Scientific
Research

A Theoretical Foundation for the Widely Linear Processing of Quaternion-Valued Data

Tohru Nitta

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
Email: tohru-nitta@aist.go.jp

ABSTRACT

In this paper, we will give a theoretical foundation for a quaternion-valued widely linear estimation framework. The estimation error obtained with the quaternion-valued widely linear estimation method is proved to be smaller than that obtained using the usual quaternion-valued linear estimation method.

Keywords: Augmented Statistics; Complex Number; Estimation; Quaternion

1. Introduction

The widely linear (WL) estimation method has been proved mathematically to be effective for estimation problems using complex-valued data. Estimation using WL uses complex conjugate parameters in addition to complex-valued parameters [1]. It has been applied to communications [2,3] and adaptive filters [4], together with so-called *augmented complex statistics*, a concept introduced by Picinbono *et al.* Moreover, an extension of the WL method to quaternion-valued case has been presented [5], which fully exploits available statistical information. A quaternion, a four-dimensional number invented by W. R. Hamilton in 1843, is an extension of a complex number. Let H be a four-dimensional vector space over R with an ordered basis, denoted by $1, i, j$ and k where R is a set of real numbers. Any quaternion $q \in H$ is expressed as $q = a + ib + jc + kd$ where $a, b, c, d \in R$. The three basis elements i, j, k satisfy the relations

$$i^2 = j^2 = k^2 = -1, \tag{1}$$

$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j. \tag{2}$$

Quaternion algebra has been used in fields such as robotics, computer vision, neural networks, signal processing, and communications (e.g. [6-8]).

In this paper, we present a theoretical foundation for quaternion-valued WL estimation: it is proved that the estimation error obtained using the quaternion-valued WL estimation method is smaller than that obtained using the usual quaternion-valued estimation method [9].

2. The WL Model

In this section, the complex-valued WL model and the quaternion-valued WL model are reviewed.

2.1. The Complex-Valued WL Model

Let $y \in C$ be a scalar complex-valued random variable to be estimated in terms of an observation that is a complex-valued random vector $x \in C^N$ where C is a set of complex numbers, and N is a natural number. That is, y is a true value and x is an observed value. In complex-valued linear mean square estimation (LMSE), the problem is to find an estimate written as

$$\hat{y}_L = h^H x, \tag{3}$$

where $h \in C^N$, and H represents the complex conjugation and transposition. Then, the objective of the problem is to find the parameter $h \in C^N$ that minimizes the estimation error $E|y - \hat{y}_L|^2$.

In the meantime, the complex-valued widely linear mean square estimation (WLMSE) problem can be stated as follows: Consider the scalar $\hat{y}$ defined as

$$\hat{y} = h^H x + g^H x^*, \tag{4}$$

where $g, h \in C^N$, and $v^* \overset{\text{def}}{=} a - bi$ is the complex conjugate of $v = a + bi \in C$. Then, the objective of the problem is to find parameters $g, h \in C^N$ that minimize the estimation error $E|y - \hat{y}|^2$.

Picinbono *et al.* gave a theoretical foundation for the complex-valued WLMSE described above: it was proved that the estimation error obtained using the complex-valued WLMSE method is smaller than that obtained using the usual complex-valued LMSE method: $E|y - \hat{y}_L|^2 \geq E|y - \hat{y}|^2$ where the equality holds only in exceptional cases [1].

Figures 1 and **2** demonstrate the effectiveness of complex-valued widely linear estimation. Samples of complex-valued random variable x and y are axisymmetric about the broken line: the white circle, the black circle, and the christcross in **Figure 1** correspond to the white circle, the black circle, and the christcross in **Figure 2**, respectively, and the three marks in **Figure 1** and the three marks in **Figure 2** are all axisymmetric about the broken line. It is therefore difficult to estimate the three samples in **Figure 1** by performing rotation, reduction, or expansion of the three samples in **Figure 2** about the origin. Nevertheless, inversion of the samples about the real axis, *i.e.*, assuming it as the samples of conjugate complex-valued random variable, allows to be estimated by rotation, reduction, and expansion about the origin. This is an intuitive explanation that the complex-valued widely linear estimation is superior to the conventional linear estimation method.
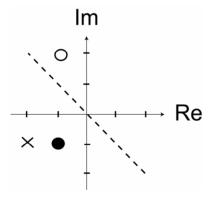

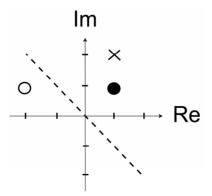**Figure 1. Three samples of a complex-valued random variable x.**

Figure 2. Three samples of a complex-valued random variable y.

