Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 87 (2007) 2303–2312
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

The matrix-restricted total least-squares problem

Amir Beck

Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel

Received 13 September 2006; accepted 8 November 2006
Available online 11 December 2006

Abstract

We present and study the *matrix-restricted total least squares* (MRTLS) devised to solve linear systems of the form $\mathbf{Ax} \approx \mathbf{b}$ where $\mathbf{A}$ and $\mathbf{b}$ are both subjected to noise and $\mathbf{A}$ has errors of the form $\mathbf{DEC}$. $\mathbf{D}$ and $\mathbf{C}$ are known matrices and $\mathbf{E}$ is unknown. We show that the MRTLS problem amounts to solving a problem of minimizing a sum of fractional quadratic terms and a quadratic function and compare it to the related restricted TLS problem of Van Huffel and Zha [The restricted total least squares problem: formulation, algorithm, and properties, SIAM J. Matrix Anal. Appl. 12(2) (1991) 292–309.]. Finally, we present an algorithm for solving the MRTLS, which is based on a reduction to a single-variable minimization problem. This reduction is shown to have the ability of eliminating local optima points.

Keywords: Structured total least-squares; Restricted total least squares; Nonconvex optimization; Generalized trust region subproblem

1. Introduction

Consider an uncertain linear system:

$$(\mathbf{A} + \widetilde{\mathbf{E}})\mathbf{x} = \mathbf{b} + \mathbf{w}, \tag{1.1}$$

in which $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$ are known and $\widetilde{\mathbf{E}} \in \mathbb{R}^{m\times n}, \mathbf{w} \in \mathbb{R}^m$ are unknown. In the case when $\widetilde{\mathbf{E}}$ and $\mathbf{w}$ are not assumed to have any underlying structure, the total least-squares (TLS) approach to this problem [1–3] is to seek a perturbation matrix $\widetilde{\mathbf{E}}$ and a perturbation vector $\mathbf{w}$ that minimize $\|\widetilde{\mathbf{E}}\|^2 + \|\mathbf{w}\|^2$ subject to the consistency equation $(\mathbf{A} + \widetilde{\mathbf{E}})\mathbf{x} = \mathbf{b} + \mathbf{w}$. It is well known [1,3] that the TLS problem can be reduced to the following problem in the $\mathbf{x}$ variables:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\|\mathbf{Ax} - \mathbf{b}\|^2}{\|\mathbf{x}\|^2 + 1}, \tag{1.2}$$

and that its solution can be expressed via the singular value decomposition of the augmented matrix $(\mathbf{A}, \mathbf{b})$.

When the perturbation matrix $\widetilde{\mathbf{E}}$ possesses some linear structure, i.e., $\mathscr{L}(\tilde{\mathbf{E}}) = \mathbf{0}$ for some linear operator $\mathscr{L}$, the corresponding problem can be expressed as

$$\min_{\mathbf{E},\mathbf{w},\mathbf{x}} \{\|\tilde{\mathbf{E}}\|^2 + \|\mathbf{w}\|^2 : (\mathbf{A} + \widetilde{\mathbf{E}})\mathbf{x} = \mathbf{b} + \mathbf{w}, \mathscr{L}(\widetilde{\mathbf{E}}) = \mathbf{0}\},$$

which is known as the *structured total least-squares* (STLS) problem. In contrast to TLS problems for which efficient solution procedures are known, STLS problems often give rise to hard nonconvex problems and current algorithms for solving these type of problems are not guaranteed to find a global solution but rather a local solution or even just a stationary point [4–9].

There are several exceptions for this state of affairs. If $\mathbf{A}$ (and therefore also $\widetilde{\mathbf{E}}$) has a block

E-mail address: becka@ie.technion.ac.il.

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2006.11.004

circulant structure, then the corresponding STLS problem can be solved by decomposing the problem into several smaller TLS problems using the discrete Fourier transform [10]. Another tractable STLS problem arises when some of the columns of $\mathbf{A}$ are error free while the other are subjected to noise. This problem is called the *generalized TLS* (GTLS) problem or *mixed LS–TLS* problem and its solution can be obtained by computing a QR factorization of $\mathbf{A}$ and then solving a TLS problem of reduced dimension [11]. A more general problem is the *restricted TLS* problem introduced in [12]. Here, it is assumed that $(\widetilde{\mathbf{E}}, \mathbf{w}) = \mathbf{D}_1\mathbf{E}\mathbf{C}_1$, where $\mathbf{D}_1 \in \mathbb{R}^{m\times p}$ and $\mathbf{C}_1 \in \mathbb{R}^{l\times(n+1)}$ are known matrices and $\mathbf{E} \in \mathbb{R}^{p\times l}$ is unknown. As was mentioned in [12], by choosing the matrices $\mathbf{D}_1$ and $\mathbf{C}_1$ appropriately, the restricted TLS problem can handle any weighted least squares (LS), generalized LS, TLS, and GTLS problems. The restricted TLS problem can be solved using the restricted singular value decomposition [13].

In this paper we introduce and analyze the *matrix-restricted TLS* (MRTLS) problem in which the unknown matrix perturbation $\widetilde{\mathbf{E}}$ has the form $\tilde{\mathbf{E}} = \mathbf{DEC}$, where $\mathbf{D} \in \mathbb{R}^{m\times p}, \mathbf{E} \in \mathbb{R}^{p\times l}, \mathbf{C} \in \mathbb{R}^{l\times n}$. The STLS problem for this structure can therefore be expressed as

$$\text{(MRTLS):} \quad \min_{\mathbf{E},\mathbf{w},\mathbf{x}} \{\|\mathbf{E}\|^2 + \|\mathbf{w}\|^2 : (\mathbf{A} + \mathbf{DEC})\mathbf{x} = \mathbf{b} + \mathbf{w}\}. \tag{1.3}$$

This problem is of course related to the restricted TLS problem, however, while it is assumed in the restricted TLS problem that the augmented perturbation matrix $(\widetilde{\mathbf{E}}, \mathbf{w})$ has the "**DEC**" structure, in the MRTLS problem only the perturbation matrix $\widetilde{\mathbf{E}}$ is of this structure. This allows us to model *different* situations than those handled by the restricted TLS. For example, the choice

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{pmatrix}, \quad \mathbf{C} = \mathbf{I}_n, \tag{1.4}$$

corresponds to the situation in which the first m_1 rows of $\mathbf{A}$ are contaminated by noise, while the remainder $m - m_1$ rows are not subjected to noise; all the components of the right-hand side vector $\mathbf{b}$ are assumed to be noisy. In other words, $\mathbf{A}$ and $\mathbf{b}$ can be decomposed:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}, \tag{1.5}$$

with $\mathbf{A}_1 \in \mathbb{R}^{m_1\times n}, \mathbf{A}_2 \in \mathbb{R}^{(m-m_1)\times n}, \mathbf{b}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{b}_2 \in \mathbb{R}^{m-m_1}$, so that the linear problem suitable for the choice (1.4) is

$$\mathbf{A}_1\mathbf{x} \approx \mathbf{b}_1, \quad \mathbf{A}_2\mathbf{x} \approx \mathbf{b}_2,$$

where $\mathbf{A}_1, \mathbf{b}_1$ and $\mathbf{b}_2$ are subjected to noise and $\mathbf{A}_2$ is error free. This is evidently a mixture of LS and TLS problems. It is different from the "mixed LS–TLS" problem introduced in [11] in which part of the *columns* of $\mathbf{A}$ are subjected to noise. This model will be called the *horizontal mixed LS– TLS problem*.

