Estimation of Symmetric, Positive Definite Matrices from Imperfect Measurements

Yixin Chen, *Student Member, IEEE*, John E. McInroy, *Member, IEEE*

This work was supported by the Ballistic Missile Defense Organization and Army Research Office under grants DAAG55-98-1-0007 and DAAD19-00-1-0153.

Yixin Chen is with the Department of Computer Science and Engineering, the Pennsylvania State University, University Park, PA 16801, USA. Research was performed when the author was with the Department of Electrical Engineering, University of Wyoming. E-mail:yixchen@cse.psu.edu.

John E. McInroy is with the Department of Electrical Engineering, University of Wyoming, Laramie, WY 82071, USA. E-mail:mcinroy@uwyo.edu.

Abstract

In a number of contexts relevant to control problems, including estimation of robot dynamics, covariance, and smart structure mass and stiffness matrices, we need to solve an over-determined set of linear equations $\mathbf{AX} \approx \mathbf{B}$ with the constraint that the matrix $\mathbf{X}$ be symmetric and positive definite. In the classical least squares method the measurements of $\mathbf{A}$ are assumed to be free of error, hence, all errors are confined to $\mathbf{B}$. Thus, the "optimal" solution is given by minimizing the optimization criterion $\|\mathbf{AX}-\mathbf{B}\|_F^2$. However, this assumption is often impractical. Sampling errors, modeling errors, and, sometimes, human errors bring inaccuracies to $\mathbf{A}$ as well. In this paper, we introduce a different optimization criterion, based on area, which takes the errors in both $\mathbf{A}$ and $\mathbf{B}$ into consideration. Under the condition that the data matrices $\mathbf{A}$ and $\mathbf{B}$ are full rank, which in practice is easy to satisfy, the analytic expression of the global optimizer is derived. A method to handle the case that $\mathbf{A}$ is full rank and $\mathbf{B}$ loses rank is also discussed. Experimental results indicate that the new approach is practical, and improves performance.

Keywords

Covariance estimation, mass estimation, stiffness estimation, symmetric positive definite matrix, educational testing problem, matrix modification problem.

I. Introduction

Estimation of symmetric positive definite matrices is required when solving a variety of control problems including robotic control, smart structure control, and intelligent control. In robotics, the mass-inertia matrix of a robotic system is in the symmetric positive definite class, and the accuracy of its estimate directly affects control performance [15][13][1]. Similarly, controlling vibrations and precise positions of "smart" structures often requires estimation of the structure's mass and stiffness matrices [11][12]; both are symmetric and positive definite. In intelligent control, control decisions are often made based on estimation of a covariance matrix [14][9][5], which is, of course, symmetric and positive definite. Estimation of symmetric positive definite matrices also appears, to a lesser extent, in fields outside control including the educational testing [2] and matrix modification problems [3]. Most of the above examples can be formulated directly or indirectly into finding an optimal solution of a set of linear equations

$$\mathbf{AX} \approx \mathbf{B} \tag{1}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$ are given, $\mathbf{X} \in \mathbb{P}$ is the fitting matrix, $\mathbb{P}$ is the set of symmetric and positive definite matrices with size $n \times n$. For example, the estimation of the joint space mass-inertia matrix of a flexure jointed hexapod (Stewart Platform) [1] and stiffness matrix directly fit into (1). The covariance matrix estimation problem and the matrix modification problem (with symmetric positive definite constraint) can be regarded as extracting a symmetric positive definite matrix ($\mathbf{C}^*$) from a symmetric but indefinite matrix ($\mathbf{C}$). Thus it can be formulated as solving $\mathbf{CX} \approx \mathbf{I}$ where $\mathbf{X} \in \mathbb{P}$, $\mathbf{I}$ being the identity matrix of size $n \times n$. The "optimal" (under certain criterion) $\mathbf{C}^*$ is given by $\mathbf{C}^* = \mathbf{X}^{-1}$.

There is a rich resource of prior work on this type of problem. Space limitations do not allow us to present a broad survey. Instead we try to emphasize some of the work that is most related to our work. Higham [4] finds an optimal symmetric estimate using the least squares approach (Symmetric Procrustes Problem). Although the positive definite constraint is not directly considered in his method, Higham shows that the estimate will be positive definite (semi-definite) if the data matrix $\mathbf{A}^T\mathbf{B} + \mathbf{B}^T\mathbf{A}$ is positive definite (semi-definite). If $\mathbf{A}^T\mathbf{B} + \mathbf{B}^T\mathbf{A}$ is indefinite, then nothing can be concluded about the definiteness of the estimate. Hu [6] presents a least squares based method to handle the positive definite constraint. In his method, the upper and lower bounds for each entry of the fitting matrix must be given explicitly as the constraint. A non-negative scalar is also introduced as a constraint, which measures the degree of positive definiteness. Using the least squares criterion, $\|\mathbf{AX} - \mathbf{B}\|_F^2$, the problem can also be cast as a semi-definite program [16] by specifying lower (and/or upper) bounds of the eigenvalues of $\mathbf{X}$.