2.2. The Quaternion-Valued WL Model

The quaternion-valued WL model is a natural extension of the complex-valued WL model described in Section 2.1. Let $y \in \boldsymbol{H}$ be a scalar quaternion-valued random variable to be estimated in terms of an observation that is a quaternion-valued random vector $\boldsymbol{x} \in \boldsymbol{H}^N$. That is, y is a true value and $\boldsymbol{x}$ is an observed value.

In quaternion-valued linear mean square estimation (LMSE), the problem is to find an estimate written as

$$\hat{y}_L = \boldsymbol{h}^H \boldsymbol{x}, \tag{5}$$

where $\boldsymbol{h} \in \boldsymbol{H}^N$, N is a natural number, and H represents the quaternionic conjugation and transposition.

The quaternion-valued widely linear mean square estimation (WLMSE) problem can be stated as follows: Consider the scalar $\hat{y}$ defined as

$$\hat{y} = \boldsymbol{h}^H \boldsymbol{x} + \boldsymbol{g}^H \boldsymbol{x}^*, \tag{6}$$

where $\boldsymbol{g}, \boldsymbol{h} \in \boldsymbol{H}^N$, N is a natural number, H represents the$_{\text{def}}$quaternionic conjugation and transposition, and $v^* = a - bi - cj - dk$ is the quaternionic conjugate of $v = a + bi + cj + dk \in \boldsymbol{H}$. Then, the objective of the problem is to find parameters $\boldsymbol{g}, \boldsymbol{h} \in \boldsymbol{H}^N$ that minimize $E|y - \hat{y}|^2$.

Took and Mandic derived an augmented quaternion least mean squares (AQLMS) algorithm for quaternion valued adaptive filters based on the quaternion-valued WL model, and confirmed its effectiveness via computer simulations [5]. Actually, the experimental results on the Lorenz attractor, real-world wind forecasting, and data fusion via quaternion spaces support the approach. Consequently, computer simulations proved that the quaternion-valued WL estimation method is superior to the usual quaternion-valued linear estimation method.

Moreover, the three perpendicular quaternion involutions can be introduced into the quaternion-valued WLMSE, which are given by

$$q^i = -iqi = a + ib - jc - kd, \tag{7}$$

$$q^j = -jqj = a - ib + jc - kd, \tag{8}$$

$$q^k = -kqk = a - ib - jc + kd, \tag{9}$$

where $q = a + ib + jc + kd \in \boldsymbol{H}$ [10-12]. The quaternion-valued WLMSE using (6) is an initial insight. However, in order to achieve the complete description of the second order statistics in $\boldsymbol{H}$, we need to consider the involutions ((7)-(9)). Actually, the quaternion-valued widely linear mean square estimation (WLMSE) problem using the involutions has been formulated as follows [12]: Consider the scalar $\hat{y}$ defined as

$$\hat{y} = \boldsymbol{h}^H \boldsymbol{x} + \boldsymbol{g}^H \boldsymbol{x}^i + \boldsymbol{u}^H \boldsymbol{x}^j + \boldsymbol{v}^H \boldsymbol{x}^k, \tag{10}$$

where $\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{H}^N$. Then, the objective of the prob-

lem is to find parameters $g, h, u, v \in H^N$ that minimize $E|y - \hat{y}|^2$.

However, no theoretical proof for the superiority of the quaternion-valued WL estimation method on estimation errors has been given to date, as it has been for the complex-valued WL estimation method. In the complex-valued setting, Picinbono *et al.* proved that the estimation error obtained with the complex-valued WLMSE is smaller than the error obtained using the complex-valued LMSE [1].

