



## GEOMETRY OF KALMAN FILTERS

MACIEJ P. WOJTKOWSKI

Communicated by Martin Schlichenmaier

---

**Abstract.** We present a geometric explanation of Kalman filters in terms of a symplectic linear space and a special quadratic form on it. It is an extension of the work of Bougerol with application of a different metric we have introduced earlier. The new results are contained in Theorem 1 and Theorem 4.

---

### 1. Kalman Filters and Two Contractions

Let us consider a linear dynamical system with noise. In the simplest case of discrete time we have a linear map  $F : \mathbb{R}^k \rightarrow \mathbb{R}^k$  with additive noise

$$X_n = FX_{n-1} + G\varepsilon_n, \quad n \geq 1$$

where  $X_n$  is the state of the system at time  $n \geq 0$  and the noise  $\{\varepsilon_n\}, n \geq 1$ , is a sequence of independent normalized gaussian vectors in  $\mathbb{R}^l$  and  $G$  is a  $k \times l$  matrix, with  $l \leq k$ . The initial state  $X_0$  has also a gaussian distribution with mean  $\widehat{X}_0$  and covariance matrix  $P_0$ .

We take measurements on the system and the result at time  $n$  is

$$Y_n = HX_n + \eta_n, \quad n \geq 1$$

where  $H$  is a  $m \times k$  matrix and  $\{\eta_n\}$  is the observational noise which is assumed again to be a sequence of independent normalized gaussian vectors. The dimension of the measurement vector ( $m$ ) is in general smaller than the dimension of the state vector ( $k$ ), e.g.,  $m = 1$ .

The problem addressed by the Kalman filter is to estimate  $X_n$  given the values of  $Y_1, \dots, Y_n$ . The ideal answer is the estimator

$$\widehat{X}_n = \mathbb{E}(X_n | \mathcal{F}_n)$$

where  $\mathcal{F}_n$  is the sigma-algebra generated by the random vectors  $Y_1, \dots, Y_n$ . The Kalman filter amounts to the remarkable recursive relation

$$\widehat{X}_n = Z_n \widehat{X}_{n-1} + T_n Y_n \quad (1)$$

where

$$Z_n = G - P_n H^* H G, \quad T_n = P_n H^*$$

and  $P_n$  is the conditional error covariance matrix

$$P_n = \mathbb{E}((X_n - \widehat{X}_n)(X_n - \widehat{X}_n)^* | \mathcal{F}_n).$$

Postponing for a moment the discussion of  $P_n$  let us note that it is unusual for  $\widehat{X}_1$  to be a good estimate of  $X_1$  since little information is available. On the other hand after some time the quality of the estimate  $\widehat{X}_n$  is much better. It follows that in the successive applications of (1) the memory of  $\widehat{X}_0$  should disappear. Hence the product

$$Z^n = Z_n \dots Z_2 Z_1$$

needs to decay, without much assumptions and sufficiently fast. We will give a purely geometric account of this phenomenon, which essentially follows Bougerol's paper [1].

The Kalman filter is completed by the recursive, and deterministic, relation for  $P_n$

$$P_n = \Phi(P_{n-1}) \quad (2)$$

where

$$\Phi(P) = (C + DP)(A + BP)^{-1} \quad (3)$$

and the  $k \times k$  matrices  $A, B, C, D$  are the blocks of the linear map  $S : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k \times \mathbb{R}^k$

$$\begin{aligned} S &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} (I + H^* H)F^{*-1} & H^* H F \\ G G^* F^{*-1} & F \end{bmatrix} \\ &= \begin{bmatrix} I & H^* H \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ G G^* & I \end{bmatrix} \begin{bmatrix} F^{*-1} & 0 \\ 0 & F \end{bmatrix}. \end{aligned} \quad (4)$$

Maps of the form (3) are given different names: *matrix linear fractional transformation* (mlft), matrix Möbius map, or discrete Riccati equation. We will use the first term. The mlft (3) can be understood as the action of the linear map  $S$  on the grassmannian (the manifold of linear subspaces of fixed dimension in a vector space). In our case it is the action of the linear map on subspaces of the form  $gP := \{(x, Px); x \in \mathbb{R}^k\}$  ( $gP$  is the graph of  $P$ ) where  $P$  is a  $k \times k$  matrix. The linear map  $S$  is quite special since  $\Phi$  has to take symmetric positive definite matrices into symmetric positive definite matrices.

