Sequential Monte Carlo scheme for Bayesian estimation in the presence of data outliers

Liang Huang[1] and Ying-Cheng Lai[1,2]
[1]*Department of Electrical Engineering, Arizona State University, Tempe, Arizona 85287, USA*
[2]*Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287, USA*
(Received 25 September 2006; revised manuscript received 16 February 2007; published 21 May 2007)

Bayesian inference has been used widely in physics, biology, and engineering for a variety of experiment- or observation-based estimation problems. Sequential Monte Carlo simulations are effective for realizing Bayesian estimations when the system and observational processes are nonlinear. In realistic applications, large disturbances in the observation, or outliers, may be present. We develop a theory and practical strategy to suppress the effect of outliers in the experimental observation and provide numerical support.

DOI: 10.1103/PhysRevE.75.056705 PACS number(s): 02.70.Uu, 05.40.Ca, 05.45.Tp

I. INTRODUCTION

The problem of estimating the time-varying state of a system based on experimental measurements or observations finds many applications in physics, biology, and engineering. Examples include quantum-state reconstruction and purity estimation [1,2], noise reduction and state reconstruction in chaotic dynamical systems [3,4], estimation of bounds on isocurvature perturbations in the Universe and on cosmic strings from cosmic microwave background and large-scale structure data [5,6], gravitational-wave signal analysis [7,8], macromolecular structure determination [9,10], prediction of protein-protein interactions from genomic data [11], tracking and positioning problems [12], etc. In general, a system model that evolves the state vector (x) in time is needed, so is an observational model that relates an observation vector (y) to the state vector. In any realistic application both noise and model uncertainties exist, rendering necessary a probabilistic treatment of the estimation problem. That is, one can evolve x according to the system model and make corrections to or update x based on the available y. The quantity of interest is the posterior probability density function (PDF) $p(x|y)$, given all available observations y. The standard approach to addressing this problem is Bayesian inference [13], which leads to the classical Kalman filter when both the system and the observational models are linear. For nonlinear problems, a viable approach is *sequential Monte Carlo simulation* (or particle filter) [14–16], which uses a set of random samples to approximate the posterior PDF $p(x|y)$. The approximated PDF evolves and is corrected by the observation based on the Bayes rule. If the number of samples is sufficiently large, the approximation approaches the optimal Bayesian estimate. Due to the constant improvement of modern computing technology, the sequential Monte Carlo approach has begun to find significant applications in science and engineering [15–17].

A fundamental question in sequential Monte Carlo simulations is how the precision of the estimated state vector depends on noise in the system and in the observation. Another issue of significant practical interest is how to deal with occasional but large disturbances, or *outliers*, in the observation. The aim of this paper is to address these two related problems. In particular, we shall derive and verify a self-consistent equation that relates the covariance matrix of the samples, which determines the precision of the state estimate, to the covariance matrices of the system noise and of the observational noise. We then propose a robust sequential Monte Carlo scheme to overcome the effect of outliers. In this regard, a previous approach includes using a heavy-tailed error distribution to improve the state-space models so that they react quite flexibly to changes in points or edges, but still provide smooth fits in other regions [18]. Leave-k-out diagnostics is used to detect a series of consecutive outliers for a linear state-space model [19]. It uses all residual observations in the time span to check whether a series of consecutive observations are jointly outlying, and thus it is actually "off line." Our idea is to detect the outliers from the previous knowledge about the system and then to eliminate them in the sequential Monte Carlo implementation. Simulations using a precise positioning problem demonstrate the power of our scheme. We expect our results to have significant impact on problems where the underlying system and/or experimental observations are subject to outliers. For example, the observation of a star or a galaxy may be corrupted by the drastic activity of another celestial body in a short period. In biological physics, macromolecular structure is inferred indirectly from various measurements—e.g., nuclear magnetic resonance spectra, x-ray reflections, or homology-derived restraints—which can easily contain outliers [9]. In global positioning system- (GPS-) based precise positioning problems, GPS signals may be disturbed by sudden and large jamming. In Sec. II, we outline the basic steps of the sequential Monte Carlo method. A self-consistent equation governing the dependence of the estimation error on noise is derived in Sec. III. Our robust sequential Monte Carlo scheme for mitigating the effect of measurement or observational outliers is presented in Sec. IV. Numerical support for the self-consistent equation and for our robust scheme is provided in Sec. V. Concluding remarks are presented in Sec. VI.

