boldsymbol{x} \,|\, \boldsymbol{x}') = \prod_{i=1}^K f_\theta(\boldsymbol{x}_i \,|\, \boldsymbol{x}'_i)$, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K$, the likelihood however takes a more sophisticated form so that additional notations are required. Let $\mathrm{Sym}(k)$ be the symmetric group over k letters and u_k be the uniform distribution over $\mathrm{Sym}(k)$, also let $\boldsymbol{q_\theta}$ be the distribution on $\{0,1\}^K$ such that $\boldsymbol{q_\theta}(\boldsymbol{d}) \doteq p_{\mathrm{D}}^{|\boldsymbol{d}|}(1-p_{\mathrm{D}})^{K-|\boldsymbol{d}|}$ for any $\boldsymbol{d} \in \{0,1\}^K$, where $|\boldsymbol{d}|$ is the 1-norm of $\boldsymbol{d}$, i.e. the number of detected targets. The variable $\boldsymbol{d}$ is such that $\boldsymbol{d}_i = 1$ if and only if target i is detected for any $i \in \{1,\dots,K\}$. The K-target likelihood $\boldsymbol{g_\theta}(\boldsymbol{y}_t \,|\, \boldsymbol{x}_t)$ of the observations $\boldsymbol{y}_t \in \mathbb{Y}^\times$ at time t given the state $\boldsymbol{x} \in \mathbb{X}^K$ is characterised by

$$
(1) \quad \boldsymbol{g_\theta}(\boldsymbol{y}_t \,|\, \boldsymbol{x}) \doteq
$$

$$
\sum_{\substack{\boldsymbol{d}\in\{0,1\}^K \\ |\boldsymbol{d}|\leq M_t}} \bigg[\mathrm{Po}_\lambda(M_t - |\boldsymbol{d}|) \sum_{\sigma\in\mathrm{Sym}(M_t)} \prod_{i=|\boldsymbol{d}|+1}^{M_t} p_\psi\big(\boldsymbol{y}_{t,\sigma(i)}\big) \prod_{i=1}^{|\boldsymbol{d}|} g_\theta\big(\boldsymbol{y}_{t,\sigma(i)} \,|\, \boldsymbol{x}_{r(i)}\big) u_{M_t}(\sigma) \boldsymbol{q_\theta}(\boldsymbol{d}) \bigg]
$$

where Po_λ denotes the Poisson distribution with parameter λ and where $r(i)$ is the index of the i^{th} detected target that is the smallest integer verifying $|\boldsymbol{d}_{1:r(i)}| = i$, or more formally $r(i) = \min\big\{k : |\boldsymbol{d}_{1:k}| = i\big\}$. This choice of the likelihood $\boldsymbol{g_\theta}$ corresponds to a marginalisation over the observation-to-track data association. Note that $|\boldsymbol{d}| \leq K$ for any $\boldsymbol{d} \in \{0,1\}^K$ so that $\boldsymbol{g_\theta}(\boldsymbol{y} \,|\, \boldsymbol{x}) = 0$ for any $\boldsymbol{x} \in \mathbb{X}^K$ if $\lambda = 0$ and if the number of observations in $\boldsymbol{y}$, denoted $\#\boldsymbol{y}$, is strictly greater than K. The law of the joint Markov chain $(\boldsymbol{X}_t, \boldsymbol{Y}_t)_t$ under the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is denoted $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}}$ when initialised by the stationary distribution and $\boldsymbol{P_\theta}(\cdot \,|\, \boldsymbol{x}_0)$ when assumed to start at the state $\boldsymbol{x}_0 \in \mathbb{X}^K$. The corresponding densities are written accordingly with lower-case letters.

The objective is to study the ratio $\boldsymbol{p_\theta}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)/\boldsymbol{p}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}'_0)$ for any $\boldsymbol{x}_0 \in \mathbb{X}^K$ and any $\boldsymbol{x}'_0 \in \mathbb{X}^{K^*}$. The assumptions that are considered for this purpose are detailed in the next section.

4.2. Assumptions and transferability. In order to bring a better understanding of multi-object systems as a combination of single-object systems corrupted by clutter, assumptions are primarily made on individuals systems. The properties of multi-object systems will be deduced from these whenever this is possible.

A.1 The constants $\tau_- = \inf_\theta \inf_{(x,x')} f_\theta(x \,|\, x')$ and $\tau_+ = \sup_\theta \sup_{(x,x')} f_\theta(x \,|\, x')$ satisfy $\tau_- > 0$ and $\tau_+ < \infty$.

The condition on τ_- in Assumption **A.1** ensures that any point of the state space can be reached from any other point in a single time step (otherwise $f_\theta(x \,|\, x') = 0$ would hold for at least one pair $(x, x') \in \mathbb{X}^2$) while the condition on τ_+ ensures the transition is sufficiently regular when compared to the reference measure μ, i.e. the transition should be diffuse (in the sense that there should be no concentration of probability mass on a single point of the state space). Under Assumption **A.1** it also holds that

$$\tau_-^K \leq \boldsymbol{f_\theta}(\boldsymbol{x} \,|\, \boldsymbol{x}') \leq \tau_+^K \tag{2}$$

for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K$, so that $\boldsymbol{f_\theta}$ straightforwardly satisfies the same type of conditions as f_θ, since S^{T} is compact and hence K is finite.

Let Π_θ be the transition kernel of the joint Markov chain $(X_t, Y_t)_t$ on $\mathbb{X} \times \mathbb{Y}$ defined as $\Pi_\theta(x, y \,|\, x', y') = g_\theta(y \,|\, x) f_\theta(x \,|\, x')$. The property (2) is sufficient to ensure that the joint kernel defined as $\boldsymbol{\Pi_\theta}(\boldsymbol{x}, \boldsymbol{y} \,|\, \boldsymbol{x}', \boldsymbol{y}') = \boldsymbol{g_\theta}(\boldsymbol{y} \,|\, \boldsymbol{x}) \boldsymbol{f_\theta}(\boldsymbol{x} \,|\, \boldsymbol{x}')$, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K$ and any $\boldsymbol{y}, \boldsymbol{y}' \in \mathbb{Y}^\times$ is positive Harris-recurrent and aperiodic.

In the next assumption, the expectations $\bar{\mathbb{E}}_{\theta^*}[\cdot]$, $\mathbb{E}_{\psi^*}[\cdot]$ and $\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}[\cdot]$ are taken with respect to $\bar{P}_{\theta^*}$, P_{ψ^*} and $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$ respectively, also Bi_p^k denotes the binomial distribution with success probability p and k trials.

A.2 The constant $\hat{b}_+^{\mathrm{T}} \doteq \sup_{(\theta,x,y)} g_\theta(y \,|\, x)$ satisfies $\hat{b}_+^{\mathrm{T}} < \infty$, the target- and clutter-related functions

$$b_-^{\mathrm{T}}(y) = \inf_\theta \int g_\theta(y \,|\, x) \mathrm{d}x \qquad \text{and} \qquad b_+^{\mathrm{T}}(y) = \sup_\theta \int g_\theta(y \,|\, x) \mathrm{d}x,$$
$$b_-^{\mathrm{C}}(y) = \inf_\psi p_\psi(y) \qquad \text{and} \qquad b_+^{\mathrm{C}}(y) = \sup_\psi p_\psi(y),$$

satisfy $b_-^{\mathrm{T}}(y) > 0$, $b_+^{\mathrm{T}}(y) < \infty$, $b_-^{\mathrm{C}}(y) > 0$ and $b_+^{\mathrm{C}}(y) < \infty$ for any $y \in \mathbb{Y}$ as well as

$$\bar{\mathbb{E}}_{\theta^*}[|\log b_-^{\mathrm{T}}(Y)|] < \infty \qquad \text{and} \qquad \mathbb{E}_{\psi^*}[|\log b_-^{\mathrm{C}}(Y)|] < \infty, \tag{4}$$

and it holds that

$$\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\big[\big|\log \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathrm{Bi}_{p_{\mathrm{D}}}^K * \mathrm{Po}_\lambda(\#\boldsymbol{Y})\big|\big] < \infty. \tag{5}$$

Assumption **A.2** ensures that all points of the observation space $\mathbb{Y}$ can be reached from at least some states in $\mathbb{X}$ although $g_\theta(y \,|\, x) = 0$ might hold for some $(x, y) \in \mathbb{X} \times \mathbb{Y}$.

Equation (4) will ensure boundedness in the calculations related to identifiability. The supremum $\hat{b}_+^{\mathrm{T}}$ of the likelihood function is also assumed to be finite so that no concentration of probability mass is allowed at any point of $\mathbb{X} \times \mathbb{Y}$. It is demonstrated in the following lemma that the upper and lower bounds considered in Assumption **A.2** for a single target and for the clutter common distribution are sufficient to guarantee the same type of result for multiple targets. The proof is in Appendix B.

LEMMA 1 (Transfer of boundedness). *Under Assumption **A.2**, it holds that the constant* $\hat{\boldsymbol{b}}_+ \doteq \sup_{(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})} \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y} \,|\, \boldsymbol{x})$ *is finite and that* $\boldsymbol{b}_-(\boldsymbol{y}) = \inf_{\boldsymbol{\theta}} \int \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ *and* $\boldsymbol{b}_+(\boldsymbol{y}) \sup_{\boldsymbol{\theta}} \int \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$*, verify* $\boldsymbol{b}_-(\boldsymbol{y}) > 0$ *and* $\boldsymbol{b}_+(\boldsymbol{y}) < \infty$ *for any* $\boldsymbol{y} \in \mathbb{Y}^\times$ *as well as* $\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}[|\log \boldsymbol{b}_-(\boldsymbol{Y})|] < \infty$.

An important result that follows from the assumptions introduced so far is the uniform forgetting of the conditional Markov chain: it can be proved under Assumptions **A.1** and **A.2** that for any $k, l \in \mathbb{N}_0$ such that $k \leq l$ and any parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, it holds that

$$\int \left| \int \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t \,|\, \boldsymbol{x}_k, \boldsymbol{y}_{k+1:l}) \boldsymbol{p}(\boldsymbol{x}_k) \mathrm{d}\boldsymbol{x}_k - \int \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t \,|\, \boldsymbol{x}_k, \boldsymbol{y}_{k+1:l}) \boldsymbol{p}'(\boldsymbol{x}_k) \mathrm{d}\boldsymbol{x}_k \right| \mathrm{d}\boldsymbol{x}_t \leq \boldsymbol{\rho}_{\boldsymbol{\theta}}^{t-k},$$

for all $t \geq k$, all probability densities $\boldsymbol{p}$, $\boldsymbol{p}'$ on $\mathbb{X}^K$ and all sequences of observations $\boldsymbol{y}_{k+1:l}$, where $\boldsymbol{\rho}_{\boldsymbol{\theta}} \doteq 1 - (\tau_-/\tau_+)^K$. The K-target forgetting rate $\boldsymbol{\rho}_{\boldsymbol{\theta}}$ will generally be smaller than the single-target rate $1 - \tau_-/\tau_+$, although mixing is still guaranteed since K is finite and hence $\boldsymbol{\rho}_{\boldsymbol{\theta}} \in [0, 1)$. It is also possible to conclude about the pointwise convergence of the log-likelihood function to the function $\ell : \boldsymbol{\theta} \in \Theta \mapsto \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\left[\ell_{\boldsymbol{Y}_{-\infty:0}}(\boldsymbol{\theta})\right]$, where $\ell_{\boldsymbol{y}_{-\infty:0}}$ is defined on $\boldsymbol{\Theta}$ for any realisation $\boldsymbol{y}_{-\infty:0}$ of the observation process as $\ell_{\boldsymbol{y}_{-\infty:0}} : \boldsymbol{\theta} \mapsto \lim_{m\to\infty} \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_0 \,|\, \boldsymbol{y}_{-m:-1}, \boldsymbol{x}_{-m-1})$, and this limit does not depend on $\boldsymbol{x}_{-m-1}$. Indeed, under Assumptions **A.1** and **A.2**, it holds for all $K \in S^{\mathrm{T}}$ and all $\boldsymbol{x}_0 \in \mathbb{X}^K$ that

$$\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{Y}_{1:n} \,|\, \boldsymbol{x}_0) = \ell(\boldsymbol{\theta}), \qquad \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}\text{-a.s.} \tag{6}$$

This result shows that for any realisation $\boldsymbol{y}_{1:\infty}$ of the observation process, the empirical average $n^{-1} \log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)$ will converge to $\ell(\boldsymbol{\theta})$ irrespectively of the assumed initial state $\boldsymbol{x}_0$. A continuity assumption is required in order to turn the pointwise convergence result of (6) into a uniform convergence result.

A.3 For all $x, x' \in \mathbb{X}$ and all $y \in \mathbb{Y}$, the mappings $\theta \mapsto f_\theta(x \,|\, x')$, $\theta \mapsto g_\theta(y \,|\, x)$ and $\psi \mapsto p_\psi(y)$ are continuous.

It follows directly from Assumption **A.3** that for all $K \in S^{\mathrm{T}}$, all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K$ and all $\boldsymbol{y} \in \mathbb{Y}^\times$, the mappings $\boldsymbol{\theta} \mapsto \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x} \,|\, \boldsymbol{x}')$ and $\boldsymbol{\theta} \mapsto \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y} \,|\, \boldsymbol{x})$ are continuous on the hyperplane of $\boldsymbol{\Theta}$ made of parameters with a number of targets equal to K, since these mappings are sums and products of continuous functions. Although the continuity for the multi-target Markov kernel and likelihood function is limited to hyperplanes, the result of [6, Lemma 4] can be extended to: for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\lim_{\delta\to 0} \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}'-\boldsymbol{\theta}|\leq\delta} \left| \ell_{\boldsymbol{Y}_{-\infty:0}}(\boldsymbol{\theta}') - \ell_{\boldsymbol{Y}_{-\infty:0}}(\boldsymbol{\theta}) \right| \right] = 0,$$

where $|\cdot|$ is the 1-norm on $\boldsymbol{\Theta}$, since $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}$ will be in the same hyperplane for δ small enough. The addition of the continuity assumption enables the derivation of the following result regarding the uniform convergence of the log-likelihood function:

Under Assumptions **A.1** to **A.3**, it can be proved by following the same steps as in [6, Proposition 2] that

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{x}_0\in\mathbb{X}^K} \left| \frac{1}{n} \log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{Y}_{1:n} \,|\, \boldsymbol{x}_0) - \ell(\boldsymbol{\theta}) \right| = 0, \qquad \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}\text{-a.s.} \tag{7}$$

Since the conditional log-likelihood function $\log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)$ is continuous and uniformly bounded, it follows from (7) that ℓ is also continuous on the hyperplanes of $\boldsymbol{\Theta}$ of constant target number.

