Linear Estimation in Krein Spaces—Part I: Theory

Babak Hassibi, Ali H. Sayed, *Member, IEEE*, and Thomas Kailath, *Fellow, IEEE*

***Abstract*— The authors develop a self-contained theory for linear estimation in Krein spaces. The derivation is based on simple concepts such as projections and matrix factorizations and leads to an interesting connection between Krein space projection and the recursive computation of the stationary points of certain second-order (or quadratic) forms. The authors use the innovations process to obtain a general recursive linear estimation algorithm. When specialized to a state-space structure, the algorithm yields a Krein space generalization of the celebrated Kalman filter with applications in several areas such as H^∞-filtering and control, game problems, risk sensitive control, and adaptive filtering.**

I. Introduction

In some recent explorations, we have found that H^∞ estimation and control problems and several related problems (risk-sensitive estimation and control, finite memory adaptive filtering, stochastic interpretation of the KYP lemma, and others) can be studied in a simple and unified way by relating them to Kalman filtering problems, not in the usual (stochastic) Hilbert space, but in a special kind of indefinite metric space known as a Krein space (see, e.g., [9], [10]). Although the two types of spaces share many characteristics, they differ in special ways that turn out to mark the differences between the linear-quadratic-Gaussian (LQG) or H^2 theories and the more recent H^∞ theories. The connections with the conventional Kalman filter theory will allow several of the newer numerical algorithms, developed over the last three decades, to be applied to the H^∞ theories [22].

In this paper the authors develop a self-contained theory for linear estimation in Krein spaces. The ensuing theory is richer than that of the conventional Hilbert space case which is why it yields a unified approach to the above mentioned problems. Applications will follow in later papers.

The remainder of the paper is organized as follows. We introduce Krein spaces in Section II and define projections in Krein spaces in Section III. Contrary to the Hilbert space case where projections always exist and are unique, the Krein-space projection exists and is unique if, and only if, a certain Gramian matrix is nonsingular. In Section IV, we first remark that while quadratic forms in Hilbert space always have minima (or maxima), in Krein spaces one can assert only that they will always have stationary points. Further conditions will have to be met for these to be minima or maxima. We explore this by first considering the problem of finding a vector k to stationarize the quadratic form $\langle \boldsymbol{z} - k^*\boldsymbol{y}, \boldsymbol{z} - k^*\boldsymbol{y}\rangle$, where $\langle \cdot, \cdot \rangle$ is an indefinite inner product, $*$ denotes conjugate transpose, $\boldsymbol{y}$ is a collection of vectors in a Krein space (which we can regard as generalized random variables), and $\boldsymbol{z}$ is a vector outside the linear space spanned by the $\boldsymbol{y}$. If the Gramian matrix $R_y = \langle \boldsymbol{y}, \boldsymbol{y}\rangle$ is nonsingular, then there is a unique stationary point $k_0^*\boldsymbol{y}$, given by the projection of $\boldsymbol{z}$ onto the linear space spanned by the $\boldsymbol{y}$; the stationary point will be a minimum if, and only if, R_y is strictly positive definite as well. In a Hilbert space, the nonsingularity of R_y and its strict positive definiteness are equivalent properties, but this is not true with $\boldsymbol{y}$ in a Krein space.

Now in the Hilbert space theory it is well known (motivated by a Bayesian approach to the problem) that a certain deterministic quadratic form $J(z, y)$, where now z and y are elements of the usual Euclidean vector space, is also minimized by $k_0^* y$ with exactly the same k as before. In the Krein-space case, $k_0^* y$ also yields a stationary point of the corresponding deterministic quadratic form, but now this point will be a minimum if, and only if, a different condition, not $R_y > 0$, but $R_z - R_{zy}R_y^{-1}R_{yz} > 0$, is satisfied. In Hilbert space, unlike Krein space, the two conditions for a minimum hold simultaneously (see Corollary 3 in Section IV). This simple distinction turns out to be crucial in understanding the difference between H^2 and H^∞ estimation, as we shall show in detail in Part II of this series of papers.

In this first part, we continue with the general theory by exploring the consequences of assuming that $\{\boldsymbol{z}, \boldsymbol{y}\}$ are based on some underlying state-space model. The major ones are a reduction in computational effort, $O(Nn^3)$ versus $O(N^3)$, where N is the number of observations and n is the number of states and the possibility of recursive solutions. In fact, it will be seen that the innovations-based derivation of the Hilbert space–Kalman filter extends to Krein spaces, except that now the Riccati variable P_i, and the innovations Gramian $R_{e,i}$ are not necessarily positive (semi)definite. The Krein space–Kalman filter continues to have the interpretation of performing the triangular factorization of the Gramian matrix of the observations, R_y; this reduces the test for $R_y > 0$ to recursively checking that the $R_{e,i} > 0$.

Similar results are expected for the corresponding indefinite quadratic form. While global expressions for the stationary point of such quadratic forms and of the minimization

Manuscript received March 4, 1994; revised June 16, 1995. Recommended by Associate Editor at Large, B. Pasik-Duncan. This work was supported in part by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Office of Scientific Research under Contract F49620-93-1-0085 and in part by a grant from NSF under award MIP-9409319.

B. Hassibi and T. Kailath are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA.

A. H. Sayed is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA.

Publisher Item Identifier S 0018-9286(96)00386-8.

condition were readily obtained, as previously mentioned, recursive versions are not easy to obtain. Dynamic programming arguments are the ones usually invoked, and they turn out to be algebraically more complex than the simple innovations (Gram–Schmidt orthogonalization) ideas available in the stochastic (Krein space) case.

Briefly, given a possibly indefinite quadratic form, our approach is to associate with it (by inspection) a Krein-space model whose stationary point will have the same gain k_0^* as for the deterministic problem. The Kalman filter (KF) recursions can now be invoked and give a recursive algorithm for the stationary point of the deterministic quadratic form; moreover, the condition for a minimum can also be expressed in terms of quantities easily related to the basic Riccati equations of the Kalman filter. These results are developed in Sections V and VI, with Theorems 5 and 6 being the major results.

While it is possible to pursue many of the results of this paper in greater depth, the development here is sufficient to solve several problems of interest in estimation theory. In the companion paper [1], we shall apply these results to H^∞ and risk-sensitive estimation and to finite memory adaptive filtering. In a future paper we shall study various dualities and apply them to obtain dual (or so-called complementary) state-space models and to solve the H^2, H^∞, and risk-sensitive control problems. We may mention that using these results we have also been able to develop the (possibly) numerically more attractive square root arrays and Chandrasekhar recursions for H^∞ problems [22], to study robust adaptive filtering [23], to obtain a stochastic interpretation of the Kalman–Yacubovich–Popov lemma, and to study convergence issues and obtain steady-state results. The point is that the many years of experience and intuition gained from the LQG or H^2 theory can be used as a guide to the corresponding H^∞ results.

A. Notation

A remark on the notation used in the paper. Elements in a Krein space are denoted by bold face letters, and elements in the Euclidean space of complex numbers are denoted by normal letters. Whenever the Krein-space elements and the Euclidean space elements satisfy the same set of constraints, we shall denote them by the same letters with the former ones being bold and the latter ones being normal. (This convention is similar to the one used in probability theory, where random variables are denoted by bold face letters and their assumed values are denoted by normal letters.)

II. On Krein Spaces

We briefly introduce the definitions and basic properties of Krein spaces, focusing on those results that we shall need later. Detailed expositions can be found in books [9]–[11]. Most readers will be familiar with finite-dimensional (often called Euclidean) and infinite-dimensional Hilbert spaces. Finite-dimensional (often called Minkowski) and infinite-dimensional Krein spaces share many of the properties Hilbert spaces but differ in some important ways that we shall emphasize in the following.

Definition 1 (Krein Spaces): An abstract vector space $\{\mathcal{K}, \langle\cdot,\cdot\rangle\}$ that satisfies the following requirements is called a *Krein Space*:

i) $\mathcal{K}$ is a linear space over $\mathcal{C}$, the complex numbers.
ii) There exists a bilinear form $\langle\cdot,\cdot\rangle \in \mathcal{C}$ on $\mathcal{K}$ such that
   a) $\langle \boldsymbol{y}, \boldsymbol{x}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle^*$.
   b) $\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{z}\rangle = a\langle \boldsymbol{x}, \boldsymbol{z}\rangle + b\langle \boldsymbol{y}, \boldsymbol{z}\rangle$
   for any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{K}$, $a, b \in \mathcal{C}$, and where $*$ denotes complex conjugation.
iii) The vector space $\mathcal{K}$ admits a direct orthogonal sum decomposition

$$\mathcal{K} = \mathcal{K}_+ \oplus \mathcal{K}_-$$

   such that $\{\mathcal{K}_+, \langle\cdot,\cdot\rangle\}$ and $\{\mathcal{K}_-, -\langle\cdot,\cdot\rangle\}$ are Hilbert spaces, and

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$$

   for any $\boldsymbol{x} \in \mathcal{K}_+$ and $\boldsymbol{y} \in \mathcal{K}_-$.

Remarks:

1) Recall that Hilbert spaces satisfy not only i), ii)-a), and ii)-b) above, but also the requirement that

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle > 0 \quad \text{when} \quad \boldsymbol{x} \neq 0.$$

2) The fundamental decomposition of $\mathcal{K}$ defines two projection operators $\mathcal{P}_+$ and $\mathcal{P}_-$ such that

$$\mathcal{P}_+\mathcal{K} = \mathcal{K}_+ \quad \text{and} \quad \mathcal{P}_-\mathcal{K} = \mathcal{K}_-.$$

   Therefore, for every $\boldsymbol{x} \in \mathcal{K}$ we can write

$$\boldsymbol{x} = P_+\boldsymbol{x} + P_-\boldsymbol{x} = \boldsymbol{x}_+ + \boldsymbol{x}_-, \boldsymbol{x}_\pm \in \mathcal{K}_\pm.$$

   Note that for every $\boldsymbol{x} \in \mathcal{K}_+$, we have $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$, but the converse is not true: $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$ does not necessarily imply that $\boldsymbol{x} \in \mathcal{K}_+$.
3) A vector $\boldsymbol{x} \in \mathcal{K}$ will be said to be positive if $\langle \boldsymbol{x}, \boldsymbol{x}\rangle > 0$, neutral if $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$, or negative if $\langle \boldsymbol{x}, \boldsymbol{x}\rangle < 0$. Correspondingly, a subspace $\mathcal{M} \subset \mathcal{K}$ can be positive, neutral, or negative, if all its elements are so, respectively.

We now focus on linear subspaces of $\mathcal{K}$. We shall define $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ as the linear subspace of $\mathcal{K}$ spanned by the elements $\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ in $\mathcal{K}$. The Gramian of the collection of elements $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ is defined as the $(N+1) \times (N+1)$ matrix

$$R_y \triangleq \left[\langle \boldsymbol{y}_i, \boldsymbol{y}_j\rangle\right]_{i,j=0:N}. \tag{1}$$

The reflexivity property, $\langle \boldsymbol{y}_i, \boldsymbol{y}_j\rangle = \langle \boldsymbol{y}_j, \boldsymbol{y}_i\rangle^*$, shows that the Gramian is a Hermitian matrix.

It is useful to introduce some matrix notation here. We shall write the column vector of the $\{\boldsymbol{y}_i\}$ as

$$\boldsymbol{y} = \text{col}\{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\}$$

and denote the above Gramian of the $\{\boldsymbol{y}_i\}$ as

$$R_y \triangleq \langle \boldsymbol{y}, \boldsymbol{y}\rangle.$$

(A useful mnemonic device for recalling this is to think of the $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ as "random variables" and their Gramian as the "covariance matrix"

$$R_y = [E\boldsymbol{y}_i\boldsymbol{y}_j^*] = E\boldsymbol{y}\boldsymbol{y}^*$$

where $E(\cdot)$ denotes "expectation." We use the quotation marks because in our context, the covariance matrix will generally be indefinite, so we are dealing with some kind of generalized "random variables." We do not pursue this interpretation here since our aim is only to provide readers with a convenient device for interpreting the shorthand notation.)

Also, if we have two sets of elements $\{\boldsymbol{z}_0, \cdots, \boldsymbol{z}_M\}$ and $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ we shall write

$$\boldsymbol{z} = \text{col}\{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\}$$

and

$$\boldsymbol{y} = \text{col}\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$$

and introduce the $(M+1) \times (N+1)$ cross-Gramian matrix

$$R_{zy} = [\langle \boldsymbol{z}_i, \boldsymbol{y}_j \rangle]_{\substack{i=0:M \\ j=0:N}} \triangleq \langle \boldsymbol{z}, \boldsymbol{y} \rangle.$$

Note the property

$$R_{zy} = R_{yz}^*.$$

We now proceed with a simple result.