II. SEQUENTIAL MONTE CARLO METHOD

Let $y(0:t)=\{y(t'),t'=t_0(=0),t_1,t_2,\ldots,t_k(=t)\}$ be the observations from time 0 to time t, which are not necessarily equidistant in time. We seek to obtain the posterior PDF $p[x(t)|y(0:t)]$. The state and observational equations are

$$x(t)=f[x(t_{k-1}),v(t_{k-1})], \tag{1}$$

$$y(t) = g[x(t), e(t)]. \quad (2)$$

Equation (1) describes the evolution of the state, and Eq. (2) maps the state to the observational vector; f and g can be nonlinear functions. The processes $v(t)$ and $e(t)$ represent random fluctuations (e.g., noise, uncertainties, outliers, etc.) in the system and in the observation, respectively. Often, in an application the distribution of the initial state can be obtained by considering the specific physics involved. It is thus reasonable to assume that this distribution be available. The PDF $p[x(t)|y(0:t)]$ can then be obtained recursively by prediction through the dynamical equation (1) and likelihood correction through the observational equation (2). In particular, given the PDF $p[x(t_{k-1})|y(0:t_{k-1})]$ at time t_{k-1}, the prediction step uses the dynamic equation (1) to obtain the prior PDF of the state at time t via the Chapman-Kolmogorov equation

$$p[x(t)|y(0:t_{k-1})] = \int dx(t_{k-1}) p[x(t)|x(t_{k-1})] p[x(t_{k-1})|y(0:t_{k-1})], \quad (3)$$

where $p[x(t)|x(t_{k-1}), y(0:t_{k-1})] = p[x(t)|x(t_{k-1})]$ is used. At time t, a new measurement $y(t)$ becomes available, which can be used to correct the prior PDF via the Bayes rule

$$p[x(t)|y(0:t)] = \frac{p[y(t)|x(t)] p[x(t)|y(0:t_{k-1})]}{p[y(t)|y(0:t_{k-1})]}, \quad (4)$$

where

$$p[y(t)|y(0:t_{k-1})] = \int p[y(t)|x(t)] p[x(t)|y(0:t_{k-1})] dx$$

depends on the likelihood function $p[y(t)|x(t)]$. The recurrence relations (3) and (4) form the basis for the optimal Bayesian solution. For Gaussian noise, when f and g are linear functions, the recurrence relation can be solved analytically, which is the classical Kalman filter. For nonlinear functions f and g, a linearization technique is viable which leads to the so-called extended Kalman filter [20]. An unscented Kalman filter deliberately selects a set of points and propagates them through nonlinearity to estimate the Gaussian approximation [21], while for more general cases, the approach of sequential Monte Carlo simulations are desirable [14–17].

Let $\{x_i(t), w_i(t)\}_{i=1}^N$ denote a random measure that characterizes the posterior PDF $p[x(t)|y(0:t)]$, where $\{x_i(t), i = 1, \ldots, N\}$ is a set of support points with associated weights $\{w_i(t), i = 1, \ldots, N\}$. The weights are normalized that $\Sigma_i w_i(t) = 1$. The posterior PDF can be approximated as

$$p[x(t)|y(0:t)] \simeq \sum_{i=1}^{N} w_i(t) \delta(x(t) - x_i(t)),$$

where $\delta(x)$ is the Dirac delta function such that $\delta(x) = 0$ if $x \neq 0$ and $\int_{x_1}^{x_2} \delta(x) dx = 1$ if (x_1, x_2) contains 0. The average of an arbitrary function $f(x)$ can be simplified as

$$\langle f[x(t)] \rangle = \int f(x) p[x|y(0:t)] dx = \sum_i w_i(t) f[x_i(t)].$$

The weights are chosen using the principle of importance sampling [22]. In particular, given an arbitrary PDF $p(x)$, it may be difficult to draw samples. Suppose for an alternative PDF $q(x)$ samples can be drawn relatively easily. Letting $x_i \sim q(x)$ $(i = 1, \ldots, N)$ be samples drawn from some $q(\cdot)$, the *importance density*, we obtain the following weighted approximation:

$$p(x) \simeq \sum_{i=1}^{N} w_i \delta(x - x_i),$$

where $w_i \propto p(x_i)/q(x_i)$ is the normalized weight of the ith sample.

