On Cumulant Techniques in Speech Processing

Vladimir Zaiats and Jordi Solé-Casals

Digital Technologies Group, Universitat de Vic, c/. Sagrada Família, 7, 08500 Vic (Barcelona), Spain,
{vladimir.zaiats,jordi.sole}@uvic.cat

Abstract. This paper analyzes applications of cumulant analysis in speech processing. A special focus is made on different second-order statistics. A dominant role is played by an integral representation for cumulants by means of integrals involving cyclic products of kernels.

Keywords: Cumulants, higher-order statistics, correlogram, speech enhancement

1 Introduction

Different methods in speech recognition use linear and non-linear procedures derived from the speech signal by matching the autocorrelation or the power spectrum, [4, 6, 5]. Many of these methods perform well for clean speech, while their performance decreases strongly if noise conditions mismatch for training and testing.

We obtain a representation for cumulants of second-order statistics containing a special type of integrals that involve cyclic products of kernels. Our techniques are based on [1–3, 7].

2 Integrals Involving Cyclic Products of Kernels

For $m \in \mathbb{N}$, define $\mathbb{N}_m := \{1, \dots, m\}$. Assume that $(\mathbb{V}, \mathcal{F}_{\mathbb{V}})$ is a measurable space and $\mu_1, \dots, \mu_m$ are σ-finite (real- or complex-valued) measures on $(\mathbb{V}, \mathcal{F}_{\mathbb{V}})$. For $m \in \mathbb{N}$, $m \geq 2$, consider the following integral:

$$\overset{\frown}{I}\,(K_1, \dots, K_m;\ \boldsymbol{\varphi}) \tag{1}$$
$$:= \int \cdots \int_{\mathbb{V}^m} \left[\prod_{p=1}^{m} K_p(v_p, v_{p+1}) \right] \boldsymbol{\varphi}(v_1, \dots, v_m) \mu_1(dv_1) \dots \mu_m(dv_m)$$

where $v_{m+1} := v_1$. Integral (1) will be called an integral involving a cyclic product of kernels (IICPK).

We will denote:

$$\overset{\frown}{\prod_{p \in \mathbb{N}_m}} K_p(v_p, v_{p+1}) := \prod_{p=1}^{m} K_p(v_p, v_{p+1})$$

with $v_{m+1} := v_1$. This function will be called a cyclic product of kernels $K_1, \dots, K_m$.

3 Cumulants of General Bilinear Forms of Gaussian Random Vectors

Suppose that $m \in \mathbb{N}$; $n_{j,1}, n_{j,2} \in \mathbb{N}$, $j \in \mathbb{N}_m$, and write

$$\boldsymbol{X}_{j,1} := (X_{j,1}(k),\ k \in \mathbb{N}_{n_{j,1}}), \qquad \boldsymbol{X}_{j,2} := (X_{j,2}(k),\ k \in \mathbb{N}_{n_{j,2}}), \quad j \in \mathbb{N}_m.$$

Assume that $\boldsymbol{X}_{j,1}$ and $\boldsymbol{X}_{j,2}$, $j \in \mathbb{N}_m$, are real-valued zero-mean random vectors and consider the following bilinear forms:

$$U_j := \sum_{k,l=1}^{n_{j,1},n_{j,2}} a_j(k,l) X_{j,1}(k) X_{j,2}(l), \quad j \in \mathbb{N}_m,$$

where

$$\sum_{k,l=1}^{n_{j,1},n_{j,2}} := \sum_{k=1}^{n_{j,1}} \sum_{l=1}^{n_{j,2}}.$$

If we put

$$(a_j(k,l)) := \begin{pmatrix} a_j(1,1) & \dots & a_j(1,n_{j,2}) \\ \vdots & \dots & \vdots \\ a_j(n_{j,1},1) & \dots & a_j(n_{j,1},n_{j,2}) \end{pmatrix}, \quad j \in \mathbb{N}_m,$$

then for any $j \in \mathbb{N}_m$

$$\begin{aligned} U_j &= \boldsymbol{X}_{j,1} \left(a_j(k,l)\right) \boldsymbol{X}_{j,1}^{\top} \\ &= (X_{j,1}(1), \dots, X_{j,1}(n_{j,1})) \begin{pmatrix} a_j(1,1) & \dots & a_j(1,n_{j,2}) \\ \vdots & \dots & \vdots \\ a_j(n_{j,1},1) & \dots & a_j(n_{j,1},n_{j,2}) \end{pmatrix} \begin{pmatrix} X_{j,2}(1) \\ \vdots \\ X_{j,2}(n_{j,2}) \end{pmatrix}. \end{aligned}$$

Consider the joint simple cumulant cum $(U_1, \dots, U_m)$ of the random variables $U_1, \dots, U_m$ assuming that this cumulant exists. By general properties of the cumulants, we obtain

$$\begin{aligned} &\operatorname{cum}(U_1, \dots, U_m) \\ &= \sum_{k_{1,1},k_{1,2}=1}^{n_{1,1},n_{1,2}} \cdots \sum_{k_{m,1},k_{m,2}=1}^{n_{m,1},n_{m,2}} \Bigg[\left(\prod_{j=1}^{m} a_j(k_{j,1}, k_{j,2}) \right) \\ &\qquad \times \operatorname{cum}\left(X_{j,1}(k_{j,1}) X_{j,2}(k_{j,2}), j \in \mathbb{N}_m\right) \Bigg]. \end{aligned}$$

Since any general bilinear form can be represented as a sum of diagonal bilinear forms, the following result holds.

