Mean position tracking of respiratory motion

Dan Ruan[a] and Jeffrey A. Fessler
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109

James M. Balter
Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109

(Received 20 June 2007; revised 13 November 2007; accepted for publication 15 November 2007; published 30 January 2008)

Modeling and predicting tumor motion caused by respiration is challenging due to temporal variations in breathing patterns. Treatment approaches such as gating or adaptive bed adjustment/alignment may not require full knowledge of instantaneous position, but might benefit from tracking the general trend of the motion. One simple method for tracking mean tumor position is to apply moving average filters with window sizes corresponding to the breathing periods. Yet respiratory motion is only semiperiodic, so such methods require reliable phase estimation, which is difficult in the presence of noise. This article describes a robust method to track the mean position of respiratory motion without explicitly estimating instantaneous phase. We form a state vector from the respiration signal values at the current instant and at a previous time, and fit an ellipse model to training data. Ellipse eccentricity and orientation potentially capture hysteresis in respiratory motion. Furthermore, we provide two recursive online algorithms for real time mean position tracking: a windowed version with an adaptive window size and another one with temporal discounting. We test the proposed method with simulated breathing traces, as well as with real time-displacement (RPM, Varian) signals. Estimation traces are compared with retrospectively generated moving average results to illustrate the performance of the proposed approach. © *2008 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2825616]

Key words: semiperiodicity, mean position estimation, ellipse fitting, algebraic distance, generalized eigen decomposition, stationary ergodic process, adaptive estimation, effective memory length

I. INTRODUCTION

Accurate modeling and prediction of tumor motion caused by breathing is a challenging problem. Previous studies[1–4] have noted the difficulty of instantaneous position tracking and prediction. Given such limitations in accuracy, and considering the actual dosimetric impact of small motion variations, treatment approaches such as gating or adaptive bed adjustment/alignment may not require instantaneous position, but might benefit from following trends of the motion, in particular mean position drifting and/or abrupt shifts. Current amplitude-based gating systems compare an instantaneous tumor location measurement with a predetermined gating window threshold and trigger the treatment beam on/off. A potential modification to such systems would incorporate real-time mean drifting information to (1) adjust bed position to compensate for continuous mild drifting; (2) trigger the treatment beam off upon detection of significant drift. Compensating for mean position drifting could increase effective delivered dose given a fixed treatment margin, or alternatively, it could allow the use of smaller margins to achieve the same dose delivery. Previously, other investigators have shown that there could be only limited gains in trying to eliminate breathing movement completely, and laid the groundwork for consideration of the methodology described here. Engelsman demonstrated that the margin needed for cyclic breathing can be represented as a Gaussian with standard deviation of 0.4 times the amplitude of motion.[5] Wolthaus demonstrated a method for efficiently selecting a mean patient representation from a four-dimensional computed tomography data set.[6] Evidence from these and other investigations[7] hint at the possibility that a "tracking" system that estimates variation in position such as the local mean may provide significant benefit by reducing or eliminating nonperiodic trends in motions, while reducing demands on temporal response and acceleration of couch or multileaf collimator-based adjustment systems.

Figure 1 illustrates how real-time knowledge of mean drifting helps to reduce gating margin for the same treatment dose delivery (90% in this simulation). In this example, mean compensation reduces the margin by about 70% compared to traditional static gating approach.

Furthermore, mean drift (or home position motion) is more stable, with slower temporal variation than instantaneous position. This makes it more practical to seek an estimator for this lower order quantity. By imposing smoothness, a good estimator should be less susceptible to noise than instantaneous position trackers.

The seemingly intuitive moving average filter is impractical for real-time application due to (1) the absence of "fu-

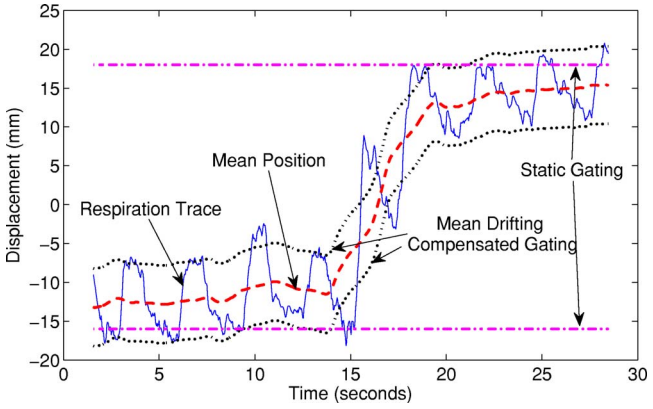

FIG. 1. Effect of drift compensation for gating system: respiration trajectory (solid line); mean position (dashed line): static gating with 90% delivery coverage (dashdot line); mean drift compensated dynamic gating with 90% coverage (dotted line).

ture" observations at the instant of estimation, and (2) the difficulty of estimating instantaneous phase online from noisy observations.

To circumvent the difficulties facing direct filtering approaches, we propose the following procedure. We form a state vector at each time instant from the current respiration value and the value from K samples earlier. We then fit a low-dimensional ellipse model to some or all of the state vectors. The state space helps capture respiration dynamics and elliptical model fitting has the advantage of being robust to noise, outliers and partially missing data, as discussed in Sec. II A. The semiperiodicity of respiratory motion promises reasonable shape inference from proper training data, while the low degrees of freedom of the ellipse model provides immunity against local noise (especially at extreme tidal positions, such as deep inhale and/or exhale). Ellipse eccentricity and orientation can potentially capture hysteresis in respiratory motion. We show that ellipse fitting can be solved using a generalized eigen decomposition. For real-time tracking, it is desirable to update the estimate upon arrival of new observations, but without recalculating all quantities. We provide algorithms for recursive fitting, using both temporal windowing and temporal discounting. Section II B discusses selection of window size and discount factor.

We tested the algorithm with both simulated data and clinical real-time position management (RPM) data. Experimental data are described in Sec. II B, and the test results are reported in Sec. III. The estimated trace is compared with retrospectively generated moving average results to illustrate the performance of our mean tracking approach. Finally, we summarize our work and discuss future directions in Sec. IV.

II. METHODS AND MATERIALS

II.A. Technical problem formulation

A strict definition of local "mean" position that reflects clinical significance is unavailable. On one hand, it is intuitive that this time varying quantity should convey local first-order statistical information; on the other hand, it is challenging to determine the intervals over which local statistics should be computed. The statistical properties of respiratory motion change with time, with semiperiodicity, which is important to consider when designing mean position estimators. As discussed in Sec. I, moving average filtering requires point-wise phase estimation which is sensitive to noise. These disadvantages motivate the search for a direct estimator for the mean statistics that is robust to noise. Computationally efficient algorithms are also necessary for online application.

For simplicity, we assume the observations to be a scalar function of time, termed "displacement." This is the scenario in RPM and many other external surrogate systems. When higher-dimensional data are available, such as internal three-dimensional tumor position readout from either gold implants or extracted from real-time imaging, we simply track the mean along each dimension separately. If desired, coupling among different directions could also be incorporated.

For our local mean estimator, we first form a state vector using the current observation and a past observation, intending to capture the (first order) system dynamics. We fit ellipses to the observed trajectory (its windowed or discounted version) in this state space. An ellipse has relatively few degrees of freedom, providing robustness to noise and to partially missing/unreliable observations. The latter property makes it possible to estimate the mean without meticulously choosing the training window size. Missing observations also can occur in practice when the imaging system loses track of the object of interest. Since we do not intend to use the ellipse fitting results to model the breathing trace, the benefit of robustness outweighs the issue of model mismatch. Our ellipse fitting method is invariant to affine transformations of the state vectors, making it well suited for unitless data such as RPM signals. Finally, orientation can indicate hysteresis; the less aligned the ellipse is with the Euclidean coordinates, the stronger the hysteresis.

