arXiv:math.ST/0612674v1 22 Dec 2006

Multivariate Lag-Windows and Group Representations*

Arthur Berg
University of California, San Diego
aberg@ucsd.edu

Abstract

Symmetries of the auto-cumulant function (the generalization of the auto-covariance function) of a kth-order stationary time series are derived through a connection with the symmetric group of degree k. Using theory of group representations, symmetries of the auto-cumulant function are demystified and lag-window functions are symmetrized to satisfy these symmetries. A generalized Gabr-Rao optimal kernel, used to estimate general k^{th}-order spectra, is also derived through the developed theory.

KEYWORDS: higher-order spectra, group representations, multivariate lag-windows, symmetry group

1 Introduction

There have been numerous approaches and references to constructing multivariate lag-windows that inherit the appropriate symmetries when estimating k^{th}-order spectra; cf. [9, 10, 11, 12, 15, 17, 18, 19, 20]. However, each approach constructs multivariate lag-windows from univariate lag-windows which severely limits their shape. The case $k = 2$, corresponding to univariate lag-windows, is trivial since the only symmetry requirement imposed is evenness. The symmetries are much more complex in the case $k = 3$, and the challenge is to fully understand these symmetries and to easily construct lag-windows that inherit the symmetries. In the process, an intimate relationship is discovered between the symmetries of the bivariate lag-window and the symmetric group S_3. Here we begin to see the makings of a beautiful connection between stationary time series and group representations of the symmetric group.

General theory of the symmetric group is pulled back via this connection to the context of stationary time series and many useful results are ascertained. Similarly, natural constructions of stationary time series are pushed forward into the theory of group representations of S_k to elicit a new realization of a familiar representation of S_k with dimension $k - 1$.

*Last Updated: December 22, 2006

Current practices of generating multivariate lag-windows are shown to be inadequate as they restrict the construction of many useful lag-windows including the Gabr-Rao optimal bivariate lag-window [17], yet this lag-window occurs naturally under the symmetrization routine described in this paper. The connection to the symmetric group via group representations allows for effortless generalization of the Gabr-Rao lag-window for estimation of the kth-order spectral density.

This is not the first time a connection between group representations and statistics has been drawn. In fact an entire book *Group Representations in Probability and Statistics* [5], written by the mathemagician Persi Diaconis, exposes many connections between group representations and statistics, but the connections drawn in this book are much different, with different motivations, from what is done the present article.

The first section defines stationarity—the driving force in the connection with the symmetric group—and introduces higher-order spectral densities and spectral density estimates with lag-windows. To solidify ideas, the case $k = 3$ is considered in Section 2, and a connection between symmetries of the auto-cumulant function (acf) and permutations of the symmetric group is obtained via a special group representation on S_3; these results are generalized for general k in Section 4. Section 5 describes the actions in the Fourier domain viewpoint which allows for the generalization of the Gabr-Rao kernel. Finally, several examples of bivariate kernels and lag-windows are produced in Section 6.

2 Stationarity

Let $\{X_t\}$ be a kth-order discrete (resp. continuous) stationary time series, that is

$$\mathrm{E}\left[X_t \prod_{t_i \in S} X_{t+x_i}\right] \tag{1}$$

is finite and the same for every $t \in \mathbb{Z}$ (resp. $t \in \mathbb{R}$) where $S \subset \mathbb{Z}$ (resp. $S \subset \mathbb{R}$) is any set of size at most $k-1$. A stronger, yet more typical, assumption is to require the time series to be strictly stationary, that is for any positive integer n and any $t, x_1, \ldots, x_n$ in $\mathbb{Z}$ (resp. $\mathbb{R}$), the random vectors $(X_{x_1}, \ldots, X_{x_n})$ and $(X_{t+x_1}, \ldots, X_{t+x_n})$ have identical joint distributions. However, this stronger version of stationarity will not be needed here since this paper only deals with expressions like those in (1).

As far as this paper is concerned, the theory for discrete time series is the same as that for continuous time series, so the remainder of this paper will focus on discrete time series only.

Let's specialize for the moment to the case of second-order (aka weakly) stationary time series. The assumption of second-order stationarity requires

(i) $\mathrm{E}(X_t^2) < \infty$ for all t;

(ii) $\mathrm{E}(X_t) = \mu$ for all t;

(iii) $C(x) = \mathrm{E}[(X_t - \mu)(X_{t+x} - \mu)]$ for all x and t.

The function $C(x)$ is called the auto-covariance function (referred to within as the

auto-cumulant function), and by replacing x with $-x$, we see that

$$C(-x) = \mathrm{E}[(X_t - \mu)(X_{t-x} - \mu)] = \mathrm{E}[(X_{t'} - \mu)(X_{t'+x} - \mu)] = C(x)$$

(using $t' = t+x$); so $C(x)$ is an even function, and this is the only symmetry requirement of $C(x)$.

Third-order stationarity is implied by the conditions (i)-(iii) above plus following two conditions:

(iv) $\mathrm{E}(X_t^4) < \infty$

(v) $C(x, y) = \mathrm{E}[(X_t - \mu)(X_{t+x} - \mu)(X_{t+y} - \mu)]$ for all x, y, and t.

In exploring the symmetries of the bivariate acf, $C(x, y)$, we can use the substitutions $t' = t - x$ and $t' = t - y$ and the commutativity property of multiplication to deduce

$$C(x, y) = C(-x, y - x) = C(y, x) = C(x - y, -y) = C(-y, x - y) = C(y - x, -x) \quad (2)$$

There is redundancy in this expression; for instance the relations $C(x, y) = C(y, x) = C(-x, y - x)$ imply $C(x, y) = C(y - x, -x)$. This begs two natural questions:

— Which equations can be removed without losing any information?