Lemma 1 (Positive and Negative Linear Subspaces): Suppose $\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N$ are linearly independent elements of $\mathcal{K}$. Then $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ is a "positive" (negative) subspace of $\mathcal{K}$ if, and only if

$$R_y > 0 (R_y < 0).$$

Proof: Since the $\boldsymbol{y}_i$ are linearly independent, for any $\boldsymbol{z} \neq 0 \in \mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ there exists a unique $k \in \mathcal{C}^{N+1}$ such that $\boldsymbol{z} = k^*\boldsymbol{y}$. Now

$$\langle \boldsymbol{z}, \boldsymbol{z} \rangle = k^* \langle \boldsymbol{y}, \boldsymbol{y} \rangle k = k^* R_y k$$

so that $\langle z, z \rangle > 0$ for all $\boldsymbol{z} \in \mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, if, and only if, $R_y > 0$. The proof for $R_y < 0$ is similar. □

Note that any linear subspace whose Gramian has mixed inertia (both positive and negative eigenvalues) will have elements in both the positive and negative subspaces.

A. A Geometric Interpretation

Indefinite metric spaces were perhaps first introduced into the solution of physical problems via the finite-dimensional Minkowski spaces of special relativity [12], and some geometric insight may be gained by considering the special three-dimensional Minkowski space of Fig. 1, defined by the inner product

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = x_1x_2 + y_1y_2 - t_1t_2$$

when

$$\boldsymbol{v}_1 = (x_1, y_1, t_1), \quad \boldsymbol{v}_2 = (x_2, y_2, t_2) \quad \text{and} \quad x_i, y_i, t_i \in \mathcal{C}.$$

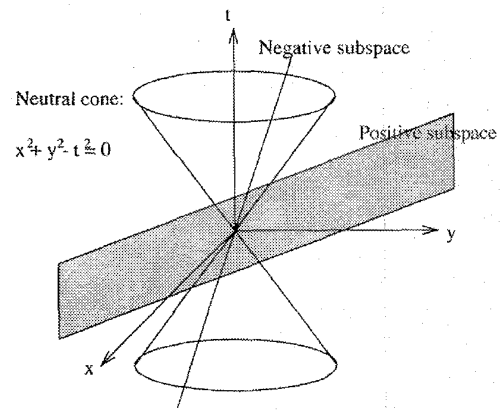


Fig. 1. Three-dimensional Minkowski space.

The (indefinite) squared norm of each vector $\boldsymbol{v} = (x, y, t)$ is equal to

$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle = x^2 + y^2 - t^2.$$

In this case, we can take $\mathcal{K}_+$ to be the $x - y$ plane and $\mathcal{K}_-$ as the t-axis. The neutral subspace is given by the cone, $x^2 + y^2 - t^2 = 0$, with points inside the cone belonging to the negative subspace, $x^2 + y^2 - t^2 < 0$, and points outside the cone corresponding to the positive subspace, $x^2 + y^2 - t^2 > 0$.

Moreover, any plane passing through the origin but lying outside the neutral cone will have positive definite Gramian, and any line passing through the origin and inside the neutral cone will have negative definite Gramian. Also, any plane passing through the origin that intersects the neutral cone will have Gramian with mixed inertia, and any plane tangent to the cone will have singular Gramian.

Two key differences between Krein spaces and Hilbert spaces are the existence of neutral and isotropic vectors. As mentioned earlier, a neutral vector is a nonzero vector that has zero length; an isotropic vector is a nonzero vector lying in a linear subspace of $\mathcal{K}$ that is orthogonal to every element in that linear subspace. There are obviously no such vectors in Euclidean or Hilbert spaces. In the Minkowski space described above, $[1\ 1\ \sqrt{2}]$ is a neutral vector, and if one considers the linear subspace $\mathcal{L}\{[1\ 1\ \sqrt{2}],\ [\sqrt{2}\ 0\ 1]\}$, then $[1\ 1\ \sqrt{2}]$ is also an isotropic vector in this linear subspace.

III. Projections in Krein Spaces

An important notion in both Hilbert and Krein spaces is that of the projection onto a subspace.

Definition 2 (Projections): Given the element $\boldsymbol{z}$ in $\mathcal{K}$ and the elements $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\}$ also in $\mathcal{K}$, we define $\hat{\boldsymbol{z}}$ to be the projection of $\boldsymbol{z}$ onto $\mathcal{L}\{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\}$ if

$$\boldsymbol{z} = \hat{\boldsymbol{z}} + \tilde{\boldsymbol{z}} \tag{2}$$

where $\hat{\boldsymbol{z}} \in \mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ and $\tilde{\boldsymbol{z}}$ satisfies the orthogonality condition

$$\tilde{\boldsymbol{z}} \perp \mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$$

or equivalently, $\langle \tilde{\boldsymbol{z}}, \boldsymbol{y}_i \rangle = 0$ for $i = 0, 1, \cdots, N$.

In Hilbert space, projections always exist and are unique. In Krein space, however, this is not always the case. Indeed we have the following result, where for simplicity we shall write $\mathcal{L}\{\boldsymbol{y}\} \triangleq \mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$.

Lemma 2 (Existence and Uniqueness of Projections): In the Hilbert space setting, projections always exist and are unique. In the Krein-space setting, however:

a) If the Gramian matrix $R_y = \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ is nonsingular, then the projection of $\boldsymbol{z}$ onto $\mathcal{L}\{\boldsymbol{y}\}$ exists, is unique, and is given by

$$\hat{\boldsymbol{z}} = \langle \boldsymbol{z}, \boldsymbol{y} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle^{-1} \boldsymbol{y} = R_{zy} R_y^{-1} \boldsymbol{y}. \tag{3}$$

b) If the Gramian matrix $R_y = \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ is singular, then

i) If $\mathcal{R}(R_{yz}) \subseteq \mathcal{R}(R_y)$ (where $\mathcal{R}(A)$ denotes the column range space of the matrix A), the projection $\hat{\boldsymbol{z}}$ exists but is nonunique. In fact, $\hat{\boldsymbol{z}} = k_0^* \boldsymbol{y}$, where k_0 is "any" solution to the linear matrix equation

$$R_y k_0 = R_{yz}. \tag{4}$$

ii) If $\mathcal{R}(R_{yz}) \not\subseteq \mathcal{R}(R_y)$, the projection $\hat{\boldsymbol{z}}$ does not exist.

Proof: Suppose $\hat{\boldsymbol{z}}$ is a projection of $\boldsymbol{z}$ onto the desired space. By (2), we can write

$$\boldsymbol{z} = k_0^* \boldsymbol{y} + \tilde{\boldsymbol{z}}$$

for some $k_0 \in \mathcal{C}^{(N+1)}$. Since $\langle \tilde{\boldsymbol{z}}, \boldsymbol{y} \rangle = 0$

$$R_{zy} = \langle \boldsymbol{z}, \boldsymbol{y} \rangle = k_0^* \langle \boldsymbol{y}, \boldsymbol{y} \rangle + 0 = k_0^* R_y. \tag{5}$$

If R_y is nonsingular, then the solution for k in (5) is unique and the projection is given by (3). If R_y is singular, two things may happen: either $\mathcal{R}(R_{yz}) \subseteq \mathcal{R}(R_y)$, in which case (5) will have a nonunique solution (since any k_1^* in the left null space of R_y can be added to k_0^*), or $\mathcal{R}(R_{yz}) \not\subseteq \mathcal{R}(R_y)$, in which case the projection does not exist since a solution to (5) does not exist.

In Hilbert spaces the projection always exists because it is always true that $\mathcal{R}(R_{yz}) \subseteq \mathcal{R}(R_y)$, or equivalently, that $\mathcal{N}(R_y) \subseteq \mathcal{N}(R_{zy})$ where $\mathcal{N}(A)$ is the right nullspace of the matrix A. To show this, suppose that $l \in \mathcal{N}(R_y)$. Then

$$\begin{aligned} R_y l = 0 &\Rightarrow l^* R_y l = 0 \\ &\Rightarrow l^* \langle \boldsymbol{y}, \boldsymbol{y} \rangle l = \langle l^* \boldsymbol{y}, l^* \boldsymbol{y} \rangle = 0 \\ &\Rightarrow l^* \boldsymbol{y} = 0 \end{aligned}$$

where the last equality follows from the fact that in Hilbert spaces $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = 0$. We now readily conclude that $\langle \boldsymbol{z}, l^* \boldsymbol{y} \rangle = R_{zy} l = 0$, i.e., $l \in \mathcal{N}(R_{zy})$ and hence $\mathcal{N}(R_y) \subseteq \mathcal{N}(R_{zy})$. Therefore a solution to (5) (and hence a projection) always exists in Hilbert spaces.

In Hilbert spaces the projection is also unique because if k_1 and k_2 are two different solutions to (5), then $(k_1 - k_2)^* R_y = 0$. But the above argument shows that we must then have $(k_1 - k_2)\boldsymbol{y} = 0$. Hence the projection

$$\hat{\boldsymbol{z}} = k_1^* \boldsymbol{y} = k_2^* \boldsymbol{y}$$

is unique. □

The proof of the above lemma shows that in Hilbert spaces the singularity of R_y implies that the $\{\boldsymbol{y}_i\}$ are linearly dependent, i.e.,

$$\det(R_y) = 0 \Leftrightarrow k^* \boldsymbol{y} = 0 \text{ for some vector } k \in \mathcal{C}^{N+1}.$$

In the Krein-space setting, all we can deduce from the singularity of R_y is that there exists a linear combination of the $\{\boldsymbol{y}_i\}$ that is orthogonal to every vector in $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, i.e., that $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ contains an isotropic vector. This follows by noting that for any complex matrix k_1, and for any k in the null space of R_y, we have

$$k_1^* R_y k = \langle k_1^* \boldsymbol{y}, k^* \boldsymbol{y} \rangle = 0$$

which shows that the linear combination $k^* \boldsymbol{y}$ is orthogonal to $k_1^* \boldsymbol{y}$, for every k_1, i.e., $k^* \boldsymbol{y}$ is an isotropic vector in $\mathcal{L}\{\boldsymbol{y}\}$.

Standing Assumption: Since existence and uniqueness will be important for all our future results, we shall make the standing assumption that the Gramian

$$R_y \quad \text{is nonsingular.}$$

A. Vector-Valued Projections

Consider the n-vector $\boldsymbol{z} = \text{col}\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n\}$ composed of elements $\boldsymbol{z}_i \in \mathcal{K}$, and the set $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, where $\boldsymbol{y}_j \in \mathcal{K}$; project each element $\boldsymbol{z}_i$ onto $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ to obtain $\hat{\boldsymbol{z}}_i$. We define $\hat{\boldsymbol{z}} = \text{col}\{\hat{\boldsymbol{z}}_1, \cdots, \hat{\boldsymbol{z}}_n\}$ as the projection of $\boldsymbol{z}$ onto $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$. (Strictly speaking, we should call $\hat{\boldsymbol{z}} \in \mathcal{K}^n$ the projection of $\boldsymbol{z} \in \mathcal{K}^n$ onto $\mathcal{L}^n\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, since it is an element of $\mathcal{L}^n\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ and not $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$. For simplicity, however, we shall generally use the looser terminology.)

It is easy to see that the results on the existence and uniqueness of projections in Lemma 2 continue to hold in the vector case as well.

In this connection, it will be useful to introduce a slight generalization of the definition of Krein spaces that was given in Section II. There, in Definition 1, we mentioned that $\mathcal{K}$ should be linear over the field of complex numbers, $\mathcal{C}$. It turns out, however, that we can replace $\mathcal{C}$ with any ring $\mathcal{S}$. In other words, the first two axioms for Krein spaces can be replaced by:

i) $\mathcal{K}$ is a linear space over the ring $\mathcal{S}$.

ii) There exists a bilinear form $\langle \cdot, \cdot \rangle \in \mathcal{S}$ on $\mathcal{K}$ such that

  a) $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle^*$

  b) $\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{z} \rangle = a\langle \boldsymbol{x}, \boldsymbol{z} \rangle + b\langle \boldsymbol{y}, \boldsymbol{z} \rangle$

  for any $\boldsymbol{x}, y, z \in \mathcal{K}$ and $a, b \in \mathcal{S}$, and where the operation $*$ depends on the ring $\mathcal{S}$.