The following identifiability assumption is considered in order to show the consistency of the maximum likelihood estimator

A.4 $\bar{P}_\theta = \bar{P}_{\theta^*}$ if and only if $\theta = \theta^*$ and $P_\psi = P_{\psi^*}$ if and only if $\psi = \psi^*$

Assumption **A.4** is fundamental since there would be no chance to discriminate the true value θ^* among all the other possible $\theta \in \Theta \setminus \{\theta^*\}$ if some of these parameters did yield the same law for the observations. For instance, if the colour of the target is considered as a parameter but if the likelihood of the observations does not depend on this characteristics of the target, e.g. if the observations come from a radar, then any θ obtained by changing the colour in θ^* would induce a law $\bar{P}_\theta$ that is equal to $\bar{P}_{\theta^*}$ and Assumption **A.4** would not be verified. It is shown in the next theorem that identifiability of the multi-target problem can be deduced from the identifiability of the single-target one under important special cases. The proof is in Appendix C.

Theorem 2 (Transfer of identifiability). *Under Assumption **A.4** it holds that*

a) if the true parameter $\boldsymbol{\theta}^*$ is in $\boldsymbol{\Theta}_{\lambda=0}$, then it holds that $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}} = \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\lambda=0}$,

b) if the true parameter $\boldsymbol{\theta}^*$ is in the subset $\boldsymbol{\Theta}|_{K=1}$ of $\boldsymbol{\Theta}$ made of parameters of the form $(\theta, K, p_{\mathrm{D}}, \lambda, \psi)$ with $K = 1$, then it holds that $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}} = \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}|_{K=1}$.

It is more challenging to prove that identifiability transfers to the whole parameter set $\boldsymbol{\Theta}$ and this property is assumed to hold rather than demonstrated.

A.5 $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}} = \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

Assumption **A.5** does not seem to be a stringent condition since Theorem 2 shows that the single-target identifiability is sufficient to ensure multi-target identifiability in some important special cases. However, Assumption **A.5** would not hold for $p^*_{\mathrm{D}} = 0$ since identifiability w.r.t. θ^* and K^* would clearly be lost in this case because of the absence of observations from the targets. The same remark can be made about $K^* = 0$ for the identifiability w.r.t. θ^* since there is obviously no way to learn about the dynamics and observation of the targets if none of them is present.

The different assumptions considered here are combined in the next section in order to prove the consistency of the maximum likelihood estimator.

4.3. Consistency and asymptotic normality. As a consequence of (6) and by the dominated convergence theorem it holds that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, any infinite observation sequence $\boldsymbol{y}_{1:\infty}$ and any initial states $\boldsymbol{x}_0 \in \mathbb{X}^K$ and $\boldsymbol{x}'_0 \in \mathbb{X}^{K^*}$

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \frac{\boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)}{\boldsymbol{p}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}'_0)} &= \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}^*) \\ &= \lim_{m\to\infty} \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\left[\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\left[\log \frac{\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-m:-1})}{\bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-m:-1})} \middle| \boldsymbol{Y}_{-m:-1} \right]\right] \le 0, \end{aligned} \tag{8}$$

where the inequality holds since the conditional expectations are Kullback-Leibler divergences. Yet, it could happen that some $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ would verify $\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{Y}_0 \,|\, \boldsymbol{y}_{-m:-1}) = \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{y}_{-m:-1})$ $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$-a.s. for all $m \in \mathbb{N}_0$ and for all $\boldsymbol{y}_{-m:-1}$, which would compromise identifiability. However, Assumption **A.5** is equivalent to the following statement

$$\boldsymbol{\theta} = \boldsymbol{\theta}^* \qquad \text{if and only if} \qquad \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\left[\log \frac{\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{Y}_{1:n})}{\bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_{1:n})}\right] = 0, \quad \forall n \geq 1.$$

The objective is to show that this, in turn, is equivalent to "$\boldsymbol{\theta} = \boldsymbol{\theta}^*$ if and only if $\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}^*) = 0$" since this is the term that appears in (8). Following the same line of arguments as [6, Proposition 3], we find that under Assumptions **A.1** to **A.3** and **A.5**, it holds that $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}^*)$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, from which we conclude that the considered approach allows for studying the identifiability of $\boldsymbol{\theta}^*$. Applying the strict Jensen inequality to the conditional expectation in the r.h.s. of (8), it indeed follows that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)}{\boldsymbol{p}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}'_0)} < 0,$$

for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, which implies that the likelihood of the observation sequence $\boldsymbol{y}_{1:n}$ under the parameter $\boldsymbol{\theta}$ decreases exponentially fast when compared to the likelihood under $\boldsymbol{\theta}^*$, irrespectively of the assumed initial states $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$. Denoting $\hat{\boldsymbol{\theta}}_{n,\boldsymbol{x}_0}$ the argument of the maximum of $\log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n} \,|\, \boldsymbol{x}_0)$, the consistency of the maximum likelihood estimator can be expressed as in Theorem 3 below. This theorem also states the asymptotic normality of the estimator which makes use of the Fisher information. The latter involves differentiation with respect to the parameter $\boldsymbol{\theta}$, however since the number of target K is a natural number, differentiations has to be performed for a fixed K. This is what is understood by default when writing $\nabla_{\boldsymbol{\theta}}$. Under the standard regularity assumptions (see Assumptions **A.6** to **A.8** in Appendix A), the FIM can be expressed as

$$\boldsymbol{I}(\boldsymbol{\theta}^*) = \lim_{n\to\infty} \frac{1}{n} \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\Big[\nabla_{\boldsymbol{\theta}} \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_{1:n}) \cdot \nabla_{\boldsymbol{\theta}} \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_{1:n})^{\mathrm{t}}\Big],$$

where $\cdot^{\mathrm{t}}$ is the matrix transposition.

THEOREM 3. *Under Assumptions **A.1** to **A.3** and **A.5**, it holds that*

$$\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_{n,\boldsymbol{x}_0} = \boldsymbol{\theta}^*$$

*for any $\boldsymbol{x}_0 \in \mathbb{X}^K$ with $K \in \mathbb{N}$. Considering additionally Assumptions **A.6** to **A.8** (see Appendix A) and assuming that $\boldsymbol{I}(\boldsymbol{\theta}^*)$ is positive definite, it holds that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\boldsymbol{x}_0} - \boldsymbol{\theta}^*) \to \mathcal{N}\big(0, \boldsymbol{I}(\boldsymbol{\theta}^*)^{-1}\big),$$

for any $\boldsymbol{x}_0 \in \mathbb{X}^K$ and any $K \in \mathbb{N}$, where $\to$ denotes the convergence in distribution as n tends to infinity and where $\mathcal{N}(0, V)$ is the normal distribution with mean 0 and variance V.

The proof of Theorem 3 follows from Lemma 1 combined with [6, Theorems 1 and 4]. It can be demonstrated that the result of Theorem 3 also holds for the special parameter sets $\boldsymbol{\Theta}_{\lambda=0}$, $\boldsymbol{\Theta}_{p_{\mathrm{D}}=1}$ and $\boldsymbol{\Theta}_{\lambda=0,p_{\mathrm{D}}=1}$. These special-parameter sets will be used to understand the behaviour of the FIM in simple cases in the next section.

5. Analysis of the Fisher information. Theorem 3 guarantees the convergence of the maximum likelihood estimator under certain conditions and proves the asymptotic normality of the estimator, the variance of the latter being the inverse of the FIM. It is therefore of interest to understand how the Fisher information behaves in different multi-target configurations.

This section is structured as follows: an equivalent observation model for which the FIM is easier to study is introduced in subsection 5.1 and yields a characterisation of the configurations in which the information loss induced by data association uncertainty and detection failures is strictly positive. Qualitative estimates of the information loss are then obtained when isolating the different sources of loss from subsection 5.2 to subsection 5.4. Each of these qualitative estimates are confirmed by numerical results on simulated data obtained by direct Monte Carlo integration of the original expression of the Fisher information, so as to confirm the validity of the derived alternative expressions.

Henceforth, if A and B are two square matrices of the same dimensions then $A \geq B$ is understood as $A - B \geq 0$, i.e. $A - B$ is positive semi-definite, and $A > B$ stand for $A - B > 0$, i.e. $A - B$ is positive definite.

EXAMPLE 1. *Assuming that $\boldsymbol{\theta}^*$ is in $\boldsymbol{\Theta}_{p_{\mathrm{D}}=1}$ and that the data association is known, the joint probability of the observations becomes*

$$\bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n}) = \prod_{t=1}^{n} \Big[\mathrm{Po}_{\lambda^*}(M_t - K^*) \prod_{i=K^*+1}^{M_t} p_{\psi^*}(\boldsymbol{y}_{t,i}) \Big] \times \int \pi_{\theta^*}^{\times K^*}(\boldsymbol{x}_0) \prod_{t=1}^{n} \prod_{i=1}^{K^*} \Big[g_{\theta^*}(\boldsymbol{y}_{t,i} \,|\, \boldsymbol{x}_{t,i}) f_{\theta^*}(\boldsymbol{x}_{t,i} \,|\, \boldsymbol{x}_{t-1,i}) \Big] \mathrm{d}\boldsymbol{x}_{0:n}.$$

The score is then found to be

$$\nabla_{\boldsymbol{\theta}} \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n}) = \sum_{i=1}^{K^*} \nabla_{\boldsymbol{\theta}} \log \bar{p}_{\theta^*}(\boldsymbol{y}_{1:n,i}) + \sum_{t=1}^{n} \Big[\frac{M_t - K^*}{\lambda^*} - 1 + \sum_{i=K^*+1}^{M_t} \nabla_{\boldsymbol{\theta}} \log p_{\psi^*}(\boldsymbol{y}_{t,i}) \Big]$$

so that $\boldsymbol{I}(\boldsymbol{\theta}^*) = K^* I(\boldsymbol{\theta}^*) + 1/\lambda^* + \lambda^* I^{\mathrm{C}}(\boldsymbol{\theta}^*)$ because of the independence between the targets and clutter, with $I(\boldsymbol{\theta}^*)$ and $I^{\mathrm{C}}(\boldsymbol{\theta}^*)$ the Fisher information for the distribution of one target and one clutter point respectively, where the gradient is taken w.r.t. $\boldsymbol{\theta}$, that is

$$I(\boldsymbol{\theta}^*) = \lim_{n \to \infty} \frac{1}{n} \bar{\mathbb{E}}_{\theta^*} \Big[\nabla_{\boldsymbol{\theta}} \log \bar{p}_{\theta^*}(Y_{1:n}) \cdot \nabla_{\boldsymbol{\theta}} \log \bar{p}_{\theta^*}(Y_{1:n})^{\mathrm{t}} \Big]$$
$$I^{\mathrm{C}}(\boldsymbol{\theta}^*) = \mathbb{E}_{\psi^*} \Big[\nabla_{\boldsymbol{\theta}} \log p_{\psi^*}(Y) \cdot \nabla_{\boldsymbol{\theta}} \log p_{\psi^*}(Y)^{\mathrm{t}} \Big].$$

In spite of its simplicity, Example 1 yields important remarks: unsurprisingly, if there is no missing information and no data association uncertainty, the information increases with the number of targets. Similarly, if the Fisher information of the clutter distribution p_{ψ^*} increases, then the overall information increases too. The interpretation for the Poisson parameter λ^* is less straightforward, the main objective is however to study the Fisher information w.r.t. the targets rather than the false alarms so that it is of interest to compute the score without differentiating with respect to ψ or λ.

Although the Fisher information becomes more difficult to compute when $p^*_{\mathrm{D}} \in (0, 1)$, some conclusions can be drawn by focusing on the cardinality. Only the term

$\mathbb{E}[\nabla_{p_{\mathrm{D}}} \log \boldsymbol{q}_{\boldsymbol{\theta}^*}(\boldsymbol{D}) \cdot \nabla_{p_{\mathrm{D}}} \log \boldsymbol{q}_{\boldsymbol{\theta}^*}(\boldsymbol{D})^{\mathrm{t}}] = K^*/(p^*_{\mathrm{D}}(1-p^*_{\mathrm{D}}))$ remains when computing the FIM since the parameter θ does not affect the cardinality, with $\boldsymbol{D}$ the random variable induced by $\boldsymbol{Y}$ on $\{0,1\}^{K^*}$. This term is minimal when $p^*_{\mathrm{D}} = 0.5$ and increases when p^*_{D} goes toward 0 or 1. This is not sufficient to conclude since the fact that information is lost when detection failures happen is not taken into account in the cardinality and the information is the same for, e.g. p^*_{D} equal to 0.99 or 0.01. Indeed, it is equally easy to estimate p^*_{D} when an observation is always or never received. For this reason, it is useful to consider the information w.r.t. θ^* only.