Now consider the joint probability $p[x(0:t)|y(0:t)]$. In the case of independent noise samples, we can write

$$p[x(0:t)|y(0:t)] \propto p[x(0)|y(0)] \prod_{j=1}^{k} p[y(t_j)|x(t_j)] p[x(t_j)|x(t_{j-1})].$$

Thus

$$p[x(0:t)|y(0:t)] = p[x(t), x(0:t_{k-1})|y(t), y(0:t_{k-1})] = \frac{p[y(t)|x(t)] p[x(t)|x(t_{k-1})]}{p[y(t)|y(0:t_{k-1})]} \times p[x(0:t_{k-1})|y(0:t_{k-1})].$$

Assume that the posterior distribution $p[x(0:t_{k-1})|y(0:t_{k-1})]$ is approximated by $\{x_i(0:t_{k-1}), w_i(t_{k-1})\}_{i=1}^N$; given a new observation $y(t)$, the objective is to obtain an approximation $\{x_i(0:t), w_i(t)\}_{i=1}^N$ for $p[x(0:t)|y(0:t)]$, such that the estimation of quantities of interest at time t can be calculated. The sequential Monte Carlo scheme is to generate a sample $x_i(t)$ and append it to $x_i(0:t_{k-1})$ to form $x_i(0:t)$, and update the weight $w_i(t_{k-1})$ to $w_i(t)$.

If the importance function $q[x(0:t)|y(0:t)]$ can be factorized as

$$q[x(0:t)|y(0:t)] = q[x(t)|x(0:t_{k-1}), y(0:t)] q[x(0:t_{k-1})|y(0:t_{k-1})],$$

and $x_i(t)$ is sampled from $q[x(t)|x_i(0:t_{k-1}), y(0:t)]$, the weight of the trajectory $x_i(0:t)$ is

$$w_i(t) \propto \frac{p[x_i(0:t)|y(0:t)]}{q[x_i(0:t)|y(0:t)]} = \frac{p[y(t)|x_i(t)] p[x_i(t)|x_i(t_{k-1})]}{q[x_i(t)|x_i(0:t_{k-1}), y(0:t)] p[y(t)|y(0:t_{k-1})]} \times \frac{p[x_i(0:t_{k-1})|y(0:t_{k-1})]}{q[x_i(0:t_{k-1})|y(0:t_{k-1})]} \propto \frac{p[y(t)|x_i(t)] p[x_i(t)|x_i(t_{k-1})]}{q[x_i(t)|x_i(0:t_{k-1}), y(0:t)]} w_i(t_{k-1}),$$

where $p[y(t)|y(0:t_{k-1})]$ is omitted since it is common to all samples. A convenient choice for the importance density is the following *prior importance density* [16,17]:

$q[x(t)|x_i(0:t_{k-1}),y(0:t)]=p[x(t)|x_i(t_{k-1})]$, with which the weight updating equation becomes

$$w_i(t) \propto w_i(t_{k-1})p[y(t)|x_i(t)], \tag{5}$$

and the posterior filtered density $p[x(t)|y(0:t)]$ can be approximated as

$$p[x(t)|y(0:t)] \simeq \sum_{i=1}^{N} w_i(t)\delta[x(t)-x_i(t)]. \tag{6}$$

From a numerical point of view, the above analysis can be implemented as follows. First generate N samples $\{x_i(0)|i=1,\ldots,N\}$ from the distribution of $y(0)$ as given in Eq. (2). Each sample has a weight of $1/N$. Each sample $x_i(0)$ then evolves according to the dynamical equation (1) by considering the noise v to get the value at time t_1—e.g., $x_i(t_1)$—and the weights are updated via Eq. (5). The estimation of the state at time t_1 is $\langle x(t_1)\rangle=\Sigma_{i=1}^{N} w_i(t_1)x_i(t_1)$. During the evolution, it may occur that there are disproportionally fewer samples about x_i than determined by the weight w_i. To avoid this, a sample importance resampling procedure [16] can be applied. That is, we can generate a new set of samples $\{x_i^*,w_i^*\}_{i=1}^N$ from the samples $\{x_i,w_i\}_{i=1}^N$ with probability being their weights; i.e., each time, the probability to draw sample x_i is its weight w_i (note that the weights are normalized so that $\Sigma_i w_i=1$). The weights w_i^* for the new samples are then set as $1/N$. As a result of this resampling procedure, the weight w_i of a sample is represented by the number of duplications of the sample, and thus the statistics of the samples—e.g., mean value, covariance, etc.—are unchanged in the large-N limit. The resampling step automatically concentrates the samples in the regions of interest and effectively discards samples with low weight. However, this may result in overlaps for some samples. For example, if one sample has a very large weight, after resampling, it may have many duplications, which leads to a degeneration problem. To overcome this difficulty, a regularization process can be applied: A small random vector is added to each sample as a perturbation: $x_i \leftarrow x_i^*+hD\epsilon_i$, where ϵ_i follows the standard normal distribution and D is such that $DD^T=S$ (D^T is the transpose of D). The matrix S is the empirical covariance matrix of the samples before resampling, and the quantity h is a regularization parameter [16,23]. The samples again propagate via the dynamical equation (1) to yield the values at the next time step. The process continues until a desired time span for estimation is reached.