  When the inner product $\langle \cdot, \cdot \rangle \in \mathcal{S}$ is positive, $\{\mathcal{K}, \langle \cdot, \cdot \rangle\}$ is referred to as a module. Thus the third axiom for Krein spaces can be replaced by iii).

iii) The vector space $\mathcal{K}$ admits a direct orthogonal sum decomposition

$$\mathcal{K} = \mathcal{K}_+ \oplus \mathcal{K}_-$$

  such that $\{\mathcal{K}_+, \langle \cdot, \cdot \rangle\}$ and $\{\mathcal{K}_-, -\langle \cdot, \cdot \rangle\}$ are modules, and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ for any $\boldsymbol{x} \in \mathcal{K}_+$ and $\boldsymbol{y} \in \mathcal{K}_-$.

The most important case for us is when $\mathcal{S}$ is a ring of complex matrices, and the operation $*$ denotes Hermitian transpose.

The point of this generalization is that we can now directly define the projection of a vector $\boldsymbol{z} \in \mathcal{K}^n$ onto $\mathcal{L}^n\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ as an element $\hat{\boldsymbol{z}} \in \mathcal{L}^n\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, such that

$$\hat{\boldsymbol{z}} = k_0^* \boldsymbol{y}, \quad k_0^* \in \mathcal{C}^{n \times N}$$

where k is such that

$$0 = \langle \boldsymbol{z} - k_0^* \boldsymbol{y}, \boldsymbol{y} \rangle \triangleq R_{zy} - k_0^* R_y$$

or

$$k_0^* R_y = R_{zy}.$$

Finally, let us remark that to avoid additional notational burden, we shall often refrain from writing $\mathcal{K}^n$ and shall simply use the notation $\mathcal{K}$ for any Krein space. The ring $\mathcal{S}$ over which the Krein space is defined will be obvious from the context.

IV. Projections and Quadratic Forms

In Hilbert space, projections extremize (minimize) certain quadratic forms, as we shall briefly first describe. In Krein spaces, we can in general only assert that projections stationarize such quadratic forms; further conditions need to be met for the stationary points to be extrema (minima). This will be elaborated in Section IV-A, in the context of (what we shall call) a stochastic minimization problem. In Section IV-B, we shall study a closely related quadratic form arising in what we shall call a partially equivalent deterministic minimization problem.

A. Stochastic Minimization Problems in Hilbert and Krein Spaces

Consider a collection of elements $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ in a Krein space $\mathcal{K}$ with indefinite inner product $\langle \cdot, \cdot \rangle$. Let $\boldsymbol{z} = \mathrm{col}\{\boldsymbol{z}_0, \cdots, \boldsymbol{z}_M\}$ be some column vector of elements in $\mathcal{K}$, and consider an arbitrary linear combination of $\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$, say $k^* \boldsymbol{y}$, where $k^* \in \mathcal{C}^{(M+1) \times (N+1)}$ and $\boldsymbol{y} = \mathrm{col}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$. A natural object to study is the error Gramian

$$P(k) = \langle \boldsymbol{z} - k^* \boldsymbol{y}, \boldsymbol{z} - k^* \boldsymbol{y} \rangle. \tag{6}$$

To motivate the subsequent discussion, let us first assume that the $\{\boldsymbol{y}_i\}$ and $\{\boldsymbol{z}_j\}$ belong to a Hilbert space of zero-mean random variables and that their variance and cross-variances are known. In this case the inner product is $\langle \boldsymbol{z}_i, \boldsymbol{y}_j \rangle_{\mathcal{H}} = E \boldsymbol{z}_i \boldsymbol{y}_j^*$ (where $E(\cdot)$ denotes expectation), and $P(k)$ is simply the mean-square-error (or error variance) matrix in estimating $\boldsymbol{z}$ using $k^* \boldsymbol{y}$, viz.

$$P(k) = E(\boldsymbol{z} - k^* \boldsymbol{y})(\boldsymbol{z} - k^* \boldsymbol{y})^* \doteq \|\boldsymbol{z} - k^* \boldsymbol{y}\|_{\mathcal{H}}^2.$$

It is well known that the linear least-mean-square estimate, which minimizes $P(k)$, is given by the projection of $\boldsymbol{z}$ on $\mathcal{L}\{\boldsymbol{y}\}$

$$\hat{\boldsymbol{z}} = k_0^* \boldsymbol{y}$$

where

$$k_0^* = E \boldsymbol{z} \boldsymbol{y}^* [E \boldsymbol{y} \boldsymbol{y}^*]^{-1} = R_{zy} R_y^{-1}.$$

The simple proof will be instructive. Thus note that

$$\begin{aligned} P(k) &= \|\boldsymbol{z} - k^* \boldsymbol{y}\|_{\mathcal{H}}^2 \\ &= \|\boldsymbol{z} - \hat{\boldsymbol{z}} + \hat{\boldsymbol{z}} - k^* \boldsymbol{y}\|_{\mathcal{H}}^2 \\ &= \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|_{\mathcal{H}}^2 + \|\hat{\boldsymbol{z}} - k^* \boldsymbol{y}\|_{\mathcal{H}}^2 \end{aligned}$$

since by the definition of $\hat{\boldsymbol{z}}$, it holds that

$$\langle \boldsymbol{z} - \hat{\boldsymbol{z}}, \hat{\boldsymbol{z}} - k^* \boldsymbol{y} \rangle_{\mathcal{H}} = 0.$$

Clearly, since $\hat{\boldsymbol{z}} = k_0^* \boldsymbol{y}$

$$P(k) \geq P(k_0)$$

with equality achieved only when $k = k_0$.

This argument breaks down, however, when the elements are in a Krein space, since then we could have

$$\|\hat{\boldsymbol{z}} - k^* \boldsymbol{y}\|^2 = \|k_0^* \boldsymbol{y} - k^* \boldsymbol{y}\|^2 = 0, \text{ even if } k_0 \neq k.$$

All we can assert is that

$$k_0^* \boldsymbol{y} - k^* \boldsymbol{y} = \text{an isotropic vector in the linear subspace spanned by } \{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}.$$

Moreover, since $\|k_0^* \boldsymbol{y} - k^* \boldsymbol{y}\|^2$ could be negative, it is not true that $P(k)$ will be minimized by choosing $k = k_0$. So a closer study is necessary.

We shall start with a definition.

Definition 3 (Stationary Point): The matrix $k_0 \in \mathcal{C}^{(N+1)} \times (M+1)$ is said to be a stationary point of an $(M+1) \times (M+1)$ matrix quadratic form in k, say

$$P(k) = A + Bk + k^* B^* + k^* C k$$

iff $k_0 a$ is a stationary point of the "scalar" quadratic form $a^* P(k) a$ for all complex column vectors $a \in \mathcal{C}^{M+1}$, i.e., iff

$$\left. \frac{\partial a^* P(k) a}{\partial k a} \right|_{k=k_0} = 0.$$

Now we can prove the following.

Lemma 3 (Condition for Minimum): A stationary point of $P(k)$ is a minimum iff for all $a \in \mathcal{C}^{M+1}$

$$\left. \frac{\partial^2 a^* P(k) a}{\partial (k a)^2} \right|_{k=k_0} \geq 0. \tag{7}$$

Moreover, it is a unique minimum iff

$$\left. \frac{\partial^2 a^* P(k) a}{\partial (k a)^2} \right|_{k=k_0} > 0. \tag{8}$$

$$a^* P(k) a = a^* P(k_0) a + \underbrace{\left. \frac{\partial a^* P(k) a}{\partial k a} \right|_{k=k_0}}_{=0} \cdot (k - k_0) a + a^* (k - k_0)^* \left. \frac{\partial^2 a^* P(k) a}{\partial (k a)^2} \right|_{k=k_0} \cdot (k - k_0) a$$

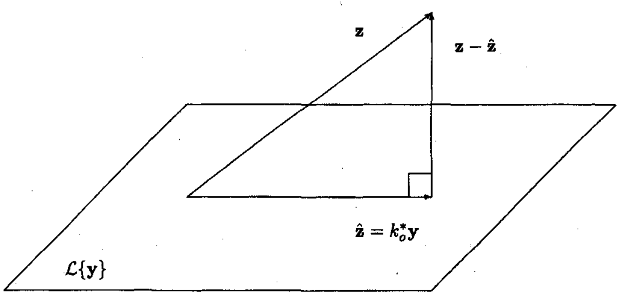


Fig. 2. The projection $\hat{z} = k_0^* \boldsymbol{y}$ stationarizes the error Gramian $P(k) = \langle z - k^*\boldsymbol{y}, z - k^*\boldsymbol{y}\rangle$ over all $k^*\boldsymbol{y} \in \mathcal{L}\{\boldsymbol{y}\}$.

Proof: Writing the Taylor series expansion of $a^*P(k)a$ around the stationary point k_0 yields (since $a^*P(k)a$ is quadratic in ka), as shown at the bottom of the previous page, or equivalently

$$\begin{aligned} &a^*P(k)a - a^*P(k_0)a \\ &\quad = a^*(k-k_0)^* \left.\frac{\partial^2 a^*P(k)a}{\partial(ka)^2}\right|_{k=k_0} \cdot (k-k_0)a. \end{aligned}$$

Using the above expression, we see that k_0 is a minimum, i.e., $a^*P(k)a - a^*P(k_0)a \geq 0$ for all $k \neq k_0$ iff (7) is satisfied. Moreover, k_0 will be a unique minimum, i.e., $a^*P(k)a - a^*P(k_0)a > 0$ for all $k \neq k_0$ iff (8) is satisfied. □

Let us now return to the error Gramian $P(k)$ in (6) and expand it as

$$P(k) = \langle \boldsymbol{z}, \boldsymbol{z}\rangle_{\mathcal{K}} - \langle \boldsymbol{z}, \boldsymbol{y}\rangle_{\mathcal{K}} k - k^*\langle \boldsymbol{y}, \boldsymbol{z}\rangle_{\mathcal{K}} + k^*\langle \boldsymbol{y}, \boldsymbol{y}\rangle_{\mathcal{K}} k \quad \text{(9a)}$$

or more compactly

$$P(k) = \begin{bmatrix} I & -k^* \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} \begin{bmatrix} I \\ -k \end{bmatrix}. \quad \text{(9b)}$$

Note that the center matrix appearing in (9b) is the Gramian of the vector $\mathrm{col}\{\boldsymbol{z}, \boldsymbol{y}\}$.

For this particular quadratic form, we can use the easily verified triangular factorization (recall our standing assumption that R_y is nonsingular)

$$\begin{aligned} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} &= \begin{bmatrix} I & R_{zy}R_y^{-1} \\ 0 & I \end{bmatrix} \\ &\quad \times \begin{bmatrix} R_z - R_{zy}R_y^{-1}R_{yz} & 0 \\ 0 & R_y \end{bmatrix} \begin{bmatrix} I & 0 \\ R_y^{-1}R_{yz} & I \end{bmatrix} \end{aligned} \quad \text{(10)}$$

to write

$$\begin{aligned} a^*P(k)a &= \begin{bmatrix} a^* & a^*k^* - a^*R_{zy}R_y^{-1} \end{bmatrix} \\ &\quad \times \begin{bmatrix} R_z - R_{zy}R_y^{-1}R_{yz} & 0 \\ 0 & R_y \end{bmatrix} \begin{bmatrix} a \\ ka - R_y^{-1}R_{yz}a \end{bmatrix}. \end{aligned} \quad \text{(11)}$$

Calculating the stationary point of $P(k)$ and the corresponding condition for a minimum is now straightforward. Note, moreover, that R_y nonsingular implies that the stationary point is unique.

Theorem 1 (Stationary Point of the Error Gramian): When R_y is nonsingular, k_0, the unique coefficient matrix in the projection of $\boldsymbol{z}$ onto $\mathcal{L}\{\boldsymbol{y}\}$

$$\hat{\boldsymbol{z}} = k_0^*\boldsymbol{y}, \quad k_0 = R_y^{-1}R_{yz}$$

yields the unique stationary point of the error Gramian

$$\begin{aligned} P(k) &\triangleq \langle \boldsymbol{z} - k^*\boldsymbol{y}, \boldsymbol{z} - k^*\boldsymbol{y}\rangle \\ &= \begin{bmatrix} I & -k^* \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} \begin{bmatrix} I \\ -k \end{bmatrix} \end{aligned} \quad \text{(12)}$$

over all $k \in \mathcal{C}^{(N+1)\times(M+1)}$. Moreover, the value of $P(k)$ at the stationary point is given by

$$P(k_0) = R_z - R_{zy}R_y^{-1}R_{yz}.$$

Proof: The claims follow easily from (11) by differentiation. □

Further differentiation and use of Lemma 3 yields the following result.

Corollary 1 (Condition for a Minimum): In Theorem 1, k_0 is a unique minimum iff

$$R_y > 0$$

i.e., R_y is not only nonsingular but also positive definite.

