MARGINALIZED CONDITIONAL POSTERIOR CRAMÉR-RAO LOWER BOUND WITH WEIGHTING

Ping ZHANG

School of Information Science and Engineering, Southeast University, Nanjing, 210096, China.

ABSTRACT

Conditional posterior Cramér-Rao lower bound (CPCRLB) gives a lower bound for the mean squared error (MSE) of a sequential Bayesian estimator that is conditional on a historical observation sequence. Yet it can be further improved in tightness, together with a reduction in computational burden. In this paper, we introduce a marginalization technique and a weighting operation to improve the CPCRLB, where marginalizing out the historical states greatly simplifies the integral without any loss of the efficiency on bounding the one-step-ahead MSE, and the weighting operation lifts the CPCRLB after finding out a suitable weighting function. An application on univariate non-stationary growth model (UNGM) presents a closed-form solution of the marginalized CPCRLB, and a high-performance weighted CPCRLB needing only an extra one-dimensional numerical maximization.

Index Terms— Bayesian estimator, Fisher information matrix (FIM), Monte Carlo integral, non-linear Gaussian, particle filter

1. INTRODUCTION

Conventional Cramér-Rao lower bound (CRLB) lower bounds the variance of any unbiased estimate of a fixed parameter, whereas the posterior CRLB (PCRLB) bounds the mean squared error (MSE) of the estimates of stochastic parameters [1]. For discrete-time filtering problems, the (unconditional) PCRLB low bounds the best achievable performance of a nonlinear dynamic system, through an elegant recursive scheme [2]. But if taking the historical observation data into consideration, one should resort to the conditional PCRLB (CPCRLB) [3].

The CPCRLB meets the requirement of many practical applications, and there exists a recursive scheme (with an inevitable approximation on expectation) to cope with high-dimensional integrals [3]. But is CPCRLB tight enough? Focusing on the tightness of the PCRLB, [4] has presented a brilliant weighting technique to control the tightness of the bound. However, due to the constructive characteristic of this approach, it is difficult to apply this technique to the high-dimensional and complicated non-linear state-space models directly.

Marginalizing out the historical states is suggested in this paper, based on which, a practically operable weighting function is introduced to improve the tightness of the CPCRLB. Notably, the marginalization operation leads to no performance degradation for bounding the one-step-ahead MSE, and is practically available, in an approximated form, from a variety of the existing Bayesian filtering algorithms. For this reason, it is possible to improve the CPCRLB by performing weighting operation on the marginalized model, only needing a deliberate selection of a weighting function that reflects a tradeoff between the tightness and the computational feasibility. An application on univariate non-stationary growth model (UNGM) demonstrates the practicability and efficiency of our marginalized and weighted CPCRLBs, in which all integrals are calculated analytically.

2. NOTATION

Distinguishing stochastic variables/vectors from their realizations (samples), using uppercase letters as [5], is helpful (and sometimes indispensable) to clarify the expectation operation related to the CPCRLB. This section also adds an extra explanation on the distribution used in computing the expectation, but it is omitted in the following sections without loss of rigor.

Throughout $X_{1:k} = [X_1^T, X_2^T, \dots, X_k^T]^T$ denotes a stochastic state sequence with the expectation

$$\mathrm{E}[X_{1:k}^T X_{1:k}] < \infty \tag{1}$$

where the expectation operation is taken with respect to (w.r.t.) the probability density function (pdf) $p(\mathbf{x}_{1:k})$, $\mathbf{x}_{1:k} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_k^T]^T$ denotes a realization of $X_{1:k}$, and $\mathbf{x}_i$, $i = 1, 2, \dots, k$, are n-dimensional vectors.

Let $\mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_k^T]^T$ be a realization of a stochastic observation sequence $Z_{1:k} = [Z_1^T, Z_2^T, \dots, Z_k^T]^T$ and assume the derivability of the conditional pdf $p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})$ w.r.t.

This work is supported in parts by the Major State Basic Research Development Program of China under Grant No. 2011CB302905, the National High Technology Research and Development Program of China under Grant No. 2009AA11Z219, Research Project of Education Administration of P. R. China under Grant No.706028, the Distinguish Research Fellow Program under Grant No. 07-E-016, and Innovation Fund of Southeast University, China.