III. NOISE DEPENDENCE OF THE ESTIMATION ERROR

When noise of the system is stationary—i.e., the covariance matrices Σ_v and Σ_e for the process noises v and e in Eqs. (1) and (2) are constant in time—the samples can evolve into a "steady" state and their covariance matrix can be obtained, which is proportional to the estimation error. Suppose that at time t_{k-1} the covariance matrix of the samples is $\Sigma_x(t_{k-1})$, which is unknown. Since the dynamical equation f is known, after propagating through Eq. (1), the covariance matrix Σ_f of the samples at time t_k can be expressed in terms of $\Sigma_x(t_{k-1})$ and Σ_v, which reads $\Sigma_f(\Sigma_x(t_{k-1}),\Sigma_v)$. To make use of the correction step [Eq. (4)], we solve x from Eq. (2): $x=g^{-1}(y,e)$. Therefore, for a given observation $y(t)$, the covariance matrix for $x(t)$, from the observational point of view, can be obtained as $\Sigma_s[y(t),\Sigma_e]$. Usually, Σ_s depends mainly on Σ_e and has little dependence on $y(t)$; thus, Σ_s is merely constant in time and can be calculated using the initial observation $y(0)$. The correction procedure is equivalent to a modulation posted by a distribution with covariance matrix Σ_s on a distribution with covariance matrix Σ_f. Suppose both distributions are Gaussian; the resulting distribution is also Gaussian, but with covariance matrix $\{\Sigma_f^{-1}+\Sigma_s^{-1}\}^{-1}$. The resampling step does not change the covariance matrix, and the regularization step simply adds a factor of $1+h^2$. Thus we have $\Sigma_x(t)=\{[\Sigma_f^{-1}+\Sigma_s^{-1}\}^{-1}(1+h^2)$. In the steady state, we have $\Sigma_x(t)\simeq\Sigma_x(t_{k-1})$, which leads to the following self-consistent equation:

$$\Sigma_f^{-1}+\Sigma_s^{-1}=(1+h^2)\Sigma_x^{-1}, \tag{7}$$

which determines the covariance matrix of the samples, or the posterior PDF $p[x(t)|y(0:t)]$, for given dynamical and observational noise levels. Note that Σ_f is a function of Σ_x and Σ_v. For certain cases, Σ_f can be expressed explicitly in terms of Σ_x and Σ_v, which can be used to further simplify the above equation. For example, for linear dynamical systems $f=\sqrt{a}x+\sqrt{b}v$ and $\Sigma_f=a\Sigma_x+b\Sigma_v$, Eq. (7) becomes

$$(a\Sigma_x+b\Sigma_v)^{-1}+\Sigma_s^{-1}=(1+h^2)\Sigma_x^{-1}. \tag{8}$$

From Eq. (7), the dependence of Σ_x on Σ_v and Σ_e can be obtained, which can be used to find out the "leading term" of the noise source—i.e., which noise term has the most influence on Σ_x and therefore on the estimation precision. This information can be useful for improving the estimation precision by suppressing the leading noise source. In practice, due to the nonlinearity of the function g, an explicit expression of Σ_s is not always possible and a Monte Carlo scheme is viable: i.e., draw a set of samples $\{e_i\}$ from the distribution of e [Eq. (2)]; for each sample e_i, calculate x_i as $x_i=g^{-1}(y(0),e_i)$, then Σ_s can be approximated by the covariance matrix of the samples $\{x_i\}$.

