RESEARCH

Using Dual Approximation for Best Linear Unbiased Estimators in Continuous Time, with Application to Continuous-Time Phase Estimation

Todd K. Moon*, Randy Christensen* and Jacob H. Gunther*

Abstract

Best linear unbiased estimator (BLUE) theory is well established for discrete, finite-dimensional vectors, where methods of vector gradients can be used on a constrained optimization problem. However, when the observation is infinite-dimensional (e.g., continuous-time functions), the gradient-based approach can be problematic. We pose the BLUE problem as an instance of a dual approximation problem, which recasts the problem into finite dimensional space employing the principle of orthogonality, requiring no gradients for solution. To demonstrate the ideas, they are first developed on a finite-dimensional problem, then applied to continuous-time. We present an example application of phase estimation from continuous-time observations.

Keywords: statistical estimation theory; Best linear unbiased estimator; Infinite dimensional optimization

1 Introduction

The theory of the best linear unbiased estimator (BLUE) is well established for discrete p-dimensional vectors (see, e.g., [1, Chapter 6]). Despite the prevalence of digital (discrete-time/discrete-vector) signal processing, there are still situations where continuous-time processing is warranted, such as for high-frequency signals where sampling and digital processing would be more expensive than continuous-time hardware. In this paper, we generalize BLUE signal processing to continuous-time signals. For an additional measure of generality, this is done for complex-valued functions.

In [1], the BLUE problem is posed as a constrained optimization problem, in which the variance is minimized subject to linear unbiased constraints, and the optimization problem is solved for the optimal estimator parameter vector using straightforward vector gradient computations. This gradient-based approach becomes problematic when the coefficient "vector" is actually a continuous-time function, resulting in an infinite-dimensional problem. A calculus (gradient)-based approach might suggest that solving the problem in continuous time would lead to solution using calculus of variations. However, the problem can be cast as a dual optimization problem [2, 3], which provides optimal solutions using the principle of orthogonality (no gradients!), turning the continuous-time problem into another problem in matrix linear algebra. In this paper, for purposes of comparison, the BLUE estimator is derived using dual optimization theory for complex vectors. Notation is then developed used for the continuous time problem, and the continuous-time BLUE estimator is developed.

In section 2, a summary of dual approximation is presented. In section 3 this theory is applied to (complex) discrete vectors. This leads to familiar results, and will help show how dual approximation applies. In section 4, the application to continuous-time functions is presented. As an application of the estimator theory, an algorithm for phase difference estimation is presented.

2 Dual Approximation Theory

Dual approximation is an optimization set in Hilbert space which provides the solution of a minimum-norm problem, subject to linear (inner product) constraints. Let S be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. Let M be a subspace of S, and let $M^{\perp}$ be the subspace of vectors orthogonal to M. Let

*Correspondence: Todd.Moon@usu.edu; Randy.Christensen@usu.edu; Jake.Gunther@usu.edu
Electrical and Computer Engineering Dept., Utah State University, Logan, UT, USA
Full list of author information is available at the end of the article

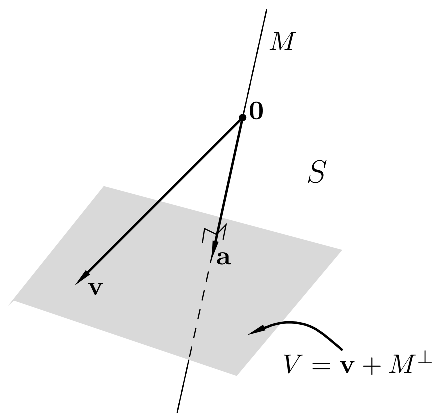

Figure 1 A linear variety $\mathbf{x}_0 + M$ and the point $\mathbf{v}_0$ of minimum norm in it

$\mathbf{v} \in S$. The set $V = \mathbf{v} + M^\perp$ is called the translation of $M^\perp$ by $\mathbf{v}$. This translated set is called a *linear variety*. The concept is portrayed in figure 1. There is a unique vector $\mathbf{a} \in V$ of minimum norm $\|\mathbf{a}\|$. By the principle of orthogonality [3, 2] and as suggested in figure 1, the minimum-norm vector $\mathbf{a}$ is orthogonal to $M^\perp$, and hence lies in M.