Nevertheless, in many applications, there is a question of the suitability of the least squares criterion $\|\mathbf{AX} - \mathbf{B}\|_F^2$. In the classical least squares approach, the measurements $\mathbf{A}$ are supposed to be free of error, hence, all errors are restricted to $\mathbf{B}$. However, this assumption is frequently impractical. Sampling errors, modeling errors, and, sometimes, human errors bring inaccuracies to $\mathbf{A}$ as well. For example, in the estimation of a flexure jointed hexapod's joint space mass-inertia matrix [1], $\mathbf{A}$ and $\mathbf{B}$ contain the measurements of payload accelerations and base forces, respectively. As a result, sampling and instrument noises appear in both $\mathbf{A}$ and $\mathbf{B}$. Similar phenomenon happens in identifying a robot dynamic model [8]. Thus, it is natural for one to expect improved performance by employing a criterion that is capable of describing the errors occurring in both measurement

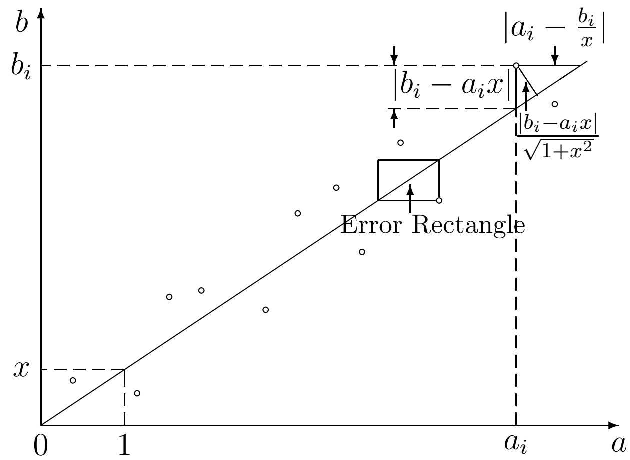

Fig. 1. Geometric interpretations of one parameter estimation using the least squares, the total least squares, and the new approaches.

matrices, rather than using the least squares criterion in which only the errors in $\mathbf{B}$ are considered.

In this paper, we present a new method of solving an over-determined set of linear equations (1) with $\mathbf{X}$ being symmetric positive definite, and both $\mathbf{A}$ and $\mathbf{B}$ containing errors.

II. Problem Formulation

A simple example will be more intuitive than a complex one for illustrating and understanding the motivation for the new optimization criterion. So let's consider the following problem with only one variable: estimating a single parameter from a set of over-determined equations

$$\vec{a}x \approx \vec{b}$$

where $\vec{a} = [a_1, a_2, \cdots, a_m]^T$, $\vec{b} = [b_1, b_2, \cdots, b_m]^T \in \mathbb{R}^m$ are known data vectors with $\vec{a}^T\vec{a} > 0$ and $\vec{b}^T\vec{b} > 0$, $x \in \mathbb{R}$ is the variable to be estimated. Using the classical least squares approach, the solution is the minimizer of the optimization criterion $(\vec{b} - \vec{a}x)^T(\vec{b} - \vec{a}x)$ or equivalently $\sum_{i=1}^{m} (b_i - a_i x)^2$. Geometrically, as shown in Figure 1, this criterion is the summation of the squared vertical "errors" (the distance from a data point (a_i, b_i) to the fitting line along the direction of b axis). This criterion is reasonable if the errors only occur in the data vector $\vec{b}$, because we are making predictions based on $\vec{a}$ that is free of error. If the errors are confined to $\vec{a}$, and $\vec{b}$ is free of error, the least squares approach is still appropriate, because we can minimize $(\vec{a}-\frac{\vec{b}}{x})^T(\vec{a}-\frac{\vec{b}}{x})$ or equivalently $\sum_{i=1}^{m}(a_i-\frac{b_i}{x})^2$, which will give the estimate of $\frac{1}{x}$. As shown in Figure 1, this time, the least squares solution minimizes the summation of the squared horizontal "errors" (the distance from a data point to the fitting line along the direction of a axis).

However, in many applications, both $\vec{a}$ and $\vec{b}$ are measurements containing errors. Under this scenario, a more appropriate approach of fitting is the total least squares method [7] (termed orthogonal regression or errors-in-variables regression in the statistical literature). For the above single parameter estimation problem, the total least squares solution minimizes $\sum_{i=1}^{m}\frac{(b_i-a_ix)^2}{(1+x^2)}$, which, as shown in Figure 1, is the summation of the squared minimum "errors" (the minimum distance from a data point to the fitting line) [1]. From the properties of the right triangle we can easily derive $\frac{(b_i-a_ix)^2}{(1+x^2)}=\frac{(b_i-a_ix)^2(a_i-\frac{b_i}{x})^2}{(b_i-a_ix)^2+(a_i-\frac{b_i}{x})^2}$, i.e., the minimum "error" contains the information of both the vertical "error" and the horizontal "error".

