To appear in Operator Theory: Advances and Applications

INERTIA CONDITIONS FOR THE MINIMIZATION OF QUADRATIC FORMS IN INDEFINITE METRIC SPACES*

A. H. Sayed, B. Hassibi, and T. Kailath

We study the relation between the solutions of two minimization 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, which is the case of classical least-squares problems in both the deterministic and stochastic frameworks, they nevertheless turn out to mark the differences between the two optimization problems in indefinite metric spaces. Applications to $\mathrm{H}^{\infty}-$filtering, robust adaptive filtering, and approximate total-least-squares methods are included.

1 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:

$$\min_{z}\left[z^*\Pi^{-1}z + (y-Az)^*W^{-1}(y-Az)\right], \tag{1}$$

where z is a column vector of unknowns, and

$$\min_{K}\left\{\Pi - KA\Pi - \Pi A^*K^* + K[A\Pi A^* + W]K^*\right\}, \tag{2}$$

where K is a matrix. The symbol "$*$" stands for Hermitian conjugation (complex conjugation for scalars). If we denote the cost function that appears in (2) by $J(K)$,

$$J(K) \stackrel{\Delta}{=} \Pi - KA\Pi - \Pi A^*K^* + K[A\Pi A^* + W]K^*. \tag{3}$$

then by the minimization in (2) we mean finding a K^o such that for any complex column vector a, and for all K, we have $a^*J(K^o)a \leq a^*J(K)a$.

An interpretation of both optimization criteria (1) and (2) in terms of estimation problems in indefinite metric spaces is provided in the next sections. Here we only wish to emphasize that both cost functions

*This work was supported in part by a grant from the National Science Foundation under award no. MIP-9409319, and by the Army Research Office under contract DAAL03-89-K-0109.

in (1) and (2) are quadratic in the respective independent variables z and K, and that they can also be rewritten in the following revealing forms:

$$\min_{z} \begin{bmatrix} z^* & y^* \end{bmatrix} \begin{bmatrix} \Pi^{-1} + A^*W^{-1}A & -A^*W^{-1} \\ -W^{-1}A & W^{-1} \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix}, \tag{4}$$

and

$$\min_{K} \begin{bmatrix} I & -K \end{bmatrix} \begin{bmatrix} \Pi & \Pi A^* \\ A\Pi & A\Pi A^* + W \end{bmatrix} \begin{bmatrix} I \\ -K^* \end{bmatrix}, \tag{5}$$

where the central matrices

$$\begin{bmatrix} \Pi^{-1} + A^*W^{-1}A & -A^*W^{-1} \\ -W^{-1}A & W^{-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \Pi & \Pi A^* \\ A\Pi & A\Pi A^* + W \end{bmatrix}, \tag{6}$$

are in fact the inverses of each other, as detailed below.

Moreover, and contrary to standard quadratic minimization problems, the weighting matrices $\{\Pi, W\}$ in (1) and (2) are allowed to be indefinite (i.e., they are not restricted to being positive-definite). Consequently, the central matrices in (4) and (5) are generally indefinite. For this reason, solutions to (1) and (2) are not always guaranteed to exist. However, when they exist, we shall show that the expressions for the solutions, and the conditions for their existence, are very closely related. This relation will be established via a fundamental set of inertia conditions. Here, by the inertia of an invertible Hermitian matrix X, we mean a pair of integers, denoted by $I_+(X)$ and $I_-(X)$, where

$$\begin{aligned} I_+(X) &\triangleq \text{number of strictly positive eigenvalues of } X, \\ I_-(X) &\triangleq \text{number of strictly negative eigenvalues of } X. \end{aligned}$$

Note also that since X is assumed invertible, it has no zero eigenvalues and, consequently,

$$I_+(X) + I_-(X) = \text{number of columns (or rows) of the matrix } X.$$

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 quadratic problems of the form (1), with indefinite weighting matrices. A particular example that has received increasing attention in the last decade is the class of so-called H^∞-filtering and control problems, as suggested by several of the references at the end of this paper – see, e.g., the recent book [GL95] for more details and references on the topic. In this context, the Π matrix in (1) is further restricted to be positive-definite and the W matrix is indefinite but of the special form $W = \text{diag.}\{I, -\gamma^2 I\}$, for a given positive constant γ^2. Here we shall treat the general class of optimization problems suggested by (1) where both $\{\Pi, W\}$ are allowed to be arbitrary indefinite matrices. For example, the special case $\Pi = -\rho^2 I$ and $W = I$ turns out to be useful in approximate solutions of the so-called total least-squares (TLS) or errors-in-variables methods.

On the other hand, problems of the form (2) are characteristic of state-space estimation formulations, where a so-called Kalman filter procedure is available as an efficient computational scheme for determining the solution in the presence of state-space structure, as pointed out in [HSK93]. 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 available information from the solution of (2).

In the sequel, we shall use capital letters to denote matrices (e.g., A) and small letters to denote vectors.

2 An Inertia Result for Linear Transformations

We first establish a preliminary inertia result that tells us how the inertia of the matrices Π and W is affected by transformations of the form

$$(A\Pi A^* + W) \quad \text{and} \quad (\Pi^{-1} + A^*W^{-1}A), \tag{7}$$

for arbitrary matrices A of appropriate dimensions. The reason for choosing these transformations is because the positivity of the matrices in (7) will be shown later to be 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) – see Theorem 2.1 below. Also, a justification for the name *linear* transformations that appears in the title of this section will become clear further ahead, where it will be shown that the matrix A can be interpreted as the coefficient matrix of a linear model.

We start by noting that the matrices in (6) are indeed the inverses of each other and, consequently, that their inertia coincide. For this purpose, we form the square Hermitian matrix

$$G \triangleq \begin{bmatrix} \Pi & \Pi A^* \\ A\Pi & A\Pi A^* + W \end{bmatrix}, \tag{8}$$

and note that the Schur decomposition of G into a (block) *lower-diagonal-upper* triangular form leads to

$$G = \begin{bmatrix} I & 0 \\ A & I \end{bmatrix} \begin{bmatrix} \Pi & 0 \\ 0 & W \end{bmatrix} \begin{bmatrix} I & A^* \\ 0 & I \end{bmatrix}. \tag{9}$$

This establishes, in view of the assumptions on Π and W, that G is invertible. Its inverse is given by

$$G^{-1} = \begin{bmatrix} I & -A^* \\ 0 & I \end{bmatrix} \begin{bmatrix} \Pi^{-1} & 0 \\ 0 & W^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A & I \end{bmatrix} = \begin{bmatrix} \Pi^{-1} + A^*W^{-1}A & -A^*W^{-1} \\ -W^{-1}A & W^{-1} \end{bmatrix}, \tag{10}$$

which thus establishes our earlier claim that the matrices in (6) are the inverses of each other.

Note also that the Schur decompositions in (9) and (10) are in fact congruence relations. This shows, in view of Sylvester's law of inertia [Gan59], that G and G^{-1} have the same positive and negative inertia as the block diagonal matrix $(\Pi \oplus W)$,

$$I_+(G) = I_+(G^{-1}) = I_+(\Pi \oplus W), \quad I_-(G) = I_-(G^{-1}) = I_-(\Pi \oplus W).$$

Here, the notation $A \oplus B$ stands for a block diagonal matrix,

$$(A \oplus B) \triangleq \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

We state this preliminary result in the following lemma.

Lemma 2.1 (Inertia of G) *Given $\{\Pi, W\}$ Hermitian and invertible. Then, for any matrix A of appropriate dimensions, the block matrix*

$$G \triangleq \begin{bmatrix} \Pi & \Pi A^* \\ A\Pi & A\Pi A^* + W \end{bmatrix},$$

has the same positive and negative inertia as the block diagonal matrix $(\Pi \oplus W)$,

$$I_+(G) = I_+(\Pi \oplus W), \quad I_-(G) = I_-(\Pi \oplus W). \tag{11}$$

Proof: The proof is immediate from the congruence relation (9) and from Sylvester's law of inertia.

■

A less immediate inertia result follows if we instead perform a (block) *upper-diagonal-lower* triangular factorization of G. In this case, we need to further assume that the lower-right corner element of G is also invertible, viz.,

$$(A\Pi A^* + W) \quad \text{is invertible.} \tag{12}$$

It then follows that the matrix $(\Pi^{-1} + A^*W^{-1}A)$ will be invertible, as is immediate from the matrix inversion formula

$$(\Pi^{-1} + A^*W^{-1}A)^{-1} = \Pi - \Pi A^*(A\Pi A^* + W)^{-1}A\Pi. \tag{13}$$

This is in fact a useful preliminary result for our later analysis and a stronger statement is given below.

Lemma 2.2 (Invertibility Conditions) *Assume $\{\Pi, W\}$ are invertible. Then, for any matrix A of appropriate dimensions, $(A\Pi A^* + W)$ is invertible if, and only if, $(\Pi^{-1} + A^*W^{-1}A)$ is invertible.*

Proof: The result follows from the matrix inversion formulas

$$(\Pi^{-1} + A^*W^{-1}A)^{-1} = \Pi - \Pi A^*(A\Pi A^* + W)^{-1}A\Pi,$$

and

$$(A\Pi A^* + W)^{-1} = W^{-1} - W^{-1}A\left[\Pi^{-1} + A^*W^{-1}A\right]^{-1}A^*W^{-1}.$$

■

The right-hand side of (13) is simply the Schur complement of G with respect to its lower right block entry. We can therefore write the alternative Schur decomposition

$$G = \tag{14}$$

$$\begin{bmatrix} I & \Pi A^*(A\Pi A^* + W)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} (\Pi^{-1} + A^*W^{-1}A)^{-1} & 0 \\ 0 & A\Pi A^* + W \end{bmatrix} \begin{bmatrix} I & 0 \\ (A\Pi A^* + W)^{-1}A\Pi & I \end{bmatrix}.$$

It again follows from Sylvester's law of inertia, and under the additional assumption (12), that G has the same inertia as the block diagonal matrix $\left[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)\right]$. We establish a stronger statement in the following theorem.

Theorem 2.1 (Fundamental Inertia Result) *Given $\{\Pi, W\}$ Hermitian and invertible. Then, for any matrix A of appropriate dimensions, the following inertia equalities hold,*

$$\begin{aligned} I_+(\Pi \oplus W) &= I_+[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)], & (15) \\ I_-(\Pi \oplus W) &= I_-[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)], & (16) \end{aligned}$$

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

Proof: If $(A\Pi A^* + W)$ is invertible then the triangular decomposition (14) is applicable, thus leading to a congruence relation. This shows that G has the same inertia as

$$\left[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)\right].$$

The inertia equalities of the theorem then follow from (11).

Conversely, assume the inertia conditions (15) and (16) hold. Then the total number of nonzero eigenvalues of the block diagonal matrix $\left[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)\right]$ is equal to $(n + N)$, which is also the size of this block matrix. Here, n is the size of Π and N is the size of W. Consequently, none of the eigenvalues of either $(\Pi^{-1} + A^*W^{-1}A)$ or $(A\Pi A^* + W)$ can be zero. This implies that we must necessarily have an invertible matrix $(A\Pi A^* + W)$.

■

The inertia conditions (15) and (16) will play an important role in our analysis. In simple terms, they show how the inertia of the matrices $\{\Pi, W\}$ affects the inertia of the matrices $\{(A\Pi A^* + W), (\Pi^{-1} + A^*W^{-1}A)\}$, and vice-versa. In the special case of positive-definite matrices $\{\Pi, W\}$, we see that relation (16) becomes unnecessary and relation (15) is trivialized.

3 The Indefinite-Weighted Least-Squares Problem

We now return to the optimization problems (1) and (2) and proceed to a closer study of both criteria. We shall also motivate both problems by arguing that they can be related to estimation problems in indefinite metric spaces. We start with the first problem (1), which we shall refer to, for reasons to be clarified soon, as the indefinite-weighted least-squares problem (IWLS, for short).

Problem 3.1 (IWLS Problem) *Given invertible Hermitian matrices $\{\Pi, W\}$, a column vector y, and a matrix A of appropriate dimensions, we are interested in determining, if possible, the optimal $\hat{z}$ that solves the optimization problem:*

$$\min_z \left[z^*\Pi^{-1}z + (y - Az)^*W^{-1}(y - Az)\right] . \tag{17}$$

3.1 Interpretation as an Estimation Problem with an Indefinite Metric

The problem (17) admits an interpretation in terms of an estimation problem as follows. We may regard z as a column vector of n unknown parameters that is related to the vector y via a linear relation of the form

$$y = Az + v, \tag{18}$$

where v denotes the mismatch between the value of y and the value of Az. In signal processing literature, the y is called the *observation* vector, the v is called the *noise* vector, and the objective is to use the available data y in order to come up with an estimate for the unknown vector z. The problem is posed as one of minimizing a quadratic cost function of the same form as in (17) but with *positive-definite* matrices $\{\Pi, W\}$ [Hay91, PRLN92]. It is well known in such cases that for any positive-definite matrix W, and for any complex-valued column vectors a and b in $\mathcal{C}^n$, the scalar quantity $a^*W^{-1}b$ is a well-defined inner product, denoted by $< b, a >$, and, consequently, least-squares solutions can be found by orthogonally projecting onto appropriate linear subspaces – see, e.g., [SK94] for a recent survey on the topic in the positive-definite case and along the lines of this paper.

Here, however, we allow for indefinite matrices $\{\Pi, W\}$, thus leading to a least-squares problem with indefinite weighting matrices. Now a bilinear form $a^*W^{-1}b$ is not guaranteed to satisfy the positivity condition $a^*W^{-1}a > 0$ for all nonzero a. We thus say that $\mathcal{C}^n$, coupled with a bilinear form $a^*W^{-1}b$ with W indefinite, is an indefinite metric space. More generally, an indefinite metric space $\{\mathcal{K}, < .,. >_{\mathcal{K}}\}$ is defined as a vector space that satisfies two simple requirements (see, e.g., [Bog74, GLR83] for more details):

(i) $\mathcal{K}$ is linear over the field of complex numbers $\mathcal{C}$, and

(ii) $\mathcal{K}$ possesses a bilinear form, $< .,. >_{\mathcal{K}}$, such that for any $a, b, c \in \mathcal{K}$, and for any $\alpha, \beta \in \mathcal{C}$, we have

$$\begin{aligned} < \alpha a + \beta b, c >_{\mathcal{K}} &= \alpha < a, c >_{\mathcal{K}} + \beta < b, c >_{\mathcal{K}}, \\ < b, a >_{\mathcal{K}} &= < a, b >_{\mathcal{K}}^* . \end{aligned}$$

In particular, the quantity $< a, a >_{\mathcal{K}}$ is in general indefinite. This is in contrast to a Hilbert space setting, $\{\mathcal{H}, < .,. >_{\mathcal{H}}\}$, where for any $a \in \mathcal{H}$ the quantity $< a, a >_{\mathcal{H}}$ is necessarily nonnegative.

In the formulation (17), each of the terms $z^*\Pi^{-1}z$ and $(y - Az)^*W^{-1}(y - Az)$ may be indefinite. Note also that we can rewrite the cost function in (17) in the form:

$$\min_z \left(\begin{bmatrix} 0 \\ y \end{bmatrix} - \begin{bmatrix} I \\ A \end{bmatrix} z \right)^* \begin{bmatrix} \Pi & 0 \\ 0 & W \end{bmatrix}^{-1} \left(\begin{bmatrix} 0 \\ y \end{bmatrix} - \begin{bmatrix} I \\ A \end{bmatrix} z \right),$$

where the central matrix $(\Pi \oplus W)^{-1}$ is indefinite. This further highlights that the cost function in (17) is an *indefinite* quadratic cost function.

Also, in estimation problems it often happens that the linear model (18) arises as a consequence of repeated experiments. That is, one collects several observation vectors $\{y_i\}$ that are also linearly related to the same unknown z, say via

$$y_i = A_i z + v_i,$$

where the A_i are given matrices of appropriate dimensions, and the v_i are the corresponding noise components. If we collect several such observations into matrix form and write

$$\underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_N \end{bmatrix}}_{A} z + \underbrace{\begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_N \end{bmatrix}}_{v},$$

we again obtain the linear model (18) and we are back to the problem of estimating z from the y by solving (17).

