6.1. Solvable linear algebraic systems. Probabilistic interpretation. Let $A$ be a quadratic matrix (not obligatory nonsingular). Consider a system of solvable linear algebraic equation

$$Ax = y,$$  \hspace{1cm} (6.1)

that is under singular matrix $A$ system (6.1) possesses a solution. It is well known that all solutions of (6.1) under the singular matrix $A$ are described as follows:

$$x = A^\oplus y + (I - A^\oplus A)h,$$  \hspace{1cm} (6.2)

where $A^\oplus$ is the Moore-Penrose pseudoinverse matrix of $A$ and $I$ is the unite matrix. The vectors $A^\oplus y$ and $(I - A^\oplus A)h$ are orthogonal in a sense their inner product is equal to zero. In fact, since $A^\oplus A$ is symmetric matrix and $A^\oplus AA^\oplus = A^\oplus$ (see Section 5.2 in Lect. 5) we have

$$\langle A^\oplus y, (I - A^\oplus A)h \rangle = h^T (I - A^\oplus A)^T A^\oplus y = h^T (I - A^\oplus A)A^\oplus y = h^T (A^\oplus - A^\oplus AA^\oplus) y = 0.$$  

Hence, owing to an arbitrariness of $h$, the solution $A^\oplus y$ has the minimal Euclidean norm. Without of a prior information, it is impossible to find a solution corresponding to a specific $h$ (unless $A$ is nonsingular).

We give now a probabilistic description of the solution $x^\circ = A^\oplus y$. Let $X$ be a random vector with $E X = 0$ and $\text{Cov}(X, X) = I$ (here $I$ is the unite matrix) and $Y = AX$. Assume the random vector $Y$ is observable (measurable), the matrix $A$ is known and it is required to find the best estimate of $X$. Clearly that as such estimate

$$\hat{X} = \hat{\mathbf{E}}(X|Y)$$

can be taken. Notice that $\text{Cov}(X, Y) = \text{Cov}(X, X)A^T = A^T$ and $\text{Cov}(Y, Y) = A \text{Cov}(X, X)A^T = AA^T$. Therefore, by Theorem 5.1 (Lect. 5), $\hat{X} = A^T(AA^T)^\oplus$. On the other hand, since $A^T(AA^T)^\oplus = A^\oplus$ (see, Section 5.2, (6)), we have

$$\hat{X} = A^\oplus Y.$$  \hspace{1cm} (6.3)

Consider now the solution $x^\circ = A^\oplus y$. Set

$$z = X^T x^\circ.$$  \hspace{1cm} (6.4)

Then

$$\mathbf{E} \hat{X} z = \mathbf{E} A^\oplus Y z = \mathbf{E} A^\oplus AX X^T x^\circ = A^\oplus A \text{Cov}(X, X) x^\circ.$$
Further, since $\text{Cov}(X, X) = I$ and $x^o = A^\oplus y$ and $A^\oplus AA^\oplus = A^\oplus$, we have

$$E\hat{X}_z = x^o. \quad (6.5)$$

6.2. Kalman filter for $\hat{X}$. For a definiteness, assume that $A = A_{n \times n}$ and denote by $a_k$, $k = 1, \ldots, n$ rows of $A$. By $Y_k$, $k = 1, \ldots, n$ denote entries of $Y$. For notational convenience, introduce also a sequence $X_k$, $k = 0, 1, \ldots, n$ with $X_0 = X$ and $X_k = X_{k-1}$. Obviously, $X_k \equiv X$. Nevertheless, these notions allow us to present the relation $Y = AX$ in a form compatible with “noiseless version” of (5.8) (Lect. 5)

$$X_k = X_{k-1}$$
$$Y_k = a_kX_{k-1}. \quad (6.6)$$

Since the Kalman filter is also valid for the noiseless model, required only the use of pseudoinverse matrices, for $\hat{X}_k = \hat{E}(X_k|M_k)$, where $M_k$ is the linear space generated by “1, $Y_1, \ldots, Y_k$”, and

$$P_k = E(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T$$

the following Kalman filter occurs

$$\hat{X}_k = \hat{X}_{k-1} + P_{k-1}a_k^T(a_kP_{k-1}a_k^T)^\oplus(Y_k - a_k\hat{X}_{k-1})$$
$$P_k = P_{k-1} - P_{k-1}a_k^T(a_kP_{k-1}a_k^T)^\oplus a_kP_{k-1} \quad (6.7)$$

subject to the initial conditions $\hat{X}_0 = 0$ and $P_0 = I$.