The objective will therefore be to characterise how the Fisher information

$$\boldsymbol{I}(\theta^*) = \lim_{n\to\infty} \frac{1}{n} \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\Big[\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_{1:n}) \cdot \nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_{1:n})^{\mathrm{t}}\Big], \tag{10}$$

of a multi-object dynamical system behaves when compared to the information of the unperturbed system that excludes false alarms, detection failures and for which data association is known. We refer to the difference between (10) and the latter as the *information loss*. Since the Fisher information of the unperturbed system is a quantity that depends on the number of objects in the system, the aim is to express the information loss as a function of the single-object FIM $I(\theta^*)$. The FIM of the unperturbed multi-object system is clearly equal to $K^* I(\theta^*)$ because of the independence between the targets' observation in the absence of data association uncertainty. In order to compute $\boldsymbol{I}(\theta^*)$, we have to take the logarithm of the p.d.f. $\bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n}, \boldsymbol{x}_{0:n}) = \pi^{\times K^*}_{\theta^*}(\boldsymbol{x}_0) \prod_{t=1}^n \big[\boldsymbol{g}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_t \,|\, \boldsymbol{x}_t) \boldsymbol{f}_{\theta^*}(\boldsymbol{x}_t \,|\, \boldsymbol{x}_{t-1})\big]$. However, the presence of a sum in the term $\boldsymbol{g}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_t \,|\, \boldsymbol{x}_t)$ prevents from further analysing the Fisher information in a general setting. To avoid directly dealing with these sums, an equivalent observation model which depends explicitly on the assignment is introduced in the next section. This observation model is an important contribution since it allows us to understand the behaviour of the Fisher information for multi-target tracking.

5.1. Alternative observation model. Let d_H be the Hamming metric on the symmetric group $\mathrm{Sym}(k)$ characterised by letting $d_H(\sigma, \sigma')$ be the number of points moved by $\sigma' \circ \sigma^{-1}$ for any given $\sigma, \sigma' \in \mathrm{Sym}(k)$. Let $\oplus$ be the vector concatenation operator such that if $\boldsymbol{y} = [y_1, \dots, y_n]^{\mathrm{t}} \in \mathbb{Y}^n$ and $\boldsymbol{y}' = [y'_1, \dots, y'_m]^{\mathrm{t}} \in \mathbb{Y}^m$ then $\boldsymbol{y} \oplus \boldsymbol{y}' \doteq [y_1, \dots, y_n, y'_1, \dots, y'_m]^{\mathrm{t}} \in \mathbb{Y}^{n+m}$. Let $R_{\boldsymbol{d}}$ be the matrix of size $|\boldsymbol{d}| \times K^*$ such that $(R_{\boldsymbol{d}})_{i,j} = \delta_{j,r(i)}$ for any $\boldsymbol{d} \in \{0,1\}^{K^*}$, i.e. $R_{\boldsymbol{d}}$ has as many lines as there are detected targets and can be seen as a mask matrix that removes the observations of non-detected ones. Let S_σ be the permutation matrix corresponding to $\sigma \in \mathrm{Sym}(k)$ for any $k \geq 1$, i.e. $S_\sigma \doteq [\boldsymbol{e}^{\mathrm{t}}_{\sigma(1)}, \dots, \boldsymbol{e}^{\mathrm{t}}_{\sigma(K^*)}]^{\mathrm{t}}$, with $\boldsymbol{e}_i$ the row vector with 1 at the i^{th} position and 0 elsewhere. The observation model with known data association is written as $\boldsymbol{Y}_t = \boldsymbol{h}(\boldsymbol{X}_t) + \boldsymbol{\eta}$ with $\boldsymbol{h}$ and $\boldsymbol{\eta}$ the multi-target observation function and the observation noise respectively, where $\boldsymbol{\eta}$ is i.i.d. across its K^* components. The false alarms are defined as a random variable $\hat{\boldsymbol{Y}}$ in $\mathbb{Y}^\times$, independent of $\boldsymbol{Y}_t$, such that $\hat{\boldsymbol{Y}}_i \sim p_\psi$ and $\hat{\boldsymbol{Y}}_i$ is independent of $\hat{\boldsymbol{Y}}_j$ for any $1 \leq i, j \leq \#\hat{\boldsymbol{Y}}$. The observation model of interest can then be defined for given integers $\alpha > 0$ and $0 \leq \beta \leq K^*$ as

$$\mathbf{Y}^{\alpha,\beta}_t = S_\varsigma\big((R_{\boldsymbol{D}} \boldsymbol{Y}_t) \oplus \hat{\boldsymbol{Y}}\big), \tag{11}$$

where $\boldsymbol{D}$ is a random element of $B_\beta \doteq \{\boldsymbol{d}' \in \{0,1\}^{K^*} : |\boldsymbol{1} - \boldsymbol{d}'| \leq \beta\}$ having as a distribution the restriction $\boldsymbol{q}^\beta_{\boldsymbol{\theta}}$ of $\boldsymbol{q}_{\boldsymbol{\theta}}$ to B_β and where ς is a random permutation drawn from the uniform law u^α_k with $k = \#\hat{\boldsymbol{Y}} + |\boldsymbol{d}|$ on the set $A^\alpha_k \doteq \{\sigma \in \mathrm{Sym}(k) : d_H(\mathrm{id}, \sigma) \leq \alpha\}$ with id denoting the identity function. Henceforth, the letter ς will be used for a random permutation and σ for a realisation. The case $\alpha = 0$ is not considered to avoid redundancy: it holds that $A_k^0 = A_k^1 = \{\mathrm{id}\}$ for any $k \geq 1$ since permutations that are different from the identity move at least two points. The case of Example 1 is recovered by considering $\alpha = 1$ and $\beta = 0$, i.e. $\varsigma = \mathrm{id}$ and $\boldsymbol{D} = \boldsymbol{1}$ a.s., whereas the full data-association problem corresponds to the choice $\alpha = \infty$ and $\beta = \infty$.

The alternative observation model (11) brings insight about the associated FIM $\boldsymbol{I}^{\alpha,\beta}(\theta^*)$, when compared to the unperturbed case. The corresponding information loss is $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*) \doteq K^* I(\theta^*) - \boldsymbol{I}^{\alpha,\beta}(\theta^*)$. In some cases, the relative information loss $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*)/(K^* I(\theta^*))$ will be used instead. The next theorem is the central result of this section, its proof can be found in Appendix D.

Theorem 4. *Under Assumptions **A.1**, **A.2** and **A.6** to **A.8**, the information loss $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*)$ verifies $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*) \geq 0$ for any $\alpha \geq 1$ and any $\beta \geq 0$, the inequality being strict if either $\alpha > 1$ or $\beta > 0$ and if $I(\theta^*) \neq 0$.*

Notice that the condition $\alpha > 1$ would not be sufficient to make the inequality in Theorem 4 strict if λ^* were equal to 0 since data association might have no influence in some specific configurations, e.g. when the individual likelihood does not depend on the objects' state. Theorem 4 does not provide a quantitative characterisation of the information loss. Doing so is challenging in the general case, yet, the behaviour of the information loss can be analysed for special cases, and such will be the objective in the remainder of this section.

One of the advantages with the modified observation model (11) is that the Fisher identity can be utilised as an alternative way of computing the score function based on the unobserved random variables in this model:

$$\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n}) = \bar{\mathbb{E}}_{\boldsymbol{\theta}}\big[\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{Y}^{\alpha,\beta}_{1:n}, \varsigma_{1:n}, \boldsymbol{D}_{1:n}, \boldsymbol{X}_{0:n}) \,|\, \boldsymbol{Y}^{\alpha,\beta}_{1:n} = \boldsymbol{y}_{1:n}\big], \tag{12}$$

where

$$\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n}, \sigma_{1:n}, \boldsymbol{d}_{1:n}, \boldsymbol{x}_{0:n}) = \pi_\theta^{\times K}(\boldsymbol{x}_0) \prod_{t=1}^{n} \bigg[\mathrm{Po}_\lambda(M_t - |\boldsymbol{d}_t|)$$
$$\times \prod_{i=|\boldsymbol{d}_t|+1}^{M_t} p_\psi(\boldsymbol{y}_{t,\sigma_t(i)}) \prod_{i=1}^{|\boldsymbol{d}|} g_\theta(\boldsymbol{y}_{t,\sigma_t(i)} \,|\, \boldsymbol{x}_{t,r(i)}) \prod_{i=1}^{K} f_\theta(\boldsymbol{x}_{t,i} \,|\, \boldsymbol{x}_{t-1,i}) u^\alpha_{M_t}(\sigma_t) \boldsymbol{q}^\beta_{\boldsymbol{\theta}}(\boldsymbol{d}_t)\bigg].$$

The simplification of the expression of $\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n})$ is only notational. The random variables $\varsigma_{1:n}$, $\boldsymbol{D}_{1:n}$ and $\boldsymbol{X}_{0:n}$ are conditioned on the event $\boldsymbol{Y}^{\alpha,\beta}_{1:n} = \boldsymbol{y}_{1:n}$ in (12), so that their respective distributions are now the conditional distributions given the observations, which are more complex than their priors. Yet, the Fisher identity enabled to move the sums and integrals outside of the logarithm, hence making easier the analysis of the FIM.

5.2. Single static target with false alarm. Consider the case of one almost-surely detected static target with state $x \in \mathbb{X}$ which observation is corrupted by false alarms and unknown data association. The corresponding $\boldsymbol{\theta}^*$ is in the hyperplane $\boldsymbol{\Theta}_{p_{\mathrm{D}}=1}|_{K=1}$ of the special parameter set $\boldsymbol{\Theta}_{p_{\mathrm{D}}=1}$ composed of parameters for which $K = 1$. It is sufficient to study one time step since the observations at different times are now independent and it holds that $\boldsymbol{I}(\theta^*) = \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}[\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_1) \cdot \nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_1)^{\mathrm{t}}]$.

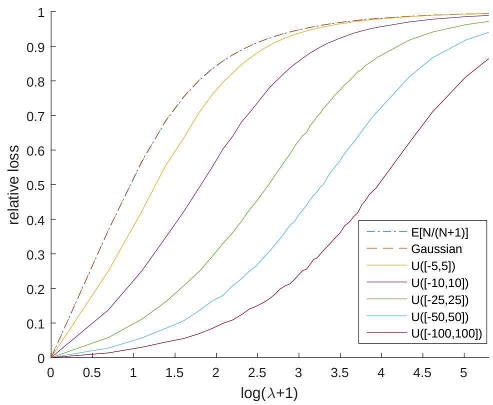


Fig. 1: Information loss as a function of the Poisson parameter λ in log-scale, calculated with 5×10^5 samples (Gaussian: worst-case scenario; $U([-a, a])$: uniform distribution over $[-a, a]$).

Making use of the Fisher identity (12), the FIM $\boldsymbol{I}^{\infty,0}(\theta^*)$ can be expressed as

$$\boldsymbol{I}^{\infty,0}(\theta^*) = \mathbb{E}_{\boldsymbol{\theta}^*}\Bigg[\sum_{i,j=1}^{M} c_i(\boldsymbol{Y})c_j(\boldsymbol{Y})\nabla_\theta \log g_{\theta^*}(\boldsymbol{Y}_i \,|\, x) \cdot \nabla_\theta \log g_{\theta^*}(\boldsymbol{Y}_j \,|\, x)^{\mathrm{t}}\Bigg],$$

where $M = \#\boldsymbol{Y}$ and where

$$c_i(\boldsymbol{y}) = \sum_{\substack{\sigma\in\mathrm{Sym}(M)\\ \sigma(1)=i}} u_M(\sigma \,|\, \boldsymbol{y}) = \frac{g_{\theta^*}(\boldsymbol{y}_i \,|\, x)/p_{\psi^*}(\boldsymbol{y}_i)}{\sum_{j=1}^{M} g_{\theta^*}(\boldsymbol{y}_j \,|\, x)/p_{\psi^*}(\boldsymbol{y}_j)}.$$

Identifying the parameter θ^* is most challenging when the distribution of the false alarm is equal to the one of the target-originated observation at θ^*, i.e. $p_{\psi^*} = g_{\theta^*}(\cdot \,|\, x)$, since all the observations will look alike for θ close to θ^*. In this case it holds that $c_i(\boldsymbol{Y}) = 1/M$ for any $1 \leq i \leq M$ so that $\boldsymbol{I}^{\infty,0}(\theta^*) = \mathbb{E}[1/(N+1)]I(\theta^*)$ where the expectation is taken w.r.t. the random variable $N \sim \mathrm{Po}_{\lambda^*}$. It follows that the relative information loss is equal to $\mathbb{E}[N/(N+1)]$ so that it is strictly increasing with λ^* and tends to 1 when λ^* tends to infinity. This result is supported by the experiments displayed in Figure 1 where the observation of one static target in $\mathbb{X} = \mathbb{R}$ at $x = 0$ is corrupted by false alarms. The observation model is assumed to be linear and Gaussian with variance θ such that $\theta^* = 1$. Note that the loss of information is indeed with respect to a parameter of the MTT model, here the variance of the observation noise. Cases where the false alarm is uniform over the subset $[-a, a]$ with $a \in \{5, 10, 25, 50, 100\}$ are also considered. The scenario where the false alarm is distributed in the same way as the target-originated observation at θ^*, i.e. $p_{\psi^*} = g_{\theta^*}(\cdot \,|\, x)$, is also confirmed to be the worst-case scenario.

Since a static target is considered in this scenario as well as in subsection 5.3 below, the CRLB becomes equal to the inverse of the asymptotic FIM. Analysis of the information loss for the targets' state due to association uncertainty, false alarms and detection failures have been conducted in this case for specific models and for a single target in [16, 28]. More specifically, [16] studies the information loss when the observation noise is assumed to be either a generalized Gaussian distribution with different shape parameters or a Johnson distribution for different kurtosis values, while [28] considers a dynamic version of the CRLB for a single target and quantifies the information loss for different values of the probability of detection and of the false-alarm rate. In contrast, our analysis here applies to the parameters of the MTT models, as opposed to the targets' state, and therefore brings additional insight on Fisher information beyond the results reported in [16, 28].

In the next two sections, the focus will be on understanding the role played specifically by unknown data association and detection failures.

