

The Kriging Update Model and Recursive Space-Time Function Estimation

William S. Kerwin, *Student Member, IEEE*, and Jerry L. Prince, *Senior Member, IEEE*

Abstract—We present a method for efficiently fitting a time series of spatial functions to observed data. The method is closely related to kriging, which is an interpolation method based on a stochastic data model. While kriging is effective and versatile for estimating individual functions from observed data, it must be extended to incorporate temporal correlation. In this paper, we introduce temporal correlation to kriging in the form of a stochastic state equation representation—the *kriging update model*. This permits a recursive solution that is akin to Kalman filtering to estimate time series of functions that avoids growing data problems associated with other space-time extensions of kriging. The state equation representation incorporates the principle assumption of universal kriging: that the mean is deterministic but unknown. We derive the estimate using best linear unbiased estimation and state the result in a concise algorithm for general use in arbitrary spatial dimensions. To demonstrate the algorithm, we apply it to two sets of test functions and provide an example application in estimating heart motion from medical images.

I. INTRODUCTION

MANY physical problems require the estimation of a function that varies in both time and space. Often, the function is observed at a sequence of times with a finite number of sensors in scattered spatial locations. The goal is then to reconstruct the complete spatial variation of the function at each of the observation times, in essence creating a time sequence of spatial estimates. Applications include determining the instantaneous flow profile of a fluid [1], reconstructing the configuration of a deforming surface [2], or mapping the atmospheric distribution of pollutants [3] over time.

Several methods for generating function estimates from space-time data have been proposed within the general area of kriging [4]. In kriging, a spatial function is modeled stochastically and then estimated using best linear unbiased estimation (BLUE)¹ (see [6]). In the most general case of universal kriging, the stochastic model assumes the function

has a known covariance but an unknown mean of the form

$$\sum_{i=1}^r d_i f_i(x) \quad (1)$$

where $f_1(x), \dots, f_r(x)$ are *known* functions of the multidimensional spatial variable x , and d_1, \dots, d_r are *unknown, deterministic* coefficients. Applying BLUE to this model results in a closed-form expression that estimates the function throughout space. When estimating a time series of functions from time series data, kriging can be used to individually estimate each function in the series from its corresponding set of observations. Alternatively, better performance may be expected from extensions of spatial kriging that utilize the temporal correlation between functions. In space-time kriging, the temporal dimension is treated as an additional spatial dimension [7], and in cokriging, the functions are treated as distinct but correlated entities with known cross-covariances [8]. However, both of these methods require the solution of a large system of linear equations that grows without bound over time.

To address this growing data problem and still utilize temporal correlation, we propose coupling the kriging model to a stochastic state model and applying computationally efficient Kalman filtering methods. Such an approach was previously proposed for the special case of zero-mean functions [9]—as in simple kriging—essentially extending the standard vector Kalman filter to address a space-time problem. More general solutions have been proposed [10], [11] for unknown means characterized by (1), where the coefficients d_1, \dots, d_r themselves obey a known stochastic model, thus extending methods for vector Kalman filtering in the presence of random bias [12]–[14]. However, neither of these approaches captures two great strengths of universal kriging. First, universal kriging requires no prior assumptions regarding the statistics of d_1, \dots, d_r , thereby avoiding bias associated with incorrect statistical assumptions. Second, the unconstrained coefficients of universal kriging are well-suited for modeling systems that undergo sudden jumps, which is a common natural occurrence.

Recognizing the need for a method that captures the strengths of universal kriging, we propose a stochastic state equation representation—we call it the *kriging update model*—that preserves the characteristic, unknown mean of universal kriging. We then apply BLUE to reconstruct the sequence of functions without statistical characterization of the mean coefficients. Finally, by paralleling the derivation of Kalman filtering, we derive a recursive BLUE formulation that incorporates temporal correlation while maintaining a

Manuscript received October 27, 1998; revised May 6, 1999. The associate editor coordinating the review of this paper and approving it for publication was Prof. Chi Chung Ko.

The authors are with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: jprince@jhu.edu).

Publisher Item Identifier S 1053-587X(99)08316-6.

¹In the kriging literature, the name best linear unbiased *prediction* (BLUP) is more commonly used. We use “estimation” in place of “prediction” to be consistent with the Kalman filtering literature, where prediction implies temporal forecasting (see [5] for a discussion on this subject).

computational burden comparable with individually kriging each function in the series. In essence, the algorithm is the equivalent of the unbiased vector Kalman filter proposed by Kitanidis [15] extended to the spatial domain of kriging theory. We therefore call the algorithm the *kriging filter*.

The correspondence is organized as follows. In Section II, we present the kriging update model. For this model, we define, in Section III, two problems in best linear unbiased estimation: 1) estimating the sequence of functions when the unknown mean is attributable to sensor bias and must be removed and 2) estimating the sequence of functions with the unknown mean included. We also characterize the solutions with a theorem we call projected orthogonality. Based on this theorem, we derive, in Sections IV and V, recursive solutions for both estimation problems. Then, in Section VI, we repackage the estimation equations into the kriging filter algorithm for computing either estimate. Finally, we conclude the paper with some numerical examples in Section VII, an application of the algorithm in Section VIII, and a discussion of the method in Section IX.

II. DATA MODEL

1) *Kriging Update Model*: Let $z_1(x), z_2(x), \dots$ be a time series of functions, where x is a spatial variable of any dimension, and the subscript is a discrete time index. We assume that each function in the series is governed by the state equation representation

$$z_n(x) = z_{n-1}(x) + u_n(x) + \mathbf{f}^T(x)\mathbf{d}_n \quad (2)$$

where

$$\mathbf{f}(x) = [f_1(x) \quad \dots \quad f_r(x)]^T \quad (3)$$

and

$$\mathbf{d}_n = [d_{n1} \quad \dots \quad d_{nr}]^T.$$

In these definitions, $u_n(x)$ is a zero-mean stochastic process with known covariance function

$$E[u_n(x_a)u_n(x_b)] = k(x_a, x_b) \quad (4)$$

$f_1(x), \dots, f_r(x)$ are r known *drift functions*, and d_{n1}, \dots, d_{nr} are unknown *drift coefficients*. With these definitions, the update $u_n(x) + \mathbf{f}^T(x)\mathbf{d}_n$ is described by the stochastic model used in universal kriging—note that $\mathbf{f}^T(x)\mathbf{d}_n$ is the vector form of (1)—hence the name kriging update model. The notation we use is similar to that used by David [16] to explain kriging.

In working with this model, it is useful to separate the individual components of $z_n(x)$. Specifically, we define

$$c_n(x) = c_{n-1}(x) + u_n(x) \quad (5)$$

and

$$\mathbf{m}_n = \mathbf{m}_{n-1} + \mathbf{d}_n. \quad (6)$$

Using these definitions, (2) can be written

$$z_n(x) = c_n(x) + \mathbf{f}^T(x)\mathbf{m}_n. \quad (7)$$

The first component $c_n(x)$ is the zero-mean, stochastic portion of $z_n(x)$. The second component \mathbf{m}_n provides the mean

$$E[z_n(x)] = \mathbf{f}^T(x)\mathbf{m}_n. \quad (8)$$

The key characteristic of this model is that the mean (8) is unknown, as in universal kriging.

Because the kriging update model extends a spatial model to a spatio-temporal one, some additional temporal assumptions are necessary. We assume that neither $k(x_a, x_b)$ nor $\mathbf{f}(x)$ varies with n . We also assume that $u_n(x)$ is white in time so that

$$E[u_n(x_a), u_m(x_b)] = 0, \quad m \neq n.$$

Finally, as an initial condition, we assume that $c_0(x)$ is a zero-mean process with a covariance function

$$E[c_0(x_a)c_0(x_b)] = \alpha k(x_a, x_b) \quad (9)$$

for some known parameter α and that $c_0(x)$ is uncorrelated with $u_n(x)$ for all n . More general initial conditions will be discussed briefly at the end of this paper. Note that no initial conditions are required for \mathbf{m}_0 . The fact that stochastic characterization of \mathbf{m}_0 is not necessary is one of the strengths of this approach.