— Is there minimal set of equations representing all of the symmetries?

We will answer both of these questions in the following sections, but first we return to the general case of kth-order stationarity.

The kth-order auto-cumulant function is defined to be the function of $k-1$ variables given by

$$C(x_1, \ldots, x_{k-1}) = \sum_{(\nu_1, \ldots, \nu_p)} (-1)^{p-1}(p-1)!\, \mu_{\nu_1} \cdots \mu_{\nu_p} \quad (3)$$

where the sum is over all partitions $(\nu_1, \ldots, \nu_p)$ of $\{0, x_1, \ldots, x_{k-1}\}$ and $\mu_{\nu_j} = \mathrm{E}\left[\prod_{x_i \in \nu_j} X_{x_i}\right]$. For instance, for $k = 3$ we have

$$\begin{aligned} C(x_1, x_2) &= \mathrm{E}(X_0 X_{x_1} X_{x_2}) - \mathrm{E}(X_0 X_{x_1})\,\mathrm{E}(X_{x_2}) - \mathrm{E}(X_0 X_{x_2})\,\mathrm{E}(X_{x_1}) \\ &\quad - \mathrm{E}(X_{x_1} X_{x_2})\,\mathrm{E}(X_0) + 2\,\mathrm{E}(X_0)\,\mathrm{E}(X_{x_1})\,\mathrm{E}(X_{x_2}) \\ &= \mathrm{E}[(X_t - \mu)(X_{t+x_1} - \mu)(X_{t+x_2} - \mu)] \end{aligned} \quad (4)$$

The last equality in (4) does not generalize for $k > 3$, that is C is not the same as the more obvious generalization

$$\widetilde{C}(x_1, \ldots, x_{k-1}) = \mathrm{E}\left[(X_t - \mu)(X_{t+x_1} - \mu) \cdots (X_{t+x_{k-1}} - \mu)\right]$$

For justification of the use of C as opposed to $\widetilde{C}$, the reader is referred to section 7 "moments or cumulants?" of [4] and section 2.3.8 "Why Cumulant Spectra and not Moment Spectra?" of [11]. Nonetheless, we can see from (3) that C posses the same symmetries as $\widetilde{C}$. The symmetries of the general cumulant function will be explored in subsequent sections.

The kth-order spectral density, $f(\boldsymbol{\omega})$, is defined as the continuous (resp. discrete) Fourier transform of the kth-order cumulant function, i.e.

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{k-1}} \sum_{\boldsymbol{t} \in \mathbb{Z}^{k-1}} C(\boldsymbol{t}) e^{-i\boldsymbol{t}\cdot\boldsymbol{\omega}}.$$

The second-order spectral density has many uses; see section 1.10, "Time Series Analysis: Use of Spectral Analysis in Practice", of [13]. Higher-order spectra are less frequently used, but still many applications exist—most notably is the use of the bispectrum in constructing linearity tests of time series [16, 8]. Other applications of general higher-order spectra can be found in [11, 12].

Estimation of higher-order spectra is typically done with lag-window estimators of the form

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{k-1}} \sum_{\|\boldsymbol{t}\|<N} \lambda(\boldsymbol{t}/M)\widehat{C}(\boldsymbol{t})e^{-i\boldsymbol{t}\cdot\boldsymbol{\omega}}$$

where λ is a lag-window function whose properties will be described below; M is a bandwidth or smoothing parameter; $\widehat{C}$ is an estimate of C produced by estimating the expectations μ_{ν_i} in (3) by sample means $\hat{\mu}_{\nu_i}$. Indeed, estimation of kth-order spectra for large k is plagued by the curse of dimensionality, so accurate estimation of high-order spectra typically requires a large sample size.

The lag-window function λ can have varying restrictions, but three consistent assumptions are:

(a) $\lambda(\mathbf{0}) = 1$;

(b) $\|\lambda\|_{L_2} < \infty$;

(c) λ inherits the symmetries of C.

It is easy to find functions to satisfy conditions (a) and (b). However, in order to satisfy condition (c), we must first determine the symmetries of C and then construct a function that possesses these same symmetries. For $k = 2$ the problem is trivial since the only symmetry condition on $C(x)$ is $C(x) = C(-x)$. We can easily impose evenness on λ by constructing $\tilde{\lambda} = h\left(\lambda(x), \lambda(-x)\right)$ where h is any symmetric function (of two variables). For instance, if $h(x, y) = x + y$ or $h(x, y) = xy$, then $\tilde{\lambda}$ will be forced to be even. The solution to the general problem will be addressed in the following sections.

3 The Case $k = 3$

Equation (2) in the previous section exposes the six symmetries of $C(x, y)$, and a general technique to arrive at these symmetries is addressed below. We start with S_3, the set of permutations of three labels. Take for instance the permutation $\sigma \in S_3$ that transposes the first two labels and fixes the third. This permutation is denoted as $\sigma = (12)(3)$, or more simply, $\sigma = (12)$. All six permutations of S_3 can be written down similarly; they constitute the set $\{e, (12), (23), (13), (123), (132)\}$ where e is the identity permutation that doesn't move any labels.

