KYBERNETIKA — VOLUME 11 (1975), NUMBER 4

LQG Problem of Estimation and Control in the Tensor Space

Evolution in the Discrete Finite Time

ANTONÍN VANĚČEK

LQG problem of estimation and control with linear dynamics, quadratic criterion, and gaussian environment leads during its application for identification naturally to the tensor formulation. Quasikalmanean tensor model and associated LQG problems are introduced. Theorems on prediction, filtering, control, stochastic control, and stochastic control/filtering are presented. The cases of filtering and stochastic control for the naturally introduced tensor models are given as the applications.

INTRODUCTION

LQG problem, see [2] for a survey, was stated and solved mainly as the problem of estimation and control of the state vector. Through the formulation of the three constraints – linear dynamics, quadratic criterion, and white gaussian environment – LQG problem covered an inportant part of questions solvable by the finite methods. In [12] we have tried to show the usefulness of the mentioned problem for identification. Nevertheless there are mainly identified the linear maps of the vectors, consequently the tensors. In the following we shall state and solve the LQG problems for the tensors. Through the intrinsic duality we shall double our relations: from the motivation point of view we take the relations for the control as the by-products.

Let us refer to one area at which the introduction of richer structures has proved useful. The Maxwell theory of electromagnetic field in vacuum can be desribed either by 24 scalar equations or by 8 vector equations or by 3 tensor equations or by 1 spinor equation. (It is to be understood this is not just through trivial direct sum.) The mentioned reduction at the same time contributed to the knowledge and the deepening of the Maxwell theory.

Tensor symbolism.

$a, e, m, u, v, w, x, y, z$	1st order tensors,
$A, B, C, D, E, H, I, M, S, U, V, W, X, Y, Z$	2nd order tensors,
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{F}, \mathcal{H}, \mathcal{I}, \mathcal{K}, \mathcal{L}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$	4th order tensors.

Definition 1. Let x, y, Z be the elements of the linear finite – dimensional spaces R^b, R^a, $R^a \times R^b$ over the field of real numbers R. The map $\otimes$: $R^b \times R^a \to R^a \times R^b$: $(x, y) \mapsto Z = y \otimes x$: $(x_\beta, y_\alpha) \mapsto Z_{\alpha\beta} = y_\alpha x_\beta$ $(\alpha = 1, \ldots, a, \beta = 1, \ldots, b)$ we shall the tensor product, [4; 9], and the result of that product, $Z = y \otimes x$, we shall call the 2nd order tensor. Further let X, Y, $\mathcal{Z}$ be the elements of the linear spaces $R^c \times R^d$, $R^a \times R^b$, $R^a \times R^b \times R^c \times R^d$ over R. The map $\otimes$: $R^c \times R^d \times R^a \times R^b \to R^a \times R^b \times R^c \times R^d$: $(X, Y) \mapsto \mathcal{Z} = Y \otimes X$: $(X_{\gamma\delta}, Y_{\alpha\beta}) \mapsto \mathcal{Z}_{\alpha\beta\gamma\delta} = Y_{\alpha\beta} X_{\gamma\delta}$ $(\alpha = 1, \ldots, a; \ldots; \delta = 1, \ldots, d)$ we shall again call the tensor product, and the result of that product, $\mathcal{Z} = Y \otimes X$, we shall call the 4th order tensor. Finally we shall call x, y the 1st order tensors.

Convention (Einstein summation rule). With the use of the component tensor symbolism, we suppose the compatible dimensions and add over the indexes occuring twice. E.g. $\mathcal{Y}_{\alpha\beta\gamma\delta} X_{\beta\delta}$ denotes $\sum_{(\beta,\delta)} \mathcal{Y}_{\alpha\beta\gamma\delta} X_{\beta\delta}$, $\mathcal{Y}_{\alpha\beta\beta\alpha}$ denotes $\sum_{(\alpha,\beta)} \mathcal{Y}_{\alpha\beta\beta\alpha}$. We use the component tensor symbolism only in Preliminaries.

Definition 2. The map $Y: R^b \to R^a : x \mapsto z = Yx : x_\beta \mapsto z_\alpha = Y_{\alpha\beta} x_\beta$ $(\alpha = 1, \ldots, a, \beta = 1, \ldots, b)$ we shall call the linear map between the 1st order tensors. Such a map Y that $\forall x : Yx = x$ we shall call unit 2nd order tensor and denote 1_{R^b}. The map: $R^b \times R^c \times R^a \times R^b \to R^a \times R^c$: $(X, Y) \mapsto Z = YX : (X_{\beta\gamma}, Y_{\alpha\beta}) \mapsto Z_{\alpha\beta} = Y_{\alpha\beta} X_{\beta\gamma}$ $(\alpha = 1, \ldots, a; \ldots; \gamma = 1, \ldots, c)$ we shall call the product of the 2nd order tensors. The map $\mathcal{Y} : R^b \times R^d \to R^a \times R^c : X \mapsto Z = \mathcal{Y}X : X_{\beta\delta} \mapsto Z_{\alpha\gamma} = \mathcal{Y}_{\alpha\beta\gamma\delta} X_{\beta\delta}$ $(\alpha = 1, \ldots, a; \ldots; \delta = 1, \ldots, d)$ we shall call the linear map between the 2nd order tensors. Such a map $\mathcal{Y}$ that $\forall X : \mathcal{Y}X = X$, we shall call the unit 4th order tensor and denote $1_{R^b \times R^d}$. The map: $R^b \times R^e \times R^d \times R^f \times R^a \times R^b \times R^c \times R^d \to R^a \times R^c \times R^e \times R^f : (\mathcal{X}, \mathcal{Y}) \mapsto \mathcal{Z} = \mathcal{Y}\mathcal{X} : (\mathcal{X}_{\beta\varepsilon\delta\varphi}, \mathcal{Y}_{\alpha\beta\gamma\delta}) \mapsto \mathcal{Z}_{\alpha\gamma\varepsilon\varphi} = \mathcal{Y}_{\alpha\beta\gamma\delta} \mathcal{X}_{\beta\varepsilon\delta\varphi}$ $(\alpha = 1, \ldots, a; \ldots; \varphi = 1, \ldots, f)$ we shall call the product of the 4th order tensors.

Definition 3. The map $\operatorname{tr} : R^a \times R^a \to R : Y \mapsto \operatorname{tr} Y : Y_{\alpha\beta} \mapsto Y_{\alpha\alpha}$ $(\alpha, \beta = 1, \ldots, a)$ we shall call the trace of the 2nd order tensor. The map $\operatorname{tr} : R^a \times R^a \times R^c \times R^c \to R : \mathcal{Y} \mapsto \operatorname{tr} \mathcal{Y} : \mathcal{Y}_{\alpha\beta\gamma\delta} \mapsto \mathcal{Y}_{\alpha\beta\beta\alpha}$ $(\alpha, \delta = 1, \ldots, a;\ \beta, \gamma = 1, \ldots, b)$ we shall call the trace of the 4th order tensor.

Definition 4. Let $x \in R^b$, $Y : R^b \to R^a$, $z \in R^a$. Such a linear map Y^*: $R^a \to R^b$ that $\forall x, Y, z : \operatorname{tr}[z \otimes (Yx)] = \operatorname{tr}[(Y^*z) \otimes x]$ we shall call adjoint to the map Y.

Further let $X \in R^b \times R^d, \mathscr{Y} : R^b \times R^d \to R^a \times R^c, Z \in R^a \times R^c$. Such a linear map $\mathscr{Y}^* : R^a \times R^c \to R^b \times R^d$ that $\forall X, \mathscr{Y}, Z : \mathrm{tr}\,[Z \otimes (\mathscr{Y}X)] = \mathrm{tr}\,[(\mathscr{Y}^*Z) \otimes X]$ we shall call adjoint to the map $\mathscr{Y}$. Finally let $X \in R^a \times R^c, \mathscr{Y} : R^a \times R^c \to R^a \times R^c$, $\ker \mathscr{Y} = 0$. Such a linear map $\mathscr{Y}^{-1} : R^a \times R^c \to R^a \times R^c$ that $\forall X, \mathscr{Y} : \mathscr{Y}^{-1}\mathscr{Y}X = X$ we shall call inverse to the map $\mathscr{Y}$ (ker denotes kernel, i. e. null space).