2) *Observation Model*: Observations of each function in the series are obtained at p spatial positions x_1, \dots, x_p ($p \gg r$), which may be arranged in any pattern. For now, we assume these positions are fixed for all observation times; in Section IX, we will briefly comment on how changing observation positions affect the algorithm. The observed values are placed into the $p \times 1$ vectors

$$\mathbf{y}_n = \mathbf{z}_n + \mathbf{v}_n, \quad n = 1, 2, \dots \quad (10)$$

where

$$\mathbf{z}_n = [z_n(x_1) \quad \dots \quad z_n(x_p)]^T$$

are the true function values at these positions, and \mathbf{v}_n is a $p \times 1$ vector of observation noise. We assume \mathbf{v}_n is zero-mean, has a known covariance matrix

$$E[\mathbf{v}_n \mathbf{v}_n^T] = \Sigma_n \quad (11)$$

and is uncorrelated with all past noise and data. Finally, we note that the mean of the data vector is given by

$$E[\mathbf{y}_n] = F^T \mathbf{m}_n \quad (12)$$

where

$$F = [\mathbf{f}(x_1) \quad \dots \quad \mathbf{f}(x_p)]. \quad (13)$$

We assume that the $r \times p$ matrix F has rank r , which amounts to a mild constraint on the choice of drift functions and the distribution of sample points. For example, in two dimensions, the samples typically cannot be arrayed along a straight line.

III. BEST LINEAR UNBIASED ESTIMATION

Given the model and the data vectors $\mathbf{y}_1, \mathbf{y}_2, \dots$, we consider two estimation problems. In one, we estimate the function $z_n(x)$, and in the other, we estimate only the zero-mean portion $c_n(x)$ —desirable when $\mathbf{f}^T(x)\mathbf{m}_n$ is a corrupting bias in (7) that must be eliminated. In either case, we assume that only past and present data vectors are available for the estimate. The primary challenge in estimating $z_n(x)$ or $c_n(x)$ is that the means of the function $z_n(x)$ and the data vector \mathbf{y}_n are unknown. Therefore, we are unable to use linear minimum square error (LMMSE) estimation; instead, we use BLUE (see [5] for an overview of BLUE).

The linearity of BLUE means that the estimate of $z_n(x)$ is given by

$$\hat{z}_n(x) = \sum_{i=1}^n \mathbf{a}_i^T(x) \mathbf{y}_i \quad (14)$$

for some set of spatially varying $p \times 1$ weighting vectors $\mathbf{a}_1(x), \dots, \mathbf{a}_n(x)$. Unbiasedness implies that the estimate must satisfy $E[\hat{z}_n(x)] = E[z_n(x)]$, which reduces to the requirement

$$\sum_{i=1}^n \mathbf{a}_i^T(x) F^T \mathbf{m}_i = \mathbf{f}^T(x) \mathbf{m}_n.$$

To derive this requirement, we evaluated the expectation of (14) using (12) and set the result equal to (8). Because this must hold for any unknown sequence $\mathbf{m}_1, \dots, \mathbf{m}_n$, we arrive at the unbiasedness constraint on the weighting vectors

$$F \mathbf{a}_i(x) = \begin{cases} \mathbf{0}, & i = 1, \dots, n-1 \\ \mathbf{f}(x), & i = n \end{cases} \quad (15)$$

Finally, the best estimate will be that which minimizes the mean square error $E[(z_n(x) - \hat{z}_n(x))^2]$ subject to the unbiasedness constraint.

Likewise, the estimate of $c_n(x)$ will be given by a linear combination

$$\hat{c}_n(x) = \sum_{i=1}^n \mathbf{a}_i^T(x) \mathbf{y}_i \quad (16)$$

for some different weighting vectors $\mathbf{a}_1(x), \dots, \mathbf{a}_n(x)$. To be unbiased, this estimate must satisfy $E[\hat{c}_n(x)] = 0$, which leads to the constraint

$$F \mathbf{a}_i(x) = \mathbf{0} \quad i = 1, \dots, n. \quad (17)$$

The best estimate $\hat{c}_n(x)$ then minimizes the mean squared error $E[(c_n(x) - \hat{c}_n(x))^2]$ subject to this constraint.

Given these definitions and the assumed statistical characteristics stated in Section II, the weights $\mathbf{a}_1(x), \dots, \mathbf{a}_n(x)$ corresponding to $\hat{z}_n(x)$ or $\hat{c}_n(x)$ may easily be derived using the method of Lagrange multipliers for constrained minimization. The derivation for $\hat{z}_n(x)$, in fact, results precisely in the cokriging equations (see [17]) for our kriging update model. Our reason for defining the model, however, was to avoid computationally intensive cokriging. We will therefore develop computationally efficient, recursive expressions for these estimates.

To derive the recursive estimators and prove they result in the identical optimal estimates defined here, we will make extensive use of a theorem we call *projected orthogonality*. This theorem is related to several classical results in estimation theory, with modifications to address the specific characteristics of the kriging update model. For purposes of generality, suppose we are estimating some value ψ —which might be $z_n(x)$ or $c_n(x)$ —with the linear combination

$$\hat{\psi} = \sum_{i=1}^n \mathbf{a}_i^T \mathbf{y}_i.$$

Suppose, in addition, that the weights in this expression are subject to the constraints

$$F \mathbf{a}_i = \mathbf{f}_i, \quad i = 1, \dots, n$$

which generalizes both (15) and (17). Then, the projected orthogonality theorem states that the error $(\psi - \hat{\psi})$ is statistically *orthogonal* to the *projection* of \mathbf{y}_i onto the null space of the matrix F . Formally, this statement is as follows.

Theorem 1—Projected Orthogonality: Let the vectors $\mathbf{y}_1, \dots, \mathbf{y}_n$ have finite covariances. If the vectors $\check{\mathbf{a}}_1, \dots, \check{\mathbf{a}}_n$ solve

$$\min_{\mathbf{a}_1, \dots, \mathbf{a}_n} E \left[\left(\psi - \sum_{i=1}^n \mathbf{a}_i^T \mathbf{y}_i \right)^2 \right] \quad \text{subject to } F \mathbf{a}_i = \mathbf{f}_i \quad i = 1, \dots, n \quad (18)$$

then

$$E \left[\left(\psi - \sum_{i=1}^n \check{\mathbf{a}}_i^T \mathbf{y}_i \right) \mathbf{b}^T \mathbf{y}_l \right] = 0 \quad (19)$$

for every \mathbf{b} satisfying $F \mathbf{b} = \mathbf{0}$ and every $l \leq n$.

Proof: We prove the necessity of (19) by contradiction. Assume $\check{\mathbf{a}}_i, i = 1, \dots, n$ solve (18) but do not satisfy (19). Then, for some $l \leq n$ and \mathbf{b} satisfying $F \mathbf{b} = \mathbf{0}$

$$E \left[\left(\psi - \sum_{i=1}^n \check{\mathbf{a}}_i^T \mathbf{y}_i \right) \mathbf{b}^T \mathbf{y}_l \right] = \beta$$

where $\beta \neq 0$. Now, consider

$$\mathbf{a}_i = \begin{cases} \check{\mathbf{a}}_i, & i \neq l \\ \check{\mathbf{a}}_l + \frac{\beta}{\eta^2} \mathbf{b}, & i = l \end{cases}$$

where $\eta^2 = E[(\mathbf{b}^T \mathbf{y}_l)^2]$. We find

$$\begin{aligned} E \left[\left(\psi - \sum_{i=1}^n \mathbf{a}_i^T \mathbf{y}_i \right)^2 \right] &= E \left[\left(\psi - \sum_{i=1}^n \check{\mathbf{a}}_i^T \mathbf{y}_i \right)^2 \right] - \frac{\beta^2}{\eta^2} \\ &< E \left[\left(\psi - \sum_{i=1}^n \check{\mathbf{a}}_i^T \mathbf{y}_i \right)^2 \right] \end{aligned}$$

where the finite covariance of \mathbf{y}_l implies $\frac{\beta^2}{\eta^2} > 0$, which yields the strict inequality. This inequality contradicts our initial assumption that $\check{\mathbf{a}}_1, \dots, \check{\mathbf{a}}_n$ solve (18); each new vector \mathbf{a}_i satisfies $F \mathbf{a}_i = \mathbf{f}_i$, and together, they produce a smaller mean square error in (18). Therefore, the necessity of (19) is proven. \square

A similar proof also shows that (19) is sufficient to ensure that $\hat{\psi}$ is the optimal estimate, but this fact is not necessary in the following derivations.

IV. RECURSIVE ESTIMATOR FOR $c_n(x)$

Turning to the derivations, we first consider the case where $\mathbf{f}^T(x)\mathbf{m}_n$ is a corrupting bias and find the estimator $\hat{c}_n(x)$. The importance of the projected orthogonality theorem is immediately apparent in that it permits us to prove the following proposition, which puts the estimator into a recursive framework.