$\mathbf{x}_{1:k}$, we have the following scoring function

$$\mathbf{s}(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}) = \frac{\partial p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})}{p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})\partial \mathbf{x}_{1:k}} = \frac{\partial \log p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})}{\partial \mathbf{x}_{1:k}}. \tag{2}$$

The Fisher information matrix (FIM) of the stochastic vector $X_{1:k}$ conditional on the observation $\mathbf{z}_{1:k-1}$ is defined based on the scoring function $\mathbf{s}(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})$ as

$$\mathbf{J}_{1:k} = \mathrm{E}\left[\mathbf{s}(X_{1:k}|\mathbf{z}_{1:k-1}, Z_k)\mathbf{s}(X_{1:k}|\mathbf{z}_{1:k-1}, Z_k)^T\right] \tag{3}$$

where the expectation is taken w.r.t. $p(\mathbf{x}_{1:k}, \mathbf{z}_k|\mathbf{z}_{1:k-1})$, and (3) also implies the existence of the expectation.

The inverse $\mathbf{J}_{1:k}^{-1}$ is called CPCRLB [3], which lower bounds the (conditional) MSE

$$\begin{aligned}\mathbf{\Sigma}_{1:k} = \mathrm{E}[&(X_{1:k} - \mathrm{E}[X_{1:k}|\mathbf{z}_{1:k-1}, Z_k])\\ &(X_{1:k} - \mathrm{E}[X_{1:k}|\mathbf{z}_{1:k-1}, Z_k])^T]\end{aligned} \tag{4}$$

as

$$\mathbf{\Sigma}_{1:k} \geq \mathbf{J}_{1:k}^{-1} \tag{5}$$

where $\mathbf{\Sigma}_{1:k} \geq \mathbf{J}_{1:k}^{-1}$ means $\mathbf{\Sigma}_{1:k} - \mathbf{J}_{1:k}^{-1}$ is positive semidefinite. Here, $\mathbf{\Sigma}_{1:k}$ is actually the optimal MSE achieved by the Bayesian estimator, and the inner and outer expectation in (4) are taken w.r.t. $p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})$ and $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$, respectively.

3. MARGINALIZATION

The computational burden (*i.e.*, high-dimensional integral) in (3) can be greatly reduced if we focus on deriving a marginalized CPCRLB for the one-step-ahead state X_k (assuming that $\mathbf{z}_k$ has not been observed yet). Under this strategy, we can derive the CPCRLBs successively, which leads to an online estimation of the one-step-ahead MSE.

Note that the one-step-ahead MSE is the lower-right n-by-n submatrix of $\mathbf{\Sigma}_{1:k}$, which can also be represented as

$$\begin{aligned}\mathbf{\Sigma}_k = \mathrm{E}[\,&(X_k - \mathrm{E}[X_k|\mathbf{z}_{1:k-1}, Z_k])\\ &(X_k - \mathrm{E}[X_k|\mathbf{z}_{1:k-1}, Z_k])^T].\end{aligned} \tag{6}$$

The standard PCRLB theory introduces an (marginalized) FIM

$$\mathbf{J}_k = \mathrm{E}\left[\mathbf{s}(X_k|\mathbf{z}_{1:k-1}, Z_k)\mathbf{s}(X_k|\mathbf{z}_{1:k-1}, Z_k)^T\right] \tag{7}$$

to lower bound $\mathbf{\Sigma}_k$ as

$$\mathbf{\Sigma}_k \geq \mathbf{J}_k^{-1}. \tag{8}$$

Then, questions are: a) what is the relationship between $\mathbf{J}_k^{-1}$ and the n-by-n lower-right submatrix of $\mathbf{J}_{1:k}^{-1}$; b) how to calculate the expectation (7).

The first question is answered by Theorem 1, which declares that $\mathbf{J}_k^{-1}$ is no lower than the n-by-n lower-right submatrix of $\mathbf{J}_{1:k}^{-1}$. Therefore, (7) reduces the nk-dimensional integral in (3) into an n-dimensional one without any loss of efficiency on bounding the one-step-ahead MSE $\mathbf{\Sigma}_k$.