Note 1. Having at our disposal matrix algebra numerical algorithms, the matrix realization of tensor algebra may be useful. Let $x \in R^b$, $Y \in R^a \times R^b$, $X \in R^c \times R^d$, $\mathscr{Y} \in R^a \times R^b \times R^c \times R^d$, $\mathscr{X} \in R^b \times R^e \times R^d \times R^f$. Then from Definitions 1, 2, 4 follow the isomorfisms ($=_{\mathrm{df}}$ stands for "denotes"):

$$x \cong \begin{bmatrix} x_1 \\ \dots \\ x_b \end{bmatrix} \in R^b \times R\,;$$

$$Y \cong \begin{bmatrix} \begin{bmatrix} Y_{11} \\ \dots \\ Y_{a1} \end{bmatrix} \dots \begin{bmatrix} Y_{1b} \\ \dots \\ Y_{ab} \end{bmatrix} \end{bmatrix} =_{\mathrm{df}} [{}^1Y \dots {}^bY] \cong \begin{bmatrix} {}^1Y \\ \dots \\ {}^bY \end{bmatrix} \in R^{ab} \times R\,;$$

$$Y \otimes X \cong \begin{bmatrix} X_{11}Y & \dots & X_{1d}Y \\ \dots & \dots & \dots \\ Y_{c1}Y & \dots & X_{cd}Y \end{bmatrix} \in R^{ac} \times R^{bd}\,;$$

$$\mathscr{Y} \cong \begin{bmatrix} \mathscr{Y}_{11} & \dots & \mathscr{Y}_{1d} \\ \dots & \dots & \dots \\ \mathscr{Y}_{c1} & \dots & \mathscr{Y}_{cd} \end{bmatrix} \in R^{ac} \times R^{bd}\,;$$

$$\mathscr{Y}^* \cong \begin{bmatrix} \mathscr{Y}_{11} & \dots & \mathscr{Y}_{1d} \\ \dots & \dots & \dots \\ \mathscr{Y}_{c1} & \dots & \mathscr{Y}^{cd} \end{bmatrix}^{\mathrm{T}} \in R^{bd} \times R^{ac}\,;$$

$$\mathscr{Y}X \cong \begin{bmatrix} \mathscr{Y}_{11} & \dots & \mathscr{Y}_{1d} \\ \dots & \dots & \dots \\ \mathscr{Y}_{c1} & \dots & \mathscr{Y}_{cd} \end{bmatrix} \begin{bmatrix} {}^1X \\ \dots \\ {}^dX \end{bmatrix} \in R^{ac} \times R\,;$$

$$\mathscr{Y}\mathscr{X} \cong \begin{bmatrix} \mathscr{Y}_{11} & \dots & \mathscr{Y}_{1d} \\ \dots & \dots & \dots \\ \mathscr{Y}_{c1} & \dots & \mathscr{Y}_{cd} \end{bmatrix} \begin{bmatrix} \mathscr{X}_{11} & \dots & \mathscr{X}_{1f} \\ \dots & \dots & \dots \\ \mathscr{X}_{1d} & \dots & \mathscr{X}_{df} \end{bmatrix} \in R^{ac} \times R^{ef}$$

where $Y \otimes X \in R^a \times R^b \times R^c \times R^d, \mathscr{Y}^* \in R^b \times R^a \times R^d \times R^c$, $\mathscr{Y}X \in R^a \times R^c$, $\mathscr{Y}\mathscr{X} \in R^a \times R^c \times R^e \times R^f$. Further let $\ker \mathscr{Y} = 0$, $b = a$, $d = c$. Then

$$\mathscr{Y}^{-1} \cong \begin{bmatrix} \mathscr{Y}_{11} & \dots & \mathscr{Y}_{1c} \\ \dots & \dots & \dots \\ \mathscr{Y}_{c1} & \dots & \mathscr{Y}_{cc} \end{bmatrix}^{-1} \in R^{ac} \times R^{ac}\,.$$

Note 2. We shall show how the tensor product and the 4th order tensors naturally arise. Let $\mathscr{A} : R^b \times R^d \to R^a \times R^c : X \mapsto Y = \mathscr{A}X = AXB + CX + XD$. Then from Definitions 1, 2, 4 or from Note 1 follows: $\mathscr{A} = A \otimes B^* + C \otimes 1_{R^d} + 1_{R^b} \otimes D^*$.

Definition 5. Let $X \in R^a \times R^c, \mathscr{W} \in R^a \times R^a \times R^c \times R^c, \mathscr{X} \in R^a \times R^e \times R^c \times R^f$. Then $\|X\|^2_{\mathscr{W}}$, $\|\mathscr{X}\|^2_{\mathscr{W}}$ will denote $\operatorname{tr}[X \otimes (\mathscr{W}X)^*] = \operatorname{tr}[X^*(\mathscr{W}X)]$, $\operatorname{tr}[\mathscr{X}^*(\mathscr{W}\mathscr{X})]$, respectively. Esp. for $\mathscr{W} = 1_{R^a \times R^c}$, $\|X\|^2$, $\|\mathscr{X}\|^2$ will denote $\operatorname{tr}(X \otimes X^*) = \operatorname{tr}(X^*X)$, $\operatorname{tr}(\mathscr{X}^*\mathscr{X})$, respectively. Further if $\forall X \neq 0$: $\|X\|^2_{\mathscr{W}} > 0$, $\geqq 0$, we shall write $\mathscr{W} > 0$, $\geqq 0$, respectively. Finally let $\mathscr{W} > 0$. Then $\|X\|_{\mathscr{W}}$, $\|\mathscr{X}\|_{\mathscr{W}}$ we shall call the quadratic norm of X, $\mathscr{X}$, respectively.

Definition 6. Let (Ω_1, F_1, P_1), (Ω_2, F_2, P_2) be the probability spaces and $X : \Omega_1 \to R^d \times R^c$, $Y: \Omega_2 \to R^a \times R^b$ be the random variables. The map: $(X, Y) \mapsto \underset{\Omega_1 \times \Omega_2}{\mathrm{E}} \{[Y(\Omega_2) - \underset{\Omega_2}{\mathrm{E}}\, Y(\Omega_2)] \otimes [X(\Omega_1) - \underset{\Omega_1}{\mathrm{E}}\, X(\Omega_1)]^*\} \in R^a \times R^b \times R^c \times R^d$ we shall call the covariance and denote $\operatorname{cov}(Y, X)$ or $\mathscr{S}_{YX}$. Esp. $\operatorname{cov}(Y, Y)$ we shall denote $\operatorname{cov}(Y, \ldots)$ or $\mathscr{S}_{YY}$ or $\mathscr{S}$. The random variable $Y: \Omega \to R^a \times R^b$ with the mean value $M_Y \in R^a \times R^b$, with the covariance $\mathscr{S}_{YY} \in R^a \times R^b \times R^b \times R^a$, $\mathscr{S}_{YY} \geqq 0$, and with the characteristic function: $R^a \times R^b \to R^2 : V \mapsto \exp\{\mathrm{j} \operatorname{tr}(V \otimes M_Y^*) - 2^{-1} \operatorname{tr}[V \otimes (\mathscr{S}_{YY}V)^*]\}$, $\mathrm{j}^2 = -1$, we shall call gaussian. We shall write $Y \sim \mathrm{N}(M_Y, \mathscr{S}_{YY})$.

Note 3. The map tr induces – together with the componentwise addition and the outer (from R) componentwise multiplication – the Hilbert space over R for the tensors. In the stochastic case the Hilbert space scalar product is in addition induced by the trace mean value.

PROBLEM STATEMENT

Definition 7 (Quasikalmanean tensor model and associated LQG problems).
State model:

$$X_{k+1} = \mathscr{A}_k X_k + \mathscr{B}_k U_k + \mathscr{D}_k W_k\,, \tag{1}$$

output model:

$$Z_k = \mathscr{C}_k X_k + V_k\,,$$

where $k = 0, 1, \ldots, N - 1$ and linear maps $\mathscr{A}_k : R^a \times R^b \to R^a \times R^b$, $\mathscr{B}_k : R^c \times R^d \to R^a \times R^b$, $\mathscr{D}_k : R^e \times R^f \to R^a \times R^b$, $\mathscr{C}_k : R^a \times R^b \to R^g \times R^h$.