Given a set one-dimensional discrete samples $x_i=s(t_i)$, we first augment it with $y_i=x_{i-k}$ where $\tau=k\Delta t$ indicates the delay for augmentation, assuming uniform sampling. The particular choice of delay length is not too crucial, and here we use $\tau=0.5$ s. (For noisy data, it is also feasible to locally smooth x_i and y_i to better reveal the system dynamics.) We first discuss the static ellipse fitting problem, where we fit an ellipse to a fixed collection of samples (x_i,y_i) for $i=1,2,\ldots,n$. We then provide recursive algorithms where our estimates evolve over time.

II.A.1. Ellipse fitting with static data

We model ellipses using a general quadratic curve equation. Let (x,y) denote the coordinates of a point in the two-dimensional state space, and define $\boldsymbol{z}=[x^2\ xy\ y^2\ x\ y\ 1]^T$, where superscript T denotes transpose. Then point (x,y) falls on the ellipse parameterized by $\boldsymbol{a}=[a\ b\ c\ d\ e\ f]^T$ if and only if they satisfy the following quadratic curve equation:

$$F(\boldsymbol{a},\boldsymbol{z})=\boldsymbol{a}^T\boldsymbol{z}=ax^2+bxy+cy^2+dx+ey+f=0 \qquad (1)$$

with negative discriminant, i.e., $b^2-4ac<0$.

The center (x_0, y_0) of the ellipse parameterized with $\boldsymbol{a}$ is given by

$$x_0 = \frac{2cd - bf}{b^2 - 4ac},$$

$$y_0 = \frac{2af - bd}{b^2 - 4ac}. \tag{2}$$

From Eq. (1), a sample $\boldsymbol{z}_i$ lies on a given ellipse parameterized by $\boldsymbol{a}$ if and only if $F(\boldsymbol{a}, \boldsymbol{z}_i) = 0$. This motivates the use of $F^2(\boldsymbol{a}, \boldsymbol{z}_i)$ as a measure of deviation of the sample from the ellipse. This is known as "algebraic distance" which coincides with Euclidean distance in the case F is a plane. It is computationally beneficial to adopt this discrepancy measure so that the collective distances for N samples can be conveniently written in standard matrix form and manipulated with classic least-squares approaches as in Eq. (3). For observed samples of the form (x_i, y_i), $i = 1, 2, \ldots, N$, we want to find the ellipse parameter $\boldsymbol{a}$ that minimizes the following cost function:

$$\sum_{i=1}^{N} F^2(\boldsymbol{a}, \boldsymbol{z}_i) = \boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}, \tag{3}$$

where we define the 6×6 empirical correlation matrix $\boldsymbol{S} = \sum_{i=1}^{N} \boldsymbol{z}_i \boldsymbol{z}_i^T$.

The minimizer of Eq. (3) is invariant to constant scaling applied to $\boldsymbol{a}$, so we impose the constraint that $b^2 - 4ac = -1$, or equivalently in matrix form $\boldsymbol{a}^T \boldsymbol{C} \boldsymbol{a} = 1$ with

$$\boldsymbol{C} = \begin{bmatrix} \tilde{\boldsymbol{C}} & 0_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} \end{bmatrix}, \tag{4}$$

where

$$\tilde{\boldsymbol{C}} \triangleq \begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix},$$

and $0_{3\times 3}$ denotes a 3×3 matrix of zeros. In other words, our ellipse fitting requires minimizing $\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}$ subject to the constraint that $\boldsymbol{a}^T \boldsymbol{C} \boldsymbol{a} = 1$. The center of the fitted ellipse given by Eq. (2) will be our mean position estimate.

Introducing the Lagrangian multiplier λ and differentiating, we need to solve the system of equations

$$\begin{cases} \boldsymbol{S}\boldsymbol{a} - \lambda \boldsymbol{C}\boldsymbol{a} = 0; \\ \boldsymbol{a}^T \boldsymbol{C} \boldsymbol{a} = 1. \end{cases} \tag{5}$$

We solve this using the generalized eigen decomposition of the pair $(\boldsymbol{S}, \boldsymbol{C})$. By Theorem 1 in Ref. 8, there is exactly one positive generalized eigenvalue and it corresponds to the unique local minimum of the Lagrangian. The corresponding eigenvector is the optimal solution to the ellipse parameter in Eq. (3). Let $(\lambda, \boldsymbol{u})$, denote the solution to the generalized eigenvalue problem with $\lambda > 0$, then $\hat{\boldsymbol{a}} = \boldsymbol{u} / \sqrt{\boldsymbol{u}^T \boldsymbol{C} \boldsymbol{u}}$ is the solution to the constrained minimization problem. The rank deficiency of $\boldsymbol{C}$ can cause instability issues if a conventional generalized eigen-decomposition algorithm were applied, e.g., Ref. 9, without caution. On the other hand, its sparsity may reduce computation. We describe next a specific algorithm for finding $\hat{\boldsymbol{a}}$ for this problem.

First, we decompose the 6×6 empirical correlation matrixes $\boldsymbol{S}$ into block form analogous to Eq. (4) as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{D} \end{bmatrix},$$

where each block is 3×3.

Then we use the following iterative algorithm to compute $\hat{\boldsymbol{a}}$:

$$\boldsymbol{a}_{n+1} = \eta_n \frac{\boldsymbol{a}_n^T \boldsymbol{W} \boldsymbol{a}_n}{\boldsymbol{a}_n^T \boldsymbol{C} \boldsymbol{a}_n} \boldsymbol{S}^{-1} \boldsymbol{C} \boldsymbol{a}_n + (1 - \eta_n) \boldsymbol{a}_n, \tag{6}$$

for iterations $n = 0, 1, \ldots$, where we define

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{E} & 0 \\ 0 & -\boldsymbol{D} \end{bmatrix}.$$

It is shown in Ref. 10 that $\boldsymbol{a}_n$ converges asymptotically to the eigenvector that corresponds to the unique positive eigenvalue of $(\boldsymbol{S}, \boldsymbol{C})$, provided that the stepsize parameter $\eta_n \in (0, 1)$ is asymptotically bounded above by $2/(\kappa + 1)$ with κ being the condition number of $(\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{C}})$, where, $\tilde{\boldsymbol{S}}$ is the Schur complement of the block $\boldsymbol{D}$ in $\boldsymbol{S}$ defined by $\tilde{\boldsymbol{S}} \triangleq \boldsymbol{E} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{B}^T$. The condition number of a generalized eigen decomposition is defined as $\kappa(\boldsymbol{A}, \boldsymbol{B}) = |\lambda_{\max}(\boldsymbol{A}, \boldsymbol{B}) / \lambda_{\min}(\boldsymbol{A}, \boldsymbol{B})|$ where $\lambda_{\max}(\boldsymbol{A}, \boldsymbol{B})$, $\lambda_{\min}(\boldsymbol{A}, \boldsymbol{B})$ denote the maximal and minimal (by moduli) generalized eigenvalues of $(\boldsymbol{A}, \boldsymbol{B})$.