3.2 Solution of the IWLS Problem

Let $J(z)$ denote the quadratic cost function that appears in (17),

$$\begin{aligned} J(z) &\stackrel{\Delta}{=} z^*\Pi^{-1}z + (y - Az)^*W^{-1}(y - Az), \\ &= z^*[\Pi^{-1} + A^*W^{-1}A]z - y^*W^{-1}Az - z^*A^*W^{-1}y \; + \; y^*W^{-1}y. \end{aligned} \tag{19}$$

Every $\hat{z}$ at which the gradient of $J(z)$ with respect to z vanishes is called a *stationary* point of $J(z)$. A stationary point $\hat{z}$ may or may not be a minimum of $J(z)$ as clarified by the following statement.

Theorem 3.1 (Solution of the IWLS Problem) *The stationary points $\hat{z}$ of $J(z)$ in (19), if they exist, are solutions of the linear system of equations*

$$[\Pi^{-1} + A^*W^{-1}A]\hat{z} = A^*W^{-1}y. \tag{20}$$

There exists a unique stationary point if, and only if, $[\Pi^{-1} + A^*W^{-1}A]$ is invertible. In this case, it is given by

$$\hat{z} = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}y, \tag{21}$$

and the corresponding value of the cost function is

$$J(\hat{z}) = y^* \left[W + A\Pi A^*\right]^{-1} y. \tag{22}$$

Moreover, this unique point is a minimun if, and only if, the coefficient matrix is positive-definite,

$$(\Pi^{-1} + A^*W^{-1}A) > 0. \tag{23}$$

Proof: It is straightforward to verify, by differentiation, that the gradient of $J(z)$ with respect to z^* is equal to $\left([\Pi^{-1} + A^*W^{-1}A]z - A^*W^{-1}y\right)$. Therefore, the stationary points of $J(z)$, when they exist, must satisfy the linear system of equations

$$[\Pi^{-1} + A^*W^{-1}A]\hat{z} = A^*W^{-1}y.$$

This has a unique solution $\hat{z}$ if, and only if, the coefficient matrix is invertible. Also, the Hessian matrix is equal to $[\Pi^{-1} + A^*W^{-1}A]$, which thus needs to be positive-definite for a unique minimum solution with respect to z.

■

Note that in contrast to positive-definite least-squares problems (i.e., when $\Pi > 0$ and $W > 0$) where $[\Pi^{-1} + A^*W^{-1}A]$ is always guaranteed to be positive for any A and, consequently, a unique minimizing solution of $J(z)$ always exists, the IWLS problem may or may not have a minimum, and actually may not even have a stationary point if a solution to (20) does not exist.

4 The Equivalent Estimation Problem

We now study the second optimization criterion (2) and also present an interpretation for it in terms of an estimation problem in an indefinite metric space. We shall refer to this problem as the equivalent estimation problem (or EE, for short).

Problem 4.1 (The EE Problem) *Given invertible Hermitian matrices $\{\Pi, W\}$, and a matrix A of appropriate dimensions, we are interested in determining, if possible, the optimal K^o that solves the optimization problem (in the sense explained after (3)):*

$$\min_K \{\Pi - KA\Pi - \Pi A^* K^* + K[A\Pi A^* + W]K^*\}. \tag{24}$$

4.1 Interpretation as an Estimation Problem with an Indefinite Metric

An interpretation for this problem is the following. We consider column vectors $\{\mathbf{y}, \mathbf{v}, \mathbf{z}\}$ that are linearly related via the expression

$$\mathbf{y} = A\mathbf{z} + \mathbf{v}, \tag{25}$$

and where the individual entries $\{\mathbf{y}_i, \mathbf{v}_i, \mathbf{z}_i\}$ of the vectors $\{\mathbf{y}, \mathbf{v}, \mathbf{z}\}$ are all elements of an indefinite metric space, say $\mathcal{K}'$.

For two vectors $\{\mathbf{a}, \mathbf{b}\}$, with entries $\{\mathbf{a}_i, \mathbf{b}_j\}$ in $\mathcal{K}'$, we write $< \mathbf{a}, \mathbf{b} >_{\mathcal{K}'}$ to denote a matrix whose entries are the individual $< \mathbf{a}_i, \mathbf{b}_j >_{\mathcal{K}'}$. In a Hilbert setting, an analogy arises with the space of scalar-valued zero-mean random variables, say $\mathcal{E}$: for two column vectors $\mathbf{p}$ and $\mathbf{q}$ of random variables, the bilinear form $E\mathbf{p}\mathbf{q}^*$ is a matrix whose individual entries are $E\mathbf{p}_i\mathbf{q}_j^*$ (see, e.g., [AM79, Kai81]). Note that to distinguish between the elements in $\mathcal{K}$ and $\mathcal{K}'$, we are using boldface letters to denote the variables of the equivalent problem.

The variables $\{\mathbf{v}, \mathbf{z}\}$ can be regarded as having Gramian matrices $\{W, \Pi\}$ and cross Gramian zero, namely

$$W \triangleq < \mathbf{v}, \mathbf{v} >_{\mathcal{K}'}, \quad \Pi \triangleq < \mathbf{z}, \mathbf{z} >_{\mathcal{K}'}, \quad < \mathbf{z}, \mathbf{v} >_{\mathcal{K}'} = 0.$$

Under these conditions, it follows from the linear model (25) that the Gramian matrix of $\mathbf{y}$ is equal to

$$< \mathbf{y}, \mathbf{y} >_{\mathcal{K}'} = A\Pi A^* + W.$$

Let $J(K)$ denote the quadratic cost function that appears in (24),

$$J(K) \triangleq \Pi - KA\Pi - \Pi A^* K^* + K[A\Pi A^* + W]K^*. \tag{26}$$

It is then immediate to see that $J(K)$ can be interpreted as the Gramian matrix of the vector difference $(\mathbf{z} - K\mathbf{y})$, viz.,

$$J(K) = < \mathbf{z} - K\mathbf{y}, \mathbf{z} - K\mathbf{y} >_{\mathcal{K}'}.$$

Every K^o at which the gradient of $a^*J(K)a$ with respect to a^*K vanishes for all a is called a *stationary* solution of $J(K)$ [Note from (26) that $a^*J(K)a$ is a function of a^*K]. A stationary point K^o may or may not be a minimum as clarified further ahead.

Hence, solving for the stationary solutions K^o can also be interpreted as solving the problem of linearly estimating $\mathbf{z}$ from $\mathbf{y}$.

Definition 4.1 (Linear Estimates) *A linear estimate of $\mathbf{z}$ given $\mathbf{y}$ is defined by*

$$\hat{\mathbf{z}} \triangleq K^o\mathbf{y}, \tag{27}$$

where K^o is a stationary solution of (24). This estimate is uniquely defined if K^o is unique. It is further said to be the optimal linear estimate if K^o is the unique minimizing solution of (24).

4.2 Solution of the EE Problem

We now state and prove the solution of (24).

Theorem 4.1 (Solution of the EE Problem) *The stationary points K^o of $J(K)$, if they exist, are solutions of the linear system of equations*

$$\Pi A^* = K^o[A\Pi A^* + W]. \tag{28}$$

There exists a unique stationary point K^o if, and only if, $(A\Pi A^* + W)$ is invertible. In this case, it is given by

$$K^o = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}, \tag{29}$$

and the corresponding value of the cost function is

$$J(K^o) = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1}. \tag{30}$$

The unique linear estimate of the corresponding $\mathbf{z}$ in (27) is

$$\hat{\mathbf{z}} = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}\mathbf{y}. \tag{31}$$

Moreover, this unique point K^o is a minimum (and, correspondingly, $\hat{\mathbf{z}}$ is optimal) if, and only if, the coefficient matrix is positive-definite,

$$(A\Pi A^* + W) > 0. \tag{32}$$

Proof: The proof follows the same lines of Theorem 3.1 when applied to the now scalar-valued cost function $a^*J(K)a$, where a is any column vector (recall the explanation below (3)). In particular, it is immediate to see that any stationary solution K^o, if it exists, must satisfy the orthogonality condition $< \mathbf{z} - K^o\mathbf{y}, \mathbf{y} >_{\mathcal{K}'} = 0$, which leads to the linear system of equations

$$\Pi A^* = K^o[A\Pi A^* + W].$$

A unique stationary point K^o then exists as long as $[A\Pi A^* + W]$ is invertible, thus leading to the expression

$$K^o = \Pi A^*[A\Pi A^* + W]^{-1}. \tag{33}$$

But in view of the matrix inversion formula, and Lemma 2.2,

$$[A\Pi A^* + W]^{-1} = W^{-1} - W^{-1}A\left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1},$$

we can also write

$$K^o = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}.$$

The necessary and sufficient condition for this solution to correspond to a minimum is $(A\Pi A^* + W) > 0$, as follows if we evaluate the Hessian matrix of $a^*J(K)a$.

■

The matrices that appear in (33) can be interpreted as follows:

$$< \mathbf{z}, \mathbf{y} >_{\mathcal{K}'} = \Pi A^*, \quad < \mathbf{y}, \mathbf{y} >_{\mathcal{K}'} = A\Pi A^* + W.$$

We therefore conclude that the following equivalent equalities also hold:

$$K^o = < \mathbf{z}, \mathbf{y} >_{\mathcal{K}'} < \mathbf{y}, \mathbf{y} >_{\mathcal{K}'}^{-1}, \tag{34}$$

$$\hat{\mathbf{z}} = < \mathbf{z}, \mathbf{y} >_{\mathcal{K}'} < \mathbf{y}, \mathbf{y} >_{\mathcal{K}'}^{-1} \mathbf{y}. \tag{35}$$

5 Relations between the IWLS and EE Problems

We now compare expressions (31) and (21). We see that if we make the identifications: $\hat{\mathbf{z}} \leftarrow \hat{z}$ and $\mathbf{y} \leftarrow y$, then both expressions coincide. This means that the IWLS problem and the equivalent problem have the same expressions for the stationary points, $\hat{z}$ and $\hat{\mathbf{z}}$. But while a minimum for the IWLS problem (17) exists as long as $(\Pi^{-1} + A^*W^{-1}A) > 0$, the equivalent problem (24), on the other hand, has a minimum at K^o if, and only if, $(W + A\Pi A^*) > 0$.

This indicates that both problems are not generally guaranteed to have simultaneous minima. In the special case of positive-definite matrices $\{\Pi, W\}$, both conditions

$$(\Pi^{-1} + A^*W^{-1}A) > 0 \quad \text{and} \quad (W + A\Pi A^*) > 0,$$

are simultaneously met. But this situation does not hold for general indefinite matrices Π and W. A question of interest then is the following: given that one problem has a unique stationary solution, say the EE problem (24), and given that this solution has been computed, is it possible to verify whether the other problem, say the IWLS problem (17) admits a minimizing solution without explicitly checking for its positivity condition $(\Pi^{-1} + A^*W^{-1}A) > 0$?

The relevance of this question is that, as we shall see in a later section, when state-space structure is further imposed on the data, an efficient recursive procedure can be derived for the solution of the equivalent problem (24). Hence, once a connection is established with the IWLS problem (17), the solution of the latter should follow immediately.

We shall see that this is indeed possible by invoking the inertia results of Sec. 2. To begin with, the following result is a consequence of Lemma 2.2.

Lemma 5.1 (Simultaneous Stationary Points) *The IWLS problem (17) has a unique stationary point $\hat{z}$ if, and only if, the equivalent problem (24) has a unique stationary point K^o.*

Proof: The IWLS problem (17) has a unique stationary point $\hat{z}$ iff $(\Pi^{-1} + A^*W^{-1}A)$ is nonsingular. Likewise, the equivalent problem (24) has a unique stationary point K^o iff $(W + A\Pi A^*)$ is nonsingular. But, according to Lemma 2.2, the nonsingularity of one matrix implies the nonsingularity of the other, which thus establishes the desired result.

■

This means that both optimization problems are always guaranteed to simultaneously have unique stationary solutions $\hat{z}$ and K^o, regardless of the invertible matrices $\{\Pi, W\}$ and for any A. That is, once we find a unique stationary solution K^o for the equivalent problem (24), we are at least guaranteed a unique stationary solution $\hat{z}$ for the IWLS problem. But we are in fact interested in a stronger result. We would like to verify whether this stationary solution $\hat{z}$ is a minimum or not. We would also like to be able to settle this question by exploiting the solution of the equivalent problem (24), and without explicitly checking the positivity condition that is required on $(\Pi^{-1} + A^*W^{-1}A)$ in the IWLS case (17).

The next statement is one of the main conclusions of this paper since it provides a set of inertia conditions that allows us to check the solvability of the IWLS problem (17) in terms of the inertia properties of the Gramian matrix $(A\Pi A^* + W)$ associated with the equivalent problem (24).

Theorem 5.1 (Fundamental Inertia Conditions) *Given invertible and Hermitian matrices Π and W, and an arbitrary matrix A of appropriate dimensions, the optimization problem (1) (i.e., the IWLS problem (17)) has a unique minimizing solution $\hat{z}$ if, and only if,*

$$\begin{aligned} I_-\left[W + A\Pi A^*\right] &= I_-\left[\Pi \oplus W\right], \\ I_+\left[W + A\Pi A^*\right] &= I_+\left[\Pi \oplus W\right] - n, \end{aligned}$$

where $n \times n$ is the size of Π.

Proof: Assume the IWLS problem has a unique minimizing solution. This means that we necessarily have

$$(\Pi^{-1} + A^*W^{-1}A) > 0.$$

We then obtain from Lemma 5.1 that $(W + A\Pi A^*)$ is also invertible.

In view of Theorem 2.1 we conclude that we must have

$$\begin{aligned} I_+(\Pi \oplus W) &= I_+[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)], \\ I_-(\Pi \oplus W) &= I_-[(\Pi^{-1} + A^*W^{-1}A) \oplus (A\Pi A^* + W)]. \end{aligned}$$

But $I_-[(\Pi^{-1} + A^*W^{-1}A)] = 0$ and $I_+[(\Pi^{-1} + A^*W^{-1}A)] = n$. Hence,

$$\begin{aligned} I_-\left[W + A\Pi A^*\right] &= I_-\left[\Pi \oplus W\right], \\ I_+\left[W + A\Pi A^*\right] &= I_+\left[\Pi \oplus W\right] - n. \end{aligned}$$

Conversely, assume the above inertia relations hold. It follows that the number of (strictly positive and strictly negative) eigenvalues of $(W + A\Pi A^*)$ is equal to the size of W. Therefore, $(W + A\Pi A^*)$ has no zero eigenvalues and is thus invertible. It follows from Lemma 2.2 that $(\Pi^{-1} + A^*W^{-1}A)$ is also invertible. We further invoke Theorem 2.1 to conclude that

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

which thus establishes that we necessarily have

$$I_-[(\Pi^{-1} + A^*W^{-1}A)] = 0.$$

Therefore, $(\Pi^{-1} + A^*W^{-1}A) > 0$ and the IWLS problem (1) has a unique minimum.

■

The importance of the above theorem is that it allows us to check whether a minimizing solution exists to the IWLS problem (17) by comparing the inertia of the Gramian matrix of the equivalent problem, viz., $(W + A\Pi A^*)$, with the inertia of $(\Pi \oplus W)$. 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 $(W + A\Pi A^*)$. In particular, the procedure will produce a sequence of matrices $\{R_{e,i}\}$ such that

$$\text{Inertia}(W + A\Pi A^*) = \text{Inertia}\ \left(R_{e,0} \oplus R_{e,1} \oplus R_{e,2} \ldots\right).$$

The theorem then shows that "all" we need to do is compare the inertia of the given matrices Π and W with that of the matrices $\{R_{e,i}\}$ that are made available via the recursive procedure.

Equally important is that this procedure will further allow us to compute the quantity $\hat{\mathbf{z}}$ in (27). But since we argued above that $\hat{\mathbf{z}}$ has the same expression as $\hat{z}$, the stationary solution of (17), then the procedure will also provide us with $\hat{z}$.

In summary, by establishing an explicit relation between both problems (17) and (24), we shall be 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 fully exploited in [SK94] 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 [SK94] that once such an equivalence relation is established, the varied forms of adaptive filtering algorithms can be obtained by writing down different variants of the so-called Kalman filter.