**Proposition 1.** *Let a linear nonsingular map  $S : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k \times \mathbb{R}^k$*

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

*act on  $k$ -dimensional subspaces of the form  $\{(x, Px); x \in \mathbb{R}^k\}$  by the mlft*

$$\Phi(P) = (C + DP)(A + BP)^{-1}.$$

*If  $\Phi(P)$  is symmetric for any  $P$  from an open set of symmetric matrices then  $A^*C$  and  $D^*B$  are symmetric and there must be a nonzero constant  $\alpha$  such that  $A^*D - C^*B = \alpha I$ .*

It follows that  $S$  is a multiple of a symplectic matrix. Namely let us consider the linear symplectic space  $\mathcal{W} = \mathbb{R}^k \times \mathbb{R}^k$  with the symplectic form  $\omega$

$$\omega(w, w') = \langle w_1, w'_2 \rangle - \langle w_2, w'_1 \rangle$$

where  $w = (w_1, w_2)$ ,  $w' = (w'_1, w'_2)$  and  $\langle \cdot, \cdot \rangle$  denotes the standard scalar product in  $\mathbb{R}^k$ . The claim in Proposition 1 is equivalent to

$$\omega(Sw, Sw') = \alpha\omega(w, w')$$

for all vectors  $w, w' \in \mathbb{R}^k \times \mathbb{R}^k$ .

In the language of the symplectic form the matrix  $P$  is symmetric if and only if  $\omega$  vanishes on the graph subspace  $gP$ , i.e.,  $gP$  is a Lagrangian subspace. Hence Proposition 1 can be reformulated to say that any nonsingular linear map  $S$  which takes Lagrangian subspaces into Lagrangian subspaces must be a multiple of a linear symplectic map. Multiplying  $S$  by a scalar does not change the respective mlft. The map  $S$  in (4) is actually symplectic, i.e.,  $\alpha = 1$ .

In the proof we will need the following

**Lemma 2.** *If for two  $k \times k$  matrices  $X, Y$  we have that  $XZ + ZY$  is symmetric for any symmetric matrix  $Z$ , then there is a constant  $\alpha$  such that  $X - Y^* = \alpha I$ .*

**Proof:** We have  $XZ + ZY = ZX^* + Y^*Z$  and hence  $(X - Y^*)Z = Z(X^* - Y)$  for any symmetric matrix  $Z$ . Taking  $Z = I$  we get that  $X - Y^* = X^* - Y$ . Further the only matrix that commutes with all symmetric matrices is a multiple of identity. ■

**Proof of Proposition 1:**

If  $\Phi(P)$  is symmetric then so is

$$\begin{aligned}\Psi(P) &= (A^* + PB^*)\Phi(P)(A + BP) = (A^* + PB^*)(C + DP) \\ &= A^*C + A^*DP + PB^*C + PB^*DP.\end{aligned}$$

Differentiating  $\Psi(P)$  at some fixed  $P_0$  we get that for any symmetric  $W$  the following matrix is also symmetric

$$\begin{aligned}\mathcal{D}_{P_0}\Psi W &= A^*DW + WB^*C + WB^*DP_0 + P_0B^*DW \\ &= (A^*D + P_0B^*D)W + W(B^*C + B^*DP_0).\end{aligned}$$

By Lemma 1 we conclude that there is  $\alpha = \alpha(P_0)$  such that

$$A^*D + P_0B^*D = C^*B + P_0D^*B + \alpha(P_0)I.$$

Differentiating the last equation with respect to  $P_0$  we conclude that  $B^*D$  is symmetric and  $\alpha$  must be constant. We obtain further that  $A^*D - C^*B = \alpha I$ .

Finally also  $A^*C = \Psi(P) - A^*DP - PB^*C - PB^*DP$  must be symmetric.  $\square$

In his groundbreaking paper Bougerol [1] addressed the consequences of the fact that  $\Phi$  must take positive definite matrices into positive definite matrices. The space of positive definite matrices carries the riemannian metric of a symmetric space and Bougerol showed that it is contracted by the action of any such  $\Phi$ . He also showed in a very general setting that each  $Z_n$  is a contraction when  $\langle P_{n-1}^{-1}, \cdot \rangle$  and  $\langle P_n^{-1}, \cdot \rangle$  are used as scalar products in the domain and target spaces respectively. These two contraction properties are the bedrock of the stability of Kalman filters. In particular  $P_n$  for large  $n$  depends very weakly on  $P_0$ . Bougerol's approach allows variable, or random  $S$  and  $\Phi$  [2].

Contraction property is understood here in the weak sense of non-expansion. It holds without any assumptions on the system. Strict contraction requires additional properties which are called controllability and observability. These two properties have transparent geometric interpretation in the language of the maps  $S$ .