The key ingredients underlying Eq. (6) (Ref. 10) are: (1) the rank deficiency of $\boldsymbol{C}$ makes it necessary to identify the "essential subspace" (corresponding to $\tilde{\boldsymbol{S}}$ and $\tilde{\boldsymbol{C}}$) and track its evolution; (2) to permit a stochastic approximation setting, the discrete time algorithm of interest is linked to a continuous time system that can be represented by an ordinary differential equation (ODE) whose convergence performance is derived. The convergence of the discrete algorithm is guaranteed if it follows the ODE close enough, which results in the condition on step size η_n.

To implement the algorithm in Eq. (6), we must choose η_n, which is essentially the step size for updating $\boldsymbol{a}$. We compute $\tilde{\boldsymbol{S}}$ from the empirical $\boldsymbol{S}$ for training data, then evaluate $\kappa(\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{C}})$, then set $\eta_n \leqslant 2/(\kappa + 1)$. Ideally, η_n should be set as close to $2/(\kappa + 1)$ as possible for fast convergence, and the speed is determined by the condition number κ of the $(\boldsymbol{S}, \boldsymbol{C})$ pair.

II.A.2. Adaptivity with new observations

The iteration in Eq. (6) provides ellipse fitting for a given set of data. For real-time use of mean tracking, we want an efficient procedure to update $\hat{\boldsymbol{a}}$ as each new data point is measured. In ellipse fitting, the matrix $\boldsymbol{C}$ describes the shape prior (constraint), and remains constant. Whenever new data become available, the empirical correlation matrix $\boldsymbol{S}$ changes. However, iteration (6) uses the inverse of $\boldsymbol{S}$, and we want fast methods for updating that inverse.

To express the time varying property of the system, we use $\boldsymbol{a}_n(i)$, $\boldsymbol{S}_n(i)$ etc. to denote the various quantities at a given acquisition time $t=i\Delta t$. For a given i, we rewrite Eq. (6) as follows:

$$\boldsymbol{a}_{n+1}(i)=\eta_n\frac{\boldsymbol{a}_n(i)^T\boldsymbol{W}(i)\boldsymbol{a}_n(i)}{\boldsymbol{a}_n(i)^T\boldsymbol{C}\boldsymbol{a}_n(i)}\boldsymbol{S}(i)^{-1}\boldsymbol{C}\boldsymbol{a}_n(i)+(1-\eta_n)\boldsymbol{a}_n(i),\quad n=0,1,\ldots,N_i-1, \tag{7}$$

where N_i denotes the number of iterations used to compute the ellipse parameters at a given time i.

When a new data sample becomes available ($i\rightarrow i+1$), we initialize the ellipse parameter as follows:

$$\boldsymbol{a}_0(i+1)=\boldsymbol{a}_{N_i}(i), \tag{8}$$

where $\boldsymbol{a}_{N_i}(i)$ is the state estimate obtained from the last inner iteration. The challenge is to compute the inverse of $\boldsymbol{S}(i+1)$ efficiently, and we provide below efficient rank-one updates for $\boldsymbol{S}^{-1}$ for both the sliding window adaptation and exponential discount adaptation.

1. **Sliding window update**

We define $\boldsymbol{S}(i)=\Sigma_{j=i-L+1}^{i}z_jz_j^T$ with L indicating the constant window size.

When a new sample pair z_{i+i} becomes available

$$\boldsymbol{S}(i+1)=\sum_{j=i+1-L+1}^{i+1}z_jz_j^T=\boldsymbol{S}(i)-z_{i-L+1}z_{i-L+1}^T+z_{i+1}z_{i+1}^T. \tag{9}$$

To compute $\boldsymbol{S}(i+1)^{-1}$ from $\boldsymbol{S}(i)^{-1}$, we denote $\boldsymbol{Q}(i)=\boldsymbol{S}(i)-z_{i-L+1}z_{i-L+1}^T$, so that $\boldsymbol{S}(i+1)=\boldsymbol{Q}(i)+z_{i+1}z_{i+1}^T$. We invoke the Woodbury matrix identity[11] to compute $\boldsymbol{S}(i+1)^{-1}$ with two step rank-one updates:

$$\begin{aligned}\boldsymbol{Q}(i)^{-1}&=(\boldsymbol{S}(i)-z_{i-L+1}z_{i-L+1}^T)^{-1}\\&=\boldsymbol{S}(i)^{-1}\\&\quad-\boldsymbol{S}(i)^{-1}z_{i-L+1}(z_{i-L+1}^T\boldsymbol{S}(i)^{-1}z_{i-L+1}-1)^{-1}\\&\quad z_{i-L+1}^T\boldsymbol{S}(i)^{-1},\end{aligned} \tag{10}$$

$$\begin{aligned}\boldsymbol{S}(i+1)^{-1}&=(\boldsymbol{Q}(i)+z_{i+1}z_{i+1}^T)^{-1}\\&=\boldsymbol{Q}(i)^{-1}\\&\quad-\boldsymbol{Q}(i)^{-1}z_{i+1}(z_{i+1}^T\boldsymbol{Q}(i)^{-1}z_{i+1}+1)^{-1}z_{i+1}^T\boldsymbol{Q}(i)^{-1}.\end{aligned} \tag{11}$$

This pair of properties provides a recursion in $\boldsymbol{S}(i)^{-1}$. Substituting into Eq. (7) yields a recursion in the estimation parameters $\boldsymbol{a}(i+1)$.

2. **Discounting update**

As an alternative to a fixed-length sliding window, we can use temporal discounting to emphasize the most recent data. In this case, we define $\boldsymbol{S}(i)=(1-\gamma)/(1-\gamma^t)\Sigma_{j=1}^{i}\gamma^{i-j}z_jz_j^T$, where $\gamma\in(0,1)$ is a user-selectable discounting parameter. We can easily write $\boldsymbol{S}(i+1)$ recursively as

$$\frac{1-\gamma^{i+1}}{1-\gamma}\boldsymbol{S}(i+1)=\gamma\frac{1-\gamma^{i}}{1-\gamma}\boldsymbol{S}(i)+z_{i+1}z_{i+1}^T. \tag{12}$$

Invoking the matrix inversion lemma yields the recursion for $\boldsymbol{S}^{-1}$

$$\begin{aligned}\boldsymbol{S}(i+1)^{-1}&=\frac{1-\gamma^{i+1}}{\gamma-\gamma^{j+1}}\boldsymbol{S}(i)^{-1}\\&\quad-\boldsymbol{S}(i)^{-1}\boldsymbol{x}_{i+1}\left\{\frac{1-\gamma}{\gamma-\gamma^{i+1}}\right.\\&\quad\left.+\boldsymbol{x}_{i+1}^T\boldsymbol{S}(i)^{-1}\boldsymbol{x}_{j+1}\right\}\boldsymbol{x}_{i+1}^T\boldsymbol{S}(i)^{-1}.\end{aligned} \tag{13}$$

Substituting this in Eq. (7) yields an adaptive ellipse fitting algorithm with temporal discounting.

The size of the window width L and the discount factor γ control the trade-off between response speed and smoothness of the tracking trace in each adaptive algorithm, respectively. Even though the ellipse fitting method is robust to missing data (e.g., a partial period), it is still desirable to react more promptly when changes are more frequent (short underlying breathing periods and/or rapid shifts in mean position) and track stably otherwise. For fixed-length sliding window adaptivity. it is preferable to choose a window size that roughly matches the "true" period of the signal. Therefore, we use a short segment of training data at the beginning of each treatment fraction, find the closest periodic function to the training segment using a subspace projection method[12] and use the derived period as the fixed window length L.