IV. ROBUST SEQUENTIAL MONTE CARLO SCHEME FOR MITIGATING OUTLIERS

The above scheme of sequential Monte Carlo simulations works well for stationary noise. In the presence of nonstationary disturbances—e.g., outliers—the weight-updating scheme needs to be improved. To be concrete, we treat outliers in the observation, which can lead to a larger covariance matrix Σ_s. This will cause a larger estimation error [Eq. (7)]. Thus, if the outliers can be detected and then discarded, the elements of the covariance matrix of the observations can be reduced. Our idea is to first calculate the empirical distribution of the Monte Carlo samples and then compare the observation with this distribution. If the prediction of the observation is close to the center of the samples, the observation is likely to be true and it should be accounted for in the estimation of the state. However, if the prediction de-

viates from the center of the samples, it is less reliable and should therefore be counted less [24]. Quantitatively, it is convenient to introduce a contribution factor α to characterize this effect. Let $\tilde{w}_i(t)=w_i(t_{k-1})p[y(t)|x_i(t)]$. We modify Eq. (5) as

$$w_i(t) = (1-\alpha)w_i(t_{k-1}) + \alpha\tilde{w}_i(t) \Big/ \sum_j \tilde{w}_j(t). \tag{9}$$

Generally, the optimal value of α depends on the distribution of the samples and on the prediction of the observation in a sophisticated way. Our strategy for choosing α is the following. After propagating the samples through the dynamical equation, we calculate Σ_x. The covariance matrix Σ_v of the dynamical noise and Σ_e of the observational noise are assumed to be known. Define $\beta \equiv \sqrt{\mathrm{Tr}(\Sigma_v)/\mathrm{Tr}(\Sigma_e)}$ and $\Delta x \equiv (x_{LS}-\langle x\rangle)/\sqrt{\mathrm{Tr}(\Sigma_x)}$, where $\langle x\rangle$ is the average of the samples and x_{LS} is the least-squares estimation of the state, which minimizes the square error of the observations (this is the case when the number of observations is more than the number of unknowns) [25]. We propose the following criterion for choosing α:

$$\alpha = \begin{cases} 1, & \|\Delta x\| \leqslant c_0, \\ (1-\beta)\dfrac{c_0}{\|\Delta x\|}\left(\dfrac{c_1-\|\Delta x\|}{c_1-c_0}\right)^2 + \beta, & c_0 < \|\Delta x\| \leqslant c_1, \\ \beta\dfrac{c_1}{\|\Delta x\|}\left(\dfrac{c_2-\|\Delta x\|}{c_2-c_1}\right)^2, & c_1 < \|\Delta x\| \leqslant c_2, \\ 0, & c_2 < \|\Delta x\|, \end{cases}$$

where $c_0 \sim 1$, $c_1 \sim 3.5$, and $c_2 \sim 7$, and the optimal values can vary for different situations. Note that $\|\Delta x\|$ is the distance between the estimation x_{LS} obtained from the observations and the mean value of the samples $\langle x\rangle$, normalized by the "standard deviation." If $\|\Delta x\| < 1$, the estimation is within the range of the standard deviation and is reliable. If $\|\Delta x\|$ is in the range of one standard deviation to three standard deviations, the observation is less reliable. Since there is noise in the dynamics, the samples may themselves contain some error. The quantity β is introduced to account for such uncertainties. If $\|\Delta x\|$ is even larger, the weight for the observation decreases, and at a certain point—say, beyond seven standard deviations—the observation is deemed as outliers and the weight α is set to zero.

V. NUMERICAL SUPPORT

To substantiate our ideas, we consider a synthetic two-dimensional positioning problem of moving object—say, a car—by using GPS observations (see Fig. 1). The origin is the center of the Earth, and the car is originally located at the surface of the Earth—$(0, R_e)$ in an Earth centered coordinate, where $R_e=6357$ km is the radius of the Earth—which is assumed to be unknown. The velocity can be read from the velocimeter and has a constant true value of 70 mph, or 31.3 m/s. The velocity is assumed to have a Gaussian measurement error with covariance matrix $\Sigma_v=\sigma_v^2\,\mathrm{diag}[1\ 1]$, where σ_v is regarded as an adjustable parameter. The direction of the velocity changes in time. The car is equipped with