Proposition 2: The best linear unbiased estimate of $c_n(x)$ given $\mathbf{y}_1, \dots, \mathbf{y}_n$ can be put into the form

$$\hat{c}_n(x) = \hat{c}_{n-1}(x) + \mathbf{g}_n^T(x)\tilde{\mathbf{y}}_n \quad (20)$$

where

$$\begin{aligned} \tilde{\mathbf{y}}_n &= \mathbf{y}_n - \hat{\mathbf{c}}_{n-1} \\ \hat{\mathbf{c}}_{n-1} &= [\hat{c}_{n-1}(x_1) \quad \dots \quad \hat{c}_{n-1}(x_p)]^T \end{aligned} \quad (21)$$

and the $p \times 1$ weight vector $\mathbf{g}_n(x)$ satisfies the constraint

$$F\mathbf{g}_n(x) = \mathbf{0}.$$

Proof: the proof is given in Appendix A. \square

We will refer to (21) as the *innovations sequence* for the data because of its strong similarity to the innovations sequence in Kalman filtering theory. We note that unlike in Kalman filtering, $\tilde{\mathbf{y}}_n$ is not a white sequence, which makes its utilization different, and the proof of Proposition 2 nontrivial.

As a result of Proposition 2, the problem of finding $\hat{c}_n(x)$ is reduced to finding a single weight vector $\mathbf{g}_n(x)$ rather than the n weight vectors needed in (16). To find $\mathbf{g}_n(x)$, we write the mean square error of the estimator as

$$\sigma_e^2 = E[(c_n(x) - \hat{c}_{n-1}(x) - \mathbf{g}_n^T(x)\tilde{\mathbf{y}}_n)^2].$$

Expanding $c_n(x)$ and $\tilde{\mathbf{y}}_n$ gives

$$\begin{aligned} \sigma_e^2 &= E[(u_n(x) + c_{n-1}(x) - \hat{c}_{n-1}(x) \\ &\quad - \mathbf{g}_n^T(x)(\mathbf{u}_n + \mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1} + F^T\mathbf{m}_n + \mathbf{v}_n))^2] \end{aligned} \quad (22)$$

where

$$\mathbf{u}_n = [u_n(x_1) \quad \dots \quad u_n(x_p)]^T$$

and

$$\mathbf{c}_{n-1} = [c_{n-1}(x_1) \quad \dots \quad c_{n-1}(x_p)]^T.$$

Squaring the argument and applying the expectation yields

$$\begin{aligned} \sigma_e^2 &= \sigma_u^2(x) + \sigma_c^2(x) + \mathbf{g}_n^T(x)(H_{n-1} + K + \Sigma_n)\mathbf{g}_n(x) \\ &\quad - 2\mathbf{g}_n^T(x)(\mathbf{k}(x) + \mathbf{h}_{n-1}(x)) \end{aligned} \quad (23)$$

where

$$\sigma_u^2(x) = E[u_n^2(x)] \quad (24)$$

$$\sigma_c^2(x) = E[(c_{n-1}(x) - \hat{c}_{n-1}(x))^2] \quad (25)$$

$$H_{n-1} = E[(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1})(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1})^T] \quad (26)$$

$$\mathbf{h}_{n-1}(x) = E[(c_{n-1}(x) - \hat{c}_{n-1}(x))(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1})] \quad (27)$$

$$K = E[\mathbf{u}_n\mathbf{u}_n^T] \quad (28)$$

$$\mathbf{k}(x) = E[u_n(x)\mathbf{u}_n] \quad (29)$$

and Σ_n is given by (11). In deriving (23), we have used the facts that $u_n(x)$ and \mathbf{v}_n are uncorrelated with past data and with each other. We also used the constraint that $F\mathbf{g}_n(x) = \mathbf{0}$ to remove the term $F^T\mathbf{m}_n$ from the expression.

The desired solution is obtained by minimizing (23) subject to $F\mathbf{g}_n(x) = \mathbf{0}$. To perform this minimization, we use the method of Lagrange multipliers, obtaining the augmented function

$$\xi(x) = \sigma_e^2(x) + \boldsymbol{\lambda}^T F\mathbf{g}_n(x)$$

where $\boldsymbol{\lambda}$ is a vector of r Lagrange multipliers. Setting the derivatives of this function with respect to $\mathbf{g}_n(x)$ and $\boldsymbol{\lambda}$ equal to zero and solving for $\mathbf{g}_n(x)$ yields the desired minimizing weight vector

$$\mathbf{g}_n(x) = (I - M_n^T F)L_n(\mathbf{k}(x) + \mathbf{h}_{n-1}(x)) \quad (30)$$

where

$$L_n = (H_{n-1} + K + \Sigma_n)^{-1} \quad (31)$$

and

$$M_n = (FL_n F^T)^{-1} FL_n. \quad (32)$$

Both (31) and (32) involve matrix inversions. To guarantee the inverses exist, we require $u_n(x_1), \dots, u_n(x_p)$ to be linearly independent so that K is positive definite. This implies that L_n is positive definite because both H_{n-1} and Σ_n are positive semi-definite covariance matrices. Then, M_n also exists because we have assumed F has rank r (see Section II).

In the solution for $\mathbf{g}_n(x)$, the noise covariance Σ_n and the update covariance $E[u_n(x_a)u_n(x_b)]$ are known (by assumption) so that K and $\mathbf{k}(x)$ are also known. However, we do not have expressions for the error covariance terms H_{n-1} or $\mathbf{h}_{n-1}(x)$. To address the unknown H_{n-1} term, we obtain a recursive expression by first evaluating the estimate (20) at x_1, \dots, x_p to obtain

$$\begin{aligned} \hat{\mathbf{c}}_n &= \hat{\mathbf{c}}_{n-1} + [\mathbf{k}(x_1) \quad \dots \quad \mathbf{k}(x_p)]^T K^{-1} G_n \tilde{\mathbf{y}}_n \\ &= \hat{\mathbf{c}}_{n-1} + G_n \tilde{\mathbf{y}}_n \end{aligned} \quad (33)$$

where

$$G_n = (K + H_{n-1})L_n(I - F^T M_n). \quad (34)$$

Then, we evaluate $E[(\mathbf{c}_n - \hat{\mathbf{c}}_n)(\mathbf{c}_n - \hat{\mathbf{c}}_n)^T]$, use the projected orthogonality theorem to simplify, and obtain the recursive expression (see Appendix B)

$$H_n = (K + H_{n-1})(I - G_n^T). \quad (35)$$

The initial condition for this expression is $H_0 = \alpha K$ as a result of our assumed initial covariance function (9).

Obtaining an expression for the only remaining unknown quantity $\mathbf{h}_{n-1}(x)$ is somewhat more challenging. Specifically, the spatial aspect of this problem precludes a useful recursive expression because the continuous function $\mathbf{h}_{n-1}(x)$ cannot be stored for every x . To address this problem, in Appendix C, we use the projected orthogonality theorem to derive the following relationship between $\mathbf{h}_{n-1}(x)$ and the known function $\mathbf{k}(x)$:

$$\mathbf{h}_{n-1}(x) = H_{n-1} K^{-1} \mathbf{k}(x). \quad (36)$$

This yields the solution

$$\mathbf{g}_n(x) = G_n^T K^{-1} \mathbf{k}(x). \quad (37)$$

Although this provides the final piece necessary to compute $\hat{c}_n(x)$, the form of (20) is impractical for implementation because the continuous function $\hat{c}_{n-1}(x)$ also cannot be stored for every x . We overcome this problem by rearranging the equations as follows. First, we observe that (20) can also be written

$$\hat{c}_n(x) = \sum_{i=1}^n \mathbf{g}_i^T(x) \tilde{\mathbf{y}}_i.$$

Using (37), we find

$$\hat{c}_n(x) = \mathbf{k}^T(x) \mathbf{w}_n \quad (38)$$

where

$$\begin{aligned} \mathbf{w}_n &= \sum_{i=1}^n K^{-1} G_i \tilde{\mathbf{y}}_i \\ &= \mathbf{w}_{n-1} + K^{-1} G_n \tilde{\mathbf{y}}_n. \end{aligned}$$

We initialize the recursive form with $\mathbf{w}_0 = \mathbf{0}$. Then, using the facts that $\tilde{\mathbf{y}}_n = \mathbf{y}_n - \hat{\mathbf{c}}_{n-1}$ and $\hat{\mathbf{c}}_{n-1} = K \mathbf{w}_{n-1}$, we find

$$\mathbf{w}_n = \mathbf{w}_{n-1} + K^{-1} G_n (\mathbf{y}_n - K \mathbf{w}_{n-1}). \quad (39)$$

This results in an update equation (39) that does not depend on x ; therefore, it can be stored at each iteration. Then, (38) is evaluated without requiring a summation. The “gain” matrix G_n is found by alternating (34) and (35) in a manner analogous to Kalman filtering.