We could choose the discount factor γ analogously by using *effective memory length*, defined by

$$\tilde{L}(i)=\sum_{j=1}^{i}\gamma^{j-j}, \tag{14}$$

because the time unit has a more intuitive physical interpretation. For large i, the efficient memory length is $\tilde{L}=1/(1-\gamma)$. In other words, we expect the performance of an adaptive mean tracker with discount factor $\gamma=1-1/L$ to behave similarly to a sliding window estimator with window size L. In general, however, the discount method should be more stable, but less responsive towards changes than the corresponding sliding window approach with $L=\tilde{L}$ because previous samples are never completely "forgotten." Thus, we use the period estimated from projection as noted before[12] to find L from 20 s of training data, then find γ such that

$$\gamma^{\alpha L}=\beta,$$

where the pair of parameters (α,β) adjusts the decay rate. It has the interpretation that the effect of a given sample decays to β after α periods. We found that in practice $\alpha=1$ and $\beta=0.05$ is a reasonable choice and we use these values in later investigations.

II.A.3. Alternative mean estimators

- **Moving average**

The simplest way to compute local first order statis-

tics is to use the empirical average of observation samples in an appropriate window around each time instant, i.e., form the neighborhood (window) and take the empirical average. As the time instant of interest proceeds, the window moves at the same speed.

The traditional method of moving average over fixed window is formulated as (we assume the window size L to be even to simplify notation).

$$x_0(i) = \frac{1}{L_i + 1} \sum_{j=i-L_i/2}^{i+L_i/2} x_i, \tag{15}$$

where L_i is the "period" near sample i.

One obvious issue with this approach is that the "future" samples $\{j=i+1,\ldots,i+L/2\}$ are not available for the computation of the instantaneous mean at time $t=i\Delta t$. Furthermore, even if the signal is regular and periodic, in which case those values can be inferred fairly accurately, this estimator is effective only if the window sizes closely match with the signal period.

- **Eigen decomposition in the original data space**

  As the ellipse fitting methods aim to exploit the scatter structure in the augmented state space $z_i=(x_i,y_i)$, one may wonder whether the conventional eigen decomposition (also known as "principle component analysis") approach applies. Eigen decomposition may capture the elliptical shape by detecting the major and minor direction of signal variation (the direction of eigen-vector) and their different energy levels, but it is not useful in mean estimation. In fact, the center of such "ellipse" from eigen decomposition would be exactly the same as the average over the data segment used, because the quantities that enter the covariance matrix are first adjusted to have zero mean. Therefore, eigen decomposition would perform exactly the same as moving average filters for the task of mean estimation.

II.B. Experiment setup

We simulated two sets of data so that we could have "ground truth" for verification purposes. For the first set of simulations, we used noise-free, strictly periodic data with both ideal sinusoid and modified cosine models.[13] In particular, the discrete sinusoidal and modified cosine waveforms were generated, respectively, with

$$x_i^{\sin} = x(i\Delta t) = x_0 + a\,\sin(\pi i\Delta t/T - \phi), \tag{16}$$

$$x_i^{\text{modified cos}} = x(i\Delta t) = x_0 - a\,\cos^{2n}(\pi i\Delta t/T - \phi), \tag{17}$$

where we used the value $n=2$. In the second test, we generated a semiperiodic sinusoid function with slow frequency drifting by modulating the *local* frequency with random offset components, as follows:

$$x_i^{\sin} = x_0 + a\,\sin\left\{\pi\Delta t\sum_{k=1}^{i}(1/T + \delta_k') - \phi\right\}, \tag{18}$$

$$x_i^{\text{modified cos}} = x_0 - a\,\cos^{2n}\left\{\pi\Delta t\sum_{k=1}^{i}(1/T + \delta_k) - \phi\right\}, \tag{19}$$

where the δ_k values were distributed via a Gaussian distribution $N(0,\sigma^2)$ with $\sigma \ll 1/T$. In the simulation, we set period $T=5$ s, $\Delta t=1/30$ s corresponding to a sampling frequency of 30 Hz, home position $x_0=0$, magnitude $a=5$ cm, and systematic phase offset $\phi=0$. Figure 2 shows typical simulation traces.

For real clinical data, we used the Real-Time Position Management (RPM, Varian Medical Systems, Palo Alto, CA) system to obtain the trajectories of external fiducials placed on the chests of 12 patients. The displacement-time relationship was recorded at 30 Hz and is assumed to be highly correlated with superior-inferior diaphragm motion,[4] which is a major source of respiratory motion for tumors in the chest or lung area. We centered and scaled the unitless RPM data so that their dynamic range corresponds to typical superior-inferior (SI) motion for chest and lung tumors.[1,14] We can thereafter consider the units to be on the order of mm for typical thorax tumor motion. Characteristic parameters for the RPM data used in our experiment are reported in Table I.

III. RESULTS AND DISCUSSION

III.A. Scatter plot in state space and robustness towards noncentered data

The fitting methods approximate data in the state space (x,y) by ellipses. It is desirable to have the center of such ellipse, which corresponds to the mean estimator, to be robust to missing data, spurious data, and to input data lengths that differ from the ideal period centered at the time instant of estimation. Figure 2 illustrates both data-abundant cases and the cases where only a segment (3 s worth) of arc data is available for fitting. The fitted ellipses are overlaid with the observation samples in the augmented state space. The second column in Fig. 2 illustrates that ellipses are reasonable approximations for the scattered observations in the state space. The difference between column 3 and 4 in Fig. 2 indicates the change of parameters in the presence of scarce and/or noncentered data. Not only does the ellipse fitting method degrade gracefully with partial data, but also the mean position estimated from this approach is reasonably stable. This empirical study illustrates the feasibility of using the proposed method in mean tracking and prediction.

III.B. Adaptive estimation

We first test the case where we use a fixed interval of the most recent data. In the real-time estimation and prediction setting, all the input samples into the estimation algorithm precede the time instant of interest. We also want to emphasize that the windowed history is used to help estimate the ellipse parameters; and it need not have integer multiples of the period. We tested the windowed ellipse fitting with 5 s and 7 s local history length, and report the results in Fig. 3. Discount adaptations yield very similar results to the win-