We apply this geometric concept to dual approximation as follows. Let $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m$ be a finite set of linearly independent vectors in S. Let $M = \operatorname{span}(\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m\})$. The set of $\mathbf{z} \in S$ such that

$$\begin{aligned} \langle \mathbf{z}, \mathbf{h}_1 \rangle &= 0 \\ \langle \mathbf{z}, \mathbf{h}_2 \rangle &= 0 \\ &\vdots \\ \langle \mathbf{z}, \mathbf{h}_m \rangle &= 0 \end{aligned}$$

forms a subspace in S, and is, in fact $M^\perp$. Now consider the set of all $\mathbf{a} \in S$ that satisfy the set of linear constraints

$$\begin{aligned} \langle \mathbf{a}, \mathbf{h}_1 \rangle &= d_1 \\ \langle \mathbf{a}, \mathbf{h}_2 \rangle &= d_2 \\ &\vdots \\ \langle \mathbf{a}, \mathbf{h}_m \rangle &= d_m \end{aligned} \tag{1}$$

If we can find any point $\mathbf{a} = \mathbf{a}_0$ that satisfies the constraints in (1), then for any vector $\mathbf{z} \in M^\perp$ the point $\mathbf{a}_0 + \mathbf{z}$ also satisfies the constraints in (1). Thus, any point in the linear variety $V = \mathbf{a}_0 + M^\perp$ is a solution to the set of constraints (1).

We impose a way of selecting a single one of the possible solutions: choose the solution $\mathbf{a}$ of minimum norm in the linear variety $\mathbf{a}_0 + M^\perp$. That is, we desire to solve the constrained optimization problem

$$\begin{aligned} &\text{minimize } \|\mathbf{a}\|^2 \\ &\text{subject to the inner product constraints in (1).} \end{aligned}$$

The discussion above indicates that $\mathbf{a}$ is orthogonal to $M^\perp$, that is, $\mathbf{a} \in M$, so

$$\mathbf{a} = \sum_{i=1}^{m} c_i \mathbf{h}_i.$$

The coefficients in this linear combination are selected to satisfy the constraints (1). This leads to the equations

$$\begin{bmatrix} \langle \mathbf{h}_1, \mathbf{h}_1 \rangle & \langle \mathbf{h}_2, \mathbf{h}_1 \rangle & \cdots & \langle \mathbf{h}_m, \mathbf{h}_1 \rangle \\ \langle \mathbf{h}_1, \mathbf{h}_2 \rangle & \langle \mathbf{h}_2, \mathbf{h}_2 \rangle & \cdots & \langle \mathbf{h}_m, \mathbf{h}_2 \rangle \\ \vdots & & & \\ \langle \mathbf{h}_1, \mathbf{h}_m \rangle & \langle \mathbf{h}_2, \mathbf{h}_m \rangle & \cdots & \langle \mathbf{h}_m, \mathbf{h}_m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix}. \tag{2}$$

With $H = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_m \end{bmatrix}$, the minimum norm solution $\mathbf{a}$ is $\mathbf{a} = H\mathbf{c}$, where $\mathbf{c}$ solves (2).

This presentation of dual approximation has been portrayed as if the Hilbert space is finite-dimensional. But the method applies if S is infinite dimensional (e.g., the space of L_2 functions). The fact that there are m inner product constraints recasts the problem as an m-dimensional problem.

3 Dual Optimization for BLUE for discrete vectors

Let the linearly independent columns of a matrix H,

$$H = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_p \end{bmatrix} \tag{3}$$

be considered as basis vectors for a signal model, where each $\mathbf{h}_i \in \mathbb{C}^N$ and where we assume that $N > p$. The columns of H form a basis for a linear signal + noise model, $\mathbf{x} = H\boldsymbol{\theta} + \boldsymbol{\nu}$, where $\boldsymbol{\theta}$ is a fixed (nonrandom) parameter vector $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \end{bmatrix}^T$. The noise is assumed to be zero-mean with covariance

$$\operatorname{cov}(\boldsymbol{\nu}) = E[\boldsymbol{\nu}\boldsymbol{\nu}^H] = C.$$

Here, ^{H} denotes conjugate transpose. The mean of the signal is $E[\mathbf{x}] = H\boldsymbol{\theta}$.