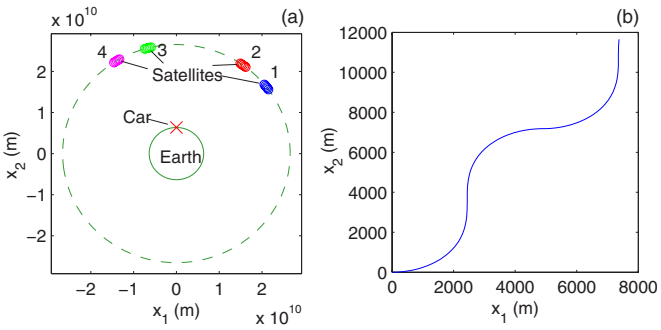

FIG. 1. (Color online) Setup of the numerical problem. (a) The tracks of the satellites and the car for the simulation time (500 s). The satellites move counterclockwise. (b) The track of the car, where the origin is shifted to the original position of the car.

a GPS receiver. Four visible satellites ($n_s=4$) are located at the altitude of 20 200 km above the Earth's surface with initial angles $\pi/5$, $7\pi/24$, $4\pi/7$, and $2\pi/3$ (in Earth-centered coordinates). The satellites orbit the Earth at a period of 12 h. The receiver on the car can receive GPS signals from each satellite at the frequency of 1 Hz ($\Delta t=t_k-t_{k-1}=1$ s), from which the distances from the satellites to the car, P^k (pseudoranges), can be measured. Assuming that the receiver has no clock offset and the satellites are not correlated, we can write the covariance matrix of the pseudoranges as $\Sigma_p=\sigma_p^2 I_{n_s}$, where I_{n_s} is the identity matrix of order n_s. A motion model which is linear in the state dynamics and nonlinear in the measurements is [12]

$$x(t) = x(t_{k-1}) + v(t_{k-1})\Delta t,$$

$$y(t) = g[x(t)] + e_p(t),$$

where y is the pseudorange measurements $y=[P^1 P^2 \cdots P^{n_s}]^T$. The measurement function is $g(x)=[R^1 R^2 \cdots R^{n_s}]^T$, where $R^j=\|X^j-x\|$ is the Euclidean distance from the car's position x to the jth satellite X^j and e_p is the pseudorange observational noise. The covariance matrices for v and e_p are Σ_v and Σ_p, respectively.

Figure 2 shows the dependence of Σ_x on Σ_v when Σ_s is given, which can be obtained numerically from the distribution of the pseudoranges (Σ_p). The symbols are obtained from direct simulations; the curves are from our theory Eq. (8). They agree quite well.

To test the robustness of our Monte Carlo strategy, we add 15 random outliers of 30 m to the measured pseudoranges of the second satellite in a time span of 500 s with measurement frequency 1Hz. The result of the position estimation is shown in Fig. 3. Three cases are presented for comparison. Figure 3(a) shows the prediction error of x_1 if we only know the initial position and the measured velocity, with a standard deviation $\sigma_v=0.5$. There is a systematic drift of errors. Figure 3(b) shows the prediction error of the least-squares method [25] if only the measured GPS pseudoranges are available (standard deviation $\sigma_p=2.5$ for each satellite). Figure 3(c) is the estimation error from our proposed Monte Carlo strategy, which apparently contains no systematic error and exhibits much smaller statistical errors.

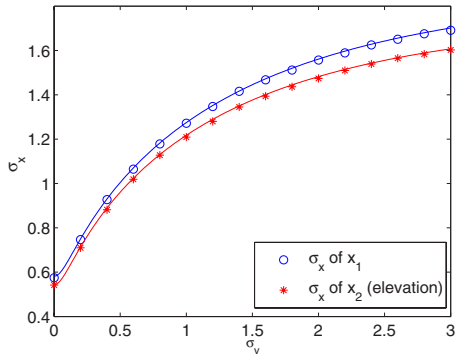

FIG. 2. (Color online) For the two-dimensional positioning problem, dependence of the standard deviations of the samples σ_x for the two coordinates on the standard deviation of the velocity σ_v. The standard deviation of the pseudoranges is $\sigma_p=2.5$, the number of samples is $N=1000$, and $h=0.3$. Symbols are obtained from numeric simulations, where each data point is the average of 10 runs, and 100 different time steps for each run are used. The curves are from our theoretical prediction, Eq. (8).