We also note that for $\mathbf{D} = \mathbf{I}_m, \mathbf{C} = \mathbf{I}_n$, the MRTLS problem reduces to the standard TLS problem and that by choosing $\mathbf{D} = \mathbf{I}_m, \mathbf{C} = (\mathbf{0}_{n_1}, \mathbf{I}_{n-n_1})$, we recover the generalized TLS problem in which the first n_1 columns of $\mathbf{A}$ are error free while the rest $n - n_1$ are noisy. Another scenario in which the MRTLS structure is suitable is when the components of the perturbation matrix $\widetilde{\mathbf{E}}$ are correlated and there exist nonsingular (square) matrices $\mathbf{D}$ and $\mathbf{C}$ for which the components of $\mathbf{D}^{-1}\widetilde{\mathbf{E}}\mathbf{C}^{-1}$ are uncorrelated with equal variance.

The paper is organized as follows. In Section 2 we derive a simplified form of the MRTLS problem (1.3). We show that the MRTLS problem amounts to solving a problem of minimizing a *sum* of fractional quadratic terms and a single quadratic term. We then compare the derived MRTLS problem to the restricted TLS problem of [12], which is a simpler problem consisting of minimizing only a *single* fractional expression. Sufficient conditions for the existence of the MRTLS solution are derived in Section 3 in the case when $\mathbf{DD}^{\mathrm{T}}$ is a projection; this case includes the horizontal LS–TLS model. Finally, in Section 4.1 we describe an algorithm for solving the MRTLS problem. The procedure is based on a reduction of the multivariate problem into a one-dimensional optimization problem. We prove and illustrate that the process of passing to the one-dimensional problem can eliminate local optima points but cannot add any new ones. For the convenience of the reader, a detailed MATLAB implementation of our code is given.

Notation: For simplicity, instead of inf/sup we use min/max; however, this does not mean that we assume that the optimum is attained and/or finite. Vectors are denoted by boldface lowercase letters, e.g., $\mathbf{y}$, and matrices are denoted by boldface uppercase letters e.g., $\mathbf{A}$. For any symmetric matrix $\mathbf{A}$ and positive definite matrix $\mathbf{B}$, we denote the corresponding generalized minimum eigenvalue by

$\lambda_{\min}(\mathbf{A},\mathbf{B})$; the generalized minimum eigenvalue has several equivalent formulations:

$$\lambda_{\min}(\mathbf{A},\mathbf{B}) = \max\{\lambda : \mathbf{A}-\lambda\mathbf{B}\succeq \mathbf{0}\} = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} = \lambda_{\min}(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}),$$

where we use the notation $\mathbf{A}\succeq 0$ ($\mathbf{A}\succ 0$) for a positive semidefinite (positive definite) matrix $\mathbf{A}$. We follow the MATLAB convention and use ";" for adjoining scalars, vectors or matrices in a column. The identity matrix of size $m\times m$ is denoted by $\mathbf{I}_m$ and $\mathbf{0}_{k\times l}$ stands for the zero matrix of size $k\times l$. For a linear space S, we denote by $\mathscr{P}_S$ the orthogonal projection onto the space S.

2. The MRTLS problem

In order to analyze and solve the MRTLS problem, we find in Section 2.1 a simplified form of the MRTLS problem which is expressed only by the $\mathbf{x}$ variables. We then compare in Section 2.2 the structure of the derived form of the MRTLS problem to the one of the restricted TLS problem.

2.1. A (more) explicit form of the MRTLS problem

The next lemma simplifies the problem by eliminating the $\mathbf{E}$ and $\mathbf{w}$ variables. This is done by minimizing first with respect to the variables $\mathbf{E}$ and $\mathbf{w}$.

Lemma 2.1. *Let* $\mathbf{A}\in\mathbb{R}^{m\times n}, \mathbf{b}\in\mathbb{R}^m, \mathbf{D}\in\mathbb{R}^{m\times p}$ *and* $\mathbf{C}\in\mathbb{R}^{l\times n}$. *Then* $(\mathbf{E},\mathbf{w},\mathbf{x})\in\mathbb{R}^{p\times l}\times\mathbb{R}^m\times\mathbb{R}^n$ *is an optimal solution of* (1.3) *if and only if* $\mathbf{x}$ *is an optimal solution of*

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{(\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\mathbf{D}\mathbf{D}^{\mathrm{T}})^{-1}(\mathbf{A}\mathbf{x}-\mathbf{b})\}. \tag{2.1}$$

Proof. Note that for a fixed $\mathbf{x}$, problem (1.3) is a linearly constrained convex problem and therefore the KKT conditions in this case are necessary and sufficient [14, Proposition 3.4.1], and we conclude that $(\mathbf{E},\mathbf{w})$ is an optimal solution of (1.3) if and only if there exists $\boldsymbol{\lambda}\in\mathbb{R}^m$ such that

$$\mathbf{E}+\mathbf{D}^{\mathrm{T}}\boldsymbol{\lambda}\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}=\mathbf{0}, \tag{2.2}$$

$$\mathbf{w}-\boldsymbol{\lambda}=\mathbf{0}, \tag{2.3}$$

$$(\mathbf{A}+\mathbf{D}\mathbf{E}\mathbf{C})\mathbf{x}=\mathbf{b}+\boldsymbol{\lambda}. \tag{2.4}$$

Substituting (2.2) and (2.3) into (2.4) we obtain:

$$\mathbf{A}\mathbf{x}-\mathbf{b}=(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\mathbf{D}\mathbf{D}^{\mathrm{T}})\boldsymbol{\lambda},$$

which yields

$$\boldsymbol{\lambda}=(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\mathbf{D}\mathbf{D}^{\mathrm{T}})^{-1}(\mathbf{A}\mathbf{x}-\mathbf{b}). \tag{2.5}$$

Therefore,

$$\begin{aligned}\|\mathbf{E}\|^2+\|\mathbf{w}\|^2 &\overset{(2.3)}{=} \|\mathbf{E}\|^2+\|\boldsymbol{\lambda}\|^2\\ &\overset{(2.2)}{=} (\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{D}\mathbf{D}^{\mathrm{T}}\boldsymbol{\lambda}+\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}\\ &= \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\mathbf{D}\mathbf{D}^{\mathrm{T}})\boldsymbol{\lambda}\\ &= (\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\mathbf{D}\mathbf{D}^{\mathrm{T}})^{-1}\\ &\quad\times(\mathbf{A}\mathbf{x}-\mathbf{b}),\end{aligned}$$

so that problem (1.3) reduces to (2.1). □

In order to analyze our main problem (2.1), we will occasionally use an even more explicit expression for its objective function. To do so, consider the spectral decomposition of $\mathbf{D}\mathbf{D}^{\mathrm{T}}$:

$$\mathbf{D}\mathbf{D}^{\mathrm{T}}=\mathbf{U}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{U}, \tag{2.6}$$

where $\mathbf{U}\in\mathbb{R}^{m\times m}$ is an orthogonal matrix, $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_k,0,\ldots,0), \lambda_i>0$ and $k=\mathrm{rank}(\mathbf{D}\mathbf{D}^{\mathrm{T}})$. Such a decomposition exists by the symmetry of $\mathbf{D}\mathbf{D}^{\mathrm{T}}$. Using the decomposition (2.6), problem (2.1) becomes