A priori distribution:

$$\begin{bmatrix} X_0 \\ W_0 \\ \cdot \\ \cdot \\ \cdot \\ W_k \\ V_0 \\ \cdot \\ \cdot \\ \cdot \\ V_k \end{bmatrix} \sim \mathrm{N} \left\{ \begin{bmatrix} M \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \begin{bmatrix} \mathscr{S} & & & & & & & \\ & \mathscr{W}_0 & & & & & & \\ & & \ddots & & & & & \\ & & & \mathscr{W}_k & & & & \\ & & & & \mathscr{V}_0 & & & \\ & & & & & \ddots & & \\ & & & & & & & \mathscr{V}_k \end{bmatrix} \right\},$$

$\mathscr{S}, \mathscr{W}_0, \ldots, \mathscr{W}_k \geqq 0, \mathscr{V}_0, \ldots, \mathscr{V}_k > 0.$

Control criterion:

$$J(\boldsymbol{U}_k) = \mathrm{tr} \left\{ \begin{bmatrix} X_N \\ X_{N-1} \\ \cdot \\ \cdot \\ \cdot \\ X_k \\ U_{N-1} \\ \cdot \\ \cdot \\ \cdot \\ U_k \end{bmatrix}^* \begin{bmatrix} \mathscr{P} & & & & & & \\ & \mathscr{Q}_{N-1} & & & & & \\ & & \ddots & & & & \\ & & & \mathscr{Q}_k & & & \\ & & & & \mathscr{R}_{N-1} & & \\ & & & & & \ddots & \\ & & & & & & \mathscr{R}_k \end{bmatrix} \begin{bmatrix} X_N \\ X_{N-1} \\ \cdot \\ \cdot \\ \cdot \\ X_k \\ U_{N-1} \\ \cdot \\ \cdot \\ \cdot \\ U_k \end{bmatrix} \right\},$$

$\mathscr{P}, \mathscr{Q}_{N-1}, \ldots, \mathscr{Q}_k, \mathscr{R}_{N-1}, \ldots, \mathscr{R}_k > 0$; future control: $\boldsymbol{U}_k =_{\mathrm{df}} \{U_k, U_{k+1}, \ldots, U_{N-1}\}$.

Past inputs: $\boldsymbol{Z}_k =_{\mathrm{df}} \{Z_k, Z_{k-1}, \ldots, Z_0, M, \mathscr{S}\}$, $\boldsymbol{Z}_{-1} =_{\mathrm{df}} \{M, \mathscr{S}\}$; past states: $\boldsymbol{X} =_{\mathrm{df}} \{X_k, X_{k-1}, \ldots, X_0\}$.

One-step prediction problem – $n = 1$, prediction problem – $n \geqq 2$, filtering problem – $n = 0$ are: Find affine map $\boldsymbol{Z}_k \mapsto \hat{X}_{k+n|k}$ minimizing $\mathrm{E}\{\|X_{k+n} - \hat{X}_{k+n|k}\|^2 | \boldsymbol{Z}_k\}$.

Control problem: For $\mathscr{D}_k V_k = \ldots = \mathscr{D}_{N-1} V_{N-1} = 0$ find a linear map $\boldsymbol{X}_k \mapsto U_k$ minimizing $J(\boldsymbol{U}_k)$.

Stochastic control problem: Find a linear map $\boldsymbol{X}_k \mapsto U_k$ minimizing $\mathrm{E}\{J(\boldsymbol{U}_k) \mid \boldsymbol{X}_k\}$.

Stochastic control/filtering problem: Find an affine map $\boldsymbol{Z}_k \mapsto U_k$ minimizing $\mathrm{E}\{J(\boldsymbol{U}_k) \mid \boldsymbol{Z}_k\}$.

Lemma. Let $\operatorname{tr}(Y \otimes Z^*) = 0$; $Y \otimes Y^*, Z \otimes Z^* > 0$. Then $\exists!$ (there exist unique) tensors $\mathscr{A}^\circ, \mathscr{B}^\circ$ minimizing $J(\mathscr{A}, \mathscr{B}) = \|X - \mathscr{A}Y - \mathscr{B}Z\|^2$. Further: $\mathscr{A}^\circ = (X \otimes Y^*)(Y \otimes Y^*)^{-1}$, $\mathscr{B}^\circ = (X \otimes Z^*)(Z \otimes Z^*)^{-1}$, $J(\mathscr{A}^\circ, \mathscr{B}^\circ) = \operatorname{tr}[X \otimes X^* - \mathscr{A}^\circ(Y \otimes X^*) - \mathscr{B}^\circ(Z \otimes X^*)]$.

Proof. We shall find the stationary point of criterion from nullity of the Gateaux differential, [6, 10]: $\delta J(\mathscr{A}, \mathscr{B}; \mathscr{H}, \mathscr{K}) = \lim_{\alpha \to 0} \alpha^{-1} [J(\mathscr{A} + \alpha\mathscr{H}, \mathscr{B} + \alpha\mathscr{K}) - J(\mathscr{A}, \mathscr{B})] = \lim_{\alpha \to 0} \alpha^{-1} \operatorname{tr} \{[X - (\mathscr{A} + \alpha\mathscr{H}) Y - (\mathscr{B} + \alpha\mathscr{K}) Z] \otimes [X - (\mathscr{A} + \alpha\mathscr{H}) . Y - (\mathscr{B} + \alpha\mathscr{K}) Z]^*\} = \operatorname{tr} \lim_{\alpha \to 0} \alpha^{-1}[-\alpha\mathscr{H}Y \otimes (X - \mathscr{A}Y)^* - \alpha(X - \mathscr{A}Y) \otimes (\mathscr{H}Y)^* - \alpha\mathscr{K}Z \otimes (X - \mathscr{B}Z)^* - \alpha(X - \mathscr{B}Z) \otimes (\mathscr{K}Z)^* + o(\alpha)] = -2 \operatorname{tr} [(X - \mathscr{A}Y) \otimes (\mathscr{H}Y)^*] - 2 \operatorname{tr} [(X - \mathscr{B}Z) \otimes (\mathscr{K}Z)^*] = -2 \operatorname{tr} [(X - \mathscr{A}Y) \otimes Y^*\mathscr{H}^*] - 2 \operatorname{tr} [(X - \mathscr{B}Z) \otimes Z^*\mathscr{K}^*] = 0 \; \forall \mathscr{H}, \mathscr{K}$. The given condition is fulfilled just for $(X - \mathscr{A}Y) \otimes Y^* = 0$, $(X - \mathscr{B}Z) \otimes Z^* = 0$, i.e. for the normal equations for the tensors $\mathscr{A}, \mathscr{B}$. From $Y \otimes Y^*, Z \otimes Z^* > 0$ follows that $\exists!$ solution of these normal equations: $\mathscr{A}^\circ = (X \otimes Y^*)(Y \otimes Y^*)^{-1}$, $\mathscr{B}^\circ = (X \otimes Z^*)(Z \otimes Z^*)^{-1}$. From the unit criterion weight follows that this solution is the minimum. Finally the value of $J(\mathscr{A}^\circ, \mathscr{B}^\circ)$ follows from the direct substitution.

Corollary 1. Let $Y \otimes Y^* > 0$. Then $\exists!$ tensor $\mathscr{A}^\circ$ minimizing $J(\mathscr{A}) = \|\mathscr{L}_1X_1 + \mathscr{L}_2X_2 - \mathscr{A}Y\|^2$. Further: $\mathscr{A}^\circ = \mathscr{L}_1\mathscr{A}_1^\circ + \mathscr{L}_2\mathscr{A}_2^\circ$ where $\mathscr{A}_i^\circ = (X_i \otimes Y^*)(Y \otimes Y^*)^{-1}$, $i = 1, 2$.