Motivated by above geometric interpretations of the least squares and the total least squares methods, we introduce a new optimization criterion, the area criterion, which is defined as the summation of the areas of the "error rectangles", i.e., $\sum_{i=1}^{m}|b_i-a_ix||a_i-\frac{b_i}{x}|$. As shown in Figure 1, the ith "error rectangle" is constructed by the ith vertical and ith horizontal "errors". Considering the symmetric and positive definite constraints (in this example, it implies $x>0$), the area criterion can be equivalently written as

$$\begin{aligned}\sum_{i=1}^{m}|b_i-a_ix|\left|a_i-\frac{b_i}{x}\right| &= \sum_{i=1}^{m}(a_ix-b_i)\left(a_i-\frac{b_i}{x}\right)\\ &= (\vec{a}y^2-\vec{b})^T\left(\vec{a}-\frac{\vec{b}}{y^2}\right)\\ &= \left\|\vec{a}y-\frac{\vec{b}}{y}\right\|_2^2\end{aligned}$$

where $y\in\mathbb{R}$, $y\neq 0$, $x=yy^T=y^2$. Note that we have transformed the positive constraint on x to the invertible constraint on y.

[1]To our knowledge, employing the symmetric and positive definite matrix constraints in the total least squares method is still an open problem.

Now let's consider the original problem given by (1). The area criterion is then extended as $Tr[(\mathbf{AX}-\mathbf{B})^T(\mathbf{A}-\mathbf{BX}^{-1})]$ where $\mathbf{AX}-\mathbf{B}$ represents the errors in $\mathbf{B}$ from the predictions based on $\mathbf{A}$, and $\mathbf{A}-\mathbf{BX}^{-1}$ represents the errors in $\mathbf{A}$ from the predictions based on $\mathbf{B}$. Using the properties of matrix calculus and the well known fact that $\mathbf{X} = \mathbf{YY}^T$ for any $\mathbf{X} \in \mathbb{P}$ where $\mathbf{Y} \in \mathbb{I}$, $\mathbb{I}$ being the set of real invertible matrices, the above extended area criterion can be equivalently written as $\|\mathbf{AY} - \mathbf{BY}^{-T}\|_F^2$. Thus, we can define an optimization problem as follows.

Definition II.1: (Symmetric Positive Definite Estimation problem, SPDE) For an over-determined set of m linear equations $\mathbf{AX} \approx \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ are given, $\mathbf{X} \in \mathbb{P}$ is the fitting matrix, let the area criterion, $f : \mathbb{I} \to \mathbb{R}$, be defined as

$$f(\mathbf{Y}) = \|\mathbf{AY} - \mathbf{BY}^{-T}\|_F^2 \tag{2}$$

with $\|\cdot\|_F$ being the Frobenius norm of a real matrix. The SPDE problem seeks to minimize the area criterion on $\mathbb{I}$. The symmetric positive definite estimate $\mathbf{X}^*$ is given by $\mathbf{X}^* = \mathbf{Y}^*\mathbf{Y}^{*T}$ where $\mathbf{Y}^*$ is a minimizer of (2).

III. Finding The Optimizer

To simplify derivations, we introduce two optimization criteria which are different to (2) by only a constant.

Lemma III.1: Let $g : \mathbb{I} \to \mathbb{R}$ and $h : \mathbb{P} \to \mathbb{R}$ be defined by

$$g(\mathbf{Y}) = Tr(\mathbf{Y}^T\mathbf{PY} + \mathbf{Y}^{-1}\mathbf{QY}^{-T}), \tag{3}$$

$$h(\mathbf{X}) = Tr(\mathbf{PX} + \mathbf{X}^{-1}\mathbf{Q}) \tag{4}$$

where $\mathbf{P} = \mathbf{A}^T\mathbf{A}$ and $\mathbf{Q} = \mathbf{B}^T\mathbf{B}$. Then minimizing $f(\mathbf{Y})$ on $\mathbb{I}$, minimizing $g(\mathbf{Y})$ on $\mathbb{I}$, and minimizing $h(\mathbf{X})$ on $\mathbb{P}$ are equivalent, i.e., $\mathbf{Y}^* \in \mathbb{I}$ minimizes $f(\mathbf{Y})$ if and only if $\mathbf{Y}^*$ minimizes $g(\mathbf{Y})$ if and only if $\mathbf{X}^* = \mathbf{Y}^*\mathbf{Y}^{*T} \in \mathbb{P}$ minimizes $h(\mathbf{X})$.

Proof: From the identities in matrix calculus, we have

$$f(\mathbf{Y}) = g(\mathbf{Y}) - 2\,Tr(\mathbf{A}^T\mathbf{B}) = h(\mathbf{X}) - 2\,Tr(\mathbf{A}^T\mathbf{B}) \tag{5}$$

where $\mathbf{X} = \mathbf{YY}^T$. □

In the following two theorems, we assume that $Rank(\mathbf{A}) = Rank(\mathbf{B}) = n$, i.e., $\mathbf{P}, \mathbf{Q} \in \mathbb{P}$. This assumption is easy to satisfy in most applications. At the end of this section, we will show that with only minor modification the results can be easily extended to the case that $\mathbf{A}$ is full rank and $\mathbf{B}$ loses rank.