5.3. Unknown data association. In order to set the focus on data association, it is assumed that $\boldsymbol{\theta}^*$ belongs to the special parameter set $\boldsymbol{\Theta}_{\lambda=0,p_{\mathrm{D}}=1}$. In these conditions, the joint probability of the observations and states becomes

$$\begin{aligned}(13)\quad \bar{\boldsymbol{p}}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n},\boldsymbol{x}_{0:n}) &= \pi_{\theta^*}^{\times K^*}(\boldsymbol{x}_0)\\ &\times \prod_{t=1}^{n}\sum_{\sigma\in\mathrm{Sym}(K^*)}\left[\prod_{i=1}^{K^*}\left[g_{\theta^*}(\boldsymbol{y}_{t,\sigma(i)}\,|\,\boldsymbol{x}_{t,i})f_{\theta^*}(\boldsymbol{x}_{t,i}\,|\,\boldsymbol{x}_{t-1,i})\right]u_{K^*}(\sigma)\right].\end{aligned}$$

The sum in the previous expression makes it difficult to directly compute the FIM. Some insight about it can however be obtained by considering static objects as in the following example.

EXAMPLE 2. *Let $x_1,\dots,x_{K^*}$ be the known position of K^* static objects. The joint distribution of the observations is then found to be*

$$\boldsymbol{p}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n}) = \prod_{t=1}^{n}\sum_{\sigma\in\mathrm{Sym}(K^*)}\left[\prod_{i=1}^{K^*} g_{\theta^*}(\boldsymbol{y}_{t,\sigma(i)}\,|\,x_i)u_{K^*}(\sigma)\right].$$

In this simplified setting, we can assume that g_{θ^*} has finite support so that the objects' state can be chosen far enough from each other for $\prod_{i=1}^{K^*} g_{\theta^*}(\boldsymbol{Y}_{t,\sigma(i)}\,|\,x_i)=0$ to hold $\mathbb{P}$-a.s. whenever $\sigma\neq\mathrm{id}$. In this case, and as expected, there is no loss of information when compared to the case with known data association. A less intuitive result can be found when all the objects' state are equal to a given $x\in\mathbb{X}$. In this situation, it holds that all permutations are equally probable so that $\boldsymbol{p}_{\boldsymbol{\theta}^*}(\boldsymbol{y}_{1:n})=\prod_{t=1}^{n}\prod_{i=1}^{K^*} g_{\theta^*}(\boldsymbol{y}_{t,i}\,|\,x)$, and once again, there is no loss of information. These two cases correspond to extreme configurations where the uncertainty on the data association is either resolvable or irrelevant.

The Fisher identity can be used to provide an expression of the Fisher information for static objects as follows. For any fixed $x_1,\dots,x_{K^*}$, the Fisher information for $\alpha=\infty$ (fully unknown association) and $\beta=0$ can be deduced from

$$\begin{aligned}\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}) &= \mathbb{E}_{\boldsymbol{\theta}}\left[\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{Y},\varsigma)\,|\,\boldsymbol{Y}=\boldsymbol{y}\right]\\ &= \sum_{\sigma\in\mathrm{Sym}(K)}\sum_{i=1}^{K}\nabla_\theta \log g_\theta(\boldsymbol{y}_{\sigma(i)}\,|\,x_i)u_K(\sigma|\boldsymbol{y}).\end{aligned}$$

The FIM $\boldsymbol{I}^{\alpha,\beta}(\theta^*)$ with $\alpha=\infty$, $\beta=0$ and without false alarm is found to be

$$\boldsymbol{I}^{\infty,0}(\theta^*) = \sum_{i,j,k,l=1}^{K^*}\mathbb{E}_{\theta^*}\left[c_{i,k}(\boldsymbol{Y})c_{j,l}(\boldsymbol{Y})\,\mathrm{Sco}_i(\boldsymbol{Y}_k)\cdot\mathrm{Sco}_j(\boldsymbol{Y}_l)^{\mathrm{t}}\right]$$

with $\mathrm{Sco}_i(y) = \nabla_\theta \log g_{\theta^*}(y \,|\, x_i)$ for any $y \in \mathbb{Y}$ and with

$$c_{i,k}(\boldsymbol{y}) = g_{\theta^*}(\boldsymbol{y}_k \,|\, x_i) \sum_{\substack{\sigma \in \mathrm{Sym}(K^*) \\ \sigma(i)=k}} \prod_{j \neq i} g_{\theta^*}(\boldsymbol{y}_{\sigma(j)} \,|\, x_j) \bigg(\sum_{\sigma \in \mathrm{Sym}(K^*)} \prod_{j=1}^{K^*} g_{\theta^*}(\boldsymbol{y}_{\sigma(j)} \,|\, x_j) \bigg)^{-1}$$

for any $\boldsymbol{y} \in \mathbb{Y}^{K^*}$ and any $i,k \in \{1,\dots,K^*\}$. The term $c_{i,k}(\boldsymbol{y})$ is the conditional probability for the object with state x_i to have generated observation k given all observations $\boldsymbol{y}$.

In order to obtain a quantitative characterisation of the information loss, a special likelihood has to be introduced. We consider an observation model of the same form as the one displayed in Figure 2, i.e. such that $\mathbb{Y}$ is compact and there exists a collection of disjoint subsets $\{B_i\}_{i=1}^{K^*}$ of $\mathbb{Y}$ such that $g_\theta(\cdot \,|\, x_i)$ uniformly distributes a probability mass $\epsilon > 0$ outside of B_i. An example of such a distribution is given in Figure 2 for two objects. Then, for K objects,

$$\boldsymbol{I}^{\infty,0}(\theta^*) = \sum_{i,j,k,l=1}^{K} E_{i,j}^{k,l}(\theta^*) \tag{17}$$

with $E_{i,j}^{k,l}(\theta^*) \doteq \mathbb{E}_{\theta^*}\big[c_{i,k}(\boldsymbol{Y}) c_{j,l}(\boldsymbol{Y}) \,\mathrm{Sco}_i(\boldsymbol{Y}_k) \cdot \mathrm{Sco}_j(\boldsymbol{Y}_l)^{\mathrm{t}}\big]$ for any $i,j,k,l \in \{1,\dots,K\}$. The objective is now to understand the behaviour of $\boldsymbol{I}^{\infty,0}(\theta^*)$ when K is large. The order of the term $c_{i,k}(\boldsymbol{y})$ is in $O(1)$ when $i = k$ and in $O(K^{-1})$ when $i \neq k$. The order of the summand in (17) can then be determined for the different values of i,j,k,l. For instance:

- If $i \neq k \neq l \neq j$ then

$$E_{i,j}^{k,l}(\theta^*) = \frac{\epsilon^2}{|\mathbb{Y} \setminus B_k|^2} \int_{C_{i,j}^{k,l}} c_{i,k}(\boldsymbol{y}) c_{j,l}(\boldsymbol{y}) \frac{\nabla_\theta\, g_{\theta^*}(\boldsymbol{y}_k \,|\, x_i) \cdot \nabla_\theta\, g_{\theta^*}(\boldsymbol{y}_l \,|\, x_j)^{\mathrm{t}}}{g_{\theta^*}(\boldsymbol{y}_k \,|\, x_i) g_{\theta^*}(\boldsymbol{y}_l \,|\, x_j)} \mathrm{d}\boldsymbol{y}, \tag{18}$$

where $C_{i,j}^{k,l} \doteq \{\boldsymbol{y} \in \mathbb{Y}^K : \boldsymbol{y}_k \in B_i,\ \boldsymbol{y}_l \in B_j\}$, because $g_{\theta^*}(y \,|\, x_k) = \epsilon / |\mathbb{Y} \setminus B_k|$ for all $y \notin B_k$ and because $B_i \cap B_k = \emptyset$ since $i \neq k$. When K increases, $\mathbb{Y}$ needs to be augmented at least linearly to ensure that the family $\{B_i\}_{i=1}^K$ is disjoint and (18) shows inverse proportionality with $|\mathbb{Y}|^2$, so that it is of order $O(K^{-4})$ at most. There are $O(K^4)$ terms of this form in the sum in the r.h.s. of (17) so that the sum of these terms is of order $O(1)$ at most.

- If $k = l$ and $i \neq j$ then $E_{i,j}^{k,l}(\theta^*) = 0$ since in this case it holds that $\mathrm{Sco}_i(\boldsymbol{y}_k) \cdot \mathrm{Sco}_j(\boldsymbol{y}_k)^{\mathrm{t}} = 0$ for any $\boldsymbol{y} \in \mathbb{Y}^K$ which follows from the facts that $\mathrm{Sco}_i(y) \neq 0$ when $y \in B_i$ only and that $B_i \cap B_j = \emptyset$.
- If $i = j = k = l$ then

$$E_{i,j}^{k,l}(\theta^*) = \int \mathbf{1}_{B_i}(\boldsymbol{y}_i) c_{i,i}(\boldsymbol{y})^2 \frac{\nabla_\theta\, g_{\theta^*}(\boldsymbol{y}_i \,|\, x_i) \cdot \nabla_\theta\, g_{\theta^*}(\boldsymbol{y}_i \,|\, x_i)^{\mathrm{t}}}{g_{\theta^*}(\boldsymbol{y}_i \,|\, x_i)} \mathrm{d}\boldsymbol{y},$$

which does not depend on K or $|\mathbb{Y}|$ and is therefore of order $O(1)$.

Following the same principles for the other values of i,j,k,l, we find that $\boldsymbol{I}^{\infty,0}(\theta^*)$ is of order $O(K)$. Since the information in the idealised observation model, i.e. when data association is known, is equal to $KI(\theta^*)$, it follows that the relative loss is constant. In other words, for a large number of targets, adding more targets increases the information at the same rate as in the idealised model.

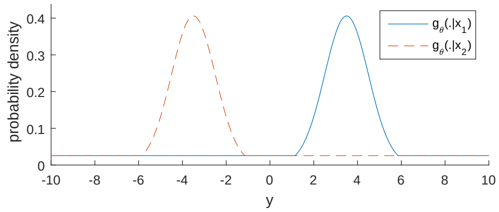


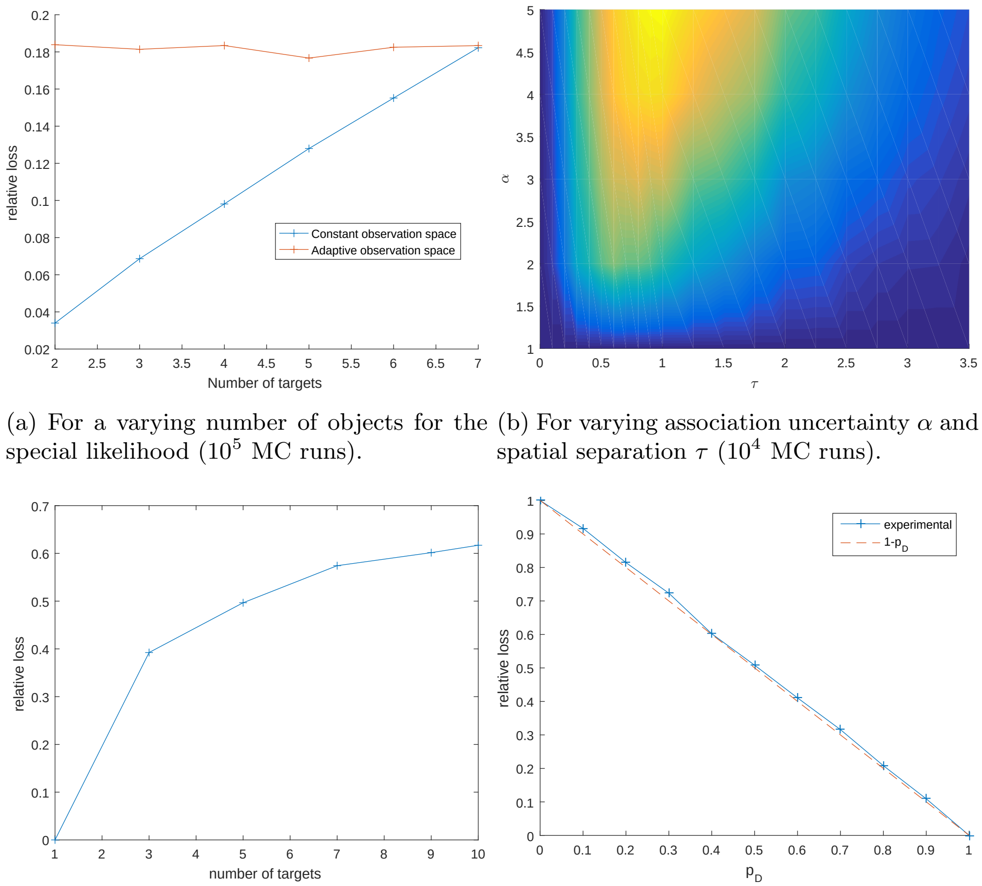
Fig. 2: Example of likelihood with two objects at states x_1 and x_2.

Validation via simulations. The special likelihood is taken of the form $g_\theta(y \mid x_k) = \mathcal{N}(y; x_k + m, 1)$ if $y \in B_k \doteq (x_k + m - r, x_k + m + r)$ and $g_\theta(y \mid x_k) = \epsilon / |\mathbb{Y} \setminus B_k|$ otherwise, with $\epsilon = 0.1$ and with r characterised by $\int_{B_k} \mathcal{N}(y; x_k, 1) \mathrm{d}y = 1 - \epsilon$ via B_k. In this case, the displacement m is considered as the parameter θ and the true value is $\theta^* = 0$. The relative information loss associated with this likelihood is displayed in Figure 3a under two different configurations. The first one (*Constant observation space* in the figure) corresponds to the case where the observation space is large enough to meet the requirements associated with (17); the relative loss can be seen to increase linearly with the number of targets. The second case (*Adaptive observation space* in the figure) corresponds to the case where the observation space has to be augmented to fit new targets and shows a constant relative information loss. This last result is consistent with the conclusion above that the information loss is of the same order as the number of targets when the observation space has to be augmented.