V. RECURSIVE ESTIMATOR FOR $z_n(x)$

A recursive estimator for $z_n(x)$ can readily be derived by building on the derivation of $\hat{c}_n(x)$. In particular, the following proposition shows a powerful relationship between $\hat{z}_n(x)$ and $\hat{c}_n(x)$.

Proposition 3: The best linear unbiased estimate of $z_n(x)$ can be put into the form

$$\hat{z}_n(x) = \hat{c}_{n-1}(x) + \mathbf{q}_n^T(x) \tilde{\mathbf{y}}_n \quad (40)$$

where

$$F \mathbf{q}_n(x) = \mathbf{f}(x)$$

and $\tilde{\mathbf{y}}_n$ is the innovation defined in (21).

Proof: The proof is given in Appendix A. \square

Thus, estimating $z_n(x)$ can be accomplished by first estimating the sequence of functions $c_1(x), c_2(x), \dots$ and then calculating the one coefficient vector $\mathbf{q}_n(x)$ in (40). To calculate $\mathbf{q}_n(x)$, we write the mean square error as

$$\sigma_e^2 = E[(z_n(x) - \hat{c}_{n-1}(x) - \mathbf{q}_n^T(x) \tilde{\mathbf{y}}_n)^2].$$

Expanding $z_n(x)$ and $\tilde{\mathbf{y}}_n$ yields

$$\begin{aligned} \sigma_e^2 &= E[(u_n(x) + c_{n-1}(x) + \mathbf{f}^T(x) \mathbf{m}_n - \hat{c}_{n-1}(x) \\ &\quad - \mathbf{q}_n^T(x)(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1} + \mathbf{u}_n + \mathbf{v}_n))^2]. \end{aligned}$$

Then, because we must obey the constraint $F \mathbf{q}_n(x) = \mathbf{f}(x)$, the terms involving \mathbf{m}_n cancel, leaving

$$\begin{aligned} \sigma_e^2 &= E[(u_n(x) + c_{n-1}(x) - \hat{c}_{n-1}(x) \\ &\quad - \mathbf{q}_n^T(x)(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1} + \mathbf{u}_n + \mathbf{v}_n))^2]. \end{aligned}$$

Comparison of this expression to (22) shows them to be identical, except that $\mathbf{g}_n(x)$ has been replaced by $\mathbf{q}_n(x)$. Thus, we will be minimizing precisely the same expression as in (23); however, the minimization is subject to a different constraint, namely, we must minimize

$$\begin{aligned} \sigma_e^2 &= \sigma_u^2(x) + \sigma_c^2(x) + \mathbf{q}_n^T(x)(H_{n-1} + K + \Sigma_n) \mathbf{q}_n(x) \\ &\quad - 2 \mathbf{q}_n^T(x)(\mathbf{k}(x) + \mathbf{h}_{n-1}(x)) \end{aligned}$$

subject to $F \mathbf{q}_n(x) = \mathbf{f}(x)$, where the terms in the mean square error expression are defined in (24)–(29).

The same Lagrange multiplier approach used to find $\mathbf{g}_n(x)$ can be used to find $\mathbf{q}_n(x)$. In this case, the augmented function is

$$\xi(x) = \sigma_e^2(x) + \lambda^T (F \mathbf{q}_n(x) - \mathbf{f}(x))$$

which is minimized by

$$\mathbf{q}_n(x) = (I - M_n^T F) L_n (\mathbf{k}(x) + \mathbf{h}_{n-1}(x)) + M_n^T \mathbf{f}(x)$$

where L_n and M_n are defined in Section IV by (31) and (32), respectively. Furthermore, using (30), we can write the solution as

$$\mathbf{q}_n(x) = \mathbf{g}_n(x) + M_n^T \mathbf{f}(x) \quad (41)$$

where $\mathbf{g}_n(x)$ is given by (37). Then, using this expression in (40), we find

$$\begin{aligned} \hat{z}_n(x) &= \hat{c}_{n-1}(x) + \mathbf{g}_n^T(x) \tilde{\mathbf{y}}_n + \mathbf{f}^T(x) M_n \tilde{\mathbf{y}}_n \\ &= \hat{c}_n(x) + \mathbf{f}^T(x) M_n (\mathbf{y}_n - K \mathbf{w}_{n-1}). \end{aligned}$$

Finally, we define

$$\hat{\mathbf{m}}_n = M_n (\mathbf{y}_n - K \mathbf{w}_{n-1}) \quad (42)$$

and use (38) to write

$$\hat{z}_n(x) = \mathbf{w}_n^T \mathbf{k}(x) + \hat{\mathbf{m}}_n^T \mathbf{f}(x). \quad (43)$$

We use the notation $\hat{\mathbf{m}}_n$ because that quantity can be shown to be the best linear unbiased estimate of \mathbf{m}_n .

VI. RECURSIVE ESTIMATOR IN ALGORITHMIC FORM

For easy reference, in this section, we package the important equations from Sections IV and V into a step-by-step algorithm—the kriging filter—for computing the estimates. As a preliminary step, we reiterate the important definitions. First, we must specify the noise covariance Σ_n given by (11). Second, we must specify the set of drift functions that define $\mathbf{f}(x)$ and F in (3) and (13). Third, we must specify $k(x_a, x_b)$ [see (4)] to determine K and $\mathbf{k}(x)$ in (28) and (29). Specifically

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_p) \\ \vdots & \ddots & \vdots \\ k(x_p, x_1) & \cdots & k(x_p, x_p) \end{bmatrix}$$

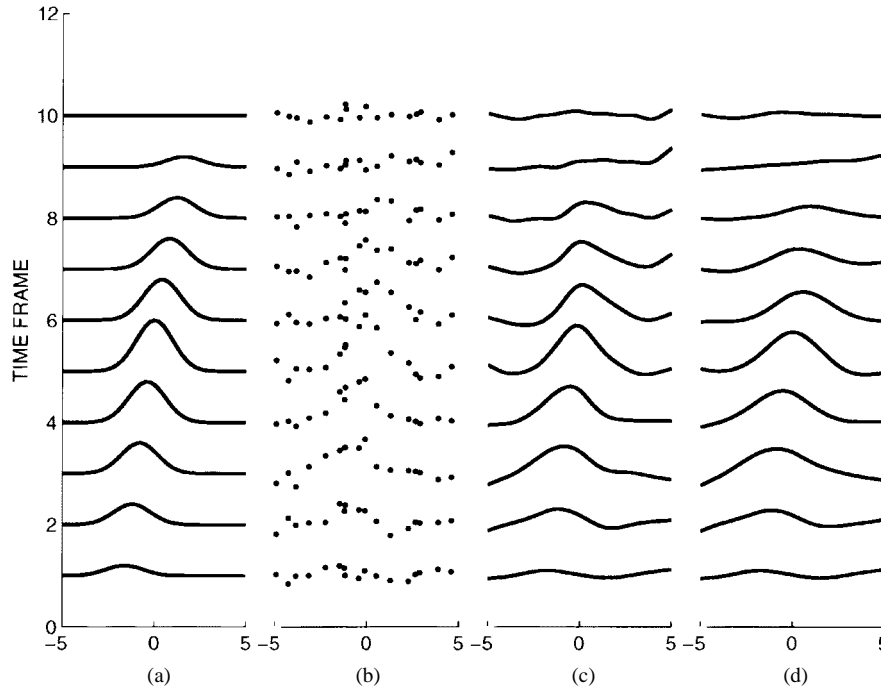


Fig. 1. Algorithm applied to a one-dimensional sequence of ten functions. (a) Actual functions. (b) Noisy samples of the functions. (c) Estimates of the functions using our algorithm. (d) Estimates of the functions using kriging.

and

$$\mathbf{k}(x) = [k(x, x_1) \quad \cdots \quad k(x, x_p)]^T.$$

Finally, we must appropriately choose α in (9), which defines the initial covariance. Then, the estimation algorithm is as follows.