We seek an estimator of the parameter vector $\boldsymbol{\theta}$ that is a linear function of the observed data $\mathbf{x}$. That is, each parameter estimator has the form

$$\hat{\theta}_i = \mathbf{a}_i^H \mathbf{x} = \mathbf{a}_i^H H\boldsymbol{\theta} + \mathbf{a}_i^H \boldsymbol{\nu},$$

where the linear coefficient vector $\mathbf{a}_i$ is to be determined. These estimates can be stacked to form a parameter estimator vector

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_p \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^H \\ \vdots \\ \mathbf{a}_p^H \end{bmatrix} \mathbf{x} = A^H \mathbf{x}$$

where A^H is a $p \times N$ matrix. The variance of $\hat{\theta}_i$ and the covariance of $\hat{\boldsymbol{\theta}}$ are

$$\mathrm{var}(\hat{\theta}_i) = \mathbf{a}_i^H C \mathbf{a}_i \qquad \mathrm{cov}(\hat{\boldsymbol{\theta}}) = A^H C A. \tag{4}$$

It will be helpful to define an inner product appropriate for this problem. Define the weighted inner product

$$\langle \mathbf{a}, \mathbf{b} \rangle_C = \mathbf{a}^H C \mathbf{b}. \tag{5}$$

In this definition we have placed the conjugate on the first element. Using this inner product define an induced norm (see [3, Chapter 2])

$$\|\mathbf{a}\|_C^2 = \langle \mathbf{a}, \mathbf{a} \rangle_C = \mathbf{a}^H C \mathbf{a}.$$

The variance of the estimator $\hat{\theta}_i$ can be expressed using this norm as $\mathrm{var}(\hat{\theta}_i) = \|\mathbf{a}_i\|_C^2$. The mean of the estimator is $E[\hat{\boldsymbol{\theta}}] = A^H H \boldsymbol{\theta}$. In order for the estimator to be unbiased it must be the case that $A^H H = I_p$, where I_p is the $p \times p$ identity matrix. That is,

$$\mathbf{a}_i^H \mathbf{h}_j = \delta_{ij} \qquad \text{or} \qquad \mathbf{h}_j^H \mathbf{a}_i = \delta_{ij}. \tag{6}$$

Using the inner product (5), these constraints can be expressed as

$$\mathbf{a}_i^H \mathbf{h}_j = \langle \mathbf{a}_i, C^{-1} \mathbf{h}_j \rangle_C = \delta_{ij}.$$

The BLUE can now be expressed as follows. For each $i = 1, 2, \ldots, p$ (that is, for each parameter):

$$\begin{aligned} &\text{minimize } \|\mathbf{a}_i\|_C^2 \\ &\text{subject to } \langle \mathbf{a}_i, C^{-1} \mathbf{h}_j \rangle_C = \delta_{ij}, \qquad j = 1, 2, \ldots, p. \end{aligned}$$

By the theory of dual approximation, the $\mathbf{a}_i$ satisfying this constrained optimization problem lies in the span of the functions appearing in the inner product constraints,

$$\mathbf{a}_i \in \mathrm{span}(\{C^{-1}\mathbf{h}_1, \ldots, C^{-1}\mathbf{h}_p\}). \tag{7}$$

so that

$$\mathbf{a}_i = C^{-1} H \mathbf{c}_i,$$

where the $p \times 1$ coefficient vector $\mathbf{c}_i$ is found to satisfy the constraints. Multiply by H^H:

$$H^H \mathbf{a}_i = H^H C^{-1} H \mathbf{c}_i$$

The LHS represents a stack of $\mathbf{h}_j^H \mathbf{a}_i$, so by (6) each is equal to δ_{ij}. The stack of δ_{ij} can be represented as the unit vector $\mathbf{e}_i$, which places 1 in the ith location. We thus have the equation

$$H^H C^{-1} H \mathbf{c}_i = \mathbf{e}_i$$

so that

$$\mathbf{c}_i = (H^H C^{-1} H)^{-1} \mathbf{e}_i.$$

The optimal coefficient vector is then

$$\mathbf{a}_i = C^{-1} H (H^H C^{-1} H)^{-1} \mathbf{e}_i,$$

with corresponding estimator

$$\hat{\theta}_i = \mathbf{e}_i^T (H^H C^{-1} H)^{-1} H^H C^{-1} \mathbf{x}.$$

The stack of estimators yields

$$\hat{\boldsymbol{\theta}} = (H^H C^{-1} H)^{-1} H^H C^{-1} \mathbf{x}. \tag{8}$$

since $\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p \end{bmatrix}$ is an identity matrix. Using (4), the variance is

$$\begin{aligned} \mathrm{var}(\hat{\theta}_i) &= \mathbf{e}_i^t (H^H C^{-1} H)^{-1} H^H C^{-1} C C^{-1} H (H^H C^{-1} H)^{-1} \mathbf{e}_i \\ &= \mathbf{e}_i^T (H^H C^{-1} H)^{-1} \mathbf{e}_i \end{aligned}$$

Similarly, stacking up all the elements,

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) = (H^H C^{-1} H)^{-1}. \tag{9}$$

Other than the notation for complex quantities, this is identical to the BLUE estimate and its covariance derived in [1], but it was derived here using the principle of orthonality, without the use of gradients.