Further simulations. Five static objects on $\mathbb{X} = \mathbb{R}$ at positions $x_i = \tau(i - 3)$ with $i \in \{1, \ldots, 5\}$ are observed via a linear Gaussian model with variance equal to 1. The objective is to understand how the FIM $\boldsymbol{I}^{\alpha,0}(\theta^*)$ evolves with α and with the position of the objects. It is assumed that θ parametrises the variance of the Gaussian observation model only, so that $\boldsymbol{I}^{\alpha,0}(\theta^*)$ is a scalar. The relative information loss is displayed in Figure 3b and confirms the intuition that the information loss increases with α, except in the case $\alpha = 1$ where there is no loss by definition since $A^1_k = \{\mathrm{id}\}$ for any $k \geq 1$ so that the data association is known in this case. Also, the loss is increased when the individual likelihoods overlap while being increasingly different and then decreases when the overlap becomes negligible. The maximum is reached when $\tau = 1$, that is when the distance $|x_i - x_{i-1}|$ between two consecutive objects is 1 for any $i \in \{2, \ldots, 5\}$. The fact that there is no loss when $\tau = 0$ follows from the irrelevance of data association uncertainty when all objects are at the same position, as explained in Example 2. To better understand the behaviour w.r.t. the number of targets, Figure 3c displays the relative information loss for 1 to 10 targets in the case of full data association uncertainty with $\tau = 1$.

The results for the two sets of simulations are consistent and show the same trend: the relative information loss increases with the number of targets but tend to stabilise. To sum up, there is no loss for 1 target by construction, the loss is linear in the number of targets when there are sufficiently many, and it increases the fastest during the transition between these two modes.

5.4. Detection failures. In this section, the case of detection failures is analysed when assuming that there are no false alarms, that is when $\boldsymbol{\theta}^*$ is in the special-parameter set $\boldsymbol{\Theta}_{\lambda=0}$. To establish our main result in this section (Theorem 5), we will use the concept of missing information (see for instance [3] in the context of



(a) For a varying number of objects for the special likelihood (10^5 MC runs).

(b) For varying association uncertainty α and spatial separation τ (10^4 MC runs).

(c) For a varying number of objects with separation $\tau = 1$ and $\alpha = \infty$ (10^4 MC runs).

(d) For a varying probability of detection p_{D}, compared to $1 - p_{\mathrm{D}}$.

Fig. 3: Information loss with association uncertainty (3a-3c) or detection failures (3d) according to the models introduced in subsection 5.3 and subsection 5.4 respectively. See relevant subsections for interpretations.

Approximate Bayesian Computation).

THEOREM 5. *Assuming $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_{\lambda=0}$, the information loss $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*)$ for known data association with unconstrained detection failures, i.e for $\alpha = 1$, $\beta = \infty$, is found to be*

$$\boldsymbol{I}^{1,\infty}_{\mathrm{loss}}(\theta^*) \doteq (1 - p^*_{\mathrm{D}})K^* I(\theta^*).$$

The proof can be found in Appendix E. It follows from Theorem 5 that in the considered configuration the FIM $\boldsymbol{I}^{1,\infty}(\theta^*)$ can be made arbitrary close to 0 by making p^*_{D} tend to 0. Also, there is no loss at all when $p^*_{\mathrm{D}} = 1$, as expected. In order to verify the result of Theorem 5 in practice, a single-object scenario with detection failures and without false alarms is considered. The object starts at time $t = 0$ from the position $x_0 = 0$ and evolves in $\mathbb{X} = \mathbb{R}$ according to a random walk with standard deviation 0.1 until time $n = 50$. The observation is linear and Gaussian with variance equal to 1.

The integral over the state space in the expression of the score is computed by Monte Carlo simulation with 10^3 samples while the expectation in the Fisher information utilises 10^4 samples. The relative information loss is displayed in Figure 3d and confirms the coefficient $1 - p_{\mathrm{D}}$ found analytically in Theorem 5. The next example shows how the Fisher information evolves in general when adding new objects without involving them in data association uncertainty.

EXAMPLE 3. *The Fisher information $\boldsymbol{I}^{\alpha,\beta}(\theta^*, K)$ of a K-object problem can be related to the information $\boldsymbol{I}^{\alpha,\beta}(\theta^*, K+N)$ where the N new objects are not perturbed by data-association uncertainties, i.e. when the random variable ς in the observation model (11) verifies $\varsigma|_D = \mathrm{id}$ a.s. with $D \doteq \{|\boldsymbol{d}_{1:K+1}|, \dots, |\boldsymbol{d}_{1:N}|\}$. It then follows from Theorem 5 that $\boldsymbol{I}^{\alpha,\beta}(\theta^*, K+N) = \boldsymbol{I}^{\alpha,\beta}(\theta^*, K) + p^*_{\mathrm{D}} N I(\theta^*)$ for any $\alpha > 0$ and any $\beta \geq 0$. This example gives an upper bound for the increase of the Fisher information when the number of objects is increased, since it depicts the case where there is no data association uncertainty for these objects. This would correspond in practice to a case where the added objects are in an area where there is no false alarm and where these objects are "far" from the existing objects as well as "far" from each other, where "far" depends on the likelihood.*

6. Conclusion. The first important result in this article is the proof of consistency of the maximum likelihood estimator for multi-target tracking under *weak* conditions, where weak means that these conditions are as often as possible applying to the single-target dynamics and observation. Asymptotic normality holds under additional assumptions and the second part of the article brings understanding to the asymptotic variance of the maximum likelihood estimate by analysing the FIM corresponding to multi-target tracking. Qualitative results are obtained in the general case, that is, the Fisher information decreases with data association uncertainty, detection failures and in the presence of false alarms. Quantitative results are also derived in important special cases: a) one static target with false alarm and unknown data association, b) multiple static targets with unknown data association under a particular observation model, and c) multiple targets with detection failures.

Future works include the study of identifiability of specific observation-to-track associations, instead of marginalising over all possibilities as considered in this article. Such an approach involves additional challenges since the parameters to be learned increase in dimensionality with time, so that it is not a special case of the results presented here.

Acknowledgement. All authors were supported by the Singapore ministry of education tier 1 grant number R-155-000-182-114.

Appendix A. Assumptions for Theorem 3. The following assumptions are required for the proof of the asymptotic normality of the maximum likelihood estimator in multi-target tracking. The norm $\|\cdot\|$ is defined as $\|M\| = \sum_{i,j} |M_{i,j}|$ for any matrix M.

A.6 For all $K \in S^{\mathrm{T}}$, all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K$ and all $\boldsymbol{y} \in \mathbb{Y}^{\times}$, the mappings $\boldsymbol{\theta} \mapsto \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x} \,|\, \boldsymbol{x}')$ and $\boldsymbol{\theta} \mapsto \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y} \,|\, \boldsymbol{x})$ are twice continuously differentiable on the hyperplane of $\boldsymbol{\Theta}$ made of parameters with a number of target equal to K

A.7 It holds that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K} \|\nabla_{\boldsymbol{\theta}} \log \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x} \,|\, \boldsymbol{x}')\| < \infty \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}^K} \|\nabla^2_{\boldsymbol{\theta}} \log \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x} \,|\, \boldsymbol{x}')\| < \infty$$

and that

$$\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\Big[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{x}\in\mathbb{X}^K}\|\nabla_{\boldsymbol{\theta}}\log \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{Y}\,|\,\boldsymbol{x})\|\Big]<\infty \;\text{ and }\; \bar{\mathbb{E}}_{\boldsymbol{\theta}^*}\Big[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{x}\in\mathbb{X}^K}\|\nabla^2_{\boldsymbol{\theta}}\log \boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{Y}\,|\,\boldsymbol{x})\|\Big]<\infty$$

A.8 For all $\boldsymbol{y}\in\mathbb{Y}^\times$, there exists an integrable function $h_{\boldsymbol{y}}:\bigcup_{k\geq 0}\mathbb{X}^k\to\mathbb{R}^+$ such that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{x})\leq h_{\boldsymbol{y}}(\boldsymbol{x})$. For all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and for all $\boldsymbol{x}\in\mathbb{X}^K$, there exist integrable functions $h^1_{\boldsymbol{x}},h^2_{\boldsymbol{x}}:\mathbb{Y}^\times\to\mathbb{R}^+$ such that $\|\nabla_{\boldsymbol{\theta}}\,\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{x})\|\leq h^1_{\boldsymbol{x}}(\boldsymbol{y})$ and $\|\nabla^2_{\boldsymbol{\theta}}\,\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{x})\|\leq h^2_{\boldsymbol{x}}(\boldsymbol{y})$

Appendix B. Proof of Lemma 1. It follows from Assumption A.2 that the supremum $\boldsymbol{b}^{\mathrm{C}}_+$ of the clutter density $\boldsymbol{p}_{(\lambda,\psi)}$ characterised for any $k\in\mathbb{N}_0$ and any $\boldsymbol{y}\in\mathbb{Y}^k$ by $\boldsymbol{p}_{(\lambda,\psi)}(\boldsymbol{y})=\mathrm{Po}_\lambda(k)\prod_{i=1}^k p_\psi(\boldsymbol{y}_i)$ verifies $\boldsymbol{b}^{\mathrm{C}}_+<\infty$ since

$$\sup_{\lambda\in S^{\mathrm{C}}}\bigg(\sum_{k\geq 0}\sup_{(\psi,\boldsymbol{y})\in\Psi\times\mathbb{Y}^k}\boldsymbol{p}_{(\lambda,\psi)}(\boldsymbol{y})\bigg)=\sup_{\lambda\in S^{\mathrm{C}}}\sum_{k\geq 0}\frac{(\lambda b^{\mathrm{C}}_+)^k e^{-\lambda}}{k!}=\sup_{\lambda\in S^{\mathrm{C}}}e^{\lambda(b^{\mathrm{C}}_+-1)}<\infty,$$

and since all the terms in the sum are positive. It then holds that

$$\hat{\boldsymbol{b}}_+\leq\sup_{(p_{\mathrm{D}},K)\in(0,1)\times S^{\mathrm{T}}}\boldsymbol{b}^{\mathrm{C}}_+(1-p_{\mathrm{D}}+p_{\mathrm{D}}b^{\mathrm{T}}_+)^K<\infty,$$

which concludes the first part of the proof. For any $k\in\mathbb{N}_0$ and any $\boldsymbol{y}\in\mathbb{Y}^k$

$$\int\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{x})\mathrm{d}\boldsymbol{x}\geq\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathrm{Bi}^K_{p_{\mathrm{D}}}*\mathrm{Po}_\lambda(k)\prod_{i=1}^k\big[b^{\mathrm{T}}_-(\boldsymbol{y}_i)\wedge b^{\mathrm{C}}_-(\boldsymbol{y}_i)\big].$$

It also holds that $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathrm{Bi}^K_{p_{\mathrm{D}}}*\mathrm{Po}_\lambda(k)>0$ for any $k\in\mathbb{N}_0$ since the support of Po_λ is $\mathbb{N}_0$ for any $\lambda\in S^{\mathrm{C}}$ which guarantees that the convolution has also $\mathbb{N}_0$ as a support so that the infimum is strictly greater than zero. It follows that $\boldsymbol{b}_-(\boldsymbol{y})>0$ and, considering (5), that $\bar{\mathbb{E}}_{\boldsymbol{\theta}^*}[|\log\boldsymbol{b}_-(\boldsymbol{Y})|]<\infty$. Similarly, for any $k\in\mathbb{N}_0$ and any $\boldsymbol{y}\in\mathbb{Y}^k$ it holds that $\boldsymbol{b}_+(\boldsymbol{y})\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathrm{Bi}^K_{p_{\mathrm{D}}}*\mathrm{Po}_\lambda(k)\prod_{i=1}^k\big[b^{\mathrm{T}}_+(\boldsymbol{y}_i)\vee b^{\mathrm{C}}_+(\boldsymbol{y}_i)\big]$, which is finite when k is finite. In the infinite case, noticing that $\mathrm{Po}_\lambda(k-K)p^K_{\mathrm{D}}$ is the leading term in the convolution, we find that

$$\lim_{k\to\infty}\sup_{\boldsymbol{y}\in\mathbb{Y}^k}\boldsymbol{b}_+(\boldsymbol{y})\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}c_{\boldsymbol{\theta}}\lim_{k\to\infty}\frac{e^{-\lambda}}{(k-K)!}\Big[\sup_{y\in\mathbb{Y}}\big(b^{\mathrm{T}}_+(y)\vee b^{\mathrm{C}}_+(y)\big)\Big]^{k-K}<\infty$$

where $c_{\boldsymbol{\theta}}$ is a finite constant, which concludes the proof of the lemma.