The discussion in this paper, while it provides a similar connection for indefinite quadratic cost functions, it shows that a satisfactory link can be established via an additional set of inertia conditions. These conditions are necessary because, contrary to the case of positive-definite quadratic cost functions, minimizing solutions are not always guaranteed to exist in the indefinite case. Note that in the positive case (i.e., Π and W positive), the inertia conditions of Theorem 5.1 are automatically satisfied.

We may finally remark that the above inertia conditions include, as special cases, the well-known conditions for the existence of H^∞-controllers and filters, as will be clarified in later sections.

6 Incorporating 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 equivalent problem (2).

More specifically, we shall pose an optimization problem that will be of the same form as (2) except that the associated A matrix will have considerable structure in it. In particular, the A matrix will be block-lower triangular and its individual entries will be further parameterized in terms of matrices $\{F_i, G_i, H_i\}$ that arise from an underlying state-space assumption. This will allow us to derive an efficient computational scheme for the solution of the corresponding optimization problem (2). The scheme is an extension to the indefinite case of a well-known Kalman filtering algorithm [HSK93].

6.1 Statement of the State-Space Problem

We consider an indefinite metric space $\mathcal{K}'$ and continue to employ the notation $<\mathbf{a}, \mathbf{b}>_{\mathcal{K}'}$ to denote a matrix with entries $<\mathbf{a}_i, \mathbf{b}_j>_{\mathcal{K}'}$, where $\{\mathbf{a}_i, \mathbf{b}_j\} \in \mathcal{K}'$ are the individual entries of the columns $\mathbf{a}$ and $\mathbf{b}$.

We further consider vectors $\{\mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i\}$, all with entries in $\mathcal{K}'$, and assume that they are related via state-space equations of the form

$$\begin{aligned} \mathbf{x}_{i+1} &= F_i \mathbf{x}_i + G_i \mathbf{u}_i, \\ \mathbf{y}_i &= H_i \mathbf{x}_i + \mathbf{v}_i\ , \quad i \geq 0, \end{aligned} \tag{36}$$

where F_i, H_i, and G_i are known $n \times n$, $p \times n$, and $n \times m$ matrices, respectively. It is further assumed that the Gramian matrices of $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ are known, say

$$<\mathbf{v}_i, \mathbf{v}_i>_{\mathcal{K}'} = R_i, \quad <\mathbf{u}_i, \mathbf{u}_i>_{\mathcal{K}'} = Q_i, \quad <\mathbf{x}_0, \mathbf{x}_0>_{\mathcal{K}'} = \Pi_0.$$

We also assume that the following relations hold for all $i \neq j$,

$$<\mathbf{v}_i, \mathbf{v}_j>_{\mathcal{K}'} = 0, \quad <\mathbf{u}_i, \mathbf{u}_j>_{\mathcal{K}'} = 0, \quad <\mathbf{v}_i, \mathbf{x}_0>_{\mathcal{K}'} = 0, \quad <\mathbf{u}_i, \mathbf{x}_0>_{\mathcal{K}'} = 0,$$

as well as $<\mathbf{v}_i, \mathbf{u}_j>_{\mathcal{K}'} = 0$ for all i, j. More compactly, we may write the above requirements in the following form

$$< \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \end{bmatrix} >_{\mathcal{K}'} = \begin{bmatrix} Q_i\delta_{ij} & 0 & 0 \\ 0 & R_i\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}, \tag{37}$$

where δ_{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.

The quantities $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ are assumed *unknown* and only the $\{\mathbf{y}_i\}$ are *known*. In other words, we assume that we have a collection of vectors $\{\mathbf{y}_i\}$ that we know arose from a state-space model of the form (36), with known $\{F_i, G_i, H_i\}$, but with no further access to the $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$, except for the knowledge of their Gramian matrices as in (37).

The state-space structure (36) leads to a linear relation between the vectors $\{\mathbf{y}_i\}$ and the vectors $\{\mathbf{x}_0, \mathbf{u}_i\}_{i=0}^{N-1}$. Indeed, if we collect the $\{\mathbf{y}_i\}_{i=0}^{N}$ and the $\{\mathbf{v}_i\}_{i=0}^{N}$ into two column vectors, $\{\mathbf{y}, \mathbf{v}\}$, respectively,

$$\mathbf{y} \triangleq \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \mathbf{v} \triangleq \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_N \end{bmatrix}, \tag{38}$$

and define the column vector,

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N-1} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u} \end{bmatrix}, \tag{39}$$

it then follows from the state-space equations that

$$\mathbf{y} = A\mathbf{z} + \mathbf{v},$$

where A is the block-lower triangular matrix

$$A \triangleq \begin{bmatrix} H_0 & & & \\ H_1 F^{[0,0]} & H_1 G_0 & & \\ H_2 F^{[1,0]} & H_2 F^{[1,1]} G_0 & & \\ \vdots & \vdots & \ddots & \\ H_N F^{[N-1,0]} & H_N F^{[N-1,1]} G_0 & \dots & H_N G_{N-1} \end{bmatrix}. \tag{40}$$

Here, the notation $F^{[i,j]}$, $i \geq j$, stands for

$$F^{[i,j]} \triangleq F_i F_{i-1} \dots F_j.$$

Moreover, the Gramian matrices of the variables $\{\mathbf{z}, \mathbf{v}, \mathbf{y}\}$ so defined are easily seen to be, in view of the assumptions (37),

$$< \mathbf{z}, \mathbf{z} >_{\mathcal{K}'} = (\Pi_0 \oplus Q_0 \dots \oplus Q_{N-1}), \tag{41}$$

$$< \mathbf{v}, \mathbf{v} >_{\mathcal{K}'} = (R_0 \oplus R_1 \oplus \dots \oplus R_N). \tag{42}$$

More compactly, we shall write

$$< \mathbf{z}, \mathbf{z} >_{\mathcal{K}'} \triangleq \Pi, \quad < \mathbf{v}, \mathbf{v} >_{\mathcal{K}'} \triangleq W, \tag{43}$$

where the $\{\Pi, W\}$ are block diagonal matrices as defined in (41) and (42). We can now pose the following problem.

Problem 6.1 (State-Space Estimation Problem) *Consider the state-space model (36) and given the $\{\mathbf{y}, A, \Pi, W\}$ as above, determine a matrix K, and conditions on $\{A, \Pi, W\}$, so as to minimize the Gramian matrix*

$$\min_K \; < \mathbf{z} - K\mathbf{y}, \mathbf{z} - K\mathbf{y} >_{\mathcal{K}'} . \tag{44}$$

The optimal solution K^o, when it exists, can be used to define $K^o\mathbf{y}$ as the optimal linear estimate for $\mathbf{z}$. We denote this by

$$\hat{\mathbf{z}} \triangleq K^o \mathbf{y}.$$

In other words, we have posed the problem of linearly estimating $\mathbf{z}$ from $\mathbf{y}$ so as to minimize the Gramian matrix of the error signal, $\mathbf{z} - K\mathbf{y}$. This Gramian matrix can be expanded and the problem is easily seen to be equivalent to

$$\min_K \left\{ \Pi - KA\Pi - \Pi A^* K^* + K[A\Pi A^* + W]K^* \right\},$$

where we have used (41) and (42).

We thus see that, given a state-space model of the form (36) and (37), the problem of linearly estimating the variables $\{\mathbf{x}_0, \mathbf{u}_0, \dots, \mathbf{u}_{N-1}\}$ from the variables $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_N\}$ leads to an optimization problem of the same form as in (2): it requires that we determine a coefficient matrix K that minimizes $J(K)$. The optimal

K^o is then used to define the optimal linear estimate of the desired variables via $\hat{\mathbf{z}} = K^o\mathbf{y}$. In case K^o is simply a unique stationary solution of $J(K)$, but not necessarily the minimum solution, we shall refer to $\hat{\mathbf{z}}$ as simply the linear estimate of $\mathbf{z}$ given $\mathbf{y}$, instead of the *optimal* linear estimate.

Using the result of Theorem 4.1, a unique linear estimate $\hat{\mathbf{z}}$ exists as long as $(A\Pi A^* + W)$ is invertible, where the matrices $\{A, \Pi, W\}$ are now as defined above. Moreover, when this happens the estimate $\hat{\mathbf{z}}$ is given by the expression

$$\hat{\mathbf{z}} = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}\mathbf{y}. \tag{45}$$

Alternatively, and using (35), we also write for later reference,

$$\hat{\mathbf{z}} = <\mathbf{z}, \mathbf{y}>_{\mathcal{K}'} <\mathbf{y}, \mathbf{y}>_{\mathcal{K}'}^{-1} \mathbf{y}. \tag{46}$$

While the expression (45) is analytically satisfactory, it however does not exploit two important facts that occur under the assumption of the state-space structure, namely that the matrices $\{\Pi, W\}$ are block diagonal and, more importantly, that the matrix A is now block-lower triangular. The entries of A are also completely parameterized by the matrices $\{F_i, G_i, H_i\}$ that describe the state-space model (36).

We shall see in the sequel that these two facts can be exploited in order to provide an alternative method for computing the solution $\hat{\mathbf{z}}$. While (45) provides a global expression for $\hat{\mathbf{z}}$, we shall argue that it will be more convenient to introduce a recursive procedure for computing $\hat{\mathbf{z}}$.

Remark on Notation. We shall from now on write $\mathbf{z}_N$ instead of $\mathbf{z}$ to indicate that it includes $\mathbf{x}_0$ and the vectors $\{\mathbf{u}_j\}$ up to time $N-1$, as defined in (39). That is, the subindex N indicates which vectors $\{\mathbf{u}_j\}$ are included in the definition of $\mathbf{z}$. We shall then write $\hat{\mathbf{z}}_{N|N}$ instead of simply $\hat{\mathbf{z}}$ to indicate that it is the estimate of $\mathbf{z}_N$ that is obtained by using the vectors $\{\mathbf{y}_i\}$ up to time N. That is, the $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$ are used in (45),

$$\hat{\mathbf{z}}_{N|N} = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}\mathbf{y}. \tag{47}$$

More generally, the estimate of $\mathbf{z}_N$ that is based on a different number of vectors $\{\mathbf{y}_j\}$, say up to time k, will be correspondingly indicated by $\hat{\mathbf{z}}_{N|k}$. In other words, the first subindex indicates which vectors $\{\mathbf{u}_j\}$ are included in the definition of the variable $\mathbf{z}$ and the second subindex indicates which vectors $\{\mathbf{y}_j\}$ are used in the estimation of $\mathbf{z}$.

These notational changes are necessary because we shall find it useful later to also define, for each i, the vector $\mathbf{z}_i$,

$$\mathbf{z}_i \triangleq \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{i-1} \end{bmatrix}, \tag{48}$$

which contains $\mathbf{x}_0$ and the vectors $\{\mathbf{u}_j\}$ up to time $(i-1)$. Correspondingly, the estimate of $\mathbf{z}_i$ that is based on vectors $\{\mathbf{y}_j\}$ up to a time k will be indicated by $\hat{\mathbf{z}}_{i|k}$.

6.2 A Strong Regularity Condition on the Gramian Matrix

Let $\hat{\mathbf{z}}_{N|i}$ denote the unique linear estimate of $\mathbf{z}_N$ that is based on the vectors $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$. That is, only the output vectors up to time i are used. By definition, this means that we should determine a coefficient matrix, say K_i^o, such that

$$\hat{\mathbf{z}}_{N|i} = K_i^o \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \end{bmatrix}, \tag{49}$$

and K_i^o is the unique stationary solution of

$$J(K_i) \triangleq < \mathbf{z}_N - K_i \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \end{bmatrix}, \mathbf{z}_N - K_i \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \end{bmatrix} >_{\mathcal{K}'} . \tag{50}$$

If we define

$$W_i \triangleq (R_0 \oplus R_1 \oplus \ldots \oplus R_i), \quad \Pi_i \triangleq (\Pi_0 \oplus Q_0 \oplus \ldots Q_{i-1}), \tag{51}$$

and

$$A_i \triangleq \begin{bmatrix} H_0 & & & \\ H_1 F^{[0,0]} & H_1 G_0 & & \\ H_2 F^{[1,0]} & H_2 F^{[1,1]} G_0 & & \\ \vdots & \vdots & \ddots & \\ H_i F^{[i-1,0]} & H_i F^{[i-1,1]} G_0 & \ldots & H_i G_{i-1} \end{bmatrix}, \tag{52}$$

then, as before, the problem (50) has a unique stationary solution K_i^o if, and only if,

$$W_i + \begin{bmatrix} A_i & 0 \end{bmatrix} \Pi \begin{bmatrix} A_i^* \\ 0 \end{bmatrix} = W_i + A_i \Pi_i A_i^* \text{ is invertible.}$$

A minimizing solution requires the positivity of this matrix. In any case, due to the block diagonal structure of $\{W, \Pi\}$ and due to the block lower-triangular structure of A, it is immediate to see that $(W_i + A_i \Pi_i A_i^*)$ is in fact a leading submatrix of $(W + A\Pi A^*)$.

To further clarify the implications of this observation, let R_y denote the Gramian matrix of the vector $\mathbf{y}$ in (27), i.e.,

$$R_y \triangleq < \mathbf{y}, \mathbf{y} >_{\mathcal{K}'} = W + A\Pi A^*. \tag{53}$$

The existence of a unique stationary solution K^o to $J(K)$ in (44) then requires the invertibility of R_y. Likewise, the existence of unique stationary solutions K_i^o in (50), for $0 \leq i < N$, requires the invertibility of the leading (block) submatrices of R_y. We shall therefore assume here that all the leading (block) submatrices of R_y are invertible in order to guarantee the existence of unique stationary solutions K_i^o to the estimation problems (50) for $0 \leq i \leq N$. In this case, we say that R_y is (block) strongly regular.

Under this assumption, we can introduce the unique (block) lower-diagonal-upper triangular factorization

$$R_y \triangleq LDL^*, \tag{54}$$

where L is chosen to have unit diagonal entries and D is a block diagonal matrix whose entries are denoted by

$$D \triangleq \{R_{e,0}, R_{e,1}, \ldots, R_{e,N}\}.$$

The sizes of the blocks $R_{e,i}$ are $p \times p$, in accordance with the $p \times 1$ dimension of each $\mathbf{y}_i$. Also, the (block) strong regularity of R_y guarantees the invertibility of the $\{R_{e,i}\}$.

6.3 Orthogonalization via the Gram-Schmidt Procedure

In this section we shall argue that, under the strong regularity condition on the Gramian matrix R_y, a recursive procedure that allows us to directly update $\hat{\mathbf{z}}_{N|i}$ to $\hat{\mathbf{z}}_{N|i+1}$ is possible without explicitly computing K_{i+1}^o. This will be first achieved by "orthogonalizing" the output vectors $\{\mathbf{y}_j\}$, as we now explain.

Introduce the variables $\{\mathbf{e}_i\}$ defined by (these variables are often known as the *innovation* variables in the signal processing literature)

$$\mathbf{e} \triangleq L^{-1}\mathbf{y} \quad \text{or} \quad L\mathbf{e} = \mathbf{y}, \tag{55}$$

where $\mathbf{e}$ denotes the collection of the $\mathbf{e}_i$,

$$\mathbf{e} \stackrel{\Delta}{=} \begin{bmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_N \end{bmatrix}.$$

It is immediate to conclude that the Gramian matrix of $\mathbf{e}$ is block diagonal since

$$< \mathbf{e}, \mathbf{e} >_{\mathcal{K}'} = < L^{-1}\mathbf{y}, L^{-1}\mathbf{y} >_{\mathcal{K}'} = L^{-1}R_y L^{-*} = D = (R_{e,0} \oplus R_{e,1} \oplus \ldots \oplus R_{e,N}).$$

Note that the vectors $\mathbf{e}$ and $\mathbf{y}$ are linearly related via an invertible transformation. They, therefore, span the same linear space. Also, and more importantly, the estimate of a variable $\mathbf{z}$ given the $\mathbf{y}$ is equal to the estimate of $\mathbf{z}$ given the $\mathbf{e}$. We prove this fact below and then discuss its ramifications.