Our purpose in this note is to show that both contraction properties can be understood purely in terms of the symplectic structure and that they are rigidly related. This is accomplished by the introduction of another metric into the space of positive definite matrices, based on the symplectic structure [5]. We will call it *the sector metric*. Several basic properties of this metric were established in [5]. In particular it is a finslerian metric, i.e., at every tangent space there is a norm, rather than a scalar product.

We will prove the following property of Kalman filters.

**Theorem 3.** *For a Kalman filter (1),(2) we have for any  $j \geq 1, i \geq 1$*

$$\begin{aligned} 1 &\geq \|Z_{j+i} \dots Z_j\|^2 = \|\mathcal{D}_{P_{j-1}} \Phi^{i+1}\| \\ &\geq \|(\mathcal{D}_{P_{j-1}} \Phi^{i+1})^{-1}\|^{-1} = \|(Z_{j+i} \dots Z_j)^{-1}\|^{-2} \end{aligned}$$

where the sector metric is used for the norm of  $\mathcal{D}\Phi$ , and the scalar products  $\langle P_{j-1}^{-1}, \cdot \rangle$  and  $\langle P_{i+j}^{-1}, \cdot \rangle$  for the other norms.

In particular we get that for  $n = 1, 2, \dots$ , both the weakest and the strongest rates of local contraction of  $\mathcal{D}_{P_{n-1}} \Phi$  are equal to the squares of the weakest and strongest rates of contraction of  $Z_n$ , when these rates are measured in appropriate way (using the sector metric for  $\mathcal{D}\Phi$ ). It shows that the two contractions are rigidly tied.

## 2. Symplectic Sectors and the Sector Metric

The relative covariance matrices  $P_n$  of a Kalman filter are positive definite matrices and hence the symplectic map  $S$  has to preserve the family of Lagrangian subspaces which are graphs of positive semi-definite matrices. The union of such Lagrangian subspaces is the cone

$$\mathcal{C} = \{w = (w_1, w_2) \in \mathbb{R}^k \times \mathbb{R}^k; \langle w_1, w_2 \rangle \geq 0\}$$

which we will call *the standard symplectic sector*. The symplectic sector is defined as the positive cone of the indefinite quadratic form  $\mathcal{Q}(w) = 2\langle w_1, w_2 \rangle$  (the factor 2 is introduced here to simplify some further formulas). The symplectic map  $S$  defined by a Kalman filter takes the standard symplectic sector into itself, we will call such a symplectic map *monotone*. It turns out that monotonicity is equivalent to  $\mathcal{Q}(Sw) \geq \mathcal{Q}(w)$  for every vector  $w$ . If this inequality is strict for nonzero vectors we say that  $S$  is *strictly monotone*. For every monotone symplectic map  $S$  we have the following factorization

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & L \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ R & I \end{bmatrix} \begin{bmatrix} D^{*-1} & 0 \\ 0 & D \end{bmatrix} \quad (5)$$

with symmetric positive semi-definite  $L$  and  $R$ , i.e.,  $L \geq 0, R \geq 0$ . In particular  $D$  (and  $A$ ) must be nonsingular. Strict monotonicity is equivalent to  $L > 0$  and  $R > 0$ , [8].

Let us introduce a symplectic sector and the form  $\mathcal{Q}$  in a geometric, i.e., coordinate free way. This was done in [8], see also [9], where one can find proofs of all the properties which are not proven here.

**Definition 4.** Given two transversal Lagrangian subspaces  $V_1$  and  $V_2$  we define the symplectic sector between  $V_1$  and  $V_2$  by

$$\mathcal{C} = \mathcal{C}(V_1, V_2) = \{w \in \mathcal{W}; \omega(v_1, v_2) \geq 0 \text{ for } w = v_1 + v_2, v_i \in V_i, i = 1, 2\}.$$

We call  $V_1$  and  $V_2$  the sides of the sector.

Equivalently we define first the quadratic form associated with an ordered pair of transversal Lagrangian subspaces by the formula  $\mathcal{Q}(w) = 2\omega(v_1, v_2)$  where  $w = v_1 + v_2, v_i \in V_i, i = 1, 2$ , is the unique decomposition. We have then  $\mathcal{C} = \{w \in \mathcal{W}; \mathcal{Q}(w) \geq 0\}$ .