B. A Partially Equivalent Deterministic Problem

We shall now consider what we call a partially equivalent deterministic problem. We refer to it as deterministic because it involves computing the stationary point of a certain scalar quadratic form over ordinary complex variables (not Krein space ones). Moreover, it is called partially equivalent since its solution, i.e., the stationary point, is given by the same expression as the projection of one suitably defined Krein-space vector onto another, while the condition for a minimum is different than that for the Krein-space projection.

To this end, consider the scalar second-order form

$$J(z, y) \triangleq \begin{bmatrix} z^* & y^* \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} \begin{bmatrix} z \\ y \end{bmatrix} \quad \text{(13)}$$

where the central matrix is the inverse of the Gramian matrix in the stochastic problem of Theorem 1 [see (9b)]. Suppose we seek the stationarizing element z_0 for a given y. [Of course now we assume not only that R_y is nonsingular, but so also the block matrix appearing in (13).] Note that z and y are no longer boldface, meaning that they are to be regarded as (ordinary) vectors of complex numbers.

Referring to the discussion at the beginning of Section IV-A on Hilbert spaces, the motivation for this problem is the fact that for jointly Gaussian random vectors $\{\boldsymbol{z}, \boldsymbol{y}\}$, the linear least-mean-squares estimate can be found as the conditional mean of the conditional density $p_{\boldsymbol{zy}}(z, y)/p_{\boldsymbol{y}}(y)$. When $\{\boldsymbol{z}, \boldsymbol{y}\}$ are zero-mean with covariance matrix $\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}$, taking logarithms of the conditional density results in the quadratic form (13) which is the negative of the so-called log-likelihood function. In this case, the relation between (13) and the projection follows from the fact that the linear least-mean-squares estimate is the same as the maximum likelihood estimate [obtained by minimizing (13)]. With this motivation, we now introduce and study the quadratic form $J(z,y)$ without any reference to $\{\boldsymbol{z},\boldsymbol{y}\}$ being Gaussian.

Theorem 2 (Deterministic Stationary Point): Suppose both R_y and the block matrix in (13) are nonsingular. Then

a) The stationary point z_0 of $J(z,y)$ over z is given by

$$z_0 = R_{zy}R_y^{-1}y.$$

b) The value of $J(z,y)$ at the stationary point is

$$J(z_0,y) = y^*R_y^{-1}y.$$

Corollary 2 (Condition for a Minimum): In Theorem 2, z_0 is a minimum iff

$$R_z - R_{zy}R_y^{-1}R_{yz} > 0.$$

Proof: We note that [see (10)]

$$\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -R_y^{-1}R_{yz} & I \end{bmatrix} \times \begin{bmatrix} R_z - R_{zy}R_y^{-1}R_{yz} & 0 \\ 0 & R_y \end{bmatrix}^{-1} \times \begin{bmatrix} I & -R_{zy}R_y^{-1} \\ 0 & I \end{bmatrix}$$

so that we can write

$$J(z,y) = [(z^* - y^*R_y^{-1}R_{yz}) \quad y^*] \times \begin{bmatrix} R_z - R_{zy}R_y^{-1}R_{yz} & 0 \\ 0 & R_y \end{bmatrix}^{-1} \begin{bmatrix} (z - R_{zy}R_y^{-1}y) \\ y \end{bmatrix}.$$

It now follows by differentiation that the stationary point of $J(z,y)$ is equal to $z_0 = R_{zy}R_y^{-1}y$, and that $J(z_0,y) = y^*R_y^{-1}y$. To prove the Corollary, we differentiate once again, and use Lemma 3. □

Remark 1: Comparing the results of Theorems 1 and 2 shows that the stationary point z_0, of the scalar quadratic form (13) is given by a formula that is exactly the same as that in Theorem 1 for the Krein-space projection of a vector $\boldsymbol{z}$ onto the linear span $\mathcal{L}\{\boldsymbol{y}\}$. In Theorem 2, however, there is no Krein space: z and y are just vectors (in general of different dimensions) in Euclidean space and z_0 is not the projection of z onto the vector y. What we have shown in Theorem 2 is that by properly defining the scalar quadratic form as in (13) using coefficient matrices R_z, R_y, R_{zy}, and R_{yz} that are arbitrary but can be regarded as being obtained from Gramians and cross-Gramians of some Krein-space vectors $\{\boldsymbol{z},\boldsymbol{y}\}$, we can calculate the stationary point using the same recipe as in Theorem 1.

Remark 2: Although the stationary points of the matrix quadratic form $P(k)$ and the scalar quadratic form $J(z,y)$ are found by the same computations, the two forms do not necessarily simultaneously have a minimum, since one requires the condition $R_y > 0$ (Corollary 1), and the other requires the condition $R_z - R_{zy}R_y^{-1}R_{yz} > 0$ (Corollary 2). This is the major difference from the classical Hilbert space context where we have

$$\left\langle \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{y} \end{bmatrix}, \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{y} \end{bmatrix} \right\rangle_{\mathcal{H}} = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} > 0. \tag{14}$$

When (14) holds, the approaches of Theorems 1 and 2 give equivalent results.

Corollary 3 (Simultaneous Minima): For vectors $\boldsymbol{z}$ and $\boldsymbol{y}$ of linear independent elements in a Hilbert space $\mathcal{H}$, the conditions $R_z - R_{zy}R_y^{-1}R_{yz} > 0$ and $R_y > 0$ occur simultaneously.

Proof: Immediate from the factorization (10). □

We shall see in more detail in Part II, and to some extent in Section VI-B of this paper, that this difference is what makes H^∞ (and risk-sensitive and finite memory adaptive filtering) results different from H^2 results. Briefly, H^∞ problems will lead directly to certain indefinite quadratic forms: to stationarize them we shall find it useful to set up the corresponding Krein-space problem and appeal to Theorem 1. While this will give an algorithm, further work will be necessary to check for the minimum condition of Theorem 2 in the H^∞ problem.

It is this difference that leads us to say that the deterministic problem is only partially equivalent to the stochastic problem of Section IV-A. (We may remark that we are making a distinction between equivalence and "duality": one can in fact define duals to both the above problems, but we defer this topic to another occasion.)

Remark 3: Finally, recall that Lemma 2 on the existence and uniqueness of the projection implies that the stochastic problem of Theorem 1 has a unique solution if, and only if, R_y is nonsingular, thus explaining our standing assumption. The following result is the analog for the deterministic problem.

Lemma 4 (Existence of Stationarizing Solutions): The deterministic problem of Theorem 2 has a unique stationarizing solution for all y if, and only if, R_y is nonsingular.

Proof: Let us denote

$$\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

so that

$$J(z,y) = [z^* \quad y^*] \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix}.$$

If $J(z,y)$ has a unique stationarizing solution for all y, then A must be nonsingular (since by differentiation the stationary point must satisfy the equation $Az_0 = By$). But the invertibility of A and the whole center matrix appearing in $J(z,y)$ imply the invertibility of the Schur complement $C - B^*A^{-1}B$. But it is easy to check that this Schur complement must be the inverse of R_y. Thus R_y must be invertible.

On the other hand if R_y is invertible, then the deterministic problem has a unique stationarizing solution as given by Theorem 2. □

C. Alternative Inertia Conditions for Minima

In many cases it can be complicated to directly check for the positivity condition of the deterministic problem, namely $R_z - R_{zy}R_y^{-1}R_{yz} > 0$. On the other hand, it is often easier to compute the inertia (the number of positive, negative, and zero eigenvalues) of R_y itself. This often suffices [24].

Lemma 5 (Inertia Conditions for Deterministic Minimization):

a) If R_y and R_z are nonsingular, then the deterministic problem of Theorem 2 will have a minimizing solution (i.e., $R_z - R_{zy}R_y^{-1}R_{yz}$ will be > 0) if, and only if

$$I_-[R_y] = I_-[R_z] + I_-[(R_y - R_{yz}R_z^{-1}R_{zy})] \tag{15}$$

where $I_-[A]$ denotes the negative inertia (number of negative eigenvalues) of A.

b) When $R_z > 0$ (rather than just being nonsingular) then we will have a minimizing solution iff

$$I_-[R_y] = I_-[R_y - R_{yz}R_z^{-1}R_{zy}] \tag{16}$$

i.e., if, and only if, R_y and $R_y - R_{yz}R_z^{-1}R_{zy}$ have the same inertia.

Proof: If R_y and R_z are both nonsingular, then equating the lower-upper and upper-lower block triangular factorizations of the Gramian matrix in (10) will yield the result that

$$\begin{bmatrix} R_z - R_{zy}R_y^{-1}R_{yz} & 0 \\ 0 & R_y \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} R_z & 0 \\ 0 & R_y R_{yz}R_z^{-1}R_{zy} \end{bmatrix}$$

are congruent. By Sylvester's Law that congruent matrices have the same inertia [16], we have

$$\begin{aligned} &I_-[R_z - R_{zy}R_y^{-1}R_{yz}] + I_-[R_y] \\ &\quad = I_-[R_z] + I_-[(R_y - R_{yz}R_z^{-1}R_{zy})]. \end{aligned}$$

Now if (15) holds, then $I_-[R_z - R_{zy}R_y^{-1}R_{yz}] = 0$, so that $R_z - R_{zy}R_y^{-1}R_{yz} > 0$.

Conversely if $I_-[R_z - R_{zy}R_y^{-1}R_{yz}] = 0$, then (15) holds.

When $R_z > 0$, we have $I_-[R_z] = 0$, and (16) follows immediately. □

The general results presented so far can be made even more explicit when there is more structure in the problems. In particular, we shall see that when we have state-space structure both R_z and $R_y - R_{yz}R_z^{-1}R_{zy}$ are block-diagonal. Moreover, a Krein space–Kalman filter will yield a direct method for computing the inertia of R_y. Thus, when we have state-space structure, it will be much easier to use the results of Lemma 5 than to directly check for the positivity of $R_z - R_{zy}R_y^{-1}R_{yz}$ [22], [24].

V. State-Space Structure

One approach at this point is to begin by assuming that the components $\{\boldsymbol{y}_j\}$ of $\boldsymbol{y}$ arise from an underlying Krein space state-space model. To better motivate the introduction of such state-space models, however, we shall start with the following (indefinite) quadratic minimization problem.

Consider a system described by the state-space equations

$$\begin{cases} x_{j+1} = F_j x_j + G_j u_j, & 0 \le j \le N \\ y_j = H_j x_j + v_j \end{cases} \tag{17}$$

where $F_j \in \mathcal{C}^{n\times n}$, $G_j \in \mathcal{C}^{n\times m}$, and $H_j \in \mathcal{C}^{p\times n}$ are given matrices and the initial state $x_0 \in \mathcal{C}^n$, the driving disturbance $u_j \in \mathcal{C}^m$, and the measurement disturbance $v_j \in \mathcal{C}^p$, are unknown complex vectors. The output $y_j \in \mathcal{C}^p$ is assumed known for all j.

In many applications one is confronted with the following deterministic minimization problem: Given $\{y_j\}_{j=0}^N$, minimize over x_0 and $\{u_j\}_{j=0}^N$ the quadratic form

$$\begin{aligned} J(x_0, u, y) = x_0^*\Pi_0^{-1}x_0 + \sum\nolimits_{j=0}^{N} &[u_j^* \quad v_j^*] \\ \times &\begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ v_j \end{bmatrix} \end{aligned} \tag{18}$$

subject to the state-space constraints (17), and where $Q_j \in \mathcal{C}^{m\times m}$, $S_j \in \mathcal{C}^{m\times p}$, $R_j \in \mathcal{C}^{p\times p}$, $\Pi_0 \in \mathcal{C}^{n\times n}$ are (possibly indefinite) given Hermitian matrices.

The above deterministic quadratic form is usually encountered in filtering problems; a special case that we shall see in the companion paper is the H^∞-filtering problem where the weighting matrices are Π_0, $Q_j = I$, and $R_j = \begin{bmatrix} I & 0 \\ 0 & -\gamma_f^2 I \end{bmatrix}$, and where H_i is now replaced by $\text{col}\{H_i, L_i\}$. Another application arises in adaptive filtering in which case we usually have $u_j \equiv 0$ and $F_j \equiv I$ [15], [23]. In the general case, however, Π_0 represents the penalty on the initial state, and $\{Q_j, R_j, S_j\}$ represents the penalty on the driving and measurement disturbances $\{u_j, v_j\}$. (There is also a "dual" quadratic form that arises in control applications which we shall study elsewhere.)

Such deterministic problems can be solved via a variety of methods, such as dynamic programming or Lagrange multipliers (see, e.g., [5]), but we shall find it easier to use the equivalence discussed in Section IV: construct a (partially) equivalent Krein space (or stochastic) problem. To do so we first need to express the $J(x_0, u, y)$ of (18) in the form of (13) of Section IV-B.