4 Blue for continuous-time functions

In this section we extend the application of dual approximation theory to BLUE for continuous-time functions. We start by introducing some notation. Let $L_2(0,T)$ be the space of real- or complex-valued square-integrable functions defined on the interval $(0,T)$. Define an inner product $\langle \cdot, \cdot \rangle : L_2(0,T) \times L_2(0,T) \to \mathbb{C}$ by

$$\langle f(t), g(t) \rangle = \int_0^T f(t)^* g(t)\, dt$$

From this inner product define the induced norm $\|\cdot\| : L_2(0,T) \to \mathbb{R}$ by

$$\|g(t)\|^2 = \langle g(t), g(t) \rangle.$$

We extend this inner product notation to vectors of functions. Let $\mathbf{H}(t)$ be a row vector of complex-valued $L_2(0,T)$ functions

$$\mathbf{H}(t) = \begin{bmatrix} h_1(t) & h_2(t) & \cdots & h_p(t) \end{bmatrix}. \tag{10}$$

Conceptually as t varies in the interval $t \in (0,T)$, $\mathbf{H}(t)$ forms a ("tall" $\times p$) "matrix". Then the inner product with $\mathbf{H}(t)$ is defined element-by-element as

$$\begin{aligned}&\langle a(t), \mathbf{H}(t) \rangle \\ &= \begin{bmatrix} \langle a(t), h_1(t) \rangle & \langle a(t), h_2(t) \rangle & \cdots & \langle a(t), h_p(t) \rangle \end{bmatrix}\end{aligned}$$

Similarly, let $\mathbf{A}(t)$ be a column vector of complex-valued L_2 functions

$$\mathbf{A}(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \\ \vdots \\ a_p(t) \end{bmatrix}. \tag{11}$$

Conceptually as t varies in the interval $t \in (0,T)$, $\mathbf{A}(t)$ forms a ($p\times$ "wide") "matrix". Then define

$$\langle \mathbf{A}(t), h(t) \rangle = \begin{bmatrix} \langle a_1(t), h(t) \rangle \\ \langle a_2(t), h(t) \rangle \\ \vdots \\ \langle a_p(t), h(t) \rangle \end{bmatrix}$$

We define $\langle \mathbf{A}(t), \mathbf{H}(t) \rangle$ as the $p \times p$ matrix with elements

$$[\langle \mathbf{A}(t), \mathbf{H}(t) \rangle]_{ij} = \langle a_i(t), h_j(t) \rangle$$

We denote this $p \times p$ matrix using the suggestive notation $\mathbf{A}^H\mathbf{H}$:

$$\mathbf{A}^H\mathbf{H} \stackrel{\triangle}{=} \langle \mathbf{A}(t), \mathbf{H}(t) \rangle \tag{12}$$

This is a notation only; $\mathbf{A}^H\mathbf{H}$ does not represent a product of matrices, but a matrix of inner products.

With this notation in place, let $x(t)$ be the p-parameter linear model

$$x(t) = h_1(t)\theta_1 + h_2(t)\theta_2 + \cdots + h_p(t)\theta_p + \nu(t),$$

where $h_i(t) \in L_2(0,T)$. With $\mathbf{H}(t)$ defined as in (10) the model can be written as

$$x(t) = \mathbf{H}(t)\boldsymbol{\theta} + \nu(t),$$

where $\boldsymbol{\theta}$ is a vector of parameters. The noise is assumed to be zero-mean, complex, and white. For the case of real signals,

$$E[\nu(t)\nu(s)] = \frac{N_0}{2}\delta(t-s)$$

where $\frac{N_0}{2}$ is the noise power spectral density. For the case of complex signals, the real and imaginary parts of the signal are assumed to be uncorrelated, and the real and imaginary parts of $n(t)$ satisfy

$$E[\mathrm{Re}(\nu(t))\,\mathrm{Re}(\nu(s))] = E[\mathrm{Im}(\nu(t))\,\mathrm{Im}(\nu(s))] = \frac{N_0}{2}\delta(t-s),$$

so that $E[\nu^*(t)\nu(t)] = N_0\delta(t-s)$. In either the real or imaginary case, write

$$E[\nu^*(t)\nu(t)] = N_{RZ}\delta(t-s)$$

The expectation of the signal is $E[x(t)] = \mathbf{H}(t)\boldsymbol{\theta}$.