A methodical procedure is now presented that produces each of the symmetries of $C(x, y)$ in (2). Under the commutativity property of multiplication, the product $X_t X_{t+x} X_{t+y}$ remains the same under any permutation of the three variables, so for each permutations in S_3, we permute the variables, according to the permutation, then adjust t so that the first variable has index t. Then the symmetry condition can be read off from the last two variables. Say for example we take the permutation $\sigma = (12)$,

then we have

$$\mathrm{E}\left[X_t X_{t+x} X_{t+y}\right] \xrightarrow{\sigma} \mathrm{E}\left[X_{t+x} X_t X_{t+y}\right] \longrightarrow \mathrm{E}\left[X_t X_{t-x} X_{t+y-x}\right] \tag{5}$$

which corresponds to the condition $C(x,y) = C(-x, y-x)$. This process can be simplified slightly by only writing down the indices and introducing the function $\psi : \mathbb{R}^3 \to \mathbb{R}^2$ given by $\psi(a,b,c) \mapsto (b-a, c-a)$. Then the example done in (5) is simplified to

$$(0,x,y) \xrightarrow{\sigma} (x,0,y) \xrightarrow{\psi} (-x, y-x)$$

Now suppose we take 2×2 matrix resulting from the coefficients of the x and y variables in each coordinate. This induces a mapping from each permutation into the set of 2×2 matrices over $\mathbb{R}$. For instance in the above example we would have the matrix $\begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}$ corresponding to the permutation $\sigma = (12)$. This also establishes a correspondence between S_3 and identities on $C(x,y)$. Writing down this correspondence for each permutation in S_3 gives

$$\begin{aligned}
e &\longleftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \longleftrightarrow C(x,y) & (13) &\longleftrightarrow \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \longleftrightarrow C(x-y, -y) \\
(12) &\longleftrightarrow \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \longleftrightarrow C(-x, y-x) & (123) &\longleftrightarrow \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \longleftrightarrow C(-y, x-y) \\
(23) &\longleftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \longleftrightarrow C(y,x) & (132) &\longleftrightarrow \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \longleftrightarrow C(y-x, -x)
\end{aligned} \tag{6}$$

We see that each of these matrices is invertible, so in fact there is a map, call it ρ, from the symmetric group S_3 to the general linear group $\mathrm{GL}_2(\mathbb{R})$, the group of invertible 2×2 matrices over $\mathbb{R}$. Amazingly, there is compatibility in the multiplication, namely if we take two permutations σ and τ and construct the composite permutation $\gamma = \sigma\tau$, then in fact $\rho(\gamma) = \rho(\sigma)\rho(\tau)$. In other words ρ is a group homomorphism, and since it maps a group to the general linear group of some dimension, it is called a group representation. For example, if we take $\sigma = (12)$ and $\tau = (13)$, then $\sigma\tau$, going from right to left, is equivalent to the permutation $\gamma = (132)$, and indeed

$$\rho(\gamma) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} = \rho(\sigma)\rho(\tau)$$

This is more than just a group homomorphism—it is an injective group homomorphism since each matrix in the image is distinct. In terms of representation theory, this is called a faithful group representation. This means there is no reduction of the group structure when viewing the group in terms of its image in the matrix group. A simple example of a non-faithful representation is the trivial representation that takes every element of the group to the number 1 (which belongs to $\mathrm{GL}_1(\mathbb{R})$); in this case all group structure is lost in the image of the representation.

Matrices similar to those in (6) appear in the first chapters of most books on group representations; for instance one can find a similar set of the matrices on page 7 of

Diaconis' book [5]. Yet the derivation of these matrices typically arise in a much different context that proceeds as follows. Let $V = \{(a, b, c) \in \mathbb{R}^3 | a + b + c = 0\}$, a two dimensional subspace of $\mathbb{R}^3$, and consider the basis $\{b_1 = (1, -1, 0), b_2 = (1, 0, -1)\}$. When the coordinates of b_1 and b_2 are permuted by a permutation in S_3, the resulting vector still lies in V since the coordinates still sum to 0. The permuted basis vectors are rewritten as a linear combination of the original basis vectors, and the coefficients are extracted to form a 2×2 matrix. For example, $\sigma = (12)$ takes b_1 to $-b_1$ and b_2 to $b_2 - b_1$ giving the matrix correspondence for σ as the one in (6). In fact this representation is identical to the one described earlier, even though the general procedures giving rise to the representations are completely different.

We come back to the two questions posed in the previous section by identifying each identity in (2) with an element of S_3 via the representation above. If we take the identities represented by the permutations $\sigma = (12)$ and $\tau = (23)$, then we automatically have the identity corresponding to $\sigma\tau = (123)$, i.e. given $C(x, y) = C(-x, y - x)$ and $C(x, y) = C(y, x)$, then $C(x, y) = C(-y, x - y)$ follows automatically by applying the first identity to the second. But more identities can be produced from just σ and τ. For instance, $(13) = \tau\sigma\tau$, so if we apply the identity $C(x, y) = C(y, x)$ again to $C(-y, x - y)$ we pick up $C(x - y, -y)$, i.e. the identity corresponding to the permutation (13). Actually every identity can be produced from just the two identities corresponding to the permutations σ and τ. This is because these two transpositions generate the entire group S_3. Therefore we can conclude that requiring

$$C(x, y) = C(y, x) = C(-x, y - x)$$

is equivalent to requiring the entire string of equalities in (2), and in general, any set of equations is sufficient as long as the corresponding permutations generate all of S_3. So we have established a three-way correspondence between elements of S_3, matrices in $\mathrm{GL}_2(\mathbb{R})$, and symmetries of $C(x, y)$.

There have been several attempts at constructing lag-window functions possessing the same symmetries as in (2), i.e. constructing a lag-window $\lambda(x, y)$ that satisfies

$$\lambda(x, y) = \lambda(-x, y - x) = \lambda(y, x) = \lambda(x - y, -y) = \lambda(-y, x - y) = \lambda(y - x, -x). \quad (7)$$

For instance, the procedures in [17, 18, 10, 15, 9] all produce bivariate lag-windows as constructed from univariate lag-windows. Specifically, given any even function $\lambda(x)$,

$$\lambda(x, y) = \lambda(x)\lambda(y)\lambda(x - y) \quad (8)$$

will satisfy the requirements of (7). However this class of functions is too restrictive; we will see below that the popular Gabr-Rao optimal lag-window is excluded from this construction. A much different approach to constructing bivariate lag-windows is considered in [20], but their focus is not on symmetrizing the lag-window function.