Lemma 6.1 (Estimation Based on the $\{\mathbf{e}_i\}$) *Let $\hat{\mathbf{z}}$ denote the unique linear estimate of $\mathbf{z}$ given $\mathbf{y}$. That is, $\hat{\mathbf{z}} = K^o\mathbf{y}$, where K^o is the unique stationary solution of $< \mathbf{z} - K\mathbf{y}, \mathbf{z} - K\mathbf{y} >_{\mathcal{K}'}$. Let also $\hat{\mathbf{z}}_e$ denote the unique linear estimate of $\mathbf{z}$ given $\mathbf{e}$. That is, $\hat{\mathbf{z}}_e = K^{o,e}\mathbf{e}$, where $K^{o,e}$ is the unique stationary solution of $< \mathbf{z} - K^e\mathbf{e}, \mathbf{z} - K^e\mathbf{e} >_{\mathcal{K}'}$. Then $\hat{\mathbf{z}} = \hat{\mathbf{z}}_e$ and $K^o = K^{o,e}L^{-1}$.*

Proof: We know from (35) that estimating a variable $\mathbf{z}$ from $\mathbf{y}$ amounts to

$$\begin{aligned} \hat{\mathbf{z}} &= < \mathbf{z}, \mathbf{y} >_{\mathcal{K}'} < \mathbf{y}, \mathbf{y} >_{\mathcal{K}'}^{-1} \mathbf{y}, \\ &= < \mathbf{z}, L\mathbf{e} >_{\mathcal{K}'} < L\mathbf{e}, L\mathbf{e} >_{\mathcal{K}'}^{-1} L\mathbf{e}, \\ &= < \mathbf{z}, \mathbf{e} >_{\mathcal{K}'} < \mathbf{e}, \mathbf{e} >_{\mathcal{K}'}^{-1} \mathbf{e}, \\ &= \hat{\mathbf{z}}_e. \end{aligned}$$

■

The result also clearly holds for estimating $\mathbf{z}$ from a subcollection $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$. In other words, we can work with the $\{\mathbf{e}_i\}$ instead of the $\{\mathbf{y}_i\}$. This corresponds to a change of basis and its main advantage is that the $\{\mathbf{e}_i\}$ are orthogonal in $\mathcal{K}'$, i.e.,

$$< \mathbf{e}_i, \mathbf{e}_j >_{\mathcal{K}'} = R_{e,i}\delta_{ij}.$$

Lemma 6.2 (Recursive Computation) *Let $\hat{\mathbf{z}}_{N|N}$ denote the unique linear estimate of $\mathbf{z}_N$ that is based on the vectors $\{\mathbf{y}_i\}$ up to time N. Then it can be recursively updated as follows:*

$$\hat{\mathbf{z}}_{N|N} = \hat{\mathbf{z}}_{N|N-1} + < \mathbf{z}_N, \mathbf{e}_N >_{\mathcal{K}'} R_{e,N}^{-1}\mathbf{e}_N. \tag{56}$$

Proof: It follows from Lemma 6.1 that

$$\begin{aligned} \hat{\mathbf{z}} = \hat{\mathbf{z}}_{N|N} &= < \mathbf{z}_N, \mathbf{e} >_{\mathcal{K}'} < \mathbf{e}, \mathbf{e} >_{\mathcal{K}'}^{-1} \mathbf{e}, \\ &= \sum_{j=0}^{N} < \mathbf{z}_N, \mathbf{e}_j >_{\mathcal{K}'} < \mathbf{e}_j, \mathbf{e}_j >_{\mathcal{K}'}^{-1} \mathbf{e}_j \ , \\ &= \sum_{j=0}^{N-1} < \mathbf{z}_N, \mathbf{e}_j >_{\mathcal{K}'} < \mathbf{e}_j, \mathbf{e}_j >_{\mathcal{K}'}^{-1} \mathbf{e}_j \ + \ < \mathbf{z}_N, \mathbf{e}_N >_{\mathcal{K}'} < \mathbf{e}_N, \mathbf{e}_N >_{\mathcal{K}'}^{-1} \mathbf{e}_N, \\ &= \hat{\mathbf{z}}_{N|N-1} + < \mathbf{z}_N, \mathbf{e}_N >_{\mathcal{K}'} R_{e,N}^{-1}\mathbf{e}_N. \end{aligned}$$

■

For this recursive scheme to be complete, we still need to show the following. Given the state-space model (36),

(i) How to compute the $\{\mathbf{e}_i\}$?

(ii) How to compute the $\{R_{e,i}\}$?

(iii) How to compute the $\{< \mathbf{z}_N, \mathbf{e}_i >_{\mathcal{K}'}\}$?

6.4 Computation of the $\{\mathbf{e}_i\}$ via a Kalman-Type Procedure

The computation of the variables $\{\mathbf{e}_i\}$ can be achieved via a standard Gram-Schmidt procedure:

- Let $\mathbf{e}_0 = \mathbf{y}_0$.
- Then form $\mathbf{e}_1$ by subtracting from $\mathbf{y}_1$ its linear estimate that is based on $\mathbf{y}_0$, written as $\hat{\mathbf{y}}_{1|0}$,

$$\mathbf{e}_1 = \mathbf{y}_1 - \hat{\mathbf{y}}_{1|0} = \mathbf{y}_1 - <\mathbf{y}_1, \mathbf{y}_0>_{\mathcal{K}'} <\mathbf{y}_0, \mathbf{y}_0>_{\mathcal{K}'}^{-1} \mathbf{y}_0 = \mathbf{y}_1 - <\mathbf{y}_1, \mathbf{e}_0>_{\mathcal{K}'} <\mathbf{e}_0, \mathbf{e}_0>_{\mathcal{K}'}^{-1} \mathbf{e}_0.$$

- Then form $\mathbf{e}_2$ by subtracting from $\mathbf{y}_2$ its linear estimate that is based on $\{\mathbf{y}_0, \mathbf{y}_1\}$, written as $\hat{\mathbf{y}}_{2|1}$,

$$\mathbf{e}_1 = \mathbf{y}_2 - \hat{\mathbf{y}}_{2|1} = \mathbf{y}_2 - <\mathbf{y}_2, \mathbf{e}_0>_{\mathcal{K}'} <\mathbf{e}_0, \mathbf{e}_0>_{\mathcal{K}'}^{-1} \mathbf{e}_0 - <\mathbf{y}_2, \mathbf{e}_1>_{\mathcal{K}'} <\mathbf{e}_1, \mathbf{e}_1>_{\mathcal{K}'}^{-1} \mathbf{e}_1.$$

More generally, we have

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_{i|i-1}, \tag{57}$$

where $\hat{\mathbf{y}}_{i|i-1}$ denotes the linear estimate of $\mathbf{y}_i$ that is based on $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i-1}\}$. It is immediate to conclude from the second line of the state-equations (36), by linearity and by the fact that $<\mathbf{v}_i, \mathbf{y}_j>_{\mathcal{K}'} = 0$ for $j < i$, that

$$\hat{\mathbf{y}}_{i|i-1} = H_i \hat{\mathbf{x}}_{i|i-1},$$

where $\hat{\mathbf{x}}_{i|i-1}$ now denotes the linear estimate of $\mathbf{x}_i$ that is based on $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i-1}\}$. We thus see that

$$\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}, \tag{58}$$

and the computation of $\mathbf{e}_i$ is reduced to that of $\hat{\mathbf{x}}_{i|i-1}$.

Theorem 6.1 (Recursive Kalman Algorithm) *Consider the state-space model (36) and assume the Gramian matrix, $R_y = W + A\Pi A^*$, of the vector $\mathbf{y}$, defined in (38), is (block) strongly regular. The variables $\{\mathbf{e}_i\}$ defined via (55) or (57) can be recursively computed as follows. Start with $\hat{\mathbf{x}}_{0|-1} = 0$, $P_0 = \Pi_0$, and repeat for $i \geq 0$:*

$$\begin{aligned}
\mathbf{e}_i &= \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}, && (59)\\
\hat{\mathbf{x}}_{i+1|i} &= F_i \hat{\mathbf{x}}_{i|i-1} + K_{p,i} \mathbf{e}_i, && (60)\\
K_{p,i} &= F_i P_i H_i^* R_{e,i}^{-1}, && (61)\\
R_{e,i} &= R_i + H_i P_i H_i^*, && (62)\\
P_{i+1} &= F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*. && (63)
\end{aligned}$$

Proof: In view of the recursive formula (56) (taking $\mathbf{x}_{i+1}$ as the variable $\mathbf{z}$) we have

$$\hat{\mathbf{x}}_{i+1|i} \;=\; \hat{\mathbf{x}}_{i+1|i-1} + <\mathbf{x}_{i+1}, \mathbf{e}_i>_{\mathcal{K}'} R_{e,i}^{-1} \mathbf{e}_i \;=\; \hat{\mathbf{x}}_{i+1|i-1} + K_{p,i} \mathbf{e}_i, \tag{64}$$

where we have defined $K_{p,i} \stackrel{\Delta}{=} <\mathbf{x}_{i+1}, \mathbf{e}_i>_{\mathcal{K}'} <\mathbf{e}_i, \mathbf{e}_i>_{\mathcal{K}'}^{-1}$. It also follows from the first line of (36), and from the fact that $<\mathbf{u}_i, \mathbf{y}_j>_{\mathcal{K}'} = 0$ for $j < i$, that

$$\hat{\mathbf{x}}_{i+1|i-1} = F_i \hat{\mathbf{x}}_{i|i-1} + G_i \hat{\mathbf{u}}_{i|i-1} = F_i \hat{\mathbf{x}}_{i|i-1} + 0 = F_i \hat{\mathbf{x}}_{i|i-1}.$$

Substituting into (64) we obtain (60). To complete the argument we still need to show how to compute the $K_{p,i}$. Define the error quantity $\tilde{\mathbf{x}}_{i|i-1} \stackrel{\Delta}{=} \mathbf{x}_i - \hat{\mathbf{x}}_{i|i-1}$, and let P_i denote its Gramian matrix, $P_i \stackrel{\Delta}{=} <\tilde{\mathbf{x}}_{i|i-1}, \tilde{\mathbf{x}}_{i|i-1}>_{\mathcal{K}'}$. Then

$$\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1} = H_i \mathbf{x}_i - H_i \hat{\mathbf{y}}_{i|i-1} + \mathbf{v}_i = H_i \tilde{\mathbf{x}}_{i|i-1} + \mathbf{v}_i. \tag{65}$$

But it is immediate to note that $< \mathbf{v}_i, \tilde{\mathbf{x}}_{i|i-1} >_{\mathcal{K}'} = 0$ and, hence, (62) follows. Moreover,

$$< \mathbf{x}_{i+1}, \mathbf{e}_i >_{\mathcal{K}'} = F_i < \mathbf{x}_i, \mathbf{e}_i >_{\mathcal{K}'} + G_i < \mathbf{u}_i, \mathbf{e}_i >_{\mathcal{K}'} . \tag{66}$$

Now

$$< \mathbf{x}_i, \mathbf{e}_i >_{\mathcal{K}'} = < \mathbf{x}_i, \tilde{\mathbf{x}}_{i|i-1} >_{\mathcal{K}'} H_i^* + < \mathbf{x}_i, \mathbf{v}_i >_{\mathcal{K}'} = P_i H_i^* + 0,$$

while

$$< \mathbf{u}_i, \mathbf{e}_i >_{\mathcal{K}'} = < \mathbf{u}_i, \tilde{\mathbf{x}}_{i|i-1} >_{\mathcal{K}'} H_i^* + < \mathbf{u}_i, \mathbf{v}_i >_{\mathcal{K}'} = 0,$$

so that we can write

$$K_{p,i} \stackrel{\Delta}{=} < \mathbf{x}_{i+1}, \mathbf{e}_i >_{\mathcal{K}'} R_{e,i}^{-1} = F_i P_i H_i^* R_{e,i}^{-1}. \tag{67}$$

Therefore $\{K_{p,i}, R_{e,i}\}$ can be determined once we have the Gramian matrices $\{P_i\}$.

The most direct method for computing the $\{P_i\}$ is to seek a recursion for $\tilde{\mathbf{x}}_{i+1|i}$ and then form P_{i+1}. In fact, from the model equations (36) and the estimator equation (60) we obtain

$$\tilde{\mathbf{x}}_{i+1|i} \quad = \quad F_{p,i} \tilde{\mathbf{x}}_{i|i-1} + \begin{bmatrix} G_i & -K_{p,i} \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix},$$

where we have defined $F_{p,i} = F_i - K_{p,i} H_i$. Now it follows that P_i obeys the recursion (63).

■

We should remark here that the above recursive formulas extend the so-called Kalman filter to an indefinite metric space [HSK93]. The recursions have exactly the same form as those of the Kalman filter, except for the fact that the Gramian matrices $\{\Pi_0, R_i, Q_i\}$ are allowed to be indefinite. Also, the recursion (63) for P_i (with (61) and (62) inserted in (63)) is known as the Riccati difference equation.

An important fall out of the above algorithm is that the inertia of the Gramian matrix $< \mathbf{y}, \mathbf{y} >_{\mathcal{K}'}$ is completely determined by the inertia of the $\{R_{e,i}\}$.

Corollary 6.1 (Inertia of the Gramian Matrix) *Consider the state-space model (36) and let R_y denote the Gramian matrix of the vector $\mathbf{y}$ defined in (38), viz.,*

$$R_y \;=\; W + A\Pi A^*,$$

where $\{W, \Pi, A\}$ are as defined in (40), (41), and (42). The R_y is further assumed (block) strongly regular. Then

$$\textit{Inertia of}\,(W + A\Pi A^*) \;=\; \textit{Inertia of}\,(R_{e,0} \oplus R_{e,1} \oplus \ldots \oplus R_{e,N}). \tag{68}$$

Proof: This follows from the congruence relation $R_y = LDL^*$, where $D = (R_{e,0} \oplus R_{e,1} \oplus \ldots \oplus R_{e,N})$.

■

6.5 Recursive Estimation of $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{N-1}\}$

We already know how to recursively evaluate the $\{\mathbf{e}_i\}$ and the corresponding Gramian matrices $\{R_{e,i}\}$. We now return to (56), viz.,

$$\hat{\mathbf{z}}_{N|N} = \hat{\mathbf{z}}_{N|N-1} + < \mathbf{z}_N, \mathbf{e}_N >_{\mathcal{K}'} R_{e,N}^{-1} \mathbf{e}_N, \tag{69}$$

and show how to evaluate the terms $\{< \mathbf{z}_N, \mathbf{e}_i >_{\mathcal{K}'}\}$. Once this is done, we shall have an algorithm for the recursive update of the estimates $\{\hat{\mathbf{z}}_{N|i}\}$. Recall that $\hat{\mathbf{z}}_{N|i}$ was defined as the unique linear estimate of $\mathbf{z}_N$ based on the $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$.

Theorem 6.2 (Recursive Smoothing Solution) *Assume R_y is (block) strongly regular. Then the stationary solution $\hat{\mathbf{z}}$ is equal to $\hat{\mathbf{z}}_{N|N}$, where $\hat{\mathbf{z}}_{N|N}$ can be recursively computed as follows: start with $\hat{\mathbf{z}}_{N|-1} = 0$ and repeat for $i = 0, 1, \ldots, N$:*

$$\hat{\mathbf{z}}_{N|i} = \hat{\mathbf{z}}_{N|i-1} + K_{z,i} H_i^* R_{e,i}^{-1} \mathbf{e}_i,$$

where

$$K_{z,i+1} = K_{z,i} \left[F_i - K_{p,i} H_i\right]^* + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} Q_i G_i^*, \quad K_{z,0} = \begin{bmatrix} \Pi_0 \\ 0 \end{bmatrix}.$$

The identity matrix in the recursion for $K_{z,i+1}$ occurs at the position that corresponds to the entry $\mathbf{u}_i$.