Notice that, whereas $X_k \equiv X$, we have $\hat{X}_k = \hat{E}(X|M_k)$. Hence $\hat{X}_n = \hat{X}$. Denote by $A_k$ the sub-matrix of $A$ containing first $k$ rows of $A$ and by $Y^k$ the sub-vector of $Y$ containing first $k$ entries of $Y$. Owing to

$$Y^k = A_kX,$$

we have $\hat{X}_k = A_k^\oplus Y^k$. So, the Kalman filter generates a sequence of projections

$$A_1^\oplus Y^1, A_2^\oplus Y^2, \ldots, A_n^\oplus Y^n (= A^\oplus Y).$$

At the same time, since

$$X_k - \hat{X}_k = X - A_ky^k$$
$$= X - A_k^\oplus A_kX$$
$$= (I - A_k^\oplus A_k)X,$$

we find

$$P_k = (I - A_k^\oplus A_k); \quad (6.8)$$

recall that $\text{Cov}(X, X) = I$, $(I - A_k^\oplus A_k)$ is symmetric matrix, and $(A_k^\oplus A_k)^2 = A_k^\oplus A_k$ (see, Section 5.2, Lect. 5).
6.3. Kalman filter for \( x^o \). Denote by \( y_k, k = 1, \ldots, y_k \) entries of \( y = Ax^o \) and by \( y^k \) the sub-vector of \( y \) containing first \( k \) entries of \( y \). Similar to (6.5), with \( z \) defined in (6.4), we find

\[
E \hat{x}_k = E \hat{x}_k X^T x^o = EA_k^\oplus y^k X^T x^o
\]

\[
= EA_k^\oplus A_k X^T x^o = A_k^\ominus A_k \text{Cov}(X, X) x^o
\]

\[
= A_k^\ominus A_k x^o = A_k^\ominus y^k := \hat{x}_k. \quad (6.9)
\]

Notice that

\[
x^o = \hat{x}_n.
\]

and

\[
EY_k = EA_k \text{Cov}(X, X) x^o = A_k x^o = y^k,
\]

\[
EY_k = EA_k \text{Cov}(X, X) x^o = a_k x = y_k. \quad (6.10)
\]

Now, taking into the consideration (6.9) and (6.10) and multiplying the left- and right-hand sides of the first equation in (6.7) by \( z \) and then taking the expectation of the expressions obtained, we transform (6.7) into the Kalman filter generating \( \hat{x}_k, k = 1, \ldots, n \):

\[
\hat{x}_k = \hat{x}_{k-1} + P_{k-1} a_k^T \left( a_k P_{k-1} a_k^T \right)^\ominus \left( y_k - a_k \hat{x}_{k-1} \right)
\]

\[
P_k = P_{k-1} - P_{k-1} a_k^T \left( a_k P_{k-1} a_k^T \right)^\ominus a_k P_{k-1}, \quad (6.11)
\]

where \( \hat{x}_0 = 0, P_0 = I \).

A natural question arises: why it makes sense to use (6.11) for a computation of \( A^\ominus y^k \)? Our arguments are the following. The algorithm given in (6.11) applies the simplest pseudoinverse operation for \( \left( a_k P_{k-1} a_k^T \right)^\ominus \) for the scalar \( a_k P_{k-1} a_k^T \), so that

\[
\left( a_k P_{k-1} a_k^T \right)^\ominus = \begin{cases} 
\frac{1}{a_k P_{k-1} a_k^T}, & \text{if } a_k P_{k-1} a_k^T > 0 \\
0, & \text{otherwise.} 
\end{cases} \quad (6.12)
\]