Note 4. Let X, Y, Z be random variables such that $M_X, M_Y, M_Z, \mathscr{S}_{YZ} = 0$; $\mathscr{S}_{YY}, \mathscr{S}_{ZZ} > 0$. Defining the criterion of estimation as $\mathrm{E}\{\|X - \mathscr{A}Y - \mathscr{B}Z\|^2 \mid Y, Z\}$, we shall call $\mathscr{A}^\circ Y + \mathscr{B}^\circ Z$ ($\mathscr{A}^\circ = \mathscr{S}_{XY}\mathscr{S}_{YY}^{-1}$, $\mathscr{B}^\circ = \mathscr{S}_{XZ}\mathscr{S}_{ZZ}^{-1}$) the (linear) optimum estimate of X conditioned by Y, Z and denote $\hat{X}_{|Y,Z}$. (For $M_X \neq 0$ we shall obtain the (affine) optimum estimate $M_X + \mathscr{A}^\circ Y + \mathscr{B}^\circ Z$.) Similarly we shall call $\mathscr{A}^\circ Y_i + \mathscr{B}^\circ Z_j$ (or $M_X + \mathscr{A}^\circ Y_i + \mathscr{B}^\circ Z_j$) the optimum estimate of X_k conditioned by Y_i, Z_j and denote $\hat{X}_{k|i,j} = \hat{X}_{k|i} + \hat{X}_{k|j}$.

Corollary 2. Let $Y \otimes Y^*, I \otimes I^* > 0$, where innovation $I = Z - (Z \otimes Y^*) . (Y \otimes Y^*)^{-1} Y$. Then: (i) $\operatorname{tr}(Y \otimes I^*) = 0$, (ii) $\|X - \mathscr{A}^\circ Y - \mathscr{B}^\circ Z\|^2 = \|X - \mathscr{C}^\circ Y - \mathscr{D}^\circ I\|^2$.

Proof. $I \otimes Y^* = Z \otimes Y^* - (Z \otimes Y^*)(Y \otimes Y^*)^{-1}(Y \otimes Y^*) = 0$, so it holds (i) and for minimization of $\|X - \mathscr{C}Y - \mathscr{D}I\|^2$, Lemma can be used. Further let us notice that both Y, Z and Y, I span the same linear space (innovation de facto took its origin as the 2nd step of Gram-Schmidt orthogonalization) so it holds (ii) and instead of the minimization of $\|X - \mathscr{A}Y - \mathscr{B}Z\|^2$, we can concern ourselves with the simpler minimization of $\|X - \mathscr{C}Y - \mathscr{D}I\|^2 : \mathscr{C}^\circ = (X \otimes Y^*)(Y \otimes Y^*)^{-1}$, $\mathscr{D}^\circ = (\tilde{X}_{|Y} \otimes I^*)(I \otimes I^*)^{-1}$, $\tilde{X}_{|Y} =_{df} X - \mathscr{C}^\circ Y$, $\|\tilde{X}_{|Y} - \mathscr{D}^\circ I\|^2 = \operatorname{tr}\{\tilde{X}_{|Y} \otimes \tilde{X}^*_{|Y} - \mathscr{D}^\circ(I \otimes \tilde{X}^*_{|Y})\}$.

Theorem 1 (One-step prediction). ∃! solution of one-step prediction problem. Further

$$\hat{X}_{k+1|k} = \mathscr{A}_k \hat{X}_{k|k-1} + \mathscr{B}_k U_k + \mathscr{L}_k I_k, \quad \hat{X}_{0|-1} = M \tag{2}$$

where the gain $\mathscr{L}_k = \mathscr{A}_k \mathscr{S}_{k|k-1} \mathscr{C}_k^* (\mathscr{C}_k \mathscr{S}_{k|k-1} \mathscr{C}_k^* + \mathscr{V}_k)^{-1}$, the innovation $I_k = Z_k - \mathscr{C}_k \hat{X}_{k|k-1}$, the one-step prediction error $X_{k+1} - \hat{X}_{k+1|k} =_{df} \tilde{X}_{k+1|k}$, and its covariance $\operatorname{cov}[\tilde{X}_{k+1|k}, \ldots] =_{df} \mathscr{S}_{k+1|k}$ is

$$\mathscr{S}_{k+1|k} = \mathscr{A}_k \mathscr{S}_{k|k-1} \mathscr{A}_k^* + \mathscr{D}_k \mathscr{W}_k \mathscr{D}_k^* - \mathscr{L}_k \mathscr{C}_k \mathscr{S}_{k|k-1} \mathscr{A}_k^*, \ \mathscr{S}_{0|-1} = \mathscr{S} . \tag{3}$$

The value of the criterion is $\operatorname{tr} \mathscr{S}_{k+1|k}$.

Proof. $\hat{X}_{k+1|k}$ from (2) is an affine map of $\mathbf{Z}_k$, we shall convince ourselves that $\mathscr{A}_k$, $\mathscr{L}_k$ optimize the one-step prediction criterion. We shall use isomorfisms between Corollary 2, Note 4 and one-step prediction problem: $X \cong X_{k+1}$, $Y \cong \mathbf{Z}_{k-1}$, $Z \cong Z_k$, $I \cong I_k = Z_k - \hat{Z}_{k|k-1} =_{df} \tilde{Z}_{k|k-1}$, $\mathscr{D}^\circ \cong \mathscr{L}_k$. Then $\hat{X}_{k+1|k} = \hat{X}_{k+1|k-1} + \mathscr{L}_k \tilde{Z}_{k|k-1}$ where $\hat{X}_{k+1|k-1} = (\mathscr{A}_k X_k + \mathscr{B}_k U_k + \mathscr{D}_k W_k)\hat{}_{|k-1} = \mathscr{A}_k X_{k|k-1} + \mathscr{B}_k U_k$, when we used Corollary 1 and Note 4. Further $\mathscr{L}_k = \operatorname{cov}[\tilde{X}_{k+1|k-1}, \tilde{Z}_{k|k-1}] \operatorname{cov}^{-1} . [\tilde{Z}_{k|k-1}, \ldots] = \operatorname{cov}[\mathscr{A}_k X_k + \mathscr{B}_k U_k + \mathscr{D}_k W_k - \mathscr{A}_k \hat{X}_{k|k-1} - \mathscr{B}_k U_k, \mathscr{C}_k \tilde{X}_{k|k-1} + V_k] . \operatorname{cov}^{-1}[\mathscr{C}_k \tilde{X}_{k|k-1} + V_k, \ldots] = \operatorname{cov}[\mathscr{A}_k \tilde{X}_{k|k-1} + \mathscr{D}_k W_k, \mathscr{C}_k \tilde{X}_{k|k-1} + V_k] . (\mathscr{C}_k \mathscr{S}_{k|k-1} . \mathscr{C}_k^* + \mathscr{V}_k)^{-1} = \mathscr{A}_k \mathscr{S}_{k|k-1} \mathscr{C}_k^* (\mathscr{C}_k \mathscr{S}_{k|k-1} \mathscr{C}_k^* + \mathscr{V}_k)^{-1}$ where $\mathscr{S}_{k|k-1} \geqq 0$, $\mathscr{V}_k > 0$ so that ∃! $(\mathscr{C}_k \mathscr{S}_{k|k-1} \mathscr{C}_k^* + \mathscr{V}_k)^{-1}$. Finally $\mathscr{S}_{k+1|k} = \operatorname{cov}[\tilde{X}_{k+1|k-1}, \ldots] - \mathscr{L}_k \operatorname{cov}[\tilde{Z}_{k|k-1}, \tilde{X}_{k+1|k-1}] = \operatorname{cov}[\mathscr{A}_k \tilde{X}_{k|k-1} + \mathscr{D}_k W_k, \ldots] - \mathscr{L}_k \operatorname{cov}[\mathscr{C}_k \tilde{X}_{k|k-1} + V_k, \mathscr{A}_k \tilde{X}_{k|k-1} + \mathscr{D}_k W_k] = \mathscr{A}_k \mathscr{S}_{k|k-1} \mathscr{A}_k^* + \mathscr{D}_k \mathscr{W}_k \mathscr{D}_k^* - \mathscr{L}_k \mathscr{C}_k \mathscr{S}_{k|k-1} \mathscr{A}_k^*$. By this we have proved that if (2, 3) hold for $k - 1$ then (2, 3) hold for $k (k > 0)$. For $k = 0$ we use $\mathbf{Z}_{-1} = \{M, \mathscr{S}\}$, and (2, 3) follow from Lemma and Note 4. Finally $\mathrm{E}\{\|X_{k+1} - \hat{X}_{k+1|k}\|^2 | \mathbf{Z}_k\} = \mathrm{E}\{\tilde{X}_{k+1|k} \otimes \tilde{X}^*_{k+1|k} \mid \mathbf{Z}_k\} = \operatorname{tr} \mathscr{S}_{k+1|k}$.