Appendix C. Proof of Theorem 2. The two cases of Theorem 2 are proved separately as follows:

a) When $\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\lambda=0}$, the joint probability of the observations when the system is initialised with its stationary distribution is characterised by

$$\bar{\boldsymbol{P}}_{\boldsymbol{\theta}}(B)=\int\mathbf{1}_B(\boldsymbol{y}_{1:n})\prod_{t=1}^n\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}_t\,|\,\boldsymbol{x}_t)\prod_{i=1}^K\bigg[\pi_\theta(\boldsymbol{x}_{0,i})\prod_{t=1}^n f_\theta(\boldsymbol{x}_{t,i}\,|\,\boldsymbol{x}_{t-1,i})\bigg]\mathrm{d}\boldsymbol{y}_{1:n}\mathrm{d}\boldsymbol{x}_{0:n},$$

for any measurable subset $B=B_1\times\cdots\times B_n$ of $(\mathbb{Y}^\times)^n$ with

$$\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{x})\doteq\sum_{\substack{\boldsymbol{d}\in\{0,1\}^K\\|\boldsymbol{d}|=m}}\bigg[\sum_{\sigma\in\mathrm{Sym}(m)}\prod_{i=1}^m g_\theta\big(\boldsymbol{y}_{\sigma(i)}\,|\,\boldsymbol{x}_{r(i)}\big)u_m(\sigma)\boldsymbol{q}_{\boldsymbol{\theta}}(\boldsymbol{d})\bigg].$$

for any $m \in \mathbb{N}_0$ and any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{X}^K \times \mathbb{Y}^m$. Assuming that B_t is a measurable subset of $\mathbb{Y}^K$ of the form $A_t \times \cdots \times A_t$ for any $1 \leq t \leq n$, then the sum over $\boldsymbol{d}$ collapses to a single term where all targets are detected and all the terms in the sum over σ are equal, so that $w^1_{\boldsymbol{\theta}} \doteq \bar{\boldsymbol{P}}_{\boldsymbol{\theta}}(B)$ with

$$\bar{\boldsymbol{P}}_{\boldsymbol{\theta}}(B) = p_{\mathrm{D}}^{Kn} \int \prod_{i=1}^{K} \Big[\pi_\theta(\boldsymbol{x}_0) \prod_{t=1}^{n} \Big[\mathbf{1}_{A_t}(\boldsymbol{y}_{t,i}) g_\theta\big(\boldsymbol{y}_{t,i} \,|\, \boldsymbol{x}_{t,i}\big) f_\theta(\boldsymbol{x}_{t,i} \,|\, \boldsymbol{x}_{t-1,i}) \Big] \Big] \mathrm{d}\boldsymbol{y}_{1:n} \mathrm{d}\boldsymbol{x}_{0:n}.$$

A second case that can be considered is when B_t represents the configuration where there are $m \leq K$ observations without considering their locations for all $1 \leq t \leq n$, i.e. $B_t = \mathbb{Y} \times \cdots \times \mathbb{Y}$, in which case it holds that $w^{2,m}_{\boldsymbol{\theta}} \doteq \bar{\boldsymbol{P}}_{\boldsymbol{\theta}}(B) = \big(\mathrm{Bi}^K_{p_{\mathrm{D}}}(m)\big)^n$ If $(K, p_{\mathrm{D}}) \neq (K^*, p^*_{\mathrm{D}})$ then we can show that $w^{2,K}_{\boldsymbol{\theta}} = w^{2,K}_{\boldsymbol{\theta}^*}$ and $w^{2,K-1}_{\boldsymbol{\theta}} = w^{2,K-1}_{\boldsymbol{\theta}^*}$ cannot hold at the same time for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\lambda=0}$. Alternatively, if $(K, p_{\mathrm{D}}) = (K^*, p^*_{\mathrm{D}})$ then $w^1_{\boldsymbol{\theta}} \neq w^1_{\boldsymbol{\theta}^*}$ follows easily from the identifiability of θ^*. These two cases considered together show that the distributions associated to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ differ in some subset of the multi-target observation space so that $\bar{\boldsymbol{P}}_{\boldsymbol{\theta}} \neq \bar{\boldsymbol{P}}_{\boldsymbol{\theta}^*}$.

b) When $\boldsymbol{\theta} \in \boldsymbol{\Theta}|_{K=1}$, the multi-target likelihood becomes

$$\boldsymbol{g}_{\boldsymbol{\theta}}(\boldsymbol{y} \,|\, x) = (1 - p_{\mathrm{D}}) \mathrm{Po}_\lambda(m) \prod_{i=1}^{m} p_{\psi^*}(\boldsymbol{y}_i) + \frac{p_{\mathrm{D}}}{m} \sum_{i=1}^{m} g_{\theta^*}\big(\boldsymbol{y}_i \,|\, x\big) \mathrm{Po}_\lambda(m-1) \prod_{\substack{1 \leq j \leq m \\ j \neq i}} p_{\psi^*}(\boldsymbol{y}_j),$$

for any $m \in \mathbb{N}_0$ and any $(x, \boldsymbol{y}) \in \mathbb{X} \times \mathbb{Y}^m$. Marginalising over the location of the observations at each time step and considering the case where there are m observations, i.e. $B_t = \mathbb{Y} \times \cdots \times \mathbb{Y}$, gives $w^m_{\boldsymbol{\theta}} \doteq \bar{\boldsymbol{P}}_{\boldsymbol{\theta}}(B) = (1 - p_{\mathrm{D}}) \mathrm{Po}_\lambda(m) + p_{\mathrm{D}} \mathrm{Po}_\lambda(m-1)$, $m \geq 1$ and $w^0_{\boldsymbol{\theta}} \doteq (1 - p_{\mathrm{D}}) e^{-\lambda}$. Assuming that $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ and considering that $(p_{\mathrm{D}}, \lambda) \neq (p^*_{\mathrm{D}}, \lambda^*)$ it follows that $w^0_{\boldsymbol{\theta}} = w^0_{\boldsymbol{\theta}^*}$, $w^1_{\boldsymbol{\theta}} = w^1_{\boldsymbol{\theta}^*}$ and $w^2_{\boldsymbol{\theta}} = w^2_{\boldsymbol{\theta}^*}$ cannot all hold at the same time, which concludes the proof.

Appendix D. Proof of Theorem 4.

LEMMA 6. *For given integers m and K, let $\boldsymbol{P}_\theta$ be a family of probability measures on $\mathbb{Y}^{mK}$ indexed by $\theta \in \Theta$ and let $\boldsymbol{p}_\theta$ denote the corresponding probability density w.r.t. a common reference measure, for all θ, on $\mathbb{Y}^{mK}$. Assume that $\boldsymbol{p}_\theta\,(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m) > 0$ for any θ and $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$.*

For any integers $\alpha \geq 1$ and $\beta \geq 0$, let the random vectors $(\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m)$ be conditionally independent given $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)$, with law $\boldsymbol{P}^{\alpha,\beta}_{\boldsymbol{Y}'_{1:m}|\boldsymbol{Y}_{1:m}} = \boldsymbol{P}^{\alpha,\beta}_{\boldsymbol{Y}'_1|\boldsymbol{Y}_1} \cdots \boldsymbol{P}^{\alpha,\beta}_{\boldsymbol{Y}'_m|\boldsymbol{Y}_m}$ and each $\boldsymbol{P}^{\alpha,\beta}_{\boldsymbol{Y}'_i|\boldsymbol{Y}_i}$ is defined as in (11) via a process of thinning, augmentation with clutter with density p_ψ on $\mathbb{Y}$ and random permutation. Assume $p_\psi > 0$.

1. *Consider any θ and (α, β) such that $\alpha > 1$ or $\beta > 0$. If $f(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m) = \mathbb{E}[f(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m) | \boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m]$ then $f(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)$ is constant a.s..*
2. *Let the probability measure of $(\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m)$ be $\boldsymbol{P}^{\alpha,\beta}_\theta$ and its corresponding probability density be $\boldsymbol{p}^{\alpha,\beta}_\theta$. Assume that the densities $\boldsymbol{p}_\theta$ and $\boldsymbol{p}^{\alpha,\beta}_\theta$ are differentiable w.r.t. θ then*

$$\begin{aligned} (24) \quad & \mathbb{E}[\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m) \cdot \nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)^{\mathrm{t}}] \\ & \qquad \geq \mathbb{E}[\nabla_\theta \log \boldsymbol{p}^{\alpha,\beta}_\theta(\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m) \cdot \nabla_\theta \log \boldsymbol{p}^{\alpha,\beta}_\theta(\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m)^{\mathrm{t}}], \end{aligned}$$

with the inequality being strict if and only if $\alpha > 1$ or $\beta > 0$ and if the l.h.s. is strictly greater than 0.

The observation model (11) does not imply the equality of the gradients of $\log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{Y}', \boldsymbol{Y})$ and $\log \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{Y})$ w.r.t. $\boldsymbol{\theta}$ since $\boldsymbol{P}_\theta(\mathrm{d}\boldsymbol{Y}' \,|\, \boldsymbol{Y})$ depends on p_{D}, λ and ψ which are parameters included in $\boldsymbol{\theta}$. The interest is however in the information loss w.r.t. θ so that the result of Lemma 6 is satisfying.

Proof of Theorem 4. The considered perturbed observation model has the same properties as the one in [3], i.e. that $\nabla_\theta \log \boldsymbol{p}^{\alpha,\beta}_\theta(\boldsymbol{Y}, \boldsymbol{Y}^{\alpha,\beta}) = \nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y})$ holds a.s.. The result of [3, Lemma 3] and [3, Remark 9] can therefore be used directly in the context of interest to give, for any integer $m \geq 1$, $\boldsymbol{I}^{\alpha,\beta}_{\mathrm{loss}}(\theta^*) = \bar{\mathbb{E}}_{\theta^*}[\boldsymbol{I}^{(m)}_{\boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}}(\theta^*)]$, where

$$\begin{aligned}
\boldsymbol{I}^{(m)}_{\boldsymbol{Y}_{\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}}(\theta^*) =& \frac{1}{m}\bar{\mathbb{E}}_{\theta^*}\Big[\nabla_\theta \log \boldsymbol{p}_{\theta^*}\big(\boldsymbol{Y}_{0:m-1} \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\big) \cdot \\
& \qquad \nabla_\theta \log \boldsymbol{p}_{\theta^*}\big(\boldsymbol{Y}_{0:m-1} \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\big)^{\mathrm{t}} \,\big|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\Big] \\
& - \frac{1}{m}\bar{\mathbb{E}}_{\theta^*}\Big[\nabla_\theta \log \boldsymbol{p}_{\theta^*}\big(\boldsymbol{Y}^{\alpha,\beta}_{0:m-1} \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\big) \cdot \\
& \qquad \nabla_\theta \log \boldsymbol{p}_{\theta^*}\big(\boldsymbol{Y}^{\alpha,\beta}_{0:m-1} \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\big)^{\mathrm{t}} \,\big|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}\Big].
\end{aligned}$$

The objective is to prove that

$$\bar{\mathbb{E}}_{\theta^*}\Big[\boldsymbol{I}^{(m)}_{\boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}}(\theta^*)\Big] = 0 \tag{26}$$

for all $m \geq 1$ implies that $I(\theta^*) = 0$. From Lemma 6 applied to the involved conditional laws, (26) implies that $\nabla_\theta \log \boldsymbol{p}_{\theta^*}(\boldsymbol{Y}_{0:m-1} \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}^{\alpha,\beta}_{m:\infty}) = 0$ a.s. for almost all $\boldsymbol{Y}_{-\infty:-1}$ and almost all $\boldsymbol{Y}^{\alpha,\beta}_{m:\infty}$. Following the same principle as in [3, Lemma 4], it follows that if (26) holds for all $m \geq 1$ then $\nabla_\theta \log \boldsymbol{p}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}) = 0$ a.s., which in turn implies that $I(\theta^*) = 0$. ◻

Proof of Part 1 of Lemma 6. The proof of the first part of Lemma 6 is lengthy so only the case where $m = 1$ is given below (and it serves as a proof sketch for $m > 1$.) For $m = 1$, the statement of Part 1 of Lemma 6 will read as follows once we drop the subscript "1": if $f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y})|\boldsymbol{Y}']$ then $f(\boldsymbol{Y})$ is constant a.s.[1]

I When $\alpha = 1$ and $\beta > 0$, which corresponds to no random permutation but only random thinning, the result follows from Lemma 7 (which main assumption is satisfied because of the positivity of the density $\boldsymbol{p}_\theta$).

II When $\alpha > 1$ and $\beta = 0$, which corresponds to no random thinning but only random permutation, the result follows from Corollary 9.

III When $\alpha \geq 1$ and $\beta > 0$, i.e. both random thinning and random permutation are present.

  i Notice that $\sigma(\boldsymbol{Y}') \subseteq \sigma(\varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}})$; see (11) for the mapping from $(\varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}})$ to $\boldsymbol{Y}'$. Thus the σ-algebra generated by $\boldsymbol{Y}'$ is coarser than the one generated by $(\varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}})$. The fact that $f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y}) \,|\, \boldsymbol{Y}']$ a.s. implies that $f(\boldsymbol{Y})$ is also $\sigma(\varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}})$ measurable or equivalently

$$f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y}) \,|\, \varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}}] \quad a.s.,$$

  ii Since $\boldsymbol{Y}$ is independent of $\hat{\boldsymbol{Y}}$ and ς, it holds that

$$\mathbb{E}[f(\boldsymbol{Y}) \,|\, \varsigma, \boldsymbol{Y}_D, \hat{\boldsymbol{Y}}] = \mathbb{E}[f(\boldsymbol{Y}) \,|\, \boldsymbol{Y}_D] \quad a.s.,$$

[1] Full details can be obtained from the authors upon request

which implies that $f(\boldsymbol{Y})$ is also $\sigma(\boldsymbol{Y}_D)$ measurable and thus is constant a.s. by Lemma 7.

The rest of the proof is concerned with the second part of Lemma 6.
Let $\boldsymbol{Y}$ be the K measurements of K targets, $\hat{\boldsymbol{Y}}$ be the clutter, ς the random permutation, $\boldsymbol{D}$ the K dimensional vector of deletions. Let $\boldsymbol{Y}' = S_\varsigma((R_{\boldsymbol{D}}\boldsymbol{Y}) \oplus \hat{\boldsymbol{Y}})$. The missing target generated observations are $\boldsymbol{Y}_\mathrm{m} = R_{\boldsymbol{D}'}\boldsymbol{Y}$ where $\boldsymbol{D}' = \boldsymbol{1} - \boldsymbol{D}$. Since the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ does not have a density w.r.t. the Lebesgue measure, the proof of loss of information has to rely on a reparametrisation from $(\boldsymbol{Y}, \boldsymbol{Y}')$ to $(\boldsymbol{Y}', \varsigma, \boldsymbol{D}, \boldsymbol{Y}_\mathrm{m})$.