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{(\widetilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}})^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\boldsymbol{\Lambda})^{-1}(\widetilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}})\},$$

where $\widetilde{\mathbf{A}}=\mathbf{U}^{\mathrm{T}}\mathbf{A}$ and $\tilde{\mathbf{b}}=\mathbf{U}^{\mathrm{T}}\mathbf{b}$. The latter problem can also be written as follows:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{(\mathbf{F}\mathbf{x}-\mathbf{g})^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x})\boldsymbol{\Lambda}_1)^{-1}(\mathbf{F}\mathbf{x}-\mathbf{g}) + \|\mathbf{P}\mathbf{x}-\mathbf{q}\|^2\}. \tag{2.7}$$

Here $\mathbf{F}\in\mathbb{R}^{k\times n}, \mathbf{P}\in\mathbb{R}^{(m-k)\times n}, \mathbf{g}\in\mathbb{R}^k, \mathbf{q}\in\mathbb{R}^{m-k}$ and $\boldsymbol{\Lambda}_1\in\mathbb{R}^{k\times k}$ are defined by the following identities:

$$\widetilde{\mathbf{A}}=\begin{pmatrix}\mathbf{F}\\ \mathbf{P}\end{pmatrix},\quad \tilde{\mathbf{b}}=\begin{pmatrix}\mathbf{g}\\ \mathbf{q}\end{pmatrix},$$

$$\boldsymbol{\Lambda}=\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}_{k\times(m-k)}\\ \mathbf{0}_{(m-k)\times k} & \mathbf{0}_{(m-k)\times(m-k)}\end{pmatrix}. \tag{2.8}$$

Problem (2.7) can also be written as a problem of minimizing a sum of fractional quadratic terms and a single quadratic term:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\sum_{i=1}^{k}\frac{(\mathbf{f}_i^{\mathrm{T}}\mathbf{x}-g_i)^2}{1+\lambda_i\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{x}}+\|\mathbf{P}\mathbf{x}-\mathbf{q}\|^2\right\}, \tag{2.9}$$

where $\mathbf{f}_1^{\mathrm{T}},\ldots,\mathbf{f}_k^{\mathrm{T}}$ are the rows of $\mathbf{F}$ and $g_1,\ldots,g_k$ are the components of $\mathbf{g}$.

Remark 2.1. A problem of a related structure to (2.9) was addressed in [15] where the Tikhonov regularization of the TLS (TRTLS) was considered. The TRTLS consists of minimizing a fractional quadratic function and a pure quadratic term:

$$\text{(TRTLS):} \ \min\left\{\frac{\|\mathbf{Ax}-\mathbf{b}\|^2}{\|\mathbf{x}\|^2+1}+\|\mathbf{Lx}\|^2\right\}. \tag{2.10}$$

While (2.9) is a reminiscent of the TRTLS problem, there are several major differences. First, problem (2.9) consists of minimizing a sum of fractional terms rather than a single fractional expression. Second, the denominators in (2.9) are not strictly convex functions since $\mathbf{C}$ is not necessarily of full row rank. Finally, the quadratic term in (2.9) is not homogenous as the one in the TRTLS problem. All these differences complicate the problem considerably.

2.2. Connection to the restricted TLS

As was mentioned in the Introduction, in the restricted TLS problem [12], both the matrix and right-hand side vector have the "restricted" structure. Namely, the corresponding optimization problem is

$$\min_{\mathbf{E},\mathbf{x}}\left\{\|\mathbf{E}\|^2:(\widehat{\mathbf{A}}+\mathbf{D}_1\mathbf{E}\mathbf{C}_1)\begin{pmatrix}\mathbf{x}\\-1\end{pmatrix}=\mathbf{0}\right\}, \tag{2.11}$$

where $\mathbf{D}_1\in\mathbb{R}^{m\times p}$ and $\mathbf{C}_1\in\mathbb{R}^{l\times(n+1)}$ (and therefore $\mathbf{E}$ is a $p\times l$ matrix). Here, $\widehat{\mathbf{A}}$ denotes the augmented matrix $(\mathbf{A},\mathbf{b})$. For the sake of simplicity we assume that $\mathbf{D}_1$ has full row rank. This assumption can be removed in the price of making a more subtle analysis. We also assume that $\mathbf{b}\notin\mathscr{R}(\mathbf{A})$; otherwise any solution to problem (2.11) will be of the form $(\mathbf{0},\mathbf{x}_0)$ where $\mathbf{x}_0$ is an arbitrary solution to the system $\mathbf{Ax}=\mathbf{b}$.

Similar to Lemma 2.1, we now derive a simpler optimization problem than (2.11) by eliminating the $\mathbf{E}$ variables.

Lemma 2.2. *Let* $\mathbf{A}\in\mathbb{R}^{m\times n},\mathbf{b}\in\mathbb{R}^m,\mathbf{D}_1\in\mathbb{R}^{m\times p}$ *and* $\mathbf{C}_1\in\mathbb{R}^{l\times(n+1)}$. *Assume that* $\mathbf{D}_1$ *has full row rank. Then* $(\mathbf{E},\mathbf{x})\in\mathbb{R}^{p\times l}\times\mathbb{R}^n$ *is an optimal solution of* (2.11) *if and only if* $\mathbf{x}$ *is an optimal solution of*

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{\tilde{\mathbf{x}}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}(\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}})^{-1}\widehat{\mathbf{A}}\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}}}:\tilde{\mathbf{x}}=(\mathbf{x};-1)\right\}, \tag{2.12}$$

where $\widehat{\mathbf{A}}=(\mathbf{A},\mathbf{b})$.

Proof. By denoting $\tilde{\mathbf{x}}=(\mathbf{x};-1)$, problem (2.11) becomes

$$\min_{\mathbf{E},\mathbf{x}}\{\|\mathbf{E}\|^2:(\widehat{\mathbf{A}}+\mathbf{D}_1\mathbf{E}\mathbf{C}_1)\tilde{\mathbf{x}}=\mathbf{0}\}. \tag{2.13}$$

We now proceed as in the proof of Lemma 2.1. By fixing $\mathbf{x}$, the problem becomes a linearly constrained problem in the variables $\mathbf{E}$ and thus $\mathbf{E}$ is an optimal solution of (2.13) if and only if there exists $\boldsymbol{\lambda}\in\mathbb{R}^m$ such that

$$\mathbf{E}+\mathbf{D}_1^{\mathrm{T}}\boldsymbol{\lambda}\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}=\mathbf{0},$$
$$\widehat{\mathbf{A}}\tilde{\mathbf{x}}+\mathbf{D}_1\mathbf{E}\mathbf{C}_1\tilde{\mathbf{x}}=\mathbf{0}.$$

Substituting the first equation into the second equation we obtain

$$\widehat{\mathbf{A}}\tilde{\mathbf{x}}-\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}}\boldsymbol{\lambda}\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}}=\mathbf{0}.$$

Since $\widehat{\mathbf{A}}\tilde{\mathbf{x}}\neq\mathbf{0}$ (by the assumption that $\mathbf{b}\notin\mathscr{R}(\mathbf{A})$), we conclude that $\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}}\neq 0$ and, by using the assumption that $\mathbf{D}_1$ has full row rank, we obtain

$$\boldsymbol{\lambda}=\frac{1}{\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}}}(\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}})^{-1}\widehat{\mathbf{A}}\tilde{\mathbf{x}},$$

so that the optimal value of problem (2.13) is equal to

$$\|\mathbf{E}\|^2=(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}}\boldsymbol{\lambda})(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}})=\frac{\tilde{\mathbf{x}}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}(\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}})^{-1}\widehat{\mathbf{A}}\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_1^{\mathrm{T}}\mathbf{C}_1\tilde{\mathbf{x}}},$$

and the result follows. □

Note that problem (2.12) is much simpler than the MRTLS problem (2.9). Indeed, the restricted TLS problem (2.12) has a form similar to the one of the TLS problem (1.2), i.e., it consists of minimizing a fractional quadratic function. Therefore, it is not surprising that the solution of this problem—similarly to the solution of the TLS problem—can be computed by using some kind of a generalization of the SVD [12]. In contrast, the MRTLS problem amounts to minimizing a *sum* of fractional quadratic functions and a quadratic term (2.9), which is a much more complicated structure. It seems impossible to solve this problem by simple means such as SVD-type methods.