3. A Theoretical Foundation of the Quaternion-Valued WL Model

In this section, we show the superiority of the quaternion-valued WLMSE method. The main result is as follows: the estimation error obtained using the quaternion-valued WL estimation method based on (6) is smaller than that obtained using the usual quarternion-valued linear estimation method, except in rare cases. The result is obtainable in the same manner described by [1]. However, the noncommutativity of the quaternion products must be considered during the analysis ($xy \neq yx$ for any $x, y \in H$).

Define

$$X \overset{\text{def}}{=} \left\{ Z(\omega) = h^H x(\omega) + g^H x^*(\omega) \middle| g, h \in H^N \right\}.$$

Therein, X is a set of scalar quaternion-valued random variables that constitutes a linear space, and which becomes a Hilbert subspace of the one-dimensional quaternion-valued Hilbert space $Y = \{z(\omega) \in H\}$ with the scalar product $\langle z, \omega \rangle \overset{\text{def}}{=} E[z\omega^*]$ ($z, \omega \in X$). Then, for a true value $y \in Y$, an observed value $x \in H^N$, and the estimate $\hat{y} \in X$, the following equations hold:

$$(y - \hat{y}) \perp x, \tag{12}$$

$$(y - \hat{y}) \perp x^*, \tag{13}$$

where $u \perp v$ means that all the components of v are orthogonal to u with the scalar product $\langle \cdot, \cdot \rangle$ $(u \in H, v \in H^N)$. From (12) and (13), we obtain

$$E[xy^*] = E[x\hat{y}^*], \tag{14}$$

$$E[x^* y^*] = E[x^* \hat{y}^*]. \tag{15}$$

Consequently, from (6), (14), (15), the following equations hold (see the appendices for the detail of the derivation):

$$\Gamma_1 h + Cg = r, \tag{16}$$

$$C^H h + \Gamma_2 g = s^*, \tag{17}$$

where $\Gamma_1 \overset{\text{def}}{=} E[xx^H]$, $\Gamma_2 \overset{\text{def}}{=} E[x^* x^T]$,

$C \overset{\text{def}}{=} E[xx^T]$, $r \overset{\text{def}}{=} E[xy^*]$, $s \overset{\text{def}}{=} E[yx]$. Equations (16) and (17) yield

$$g = (\Gamma_2 - C^H \Gamma_1^{-1} C)^{-1} \cdot (s^* - C^H \Gamma_1^{-1} r), \tag{18}$$

$$h = (\Gamma_1 - C \Gamma_2^{-1} C^H)^{-1} \cdot (r - C \Gamma_2^{-1} s^*) \tag{19}$$

where we assume that Γ_1^{-1} Γ_2^{-1} and $(\Gamma_2 - C^H \Gamma_1^{-1} C)^{-1}$ exist (see the appendices for the detail of the derivation). Then, the estimation error ε_{WL} is calculated from (6), (16), and (17) as follows:

$$\varepsilon_{WL}^2 \overset{\text{def}}{=} E|y - \hat{y}|^2 = E|y|^2 - (r^H h + s^T g). \tag{20}$$

The estimation error ε_L in the quaternion-valued LMSE can be obtained using (5) as shown below.

$$\varepsilon_L^2 \overset{\text{def}}{=} E|y - \hat{y}_L|^2 = E|y|^2 - r^H \Gamma_1^{-1} r. \tag{21}$$

Then, from (16) and (18), (20), (21), the quantity $\delta\varepsilon^2$ can be expressed as

$$\delta\varepsilon^2 \overset{\text{def}}{=} \varepsilon_L^2 - \varepsilon_{WL}^2 = (s^* - C^H \Gamma_1^{-1} r)^H \\ \cdot (\Gamma_2 - C^H \Gamma_1^{-1} C)^{-1} \cdot (s^* - C^H \Gamma_1^{-1} r), \tag{22}$$

which is the difference between the estimation error of the quaternion-valued LMSE and that of the quarternion-valued WLMSE. Equation (22) is nonnegative because the matrix $\Gamma_2 - C^H \Gamma_1^{-1} C$ is positive-semidefinite (see the appendices for the proof). Furthermore, (22) is equal to zero only when one of the following conditions holds:

$$s^* - C^H \Gamma_1^{-1} r = \mathbf{0}, \tag{23}$$

$$\hat{y} = y. \tag{24}$$

Equation (23) is an exceptional case, and (24) means that the true value y can be estimated with probability of one, which is a rare case.

