Inertia Properties of Indefinite Quadratic Forms

Ali H. Sayed, Babak Hassibi, and Thomas Kailath

Abstract—We study the relation between the solutions of two estimation problems with indefinite quadratic forms. We show that a complete link between both solutions can be established by invoking a fundamental set of inertia conditions. While these inertia conditions are automatically satisfied in a standard Hilbert space setting, they nevertheless turn out to mark the differences between the two estimation problems in indefinite metric spaces. They also include, as special cases, the well-known conditions for the existence of $H^\infty$-filters and controllers.

I. INTRODUCTION

Given two invertible Hermitian matrices $\{\Pi, W\}$, a column vector $y$, and an arbitrary matrix $A$ of appropriate dimensions, we study the relation between the following two minimization problems with quadratic cost functions:

$$\min_z [z^*\Pi^{-1}z + (y - Az)^*W^{-1}(y - Az)]$$

(1)

where $z$ is a column vector of unknowns, and

$$\min_K (\Pi - K\Pi K - \Pi A^*K^* + K[\Pi A A^* + W]K^*)$$

(2)

where $K$ is a matrix. The symbol "**" stands for Hermitian conjugation. Moreover, and contrary to standard quadratic minimization problems, the weighting matrices $\Pi$ and $W$ are allowed to be indefinite. For this reason, solutions to (1) and (2) are not always guaranteed to exist. However, when they exist, the expressions for the solutions, and the conditions for their existence, can be related via a fundamental set of inertia conditions. Here, by the inertia of an invertible Hermitian matrix $X$, we mean a pair of integers, $I_+(X)$ and $I_-(X)$, that are equal to the number of strictly positive and negative eigenvalues of $X$, respectively.

The significance of the relations to be established between problems (1) and (2) is the following. It often happens in applications that one is interested in solving problems of the form (1), with indefinite weighting matrices. A prominent example that has received increasing attention in the last decade is the class of $H^\infty$-filtering and control problems [1]; in this context, the $\Pi$ matrix in (1) is restricted to be positive-definite, and the $W$ matrix is of, the special diagonal form $W = \text{diag.}\{\gamma I, I\}$, for a given positive constant $\gamma$. Here we shall allow for arbitrary indefinite weighting matrices $\{\Pi, W\}$.

On the other hand, problems of the form (2) are characteristic of state-space estimation formulations, where so-called Kalman filter procedure is available as an efficient computational scheme for determining the solution in the presence of state-space structure [2]. By relating the solutions of (1) and (2), we shall then be able to apply Kalman-type algorithms to the solution of (1), as well as obtain a complete set of inertia conditions that will automatically test for the existence of solutions to (1), without discarding the information from the solution of (2).

The inertia conditions follow as special cases of a basic result that tells us how the inertia of the matrices $\Pi$ and $W$ is affected by transformations of the form $(\Pi A A^* + W)$ and $(\Pi^{-1} + A^* W^{-1} A)$. The reason for choosing these transformations is because the positivity of these matrices is equivalent to necessary and sufficient conditions for the solvability of the problems (1) and (2). Hence, by studying how their inertia depends on $\{\Pi, W\}$, we shall be able to conclude how the choice of $\{\Pi, W\}$ affects the solvability of problems (1) and (2). The next conclusion follows by invoking Sylvester’s law of inertia [3] and the matrix inversion formula. Also $(\Pi \oplus W)$ denotes diagonal $\{\Pi, W\}$.

Theorem 1: The following inertia equalities hold:

$$I_+(\Pi \oplus W) = I_+[(\Pi^{-1} + A^* W^{-1} A) \oplus (\Pi A A^* + W)]$$

$$I_-(\Pi \oplus W) = I_-[(\Pi^{-1} + A^* W^{-1} A) \oplus (\Pi A A^* + W)]$$

if, and only if, $(\Pi A A^* + W)$ is invertible.

II. TWO INDEFINITE ESTIMATION PROBLEMS

We shall refer to problem (1) as the indefinite-weighted least-squares problem (IWLS, for short). The indefiniteness arises from the presence of the indefinite weighting matrices $\{\Pi, W\}$. Consequently, a bilinear form $a^* W^{-1} b$ is not guaranteed to satisfy the positivity condition $a^* W^{-1} a > 0$ for all nonzero column vectors $a$. We thus say that $C^*$, coupled with a bilinear form $a^* W^{-1} b$ with $W$ indefinite, is an indefinite metric space [4].