Algorithm 1

The Kriging Filter

- 1) Initialize $H_0 = \alpha K$ and $\mathbf{w}_0 = \mathbf{0}$ ($p \times 1$). Set $n = 1$.
- 2) Calculate the matrices

$$\begin{aligned} L_n &= (K + \Sigma_n + H_{n-1})^{-1} \\ M_n &= (FL_n F^T)^{-1} FL_n \\ G_n &= (K + H_{n-1})L_n(I - F^T M_n) \\ H_n &= (K + H_{n-1})(I - G_n^T). \end{aligned}$$

- 3) Calculate the coefficient vectors

$$\begin{aligned} \hat{\mathbf{m}}_n &= M_n(\mathbf{y}_n - K\mathbf{w}_{n-1}) \\ \mathbf{w}_n &= \mathbf{w}_{n-1} + K^{-1}G_n(\mathbf{y}_n - K\mathbf{w}_{n-1}). \end{aligned}$$

- 4) Calculate the desired estimate using

$$\hat{z}_n(x) = \mathbf{k}^T(x)\mathbf{w}_n + \mathbf{f}^T(x)\hat{\mathbf{m}}_n$$

or

$$\hat{c}_n(x) = \mathbf{k}^T(x)\mathbf{w}_n.$$

- 5) Increment n , and go to 2.

VII. EXAMPLES

We demonstrate the effectiveness of the kriging filter with two simple test cases: one in one dimension and the other in two dimensions. In both cases, we reconstruct known analytical functions from observations corrupted by independent Gaussian noise. We also compare the results of the kriging filter to other function reconstruction methods.

The first test function we chose is

$$z_n(x) = (1 - |\alpha_n|)e^{-\frac{1}{2}(x - \alpha_n)^2}$$

where $\alpha_n = n/5 - 1$. Each function in the series is a Gaussian curve centered at α_n with a peak value of $1 - |\alpha_n|$. The first ten functions in the series are depicted in Fig. 1(a) and arranged from bottom to top. Over the ten time frames, the Gaussian curve grows to a height of 1 before shrinking back to 0. In addition, the center of the curve shifts from left to right. We randomly chose 17 points in the interval $x \in (-5, 5)$ at which to observe the functions. Each observation included independent, zero-mean Gaussian noise with a standard deviation of 0.1. The noisy samples are shown in Fig. 1(b) for all ten time frames.

We applied the kriging filter with model parameters $k(x_a, x_b) = |x_a - x_b|^3$ (a generalized covariance function [18]), $\mathbf{f}(x) = [1 \ x]^T$, $\Sigma_n = 10I$, and $\alpha = 0$. The resulting function estimates are depicted in Fig. 1(c). The same parameter choices may be used in kriging, where the noise variance models the so-called nugget effect. The results from individually kriging each function in the series are shown in Fig. 1(d). Note that kriging dramatically broadens the peak, which is a phenomenon attributable to ignoring temporal information. In addition, note that in the first time frame, our estimate and the kriging estimate are identical, indicating

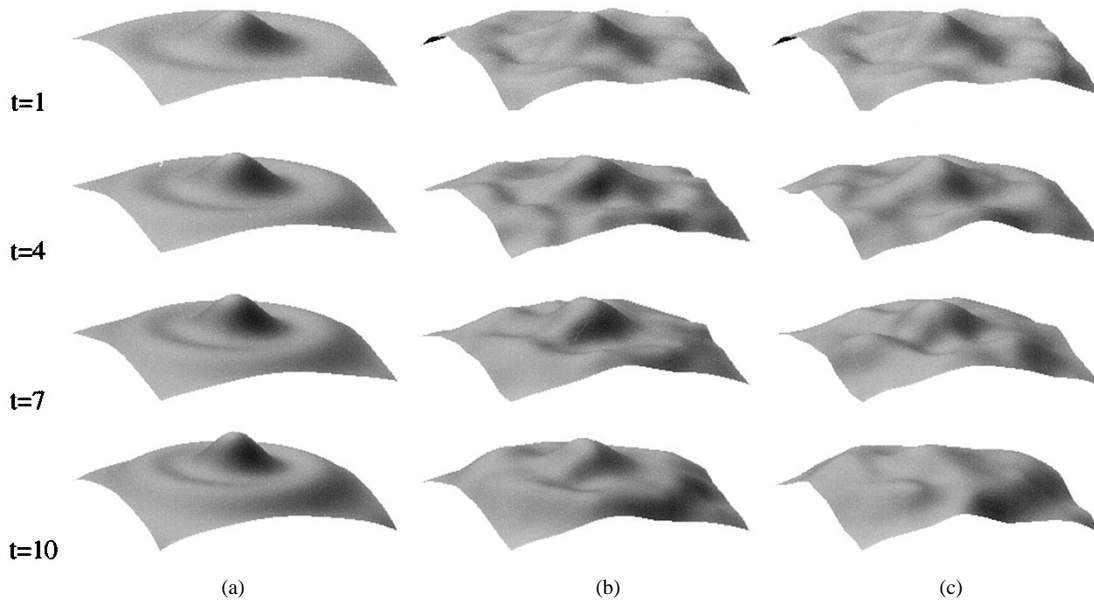


Fig. 2. Algorithm applied to a 2-D series of functions (times 1, 4, 7, and 10 are shown). (a) Original functions. (b) Functions resulting from our algorithm. (c) Smoothing thin-plate spline functions.

that with only one time frame available, our method simply performs kriging.

Next, we applied the kriging filter to a series of two-dimensional (2-D) functions. In 2-D, each function to be estimated is expressed as $z_n(x, y)$, where x and y are the two spatial variables. Several members of the chosen test series are depicted in Fig. 2(a), each of which consists of a sinc function added to a half sphere. Over time, the radius of the sphere decreases and the width of the sinc function contracts. Each function was sampled over a regular 11×11 grid, and to each sample, independent Gaussian noise was added, with a standard deviation equal to 0.25 times the peak height of the sinc function.

To estimate the original functions from the noisy samples, we applied the kriging filter with data covariance function

$$k(x_a, y_a; x_b, y_b) = \|(x_a, y_a) - (x_b, y_b)\|^2 \log \|(x_a, y_a) - (x_b, y_b)\|$$

and drift functions

$$\mathbf{f}(x, y) = [1 \quad x \quad y]^T.$$

We also chose $\Sigma_n = \sigma^2 I$ and $\alpha = 20$. The results are shown in Fig. 2(b) for $\sigma^2 = 100$.

These parameter choices relate our method to the thin-plate spline, which is a special case of kriging [19]. In fact, the above choices for K , $\mathbf{k}(x)$, and $\mathbf{f}(x)$, when used in kriging, result in a 2-D thin-plate spline. We therefore compare the results of the kriging filter with those obtained using thin-plate spline methods. For fairer comparison, we use the smoothing thin-plate spline [20], which produces an estimate $\hat{z}_n(x)$ that

minimizes

$$\psi \iint \left[\left(\frac{\partial^2}{\partial x^2} \hat{z}_n(x, y) \right)^2 + 2 \left(\frac{\partial}{\partial x} \frac{\partial}{\partial y} \hat{z}_n(x, y) \right)^2 + \left(\frac{\partial^2}{\partial y^2} \hat{z}_n(x, y) \right)^2 \right] dx dy + \|\hat{\mathbf{z}}_n - \mathbf{y}_n\|^2$$

where ψ is a smoothness parameter. The standard thin-plate spline is obtained as the limiting case when $\psi \rightarrow 0$.

We individually applied the smoothing thin-plate spline for each time frame, and the results are depicted in Fig. 2(c) for $\psi = 5/(4\pi^2)$. This value was chosen so that the estimates from the smoothing thin-plate spline and the kriging filter are identical at the first time frame. At later time frames, the kriging filter produces qualitatively better results than the smoothing thin-plate spline, particularly at time frame 10, where the smoothing thin-plate spline has greatly attenuated the peak of the function. Although the results from the thin-plate spline can be altered by adjusting ψ , there is a tradeoff between smoothness of the estimate and attenuation of the peak. If the smoothness is relaxed to reduce the attenuation of the peak, the complete estimate becomes significantly rougher. Therefore, the smoothing thin-plate spline is unable to produce an estimate that is as good as the one from the kriging filter, which uses temporal information to maintain the peak while allowing a considerable degree of smoothing.

VIII. AN APPLICATION IN MEDICAL IMAGING

In this section, we describe a physical problem in medical imaging that is particularly well suited for the kriging filter. The problem illustrates the rationale behind the kriging update model and how model parameters might be chosen. Here, we provide only a brief description of the problem; detailed results can be found in [21].