3. Existence of the MRTLS solution when $\mathbf{DD}^{\mathrm{T}}$ is a projection

Problem (2.7) is bounded below by zero so that it must have an infimum. However, the infimum is not necessarily attained. An example with one fractional term can be found in Section 3 of [15]. In this section we derive a sufficient condition under which the minimum is guaranteed to exist. We will restrict the discussion to the case when $\mathbf{DD}^{\mathrm{T}}$ is a projection, i.e., a matrix whose eigenvalues are zero or one [16, Theorem 4.1]. In this case $\mathbf{\Lambda}_1 = \mathbf{I}$ and problem 2.7 reduces to

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{\|\mathbf{F}\mathbf{x}-\mathbf{g}\|^2}{1+\mathbf{x}^T\mathbf{C}^T\mathbf{C}\mathbf{x}}+\|\mathbf{P}\mathbf{x}-\mathbf{q}\|^2\right\}. \tag{3.1}$$

We note that in the horizontal mixed LS–TLS model, $\mathbf{D}\mathbf{D}^T$ is a projection by the definition of $\mathbf{D}$ (1.4). Therefore, the results of this section can be applied to the horizontal mixed LS–TLS problem which takes the form:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\underbrace{\frac{\|\mathbf{A}_1\mathbf{x}-\mathbf{b}_1\|^2}{\|\mathbf{x}\|^2+1}}_{\text{TLS term}}+\underbrace{\|\mathbf{A}_2\mathbf{x}-\mathbf{b}_2\|^2}_{\text{LS term}}\right\}, \tag{3.2}$$

where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{b}_1$ and $\mathbf{b}_2$ are defined by (1.5).

Theorem 3.1. *Consider problem* (3.1) *with* $\mathbf{F}\in\mathbb{R}^{k\times n}, \mathbf{P}\in\mathbb{R}^{(m-k)\times n}, \mathbf{g}\in\mathbb{R}^k$ *and* $\mathbf{q}\in\mathbb{R}^{m-k}$. *Assume that* $\text{Null}(\mathbf{C})\cap\text{Null}(\mathbf{P}) = \{0\}$. *Let* $\mathbf{N}$ *be a matrix whose columns form an orthonormal basis for the null space of* $\mathbf{P}$ *and let* $\mathbf{x}_0$ *be any solution to the system* $\mathbf{P}\mathbf{x} = \mathscr{P}_{\text{Im}(\mathbf{P})}(\mathbf{q})$. *Suppose that the following condition is satisfied*:

$$\lambda_{\min}(\mathbf{M}_1, \mathbf{M}_2) < \lambda_{\min}(\mathbf{N}^T\mathbf{F}^T\mathbf{F}\mathbf{N}, \mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}), \tag{3.3}$$

where

$$\mathbf{M}_1 = \begin{pmatrix} \mathbf{N}^T\mathbf{F}^T\mathbf{F}\mathbf{N} & \mathbf{N}^T\mathbf{F}(\mathbf{F}\mathbf{x}_0-\mathbf{g}) \\ (\mathbf{F}\mathbf{x}_0-\mathbf{g})^T\mathbf{F}^T\mathbf{N} & \|\mathbf{F}\mathbf{x}_0-\mathbf{g}\|^2 \end{pmatrix},$$

$$\mathbf{M}_2 = \begin{pmatrix} \mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N} & \mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 \\ \mathbf{x}_0^T\mathbf{C}^T\mathbf{C}\mathbf{N} & 1+\mathbf{x}_0^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 \end{pmatrix}.$$

Then the minimum of (3.1) *is attained.*

Remark 3.1. Before proceeding with the proof of the theorem, it is important to establish the positive definiteness of the matrices $\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}$ and $\mathbf{M}_2$ since otherwise the generalized minimum eigenvalues in (3.3) would not be well defined. The matrix $\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}$ is positive definite by the assumption that $\text{Null}(\mathbf{C})\cap\text{Null}(\mathbf{P}) = \{\mathbf{0}\}$. Moreover, for every $\mathbf{w}\in\mathbb{R}^r, t\in\mathbb{R}$ such that $(\mathbf{w}; t)\neq\mathbf{0}_{r+1}$ (here r is the dimension of the null space of $\mathbf{P}$), we have

$$\begin{aligned}
&(\mathbf{w}; t)^T\mathbf{M}_2(\mathbf{w}; t)^T \\
&\quad= (\mathbf{w}^T, t)\begin{pmatrix} \mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N} & \mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 \\ \mathbf{x}_0^T\mathbf{C}^T\mathbf{C}\mathbf{N} & 1+\mathbf{x}_0^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 \end{pmatrix}\begin{pmatrix} \mathbf{w} \\ t \end{pmatrix} \\
&\quad= \mathbf{w}^T\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}\mathbf{w} + 2\mathbf{w}^T\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 t \\
&\qquad+ \mathbf{x}_0^T\mathbf{C}^T\mathbf{C}\mathbf{x}_0 t^2 + t^2 \\
&\quad= \|\mathbf{C}\mathbf{N}\mathbf{w} + t\mathbf{x}_0\|^2 + t^2.
\end{aligned} \tag{3.4}$$

If $t\neq 0$ then (3.4) is positive. On the other hand, if $t = 0$ then $\mathbf{w}$ is nonzero and the expression $\|\mathbf{C}\mathbf{N}\mathbf{w} + t\mathbf{x}_0\|^2 + t^2 = \mathbf{w}^T\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}\mathbf{w}$ is positive by the positive definiteness property of $\mathbf{N}^T\mathbf{C}^T\mathbf{C}\mathbf{N}$. We thus conclude that $\mathbf{M}_2$ is positive definite.

Proof of Theorem 3.1. Consider the decomposition

$$\mathbf{q} = \mathscr{P}_{\text{Im}(\mathbf{P})}(\mathbf{q}) + \mathscr{P}_{\text{Null}(\mathbf{P}^T)}(\mathbf{q}).$$

Then by the orthogonality of $\mathscr{P}_{\text{Im}(\mathbf{P})}(\mathbf{q})$ and $\mathscr{P}_{\text{Null}(\mathbf{P}^T)}(\mathbf{q})$ we have

$$\|\mathbf{P}\mathbf{x}-\mathbf{q}\|^2 = \|\mathbf{P}\mathbf{x}-\mathscr{P}_{\text{Im}(\mathbf{P})}(\mathbf{q})\|^2 + \|\mathscr{P}_{\text{Null}(\mathbf{P}^T)}(\mathbf{q})\|^2.$$

Let $\mathbf{x}_0\in\mathbb{R}^n$ be, as defined in the premise of the theorem, an arbitrary solution of the consistent system $\mathbf{P}\mathbf{x} = \mathscr{P}_{\text{Im}(\mathbf{P})}(\mathbf{q})$. By making the change of variables $\mathbf{y} = \mathbf{x}-\mathbf{x}_0$ and omitting the constant term $\|\mathscr{P}_{\text{Null}(\mathbf{P}^T)}(\mathbf{q})\|^2$, problem (3.1) transforms to

$$\min_{\mathbf{y}\in\mathbb{R}^n}\left\{f(\mathbf{y}) \equiv \frac{\|\mathbf{F}\mathbf{y}+\mathbf{F}\mathbf{x}_0-\mathbf{g}\|^2}{1+(\mathbf{y}+\mathbf{x}_0)^T\mathbf{C}^T\mathbf{C}(\mathbf{y}+\mathbf{x}_0)}+\|\mathbf{P}\mathbf{y}\|^2\right\}. \tag{3.5}$$