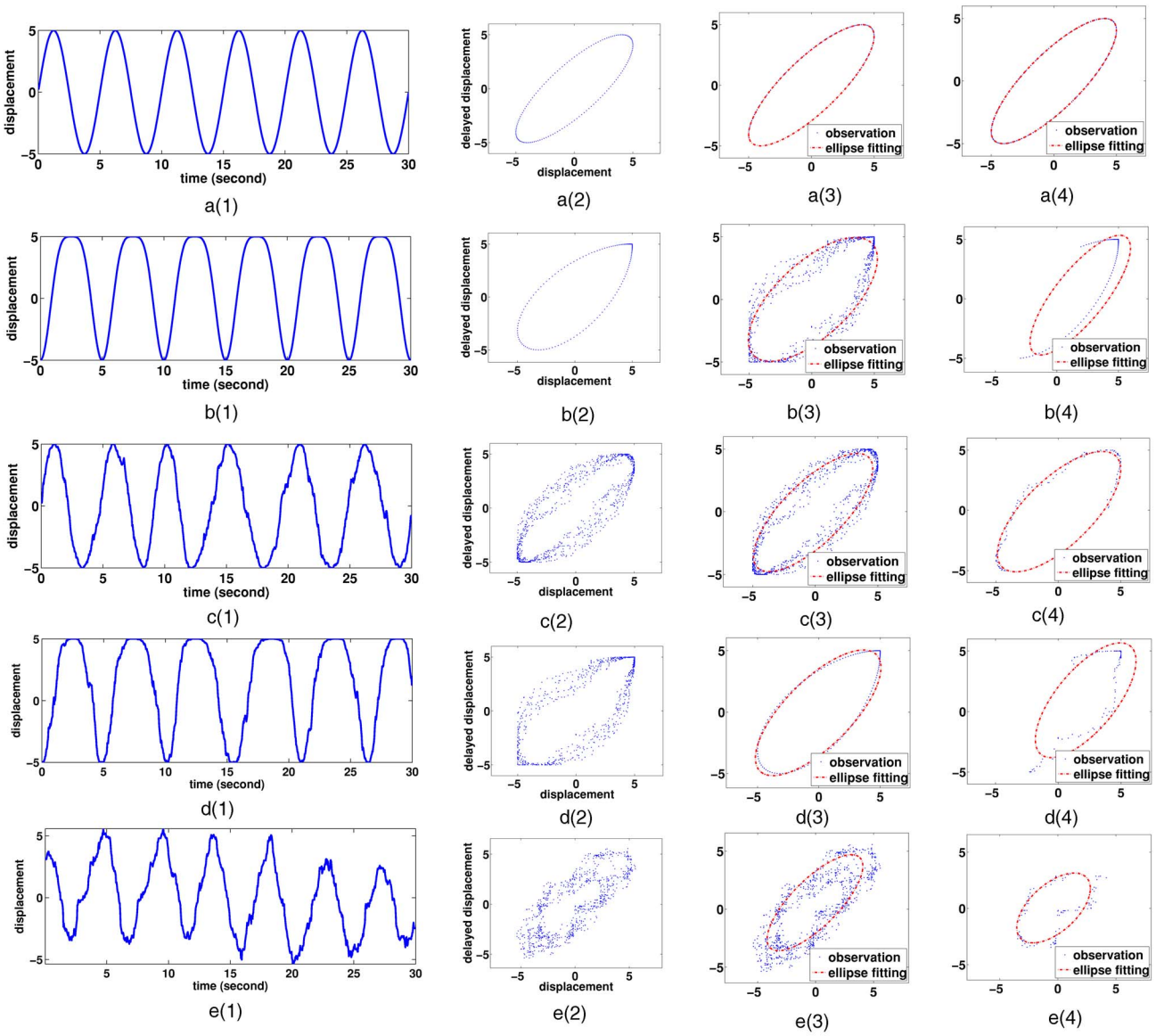

FIG. 2. Illustration of ellipse fitting performance of the proposed method. Each row corresponds to a different data source: row 1 (aX) ideal sinusoid; row 2 (bX) ideal modified cosine; (cX) locally modulated (noisy) sinusoid; (dX) locally modulated (noisy) modified cosine; (eX) clinical RPM trace scaled so that P-P $\approx$ 10 mm to mimic SI motion. Columnwise: X(1) time-displacement graph; X(2) augmented state space with displacement and its delay (τ=0.5 s); X(3) ellipse fitting (dashed line) applied to complete dataset; X(4) ellipse fitting (dashed line) applied to partial dataset.

dowed fitting, resulting in virtually overlapping real-time mean tracking curves. We omit them from the figures for visualization clarity. We also plot the outputs of two simple moving average filters with fixed window lengths.

We constructed our simulations to have frequency 0.2 Hz for deterministic cases or centered around that for the randomly frequency modulated realizations. Therefore, the ground-truth mean motion was zero for all the simulations.

TABLE I. RPM Dataset information.

ID V.S. Parameter	1	2	3	4	5	6	7	8	9	10	11	12
						Data characterization[a]						
STD	2.91	6.47	13.05	2.83	4.86	2.78	4.30	7.61	2.08	7.72	13.04	6.56
P-P	10.93	25.03	48.91	9.02	13.09	11.47	17.77	26.93	13.14	37.44	38.97	32.54
Period (sec)	4.5	4.6	7.2	5.6	4.4	5.4	4.7	9.7	4.7	4.1	3.1	5.2

[a]The data were adjusted to have globally zero mean; average periods were estimated with subspace projection method (see Ref. 12).