Theorem 2 (n-step prediction). ∃! solution of n-step prediction problem. Further

$$\hat{X}_{k+n|k} = \mathscr{F}_{k+n,k+1} \hat{X}_{k+1|k} + \textstyle\sum_{(j)} \mathscr{F}_{k+n,j+1} \mathscr{B}_j U_j ,$$

$j = k + 1, \ldots, k + n - 1$, $\mathscr{F}_{m,i} = \mathscr{A}_{m-1} \mathscr{A}_{m-2}, \ldots, \mathscr{A}_i$, $\mathscr{F}_{i,i} = 1_{R^a \times R^b}$ and $\hat{X}_{k+1|k}$ is one-step prediction from Theorem 1. The n-step prediction error $X_{k+n} - \hat{X}_{k+n|k} =_{df} =_{df} \tilde{X}_{k+n|k}$, its covariance $\operatorname{cov}[\tilde{X}_{k+n|k}, \ldots] =_{df} \mathscr{S}_{k+n|k}$ is

$$\mathscr{S}_{k+n|k} = \mathscr{F}_{k+n,k+1} \mathscr{S}_{k+1|k} \mathscr{F}^*_{k+n,k+1} + \textstyle\sum_{(j)} \mathscr{F}_{k+n,j+1} \mathscr{D}_j \mathscr{W}_j \mathscr{D}_j^* \mathscr{F}^*_{k+n,j+1}$$

where $\mathscr{S}_{k+1|k}$ is the one-step prediction covariance-error from Theorem 1. The value of the criterion is $\operatorname{tr} \mathscr{S}_{k+n|k}$.

Proof. We apply Corollary 1 and the definition of $\mathcal{S}_{k+n|k}$ directly to $X_{k+n} = \mathcal{F}_{k+n,k+1}X_{k+1} + \sum_{(j)}\mathcal{F}_{k+n,j+1}(\mathcal{B}_j U_j + \mathcal{D}_j W_j)$.

Theorem 3 (Filtering). ∃! solution of the filtering problem. Further

$$(4) \qquad \hat{X}_{k|k} = \hat{X}_{k|k-1} + \mathcal{L}_k I_k\,, \quad \hat{X}_{0|0} = M + \mathcal{L}_0 I_0$$

where one-step prediction based on the old filtered value

$$(5) \qquad \hat{X}_{k|k-1} = \mathcal{A}_{k-1}\hat{X}_{k-1|k-1} + \mathcal{B}_{k-1}U_{k-1}\,, \quad \hat{X}_{0|-1} = M\,,$$

the gain $\mathcal{L}_k = \mathcal{S}_{k|k-1}\mathcal{C}_k^*(\mathcal{C}_k\mathcal{S}_{k|k-1}\mathcal{C}_k^* + \mathcal{V}_k)^{-1}$, $\mathcal{L}_0 = \mathcal{S}\mathcal{C}_0^*(\mathcal{C}_0\mathcal{S}\mathcal{C}_0^* + \mathcal{V}_0)^{-1}$; the innovation $I_k = Z_k - \mathcal{C}_k\hat{X}_{k|k-1}$, $I_0 = Z_0 - \mathcal{C}_0 M$; the filtering error $X_k - \hat{X}_{k|k} =_{\text{df}} =_{\text{df}} \tilde{X}_{k|k}$, its covariance, $\text{cov}\,[\tilde{X}_{k|k}, \ldots] =_{\text{df}} \mathcal{S}_{k|k}$, is

$$(6) \qquad \mathcal{S}_{k|k} = \mathcal{S}_{k|k-1} - \mathcal{L}_k\mathcal{C}_k\mathcal{S}_{k|k-1}\,, \quad \mathcal{S}_{0|0} = \mathcal{S} - \mathcal{L}_0\mathcal{C}_0\mathcal{S}$$

where the error of one-step prediction based on the old filtered value $X_k - \hat{X}_{k|k-1} =_{\text{df}} =_{\text{df}} \tilde{X}_{k|k-1}$ and its covariance, $\text{cov}\,[\tilde{X}_{k|k-1}, \ldots] =_{\text{df}} \mathcal{S}_{k|k-1}$, is

$$(7) \qquad \mathcal{S}_{k|k-1} = \mathcal{A}_{k-1}\mathcal{S}_{k-1|k-1}\mathcal{A}_{k-1}^* + \mathcal{D}_{k-1}\mathcal{W}_{k-1}\mathcal{D}_{k-1}^*\,, \quad \mathcal{S}_{0|-1} = \mathcal{S}\,.$$

The value of the criterion is $\text{tr}\, \mathcal{S}_{k|k}$.

Proof. $\hat{X}_{k|k}$ from (4) is an affine map of $\mathbf{Z}_k$, we shall convince ourselves that $\mathcal{A}_{k-1}$, $\mathcal{L}_k$ optimize the filtering criterion. We shall use isomorfisms between Corollary 2, Note 4 and filtering problem: $X \cong X_k$, $Y \cong \mathbf{Z}_{k-1}$, $Z \cong Z_k$, $I \cong I_k = \tilde{Z}_{k|k-1}$, $\mathcal{D}^\circ \cong \mathcal{L}_k$. Then $\hat{X}_{k|k} = \hat{X}_{k|k-1} + \mathcal{L}_k I_k$, where $\hat{X}_{k|k-1} = (\mathcal{A}_{k-1}X_{k-1} + \mathcal{B}_{k-1}U_{k-1} + \mathcal{D}_{k-1}W_{k-1})\hat{}_{|k-1} = \mathcal{A}_{k-1}X_{k-1|k-1} + \mathcal{B}_{k-1}U_{k-1}$. Further $\mathcal{L}_k = \text{cov}\,[\tilde{X}_{k|k-1}, \tilde{Z}_{k|k-1}]\,.\,\text{cov}^{-1}\,[\tilde{Z}_{k|k-1}, \ldots] = \text{cov}\,[\tilde{X}_{k|k-1}, \mathcal{C}_k\tilde{X}_{k|k-1} + V_k]\,.\,\text{cov}^{-1}\,[\mathcal{C}_k\tilde{X}_{k|k-1} + V_k, \ldots] = \mathcal{S}_{k|k-1}\mathcal{C}_k^*(\mathcal{C}_k\mathcal{S}_{k|k-1}\mathcal{C}_k^* + \mathcal{V}_k)^{-1}$, where $\mathcal{S}_{k|k-1} \geqq 0$, $\mathcal{V}_k > 0$ so that ∃! $(\mathcal{C}_k\mathcal{S}_{k|k-1}\mathcal{C}_k^* + \mathcal{V}_k)^{-1}$. Finally $\mathcal{S}_{k|k} = \text{cov}\,[\tilde{X}_{k|k-1}, \ldots] - \mathcal{L}_k\,\text{cov}\,[\tilde{Z}_{k|k-1}, \tilde{X}_{k|k-1}] = \text{cov}\,[\mathcal{A}_{k-1}X_{k-1} + \mathcal{B}_{k-1}U_{k-1} + \mathcal{D}_{k-1}W_{k-1} - \mathcal{A}_{k-1}X_{k-1|k-1} - \mathcal{B}_{k-1}\,.\,U_{k-1}, \ldots] - \mathcal{L}_k\,\text{cov}\,[\mathcal{C}_k\tilde{X}_{k|k-1} + V_k, \tilde{X}_{k|k-1}] = \text{cov}\,[\mathcal{A}_{k-1}\tilde{X}_{k-1|k-1} + \mathcal{D}_{k-1}\,.\,W_{k-1}, \ldots] - \mathcal{L}_k\mathcal{C}_k\mathcal{S}_{k|k-1} = \mathcal{A}_{k-1}\mathcal{S}_{k-1|k-1}\mathcal{A}_{k-1}^* + \mathcal{D}_{k-1}\mathcal{W}_{k-1}\mathcal{D}_{k-1}^* - \mathcal{L}_k\mathcal{C}_k\,.\,\mathcal{S}_{k|k-1} = \mathcal{S}_{k|k-1} - \mathcal{L}_k\mathcal{C}_k\,\mathcal{S}_{k|k-1}$. By this we have proved that if (4, …, 7) hold for $k - 1$ then (4, …, 7) hold for k ($k > 0$). For $k = 0$ we use $\mathbf{Z}_{-1}$, and (4, …, 7) follow from Lemma and Note 4.