Lemma III.1 implies that it is sufficient to derive the normal equation for one of the optimization criteria ($f(\mathbf{Y})$, $g(\mathbf{Y})$, or $h(\mathbf{X})$). We derive the normal equation for $g(\mathbf{Y})$ as follows.

Theorem III.2: Let $g(\mathbf{Y})$ be defined by (3). If $\mathbf{Y}^*$ is a minimizer of $g(\mathbf{Y})$, then it satisfies

$$\mathbf{Y}^*\mathbf{Y}^{*T}\mathbf{P}\mathbf{Y}^*\mathbf{Y}^{*T} = \mathbf{Q}. \tag{6}$$

Proof: Let $g' : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ be defined as

$$g'(\mathbf{Y}, \mathbf{Z}) = Tr(\mathbf{Y}^T\mathbf{P}\mathbf{Y} + \mathbf{Z}\mathbf{Q}\mathbf{Z}^T)$$

where $\mathbb{X}$ is the set of real $n \times n$ matrices. Then minimizing $g(\mathbf{Y})$ on $\mathbb{I}$ is equivalent to minimizing $g'(\mathbf{Y}, \mathbf{Z})$ on $\mathbb{X} \times \mathbb{X}$ with the constraint $\mathbf{YZ} = \mathbf{I}$ where $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix.

Let $\mathbf{Y}, \mathbf{Z}, \mathbf{\Psi} \in \mathbb{X}$, $\mathbf{Y} = [\vec{y}_1, \vec{y}_2, \cdots, \vec{y}_n]^T$, and $\mathbf{Z} = [\vec{z}_1, \vec{z}_2, \cdots, \vec{z}_n]$. Let y_{ij}, z_{ij}, and ψ_{ij} be the ijth entries of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{\Psi}$, respectively. The Lagrangian, $L : \mathbb{X} \times \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$, associated with the constraint $\mathbf{YZ} = \mathbf{I}$ is defined as

$$L(\mathbf{Y}, \mathbf{Z}, \mathbf{\Psi}) = Tr[\mathbf{Y}^T\mathbf{P}\mathbf{Y} + \mathbf{Z}\mathbf{Q}\mathbf{Z}^T + \mathbf{\Psi}(\mathbf{YZ} - \mathbf{I})].$$

Setting the partial derivatives of L with respect to y_{ij}, z_{ij}, and ψ_{ij} to 0's for all $1 \leq i, j \leq n$ gives,

$$2\mathbf{Y}^T\mathbf{P} + \mathbf{Z}\mathbf{\Psi} = \mathbf{0}, \tag{7}$$

$$2\mathbf{Q}\mathbf{Z}^T + \mathbf{\Psi}\mathbf{Y} = \mathbf{0}, \tag{8}$$

$$\mathbf{YZ} = \mathbf{I}. \tag{9}$$

Solving (7-9) for $\mathbf{Y}$ gives (6). $\square$

Theorem III.2 and Lemma III.1 imply two facts:

1. Any symmetric and positive definite estimate, $\mathbf{X}^*$, of the SPDE problem must satisfy

$$\mathbf{X}^*\mathbf{P}\mathbf{X}^* = \mathbf{Q} \tag{10}$$

where $\mathbf{X}^* = \mathbf{Y}^*\mathbf{Y}^{*T}$, $\mathbf{Y}^*$ is a solution of (6).

2. Any minimizer for (4) must also satisfy (10).

However, we still need to show that the solutions (or a solution) of (6) minimize(s) (3). From Lemma III.1 and the above facts, this is equivalent to verifying that the solutions (or a solution) of (10) minimize(s) (4), which is proven in the following theorem.

Theorem III.3: The unique minimizer of (4), which is the unique solution of (10), is given by

$$\mathbf{X}^* = \mathbf{U_P}\mathbf{\Sigma_P}^{-1}\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma_P}^{-1}\mathbf{U_P}^T \tag{11}$$

where

$$\mathbf{P} = \mathbf{U_P}\mathbf{\Sigma_P}^2\mathbf{U_P}^T, \tag{12}$$

$$\tilde{\mathbf{Q}} = \mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{Q}\mathbf{U_P}\mathbf{\Sigma_P} = \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}^2\mathbf{U}_{\tilde{\mathbf{Q}}}^T \tag{13}$$

are the Schur decomposition of $\mathbf{P}$ and $\tilde{\mathbf{Q}}$ respectively, and

$$\begin{aligned}\mathbf{\Sigma_P} &= diag[\sqrt{\lambda_{\mathbf{P}}^1}, \sqrt{\lambda_{\mathbf{P}}^2}, \cdots, \sqrt{\lambda_{\mathbf{P}}^n}],\\ \mathbf{\Sigma}_{\tilde{\mathbf{Q}}} &= diag[\sqrt{\lambda_{\tilde{\mathbf{Q}}}^1}, \sqrt{\lambda_{\tilde{\mathbf{Q}}}^1}, \cdots, \sqrt{\lambda_{\tilde{\mathbf{Q}}}^n}]\end{aligned}$$

where $\lambda_{\mathbf{P}}^i$'s and $\lambda_{\tilde{\mathbf{Q}}}^j$'s are eigenvalues of $\mathbf{P}$ and $\tilde{\mathbf{Q}}$, respectively.