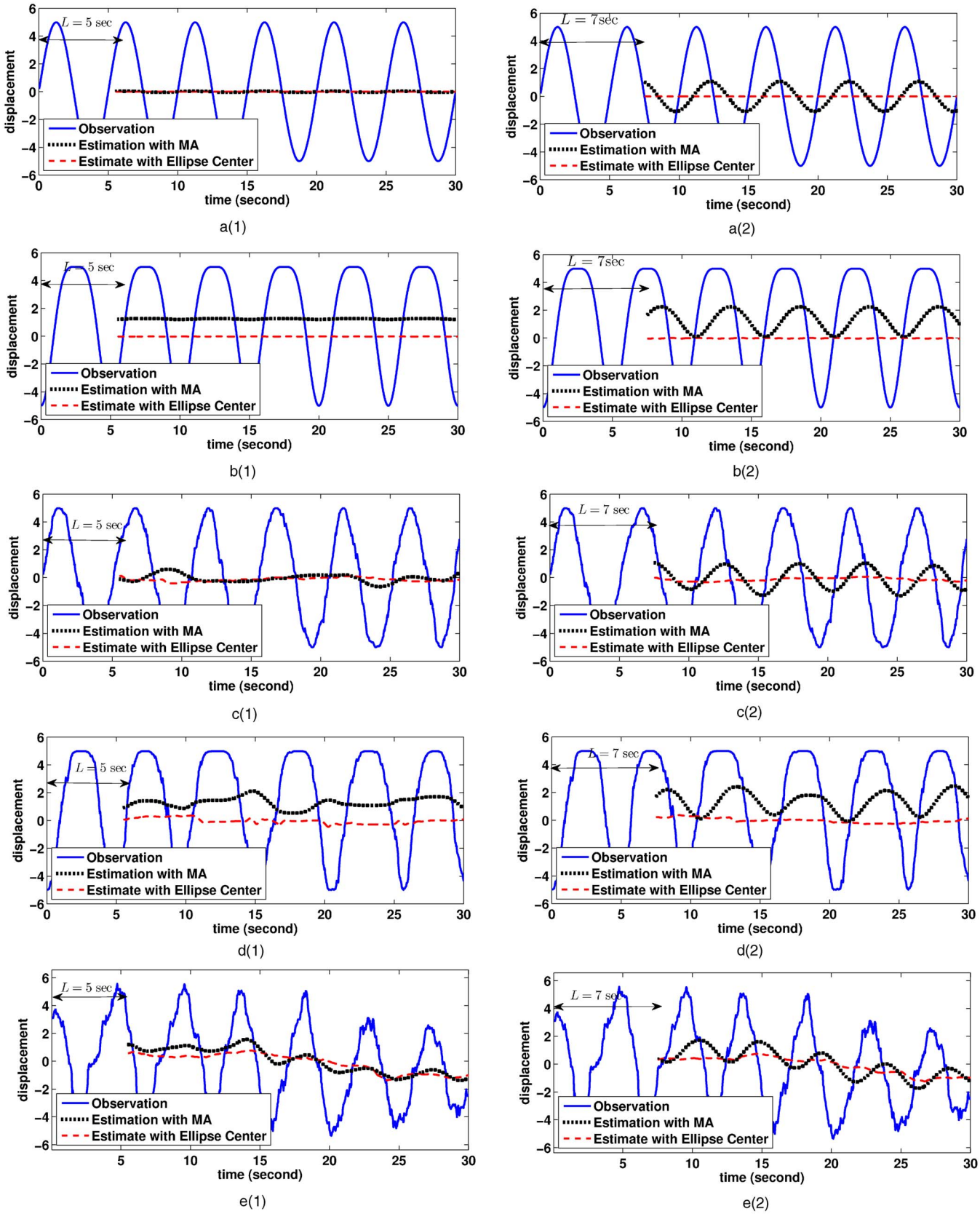

FIG. 3. Comparison of moving average (MA) and ellipse fitting estimator for mean position tracking: left column X(1): oracle history window length: L =5 s matches the underlying signal periods exactly; right column X(2): history window length L=7 s disagrees with the signal periods. Rows correspond to different data source as in Fig. 2. Solid line: observation signal; dotted line: moving average output; dashdot line: output from the ellipse fitting algorithm.

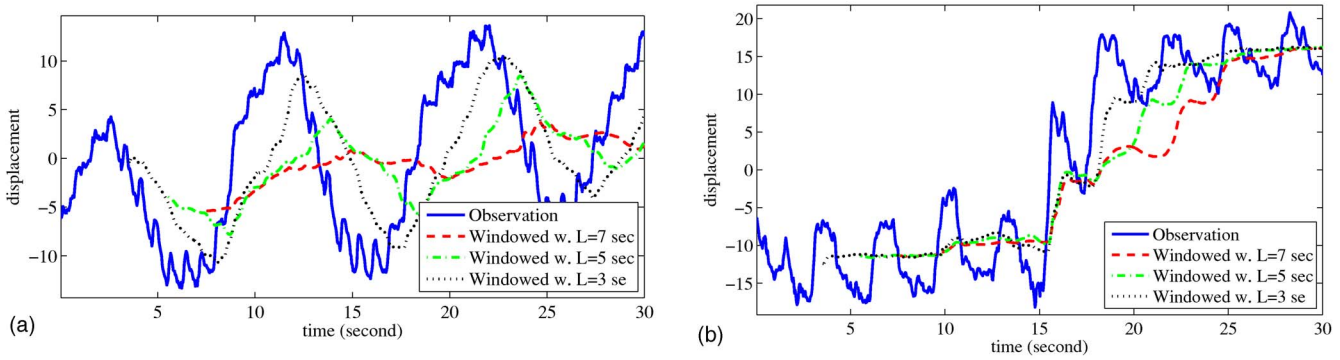

FIG. 4. Effect of window length L on tracking performance. Solid line: observation; dashed line: $L=7$ s; dashdot line: $L=5$ s; dotted line: $L=3$ s. (a) RPM with relatively long period; (b) RPM with relatively short period.

The clinical RPM data (Patient 1 in both Table I and Fig. 6) also has approximately the same frequency. Since both the simulated and clinical data lack mean drifting, a good estimator for the mean position should yield very stable (flat) output. When we select the training window size to be the "oracle" (ground-truth value unknown to the algorithm) value of $L=5$ s, which coincides with the signal period, outputs are stable from both the moving average operator and the proposed method [A constant offset, as observed in the modified cosine case, has marginal clinical effect, as long as it is consistent], as illustrated by the the left column in Fig. 3. On the other hand, it is impossible to guarantee that the history window size will always match the true period. We illustrate the effect of a disagreement, where window size $L=7$ s in the right column in Fig. 3. The moving average filter exhibits undesirable oscillations, whereas the ellipse fitting method provides comparable results as in the case of perfect period match.

As discussed in Sec. II A 2, the size of the sliding window and the discount factor must compromise between response speed (tracking efficiency) and robustness (tracking stability). Even though the ellipse fitting method is not too sensitive to the window size, it is helpful to choose window length L and discount factor γ from a short segment of training data. Figure 4 illustrates the effect of various choices of window length parameter L on mean estimation performance with some RPM data and Fig. 5 illustrates the effect of the discount factor γ. For RPM data with relatively long period and slow drifting, as in Figs. 4(a) and 5(a), it is desirable to use a larger window size (and correspondingly weaker discounting, large γ) to take advantage of its robustness. On the other hand, for breathing signals that have relatively short periods and rapid shifts in mean position, such as the one illustrated in Figs. 4(b) and 5(b), shorter window lengths and small discount factors are preferable for prompt response to mean changes.

To automatically adjust the sliding window length and the discount parameter, we take a short segment of training data at the beginning of each treatment fraction, and apply a subspace projection-based period estimation method.[12] For the signals in Fig. 4, the signal in subplot (a) yields a period estimate of 9.7 s and the signal in subplot (b) yields a period estimate of 3.1 s. Using the estimated period as the sliding window length and choosing the corresponding discount factor appear to be reasonable based on Figs. 4 and 5. We apply this scheme to automatically choose the adaptive parameters for all of the 12 RPM datasets and report the results in Fig. 6. For base line comparison, we collect the complete trajectory, and apply a moving average filter according to Eq. (15) with the oracle window size L to obtain a reasonable ground truth. The deviation of the two adaptive real-time mean position estimators from this "gold standard" (with constant offset

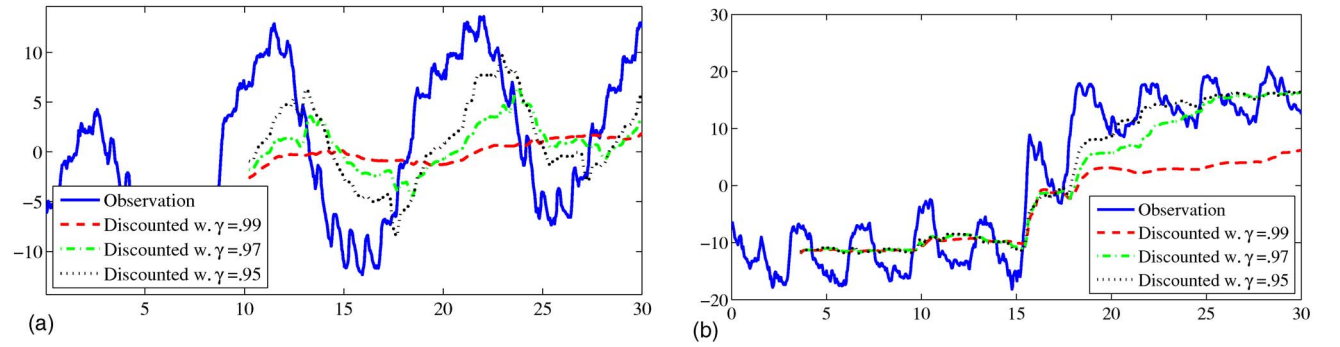

FIG. 5. Effect of discount factor γ on tracking performance. Solid line: observation; dashed line: $\gamma=0.99$; dashdot line: $\gamma=0.97$; dotted line: $\gamma=0.95$. (a) RPM with relatively long period; (b) RPM with relatively short period.