Theorem 1 *Under the assumptions of (1) and (3), we have*

$$\mathbf{J}_k^{-1} \geq \left(\mathbf{J}_{1:k}^{-1}\right)_{k,k} \tag{9}$$

where $\left(\mathbf{J}_{1:k}^{-1}\right)_{k,k}$ *denotes the* n*-by-*n *lower-right submatrix of* $\mathbf{J}_{1:k}^{-1}$.

The proof of Theorem 1 is referred to Proposition 1 in [4].

The second question is somewhat complicated. In theory, the expectation in (7) needs the marginal pdf $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k}) = p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k)$, which is not easy to get (otherwise one may compute $\mathbf{\Sigma}_k$ directly). But in practice, $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is referred to a marginalized prediction pdf, which is approximated in most existing Bayesian filtering algorithms, including a Gaussian density approximation in (extended or uncensored) Kalman filtering and Gaussian particle filtering, a Gaussian mixture approximation in Gaussian sum particle filtering, and (weighted) particles approximation in particle filtering. In summary, $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k})$ is practically available in most filtering processes, which can be used to calculate (7) straightforwardly, even admitting analytical solutions.

4. WEIGHTING

Weighting is a critical technique that controls the tightness of the PCRLB, which can be used to improve the efficiency of the PCRLB. This section presents how to improve our marginalized CPCRLB through weighting.

For clarity, we summarize the operations and properties for our weighted CPCRLB in Theorem 2. The proof is omitted since it is paralleled to Proposition 2 and 3 in [4]. But our weighting function $\mathbf{q}(\cdot)$ differs from the one in [4] in bounding the one-step-ahead MSE rather than the current MSE.

Theorem 2 *Let* $\mathbf{q}(\mathbf{x}_k, \mathbf{z}_k) = [q_1(\mathbf{x}_k, \mathbf{z}_k), \ldots, q_n(\mathbf{x}_k, \mathbf{z}_k)]^T$ *be a vector-valued function where* n *is the dimension of* $\mathbf{x}_k$*. For each* $i = 1, 2, \ldots, n$*:*

1. $q_i(\mathbf{x}_k, \mathbf{z}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k})$ *is continuously differentiable.*
2. $\mathrm{E}\left[\|q_i(\mathbf{X}_k, \mathbf{Z}_k)\|\,|\mathbf{z}_{1:k-1}\right] < \infty$.
3. $\mathrm{E}\left[q_i(\mathbf{X}_k, \mathbf{Z}_k)|\mathbf{z}_{1:k-1}\right] \neq 0$.

Then, define $\mathbf{s}_{\mathbf{q}}(\mathbf{x}_k|\mathbf{z}_{1:k}) = [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n]^T$ *where*

$$\tilde{s}_i = \frac{\partial q_i(\mathbf{x}_k, \mathbf{z}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k})}{p(\mathbf{x}_k|\mathbf{z}_{1:k})\partial x_{k,i}}, \quad i = 1, 2, \ldots, n \tag{10}$$

have finite second moments. This induces

$$\mathbf{J}_{\mathbf{q},k} = \mathrm{E}\left[\mathbf{s}_{\mathbf{q}}(\mathbf{X}_k|\mathbf{z}_{1:k-1}, Z_k)\mathbf{s}_{\mathbf{q}}(\mathbf{X}_k|\mathbf{z}_{1:k-1}, Z_k)^T\right]. \tag{11}$$

At last, we have

$$\mathbf{\Sigma}_k \geq \mathbf{Q}\mathbf{J}_{\mathbf{q},k}^{-1}\mathbf{Q} \tag{12}$$

where $\mathbf{Q} = \mathrm{diag}\left(\mathrm{E}\left[\mathbf{q}(X_k, Z_k)|\mathbf{z}_{1:k-1}\right]\right)$ *is a diagonal matrix with the diagonal entries* $\mathrm{E}\left[\mathbf{q}_i(X_k, Z_k)|\mathbf{z}_{1:k-1}\right]$*. And the equality holds if and only if for each* i*,* $q_i(\mathbf{x}_k, \mathbf{z}_k)$ *has the form*

