

# Solving a Type of the Tikhonov Regularization of the Total Least Squares by a New S-Lemma

Huu-Quang Nguyen<sup>1,2</sup>, Ruey-Lin Sheu<sup> $2(\boxtimes)$ </sup>, and Yong Xia<sup>3</sup>

<sup>1</sup> Department of Mathematics, Vinh University, Vinh, Vietnam quangdhv@gmail.com

<sup>2</sup> Department of Mathematics, National Cheng Kung University, Tainan, Taiwan rsheu@mail.ncku.edu.tw

<sup>3</sup> State Key Laboratory of Software Development Environment

School of Mathematics and System Sciences, Beihang University, Beijing, China

yxia@buaa.edu.cn

Abstract. We present a new S-lemma with two quadratic equalities and use it to minimize a special type of polynomials of degree 4. As a result, by the Dinkelbach approach with 2 SDP's (semidefinite programming), the minimum value and the minimum solution to the Tikhonov regularization of the total least squares problem with L = I can be nicely obtained.

**Keywords:** S-lemma with equality  $\cdot$  Tikhonov regularization  $\cdot$  Total least squares  $\cdot$  Dinkelbach method

## 1 Introduction

The well-known S-lemma due to Yakubovich [15] is a fundamental tool in control theory, optimization and robust analysis. Given two quadratic functions  $f(x) = x^T P x + 2p^T x + p_0$  and  $g(x) = x^T Q x + 2q^T x + q_0$  having symmetric matrices P and Q, the S-lemma asserts that, if  $g(x) \leq 0$  satisfies Slater's condition (i.e.,  $g(\bar{x}) < 0$  for some  $\bar{x}$ ), the following two statements are always equivalent  $((S_1) \sim (S_2))$ :

- (S<sub>1</sub>)  $(\forall x \in \mathbb{R}^n) g(x) \le 0 \implies f(x) \ge 0.$
- (S<sub>2</sub>) There exists a  $\lambda \ge 0$  such that  $f(x) + \lambda g(x) \ge 0, \ \forall x \in \mathbb{R}^n$ .

The S-lemma can be extended to deal with the equality g(x) = 0 along a series approaches, for example, please see [2,5,6,14]. They try to answer, for what pairs of (f(x), g(x)), the following two statements can become equivalent  $((E_1) \sim (E_2))$ :

(E<sub>1</sub>)  $(\forall x \in \mathbb{R}^n) g(x) = 0 \implies f(x) \ge 0.$ 

(E<sub>2</sub>) There exists a  $\lambda \in \mathbb{R}$  such that  $f(x) + \lambda g(x) \ge 0, \ \forall x \in \mathbb{R}^n$ .

© Springer Nature Switzerland AG 2020

H. A. Le Thi et al. (Eds.): WCGO 2019, AISC 991, pp. 221–227, 2020. https://doi.org/10.1007/978-3-030-21803-4\_23

222 H.-Q. Nguyen et al.

The complete necessary and sufficient conditions for the pair of quadratic functions (f(x), g(x)) under which  $(E_1) \sim (E_2)$  were established by Xia et. al. [13] with new applications to both quadratic optimization and the convexity of the joint numerical range. As a further extension, Wang and Xia [12] established the so-called S-lemma with interval bounds:

 $\begin{array}{l} (\mathrm{I}_1) \ (\forall x \in \mathbb{R}^n) \ (u_0 \leq g(x) \leq v_0 \Longrightarrow f(x) \geq 0); \\ (\mathrm{I}_2) \ \mathrm{There \ exists \ a \ } \lambda \in \mathbb{R} \ \mathrm{such \ that} \ f(x) + \lambda_+(g(x) - v_0) + \lambda_-(g(x) - u_0) \geq \\ 0, \ \forall x \in \mathbb{R}^n. \end{array}$ 

It has direct applications in the extended trust-region subproblem [7,11]. More importantly, it helps to guarantee the strong Lagrangian duality under the most mild assumptions [12]. Other extensions, such as the one by Polyak [8], focused on the system of three homogeneous quadratic forms. More discussions can be found in the survey paper [3].