Next, we study the cases with non-Gaussian noise in the GPS pseudorange observations. This might be the case when the car is moving in a city or in the forests, where the GPS signal may be blocked by buildings or trees, causing difficulty in distinguishing multipath signals from the original signal, which may introduce systematic biases [26]. Furthermore, because of the complexity of the environment, at certain moments the original signal may be unavailable. We assume that the distribution of noise in the pseudoranges consists of two Gaussians with different mean values. The probability density function is

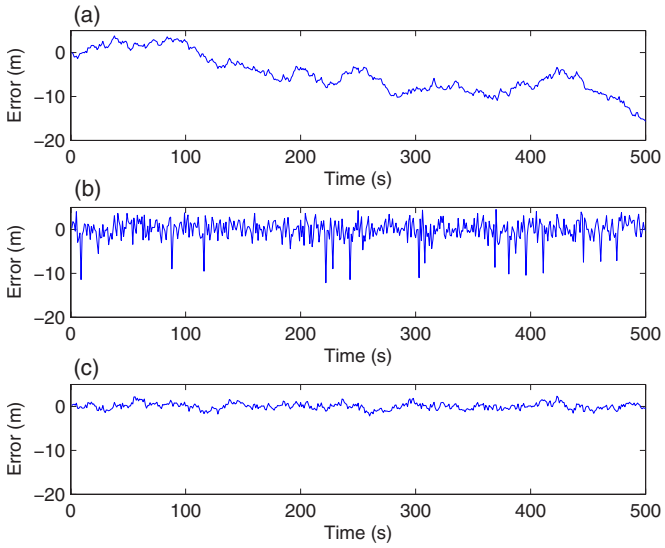

FIG. 3. (Color online) A comparison of the errors of the first coordinate x_1 of the position prediction by velocity information only (a), by least squares estimation of GPS distances with outliers only (b), and by our strategy (c). The parameter values used in the simulation are $\sigma_v=0.5$, $\sigma_p=2.5$, $c_0=0.7$, $c_1=4.2$, $c_2=7$, $N=1000$, and $h=0.3$.

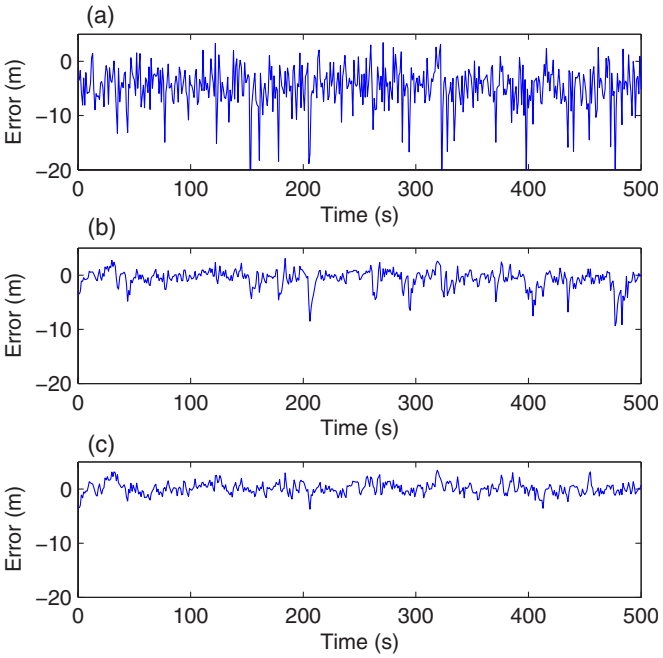

FIG. 4. (Color online) A comparison of the errors of the second coordinate x_2 of the position prediction only by least squares estimation of GPS distances with outliers (a), by regularized sequential Monte Carlo simulation without the robust strategy (b), and by our strategy (c). The parameter values used in the simulation are $\sigma_v=1$, $\sigma_p=2.5$, $c_0=1.5$, $c_1=4.2$, $c_2=8$, $N=1000$, and $h=0.3$.

$$f(x) = b\frac{1}{\sqrt{2\pi}\sigma_1}e^{-x^2/2\sigma_1^2} + (1-b)\frac{1}{\sqrt{2\pi}\sigma_2}e^{-(x-x_0)^2/2\sigma_2^2}, \tag{10}$$

where b is a weight factor and we take $b=0.6$ in our simulation. Other parameters are $\sigma_1=\sigma_p=2.5$, $\sigma_2=1$, and $x_0=8$.