Proof: Recall that $\mathbf{e}_i = H_i \tilde{\mathbf{x}}_{i|i-1} + \mathbf{v}_i$. Therefore,

$$\begin{aligned} \hat{\mathbf{z}}_{N|i} &= \hat{\mathbf{z}}_{N|i-1} + < \mathbf{z}_N, \mathbf{e}_i >_{\mathcal{K}'} R_{e,i}^{-1} \, \mathbf{e}_i \,, \\ &= \hat{\mathbf{z}}_{N|i-1} + < \mathbf{z}_N, \tilde{\mathbf{x}}_{i|i-1} >_{\mathcal{K}'} H_i^* R_{e,i}^{-1} \, \mathbf{e}_i . \end{aligned}$$

We now define $K_{z,i} \stackrel{\Delta}{=} < \mathbf{z}_N, \tilde{\mathbf{x}}_{i|i-1} >_{\mathcal{K}'}$, and note that

$$\begin{aligned} K_{z,i+1} = < \mathbf{z}_N, \tilde{\mathbf{x}}_{i+1|i} >_{\mathcal{K}'} &= < \mathbf{z}_N, \left[F_i \tilde{\mathbf{x}}_{i|i-1} - K_{p,i}\mathbf{e}_i + G_i \mathbf{u}_i\right] >_{\mathcal{K}'} \,, \\ &= K_{z,i} \left[F_i - K_{p,i} H_i\right]^* + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} Q_i G_i^* . \end{aligned}$$

■

A remark is due here. Recall that we have defined $\hat{\mathbf{z}}_{N|i}$ in (50) as the unique linear estimator of $\mathbf{z}_N$ that is based on the vectors $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$. Now $\mathbf{z}_N$ is a vector containing the $\{\mathbf{x}_0, \mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}\}$. By linearity, it follows that the entries of $\hat{\mathbf{z}}_{N|i}$ can be interpreted as the linear estimates of the corresponding entries of $\mathbf{z}_N$ given the $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$. That is, we have

$$\hat{\mathbf{z}}_{N|i} = \begin{bmatrix} \hat{\mathbf{x}}_{0|i} \\ \hat{\mathbf{u}}_{0|i} \\ \hat{\mathbf{u}}_{1|i} \\ \vdots \\ \hat{\mathbf{u}}_{N-1|i} \end{bmatrix},$$

where the notation $\hat{\mathbf{x}}_{0|i}$ denotes the linear estimate of $\mathbf{x}_0$ that is based on $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$. Likewise, $\hat{\mathbf{u}}_{j|i}$ denotes the linear estimate of $\mathbf{u}_j$ that is based on the same vectors $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$. But it follows from (37) that $< \mathbf{u}_j, \mathbf{y}_k >_{\mathcal{K}'} = 0$ for all $j \geq k$. This implies that

$$\hat{\mathbf{u}}_{i|i} = \hat{\mathbf{u}}_{i+1|i} = \ldots = \hat{\mathbf{u}}_{N-1|i} = 0.$$

Consequently, the last entries of $\hat{\mathbf{z}}_{N|i}$ are in fact zero,

$$\hat{\mathbf{z}}_{N|i} = \begin{bmatrix} \hat{\mathbf{x}}_{0|i} \\ \hat{\mathbf{u}}_{0|i} \\ \vdots \\ \hat{\mathbf{u}}_{i-1|i} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{70}$$

If we introduce the definition of $\mathbf{z}_i$ as in (48), i.e., a vector composed of $\mathbf{x}_0$ and the $\{\mathbf{u}_j\}$ up to time $(i-1)$, then we can rewrite (70) more compactly as follows:

$$\hat{\mathbf{z}}_{N|i} = \begin{bmatrix} \hat{\mathbf{z}}_{i|i} \\ 0 \end{bmatrix}. \tag{71}$$

That is, the leading nonzero entries of the successive $\hat{\mathbf{z}}_{N|i}$ are precisely the entries of $\hat{\mathbf{z}}_{i|i}$.

7 A Recursive IWLS Problem in the Presence of State-Space Structure

In order to further appreciate the results of the earlier sections, let us first summarize what has been concluded in the state-space context.

Starting with a state-space model (36), with entries in an indefinite metric space $\mathcal{K}'$, we defined two vectors $\mathbf{z}$ and $\mathbf{y}$ as in (38) and (39). The vector $\mathbf{y}$ contained the output vectors $\{\mathbf{y}_i\}$ and the vector $\mathbf{z}$ contained the vectors $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{N-1}\}$. We then used $\mathbf{z}$ and $\mathbf{y}$ as a motivation to introduce a quadratic minimization problem. This was achieved by defining the linear estimate of $\mathbf{z}$ given $\mathbf{y}$ as the vector $\hat{\mathbf{z}}$ obtained via $\hat{\mathbf{z}} = K^o\mathbf{y}$, where K^o was defined as the unique stationary solution of the cost function

$$J(K) = <\mathbf{z} - K\mathbf{y}, \mathbf{z} - K\mathbf{y}>_{\mathcal{K}'} = \Pi - KA\Pi - \Pi A^*K^* + K[A\Pi A^* + W]K^*. \tag{72}$$

We then observed that $J(K)$ is a special case of the optimization problem (2) introduced earlier in the paper, and hence the solution $\hat{\mathbf{z}}$, also denoted by $\hat{\mathbf{z}}_{N|N}$, could be obtained via the global expression (45),

$$\hat{\mathbf{z}} = \left[\Pi^{-1} + A^*W^{-1}A\right]^{-1} A^*W^{-1}\mathbf{y}.$$

But we further showed that in this case, and due to the state-space assumptions (36) and (37), the matrices $\{\Pi, W, A\}$ have extra structure in them. In particular, the $\{\Pi, W\}$ were shown to be diagonal matrices in (41) and (42), and the A matrix was shown to be block lower triangular in (40). As a result, we then argued that this structure can in fact be exploited in order to derive a recursive scheme that would allow us to directly update the estimate $\hat{\mathbf{z}}_{N|i}$ to $\hat{\mathbf{z}}_{N|i+1}$, starting with $\hat{\mathbf{z}}_{N|-1} = 0$ and ending with the desired solution $\hat{\mathbf{z}}_{N|N}$. This was achieved by the recursions of Theorem 6.2, which in turn rely on the recursions of Theorem 6.1. These recursions assume that the Gramian matrix R_y is (block) strongly regular so that the stationary solutions K_i^o that correspond to each estimate $\hat{\mathbf{z}}_{N|i}$ are uniquely defined.

Now, in view of the discussion at the beginning of Sec. 5, the above solution $\hat{\mathbf{z}}_{N|N}$ has the same expression as the solution $\hat{z}$ of a related minimization problem of the form (1). Indeed, it is rather immediate to write down the IWLS problem whose stationary point matches the above $\hat{\mathbf{z}}$ (or $\hat{\mathbf{z}}_{N|N}$). We simply use (72) to conclude that the related problem of the form (1) is the following:

$$\min_{z=\begin{bmatrix} x_0 \\ u \end{bmatrix}} \left\{ \begin{bmatrix} x_0 \\ u \end{bmatrix}^* \Pi^{-1} \begin{bmatrix} x_0 \\ u \end{bmatrix} + \left(y - A\begin{bmatrix} x_0 \\ u \end{bmatrix}\right)^* W^{-1} \left(y - A\begin{bmatrix} x_0 \\ u \end{bmatrix}\right) \right\}. \tag{73}$$

Equivalently, using (42), (41), and (40), this can also be written as

$$\min_{\{x_0, u_0, \ldots, u_{N-1}\}} \left[x_0^*\Pi_0^{-1}x_0 + \sum_{j=0}^{N}(y_j - H_jx_j)^*R_j^{-1}(y_j - H_jx_j) + \sum_{j=0}^{N-1} u_j^*Q_j^{-1}u_j \right], \tag{74}$$

subject to

$$x_{j+1} = F_jx_j + G_ju_j. \tag{75}$$

Likewise, the IWLS problem whose stationary solution $\hat{z}_i$ matches the $\hat{\mathbf{z}}_{i|i}$ is

$$\min_{\{x_0,u_0,\ldots,u_{i-1}\}} \quad \left[x_0^*\Pi_0^{-1}x_0 + \sum_{j=0}^{i}(y_j - H_jx_j)^*R_j^{-1}(y_j - H_jx_j) + \sum_{j=0}^{i-1} u_j^*Q_j^{-1}u_j \right], \tag{76}$$

subject to $x_{j+1} = F_jx_j + G_ju_j$. That is, only vectors $\{y_j\}$ up to time i are included. The stationary solution $\hat{z}_{i|i}$ exists and is unique if, and only if, using (51) and (52),

$$\Pi_i^{-1} + A_i^*W_i^{-1}A_i \text{ is invertible.}$$

This implies, in view of Lemma 2.2, that $(W_i + A_i\Pi A_i^*)$ is also invertible. We thus have the following preliminary conclusion, which shows that the strong regularity assumption that we imposed earlier on the Gramian matrix R_y is not a restriction. It is in fact necessary if we are interested in all the stationary solutions $\{\hat{z}_{i|i}\}$.

Lemma 7.1 (Strong Regularity) *The stationary solutions $\hat{z}_{i|i}$ are uniquely defined for all $0 \leq i \leq N$ if, and only if, the matrix $(W + A\Pi A^*)$ is (block) strongly regular.*

Proof: Since $\{W, \Pi\}$ are block diagonal and A is block lower triangular, the (block) leading submatrices of $(W+A\Pi A^*)$ are of the form $(W_i+A_i\Pi A_i^*)$. But we argued above that $\hat{z}_{i|i}$ is uniquely defined iff $(W_i+A_i\Pi A_i^*)$ is invertible. Since this holds for all $0 \leq i \leq N$, we conclude that $(W + A\Pi A^*)$ is necessarily (block) strongly regular.

■

In other words, recall that we have established earlier in Lemma 5.1 that the standard optimization problems (1) and (2) are always guaranteed to simultaneously have unique stationary solutions $\hat{z}$ and K^o (and also $\hat{\mathbf{z}}$). The above result then extends this conclusion to the successive solutions $\{\hat{z}_{i|i}, \hat{\mathbf{z}}_{i|i}\}$ of (50) and (76). That is, when state-space structure is incorporated into both optimization criteria, and recursive stationarization is employed, it also holds that the criteria have simultaneous stationary points.

Problem 7.1 (The IWLS State-Space Problem) *For each i, define the quadratic cost function*

$$J_i(x_0, u_0, \ldots, u_{i-1}) \triangleq \left[x_0^*\Pi_0^{-1}x_0 + \sum_{j=0}^{i}(y_j - H_jx_j)^*R_j^{-1}(y_j - H_jx_j) + \sum_{j=0}^{i-1} u_j^*Q_j^{-1}u_j \right]. \tag{77}$$

We are interested in minimizing, when possible, the J_i over $(x_0, u_0, \ldots, u_{i-1})$, for all $0 \leq i \leq N$, and subject to the state-space constraint $x_{i+1} = F_ix_i + G_iu_i$.

Before stating the conditions that would allow us to check whether the existence of minima for all J_i exist, we shall first consider the following:

(i) We shall show how to recursively compute the unique stationary points $\{\hat{z}_{i|i}\}$ when they exist.

(ii) We shall then derive conditions for these points to be minima.

In order to highlight the possibilities that may occur in the indefinite case, let us assume for now that the $\{J_i\}$ have unique stationary points $\{\hat{z}_{i|i}\}$, so that $(W + A\Pi A^*)$ is guaranteed to be (block) strongly regular, as proven in Lemma 7.1.

Now, each one of the stationary points $\hat{z}_{i|i}$ may or may not be a minimum in its own right, and this is independent of whether among the earlier solutions $\{\hat{z}_{j|j}\}_{j<i}$ we have minima or not. This is in contrast to the recursive minimization of quadratic cost functions with positive-definite weighting matrices, where all the solutions $\hat{z}_{i|i}$ are guaranteed to be minima. In the indefinite case however, it may happen that at a particular time instant, say the i^{th} instant, the $\hat{z}_{i|i}$ is a minimum of J_i, while in the next time instant, the $\hat{z}_{i+1|i+1}$ is

not a minimum of J_{i+1}. This is because, the minimality of one requires the positivity of $(\Pi_i^{-1} + A_i^* W_i^{-1} A_i)$, while the minimality of the other requires the positivity of $(\Pi_{i+1}^{-1} + A_{i+1}^* W_{i+1}^{-1} A_{i+1})$, and the positivity of these two matrices do not imply each other. In particular, the second matrix contains new entries, such as Q_i, R_{i+1}, and an extra row in A_{i+1}. These entries can destroy the positivity of $(\Pi_{i+1}^{-1} + A_{i+1}^* W_{i+1}^{-1} A_{i+1})$. This situation does not occur with positive-definite quadratic forms because, in this case, the weighting matrices $\{\Pi, W\}$ are positive-definite and hence, $(\Pi_i^{-1} + A_i^* W_i^{-1} A_i)$ is positive-definite for all i.

7.1 Fundamental Inertia Conditions

The following result, for example, establishes under what condition J_N has a minimum at $\hat{z}_{N|N}$.

Lemma 7.2 (Minimization of J_N) *Consider a quadratic cost function as in (77) and subject to $x_{i+1} = F_i x_i + G_i u_i$. The quantities $\{x_0, u_0, \ldots, u_{N-1}\}$ are the unknowns. Let $m \times m$ denote the size of each Q_i. Likewise, let $n \times n$ denote the size of Π_0. Define*

$$\Pi \triangleq (\Pi_0 \oplus Q_0 \ldots \oplus Q_{N-1}), \quad W \triangleq (R_0 \oplus R_1 \oplus \ldots \oplus R_N).$$

Assume $(W + A\Pi A^*)$ is (block) strongly regular (i.e., the J_i are guaranteed to have unique stationary points $\hat{z}_{i|i}$ for all $0 \le i \le N$). Then J_N has a minimum with respect to these unknowns (i.e., the last stationary point $\hat{z}_{N|N}$ is a minimum) if, and only if,

$$\begin{aligned} I_-[\Pi \oplus W] &= I_-\{R_{e,0} \oplus \ldots \oplus R_{e,N}\}, \\ I_+[\Pi \oplus W] &= I_+\{R_{e,0} \oplus \ldots \oplus R_{e,N}\} + n + mN, \end{aligned}$$

where the matrices $\{R_{e,i}\}$ are recursively computed as follows:

$$R_{e,i} = H_i P_i H_i^* + R_i,$$

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0,$$

$$K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1}.$$

Proof: Note here that the size of Π is $(n+mN) \times (n+mN)$. It then follows from Theorem 5.1 that problem (74) (or, equivalently, (73)) has a minimum if, and only if,

$$\begin{aligned} I_-[W + A\Pi A^*] &= I_-[\Pi \oplus W], \\ I_+[W + A\Pi A^*] &= I_+[\Pi \oplus W] - n - mN, . \end{aligned}$$

The result of the lemma now follows by invoking Corollary 6.1, which states that the matrix $(W + A\Pi A^*)$ has the same inertia as $\{R_{e,0} \oplus \ldots \oplus R_{e,N}\}$. This last statement holds as a result of the strong regularity of $(W + A\Pi A^*)$.

∎

An immediate conclusion is the following special case where the Π matrix is itself positive-definite and, hence, its negative inertia is zero while its positive-inertia is equal to the number of its columns (or rows), $n + mN$.

Corollary 7.1 (A Special Case) *Consider the same setting of Lemma 7.2. Assume further that $\Pi_0 > 0$ and the $\{Q_i\}_{i=0}^{N-1}$ are positive-definite. Then J_N has a minimum with respect to z_N if, and only if,*

$$\begin{aligned} I_-\{R_0 \oplus \ldots \oplus R_N\} &= I_-\{R_{e,0} \oplus \ldots \oplus R_{e,N}\}, \\ I_+\{R_0 \oplus \ldots \oplus R_N\} &= I_+\{R_{e,0} \oplus \ldots \oplus R_{e,N}\}. \end{aligned}$$

The above results were concerned with the existence of a minimum for the last cost function J_N. More generally, we are interested in checking whether each $\hat{z}_{i|i}$ is a minimum of the corresponding J_i. This is addressed in the following statement.

Theorem 7.1 (Recursive Minimization of $\{J_i\}$) *Consider a quadratic cost function as in (77) and subject to $x_{i+1} = F_i x_i + G_i u_i$. The quantities $\{x_0, u_0, \ldots, u_{N-1}\}$ are the unknowns. Let $m \times m$ denote the size of each Q_i. Likewise, let $n \times n$ denote the size of Π_0. Define*

$$\Pi \triangleq (\Pi_0 \oplus Q_0 \ldots \oplus Q_{N-1}), \quad W \triangleq (R_0 \oplus R_1 \oplus \ldots \oplus R_N).$$

Then each J_i has a minimum with respect to $\{x_0, u_0, \ldots, u_{i-1}\}$ if, and only if,

$$\begin{aligned} I_-\left[\Pi_0 \oplus R_0\right] &= I_-\{R_{e,0}\}, & (78)\\ I_+\left[\Pi_0 \oplus R_0\right] &= I_+\{R_{e,0}\} + n, & (79) \end{aligned}$$

and, for $i = 1, 2, \ldots, N$,

$$\begin{aligned} I_-\{Q_{i-1} \oplus R_i\} &= I_-\{R_{e,i}\}, & (80)\\ I_+\{Q_{i-1} \oplus R_i\} &= I_+\{R_{e,i}\} + m. & (81) \end{aligned}$$

Moreover, when the stationary solutions (or minima) of the J_i are uniquely defined, the value of each J_i at its unique stationary solution (or minimum) $\hat{z}_{i|i}$ is given by

$$J_i(\hat{z}_{i|i}) = \sum_{j=0}^{i} e_i^* R_{e,i}^{-1} e_i, \tag{82}$$

where $e_i = (y_i - H_i \hat{x}_{i|i-1})$.