The function $x(t)$ is observed over the interval $t \in (0,T)$. The estimators of the parameters are assumed to be linear functions of the observations. Thus, for some function $a_i(t) \in L_2(0,T)$ to be determined,

$$\begin{aligned}\hat{\theta}_i &= \int_0^T a_i^*(t)x(t)\,dt, \qquad i = 1,\ldots,p \\ &= \langle a_i(t), x(t) \rangle = \langle a_i(t), \mathbf{H}(t)\boldsymbol{\theta} + n(t) \rangle \\ &= \langle a_i(t), \mathbf{H}(t) \rangle \boldsymbol{\theta} + \langle a_i(t), n(t) \rangle\end{aligned} \tag{13}$$

Using $\mathbf{A}(t)$ defined in (11), the stack of these estimators is

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \int_0^T \begin{bmatrix} a_1(t) \\ a_2(t) \\ \vdots \\ a_p(t) \end{bmatrix} x(t)\,dt = \begin{bmatrix} \langle a_1(t), x(t) \rangle \\ \langle a_2(t), x(t) \rangle \\ \vdots \\ \langle a_p(t), x(t) \rangle \end{bmatrix} \\ &= \langle \mathbf{A}(t), x(t) \rangle \\ &= \langle \mathbf{A}(t), \mathbf{H}(t)\boldsymbol{\theta} \rangle = \langle \mathbf{A}(t), \mathbf{H}(t) \rangle \boldsymbol{\theta} + \langle \mathbf{A}(t), n(t) \rangle.\end{aligned}$$

Using the notation in (12),

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^H\mathbf{H}\boldsymbol{\theta} + \langle \mathbf{A}(t), n(t) \rangle.$$

The expected value of the estimate is $E[\hat{\boldsymbol{\theta}}] = \mathbf{A}^H\mathbf{H}\boldsymbol{\theta}$. In order for the estimate to be unbiased it must be the case that this expectation is equal $\boldsymbol{\theta}$, so that

$$\mathbf{A}^H\mathbf{H} = I_p \qquad \text{(where } I_p \text{ is the } p \times p \text{ identity).}$$

In other words, the elements of $\mathbf{A}^H\mathbf{H}$ satisfy the linear constraints

$$\langle a_i(t), h_j(t) \rangle = \delta_{ij}, \; i,j = 1,2,\ldots,p. \tag{14}$$

Since $\hat{\theta}_i - E[\hat{\theta}_i] = \langle a_i(t), n(t)\rangle$, the variance of the estimate is

$$\begin{aligned}\operatorname{var}(\hat{\theta}_i) &= E[(\hat{\theta}_i - E[\hat{\theta}_i])^*(\hat{\theta}_i - E[\hat{\theta}_i])] \\ &= N_{RZ}\langle a_i(t), a_i(t)\rangle = N_{RZ}\|a_i(t)\|_2^2.\end{aligned}$$

Adopting the BLUE paradigm, each estimator coefficient function $a_i(t)$ is to be chosen to minimize the variance of the estimate, while providing an unbiased estimate. Since $a_i(t)$ is a function over $(0, T)$, the minimization might appear to require tools such as calculus of variations to solve. However, this is not the case.

The BLUE problem can be expressed as the following constrained optimization problem. For each $i = 1, 2, \ldots, p$, separately:

$$\begin{aligned}&\text{minimize } N_{RZ}\|a_i(t)\|_2^2 \\ &\text{subject to } \langle a_i(t), h_j(t)\rangle = \delta_{ij}, \qquad j = 1, 2, \ldots, p\end{aligned}$$

Again, this is expressed in the form of a dual approximation problem. As a dual approximation problem, the norm-minimizing function $a_i(t)$ lies in span($\{h_1(t), \ldots, h_p(t)\}$), so that

$$a_i(t) = \sum_{k=1}^{p} c_{ik} h_k(t) = \mathbf{H}(t)\mathbf{c}_i, \tag{15}$$

where the $p\times 1$ coefficient vector $\mathbf{c}_i = \begin{bmatrix} c_{i1} & c_{i2} & \cdots & c_{ip}\end{bmatrix}^T$ is determined to as to satisfy the inner product constraints (14). Multiplying (15) through by $h_j(t)^*$, then integrating both sides $\int_0^T$ produces the equations

$$\langle h_j(t), a_i(t)\rangle = \sum_{k=1}^{p} \langle h_j(t), h_k(t)\rangle c_{ik} \tag{16}$$