Let $J(z)$ denote the quadratic cost function that appears in (1). Every $z$ at which the gradient of $J(z)$ with respect to $z$ vanishes is called a stationary point of $J(z)$. A stationary point $z$ may or may not be a minimum of $J(z)$; it is unique if, and only if, $[\Pi^{-1} + A^* W^{-1} A]$ is invertible, in which case, it is given by $z = [\Pi^{-1} + A^* W^{-1} A]^{-1} A^* W^{-1} y$. This point will be a minimum iff $(\Pi^{-1} + A^* W^{-1} A) > 0$.

We shall refer to (2) as the equivalent estimation problem (or EE, for short). It arises in the following context. Consider column vectors $\{y, \nu, z\}$ that are linearly related via the expression $y = Az + \nu$, for some $A$, and where the
individual entries \( \{ y_i, v_i, z_i \} \) of \( \{ y, v, z \} \) are all elements of an indefinite metric space, say, \( K' \) (note that we are using boldface letters to denote the variables of the EE problem). The variables \( \{ v, z \} \) can be regarded as having Gramian matrices \( \{ W, \Pi \} \) and cross-Gramian zero, \( W = (v, u)_{K'}, \Pi = (z, z)_{K'}, \) and \( (v, z)_{K'} = 0. \) Under these conditions, it follows from the linear model that the Gramian matrix of \( y \) is equal to \( (y, y)_{K'} = A^*A + W. \) Let \( J(K) \) denote the quadratic cost function that appears in (2). It is then immediate to see that \( J(K') \) can be interpreted as the Gramian matrix of the vector difference \( (z - Ky, \text{viz}_z, \ J(K) = (z - Ky, z - Ky)_{K'}. \) Every \( K' \) at which the gradient of \( \alpha^*J(K)a \) with respect to \( \alpha^*A \) vanishes for all columns \( a \) is called a stationary solution of \( J(K'). \) A stationary point \( K' \) may or may not be a minimum.

Hence, solving for the stationary solutions \( K' \) can also be interpreted as solving the problem of linearly estimating \( z \) from \( y, \) which is denoted by \( \hat{z} = K' y. \) This estimate is uniquely defined if \( K' \) is unique. It is said to be the optimal linear estimate if \( K' \) is the unique minimizing solution.

A unique stationary solution \( K' \) exists iff \( (A^*A + W) \) is invertible. This gives \( K' = \frac{1}{\Pi} = (A^*W^{-1} - A^*W^{-1} A^*W^{-1} A^*W^{-1} A). \) Moreover, \( K' \) is a minimum (and \( \hat{z} \) is optimal) iff \( (A^*A + W) > 0. \)

Comparing the expressions for \( \hat{z} \) and \( \hat{y} \), we see that if we make the identifications \( \hat{z} \leftrightarrow \hat{z} \) and \( \hat{y} \leftrightarrow \hat{y}, \) then both expressions coincide. This means that the IVLS problem and the EE problem have the same expressions for the stationary points, \( \hat{z} \) and \( \hat{y}. \) But while a minimum for the IVLS problem (1) exists as long as \( (\Pi - 1 + A^*W^{-1} A) > 0, \) the equivalent problem (2) requires \( (W + A^*A) > 0. \)

A question of interest then is the following: given that one problem has a unique stationary solution, say, the EE problem, and given that this solution has been computed, is it possible to verify whether the other problem, say, the IVLS problem, admits a minimizing solution without explicitly checking for its positivity condition? The answer follows from Theorem 1: the IVLS problem (1) will have a unique minimum \( \hat{z} \) if (i.e., \((\Pi - 1 + A^*W^{-1} A) > 0, \) iff

\[
I_- [W + A^*A] = I_- [\Pi + W] \quad (3)
\]

\[
I_+ [W + A^*A] = I_+ [\Pi + W] - n. \quad (4)
\]

where \( n \times n \) is the size of \( \Pi. \)

The importance of the above conclusion is that it allows us to check whether a minimizing solution exists to the IVLS problem (1) by comparing the inertia of the Gramian matrix of the equivalent problem, viz., \( (W + A^*A^*), \) with the inertia of \( (\Pi + W). \) This is relevant because, as we shall see in the next section, when state-space structure is further imposed, we can derive an efficient procedure that allows us to keep track of the inertia of \( (W + A^*A^*). \) In particular, the procedure will produce a sequence of matrices \( \{ R_{e_i} \} \) such that

\[
\text{Inertia} (W + A^*A^*) = \text{Inertia} (R_{e_0} + R_{e_1} + R_{e_2} + \cdots).
\]

Therefore, “all” we need to do is compare the inertia of the given matrices \( \Pi \) and \( W \) to that of the matrices \( \{ R_{e_i} \} \) that are made available via the recursive procedure.