In this paper, we obtain a new variant of the S-lemma. Given  $a, b, c \in \mathbb{R}^m$ and  $\Theta = \Theta^T = \Theta = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \in \mathbb{R}^{2 \times 2}, \theta \in \mathbb{R}^2, \gamma \in \mathbb{R}$ , this new version asks, when the following two statements can become equivalent  $((G_1) \sim (G_2))$ :

$$(\mathbf{G}_1) \quad (\forall x \in \mathbb{R}^n, z = (z_1, z_2)^T \in \mathbb{R}^2)$$

$$f(x) - z_1 = 0, g(x) - z_2 = 0, z_1a + z_2b \le c \Longrightarrow z^T \Theta z + \theta^T z - \gamma \ge 0;$$

(G<sub>2</sub>) There exist  $\alpha, \beta \in \mathbb{R}$  and  $\mu \in \mathbb{R}^m_+$  such that,  $\forall (z, x) \in \mathbb{R}^{n+2}$ ,

$$z^{T}\Theta z + \theta^{T}z - \gamma + \alpha(f(x) - z_{1}) + \beta(g(x) - z_{2}) + \mu^{T}(z_{1}a + z_{2}b - c) \ge 0.$$

Our main result is a sufficient condition for  $(G_1) \sim (G_2)$  as follows:

**Theorem 1.** Under the following assumptions

$$\exists \zeta, \eta \in \mathbb{R} : \ \zeta P + \eta Q \succ 0, \tag{1}$$

$$\Theta = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \succeq 0, \tag{2}$$

there is  $(G_1) \sim (G_2)$ .

Our most interest in this paper is to apply Theorem 1 to optimize a special class of polynomials of degree 4

$$\min_{x \in \mathbb{R}^n} G(x) = \theta_1 f(x)^2 + 2\theta_2 f(x)g(x) + \theta_3 g(x)^2 + \theta_4 f(x) + \theta_5 g(x)$$
(PoD4)

under the condition (2). Then, use the result from minimizing (PoD4) to solve a type of the Tikhonov regularization of the total least squares (TRTLS) proposed by Beck and Ben-Tal in [1]. The purpose of resolving (TRTLS) is to stabilize, via the Tikhonov regularization, the total least square solution for fitting an overdetermined linear system Ax = b. It was formulated in [1] as follows. Given

the regularization matrix  $L \in \mathbb{R}^{k \times n}$  and  $\rho > 0$  is a penalty parameter, consider the following problem

$$\min_{E,r,x} \{ \|E\|^2 + \|r\|^2 + \rho \|Lx\|^2 : (A+E)x = b+r \}$$
(TRTLS)

where  $E \in \mathbb{R}^{m \times n}$ , r and  $x \in \mathbb{R}^n$ . Then, (TRTLS) can be transformed to the following sum-of-ratios problem:

$$\min_{E,r,x} \{ \|E\|^2 + \|r\|^2 + \rho \|Lx\|^2 : (A+E)x = b+r \} 
= \min_{x} \left\{ \min_{E,r} \{ \|E\|^2 + \|r\|^2 + \rho \|Lx\|^2 : (A+E)x = b+r \} \right\} 
= \min_{x \in \mathbb{R}^n} \frac{\|Ax - b\|^2}{\|x\|^2 + 1} + \rho \|Lx\|^2$$
(3)

For L = I, Beck and Ben-Tal in [1] then used the Dinkelbach method [4] incorporating with the bisection search method to solve (3). We show, in Sect. 3, that (3) can be resolved by solving two SDP's, with one SDP to obtain its optimal value and the other one for the optimal solution. There is no need for any bisection method.

The remainder of this study is organized as follows: In Sect. 2, we provide the proof for Theorem 1 and solve Problem (PoD4). In Sect. 3, we use the Dinkelbach method incorporating two SDP's to solve (TRTLS) for the case L = I. Finally, we have a short discussion in Sect. 4 for future extensions.

#### 2 Proof for the New Version of the S-Lemma

The proof was done by using an important result by Polyak [8] that, under Condition (1), the joint numerical range (f(x), g(x)) is a convex subset in  $\mathbb{R}^2$ .

*Proof.*  $(G_1) \Longrightarrow (G_2)$ : By a result in [8, Theorem 2.2], the set

$$D_1 = \{ (z_1, z_2) \mid f(x) - z_1 = 0, g(x) - z_2 = 0, \ x \in \mathbb{R}^n \} \subset \mathbb{R}^2,$$
(4)

is convex. Let

$$D_2 = \{ (z_1, z_2) | \ z_1 a + z_2 b \le c \}.$$
(5)

and it is easy to see that  $D_2 \subset \mathbb{R}^2$  is also convex. Then, the statement (G<sub>1</sub>) can be recast as

$$(z_1, z_2) \in D_1 \cap D_2 \Rightarrow F(z) - \gamma = (z^T \Theta z + \theta^T z - \gamma) \ge 0.$$