Proof: Substituting (12) into (10) gives

$$\mathbf{X}^*\mathbf{U_P}\mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{U_P}\mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{X}^* = \mathbf{Q}. \tag{14}$$

Left multiplying both sides of (14) by $\mathbf{\Sigma_P}\mathbf{U_P}^T$, right multiplying both sides of (14) by $\mathbf{U_P}\mathbf{\Sigma_P}$, substituting (13) into (14), and collecting terms, we have

$$(\mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{X}^*\mathbf{U_P}\mathbf{\Sigma_P})^2 = (\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T)^2.$$

It is clear that $(\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T)^2 \in \mathbb{P}$, and $\mathbf{X}^* \in \mathbb{P}$ if and only if $\mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{X}^*\mathbf{U_P}\mathbf{\Sigma_P} \in \mathbb{P}$. Since a symmetric positive definite matrix has a unique symmetric positive definite square root, we have

$$\mathbf{\Sigma_P}\mathbf{U_P}^T\mathbf{X}^*\mathbf{U_P}\mathbf{\Sigma_P} = \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T. \tag{15}$$

Solving (15) gives (11).

Next we will show that the $\mathbf{X}^*$ given by (11) minimizes $h(\mathbf{X})$. Let the Schur decomposition of $\mathbf{X} \in \mathbb{P}$ be

$$\mathbf{X} = \mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^2\mathbf{U}_\mathbf{X}^T \tag{16}$$

where $\mathbf{\Sigma_X} = diag[\sqrt{\lambda_\mathbf{X}^1}, \sqrt{\lambda_\mathbf{X}^2}, \cdots, \sqrt{\lambda_\mathbf{X}^n}]$ with $\lambda_\mathbf{X}^i$ being the ith eigenvalue of $\mathbf{X}$. Equation (13) can be written as

$$\mathbf{Q} = \mathbf{U_P}\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T. \tag{17}$$

Substituting equations (12,16,17) into (4), we have

$$\begin{aligned}
h(\mathbf{X}) &= Tr(\mathbf{U_P}\mathbf{\Sigma}_\mathbf{P}^2\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^2\mathbf{U}_\mathbf{X}^T + \mathbf{U_P}\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{U}_{\tilde{\mathbf{Q}}} \\
&\quad \mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-2}\mathbf{U}_\mathbf{X}^T) \\
&= Tr[(\mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X})(\mathbf{\Sigma_X}\mathbf{U}_\mathbf{X}^T\mathbf{U_P}\mathbf{\Sigma_P}) + (\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T \\
&\quad \mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1})(\mathbf{\Sigma}_\mathbf{X}^{-1}\mathbf{U}_\mathbf{X}^T\mathbf{U_P}\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T)] \\
&= Tr[(\mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X} - \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1}) \\
&\quad (\mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X} - \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1})^T + \\
&\quad \mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X}\mathbf{\Sigma}_\mathbf{X}^{-1}\mathbf{U}_\mathbf{X}^T\mathbf{U_P}\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T + \\
&\quad \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1}\mathbf{\Sigma_X}\mathbf{U}_\mathbf{X}^T\mathbf{U_P}\mathbf{\Sigma_P}] \\
&= \|\mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X} - \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1}\|_F^2 \\
&\quad +2\,Tr(\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}).
\end{aligned} \tag{18}$$

It is clear that $h(\mathbf{X})$ achieves the global minimum when

$$\mathbf{\Sigma_P}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma_X} = \mathbf{U}_{\tilde{\mathbf{Q}}}\mathbf{\Sigma}_{\tilde{\mathbf{Q}}}\mathbf{U}_{\tilde{\mathbf{Q}}}^T\mathbf{\Sigma}_\mathbf{P}^{-1}\mathbf{U}_\mathbf{P}^T\mathbf{U_X}\mathbf{\Sigma}_\mathbf{X}^{-1}, \tag{19}$$

and $\mathbf{X}^*$ is the only solution to (19). □

Corollary III.4: The symmetric positive definite estimate, $\mathbf{X}^*$, of the SPDE problem is given by equation (11). The minimum of the area criterion, $f(\mathbf{Y})$, is $2\,Tr(\mathbf{\Sigma}_{\tilde{\mathbf{Q}}} - \mathbf{A}^T\mathbf{B})$.