The problem is bounded below by zero and thus has an infimum. In order to show the attainment of the infimum (i.e., existence of a minimum), let us assume in contradiction that the minimum is not attained. In that case there must exist a sequence $\{\mathbf{y}_n\}$ such that $\|\mathbf{y}_n\|\to\infty$ and $f(\mathbf{y}_n)\to f^*$ where f^* is the infimum of problem (3.5). The sequence $\{\mathbf{y}_n/\|\mathbf{y}_n\|\}$ is comprised of unit-norm vectors and therefore has a subsequence $\{\mathbf{y}_{n_k}/\|\mathbf{y}_{n_k}\|\}$ that converges to a unit-norm vector $\mathbf{d}\in\mathbb{R}^n$. Since $\{f(\mathbf{y}_{n_k})\}$ is a convergent sequence we have

$$\frac{f(\mathbf{y}_{n_k})}{\|\mathbf{y}_{n_k}\|^2}\to 0. \tag{3.6}$$

On the other hand, $f(\mathbf{y}) = f_1(\mathbf{y}) + f_2(\mathbf{y})$ where

$$f_1(\mathbf{y}) = \frac{\|\mathbf{F}\mathbf{y}+\mathbf{F}\mathbf{x}_0-\mathbf{g}\|^2}{1+(\mathbf{y}+\mathbf{x}_0)^T\mathbf{C}^T\mathbf{C}(\mathbf{y}+\mathbf{x}_0)},$$

$$f_2(\mathbf{y}) = \|\mathbf{P}\mathbf{y}\|^2.$$

Combining (3.6) with the nonnegativity of the functions f_1, f_2 we conclude that

$$\frac{f_1(\mathbf{y}_{n_k})}{\|\mathbf{y}_{n_k}\|^2}\to 0, \tag{3.7}$$

$$\frac{f_2(\mathbf{y}_{n_k})}{\|\mathbf{y}_{n_k}\|^2} \to 0. \tag{3.8}$$

The second limit (3.8) together with $\mathbf{y}_{n_k}/\|\mathbf{y}_{n_k}\| \to \mathbf{d}$ implies that $\|\mathbf{d}\|^2 = 1$ and $\|\mathbf{Pd}\|^2 = 0$ so that $\mathbf{d} \in \text{Null}(\mathbf{P})$; a direct consequence of the latter inclusion is that $\mathbf{d} = \mathbf{Nv}$ where $\mathbf{N}$ is defined in the premise of the theorem and $\mathbf{v} \in \mathbb{R}^r$ is nonzero (r being the dimension of the null space of $\mathbf{P}$). Therefore,

$$\begin{aligned}
f^* &= \lim_{k\to\infty} f(\mathbf{y}_{n_k}) \\
&\geqslant \lim_{k\to\infty} f_1(\mathbf{y}_{n_k}) \\
&= \lim_{k\to\infty} \frac{\|\mathbf{F}\mathbf{y}_{n_k} + \mathbf{F}\mathbf{x}_0 - \mathbf{g}\|^2}{1 + (\mathbf{y}_{n_k} + \mathbf{x}_0)^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}(\mathbf{y}_{n_k} + \mathbf{x}_0)} \\
&= \lim_{k\to\infty} \frac{\|\mathbf{F}\mathbf{y}_{n_k}/\|\mathbf{y}_{n_k}\| + (\mathbf{F}\mathbf{x}_0 - \mathbf{g})/\|\mathbf{y}_{n_k}\|\|^2}{1/\|\mathbf{y}_{n_k}\|^2 + ((\mathbf{y}_{n_k} + \mathbf{x}_0)/\|\mathbf{y}_{n_k}\|)^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}((\mathbf{y}_{n_k} + \mathbf{x}_0)/\|\mathbf{y}_{n_k}\|)} \\
&= \frac{\mathbf{d}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{d}}{\mathbf{d}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{d}} = \frac{\mathbf{v}^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{N}\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{N}\mathbf{v}} \\
&\geqslant \lambda_{\min}(\mathbf{N}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{N}, \mathbf{N}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{N}).
\end{aligned}$$

Now, combining the derived inequality $f^* \geqslant \lambda_{\min}(\mathbf{N}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{N}, \mathbf{N}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{N})$ with condition (3.3), we obtain

$$f^* > \lambda_{\min}(\mathbf{M}_1, \mathbf{M}_2). \tag{3.9}$$

On the other hand,

$$\begin{aligned}
f^* &= \min_{\mathbf{y}\in\mathbb{R}^n} f(\mathbf{y}) \\
&\leqslant \min_{\mathbf{y}\in\mathbb{R}^n} \{f(\mathbf{y}) : \mathbf{y} \in \text{Null}(\mathbf{P})\} = \min_{\mathbf{v}\in\mathbb{R}^r} \{f(\mathbf{Nv})\} \\
&= \min_{\mathbf{v}\in\mathbb{R}^r} \left\{ \frac{\|\mathbf{FNv} + \mathbf{F}\mathbf{x}_0 - \mathbf{g}\|^2}{1 + (\mathbf{Nv} + \mathbf{x}_0)^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}(\mathbf{Nv} + \mathbf{x}_0)} \right\} \\
&= \min_{\mathbf{v}\in\mathbb{R}^r, t\in\mathbb{R}} \left\{ \frac{\|\mathbf{FNv} + t\mathbf{F}\mathbf{x}_0 - t\mathbf{g}\|^2}{t^2 + (\mathbf{Nv} + t\mathbf{x}_0)^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}(\mathbf{Nv} + t\mathbf{x}_0)} : t \neq 0 \right\}.
\end{aligned} \tag{3.10}$$

We now claim that the value of problem (3.10) is equal to the value of

$$\min_{\mathbf{v}\in\mathbb{R}^r, t\in\mathbb{R}} \left\{ \frac{\|\mathbf{FNv} + t\mathbf{F}\mathbf{x}_0 - t\mathbf{g}\|^2}{t^2 + (\mathbf{Nv} + t\mathbf{x}_0)^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}(\mathbf{Nv} + t\mathbf{x}_0)} : (\mathbf{v}; t) \neq \mathbf{0}_{m-k+1} \right\}, \tag{3.11}$$

which, by the definition of generalized minimum eigenvalues, is equal to $\lambda_{\min}(\mathbf{M}_1, \mathbf{M}_2)$. To show the equality between the values of (3.10) and (3.11), suppose on the contrary that the value of the minimization problem (3.11) is strictly less than the value of (3.10). Then in that case $\lambda_{\min}(\mathbf{M}_1, \mathbf{M}_2)$, which is the optimal value of problem (3.11), is equal to

$$\lambda_{\min}(\mathbf{N}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{N}, \mathbf{N}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{N}) = \min_{\mathbf{v}\in\mathbb{R}^r} \left\{ \frac{\|\mathbf{FNv}\|^2}{(\mathbf{Nv})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}(\mathbf{Nv})} : \mathbf{v} \neq \mathbf{0}_r \right\},$$

which is a contradiction to (3.3). Therefore, we have shown that

$$f^* \leqslant \text{val}(3.11) = \lambda_{\min}(\mathbf{M}_1, \mathbf{M}_2).$$

However, this contradicts inequality (3.9), and we thus conclude that the solution of (3.1) is attained. □

Remark 3.2. Weak inequality is always satisfied in (3.3): the matrix in the right-hand side of (3.3) is a principal submatrix of the one in the left-hand side. Hence, by the interlacing theorem of eigenvalues [16, Theorem 7.8], weak inequality holds.

