UC Riverside
UC Riverside Previously Published Works

Title
Generalized Karhunen-Loeve Transform

Permalink
https://escholarship.org/uc/item/3312c23v

Journal
The IEEE Signal Processing Letters, 5(6)

ISSN
1070-9908

Authors
Hua, Yingbo
Liu, Wanquan

Publication Date
1998-06-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Generalized Karhunen–Loeve Transform

Yingbo Hua, *Senior Member, IEEE*, and Wanquan Liu, *Member, IEEE*

***Abstract*— We present a novel generic tool for data compression and filtering: the generalized Karhunen–Loeve (GKL) transform. The GKL transform minimizes a distance between any given reference and a transformation of some given data where the transform has a predetermined maximum possible rank. The GKL transform is also a generalization of the relative Karhunen–Loeve (RKL) transform by Yamashita and Ogawa where the latter assumes that the given data consist of the given reference (signal) and an independent noise. This letter provides a very simple and yet complete description of the GKL transform and shows useful engineering insights into the GKL transform.**

***Index Terms*— Data compression, data filtering, Karhunen–Loeve transform, rank reduction, subspace decomposition, SVD, Wiener filter.**

I. Review of the KL Transform

The Karhunen–Loeve (KL) transform is a well-known signal processing technique for data compression and filtering. A simple description of the KL transform is as follows. Given a (complex) random vector $\boldsymbol{x}$ of dimension $n \times 1$, the KL transform of $\boldsymbol{x}$ is represented by a square matrix $\mathbf{T}_{\mathrm{KL}}$ of maximum possible rank $m\,(\leq n)$ that minimizes

$$J_{\mathrm{KL}}(\mathbf{T}) = E\{\|\mathbf{x} - \mathbf{T}\mathbf{x}\|^2\} \tag{1}$$

where E denotes expectation, and $\|\,\|$ the Frobenius norm [1]. The matrix $\mathbf{T}_{\mathrm{KL}}$ is known to be the projection matrix onto the rank-m principal subspace of the covariance matrix $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$, where the superscript ^{H} denotes conjugate transpose. More specifically, if the eigendecomposition of $\mathbf{R}_x$ is expressed as

$$\mathbf{R}_x = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i \mathbf{e}_i^H \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then $\mathbf{T}_{\mathrm{KL}} = \mathbf{P}\mathbf{P}^H$ where $\mathbf{P} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_m]$. The matrix $\mathbf{T}_{\mathrm{KL}}$ implies two companion operations: compression, i.e., $\mathbf{x}_c = \mathbf{Q}^{-1}\mathbf{P}^H\mathbf{x}$, and reconstruction, i.e., $\mathbf{x}_r = \mathbf{P}\mathbf{Q}\mathbf{x}_c$, where $\mathbf{Q}$ can be chosen to be any nonsingular matrix but $\mathbf{Q} = \mathbf{I}$ is the most popular. The data compression ratio is given by m/n.

II. The GKL Transform

The generalized Karhunen–Loeve (GKL) transform is represented by a (possibly nonsquare) matrix $\mathbf{T}_{\mathrm{GKL}}$ of maximum possible rank $m\,(\leq n)$ that minimizes

$$J_{\mathrm{GKL}}(\mathbf{T}) = E\{\|\mathbf{s} - \mathbf{T}\mathbf{x}\|^2\} \tag{3}$$

where $\mathbf{s}$ can be any desired random vector of a dimension possibly different from (in practice, however, not larger than) that of $\mathbf{x}$. The choice of $\mathbf{s}$ depends on some prior knowledge about the data $\mathbf{x}$, which is, of course, highly application dependent. In the setting up of the relative Karhunen–Loeve (RKL) transform [2], $\mathbf{x}$ is assumed to be the sum of $\mathbf{s}$ and an independent noise vector $\mathbf{n}$.

Without the rank constraint on $\mathbf{T}$, the above minimization is associated with the well-known concept of Wiener filtering, and the optimum transform is given by [easy to prove either directly or by using (5)]

$$\mathbf{T}_0 = \mathbf{R}_{sx}\mathbf{R}_x^{+} \tag{4}$$

where $\mathbf{R}_{sx} = E\{\mathbf{s}\mathbf{x}^H\}$, and the superscript $^+$ denotes the Moore–Penrose pseudoinverse [1].