$$q_i(\mathbf{x}_k, \mathbf{z}_k) = \frac{c_i}{p(\mathbf{x}_k|\mathbf{z}_{1:k})} \int_{\mathbf{x}_{k,i}}^{\infty} \gamma_i p(\mathbf{t}|\mathbf{z}_{1:k}) dt_i \Bigg|_{\mathbf{t}_{i^*} = \mathbf{x}_{k,i^*}} \tag{13}$$

where

$$\boldsymbol{\gamma} = \mathbf{\Sigma}_k^{-1}\left(\mathbf{x}_k - \mathrm{E}\left[X_k|Z_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

c_i *is an arbitrary non-zero constant, and* i^* *indicates the multiindex by eliminating* i *from* $[1:n]$.

$$\boldsymbol{\Sigma}_k \geq \frac{\mathrm{E}^2\left[q(X_k)|\mathbf{z}_{1:k-1}\right]}{\mathrm{E}\left[\left(q'(X_k)+q(X_k)\frac{p'(X_k|z_{1:k-1})}{p(X_k|z_{1:k-1})}\right)^2+\frac{1}{\rho^2}q^2(X_k)\left(g'(X_k)\right)^2\middle|\mathbf{z}_{1:k-1}\right]} \tag{14}$$

Gaussian additive noise assumption on the observation $\mathbf{z}_k$ may greatly simplify the expression (12). A useful example with scalar states and observations was provided in [4], where the observation z_k is assumed to follow $z_k = g(x_k) + \epsilon$ with Gaussian noise $\epsilon \sim \mathcal{N}(0, \rho^2)$. In this case, after restricting $q(x_k, z_k) = q(x_k)$ independent of z_k, (12) can be simplified into (14). This is utilized to solve the following example.

Example This example is closely related to the UNGM model investigated in Section 5. In our example, the observation function $z_k = x_k^2 + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \rho^2)$, and the prediction pdf is modeled as $p(x_k|\mathbf{z}_{1:k-1}) = \mathcal{N}(x_k; \mu, \sigma^2)$.

For easy calculation, we restrict $q(x_k) = \exp(\lambda x_k^2)$ (other forms are also admissible), where λ is an adjustable parameter. After performing analytical integrals (the process is omitted here), we get a family of weighted CPCRLBs as

$$\begin{aligned}&\boldsymbol{\Sigma}_k \geq c(\lambda)\\&= \frac{\frac{\sqrt{1-4\lambda\sigma^2}}{1-2\lambda\sigma^2}\exp\left(-\frac{\mu^2}{2\sigma^2}\left(1+\frac{1}{1-4\lambda\sigma^2}-\frac{2}{1-2\lambda\sigma^2}\right)\right)}{\frac{(2\lambda\sigma^2-1)^2/\sigma^2+4\sigma^2/\rho^2}{1-4\lambda\sigma^2}+\frac{4\mu^2(\lambda^2+1/\rho^2)}{(1-4\lambda\sigma^2)^2}}.\end{aligned} \tag{15}$$

After maximizing the univariate function $c(\lambda)$ numerically, the optimal weighted CPCRLB can be obtained, as plotted in Fig. 1. In the rest of this paper, the weighted CPCRLB always indicates the optimal one.

Fig. 1 plots the MSE, weighted CPCRLB, and (marginalized) CPCRLB as the functions of μ, where $\sigma = 1$, and ρ is set as 0.1, 1, and 10, respectively. Obviously, the CPCRLB diverges from the MSE as μ is close to zero and/or ρ is relatively small. This suggests the bimodal characteristic of the observation equation as an origin of the divergence. By using the weighting operation, the divergence has been reduced a great deal, which is more significant when μ is not too close to zero. Hence, it is attracting to use the weighting technique to improve the CPCRLB, by an elaborate construction of $\mathbf{q}(\cdot)$.