Let $\boldsymbol{p}_{\boldsymbol{Y}',\boldsymbol{Y}_\mathrm{m},\boldsymbol{D},\varsigma}(\boldsymbol{y}', \boldsymbol{y}_m, \boldsymbol{d}, \sigma)$ denote the joint pdf/pmf of $(\boldsymbol{Y}', \boldsymbol{Y}_\mathrm{m}, \boldsymbol{D}, \varsigma)$ that depends implicitly on θ. Using the change of variable formula, noting that $(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = F(\boldsymbol{Y}', \boldsymbol{Y}_\mathrm{m}, \boldsymbol{D}, \varsigma)$ where the mapping $F(\cdot, \cdot, \boldsymbol{d}, \sigma)$ is a permutation of $(\boldsymbol{Y}', \boldsymbol{Y}_\mathrm{m})$ for any given $\boldsymbol{d}$ and σ and hence the Jacobian of the transformation has determinant 1, it follows that $\boldsymbol{p}_{\boldsymbol{Y}',\boldsymbol{Y}_\mathrm{m},\boldsymbol{D},\varsigma}(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma) = \boldsymbol{p}_{\boldsymbol{Y},\hat{\boldsymbol{Y}},\boldsymbol{D},\varsigma}(F(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma), \boldsymbol{d}, \sigma)$ holds for $\boldsymbol{y}' = S_\sigma((R_{\boldsymbol{d}}\boldsymbol{y}) \oplus \hat{\boldsymbol{y}})$, $\boldsymbol{y}_\mathrm{m} = R_{\boldsymbol{d}'}\boldsymbol{y}$ where $\boldsymbol{d}' = \boldsymbol{1} - d$. Since only the law of $\boldsymbol{Y}$ depends on θ, it holds that $\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y},\hat{\boldsymbol{Y}},\boldsymbol{D},\varsigma}(\boldsymbol{y}, \hat{\boldsymbol{y}}, \boldsymbol{d}, \sigma) = \nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}}(\boldsymbol{y})$ so that

$$\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}',\boldsymbol{Y}_\mathrm{m},\boldsymbol{D},\varsigma}(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma) = \nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}}(F_\mathrm{T}(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma))$$

where $F_\mathrm{T}(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma)$ is the projection of $F(\boldsymbol{y}', \boldsymbol{y}_\mathrm{m}, \boldsymbol{d}, \sigma)$ on the coordinates describing $\boldsymbol{Y}$. It follows that $\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}',\boldsymbol{Y}_\mathrm{m},\boldsymbol{D},\varsigma}(\boldsymbol{Y}', \boldsymbol{Y}_\mathrm{m}, \boldsymbol{D}, \varsigma) = \nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}}(\boldsymbol{Y})$ a.s.. Let $\boldsymbol{p}_{\boldsymbol{Y}'}$ denote the density of $\boldsymbol{Y}'$. Then, using the Fisher identity, it follows that

$$\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}'}(\boldsymbol{Y}') = \mathbb{E}_\theta[\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}',\boldsymbol{Y}_\mathrm{m},\boldsymbol{D},\varsigma}(\boldsymbol{Y}', \boldsymbol{Y}_\mathrm{m}, \boldsymbol{D}, \varsigma)|\boldsymbol{Y}'] = \mathbb{E}_{\boldsymbol{P}_\theta}[\nabla_\theta \log \boldsymbol{p}_{\boldsymbol{Y}}(\boldsymbol{Y})|\boldsymbol{Y}']. \tag{27}$$

Applying (27) to the joint random variables $\boldsymbol{Y}_{1:m}$ and $\boldsymbol{Y}'_{1:m}$ defined in the lemma, it follows that $\nabla_\theta \log \boldsymbol{p}_\theta^{\alpha,\beta}(\boldsymbol{Y}'_{1:m}) = \mathbb{E}_{\boldsymbol{P}_\theta}\left[\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) \,|\, \boldsymbol{Y}'_{1:m}\right]$. Let $v \in \mathbb{R}^{d_\Theta}$, then Jensen's inequality applied to the function $x \mapsto x^2$ and to the random variable $v^\mathrm{t}\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})$ yields

$$\mathbb{E}_{\boldsymbol{P}_\theta}\left[v^\mathrm{t}\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) \,|\, \boldsymbol{Y}'_{1:m}\right]^2 \le \mathbb{E}_{\boldsymbol{P}_\theta}\left[\left(v^\mathrm{t}\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})\right)^2 \,|\, \boldsymbol{Y}'_{1:m}\right]$$

a.s., so that

$$\begin{aligned} &v^\mathrm{t}\mathbb{E}_{\boldsymbol{P}_\theta^{\alpha,\beta}}[\nabla_\theta \log \boldsymbol{p}_\theta^{\alpha,\beta}(\boldsymbol{Y}_{1:m}) \cdot \nabla_\theta \log \boldsymbol{p}_\theta^{\alpha,\beta}(\boldsymbol{Y}_{1:m})^\mathrm{t}]v \\ &\qquad\qquad \le v^\mathrm{t}\mathbb{E}_{\boldsymbol{P}_\theta}[\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) \cdot \nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})^\mathrm{t}]v \end{aligned}$$

which proves (24). Since Jensen's inequality has been applied to a strictly convex function, the case of equality:

$$\begin{aligned} &v^\mathrm{t}\mathbb{E}_{\boldsymbol{P}_\theta^{\alpha,\beta}}[\nabla_\theta \log \boldsymbol{p}_\theta^{\alpha,\beta}(\boldsymbol{Y}_{1:m}) \cdot \nabla_\theta \log \boldsymbol{p}_\theta^{\alpha,\beta}(\boldsymbol{Y}_{1:m})^\mathrm{t}]v \\ &\qquad\qquad = v^\mathrm{t}\mathbb{E}_{\boldsymbol{P}_\theta}[\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) \cdot \nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})^\mathrm{t}]v, \end{aligned}$$

holds if and only if, for all $v \in \mathbb{R}^{d_\Theta}$, $v^\mathrm{t}\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) = \mathbb{E}[v^\mathrm{t}\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m}) \,|\, \boldsymbol{Y}'_{1:m}]$ is $\sigma(\boldsymbol{Y}'_{1:m})$-measurable. Part 1 of the lemma yields that $\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})$ is $\sigma(\boldsymbol{Y}'_{1:m})$-measurable if and only if it is constant a.s.. Given that $\mathbb{E}_{\boldsymbol{P}_\theta}[\nabla_\theta \log \boldsymbol{p}_\theta(\boldsymbol{Y}_{1:m})] = 0$ it follows that the function itself is equal to 0 since it is constant, hence proving the lemma. ∎

Lemma 7 (Multiple deletions for $K > 2$.). *Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ be a random vector, $D \subseteq \{1, \ldots, K\}$ and $\boldsymbol{Y}_D$ denote the thinned version where components not in*

D have been removed. Assume $0 < \mathbb{P}(D = \sigma) < 1$ for all subsets $\sigma \subset \{1, \dots, K\}$ such that $|\sigma| = K - 1$. Furthermore, assume the following:

- *For each $i, j \in \{1, \dots, K\}$, $i \neq j$, let $Z \subseteq Y_{1:K\setminus\{i,j\}}$. If $\mathbb{P}((Y_i, Z) \in A) < 1$ and $\mathbb{I}_A(Y_i, Z) \neq \mathbb{E}[\mathbb{I}_A(Y_i, Z)|Z]$ then $\mathbb{P}((Y_i, Z) \in A|(Y_j, Z) \in A) < 1$. (Here $f \neq g$ means $\mathbb{P}(f \neq g) > 0$.)*

Then $f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y})|\boldsymbol{Y}_D]$ implies $f(\boldsymbol{Y}) = c$ a.s. for some constant c.

The main assumption of Lemma 7 is satisfied if $\nu_1 \times \cdots \times \nu_K \ll \mathbb{P}_{\boldsymbol{Y}} \ll \nu_1 \times \cdots \times \nu_K$ where ν_i are probability measures, i.e. $\mathbb{P}_{\boldsymbol{Y}}$ and $\nu_1 \times \cdots \times \nu_K$ are mutually absolutely continuous.

Proof of Lemma 7. The random variable $\boldsymbol{Y}_D$ belongs to $\mathbb{Y}^\times$, that is to the disjoint union $\cup_{k=0}^K \mathbb{Y}^k$ with $\mathbb{Y}^0 \equiv \emptyset$. Thus we can write

$$0 = \mathbb{E}[|f(\boldsymbol{Y}) - g(\boldsymbol{Y}_D)|] = \sum_{i=0}^{K} \sum_{\sigma:|\sigma|=i} \mathbb{E}[|f(\boldsymbol{Y}) - g_i(\boldsymbol{Y}_\sigma)|]\mathbb{P}(D = \sigma)$$

where g_0 is a constant, $g_i : \mathbb{Y}^i \to \mathbb{R}$ are measurable functions and independence of D and $\boldsymbol{Y}$ has been invoked. If $\mathbb{P}(D = \emptyset) > 0$ then it is trivial since this implies $\mathbb{E}[|f(\boldsymbol{Y}) - g_0|] = 0$. So assume $\mathbb{P}(D = \emptyset) = 0$. Having assumed $0 < \mathbb{P}(D = \sigma) < 1$ for all subsets $\sigma \subset \{1, \dots, K\}$ such that $|\sigma| = K - 1$, we focus on these terms only: i.e. $\sum_{\sigma:|\sigma|=K-1} \mathbb{E}[|f(\boldsymbol{Y}) - g_{K-1}(\boldsymbol{Y}_\sigma)|]\mathbb{P}(D = \sigma)$ which also implies

$$g_{K-1}(\boldsymbol{Y}_\sigma) = g_{K-1}(\boldsymbol{Y}_{\sigma'}) \quad \text{or} \quad \mathbb{I}_A(\boldsymbol{Y}_\sigma) = \mathbb{I}_A(\boldsymbol{Y}_{\sigma'}) \quad \text{a.s.} \tag{29}$$

for all σ, σ' and $A = g_{K-1}^{-1}(B)$ for a measurable set B in $\mathbb{R}$. For example, when $\sigma = (1, 3, \dots, K)$, $\sigma = (2, 3, \dots, K)$ and $Z = (Y_3, \dots, Y_K)$, we get

$$\mathbb{P}((Y_1, Z) \in A) = \mathbb{P}((Y_1, Z) \in A, (Y_2, Z) \in A) = \mathbb{P}((Y_2, Z) \in A).$$

Henceforth we refer to g_{K-1} simply as g. We need to show that $g(\boldsymbol{Y}_\sigma) = c$, for some constant c, a.s.. If this is not the case then there exists subsets of variables $Y_i \in \boldsymbol{Y}_\sigma$, $Z \subset \boldsymbol{Y}_\sigma$ and $Y_i \notin Z$ (recall $\sigma \subset \{1, \dots, K\}$ with $|\sigma| = K - 1$) such that

$$g(\boldsymbol{Y}_\sigma) = \mathbb{E}[g(\boldsymbol{Y}_\sigma)|Y_i, Z] \quad \text{a.s. and} \quad g(\boldsymbol{Y}_\sigma) \neq \mathbb{E}[g(\boldsymbol{Y}_\sigma)|Z] \quad \text{a.s.} \tag{30}$$

The interpretation is that $g(\boldsymbol{Y}_\sigma)$ can potentially be a function of the reduced set of variables (Y_i, Z) (as asserted by the first equality) but it must genuinely be a function of at least the variable Y_i. For clarity and simplicity assume $i = 1$ and $Z = (Y_3, \dots, Y_K)$. Consider the terms in the sum due to $\sigma = (1, 3, \dots, K)$ and $\sigma = (2, 3, \dots, K)$. We assume that there exists a measurable set $A = g^{-1}(B)$ such that $0 < \mathbb{P}(A) < 1$ and $\mathbb{I}_A(Y_1, Z) \neq \mathbb{E}[\mathbb{I}_A(Y_1, Z)|Z]$. But (29) implies $\mathbb{P}((Y_1, Z) \in A|(Y_2, Z) \in A) = 1$ which violates the main assumption of the lemma. ∎

LEMMA 8 (Randomly permuting a random vector.). *Let $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ and $\hat{\boldsymbol{Y}} = (Y_{n+1}, \dots, Y_{n+m})$. Let ς denote the randomised permutation which is independent of $(\boldsymbol{Y}, \hat{\boldsymbol{Y}})$ and let $\mathbf{Z} = (Y_{\varsigma(1)}, \dots, Y_{\varsigma(m+n)})$. Assume ς permits, at the least, the exchange of any two indices, i.e. $\mathbb{P}(\varsigma(i) = j, \varsigma(j) = i, \{\varsigma(k) = k : k \neq i, j\}) > 0$ for all i, j. Furthermore, $\mathbb{P}(\varsigma = (1, \dots, n)) > 0$. Assume the law of $(\boldsymbol{Y}, \hat{\boldsymbol{Y}})$ satisfies $\nu^{n+m} \gg \mathbb{P}_{\boldsymbol{Y},\hat{\boldsymbol{Y}}} \gg \nu^{n+m}$ where ν is some probability measure and ν^{n+m} the product probability measure on $\mathbb{Y}^{n+m}$. If $f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y})|\mathbf{Z}]$, then $f(\boldsymbol{Y})$ is a constant a.s..*

Proof of Lemma 8. The proof is completed for the case $m = 1$ and easily generalised to $m > 1$. Let $g(\mathbf{Z}) = \mathbb{E}[f(\boldsymbol{Y})|\mathbf{Z}]$. For any σ such that $\mathbb{P}(\varsigma = \sigma) > 0$,

$$\mathbb{E}[|f(\boldsymbol{Y}) - g(Y_{\sigma(1)}, \ldots, Y_{\sigma(n+1)})|\mathbb{I}_{[\varsigma=\sigma]}] = 0.$$

Since ς is independent of $(\boldsymbol{Y}, \hat{\boldsymbol{Y}})$, we have $\mathbb{E}[|f(\boldsymbol{Y}) - g(Y_{\sigma(1)}, \ldots, Y_{\sigma(n+1)})|] = 0$ and thus $g(Y_{\sigma(1)}, \ldots, Y_{\sigma(n+1)}) = g(Y_{\sigma'(1)}, \ldots, Y_{\sigma'(n+1)})$ a.s. for any other σ' such that $\mathbb{P}(\varsigma = \sigma') > 0$. To present the arguments we employ in the clearest way, we consider the case $(n = 2, m = 1)$. The preceding statements imply, $\mathbb{P}_{\boldsymbol{Y},\hat{\boldsymbol{Y}}}$ almost everywhere (and hence ν^3 almost everywhere),

$$g(Y_1, Y_2, Y_3) = f(Y_1, Y_2) \tag{31}$$

$$g(Y_1, Y_2, Y_3) = g(Y_3, Y_2, Y_1) \tag{32}$$

$$g(Y_1, Y_2, Y_3) = g(Y_1, Y_3, Y_2). \tag{33}$$

We will show that the further implication

$$f(Y_1, Y_2) = f(Y_3, Y_2) = f(Y_1, Y_3) \tag{34}$$

holds ν^3 almost everywhere may be derived. Once this is done, to complete the proof, we will further manipulate (34) under the assumption that the random variables Y_i are independently and identically distributed with respect to measure ν to show that $f = c$, for some constant c, ν^3 almost everywhere. From the first equality of (34),

$$f(Y_1, Y_2) = \mathbb{E}_{\nu^3}(f(Y_3, Y_2)|Y_1, Y_2) = \mathbb{E}_{\nu^3}(f(Y_3, Y_2)|Y_2) = h(Y_2)$$

for some function h. That is $f(Y_1, Y_2)$ collapses to a function of variable Y_2 only, which is denoted by $h(Y_2)$. Using the second equality of (34), $h(Y_2) = f(Y_1, Y_3)$ and thus it must be that h is a constant as Y_i are independent. We now verify (31)-(32) implies $f(Y_1, Y_2) = f(Y_3, Y_2)$ of (34). We have $\mathbb{E}_{\nu^3}[|f(Y_1, Y_2) - g(Y_3, Y_2, Y_1)|] = 0$ and a change of variable gives $\mathbb{E}_{\nu^3}[|f(Y_3, Y_2) - g(Y_1, Y_2, Y_3)|] = 0$. The same procedure applied to (31)-(33) shows the second equality of (34). ∎

Corollary 9 extends Lemma 8 to the situation when $\hat{\boldsymbol{Y}}$ therein follows the law of a clutter process as defined in (11).