For this, we first introduce some vector notation. Note that the states $\{x_j\}$ and the outputs $\{y_j\}$ are linear combinations of the fundamental quantities $\{x_0, \{u_j, v_j\}_{j=0}^N\}$. We introduce (the state transition matrix)

$$\Phi(j,k) = F_{j-1}\cdots F_{k+1}, j > k, \quad \Phi(j,j) = I$$

and define

$$h_{jk} \triangleq H_j F_{j-1}\cdots F_{k+1}G_k = H_j\Phi(j,k)G_k$$

as the response at time j to an impulse at time $k < j$ (assuming both $x_0 = 0$ and $v_k \equiv 0$).

Then with

$$\begin{aligned} y &\triangleq \text{col}\{y_0, \cdots, y_N\} \\ u &\triangleq \text{col}\{u_0, \cdots, u_N\} \\ v &\triangleq \text{col}\{v_0, \cdots, v_N\} \end{aligned}$$

the state-space equations (17) allow us to write

$$y = \mathcal{O}x_0 + \Gamma u + v = [\mathcal{O} \quad \Gamma]\begin{bmatrix} x_0 \\ u \end{bmatrix} + v \tag{19}$$

where $\mathcal{O}$ and Γ are the observability map and the impulse response matrix, respectively

$$\mathcal{O} = \begin{bmatrix} H_0 \\ H_1\Phi(1,0) \\ H_2\Phi(2,0) \\ \vdots \\ H_N\Phi(N,0) \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} 0 & & & & \\ h_{10} & 0 & & & \\ h_{20} & h_{21} & 0 & & \\ h_{30} & h_{31} & h_{32} & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

With these definitions we can rewrite $J(x_0, u, y)$ as

$$J(x_0,u,y) = [x_0^* \quad u^* \quad v^*] \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q & S \\ 0 & S^* & R \end{bmatrix}^{-1} \begin{bmatrix} x_0 \\ u \\ v \end{bmatrix} \tag{20}$$

where we have defined

$$Q \triangleq Q_0 \oplus \cdots \oplus Q_N$$
$$R \triangleq R_0 \oplus \cdots \oplus R_N$$
$$S \triangleq S_0 \oplus \cdots \oplus S_N.$$

Finally we make the change of coordinates

$$\begin{bmatrix} x_0 \\ u \\ y \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix} \begin{bmatrix} x_0 \\ y \\ v \end{bmatrix}$$

to obtain

$$\begin{aligned} J(x_0,u,y) &= \begin{bmatrix} x_0 \\ u \\ y \end{bmatrix}^* \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\mathcal{O} & -\Gamma & I \end{bmatrix}^* \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q & S \\ 0 & S^* & R \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\mathcal{O} & -\Gamma & I \end{bmatrix} \begin{bmatrix} x_0 \\ u \\ y \end{bmatrix} \\ &= \begin{bmatrix} x_0 \\ u \\ y \end{bmatrix}^* \left\{ \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix} \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q & S \\ 0 & S^* & R \end{bmatrix} \right. \\ &\quad \left. \cdot \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix}^* \right\}^{-1} \begin{bmatrix} x_0 \\ u \\ y \end{bmatrix}. \end{aligned} \tag{21}$$

This is now of the desired form (13) (with $z \triangleq \text{col}\{x_0, u\}$). Therefore, comparing with (12) in Theorem 1, we introduce a Krein space state-space model

$$\begin{cases} \boldsymbol{x}_{j+1} = F_j \boldsymbol{x}_j + G_j \boldsymbol{u}_j, & 0 \le j \le N \\ \boldsymbol{y}_j = H_j \boldsymbol{x}_j + \boldsymbol{v}_j \end{cases} \tag{22a}$$

where the initial state, $\boldsymbol{x}_0$, and the driving and measurement disturbances, $\{\boldsymbol{u}_j\}$ and $\{\boldsymbol{v}_j\}$, are such that

$$\left\langle \begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{v}_j \\ \boldsymbol{x}_0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{v}_k \\ \boldsymbol{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix} \delta_{jk} & 0 \\ 0 & \Pi_0 \end{bmatrix}. \tag{22b}$$

The condition (22b) is the Krein-space version of the usual assumption made in the stochastic (Hilbert space) state-space models, viz., that the initial condition $\boldsymbol{x}_0$ and the driving and measurement disturbances $\{\boldsymbol{u}_i, \boldsymbol{v}_i\}$ are zero-mean uncorrelated random variables with variance matrices Π_0 and $\begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}$, respectively, and that the $\{\boldsymbol{u}_i, \boldsymbol{v}_i\}$ form a white (uncorrelated) sequence. As mentioned before, the Krein-space elements can be thought of as some kind of generalized random variables.

Now if, as was done earlier, we define

$$\boldsymbol{y} = \text{col}\{\boldsymbol{y}_0, \cdots \boldsymbol{y}_N\}$$
$$\boldsymbol{u} = \text{col}\{\boldsymbol{u}_0, \cdots \boldsymbol{u}_N\}$$
$$\boldsymbol{v} = \text{col}\{\boldsymbol{v}_0, \cdots \boldsymbol{v}_N\}$$

then we can use the state-space model (22a) to write

$$\begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{u} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \tag{23}$$

and to see that

$$\begin{aligned} &\left\langle \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{u} \\ \boldsymbol{y} \end{bmatrix}, \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{u} \\ \boldsymbol{y} \end{bmatrix} \right\rangle = \\ &\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix} \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q & S \\ 0 & S^* & R \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \mathcal{O} & \Gamma & I \end{bmatrix}^* \end{aligned} \tag{24}$$

which is exactly the inverse of the central matrix appearing in expression (21) for $J(x_0, u, y)$. Therefore, referring to Theorems 1 and 2, the main point is that to find the stationary point of $J(x_0, u, y)$ over $\{x_0, u\}$, we can alternatively find the projection of $\{\boldsymbol{x}_0, \boldsymbol{u}\}$ onto $\mathcal{L}\{\boldsymbol{y}\}$ in the Krein-space model (22a).

Now that we have identified the stochastic and deterministic problems when a state-space structure is assumed, we can give the analogs of Theorems 1 and 2.

Lemma 6 (Stochastic Interpretation): Suppose $z = \text{col}\{\boldsymbol{x}_0, \boldsymbol{u}\}$ and $\boldsymbol{y}$ are related through the state-space model (22a) and (22b), and that R_y given by (27) is nonsingular. Then the stationary point of the error Gramian

$$\langle z - k^*\boldsymbol{y}, z - k^*\boldsymbol{y} \rangle \tag{25}$$

over all $k^*\boldsymbol{y}$ is given by the projection

$$\begin{bmatrix} \hat{\boldsymbol{x}}_{0|N} \\ \hat{\boldsymbol{u}}_{|N} \end{bmatrix} = \begin{bmatrix} \Pi_0 \mathcal{O}^* \\ Q\Gamma^* + S \end{bmatrix} R_y^{-1} \boldsymbol{y} \tag{26}$$

where

$$R_y = \mathcal{O}\Pi_0\mathcal{O}^* + [\Gamma \quad I] \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} \Gamma^* \\ I \end{bmatrix}. \tag{27}$$

Moreover this stationary point is a minimum if, and only if, $R_y > 0$.

We can now also give the analog result to Theorem 2.

Lemma 7 (Deterministic Quadratic Form): The expression

$$\begin{bmatrix} \hat{x}_{0|N} \\ \hat{u}_{|N} \end{bmatrix} = \begin{bmatrix} \Pi_0 \mathcal{O}^* \\ Q\Gamma^* + S \end{bmatrix} R_y^{-1} \boldsymbol{y} \tag{28}$$

yields the stationary point of the quadratic order form

$$\begin{aligned} J(x_0,u,y) &= x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{N} [u_j^* \quad (y_j - H_j x_j)^*] \\ &\quad \times \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix} \end{aligned} \tag{29a}$$

over x_0 and $u = \text{col}\{u_0, \cdots, u_N\}$, and subject to the state-space constraints

$$\begin{cases} x_{j+1} = F_j x_j + G_j u_j, & 0 \le j \le N \\ y_j = H_j x_j + v_j. \end{cases}$$

In particular, when $S_j \equiv 0$, the quadratic form is

$$J(x_0, u, y) = x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{N} u_j^* Q_j^{-1} u_j + \sum_{j=0}^{N} (y_j - H_j x_j)^* R_j^{-1} (y_j - H_j x_j). \quad (29b)$$

The value of $J(x_0, u, y)$ (with either $S_j \equiv 0$ or $S_j \neq 0$) at the stationary point is

$$J(\hat{x}_{0|N}, \hat{u}_{|N}, y) = y^* R_y^{-1} y.$$

A. The Conditions for a Minimum

As mentioned earlier, the important point is that the conditions for minima in these two problems are different: $R_y > 0$ in the stochastic problem, and

$$M \triangleq R_z - R_{zy} R_y^{-1} R_{yz} > 0 \quad \text{where} \quad \boldsymbol{z} = \text{col}\{\boldsymbol{x}_0, \boldsymbol{u}\}$$

in the deterministic problem. In the state-space case R_y is given by (27). In this section we shall explore the condition for a deterministic minimum under the state-space assumption. First note that for M we have (30) as shown at the bottom of the page.

Now we know that $M > 0$ iff both the (1, 1) block entry in (30) and its Schur complement are positive definite. The (1, 1) block entry may be identified as the Gramian of the error $\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|N}$, i.e.,

$$\Pi_0 - \Pi_0 \mathcal{O}^* R_y^{-1} \mathcal{O} \Pi_0 = \langle \boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|N}, \boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|N} \rangle \triangleq P_{0|N}. \quad (31)$$

To obtain a nice form for the Schur complement of the (1, 1) block entry, say Δ, we have to use a little matrix algebra. Recall that

$$\begin{aligned} R_y &= \mathcal{O}\Pi_0\mathcal{O}^* + [\Gamma \quad I] \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} \Gamma^* \\ I \end{bmatrix} \\ &= [\mathcal{O} \quad \Gamma + S^* Q^{-1}] \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \mathcal{O}^* \\ \Gamma^* + Q^{-1} S \end{bmatrix} \\ &\quad + R - S^* Q^{-1} S. \end{aligned}$$

Using the second expression for R_y and a well-known matrix inversion formula leads to the expression

$$M^{-1} = \begin{bmatrix} \Pi_0^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} + \begin{bmatrix} \mathcal{O}^* \\ \Gamma^* + Q^{-1} S \end{bmatrix} \times (R - S^* Q^{-1} S)^{-1} [\mathcal{O} \quad \Gamma + S^* Q^{-1}]. \quad (32)$$

Now we use another well-known fact: the (2, 2) block element of M^{-1} is just Δ^{-1} (where Δ^{-1} exists since M is positive-definite). Therefore the condition now becomes

$$Q^{-1} + (\Gamma^* + Q^{-1} S)(R - S^* Q^{-1} S)^{-1} (\Gamma + S^* Q^{-1}) > 0$$

so that we have the following result.

Lemma 8 (A Condition for a Minimum): If Q and $R - S^* Q^{-1} S$ are invertible, a necessary and sufficient condition for the stationary point of Lemma 7 to be a minimum is that

i) $P_{0|N} > 0$.

ii) $Q^{-1} + (\Gamma^* + Q^{-1} S)(R - S^* Q^{-1} S)^{-1} (\Gamma + S^* Q^{-1}) > 0$.

When $S \equiv 0$, the second condition becomes $Q^{-1} + \Gamma^* R^{-1} \Gamma > 0$.

The conditions of Lemma 8 need to be reduced further to provide useful computational tests. This can be done in several ways, leading to more specific tests. One interesting way is by showing that $Q^{-1} + (\Gamma^* + Q^{-1} S)(R - S^* Q^{-1} S)^{-1} (\Gamma + S^* Q^{-1})$ may be regarded as the Gramian matrix of the output of a so-called backward dual state-space model. This identification will be useful in studying the H^∞-control problem (and in other ways), but we shall not pursue it here.