4. Discussions

In the previous section, a theoretical foundation of the quaternion-valued WL model based on (6) is given, which guarantees the superiority of the quaternion-valued WLMSE method. It is difficult to analyze the quaternion-value WLMSE method using (7)-(9) which contains all the necessary second order statistical information because the process of the analysis is very complicated due to the three involutions if we adopt the way of proof used in [1]. To solve the problem, another approach would be needed.

5. Conclusion

We have presented a theoretical foundation for the qua-

ternion-valued WL estimation method. It was proved that the estimation error obtained using the quaternion-valued WL estimation method is smaller than that obtained using the usual quaternion-valued linear estimation method, except in rare cases. In our future studies, we will proceed with analyses of the WL estimation based on the Clifford algebra [13].

6. Acknowledgements

The author extends special thanks to Prof. B. Picinbono, the Laboratoire des Signaux et Systèmes, Supélec, Gifsur Yvette, France for help in resolving several questions, and the anonymous reviewers for valuable comments.

REFERENCES

[1] B. Picinbono and P. Chevalier, "Widely Linear Estimation with Complex Data," *IEEE Transactions on Signal Processing*, Vol. 43, No. 8, 1995, pp. 2030-2033. http://dx.doi.org/10.1109/78.403373

[2] H. Gerstacker, R. Schober and A. Lampe, "Receivers with Widely Linear Processing for Frequency-Selective Channels," *IEEE Transactions on Communications*, Vol. 51, No. 9, 2003, pp. 1512-1523. http://dx.doi.org/10.1109/TCOMM.2003.816992

[3] R. Schober, W. H. Gerstacker and L. H.-J. Lampe, "Data-Aided and Blind Stochastic Gradient Algorithms for Widely Linear MMSE MAI Suppression for DS-CDMA," *IEEE Transactions on Signal Processing*, Vol. 52, No. 3, 2004, pp. 746-756. http://dx.doi.org/10.1109/TSP.2003.822359

[4] D. P. Mandic and V. S. L. Goh, "Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models," John Wiley and Sons Ltd., Hoboken, 2009. http://dx.doi.org/10.1002/9780470742624

[5] C. C. Took and D. P. Mandic, "The Quaternion LMS Algorithm for Adaptive Filtering of Hypercomplex Processes," *IEEE Transactions on Signal Processing*, Vol. 57,

No. 4, 2009, pp. 1316-1327. http://dx.doi.org/10.1109/TSP.2008.2010600

[6] T. Nitta, "A Quaternary Version of the Back-Propagation Algorithm," *Proceedings of the IEEE International Conference on Neural Networks* (ICNN'95), Perth, 27 November-1 December, 1995, pp. 2753-2756.

[7] P. Arena, L. Fortuna, G. Muscato and M. G. Xibilia, "Neural Networks in Multidimensional Domains," Springer, London, 1998.

[8] T. Isokawa, N. Matsui and H. Nishimura, "Quaternionic Neural Networks: Fundamental Properties and Applications," In: T. Nitta, Ed., *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters*, Information Science Reference (IGI Global), Hershey, New York, 2009, pp. 411-439. http://dx.doi.org/10.4018/978-1-60566-214-5.ch016

[9] T. Nitta, "Widely Linear Processing of Hypercomplex Signals," In: B.-L. Lu, L. Zhang and J. Kwok, Eds., *Neural Information Processing*, *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2011, pp. 519-525.

