Target Tracking Using a Bank of MAP Estimators

Guoquan P. Huang, and Stergios I. Roumeliotis

MARS LAB

Multiple Autonomous
Robotic Systems Labroratory

Technical Report
Number -2010-0002
August 2010

Dept. of Computer Science & Engineering
University of Minnesota
4-192 EE/CS Building
200 Union St. S.E.
Minneapolis, MN 55455
Tel: (612) 625-2217
Fax: (612) 625-0572
URL: http://www.cs.umn.edu/~ghuang

Target Tracking Using a Bank of MAP Estimators

Guoquan P. Huang, and Stergios I. Roumeliotis

Multiple Autonomous Robotic Systems Laboratory, TR-2010-0002

August 2010

Abstract

Nonlinear estimation problems, such as range-only and bearing-only target tracking, are often addressed using linearized estimators (e.g., the extended Kalman filter (EKF)). The resulting linearization errors, as well as these estimators' inability to track multimodal probability density functions (pdfs) greatly diminish their performance (accuracy and consistency). In this work, we employ relinearization, provided by the batch maximum a posteriori (MAP) estimator, to reduce linearization errors. Moreover, we introduce a general framework (i.e., a bank of MAP estimators) to track multiple modes of the posterior pdf. In particular, we design a hypothesis generation scheme that effectively focuses available resources on the most probable hypotheses of the state trajectory, by analytically computing all modes of the posterior pdf at the current time step. This is efficiently carried out by converting the nonlinear cost function into polynomial form, and subsequently employing algebraic geometry techniques to compute all its stationary points. Each local minimum for the current state is combined with an estimate of all previous states and then used as a high-quality initialization point for iterative MAP optimization, whose result is maintained as one element in a bank of MAP estimators. In addition, we employ pruning and marginalization to control the computational cost.

1 Introduction

In this paper, we introduce a general framework to improve estimation performance for problems with (nonlinear) measurement functions that can be transformed into polynomial or rational form. We apply our method to the particular problem of target tracking, i.e., estimating the kinematic state of a moving target using only range or bearing measurements from a mobile sensor (robot) whose position and orientation are known. Target tracking has attracted significant interest over the past decades, since it arises in a variety of practical applications, such as environmental monitoring [1, 2], submarine tracking with towed-array sonar [3], and aircraft surveillance using radar [4]. Examples of recent research include design of new estimation algorithms and adaptive optimal control of sensor motions [5–8].

Nonlinear estimation problems such as the aforementioned target tracking, are often addressed using linearized estimators, e.g., the extended Kalman filter (EKF) [3, 4]. These estimators suffer from linearization errors and the inability to track multimodal probability density functions (pdfs), which often arise in nonlinear estimation problems. Several methods have been proposed to reduce the effect of the linearization errors. The iterated EKF (IEKF) [4], for example, iterates the filter update till convergence, by relinearizing the measurement function at each iteration. Alternatively, the unscented Kalman filter (UKF) [9] deterministically samples the nonlinear function around the state estimate, thus improving the linear approximation. However, any (explicit or implicit) linearization-based filtering approach marginalizes all but the current state, and is hence unable to refine past linearization points. In contrast, a batch maximum a posteriori (MAP) estimator [10] computes the estimates for the states at all time steps using all available measurements. This allows continuous relinearization of the entire state trajectory, which greatly reduces the linearization errors. However, just as the IEKF and its variants, the batch-MAP estimator can only track one of the potentially many modes of the posterior pdf. Only a few estimators, such as the multi-hypothesis EKF (MHEKF) [6], and the particle filter (PF) [11], are specifically designed to treat multimodal distributions by simultaneously tracking a set of different state estimates. However, in most cases these hypotheses are generated randomly, wasting in effect a considerable portion of the computational resources.

The proposed estimation framework provides both continuous relinearization and multi-hypothesis tracking (estimation), together with a highly efficient hypothesis generation scheme. The ideal approach to the batch-MAP estimation problem would be to compute all modes of the posterior pdf, thus ensuring a globally optimal estimate. However,

as our analysis shows later, this approach is computationally intractable due to the growing size of the state vector. We therefore relax the problem, and optimize only for the current target's state at each time step, treating the state history of each hypothesis as a constant prior. To solve this subproblem, we first convert its nonlinear cost function into *polynomial* or *rational* form, and subsequently employ algebraic geometry techniques [12] to *analytically* compute *all* local minima and thus all modes of the pdf. Each mode is used to initialize a new MAP estimator in the bank, thus allowing to track the *most* probable hypotheses of the state trajectory, and in turn greatly improving the accuracy of the MAP estimate. At the same time, we achieve low, resource-adaptive computational cost through pruning and marginalization. The former controls the exponential growth of hypotheses, while the latter limits the size of the state vector. Note that in our previous conference publications [13, 14], we have successfully applied this idea to range-only and bearing-only target tracking, respectively. In this paper, we present the theoretical analysis in more detail and test the proposed algorithm more thoroughly in both Monte-Carlo simulations and real-world experiments. At this point, we also stress that apart from the particular application of target tracking treated in this paper, the proposed analytical methodology is applicable to a broad class of nonlinear estimation problems in robotics and computer vision that can be expressed in (or converted into) polynomial form (e.g., see [15–18]).

2 Related Work

The problem of target tracking has been studied for decades and many different estimators of both batch and recursive types have been proposed in the literature [4, 19]. Among the available algorithms, the EKF is one of the most widely used methods [4]. However, the EKF is unable to refine the past linearization points when new measurements are available and thus becomes vulnerable to large linearization errors, which degrade its performance. This has given rise to refinements of the EKF or the UKF; for instance, the modified polar coordinates EKF [20] and the shifted Rayleigh filter (SRF) [5], which were developed specifically for bearing-only target tracking, and the iterated UKF (IUKF) proposed for passive target tracking [21]. However, all these EKF/UKF variants can only track a single mode, or the mean, of the posterior pdf of the target state and thus suffer from the same problem as the EKF, i.e., they can potentially track an inaccurate mode of the pdf and hence become inconsistent or even diverge.

To mitigate the aforementioned issue, an MHEKF was proposed specifically for bearing-only tracking in [6] to track multiple hypotheses of the target state. The MHEKF makes an assumption about the minimum and maximum distance between the sensor and target and partitions this range interval into a number of subintervals, each representing a hypothesis regarding the true range of the target. A bank of independently operating range-parameterized EKFs are thus created, each corresponding to one of the hypotheses and receiving the same bearing measurements. In [6], the MHEKF determines a fixed number of EKFs at the first available measurement, while in [22] this idea was extended so that the filter bank can dynamically change its size at each time step based on the current measurement likelihood. Since no filter in the MHEKF can guarantee computing the globally optimal estimate (due to the multimodal nature of the distribution as well as its inability to relinearize the nonlinear measurement function), this approach can also become inconsistent and diverge. Moreover, this method assumes prior knowledge about the range interval, which may not always be available in, e.g., bearing-only target tracking. Most importantly, this approach does not provide a measurable criterion about how many partitions are needed in the assumed range interval, or where to choose them. In contrast, the proposed bank-of-MAP estimator selects the most probable hypotheses of the target trajectory based on local optimality at each time step.

Considerable attention has recently been paid to the PF for both bearing-only and range-only target tracking [11, 23–27], because of its ability of solving nonlinear estimation problems involving multimodal pdfs. In the standard (bootstrap) PF, each particle represents a hypothesis of the target state, weighted and normalized based on its measurement likelihood [25]. If the particles sample the state space adequately, the PF will converge to the true distribution. However, the particles are usually initialized randomly, and if far from a mode of the pdf, their weights can decay quickly and lead to particle depletion (i.e., only few particles having significant weights) and thus inconsistency, even if a resampling scheme (e.g., stratified resampling [28]) is employed. This is due to the fact that the very few surviving particles may not be sufficient to represent the underlying multimodal pdf. Therefore, in order to converge to meaningful estimates, the PF requires using a large number of particles, thus exacerbating its computational demands. In contrast, the proposed bank-of-MAP estimator *analytically* computes all modes of the posterior pdf at the current time step and *efficiently* focuses the computational resources on the most probable hypotheses of the target state.

3 Problem Formulation

Consider a single sensor moving in a plane and estimating the state (position, velocity, etc.) of a moving target, by processing the available range or bearing measurements. In this work, we study the case of *global* tracking, i.e., the position of the target is expressed with respect to a fixed (global) frame of reference, instead of a relative *sensor-centered* one. We hereafter assume that the pose (position and orientation) of the tracking sensor is known with high accuracy in the global frame of reference (e.g., from GPS and compass measurements). The state vector of the target at time-step k, is defined as a vector of dimension $2N$, where $N-1$ is the highest-order time derivative of the target's position described by a known stochastic motion model, and comprises states such as position, velocity, acceleration, etc.:

$$\mathbf{x}_k = \begin{bmatrix} x_{T_k} & y_{T_k} & \dot{x}_{T_k} & \dot{y}_{T_k} & \ddot{x}_{T_k} & \ddot{y}_{T_k} & \cdots \end{bmatrix}^T \tag{1}$$

$$= \begin{bmatrix} \mathbf{p}_{T_k}^T & \mathbf{d}_{T_k}^T \end{bmatrix}^T \tag{2}$$

where $\mathbf{p}_{T_k} \triangleq \begin{bmatrix} x_{T_k} & y_{T_k} \end{bmatrix}^T$ is the target's position, and $\mathbf{d}_{T_k} \triangleq \begin{bmatrix} \dot{x}_{T_k} & \dot{y}_{T_k} & \ddot{x}_{T_k} & \ddot{y}_{T_k} & \cdots \end{bmatrix}^T$ denotes all the higher-order time derivatives of the target's position. Note that in this paper we consider a generic tracking scenario where the target's trajectory is observable [22, 29, 30].

In the following, we present the target stochastic motion model and the sensor measurement models that will be used throughout the paper.

3.1 Motion Model

We consider the case where the target moves randomly but assume that a stochastic kinematic model describing its motion (e.g., constant acceleration or constant velocity [4]) is known. In particular, the discrete-time state propagation equation is generically given by the following linear form:

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

where $\mathbf{w}_{k-1}$ is zero-mean white Gaussian noise with covariance $\mathbf{Q}_{k-1}$. The state transition matrix, $\mathbf{\Phi}_{k-1}$, and the process noise Jacobian, $\mathbf{G}_{k-1}$, that appear in the preceding expression depend on the particular motion model used [4]. We will make no further assumptions on these matrices other than that their values are known.

3.2 Measurement Models

In this work, we are interested in the case in which a single sensor measures the relative distance or bearing angle to the target.[1] The corresponding measurement equations are described below.

3.2.1 Range-only measurement

The range-only measurement at time-step k is given by:

$$z_k = \sqrt{(x_{T_k} - x_{S_k})^2 + (y_{T_k} - y_{S_k})^2} + n_{\rho_k} \tag{4}$$

$$\triangleq h_\rho(\mathbf{x}_k) + n_{\rho_k} \tag{5}$$

where $\mathbf{x}_{S_k} \triangleq [\mathbf{p}_{S_k}^T \ \phi_{S_k}]^T \triangleq [x_{S_k} \ y_{S_k} \ \phi_{S_k}]^T$ is the known sensor pose expressed in the global frame of reference, and n_{ρ_k} is zero-mean white Gaussian measurement noise, with variance $\sigma_{\rho_k}^2$, i.e., $n_{\rho_k} \sim \mathcal{N}(0, \sigma_{\rho_k}^2)$.

3.2.2 Bearing-only measurement

Similarly, the bearing measurement at time-step k is given by:

$$z_k = \operatorname{atan2}\left((y_{T_k} - y_{S_k}), (x_{T_k} - x_{S_k})\right) - \phi_{S_k} + n_{\theta_k} \tag{6}$$

$$\triangleq h_\theta(\mathbf{x}_k) + n_{\theta_k} \tag{7}$$

where n_{θ_k} is zero-mean white Gaussian measurement noise, with variance $\sigma_{\theta_k}^2$, i.e., $n_{\theta_k} \sim \mathcal{N}(0, \sigma_{\theta_k}^2)$.

[1] It should be noted that although we here focus only on the single-sensor case, the methodology presented in this paper can serve as a building block for a more complex, multi-sensor case, i.e., at each time step multiple measurements may be processed sequentially using the proposed analytical approach (see Section 4) in order to find high-quality hypotheses.

3.3 Batch-MAP Estimator

The batch-MAP estimator utilizes all available information to estimate the *entire* target's trajectory that is represented by stacking all states in the time interval $[0, k]$ [see (1)]:

$$\mathbf{x}_{0:k} = \begin{bmatrix} \mathbf{x}_0^T & \mathbf{x}_1^T & \cdots & \mathbf{x}_k^T \end{bmatrix}^T \tag{8}$$

Specifically, the batch-MAP estimator seeks to determine the entire state-space trajectory estimate $\hat{\mathbf{x}}_{0:k|k}$ that maximizes the following posterior pdf:[2]

$$p(\mathbf{x}_{0:k}|z_{1:k}) \propto p(\mathbf{x}_0) \prod_{\kappa=1}^{k} p(\mathbf{x}_\kappa|\mathbf{x}_{\kappa-1}) p(z_\kappa|\mathbf{x}_\kappa) \tag{9}$$

where $p(\mathbf{x}_0) = \mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$ is the prior distribution, and $z_{1:k}$ denotes all the sensor measurements in the time interval $[1, k]$. In the above expression, we have employed the Bayes rule, by assuming that the state and measurement noise are independent and employed the Markovian property of the target motion [see (3), and (5) or (7), respectively]. Moreover, using the assumption of Gaussian noise, the above posterior pdf (9) can be written as:

$$\begin{aligned} & p(\mathbf{x}_{0:k}|z_{1:k}) \propto \\ & \frac{1}{\sqrt{(2\pi)^{2N}|\mathbf{P}_{0|0}|}} \exp\left(-\frac{1}{2}||\mathbf{x}_0 - \hat{\mathbf{x}}_{0|0}||^2_{\mathbf{P}_{0|0}}\right) \times \\ & \prod_{\kappa=1}^{k} \frac{1}{\sqrt{(2\pi)^{2N}|\mathbf{Q}'_{\kappa-1}|}} \exp\left(-\frac{1}{2}||\mathbf{x}_\kappa - \mathbf{\Phi}_{\kappa-1}\mathbf{x}_{\kappa-1}||^2_{\mathbf{Q}'_{\kappa-1}}\right) \times \\ & \prod_{\kappa=1}^{k} \frac{1}{\sqrt{2\pi\sigma_\kappa^2}} \exp\left(-\frac{1}{2}||z_\kappa - h(\mathbf{x}_\kappa)||^2_{\sigma_\kappa^2}\right) \end{aligned} \tag{10}$$

where $h(\cdot) = h_\rho(\cdot)$ and $\sigma_k = \sigma_{\rho_k}$ if range-only measurements are used, while $h(\cdot) = h_\theta(\cdot)$ and $\sigma_k = \sigma_{\theta_k}$ for bearing-only measurements. In the above expression, we have also employed the notations: $||\mathbf{a}||^2_{\mathbf{M}} \triangleq \mathbf{a}^T\mathbf{M}^{-1}\mathbf{a}$, $\mathbf{Q}'_k \triangleq \mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T$ [see (3)], and $|\mathbf{X}| \triangleq \det(\mathbf{X})$. Due to the monotonicity of the negative logarithm, the maximization of (10) is equivalent to the minimization of the following cost function:

$$c(\mathbf{x}_{0:k}) = \frac{1}{2}||\mathbf{x}_0 - \hat{\mathbf{x}}_{0|0}||^2_{\mathbf{P}_{0|0}} + \sum_{\kappa=1}^{k} \frac{1}{2}||\mathbf{x}_\kappa - \mathbf{\Phi}_{\kappa-1}\mathbf{x}_{\kappa-1}||^2_{\mathbf{Q}'_{\kappa-1}} + \sum_{\kappa=1}^{k} \frac{1}{2}||z_\kappa - h(\mathbf{x}_\kappa)||^2_{\sigma_\kappa^2} \tag{11}$$

Clearly, due to the nonlinear measurement model, $c(\mathbf{x}_{0:k})$ is a function with potentially multiple local minima. A standard approach for its optimization is to employ Newton-Raphson iterative minimization [31], which, however, is only able to converge to one local minimum within the basin of attraction of the initial estimate. The solution therefore heavily depends on the quality of the initial estimate. At the ℓ-th iteration of the Newton-Raphson method, a correction, $\delta\mathbf{x}^{(\ell)}_{0:k}$, to the current estimate, $\hat{\mathbf{x}}^{(\ell)}_{0:k|k}$, is computed by minimizing the second-order Taylor-series approximation of the cost function given by:

$$c(\hat{\mathbf{x}}^{(\ell)}_{0:k|k} + \delta\mathbf{x}^{(\ell)}_{0:k}) \simeq c(\hat{\mathbf{x}}^{(\ell)}_{0:k|k}) + \mathbf{b}^{(\ell)^T}\delta\mathbf{x}^{(\ell)}_{0:k} + \frac{1}{2}\delta\mathbf{x}^{(\ell)^T}_{0:k}\mathbf{A}^{(\ell)}\delta\mathbf{x}^{(\ell)}_{0:k} \tag{12}$$

where

$$\mathbf{b}^{(\ell)} \triangleq \nabla_{\mathbf{x}_{0:k}} c(\cdot)\Big|_{\mathbf{x}_{0:k}=\hat{\mathbf{x}}^{(\ell)}_{0:k|k}} \tag{13}$$

$$\mathbf{A}^{(\ell)} \triangleq \nabla^2_{\mathbf{x}_{0:k}} c(\cdot)\Big|_{\mathbf{x}_{0:k}=\hat{\mathbf{x}}^{(\ell)}_{0:k|k}} \tag{14}$$

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

are the Jacobian and Hessian of $c(\cdot)$ with respect to $\mathbf{x}_{0:k}$, evaluated at the latest state estimates, $\hat{\mathbf{x}}_{0:k|k}^{(\ell)}$, respectively.