If we take  $V_1 = \mathbb{R}^k \times \{0\}$  (the horizontal subspace) and  $V_2 = \{0\} \times \mathbb{R}^k$  (the vertical subspace) we get

$$\mathcal{Q}((w_1, w_2)) = 2\langle w_1, w_2 \rangle$$

and the standard symplectic sector  $\mathcal{C}$ . Since any two pairs of transversal Lagrangian subspaces are symplectically equivalent we can consider only this case without loss of generality. However the coordinate free formulations expose better the geometric contents.

The sides of a sector  $\mathcal{C}$  are singled out among all Lagrangian subspaces contained in the boundary of  $\mathcal{C}$  by being the only isolated points [8]. It follows that a symplectic sector determines uniquely its sides.

All our constructions are now based on the symplectic form  $\omega$  and a chosen symplectic sector (defined by a pair of transversal Lagrangian subspaces or equivalently by the quadratic form  $\mathcal{Q}$ ).

We introduce first two monotonicity properties of a linear symplectic map.

**Definition 5.** Given a symplectic sector  $\mathcal{C}$  we call a linear symplectic map  $S$  monotone if

$$S\mathcal{C} \subset \mathcal{C}$$

and strictly monotone if

$$S\mathcal{C} \setminus \{0\} \subset \text{int } \mathcal{C}.$$

An equivalent property is given by

**Theorem 6.** ([8])  $S$  is (strictly) monotone if and only if  $\mathcal{Q}(Sw) \geq \mathcal{Q}(w)$  for every  $w \in \mathcal{W}$  ( $\mathcal{Q}(Sw) > \mathcal{Q}(w)$  for every  $w \in \mathcal{W}, w \neq 0$ ).

Following [5] we define a symplectic angle between two vectors in one symplectic sector (using only the symplectic form  $\omega$  and the form  $\mathcal{Q}$ ).

**Definition 7.** ([5]) *The symplectic angle between two vectors  $w, w' \in \text{int } \mathcal{C}$  is the real number  $\Theta = \Theta(w, w')$  such that*

$$\omega(w, w') = \sqrt{\mathcal{Q}(w)}\sqrt{\mathcal{Q}(w')} \sinh \Theta.$$

For any monotone symplectic map  $S$  and two vectors  $w, w' \in \text{int } \mathcal{C}$  we have

$$\sinh \Theta(Sw, Sw') = \sinh \Theta(w, w') \sqrt{\frac{\mathcal{Q}(w)}{\mathcal{Q}(Sw)}} \sqrt{\frac{\mathcal{Q}(w')}{\mathcal{Q}(Sw')}}.$$

Hence by Theorem 2 a monotone map  $S$  does not increase the absolute value of the symplectic angle and the contraction of the symplectic angle can be estimated by the least coefficient of the expansion of  $\mathcal{Q}$  under the monotone map. By definition if the symplectic form vanishes on a pair of vectors then the symplectic angle between them is zero.

We say that a Lagrangian subspace is contained strictly in the sector  $\mathcal{C}$  (or that it is positive) if the restriction of the form  $\mathcal{Q}$  to the subspace is positive definite, or equivalently, if all of its nonzero vectors are contained in the interior of  $\mathcal{C}$ . We consider the manifold of all Lagrangian subspaces which are strictly contained in the sector  $\mathcal{C}$  and denote it by  $\mathcal{L}$ . In the case of the standard symplectic sector  $\mathcal{L}$  is the space of positive definite matrices.

**Definition 8.** ([5]) *The distance  $s(L, L')$  of two positive Lagrangian subspaces  $L$  and  $L'$  is equal to*

$$s(L, L') = \sup_{0 \neq w \in L, 0 \neq w' \in L'} |\Theta(w, w')|.$$

The contraction property for the mlft of a Kalman filter is built into the definition of this distance function. What is less obvious is that  $s(\cdot, \cdot)$  is indeed a metric. It can be checked by computing it in coordinates.

**Theorem 9.** ([5]) *The distance  $s(P, P')$  of two positive definite matrices  $P$  and  $P'$ , interpreted as two positive Lagrangian subspaces  $gP$  and  $gP'$  in the standard symplectic sector, is equal to*

$$\begin{aligned} s(P, P') &= \max\left\{\frac{|\ln \lambda|}{2}; \lambda \text{ is an eigenvalue of } P'P^{-1}\right\} \\ &= \sup_{0 \neq u \in \mathbb{R}^k} \frac{1}{2} |\ln \langle P'u, u \rangle - \ln \langle Pu, u \rangle|. \end{aligned}$$

**Corollary 10.** *The function  $s(\cdot, \cdot)$  is a complete metric on the space of positive definite matrices. The topology defined by this metric coincides with the standard topology.*

We will call this metric *the sector metric*.