Equivalently, it means that  $(D_1 \cap D_2) \cap \{(z_1, z_2) \mid F_{\gamma}(z_1, z_2) < 0\} = \emptyset$ . Due to Condition (2) that  $\Theta \succeq 0$ , the set  $\{(z_1, z_2) \mid F(z_1, z_2) - \gamma < 0\}$  is convex. Therefore, there exist  $\bar{\alpha}, \bar{\beta}$  such that  $\{(z_1, z_2) \mid \bar{\alpha}z_1 + \bar{\beta}z_2 + \bar{\gamma} = 0\}$  separates  $D_1 \cap D_2$  and  $\{(z_1, z_2) \mid F(z_1, z_2) - \gamma < 0\}$ . Without loss the generality, we assume that

$$\bar{\alpha}z_1 + \bar{\beta}z_2 + \bar{\gamma} \ge 0, \ \forall \ (z_1, z_2) \in D_1 \cap D_2, \tag{6}$$

224 H.-Q. Nguyen et al.

$$\bar{\alpha}z_1 + \bar{\beta}z_2 + \bar{\gamma} < 0, \ \forall \ (z_1, z_2) \in \{(z_1, z_2) \mid F(z_1, z_2) - \gamma < 0\}.$$
(7)

From (7), it implies that

$$\bar{\alpha}z_1 + \bar{\beta}z_2 + \bar{\gamma} \ge 0 \Rightarrow F(z_1, z_2) - \gamma \ge 0.$$

By S-lemma, there exists  $t \geq 0$  such that

$$F(z_1, z_2) - \gamma - t(\bar{\alpha}z_1 + \bar{\beta}z_2 + \bar{\gamma}) \ge 0, \ \forall \ (z_1, z_2) \in \mathbb{R}^2.$$

$$(8)$$

If t = 0, then with  $\alpha = \beta = 0$ ,  $\mu = 0$ , (G<sub>2</sub>) holds. If t > 0, by (6), the system

$$t\bar{\alpha}z_1 + t\bar{\beta}z_2 + t\bar{\gamma} < 0,$$
  

$$z_1a + z_2b - c \le 0,$$
  

$$(z_1, z_2) \in D_1$$

is not solvable. By Farkas theorem (see [9, Theorem 21.1], [10, Sect. 6.10 21.1], [6, Theorem 2.1]), there exists  $\mu \in \mathbb{R}^m_+$  such that

$$t\bar{\alpha}z_1 + t\bar{\beta}z_2 + t\bar{\gamma} + \mu^T(z_1a + z_2b - c) \ge 0, \ \forall \ (z_1, z_2) \in D_1.$$

Therefore, we have

$$t\bar{\alpha}f(x) + t\bar{\beta}g(x) + t\bar{\gamma} + \mu^T(f(x)a + g(x)b - c) \ge 0, \ \forall \ x \in \mathbb{R}^n.$$
(9)  
Let  $\alpha = \mu^T a + t\bar{\alpha}, \beta = \mu^T b + t\bar{\beta}.$  Then,

$$(9) \Leftrightarrow (\mu^{T}a + t\bar{\alpha})f(x) + (\mu^{T}b + t\bar{\beta})g(x) + t\bar{\gamma} - \mu^{T}c \ge 0$$
  
$$\Leftrightarrow \alpha f(x) + \beta g(x) + (\mu^{T}a + t\bar{\alpha} - \alpha)z_{1} + (\mu^{T}b + t\bar{\beta} - \beta)z_{2} + t\bar{\gamma} - \mu^{T}c \ge 0$$
  
$$\Leftrightarrow \alpha (f(x) - z_{1}) + \beta (g(x) - z_{2}) + \mu^{T}(z_{1}a + z_{2}b - c) \ge -t\bar{\alpha}z_{1} - t\bar{\beta}z_{2} - t\bar{\gamma}.$$
(10)

Combining (8) and (10), we get  $(G_2)$ .  $(G_2) \Longrightarrow (G_1)$ : It is trivial.

### 2.1 Optimizing a Class of Polynomials of Degree 4 (PoD4)

Applying Theorem 1, we can now solve the problem (PoD4) by solving the SDP (11) below under the assumption that f, g satisfy Condition (1) whereas  $\theta_1, \theta_2, \theta_3 \in \mathbb{R}$  satisfy condition (2).

