Mechanical Systems and Signal Processing 84 (2017) 462–468

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

# On computations of variance, covariance and correlation for interval data

Masako Kishida [1]

*National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*

ARTICLE INFO

*Article history:*
Received 7 April 2015
Received in revised form
4 February 2016
Accepted 25 July 2016

*Keywords:*
Complexity theory
Uncertainty modeling
Interval data
Skewed structured singular value

ABSTRACT

In many practical situations, the data on which statistical analysis is to be performed is only known with interval uncertainty. Different combinations of values from the interval data usually lead to different values of variance, covariance, and correlation. Hence, it is desirable to compute the endpoints of possible values of these statistics. This problem is, however, NP-hard in general.

This paper shows that the problem of computing the endpoints of possible values of these statistics can be rewritten as the problem of computing skewed structured singular values $\nu$, for which there exist feasible (polynomial-time) algorithms that compute reasonably tight bounds in most practical cases. This allows one to find tight intervals of the aforementioned statistics for interval data.



## 1. Introduction

This paper shows that solutions to some NP-hard problems with interval data can be explicitly expressed by using a skewed structured singular value $\nu$ [1]. Specifically, such a reformulation is considered for the problems of computing variance, covariance, and correlation for interval data. When data are known with interval uncertainty, each of these statistics cannot be expressed by a single value, and only an interval of possible values of the statistic can be obtained. In [2], it was proven that it is NP-hard to compute the upper endpoint of the exact interval of variance (the lower endpoint can be computed by a feasible algorithm), and both endpoints of the exact intervals corresponding to covariance and to correlation for interval data. In this paper, it is shown that all these computationally challenging endpoints can be expressed as skewed structured singular values, $\nu$ – with an exception of one of the lower and upper endpoints of the correlation with particular data sets, for which, polynomial time algorithm already exists [3].

Although in general, computation of $\nu$ is known to be NP-hard [4], there exist polynomial time algorithms that compute reasonably tight lower and upper bound on $\nu$ in most practical cases. These methods include power iterations and linear matrix inequalities [5–7]. $\nu$ has been used to analyze the effect of uncertainties in control communities [8], and toolboxes such as [9,10] are available for computing its lower and upper bounds.

The approach using $\nu$ also allows one to compute a guaranteed interval in which the endpoint lies by the lower and upper bounds on $\nu$.

*E-mail address:* kishida@nii.ac.jp

[1] The work was performed when the author was with University of Canterbury, Christchurch 8041, New Zealand, and Otto-von-Guericke-Universität Magdeburg, Institute for Automation Engineering, 39106 Magdeburg, Germany.

http://dx.doi.org/10.1016/j.ymssp.2016.07.037
0888-3270/

## 2. Mathematical preliminaries

This section provides definitions and lemmas needed to understand the later sections of this paper. Readers are referred to robust control textbooks (e.g., [8]) for a full understanding of the topic.

The set of real numbers and real matrices of size $n \times m$ are denoted by $\mathbb{R}$, and $\mathbb{R}^{n\times m}$, respectively. For a vector $v$, $v_i$ denotes the $i$-th element of $v$. $\mathbf{1}$ is a vector of ones. For a real matrix $M$, $M^T$ denotes the transpose of the matrix $M$. $\det(M)$ is the determinant of $M$. $\| M \|_\infty$ is the spectral norm of $M$, which is the largest singular value of $M$. $\operatorname{diag}[M_1, \ldots, M_n]$ is a block-diagonal matrix with $M_i$ on the diagonal. $I_n$ is an identity matrix of size $n$.

**Definition 2.1** (*Uncertain structures*). For a given matrix block structure $\mathcal{K} = (r_1, \ldots, r_{n_p})$, define the sets of matrices of size $r \times r$ for $r = \sum_{i=1}^{n_p} r_i$,

$$\mathbf{\Delta}_{\mathcal{K}} := \{\operatorname{diag}[\delta_1 I_{r_1}, \ldots, \delta_{n_p} I_{r_{n_p}}]:\ \delta_i \in \mathbb{R}\},$$
$$\mathbf{B\Delta}_{\mathcal{K}} := \{\Delta \in \mathbf{\Delta}_{\mathcal{K}}:\ \| \Delta \|_\infty \le 1\},\ \text{ and}$$
$$k\mathbf{B\Delta}_{\mathcal{K}} := \{\Delta \in \mathbf{\Delta}_{\mathcal{K}}:\ \| \Delta \|_\infty \le k\}.$$

Namely, $\mathcal{K}$ defines the sizes of the diagonal blocks, where each of the $n_p$ blocks of size $r_i$ is scaled by a real scalar $\delta_i$.