The sector metric is actually the restriction of the Kobayashi metric of the Siegel's upper half-plane to the "imaginary axis" of positive definite matrices, [4], [6], [9]. In particular it is a finslerian metric. More precisely let us introduce norms  $|\cdot|$  in the tangent spaces  $T_P\mathcal{L}$  by the formula

$$\begin{aligned} |W| &= \frac{1}{2} \|P^{-\frac{1}{2}}WP^{-\frac{1}{2}}\| = \frac{1}{2} \sup_{0 \neq u \in \mathbb{R}^k} \frac{|\langle Wu, u \rangle|}{\langle Pu, u \rangle} \\ &= \max\left\{\frac{|\lambda|}{2}; \lambda \text{ is an eigenvalue of } WP^{-1}\right\} \end{aligned}$$

for  $W \in T_P\mathcal{L}$ . It is proven in [5] that the sector metric is the distance function defined by this finslerian metric.

In a different context the sector metric was also introduced by Vesentini [7]. It is a natural generalization of the Hilbert projective metric since it satisfies the Birkhoff identity [3]: *the coefficient of least contraction is equal to the hyperbolic tangent of the diameter of the image*, [5].

The advantage of the sector metric is revealed in the following

**Theorem 11.** *For a fixed positive definite matrix  $P$  and its graph, the Lagrangian subspace  $L = gP$  we have*

$$\|\mathcal{D}_P\Phi\| = \|(S|_L)^{-1}\|^2 \leq 1 \leq \|(\mathcal{D}_P\Phi)^{-1}\| = \|S|_L\|^2$$

where the norm of  $\mathcal{D}_P\Phi$  is defined by the sector metric and the norm of  $S$  by the restriction of the quadratic form  $\mathcal{Q}$  to the respective Lagrangian subspaces.

It follows that  $\mathcal{D}\Phi$ , and hence also  $\Phi$ , is a contraction (in the weak sense) of the sector metric. We postpone the discussion of strict contraction.

**Proof:** The isometries of the sector metric are given by the block diagonal matrices in (5). Hence they act transitively on the space of positive Lagrangian subspaces  $\mathcal{L}$  and we can assume without loss of generality that  $P = I$  and that this subspace is a fixed point of  $\Phi$ , i.e.,  $\Phi(I) = I$ . This leads to  $A + B = C + D = K$  with nonsingular  $K$ .

The tangent space  $T_I\mathcal{L}$  is the space of symmetric matrices. For  $W \in T_I\mathcal{L}$  we have

$$s(e^{\epsilon W}, I) = \epsilon|W| + o(\epsilon), \quad s(\Phi(e^{\epsilon W}), \Phi(I)) = \epsilon|\mathcal{D}_I\Phi W| + o(\epsilon). \quad (6)$$



By the definition of the sector metric we have

$$\begin{aligned} s(e^{\epsilon W}, I) &= \sup_{u \neq 0, u' \neq 0} \sinh^{-1} \frac{|\omega((u, e^{\epsilon W} u), (u', u'))|}{2\sqrt{\langle u, e^{\epsilon W} u \rangle} \sqrt{\langle u', u' \rangle}} \\ &= \epsilon \sup_{u \neq 0, u' \neq 0} \frac{|\langle W u, u' \rangle|}{2\sqrt{\langle u, u \rangle} \sqrt{\langle u', u' \rangle}} + o(\epsilon). \end{aligned} \quad (7)$$

In the same way using (6) and (7) we have also

$$\begin{aligned} s(\Phi(e^{\epsilon W}), \Phi(I)) &= \sup_{u \neq 0, u' \neq 0} \sinh^{-1} \frac{|\omega(S(u, e^{\epsilon W} u), S(u', u'))|}{\sqrt{\mathcal{Q}(S(u, e^{\epsilon W} u))} \sqrt{\mathcal{Q}(S(u', u'))}} \\ &= \epsilon \sup_{u \neq 0, u' \neq 0} \frac{|\langle W u, u' \rangle|}{\sqrt{\mathcal{Q}(S(u, u))} \sqrt{\mathcal{Q}(S(u', u'))}} + o(\epsilon). \end{aligned} \quad (8)$$

It follows from (6) and (8) that

$$\begin{aligned} |\mathcal{D}_I \Phi W| &= \sup_{u \neq 0, u' \neq 0} \frac{|\langle W u, u' \rangle|}{\sqrt{\mathcal{Q}(S(u, u))} \sqrt{\mathcal{Q}(S(u', u'))}} \\ &\leq \sup_{u \neq 0, u' \neq 0} \frac{\sqrt{\mathcal{Q}((u, u))} \sqrt{\mathcal{Q}((u', u'))}}{\sqrt{\mathcal{Q}(S(u, u))} \sqrt{\mathcal{Q}(S(u', u'))}} |W| \\ &= \sup_{u \neq 0} \frac{\mathcal{Q}((u, u))}{\mathcal{Q}(S(u, u))} |W| = \|(S|_L)^{-1}\|^2 |W|. \end{aligned}$$

Since this inequality becomes equality for  $W = I$  we obtain the first part of Theorem 4.