We now examine the structure of the Jacobian and Hessian matrices that will be useful for the ensuing analysis. Specifically, at the ℓ-th iteration, $\mathbf{b}^{(\ell)}$ can be obtained as:

$$\begin{aligned}\mathbf{b}^{(\ell)} = \mathbf{\Pi}^T \mathbf{P}_{0|0}^{-1} \left(\hat{\mathbf{x}}_{0|k}^{(\ell)} - \hat{\mathbf{x}}_{0|0}\right) \; + \\ \sum_{\kappa=1}^{k} \mathcal{F}_{\kappa-1}^{(\ell)^T} \mathbf{Q}_{\kappa-1}'^{-1} \left(\hat{\mathbf{x}}_{\kappa|k}^{(\ell)} - \mathbf{\Phi}_{\kappa-1} \hat{\mathbf{x}}_{\kappa-1|k}^{(\ell)}\right) \; + \\ \sum_{\kappa=1}^{k} \sigma_{\kappa}^{-2} \mathcal{H}_{\kappa}^{(\ell)^T} \left(z_\kappa - h(\hat{\mathbf{x}}_{\kappa|k}^{(\ell)})\right)\end{aligned} \tag{15}$$

where $\mathbf{\Pi} \triangleq \begin{bmatrix}\mathbf{I}_{2N} & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix}$ is used to adjust the dimension of the $2N$-dimensional prior estimate to the dimension of the entire state $\mathbf{x}_{0:k}$. In the above expression, $\mathcal{F}_{\kappa-1}^{(\ell)}$ and $\mathcal{H}_{\kappa}^{(\ell)}$, are the Jacobians of the motion and measurement models [see (3), and (5) or (7), respectively], with respect to the entire state $\mathbf{x}_{0:k}$, evaluated at $\hat{\mathbf{x}}_{0:k|k}^{(\ell)}$. It is important to note that both the target motion model and the measurement function involve only a few states, i.e., the target motion only depends on two consecutive states, while the measurement only depends on the target's position where it is observed. Thus, $\mathcal{F}_{\kappa-1}$ and $\mathcal{H}_{\kappa}$ have the following sparse structure (for simplicity, the iteration index is dropped off here):

$$\mathcal{F}_{\kappa-1} = \begin{bmatrix}\mathbf{0}_{2N\times 2N} & \cdots & -\mathbf{\Phi}_{\kappa-1} & \mathbf{I}_{2N} & \cdots & \mathbf{0}_{2N\times 2N}\end{bmatrix} \tag{16}$$

$$\mathcal{H}_{\kappa} = \begin{bmatrix}\mathbf{0}_{1\times 2N} & \cdots & -\mathbf{H}_{\kappa} & \cdots & \mathbf{0}_{1\times 2N}\end{bmatrix} \tag{17}$$

where $\mathbf{H}_\kappa$ is the measurement Jacobian matrix at time-step κ, given by [see (5) and (7)]:

$$\mathbf{H}_\kappa = \begin{cases} \begin{bmatrix} \frac{(\hat{\mathbf{p}}_{T_{\kappa|k}} - \mathbf{p}_{S_\kappa})^T}{||\hat{\mathbf{p}}_{T_{\kappa|k}} - \mathbf{p}_{S_\kappa}||} & \mathbf{0}_{1\times(2N-2)} \end{bmatrix}, & \text{if range}-\text{only} \\ \begin{bmatrix} \frac{(\hat{\mathbf{p}}_{T_{\kappa|k}} - \mathbf{p}_{S_\kappa})^T \mathbf{J}^T}{||\hat{\mathbf{p}}_{T_{\kappa|k}} - \mathbf{p}_{S_\kappa}||^2} & \mathbf{0}_{1\times(2N-2)} \end{bmatrix}, & \text{if bearing}-\text{only} \end{cases} \tag{18}$$

and $\mathbf{J} \triangleq \begin{bmatrix}0 & -1 \\ 1 & 0\end{bmatrix}$. On the other hand, the Hessian matrix, $\mathbf{A}^{(\ell)}$, is approximated in the Gauss-Newton method by:

$$\mathbf{A}^{(\ell)} \simeq \mathbf{\Pi}^T \mathbf{P}_{0|0}^{-1} \mathbf{\Pi} + \sum_{\kappa=1}^{k} \mathcal{F}_{\kappa-1}^{(\ell)^T} \mathbf{Q}_{\kappa-1}'^{-1} \mathcal{F}_{\kappa-1}^{(\ell)} + \sum_{\kappa=1}^{k} \sigma_{\kappa}^{-2} \mathcal{H}_{\kappa}^{(\ell)^T} \mathcal{H}_{\kappa}^{(\ell)} \tag{19}$$

which is a reasonably good approximation for small-residual problems [31].

The value $\delta\mathbf{x}_{0:k}^{(\ell)}$ that minimizes (12) is found by solving the following linear system:

$$\mathbf{A}^{(\ell)} \delta\mathbf{x}_{0:k}^{(\ell)} = -\mathbf{b}^{(\ell)} \tag{20}$$

The Hessian $\mathbf{A}^{(\ell)}$ has dimension $2N(k+1) \times 2N(k+1)$ [see (1) and (8)]. However, due to the sparse structure of the matrices $\mathcal{H}_{\kappa}^{(\ell)}$ and $\mathcal{F}_{\kappa-1}^{(\ell)}$, the matrix $\mathbf{A}^{(\ell)}$ is also sparse. Most importantly, it has a banded structure with upper and lower bandwidth of $4N$ due to (i) the Markov motion model and (ii) the fact that the range (or bearing) measurement model only involves the target's position at one time step. We can exploit this sparse banded structure to reduce the computational complexity of solving (20) to $\mathcal{O}(N^3 k)$, instead of $\mathcal{O}\left(N^3 k^3\right)$ [32]. Once $\delta\mathbf{x}_{0:k}^{(\ell)}$ is found, the new state estimate is computed as:

$$\hat{\mathbf{x}}_{0:k|k}^{(\ell+1)} = \hat{\mathbf{x}}_{0:k|k}^{(\ell)} + \delta\mathbf{x}_{0:k}^{(\ell)} \tag{21}$$

Given an initial estimate $\hat{\mathbf{x}}_{0:k|k}^{(0)}$, this iterative algorithm only computes one local minimum for the entire target's trajectory given all measurements up to time-step k.

4 Incrementally Solving the Batch MAP Optimization Problem

In this section, we present an incremental solution to the batch MAP problem of minimizing (11). Since it is, in general, intractable to analytically solve the batch MAP problem, we relax the problem and incrementally solve a one-step minimization problem for the current state estimate *analytically*, at every time step when a new measurement becomes available. This analytic optimization is carried out by converting the nonlinear cost function into polynomial or rational form which is then solved using algebraic geometry techniques.

4.1 Relaxation of the Batch MAP Problem

As will become clear from the next section, by transforming the nonlinear measurement function into polynomial or rational form, we can convert the KKT optimality conditions of the batch MAP problem into a polynomial system that has $2N(k+1)$ variables (see Section 4.2). The number of variables increases linearly with respect to the time horizon k. However, since the complexity of solving multivariate polynomial systems is exponential in the number of variables [33], it is, in general, computationally intractable to solve the batch MAP problem analytically.

For this reason, we relax the batch MAP problem and solve it incrementally. We start by noting that the cost function (11) can be written as:

$$c(\mathbf{x}_{0:k}) = c(\mathbf{x}_{0:k-1}) + \frac{1}{2}||\mathbf{x}_k - \mathbf{\Phi}_{k-1}\mathbf{x}_{k-1}||^2_{\mathbf{Q}'_{k-1}} + \frac{1}{2}||z_k - h(\mathbf{x}_k)||^2_{\sigma_k^2} \tag{22}$$

Since the motion model (3) is linear, at every time step when a new measurement becomes available, we can combine all previous information about the target's trajectory to provide a prior for the new current state, $\mathbf{x}_k$, based on the following propagation equations:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{\Phi}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \tag{23}$$

$$\mathbf{P}_{k|k-1} = \mathbf{\Phi}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{\Phi}^T_{k-1} + \mathbf{G}_{k-1}\mathbf{Q}_{k-1}\mathbf{G}^T_{k-1} \tag{24}$$

Thus, the new state estimates can be computed *incrementally* by solving the following one-step minimization problem:

$$\min_{\mathbf{x}_k} \left[\frac{1}{2}||\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}||^2_{\mathbf{P}_{k|k-1}} + \frac{1}{2}||z_k - h(\mathbf{x}_k)||^2_{\sigma_k^2}\right] \tag{25}$$

4.2 Analytic Determination of Local Minima

Observing that both relative range and bearing measurements depend only on the target position (see (5) and (7)), we can decouple the target position $\mathbf{p}_{T_k}$ and the remaining states $\mathbf{x}_{T_{\mathbf{d},k}}$ in solving (25), so as to simplify the ensuing derivations. Specifically, using the following partitioning of the information matrix, $\mathbf{P}^{-1}_{k|k-1} \triangleq \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{pp}_{k|k-1}} & \mathbf{\Sigma}_{\mathbf{pd}_{k|k-1}} \\ \mathbf{\Sigma}_{\mathbf{dp}_{k|k-1}} & \mathbf{\Sigma}_{\mathbf{dd}_{k|k-1}} \end{bmatrix}$, the cost function of (25) can be written as:

$$\begin{aligned} c(\mathbf{x}_k) =& \frac{1}{2}(\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}})^T \mathbf{\Sigma}_{\mathbf{pp}_{k|k-1}} (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}}) + \\ & \frac{1}{2}(\mathbf{x}_{T_{\mathbf{d},k}} - \hat{\mathbf{x}}_{T_{\mathbf{d},k|k-1}})^T \mathbf{\Sigma}_{\mathbf{dd}_{k|k-1}} (\mathbf{x}_{T_{\mathbf{d},k}} - \hat{\mathbf{x}}_{T_{\mathbf{d},k|k-1}}) + \\ & (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}})^T \mathbf{\Sigma}_{\mathbf{pd}_{k|k-1}} (\mathbf{x}_{T_{\mathbf{d},k}} - \hat{\mathbf{x}}_{T_{\mathbf{d},k|k-1}}) + \\ & \frac{1}{2\sigma_k^2}\left(z_k - h(\mathbf{p}_{T_k})\right)^2 \end{aligned} \tag{26}$$

We note that

$$\min_{\mathbf{p}_{T_k}, \mathbf{x}_{T_{\mathbf{d},k}}} c(\mathbf{p}_{T_k}, \mathbf{x}_{T_{\mathbf{d},k}}) \;=\; \min_{\mathbf{p}_{T_k}} \left(\min_{\mathbf{x}_{T_{\mathbf{d},k}}} c(\mathbf{p}_{T_k}, \mathbf{x}_{T_{\mathbf{d},k}}) \right)$$

Thus, we first solve for $\mathbf{x}_{T_{\mathbf{d},k}}$ by setting the gradient of (26) with respect to $\mathbf{x}_{T_{\mathbf{d},k}}$ to zero, and obtain:

$$\mathbf{x}_{T_{\mathbf{d},k}} = \hat{\mathbf{x}}_{T_{\mathbf{d},k|k-1}} - \mathbf{\Sigma}^{-1}_{\mathbf{dd}_{k|k-1}} \mathbf{\Sigma}_{\mathbf{dp}_{k|k-1}} (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}}) \tag{27}$$

Substitution of (27) into (26) yields:

$$c(\mathbf{p}_{T_k}) = \frac{1}{2}(\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}})^T \mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}}) + \frac{1}{2\sigma_k^2} \left(z_k - h(\mathbf{p}_{T_k})\right)^2 \tag{28}$$

where $\mathbf{P}_{\mathbf{pp}_{k|k-1}}$ is the covariance matrix corresponding to the target position, obtained by partitioning the covariance matrix as $\mathbf{P}_{k|k-1} \triangleq \begin{bmatrix} \mathbf{P}_{\mathbf{pp}_{k|k-1}} & \mathbf{P}_{\mathbf{pd}_{k|k-1}} \\ \mathbf{P}_{\mathbf{dp}_{k|k-1}} & \mathbf{P}_{\mathbf{dd}_{k|k-1}} \end{bmatrix}$. In the above expression (28), we have employed the following identity:

$$\mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} = \mathbf{\Sigma}_{\mathbf{pp}_{k|k-1}} - \mathbf{\Sigma}_{\mathbf{pd}_{k|k-1}} \mathbf{\Sigma}^{-1}_{\mathbf{dd}_{k|k-1}} \mathbf{\Sigma}_{\mathbf{dp}_{k|k-1}} \tag{29}$$

which can easily be verified using the block matrix inversion lemma [32].

We thus see that solving (25) becomes equivalent to minimizing (28). It is important to note that the size of the nonlinear problem has dramatically decreased from $2N$ for (25) to a constant size of 2 for minimizing (28). Moreover, the analytic solution for the target position is independent of its higher-order time derivatives, regardless of the stochastic target motion model. In the following we present our algebraic geometry approaches for solving the problem of minimizing (28) analytically, in the cases of range-only tracking and bearing-only tracking, respectively.