Theorem 4 (Control). ∃! solution of control problem. Further

$$U_k^\circ = \mathcal{K}_k X_k$$

where the gain $\mathscr{K}_k = -(\mathscr{B}_k^* \mathscr{P}_{k+1} \mathscr{B}_k + \mathscr{R}_k)^{-1} \mathscr{B}_k^* \mathscr{P}_{k+1} \mathscr{A}_k$,

$$\mathscr{P}_k = (\mathscr{A}_k + \mathscr{B}_k \mathscr{K}_k)^* \mathscr{P}_{k+1} (\mathscr{A}_k + \mathscr{B}_k \mathscr{K}_k) + \mathscr{Q}_k + \mathscr{K}_k^* \mathscr{R}_k \mathscr{K}_k, \quad \mathscr{P}_N = \mathscr{P}.$$

The value of the criterion is $\|X_k\|^2_{\mathscr{P}_k}$.

Proof. From the criterion definition we shall succesively obtain: $J(\mathbf{U}_k^\circ) = \min_{\mathbf{U}_k} \{\|X_N\|^2_{\mathscr{P}} + \sum_{(i)} \|X_i\|^2_{\mathscr{Q}_i} + \|U_i\|^2_{\mathscr{R}_i}\} = \min_{U_k} \min_{\mathbf{U}_{k+1}} \{\|X_k\|^2_{\mathscr{Q}_k} + \|U_k\|^2_{\mathscr{R}_k} + J(\mathbf{U}_{k+1})\} = \min_{U_k} \{\|X_k\|^2_{\mathscr{Q}_k} + \|U_k\|^2_{\mathscr{R}_k} + J(\mathbf{U}_{k+1}^\circ)\}$ for $i = k, \ldots, N-1$, $k < N-1$, respectively $\min_{U_k} \{\|X_k\|^2_{\mathscr{Q}_k} + \|U_k\|^2_{\mathscr{R}_k} + \|X_N\|^2_{\mathscr{P}}\}$ for $k = N-1$. We shall start solving the recursive equation for U_k at time $N-1$. We shall minimize the criterion $J(U_{N-1}) = \|X_N\|^2_{\mathscr{P}} + \|X_{N-1}\|^2_{\mathscr{Q}_{N-1}} + \|U_{N-1}\|^2_{\mathscr{R}_{N-1}} = \operatorname{tr} \{(\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1})^* \mathscr{P} \,.\, (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1}) + X_{N-1}^* \mathscr{Q}_{N-1} X_{N-1} + U_{N-1}^* \mathscr{R}_{N-1} U_{N-1}\}$. We shall find the stationary point of our criterion from the Gateaux differential: $\delta J(U_{N-1}; H) = \lim_{\alpha \to 0} \alpha^{-1} \{[A_{N-1} X_{N-1} + \mathscr{B}_{N-1}(U_{N-1} + \alpha H)]^* \mathscr{P} [\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1}(U_{N-1} + \alpha H)] - (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1})^* \mathscr{P} (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1}) + X_{N-1}^* \,.\, \mathscr{Q}_{N-1} X_{N-1} - X_{N-1}^* \mathscr{Q}_{N-1} X_{N-1} + (U_{N-1} + \alpha H)^* \mathscr{R}_{N-1} (U_{N-1} + \alpha H) - U_{N-1}^* \,.\, \mathscr{R}_{N-1} U_{N-1}\} = 2 \operatorname{tr} \{(\mathscr{B}_{N-1} H)^* \mathscr{P} (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1}) + H^* \mathscr{R}_{N-1} U_{N-1}\} = 2 \operatorname{tr} \{H^* [\mathscr{B}_{N-1}^* \mathscr{P} (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1}) + \mathscr{R}_{N-2} U_{N-1}]\} = 0 \ \forall H$. Further $\mathscr{B}_{N-1}^* \mathscr{P} \mathscr{A}_{N-1} X_{N-1} + (\mathscr{B}_{N-1}^* \mathscr{P} \mathscr{B}_{N-1} + \mathscr{R}_{N-1}) U_{N-1} = 0$. $\mathscr{P} \geqq 0$, $\mathscr{R}_{N-1} > 0$ guarantee that $\exists!\ (\mathscr{B}_{N-1}^* \mathscr{P} \mathscr{B}_{N-1} + \mathscr{R}_{N-1})^{-1}$, so the unique stationary point: $U_{N-1}^\circ = -(\mathscr{B}_{N-1}^* \mathscr{P} \mathscr{B}_{N-1} + \mathscr{R}_{N-1})^{-1} \mathscr{B}_{N-1}^* \mathscr{P} \mathscr{A}_{N-1} X_{N-1} =_{\mathrm{df}} \mathscr{K}_{N-1} X_{N-1}$. Then $J(U_{N-1}^\circ) = \operatorname{tr} \{[(\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) X_{N-1}]^* \mathscr{P} [(\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) X_{N-1}] + X_{N-1}^* \mathscr{Q}_{N-1} X_{N-1} + (\mathscr{K}_{N-1} X_{N-1})^* \mathscr{R}_{N-1} (\mathscr{K}_{N-1} X_{N-1})\} = \operatorname{tr} \{X_{N-1}^* [(\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1})^* \mathscr{P} (\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) + \mathscr{Q}_{N-1} + \mathscr{K}_{N-1}^* \mathscr{R}_{N-1} \mathscr{K}_{N-1}] X_{N-1}\} = \|X_{N-1}\|^2_{\mathscr{P}_{N-1}}$ where $\mathscr{P}_{N-1} =_{\mathrm{df}} (\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1})^* \mathscr{P} (\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) + \mathscr{Q}_{N-1} + \mathscr{K}_{N-1}^* \mathscr{R}_{N-1} \mathscr{K}_{N-1}$. $\mathscr{P}, \mathscr{Q}_{N-1}, \mathscr{R}_{N-1} > 0$ guarantee that $\mathscr{P}_{N-1} > 0$ and the stationary point is the unique minimum. If we introduce $\mathscr{P}_N = \mathscr{P}$ and further the maps $N-1 \mapsto k$, $N \mapsto k+1$ then from the premise that Theorem 4 holds for $k+1$ we obtain that it holds for k ($k \geqq 0$) and thus we obtain the assertion of Theorem 4.

Theorem 5 (Stochastic control). $\exists!$ solution of stochastic control problem. Further U_k°, $\mathscr{K}_k$, $\mathscr{P}_k$ ($k \leqq N-1$), $\mathscr{P}_N$ are given by the same relations as in control problem. The value of the criterion is $\|X_k\|^2_{\mathscr{P}_k} + \sum_{(i)} \|\mathscr{W}_i\|^2_{\mathscr{D}_i^* \mathscr{P}_{i+1} \mathscr{D}_i}$ where $i = k, \ldots, N-1$.

Proof. From the criterion definition: $\mathrm{E}\, J(\mathbf{U}_k^\circ) = \min_{U_k} \mathrm{E}\{\|X_k\|^2_{\mathscr{Q}_k} + \|U_k\|^2_{\mathscr{R}_k} + \mathrm{E}\, J(\mathbf{U}_{k+1}^\circ)\}$ for $k < N-1$, respectively $\min_{U_k} \mathrm{E}\{\|X_k\|^2_{\mathscr{Q}_k} + \|U_k\|^2_{\mathscr{R}_k} + \|X_N\|^2_{\mathscr{P}}\}$ for $k = N-1$. Further $\mathrm{E}\, J(\mathbf{U}_{N-1}) = \mathrm{E} \operatorname{tr} \{(\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1} + \mathscr{D}_{N-1} \,.$