Instead we shall use the alternative inertia conditions of Lemma 5 to circumvent the need for direct analysis of the matrix $R_z - R_{zy} R_y^{-1} R_{yz}$. Recall from Lemma 5 that if $R_z > 0$, a unique minimizing solution to the deterministic problem of Theorem 2 exists if, and only if, R_y and $R_y - R_{yz} R_z^{-1} R_{zy}$ have the same inertia. For the state-space structure that we are considering, however

$$R_z = \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q \end{bmatrix}$$

so that after some simple algebra we have

$$\begin{aligned} R_y &- R_{yz} R_z^{-1} R_{zy} \\ &= R - S^* Q^{-1} S \\ &= (R_0 - S_0^* Q_0^{-1} S_0) \oplus \cdots \oplus (R_N - S_N^* Q_N^{-1} S_N). \quad (33) \end{aligned}$$

Thus $R_y - R_{yz} R_z^{-1} R_{zy}$ is block-diagonal, and we have the following result.

Lemma 9 (Inertia Condition for Minimum): If $\Pi_0 > 0$ and $Q > 0$, then a necessary and sufficient condition for the stationary point of Lemma 7 to be a minimum is that the matrices R_y and $R - S^* Q^{-1} S$ have the same inertia. In particular, if $S \equiv 0$, then R_y and R must have the same inertia.

As we shall see in the next section, the Krein space–Kalman filter provides the block triangular factorization of R_y, and thereby allows one to easily compare the inertia of R_y and $R - S^* Q^{-1} S$.

$$\begin{aligned} M &= \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q \end{bmatrix} - \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \mathcal{O}^* \\ \Gamma^* + Q^{-1} S \end{bmatrix} \times R_y^{-1} [\mathcal{O} \quad \Gamma + S^* Q^{-1}] \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q \end{bmatrix} \\ &= \begin{bmatrix} \Pi_0 - \Pi_0 \mathcal{O}^* R_y^{-1} \mathcal{O} \Pi_0 & -\Pi_0 \mathcal{O}^* R_y^{-1} (\Gamma Q + S^*) \\ -(Q\Gamma^* + S) R_y^{-1} \mathcal{O} \Pi_0 & Q - (Q\Gamma^* + S) R_y^{-1} (\Gamma Q + S^*) \end{bmatrix}. \end{aligned} \quad (30)$$

VI. Recursive Formulas

So far we have obtained global expressions for computing projections and for checking the conditions for deterministic and stochastic minimization. Computing the projection requires inverting the Gramian matrix R_y and checking for the minimization conditions requires checking the inertia of R_y, both of which require $O(N^3)$ (where N is the dimension of R_y) computations.

The key consequence of state-space structure in Hilbert space is that the computational burden of finding projections can be significantly reduced, to $O(Nn^3)$ (where n is the dimension of the state-space model), by using the Kalman filter recursions. Moreover, the Kalman filter also recursively factors the positive definite Gramian matrix R_y as LDL^*, L lower triangular with unit diagonal, and D diagonal.

We shall presently see that similar recursions hold in Krein space as well, provided

$$R_y \text{ is strongly nonsingular (or strongly regular)} \tag{34}$$

in the sense that all its (block) leading minors are nonzero. Recall that in Hilbert space if the $\{\boldsymbol{y}_i\}$ are linearly independent, then R_y is strictly positive definite; so that (34) holds automatically. In the Krein-space theory, we have so far only assumed that R_y is invertible which does not necessarily imply (34). Recursive projection, i.e., projection onto $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_i\}$ for all i, however, requires that all the (block) leading submatrices of R_y are nonsingular; recall also that (34) implies that R_y has a unique triangular decomposition

$$R_y = LDL^*. \tag{35}$$

Therefore, $\mathrm{In}(R_y) = \mathrm{In}(D)$, and in particular, $R_y > 0$ iff $D > 0$. This is the standard way of recursively computing the inertia of R_y.

The standard method of recursive estimation, which also gives a very useful geometric insight into the triangular factorization of R_y, is to introduce the innovations

$$\boldsymbol{e}_j = \boldsymbol{y}_j - \hat{\boldsymbol{y}}_j, \quad 0 \le j \le N \tag{36}$$

where $\hat{\boldsymbol{y}}_j \triangleq \hat{\boldsymbol{y}}_{j|j-1}$ = the projection of $\boldsymbol{y}_j$ onto $\mathcal{L}\ \{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{j-1}\}$.

Note that due to the construction (36), the innovations form an orthogonal basis for $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ (with respect to the Krein-space inner product) which simplifies the calculation of projections. For example, we can express the projection of the fundamental quantities $\boldsymbol{x}_0$ and $\boldsymbol{u}_j$ onto $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$ as

$$\hat{\boldsymbol{x}}_{0|N} = \sum_{i=0}^{N} \langle \boldsymbol{x}_0, \boldsymbol{e}_i \rangle \langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle^{-1} \boldsymbol{e}_i \tag{37}$$

and

$$\hat{\boldsymbol{u}}_{j|N} = \sum_{i=0}^{N} \langle \boldsymbol{u}_j, \boldsymbol{e}_i \rangle \langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle^{-1} \boldsymbol{e}_i \tag{38}$$

where the state-space structure may be used to calculate the above inner products recursively.

Before proceeding to show this, however, let us note that any method for computing the innovations yields the triangular factorization of the Gramian R_y. To this end, let us write

$$\begin{aligned}\boldsymbol{y}_i &= \hat{\boldsymbol{y}}_i + \boldsymbol{e}_i \\ &= \langle \boldsymbol{y}_i, \boldsymbol{e}_0 \rangle R_{e,0}^{-1} \boldsymbol{e}_0 + \cdots + \langle \boldsymbol{y}_i, \boldsymbol{e}_{i-1} \rangle R_{e,i-1}^{-1} \boldsymbol{e}_{i-1} + \boldsymbol{e}_i\end{aligned}$$

and collect such expressions in matrix form

$$\begin{aligned}\boldsymbol{y} &= \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} \\ &= \begin{bmatrix} I & & & \\ \langle \boldsymbol{y}_1, \boldsymbol{e}_0 \rangle R_{e,0}^{-1} & I & & \\ \vdots & & \ddots & \\ \langle \boldsymbol{y}_N, \boldsymbol{e}_0 \rangle R_{e,0}^{-1} & \langle \boldsymbol{y}_N, \boldsymbol{e}_1 \rangle R_{e,1}^{-1} & \cdots & I \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_0 \\ \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_N \end{bmatrix} = L\boldsymbol{e}\end{aligned}$$

where L is lower triangular with unit diagonal. Therefore, since the $\boldsymbol{e}_i$ are orthogonal, the Gramian of $\boldsymbol{y}$ is

$$R_y = LR_eL^*, \quad \text{where} \quad R_e = R_{e,0} \oplus R_{e,1} \oplus \cdots \oplus R_{e,N}.$$

We thus have the following result.

Lemma 10 (Inertia of R_y): The Gramian R_y of $\boldsymbol{y}$ has the same inertia as the Gramian of the innovations, R_e. The strong regularity of R_y implies the nonsingularity of $R_{e,i}$, $0 \le i \le N$. In particular, $R_y > 0$, if and only if

$$R_{e,i} > 0, \quad \text{for all} \quad i = 0, 1, \cdots, N.$$

We should also point out that the value at the stationary point of the quadratic form in Theorem 2 can also be expressed in terms of the innovations

$$\begin{aligned}J(z_0, y) &= y^* R_y^{-1} y = y^* L^{-*} R_e^{-1} L^{-1} y \\ &= e^* R_e^{-1} e = \sum_{j=0}^{N} e_j^* R_{e,j}^{-1} e_j.\end{aligned} \tag{39}$$

A. The Krein Space–Kalman Filter

Now we shall show that the state-space structure allows us to efficiently compute the innovations by an immediate extension of the Kalman filter.

Theorem 3 (Kalman Filter in Krein Space): Consider the Krein-space state equations

$$\begin{cases} \boldsymbol{x}_{i+1} = F_i \boldsymbol{x}_i + G_i \boldsymbol{u}_i, & 0 \le i \le N \\ \boldsymbol{y}_i = H_i \boldsymbol{x}_i + \boldsymbol{v}_i \end{cases} \tag{40}$$

with

$$\left\langle \begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{v}_j \\ \boldsymbol{x}_0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{v}_k \\ \boldsymbol{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix} \delta_{jk} & 0 \\ 0 & \Pi_0 \end{bmatrix}.$$

Assume that $R_y = [\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle]$ is strongly regular. Then the innovations can be computed via the formulas

$$\boldsymbol{e}_i = \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_i, \quad 0 \le i \le N \tag{41}$$

$$\hat{\boldsymbol{x}}_{i+1} = F_i \hat{\boldsymbol{x}}_i + K_{p,i}(\boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_i), \quad \hat{\boldsymbol{x}}_0 = 0 \tag{42}$$

$$K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1} \tag{43}$$

where

$$R_{e,i} = \langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle = R_i + H_i P_i H_i^*$$

and the $\{P_i\}$ can be recursively computed via the Riccati recursion

$$P_{i+1} = F_i P_i F_i^* - K_{p,i} R_{e,i} K_{p,i}^* + G_i Q_i G_i^*, \quad P_0 = \Pi_0. \quad (44)$$

The number of computations is dominated by those in (44) and is readily seen to be $O(n^3)$ per iteration.

Remark: The only difference from the conventional Kalman filter expressions is that the matrices P_i and $R_{e,i}$ (and, by assumption, Π_0, Q_i and R_i) may now be indefinite.

Proof: The same as in the usual Kalman filter theory (see, e.g., [13]). For completeness and to show the power of the geometric viewpoint, however, we present a simple derivation. There is absolutely no formal difference between the steps in the (usual) Hilbert space case and in the Krein-space case.

Begin by noting that

$$\begin{aligned} \boldsymbol{e}_i = \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i &= \boldsymbol{y}_i - (H_i \hat{\boldsymbol{x}}_i + \hat{\boldsymbol{v}}_i) \\ &= \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_i = H_i \tilde{\boldsymbol{x}}_i + \boldsymbol{v}_i \end{aligned} \quad (45)$$

where $\hat{\boldsymbol{x}}_i$ is the projection of $\boldsymbol{x}_i$ on $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{i-1}\}$ and where we have defined $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i$. It follows readily that

$$R_{e,i} = \langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle = R_i + H_i P_i H_i^*, \quad P_i \triangleq \langle \tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_i \rangle. \quad (46)$$

Recall (see Lemma 10) that the strong nonsingularity (all leading minors nonzero) of R_y implies that the $\{R_{e,i}\}$ are nonsingular (rather than positive-definite, as in the Hilbert space case). The Kalman filter can now be readily derived by using the orthogonality of the innovations and the state-space structure. Thus we first write

$$\hat{\boldsymbol{x}}_{i+1|i} = \hat{\boldsymbol{x}}_{i+1} = \sum_{j=0}^{i} \langle \boldsymbol{x}_{i+1}, \boldsymbol{e}_j \rangle \langle \boldsymbol{e}_j, \boldsymbol{e}_j \rangle_K^{-1} \boldsymbol{e}_j$$

and to seek a recursion we decompose the above as

$$\begin{aligned} \hat{\boldsymbol{x}}_{i+1} &= \sum_{j=0}^{i-1} \langle \boldsymbol{x}_{i+1}, \boldsymbol{e}_j \rangle R_{e,j}^{-1} \boldsymbol{e}_j + K_{p,i} \boldsymbol{e}_i \\ K_{p,i} &\triangleq \langle \boldsymbol{x}_{i+1}, \boldsymbol{e}_i \rangle R_{e,i}^{-1}. \end{aligned}$$

Now

$$\begin{aligned} \langle \boldsymbol{x}_{i+1}, \boldsymbol{e}_i \rangle &= F_i \langle \boldsymbol{x}_i, \boldsymbol{e}_i \rangle + G_i \langle \boldsymbol{u}_i, \boldsymbol{e}_i \rangle \\ &= F_i \langle \boldsymbol{x}_i, H_i \tilde{\boldsymbol{x}}_i + \boldsymbol{v}_i \rangle + G_i \langle \boldsymbol{u}_i, H_i \tilde{\boldsymbol{x}}_i + \boldsymbol{v}_i \rangle \\ &= F_i \langle \tilde{\boldsymbol{x}}_i, H_i \tilde{\boldsymbol{x}}_i \rangle + 0 + 0 + G_i \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \\ &= F_i P_i H_i^* + G_i S_i. \end{aligned}$$

Note also that the first summation can be rewritten as

$$F_i \sum_{j=0}^{i-1} \langle \boldsymbol{x}_i, \boldsymbol{e}_j \rangle R_{e,j}^{-1} \boldsymbol{e}_j + G_i \sum_{j=0}^{i-1} \langle \boldsymbol{u}_i, \boldsymbol{e}_j \rangle R_{e,j}^{-1} \boldsymbol{e}_j = F_i \hat{\boldsymbol{x}}_i + 0.$$

Combining these facts we find

$$\hat{\boldsymbol{x}}_{i+1} = F_i \hat{\boldsymbol{x}}_i + K_{p,i} \boldsymbol{e}_i \quad (47)$$

and

$$K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}. \quad (48)$$

It now remains to find a recursion for P_i. To this end, note that if we define the Gramians $\Pi_i = \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle$ and $\Sigma_i = \langle \hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_i \rangle$, then the orthogonality of the $\hat{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}_i$ yields

$$P_i = \Pi_i - \Sigma_i.$$

The state-space equations (22a) show that the state variance Π_i, obeys the recursion

$$\Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i^* G_i^*.$$

Likewise, the orthogonality of the innovations implies that (47) will yield

$$\Sigma_{i+1} = F_i \Sigma_i F_i^* + K_{p,i} R_{e,i} K_{p,i}^*, \quad \Sigma_0 = 0.$$

Subtracting the above two equations yields the desired Riccati recursion for P_i,

$$\begin{aligned} P_{i+1} &= F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^* \\ P_0 &= \Pi_0. \end{aligned} \quad (49)$$

Equations (46)–(49) constitute the Kalman filter of Theorem 3. □

In Kalman filter theory there are many variations of the above formulas and we note one here. Let us define the filtered estimate, $\hat{\boldsymbol{x}}_{i|i}$ = the projection of $\boldsymbol{x}_i$ onto $\mathcal{L}\{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_i\}$.