4.2.1 Range-only tracking

Now, we solve the problem (28) analytically. By introducing a new variable $d = h(\mathbf{x}_k)$, the problem (28) in the case of single sensor is equivalent to:

$$\min_{\mathbf{p}_{T_{k+1}}, d} \ \frac{1}{2}(\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}})^T \mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}}) + \frac{1}{2}(z_k - d)^T \mathbf{R}_k^{-1} (z_k - d) \tag{30}$$

$$\text{s.t. } d^2 = (x_{S_k} - x_{T_k})^2 + (y_{S_k} - y_{T_k})^2 \ , \ d \geq 0 \tag{31}$$

We employ the Lagrangian multiplier to solve the above problem (30)-(31). Using $\mathbf{R}_k = \sigma^2_{d_k}$, the Lagrangian function can be constructed as follows (by temporarily ignoring the inequality constraint of d, which will be used later to verify the legal solutions):

$$L(\mathbf{p}_{T_k}, d, \lambda) = \frac{1}{2}(\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}})^T \mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} (\mathbf{p}_{T_k} - \hat{\mathbf{p}}_{T_{k|k-1}}) + \frac{(z_k - d)^2}{2\sigma^2_{d_k}} + \lambda \left(d^2 - (x_{S_k} - x_{T_k})^2 - (y_{S_k} - y_{T_k})^2\right) \tag{32}$$

where λ is the Lagrangian multiplier. Without loss of generality, we assume $\mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} = \mathbf{Diag}(s_1, s_2)$, since we always can diagonalize $\mathbf{P}^{-1}_{\mathbf{pp}}$ by applying a rotational transformation. Thus, the Lagrangian function (32) becomes:

$$\mathcal{L}(x_{T_k}, y_{T_k}, d, \lambda) = \frac{s_1}{2}(x_{T_k} - \hat{x}_{T_{k|k-1}})^2 + \frac{s_2}{2}(y_{T_k} - \hat{y}_{T_{k|k-1}})^2 + \frac{(z_k - d)^2}{2\sigma^2_{d_k}} + \lambda \left(d^2 - (x_{S_k} - x_{T_k})^2 - (y_{S_k} - y_{T_k})^2\right) \tag{33}$$

where λ is the Lagrangian multiplier. By setting the respective derivatives of $L(x_{T_k}, y_{T_k}, d, \lambda)$ with respect to the four unknown variables to zeros, and with algebraic manipulations, we have:

$$\frac{\partial \mathcal{L}}{\partial x_T} = s_1(x_{T_k} - \hat{x}_{T_{k|k-1}}) - 2\lambda(x_{T_k} - x_{S_k}) = 0 \Rightarrow x_{T_k} = \frac{s_1 \hat{x}_{T_{k|k-1}} - 2\lambda x_{S_k}}{s_1 - 2\lambda} \tag{34}$$

$$\frac{\partial \mathcal{L}}{\partial y_T} = s_2(y_{T_k} - \hat{y}_{T_{k|k-1}}) - 2\lambda(y_{T_k} - y_{S_k}) = 0 \Rightarrow y_{T_k} = \frac{s_2 \hat{y}_{T_{k|k-1}} - 2\lambda y_{S_k}}{s_2 - 2\lambda} \tag{35}$$

$$\frac{\partial \mathcal{L}}{\partial d} = -\frac{1}{\sigma^2_{d_k}}(z - d) + 2\lambda d = 0 \Rightarrow d = \frac{z}{1 + 2\sigma^2_{d_k}\lambda} \tag{36}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = d^2 - (x_{S_k} - x_{T_k})^2 - (y_{S_k} - y_{T_k})^2 = 0 \tag{37}$$

Substitution of (34)-(36) into (37) yields:

$$0 = \left(\frac{z}{1+2\sigma_{d_k}^2\lambda}\right)^2 - \left(\frac{s_1 x_{S_k} - s_1\hat{x}_{T_{k|k-1}}}{s_1 - 2\lambda}\right)^2 - \left(\frac{s_2 y_{S_k} - s_2\hat{y}_{T_{k|k-1}}}{s_2 - 2\lambda}\right)^2$$

Multiplying both sides by $(1 + 2\sigma_{d_k}^2\lambda)^2(s_1 - 2\lambda)^2(s_2 - 2\lambda)^2$, we obtain a fourth-order univariate polynomial in variable λ:

$$0 = f(\lambda) = \sum_{i=0}^{4} a_i \lambda^i \tag{38}$$

where a_i, $i = 0, 1, \ldots, 4$, are the coefficients expressed in terms of $s_1, s_2, z, \sigma_d, \hat{x}_T, \hat{y}_T, x_S$, and y_S:

$$a_0 = z^2 s_1^2 s_2^2 - (s_1\hat{x}_T - s_1 x_S)^2 s_2^2 - (s_2\hat{y}_T - s_2 y_S)^2 s_1^2 \tag{39}$$

$$a_1 = 4(s_2\hat{y}_T - s_2 y_S)^2 s_1 - 4(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^2 s_1^2 - 4z^2 s_1^2 s_2 - 4z^2 s_1 s_2^2 + 4(s_1\hat{x}_T - s_1 x_S)^2 s_2 - 4(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^2 s_2^2 \tag{40}$$

$$a_2 = -4(s_2\hat{y}_T - s_2 y_S)^2 + 16(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^2 s_1 - 4(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^4 s_1^2 - 4(s_1\hat{x}_T - s_1 x_S)^2 + 16(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^2 s_2 - 4(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^4 s_2^2 + 4z^2 s_1^2 + 16z^2 s_1 s_2 + 4z^2 s_2^2 \tag{41}$$

$$a_3 = -16(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^2 + 16(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^4 s_2 - 16(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^2 + 16(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^4 s_1 - 16z^2 s_1 - 16z^2 s_2 \tag{42}$$

$$a_4 = 16z^2 - 16(s_1\hat{x}_T - s_1 x_S)^2\sigma_d^4 - 16(s_2\hat{y}_T - s_2 y_S)^2\sigma_d^4 \tag{43}$$

The roots of $f(\lambda)$ are the eigenvalues of the corresponding 4×4 companion matrix. Although there are 4 solutions for λ and thus 4 solutions for x_T, y_T and d (since x_T, y_T and d are linear with respect to λ (cf. (34)-(36))), we only need to consider the solutions of x_T and y_T that are corresponding to real eigenvalues of the companion matrix as well as nonnegative d (i.e., $d \geq 0$). Note that since these solutions also could be the local maxima, the second derivative test [34] is employed to extract the minima. Finally, once we determine all the local minima of the target position, we compute the corresponding solutions of the higher-order position derivatives via (27).

Roots of the 4-th order polynomial. The roots for the general 4-th order polynomial $f(x) = x^4 + ax^3 + bx^2 + cx + d$:

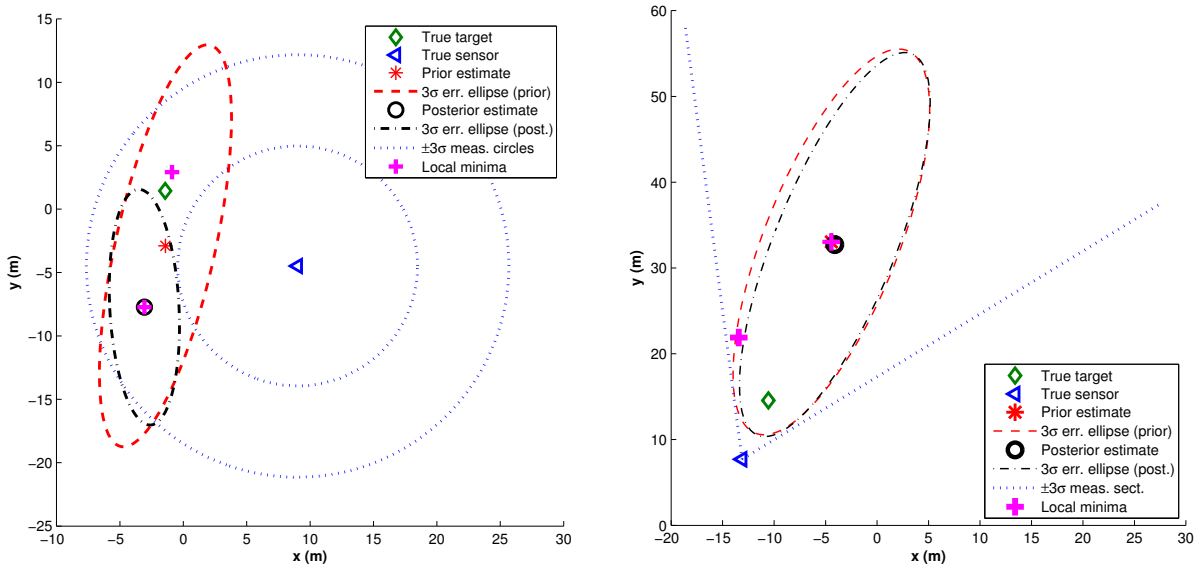
$$x_1 = \frac{-a}{4} - \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big\} - \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{2} - \frac{4b}{3} - \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} - \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}} - \left(-a^3 + 4ab - 8c\right)\Big/\Big[4\,\mathbf{sqrt}\Big[\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big]\Big]\Big\} \tag{44}$$

$$x_2 = \frac{-a}{4} - \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big\} + \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{2} - \frac{4b}{3} - \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} - \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}} - \left(-a^3 + 4ab - 8c\right)\Big/\Big[4\,\mathbf{sqrt}\Big[\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big]\Big]\Big\} \tag{45}$$

$$x_3 = \frac{-a}{4} + \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big\} - \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{2} - \frac{4b}{3} - \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} - \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}} + \left(-a^3 + 4ab - 8c\right)\Big/\Big[4\,\mathbf{sqrt}\Big[\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big]\Big]\Big\} \tag{46}$$

$$x_4 = \frac{-a}{4} + \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big\} + \frac{1}{2}\mathbf{sqrt}\Big\{\frac{a^2}{2} - \frac{4b}{3} - \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} - \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}} + \left(-a^3 + 4ab - 8c\right)\Big/\Big[4\,\mathbf{sqrt}\Big[\frac{a^2}{4} - \frac{2b}{3} + \frac{2^{\frac{1}{3}}\left(b^2 - 3ac + 12d\right)}{3\left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}\right)^{\frac{1}{3}}} + \left(\frac{2b^3 - 9abc + 27c^2 + 27a^2d - 72bd + \sqrt{-4\left(b^2 - 3ac + 12d\right)^3 + \left(2b^3 - 9abc + 27c^2 + 27a^2d - 72bd\right)^2}}{54}\right)^{\frac{1}{3}}\Big]\Big]\Big\} \tag{47}$$

Since the maximum number of local minima for the problem (25) will significantly impact the computational complexity of our proposed algorithm (see Section 5), we seek a tighter upper bound for it. In particular, based on the finite dimensional Mountain Pass Theorem (MPT) (see Theorem 5.2 in [35]), we can show the following lemma:



(a) Range-only tracking

(b) Bearing-only tracking

Figure 1: Illustrative problem for single time step target tracking: The magenta crosses indicate the location of the analytically computed local minima. It is clear that the MAP estimate initialized with the prior estimate converges to the local minimum with larger error with respect to ground truth. Note that in bearing-only tracking, the MAP estimate is computed based on the original (not inferred) measurements, and the approximation (50) used in the inferred measurements introduces a slight offset in the analytic local minima.

Lemma 4.1. *There are at most 2 local minima for the problem of minimizing* (30).

Proof. See Appendix A. □

Thus, we see from this lemma that the total number of local minima for the one-step MAP problem (25) for range-only tracking in worst case, can grow exponentially over time, in an order of 2^k, instead of 4^k. Fig. 1(a) shows a typical example where two local minima for the current state occur while the MAP estimate erroneously converges to a local minimum with larger error.

4.2.2 Bearing-only tracking

Now we focus on the analytic solution of (28). In order to use an algebraic geometry approach, we transform the bearing measurement function (7) into a *rational* form. Specifically, after moving the sensor orientation term to the right left side of (7), we apply the tangent function on both sides and obtain the following transformed measurement:

$$\begin{aligned} \check{z}_k &\triangleq \tan(z_k + \phi_{S_k}) \\ &= \tan\left(\operatorname{atan2}\left((y_{T_k} - y_{S_k}), (x_{T_k} - x_{S_k})\right) + n_k\right) \end{aligned} \tag{48}$$

By denoting $\xi_k \triangleq \operatorname{atan2}\left((y_{T_k} - y_{S_k}), (x_{T_k} - x_{S_k})\right)$, considering $z_k + \phi_{S_k} \in (-\pi, \pi]$, and following the standard formulas to compute the pdf of functions of random variables [36], the pdf of the transformed measurement is given by:

$$p(\check{z}_k|\mathbf{x}_k) = \begin{cases} \frac{\mathcal{N}(\tan^{-1}(\check{z}_k);\xi_k,\sigma^2_{\theta_k}) + \mathcal{N}(\tan^{-1}(\check{z}_k)-\pi;\xi_k,\sigma^2_{\theta_k})}{1+\check{z}_k^2}, & \text{if } \check{z}_k \geq 0 \\ \frac{\mathcal{N}(\tan^{-1}(\check{z}_k);\xi_k,\sigma^2_{\theta_k}) + \mathcal{N}(\tan^{-1}(\check{z}_k)+\pi;\xi_k,\sigma^2_{\theta_k})}{1+\check{z}_k^2}, & \text{if } \check{z}_k < 0 \end{cases} \tag{49}$$

where $\mathcal{N}(a;\mu,\sigma^2)$ denotes the Gaussian pdf of a random variable a with mean μ and variance σ^2. Clearly, $p(\check{z}_k)$ is not Gaussian (it results from the tangent of a Gaussian random variable), but it can be well approximated by a Gaussian

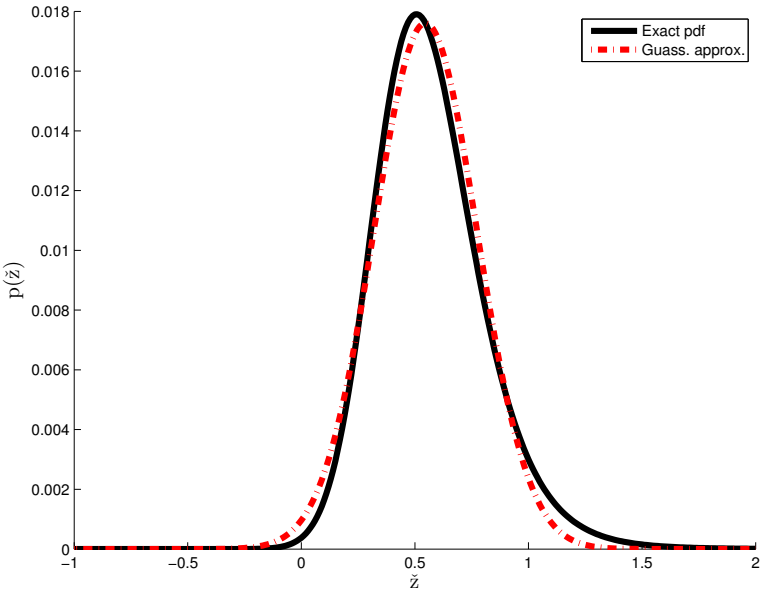


Figure 2: An example of approximating the pdf of transformed measurements by a Gaussian pdf. In this case, $\xi = 0.5$ and $\sigma_\theta = 10$ deg. In addition, the Kullback-Leibler distance (KLD) between these two pdfs is only 0.0447, which indicates the difference between the two distributions is negligible.

pdf by matching first and second order moments. This is done by linearizing (48) around the expected value of the noise, i.e.,

$$\check{z}_k \simeq \frac{y_{T_k} - y_{S_k}}{x_{T_k} - x_{S_k}} + \bar{n}_k \triangleq \bar{z}_k \tag{50}$$

where $\bar{n}_k \triangleq \sec^2(z_k + \phi_{S_k}) n_k$, is zero-mean white Gaussian noise with variance $\sigma_k^2 \triangleq \sec^4(z_k + \phi_{S_k})\sigma_{\theta_k}^2$, i.e., $\bar{n}_k \sim \mathcal{N}(0, \sigma_k^2)$. We term this approximation (50) an *inferred* measurement which is in the desired rational form. As illustrated in Fig. 2, this approximation is reasonably accurate, particularly for high signal-to-noise ratios. Moreover, the local minimum of (28) attained based on the inferred measurement is very close to that based on the corresponding original bearing measurement. This can be seen from Fig. 1(b), where one of the local minima using the inferred measurement almost coincides with the IEKF posterior estimate (which corresponds to the local minimum determined by applying the Gauss-Newton method [37, 38]) using the original bearing measurement. This further shows that the inferred measurement is a reasonably good approximation to the original bearing measurement.[3] Moreover, the inferred measurement is used *only* for finding hypotheses of the trajectory, not for estimating the state per se, whose estimates will be updated by the batch MAP estimators using all available *original* bearing measurements.