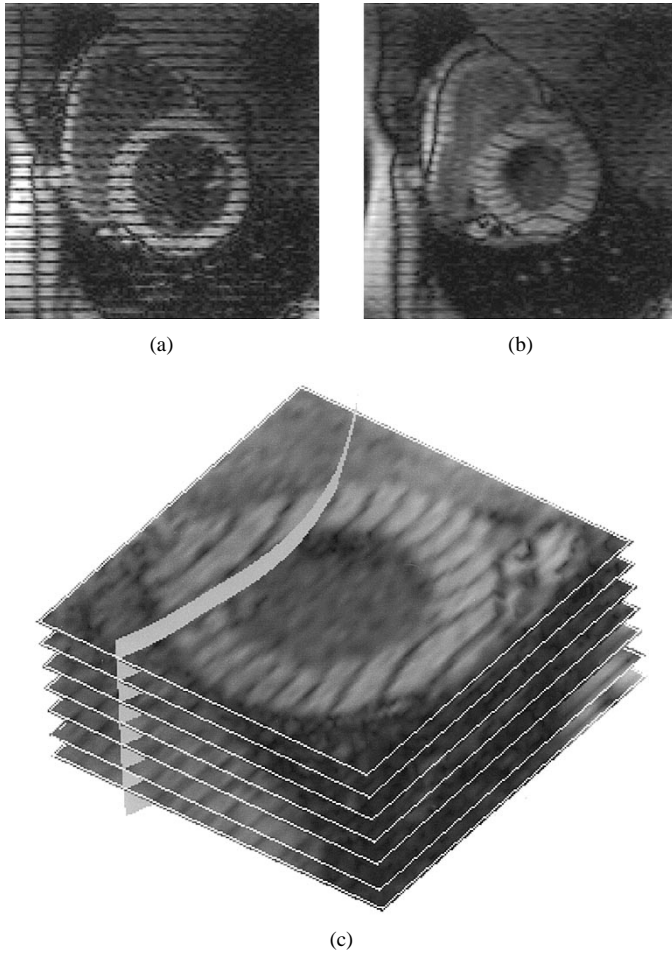


Fig. 3. Tagged MR images of the LV (bright ring of tissue). (a) One image immediately after applying parallel planar tags. (b) Same image after tag deformation due to normal LV contraction. (c) Stack of tagged images with one tag surface estimate depicted. Each dark line in each image corresponds to its own tag surface.

The application involves magnetic resonance (MR) tagging [22], which is used to study the motion of the heart, particularly the left ventricle (LV). In MR tagging, the magnetization properties of planes of tissue (so-called “tag planes”) are altered, and as a result, tissue lying within those planes appears dark in subsequent MR images. An example of a tagged image is shown in Fig. 3(a), in which a parallel set of tag planes has been applied orthogonal to the image plane, producing a stripe pattern. The advantage of using MR tagging can be seen in Fig. 3(b), which shows that the tags move with the tissue, providing a visual motion cue. Note the distortion of the stripe pattern.

Using tagged MR images, quantitative analysis of LV function can be performed by tracking the deformation of tag planes [2], [23]. However, this poses a challenge because tag surfaces (deformed tag planes) exist in three-dimensional space, and images are inherently 2-D. For example, in Figs. 3(a) and (b), the tag surfaces extend directly out of the image planes. In order to identify the complete extent of a tag surface, images must be obtained at several spatial positions, and the tag surface must be estimated from the resulting stack of images, as depicted in Fig. 3(c). Such stacks

are typically obtained at a series of time frames during heart contraction.

Thus, MR tagging presents an estimation problem with data distributed in both time and space. In order to frame tag surface estimation as a function estimation problem, we define a time series of 2-D functions $z_1(x, y), z_2(x, y), \dots$, where each function defines the position of the tag surface at one time in the series. The spatial variables x and y define a reference plane that is parallel to the initial tag plane, and each tag surface is then characterized as a height function relative to this plane. These functions are smooth and continuous within the LV tissue.

Estimating the functions from the images is a problem well suited for the kriging filter for several reasons. First, the image stacks represent an observation process that samples the tag surface location at discrete points in space and time. Spatially, the observation locations are defined by the positions of the image planes and the separation between pixels within the images. In addition, only observations within the LV are useful, which leads to an irregular pattern of observations because the LV is irregular in shape. The resulting data vector therefore requires the use of an estimation approach that allows an irregular distribution of data as in the kriging filter.

Second, our noise model is appropriate for this problem. To obtain observations of the tag surface position, points along the dark tag stripes in the images must be identified. This process is affected by image noise and the underlying pixelization of the images, leading to uncertainty in the observations. Because the uncertainty has known statistical characteristics [24], Σ_n can be accurately determined in our model.

Perhaps the best incentive for using the kriging filter on this problem is the presence of the drift vector $\mathbf{f}(x, y)$ in the model. We have found that deforming tag surfaces in the LV can be closely approximated by quadratic functions [21]. Therefore, an appropriate choice of drift vector is

$$\mathbf{f}(x, y) = [1 \quad x \quad y \quad x^2 \quad y^2 \quad xy]^T.$$

Furthermore, even though this quadratic model generally applies, the behavior of the drift coefficients $\mathbf{d}_1, \mathbf{d}_2, \dots$ changes with position in the LV and time in the contraction cycle. In addition, there can be dramatic differences in the drift coefficients between normal and diseased hearts. Therefore, defining accurate statistical properties of the drift coefficients may be impossible. The kriging filter provides a unique mechanism to take advantage of the well-characterized drift vector without further assumptions on the poorly characterized drift coefficients.

Finally, the initial tag surface is known to be a plane, so that $\alpha = 0$ and $z_0(x, y)$ is a constant. However, the value of the constant is not well known as a result of imaging uncertainties. Fortunately, the kriging filter produces identical results for any assumed initial location. This is because a plane is within the space spanned by the drift functions. Therefore, the initial position may be attributed to \mathbf{m}_0 , which need not be specified.

IX. DISCUSSION AND CONCLUSION

The kriging filter presented here is a practical method for reconstructing a time series of spatial functions from

scattered data. Set in the concise framework of Section VI, the algorithm is easily adaptable to a variety of problems. The keys to deriving the filter are Propositions 2 and 3, which establish the optimality of the recursive formulation. The key to implementation is the development of (36) through (39), which permit a finite amount of information to be stored at each time. This is the primary difference between this result and the vector filter of Kitanidis [15]. That result cannot be applied to this problem because it would require the estimate, and the error variance to be stored for every point in space at each time.

To use our algorithm, we must specify appropriate drift functions, a data covariance function, and a noise covariance matrix. Choosing these parameters can prove challenging, which is an issue discussed in the literature (see [6] and [25]). We suggest using monomial drift functions, which are commonly used in universal kriging [16]. For the covariance function, Watson [26] states that the choice is not critical if enough data points are available. On the other hand, Matheron [18] advocates the use of odd polynomials in $\|x_a - x_b\|$. Finally, choosing the noise covariance is analogous to selecting the smoothness parameters for smoothing splines, as addressed by Wahba [20].

Because our data model is based on universal kriging and our derivation parallels that of Kalman filtering, our algorithm is closely related to both. In fact, if we assume no measurement noise, so that $\Sigma_n = 0$, the kriging filter is reduced to individually kriging each function in the series using the data from that time frame only. Thus, if perfect measurements are available, we should simply employ kriging rather than our method. Second, if no drift terms are assumed, the series of functions will be zero mean. Then, the kriging filter reduces to the space-time Kalman filter discussed in [9]. In that sense, our algorithm is a bridge between Kalman filtering and universal kriging.

The method is also closely related to space-time kriging [7] and cokriging [8]. In fact, use of either method with the kriging update model results in an estimate identical to ours, which is a fact that we used to verify that our algorithm was functioning properly. The advantage of our method is that it requires smaller matrix inversions that are faster to compute and less subject to numerical instabilities. The disadvantage is that the kriging update model is more restrictive than the general models that may be used with these other methods. As in cokriging and space-time kriging, our method could also be adapted to estimate past values of the function in a manner analogous to Kalman smoothing. The approach is not, however, suited for temporal forecasting because the lack of a specified model for \mathbf{m}_n precludes the prediction of future functions $z_n(x)$. Future functions $c_n(x)$ may be predicted, but the unsatisfying result is that all future estimates are the same as the present estimate.

In the various kriging methods, the semivariogram is often used rather than the covariance because it is generally easier to model. The semivariogram of $u_n(x)$ is defined as $\gamma_n(x_a, x_b) = \frac{1}{2}E\{(u_n(x_a) - u_n(x_b))^2\}$. When estimating functions using kriging, the semivariogram and covariance can be interchanged without affecting the result, as long as the drift includes a

constant component ($f_1(x) = 1$), and $u_n(x)$ is wide-sense stationary. Given the same caveats, the kriging filter can also employ the semivariogram in place of the covariance function, with one additional alteration. In the definition of L_n , the matrix Σ_n must be replaced by $-\Sigma_n$, giving $L_n = (H_{n-1} + \Gamma - \Sigma_n)^{-1}$, where Γ is the semivariogram matrix. Thus, when using the kriging filter, we must take care to distinguish between semivariograms and the generalized covariances of intrinsic random processes, as discussed by Matheron [18] and Kent and Mardia [19].