Theorem 4 (Measurement and Time Updates): Consider the Krein state-space equations of Theorem 3 and assume that R_y is strongly regular. Then when $S_i \equiv 0$, the filtered estimates $\hat{\boldsymbol{x}}_{i|i}$ can be computed via the following (measurement and time update) formulas

$$\begin{aligned} \hat{\boldsymbol{x}}_{i+1|i+1} &= \hat{\boldsymbol{x}}_{i+1} + K_{f,i+1} \boldsymbol{e}_{i+1}, \quad \hat{\boldsymbol{x}}_0 = 0 \\ K_{f,i+1} &= P_{i+1} H_{i+1}^* R_{e,i+1}^{-1} \quad (50) \\ \hat{\boldsymbol{x}}_{i+1} &= F_i \hat{\boldsymbol{x}}_{i|i} \quad (51) \end{aligned}$$

where $\boldsymbol{e}_i$, $R_{e,i}$, and P_i are as in Theorem 3.

Corollary 4 (Filtered Recursions): The two step recursions of Theorem 4 can be combined into the single recursion

$$\hat{\boldsymbol{x}}_{i+1|i+1} = F_i \hat{\boldsymbol{x}}_{i|i} + K_{f,i+1}(\boldsymbol{y}_{i+1} - H_{i+1} F_i \hat{\boldsymbol{x}}_{i|i}), \quad \hat{\boldsymbol{x}}_{-1|-1} = 0. \quad (52)$$

For numerical reasons, certain square-root versions of the KF are now more often used in state-space estimation. Furthermore, for constant systems or in fact for systems where the time-variation is structured in a certain way, the Riccati recursions and the square-root recursions, both of which take $O(n^3)$ elementary computations (flops) per iteration, can be replaced by the more efficient Chandrasekhar recursions which require only $O(n^2)$ flops per iteration [17], [18]. The square-root and Chandrasekhar recursions can both be extended to the Krein-space setting, as described in [22].

Before closing this section we shall note how the innovations computed in Theorem 3 can be used to determine the projections $\hat{\boldsymbol{x}}_{0|N}$ and $\hat{\boldsymbol{u}}_{|N}$ using the formulas (37) and (38).

Lemma 11 (Computation of Inner Products): We can write

$$\langle \boldsymbol{x}_0, \boldsymbol{e}_i \rangle = \Pi_0 \Phi_{F-KH}^*(i,0) H_i^* \tag{53}$$

and

$$\langle \boldsymbol{u}_j, \boldsymbol{e}_i \rangle = \begin{cases} Q_j G_j^* \Phi_{F-KH}^*(i,j+1) H_i^* + S_i \delta_{ij} & j \le i \\ 0 & j > i \end{cases} \tag{54}$$

where

$$\Phi_{F-KH}(i,j) \triangleq \prod_{k=j}^{i-1} (F_k - K_{p,k} H_k).$$

These lead to the recursions

$$\hat{\boldsymbol{x}}_{0|i} = \hat{\boldsymbol{x}}_{0|i-1} + \Pi_0 \Phi_{F-KH}^*(i,0) H_i^* R_{e,i}^{-1} \boldsymbol{e}_i, \quad \hat{\boldsymbol{x}}_{0|-1} = 0 \tag{55}$$

and (56), found at the bottom of the page, where $\Phi_{F-KH}^*(i,j)$ $(i \ge j)$ satisfies the recursion

$$\Phi_{F-KH}^*(i+1,j) = \Phi_{F-KH}^*(i,j)(F_i - K_{p,i} H_i)^*$$
$$\Phi_{F-KH}^*(j,j) = I.$$

Proof: Straightforward computation. □

B. Recursive State-Space Estimation and Quadratic Forms

Theorems 5 and 6 below are essentially restatements of Theorems 1 and 2 when a state space model is assumed and a recursive solution is sought.

The error Gramian associated with the problem of projecting $\{\boldsymbol{x}_0, \boldsymbol{u}\}$ onto $\mathcal{L}\{\boldsymbol{y}\}$ has already been identified in Lemma 6 and (55), and (56) furnishes a recursive procedure for calculating this projection. The condition for a minimum is $R_y > 0$, where R_y has been shown to be congruent to the diagonal matrix R_e. This gives the following theorem.

Theorem (Stochastic Problem): Suppose $\boldsymbol{z} = \operatorname{col}\{\boldsymbol{x}_0, \boldsymbol{u}\}$ and $\boldsymbol{y}$ are related through the state-space model (22a) and (22b) and that R_y is strongly regular. Then the state-space estimation algorithm (55), (56) recursively computes the stationary point of the error Gramian

$$\langle \boldsymbol{z} - k^* \boldsymbol{y}, \boldsymbol{z} - k^* \boldsymbol{y} \rangle$$

over all $k^* \boldsymbol{y}$. Moreover, this stationary point is a minimum if, and only if

$$R_{e,j} > 0 \quad \text{for} \quad j = 0, \cdots, i.$$

Similarly, the scalar quadratic form associated with the (partially) equivalent deterministic problem has already been identified in Lemma 7

$$J_N(x_0, u, y) = x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{N} [u_j^* \quad (y_j - H_j x_j)^*] \times \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix}. \tag{57}$$

In particular, $\hat{x}_{0|N}$ and $\hat{u}_{j|N}$ are the stationary points of $J_N(x_0, u, y)$ over x_0 and u_j and subject to the state-space constraints $x_{j+1} = F_j x_j + G_j u_j$, $j = 0, \cdots, N$. In the recursions, for each time i, we find $\hat{x}_{0|i}$ and $\hat{u}_{j|i}$ which are the stationary points of

$$J_i(x_0, u, y) = x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} [u_j^* \quad (y_j - H_j x_j)^*] \times \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix}. \tag{58}$$

Theorem 6 (Deterministic Problem): If R_y is strongly regular, the stationary point of the quadratic form

$$J_i(x_0, u, y) = x_0^* \Pi_0^{-1} x_0 + \sum_{j=0}^{i} [u_j^* \quad (y_j - H_j x_j)^*] \times \begin{bmatrix} Q_j & S_j \\ S_j^* & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix} \tag{59}$$

over x_0 and u_j, subject to the state-space constraints $x_{j+1} = F_j x_j + G_j u_j$, $j = 0, 1, \cdots, i$ can be recursively computed as

$$\hat{x}_{0|i} = \hat{x}_{0|i-1} + \Pi_0 \Phi_{F-KH}^*(i,0) H_i^* R_{e,i}^{-1} e_i, \quad \hat{x}_{0|-1} = 0$$

and see (x), shown at the bottom of the page, where the innovations e_j can be computed via the recursions

$$\hat{x}_{i+1} = F_i \hat{x}_i + K_{p,i} e_i, \quad \hat{x}_0 = 0$$

with $K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$, $R_{e,i} = R_i + H_i P_i H_i^*$, $e_i = y_i - H_i \hat{x}_i$, and P_i satisfying the Riccati recursion

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

Moreover, the value of $J_i(x_0, u, y)$ at the stationary point is given by

$$J_i(\hat{x}_{0|i}, \hat{u}_{|i}, y) = \sum_{j=0}^{i} e_j^* R_{e,j}^{-1} e_j.$$

$$\hat{\boldsymbol{u}}_{j|i} = \begin{cases} \hat{\boldsymbol{u}}_{j|i-1} + Q_j G_j^* \Phi_{F-KH}^*(i,j+1) H_i^* R_{e,i}^{-1} \boldsymbol{e}_i, & \hat{\boldsymbol{u}}_{j|j} = S_j R_{e,j}^{-1} \boldsymbol{e}_i \quad j \le i \\ 0 & j > i \end{cases} \tag{56}$$

$$\hat{u}_{j|i} = \begin{cases} \hat{u}_{j|i-1} + Q_j G_j^* \Phi_{F-KH}^*(i,j+1) H_i^* R_{e,i}^{-1} e_i, & \hat{u}_{j|j} = S_j R_{e,j}^{-1} e_i \quad j \le i \\ 0 & j > i \end{cases} \tag{x}$$

Proof: The proof follows from the basic equivalence between the deterministic and stochastic problems. The recursions for $\hat{x}_{0|i}$ and $\hat{u}_{j|i}$ are the same as those in the stochastic problem of Lemma 11, and the innovations e_i are found via the Krein space–Kalman filter of Theorem 3. □

As mentioned earlier, the deterministic quadratic form of Theorem 6 is often encountered in estimation problems. By appeal to Gaussian assumptions on the v_i, u_i, and x_0, and maximum likelihood arguments, it is well known that state estimates can be obtained via a deterministic quadratic minimization problem. Here we have shown this result using simple projection arguments and have generalized it to indefinite quadratic forms.

The result of Theorem 6 is probably the most important result of this paper, and we shall make frequent use of it in the companion paper [1] to solve the problems of H^∞ and risk-sensitive estimation and finite-memory adaptive filtering. In those problems we shall also need to recursively check for the condition for a minimum, and therefore we will now study these conditions in more detail.

Recall from Lemma 9 that the above deterministic problem has a minimum iff, R_y and $R-S^*Q^{-1}S$ have the same inertia. Since R_y is congruent to the block diagonal matrix R_e, and since $R-S^*Q^{-1}S$ is also block diagonal, the solution of the recursive stationarization problem will give a minimum at each step if and only if all the block diagonal elements of R_e and $R-S^*Q^{-1}S$ have the same inertia. This leads to the following result.

Lemma 12 (Inertia Conditions for a Minimum): If $\Pi_0 > 0$, $Q > 0$, and R is nonsingular, then the (unique) stationary points of the quadratic forms (59), for $i = 0, 1, \cdots N$, will each be a unique minimum iff the matrices

$$R_{e,j} \quad \text{and} \quad R_j - S_j^* Q_j^{-1} S_j$$

have the same inertia for all $j = 0, 1, \cdots N$. In particular, when $S_j \equiv 0$, the condition becomes that $R_{e,j}$, and R_j should have the same inertia for all $j = 0, 1, \cdots N$.

The conditions of the above Lemma are easy to check since the Krein space–Kalman filter used to compute the stationary point also computes the matrices $R_{e,j}$. There is another condition, more frequently quoted in the H^∞ literature, which we restate here (see, e.g., [4]).

Lemma 13 (Condition for a Minimum): If $\Pi_0 > 0$, $Q > 0$, R is invertible, $Q - SR^{-1}S^* > 0$, and $[F_j \; G_j]$ has full rank for all j, then the quadratic forms (59) will each have a unique minimum if, and only if

$$P_{j|j}^{-1} = P_j^{-1} + H_j^* R_j^{-1} H_j > 0 \quad j = 0, 1, \cdots, N.$$

It also follows in the minimum case that $P_{j+1} > 0$ for $j = 0, 1, \cdots, N$.

Remark: In comparison to our result in Lemma 12, we here have the additional requirement that the $[F_j \quad G_j]$ must be full rank. Furthermore, we not only have to compute the P_j (which is done via the Riccati recursion of the Kalman filter), but we also have to invert P_j (and R_j) at each step and then check for the positivity of $P_j^{-1} + H_j^* R_j^{-1} H_j$. The test of Lemma 12 uses only quantities already present in the Kalman filter recursion, viz. $R_{e,j}$ and R_j. Moreover, these are $p \times p$ matrices (as opposed to $P_{j|j}^{-1}$ which is $n \times n$) with p typically less than n and whose inertia is easily determined via a triangular factorization. Furthermore it can be shown [22] that even this computation can be effectively blended into the filter recursions by going to a square-root-array version of the Riccati recursion. Here, however, for completeness we shall show how Lemma 13 follows from our Lemma 12.