In what follows we use the inferred measurement (50) (instead of (7)) to compute the analytic solutions for minimizing (28). In particular, with $\mathbf{P}^{-1}_{\mathbf{pp}_{k|k-1}} = \begin{bmatrix} s_1 & s_3 \\ s_3 & s_2 \end{bmatrix}$, (28) can be written as:

$$\begin{aligned} c(x_{T_k}, y_{T_k}) = \frac{1}{2}\Big(& s_1(x_{T_k} - \hat{x}_{T_{k|k-1}})^2 + s_2(y_{T_k} - \hat{y}_{T_{k|k-1}})^2 + \\ & 2s_3(x_{T_k} - \hat{x}_{T_{k|k-1}})(y_{T_k} - \hat{y}_{T_{k|k-1}})\Big) + \frac{1}{2\sigma_k^2}\left(\bar{z}_k - \frac{y_{T_k} - y_{S_k}}{x_{T_k} - x_{S_k}}\right)^2 \end{aligned} \tag{51}$$

Setting the derivatives of $c(x_{T_k}, y_{T_k})$ with respect to the two optimization variables to zero, and performing simple

[3]Note also that this inferred measurement model (50) does not consider the special case of $x_{T_k} = x_{S_k}$. However, this case has low probability of occurrence in practice, and can be easily avoided through an appropriate coordinate transformation.

algebraic manipulations, we have:

$$\frac{\partial c}{\partial x_{T_k}} = s_1(x_{T_k} - \hat{x}_{T_{k|k-1}}) + s_3(y_{T_k} - \hat{y}_{T_{k|k-1}}) + \frac{1}{\sigma_k^2}\left[\frac{\bar{z}_k(y_{T_k} - y_{S_k})}{(x_{T_k} - x_{S_k})^2} - \frac{(y_{T_k} - y_{S_k})^2}{(x_{T_k} - x_{S_k})^3}\right] = 0$$

$$\Rightarrow \; s_1(x_{T_k} - \hat{x}_{T_{k|k-1}})(x_{T_k} - x_{S_k})^3 + s_3(x_{T_k} - x_{S_k})^3(y_{T_k} - \hat{y}_{T_{k|k-1}}) + \frac{1}{\sigma_k^2}\left[\bar{z}_k(x_{T_k} - x_{S_k})(y_{T_k} - y_{S_k}) - (y_{T_k} - y_{S_k})^2\right] = 0 \tag{52}$$

$$\frac{\partial c}{\partial y_{T_k}} = s_2(y_{T_k} - \hat{y}_{T_{k|k-1}}) + s_3(x_{T_k} - \hat{x}_{T_{k|k-1}}) - \frac{1}{\sigma_k^2}\left[\frac{\bar{z}_k}{x_{T_k} - x_{S_k}} - \frac{(y_{T_k} - y_{S_k})}{(x_{T_k} - x_{S_k})^2}\right] = 0$$

$$\Rightarrow \; s_2(x_{T_k} - x_{S_k})^2(y_{T_k} - \hat{y}_{T_{k|k-1}}) + s_3(x_{T_k} - x_{S_k})^2(x_{T_k} - \hat{x}_{T_{k|k-1}}) - \frac{1}{\sigma_k^2}\left[\bar{z}_k(x_{T_k} - x_{S_k}) - (y_{T_k} - y_{S_k})\right] = 0 \tag{53}$$

From (53), we can compute y_{T_k} in terms of x_{T_k} as follows:

$$y_{T_k} = \frac{-\sigma_k^2 s_3(x_{T_k} - x_{S_k})^2(x_{T_k} - \hat{x}_{T_{k|k-1}}) + \bar{z}_k(x_{T_k} - x_{S_k})}{1 + \sigma_k^2 s_2(x_{T_k} - x_{S_k})^2} + \frac{\sigma_k^2 s_2(x_{T_k} - x_{S_k})^2 \hat{y}_{T_{k|k-1}} + y_{S_k}}{1 + \sigma_k^2 s_2(x_{T_k} - x_{S_k})^2} \tag{54}$$

Substitution of (54) into (52) yields a rational equation, whose denominator is always non-zero. Thus, we only need to focus on the numerator which is an eighth-order univariate polynomial in x_{T_k}:

$$0 = f(x_{T_k}) = \sum_{i=0}^{8} a_i x_{T_k}^i \tag{55}$$

where a_i, $i = 0, 1, \ldots, 8$, are the coefficients expressed in terms of $\bar{z}_k, s_1, s_2, s_3, \hat{x}_{T_{k|k-1}}, \hat{y}_{T_{k|k-1}}, x_{S_k}$, and y_{S_k}:

$$
\begin{aligned}
a_0 = {} & s_2 y_T x_S^3 z + 2 s_2^2 y_T x_S^4 y_S - s_3 y_S x_S^3 + s_3 s_2 y_T x_S^5 - z s_2 x_S^5 s_3 x_T - \\
& s_3^2 x_T x_S^7 s_2 + s_3 x_T x_S^3 z - y_S s_2 x_S^3 z - s_2^2 y_T^2 x_S^4 + z s_2^2 x_S^5 y_S - \\
& s_3 y_S s_2 x_S^5 + 2 s_3 x_T x_S^4 y_S s_2 - y_S^2 s_2^2 x_S^4 + s_1 x_T s_2^2 x_S^7 + 2 s_1 x_T s_2 x_S^5 + \\
& z^2 s_2 x_S^4 - s_3^2 x_T x_S^5 + s_3 z s_2 x_S^6 + s_3 z x_S^4 - s_3^2 x_T^2 x_S^4 - 2 s_2 y_T x_S^4 s_3 x_T - \\
& z s_2^2 x_S^5 y_T + s_1 x_T x_S^3 + s_3 y_T x_S^3 \qquad (56) \\
a_1 = {} & 5 z s_2^2 x_S^4 y_T - 3 s_3 x_T x_S^2 z + 3 y_S s_2 x_S^2 z - 7 s_1 x_T s_2^2 x_S^6 - 10 s_1 x_T s_2 x_S^4 + \\
& 7 s_3^2 x_T x_S^6 s_2 - 3 s_3 s_2 y_T x_S^4 + 3 s_3 y_S s_2 x_S^4 - 5 s_3 z s_2 x_S^5 + 3 s_3 y_S x_S^2 - \\
& 2 s_1 s_2 x_S^5 + 7 s_3^2 x_T x_S^4 - 3 s_3 y_T x_S^2 - 8 s_2^2 y_T x_S^3 y_S - 3 s_2 y_T x_S^2 z - \\
& 5 z s_2^2 x_S^4 y_S - s_1 x_S^3 + s_3^2 x_S^5 - s_1 s_2^2 x_S^7 + 4 s_3^2 x_T^2 x_S^3 - 5 s_3 z x_S^3 + \\
& s_3^2 x_S^7 s_2 - 4 z^2 s_2 x_S^3 + 4 s_2^2 y_T^2 x_S^3 + 4 y_S^2 s_2^2 x_S^3 - 3 s_1 x_T x_S^2 + \\
& 5 z s_2 x_S^4 s_3 x_T + 8 s_2 y_T x_S^3 s_3 x_T - 8 s_3 x_T x_S^3 y_S s_2 \qquad (57) \\
a_2 = {} & 10 s_1 s_2 x_S^4 - 18 s_3^2 x_T x_S^3 - 7 s_3^2 x_S^6 s_2 + 9 s_3 z x_S^2 + 6 z^2 s_2 x_S^2 - 6 y_S^2 s_2^2 x_S^2 - \\
& 6 s_3^2 x_T^2 x_S^2 - 6 s_2^2 y_T^2 x_S^2 + 3 s_1 x_T x_S + 10 z s_2^2 x_S^3 y_S - 6 s_3^2 x_S^4 + 3 s_1 x_S^2 - \\
& 10 z s_2 x_S^3 s_3 x_T - 12 s_2 y_T s_3 x_T x_S^2 + 12 s_3 x_T y_S s_2 x_S^2 + 7 s_1 s_2^2 x_S^6 + 3 s_3 y_T x_S - \\
& 3 s_3 y_S x_S + 10 s_3 z s_2 x_S^4 - 10 z s_2^2 x_S^3 y_T + 3 s_2 y_T z x_S + 12 s_2^2 y_T y_S x_S^2 + 3 s_3 x_T z x_S - \\
& 3 y_S s_2 z x_S + 21 s_1 x_T s_2^2 x_S^5 + 20 s_1 x_T s_2 x_S^3 - 21 s_3^2 x_T s_2 x_S^5 + 2 s_3 s_2 y_T x_S^3 - 2 s_3 y_S s_2 x_S^3 \qquad (58) \\
a_3 = {} & -s_1 x_T - 7 s_3 z x_S - 20 s_1 s_2 x_S^3 + 22 s_3^2 x_T x_S^2 + 21 s_3^2 s_2 x_S^5 - 10 z s_2^2 y_S x_S^2 - \\
& 3 s_1 x_S + s_3 y_S + 10 z s_2 s_3 x_T x_S^2 + 8 s_2 y_T s_3 x_T x_S - 8 s_3 x_T y_S s_2 x_S + 10 z s_2^2 y_T x_S^2 + \\
& 2 s_3 s_2 y_T x_S^2 - 2 s_3 y_S s_2 x_S^2 - 10 s_3 z s_2 x_S^3 - 8 s_2^2 y_T y_S x_S - 35 s_1 x_T s_2^2 x_S^4 - \\
& 20 s_1 x_T s_2 x_S^2 + 35 s_3^2 x_T s_2 x_S^4 + 14 s_3^2 x_S^3 - s_3 y_T - 21 s_1 s_2^2 x_S^5 - 4 z^2 s_2 x_S + \\
& 4 s_2^2 y_T^2 x_S - s_2 y_T z + y_S s_2 z - s_3 x_T z + 4 s_3^2 x_T^2 x_S + 4 y_S^2 s_2^2 x_S \qquad (59) \\
a_4 = {} & 35 s_1 s_2^2 x_S^4 - 35 s_3^2 x_T s_2 x_S^3 + 5 s_3 z s_2 x_S^2 + 35 s_1 x_T s_2^2 x_S^3 - 2 s_2 y_T s_3 x_T - \\
& 5 z s_2 s_3 x_T x_S + s_1 + 2 s_3 x_T y_S s_2 - y_S^2 s_2^2 + 2 s_2^2 y_T y_S - 13 s_3^2 x_T x_S + 20 s_1 s_2 x_S^2 + \\
& 10 s_1 x_T s_2 x_S + 2 s_3 z + 5 z s_2^2 y_S x_S - 3 s_3 s_2 y_T x_S + 3 s_3 y_S s_2 x_S - 16 s_3^2 x_S^2 - \\
& s_2^2 y_T^2 - s_3^2 x_T^2 - 35 s_3^2 s_2 x_S^4 - 5 z s_2^2 y_T x_S + z^2 s_2 \qquad (60) \\
a_5 = {} & 21 s_3^2 x_T s_2 x_S^2 + 35 s_3^2 s_2 x_S^3 - 10 s_1 s_2 x_S - 2 s_1 x_T s_2 + 9 s_3^2 x_S - s_3 z s_2 x_S - 21 s_1 x_T s_2^2 x_S^2 + \\
& s_3 s_2 y_T + 3 s_3^2 x_T - 35 s_1 s_2^2 x_S^3 + z s_2^2 y_T - s_3 y_S s_2 - z s_2^2 y_S + s_3 z s_2 x_T \qquad (61) \\
a_6 = {} & 2 s_1 s_2 + 7 s_1 x_T s_2^2 x_S - 2 s_3^2 - 7 s_3^2 x_T s_2 x_S - 21 s_3^2 s_2 x_S^2 + 21 s_1 s_2^2 x_S^2 \qquad (62) \\
a_7 = {} & -7 s_1 s_2^2 x_S + 7 s_3^2 s_2 x_S - s_1 x_T s_2^2 + s_3^2 x_T s_2 \qquad (63) \\
a_8 = {} & -s_3^2 s_2 + s_1 s_2^2 \qquad (64)
\end{aligned}
$$

The roots of $f(x_{T_k})$ are the eigenvalues of the corresponding 8×8 companion matrix [39].

Although there exist 8 solutions for x_{T_k} and thus 8 solutions for y_{T_k}, as it depends injectively on x_{T_k} (see (54)), we only need to consider the pairs (x_{T_k}, y_{T_k}) that correspond to real eigenvalues of the companion matrix. Moreover, since some of these solutions could be local maxima or saddle points, the second-order derivative test [34] is employed to extract the minima. Finally, once we determine all the local minima for the target position, we compute the corresponding estimates for the higher-order position derivatives via (27).

The following lemma provides a smaller upper bound for the maximum number of local minima, which is important since it significantly impacts the computational complexity of our proposed method (see Section 5).

Lemma 4.2. *There are at most 7 local minima for* (51)*.*

Proof. Using the Finite Dimensional Mountain Pass Theorem (see Theorem 5.2 in [35]), which states that there exists

a third critical point which is not a local minimum between any two strict local minima for a coercive $\mathcal{C}^1$ function,[4] it is straightforward to show that our cost function (51) $\in \mathcal{C}^1(\mathbb{R}^2 \backslash \{x_{T_k} = x_{S_k}\})$ is coercive, and thus at least one of the 8 critical points cannot be a local minimum, leaving a maximum number of 7 local minima. □

Note that due to its rational form, the inferred measurement (50) is symmetric with respect to the sensor, which is not the case for the original bearing measurement (7). This can result in more local minima of (51) than those of (28). To discard the spurious local minima due to the symmetry of the inferred measurement, we employ the Mahalanobis distance test [4]. As a result, we have never observed more than *four* local minima in practice.

5 Bank-of-MAP Estimator

As discussed in the preceding section, due to the nonlinearity of range and bearing measurements, the incremental one-step MAP problem (25), and thus the original batch multi-step MAP problem (11), may have multiple local minima that correspond to the modes of the posterior pdf. Any iterative algorithm such as Gauss-Newton used in the batch-MAP estimator is only able to converge to the global optimum (i.e., the true MAP estimate), when the initial estimate $\hat{\mathbf{x}}^{(0)}_{0:k|k}$ is within its region of attraction. However, in general, there exists *no* systematic method for determining an initial estimate that can always ensure convergence to the global optimum. As a result, the standard batch-MAP estimator when used for target tracking can become inconsistent and even diverge if no good initial estimate is provided. This is confirmed by the simulation and experimental results presented in Sections 6.