Proof: The proof is by induction. Minimizing J_0 over x_0 requires the inertia conditions (78) and (79), as is obvious for example from Lemma 7.2 specialized to $N = 0$. Likewise, the minimization of J_1 requires

$$\begin{aligned} I_-\left[\Pi_0 \oplus Q_0 \oplus R_0 \oplus R_1\right] &= I_-\{R_{e,0} \oplus R_{e,1}\},\\ I_+\left[\Pi_0 \oplus Q_0 \oplus R_0 \oplus R_1\right] &= I_+\{R_{e,0} \oplus R_{e,1}\} + n + m, \end{aligned}$$

which by virtue of (78) and (79) yield (80) and (81) for $i = 1$. Continuing in this fashion we establish the result for $i > 1$.

To establish (82) we recall that the value of a quadratic cost function of the form (1) at its stationary solution is given by (22), which in the present context translates to

$$J_i(\hat{z}_{i|i}) = \begin{bmatrix} y_0^* & y_1^* & \ldots & y_i^* \end{bmatrix} \left[W_i + A_i \Pi_i A_i^*\right]^{-1} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_i \end{bmatrix}.$$

But we know from the discussion in the earlier section (viz., (53) and (55)) that if we introduce the triangular factorization of the matrix $(W_i + A_i \Pi_i A_i^*)$, say

$$(W_i + A_i \Pi_i A_i^*) = L_i D_i L_i^*,$$

then

$$L_i \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_i \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_i \end{bmatrix},$$

and $D_i = (R_{e,0} \oplus \ldots \oplus R_{e,i})$. Consequently,

$$J_i(\hat{z}_{i|i}) = \begin{bmatrix} e_0^* & e_1^* & \ldots & e_i^* \end{bmatrix} D_i^{-1} \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_i \end{bmatrix} = \sum_{j=0}^{i} e_i^* R_{e,i}^{-1} e_i .$$

■

It is also clear from the discussions in Sec. 5 that the recursions of Theorem 6.2, with the proper identifications $\hat{\mathbf{z}}_{N|i} \leftarrow \hat{z}_{N|i}$, $\mathbf{y}_i \leftarrow y_i$, $\hat{\mathbf{x}}_{i|i-1} \leftarrow \hat{x}_{i|i-1}$, $\mathbf{u}_i \leftarrow u_i$, can be used to compute the stationary solutions $\hat{z}_{i|i}$ of (77). In particular, and according to the discussions that led to (70), we also have that the stationary solutions $\hat{z}_{i|i}$ are related to the $\hat{z}_{N|i}$, given below in the statement of the theorem, as follows:

$$\hat{z}_{N|i} = \begin{bmatrix} \hat{x}_{0|i} \\ \hat{u}_{0|i} \\ \vdots \\ \hat{u}_{i-1|i} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{z}_{i|i} \\ 0 \end{bmatrix} . \tag{83}$$

That is, the leading entries of $\hat{z}_{Ni|i}$ denote the stationary solution of J_i with respect to $\{x_0, u_0, \ldots, u_{i-1}\}$.

Theorem 7.2 (Recursive Solution of (77)) *Consider a quadratic cost function as in (77) and subject to $x_{i+1} = F_i x_i + G_i u_i$. The quantities $\{x_0, u_0, \ldots, u_{N-1}\}$ are the unknowns. Assume $(W + A\Pi A^*)$ is (block) strongly regular with $\{W, A, \Pi\}$ defined as in (40), (41), and (42). Let*

$$z_N \triangleq \begin{bmatrix} x_0 \\ u_0 \\ \vdots \\ u_{N-1} \end{bmatrix} .$$

The stationary solution, $\hat{z}_{i|i}$, of

$$\min_z \left[x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} (y_j - H_j x_j)^* R_j^{-1} (y_j - H_j x_j) + \sum_{j=0}^{i-1} u_j^* Q_j^{-1} u_j \right] , \tag{84}$$

can be recursively computed as follows: start with $\hat{z}_{N|-1} = 0$ and repeat for $i = 0, 1, \ldots, N$:

$$\hat{z}_{N|i} = \hat{z}_{N|i-1} + K_{z,i} H_i^* R_{e,i}^{-1} \, (y_i - H_i \hat{x}_{i|i-1}),$$

where

$$K_{z,i+1} = K_{z,i} \left[F_i - K_i R_{e,i}^{-1} H_i \right]^* + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} Q_i G_i^* , \quad K_{z,0} = \begin{bmatrix} \Pi_0 \\ 0 \end{bmatrix} ,$$

and

$$\hat{x}_{i+1|i} = F_i \hat{x}_{i|i-1} + K_{p,i} (y_i - H_i \hat{x}_{i|i-1}), \quad \hat{x}_{0|-1} = 0.$$

Remark. It may happen that the last term in the definition of the quadratic cost function J_i in (77) also includes the extra tem $u_i^* Q_i^{-1} u_i$, say

$$J_i(x_0, u_0, \ldots, u_{i-1}) \triangleq \left[x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} (y_j - H_j x_j)^* R_j^{-1} (y_j - H_j x_j) + \sum_{j=0}^{i} u_j^* Q_j^{-1} u_j \right]. \tag{85}$$

In this case, the unknown variable u_i only appears in the quadratic term $u_i^* Q_i^{-1} u_i$, and it thus follows that minimization with respect to the u_i requires the positivity of Q_i. Hence, successive minimization of the J_i would additionally require that the $\{Q_i\}$ be positive-definite, which is a special case that often arises in the context of H^∞-problems, with the additional constraint $\Pi_0 > 0$. It is thus rather immediate to handle this case. All we need to do is to simply impose a positivity condition on the $\{Q_i\}$. This motivates us to consider the following two corollaries.

Corollary 7.2 (Some Positive Weighting Matrices) *Consider the same setting as in Theorem 7.1 and further assume that the* $\{Q_i\}_{i=0}^{N-1}$ *are positive-definite. Assume also that* $\Pi_0 > 0$. *Then each* J_i *has a minimum with respect to* $\{x_0, u_0, \ldots, u_{i-1}\}$ *if, and only if, for all* i,

$$\textit{Inertia}\{R_i\} = \textit{Inertia}\{R_{e,i}\}. \tag{86}$$

In this case, it follows that

$$P_i \geq 0 \quad \textit{for} \quad 0 \leq i \leq N. \tag{87}$$

[In fact, P_0 *is strictly positive since it is equal to* Π_0*].*

Proof: The inertia conditions (86) follow immediately as a special case of Theorem 7.1. We now establish the nonnegativity of the Riccati variables $\{P_i\}$. This is achieved by induction. Assume the result is valid up to time j, i.e., $\{P_0, P_1, \ldots, P_j\}$ are nonnegative-definite and let us prove that P_{j+1} is also nonnegative-definite.

It follows from (86) that $R_{e,j} = (R_j + H_j P_j H_j^*)$ and R_j must have the same inertia and, consequently, that $(R_j + H_j P_j H_j^*)$ is invertible.

Since P_j is nonnegative-definite, we can factor it into $P_j = M_j M_j^*$, where the number of columns of M_j is equal to the rank of P_j. Defining $\bar{H}_j \triangleq H_j M_j$ we can write $(R_j + H_j P_j H_j^*) = (R_j + \bar{H}_j \bar{H}_j^*)$.

The invertibility of $(R_j + \bar{H}_j \bar{H}_j^*)$ now implies, by virtue of Lemma 2.2, that $(I + \bar{H}_j^* R_j^{-1} H_j)$ is also invertible. Using the result of Theorem 2.1 we have that

$$\begin{aligned} I_+(I \oplus R_j) &= I_+[(I + \bar{H}_j^* R_j^{-1} \bar{H}_j) \oplus (R_j + \bar{H}_j \bar{H}_j^*)], \\ I_-(I \oplus R_j) &= I_-[(I + \bar{H}_j^* R_j^{-1} \bar{H}_j) \oplus (R_j + \bar{H}_j \bar{H}_j^*)]. \end{aligned}$$

But since

$$\text{Inertia}\{R_j + \bar{H}_j \bar{H}_j^*\} = \text{Inertia}\{R_j\},$$

we conclude that I and $(I + \bar{H}_j^* R_j^{-1} \bar{H}_j)$ must have the same inertia and, hence, $(I + \bar{H}_j^* R_j^{-1} H_j) > 0$. Now the Riccati recursion (63) implies that

$$\begin{aligned} P_{j+1} &= F_j \left[P_j - P_j H_j^* (R_j + H_j P_j H_j^*)^{-1} H_j P_j \right] F_j^* + G_j Q_j G_j^*, \\ &= F_j M_j \left[I - \bar{H}_j^* (R_j + \bar{H}_j \bar{H}_j^*)^{-1} \bar{H}_j \right] M_j^* F_j^* + G_j Q_j G_j^*, \\ &= F_j M_j \left[I + \bar{H}_j^* R_j^{-1} \bar{H}_j \right]^{-1} M_j^* F_j^* + G_j Q_j G_j^*. \end{aligned}$$

But since $(I + \bar{H}_j^* R_j^{-1} \bar{H}_j) > 0$ and $G_j Q_j G_j^* \geq 0$, we conclude that $P_{j+1} \geq 0$. ■

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

Corollary 7.3 (Positive Weights and Invertible $\{F_i\}$) *Consider the same setting as in Theorem 7.1 and further assume that the $\{Q_i\}_{i=0}^{N-1}$ are positive-definite. Assume also that $\Pi_0 > 0$ and that the $\{F_i\}$ are invertible. Then the following two statements provide equivalent necessary and sufficient conditions for each J_i to have a minimum with respect to $\{x_0, u_0, \ldots, u_{i-1}\}$.*

(i) All $\{J_i\}$ have minima iff, for $0 \le i \le N$,

$$P_i^{-1} + H_i^* R_i^{-1} H_i > 0. \tag{88}$$

(ii) All $\{J_i\}$ have minima iff, for $0 \le i \le N$,

$$P_{i+1} - G_i Q_i G_i^* > 0\ . \tag{89}$$

It further follows in the minimum case that, for all i,

$$P_{i+1} > 0. \tag{90}$$

Proof: A simple inductive argument establishes the result. It follows from Corollary 7.2 that $R_{e,0} = (R_0 + H_0 \Pi_0 H_0^*)$ and R_0 must have the same inertia and, consequently, that $(R_0 + H_0 \Pi_0 H_0^*)$ is invertible. Lemma 2.2 then implies that $(\Pi_0^{-1} + H_0^* R_0^{-1} H_0)$ is also invertible. Using the result of Theorem 2.1 we have that

$$\begin{aligned} I_+(\Pi_0 \oplus R_0) &= I_+[(\Pi_0^{-1} + H_0^* R_0^{-1} H_0) \oplus (H_0 \Pi_0 H_0^* + R_0)], \\ I_-(\Pi_0 \oplus R_0) &= I_-[(\Pi_0^{-1} + H_0^* R_0^{-1} H_0) \oplus (H_0 \Pi_0 H_0^* + R_0)]. \end{aligned}$$

But since

$$\text{Inertia}\{R_0 + H_0 \Pi_0 H_0^*\} = \text{Inertia}\{R_0\},$$

we conclude that Π_0 and $(\Pi_0^{-1} + H_0^* R_0^{-1} H_0)$ must have the same inertia and, hence, $(\Pi_0^{-1} + H_0^* R_0^{-1} H_0) > 0$ since $\Pi_0 > 0$. Now the Riccati recursion (63) implies that

$$\begin{aligned} P_1 &= F_0 \left[\Pi_0 - \Pi_0 H_0^* (R_0 + H_0 \Pi_0 H_0^*)^{-1} H_0 \Pi_0\right] F_0^* + G_0 Q_0 G_0^*, \\ &= F_0 \left[\Pi_0^{-1} + H_0^* R_0^{-1} H_0\right]^{-1} F_0^* + G_0 Q_0 G_0^*. \end{aligned}$$

The invertibility of F_0 guarantees the positive-definiteness of $F_0 \left[\Pi_0^{-1} + H_0^* R_0^{-1} H_0\right]^{-1} F_0^*$. But since $Q_0 > 0$ we also have that $G_0 Q_0 G_0^* \ge 0$. Consequently, $P_1 > 0$. We can now repeat the argument to conclude that the conditions (88) hold for all i.

The equivalence of conditions (88) and (89) follow from the fact that for all i we have

$$P_{i+1} - G_i Q_i G_i^* \quad = \quad F_i \left[P_i^{-1} + H_i^* R_i^{-1} H_i\right]^{-1} F_i^*.$$

■

Conditions of the form (88) are the ones most cited in $H^\infty-$applications (e.g., [YS91]). Here we see that they are related to the inertia conditions (86). These inertia conditions also arise in the $H^\infty-$context (see, e.g., [GL95][p. 495] and Lemma 8.1 further ahead), where R_i has the additional structure $R_i = (-\gamma^2 I \oplus I)$. Here, we have derived these conditions as special cases of the general statement of Theorem 7.1, which holds for arbitrary indefinite matrices $\{\Pi_0, Q_i, R_i\}$, while the $H^\infty-$results hold only for positive-definite matrices $\{\Pi_0, Q_i\}$ and for matrices R_i of the above form. Note also that testing for (88) not only requires that we compute the P_i (via a Riccati recursion (63)), but also that we invert P_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 given by (86), 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, as pursued in [HSK94].

8 An Application to $\mathbf{H}^\infty$-Filtering

We now illustrate the applicability of the earlier results to a problem in H^∞-filtering. For this purpose, we consider a state-space model of the form

$$x_{i+1} = F_i x_i + G_i u_i \ , \quad y_i = H_i x_i + v_i \ , \tag{91}$$

where $\{x_0, u_i, v_i\}$ are unknown deterministic signals and $\{y_i\}_{i=0}^N$ are known (or measured) signals. Let $s_j = L_j x_j$ be a linear transformation of the state-vector x_j, where L_j is a known matrix.

Let $\hat{s}_{j|j}$ denote a function of the $\{y_k\}$ up to and including time j. For every time instant i we define the quadratic cost function

$$J_i(x_0, u_0, \ldots, u_i) \triangleq x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} u_j^* Q_j^{-1} u_j + \sum_{j=0}^{i} v_j^* v_j - \gamma^{-2} \sum_{j=0}^{i} (\hat{s}_{j|j} - L_j x_j)^* (\hat{s}_{j|j} - L_j x_j), \tag{92}$$

where $\{\Pi_0, Q_j\}$ are given positive-definite matrices, and γ is a given positive real number.

Problem 8.1 (An $\mathbf{H}^\infty$−Filtering Problem) *Determine, if possible, functions*

$$\{\hat{s}_{0|0}, \hat{s}_{1|1}, \ldots, \hat{s}_{N|N}\},$$

in order to guarantee that

$$J_i > 0 \quad for \quad i = 0, 1, \ldots, N. \tag{93}$$

The positivity requirement (93) can be interpreted as imposing an upper bound on the following ratios (for nonzero denominators)

$$\frac{\sum_{j=0}^{i} (\hat{s}_{j|j} - L_j x_j)^* (\hat{s}_{j|j} - L_j x_j)}{x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} u_j^* Q_j^{-1} u_j + \sum_{j=0}^{i} v_j^* v_j} \quad < \quad \gamma^2, \ \text{ for } \ 0 \leq i \leq N.$$

Using $v_j = y_j - H_j x_j$, we can rewrite the expression for J_i in the equivalent form

$$J_i \quad = \quad x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} \left(\begin{bmatrix} \hat{s}_{j|j} \\ y_j \end{bmatrix} - \begin{bmatrix} L_j \\ H_j \end{bmatrix} x_j \right)^* \begin{bmatrix} -\gamma^{-2} I & 0 \\ 0 & I \end{bmatrix} \left(\begin{bmatrix} \hat{s}_{j|j} \\ y_j \end{bmatrix} - \begin{bmatrix} L_j \\ H_j \end{bmatrix} x_j \right) + \sum_{j=0}^{i} u_j^* Q_j^{-1} u_j,$$

which is a quadratic cost function in the unknowns $\{x_0, u_0, \ldots, u_i\}$ since the $\{y_j, \hat{s}_{j|j}\}_{j=0}^i$ can be expressed in terms of $\{x_0, u_0, \ldots, u_i\}$. Therefore, each J_i will be positive if, and only if, it has a minimum with respect to $\{x_0, u_0, \ldots, u_i\}$ and, moreover, the value of J_i at its minimum is positive.