5. SIMULATIONS

Simulations are performed on the univariate non-stationary growth model (UNGM), which is given by

$$x_k = \alpha x_{k-1} + \beta\frac{x_{k-1}}{1+x_{k-1}^2} + \gamma\cos(1.2(k-1)) + u_k \tag{16}$$

$$z_k = \kappa x_k^2 + v_k, \qquad k = 1, 2, \ldots, 20 \tag{17}$$

where $x_0 \sim \mathcal{N}(0,1)$, $u_k \sim \mathcal{N}(0,1)$, $v_k \sim \mathcal{N}(0,\rho^2)$, and $\gamma = 8$, $\kappa = 1/20$. The other parameters are set respectively in three cases (with varying non-linearities and bimodalities), those are: a) $\alpha = 1$, $\beta = 5$, and $\rho = 1$; b) $\alpha = 1$, $\beta = 0.1$, and $\rho = 0.1$; c) $\alpha = 0.5$, $\beta = 25$, and $\rho = 0.1$.

The particle filtering scheme is used to produce a Gaussian density approximation of the marginalized prediction pdf as $p(x_k|\mathbf{z}_{1:k-1}) = \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$, which leads to our marginalized CPCRLB $4\kappa^2\rho^{-2}(\hat{\mu}^2 + \hat{\sigma}^2) + \hat{\sigma}^{-2}$, and the weighed CPCRLB derived by maximizing $c(\lambda)$ that is obtained by replacing μ, σ, and ρ in (15) with $\hat{\mu}$, $\hat{\sigma}$, and ρ/κ, respectively. The original CPCRLB is calculated using the iterative scheme (involving Monte-Carlo integrals) in [3].

Fig. 3 exhibits the improvement of our weighted CPCRLB and the performance of our marginalized CPCRLB through a comparison with the original CRLB and MSE. Three cases, given in Fig. 2, are considered, where the last two indicate a high-bimodality and a both high-bimodality and high-nonlinearity cases, respectively. Obviously, divergence happens on the CPCRLB under the last two cases, and our weighted CPCRLB substantially lifts the bound, extremely well in the high-bimodality case (b).

6. CONCLUSION

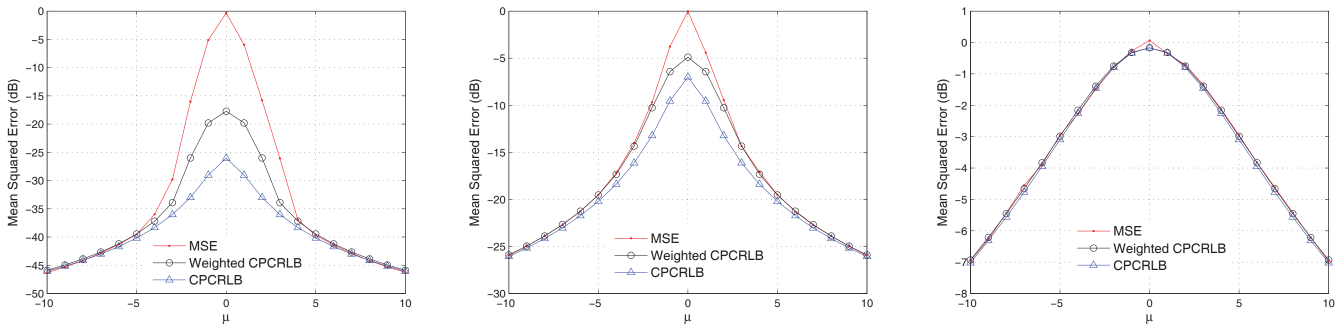
We have improved the CPCRLB in tightness and computational complexity, through adopting a weighting operation after marginalizing out the historical states. This approach makes full use of the outputs of the existing Bayesian filtering algorithms, and thus provides a low-complexity but also high-performance lower bound for the one-step-ahead MSE. Note that the derivation of the weighted CPCRLB is independent of the form of the state equation. Thus the proposed approach can be applied to a variety of non-linear filtering problems with complicated state equations, needing only to construct a weighting function suitable to the observation equation.

7. REFERENCES

[1] H. L. Van Trees, *Detection, estimation, and modulation theory, radar-sonar signal processing and gaussian signals in noise*, Wiley-Interscience, 2003.