One limitation of the kriging filter as presented is the assumption that the initial value of $c_0(x)$ is zero-mean with a covariance function $E[c_0(x_a)c_0(x_b)] = \alpha k(x_a, x_b)$. Of course, a known nonzero mean can be addressed by simply subtracting that mean from all data vectors, then performing the estimation assuming a mean of zero, and finally adding the mean back in at the end. More general covariance functions, however, require additional complexity. We chose (9) because it captures two of the most likely cases, without adding significant complexity to the description of the algorithm. Those two cases are 1) $c_0(x)$ is known so that $\alpha = 0$, and 2) $c_0(x)$ is determined by the same process that determines $u_n(x)$, in which case, α is determined by the length of time that process has been running. On the other hand, arbitrary initial covariance functions can also be used, but they require additional terms in the algorithm that depend on the chosen covariance function.

Another limitation of the kriging filter is that we assume the observation points are fixed for all times. In typical settings, sensors may fail, or new sensors may be added. In fact, the tag surface estimation example is a case where observation points change because the LV moves relative to the fixed image planes. We therefore addressed changing observation locations in [21], and the required changes to the algorithm are discussed therein. The basic forms of (20), (37), and (43) do not change, but careful bookkeeping of the error covariance terms is required. In the general case, where all observations points change at every time, an answer of the form (38) is not possible.

APPENDIX A PROOFS OF PROPOSITIONS

Proposition 3 can be viewed as a generalization of Proposition 2. We will therefore prove Proposition 3 first and then show how this also proves Proposition 2.

Proof of Proposition 3: The linear combination (14) may equivalently be written

$$\hat{z}_n(x) = \hat{c}_{n-1}(x) + \mathbf{q}_n^T(x)\tilde{\mathbf{y}}_n + \sum_{i=1}^{n-1} \mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i \quad (44)$$

because the linear span of the innovations $\tilde{\mathbf{y}}_i$ is the same as that of the original data and because $\hat{c}_{n-1}(x)$ is within that span. For (44), the appropriate unbiasedness constraints are

$$F\mathbf{q}_n(x) = \mathbf{f}(x) \quad (45)$$

$$F\mathbf{b}_i(x) = 0, \quad i = 1, \dots, n-1. \quad (46)$$

Thus, calculating $\hat{z}_n(x)$ by minimizing the mean square error of (14) subject to (15) is equivalent to minimizing the mean square error of (44) subject to (45) and (46).

Then, to prove the proposition, we need only show that this minimization leads to the solution

$$\mathbf{b}_1(x) = \cdots = \mathbf{b}_{n-1}(x) = \mathbf{0}. \quad (47)$$

To establish this, we note that for any $i < n$, the term $\mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i$ is statistically orthogonal to both $(c_{n-1}(x) - \hat{c}_{n-1}(x))$ and $(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1})$ by the projected orthogonality theorem. It is also orthogonal to $u_n(x)$, \mathbf{u}_n , and \mathbf{v}_n , which are assumed to be white. Finally, it is orthogonal to the deterministic quantities $\mathbf{f}^T(x)\mathbf{m}_n$ and $F^T\mathbf{m}_n$ because $E[\mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i] = 0$ for any $\mathbf{b}_i(x)$ satisfying (46). These facts combine to show

$$\begin{aligned} E[(z_n(x) - \hat{c}_{n-1}(x))\mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i] &= 0 \\ E[(\tilde{\mathbf{y}}_n)\mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i] &= 0 \end{aligned}$$

because $z_n(x) - \hat{c}_{n-1}(x) = u_n(x) + (c_{n-1}(x) - \hat{c}_{n-1}(x)) + \mathbf{f}^T(x)\mathbf{m}_n$, and $\tilde{\mathbf{y}}_n = \mathbf{u}_n + (\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1}) + F^T\mathbf{m}_n + \mathbf{v}_n$. This, in turn, leads to

$$\begin{aligned} E[(z_n(x) - \hat{z}_n(x))^2] &= E[(z_n(x) - \hat{c}_{n-1}(x) - \mathbf{q}_n^T(x)\tilde{\mathbf{y}}_n)^2] \\ &\quad + E\left[\left(\sum_{i=1}^{n-1} \mathbf{b}_i^T(x)\tilde{\mathbf{y}}_i\right)^2\right]. \end{aligned}$$

Minimizing this expression can be accomplished by individually minimizing both the first term, which depends only on $\mathbf{q}_n(x)$, and the second term, which depends only on $\mathbf{b}_i(x)$. Clearly, the non-negative second term is minimized by (47), which also satisfies (46). \square

Proof of Proposition 2: Follow the proof of Proposition 3, replacing $z_n(x)$ with $c_n(x)$, $\mathbf{q}_n(x)$ with $\mathbf{g}_n(x)$, and $\mathbf{f}(x)$ with $\mathbf{0}$.

APPENDIX B AN EXPRESSION FOR H_n

In order to derive an expression for

$$H_n = E[(\mathbf{c}_n - \hat{\mathbf{c}}_n)(\mathbf{c}_n - \hat{\mathbf{c}}_n)^T]$$

we use (33) to write

$$H_n = E[(\mathbf{c}_n - \hat{\mathbf{c}}_{n-1} - G_n\tilde{\mathbf{y}}_n)(\mathbf{c}_n - \hat{\mathbf{c}}_n)^T].$$

We observe that $G_n F^T = 0$, which allows us to apply the projected orthogonality theorem to obtain

$$E[(G_n\tilde{\mathbf{y}}_n)(\mathbf{c}_n - \hat{\mathbf{c}}_n)^T] = \mathbf{0}.$$

Therefore

$$\begin{aligned} H_n &= E[(\mathbf{c}_n - \hat{\mathbf{c}}_{n-1})(\mathbf{c}_n - \hat{\mathbf{c}}_n)^T] \\ &= E[(\mathbf{c}_n - \hat{\mathbf{c}}_{n-1})(\mathbf{c}_n - \hat{\mathbf{c}}_{n-1} - G_n\tilde{\mathbf{y}}_n)^T]. \end{aligned}$$

We then expand $\tilde{\mathbf{y}}_n$ to obtain

$$\begin{aligned} H_n &= E[\{\mathbf{c}_n - \hat{\mathbf{c}}_{n-1}\}\{\mathbf{c}_n - \hat{\mathbf{c}}_{n-1} - G_n(\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1} \\ &\quad + \mathbf{u}_n + \mathbf{v}_n + F^T\mathbf{m}_n)\}^T] \\ &= E[\{\mathbf{u}_n + \mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1}\}\{(\mathbf{u}_n + \mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1})^T \\ &\quad \times (I - G_n^T)\}] \end{aligned}$$

where we use the facts that $G_n F^T = 0$, \mathbf{v}_n is uncorrelated with $\mathbf{c}_n - \hat{\mathbf{c}}_{n-1}$, and $\mathbf{c}_n = \mathbf{u}_n + \mathbf{c}_{n-1}$. This expression depends only on the terms \mathbf{u}_n and $\mathbf{c}_{n-1} - \hat{\mathbf{c}}_{n-1}$, which are uncorrelated. Evaluating the expectation, we find

$$H_n = (K + H_{n-1})(I - G_n^T).$$

APPENDIX C AN EXPRESSION FOR $\mathbf{h}_n(x)$

In order to derive an expression for

$$\mathbf{h}_n(x) = E[(c_n(x) - \hat{c}_n(x))(\mathbf{c}_n - \hat{\mathbf{c}}_n)]$$

we first observe that

$$E[\hat{c}_n(x)(\mathbf{c}_n - \hat{\mathbf{c}}_n)] = \mathbf{0}$$

by the projected orthogonality theorem. This theorem applies because $\hat{c}_n(x)$ is a linear combination of the data where the coefficient vectors in that linear combination satisfy (17). Thus

$$\mathbf{h}_n(x) = E[c_n(x)\mathbf{c}_n] - E[c_n(x)\hat{\mathbf{c}}_n]. \quad (48)$$