[10] C. C. Took and D. P. Mandic, "A Quaternion Widely Linear Adaptive Filter," *IEEE Transactions on Signal Processing*, Vol. 58, No. 8, 2010, pp. 4427-4431. http://dx.doi.org/10.1109/TSP.2010.2048323

[11] J. Via, D. Ramirez and I. Santamaria, "Properness and Widely Linear Processing of Quaternion Random Vectors," *IEEE Transactions on Information Theory*, Vol. 56, No. 7, 2010, pp. 3502-3515. http://dx.doi.org/10.1109/TIT.2010.2048440

[12] D. P. Mandic, C. Jahanchahi and C. C. Took, "A Quaternion Gradient Operator and its Applications," *IEEE Signal Processing Letters*, Vol. 18, No. 1, 2011, pp. 47-50. http://dx.doi.org/10.1109/LSP.2010.2091126

[13] K. Gürlebeck, K. Habetha and W. Sprößig, "Holomorphic Functions in the Plane and N-Dimensional Space," Birkhäuser, Basel, 2008.

Appendices

Proof of Equations (16) and (17):

$$0 = E\left[\boldsymbol{x}\left(y-\hat{y}\right)^{*}\right] \quad \left(\text{from Equation }(12)\right)$$
$$= E\left[\boldsymbol{x}y^{*}\right] - E\left[\boldsymbol{x}\hat{y}^{*}\right]. \tag{25}$$

So, we obtain

$$E\left[\boldsymbol{x}y^{*}\right] = E\left[\boldsymbol{x}\hat{y}^{*}\right]. \tag{26}$$

Similarly, from Equation (13) we obtain

$$E\left[\boldsymbol{x}^{*}y^{*}\right] = E\left[\boldsymbol{x}^{*}\hat{y}^{*}\right]. \tag{27}$$

Here,

$$\hat{y}^{*} = \left(\boldsymbol{h}^{H}\boldsymbol{x}+\boldsymbol{g}^{H}\boldsymbol{x}^{*}\right)^{*} \quad \left(\text{from Equation }(6)\right)$$
$$= \left(\boldsymbol{h}^{H}\boldsymbol{x}\right)^{*}+\left(\boldsymbol{g}^{H}\boldsymbol{x}^{*}\right)^{*} \tag{28}$$
$$= \boldsymbol{x}^{H}\boldsymbol{h}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{g}.$$

Note that $\boldsymbol{x}^{H}\boldsymbol{h} \neq \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}^{*}$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{g} \neq \boldsymbol{g}^{\mathrm{T}}\boldsymbol{x}$ due to the non-commutativity of the quaternion products. Thus, from Equations (26) and (28),

$$E\left[\boldsymbol{x}y^{*}\right] = E\left[\boldsymbol{x}\left(\boldsymbol{x}^{H}\boldsymbol{h}\right)\right] + E\left[\boldsymbol{x}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{g}\right)\right]$$
$$= E\left[\left(\boldsymbol{x}\boldsymbol{x}^{H}\right)\boldsymbol{h}\right] + E\left[\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right)\boldsymbol{g}\right] \tag{29}$$
$$= E\left[\boldsymbol{x}\boldsymbol{x}^{H}\right]\boldsymbol{h} + E\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right]\boldsymbol{g},$$

which means that Equation (16) holds.

It also follows from Equations (27) and (28),

$$E\left[\boldsymbol{x}^{*}y^{*}\right] = E\left[\boldsymbol{x}^{*}\left(\boldsymbol{x}^{H}\boldsymbol{h}\right)\right] + E\left[\boldsymbol{x}^{*}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{g}\right)\right]$$
$$= E\left[\boldsymbol{x}^{*}\boldsymbol{x}^{H}\right]\boldsymbol{h} + E\left[\boldsymbol{x}^{*}\boldsymbol{x}^{\mathrm{T}}\right]\boldsymbol{g}$$
$$= E\left[\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right)^{H}\right]\boldsymbol{h} + \Gamma_{2}\boldsymbol{g} \tag{30}$$
$$= C^{H}\boldsymbol{h} + \Gamma_{2}\boldsymbol{g}.$$