Proof: It follows directly from the Definition II.1 and equations (5) and (18). □

Remark III.5: Actually, the set of linear equations (1) to be solved need not be overdetermined. All the above results still hold when $m = n$ provided that $Rank(\mathbf{A}) = Rank(\mathbf{B}) = n$.

Remark III.6: Theorem III.3 says that $h(\mathbf{X})$ has a unique minimizer on $\mathbb{P}$. But the minimizers of $f(\mathbf{Y})$ or $g(\mathbf{Y})$ on $\mathbb{I}$ are not unique. In fact, it is easy to show that if $\mathbf{Y}^* \in \mathbb{I}$ is a minimizer of $f(\mathbf{Y})$ or $g(\mathbf{Y})$ then $\mathbf{Y}^*\mathbf{U}$ is also a minimizer of $f(\mathbf{Y})$ and $g(\mathbf{Y})$ for any orthonormal matrix $\mathbf{U}$. Thus $f(\mathbf{Y})$ and $g(\mathbf{Y})$ have infinitely many minimizers on $\mathbb{I}$. Moreover, all these minimizers are related to the unique minimizer, $\mathbf{X}^*$, of $h(\mathbf{X})$ on $\mathbb{P}$ by $\mathbf{Y}^*\mathbf{Y}^{*T} = \mathbf{X}^*$. Consequently, the symmetric positive definite estimate of the SPDE problem is unique.

In the above discussions the data matrices $\mathbf{A}$ and $\mathbf{B}$ are assumed to be full rank. If either $\mathbf{A}$ or $\mathbf{B}$ lose rank, the method described above can not produce a symmetric positive definite optimizer. However, if $\mathbf{B}$ loses rank and $\mathbf{A}$ remains full rank, i.e., $\mathbf{P} \in \mathbb{P}$ and $\mathbf{Q} \in \overline{\mathbb{P}}$ (the set of symmetric positive semi-definite matrices), we can still find a positive semi-definite optimizer provided that $h(\mathbf{X})$ is optimized on $\overline{\mathbb{P}}_{Rank(\mathbf{Q})}$ (the set of symmetric positive semi-definite matrices with rank equal to the rank of $\mathbf{Q}$), and $\mathbf{X}^{-1}$ in $h(\mathbf{X})$ is replaced by $\mathbf{X}^+$ (Moore-Penrose pseudo-inverse of $\mathbf{X}$). The result is given as follows.

Corollary III.7: If $\mathbf{P} \in \mathbb{P}$, $\mathbf{Q} \in \overline{\mathbb{P}}$, $Rank(\mathbf{Q}) = r$, then the $\mathbf{X}^* \in \overline{\mathbb{P}}_r$ given by equation (11) minimizes the optimality criterion $\overline{h}(\mathbf{X}) = Tr(\mathbf{PX} + \mathbf{X}^+\mathbf{Q})$. The global minimum of $\overline{h}(\mathbf{X})$ on $\overline{\mathbb{P}}_r$ equals $2\,Tr(\mathbf{\Sigma}_{\tilde{\mathbf{Q}}})$.

Proof: The proof is similar to that of Theorem III.3. □

IV. Numerical Results

In this section, two numerical examples of estimating symmetric positive definite matrices are given. The least squares (LS) estimates [4], the total least squares (TLS) estimates [7], and the estimates using the new method (SPDE method) are compared.

The first example is the identification of the joint space mass-inertia matrix, $\mathbf{M}$, of a University of Wyoming (UW) flexure jointed hexapod [1]. In the vibration isolation control of the flexure jointed hexapod, the performance depends critically on the precision of the decoupling matrix which is calculated from the joint space mass-inertia matrix of the hexapod. Although $\mathbf{M}$ can be calculated from the design parameters of the hexapod, it is laborious to do so and can introduce errors due to manufacturing variances and payload changes. Thus a better approach is to estimate $\mathbf{M}$ from the measured payload accelerations and base forces. The relationship between payload accelerations and base

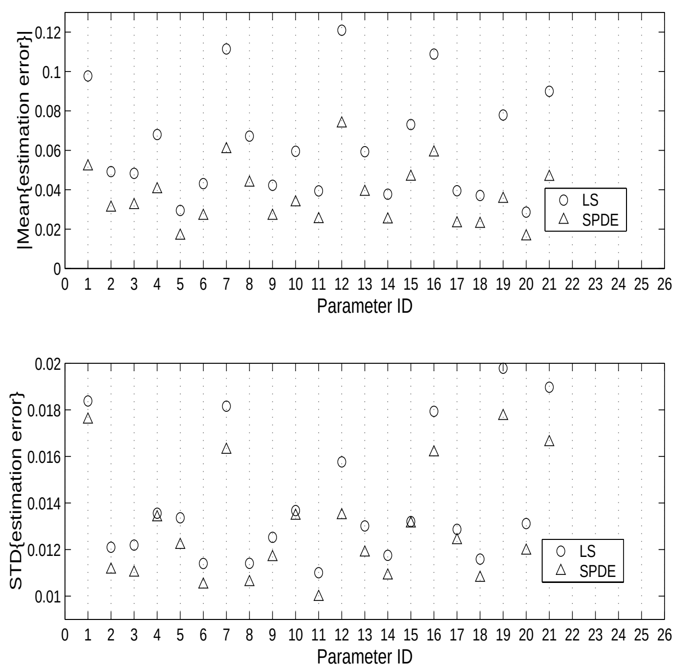

Fig. 2. Comparison of the LS and the SPDE methods: absolute mean and standard deviation of the estimation errors of $\mathbf{M}$'s entries. Mean{} and STD{} stand for the mean and the standard deviation, respectively.