Remark 3.3. For the unstructured TLS problem ($\mathbf{D} = \mathbf{I}_m, \mathbf{C} = \mathbf{I}_n$), condition (3.3) reduces to the well-known attainability condition for the TLS problem: $\sigma_{\min}(\widehat{\mathbf{A}}) < \sigma_{min}(\mathbf{A})$, where $\widehat{\mathbf{A}}$ is the augmented matrix $(\mathbf{A}, \mathbf{b})$.

4. Solving the MRTLS problem one-dimensional solvers

In this section we present an algorithm for solving the MRTLS problem. The algorithm is based on converting the problem into a single-variable minimization problem and then invoking a one-dimensional solver. We show that the multivariate problem has at least as many local optima points as the one-dimensional problem and demonstrate that local optima points tend to vanish in the passage to the one-dimensional problem. The section ends with the presentation of a detailed MATLAB code for solving the MRTLS problem. We assume in this section that the matrix $\mathbf{A}$ is of full column rank.

4.1. Reduction to a single-variable optimization problem

In order to solve problem (2.1), we use a similar methodology to the one used in [15] and convert the problem into a problem of minimizing a single-variable function:

$$\min_{\alpha\geqslant 0} \mathscr{G}(\alpha), \tag{4.1}$$

where $\mathcal{G}(\alpha)$ is defined as

$$\mathcal{G}(\alpha) = \min_{\mathbf{x}\in\mathbb{R}^n}\{(\mathbf{Ax}-\mathbf{b})^{\mathrm{T}}(\mathbf{I}_m+\alpha\mathbf{DD}^{\mathrm{T}})^{-1}(\mathbf{Ax}-\mathbf{b}): \mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}=\alpha\}. \quad (4.2)$$

Calculating function values of $\mathcal{G}$ requires solving a minimization problem with a quadratic objective function and a quadratic equality constraint. This problem is nonconvex due to the nonconvexity of its equality constraint but nonetheless can be solved efficiently; for details see Section 4.2. It can be shown that the function $\mathcal{G}$ is in fact continuous on $[0,\infty]$; the proof of this result, which relies on sensitivity analysis for optimization problems, is behind the scope of this paper and is thus omitted. The function $\mathcal{G}$ might have several local optima but in practice we observed—through numerous random problems—that it is almost always an unimodal function, namely a function with a single local optimum (which is also the global optimum).

A theoretical justification for the latter empirical observation is given in Theorem 4.1 that shows that each local optimum of the one-dimensional problem (4.1) corresponds to at least one local optimum of the multivariate problem (2.1). Interestingly, the reverse claim does not hold true in general. Therefore, local optima points of the multivariate problem (2.1) might vanish in the transition to the one-dimensional problem (4.1).

Theorem 4.1. *Suppose that α_0 is a local optimum of the single-variable problem* (4.1) *and let* $\mathbf{x}$ *be an optimal solution of* (4.2) *with $\alpha=\alpha_0$. Then $\mathbf{x}$ is a local optimum solution of the MRTLS problem* (2.1).

Proof. Since α_0 is a local optimum solution of (4.1) it follows that there exists an interval $I=(\alpha_0-\delta,\alpha+\delta)$ such that $\alpha_0\in I$ and $\mathcal{G}(\alpha_0)\leqslant\mathcal{G}(\alpha)$ for every $\alpha\in I\cap[0,\infty)$. Now, as stated in the premise of the theorem, let $\mathbf{x}_0$ be an optimal solution of problem (4.2) with $\alpha=\alpha_0$. We will show that $\mathbf{x}_0$ is a local optimum solution of (2.1). Indeed, let $\mathbf{x}\in\mathbb{R}^n$ satisfy $\|\mathbf{x}-\mathbf{x}_0\|\leqslant\rho$ where $\rho=\min\{1,\delta/2\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})(\|\mathbf{x}_0\|+1)\}$. Then by the mean value theorem, we obtain that there exist $\lambda\in[0,1]$ such that

$$\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}-\mathbf{x}_0^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}_0 = 2(\mathbf{x}_0+\lambda(\mathbf{x}-\mathbf{x}_0))^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}}\mathbf{C})(\mathbf{x}-\mathbf{x}_0),$$

so that

$$\begin{aligned}|\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}-\mathbf{x}_0^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}_0| &\leqslant 2\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})\|\mathbf{x}_0+\lambda(\mathbf{x}-\mathbf{x}_0)\|\cdot\|\mathbf{x}-\mathbf{x}_0\|\\ &\leqslant 2\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})(\|\mathbf{x}_0\|+\|\mathbf{x}-\mathbf{x}_0\|)\cdot\|\mathbf{x}-\mathbf{x}_0\|\\ &\leqslant 2\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})(\|\mathbf{x}_0\|+1)\cdot\rho\\ &\leqslant\delta.\end{aligned}$$

Therefore, since $\mathbf{x}_0^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}_0=\alpha_0$, it follows that $\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}\in I\cap[0,\infty)$ and as a result

$$\mathcal{G}(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx})\geqslant\mathcal{G}(\alpha_0).$$

Finally, denoting the objective function in (2.1) by

$$g(\mathbf{y})\equiv(\mathbf{Ay}-\mathbf{b})^{\mathrm{T}}(\mathbf{I}_m+(\mathbf{y}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cy})\mathbf{DD}^{\mathrm{T}})^{-1}(\mathbf{Ay}-\mathbf{b}),$$

we obtain that

$$g(\mathbf{x})\geqslant\mathcal{G}(\mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx})\geqslant\mathcal{G}(\alpha_0)=\mathcal{G}(\mathbf{x}_0^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx}_0)=g(\mathbf{x}_0),$$

for every $\mathbf{x}$ satisfying $\|\mathbf{x}-\mathbf{x}_0\|\leqslant\rho$, proving the local optimality of $\mathbf{x}_0$. □

The result of Theorem 4.1 implies that the transition of the multivariate problem into a one-dimensional problem can be viewed as process of eliminating local optima points. We illustrate this attractive property by an example.

Example. We consider the horizontal mixed LS–TLS problem (3.2) with randomly chosen $\mathbf{A}_1,\mathbf{A}_2\in\mathbb{R}^{3\times2},\mathbf{b}_1,\mathbf{b}_2\in\mathbb{R}^3$. A mesh and contour plots of the objective function of (3.2) are plotted in Fig. 1. The global optimum of this two-dimensional function is attained at $\mathbf{v}=(0.443,1.173)$ (marked by a triangle) and it has an additional local optimum at $\mathbf{w}=(-0.214,-2.983)$ (marked by a square). However, the corresponding one-dimensional function $\mathcal{G}$, plotted in Fig. 2, has *only one local optimum point*[1] that is attained at $\|\mathbf{v}\|^2=1.572$. Note that the function $\mathcal{G}$ does not have a local solution at $\|\mathbf{w}\|^2=8.943$ so that the local optimum point $\mathbf{w}$ vanishes in the process of passing to the one-dimensional problem.