The optimal lag-window, λ_{opt}, introduced in [17], is optimal in the sense that the variance of the bispectral estimators using this lag-window is least among a certain class of lag-windows; refer to Theorem 3 in the next section for generalization of this property. λ_{opt} is defined by the inverse Fourier transform of the optimal kernel Λ_{opt}

given by

$$\Lambda_{\text{opt}}(\omega_1,\omega_2) = \begin{cases} \frac{\sqrt{3}}{\pi^3}\left(1 - \frac{1}{\pi^2}\left(\omega_1^2+\omega_2^2+\omega_1\omega_2\right)\right), & \text{if } \omega_1^2+\omega_2^2+\omega_1\omega_2 \le \pi^2 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

This kernel, as with every kernel, is unique up to scale, i.e. the kernel $\Lambda_{\text{opt}}(x,y)$ is equivalent to the kernel $h\Lambda_{\text{opt}}(hx,hy)$ for any $h>0$. Only an approximation of the inverse Fourier transform of Λ_{opt} is provided in [17], but in [14], λ_{opt} is shown to be

$$\lambda_{\text{opt}}(\tau_1,\tau_2) = \frac{8}{\alpha(\tau_1,\tau_2)^2} J_2(\alpha(\tau_1,\tau_2))$$

where

$$\alpha(x,y) = \frac{2\pi}{\sqrt{3}}\sqrt{x^2-xy+y^2}$$

and J_2 is the second-order Bessel function of the first kind, i.e.

$$J_2(x) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{2^{2\ell+2}\ell!(2+\ell)!} x^{2\ell+2}.$$

We now ask the question: does $\lambda_{\text{opt}}(x,y)$ admit a decomposition of the form (8)? It is immediately true that no such decomposition exists for $\alpha(x,y)$ since otherwise $\alpha(0,0)=0$ would imply $\lambda(0)=0$, but then $\alpha(1,1)$ would be forced to be zero (when it's not). We extend the answer to $\lambda_{\text{opt}}(x,y)$ with the following theorem.

Theorem 1. *Suppose $\lambda(x,y)$ is a continuous function that vanishes on a set of measure zero, and suppose λ does vanish on some smooth nonlinear curve $\mathcal{C}\subset\mathbb{R}^2$. Additionally suppose $\lambda(0,0)\neq 0$. Then there is no function f such that $\lambda(x,y)=f(x)f(y)f(x-y)$.*

Proof. Suppose there is a function f such that $\lambda(x,y)=f(x)f(y)f(x-y)$. If $f(x)=0$ on some interval I, then $\lambda(x,y)=0$ for every $(x,y)\in I\times I$ contradicting the assumption that λ vanishes on a set of measure zero. We will now show that $f(x)$ is zero on some interval. Since $\lambda(0,0)=f(0)^3\neq 0$, we get that $f(0)\neq 0$. Also, $\lambda(x,x)=f(x)^2f(0)$ is continuous so $f(x)^2$ is continuous. Let (x^*,y^*) be any point on C, then $f(x^*)f(y^*)f(x^*-y^*)=0$, so at least one of $f(x^*)$, $f(y^*)$, $f(x^*-y^*)$ must be zero. As (x^*,y^*) moves continuously along $\mathcal{C}$, the smooth and nonlinear properties of $\mathcal{C}$ guarantee that x^*, y^*, x^*-y^* each simultaneously sweep out some interval of points. Continuity of f^2 implies the continuity of $|f(x^*)|$, $|f(y^*)|$, and $|f(x^*-y^*)|$, but since their product is always zero on $\mathcal{C}$, there must be some interval in which f is zero. □

Since λ_{opt} is continuous, vanishes on a set of measure zero, and vanishes on the ellipse $x^2-xy+y^2=2$, we have the following corollary.

Corollary 1. *The optimal lag-window, λ_{opt}, cannot be written as $\lambda_{opt}(x,y)=f(x)f(y)f(x-y)$ for any choice of function f.*

Now we will construct a lag-window satisfying (7) from any bivariate function $f(x,y)$. We have already seen a symmetrizing technique to create an even function in the previous section, and in taking a similar approach to this, we symmetrize $f(x,y)$ by

$$\tilde{f}(x,y) = h(f(x,y), f(-x,y-x), f(y,x), f(x-y,-y), f(-y,x-y), f(y-x,-x)) \quad (10)$$

where h is any symmetric function of its six arguments; for instance h could be a power mean like the arithmetic mean or geometric mean. The above construction can be made slightly more general by replacing $f(x,y)$ in (10) with $k(f(x,y))$ where k is any (univariate) function. On the other hand if we take just the first coordinates in each f in the above formula for $\tilde{f}$, this also produces a function that satisfies the required symmetries. That is, if we let

$$\tilde{g}(x,y) = h(g(x), g(-x), g(y), g(x-y), g(-y), g(y-x))$$

then $\tilde{g}(x,y)$ is another symmetrization built up from a generic univariate function g, and if g is assumed to be an even function, then $\tilde{g}(x,y)$ reduces to

$$\tilde{\tilde{g}}(x,y) = \tilde{h}\left(g(x), g(y), g(x-y)\right)$$

where $\tilde{h}$ is any symmetric function of three variables. Therefore we automatically get the construction (8) with the special case $\tilde{h}(x,y,z) = xyz$.