$$\begin{split} \min_{x \in \mathbb{R}^n} G(x) &= \theta_1 f(x)^2 + 2\theta_2 f(x) g(x) + \theta_3 g(x)^2 + \theta_4 f(x) + \theta_5 g(x) \\ &= \min_{\{f(x) = z_1, \, g(x) = z_2\}} F(z_1, z_2) \\ &= \max\{\gamma \ : \ \{(z_1, z_2, x) \mid f(x) = z_1, \, g(x) = z_2, F(z_1, z_2) - \gamma < 0\} = \emptyset\} \\ &= \max\{\gamma \ : \ \{(f(x) = z_1, \, g(x) = z_2)\} \Rightarrow \{F(z_1, z_2) - \gamma \ge 0\}\} \\ &= \max_{\gamma, \, \alpha, \, \beta \in \mathbb{R}} \left\{\gamma \ : \ F(z_1, z_2) - \gamma + \alpha(f(x) - z_1) + \beta(g(x) - z_2) \ge 0\} \\ &= \max_{\gamma, \, \alpha, \, \beta \in \mathbb{R}} \left\{\gamma \ : \ \begin{pmatrix} \theta_1 \ \theta_2 \ \theta_3 \ \theta_2 \ \theta_3 \ \theta_5 - \beta \ \theta_2 \ \phi_5 - \beta \ \theta_2 \ \phi_5 - \beta \ \theta_2 \ \phi_5 - \beta \ \phi_7 \ \theta_7 \ \phi_7 + \beta q \ \phi_7 \$$

# 3 Dinkelbach Method for Solving (TRTLI)

It is interesting to see that problem (PoD4) allows us to solve the total least squares with Tikhonov identical regularization problem (see [1,16]) via solving two SDPs. Let us consider the following sum-of-quadratic-ratios problem.

$$\min_{x \in \mathbb{R}^n} \frac{\theta_1 f(x)^2 + \theta_4 f(x) + \theta}{g(x) + \gamma} + \theta_3 g(x) + 2\theta_2 f(x)$$

$$= \min_{x \in \mathbb{R}^n} \frac{\theta_1 f(x)^2 + 2\theta_2 f(x) g(x) + \theta_3 g(x)^2 + (\theta_4 + 2\gamma \theta_2) f(x) + \gamma \theta_3 g(x) + \theta}{g(x) + \gamma}$$

$$= \min_{x \in \mathbb{R}^n} \frac{h(x)}{l(x)} \tag{12}$$

where f, g are quadratic functions,  $\theta_1, \theta_2, \theta_3 \in \mathbb{R}$  satisfy the following condition:

Matrix 
$$\begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \succeq 0, Q \succ 0 \text{ and } \gamma > 0.$$
 (C 2)

In fact, the problem (12) covers the problem (TRTLSI) in [1,16] as a special case. With  $\gamma = 1, \theta = 0, \theta_1 = 0, \theta_2 = 0, \theta_3 = \rho, \theta_4 = 1, f(x) = ||Ax + b||^2, g(x) = ||x||^2$  then (12) reduces to (TRTLSI).

Notice that the form (12) is a single-ratio h(x)/l(x) fractional programming problem. It can be solved by the well-known Dinkelbach method [4]. To this end, define

$$\pi(t) = \min_{x \in \mathbb{R}^n} \{h(x) - tl(x)\} = \min_{x \in \mathbb{R}^n} \{\theta_1 f(x)^2 + 2\theta_2 f(x)g(x) + \theta_3 g(x)^2 + (\theta_4 + 2\gamma \theta_2)f(x) + (\gamma \theta_3 - t)g(x) + \theta - t\gamma\}.$$

It has been proved in [4] that  $\pi(t)$  is strictly decreasing and

$$\min_{x \in \mathbb{R}^n} \{ \frac{h(x)}{l(x)} \} = t^* \quad \text{if and only if} \quad \min_{x \in \mathbb{R}^n} \{ h(x) - t^* l(x) \} = \pi(t^*) = 0 \quad (13)$$

Since  $\pi(t)$  is strictly decreasing, then we conclude that  $t^*$  is maximum of all t such that  $\pi(t) \ge 0$ . Then, we can recast (12) to become

$$t^{*} = \max_{t \in \mathbb{R}} \{t : \pi(t) \ge 0\} = \max_{t \in \mathbb{R}} \{t : \min_{x \in \mathbb{R}^{n}} (h(x) - tl(x) \ge 0)\}$$
  

$$= \max_{t \in \mathbb{R}} \{t : h(x) - tl(x) \ge 0, \forall x \in \mathbb{R}^{n}\}$$
  

$$= \max_{t \in \mathbb{R}} \{t : \theta_{1}f(x)^{2} + 2\theta_{2}f(x)g(x) + \theta_{3}g(x)^{2} + (\theta_{4} + 2\gamma\theta_{2})f(x) + (\gamma\theta_{3} - t)g(x) + \theta - t\gamma \ge 0, \forall x \in \mathbb{R}^{n}\}$$
  