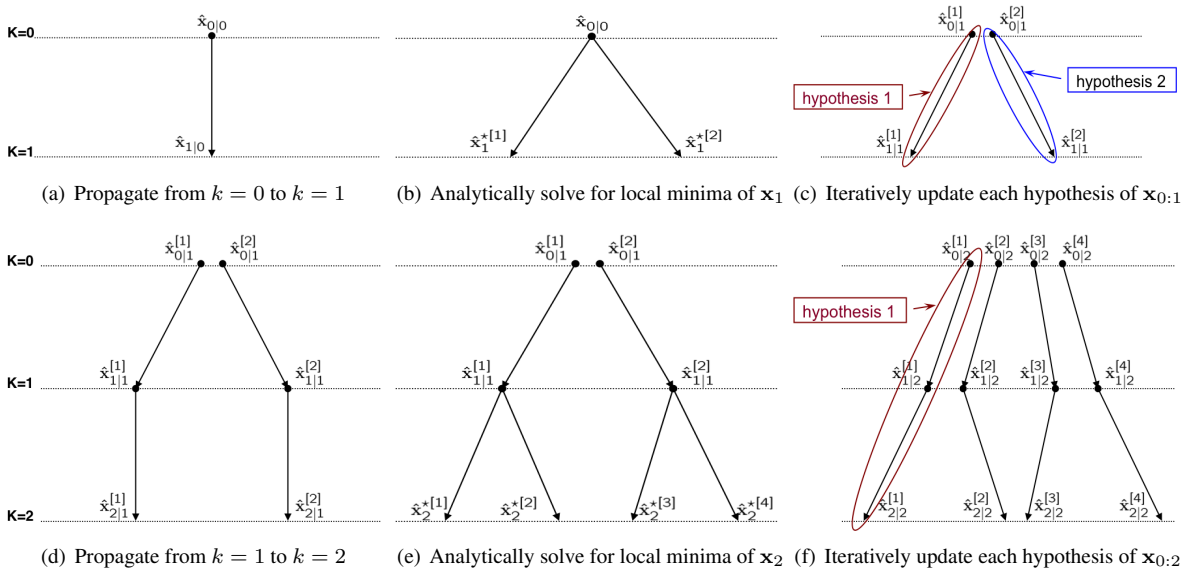
5.1 The Proposed Algorithm

To mitigate the aforementioned issue, in the following, we propose a general estimation framework for tracking multiple local minima (modes). Within this framework, we develop a bank-of-MAP estimator for the particular problem of target tracking. The key idea of our approach is to use the analytically-computed local minima at each time step (see Section 4.2) as guidance to find and track the most probable hypotheses of the target-state trajectory, thus improving estimation performance.

Specifically, at time-step $k-1$, based on (23) and (24), we first propagate the current state estimate corresponding to the i-th solution and its covariance matrix, $\hat{\mathbf{x}}^{[i]}_{k-1|k-1}$ and $\mathbf{P}^{[i]}_{k-1|k-1}$, $i = 1, \ldots, m$, where m is the number of estimators in the bank at time-step $k-1$. Then, once a new measurement becomes available, the propagated state estimate and covariance, $\hat{\mathbf{x}}^{[i]}_{k|k-1}$ and $\mathbf{P}^{[i]}_{k|k-1}$, are used as the prior in (25). Next, we use the algebraic-geometry methods presented in Section 4.2 to analytically determine all the local minima of (25), denoted by $\hat{\mathbf{x}}^{\star[j]}_{k}$, $1 \leq j \leq \alpha m$ (see Lemmas 4.1 and 4.2, $\alpha = 2$ and 7, respectively). Finally, for each of these solutions, we employ Gauss-Newton iterations to refine the *entire* state estimates up to current time-step k, $\hat{\mathbf{x}}^{[j]}_{0:k|k}$ [see (11)]. In particular, the iterative Gauss-Newton approach uses the latest estimate of the trajectory from time-step zero to $k-1$ augmented by the analytically-computed local minimum at time-step k, $\begin{bmatrix} \hat{\mathbf{x}}^{[j]}_{0:k-1|k-1} \\ \hat{\mathbf{x}}^{\star[j]}_{k} \end{bmatrix}$, as the initial value, and processes all the available original bearing or distance measurements. Fig. 3 visualizes this process by considering the first two time steps, where we assume $\alpha = 2$ for every time step. We should stress that at each time step, it is the analytically-computed local minima that guide the selection of initial values for the Gauss-Newton-based batch-MAP estimators. It should also be noted that at very beginning the proposed estimator can be initialized, either based on *a priori* knowledge about the target state, or based on the first set of measurements, for example, by employing the geometrically-constrained optimization approaches proposed in [40–42].

This procedure recursively evolves over time, and at every time step, generates at most αm trajectory estimates. In the end, we will have multiple candidates of the batch-MAP estimate, among which the one with the least cost [corresponding to the most probable hypothesis (see (10) and (11))] is selected as the best estimate for the global optimum and thus for the true state. Algorithm 1 summarizes the main steps of the proposed bank-of-MAP estimator for target tracking.

[4] For a function $F : \mathbb{R}^N \to \mathbb{R}$ to be coercive, it is bounded from below and is proper in the sense that $F(\mathbf{x}) \to \infty$ for $||\mathbf{x}|| \to \infty$.



(a) Propagate from $k = 0$ to $k = 1$ (b) Analytically solve for local minima of $\mathbf{x}_1$ (c) Iteratively update each hypothesis of $\mathbf{x}_{0:1}$

(d) Propagate from $k = 1$ to $k = 2$ (e) Analytically solve for local minima of $\mathbf{x}_2$ (f) Iteratively update each hypothesis of $\mathbf{x}_{0:2}$

Figure 3: Visualization of the proposed bank-of-MAP estimator for target tracking. In this example, we consider only the first two time steps to illustrate the key ideas of the proposed algorithm, and assume there are two analytical local minima ($\alpha = 2$) at every time step. In these plots, (a)-(c) show the three main steps (i.e., propagation of current state estimate, analytical determination of local minima, and iterative Gauss-Newton batch update of all hypotheses) for time-step $k = 1$, and (d)-(f) show the same steps for time-step $k = 2$. It is clear that due to the existence of multiple local minima, at each time step we have multiple hypotheses (i.e., 2 for $k = 1$, and 4 for $k = 2$) for the entire trajectory. Note also that since we in fact solve different batch-MAP problems at different time steps due to different cost functions [see (**??**)], the optimal solution at one time is not necessarily optimal for the next time, i.e., if hypothesis 1 is the best at $k = 1$, it may not be the best at $k = 2$ (e.g., see Fig. 4). Simply tracking one of the available hypotheses (even if the best one) at every time step may not be sufficient, and thus we aim at tracking the most probable several (if not all) hypotheses to ensure improved performance.

5.2 Computational Cost Reduction

In the worst case, the total number of analytical solutions, and thus MAP estimators in the bank, grows exponentially with time. In addition, as the target continuously moves, the size of the state vector $\mathbf{x}_{0:k}^{[i]}$ of each MAP estimator increases linearly with time. In order to make the algorithm suitable for real-time applications, in what follows, we present an effective pruning scheme, as well as the process of marginalization of old, "matured" states,[5] to reduce the computational cost of the proposed algorithm.

5.2.1 Pruning least probable hypotheses

In practice, the number of physically different trajectory hypotheses is significantly lower than the exponential number of hypotheses generated by the estimator, since many different initial guesses reside within the same basin of attraction. Additionally, we observe that in general, if two MAP estimators in the bank have similar costs, the trajectory estimates are also close (e.g., see Fig. 4). Therefore, we first aggregate the trajectory estimates of which the corresponding costs are equal within a tolerance, and retain one representative trajectory of each such group while discarding the others. In addition, we also employ the k-means algorithm [43] to cluster the remaining estimated trajectories into two groups based on their costs and remove the (outlier) group which has larger costs. These two steps, aggregation and clustering, are repeated, until the number of MAP estimators in the bank is within a threshold denoted by m_{max}, even though simulation results have shown that the aggregation is so effective that most of the time there is no need to perform the clustering. Note that we have opted for pruning hypotheses based on the batch-MAP costs, instead of the high-dimensional state-trajectory estimates, in part because the one-dimensional (scalar) costs are more well-suited for fast aggregation and clustering.

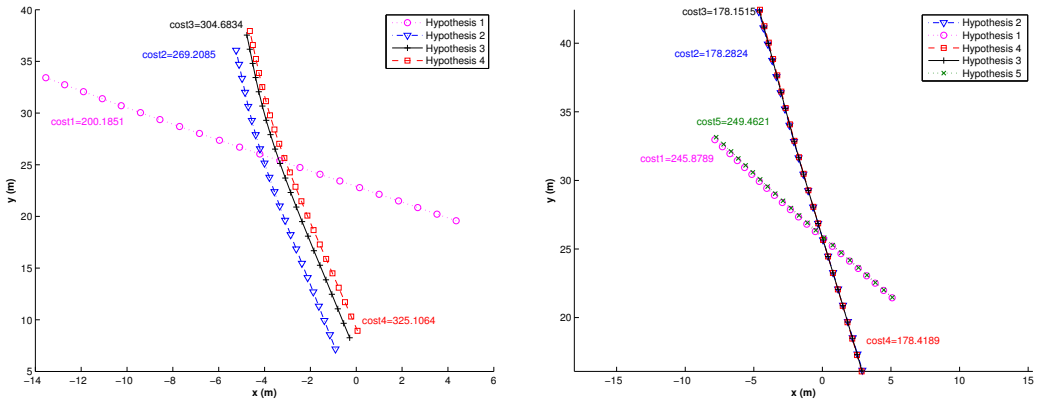
[5] Here "matured" refers to past states which will not be significantly affected by a new measurement available at the current target's position.

Algorithm 1 A bank-of-MAP estimator for target tracking

Require: Initial state estimate and covariance
At each time-step k:

- Propagate the current target state estimate and covariance via (23) and (24).
- Analytically determine all the local minima of (25).
- For each of the local minima, refine the corresponding state-trajectory estimates and covariance, by employing the Gauss-Newton iterations that use the latest state estimates corresponding to this minimum as the initial guess, and compute the batch-MAP cost (11).

In the end, select the estimate in the bank with the least cost as the resulting MAP estimate.



(a) $k = 22$ (b) $k = 23$

Figure 4: Illustration of multiple hypotheses in the proposed bank-of-MAP estimator evolving over time. In this case, a typical scenario for range-only tracking, due to multiple local minima found by the proposed analytical approach, two hypotheses (1 and 5) are generated at $k = 23$ corresponding to hypothesis 1 at $k = 22$. It becomes clear that if hypotheses have similar costs [e.g., hypotheses 2, 3 and 4, or hypotheses 1 and 5 in (b)], the corresponding target's trajectory estimates are close. Note that some estimates are so close that the corresponding lines are difficult to distinguish. Note also that hypothesis 1 at $k = 22$ has the lower cost than the other hypotheses, while it (as well as hypothesis 5) becomes less probable (of higher cost) at $k = 23$.

5.2.2 Marginalizing old states

To further reduce the computational complexity, we employ a marginalization process that removes the old, matured states from the state vector of each batch-MAP estimator in the bank (see [31]). In particular, suppose that we currently have the active states $\mathbf{x}_{k_1:k_3} = \mathbf{x}_{k_1:k_2} \cup \mathbf{x}_{k_2+1:k_3}$, and want to marginalize out $\mathbf{x}_{k_1:k_2}$. To do so, we employ the Schur complement to remove $\mathbf{x}_{k_1:k_2}$ from the state vector and discard all the measurements involving them [31]. However, since some of the discarded measurements also involve the remaining states, $\mathbf{x}_{k_2+1:k_3}$, we should retain and utilize such information in the future batch-MAP estimation after marginalization.

Toward this end, we note first that the batch-MAP cost function can be decomposed as follows [see (11)]:

$$c(\mathbf{x}_{k_1:k_3}) = c(\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1:k_3}) = c_m(\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1}) + c_c(\mathbf{x}_{k_2+1:k_3}) \tag{65}$$

where c_m consists of all the terms involving the marginalized states $\mathbf{x}_{k_1:k_2}$ (i.e., corresponding to the discarded measurements), while all the remaining terms comprise c_c. It is important to note that due to the target's Markov motion model (3), there is a term in c_m, $||\mathbf{x}_{k_2+1} - \mathbf{\Phi}_{k_2}\mathbf{x}_{k_2}||^2_{\mathbf{Q}'_{k_2}}$, also involving $\mathbf{x}_{k_2+1}$, and this is why it is explicitly highlighted in (65). To perform marginalization, we express the minimization problem as two nested ones as:

$$\min_{\mathbf{x}_{k_1:k_3}} c(\mathbf{x}_{k_1:k_3}) = \min_{\mathbf{x}_{k_2+1:k_3}} \left(c_c(\mathbf{x}_{k_2+1:k_3}) + \min_{\mathbf{x}_{k_1:k_2}} c_m(\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1}) \right) \tag{66}$$

Then, we first solve $\arg\min_{\mathbf{x}_{k_1:k_2}} c_m(\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1})$ based on Gauss-Newton iterations, i.e., by minimizing its second-order Taylor-series expansion [see (12)], and the optimal solution is computed as [44]:

$$\mathbf{x}_{k_1:k_2} = \hat{\mathbf{x}}_{k_1:k_2|k_2} - \mathbf{A}_{11}^{-1}\left(\mathbf{b}_1 + \mathbf{A}_{12}(\mathbf{x}_{k_2+1} - \hat{\mathbf{x}}_{k_2+1|k_2})\right) \tag{67}$$

where we have used the following decompositions of the Hessian and Jacobian matrices corresponding to $\mathbf{x}_{k_1:k_2}$ and $\mathbf{x}_{k_2+1}$ [see (19) and (15)]:

$$\mathbf{A} \triangleq \nabla^2_{\{\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1}\}} c_m(\cdot) = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$\mathbf{b} \triangleq \nabla_{\{\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1}\}} c_m(\cdot) = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

It is clear that this optimal solution (67) depends *only* on $\mathbf{x}_{k_2+1}$, and substituting it in (66) results in an approximately equivalent cost function that is *independent* of $\mathbf{x}_{k_1:k_2}$, denoted by $c'_c(\mathbf{x}_{k_2+1:k_3})$, in analogy to, but *different* from $c_c(\mathbf{x}_{k_2+1:k_3})$. The difference is that $c'_c(\mathbf{x}_{k_2+1:k_3})$ now also incorporates the information about $\mathbf{x}_{k_2+1}$ retained from the discarded measurements involving the marginalized states $\mathbf{x}_{k_1:k_2}$. On the other hand, $c_m(\mathbf{x}_{k_1:k_2}, \mathbf{x}_{k_2+1})$ in (66) is permanently approximated by its second-order Taylor-series expansion at $\hat{\mathbf{x}}_{k_1:k_2|k_2}$ and becomes a function of $\mathbf{x}_{k_2+1}$.

From this point on, i.e., after marginalization is complete, we solve $\arg\min_{\mathbf{x}_{k_2+1:k_3}} c'_c(\mathbf{x}_{k_2+1:k_3})$ for the MAP estimates, using the Gauss-Newton method (see Section 3.3 and [44]).

The advantage the marginalization brings in is that each batch-MAP estimator in the bank has constant computational requirements, which depend linearly on the fixed size of the sliding window. This, along with pruning (see Section 5.2.1), results in *constant* computational complexity for the proposed bank-of-MAP estimator, compared to *linear* for the standard batch-MAP estimator (see Section 3.3). Note also that in order to further speed up the batch-MAP solver, an incremental, instead of batch, Gauss-Newton iterations can be used [45], which, however, provides only an approximate solution (besides the approximation due to linearization errors).