$\cdot W_{N-1})^* \mathscr{P} (\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1} + \mathscr{D}_{N-1} W_{N-1}) + X^*_{N-1} \mathscr{Q}_{N-1} X_{N-1} + U^*_{N-1} \mathscr{R}_{N-1} U_{N-1}\}$. The Gateaux differential: $\delta \, \mathrm{E} \, J(U_{N-1}; H) = 2 \, \mathrm{E} \, \mathrm{tr} \{H^* \cdot [\mathscr{B}^*_{N-1} \mathscr{P}(\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1} + \mathscr{D}_{N-1} W_{N-1}) + \mathscr{R}_{N-1} U_{N-1}]\} = 2 \, \mathrm{tr} \{H^* 0 \cdot [\mathscr{B}^*_{N-1} \mathscr{P}(\mathscr{A}_{N-1} X_{N-1} + \mathscr{B}_{N-1} U_{N-1}) + \mathscr{R}_{N-1} U_{N-1}]\} = 0 \; \forall H$. $\mathscr{P}, \mathscr{R}_{N-1} > ..$ guarantee that $\exists!$ stationary point: $U^\circ_{N-1} = -(\mathscr{B}^*_{N-1} \mathscr{P} \mathscr{B}_{N-1} + \mathscr{R}_{N-1})^{-1} \cdot \mathscr{B}^*_{N-1} \mathscr{P} \mathscr{A}_{N-1} X_{N-1} =_{\mathrm{df}} \mathscr{K}_{N-1} X_{N-1}$. Then $\mathrm{E} \, J(U^\circ_{N-1}) = \mathrm{E} \, \mathrm{tr} \{[(\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \cdot \mathscr{K}_{N-1}) X_{N-1}]^* \mathscr{P} [(\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) X_{N-1}] + (\mathscr{D}_{N-1} W_{N-1})^* \mathscr{P} \mathscr{D}_{N-1} W_{N-1} + X^*_{N-1} \mathscr{Q}_{N-1} X_{N-1} + (\mathscr{K}_{N-1} X_{N-1})^* \mathscr{R}_{N-1} (\mathscr{K}_{N-1} X_{N-1})\} = \|X_{N-1}\|^2_{\mathscr{P}_{N-1}} + \|\mathscr{W}_{N-1}\|^2_{\mathscr{D}^*_{N-1} \mathscr{P} \mathscr{D}_{N-1}}$ where $\mathscr{P}_{N-1} =_{\mathrm{df}} (\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1})^* \mathscr{P} (\mathscr{A}_{N-1} + \mathscr{B}_{N-1} \mathscr{K}_{N-1}) + \mathscr{Q}_{N-1} + \mathscr{K}^*_{N-1} \mathscr{R}_{N-1} \mathscr{K}_{N-1}$. $\mathscr{P}, \mathscr{Q}_{N-1}, \mathscr{R}_{N-1} > 0$ guarantee that $\mathscr{P}_{N-1} > 0$ and the stationary point is the unique minimum. Introducing $\mathscr{P}_N = \mathscr{P}$ and the maps $N - 1 \mapsto k$, $N \mapsto k + 1$, the assertion of the theorem already follows.

Theorem 6 (Stochastic control/filtering. Separability principle). $\exists!$ solution of stochastic control/filtering problem. Further

$$U^\circ_k = \mathscr{K}_k \hat{X}_{k|k}$$

where for $\hat{X}_{k|k}, \mathscr{L}_k, I_k, \mathscr{S}_{k|k}$ $(k \leqq N - 1)$, $\mathscr{S}_{N|N}$ hold the relations from filtering problem and for $\mathscr{K}_k, \mathscr{P}_k$ $(k \leqq N - 1)$, $\mathscr{P}_N$ hold the relations from control problem. The value of the criterion is $\|\hat{X}_{k|k}\|^2_{\mathscr{P}_k} + \sum_{(i)} \|\mathscr{I}_{i+1}\|^2_{\mathscr{L}^*_{i+1} \mathscr{P}_{i+1} \mathscr{L}_{i+1}} + \|\mathscr{S}_{i|i}\|^2_{\mathscr{Q}_i} + \|\mathscr{S}_{N|N}\|^2_{\mathscr{P}}$ where $i = k, \ldots, N - 1$ and the innovation covariance $\mathrm{cov}\,[I_i, \ldots] =_{\mathrm{df}} =_{\mathrm{df}} \mathscr{I}_i = \mathscr{C}_i (\mathscr{A}_{i-1} \mathscr{S}_{i-1|i-1} \mathscr{A}^*_{i-1} + \mathscr{D}_{i-1} \mathscr{W}_{i-1} \mathscr{D}^*_{i-1}) \mathscr{C}^*_i$.

Proof. We shall use isomorfisms between state model (1) and filter (4; 5): $X_k \cong \tilde{X}_{k|k}, \mathscr{D}_k \cong \mathscr{L}_{k+1}, W_k \cong I_{k+1}, \mathscr{W}_k \cong \mathscr{I}_{k+1} = \mathrm{cov}\,[\mathscr{C}_{k+1}(\mathscr{A}_k X_k + \mathscr{B}_k U_k + \mathscr{D}_k W_k) + V_{k+1} - \mathscr{C}_{k+1}(\mathscr{A}_k \hat{X}_{k|k} + \mathscr{B}_{k+1} U_{k+1}), \ldots] = \mathrm{cov}\,[\mathscr{C}_{k+1}(\mathscr{A}_k \tilde{X}_{k|k} + \mathscr{D}_k W_k) + V_{k+1}, \ldots] = \mathscr{C}_{k+1}(\mathscr{A}_k \mathscr{S}_{k|k} \mathscr{A}^*_k + \mathscr{D}_k \mathscr{W}_k \mathscr{D}^*_k) \mathscr{C}^*_{k+1} + \mathscr{V}_{k+1}$. Applying stochastic control problem to the model (4, 5) we directly obtain that the control from Theorem 5 with $\mathscr{K}_k, \mathscr{P}_k$ $(k \leqq N - 1)$, $\mathscr{P}_N$ preceded by filtering from Theorem 3 with $\hat{X}_{k|k}, \mathscr{L}_k, I_k, \mathscr{S}_{k|k}$ $(k \leqq N - 1)$, $\mathscr{S}_{N|N}$ solves uniquelly stochastic control/filtering problem. It remains only to obtain the criterion value. From the criterion definition: $\mathrm{E}\{J(\boldsymbol{U}^\circ_k) \mid \boldsymbol{Z}_k\} = \mathrm{E}\{\|\hat{X}_{N|N} + \tilde{X}_{N|N}\|^2_{\mathscr{P}} + \sum_{(i)} \|\hat{X}_{i|i} + \tilde{X}_{i|i}\|^2_{\mathscr{Q}_i} + \|U_i\|^2_{\mathscr{R}_i}\}$. From the expression of the optimal estimate given by Lemma and from its stochastic interpretation given by Note 2 follows that $\hat{X}_{i|i}$ and $\tilde{X}_{i|i}$ are mutually uncorrelated. So the criterion is $\mathrm{E}\{\|\hat{X}_{N|N}\|^2_{\mathscr{P}} + \|\tilde{X}_{N|N}\|^2_{\mathscr{P}} + \sum_{(i)} \|\hat{X}_{i|i}\|^2_{\mathscr{Q}_i} + \|\tilde{X}_{i|i}\|^2_{\mathscr{Q}_i} + \|U_i\|^2_{\mathscr{R}_i}\} = \mathrm{E}\{\|\hat{X}_{N|N}\|^2_{\mathscr{P}} + \sum_{(i)} \|\hat{X}_{i|i}\|^2_{\mathscr{Q}_i} + \|U_i\|^2_{\mathscr{R}_i}\} + \|\mathscr{S}_{i|i}\|^2_{\mathscr{Q}_i} + \|\mathscr{S}_{N|N}\|^2_{\mathscr{P}}$. Applying before mentioned isomorfisms to the value of the criterion from Theorem 5, we directly obtain that our criterion value is as asserted.

Case 1 (Covariance control). Covariance evolution of the model

$$x_{k+1} = A_k x_k + D_k w_k ,$$

$k = 0, \ldots, N-1$ where

$$\begin{bmatrix} x_0 \\ w_0 \\ \vdots \\ w_k \end{bmatrix} \sim N \left\{ \begin{bmatrix} m \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} S & & & \\ & W_0 & & \\ & & \ddots & \\ & & & W_k \end{bmatrix} \right\},$$

$S, W_0, \ldots, W_k \geqq 0$, is

$$S_{k+1} = A_k S_k A_k^* + D_k W_k D_k^* .$$

Using Note 2:

$$S_{k+1} = \mathscr{A}_k S_k + \mathscr{D}_k W_k$$

where $\mathscr{A}_k = A_k \otimes A_k$, $\mathscr{D}_k = D_k \otimes D_k$. Covariance S_k control problem: Find a linear map $\mathbf{S}_k \mapsto W_k$ minimizing $\|S_N\|^2_{\mathscr{P}} + \sum_{(k)} \|S_k\|^2_{\mathscr{Q}_k} + \|W_k\|^2_{\mathscr{R}_k}$ where $\mathscr{P}, \mathscr{Q}_{N-1}, \ldots, \mathscr{Q}_k, \mathscr{R}_{N-1}, \ldots, \mathscr{R}_k > 0$. Case 1 can be of importance for start-up control, stabilization, and shut-down control in the presence of the stochastic disturbances having manipulative covariance. Solution: through direct use of Theorem 5.