Similarly

$$|\mathcal{D}_I \Phi W| \geq \|S|_L\|^{-2} \sup_{u \neq 0, u' \neq 0} \frac{|\langle W u, u' \rangle|}{\sqrt{\mathcal{Q}((u, u))} \sqrt{\mathcal{Q}((u', u'))}} = \|S|_L\|^{-2} |W|.$$

We get equality again if we choose  $W = K^* K$  that produces the rest of Theorem 4. ■

We proceed to prove Theorem 1. Let  $L_0 = gP_0$  be the graph of  $P_0$  and  $L_n = S^n L_0$  then  $Z_n^*$  is the projection on the first coordinate of the restriction of  $S^{-1}$  to  $L_n$ , i.e., the following diagram commutes

$$\begin{array}{ccccccc} \mathbb{R}^k \times \mathbb{R}^k & \xrightarrow{S^{-1}} & \mathbb{R}^k \times \mathbb{R}^k & \dots & \mathbb{R}^k \times \mathbb{R}^k & \xrightarrow{S^{-1}} & \mathbb{R}^k \times \mathbb{R}^k \\ \cup & & \cup & \dots & \cup & & \cup \\ L_n & \xrightarrow{S^{-1}} & L_{n-1} & \dots & L_1 & \xrightarrow{S^{-1}} & L_0 \\ \downarrow \pi_1 & & \downarrow \pi_1 & \dots & \downarrow \pi_1 & & \downarrow \pi_1 \\ \mathbb{R}^k & \xrightarrow{Z_n^*} & \mathbb{R}^k & \dots & \mathbb{R}^k & \xrightarrow{Z_1^*} & \mathbb{R}^k \end{array}$$

where  $\pi_1 : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k$  is the projection on the first coordinate. Hence if we use  $\langle P_n \cdot, \cdot \rangle$  and  $\langle P_{n-1} \cdot, \cdot \rangle$  as scalar product in the domain and target space respectively we obtain by Theorem 4 that  $Z_n^*$  is a contraction and that the squares of the weakest and the strongest rates are equal to the rates of contraction of  $\mathcal{D}_{P_{n-1}} \Phi$ . The contraction properties of  $Z_n$  are the same as the contraction properties of  $Z_n^*$  if we use  $\langle P_{n-1}^{-1} \cdot, \cdot \rangle$  and  $\langle P_n^{-1} \cdot, \cdot \rangle$  as scalar products in the domain and target space respectively. The same can be claimed about the compositions  $Z_{i+j} \dots Z_j$  and  $\mathcal{D}_{P_{j-1}} \Phi^{i+1}$ , in view of the general nature of Theorem 4. This establishes Theorem 1.

Let us finally discuss the important issue of strict contraction.

**Theorem 12.** ([5]) *The mlft  $\Phi$  defined by a monotone map  $S$  is a strict contraction of the sector metric if and only if  $S$  is strictly monotone.*

A monotone map  $S$  is strictly monotone if and only if  $\Phi$  takes the sides of the standard sector into positive subspaces, i.e., if and only if the matrices  $B$  and  $C$  in (5) are of full rank.

For the mlft of the Kalman filter we do not get strict monotonicity for time one because usually the matrices  $GG^*$  and/or  $H^*H$  are not of full rank. However for some composition  $S^n$  the matrices  $B$  and  $C$  in (5) may become nondegenerate. By definition the system is observable if after some time the matrix  $B$  becomes nondegenerate. It is controllable if after some time the matrix  $C$  becomes nondegenerate.

By Theorem 4 if  $S^n$  is strictly monotone, then the map  $Z^n = Z_n \dots Z_1$  is a strict contraction. However the following weaker condition is sufficient.