It is clear that (6.12), as any pseudoinverse operation, is ill-posed, since \( a_k P_{k-1} a_k^T \) decreases with increasing \( k \) and being positive might be too close to zero that the fractal \( \frac{1}{a_k P_{k-1} a_k^T} \) can not be computed correctly. Nevertheless, a structure of (6.12) allows to control the correctness of such type computation. A verification of the computation correctness heavily uses the fact that \( a_k P_{k-1} a_k^T \) is the trace of the matrix \( P_{k-1} a_k^T a_k P_{k-1} \) (in fact, by (6.8), \( \text{trace}(P_{k-1} a_k^T a_k P_{k-1}) = a_k P_{k-1} a_k^T = a_k P_{k-1} a_k^T) \). Consequently, we have

\[
\text{trace}(P_{k-1} a_k^T a_k P_{k-1})^\ominus a_k P_{k-1} = \begin{cases} 
1, & \text{if } a_k P_{k-1} a_k^T > 0 \\
0, & \text{otherwise} 
\end{cases}
\]

and any “essential deviation” from 0 or 1 fixes the incorrectness.

The next question is what indicates \( a_k P_{k-1} a_k^T = 0 \)?
Lemma 6.1. \( a_k P_{k-1}a_k^T = \min_{c_1,\ldots,c_{k-1}} \|a_k - \sum_{j=1}^{k-1} c_j a_j\|^2 \). Particularly, if \( a_k P_{k-1}a_k^T = 0 \), then the row \( a_k \) is a linear combination of \( a_1,\ldots,a_k \).

Proof. Set \( \Delta = a_k - \sum_{j=1}^{k-1} c_j a_j \), where numbers \( a_j \)'s are chosen such that to minimize the Euclidean norm \( \|\Delta\|^2 \). Let \( c \) be the vectorrow with entries \( c_1,\ldots,c_k \) and notice that \( \Delta = a_k - c A_{k-1} \) and so

\[
\|\Delta\|^2 = \|a_k\|^2 - 2a_k A_{k-1}^T c + c A_{k-1} A_{k-1}^T c^T.
\]

Since \( \|\Delta\|^2 \) is the minimal in \( c \), the vector \( c \) solves the linear equation \( \nabla_c \|\Delta\| = 0 \), that is \( c(A_{k-1}A_{k-1}^T) = a_k A_{k-1}^T \) and so

\[
c = a_k A_{k-1}^T (A_{k-1} A_{k-1}^T)^{-1} = a_k A_{k-1}^\oplus.
\]

Hence, \( \Delta = a_k (I - A_{k-1} A_{k-1}^\oplus) \) and therefore

\[
\|\Delta\|^2 = a_k (I - 2A_{k-1} A_{k-1}^\oplus + (A_{k-1} A_{k-1}^\oplus)^2) a_k^\oplus = a_k (I - A_{k-1} A_{k-1}) a_k^\oplus = a_k P_{k-1} a_k^T
\]

(for more details see Section 5.2 Lect. 5).

**KARHUNEN-LOEVE DECOMPOSITION**

Let \( \xi_k, k = 1,\ldots,n \) be zero mean sequence of random variables with the correlation function \( R(k, \ell) \). From point of view of a theoretical consideration and simulation as well, it is useful if \( \xi_k \)'s are generated with a help of linear transformation of a white noise similar to different models of stationary random sequences in the wide sense given in Lect. 1.

6.4. Orthonormal eigenvectors vectors of correlation matrix. A matrix \( \textbf{R} = \textbf{R}_{n \times n} \) with entries \( R(k, \ell) \) is the correlation matrix of \( \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} \), that is \( \textbf{R} = E \xi \xi^T \) and so \( \textbf{R} \) is nonnegative definite matrix (in fact, \( x^T \textbf{R} x = E x^T \xi \xi^T x = E \|x^T \xi\|^2 \geq 0 \)). It is well know from “Theory of matrices” that eigenvectors of a nonnegative definite matrix can be transformed in a system of orthonormal vectors. In other words, eigenvectors of \( \textbf{R} \)

\[
\varphi_1 = (\varphi(1,1),\ldots,\varphi(1,n)) \\
\varphi_2 = (\varphi(2,1),\ldots,\varphi(2,n)) \\
\vdots \\
\varphi_N = (\varphi(N,1),\ldots,\varphi(n,n))
\]
Proof. Assume first that (6.15) is implied by (6.16) under nonsingular $R$ that is entries of random vector $\text{diag}$ positive. Then the diagonal. Since $\text{diag}$ where $R$ the diagonal. Since $\text{diag}$ the quadratic matrix with rows $\varphi_1, \varphi_2, \ldots, \varphi_n$ and notice that (6.13) provides $S^T S = I$, where $I$ is the unite matrix. Hence, $S^T = S^{-1}$ and $S$ is called “orthogonal matrix”. Denote by $\lambda_j$ the right eigenvector of $R$ corresponding to $\varphi_j$: $R \varphi_j = \lambda_j \varphi_j$. Owing to properties of $S$, we have

$$S^T R S = S^T R \begin{pmatrix} \lambda_1 \varphi_1 \\ \lambda_2 \varphi_2 \\ \vdots \\ \lambda_n \varphi_n \end{pmatrix} = \text{diag}(\lambda_1, \ldots, \lambda_n),$$