Again, to test the robustness of our algorithm, we add outliers to the GPS pseudorange observations: 20 outliers of 40 m are added to satellite 2 randomly. Figure 4 compares the errors in position estimation by three methods: the least-squares estimation from the GPS pseudoranges with outliers (a), the estimation by the regularized sequential Monte Carlo scheme (b), and the estimation by our robust sequential Monte Carlo scheme (c). The least-squares estimation from the pseudoranges has large errors and can have systematic deviations (the average of the error is not zero), as shown in Fig. 4(a). The regularized sequential Monte Carlo scheme can remove these systematic deviations caused by the non-Gaussian distribution but is affected by the outliers, which can be seen from the spikes in Fig. 4(b). Our robust sequential Monte Carlo scheme can recover from both the systematic deviations and the outliers [Fig. 4(c)]. In fact, the average absolute value of the errors can be 30% smaller.

The current robust scheme deals with observational outliers. If there are dynamical outliers—e.g., the outliers appearing in v—the current scheme needs to be modified to cope with the problem. Observations after such an event will be needed to identify an outlier.

VI. CONCLUSIONS

In conclusion, we have obtained a self-consistent equation for the estimation precision of the Bayesian inference in terms of the dynamical noise and the observational noise levels. The equation may provide insights into designing improved sequential Monte Carlo simulations with higher precision. We have also proposed a strategy to deal with sudden, large disturbances that are inevitable in physical observations. The effectiveness of our method has been tested numerically. Sequential Monte Carlo simulations have begun to be used widely in various estimation problems in science and engineering. Our contribution provides a robust strategy for improving the estimation precision when experimental observations are nonstationary or even temporally interrupted.

ACKNOWLEDGMENT

This work is supported by AFOSR under Grants Nos. FA9550-04-1-0115 and FA9550-06-1-0024.

[1] A. Silberfarb, P. S. Jessen, and I. H. Deutsch, Phys. Rev. Lett. **95**, 030402 (2005).
[2] E. Bagan, M. A. Ballester, R. Muñoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **95**, 110504 (2005).
[3] J. P. M. Heald and J. Stark, Phys. Rev. Lett. **84**, 2366 (2000).
[4] R. Meyer and N. Christensen, Phys. Rev. E **65**, 016206 (2001).
[5] P. Crotty, J. García-Bellido, J. Lesgourgues, and A. Riazuelo, Phys. Rev. Lett. **91**, 171301 (2003).
[6] M. Wyman, L. Pogosian, and I. Wasserman, Phys. Rev. D **72**, 023513 (2005).
[7] N. Christensen and R. Meyer, Phys. Rev. D **64**, 022001 (2001).
[8] N. J. Cornish and J. Crowder, Phys. Rev. D **72**, 043005 (2005).
[9] W. Rieping, M. Habeck, and M. Nilges, Science **309**, 303 (2005).
[10] V. Baldazzi, S. Cocco, E. Marinari, and R. Monasson, Phys. Rev. Lett. **96**, 128102 (2006).
[11] R. Jansen *et al.*, Science **302**, 449 (2003).
[12] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, IEEE Trans. Signal Process. **50**, 425 (2002).
[13] E. T. Jaynes, in *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst (Cambridge University Press, New York, 2003).
[14] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
[15] *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon (Springer-Verlag, New York, 2001).
[16] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, IEEE Trans. Signal Process. **50**, 174 (2002).
[17] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, IEEE Signal Process. Mag. **20**(3), 19 (2003).
[18] L. Fahrmeir and R. Künstler, Metrika **49**, 173 (1999).
[19] T. Proietti, J. Time Ser. Anal. **24**, 221 (2003).
[20] A. H. Jazwinsky, *Stochastic Processes and Filtering Theory* (Academic Press, New York, 1970).
[21] E. A. Wan and R. Van der Merwe, in *Kalman Filtering and Neural Networks* (Wiley, New York, 2001), Chap. 7.
[22] A. Doucet, S. Godsill, and C. Andrieu, Stat. Comput. **10**, 197 (2000).
[23] C. Musso, N. Oudjane, and F. LeGland, in *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon (Springer, New York, 2001).
[24] Y. Yang, H. He, and G. Xu, J. Geod. **75**, 109 (2001).
[25] G. Strang and K. Borre, *Linear Algebra, Geodesy, and GPS* (Wellesley-Cambridge Press, Wellesley, MA, 1997).
[26] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *GPS: Theory and Practice* (Springer-Verlag, Wien, 2001).