4.2. *Evaluating function values of $\mathcal{G}$*

Calculating function values of $\mathcal{G}$ amounts to solving a minimization problem of the form

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{\mathbf{x}^{\mathrm{T}}\mathbf{Bx}-2\mathbf{c}^{\mathrm{T}}\mathbf{x}+d:\mathbf{x}^{\mathrm{T}}\mathbf{Gx}=h\}, \quad (4.3)$$

[1]Which is also the global optimum.

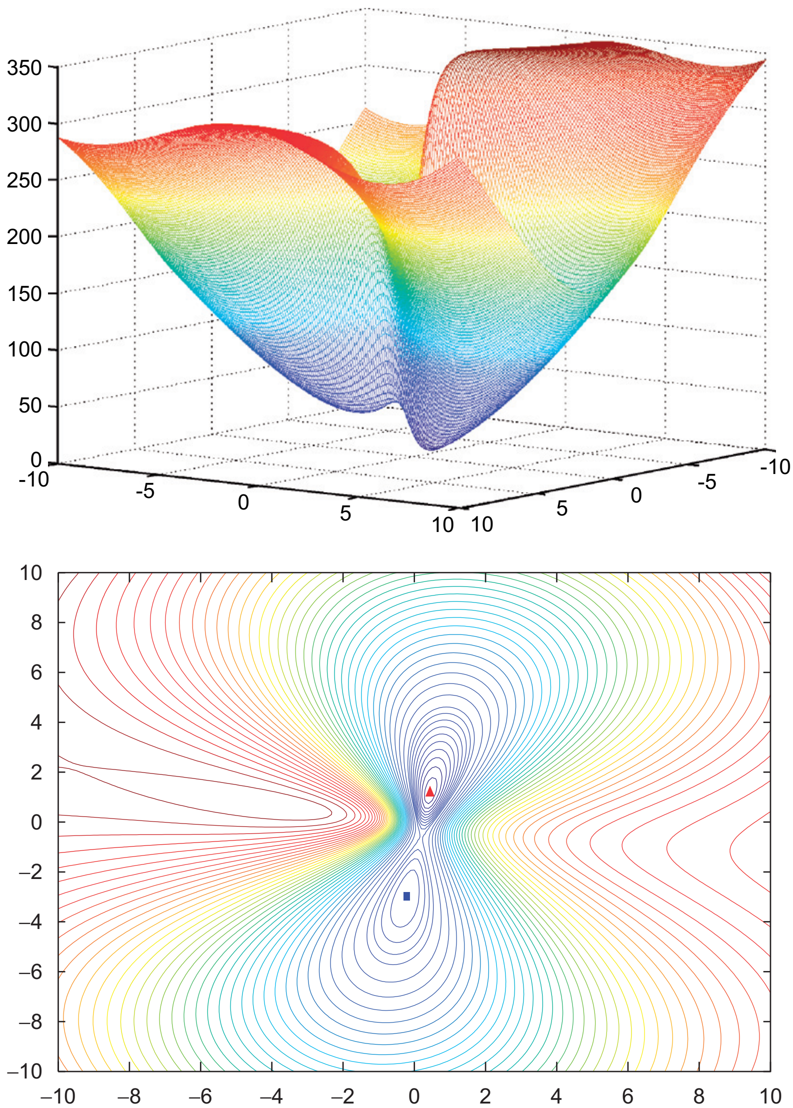

Fig. 1. Mesh and contour plots of the function $\|\mathbf{A}_1\mathbf{x}-\mathbf{b}_1\|^2/\|\mathbf{x}\|^2+1+\|\mathbf{A}_2\mathbf{x}-\mathbf{b}_2\|^2$.

where $\mathbf{B}\in\mathbb{R}^{n\times n}$ is a symmetric positive definite matrix,[2] $\mathbf{G}$ is a positive semidefinite matrix, $\mathbf{c}\in\mathbb{R}^n, d$ and h is a positive number. This is a special case of the *generalized trust region subproblem* (GTRS) [17] which consists of minimizing a general quadratic function (possible indefinite) subject to a general quadratic constraint. Under some mild condition, it is known that this class of problems possesses necessary and sufficient optimality conditions and that—as a result—the problem can be efficiently solved [17]. In particular, by Theorem 3.2 of [17], $\mathbf{x}$ is an optimal solution of (4.3) if and only if there exists $\lambda\in\mathbb{R}$ such that[3]

$$(\mathbf{B}-\lambda\mathbf{G})\mathbf{x}=\mathbf{c},$$
$$\mathbf{x}^{\mathrm{T}}\mathbf{G}\mathbf{x}=h,$$
$$\mathbf{B}-\lambda\mathbf{G}\succeq\mathbf{0}.$$

We will make a standard assumption that in fact $\mathbf{B}-\lambda\mathbf{G}\succ\mathbf{0}$ at the optimal λ. It follows directly from the optimality conditions that the optimal solution of (4.3) is given by $\mathbf{x}=(\mathbf{B}-\lambda\mathbf{G})^{-1}\mathbf{c}$ where λ is a

[2] $\mathbf{B}$ is positive definite (and not only positive semidefinite) since $\mathbf{A}$ is assumed to have full column rank.

[3] All the conditions of Theorem 3.2 of [17] are automatically satisfied for problem (4.3).

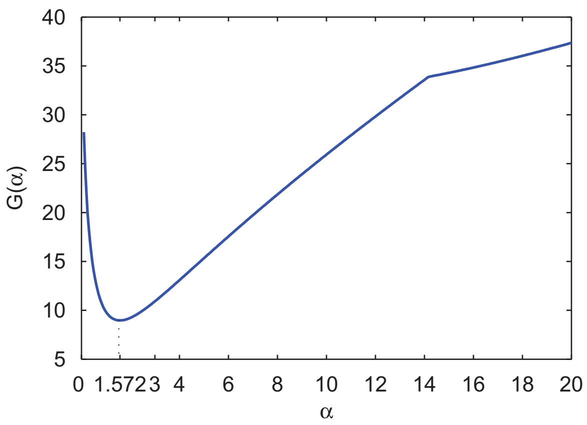

Fig. 2. A plot of the function $\mathscr{G}(\alpha) = \min\{\|\mathbf{A}_1\mathbf{x} - \mathbf{b}_1\|^2/(\alpha+1) + \|\mathbf{A}_2\mathbf{x} - \mathbf{b}_2\|^2 : \|\mathbf{x}\|^2 = \alpha\}$.

solution to the scalar equation

$$\phi(\lambda) = h, \quad \lambda < \frac{1}{\lambda_{\max}(\mathbf{G},\mathbf{B})},$$

where

$$\phi(\lambda) \equiv \mathbf{c}^{\mathrm{T}}(\mathbf{B} - \lambda\mathbf{G})^{-1}\mathbf{G}(\mathbf{B} - \lambda\mathbf{G})^{-1}\mathbf{c}.$$

In order to solve the problem, it was suggested in [18–20] to employ Newton's method on the equivalent problem of finding $\lambda < 1/\lambda_{\max}(\mathbf{G},\mathbf{B})$ satisfying $\phi^{-1/2}(\lambda) = h^{-1/2}$. It is well known [20] that in this problem Newton's method is guaranteed to have a *quadratic global convergence* if the initial point is chosen right to the optimal solution and left to the upper bound $1/\lambda_{\max}(\mathbf{G},\mathbf{B})$. We will now describe in detail the algorithm for solving (4.3). In our calculation we use the following formula for the derivative of ϕ:

$$\phi'(\lambda) = 2\mathbf{c}^{\mathrm{T}}(\mathbf{B} - \lambda\mathbf{G})^{-1}\mathbf{G}(\mathbf{B} - \lambda\mathbf{G})^{-1}\mathbf{G}(\mathbf{B} - \lambda\mathbf{G})^{-1}\mathbf{c}.$$

Algorithm FUNVAL

Input: $(\mathbf{B},\mathbf{c},d,\mathbf{G},h)$, where $\mathbf{B} \in \mathbb{R}^{n\times n}$ is a positive definite matrix, $\mathbf{G} \in \mathbb{R}^{n\times n}$ is a positive semidefinite matrix, $\mathbf{c} \in \mathbb{R}^n, d \in \mathbb{R}$ and $h > 0$.