**Proposition 13.** *The restriction of  $S$  to a positive subspace  $L = gP$  is a strict expansion if and only if the kernels of  $BP$  and  $C$  have only trivial intersection.*

**Proof:** Let us consider the factorization (5) of  $S$ . Since the last factor does not change the value of the form  $\mathcal{Q}$  we can assume without loss of generality that  $D = I$ . Now if there is a vector  $u \in \mathbb{R}^k$  such that

$$\mathcal{Q}(S(u, Pu)) = \langle u, Pu \rangle \quad (9)$$

then  $\mathcal{Q}((u, Ru + Pu)) = \langle u, Pu \rangle$  and we get  $\langle u, Ru \rangle = 0$ , and finally  $Ru = 0$  because  $R \geq 0$ .

Having established that, we get  $S(u, Pu) = (u + LPu, Pu)$  and using again (9) we conclude that  $LPu = 0$ . ■

In particular observability or controllability alone is sufficient to guarantee strict contraction in Theorem 1 and Theorem 4.

In the case of time dependent systems, or the case of random coefficients considered in [2], the explicit conditions of observability and controllability may get cumbersome, but their geometric meaning does not change, it is about the horizontal or vertical side of the standard sector entering eventually the interior (mapping onto a positive Lagrangian subspace).

If the system is both controllable and observable then  $\Phi$  is a strict contraction. Its weakest rate of contraction is connected to the weakest rate of expansion of the form  $\mathcal{Q}$  by the symplectic monotone map  $S$ . We have

**Theorem 14.** ([5]) *The coefficient of least contraction of the sector metric under the action of the mft  $\Phi$  defined by a strictly monotone map  $S$  is equal to the inverse square of the coefficient of least expansion of the form  $\mathcal{Q}$  over the interior of the symplectic sector. This coefficient is also equal to  $\tanh \frac{\text{diam}(\Phi\mathcal{L})}{2}$ , and  $\text{diam}(\Phi\mathcal{L})$  is equal to the distance of the image under  $\Phi$  of the vertical and horizontal subspaces respectively, i.e.,  $\text{diam}(\Phi\mathcal{L}) = \frac{1}{2} \ln \lambda$  where  $\lambda$  is the largest eigenvalue of  $CA^{-1}BD^{-1}$  (which is automatically positive).*

Hence our Theorem 4 can be understood as the local version of Theorem 6.

### 3. Appendix. Derivation of the Kalman Filter

For the convenience of the reader we include a brief derivation of the Kalman filter.

Let us recall first two basic properties of gaussian vectors.

- (\*) If  $X \in \mathbb{R}^k$  is a gaussian vector with mean  $\widehat{X}$  and covariance matrix  $\text{Cov}(X, X) = M$  and  $A$  is a  $k \times k$  matrix then  $AX$  is a Gaussian vector with the mean  $A\widehat{X}$  and the covariance matrix  $AMA^*$ .
- (\*\*) If  $(X, Y) \in \mathbb{R}^k \times \mathbb{R}^m$  is a gaussian vector then  $X$  and  $Y$  are gaussian vectors in their own right and the conditional distribution of  $X$  given  $Y$  is also gaussian with  $\mathbb{E}(X|Y) = \widehat{X} + SQ^{-1}(Y - \widehat{Y})$  and  $\text{Cov}(X, X|Y) = \text{Cov}(X, X) - SQ^{-1}S^*$  where  $S = \text{Cov}(X, Y)$  and  $Q = \text{Cov}(Y, Y)$ .

The crucial claim is that given  $Y_1, Y_2, \dots, Y_{n-1}$  the random vector  $(X_n, Y_n)$  is Gaussian. The proof in this case is by induction. Since  $X_0$  is Gaussian, we get

that  $(X_1, Y_1)$  is Gaussian. Let us assume that given  $Y_1, Y_2, \dots, Y_{n-2}$  the random vector  $(X_{n-1}, Y_{n-1})$  is Gaussian. We have also that  $(X_{n-1}, Y_{n-1}, \varepsilon_n, \eta_n)$  is Gaussian, given  $Y_1, \dots, Y_{n-2}$ . We have

$$X_n = FX_{n-1} + G\varepsilon_n \quad (10)$$

$$Y_n = HF X_{n-1} + HG\varepsilon_n + \eta_n. \quad (11)$$

By the property (\*) we conclude that  $(X_n, Y_n, Y_{n-1})$  is a Gaussian vector, given  $Y_1, \dots, Y_{n-2}$ . The inductive step now follows from the property (\*\*). Indeed including  $Y_{n-1}$  in the condition results in a Gaussian vector  $(X_n, Y_n)$ , given  $Y_1, \dots, Y_{n-1}$ .