$$= \max_{t \in \mathbb{R}} \{t : \theta_{1}z_{1}^{2} + 2\theta_{2}z_{1}z_{2} + \theta_{3}z_{2}^{2} + (\theta_{4} + 2\gamma\theta_{2})z_{1} + (\gamma\theta_{3} - t)z_{2} + \theta - t\gamma \ge 0, (z_{1} = f(x), z_{2} = g(x))\}$$
  

$$= \max_{t, \alpha, \beta \in \mathbb{R}} \{t : \theta_{1}z_{1}^{2} + 2\theta_{2}z_{1}z_{2} + \theta_{3}z_{2}^{2} + (\theta_{4} + 2\varrho\theta_{2})z + (\gamma\theta_{3} - t)z_{2} + \theta - t\gamma + \alpha(f(x) - z_{1}) + \beta(g(x) - z_{2}) \ge 0\}$$
(14)

226 H.-Q. Nguyen et al.

where the last equation (14) is due to Theorem 1 by re-defining the notations as  $\theta_4 + 2\gamma\theta_2 := \theta_4, \gamma\theta_3 - t := \theta_5, \theta - t\gamma := -\gamma$ . Moreover, we can write (14) as the following SDP:

$$t^* = \begin{pmatrix} \theta_1 & \theta_2 & [0] & \frac{\theta_4 + 2\gamma\theta_2 - \alpha}{2} \\ \theta_2 & \theta_3 & [0] & \frac{\gamma\theta_3 - t - \beta}{2} \\ [0] & \alpha P + \beta Q & \alpha p + \beta q \\ \frac{\theta_4 + 2\gamma\theta_2 - \alpha}{2} & \frac{\gamma\theta_3 - t - \beta}{2} & \alpha p^T + \beta q^T & \xi \end{pmatrix} \succeq 0,$$
(15)

where  $\xi = \alpha p_0 + \beta q_0 + \theta - t\gamma$ . In other words, the optimal value  $t^*$  of (12), and thus the optimal value of the problem (TRTLSI), can be computed through solving the SDP (15).

After getting the optimal value  $t^*$  of (12) from (15), by (13), we can find the corresponding optimal solution  $x^*$  by solving the following problem

$$\min_{x \in \mathbb{P}^n} \{h(x) - t^* l(x)\}\tag{16}$$

where  $h(x) - t^* l(x) = \theta_1 f(x)^2 + 2\theta_2 f(x)g(x) + \theta_3 g(x)^2 + (\theta_4 + 2\gamma \theta_2)f(x) + (\gamma \theta_3 - t^*)g(x) + \theta - t^*\gamma$ . Since (16) is a special form of (PoD4), therefore we are able to get  $x^*$  by solving another SDP similar to (11).

## 4 Discussion

In this paper, we propose a set of sufficient conditions (1)–(2) under which  $(G_1) \sim (G_2)$ . It can be easily verified that, when  $m = 1, a = 1, b = c = \theta_1 = \ldots = \theta_4 = \gamma = 0, \theta_5 = 1, (G_1) \sim (G_2)$  reduces to  $(S_1) \sim (S_2)$  and we get the classical S-lemma. Similarly,  $(G_1) \sim (G_2)$  covers  $(I_1) \sim (I_2)$  with  $m = 2, a = (1,-1)^T$ ,  $b = (0,0)^T$ ,  $c = (v_0,-u_0)^T$ ,  $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \gamma = 0$  and  $\theta_5 = 1$ . Moreover, if we further have  $u_0 = v_0 = 0$ ,  $(G_1) \sim (G_2)$  becomes  $(E_1) \sim (E_2)$ . In other words, if the sufficient conditions (1)–(2) van be removed,  $(G_1) \sim (G_2)$  would be the most general results summarizing all previous results on S-lemma so far.

#### References

- 1. Beck, A., Ben-Tal, A.: On the solution of the Tikhonov regularization of the total least squares problem. SIAM J. Optim. **17**(1), 98–118 (2006)
- Beck, A., Eldar, Y.C.: Strong duality in nonconvex quadratic optimization with two quadratic constraint. SIAM J. Optim. 17(3), 844–860 (2006)
- Derinkuyu, K., Pınar, M.Ç.: On the S-procedure and some variants. Math. Methods Oper. Res. 64(1), 55–77 (2006)
- Dinkelbach, W.: On nonlinear fractional programming. Manag. Sci. 13, 492–498 (1967)
- Nguyen, V.B., Sheu, R.L., Xia, Y.: An SDP approach for quadratic fractional problems with a two-sided quadratic constraint. Optim. Methods Softw. **31**(4), 701–719 (2016)