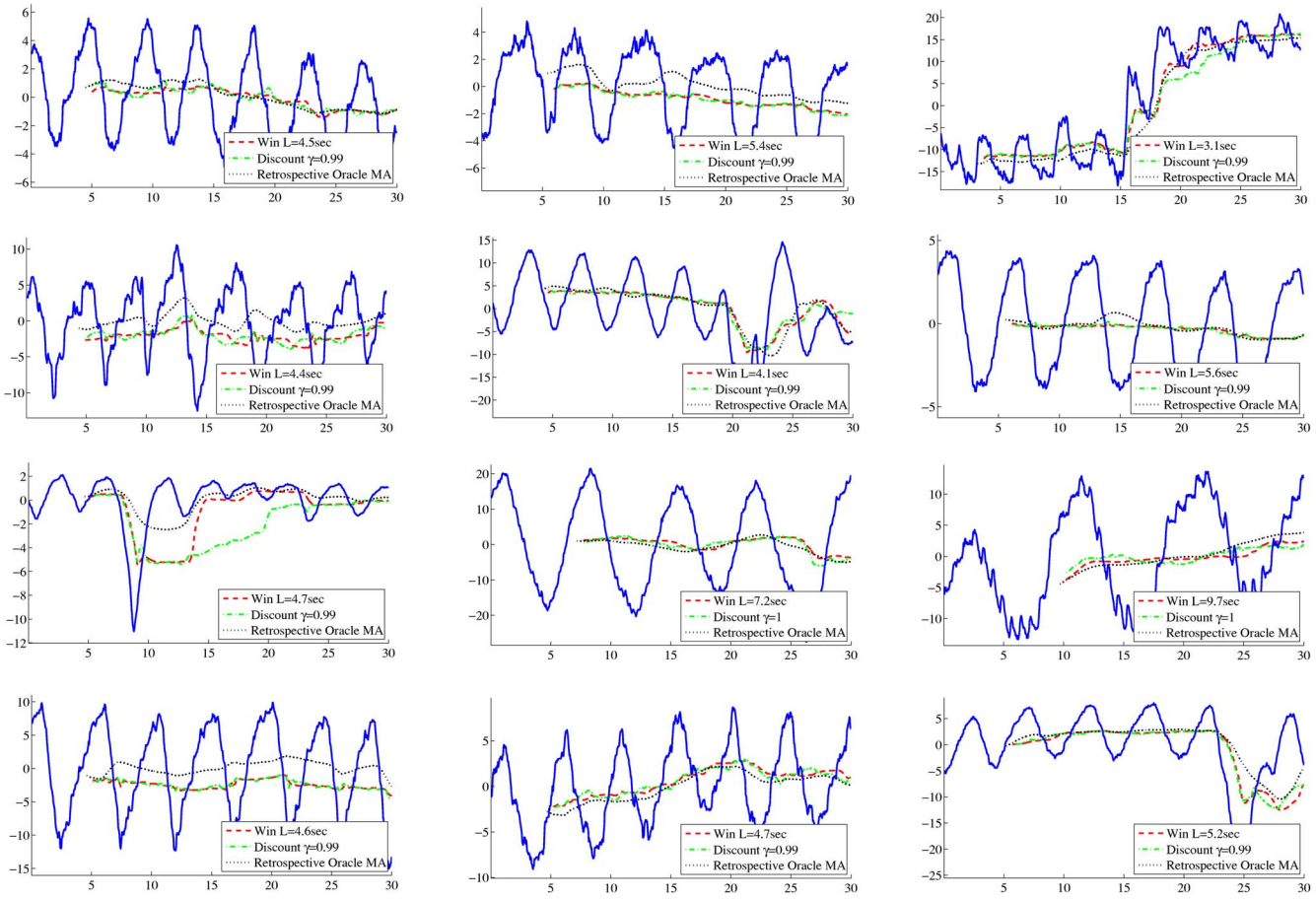

FIG. 6. Mean tracking for RPM data with window size determined by period estimator. Solid line: observed data; dotted line: retrospective moving average mean estimation with oracle period; dash line: sliding window mean estimator with window size L chosen with period estimation during training phase; dashdot line: discounting estimator with discount factor γ chosen such that $\gamma^{L/\Delta t}=1/20$.

compensated) is reported in terms of root mean squared error (RMSE) in Table II. Both adaptive methods demonstrate reasonable agreement with the retrospectively obtained ground truth.

III.B.1. Sensitivity to sampling rate

In some cases, it is preferable to obtain observations at a low frequency. This is particularly true when internal tumor motion is extracted from real-time imaging devices that would incur radiation dose. Sparse sampling poses a particular challenge to the conventional mean estimator based on a moving average filter, which is more vulnerable to miss calculation of period length when there are very few samples, resulting in intolerably high variance in mean estimation. We tested the use of sparse real-time observations by subsampling from the 30 Hz signal, applying both windowed and discounted adaptive algorithms to estimate the mean target position, and comparing with the retrospectively generated "true" mean from densely sampled data. Figure 7 illustrates how different observation rates affect overall RMS error across all patients. Both adaptive approaches are quite robust to low sampling rate. In particular, as the windowed adaptation only used historical samples that are within one period, which is normally about 4–6 s, the observable "break down" at 1 Hz in Fig. 7(a) corresponds to estimating the ellipse from 4–6 samples only, which is somewhat expected. On the

TABLE II. Mean estimation performance.

ID V.S. Parameter	1	2	3	4	5	6	7	8	9	10	11	12
Period L (sec)	4.5	4.6	7.2	5.6	4.4	5.4	4.7	9.7	4.7	4.1	3.1	5.2
Sliding Win RMSE	0.35	0.77	0.96	0.23	0.68	0.36	0.35	0.90	1.09	1.22	1.21	1.40
Discount factor γ	0.978	0.979	0.986	0.982	0.978	0.982	0.979	0.990	0.979	0.976	0.968	0.981
Discounting RMSE	0.36	0.77	1.08	0.24	0.71	0.35	0.44	1.22	1.54	1.55	2.18	1.39

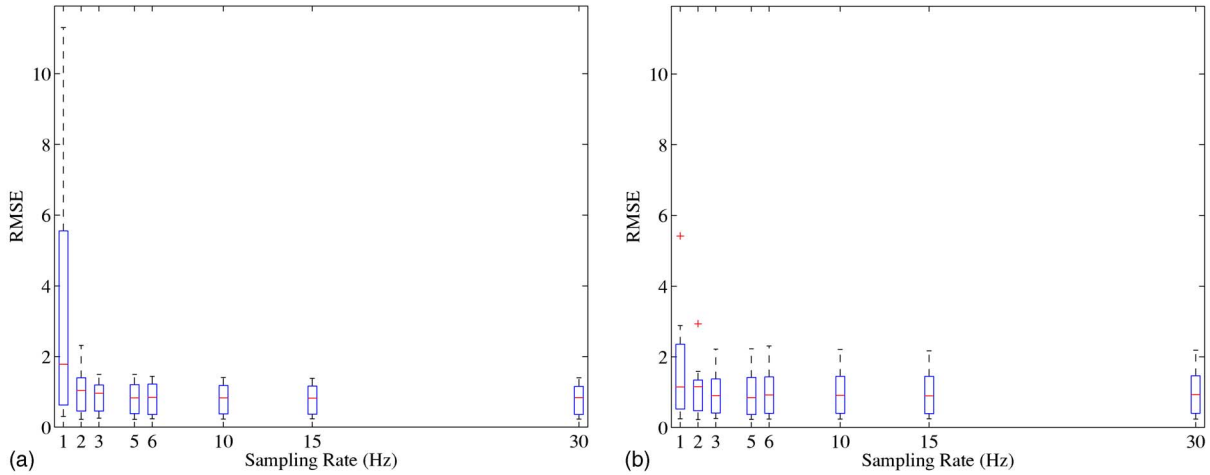

FIG. 7. Overall RMS error (across all patients) as a function of sampling rate: (a) with windowed ellipse fitting adaptivity; (b) with discounted ellipse fitting adaptivity with discount factor γ chosen such that $\gamma^{L/\Delta t}=1/20$. Both methods are robust above 2 Hz sampling rates.

other hand, the discounted adaptation utilizes all previous samples in a weighted fashion, and is naturally less affected by sparse sampling as shown in Fig. 7(b).