Averaging (10) under the condition  $Y_1, \dots, Y_{n-1}$  we get

$$\mathbb{E}(X_n | \mathcal{F}_{n-1}) = F\widehat{X}_{n-1} + GG^*.$$

We apply the property (\*\*) to the Gaussian vector  $(X_n, Y_n)$ . We conclude that

$$\widehat{X}_n = \mathbb{E}(X_n | \mathcal{F}_n) = \mathbb{E}(X_n | \mathcal{F}_{n-1}) + S_n Q_n^{-1} (Y_n - \widehat{Y}_n)$$

where  $S_n = \text{Cov}(X_n, Y_n | \mathcal{F}_{n-1})$  and  $Q_n = \text{Cov}(Y_n, Y_n | \mathcal{F}_{n-1})$ .

Let us introduce  $P_n = \text{Cov}(X_n, X_n | \mathcal{F}_n)$  and  $R_n = \text{Cov}(X_n, X_n | \mathcal{F}_{n-1})$ . We get from (10) and the property (\*\*) that

$$R_n = FP_{n-1}F^* + GG^* \quad (12)$$

$$P_n = R_n - S_n Q_n^{-1} S_n^*. \quad (13)$$

(12) is an mlft connecting  $R_n$  and  $P_{n-1}$ . Our goal is to show that also (13) is an mlft relating  $R_n$  and  $P_n$ . By the group property of mlfts we will obtain then the mlft relating  $P_n$  and  $P_{n-1}$ .

We get from (10) and (11)

$$S_n = FP_{n-1}F^*H^* + GG^*H^* = R_nH^* = P_nH^* + S_nQ_n^{-1}S_n^*H^* \quad (14)$$

$$Q_n = HFP_{n-1}F^*H^* + HGG^*H^* + I. \quad (15)$$

We get further from (14) and (15) that  $HS_n = Q_n - I = S_n^*H^*$ . Substituting the last formula into (14) we get  $P_nH^* = S_nQ_n^{-1}$ . Using the last equality and  $S_n^* = HR_n$  we get from (13)

$$R_n = P_n + S_nQ_n^{-1}S_n^* = P_n(I + H^*HR_n).$$

We conclude that

$$P_n = R_n(I + H^*HR_n)^{-1}.$$

The derivation of the Kalman filter is completed.

## Acknowledgements

While writing this paper in the Spring of 2007 we enjoyed the hospitality of MSRI in Berkeley. We especially benefitted from the discussions with the Data Assimilation Group: Amit Apte, Hamzi Boumediene, Chris Jones, Art Krener, Jonathan Mattingly and Sri Namachchivaya.

## References

- [1] Bougerol P., *Filtre de Kalman Bucy et exposants de Lyapunov*, Lyapunov Exponents, Proceedings, Oberwolfach 1990, L. Arnold, H. Crauel and J.-P. Eckmann (Eds) Lecture Notes in Math. **1486** (1991) pp. 112 – 122.
- [2] Bougerol P., *Kalman Filtering with Random Coefficients and Contractions*, Siam J. Control & Opt. **31** (1993) 942 - 959.
- [3] Birkhoff G., *Extensions of Jentzsch's Theorem*, Trans. Amer. Math. Soc. **85** (1957) 219 – 227.
- [4] Isaev A. and Krantz S., *Invariant Distances and Metrics in Complex Analysis*, Notices AMS **47** (2000) 546 -553.
- [5] Liverani C. and Wojtkowski M., *Generalization of the Hilbert Metric to the Space of Positive Definite Matrices*, Pac. J. Math. **166** (1994) 339 -355.
- [6] Suzuki M., *The Intrinsic Metrics on the Circular Domains in  $\mathbb{C}^n$* , Pac. J. Math. **112** (1984) 249-256.
- [7] Vesentini E., *Invariant Metrics on Convex Cones*, Ann. Sc. Norm. Sup. Pisa **3** (1976) 671 – 696.
- [8] Wojtkowski M., *Systems of Classical Interacting Particles with Nonvanishing Lyapunov Exponents*, Lyapunov Exponents, Proceedings, Oberwolfach 1990, L. Arnold, H. Crauel and J.-P. Eckmann (Eds) Lecture Notes in Math. **1486** (1991) 243 – 262.
- [9] Wojtkowski M., *Monotonicity,  $\mathcal{J}$ -Algebra of Potapov and Lyapunov Exponents*, Smooth Ergodic Theory and Its Applications, Proc. Symp. Pure Math., AMS, **69** (2001) 499 - 521.

Maciej P. Wojtkowski

Department of Mathematics and Informatics

University of Warmia and Mazury

ul. Zolnierska 14

10-561 Olsztyn

POLAND

*E-mail address:* maciejw@math.arizona.edu