Proof of Lemma 13: We shall prove the lemma by induction. Consider the matrix

$$\begin{bmatrix} -\Pi_0^{-1} & 0 & H_0^* \\ 0 & -Q_0^{-1} & Q_0^{-1} S_0 \\ H_0 & S_0^* Q_0^{-1} & R_0 - S_0^* Q_0^{-1} S_0 \end{bmatrix}.$$

Two different triangular factorizations (lower-upper and upper-lower) of the above matrix show that

$$\begin{bmatrix} -\Pi_0^{-1} & 0 & 0 \\ 0 & -Q_0^{-1} & 0 \\ 0 & 0 & R_0 + H_0 \Pi_0 H_0^* \end{bmatrix}$$

and (y), shown at the bottom of the page, have the same inertia. Thus, since $\Pi_0 > 0$, $Q_0 > 0$, and $Q_0 - S_0 R_0^{-1} S_0^* > 0$, then the matrices $R_{e,0} = R_0 + H_0 \Pi_0 H_0^*$ and $R_0 - S_0^* Q_0^{-1} S_0$ will have the same inertia (and we will have a minimum for J_0) iff

$$\Pi_0^{-1} + H_0^* R_0^{-1} H_0 > 0.$$

Now with some effort we may write the first step of the Riccati recursion as

$$P_1 = [F_0 \quad G_0] \left(\begin{bmatrix} \Pi_0^{-1} & 0 \\ 0 & Q_0^{-1} \end{bmatrix} + \begin{bmatrix} H_0^* \\ Q_0^{-1} S_0 \end{bmatrix} \right.$$
$$\left. \times (R_0^{-1} - S_0^* Q_0^{-1} S_0)^{-1} [H_0 \quad S_0^* Q_0^{-1}] \right)^{-1} \begin{bmatrix} F_0^* \\ G_0^* \end{bmatrix}.$$

Moreover, the center matrix appearing in the above expression is congruent to

$$\begin{bmatrix} \Pi_0^{-1} + H_0^* R_0^{-1} H_0 & 0 \\ 0 & (Q_0 - S_0 R_0^{-1} S_0^*)^{-1} \end{bmatrix}$$

and hence is positive definite. Thus if $[F_0 \; G_0]$ has full rank, we can conclude that $P_1 > 0$. We can now repeat the argument for the next time instant and so on. □

We close this section with yet another condition which will be useful in control problems.

$$\begin{bmatrix} -(\Pi_0^{-1} + H_0^* R_0^{-1} H_0) & 0 & 0 \\ 0 & -(Q_0 - S_0 R_0^{-1} S_0^*)^{-1} & 0 \\ 0 & 0 & R_0 - S_0^* Q_0^{-1} S_0 \end{bmatrix} \quad \text{(y)}$$

$$\begin{aligned}
&\begin{bmatrix} P_{i+1} & 0 \\ 0 & (Q_i - S_iR_i^{-1}S_i^*)^{-1} - G_i^*P_{i+1}^{-1}G_i \end{bmatrix} \\
&\sim \begin{bmatrix} P_{i+1} & G_i \\ G_i^* & (Q_i - S_iR_i^{-1}S_i^*)^{-1} \end{bmatrix} \\
&\sim \begin{bmatrix} P_{i+1} - G_i(Q_i - S_iR_i^{-1}S_i^*)G_i^* & 0 \\ 0 & (Q_i - S_iR_i^{-1}S_i^*)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (F_i - G_iS_iR_i^{-1}H_i)(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}(F_i - G_iS_iR_i^{-1}H_i)^* & 0 \\ 0 & (Q_i - S_iR_i^{-1}S_i^*)^{-1} \end{bmatrix} \\
&\sim \begin{bmatrix} (P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1} & 0 \\ 0 & (Q_i - S_iR_i^{-1}S_i^*)^{-1} \end{bmatrix}.
\end{aligned} \tag{z}$$

Lemma 14 (Condition for a Minimum): If in addition to the conditions of Lemma 13, the matrices $F_j - G_jS_jR_j^{-1}H_j$ are invertible for all j, then the deterministic problems of Theorem 6 will each have a unique minimum iff $P_{N+1} > 0$ and

$$(Q_j - S_jR_j^{-1}S_j^*)^{-1} - G_j^*P_{j+1}G_j > 0 \quad j = 0, 1, \cdots, N.$$

Proof: Let us first note that the Riccati recursion can be rewritten as

$$\begin{aligned}
P_{i+1} &= F_iP_iF_i^* + G_iQ_iG_i^* - (F_iP_iH_i^* + G_iS_i) \\
&\quad \times (R_i + H_iP_iH_i^*)^{-1}(F_iP_iH_i^* + G_iS_i)^* \\
&= (F_i - G_iS_iR_i^{-1}H_i)P_i(F_i - G_iS_iR_i^{-1}H_i)^* \\
&\quad + G_i(Q_i - S_iR_i^{-1}S_i^*)G_i^* - (F_i - G_iS_iR_i^{-1}H_i)P_iH_i^* \\
&\quad \times (R_i + H_iP_iH_i^*)^{-1}H_iP_i(F_i - G_iS_iR_i^{-1}H_i)^* \\
&= (F_i - G_iS_iR_i^{-1}H_i)(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1} \\
&\quad \times (F_i - G_iS_iR_i^{-1}H_i)^* + G_i(Q_i - S_iR_i^{-1}S_i^*)G_i^*.
\end{aligned}$$

The proof, which uses the last of the above equalities, now follows from the sequence of congruences, found in (z) at the top of the page, and Lemma 13. □

VII. Concluding Remarks

We developed a self-contained theory for linear estimation in Krein spaces. We started with the notion of projections and discussed their relation to stationary points of certain quadratic forms encountered in a pair of partially equivalent stochastic and deterministic problems. By assuming an additional state-space structure, we showed that projections could be recursively computed by a Krein space–Kalman filter, several applications for which are described in the companion paper [1].

The approach, in all these applications, is that given an indefinite deterministic quadratic form to which H^∞, risk-sensitive, and finite-memory problems lead almost by inspection, one can relate them to a corresponding Krein-space stochastic problem for which the Kalman filter can be written down immediately and used to obtain recursive solutions of the above problems.

Acknowledgment

The authors would like to thank P. P. Khargonekar and D. J. N. Limebeer for helpful discussions during the preparation of this manuscript. Seminars by P. Park on the KYP Lemma were also helpful in leading us to begin the research.

References

[1] B. Hassibi, A. H. Sayed, and T. Kailath, "Linear estimation in Krein spaces—Part II: Applications," this issue, pp. 34–49.

[2] P. P. Khargonekar and K. M. Nagpal, "Filtering and smoothing in an H^∞-setting," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 151–166, 1991.

[3] M. J. Grimble, "Polynomial matrix solution of the H^∞-filtering problem and the relationship to Riccati equation state-space results," *IEEE Trans. Signal Processing*, vol. 41, no. 1, pp. 67–81, Jan. 1993.

[4] U. Shaked and Y. Theodor, "H^∞-optimal estimation: A tutorial," in *Proc. IEEE Conf. Decision Contr.*, Tucson, AZ, Dec. 1992, pp. 2278–2286.

[5] T. Basar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems—A Dynamic Game Approach.* Boston, MA: Birkhauser, 1991.

[6] D. Limebeer, B. D. O. Anderson, P. P. Khargonekar, and M. Green, "A game theoretic approach to H^∞ control for time varying systems," *SIAM J. Contr. Optimization*, vol. 30, pp. 262–283, 1992.

[7] G. Tadmor, "H^∞ in the time domain: The standard problem," in *Am. Contr. Conf.*, 1989, pp. 772–773.

[8] P. Whittle, *Risk Sensitive Optimal Control.* New York: Wiley, 1990.

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

[10] V. I. Istratescu, *Inner Product Structures, Theory and Applications, Mathematics and Its Applications.* Dordrecht, Holland: Reidel, 1987.

[11] I. S. Iohvidov, M. G. Krein, and H. Langer, "Introduction to the spectral theory of operators in spaces with an indefinite metric," in *Mathematical Research.* Berlin, Germany: Akademie-Verlag, 1982.

[12] A. Einstein, *Relativity: The Special and General Theory*, transl. by R. W. Lawson. New York: Crown, 1931.

[13] T. Kailath, *Lectures on Wiener and Kalman Filtering.* Berlin, Germany: Springer-Verlag, 1981.

[14] M. Green and D. J. N. Limebeer, *Linear Robust Control.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

[15] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, pp. 18–60, July 1994.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: Johns Hopkins Univ., 1989.

[17] M. Morf, G. S. Sidhu, and T. Kailath, "Some new algorithms for recursive estimation in constant, linear, discrete-time systems," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 315–323, 1974.

[18] A. H. Sayed and T. Kailath, "Extended Chandrasekhar recursions," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 2265–2269, Nov. 1994.

[19] A. E. Bryson and Y. C. Ho, *Applied Optimal Control.* Blaisdell, 1969.

[20] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* New York: Academic, 1970.

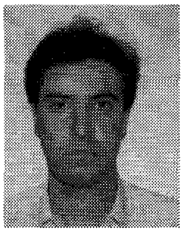
[21] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[22] B. Hassibi, A. H. Sayed, and T. Kailath, "Square-root arrays and Chandrasekhar recursions for H^∞ problems," submitted to *IEEE Trans.*

Automat. Contr., also in *Proc. 33rd IEEE Conf. Dec. Cont.*, 1994, pp. 2237–2243.

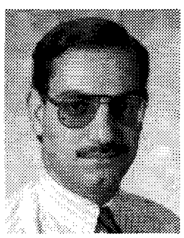
[23] ——, "H^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Processing*, to appear. Also in the *Proc. 32rd IEEE Conf. Dec. and Cont.*, 1993, pp. 74–80.

[24] ——, "Fundamental inertia conditions for the solution of H^∞ problems," in *Proc. ACC*, June 1995.



Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. degree from Stanford University, Stanford, CA, in 1993, both in electrical engineering. He is currently pursuing the Ph.D. degree at Stanford University.

From June 1992 to September 1992 he was a Summer Intern with Ricoh, California Research Center, Menlo Park, CA, and from August 1994 to December 1994 he was a short-term Research Fellow at the Indian Institute of Science, Bangalore, India. His research interests include robust estimation and control, adaptive signal processing and neural networks, blind equalization of communication channels, and linear algebra.



Ali H. Sayed (S'90–M'92) was born in São Paulo, Brazil, in 1963. He received the B.S. and M.S. degrees in electrical engineering from the University of São Paulo, in 1987 and 1989, respectively. In 1992 he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

From September 1992 to August 1993, he was a Research Associate with the Information Systems Laboratory at Stanford University, after which he joined the Department of Electrical and Computer Engineering at the University of California, Santa Barbara, as an Assistant Professor. His research interests include adaptive and statistical signal processing, robust filtering and control, interplays between signal processing and control methodologies, interpolation theory, and structured computations in systems and mathematics.

Dr. Sayed is a recipient of the Institute of Engineering Prize, 1987 (Brazil), the Conde Armando Alvares Penteado Prize, 1987 (Brazil), and a 1994 NSF Research Initiation Award.

Thomas Kailath (S'57–M'62–F'70) received the S.M. degree in 1959 and the Sc.D degree in 1961 from the Massachusetts Institute of Technology.

From October 1961 to December 1962, he worked at the Jet Propulsion Laboratories, Pasadena, CA, where he also taught part-time at the California Institute of Technology. He then went to Stanford University, where he served as Director of the Information Systems Laboratory from 1971 through 1980, as Associate Department Chairman from 1981 to 1987, and currently holds the Hitachi America Professorship in Engineering. He has held short-term appointments at several institutions around the world. His recent research interests include applications of signal processing, computation and control to problems in semiconductor manufacturing, and wireless communications. He is the author of *Linear Systems* (Englewood Cliffs, NJ: Prentice Hall, 1980) and *Lectures on Wiener and Kalman Filtering* (New York: Springer-Verlag, 1981).

Dr. Kailath is a fellow of the Institute of Mathematical Statistics and is a member of the National Academy of Engineering and the American Academy of Arts and Sciences. He has held Guggenheim, Churchill, and Royal Society fellowships, among others, and received awards from the IEEE Information Theory Society and the American Control Council, in addition to the Technical Achievement and Society Awards of the IEEE Signal Processing Society. He served as President of the IEEE Information Theory Society in 1975.