Stacking this up for $j = 1, 2, \ldots, p$, we form the matrix

$$\begin{bmatrix} \langle h_1(t), h_1(t)\rangle & \langle h_1(t), h_2(t)\rangle & \ldots & \langle h_1(t), h_p(t)\rangle \\ \langle h_2(t), h_1(t)\rangle & \langle h_2(t), h_2(t)\rangle & \ldots & \langle h_2(t), h_p(t)\rangle \\ \vdots & & & \\ \langle h_p(t), h_1(t)\rangle & \langle h_p(t), h_1(t)\rangle & \ldots & \langle h_p(t), h_p(t)\rangle \end{bmatrix}$$

Let us denote this $p \times p$ matrix as $\mathbf{H}^H\mathbf{H}$. This can also be represented as $\langle \mathbf{H}^T(t), \mathbf{H}(t)\rangle$. By (14), (16) becomes $\mathbf{H}^H\mathbf{H}\mathbf{c}_i = \mathbf{e}_i$ so $\mathbf{c}_i = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_i$, and the estimator function in (15) is

$$a_i(t) = \mathbf{H}(t)(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_i = \mathbf{e}_i^T(\mathbf{H}^H\mathbf{H})^{-T}\mathbf{H}(t)^T$$

From (13),

$$\begin{aligned}\hat{\theta}_i &= \langle \mathbf{e}_i^T(\mathbf{H}^H\mathbf{H})^{-T}\mathbf{H}(t)^H, x(t)\rangle \\ &= \mathbf{e}_i^T(\mathbf{H}^H\mathbf{H})^{-1}\langle \mathbf{H}(t)^T, x(t)\rangle.\end{aligned}$$

Stacking this up

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^H\mathbf{H})^{-1}\langle \mathbf{H}(t)^T, x(t)\rangle, \tag{17}$$

since $\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p\end{bmatrix}$ is the identity matrix. Comparing (17) with (8) reveals their structural similarity. (Recall that in (17) the inner product $\langle \mathbf{H}(t), x(t)\rangle$ conjugates the first element, so this is analogous to the $H^H C^{-1}\mathbf{x}$.)

The variance of the estimate is

$$\begin{aligned}\operatorname{var}(\hat{\boldsymbol{\theta}}_i) &= N_{RZ}\langle \mathbf{e}_i^T(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}(t)^H, \mathbf{H}(t)(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_i\rangle \\ &= N_{RZ}\mathbf{e}_i^T(\mathbf{H}^H\mathbf{H})^{-1}\langle \mathbf{H}(t)^H, \mathbf{H}(t)\rangle(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_i \\ &= N_{RZ}\mathbf{e}_i(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_i\end{aligned}$$

Based on this, stacking all the covariances we obtain

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}) = N_{RZ}(\mathbf{H}^H\mathbf{H})^{-1}. \tag{18}$$

This is clearly analogous to (9).

5 Application: Phase Difference Estimation

As an example application, the continuous-time BLUE theory developed above is applied to the problem of estimation of the phase difference between two signals. The phase estimation problem has been widely studied. For example [4] describe phase estimation and variance of the estimate. But their results, derived for sampled signals, are not fully applicable to continuous-time signals. Phase estimation is also discussed in many other sources, such as [5, 6, 7, 8, 9], all of which employ discrete-time data. Despite all of excellent and the prevalence of discrete-time methods, phase estimation in continuous time is still of interest. For example, for a robotic navigation system, it may be desirable to estimate the phase difference in a short received burst using low-cost traditional hardware. So, while this example hardly breaks new ground, it provides a meaningful demonstration of the methods described in this paper.

Let the observed signals be the real continuous-time signals

$$x_i(t) = s_i(t) + \nu_i(t), \qquad i = 1, 2, \qquad 0 \le t \le T \tag{19}$$

where

$$s_i(t) = A_i\cos(\omega_0 t + \phi_i). \tag{20}$$

The goal is to estimate the phase difference $\alpha = \phi_1 - \phi_2$ based on an observation of the signals $x_i(t), t \in$

$(0, T)$. The frequency ω_0 is assumed to be known. The additive noise is assumed to be white Gaussian with

$$E[\nu_i(t)\nu_i(s)] = \sigma_i^2 \delta(t-s)$$

and the noise signals $\nu_1(t)$ and $\nu_2(t)$ are assumed to be uncorrelated, $E[\nu_1(t)\nu_2(s)] = 0\ \forall t, s$.