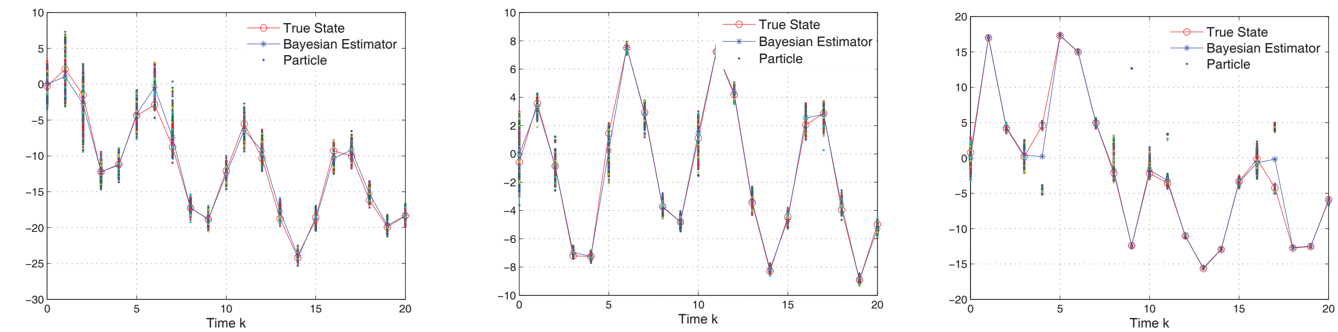
[2] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[3] L. Zuo, R. X. Niu, and P. K. Varshney, "Conditional posterior Cramér-Rao lower bounds for nonlinear sequential bayesian estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 1–14, Jan. 2011.



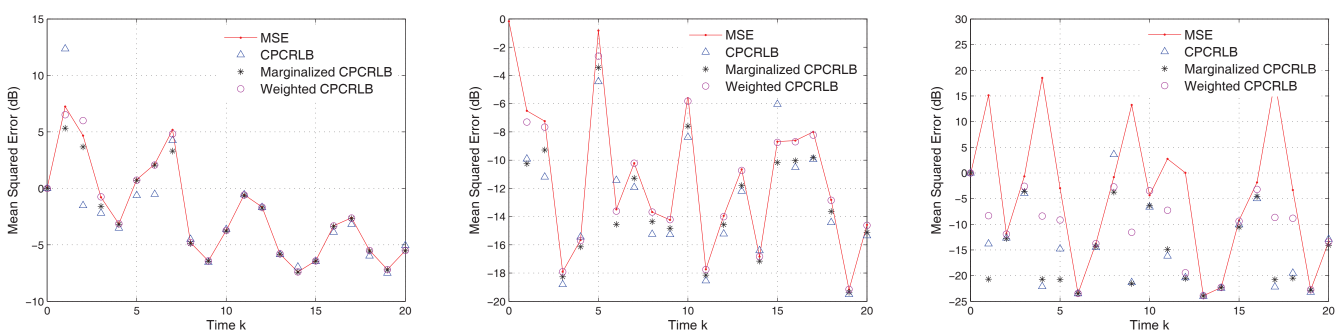
(a) $\rho = 10^{-1}$ (b) $\rho = 1$ (c) $\rho = 10$

Fig. 1: Comparisons of (marginalized) CPCRLB, weighted CPCRLB, and MSE under different μ and ρ.



(a) $\alpha = 1, \beta = 5, \rho = 1$ (b) $\alpha = 1, \beta = 0.1, \rho = 0.1$ (c) $\alpha = 0.5, \beta = 25, \rho = 0.1$

Fig. 2: Tracking a moving target under univariate non-stationary growth model (UNGM) with particle filter, where the parameters are tuned to realize different non-linearities and bimodalities.



(a) $\alpha = 1, \beta = 5, \rho = 1$ (b) $\alpha = 1, \beta = 0.1, \rho = 0.1$ (c) $\alpha = 0.5, \beta = 25, \rho = 0.1$

Fig. 3: Comparisons of MSE, CPCRLB, marginalized CPCRLB and weighted CPCRLB for the three trajectories in Fig. 2, respectively.

[4] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér-Rao bounds," *The Annals of Statistics*, vol. 15, no. 4, pp. 1421–1438, Dec. 1987.

[5] E. L. Lehmann and G. Casella, *Theory of point estimation*, Springer Verlag, 1998.