Theorem 1. *Let $m \in \mathbb{N}$; $n_{j,1} = n_{j,2} = n_j \in \mathbb{N}$, $j \in \mathbb{N}_m$. Assume that $(\boldsymbol{X}_{j,1}, \boldsymbol{X}_{j,2}, j \in \mathbb{N}_m)$ is a jointly Gaussian family of zero-mean random variables and suppose that for any $j, \tilde{j} \in \mathbb{N}_m$ and any $\alpha, \tilde{\alpha} \in \{1, 2\}$ there exists a complex-valued measure $M_{j,\tilde{j}}^{\alpha,\tilde{\alpha}}$ such that*

$$\mathsf{E} X_{j,\alpha}(k) X_{\tilde{j},\tilde{\alpha}}(\tilde{k}) = \int_{\mathbb{D}} e^{i(k-\tilde{k})\lambda} M_{j,\tilde{j}}^{\alpha,\tilde{\alpha}}(d\lambda).$$

Then

$$\begin{aligned} &\operatorname{cum}(U_1, \ldots, U_m) \\ &= \sum_{\boldsymbol{l} \in \mathcal{L}_m(n_j, j \in \mathbb{N}_m)} \sum_{(\boldsymbol{\jmath},\boldsymbol{\alpha}) \in \{P,2\}_{m-1}} \int \cdots \int_{\mathbb{D}^m} \left[\overset{\frown}{\prod_{p \in \mathbb{N}_m}} K_p^{(\boldsymbol{\jmath},\boldsymbol{\alpha},\boldsymbol{l})}(v_p - v_{p+1}) \right] \\ &\qquad\qquad \times \mu_1^{(\boldsymbol{\jmath},\boldsymbol{\alpha},\boldsymbol{l})}(dv_1) \ldots \mu_m^{(\boldsymbol{\jmath},\boldsymbol{\alpha},\boldsymbol{l})}(dv_m), \end{aligned}$$

that is $\operatorname{cum}(U_1, \ldots, U_m)$ *is represented as a finite sum of integrals involving cyclic products of kernels. Here, $\boldsymbol{\jmath} := (j_1, j_2, \ldots, j_m)$, $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \ldots, \alpha_m)$, $j_{m+1} = j_1 = 1$, $\alpha_{m+1} = \alpha_1 = 2$, and the sum $\sum_{(\boldsymbol{\jmath},\boldsymbol{\alpha})}$ is extended to all*

$$((j_2, \ldots, j_m), (\alpha_2, \ldots, \alpha_m)) \in \operatorname{Perm}\{2, \ldots, m\} \times \{1, 2\}^{m-1}. \tag{2}$$

The notation $(\boldsymbol{\jmath}, \boldsymbol{\alpha}) \in \{P, 2\}_{m-1}$ for fixed $j_1 = 1$ and $\alpha_1 = 2$ is equivalent to (2).

Here, we put $\mathbb{Z}_{|n_j-1|} := \{-(n_j - 1), \ldots, -1, 0, 1, \ldots, n_j - 1\}$ for $j \in \mathbb{N}_m$ and $\mathcal{L}_m(n_j, j \in \mathbb{N}_m) := \mathbb{Z}_{|n_1-1|} \times \ldots \times \mathbb{Z}_{|n_m-1|}$.

4 Applications

We apply the above obtained integral representations to some problems in speech recognition. Let us consider a setting where sample correlograms and sample cross-correlograms of stationary time series appear.

Let $\boldsymbol{Y}(t) := (Y_1(t), Y_2(t))$, $t \in \mathbb{Z}$, be a weak sense stationary zero-mean bidimensional vector-valued stochastic process with real-valued components whose matrix-valued autocovariance function is as follows:

$$\mathbf{C}_{\boldsymbol{Y}}(t) := \begin{pmatrix} C_{11}(t) & C_{12}(t) \\ C_{21}(t) & C_{22}(t) \end{pmatrix}, \quad t \in \mathbb{Z},$$

and let

$$\mathbf{F}_{\boldsymbol{Y}}(\lambda) := \begin{pmatrix} F_{11}(\lambda) & F_{12}(\lambda) \\ F_{21}(\lambda) & F_{22}(\lambda) \end{pmatrix}, \quad \lambda \in [-\pi, \pi],$$

be the matrix-valued spectral function of the vector-valued process $\boldsymbol{Y}(t), t \in \mathbb{Z}$. Let $\gamma, \delta \in \{1,2\}$. Consider the following random variables:

$$\hat{C}_{\gamma\delta}(\tau; N) := \sum_{k=1}^{N} b_{\gamma\delta}(k;\ \tau, N) Y_\gamma(k+\tau) Y_\delta(k), \quad \tau \in \mathbb{Z}, \quad N \in \mathbb{N},$$

where $b_{\gamma\delta}(k;\ \tau, N)$, $k \in \mathbb{N}_N$, $\tau \in \mathbb{Z}$, $N \in \mathbb{N}$, are non random real-valued weights. It is often assumed that

$$\sum_{k=1}^{N} b_{\gamma\delta}(k;\ \tau, N) = 1, \quad \tau \in \mathbb{Z}, \quad N \in \mathbb{N}. \tag{3}$$

For example, let $N \in \mathbb{N}$ be given and let

$$b_{\gamma\delta}(k;\ \tau, N) = \frac{1}{N}\,, \quad k \in \mathbb{N}_N, \quad \tau \in \mathbb{Z}.$$

Then (3) holds and

$$\hat{C}_{\gamma\delta}(\tau; N) = \frac{1}{N} \sum_{k=1}^{N} Y_\gamma(k+\tau) Y_\delta(k), \quad \tau \in \mathbb{Z}, \quad N \in \mathbb{N}. \tag{4}$$

The following sample correlograms are also often used in spectral analysis and speech recognition as estimates of $C_{\gamma\delta}(\cdot)$, $\gamma, \delta \in \{1,2\}$:

$$\tilde{C}_{\gamma\delta}(\tau; N) = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{k=1}^{N-|\tau|} Y_\gamma(k+|\tau|) Y_\delta(k), & \text{for } |\tau| < N; \\ 0, & \text{for } |\tau| \geq N, \end{cases}$$

$$\tilde{\tilde{C}}_{\gamma\delta}(\tau; N) = \begin{cases} \dfrac{1}{N-|\tau|} \displaystyle\sum_{k=1}^{N-|\tau|} Y_\gamma(k+|\tau|) Y_\delta(k), & \text{for } |\tau| < N; \\ 0, & \text{for } |\tau| \geq N. \end{cases}$$

Let $\gamma, \delta \in \{1,2\}$, $N \in \mathbb{N}$, $m \in \mathbb{N}$, and $\tau_j \in \mathbb{Z}$, $j \in \mathbb{N}_m$. Put

$$\begin{aligned}
&\mathrm{cum}_{\gamma\delta}^{(N)}(\tau_1, \ldots, \tau_m) := \mathrm{cum}(\hat{C}_{\gamma\delta}(\tau_j; N),\ j \in \mathbb{N}_m); \\
&n_j = N, \quad j \in \mathbb{N}_m; \qquad a_j(k) = b_{\gamma\delta}(k;\ \tau_j, N), \quad k \in \mathbb{N}_N,\ j \in \mathbb{N}_m; \\
&X_{j,1}(k) = Y_\gamma(k+\tau_j), \quad k \in \mathbb{N}_N,\ j \in \mathbb{N}_m; \\
&X_{j,2}(k) = Y_\delta(k), \quad k \in \mathbb{N}_N,\ j \in \mathbb{N}_m.
\end{aligned}$$

Under these conditions the results obtained in Section 3 can be applied to the cumulants. These results imply that the Gaussian component of the cumulant $\mathrm{cum}_{\gamma\delta}^{(N)}(\tau_1, \ldots, \tau_m)$ is represented as a finite sum of integrals involving cyclic products of kernels.

Acknowledgements. We acknowledge the support of the Universitat de Vic under grant R0912. Some parts of this paper are based on joint research with V. Buldygin (National Technical University of Ukraine) and F. Utzet (Universitat Autònoma de Barcelona).

References

1. Avram, F.: Generalized Szego theorems and asymptotics of cumulants by graphical methods. Trans. Amer. Math. Society, 330, 637–649 (1992)
2. Avram, F.; Taqqu, M.: On the generalized Brascamp-Lieb-Barthe inequality, a Szegő type limit theorem, and the asymptotic theory of random sums, integrals and quadratic forms. Manuscript. March 6, 2005
3. Grenander, U.; Szegő, G. : Toeplitz forms and their applications. University of California Press (1958)
4. Makhoul, J.: Linear prediction: A tutorial review. Proc. IEEE, 63, 561–580 (1975)
5. Nemer, E., Goubran, R., Mahmoud, S.: Speech enhancement using fourth-order cumulants and optimum filters in the subband domain. Speech Comm., 36, 219–246 (2002)
6. Paliwal, K. K., Sondhi, M. M.: Recognition of noisy speech using cumulant-based linear prediction analysis. Proc. ICASSP, 429–432 (1991)
7. Sugrailis, D.: On multiple Poisson stochastic integrals and associated Markov semigroups. Probab. Math. Statist., 3, 217–239 (1984)