To prove that these symmetrizations $\tilde{f}$ and $\tilde{g}$ do satisfy the required symmetries, we return to the connection with group representations and also introduce the concept of group actions. If we fix any permutation in S_3 and multiply it with all the elements of S_3, the resulting action just permutes the six elements of S_3; this is a type of group action where S_3 "acts on itself". To see that it is just a permutation of the group elements, suppose $\sigma\tau = \sigma\gamma$, then multiplying both sides by σ^{-1} gives $\tau = \gamma$. Suppose we wish to test a symmetry like $\tilde{f}(x,y) = \tilde{f}(y,x)$. We see from (6) that this condition corresponds to the permutation (23), and multiplying (23) with each element of S_3 just reorders the elements of S_3, i.e.

$$(23) \cdot \{e, (12), (23), (13), (123), (132)\} = \{(23), (132), e, (123), (13), (12)\} \quad (11)$$

The righthand side corresponds to the equation

$$\tilde{f}(y,x) = h(f(y,x), f(y-x,-x), f(x,y), f(-y,x-y), f(x-y,-y), f(-x,y-x))$$

and since h is a symmetric function, this is equivalent to the original equation $\tilde{f}(x,y)$. In general, since each identity required just corresponds to a permutations of the arguments of h, each identity will be satisfied. To justify the symmetries of $\tilde{\tilde{f}}$, we note that taking the first coordinate corresponds to multiplying the matrix representation with the vector $(1 \;\; 0)$. So instead of identifying each permutation with the matrix in (6), it is identified by the 1×2 vector that is the top row of each matrix. Now if we wish to show $\tilde{g}(x,y) = \tilde{g}(y,x)$, we see that this is equivalent to multiplying each 2×1

vector by the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. So again we see that this just induces a permutation of the elements, i.e.

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \right\} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is just a permutation of the elements

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \right\}$$

So we see the same theory as before applies as every identity is associated with a permutation the arguments of $\tilde{h}$ leaving the value fixed. There is nothing special about the vector $(1\ \ 0)$, and different vectors (or more generally $n \times 2$ matrices) lead to similar formulas. For instance, using the vector $(1\ \ 1)$, we see that the function

$$g(x+y)g(x-2y)g(y-2x)$$

also satisfies the required symmetries.

4 The General Case

Here we move from the symmetric group of degree three to the symmetric group of general degree k. It is well known that any permutation can be written as a product of cycles and every cycle is a product of transpositions (2-cycles), therefore the permutations can be generated by just the transpositions. But more is true: every permutation is the product of transpositions of the form $(i, i+1)$ for $i = 1, \ldots, k-1$. To see this, suppose $i < j$, then

$$(i, j+1) = (j, j+1)(i, j)(j, j+1)$$

So any transposition, after applying the above formula enough times, can be reduced to a product of transpositions of consecutive labels. Thus the group S_k of $k!$ permutations can be represented by a much smaller subset of only $k-1$ transpositions.

Now a representation of S_k is constructed for general k. We generalize the ψ function to $\psi : \mathbb{R}^k \longrightarrow \mathbb{R}^{k-1}$ given by

$$(x_1, x_2, \ldots, x_k) \mapsto (x_2 - x_1, \ldots, x_k - x_1)$$

and the representation on S_k is produced by composing the permutation with the ψ function acted on $(0, x_1, \ldots, x_{k-1})$, then extracting the $(k-1)$-dimensional square matrix of coefficients from the image in the natural way. As an example, let $\sigma = (1234) \in S_4$, then

$$(0, x, y, z) \xrightarrow{\sigma} (z, 0, x, y) \xrightarrow{\psi} (-z, x-z, y-z)$$

So we have the correspondence

$$(1234) \longleftrightarrow \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$$

This process clearly induces a map from S_k to $(k-1)$-dimensional square matrices, but it is not clear a priori that the matrices are invertible or if this operation establishes a group homomorphism. The following theorem proves this operation is indeed a group representation, and moreover, is a faithful group representation.

Theorem 2. *The map, referred to now as ρ, from S_k to $(k-1)$-dimensional square matrices described above is a faithful group representation of dimension $k-1$.*

Proof. Let P_σ be the $k \times k$ permutation matrix associated to σ, i.e. the (i,j)-entry is 1 if σ maps i to j and 0 otherwise. It is well known that permutation matrices are invertible ($P_\sigma^{-1} = P_{\sigma^{-1}}$), in fact the map $\sigma \mapsto P_\sigma$ is an k-dimensional group representation of S_k. Let A be the matrix with ones in the first column and zeros elsewhere. Then the above mapping can be described by associated the permutation σ with P_σ then the operation of the ψ function is represented by subtracting each row of P_σ by the first row of P_σ and then discarding the top row and leftmost column. Hence the representation ρ is described compactly by the map

$$\sigma \mapsto \langle P_\sigma - AP_\sigma \rangle = \langle (I-A)P_\sigma \rangle \tag{12}$$

where the notation $\langle B \rangle$ denotes truncation of the top row and leftmost column of the matrix B. From basic block multiplication of matrices, we see that

$$\langle B_1 \rangle \langle B_2 \rangle = \langle B_1 B_2 \rangle \tag{13}$$

for any two matrices B_1 and B_2 the multiply compatibly. Also from block multiplication, given any matrix B,

$$\langle B(I-A) \rangle = \langle B \rangle \tag{14}$$

as long as the multiplication is compatible. We first show that the image of a permutation under ρ is invertible, i.e. an element of $\mathrm{GL}_{k-1}(\mathbb{R})$, by exhibiting an inverse. Specifically, we show that the inverse of the matrix $\langle (I-A)P_\sigma \rangle$ is $\left\langle (I-A)P_\sigma^{-1} \right\rangle$. Applying properties (13) and (14) shows

$$\begin{aligned}
\langle (I-A)P_\sigma \rangle \left\langle (I-A)P_\sigma^{-1} \right\rangle &= \langle I-A \rangle \langle P_\sigma (I-A) \rangle \left\langle P_\sigma^{-1} \right\rangle \\
&= \langle I-A \rangle \langle P_\sigma \rangle \left\langle P_\sigma^{-1} \right\rangle \\
&= \langle I-A \rangle \\
&= \langle I \rangle .
\end{aligned}$$