Case 2 (Coupled state-vector filtering and gain-matrix continuously running identification). State equation:

$$x_{k+1} = A_{1,k} x_k + B_k u_k + D_k w_k , \quad x_k \in R^n ,$$

output equation:

$$z_k = C_k x_k + v_k , \quad C_k : R^n \to R^s .$$

State model of gain-matrix evolution:

$$B_{k+1} = A_{2,k} x_k a_k^* + \mathscr{A}_k B_k + \mathscr{D}_k W_k ,$$

output model of gain-matrix evolution:

$$Z_k = \mathscr{C}_k B_k + V_k , \quad Z_k \in R^s \times R^p .$$

Using Note 2 we first notice that $B_k u_k = (1_{R^n} \otimes u_k^*) B_k$; $A_{2,k} x_k a_k^* = (A_{2,k} \otimes a_k) x_k$. Then coupled state equation:

$$[x_{k+1} \ B_{k+1}] = \begin{bmatrix} A_{1,k} & 1_{R^n} \otimes u_k^* \\ A_{2,k} \otimes a_k & \mathscr{A}_k \end{bmatrix} [x_k \ B_k] + \begin{bmatrix} D_k & 0 \\ 0 & \mathscr{D}_k \end{bmatrix} [w_k \ W_k] ,$$

 coupled output equation:

$$[z_k \; Z_k] = \begin{bmatrix} C_k & 0 \\ 0 & \mathscr{C}_k \end{bmatrix} [x_k \; B_k] + [v_k \; V_k] .$$

For the interpretation of the partitioned tensors see Note 1! With application of a reduced Case 2 we were concerned in [12]. Solution: through direct application of Theorem 3.

Case 3 (Control, resp. continuously-running identification of nonstationary multivariable stochastic regression model). Let the model external description be:

$$z_{k+1} = \sum_{(j)} A_{j,k} z_{k-j+1} + B_{j,k} u_{k-j+1} + e_k$$

where $j = 1, \ldots, n$; further where $u_{(\cdot)} \in R^r$ is the measured (known) input and $z_{(\cdot)} \in R^r$ is the measured (known) output, and $e_k \sim \mathrm{N}(0, E_k)$. $E_k \geqq 0$ for the control, $E_k > 0$ for the identification.

At first we shall be concerned with the control. For the control-model internal description, the control-state $X_k = [u_{k-n+1}, \ldots, u_{k-1}, z_{k-n+1}, \ldots, z_k]$, further $U_k = u_k$, $W_k = e_k$. Finally the control-state equation:

$$X_{k+1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1_{R^{(n-1)r}} \\ B_{n,k} & B_{n-1,k}, \ldots, B_{2,k} & A_{n,k} & A_{n-1,k}, \ldots, A_{1,k} \end{bmatrix} X_k +$$

$$+ \begin{bmatrix} 0 \\ 1_{R^r} \\ 0 \\ B_{1,k} \end{bmatrix} U_k + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1_{R^r} \end{bmatrix} W_k .$$

For the interpretation of the partitioned tensors see Note 1.

Now we shall be concerned with the identification. For the identification-model internal description, the identification-state $X_k = [A_{1,k}, \ldots, A_{n,k}, B_{1,k}, \ldots, B_{n,k}]$, the identification-output $Z_k = z_{k+1}$, further $V_k = e_{k+1}$. Finally the identification-output equation:

$$Z_k = \{1_{R^r} \otimes [z_k^*, \ldots, z_{k-n+1}^*, u_k^*, \ldots, u_{k-n+1}^*]\} X_k + V_k,$$

and the identification-state equation which models the parameters evolution is:

$$X_{k+1} = \mathscr{A}_k X_k + \mathscr{D}_k W_k . \tag{8}$$

Applications of Case 3 follow from derivation and explication of the multivariable stochastic regression model, [11]. Solution of control-case: through direct application of Theorem 5. Solution of identification-case: through direct application of Theorem 3.

An identification part of Case 3 can easily be generalized. Let the internal description be: $x_{k+1} = A_k x_k + B_k u_k + D_k w_k$ where $u_{(\cdot)}$, and $x_{(\cdot)} \in R^n$ are measured (known) and $w_k \sim \mathrm{N}(0, W_k)$, $D_{k+1} W_{k+1} D_{k+1}^* > 0$. We introduce the state: $X_k = [A_k, B_k]$, output $Z_k = x_{k+1}$, measurement disturbance $V_k = D_{k+1} w_{k+1}$. Then output equation: $Z_k = \{1_{R^n} \otimes [x_k^*, u_k^*]\} X_k + V_k$. (8) is again the state equation. Applications can be of importance in area of electric actuators where the state is usually measurable. Solution: through direct application of Theorem 3.

CONCLUSION

There exist over 10^3 contributions concerned with LQG problem of estimation and control. Nevertheless we have been convinced that its version concerned with the nonstationary internal description has not been closed yet. This is because of not fully digged algebraic structure of the mentioned version of LQG problem.

Comparisons of the solution of LQG problem in the tensor space with its solution in the vector space, [7, 8, 5, 1; 2], suggest that our solution is a tensor realization of the solution in the abstract linear space. Nevertheless, in all contributions we are aware of, the solutions has been hitherto both derived and interpreted as the solutions in a concrete vector space, i.e. with maps taken to be matrices and not abstract linear spaces homomorfisms. This tensor realization follows also from the fact that during derivation it was virtually sufficient for us to consider the 2nd order tensors as the elements of the linear space and not of the linear associative algebra. Because of the mentioned correspondence and because of the space limitations, we fully passed the analysis of the properties of our LQG problem in the tensor space.

(Received June 14, 1974.)

REFERENCES

[1] K. J. Åström, R. W. Koepcke, F. Tung: On the control of linear discrete dynamic systems with quadratic loss. Res. Rep. RJ-222, IBM, San Jose Res. Lab., San Jose, Calif., 1962.

[2] M. Athans (Ed.): Linear-quadratic-Gaussian problem in estimation and control. Dedicated to prof. R. E. Kalman. IEEE Trans. Aut. Contr. *AC-16* (Dec. 1971), 6, 529—869.

[3] R. Bellman: Dynamic programming. Princeton Univ. Press, Princeton 1957.

[4] N. Bourbaki: Algèbre. Livre II. Hermann, Paris 1948.

[5] P. D. Joseph, J. T. Tou: On linear control theory. AIEE Trans. (Appl. Indust.) *80* (Sept. 1961), 193—196.

[6] E. Hille, R. S. Phillips: Functional analysis and semigroups. Amer. Math. Soc. Colloq. Publ. *XXXI*. Rev. ed., Providence 1957.

[7] R. E. Kalman, R. W. Koepcke: Optimal synthesis of linear sampling control system using generalized performance indexes. Trans. ASME *80* (Nov. 1958), 1820—1826.

[8] R. E. Kalman: A new approach to linear filtering and prediction problems. Trans. ASME, Ser. D, J. Bas. Eng. *83* (March 1960), 35—45.

[9] S. MacLane, G. Birkhoff: Algebra, Macmillan, New York 1967.

[10] D. G. Luenberger: Optimization by vector space methods. Wiley, New York 1969.

[11] V. Peterka: Mathematical models for digital control of multivariable technological processes. (In Czech.) Res. Rep. No. 540, ÚTIA ČSAV, Prague 1974.

[12] A. Vaněček: Coupled air/fuel dynamic optimization and temperature regulation. A case study in stochastic control. Preprints IFAC Symp. Stochastic Control, Budapest, Sept. 1974, 573—580.

Ing. Antonín Vaněček, CSc.; Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 180 76 Praha 8. Czechoslovakia.