In summary, by establishing an explicit relation between both problems (1) and (2), we are capable of solving either problem via the solution of the other. In the special case of positive-definite quadratic cost functions, this point of view was exploited in [5] to great effect in order to establish a close link between known results in Kalman filtering theory and more recent results in adaptive filtering theory. In particular, it was shown in [5] that once such an equivalence relation is established, the various forms of (fixed- and order-recursive) adaptive filtering algorithms can be obtained by writing down different variants of the Kalman filter. The discussion in this letter, while it provides a similar connection for indefinite quadratic cost functions, shows that a satisfactory link requires an additional set of inertia conditions.

III. INTEGRATING STATE-SPACE STRUCTURE

Now that we have established the exact relationship between the two basic optimization problems (1) and (2), we shall proceed to study an important special case of the EE problem (2) that arises in a state-space context.

We consider vectors \( \{ y_i, x_i, u_i, v_i \} \), all with entries in \( K' \), and assume that they are related via the state-space equations

\[
x_{i+1} = F_i x_i + G_i u_i, \quad y_i = H_i x_i + v_i, \quad i \geq 0 \quad (5)
\]

where \( \delta_{ij} \) is the Kronecker delta function that is equal to unity when \( i = j \) and zero otherwise. The matrices \( \{ Q_i, R_i, \Pi_0 \} \) are possibly indefinite: \( \Pi_0 = n \times n, \) \( Q_i = m \times m, \) and \( R_i \) is \( p \times p. \)

The state-space structure (5) leads to a linear relation between the vectors \( \{ y_i \}_{i=0}^{\infty} \) and the vectors \( \{ x_0, u_0, \cdots, v_N \}, \) \( n \times n. \) More explicitly, if we collect the \( \{ y_i \} \) and the \( \{ v_i \} \) into two column vectors, say, \( y = \{ y_0, \cdots, y_N \} \) and \( v = \{ v_0, \cdots, v_N \}, \) and define \( z_N = \{ z_0, u_0, \cdots, u_{N-1} \}, \) then it follows from the state-space equations (5) that \( y = A z_N + \nu, \) where \( A \) is a block-lower triangular matrix. Moreover, the Gramian matrices of the variables \( \{ z_N, v \} \) so defined are \( (z_N, z_N)_{K'} = (\Pi_0 + Q_0 \cdots + Q_{N-1}), \) and \( (v, v)_{K'} = (R_0 \cdots + R_N). \) More compactly, we shall write \( (z_N, z_N)_{K'} = \Pi_0 \) and \( (v, v)_{K'} = W. \)

Let \( R_y = (y, y)_{K'} \) denote the Gramian matrix of the vector \( y, R_y = W + A^*A^*, \) and assume that \( R_y \) is (block) strongly regular. We have shown in [2] that the matrix \( R_y \) is congruent to \( (R_{e_0} \otimes R_{e_1} \otimes \cdots \otimes R_{e_N}), \) where the \( R_{e_i} \) are computed via a Kalman-type procedure: Start with \( P_0 = \Pi_0 \) and repeat: \( K_{p,i} = F_i P_i H_i^T R_{e_i}^{-1}, R_{e_i} = R_{e_i} + H_i P_i H_i^T, \) and \( R_{e+1} = F_i P_i F_i^T + G_i Q_i G_i^T - K_{p,i} K_{p,i}^T R_{e_i} R_{e_i}^{-1} \) (These recursions can also be used to recursively update the estimate of \( \hat{z}_N \), but
we shall omit the details here and focus only on the solvability conditions (3) and (4); see (7)).

Now, in view of the discussion in Section II, the solution \( \tilde{z}_N \) has the same expression as the solution \( \tilde{z}_N \) of a related minimization problem of the form (1), which, in view of the state-space relations (5), can be seen to be given by

\[
\min_{z_N = \col \{ x_0, u_0, \ldots, u_{N-1} \}} \left[ x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{N} (y_j - H_j x_j)^* R_j^{-1} (y_j - H_j x_j) + \sum_{j=0}^{N-1} u_j^* Q_j^{-1} u_j \right]
\]

subject to \( x_{j+1} = F_j x_j + G_j u_j \). We shall denote the above cost function by \( J_N \). Accordingly, we shall write \( J_i \) to denote a similar cost function where the two sum indices run up to \( i \) and \( (i-1) \), respectively, and where the independent variable is \( z_i = \col \{ x_0, u_0, \ldots, u_{i-1} \} \).