8.1 Solvability Conditions

We thus see that we are faced with the problem of minimizing a quadratic cost function of the same general form as in (85), and also (84), where the column vector

$$\begin{bmatrix} \hat{s}_{j|j} \\ y_j \end{bmatrix},$$

and the block matrices

$$\begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \text{ and } \begin{bmatrix} L_j \\ H_j \end{bmatrix},$$

now play the roles of $\{y_j, R_j, H_j\}$ in (85). That is, the auxiliary state-space model that we may invoke here, with variables in an indefinite space $\mathcal{K}'$, takes the form

$$\begin{aligned} \mathbf{x}_{i+1} &= F_i \mathbf{x}_i + G_i \mathbf{u}_i, \\ \begin{bmatrix} \hat{\mathbf{s}}_{j|j} \\ \mathbf{y}_j \end{bmatrix} &= \begin{bmatrix} L_j \\ H_j \end{bmatrix} \mathbf{x}_i + \bar{\mathbf{v}}_i, \end{aligned}$$

with

$$< \begin{bmatrix} \mathbf{u}_i \\ \bar{\mathbf{v}}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \bar{\mathbf{v}}_j \\ \mathbf{x}_0 \end{bmatrix} >_{\mathcal{K}'} = \begin{bmatrix} Q_i \delta_{ij} & 0 & 0 \\ 0 & (-\gamma^2 I \oplus I)\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}.$$

We then conclude from Corollary 7.2, and according to the remark after Theorem 7.2, that each J_i will admit a minimizing solution if, and only if, the corresponding $R_{e,i}$ and R_i have the same inertia. In the present context, we have

$$R_i \triangleq \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad R_{e,i} \triangleq \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} L_i \\ H_i \end{bmatrix} P_i \begin{bmatrix} L_i \\ H_i \end{bmatrix}^*,$$

where P_i satisfies the Riccati difference equation

$$\begin{aligned} P_{i+1} &= F_i \left[P_i - P_i \begin{bmatrix} L_i \\ H_i \end{bmatrix}^* \left\{ \begin{bmatrix} L_i \\ H_i \end{bmatrix} P_i \begin{bmatrix} L_i \\ H_i \end{bmatrix}^* + \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \right\}^{-1} \begin{bmatrix} L_i \\ H_i \end{bmatrix} P_i \right] F_i^* + G_i Q_i G_i^*, \\ &= F_i \left[P_i^{-1} + \begin{bmatrix} L_i^* & H_i^* \end{bmatrix} \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} L_i \\ H_i \end{bmatrix} \right]^{-1} F_i^* + G_i Q_i G_i^*, \\ &= F_i \left[P_i^{-1} + H_i^* H_i - \gamma^{-2} L_i^* L_i \right]^{-1} F_i^* + G_i Q_i G_i^*. \end{aligned}$$

Lemma 8.1 (Inertia Conditions) *The J_i in (92) admit unique minima with respect to $\{x_0, u_0, \ldots, u_i\}$ if, and only if, the matrices*

$$\begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -\gamma^2 I + L_i P_i L_i^* & L_i P_i H_i^* \\ H_i P_i L_i^* & I + H_i P_i H_i^* \end{bmatrix}, \tag{94}$$

have the same inertia for all i. In this case, it also follows that all the leading submatrices of the above two matrices have the same inertia, i.e.,

$$\begin{aligned} I + H_i P_i H_i^* &> 0, \\ (-\gamma^2 I + L_i P_i L_i^*) - L_i P_i H_i^* (I + H_i P_i H_i^*)^{-1} H_i P_i L_i^* &< 0. \end{aligned}$$

Proof: The first part of the Lemma follows from Corollary 7.2. But recall also from the statement of the Corollary that the resulting P_i are further guaranteed to be nonnegative-definite, i.e., $P_i \geq 0$. It thus follows that $(I + H_i P_i H_i^*) > 0$. That is, the lower-right corner elements of both matrices in (94) have the same positive inertia. Consequently, it also holds that all the leading submatrices of the two matrices in (94) have the same inertia.

■

If the F_i are further assumed invertible, then we also conclude from Corollary 7.3 that the following alternative conditions can be used to guarantee the existence of minima for the J_i in (92),

$$P_i^{-1} + H_i^* H_i - \gamma^{-2} L_i^* L_i > 0, \quad \text{for} \quad 0 \leq i \leq N. \tag{95}$$

8.2 Construction of a Solution

To end our discussion, we still need to show how to determine the estimates $\hat{s}_{j|j}$ once the existence of minima for the J_i are guaranteed. These estimates have to be chosen so as to guarantee that the values of the successive J_i at their minima are positive.

We shall illustrate the construction by induction. Assume that the $\{\hat{s}_{0|0}, \ldots, \hat{s}_{i-1|i-1}\}$ have already been chosen and that the values of the $\{J_0, J_1, \ldots, J_{i-1}\}$ are positive at their respective minima (recall expression (82)). In particular,

$$\sum_{j=0}^{i-1} e_j^* R_{e,j}^{-1} e_j \; > 0.$$

In order to guarantee $J_i > 0$ we need to choose $\hat{s}_{i|i}$ so as to result in

$$e_i^* R_{e,i}^{-1} e_i + \sum_{j=0}^{i-1} e_j^* R_{e,j}^{-1} e_j \; > \; 0.$$

This can be achieved in many ways and the choice is nonunique. One possibility is to choose $\hat{s}_{i|i}$ so as to meet the condition

$$e_i^* R_{e,i}^{-1} e_i > 0, \tag{96}$$

or, equivalently,

$$\begin{bmatrix} e_{i,s}^* & e_{i,y}^* \end{bmatrix} \begin{bmatrix} -\gamma^2 I + L_i P_i L_i^* & L_i P_i H_i^* \\ H_i P_i L_i^* & I + H_i P_i H_i^* \end{bmatrix}^{-1} \begin{bmatrix} e_{i,s} \\ e_{i,y} \end{bmatrix} > 0, \tag{97}$$

where we have partitioned the e_i accordingly, viz.,

$$e_i \triangleq \begin{bmatrix} \hat{s}_{i|i} \\ y_i \end{bmatrix} - \begin{bmatrix} L_i \\ H_i \end{bmatrix} \hat{x}_{i|i-1} \triangleq \begin{bmatrix} e_{i,s} \\ e_{i,y} \end{bmatrix}.$$

Here $\hat{x}_{i|i-1}$ is constructed recursively as indicated in Theorem 7.2,

$$\hat{x}_{i+1|i} = F_i \hat{x}_{i|i-1} + K_{p,i} \left(\begin{bmatrix} \hat{s}_{i|i} \\ y_i \end{bmatrix} - \begin{bmatrix} L_i \\ H_i \end{bmatrix} \hat{x}_{i|i-1} \right), \quad \hat{x}_{0|-1} = 0, \tag{98}$$

with

$$R_{e,i} = \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} L_i \\ H_i \end{bmatrix} P_i \begin{bmatrix} L_i \\ H_i \end{bmatrix}^*, \quad K_{p,i} = F_i P_i \begin{bmatrix} L_i^* & H_i^* \end{bmatrix} R_{e,i}^{-1}.$$

We may now introduce the lower-diagonal-upper factorization of the central matrix in (97), viz.,

$$\begin{bmatrix} -\gamma^2 I + L_i P_i L_i^* & L_i P_i H_i^* \\ H_i P_i L_i^* & I + H_i P_i H_i^* \end{bmatrix}^{-1} = \tag{99}$$

$$\begin{bmatrix} I & 0 \\ -(I + H_i P_i H_i^*)^{-1} H_i P_i L_i^* & I \end{bmatrix} \begin{bmatrix} \Delta^{-1} & 0 \\ 0 & (I + H_i P_i H_i^*)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -(I + H_i P_i H_i^*)^{-1} H_i P_i L_i^* & I \end{bmatrix}^*,$$

where we have defined, for compactness of notation,

$$\Delta \triangleq (-\gamma^2 I + L_i P_i L_i^*) - L_i P_i H_i^* (I + H_i P_i H_i^*)^{-1} H_i P_i L_i^*,$$

which we know, from Lemma 8.1, to be a negative definite matrix.

We can then rewrite (97) in the form

$$\begin{bmatrix} e_{i,s}^* - e_{i,y}^* (I + H_i P_i H_i^*)^{-1} H_i P_i L_i^* & e_{i,y}^* \end{bmatrix} \begin{bmatrix} \Delta^{-1} & 0 \\ 0 & (I + H_i P_i H_i^*)^{-1} \end{bmatrix} \begin{bmatrix} e_{i,s} - L_i P_i H_i^* (I + H_i P_i H_i^*)^{-1} e_{i,y} \\ e_{i,y} \end{bmatrix}.$$

This is a quadratic expression in the variable $e_{i,s} = \hat{s}_{i|i} - L_i\hat{x}_{i|i-1}$, and since $\Delta < 0$ and $(I + H_iP_iH_i^*) > 0$, the positivity condition (96) can be met by setting

$$e_{i,s} - L_iP_iH_i^*(I + H_iP_iH_i^*)^{-1}e_{i,y} = 0,$$

or, equivalently,

$$\hat{s}_{i|i} - L_i\hat{x}_{i|i-1} = L_iP_iH_i^*(I + H_iP_iH_i^*)^{-1}[y_i - H_i\hat{x}_{i|i-1}].$$

Therefore, a possible choice for $\hat{s}_{i|i}$ is the following

$$\hat{s}_{i|i} = L_i\left[\hat{x}_{i|i-1} + P_iH_i^*(I + H_iP_iH_i^*)^{-1}(y_i - H_i\hat{x}_{i|i-1})\right].$$

This choice simplifies (98) to the following (using the factorization (99) for $R_{e,i}^{-1}$ in the expression for $K_{p,i}$)

$$\hat{x}_{i+1|i} = F_i\left[\hat{x}_{i|i-1} + P_iH_i^*(I + H_iP_iH_i^*)^{-1}(y_i - H_i\hat{x}_{i|i-1})\right].$$

We summarize the results in the following statement.

Lemma 8.2 (A Solution of the H$^\infty$−Problem) *Problem 8.1 has a solution if, and only if, for all* $0 \leq i \leq N$, *the matrices*

$$\begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \quad and \quad \begin{bmatrix} -\gamma^2 I + L_iP_iL_i^* & L_iP_iH_i^* \\ H_iP_iL_i^* & I + H_iP_iH_i^* \end{bmatrix}, \tag{100}$$

have the same inertia. In this case, one possible construction for the estimates $\{\hat{s}_{i|i}\}$ *is the following:*

$$\hat{s}_{i|i} \;=\; L_i\left[\hat{x}_{i|i-1} + P_iH_i^*(I + H_iP_iH_i^*)^{-1}(y_i - H_i\hat{x}_{i|i-1})\right], \tag{101}$$

where the $\hat{x}_{i|i-1}$ *is constructed recursively via*

$$\hat{x}_{i+1|i} = F_i\left[\hat{x}_{i|i-1} + P_iH_i^*(I + H_iP_iH_i^*)^{-1}(y_i - H_i\hat{x}_{i|i-1})\right], \;\; \hat{x}_{0|-1} = 0, \tag{102}$$

and

$$P_{i+1} = F_i\left[P_i - P_i\begin{bmatrix} L_i \\ H_i \end{bmatrix}^*\left\{\begin{bmatrix} L_i \\ H_i \end{bmatrix}P_i\begin{bmatrix} L_i \\ H_i \end{bmatrix}^* + \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix}\right\}^{-1}\begin{bmatrix} L_i \\ H_i \end{bmatrix}P_i\right]F_i^* + G_iQ_iG_i^*, \tag{103}$$

with the initial condition $P_0 = \Pi_0$.

9 An Application to Robust Adaptive Filters

We now consider another example that can, in effect, be regarded as a special case of the H^∞−problem studied in Sec. 8. Here, however, some simplifications occur that are worth considering separately.

We therefore assume that we have the following special state-space model

$$x_{i+1} = x_i\;, \;\; y_i = H_ix_i + v_i\;, \tag{104}$$

where $\{x_0, v_i\}$ are unknown deterministic signals and $\{y_i\}_{i=0}^N$ are known (or measured) signals. Compared with the model (91) we see that we are now assuming $u_i = 0$ and $F_i = I$. In fact, the arguments that follow can also be applied to any invertible matrix F_i (especially the arguments after Lemma 9.1).

The equations (104) show that the vector x_i does not change with time and is therefore equal to the initial unknown vector x_0. That is, we can as well regard the equations (104) as representing a collection of measured vectors $\{y_i\}$ that are linearly related to an unknown vector x_0,

$$y_i = H_ix_0 + v_i,$$

and the objective is to estimate the x_0 in a certain sense. A classical criterion is to solve a positive-definite least-squares problem of the form (see, e.g., [SK94])

$$\min_{x_0} \left[x_0^* \Pi_0^{-1} x_0 \; + \; \sum_{i=0}^{N} (y_i - H_i x_0)^* W_i^{-1} (y_i - H_i x_0) \right], \tag{105}$$

where $\{\Pi_0, W_i\}$ are given positive-definite weighting matrices. In this case, a minimizing solution is always guaranteed to exist and, under some extra conditions on the matrices $\{\Pi_0, H_i\}$, a recursive scheme is in fact possible, thus leading to the famed Recursive-Least-Squares (RLS) algorithm.

Here, however, we allow for indefinite weighting matrices $\{\Pi_0, W_i\}$, along the same lines studied in Sec. 8. More specifically, we let $\hat{x}_{j|j}$ denote a function of the $\{y_k\}$ up to and including time j. Since $x_j = x_0$, we shall also write $\hat{x}_{0|j}$ instead of $\hat{x}_{j|j}$.

For every time instant i we also define the quadratic cost function

$$J_i(x_0) \triangleq x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} v_j^* v_j - \gamma^{-2} \sum_{j=0}^{i} (\hat{x}_{0|j} - x_0)^* (\hat{x}_{0|j} - x_0), \tag{106}$$

where $\{\Pi_0\}$ is a given positive-definite matrix, and γ is a given positive number.

Problem 9.1 (A Robust Adaptive Filter) *Determine, if possible, functions*

$$\{\hat{x}_{0|0}, \hat{x}_{0|1}, \ldots, \hat{x}_{0|N}\},$$

in order to guarantee that

$$J_i > 0 \quad \textit{for} \quad i = 0, 1, \ldots, N. \tag{107}$$

The positivity requirement (107) can be interpreted as imposing an upper bound on the following ratios (for nonzero denominators)

$$\frac{\sum_{j=0}^{i} (\hat{x}_{0|j} - x_0)^* (\hat{x}_{0|j} - x_0)}{x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} v_j^* v_j} \; < \; \gamma^2, \; \text{ for } \; 0 \le i \le N.$$

Using $v_j = y_j - H_j x_0$, we can also write the above ratios in the form

$$\frac{\sum_{j=0}^{i} \|\hat{x}_{0|j} - x_0\|^2}{x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} \|y_j - H_j x_0\|^2} \; < \; \gamma^2, \; \text{ for } \; 0 \le i \le N. \tag{108}$$

Comparing with (105), we see that the cost function of (105) now appears in the denominator of (108) (with $W_i = I$). Hence, instead of minimizing (105) over x_0, we are now interested in determining estimates for x_0 in order to guarantee that the energy in the error due to estimating x_0 is upper-bounded by γ^2 times the energy of the uncertainties, viz., the denominator in (108).