With the rank constraint on $\mathbf{T}$, i.e., rank$(\mathbf{T}) \leq m$, we now consider an equivalent cost function of (3), as follows:

$$J_{\mathrm{GKL}}(\mathbf{T}) - J_{\mathrm{GKL}}(\mathbf{T}_0) = \mathrm{tr}\{(\mathbf{T} - \mathbf{T}_0)\mathbf{R}_x(\mathbf{T} - \mathbf{T}_0)^H\} \tag{5}$$

which can be easily verified by using the fact $\mathbf{R}_{sx}\mathbf{R}_x\mathbf{R}_x^{+} = \mathbf{R}_{sx}$ and the four Moore–Penrose equations of $\mathbf{R}_x^{+}$ [1]. Let $\mathbf{R}_x$ be factorized as $\mathbf{R}_x = \mathbf{R}_x^{1/2}\mathbf{R}_x^{1/2^H}$. Then, it follows that

$$\begin{aligned} J_{\mathrm{GKL}}(\mathbf{T}) - J_{\mathrm{GKL}}(\mathbf{T}_0) &= \|(\mathbf{T} - \mathbf{T}_0)\mathbf{R}_x^{1/2}\|^2 \\ &= \|\mathbf{T}\mathbf{R}_x^{1/2} - \mathbf{T}_0\mathbf{R}_x^{1/2}\|^2. \end{aligned} \tag{6}$$

Minimizing the above norm by a rank-m matrix $\mathbf{T}\mathbf{R}_x^{1/2}$ is known [1] to satisfy

$$\mathbf{T}\mathbf{R}_x^{1/2} = trun_m\{\mathbf{T}_0\mathbf{R}_x^{1/2}\} \tag{7}$$

where $trun_m$ denotes rank-m singular value decomposition (SVD) truncation, i.e., if the following SVD holds:

$$\mathbf{T}_0\mathbf{R}_x^{1/2} = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^H \tag{8}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, then

$$trun_m\{\mathbf{T}_0\mathbf{R}_x^{1/2}\} = \sum_{i=1}^{m} \sigma_i \mathbf{u}_i \mathbf{v}_i^H. \tag{9}$$

It is clear that the row space of $trun_m\{\mathbf{T}_0\mathbf{R}_x^{1/2}\}$ belongs to the row space of $\mathbf{R}_x^{1/2}$, and hence (7) has at least one solution for $\mathbf{T}$ and the minimum norm solution is given [1] by

$$\begin{aligned} \mathbf{T}_{\mathrm{GKL}} &= trun_m\{\mathbf{T}_0\mathbf{R}_x^{1/2}\}\mathbf{R}_x^{-1/2} \\ &= trun_m\{\mathbf{R}_{sx}\mathbf{R}_x^{-1/2}\}\mathbf{R}_x^{-1/2} \end{aligned} \tag{10}$$

where $\mathbf{R}_x^{-1/2} = (\mathbf{R}_x^{1/2})^{+}$.

Manuscript received October 21, 1997. This work was supported by the Australian Research Council and the Australian Cooperative Research Centre for Sensor Signal and Information Processing. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. R. M. Mersereau.

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Vic. 3052, Australia (e-mail: yhua@ee.mu.oz.au).

Publisher Item Identifier S 1070-9908(98)04340-5.

1070–9908/98$10.00 © 1998 IEEE

Note that the GKL transform requires the knowledge of the cross correlation matrix $\mathbf{R}_{sx}$. The GKL transform is identical to the RKL transform [2] if $\mathbf{R}_{sx} = \mathbf{R}_s = E\{\mathbf{s}\mathbf{s}^H\}$.

III. Discussions

As in the case of the KL transform, the GKL transform implies two companion operations: compression and reconstruction. While the data compression ratio is governed by the predetermined rank m, there is no unique choice for a pair of compression matrix $\mathbf{M}$ (of m rows) and reconstruction matrix $\mathbf{N}$ (of m columns) where $\mathbf{T}_{\text{GKL}} = \mathbf{NM}$. However, a natural choice [using (9)] can be

$$\begin{aligned}\mathbf{M} &= [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_m]^H \mathbf{R}_x^{-1/2} \\ \mathbf{N} &= [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m]\,\text{diag}\,\{\sigma_1 \quad \sigma_2 \quad \cdots \quad \sigma_m\}.\end{aligned} \tag{11}$$