Let $\mathcal{K}_m^n = (m, \ldots, m)$, and $\mathcal{K}_1 = (1)$, i.e.,

$$\mathbf{B\Delta}_{\mathcal{K}_m^n} = \{\Delta = \operatorname{diag}[\delta_1 I_m, \cdots, \delta_n I_m]:\ \delta_i \in \mathbb{R},\ |\delta_i| \le 1\},$$
$$\mathbf{B\Delta}_{\mathcal{K}_1} = \{\Delta = \delta:\ \ \delta \in \mathbb{R},\ |\delta| \le 1\},$$

and similarly for $k\mathbf{B\Delta}_{\mathcal{K}_m^n}$ and $k\mathbf{B\Delta}_{\mathcal{K}_1}$.

**Definition 2.2** (*Skewed structured singular value $\nu$ [1]*). For a matrix $M \in \mathbb{R}^{m\times m}$, and two block structures $\mathcal{K}_p$ and $\mathcal{K}_q$, the skewed structured singular value $\nu$ is defined as

$$\nu_{\mathcal{K}_p,\mathcal{K}_q}(M) = \frac{1}{\min\left\{\begin{array}{c} k \ge 0:\ \Delta_p \in \mathbf{B\Delta}_{\mathcal{K}_p},\ \Delta_q \in k\mathbf{B\Delta}_{\mathcal{K}_q}, \\ \Delta = \operatorname{diag}[\Delta_p, \Delta_q],\ \text{ s. t. }\ \det(I - M\Delta) = 0 \end{array}\right\}}$$

unless no $k$ exists for $\Delta_q \in k\mathbf{B\Delta}_{\mathcal{K}_q}$ that makes $I - M\Delta$ singular, in which case $\nu_{\mathcal{K}_p,\mathcal{K}_q}(M) = 0$.

**Definition 2.3** (*Linear fractional transform [8]*). For matrices $M = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right]$ and $\Delta$ of compatible dimensions, the (upper) linear fractional transform (LFT) is

$$F_u(M, \Delta) := M_{22} + M_{21}\Delta\left(I - M_{11}\Delta\right)^{-1} M_{12},$$

provided that the inverse $\left(I - M_{11}\Delta\right)^{-1}$ exists. An LFT is said to be well-posed if $I - M_{11}\Delta$ is invertible.

Formulas for the summation, multiplication and inverse of LFT are found in [8]. For example, in the case of summation, the formula gives expressions for $N$ and $\Delta$ for given $M$, $Q$, $\Delta_1$ and $\Delta_2$ such that

$$F_u(M, \Delta_1) + F_u(Q, \Delta_2) = F_u(N, \Delta)$$

holds.

**Lemma 2.4** (*Scaled main loop theorem [11]*). *If an LFT $F_u(M, \Delta) \in \mathbb{R}^n$ is well-posed for all $\Delta \in \mathbf{B\Delta}_{\mathcal{K}}$ for a given block structure $\mathcal{K}$, then*

$$\sup_{\Delta \in \mathbf{B\Delta}_{\mathcal{K}}} |F_u(M, \Delta)| = \nu_{\mathcal{K},\mathcal{K}_1}(M).$$

**Lemma 2.5** (*[12,13]*). *For $I_n$ and $M \in \mathbb{R}^{m\times m}$, it holds that*

$$M \otimes I_n = S_{n,m}(I_n \otimes M) S_{n,m}^T,$$

*where*

$$S_{n,m} = \sum_{i=1}^{n} \left(e_i^T \otimes I_m \otimes e_i\right) = \sum_{j=1}^{m} \left(e_j \otimes I_n \otimes e_j^T\right),$$

*$\otimes$ denotes Kronecker product and $e_i$ denotes the $i$-th column of the identity matrix of appropriate dimension.*

*Notations*: Given interval ranges $[\underline{x}_i, \bar{x}_i]$ of values $x_1, \ldots, x_n$, let