We can again rewrite the expression for J_i in the equivalent form

$$J_i \;\; = \;\; x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} \left(\begin{bmatrix} \hat{x}_{0|j} \\ y_j \end{bmatrix} - \begin{bmatrix} I \\ H_j \end{bmatrix} x_0 \right)^* \begin{bmatrix} -\gamma^{-2} I & 0 \\ 0 & I \end{bmatrix} \left(\begin{bmatrix} \hat{x}_{0|j} \\ y_j \end{bmatrix} - \begin{bmatrix} I \\ H_j \end{bmatrix} x_0 \right),$$

which is a quadratic cost function in the unknown $\{x_0\}$. We can now use Lemma 8.2 to conclude the following (by setting $L_i = I, F_i = I, G_i = 0, Q_i = 0, x_i = x_0$).

Lemma 9.1 (Solution of the Adaptive Problem) *Problem 9.1 has a solution if, and only if, for all* $0 \leq i \leq N$, *the matrices*

$$\begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \quad and \quad \begin{bmatrix} -\gamma^2 I + P_i & P_i H_i^* \\ H_i P_i & I + H_i P_i H_i^* \end{bmatrix}, \tag{109}$$

have the same inertia. In this case, one possible construction for the estimates $\{\hat{x}_{0|i}\}$ *is the following:*

$$\hat{x}_{0|i} \;=\; \hat{x}_{0|i-1} + P_i H_i^* (I + H_i P_i H_i^*)^{-1} (y_i - H_i \hat{x}_{0|i-1}), \;\; \hat{x}_{0|-1} = 0, \tag{110}$$

where

$$P_{i+1} = \left[P_i - P_i \begin{bmatrix} I \\ H_i \end{bmatrix}^* \left\{ \begin{bmatrix} I \\ H_i \end{bmatrix} P_i \begin{bmatrix} I \\ H_i \end{bmatrix}^* + \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \right\}^{-1} \begin{bmatrix} I \\ H_i \end{bmatrix} P_i \right], \tag{111}$$

with the initial condition $P_0 = \Pi_0$.

We now argue that the solvability condition can in fact be simplified in the adaptive case. For this purpose, we shall invoke the conclusions of Corollary 7.3. Indeed, it follows from the statement of the corollary that Problem 9.1 has a solution if, and only if, for all $0 \leq i \leq N$,

$$P_i^{-1} + \begin{bmatrix} I & H_i^* \end{bmatrix} \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ H_i \end{bmatrix} > 0,$$

or, equivalently,

$$P_i^{-1} + H_i^* H_i - \gamma^2 I > 0. \tag{112}$$

A simpler statement is the following.

Lemma 9.2 (A Solvability Condition for the Adaptive Problem) *Problem 9.1 has a solution if, and only if,*

$$P_{i+1} > 0 \;\; for \;\; 0 \leq i \leq N. \tag{113}$$

Proof: This follows from second condition of Corollary 7.3, using $G_i = 0$.

∎

The condition (113) is indeed natural in the adaptive context. To clarify this, we note that it follows from the Riccati recursion (111) that

$$P_{i+1}^{-1} = P_i^{-1} + \begin{bmatrix} I & H_i^* \end{bmatrix} \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ H_i \end{bmatrix}, \tag{114}$$

with initial condition $P_0^{-1} = \Pi_0^{-1}$. This implies, by recurrence, that

$$P_{i+1}^{-1} = \Pi_0^{-1} + \sum_{j=0}^{i} \begin{bmatrix} I & H_j^* \end{bmatrix} \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ H_j \end{bmatrix}, \tag{115}$$

which, in view of expression (20) in Theorem 3.1, is precisely the coefficient matrix of the linear system of equations that provides us with $\hat{x}_{0|j}$. The conclusion (113) is then immediate once we also recall from the statement of Theorem 3.1 that a minimum is guaranteed as long as the coefficient matrix is positive-definite.

10 An Application to Total Least-Squares Methods

We now consider a third application that deals with the so-called total-least-squares (or errors-in-variables) method for the solution of linear systems of equations, $Ax \approx b$ (e.g., [LS83, HV91]). The notation $Ax \approx b$ means that due to possible errors (measurement errors, modelling errors, etc) the vector b does not necessarily lie in the range space of the matrix A, denoted by $\mathcal{R}(A)$. If indeed we had $b \in \mathcal{R}(A)$, then a solution x would exist to the equations $Ax = b$. In general, however, one has to settle for an approximate solution $\hat{x}$. In least-squares methods, it is often assumed that the vector b is possibly erroneous, while the matrix A is known and one proceeds to solve for the vector $\hat{x}$ that minimizes the Euclidean distance between $A\hat{x}$ and b, say

$$\min_x \|Ax - b\|^2. \tag{116}$$

This is clearly a special case of the quadratic cost function (1) with $\Pi \to \infty I$, $W = I$, and the notational changes $y \leftarrow b$, $z \leftarrow x$. All solutions $\hat{x}$ are well-known to satisfy the so-called normal system of equations

$$(A^*A)\hat{x} = A^*b. \tag{117}$$

Total least-squares (TLS, for short) methods, on the other hand, allow us to also handle possible errors in the matrix A itself. For this reason, they have been receiving increasing attention, especially in the signal processing community. The TLS problem seeks a matrix $\hat{M}$ and a vector $\hat{x}$ that minimize the following Frobenius norm:

$$\min_{M,x} \left\| \begin{bmatrix} M - A & Mx - b \end{bmatrix} \right\|_F^2. \tag{118}$$

Here, M is regarded as an approximation for A, which in its turn is used to determine an $\hat{x}$ that guarantees $b \in \mathcal{R}(\hat{M})$.

The solution of the above TLS problem is well-known and is given by the following construction [HV91][p. 36]. Assume A is $(N+1) \times n$ with $N \geq n$, as is often the case. Let $\{\sigma_1, \ldots, \sigma_n\}$ denote the singular values of A, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. Let also $\{\bar{\sigma}_1, \ldots, \bar{\sigma}_n, \bar{\sigma}_{n+1}\}$ denote the singular values of the extended matrix $\begin{bmatrix} A & b \end{bmatrix}$, with $\bar{\sigma}_i \geq 0$. If $\bar{\sigma}_{n+1} < \sigma_n$, then the unique solution $\hat{x}$ of (118) is given by

$$\hat{x} = (A^*A - \bar{\sigma}_{n+1}^2 I)^{-1} A^* b. \tag{119}$$

Moreover, the matrix $\hat{M}$ is constructed from the SVD of $\begin{bmatrix} A & b \end{bmatrix}$. In fact, a similar construction for $\hat{x}$ also exists in terms of the data available from the SVD. But here we shall instead focus on the representation (119) of the solution $\hat{x}$. Note also that the condition $\bar{\sigma}_{n+1} < \sigma_n$ assures that $(A^*A - \bar{\sigma}_{n+1}^2 I)$ is a positive-definite matrix, since σ_n^2 is the smallest eigenvalue of A^*A.

Comparing (119) with the solution of the indefinite quadratic problem (1), as given in Theorem 3.1, expression (20), we see that we can make the identifications

$$\Pi \leftarrow -\bar{\sigma}_{n+1}^{-2} I \quad \text{and} \quad W \leftarrow I,$$

along with $y \leftarrow b$ and $z \leftarrow x$. That is, we can regard (119) as the solution of the following indefinite problem

$$\min_x \left[-\bar{\sigma}_{n+1}^2 x^* x + (b - Ax)^*(b - Ax) \right], \tag{120}$$

which is clearly a special case of (1) in two respects: the Π matrix is negative-definite and a multiple of the identity, and the W matrix is simply the identity. Indeed, the minimum of (120) exists as long as $(-\bar{\sigma}_{n+1}^2 I + A^*A)$ is positive-definite, which is guaranteed by the assumption $\bar{\sigma}_{n+1} < \sigma_n$.

Note though that the solution $\hat{x}$ of the TLS problem (119) requires a singular value decomposition (SVD), which may be computationally expensive. But more important perhaps, is that this hinders the possibility of recursive updates of the solution $\hat{x}$. More specifically, if an extra row is added to the matrix A and, correspondingly, if an extra entry is added to the vector b, then the SVD of the new extended matrix $\begin{bmatrix} A & b \end{bmatrix}$ will need to be computed again in order to evaluate the new solution $\hat{x}$.

An examination of expression (119), however, shows that the SVD step only affects the choice of the Π matrix. This suggests that a recursive scheme should be possible if one relaxes the criterion (118) and allows for other choices of the Π matrix in (120), say

$$\Pi^{-1} = -\rho^2 I,$$

for a nonnegative real number ρ^2 that is chosen by the user. In particular, any choice that satisfies $\rho^2 < \sigma_n$ will still result in a positive-definite matrix $[-\rho^2 I + A^*A]$. We may also employ a diagonal matrix of the form

$$\Pi^{-1} = -\text{diagonal}\,\{\rho_0^2, \rho_1^2, \ldots, \rho_{n-1}^2\},$$

with several nonnegative entries $\{\rho_i^2\}$. This would allow us to give different weights to the different entries of x and will also give us more freedom in controlling the existence of solutions to the recursive procedure described below.

We may also remark that the idea of replacing an optimal problem by a suboptimal one is frequent in many areas, including for example $H^\infty-$problems, and this is often due to the computational burden that may be required by an optimal formulation.

Problem 10.1 (Approximate TLS Problem) *Consider a matrix A, with rows $\{a_i\}_{i=0}^N$, a vector b with entries $\{b(i)\}_{i=0}^N$, and a diagonal matrix $\Pi^{-1} = -diag\{\rho_i^2\}$. Define, for each i, the quadratic cost function*

$$J_i \triangleq \left[x^*\Pi^{-1}x + \sum_{j=0}^{i} |b(j) - a_j x|^2 \right].$$

Let $\hat{x}_i$ denote a stationary solution of J_i. We are interested in the following:

(i) A recursive update that relates $\hat{x}_i$ to $\hat{x}_{i+1}$. For this purpose, we shall assume that (recall Lemma 7.2) $[I + A\Pi A^*]$ is strongly regular. This suggests a criterion for choosing the Π matrix.

(ii) A condition that guarantees that the last estimate $\hat{x}_N$ is indeed a minimum of J_N.

The answers to the above questions are rather immediate if we invoke the results of Sec. 7 and, in particular, Lemma 7.2 and its corollary, and Theorems 7.1 and 7.2.

Lemma 10.1 (Solution of the Approximate TLS Problem) *A recursive construction of the solution can be obtained as follows, assuming $[I + A\Pi A^*]$ is strongly regular:*

(i) The successive stationary solutions are related via

$$\hat{x}_i = \hat{x}_{i-1} + \frac{P_i a_i^*}{1 + a_i P_i a_i^*}\left[b(i) - a_i \hat{x}_{i-1}\right], \quad \hat{x}_{-1} = 0, \tag{121}$$

$$P_{i+1} = P_i - \frac{P_i a_i^* a_i P_i}{1 + a_i P_i a_i^*}, \quad P_0 = \Pi = -diag\{\rho_i^2\}. \tag{122}$$

(ii) J_N has a minimum at $\hat{x}_N$ if, and only if, the matrix $[\Pi^{-1} + A^*A]$ is positive-definite. Under the assumption of strong regularity of $[I + A\Pi A^*]$, this positivity condition is also equivalent to $P_{N+1} > 0$ since, as argued after the proof of Lemma 9.1, we can also verify here that P_{N+1} is the inverse of $[\Pi^{-1} + A^*A]$. Indeed, from (122) we obtain

$$P_{i+1}^{-1} = P_i^{-1} + a_i^* a_i, \quad P_0 = \Pi^{-1}. \tag{123}$$

We emphasize, however, that the above is only a special case of the quadratic forms studied in this paper. For example, one may choose other forms for the diagonal matrices Π and W, such as allowing for positive entries in Π and for negative entries in W, or other convenient combinations.

11 Concluding Remarks

We have posed two minimization problems in indefinite metric spaces and established a link between their solutions via a fundamental set of inertia conditions. These conditions were derived under very general assumptions and later specialized to important special cases that arise in $H^\infty-$filtering, robust adaptive filtering, and approximate TLS methods. In the $H^\infty-$context, for instance, the inertia results of Corollary 7.2 can be used as the basis for alternative computational variants that are based on square-root ideas. This point of view is detailed in [HSK94]. More generally, the inertia conditions of Theorem 7.1 can also form the basis for general square-root algorithms and this will be discussed elsewhere.

Further connections with system theory and recent applications to problems in linear and nonlinear adaptive filtering can be found in [SR95a, SR95b, SR95c, RS95].

References

[AM79] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall Inc., NJ, 1979.

[Bog74] J. Bognar. *Indefinite Inner Product Spaces.* Springer-Verlag, New York, 1974.

[DGKF89] J. C. Doyle, K. Glover, P. Khargonekar, and B. Francis. State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control*, 34(8):831–847, August 1989.

[Gan59] F. R. Gantmacher. *The Theory of Matrices.* Chelsea Publishing Company, NY, 1959.

[GL95] M. Green and D. J. N. Limebeer. *Linear Robust Control.* Prentice Hall, NJ, 1995.

[GLR83] I. Gohberg, P. Lancaster, and L. Rodman. *Matrices and Indefinite Scalar Products.* Birkhäuser Verlag, Basel, 1983.

[Gri93] M. J. Grimble. Polynomial matrix solution of the H^∞ filtering problem and the relationship to Riccati equation state-space results. *IEEE Trans. on Signal Processing*, 41(1):67–81, January 1993.

[Hay91] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ, second edition, 1991.

[HSK93] B. Hassibi, A. H. Sayed, and T. Kailath. Recursive linear estimation in Krein spaces - Part I: Theory. In *Proc. Conference on Decision and Control*, vol. 4, pages 3489–3494, San Antonio, Texas, December 1993. To appear also in *IEEE Transactions on Automatic Control.*

[HSK94] B. Hassibi, A. H. Sayed, and T. Kailath. Square-root arrays and Chandrasekhar recursions for H^∞ problems. In *Proc. Conference on Decision and Control*, Orlando, FL, December 1994.

[HV91] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 1991.

[Kai81] T. Kailath. *Lectures on Wiener and Kalman Filtering.* Springer-Verlag, NY, second edition, 1981.

[KN91] P.P. Khargonekar and K. M. Nagpal. Filtering and smoothing in an $H^\infty-$ setting. *IEEE Trans. on Automatic Control*, AC-36:151–166, 1991.

[LS83] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA, 1983.

[LS91] D. J. Limebeer and U. Shaked. New results in H^∞-filtering. In *Proc. Int. Symp. on MTNS*, pages 317–322, June 1991.

[PRLN92] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias. *Advanced Digital Signal Processing.* Macmillan Publishing Co., New York, NY, 1992.

[SK94] A. H. Sayed and T. Kailath. A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine*, 11(3):18–60, July 1994.

[YS91] I. Yaesh and U. Shaked. $H^\infty-$optimal estimation: The discrete time case. In *Proc. Inter. Symp. on MTNS*, pages 261–267, Kobe, Japan, June 1991.

[SR95a] A. H. Sayed and M. Rupp. A time-domain feedback analysis of adaptive gradient algorithms via the Small Gain Theorem. *Proc. SPIE Conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations* , F.T. Luk, ed., vol. 2563, pp. 458–469, San Diego, CA, July 1995.

[SR95b] A. H. Sayed and M. Rupp. A class of adaptive nonlinear H^∞-filters with guaranteed l_2-stability. *Proc. IFAC Symposium on Nonlinear Control System Design*, vol. 1, pp. 455–460, Tahoe City, CA, June 1995.

[SR95c] A. H. Sayed and M. Rupp. A feedback analysis of Perceptron learning for neural networks. To appear in *Proc. 29th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 1995.

[RS95] M. Rupp and A. H. Sayed. A robustness analysis of Gauss-Newton recursive methods. To appear in *Proc. Conference on Decision and Control*, New Orleans, LA, Dec. 1995.

Ali H. Sayed
Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106
USA

Thomas Kailath
Information Systems Laboratory
Stanford University
Stanford, CA 94305
USA

Babak Hassibi
Information Systems Laboratory
Stanford University
Stanford, CA 94305
USA

AMS Classification. 93E24, 93E11, 93E10, 93B36, 93B40, 60G35.