(6.14)

where $\text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix with eigenvalues of $R$ on the diagonal. Since $R$ is the nonnegative matrix, its eigenvalues are nonnegative as well.

6.5. Karhunen-Loeve theorem.

Theorem 6.2. There exists zero mean random vector with orthogonal entries $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ with $E \varepsilon_j^2 \equiv 1$ such that for every $k = 1, 2, \ldots, n$

$$\xi_k = \sum_{j=1}^n \sqrt{\lambda_j} \varphi(k, j) \varepsilon_j.$$  

(6.15)

Proof. Assume first that $R$ is nonsingular matrix, that is all $\lambda_j$’s are positive. Then $\text{diag}^{-1}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1})$.

Set

$$\varepsilon = \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) S^T \xi,$$  

(6.16)

Notice that $E \xi = 0$ and

$$E \xi \xi^T = \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) \times S^T R \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) = \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) \times \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) = I,$$

that is entries of random vector $\varepsilon$ are zero mean and orthogonal.

Further, since $S^T = S^{-1}$ and

$$\text{diag}^{-1}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_n^{-1/2}) = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}),$$

(6.15) is implied by (6.16) under nonsingular $R$, that is

$$\xi = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}) S \varepsilon.$$  

(6.17)

If $R$ is singular, then, instead of (6.16), let us introduce

$$\varepsilon' = \text{diag}^\oplus(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}) S^T \xi,$$  

(6.18)
here $\text{diag}^\oplus$ is defined similarly to $\text{diag}^{-1}$ with zeros located on same place where $\lambda$’s are zeros. Now, introduce zero mean random vector (column of $\xi$ size) $\eta$ with orthogonal entries of the unite variance. Also assume that $\eta$ and $\xi$ are orthogonal and introduce

$$
\varepsilon'' = \left( I - \text{diag}^\oplus(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}) \text{diag}^\oplus(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}) \right). 
$$

(6.19)

Henceforth, omitting arguments in $\text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2})$, write

$$
\varepsilon = \varepsilon' + \varepsilon'' = \text{diag}^\oplus S^T \xi + \left( I - \text{diag}^\oplus \text{diag}^\oplus \right) \eta. 
$$

(6.20)

Taking into consideration of the pseudoinverse matrix properties (see, Section 5.2 in Lect. 5) it is readily to check that

$$
E\varepsilon = 0 \quad \text{and} \quad E\varepsilon \varepsilon^T = I.
$$

Under singular $\mathbf{R}$, the formula, given in (6.17), remains with $\varepsilon$ from (6.20). Indeed, denote by $\tilde{\xi}$ the right hand of (6.17) with that $\varepsilon$ and show that $\xi = \tilde{\xi}$ with probability one. For this purpose, it suffices to prove that $E\|\xi - \tilde{\xi}\|^2 = 0$. Write

$$
E\|\xi - \tilde{\xi}\|^2 = E(\xi - \tilde{\xi})(\xi - \tilde{\xi})^T \\
= (I - S\text{diag}(\cdot)\text{diag}^\oplus(\cdot)S^T)\mathbf{R}(I - S\text{diag}(\cdot)\text{diag}^\oplus(\cdot)S^T)^T \\
= (I - S\text{diag}(\cdot)\text{diag}^\oplus(\cdot)S^T)S\text{diag}^2(\cdot)S^T(I - S\text{diag}(\cdot)\text{diag}^\oplus(\cdot)S^T) = 0.
$$

Remark 6.3. The Karhunen-Loeve decomposition holds true for a random vector $\xi$ with entries from a countable set $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, provided that

$$
\sum_{k=1}^\infty R(k, k) < \infty.
$$