To simplify (48), we write in terms of the original data vectors and not the innovations

$$\hat{\mathbf{c}}_n = \sum_{i=1}^n B_i \mathbf{y}_i.$$

Knowing the matrices B_1, \dots, B_n is unimportant for this derivation. Using $\mathbf{y}_i = \mathbf{c}_i + \mathbf{v}_i + F^T\mathbf{m}_i$, we obtain

$$\hat{\mathbf{c}}_n = \sum_{i=1}^n B_i(\mathbf{c}_i + \mathbf{v}_i + F^T\mathbf{m}_i)$$

which yields

$$\begin{aligned} E[c_n(x)\hat{\mathbf{c}}_n] &= \sum_{i=1}^n B_i E[c_n(x)(\mathbf{c}_i + \mathbf{v}_i + F^T\mathbf{m}_i)] \\ &= \sum_{i=1}^n B_i E[c_n(x)\mathbf{c}_i] \end{aligned}$$

where we use the facts that the zero-mean process $c_n(x)$ is uncorrelated with \mathbf{v}_i and that $B_i F^T = 0$ to ensure unbiasedness. Using this result in (48), we find

$$\mathbf{h}_n(x) = E[c_n(x)\mathbf{c}_n] - \sum_{i=1}^n B_i E[c_n(x)\mathbf{c}_i].$$

We now need to know $E[c_n(x)\mathbf{c}_i]$ for $i \leq n$. We use

$$c_n(x) = c_0(x) + \sum_{l=1}^n u_l(x)$$

to obtain

$$\begin{aligned} E[c_n(x)\mathbf{c}_i] &= E\left[\left(c_0(x) + \sum_{l=1}^n u_l(x)\right)\left(\mathbf{c}_0 + \sum_{m=1}^i \mathbf{u}_m\right)\right] \\ &= (\alpha + i)\mathbf{k}(x), \quad i \leq n. \end{aligned}$$

In this result, we use (9) and the fact that $u_l(x)$ is white in time and uncorrelated with $c_0(x)$. Thus

$$\mathbf{h}_n(x) = (\alpha + n)\mathbf{k}(x) - \sum_{i=1}^n (\alpha + i)B_i\mathbf{k}(x). \quad (49)$$

Next, we replace $\mathbf{k}(x)$ by $KK^{-1}\mathbf{k}(x)$ to obtain

$$\begin{aligned} \mathbf{h}_n(x) &= \left[(\alpha + n)K - \sum_{i=1}^n (\alpha + i)B_iK \right] K^{-1}\mathbf{k}(x) \\ &= H_n K^{-1}\mathbf{k}(x). \end{aligned}$$

The simplification is possible because the j th column of the expression in brackets is (49) evaluated at x_j . Thus, the bracketed expression is H_n .

REFERENCES

- [1] M. Uchiyama and K. Hakomori, "Measurement of instantaneous flow rate through estimation of velocity profiles," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 380–388, 1983.
- [2] W. S. Kerwin and J. L. Prince, "Cardiac material markers from tagged MR images," *Med. Image Anal.*, vol. 2, no. 4, pp. 339–353, 1998.
- [3] G. D. Egbert and D. P. Lettenmaier, "Stochastic modeling of space-time structure of atmospheric chemical deposition," *Water Res.*, vol. 22, no. 2, pp. 165–179, 1986.
- [4] P. Bogaert, "Comparison of kriging techniques in a space-time context," *Math. Geo.*, vol. 28, no. 1, pp. 73–86, 1996.
- [5] G. K. Robinson, "That BLUP is a good thing: The estimation of random effects," *Stat. Sci.*, vol. 6, no. 1, pp. 15–32, 1991.
- [6] R. Christensen, *Linear Models for Multivariate, Time Series, and Spatial Data*. New York: Springer-Verlag, 1991, ch. 6.
- [7] S. Rouhani and D. E. Myers, "Problems in space-time kriging of geohydrological data," *Math. Geo.*, vol. 22, no. 5, pp. 611–623, 1990.
- [8] A. Papritz and H. Fluhler, "Temporal change of spatially autocorrelated soil properties: Optimal estimation by cokriging," *Geoderma*, vol. 62, pp. 29–43, 1994.
- [9] H.-C. Huang and N. Cressie, "Spatio-temporal prediction of snow water equivalent using the Kalman filter," *Comput. Stat. Data Anal.*, vol. 22, no. 2, pp. 159–175, 1996.
- [10] O. Berke, "On spatiotemporal prediction for on-line monitoring data," *Commun. Statist.—Theory Meth.*, vol. 27, no. 9, pp. 2343–2369, 1998.
- [11] C. Goodall and K. V. Mardia, "Challenges in multivariate spatio-temporal modeling," in *Proc. XVIIth Int. Biometric Conf.*, Hamilton, Ont., Canada, Aug. 1994, vol. 1, pp. 1–17.
- [12] B. Friedland, "Treatment of bias in recursive filtering," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 359–367, 1969.
- [13] M. B. Ignagni, "Separate-bias Kalman estimator with bias state noise," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 338–341, 1990.
- [14] J. Y. Keller and M. Darouach, "Optimal two-stage Kalman filter in the presence of random bias," *Automatica*, vol. 33, no. 9, pp. 1745–1748, 1997.
- [15] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.
- [16] M. David, *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*. Amsterdam, The Netherlands: Elsevier, 1988.
- [17] D. E. Myers, "Matrix formulation of co-kriging," *Math. Geo.*, vol. 14, no. 3, pp. 249–257, 1982.
- [18] G. Matheron, "The intrinsic random functions and their applications," *Adv. Appl. Prob.*, vol. 5, pp. 439–468, 1973.
- [19] J. T. Kent and K. V. Mardia, "The link between kriging and thin-plate splines," in *Probability, Statistics and Optimization*, F. P. Kelly, Ed. New York: Wiley, 1994, ch. 24, pp. 325–339.
- [20] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1990.
- [21] W. S. Kerwin and J. L. Prince, "Tracking MR tag surfaces using a spatiotemporal filter and interpolator," *Int. J. Imag. Syst. Tech.*, vol. 10, no. 2, pp. 128–142, 1999.
- [22] E. A. Zerhouni, D. M. Parish, W. J. Rogers, A. Yang, and E. P. Shapiro, "Human heart: Tagging with MR imaging—A method for noninvasive assessment of myocardial motion," *Radiol.*, vol. 169, pp. 59–63, 1988.
- [23] M. J. Moulton *et al.*, "Spline surface interpolation for calculating 3-D ventricular strains from MRI tissue tagging," *Amer. J. Physiol., Heart Circ. Physiol.* 39, vol. 270, pp. H281–H297, 1996.
- [24] E. Atalar and E. McVeigh, "Optimization of tag thickness for measuring position with magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 13, pp. 152–160, Mar. 1994.
- [25] N. Cressie, *Statistics for Spatial Data*, 2nd. ed. New York: Wiley, 1993.
- [26] G. S. Watson, "Smoothing and interpolating by kriging and with splines," *Math. Geo.*, vol. 16, no. 6, pp. 601–615, 1984.



William S. Kerwin (S'95) was born in Rochester, NY, in 1967. He received the B.S. degree in electrical engineering, biomedical engineering option, from Carnegie Mellon University, Pittsburgh, PA, in 1989 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1992. He is currently pursuing the Ph.D. degree in electrical engineering at The Johns Hopkins University, Baltimore, MD.

From 1989 to 1995, he served in the United States Air Force, achieving the rank of Captain in 1993. Since 1995, he has been a Research Assistant with the Image Analysis and Communications Laboratory, Johns Hopkins University, where he studies cardiac motion using magnetic resonance imaging. His doctoral studies are supported by a Whitaker Foundation Graduate Fellowship in biomedical engineering.



Jerry L. Prince (S'78–M'83–SM'96) received the B.S. degree from the University of Connecticut, Storrs, in 1979 and the S.M., E.E., and Ph.D. degrees in 1982, 1986, and 1988, respectively, from the Massachusetts Institute of Technology, Cambridge, all in electrical engineering and computer science.

After receiving the Ph.D. on the subject of geometric reconstruction in computed tomography, he joined the Technical Staff at The Analytic Sciences Corporation (TASC), Reading, MA, where he contributed to the design of an automated vision system for synthetic aperture radar imaging. He joined the faculty at the Johns Hopkins University, Baltimore, MD, in 1989 and has been an Associate Professor with the Department of Electrical and Computer Engineering since 1994. He holds joint appointments in the Department of Radiology and the Department of Biomedical Engineering. His current research interests are in image processing and computer vision with primary application to medical imaging. Major projects include magnetic resonance imaging of cardiac motion, three-dimensional brain image analysis, and vector tomography.

Dr. Prince is a member of the Sigma Xi professional society and the Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi honor societies. He is a former Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, is a recipient of the 1993 National Science Foundation Presidential Faculty Fellows Award, and was Maryland's 1997 Outstanding Young Engineer.