Output: v- the optimal value of (4.3) (up to some tolerance ε_2).

Step 1: Calculate ρ - the generalized maximum eigenvalue of the matrix pair $(\mathbf{G},\mathbf{B})$.

Step 2: Set $\eta_0 = \frac{1}{\rho} - \varepsilon_1$ and $k = 0$.

Step 3: Repeat the following steps until $|\phi(\eta_k) - h| < \varepsilon_2$

Step 3.a Calculate a Cholesky factorization: $\mathbf{B} - \lambda\mathbf{G} = \mathbf{L}^{\mathrm{T}}\mathbf{L}$.

Step 3.b Set $\mathbf{y} = L^{-1}L^{-\mathrm{T}}\mathbf{c}$ and calculate $\phi(\eta_k) = \mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y}$.

Step 3.c Set $\mathbf{z} = L^{-\mathrm{T}}\mathbf{G}\mathbf{y}$ and calculate $\phi'(\eta_k) = 2\mathbf{z}^{\mathrm{T}}\mathbf{G}\mathbf{z}$.

Step 3.d Set

$$\eta_{k+1} = \eta_k + 2\frac{\phi^{-1/2}(\eta_k) - h^{-1/2}}{\phi^{-3/2}(\eta_k)\phi'(\eta_k)}.$$

Step 4. Set $v = \mathbf{y}^{\mathrm{T}}\mathbf{B}\mathbf{y} - 2\mathbf{c}^{\mathrm{T}}\mathbf{y} + d$.

In our implementation the tolerance parameters ε_1 and ε_2 take the values $\varepsilon_1 = 10^{-5}, \varepsilon_2 = 10^{-10}$. The Newton steps in the above algorithm converge to the optimal λ in very few iterations (usually not more than 6).

4.3. MATLAB implementation

We now present a MATLAB implementation of the function gfun that calculates function values of $\mathscr{G}$.

```
function val = gfun(alpha,A,b,DDT,CTC);
%   solves:
%   min (Ax-b)^T*(I+alpha*DDT)^{-1}*
%   (Ax-b)  s.t.  x^T*CTC*x = alpha
%   input:
%   alpha ............ real number
%   A ............ m*n matrix
%   assumes to have full column rank
%   b ............ m vector
%   DDT ............ an m*m matrix
%   that stands for D*D^T
%   CTC ............ an n*n matrix
%   that stands for C^T*C
%   output:
%   val ............ the optimal
%   value of the problem
[m,n] = size(A);
Dinv = inv(eye(m)+alpha*DDT);
B = A'*Dinv*A;
c = A'*Dinv*b;
d = b'*Dinv*b;
G = CTC;
G = (G+G')/2;
B = (B+B')/2;

opts.tol = 1e-6;
opts.disp = 0;
```

```
epsilon1 = 1e-5;
epsilon2 = 1e-10;

eta = 1/eigs(G,B,1,'la',opts)-epsilon1;
y = zeros(length(B),1);

while (abs(y'*G*y-alpha) > epsilon2)
  R = chol(B-eta*G);
  y = R\(R'\ c);
  z = R'\G*y);
  phi = y'*G*y;
  phid = 2*z'*z;
  eta = eta+2*(phi-phi^(1.5)/
sqrt(alpha))/phid;
end

val = y'*B*y-2*c'*y+d;
```

Now, in order to solve the MRTLS problem we can use any one-dimensional solver. In the example of Section 4.1 we used the MATLAB function fminbnd that finds a local optimum of a one-dimensional function over an interval. This MATLAB function is based on golden section search and parabolic interpolation and is guaranteed to find the global optimum when the function is unimodal. The MATLAB command is:

```
fminbnd(@(alpha)gfun(alpha,A,b,D*D',
C'*C),l,u,optimset('Display','iter'))
```

where l and u are lower and upper bounds on the value of $\mathbf{x}^T\mathbf{C}^T\mathbf{C}\mathbf{x}$ at an optimal solution $\mathbf{x}$.

References

[1] G.H. Golub, C.F. Van Loan, An analysis of the total least-squares problem, SIAM J. Numer. Anal. 17 (6) (December 1980) 883–893.

[2] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[3] S. Van Huffel, J. Vandewalle, The total least-squares problem: computational aspects and analysis, in: Frontier in Applied Mathematics, vol. 9, SIAM, Philadelphia, PA, 1991.

[4] J.B. Rosen, H. Park, J. Glick, Total least norm formulation and solution for structured problems, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 110–126.

[5] B. De Moor, Total least squares for affinely structured matrices and the noisy realization problem, IEEE Trans. Signal Process. 42 (11) (November 1994) 3104–3113.

[6] T.J. Abatzoglou, J.M. Mendel, G.A. Harada, The constrained total least squares technique and its applications to harmonic supperresolution, IEEE Trans. Signal Process. 39 (5) (May 1991) 1070–1087.

[7] B. De Moor, Structured total least squares and L_2 approximation problems, Linear Algebra Appl. 188-189 (1993) 163–207.

[8] N. Mastronardi, P. Lemmerling, S. Van Huffel, Fast structured total least squares algorithm for solving the basic deconvolution problem, SIAM J. Matrix Anal. Appl. 22 (2) (2000) 533–553.

[9] P. Lemmerling, S. Van Huffel, Analysis of the structured total least squares problem for Hankel/Toeplitz matrices, Numer. Algorithms 27 (2001) 89–114.

[10] A. Beck, A. Ben-Tal, A global solution for the structured total least squares problem with block circulant matrices, SIAM J. Matrix Anal. Appl. 27 (1) (2005) 238–255.

[11] S. Van Huffel, J. Vandewalle, Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error, SIAM J. Matrix Anal. Appl. 10 (3) (1989) 294–315.

[12] S. Van Huffel, H.Y. Zha, The restricted total least squares problem: formulation, algorithm, and properties, SIAM J. Matrix Anal. Appl. 12 (2) (1991) 292–309.

[13] H.Y. Zha, The restricted singular value decomposition of matrix triplets, SIAM J. Matrix Anal. Appl. 12 (1) (1991) 172–194.

[14] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, Belmont, MA, 1999.

[15] A. Beck, A. Ben-Tal, On the solution of the Tikhonov regularization of the total least squares problem, SIAM J. Optim. 17 (1) (2006) 98–118.

[16] F. Zhang, in: Basic results and techniques, Matrix Theory, Universitext, Springer, New York, 1999.

[17] J.J. Moré, Generalizations of the trust region subproblem, Optim. Method Softw. 2 (1993) 189–209.

[18] J.J. Moré, D.C. Sorensen, Computing a trust region step, SIAM J. Sci. Statist. Comput. 4 (3) (1983) 553–572.

[19] C. Fortin, H. Wolkowicz, The trust region subproblem and semidefinite programming, Optim. Method Softw. 19 (1) (2004) 41–67.

[20] A. Melman, A unifying convergence analysis of second-order methods for secular equations, Math. Comput. 66 (217) (1997) 333–344.