It now follows from the inertia conditions (3) and (4) that \( J_N \) has a minimum with respect to \( \{ x_0, u_0, \ldots, u_{N-1} \} \) if and only if

\[
I_\omega (\Pi \oplus W) = I_\omega \{ R_{e_0} \oplus \cdots \oplus R_{e_N} \}
\]

and

\[
I_\alpha (\Pi \oplus W) = I_\alpha \{ R_{e_0} \oplus \cdots \oplus R_{e_N} \} + n + mN.
\]

For example, in the special case \( \Pi_0 > 0 \) and \( Q_i > 0 \), we get \( \Pi > 0 \) and \( I_\omega (\Pi) = n + mN \), and the above becomes

\[
I_\omega \{ R_0 \oplus \cdots \oplus R_N \} = I_\omega \{ R_{e_0} \oplus \cdots \oplus R_{e_N} \}
\]

and

\[
I_\alpha \{ R_0 \oplus \cdots \oplus R_N \} = I_\alpha \{ R_{e_0} \oplus \cdots \oplus R_{e_N} \} + n + mN.
\]

More generally, we are interested in checking whether all the \( J_i \) have minima. The following follows from a recursive application of (3) and (4).

**Theorem 2:** All \( J_i \) have minima if, and only if,

\[
I_\omega (\Pi_0 \oplus R_0) = I_\omega \{ R_{e_0} \}\]

and

\[
I_\alpha (\Pi_0 \oplus R_0) = I_\alpha \{ R_{e_0} \} + n
\]

for \( i = 1, 2, \ldots, N \),

\[
I_\omega \{ Q_{i-1} \oplus R_i \} = I_\omega \{ R_{e_i} \}
\]

and

\[
I_\alpha \{ Q_{i-1} \oplus R_i \} = I_\alpha \{ R_{e_i} \} + m.
\]

It may happen in some applications that the last term in the definition of the quadratic cost function \( J_i \) also includes the extra term \( u_i^* Q^{-1} u_i \), say \( \sum_{j=0}^{N-1} u_j^* Q_j^{-1} u_j \). In this case, the unknown variable \( u_i \) only appears in the quadratic term \( u_i^* Q^{-1} u_i \), and it thus follows that minimization with respect to the \( u_i \) further requires the positivity of \( Q_i \). This motivates us to consider the following two corollaries.

**Corollary 1:** Assume further that the \( \{ Q_i \}_{i=0}^{N-1} \) are positive-definite and that \( \Pi_i > 0 \). Then all \( J_i \) have minima if, and only if, \( \text{Inertia} \{ R_i \} = \text{Inertia} \{ R_{e_i} \} \) for \( 0 \leq i \leq N \). In this case, it also follows that \( P_i \geq 0 \) for \( 0 \leq i \leq N \).

The next statement further assumes that the \( \{ F_i \} \) are invertible.

**Corollary 2:** Consider the same setting as in Corollary 1, and assume further that the \( \{ F_i \} \) are invertible. Then the following two statements provide equivalent necessary and sufficient conditions for all the \( J_i \) to have minima:

i) \( P_i^{-1} + H_i^* R_i^{-1} H_i > 0 \), for \( 0 \leq i \leq N \).

ii) \( P_{i+1} - G_i Q_i G_i^* > 0 \), for \( 0 \leq i \leq N \).

It follows in the minimum case that, for all \( i \), \( P_{i+1} > 0 \).

Conditions of the form i) in Corollary 2 are the ones most cited in \( H^\infty \)-applications. Here, we see that they are related to the inertia conditions of Corollary 1 and, more generally, to the conditions of Theorem 2. The inertia conditions of Corollary 1 also arise in the \( H^\infty \)-context [1] where \( R_i \) is further restricted to the form \( R_i = (I_q^2 I \oplus I) \). Here, we have derived these conditions as special cases of the general statement of Theorem 2, which holds for arbitrary indefinite matrices \( \{ \Pi_0, Q_i, R_i \} \).

Note also that testing for i) in Corollary 2 not only requires that we compute the \( P_i \) via a Riccati recursion, but also that we invert \( F_i \) and \( R_i \) at each step and then check for the positivity of \( P_i^{-1} + H_i^* R_i^{-1} H_i \). The inertia tests of Corollary 1 or Theorem 2, on the other hand, employ the quantities \( R_{e_i} \) and \( R_i \), which are \( p \times p \) matrices (as opposed to \( P_i \), which is \( n \times n \)). These tests can be used as the basis for alternative computational variants that are based on square-root ideas [6].

**REFERENCES**