III.C. Discussion

Although we assumed uniform sampling for simplicity, the proposed approaches easily generalize to nonuniform sampling scenarios, thanks to the robustness of the fitting process. Lower sampling rates should affect the estimation less than the partial datasets tested in Fig. 2. Nearly uniform but sparse sampling along the ellipse would increase estimator variance, but should not introduce bias, unlike the partial data case where all the samples are concentrated along an arc segment.

Unlike simple filtering methods, the ellipse fitting method is more objective oriented: it is specifically designed for estimating time-variant mean of breathing signals. The ellipse model reflects the semiperiodicity of respiratory motion. The fitting process is flexible enough to capture changing trends yet is robust enough to control noisy oscillations. The adaptive algorithms provide efficient updates of the ellipses and allow the users to determine the update rates of the fitting. For adaptive methods using either sliding window or discounting factor, parameter selection involves the trade-off between system response speed and stability. We have suggested one way to adjust the sliding window length L based on the estimated nominal period length, and discussed a connection between the discount factor γ and the "effective memory length" $\tilde{L}$ to provide some guidance about the choice of those parameters. Fast drifting sequences require a more responsive system, and this should be reflected in the corresponding parameter settings. Even though the mean drifting pattern and the respiratory frequency are very often closely correlated, a slow (and regular) breathing pattern may still exhibit abrupt changes, as observed in the upper-left corner of Fig. 6. It is possible to resolve this issue by applying the proposed method on a training segment and then investigating the variation pattern of the estimated mean position to further decouple the different causes of the mean position changes. As relation (14) only holds asymptotically, and the discounting method is less forgetful than its sliding window counterpart, the discount factor may need to be further reduced to accommodate the more rapidly changing trends.

IV. CONCLUSION AND FUTURE WORK

We have proposed a robust method to track the mean position of respiratory motion. Modeling the breathing trajectory in the augmented state space as motion along ellipses incorporates the semiperiodicity and possible hysteresis of respiration. A shape prior that translates to constraints in the data fitting problem circumvents the challenge of precise phase estimation, and makes the overall method robust to partial data. We solve the optimization in terms of ellipse parameters with generalized eigen decomposition, and provide computationally efficient iterative algorithms for both static data and the adaptive case. In particular, we investigated two typical adaptive approaches: sliding window and exponential discounting. We derived the corresponding rank-one update for each setting, respectively, and studied ways to choose update parameters (window size and discounting factor). Experiments with both simulation and clinical RPM data illustrated the feasibility of the proposed method.

Our algorithms generalize easily to nonuniformly sampled observations and higher-dimensional cases. Commercial solvers for some intermediate steps, such as generalized eigen decomposition, are available. Clinical experience and physical prior knowledge can help guide choosing either the proper sliding window size or discount factor. In general, both the window size and the discount factor allow real-time adjustment (at the possible cost of more complicated update rules), and could even be tuned intra-fraction, if necessary. The intuitive interpretation of the parameters in terms of

window size L, effective memory length $\tilde{L}$ and decaying parameters (α,β) makes the control of those parameters practical.

In the future, we will investigate approaches to learn mean position drifting rate, detect abnormal abrupt changes, and properly adjust the adaptivity pace accordingly. This concerns the clinically significant question of "how far we can reliably extrapolate into future based on current observations." We will further validate the proposed method on alternative external surrogates and internal tumor trajectories upon availability of such data, as they may bear different noise properties. We would also like to study the application of the proposed method to different treatment methods, and look into software-hardware cooperation issues.

ACKNOWLEDGMENT

This work was supported in part by NIH Grant No. P01-CA59827.

[a]Electronic mail: druan@umich.edu

[1]G. C. Sharp, S. B. Jiang, S. Shimizu, and H. Shirato, "Prediction of respiratory tumour motion for real-time image-guided radiotherapy," Phys. Med. Biol. **49**, 425–440 (2004).

[2]D. P. Gierga, G. T. Y. Chen, J. H. Kung, M. Betke, J. Lombardi, and C. G. Willett, "Quantification of respiration-induced abdominal tumor motion and its impact on IMRT dose distributions," Int. J. Radiat. Oncol., Biol., Phys. **58**, 1584–1595 (2004).

[3]S. Jiang, C. Pope, K. Al Jarrah, J. Kung, T. Bortfeld, and G. Chen, "An experimental investigation on intra-fractional organ motion effects in lung IMRT treatments," Phys. Med. Biol. **48**, 1773–1784 (2003).

[4]S. S. Vedam, P. J. Keall, A. Docef, D. A. Todor, V. R. Kini, and R. Mohan, "Predicting respiratory motion for four-dimentional radiotherapy," Med. Phys. **31**, 2274–2283 (2004).

[5]M. Engelsman, G. C. Sharp, T. Bortfeld, R. Onimaru, and H. Shirato, "How much margin reduction is possible through gating on breath hold?," Phys. Med. Biol. **50**, 477–490 (2005).

[6]J. W. Wolthaus, C. Schneider, J. J. Sonke, M. van Herk, J. S. Belderbos, M. M. Rossi, J. V. Lebesque, and E. M. Damen, "Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients," Int. J. Radiat. Oncol., Biol., Phys. **65**, 1560–1571 (2006).

[7]G. D. Hugo, D. Yan, and J. Liang, "Population and patient-specific target margins for 4D adaptive radiotherapy to account for intra- and inter-fraction variation in lung tumour position," Phys. Med. Biol. **52**, 257–274 (2007).

[8]A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 476–480 (1999).

[9]Y. N. Rao, J. C. Principe, and T. F. Wong, "Fast RLS-like algorithm for generalized eigendecomposition and its applications," J. VLSI Signal Proc. Syst. Signal, Image, Video Technol. **37**, 333–344 (2004).

[10]D. Ruan and J. A. Fessler, "Adaptive ellipse tracking and a convergence proof," Technical Report No. 382, Comm. and Sign. Proc. Lab., Dept. of EECS, University of Michigan, Ann Arbor, MI, 48109-2122, May 2007.

[11]G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. (Johns Hopkins University Press, 1989).

[12]D. Ruan, J. A. Fessler, J. M. Balter, and J.-J. Sonke, "Exploring breathing pattern irregularity with projection-based method," Med. Phys. **33**, 2491–2499 (2006).

[13]A. E. Lujan, J. M. Balter, and R. K. T. Haken, "A method for incorporating organ motion due to breathing into 3D dose calculations in the liver: Sensitivity to variations in motion," Med. Phys. **30**, 2643–2649 (2003).

[14]Y. Seppenwoolde, H. Shirato, K. Kitamura, S. Shimizu, M. Herk, J. V. Lebesque, and K. Miyasaka, "Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **53**, 822–834 (2002).