Unlike the KL transform, the GKL transform is not a projection matrix as defined in [1]. In fact, for the special case where the data vector $\mathbf{x}$ consists of the signal vector $\mathbf{s}$ and the (independent) white noise vector $\mathbf{n}$ of variance p_n, we know that $\mathbf{R}_x = \mathbf{R}_s + p_n\mathbf{I}$, $\mathbf{R}_{sx} = \mathbf{R}_s$ and hence [easy to show using (2)]

$$\begin{aligned}\mathbf{T}_{\text{GKL}} &= trun_m\{\mathbf{R}_s\mathbf{R}_x^{-1/2}\}\mathbf{R}_x^{-1/2} \\ &= \left(\sum_{i=1}^{m} \frac{\lambda_i - p_n}{\lambda_i^{1/2}}\,\mathbf{e}_i\mathbf{e}_i^H\right)\left(\sum_{i=1}^{n} \frac{1}{\lambda_i^{1/2}}\,\mathbf{e}_i\mathbf{e}_i^H\right) \\ &= \sum_{i=1}^{m} \frac{\lambda_i - p_n}{\lambda_i}\,\mathbf{e}_i\mathbf{e}_i^H\end{aligned} \tag{12}$$

which is not a project matrix since $\mathbf{T}_{\text{GKL}} \neq \mathbf{T}_{\text{GKL}}\mathbf{T}_{\text{GKL}}$. As the noise variance p_n goes to zero, however, this $\mathbf{T}_{\text{GKL}}$ approaches $\mathbf{T}_{\text{KL}}$.

The GKL transform is the best rank-m transform in minimizing (3). There are, of course, various heuristic rank-m transforms, and some of them may be more efficient in computation. One such example is a rank-m SVD truncation of the Wiener transform $\mathbf{T}_0 = \mathbf{R}_{sx}\mathbf{R}_x^+$. For the special case considered previously, it can be shown that the truncated Wiener transform is identical to the GKL transform, i.e.,

$$trun_m\{\mathbf{T}_0\} = \mathbf{T}_{\text{GKL}}. \tag{13}$$

However, as the noise variance p_n goes to zero, $trun_m\{\mathbf{T}_0\}$ becomes ill-conditioned as all the eigenvalues of $\mathbf{T}_0$ become equal to one (provided $\mathbf{R}_x$ is nonsingular).

The potential applications of the GKL transform are abundant. One example was shown in [2] where the RKL transform (a special form of the GKL transform) is applied to image compression. To show another example, we now briefly discuss a blind channel equalization problem [3], [4] where the available data are given by

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t) \tag{14}$$

in which $\mathbf{x}(t)$ is the available channel output vector, $\mathbf{H}$ the unknown channel response matrix, $\mathbf{s}(t)$ the unknown channel input vector (the desired signal), and $\mathbf{n}(t)$ the (independent) noise. Here, we add the variable t as a convention to emphasize the time dependence. For this problem, the best rank-m linear equalizer for estimating $\mathbf{s}(t)$ is given by

$$\hat{\mathbf{s}}(t) = \mathbf{T}_{\text{GKL}}\mathbf{x}(t) \tag{15}$$

where $\mathbf{T}_{\text{GKL}}$ is shown in (10) with $\mathbf{R}_{sx} = \mathbf{R}_s\mathbf{H}^H$ and $\mathbf{R}_x = \mathbf{H}\mathbf{R}_s\mathbf{H}^H + \mathbf{R}_n$. The channel response matrix $\mathbf{H}$ can be estimated from $\mathbf{x}(t)$ by a fast maximum likelihood method [3], the input correlation matrix $\mathbf{R}_s$ is an identity matrix with a known scale when the input corresponds to some pseudorandom noise as in the case of code division multiple access (CDMA) communications, and the noise correlation matrix $\mathbf{R}_n$ is also measureable off-line. The rank constraint in this context corresponds to that on the complexity of the linear equalizer. But perhaps more importantly, the rank constraint makes the linear equalizer more robust to noise in comparison to the Wiener filter, especially when $\mathbf{R}_x$ is singular or near singular. Note that when $\mathbf{R}_x$ is singular or near singular, the pseudoinverse of $\mathbf{R}_x$ as required in (4) without a predetermined rank is ill-defined and very sensitive to noise. The relation shown in (15) together with some known statistics of $\mathbf{s}(t)$ can also be used to blindly estimate $\mathbf{H}$. Further exploration along this direction is underway and will be reported elsewhere in the near future.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: Johns Hopkins Univ. Press, 1983.

[2] Y. Yamashita and H. Ogawa, "Relative Karhunen–Loeve transform," *IEEE Trans. Signal Processing,* vol. 44, pp. 371–378, Feb. 1996.

[3] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing,* vol. 44, pp. 661–672, Mar. 1996.

[4] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE,* vol. 85, pp. 1310–1322, Aug. 1997.