The signal (20) can be written as

$$s_i(t) = \theta_{i1}\cos(\omega_0 t) + \theta_{i2}\sin(\omega_0 t)$$

where

$$\theta_{i1} = A_i \cos(\phi_i) \qquad \theta_{i2} = -A_i \sin(\phi_i).$$

The signal $x_i(t)$ is thus seen as an instance of a linear model with basis functions $h_1(t) = \cos(\omega_0 t)$ and $h_2(t) = \sin(\omega_0 t)$. For notational convenience define

$$c = \int_0^T \cos^2(\omega_0 t)\, dt \qquad s = \int_0^T \sin^2(\omega_0 t)\, dt$$

$$d = \int_0^T \cos(\omega_0 t)\sin(\omega_0 t)\, dt$$

and

$$x_{ci} = \textstyle\int_0^T x_i(t)\cos(\omega_0 t)\, dt \quad x_{si} = \int_0^T x_i(t)\sin(\omega_0 t)\, dt.$$

Then from (17)

$$\hat{\boldsymbol{\theta}}_i = \begin{bmatrix}\hat{\theta}_{i1}\\ \hat{\theta}_{i2}\end{bmatrix} = \mathbf{H}^T\mathbf{H}^{-1}\langle \mathbf{H}^T(t), x(t)\rangle = \begin{bmatrix} c & d\\ d & s\end{bmatrix}^{-1}\begin{bmatrix}x_{ci}\\ x_{si}\end{bmatrix}$$
$$= \frac{1}{cs-d^2}\begin{bmatrix} s x_{ci} - d x_{si}\\ c x_{si} - d x_{ci}\end{bmatrix}$$

Under reasonable approximations for large ω_0,

$$c \approx \frac{T}{2} \qquad s \approx \frac{T}{2} \qquad d \approx 0 \tag{21}$$

so that

$$\begin{bmatrix}\hat{\theta}_{i1}\\ \hat{\theta}_{i2}\end{bmatrix} = \frac{2}{T}\begin{bmatrix}x_{ci}\\ x_{si}\end{bmatrix}$$

From (18), the covariance of $\hat{\boldsymbol{\theta}}_i$ is

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}_i) = \sigma_i^2 \begin{bmatrix} c & d\\ d & s\end{bmatrix}^{-1} \approx \sigma_i^2 \frac{2}{T} I.$$

The estimate of ϕ_i is

$$\hat{\phi}_i = \tan^{-1}\frac{-\theta_{i2}}{\theta_{i1}}$$

The phase difference is estimated by $\hat{\alpha} = \hat{\phi}_1 - \hat{\phi}_2$. (It can be shown that this estimate is equal to the maximum likelihood estimate.)

To compute the variance of the phase difference estimate, we first do a first-order perturbation

$$\begin{aligned} d\alpha = {} & \frac{\partial \tan^{-1}(-\theta_{12}/\theta_{11})}{\partial \theta_{12}} d\theta_{12} + \frac{\partial \tan^{-1}(-\theta_{12}/\theta_{11})}{\partial \theta_{11}} d\theta_{11} - \\ & \frac{\partial \tan^{-1}(-\theta_{22}/\theta_{21})}{\partial \theta_{12}} d\theta_{22} + \frac{\partial \tan^{-1}(-\theta_{22}/\theta_{21})}{\partial \theta_{21}} d\theta_{21} \\ = {} & \frac{-\cos(\phi_1)d\theta_{12} - \sin(\phi_1)d\theta_{11}}{A_1} - \\ & \frac{-\cos(\phi_2)d\theta_{22} - \sin(\phi_2)d\theta_{21}}{A_2} \end{aligned}$$

The increment in estimated phase difference is

$$\begin{aligned}\Delta\hat{\alpha} \approx {} & \frac{-\cos(\phi_1)\Delta\hat{\theta}_{12} - \sin(\phi_1)\Delta\hat{\theta}_{11}}{A_1} \\ & - \frac{-\cos(\phi_2)\Delta\hat{\theta}_{22} - \sin(\phi_2)\Delta\hat{\theta}_{21}}{A_2}\end{aligned}$$