In order to prove ρ is a representation, it must be shown that for any $\sigma, \tau \in S_k$, $\rho(\sigma)\rho(\tau) = \rho(\sigma\tau)$. This is equivalent to

$$\langle (I-A)P_\sigma \rangle \langle (I-A)P_\tau \rangle = \langle (I-A)P_\sigma P_\tau \rangle \tag{15}$$

By multiplying both sides on the right by $\langle P_\tau^{-1} \rangle$ and making free use of (13), equation (15) becomes

$$\langle (I-A)P_\sigma (I-A) \rangle = \langle (I-A)P_\sigma \rangle \tag{16}$$

(16) is clearly true by noting $\langle P_\sigma (I-A) \rangle = \langle P_\sigma \rangle$. Therefore ρ is indeed a group representation.

The final task is to show ρ is faithful, i.e. ρ is injective. One easy way to prove this is to show that the only permutation being mapped to $\langle I \rangle$ is e, the identity permutation. In the language of group theory, this says that we need to show the kernel of ρ is the identity. So suppose $\langle (I - A)P_\sigma \rangle = \langle I \rangle$, then by multiplying both sides by $\langle P_\sigma^{-1} \rangle$ gives $\langle I \rangle = \langle P_\sigma^{-1} \rangle = \langle P_{\sigma^{-1}} \rangle$. This says the permutation σ^{-1} fixes the labels 2 through k, so it must also fix label 1. Hence $\sigma^{-1} = e = \sigma$, and the only permutation in the kernel of ρ is e.

□

So now that we have established ρ is a true group representation, we can extend the results from the previous section to general k. For instance if $\rho(\sigma) = (a_{i,j})$ for some $\sigma \in S_k$, then the acf of $k-1$ variables satisfies the identity

$$C(x_1, \ldots, x_{k-1}) = C\left(\sum_{j=1}^{k-1} a_{1,j} x_j \, , \, \ldots \, , \sum_{j=1}^{k-1} a_{k-1,j} x_j\right), \tag{17}$$

Also, since the transpositions $(12), (23), \ldots, (k-1, k)$ generate S_k, all of the symmetries of C can be described by (17) with just these permutations. If we consider the permutations that fix the first label, then we see that the identities in (17) induced by these $(k-1)!$ permutations are just permutations of the arguments. Therefore C, in particular, is a symmetric function.

Now with the results of Theorem 2, we can construct lag-windows functions that satisfy all of the required symmetries.

Corollary 2. *Given any function $f : \mathbb{R}^{k-1} \longrightarrow \mathbb{R}$, a symmetrization of f is constructed by starting with any symmetric function, h, of $n!$ variables and plugging-in f_σ into the variables of h (using every $\sigma \in S_k$) where, like in (17),*

$$f_\sigma(x_1, \ldots, x_{k-1}) = f\left(\sum_{j=1}^{k-1} a_{1,j} x_j \, , \, \ldots \, , \sum_{j=1}^{k-1} a_{k-1,j} x_j\right).$$

5 The Fourier Transform

We now come back to the motivation of this theory which is to construct higher-order spectral density estimates using lag-windows as in (2), i.e. estimates of the form

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{k-1}} \sum_{\|\boldsymbol{t}\| < N} \lambda(\boldsymbol{t}/M) \widehat{C}(\boldsymbol{t}) e^{-i\boldsymbol{t}\cdot\boldsymbol{\omega}}$$

There is an equivalent expression to this estimator in the frequency domain given by

$$\hat{f}(\boldsymbol{\omega}) = \Lambda * P(\boldsymbol{\omega}) = \int_{\mathbb{R}^{k-1}} \Lambda(\boldsymbol{t}/M) P(\boldsymbol{\omega} - \boldsymbol{t}) \, d\boldsymbol{t} \tag{18}$$

where Λ is the Fourier transform of λ and P is the $(k-1)^{\text{th}}$ order periodogram; namely,

$$\Lambda(\boldsymbol{t}) = \int_{\mathbb{R}^{k-1}} \lambda(\boldsymbol{t}) e^{-i\boldsymbol{\omega}\cdot\boldsymbol{t}} \, d\boldsymbol{t}$$

and

$$P(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{k-1}} \sum_{\boldsymbol{t}\in\mathbb{Z}^{k-1}} \widehat{C}(\boldsymbol{t}) e^{-i\boldsymbol{t}\cdot\boldsymbol{\omega}}$$

If we were to use the less practical estimator given in (18), then we would need a kernel Λ instead of a lag-window λ, so an understanding of the symmetries of Λ is also important. To begin, we start by considering an example in the case $k = 3$. The permutation $\sigma = (12)$ corresponds to the identity $\lambda(x, y) = \lambda(-x, y-x)$, so we consider what the substitutions $\omega_1 \mapsto -\omega_1$ and $\omega_2 \mapsto \omega_2 - \omega_1$ do to Λ;

$$\begin{aligned}\Lambda(-\omega_1, \omega_2 - \omega_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x, y) e^{-ix(-\omega_1)} e^{-iy(\omega_2-\omega_1)}\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x, y) e^{-i\omega_1(-x-y)} e^{-i\omega_2(y)}\, dx\, dy.\end{aligned}$$