forces is described as $\mathbf{AM} \approx \mathbf{B}$ where $\mathbf{A}$ contains the payload accelerations, $\mathbf{B}$ contains the base forces, and there are sampling and instrument noises in both $\mathbf{A}$ and $\mathbf{B}$. The matrix data given below is calculated from the real design parameters of the UW's flexure jointed hexapod.

$$\mathbf{M} = \begin{bmatrix} 4.688 & 0.198 & -0.404 & 1.798 & -0.405 & 0.611 \\ 0.198 & 4.688 & 0.611 & -0.405 & 1.798 & -0.404 \\ -0.404 & 0.611 & 4.688 & 0.198 & -0.404 & 1.798 \\ 1.798 & -0.405 & 0.198 & 4.688 & 0.611 & -0.404 \\ -0.405 & 1.798 & -0.404 & 0.611 & 4.688 & 0.198 \\ 0.611 & -0.404 & 1.798 & -0.404 & 0.198 & 4.688 \end{bmatrix}.$$

Six PCB load cells measure force and six Kistler accelerometers measure acceleration to provide the data. For both methods, 100 experiments were performed and the absolute mean and the standard deviation of the estimation errors for 21 independent parameters (since $\mathbf{M}$ is a 6×6 symmetric matrix) are shown in Figure 2. Compared with the LS method, the SPDE method provides more accurate estimates for all 21 parameters.

In the second numerical example, we are trying to estimate a 2×2 symmetric positive definite matrix, $\mathbf{X}$, from a set of linear equations $(\mathbf{A}+\mathbf{V}_a)\mathbf{X} \approx \mathbf{B}+\mathbf{V}_b$ where $\mathbf{A}+\mathbf{V}_a$ and $\mathbf{B}+\mathbf{V}_b$ are noise-corrupted data matrices,$\mathbf{V}_b$ contains normal distributed noises with 0 mean and standard deviation $\text{STD}\{v_b\} = 1$, $\mathbf{V}_a$ contains normal distributed noises with 0 mean and standard deviation, $\text{STD}\{v_a\}$, varying from 0 to 1.9 in the experiments, $\mathbf{X} = [x_{ij}] = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 5 & -3 & 1 & -1 \\ 2 & 1 & -2 & 1 \end{bmatrix}^T$, $\mathbf{B} = \begin{bmatrix} 17 & -8 & 1 & -2 \\ 7 & -2 & -1 & 0 \end{bmatrix}^T$. At each value of $\text{STD}\{v_a\}$, the absolute mean and the standard deviation of the estimation errors for x_{11}, x_{12}, and x_{22} are calculated for all three methods based on $10,000$ experiments [2].

As shown in Figure 3, the SPDE method outperforms the LS method significantly at large values of $\text{STD}\{v_a\}$. This is reasonable because the area criterion includes the information of both $\mathbf{V}_b$ and $\mathbf{V}_a$ while the least squares criterion only considers $\mathbf{V}_b$. For the same reason, we can't expect performance improvements when $\text{STD}\{v_a\}$ equals 0 or is much smaller than $\text{STD}\{v_b\}$, which is also verified by Figure 3. Compared with the TLS approach, the SPDE method also produces significantly more stable estimates when $\text{STD}\{v_a\}$ varies (note that in Figure 3 the estimation errors are displayed on a log scale for the TLS/SPDE comparision). We argue that this is because the positive definite constraint is not enforced (recall that employing symmetric and positive definite constraints in the TLS method is still an open problem).

V. Conclusions and Future Work

A new method (SPDE) of solving an over-determined set of linear equations $\mathbf{AX} \approx \mathbf{B}$ with a symmetric positive definite constraint and errors in both data matrices $\mathbf{A}$ and $\mathbf{B}$ is proposed. This type of problem arises in a number of contexts relevant to control problems, including estimation of mass-inertia and covariance matrices. The SPDE method transforms the original problem into an optimization problem seeking to minimize the so called area criterion. Compared with the least squares method, the new method improves the estimation accuracy because it takes errors in both $\mathbf{A}$ and $\mathbf{B}$ into consideration. In addition, no prior knowledge of the upper and lower bounds for the entries or eigenval-

[2]The TLS estimate is computed from the closed-form expression given by the Theorem 2.7 of [7]. The positive definite constraint is not taken into consideration.