6 Simulation Results

To validate the capability of the proposed bank-of-MAP estimator to improve tracking performance, we performed 500 Monte Carlo simulations, and compared four different estimators. During each Monte Carlo run, all the estimators process the same data, to ensure a fair comparison. The compared estimators are: (i) the standard EKF, (ii) the standard batch-MAP estimator that incrementally uses the EKF estimates (i.e., the current EKF estimate along with the MAP estimates of the past states) as the initial value, and employs the same marginalization process as the proposed bank-of-MAP estimator, (iii) the sampling importance resampling (SIR)-PF with 3000 particles [11], and (iv) the proposed bank-of-MAP estimator with pruning ($m_{\max} = 10$) and marginalization (sliding window of 25 time steps). Note that in both MAP estimators, the maximum number of Gauss-Newton iterations allowed was set to 20. Note also that depending on different resampling schemes and different proposal distributions used by the PF, many variants exist (e.g., auxiliary PF, regularized PF, likelihood PF, etc.). Interested readers are referred to [11,27,46,47] for an overview of the PF. In this simulation, we implemented the standard (bootstrap) SIR-PF [25] that uses the prior distribution as the proposal distribution to draw particles and employs systematic resampling at every time step. Moreover, in order to alleviate the particle depletion issue, we have dithered the sensor noise (i.e., increased the measurement noise covariance). In addition, we have examined different resampling schemes such as Ripley's and stratified resampling [27,28], while finding negligible performance difference.

For the results presented in this section, we adopted a zero-acceleration motion model for the target [4]:

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}\mathbf{w}(t) \tag{68}$$

where

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ \mathbf{G} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},\ \mathbf{x}(t) = \begin{bmatrix} x_T(t) \\ y_T(t) \\ \dot{x}_T(t) \\ \dot{y}_T(t) \end{bmatrix}$$

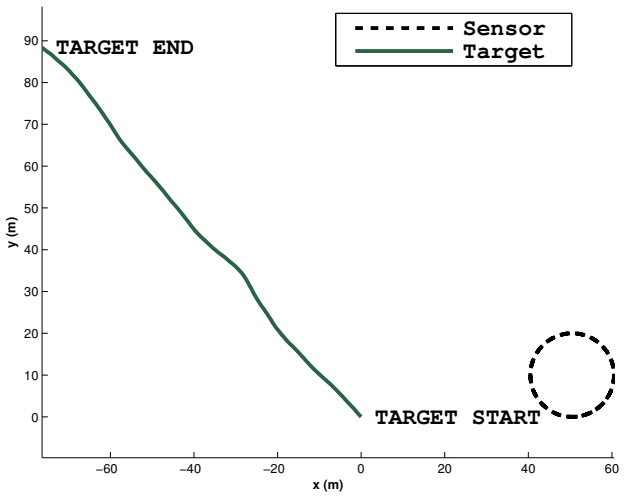


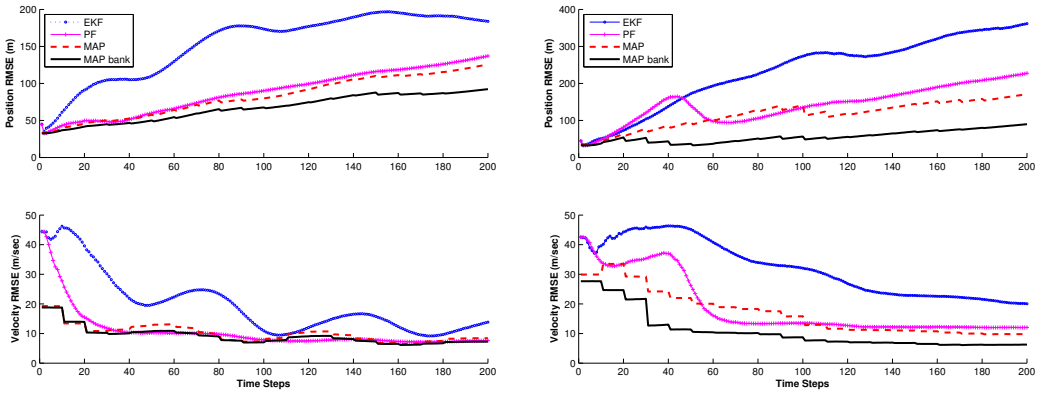
Figure 5: The trajectories of the target and the mobile sensor obtained from one typical realization of the 500 Monte Carlo simulations.

and $\mathbf{w}(t) = \begin{bmatrix} w_x(t) & w_y(t) \end{bmatrix}^T$ is zero-mean white Gaussian noise with covariance $\mathbb{E}\left[\mathbf{w}(t)\mathbf{w}(\tau)^T\right] = q\mathbf{I}_2\delta(t-\tau)$, where $q = 2\left(\frac{\text{m}}{\text{sec}^2}\right)^2 \frac{1}{\text{Hz}}$, and $\delta(t-\tau)$ is the Dirac delta function. In the implementation, we discretize this continuous-time system model (68) with time step $\delta t = 0.1$ sec. The initial true target state is $\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & -5 & 5 \end{bmatrix}^T$, while the initial estimate of the target state is randomly generated from a prior Gaussian pdf, $\mathcal{N}(\mathbf{x}_0, \mathbf{P}_{0|0})$, where $\mathbf{P}_{0|0} = 10^3\mathbf{I}_4$ is the initial covariance of the state estimate. Similar to [19], we chose a circular sensor trajectory with perfectly known poses for this test. Fig. 5 shows the trajectories of the target and the mobile sensor in one typical realization of Monte Carlo simulations.

6.1 Scenario I: Large noise

We first present the comparison results for both range-only and bearing-only target tracking in the case of large sensor noise. Specifically, the standard deviation of the distance-measurement noise was equal to 10% of the sensor-to-target distance, while the standard deviation of the bearing-measurement noise was set to 10 deg. These adverse sensor-noise levels selected are larger than what is typically encountered in practice, despite the fact that the PF is favorable to large, rather than small, noise. This was done purposefully, since higher noise levels lead to larger estimation errors, which can make the comparison more apparent.

In particular, Fig. 6 shows the average root mean squared errors (RMSE) of the four estimators, and Fig. 7 depicts the normalized (state) estimation error squared (NEES), a metric for evaluating the estimator's consistency [4]. These results are computed by averaging the corresponding RMSE and NEES over all the Monte-Carlo runs. As evident from this figure, the standard EKF estimates are inaccurate, and diverge from the ground truth. The standard batch-MAP estimator, incrementally using the EKF estimate as the initial guess, has significantly improved performance compared to the EKF, mostly due to the continuous relinearization of the past trajectory. As expected, the PF has also attained better estimation accuracy than the EKF. This is due to the fact that each particle in the PF essentially represents a hypothesis of the target state, and thus the PF is more likely to converge to the optimal solution. However, it does not always work as well as the standard batch-MAP estimator (see Fig. 6), in part because it does not allow smoothing the old state estimates using newly available measurements. Although particle-based smoothers exist, due to the curse of dimensionality, their computational requirements are significantly higher [27]. Note also that the NEES of the PF is not necessarily better than that of the EKF, primarily due to the numerical issue incurred in the simulation that the covariance matrices of the PF computed from particles become ill-conditioned. On the other hand, the performance of the proposed bank-of-MAP estimator is substantially better than that of the competing approaches (i.e., the standard EKF, the PF, and the standard batch-MAP), in terms of both accuracy (RMSE) and consistency (NEES), which is attributed to the accurate initial estimates obtained analytically through the algebraic methods presented in Section 4.2.



(a) Range-only tracking (b) Bearing-only tracking

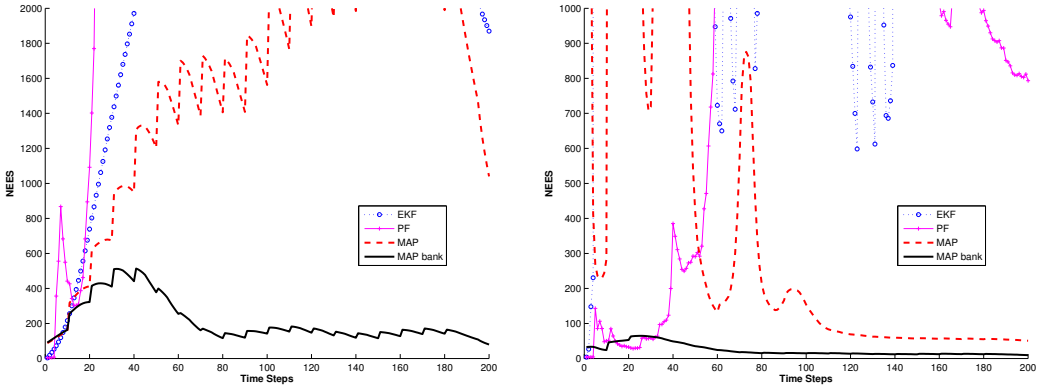
Figure 6: Monte-Carlo simulation results of average RMSE. In these plots, the dotted lines with circles correspond to the standard EKF, the solid lines with crosses to the bootstrap PF, the dashed lines to the standard batch-MAP estimator, and the solid lines to the proposed bank-of-MAP estimator. It is clear that the proposed algorithm performs more accurately than its competitors in both range-only and bearing-only target tracking.

6.1.1 Computational cost

Using the same simulation setup as described above, we have also compared the computational requirements of the proposed bank-of-MAP estimator against the three competing algorithms. We counted the CPU running time for a complete update of the EKF, the PF, the batch-MAP estimator, and the proposed bank-of-MAP estimator (including the analytical determination of all local minima, batch-MAP refinement, pruning, and marginalization). Our Matlab implementation running on a Core i7 CPU of 2.90 GHz required an average execution time for each estimator shown in Table 1. These results were obtained by averaging the CPU running time over all Monte Carlo runs and over all time steps. As expected, the EKF and the standard MAP estimator, which only track a single hypothesis of the target's trajectory, are computationally more efficient than both the PF and the proposed bank-of-MAP estimator which instead track multiple hypotheses of the target's trajectory. However, their tracking performance is substantially worse than the proposed algorithm. Moreover, as compared to the PF, the proposed bank-of-MAP estimator not only is less computationally demanding, but also achieves significantly better performance (see Fig. 6 and Table 1). Specifically, in the case of range-only tracking, as compared to the PF, the bank-of-MAP estimator reduces the position error by 25% on average and the velocity error by 10%, at 50% less computational cost. Similarly, in the case of bearing-only tracking, it reduces the position and velocity errors on average by 60% and 40%, respectively, at 20% less computational cost.

6.1.2 PFs with different numbers of particles

In order to further compare the two multi-hypothesis tracking approaches: the PF and the proposed bank-of-MAP estimator, we have performed various 50 Monte-Carlo simulations of range-only tracking, using the same setup as before but different numbers of particles (i.e., using 1000, 4000, 8000, and 12000 particles, respectively). Fig. 8 shows the comparison results. As expected, as more particles are used, the PF performs better; however, such performance gain is offsetted by the increased computational expense, in particular, when using a large number of particles. For example, in this test, the improvement of estimation accuracy is marginal from using 8000 to 12000 particles [see Fig. 8(a)], while the increase of computational overhead is significant [see Fig. 8(b)]. Interestingly, the PF outperforms the proposed bank-of-MAP when using 8000 or more particles, while this superior performance comes with a price on substantially higher computational cost. It becomes clear that the proposed bank-of-MAP estimator is more favorable than the PF with respect to the trade-off between the estimation accuracy and the computational cost, yielding better qualitative performance with affordable computational effort.



(a) Range-only tracking

(b) Bearing-only tracking

Figure 7: Monte-Carlo simulation results of average NEES. In these plots, the dotted lines with circles correspond to the standard EKF, the solid lines with crosses to the bootstrap PF, the dashed lines to the standard batch-MAP estimator, and the solid lines to the proposed bank-of-MAP estimator. It is clear that the proposed algorithm achieves better consistency than its competitors in both range-only and bearing-only target tracking. Note that for clarity of presentation, only portions of the NEES lines that are within certain thresholds are plotted.

6.2 Scenario II: Small noise

In this test, we compared the same four estimators as before in the scenario where small sensor noise is used. Specifically, the standard deviation of the distance-measurement noise was equal to 1% of the sensor-to-target distance, and the standard deviation of the bearing-measurement noise was 1 deg. The same setup as in the preceding test is employed, while the simulation duration is doubled and 50 Monte-Carlo simulations are performed. Fig. 9 shows the comparative results, which clearly follow the same trends as those in the case of large noise: The proposed bank-of-MAP algorithm outperforms the other estimators in both range-only and bearing-only tracking.

7 Conclusions

In order to improve its performance, a nonlinear estimator should be able to track multimodal pdfs which often occur in nonlinear problems. However, this is not the case for many existing estimators (e.g., the EKF, and the MAP estimator). In this paper, we have introduced a general estimation framework, a bank-of-MAP estimator, which simultaneously allows tracking multiple modes of the posterior pdf, and reduces linearization errors through relinearization of past measurements. We have applied it to the particular problems of both range-only and bearing-only target tracking. Due to the computational intractability of analytically solving the batch-MAP problem, we have employed a relaxation scheme that keeps past state estimates temporarily constant and incrementally solves a one-step minimization problem for the current state at every time step. This minimization is solved analytically and thus efficiently using algebraic geometry methods. The analytically-computed local minima are then used to find accurate initial values for the bank-of-MAP estimator, thus focusing the available resources on tracking the most probable hypotheses of the target's trajectory. Additionally, to reduce the computational cost of the proposed algorithm, we have employed hypothesis pruning along with marginalization of old states. Simulation and experimental results have shown that the proposed algorithm significantly outperforms the standard EKF, the batch-MAP estimator, as well as the SIR-PF.

Appendices

A Proof of Lemma 4.1

We first note that the following finite dimensional Mountain Pass Theorem (MPT) will be useful for the ensuing proof.