After the simplification in the second line, the actions of the Fourier transform begin to unfold, i.e. the exponential kernel of the Fourier transform induced the transposed representation $x \mapsto -x - y$ and $y \mapsto y$. Now if we start with this transposed identity, i.e. $\omega_1 \mapsto -\omega_1 - \omega_2$ and $\omega_2 \mapsto \omega_2$, we find this produces a symmetry of Λ:

$$\begin{aligned}\Lambda(-\omega_1 - \omega_2, \omega_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x, y) e^{-ix(-\omega_1-\omega_2)} e^{-iy\omega_2}\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x, y) e^{-i\omega_1(-x)} e^{-i\omega_2(y-x)}\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(-x, y-x) e^{-i\omega_1 x} e^{-i\omega_2 y}\, dx\, dy \\ &= \Lambda(\omega_1, \omega_2)\end{aligned}$$

Thus we have a pseudo-representation $\tilde{\rho}$ which maps $\sigma \in S_k$ to $\rho(\sigma)'$, the transpose of $\rho(\sigma)$. For $\sigma, \tau \in S_k$, This representation satisfies

$$\tilde{\rho}(\sigma\tau) = \rho(\sigma\tau)' = (\rho(\sigma)\rho(\tau))' = \rho(\tau)'\rho(\sigma)' = \tilde{\rho}(\tau)\tilde{\rho}(\sigma)$$

which is close to the requirement of a group representation, just reversed. However if we consider multiplication of two permutations in S_k in the opposite order, from left to right, then in fact $\tilde{\rho}$ is a true group representation with the same properties as ρ. For example the multiplication (12)(13) from left to right becomes (123) and not (132) as in the right to left situation. Therefore we see that Λ possesses the "transposed symmetries" as those of λ, i.e. if $\tilde{\rho}(\sigma) = (b_{i,j})$ for some $\sigma \in S_k$, then Λ satisfies

$$\Lambda(\omega_1, \ldots, \omega_{k-1}) = \Lambda\left(\sum_{j=1}^{n-1} b_{1,j}\omega_j\,, \ldots, \sum_{j=1}^{n-1} b_{n-1,j}\omega_j\right). \tag{19}$$

In particular, like in the case of the lag-window λ, Λ is a symmetric function.

As an example, the symmetries for $k = 3$ are given by

$$\Lambda(x, y) = \Lambda(-x - y, y) = \Lambda(y, x) = \Lambda(x, -x - y) = \Lambda(y, -x - y) = \Lambda(-x - y, y)$$

A construction of the bivariate kernels from univariate kernels similar to (8) is easily derived using the theory described to become $\Lambda(x,y) = \Lambda(x)\Lambda(y)\Lambda(-x-y)$ where the only requirement on the univariate $\Lambda(x)$ function is nonnegativity.

We shall now apply this theory to generalize the Gabr-Rao optimal window, Λ_{opt}, for estimation of the trispectrum and general k^{th}-order spectra. The window must integrate to one and satisfy (19) for every $\sigma \in S_k$. Observing that the usual Gabr-Rao window is a constant plus a homogeneous polynomial of degree 2 (inside its support), we start with the construction

$$\tilde{\Lambda}_{\text{opt}}(\omega_1, \ldots, \omega_{k-1}) = c + \sum_{i \le j} c_{ij}\, \omega_i\, \omega_j \tag{20}$$

for some constants c and c_{ij}. Since Λ_{opt} is a symmetric function, equation (20) can be greatly simplified to

$$\tilde{\Lambda}_{\text{opt}}(\omega_1, \ldots, \omega_{k-1}) = c + c_1 \sum_{i=1}^{k-1} \omega_i^2 + c_2 \sum_{i<j} \omega_i\, \omega_j \tag{21}$$

Here we see the convenience in associating the symmetries of Λ to S_k. We showed earlier that it is not necessary to test all of the possible identities of Λ, just the ones that correspond to the permutations that generate S_k. Since equation (21) accounts for the symmetries corresponding to the permutations that fix 1, all that is left is to check the identity corresponding to the permutation (12), i.e. if $\tilde{\Lambda}_{\text{opt}}$ satisfies

$$\tilde{\Lambda}_{\text{opt}}(\omega_1, \ldots, \omega_{k-1}) = \tilde{\Lambda}_{\text{opt}}\left(-\sum_{i=1}^{k-1} \omega_i\,,\ \omega_2\,,\ \ldots\,,\ \omega_{k-1} \right) \tag{22}$$

then Λ_{opt} satisfies all of the necessary symmetries. Applying (22) to (21) gives

$$\tilde{\Lambda}_{\text{opt}}(\omega_1, \ldots, \omega_{k-1}) = c + c_1 \sum_{i=1}^{k-1} \omega_i^2 + c_2 \sum_{i<j} \omega_i \omega_j \tag{23a}$$

$$= c + c_1 \omega_1^2 + c_2 \omega_1 \sum_{j=2}^{k-1} \omega_j + c_1 \sum_{i=1}^{k-1} \omega_i^2 + c_2 \sum_{1<i<j} \omega_i \omega_j \tag{23b}$$

$$= c + c_1 \left(\sum_{i=1}^{k-1} \omega_i \right)^2 - c_2 \left(\sum_{i=1}^{k-1} \omega_i \right) \sum_{j=2}^{k-1} \omega_j + c_1 \sum_{i=1}^{k-1} \omega_i^2 + c_2 \sum_{1<i<j} \omega_i \omega_j \tag{23c}$$