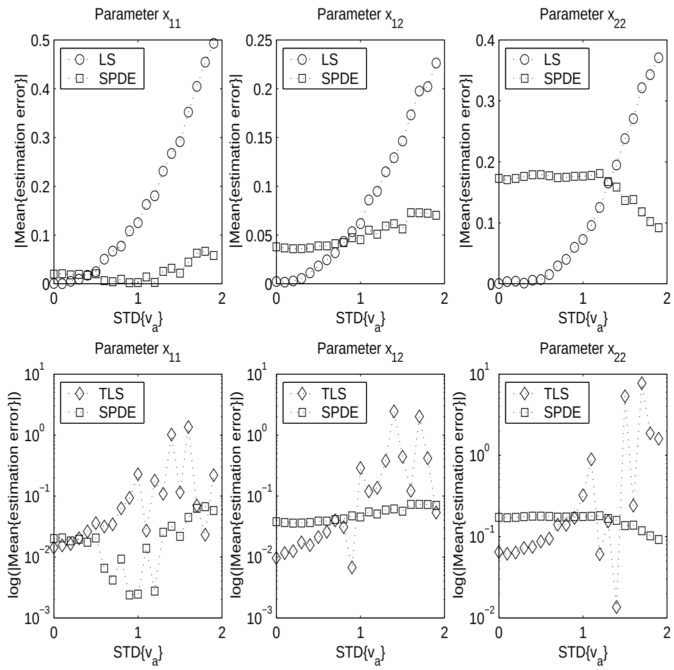

Fig. 3. Comparison of the LS, the TLS, and the SPDE methods: absolute mean and standard deviation of the estimation errors of **X**'s entries. Mean{} and STD{} stand for the mean and the standard deviation, respectively.

ues of the fitting matrix are needed for the new method, which makes the new method easy to apply. Experiment results demonstrate the superiority of the new algorithm. The statistical properties of the SPDE method are under further investigation.

VI. Acknowledgments

The authors would like to thank Dr. Siqi M. Fu for helpful technical discussions, and several anonymous reviewers for their comments that led to improvements in this paper.

References

[1] Y. Chen and J. E. McInroy, "Identification and decoupling control of flexure jointed hexapods," in *Proc. IEEE Int. Conf. Robotics and Automation*. San Francisco, CA USA, April 2000, pp. 1936–1941.

[2] R. Fletcher, "A nonlinear programming problem in statistics (educational testing)," *SIAM J. Scientific and Statistical Computing*, vol. 2, no. 3, pp. 257–267, September 1981.

[3] R. Fletcher, "Semi-definite matrix constraints in optimization," *SIAM J. Control and Optimization*, vol. 23, no. 4, pp. 493–513, 1985.

[4] N. J. Higham, "The symmetric procrustes problem," *BIT*, vol. 28, pp. 133–143, 1988.

[5] H. Hu and I Olkin, "A numerical procedure for finding the positive definite matrix closest to a patterned matrix," *Statistical and Probability Letters*, vol. 12, pp. 511–515, 1991.

[6] H. Hu, "Positive definite constrained least-squares estimation of matrices," *Linear Algebra and it's Applications*, vol. 229, pp. 167–174, 1995.

[7] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem Computational Aspects and Analysis*, SIAM Philadelphia, 1991.

[8] K. Kozlowski, *Modelling and Identification in Robotics*, Springer-Verlag, London, 1998.

[9] J. E. McInroy, J. C. Musto, and G. N. Saridis, *Reliable Plan Selection by Intelligent Machines*, World Scientific Publishing, River Edge, NJ, 1995.

[10] J. E. McInroy and J. C. Hamann, "Design and control of flexure jointed hexapods," *IEEE Trans. Robotics and Automation*, vol. 16, no. 4, pp. 372–381, August 2000.

[11] S.O.R. Moheimani and G.C. Goodwin, "Special issue on dynamics and control of smart structures," *IEEE Trans. on Control Systems Technology*, vol. 9, no. 1, January 2001.

[12] M. J. P. Musgrave, "On the constraints of positive-definite strain energy in anisotropic elastic media," *Quarterly Journal of Mechanics and Applied Mathematics*, vol. 43, no. 4, pp. 605–621, 1990.

[13] P. Poignet and M. Gautier, "Comparison of weighted least squares and extended kalman filtering methods for dynamic identification of robots," in *Proc. IEEE Int. Conf. Robotics and Automation*. San Francisco, CA USA, April 2000, pp. 3622–3627.

[14] J. C. Spall, "The information matrix in control: Computation and some applications," in *IEEE Int. Conf. Decision and Control*, December 1999, pp. 2367–2372.

[15] J. Swevers, C. Ganseman, D. Bilgin, J. D. Schutter, and H. V. Brussel, "Optimal robot excitation and identification," *IEEE Trans. Robotics and Automation*, vol. 13, no. 5, pp. 730–740, October 1997.

[16] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.