The left hand side of Equation (30) is given as:

$$E\left[\boldsymbol{x}^{*}y^{*}\right] = E\left[\left(y\boldsymbol{x}\right)^{*}\right] = \left(E\left[y\boldsymbol{x}\right]\right)^{*} = s^{*}. \tag{31}$$

Equations (30) and (31) complete the proof of Equation (17). ∎

Proof of Equations (18) and (19):

From Equation (16), we obtain

$$\boldsymbol{h} = \Gamma_{1}^{-1}\left(r - C\boldsymbol{g}\right). \tag{32}$$

Equations (17) and (34) lead Equation (18). And also, from Equation (17), we obtain

$$C\boldsymbol{g} = C\Gamma_{2}^{-1}\left(s^{*} - C^{H}\boldsymbol{h}\right). \tag{33}$$

Equations (16) and (33) lead Equation (19). ∎

Proof that the matrix $\Gamma_{2} - C^{H}\Gamma_{1}^{-1}C$ **is positive-semidefinite:**

Consider the following problem. Let $\boldsymbol{x} \in \boldsymbol{H}^{N}$ and $\boldsymbol{y} \in \boldsymbol{H}^{N}$ be two random vectors such that $E\left[\boldsymbol{x}\right] = E\left[\boldsymbol{y}\right] = \boldsymbol{0}$. The problem is to estimate linearly $\boldsymbol{y}$ in terms of $\boldsymbol{x}$, which means to find the N-dimensional quaternion valued matrix M such that $\hat{\boldsymbol{y}} = M\boldsymbol{x}$ is the best linear mean square (LMS) estimation of $\boldsymbol{y}$ in terms of $\boldsymbol{x}$. The solution is given by the orthogonality principle saying that $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to the observation $\boldsymbol{x}$ or $E\left[\left(\boldsymbol{y}-\hat{\boldsymbol{y}}\right)\boldsymbol{x}^{H}\right] = \boldsymbol{0}$. That is, $\Gamma_{yx} = M\Gamma_{xx}$ where $\Gamma_{yx} \overset{\text{def}}{=} E\left[\boldsymbol{y}\boldsymbol{x}^{H}\right]$ and $\Gamma_{xx} \overset{\text{def}}{=} E\left[\boldsymbol{x}\boldsymbol{x}^{H}\right]$. Assuming that Γ_{xx}^{-1} exists, the solution is obviously

$$M = \Gamma_{yx}\Gamma_{xx}^{-1}. \tag{34}$$

Here, the matrix $\Gamma_{\epsilon} \overset{\text{def}}{=} E\left[\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^{H}\right]$ is positive-semidefinite because for any $\boldsymbol{u} \in \boldsymbol{H}^{N}$ such that $\boldsymbol{u} \neq \boldsymbol{0}$, $\boldsymbol{u}^{H}\Gamma_{\epsilon}\boldsymbol{u} = E\left|\tilde{\boldsymbol{y}}^{H}\boldsymbol{u}\right|^{2} \geq 0$. Using the definition $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ and Equation (34) yield by simple algebra

$$\Gamma_{\epsilon} = \Gamma_{yy} - \Gamma_{yx}\Gamma_{xx}^{-1}\Gamma_{xy} \tag{35}$$

where $\Gamma_{yy} \overset{\text{def}}{=} E\left[\boldsymbol{y}\boldsymbol{y}^{H}\right]$ and $\Gamma_{xy} \overset{\text{def}}{=} E\left[\boldsymbol{x}\boldsymbol{y}^{H}\right]$.

Suppose now that $\boldsymbol{y} = \boldsymbol{x}^{*}$. This yields $\Gamma_{yy} = \Gamma_{2}$, $\Gamma_{yx} = C^{H}$, $\Gamma_{xx}^{-1} = \Gamma_{1}^{-1}$ and $\Gamma_{xy} = C$. Thus we obtain from Equation (35)

$$\Gamma_{\epsilon} = \Gamma_{2} - C^{H}\Gamma_{1}^{-1}C. \tag{36}$$

As seen above, since Γ_{ϵ} is positive-semidefinite, the right-hand side of Equation (36) is positive-semidefinite. ∎