The variance is obtained from $\operatorname{var}(\hat{\alpha}) = E[(\Delta\alpha)^2]$. Since $(\hat{\theta}_{11}, \hat{\theta}_{12})$ is uncorrelated with $(\hat{\theta}_{21}, \hat{\theta}_{22})$,

$$\begin{aligned}\operatorname{var}(\hat{\alpha}) \approx {} & E\left[\left(\frac{-\cos(\phi_1)\Delta\hat{\theta}_{12} - \sin(\phi_1)\Delta\hat{\theta}_{11}}{A_1}\right)^2\right] \\ & + E\left[\left(\frac{-\cos(\phi_2)\Delta\hat{\theta}_{22} - \sin(\phi_2)\Delta\hat{\theta}_{21}}{A_2}\right)\right] \\ = {} & \frac{1}{A_1^2}\Big[\cos^2(\phi_1)\operatorname{cov}(\hat{\theta}_{12}, \hat{\theta}_{12}) + \\ & \sin^2(\phi_1)\operatorname{cov}(\hat{\theta}_{11}, \hat{\theta}_{11}) + \\ & 2\cos(\phi_1)\sin(\phi_1)\operatorname{cov}(\hat{\theta}_{11}, \hat{\theta}_{12})\Big] + \\ & \frac{1}{A_2^2}\Big[\cos^2(\phi_2)\operatorname{cov}(\hat{\theta}_{22}, \hat{\theta}_{22}) + \\ & \sin^2(\phi_2)\operatorname{cov}(\hat{\theta}_{21}, \hat{\theta}_{21}) + \\ & 2\cos(\phi_2)\sin(\phi_2)\operatorname{cov}(\hat{\theta}_{21}, \hat{\theta}_{22})\Big] \\ = {} & \frac{2}{T}\left[\frac{\sigma^2}{A_1^2} + \frac{\sigma^2}{A_2^2}\right],\end{aligned}$$

where the approximations in (21) have been used at the last step.

6 Conclusion

This paper has demonstrated that the theory of dual approximation provides a natural way to solve BLUE problems, allowing even infinite dimensional problems (estimation from continuous-time functions) to be recast as finite-dimensional problems. The parameter estimation problems turns into a straightforward problem of matrix linear algebra and the covariance of the estimates can be readily computed using a matrix of inner products.

The method was applied to the important problem of phase estimation, using continuous-time observations.

Funding
Partial support for this research was provided by USU Space Dynamics Laboratory

Abbreviations
BLUE – best linear unbiased estimator
Text for this section...

Availability of data and materials
NA

Ethics approval and consent to participate
NA

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Text for this section...

Authors' contributions
T.K.M.: conceptualization, writing; R.C.: problem setting discussion, funding; J.H.G. discussion, writing.

Author details
Electrical and Computer Engineering Dept., Utah State University, Logan, UT, USA.

References

1. Kay, S.M.: Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall, Upper Saddle River, NJ (1993)
2. Luenberger, D.G.: Optimization by Vector Space Methods. Wiley, New York (1969)
3. Moon, T.K., Stirling, W.C.: Mathematical Methods and Algorithms for Signal Processing. Prentice-Hall, Upper Saddle River, NJ (2000)
4. So, H.C., Zhou, Z.: Two accurate phase-difference estimators for dual dual channel sine-wave model. EURASIP Journal on Advances in Signal Processing **122** (2013)
5. Handel, P., Host-Madsen, A.: Estimation of velocity and size of particles from two channel laser anemometry measurements. Measurement **21**(3), 113–123 (1997)
6. Ramos, P.M., da Silva, M.F., Serra, A.C.: Low frequency impedance measurement using sin-fitting. Measurement **35**(1), 89–96 (2004)
7. Vucijak, N.M., Saranovac, L.V.: A simple algorithm for the estimation of phase difference between two sinusoidal voltages. IEEE Transactions on Instrumentation and Measurement **59**(12), 3152–3158 (2010). doi:10.1109/TIM.2010.2047155
8. Handel, P.: Parameter estimation employing a dual-channel sine-wave model under a gaussian assumption. IEEE Trans. Instrumentation and Measurement **57**(8), 1661–1669 (2008)
9. So, H.C.: A comparative study of two discrete-time phase delay estimators. IEEE Transactions on Instrumentation and Measurement **54**(6), 2501–2504 (2005). doi:10.1109/TIM.2005.859143