COROLLARY 9. *Let $(\hat{Y}_1, \hat{Y}_2, \ldots)$ be an infinite sequence of independent $\mathbb{Y}$-valued random variables with $\hat{Y}_i \sim \hat{P}$. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ be a vector of $\mathbb{Y}$-valued random variables which is independent of $(\hat{Y}_1, \hat{Y}_2, \ldots)$. Let $\hat{M} \in \mathbb{N}_0$ be non-negative random variable independent of $(\boldsymbol{Y}, \hat{Y}_{1:\infty})$. Let $\mathbf{Z} = S_\varsigma(\boldsymbol{Y} \oplus \hat{\boldsymbol{Y}})$ where $\hat{\boldsymbol{Y}} = \hat{Y}_{1:\hat{M}}$ and S_ς is the random permutation matrix defined as in 11. Assume $(\hat{P})^K \gg \mathbb{P}_{\boldsymbol{Y}} \gg (\hat{P})^K$. If $f(\boldsymbol{Y}) = \mathbb{E}[f(\boldsymbol{Y})|\mathbf{Z}]$, then $f(\boldsymbol{Y})$ is a constant a.s..*

Proof. Let $g(\mathbf{Z}) = \mathbb{E}[f(\boldsymbol{Y})|\mathbf{Z}]$ then $\mathbb{E}[|f(\boldsymbol{Y}) - g(\mathbf{Z})||\hat{M} = m] = 0$ for all m such that $\mathbb{P}(\hat{M} = m) > 0$. Since $\hat{M}$ is independent of $(\boldsymbol{Y}, \hat{Y}_{1:\infty})$ and the random permutation matrix is itself independent of $(\boldsymbol{Y}, \hat{Y}_{1:\infty})$ given $\hat{M} = m$, the law of $(\boldsymbol{Y}, \mathbf{Z})$ conditioned on $\hat{M} = m$ satisfies the assumptions of Lemma 8. Thus, by Lemma 8, $\mathbb{E}[|f(\boldsymbol{Y}) - g(\mathbf{Z})||\hat{M} = m] = 0$ implies $f(\boldsymbol{Y}) = c_m$ a.s. for some constant c_m. (It is clear that c_m is independent of m.) ∎

Appendix E. Proof of Theorem 5. The case where $K^* = 1$ is first considered so that the number of observations M_t at time t can only be equal to zero or one.

The joint probability of the observations and states becomes

$$\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n}, x_{0:n}) = \pi_\theta(x_0) \prod_{t=1}^{n} \big[(1-p_{\mathrm{D}})^{1-M_t}(p_{\mathrm{D}} g_\theta(\boldsymbol{y}_t \,|\, x_t))^{M_t} f_\theta(x_t \,|\, x_{t-1})\big],$$

where $\boldsymbol{y}_t$ is the empty sequence when $M_t = 0$. The size of $\boldsymbol{y}_t$ at any time t can be made explicit in this expression for the sake of clarity as follows

$$\bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1:n}, x_{0:n}, m_{1:n}) = \pi_\theta(x_0) \prod_{t=1}^{n} \big[(1-p_{\mathrm{D}})^{1-m_t}(p_{\mathrm{D}} g_\theta(\boldsymbol{y}_t \,|\, x_t))^{m_t} f_\theta(x_t \,|\, x_{t-1})\big].$$

Let Y_t^ϵ be a noisy version of the original observation Y_t for any $t \geq 1$ so that the HMM $(X_t, Y_t^\epsilon)_t$ is equal in law to the HMM $(X_t, Y_t + \epsilon Z_t)_t$ where $(Z_t)_t$ is an i.i.d. sequence of random variables which common law is the uniform distribution over the ball of radius 1 and centre 0. A switching process $(s_t)_t$ is also introduced as follows: $s_t = 1$ when the target is detected and $s_t = 0$ otherwise. In order to study the Fisher information more easily, we introduce an alternative observation model where a detection failure at time t is replaced by an observation Y_t^ϵ from the target. The law of this observation model is

$$\bar{p}_\theta^\epsilon(\tilde{y}_{1:n}, x_{0:n}, s_{1:n}) = \pi_\theta(x_0) \prod_{t=1}^{n} \Big[[p_{\mathrm{D}} g_\theta(y_t \,|\, x_t)]^{s_t}[(1-p_{\mathrm{D}}) g_\theta(y_t^\epsilon \,|\, x_t)]^{1-s_t} f_\theta(x_t \,|\, x_{t-1})\Big]$$

where $\tilde{y}_t = y_t$ if $s_t = 1$ and $\tilde{y}_t = y_t^\epsilon$ if $s_t = 0$. The quantity of interest is

$$\begin{aligned}\bar{p}_\theta^\epsilon(\tilde{y}_0, s_0 \,|\, y_{-\infty:-1}, y_{1:\infty}^\epsilon) &= \\ &[p_{\mathrm{D}} g_\theta(y_0 \,|\, x_0)]^{s_0}[(1-p_{\mathrm{D}}) g_\theta(y_0^\epsilon \,|\, x_0)]^{1-s_0} \bar{p}_\theta(x_0 \,|\, y_{-\infty:-1}, y_{1:\infty}^\epsilon),\end{aligned}$$

which we compare with $\bar{p}_\theta^\epsilon(y_0 \,|\, y_{-\infty:-1}, y_{1:\infty}^\epsilon) = g_\theta(y_0 \,|\, x_0)\bar{p}_\theta(x_0 \,|\, y_{-\infty:-1}, y_{1:\infty}^\epsilon)$, i.e. the full-detection case. To justify the equivalence of the two observation model for the considered purpose, we can verify that the score $\nabla_\theta \log \bar{\boldsymbol{p}}_{\boldsymbol{\theta}}(\boldsymbol{y}_0, m_0 \,|\, y_{-\infty:-1})$ is equal to the score $\nabla_\theta \log \bar{p}_\theta^\epsilon(\tilde{y}_0, s_0 \,|\, y_{-\infty:-1}, y_{1:\infty}^\epsilon)$ when $\epsilon \to \infty$. With the required modifications and after [3, Theorem 5], it follows that the loss of information $I_{\mathrm{loss}}^\epsilon(\theta^*)$ when replacing the original observations by the ϵ-perturbed ones can be expressed as

$$\begin{aligned}I_{\mathrm{loss}}^\epsilon(\theta^*) = \bar{\mathbb{E}}_{\theta^*}&\Big[\nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon) \cdot \nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon)^{\mathrm{t}}\Big] \\ - p_{\mathrm{D}}\bar{\mathbb{E}}_{\theta^*}&\Big[\nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon) \cdot \nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon)^{\mathrm{t}}\Big] \\ - (1-p_{\mathrm{D}})\bar{\mathbb{E}}_{\theta^*}&\Big[\nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0^\epsilon \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon) \cdot \nabla_\theta \log \bar{p}_{\theta^*}(\boldsymbol{Y}_0^\epsilon \,|\, \boldsymbol{Y}_{-\infty:-1}, \boldsymbol{Y}_{1:\infty}^\epsilon)^{\mathrm{t}}\Big].\end{aligned}$$

Considering the limit $\epsilon \to \infty$, it follows that $I_{\mathrm{loss}}(\theta^*) \doteq \lim_{\epsilon\to\infty} I_{\mathrm{loss}}^\epsilon(\theta^*)$ verifies

$$I_{\mathrm{loss}}(\theta^*) = (1-p_{\mathrm{D}})\bar{\mathbb{E}}_{\theta^*}\Big[\nabla_\theta \log \bar{\boldsymbol{p}}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1}) \cdot \nabla_\theta \log \bar{\boldsymbol{p}}_{\theta^*}(\boldsymbol{Y}_0 \,|\, \boldsymbol{Y}_{-\infty:-1})^{\mathrm{t}}\Big].$$

In the multi-target case, it simply holds that the information loss is equal to $(1 - p_{\mathrm{D}}^*)K^* I(\theta^*)$ since targets' detection are independent when the data association is known, which terminates the proof of the proposition.

REFERENCES

[1] S. S. Blackman, *Multiple-target tracking with radar applications*, Artech House, 1986.
[2] N. Chenouard et al., *Objective comparison of particle tracking methods*, Nature methods, 11 (2014).
[3] T. A. Dean, S. S. Singh, A. Jasra, and G. W. Peters, *Parameter estimation for hidden Markov models with intractable likelihoods*, Scandinavian Journal of Statistics, 41 (2014).
[4] P. Del Moral, *Mean field simulation for Monte Carlo integration*, 2013.
[5] P. Del Moral and J. Houssineau, *Particle association measures and multiple target tracking*, in Theoretical Aspects of Spatial-Temporal Modeling, Springer, 2015, pp. 1–30.
[6] R. Douc, E. Moulines, and T. Ryden, *Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime*, The Annals of Statistics, 32 (2004).
[7] A. Doucet, B.-N. Vo, C. Andrieu, and M. Davy, *Particle filtering for multi-target tracking and sensor management*, in Proceedings of the Fifth International Conference on Information Fusion, vol. 1, IEEE, 2002, pp. 474–481.
[8] L. Frenkel and M. Feder, *Recursive expectation-maximization (EM) algorithms for time-varying parameters with applications to multiple target tracking*, IEEE T. on Signal Processing, 47 (1999), pp. 306–320.
[9] C. Hue, J.-P. Le Cadre, and P. Perez, *Posterior Cramer-Rao bounds for multi-target tracking*, IEEE T. on Aerospace and Electronic Systems, 42 (2006), pp. 37–49.
[10] L. Jiang, S. S. Singh, and S. Yildirim, *Bayesian tracking and parameter learning for non-linear multiple target tracking models*, IEEE T. on Signal Processing, 63 (2015).
[11] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, N. Chopin, et al., *On particle methods for parameter estimation in state-space models*, Statistical science, 30 (2015), pp. 328–351.
[12] B. G. Leroux, *Maximum-likelihood estimation for hidden Markov models*, Stochastic processes and their applications, 40 (1992).
[13] R. P. S. Mahler, *Multitarget Bayes filtering via first-order multitarget moments*, IEEE T. on Aerospace and Electronic Systems, 39 (2003).
[14] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer Science & Business Media, 2012.
[15] J. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, *A random-finite-set approach to Bayesian SLAM*, IEEE T. on Robotics, 27 (2011).
[16] R. Niu, P. Willett, and Y. Bar-Shalom, *Matrix CRLB scaling due to measurements of uncertain origin*, IEEE T. on Signal Processing, 49 (2001), pp. 1325–1335.
[17] S. Oh, S. Russell, and S. Sastry, *Markov chain Monte Carlo data association for multi-target tracking*, IEEE T. on Automatic Control, 54 (2009).
[18] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe, *A boosted particle filter: Multitarget detection and tracking*, in European Conference on Computer Vision, Springer, 2004, pp. 28–39.
[19] M. Pace and P. Del Moral, *Mean-field PHD filters based on generalized Feynman-Kac flow*, Journal of selected topics in signal processing, special issue on multi-target tracking, (2013).
[20] Y. Pailhas, J. Houssineau, Y. Petillot, and D. E. Clark, *Tracking with MIMO sonar systems: applications to harbour surveillance*, IET Radar, Sonar & Navigation, (2016).
[21] G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle approximations of the score and observed information matrix in state space models with application to parameter estimation*, Biometrika, 98 (2011), pp. 65–80.
[22] S. S. Singh, N. Whiteley, and S. J. Godsill, *Approximate likelihood estimation of static parameters in multi-target models*, in Bayesian time series models, 2011.
[23] C. B. Storlie, J. Hannig, T. C. Lee, et al., *Statistical consistency of the data association problem in multiple target tracking*, Electronic Journal of Statistics, 5 (2011).
[24] P. Tichavsky, C. H. Muravchik, and A. Nehorai, *Posterior Cramér-Rao bounds for discrete-time nonlinear filtering*, IEEE T. on signal processing, 46 (1998), pp. 1386–1396.
[25] B.-N. Vo and W.-K. Ma, *The Gaussian mixture probability hypothesis density filter*, IEEE T. on Signal Processing, 54 (2006).
[26] B.-N. Vo, S. Singh, and A. Doucet, *Sequential Monte Carlo methods for multitarget filtering with random finite sets*, IEEE T. on Aerospace and Electronic Systems, 41 (2005).
[27] S. Yildirim, L. Jiang, S. S. Singh, and T. A. Dean, *Calibrating the Gaussian multi-target tracking model*, Statistics and Computing, 25 (2015), pp. 595–608.
[28] X. Zhang, P. Willett, and Y. Bar-Shalom, *Dynamic Cramer-Rao bound for target tracking in clutter*, IEEE T. on Aerospace and Electronic Systems, 41 (2005), pp. 1154–1167.