Table 1: Monte-Carlo simulation results of computational cost versus estimation accuracy

	Runtime (sec)	Pos. Est. Err. (m)	Vel. Est. Err. (m/sec)
Range-only tracking			
EKF	0.0014	153.8764	19.7173
MAP	0.0569	82.0346	10.6915
PF	0.4756	88.1905	10.1034
MAP bank	0.2447	66.6276	9.1824
Bearing-only tracking			
EKF	0.0015	233.7407	32.0778
MAP	0.0695	116.0235	16.9055
PF	0.6289	143.0785	18.7006
MAP bank	0.5133	56.0722	10.7231

Theorem A.1. [Theorem 5.2, [35]] *Suppose that a continuous function $f \in \mathcal{C}^1(\mathbb{R}^N;\mathbb{R})$ is coercive and possesses two distinct strict relative minima $\mathbf{x}_1$ and $\mathbf{x}_2$.*[6] *Then f possesses a third critical point $\mathbf{x}_3$, which is distinct from $\mathbf{x}_1$ and $\mathbf{x}_2$, and characterized by:*

$$f(\mathbf{x}_3) = \inf_{\Sigma\in\Gamma}\max_{\mathbf{x}\in\Sigma} f(\mathbf{x}) \tag{69}$$

where $\Gamma = \{\Sigma \subset \mathbb{R}^N; \Sigma$ is compact and connected, and $\mathbf{x}_1, \mathbf{x}_2 \in \Sigma\}$. Moreover, $\mathbf{x}_3$ is not a relative minimizer; that is, in every neighborhood of $\mathbf{x}_3$, there exists a point $\mathbf{x}$ such that $f(\mathbf{x}) < f(\mathbf{x}_3)$.[7]

To preserve the clarity of presentation, without loss of generality, we hereafter translate the global frame of reference to the sensor's local frame (i.e., $\mathbf{p}_S = \mathbf{0}$), and also drop the time indices as well as the subscript " T " denoting target. Thus, the minimization of (28) can be written as:

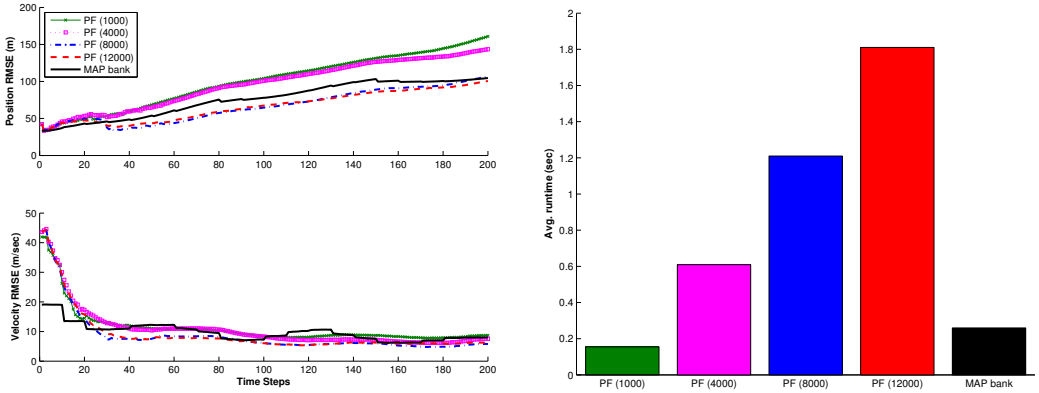
$$\begin{aligned}\min_{\mathbf{p}} c(\mathbf{p}) &= \frac{1}{2}(\mathbf{p}-\hat{\mathbf{p}})^T\mathbf{P}_{\mathbf{pp}}^{-1}(\mathbf{p}-\hat{\mathbf{p}}) + \frac{1}{2\sigma_\rho^2}\left(z-\sqrt{x^2+y^2}\right)^2 \\ &\triangleq c_1(\mathbf{p}) + c_2(\mathbf{p})\end{aligned} \tag{70}$$

Recall that $\mathbf{p} = \begin{bmatrix} x \\ y \end{bmatrix}$ denotes the target position. In (70), given a function value $c(x,y) \triangleq \alpha$, $c_1(x,y) \triangleq \beta$ is an ellipse centered at $\hat{\mathbf{p}} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}$, while depending on the available measurement value, $c_2(x,y) = \alpha - \beta \triangleq \gamma$ represents one or two circles centered at the origin, i.e., $x^2 + y^2 = (z \pm \sigma_\rho\sqrt{2\gamma})^2$. For simplicity of analysis, in the following, we assume only one circle is associated with $c_2(x,y)$, while the analysis readily holds for two circles. Notice that the cost function $c(\mathbf{p})$ (70) is coercive but *not* $\mathcal{C}^1$ since it is not differentiable at the origin. Hence Theorem A.1 is not applicable directly.

Nevertheless, in what follows, we first show that the inf-max point cannot be the origin. Based on that, we will use the MPT to show that there are at most 3 local minima, since $c(\mathbf{p})$ is coercive *and* $\mathcal{C}^1$ in $\mathbb{R}^2\backslash\{\mathbf{0}\}$. Subsequently, we show that there are at most 2 local minima to complete the proof.

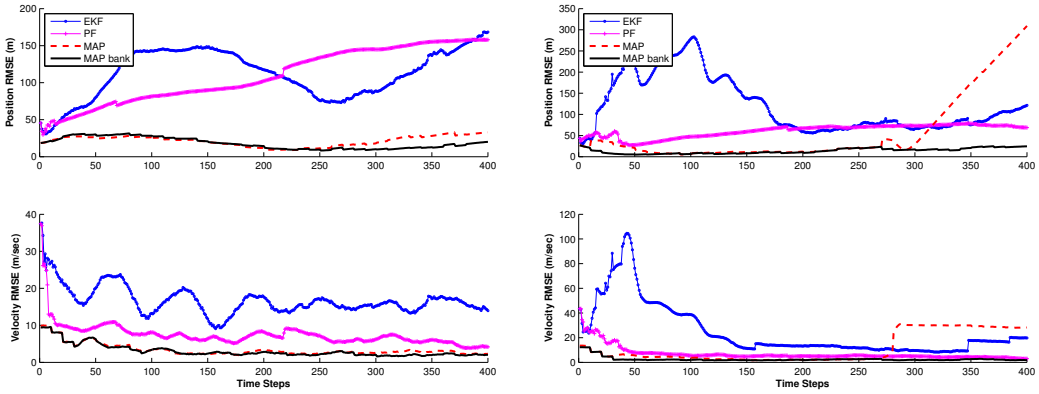
[6] A function $f : \mathbb{R}^N \to \mathbb{R}$ is coercive iff it is bounded from below and is proper in the sense that $f(\mathbf{x}) \to \infty$ for $||\mathbf{x}|| \to \infty$.

[7] Γ is the class of paths (or curves) connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ (see Theorem 1.1, Ch. II, [48]). Note that, based on the proof of this theorem (see [35]), such a critical point $\mathbf{x}_3$ (69) also exists for $f \in \mathcal{C}^0(\mathbb{R}^N;\mathbb{R})$, though, in this case, $\mathbf{x}_3$ is not necessary to be a non-minimum point. It is also important to notice that, an isolated inf-max critical point (i.e., mountain pass point), $\mathbf{x}_3$, is necessarily a saddle point (see Ch. 12, [35]).



(a) Average RMSE (b) Average runtime of each step

Figure 8: Comparison results of the proposed bank-of-MAP and the PF with different numbers of particles in the case of range-only tracking. These results suggest that the proposed estimator performs favorably against the standard PF in terms of the trade-off between estimation accuracy and computational cost.



(a) Range-only tracking (b) Bearing-only tracking

Figure 9: Monte-Carlo simulation results of average RMSE. In these plots, the solid lines with circles correspond to the standard EKF, the solid lines with crosses to the bootstrap PF, the dashed lines to the standard batch-MAP estimator, and the solid lines to the proposed bank-of-MAP estimator. It is clear that the proposed algorithm performs more accurately than its competitors in both range-only and bearing-only target tracking.

A.1 Proof that there are at most 3 local minima

Remember that there are up to 4 distinct critical (stationary) points in $\mathbb{R}^2\backslash\{\mathbf{0}\}$ (see Section **??**). Suppose that they are all strict local minima, denoted by $\mathbf{p}_i$, $i = 1, \ldots, 4$. By the MPT (Theorem A.1) in $\mathbb{R}^2\backslash\{\mathbf{0}\}$, we have an inf-max point $\mathbf{p}$ for any pair of $\mathbf{p}_i$. If this $\mathbf{p}$ is different from any $\mathbf{p}_i$, we will have (at least) 5 critical points, which contradicts the fact that there are at most 4 critical points in $\mathbb{R}^2\backslash\{\mathbf{0}\}$. Therefore, any pair of $\mathbf{p}_i$ must share a common inf-max point at the origin which is a non-differential critical point. However, we now show by contradiction that this is not the case.

Specifically, suppose that the origin, $\mathbf{p} = \mathbf{0}$, is the desired common inf-max point. From (69), we know that $c(\mathbf{p}_i) < c(\mathbf{0}) \triangleq \alpha$, $i = 1, \ldots, 4$. By continuity of $c(\cdot)$, we define the following level set:

$$\mathcal{S} = c^{-1}(\alpha) = \{\mathbf{p} \in \mathbb{R}^2 : c(\mathbf{p}) = c(\mathbf{0})\} \tag{71}$$

We now show some important properties of this level set $\mathcal{S}$ that will be useful for our proof. First of all, from the implicit function theorem (see Lemma 4 of Ch. 2 in [49]), $\mathcal{S}\backslash(0,0)$ is a smooth 1-dimensional manifold. A smooth connected 1-dimensional manifold is diffeomorphic either to a circle or to some interval of real numbers (see Appendix

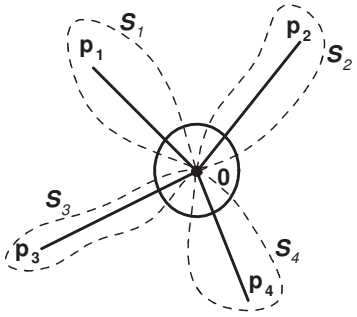


Figure 10: Illustration of the topological configuration of four local minima, $\mathbf{p}_i$ ($i = 1, \ldots, 4$), and the level set $\mathcal{S}$. In this plot, the dashed curves represent the level set $\mathcal{S}$, and the solid lines are the paths which connects the local minima $\mathbf{p}_i$ and $\mathbf{p}_j$ and attains its maximum at the origin. Note that each of the paths is contained exactly in one interior of the connected components of $\mathcal{S}$. It is clear that the circle $c_2(\mathbf{p})$ intersects $\mathcal{S}$ at 8 different points.

of [49]).[8] Thus, from differential topology [49], a 1-dimensional manifold and thus $\mathcal{S}\backslash(0,0)$ is a *union of disjoint smooth curves (lines and circles)*. This also implies that $\mathcal{S}\backslash(0,0)$ has no isolated points.

Secondly, $\mathcal{S}\backslash(0,0)$ has *no closed* curves. To see this, by contradiction, suppose that one curve component of $\mathcal{S}\backslash(0,0)$ is closed, denoted by $\mathcal{S}_1$. By continuity of $c(\cdot)$ and compactness of $\mathcal{S}_1$, and based on the Weierstrass theorem [50], there will exist (at least) one local minimum inside $\mathcal{S}_1$, which has to be one of the points $\mathbf{p}_i$ since no other critical point exists. Note that such a local minimum cannot be on the boundary of $\mathcal{S}_1$ since $c(\mathbf{p}_i) < \alpha$. Without loss of generality, we assume $\mathbf{p}_1$ is the only local minimum inside $\mathcal{S}_1$. Note that there cannot be more than one local minima inside $\mathcal{S}_1$ (see the next paragraph below). Based on continuity of $c(\cdot)$ and compactness of $\mathcal{S}_1$, we can find a sufficiently small $\epsilon > 0$ so that $\mathcal{S}_1$ is contained in one closed component of $c^{-1}(\alpha+\epsilon)$, which is disjoint from $\mathcal{S}\backslash\mathcal{S}_1$ as they have different function values ($\alpha+\epsilon$ and α, respectively). Clearly, in this case, the inf-max value between $\mathbf{p}_1$ and $\mathbf{p}_i$ ($i \neq 1$) will be larger than α and is attained at a point other than $(0,0)$, which gives the contradiction. Therefore, all curves of $\mathcal{S}\backslash(0,0)$ are open-ended. Moreover, by continuity, $c(\cdot)$ attains α at any limit point (i.e., the open-end point) of the curves of $\mathcal{S}\backslash(0,0)$. However, $(0,0)$ is the only point where $c(\cdot)$ attains α except the curves of $\mathcal{S}\backslash(0,0)$, and thus the *only* common limit of all open-ended curves of $\mathcal{S}\backslash(0,0)$.

Lastly, let us consider the properties of the interior of $\mathcal{S}$. We first define the following set $\boldsymbol{\Theta}$ whose boundary is $\mathcal{S}$, i.e., $\boldsymbol{\Theta} = \{\mathbf{p} \in \mathbb{R}^2 : c(\mathbf{p}) < \alpha\} = \cup_j \boldsymbol{\Theta}_j$, where $\boldsymbol{\Theta}_j$ is j-th component of $\boldsymbol{\Theta}$. It is clear that $\mathbf{p}_i$ ($i = 1, \ldots, 4$) is contained in $\boldsymbol{\Theta}_j$ for some j, since $c(\mathbf{p}_i) < \alpha$. Note that, $\boldsymbol{\Theta}_j$ cannot contain (more than) two $\mathbf{p}_i$, since, otherwise, the inf-max point of these two $\mathbf{p}_i$ will be different from $(0,0)$ and will have a function value less than α. On the other hand, if $\boldsymbol{\Theta}_j$ does not contain any $\mathbf{p}_i$, similarly, based on the Weierstrass theorem [50], there will exist a new local minimum in the corresponding closure, $\bar{\boldsymbol{\Theta}}_j = \boldsymbol{\Theta}_j \cup \partial\boldsymbol{\Theta}_j$. This contradicts the fact that we have only 4 critical points different from $(0,0)$. Therefore, there are exactly 4 connected components $\boldsymbol{\Theta}_i$, each of which contains one $\mathbf{p}_i$. Furthermore, by continuity of $c(\cdot)$, the i-th boundary, $\partial\boldsymbol{\Theta}_i \subset \mathcal{S}$, contains at least one curve of the connected components of $\mathcal{S}\backslash(0,0)$. Since $(0,0)$ is a limit point of all curves, $(0,0) \in \partial\boldsymbol{\Theta}_i$ for every i, and thus,

$$\partial\boldsymbol{\Theta}_i = \cup_{j=1}^{n_i} \mathcal{S}_j \cup \{(0,0)\} \tag{72}$$

where $\mathcal{S}_j$ is one curve of $\mathcal{S}\backslash(0,0)$ and n_i is the number of $\mathcal{S}_j$ contained in $\partial\boldsymbol{\Theta}_i$. Note that any two $\partial\boldsymbol{\Theta}_i$ do not share a common curve, since if they do, by definition, the shared curve will consist of (infinitely many) critical points. As a result, there are at least 4 distinct open curves in $\mathcal{S}\backslash(0,0)$ having $(0,0)$ as a common limit point (see Fig. 10).

As seen from Fig. 10, for a given measurement z, the circle $c_2(\mathbf{p})$ intersects the level set $\mathcal{S}$ at (at least) 8 intersection points denoted by $\mathbf{q}_i$, $i = 1, \ldots, 8$. We can show that all these 8 points belong to the same ellipse $c_1(\mathbf{p})$, by noting that

$$\begin{aligned} c(\mathbf{q}_i) =& c_1(\mathbf{q}_i) + c_2(\mathbf{q}_i) \\ \Rightarrow\ c_1(\mathbf{q}_i) =& c(\mathbf{q}_i) - c_2(\mathbf{q}_i) \triangleq \alpha - \gamma \triangleq \beta \end{aligned} \tag{73}$$

[8] A map $f:\ x \to y$ is called a *diffeomorphism* if f carries x homeomorphically onto y and if both f and f^{-1} are smooth.

This indicates that the 8 intersection points, $\mathbf{q}_i$, belong to the same ellipse $c_1(\mathbf{p}) = \beta$. This results in 8 intersection points between the ellipse c_1 and the circle c_2. However, it is known from geometry that there are at most 4 intersection points between any ellipse and circle. Therefore, we conclude that $(0, 0)$ cannot be an inf-max point.

By the MPT, there must exist an inf-max point between any two local minima among the 4 stationary points. Therefore, there are at most 3 local minima (by sharing one stationary point as the common inf-max point).

A.2 Proof that there are at most 2 local minima

We now prove that there are at most 2 local minima by showing that the assumption of 3 local minima (i.e., $\mathbf{p}_i$, $i = 1, 2, 3$.) leads to a contradiction. Denote the inf-max point, $\mathbf{m}_i$, corresponding to the two local minima, $\mathbf{p}_i$ and $\mathbf{p}_j$ ($i, j = 1, 2, 3$ and $i \neq j$),

$$c(\mathbf{m}_i) = \inf_{\Sigma_{ij} \in \Gamma} \max_{\mathbf{p} \in \Sigma_{ij}} c(\mathbf{p}) \tag{74}$$

Recall that we have at most 5 critical points in total (i.e., 1 non-differentiable point at the origin and 4 stationary points). So, it is clear that there are only four possible cases that we need to examine in terms of $\mathbf{m}_i$ ($i = 1, 2, 3$):

- Case I: $\mathbf{m}_1 = \mathbf{m}_2 \neq \mathbf{m}_3 = \mathbf{0}$
- Case II: $\mathbf{m}_1 \neq \mathbf{m}_2 = \mathbf{m}_3 = \mathbf{0}$
- Case III: $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}_3 = \mathbf{0}$
- Case IV: $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}_3 \neq \mathbf{0}$

Our goal is to prove that all these four cases are impossible to occur and thus there are at most 2 local minima. In what follows, we first show that the first three cases with a zero inf-max point (i.e., Cases I, II and III) cannot occur, and then disprove Case IV.

We start by considering a special case where the prior estimate coincides with the sensor position (i.e., $\hat{\mathbf{p}} = \mathbf{0}$). By denoting $\mathbf{P}_{\mathbf{pp}}^{-\mathbf{1}} = \begin{bmatrix} s_1 & s_3 \\ s_3 & s_2 \end{bmatrix}$, we expand the cost function (70) in a neighborhood of the origin as follows:

$$\begin{aligned}
c(x, y) &= \frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} s_1 & s_3 \\ s_3 & s_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2\sigma_\rho^2} \left(z - \sqrt{x^2 + y^2} \right)^2 \\
&= \left(\frac{s_1 \sigma_\rho^2 + 1}{2\sigma_\rho^2} \right) x^2 + \left(\frac{s_2 \sigma_\rho^2 + 1}{2\sigma_\rho^2} \right) y^2 + s_3 xy - \frac{z}{\sigma_\rho^2} \sqrt{x^2 + y^2} + \frac{z^2}{2\sigma_\rho^2} \\
&\triangleq Ax^2 + By^2 + Cxy - E\sqrt{x^2 + y^2} + D \qquad (75) \\
&= \sqrt{x^2 + y^2} \left(\frac{Ax^2 + By^2 + Cxy}{\sqrt{x^2 + y^2}} - E \right) + D \\
&\leq \sqrt{x^2 + y^2} \left(A|x| + B|y| + |C||x| - E \right) + D \qquad (76)
\end{aligned}$$

where $E \triangleq \frac{z}{\sigma_\rho^2} > 0$, due to the positive distance measurement z. In the above expressions, we have employed the inequalities $|x| \leq \sqrt{x^2 + y^2}$ and $|y| \leq \sqrt{x^2 + y^2}$. Clearly, there exists a neighborhood of $(0, 0)$ such that $A|x| + B|y| + |C||x| - E < 0$ and hence $c(x, y) < D = c(0, 0)$, if $(x, y) \neq (0, 0)$. By definition, $(0, 0)$ is a local maximum and thus cannot be an inf-max point.[9] This is the contradiction, and therefore Cases I, II, and III cannot happen when $\hat{\mathbf{p}} = \mathbf{0}$.

Now consider the general case where $\hat{\mathbf{p}} \neq \mathbf{0}$. First, in Case I ($\mathbf{m}_1 = \mathbf{m}_2 \neq \mathbf{m}_3 = \mathbf{0}$), as in the previous proof of at most 3 local minima (see Section A.1), we can show that there are at least 4 intersection points between the circle $c_2(\mathbf{p})$ for a given z and the level set $\mathcal{S}$ (see Fig. 11). If the circle c_2 collapses to a single point (i.e., with zero radius), clearly there is only 1 intersection point between the circle c_2 and the level set $\mathcal{S}$, and hence 1 intersection point between the

[9] Note that if $z = 0$, then (75) becomes quadratic and has a unique global minimum at the origin (by noting that $A > 0$), which clearly contradicts the assumption of three local minima.

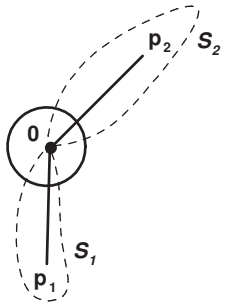


Figure 11: Illustration of the circle $c_2(\mathbf{p}) = \alpha$ intersecting the level set $\mathcal{S}$ as well as the ellipse c_1 by a small perturbation on c_2. In this plot, the dashed curves represent the level set, the solid lines are the path which connects the local minima $\mathbf{p}_1$ and $\mathbf{p}_2$ and attains its maximum at the origin, and the solid circle represents c_2 after perturbation.

circle c_2 and the ellipse c_1.[10] Importantly, by continuity of c_2 and compactness of $\mathcal{S}$, if perturbing c_2 by an arbitrarily small number in the neighborhood of origin, there are always at least 4 intersection points between the circle c_2 and the level set $\mathcal{S}$, and thus at least 4 intersection points between the circle c_2 and the ellipse c_1 [see (73)]. This perturbation results in the dynamics of the number of intersection points between the circle and the ellipse, changing from 1 to 4. However, this is not the case, since we know from geometry that by continuity of the circle and ellipse, if applying a small perturbation on the circle, 1 intersection point between the circle and ellipse can only dynamically change 0, 1 or 2 (instead of 4) intersection points. Moreover, as compared to Case I, in Case II there will be at least 6 (instead of 4) intersection points between the circle $c_2(\mathbf{p})$ for a given z and the level set $\mathcal{S}$, and thus we can show in a similar way that this case is also impossible.

In Case III ($\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}_3 = \mathbf{0}$), interestingly, no matter whether $\hat{\mathbf{p}} = \mathbf{0}$ or $\hat{\mathbf{p}} \neq \mathbf{0}$, proceeding similarly as in Cases I and II and as in Section A.1, we can derive the contradiction that there will be 6 intersection points between the circle c_2 and the level set $\mathcal{S}$ (and thus the ellipse c_1), and hence show that this case is also impossible to occur.

At this point, we have ruled out Cases I-III. We will now disprove Case IV. Specifically, to simplify notations, we denote the common inf-max point by $\mathbf{m} \triangleq \mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}_3$. Then, we first consider a special case where the prior estimate coincides with the inf-max point, i.e., $\hat{\mathbf{p}} = \mathbf{m}$. In this case, we know that $\nabla c(\mathbf{p}) \mid_{\mathbf{p}=\mathbf{m}} = \mathbf{0}$, since $\mathbf{p} = \mathbf{m}$ is a critical (inf-max) point. Therefore, we have:

$$\begin{aligned} \mathbf{0} &= \nabla c(\mathbf{p}) \mid_{\mathbf{p}=\mathbf{m}} = \nabla c_1(\mathbf{p}) \mid_{\mathbf{p}=\mathbf{m}} + \nabla c_2(\mathbf{p}) \mid_{\mathbf{p}=\mathbf{m}} \\ &= \frac{1}{\sigma_\rho^2}(z - ||\mathbf{m}||)\frac{\mathbf{m}}{||\mathbf{m}||} \;\Rightarrow\; z = ||\mathbf{m}|| \;\Rightarrow\; c(\mathbf{m}) = 0 \end{aligned} \tag{77}$$

which clearly shows that $\mathbf{p} = \mathbf{m}$ is a global minimum (by noting the quadratic and nonnegative cost function (70)). This contradicts the assumption that $\mathbf{p} = \mathbf{m}$ is an inf-max point.

We finally consider the general scenario of Case IV where $\hat{\mathbf{p}} \neq \mathbf{m}$. We define the following level set:

$$\mathcal{S}' = \{\mathbf{p} \in \mathbb{R}^2 : c(\mathbf{p}) = c(\mathbf{m}) \triangleq \alpha'\} \tag{78}$$

Similar to Case III and Section A.1 (where instead the level set $\mathcal{S}$ is considered), the corresponding interior, $\mathbf{\Theta}' = \{\mathbf{p} \in \mathbb{R}^2 : c(\mathbf{p}) < \alpha'\}$, has three curve components, each of which contains exactly one $\mathbf{p}_i$ (see Fig. 12). Consider the scenario where $\mathbf{p} = \mathbf{m}$ is an intersection point between the circle and the ellipse (by noting $c(\mathbf{m}) = c_1(\mathbf{m}) + c_2(\mathbf{m})$). In a neighborhood of $\mathbf{p} = \mathbf{m}$, if perturbing the circle $c_2(\mathbf{p}) = c_2(\mathbf{m}) \triangleq \gamma'$ by an arbitrarily small (positive or negative) value, the circle will either shrink or expand. Since there are 6 branches belonging to the level set $\mathcal{S}'$, by continuity and compactness of c_2 and $\mathcal{S}'$, in the *neighborhood* of $\mathbf{p} = \mathbf{m}$, there will exist at least 3 intersection points between the circle and the level set, and thus at least 3 intersection points between the circle and the ellipse [see (73)], either when the circle shrinks or expands. Thus, there exists an (arbitrarily) small perturbation on c_2 so that the number of intersection points between the circle and the ellipse, in the neighborhood of $\mathbf{p} = \mathbf{m}$, dynamically changes from 1 to (more than) 3. However, we know from geometry that this is not the case since 1 intersection point between a circle and an ellipse, by perturbation, can change to at most 2 in its *neighborhood*.

[10] Although depending on the measurement z, another circle possibly exists and thus may result in more intersection points between the circle c_2 and the ellipse c_1, we here consider the dynamics of the intersection point in a *neighborhood* of the origin.

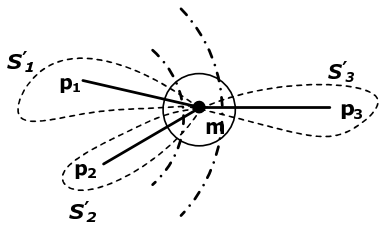


Figure 12: Illustration of the circle $c_2(\mathbf{p}) = \alpha'$ intersecting the level set $\mathcal{S}'$ by a small perturbation on c_2. In this plot, the dashed curves represent the level set, the solid lines are the path which connects the local minima $\mathbf{p}_i$ and $\mathbf{p}_j$ and attains its maximum at $\mathbf{p} = \mathbf{m}$, the dash-dotted arcs (of the circles) represent c_2 before and after perturbation, and the solid circle denotes the neighborhood of $\mathbf{p} = \mathbf{m}$.

Thus far, we have proven that all the four cases (i.e., Cases I, II, III, and IV) are impossible to occur if there are 3 local minima. As a result, there are at most 2 local minima. This completes the proof.

References

[1] J. Polastre, "Design and implementation of wireless sensor networks for habitat monitoring," Master's thesis, University of California, Berkeley, Berkeley, CA, May 2003.

[2] L. E. Parker, B. Birch, and C. Reardon, "Indoor target intercept using an acoustic sensor network and dual wavefront path planning," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV, Oct. 27-31 2003, pp. 278 – 283.

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

[4] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[5] J. M. C. Clark, R. B. Vinter, and M. M. Yaqoob, "Shifted Rayleigh filter: a new algorithm for bearings-only tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1373–1384, Oct. 2007.

[6] T. Kronhamn, "Bearings-only target motion analysis based on a multihypothesis Kalman filter and adaptive ownship motion control," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 145, no. 4, pp. 247 –252, Aug. 1998.

[7] K. Zhou and S. I. Roumeliotis, "Optimal motion strategies for range-only constrained multisensor target tracking," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1168–1185, Oct. 2008.

[8] ——, "Multirobot active target tracking with combinations of relative observations," *IEEE Transactions on Robotics*, vol. 27, no. 4, pp. 678 –695, Aug. 2011.

[9] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

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

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[12] D. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*. Springer, 2005.

[13] G. P. Huang, K. X. Zhou, N. Trawny, and S. I. Roumeliotis, "A bank of MAP estimators for single-sensor range-only target tracking," in *Proc. of the American Control Conference*, Baltimore, MD, Jun. 30 – Jul. 2, 2010, pp. 6974–6980.

[14] ——, "Bearing-only tracking using a bank of MAP estimators," in *Proc. of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 4998–5005.

[15] N. Trawny and S. I. Roumeliotis, "On the global optimum of planar, range-based robot-to-robot relative pose estimation," in *Proc. IEEE International Conference on Robotics and Automation*, Anchorage, AK, May3–8, 2010, pp. 3200–3206.

[16] N. Trawny, X. S. Zhou, K. Zhou, and S. I. Roumeliotis, "Interrobot transformations in 3-D," *IEEE Transactions on Robotics*, vol. 26, no. 2, pp. 226 –243, Apr. 2010.

[17] I. Shames, P. Bibalan, B. Fidan, and B. D. O. Anderson, "Polynomial methods in noisy network localization," in *Proc. of the 17th Mediterranean Conference on Control and Automation*, Thessaloniki, Greece, Jun. 24–26, 2009, pp. 1307–1312.

[18] M. Byröd, K. Josephson, and K. Aström, "Fast and stable polynomial equation solving and its application to computer vision," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 237–256, 2009.

[19] A. Farina, "Target tracking with bearing-only measurements," *Signal Processing*, vol. 78, pp. 61–78, Oct. 1999.

[20] V. Aidala and S. Hammel, "Utilization of modified polar coordinates for bearings-only tracking," *IEEE Transactions on Automatic Control*, vol. 28, no. 3, pp. 283–294, Mar. 1983.

[21] R. Zhan and J. Wan, "Iterated unscented Kalman filter for passive target tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 1155–1163, Jul. 2007.

[22] D. Mušicki, "Bearings only single-sensor target tracking using Gaussian mixtures," *Automatica*, vol. 45, no. 9, pp. 2088–2092, Sep. 2009.

[23] V. Cevher, R. Velmurugan, and J. H. McClellan, "A range-only multiple target particle filter tracker," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 14–19, 2006.

[24] F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, Feb. 2002.

[25] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F: Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[26] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[27] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 53 –82, Jul. 2010.

[28] R. Douc and O. Cappe, "Comparison of resampling schemes for particle filtering," in *Proc. of the International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, Sep. 15–17, 2005, pp. 64–69.

[29] T. L. Song, "Observability of target tracking with bearings-only measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1468–1472, Oct. 1996.

[30] ——, "Observability of target tracking with range-only measurements," *IEEE Journal of Oceanic Engineering*, vol. 24, no. 3, pp. 383–387, Jul. 1999.

[31] B. Triggs, P. McLauchlan, R. Hartley, and Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice*. Springer Verlag, 2000, pp. 298–375.

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

[33] I. Z. Emiris and A. Rege, "Monomial bases and polynomial system solving," in *Proc. of the International Symposium on Symbolic and Algebraic Computation*, Oxford, UK, Jul. 20–22, 1994, pp. 114–122.

[34] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[35] Y. Jabri, *The Mountain Pass Theorem: Variants, Generalizations and Some Applications*. Cambridge University Press, 2003.

[36] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall, 2001.

[37] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[38] B. M. Bell and F. W. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[39] A. Edelman and H. Murakami, "Polynomial roots from companion matrix eigenvalues," *Mathematics of Computation*, vol. 64, no. 210, pp. 763–776, Apr. 1995.

[40] A. Bishop, B. Fidan, B. D. O. Anderson, K. Dogancay, and P. Pathirana, "Optimal range-difference-based localization considering geometrical constraints," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 3, pp. 289–301, Jul. 2008.

[41] A. Bishop, B. D. O. Anderson, B. Fidan, P. Pathirana, and G. Mao, "Bearing-only localization using geometrically constrained optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 308–320, Jan. 2009.

[42] M. Cao, B. D. Anderson, and A. S. Morse, "Sensor network localization with imprecise distances," *Systems and Control Letters*, vol. 55, no. 11, pp. 887–893, 2006.

[43] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2000.

[44] G. P. Huang and S. I. Roumeliotis, "Target tracking using a bank of MAP estimators," University of Minnesota, Tech. Rep., Aug. 2010. [Online]. Available: http://people.csail.mit.edu/ghuang/paper/tr/TR_BMAPTT.pdf

[45] G. Huang, M. Kaess, J. Leonard, and S. Roumeliotis, "Analytically-selected multi-hypothesis incremental MAP estimation," in *Proc. of the International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26–31, 2013, pp. 6481–6485.

[46] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[47] A. Doucet and A. M. Johansen, *A tutorial on particle filtering and smoothing: fifteen years later*. University Press, 2009.

[48] M. Struwe, *Variational Methods: Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*. Springer, 2008.

[49] J. W. Milnor, *Topology from the Differentiable Viewpoint*. Princeton University Press, 1997.

[50] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill, 1986.