Comparing (23b) with (23c) gives the identity

$$c_1 \omega_1^2 + c_2 \omega_1 \sum_{j=2}^{k-1} \omega_j = c_1 \left(\sum_{i=1}^{k-1} \omega_i \right)^2 - c_2 \left(\sum_{i=1}^{k-1} \omega_i \right) \sum_{j=2}^{k-1} \omega_j \tag{24}$$

There are many different ways to massage (24) into conditions on c_1 and c_2. Here we differentiate both sides of (24) with respect to ω_1 producing

$$2c_1 \omega_1 + c_2 \sum_{i=2}^{k-1} \omega_i = 2c_1 \sum_{i=1}^{n} \omega_i - c_2 \sum_{i=2}^{k-1} \omega_i$$

which is equivalent to

$$2c_1 \sum_{i=2}^{k-1} \omega_i = 2c_2 \sum_{i=2}^{k-1} \omega_i$$

Therefore we must have $c_1 = c_2$, and it is easily seen that (24) is satisfied under this condition. Thus

$$\tilde{\Lambda}_{\text{opt}}(\omega_1, \ldots, \omega_{k-1}) = \alpha \left(1 - \beta \left(\sum_{i=1}^{k-1} \omega_i^2 + \sum_{i<j} \omega_i \omega_j \right) \right) \tag{25}$$

satisfies all of the necessary symmetries. We now define $\Lambda_{\text{opt}}(\boldsymbol{\omega})$ as

$$\Lambda_{\text{opt}}(\boldsymbol{\omega}) = \tilde{\Lambda}(\boldsymbol{\omega})^+ = \begin{cases} \tilde{\Lambda}(\boldsymbol{\omega}), & \text{if } \tilde{\Lambda}(\boldsymbol{\omega}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

where β is any positive constant (kernels are unique up to scale) and α is chosen such that Λ_{opt} integrates to one.

Theorem 3. *Let $\Lambda(\boldsymbol{\omega})$ be any nonnegative kernel that integrates to one and satisfies all the necessary symmetries, i.e. satisfies* (19) *for all $\sigma \in S_k$. Also assume*

$$\int_{\mathbb{R}^{k-1}} \omega_j^2 \Lambda(\boldsymbol{\omega})\, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \omega_j^2 \Lambda_{\text{opt}}(\boldsymbol{\omega})\, d\boldsymbol{\omega} \tag{27}$$

for $j = 1, \ldots, n-1$. Then $\|\Lambda\|_{L_2} \geq \|\Lambda_{\text{opt}}\|_{L_2}$, i.e.

$$\int_{\mathbb{R}^{k-1}} \Lambda(\boldsymbol{\omega})^2\, d\boldsymbol{\omega} \geq \int_{\mathbb{R}^{k-1}} \Lambda_{\text{opt}}(\boldsymbol{\omega})^2\, d\boldsymbol{\omega}.$$

Proof. Let $\Lambda(\boldsymbol{\omega}) = \Lambda_{\text{opt}}(\boldsymbol{\omega}) + \varepsilon(\boldsymbol{\omega})$, then plugging this substitution into (27) gives

$$\int_{\mathbb{R}^{k-1}} \omega_j^2 \Lambda(\boldsymbol{\omega})\, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \omega_j^2 \left(\Lambda_{\text{opt}}(\boldsymbol{\omega}) + \varepsilon(\boldsymbol{\omega}) \right)\, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \omega_j^2 \Lambda_{\text{opt}}(\boldsymbol{\omega})\, d\boldsymbol{\omega}$$

Therefore

$$\int_{\mathbb{R}^{k-1}} \omega_j^2\, \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} = 0 \tag{28}$$

for all j. Again making use of the identity in (22) gives

$$\begin{aligned} 0 &= \int_{\mathbb{R}^{k-1}} \omega_1^2\, \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \left(\sum_{j=1}^{k-1} \omega_j \right)^2 \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} \\ &= \sum_{j=1}^{k-1} \int_{\mathbb{R}^{k-1}} \omega_j^2\, \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} + 2 \sum_{i<j} \int_{\mathbb{R}^{k-1}} \omega_i\, \omega_j\, \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} \\ &= 2 \sum_{i<j} \int_{\mathbb{R}^{k-1}} \omega_i\, \omega_j\, \varepsilon(\boldsymbol{\omega})\, d\boldsymbol{\omega} \end{aligned}$$

Therefore

$$\sum_{i<j} \int_{\mathbb{R}^{k-1}} \omega_i \, \omega_j \, \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega} = 0. \tag{29}$$

Since Λ and Λ_{opt} both integrate to one, ε must integrate to zero, i.e.

$$\int_{\mathbb{R}^{k-1}} \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega} = 0. \tag{30}$$

Computing the L_2-norm of $\Lambda(\omega)$ gives

$$\int_{\mathbb{R}^{k-1}} \Lambda(\boldsymbol{\omega})^2 \, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \Lambda_{\text{opt}}(\boldsymbol{\omega})^2 \, d\boldsymbol{\omega} + \int_{\mathbb{R}^{k-1}} \varepsilon(\boldsymbol{\omega})^2 \, d\boldsymbol{\omega} + 2 \int_{\mathbb{R}^{k-1}} \Lambda_{\text{opt}}(\boldsymbol{\omega}) \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega}$$

So if we can show that last summand is nonnegative, we will have proved the theorem. Referring to the definition of $\tilde{\Lambda}(\boldsymbol{\omega})$ in (25), we have

$$\int_{\mathbb{R}^{k-1}} \Lambda_{\text{opt}}(\boldsymbol{\omega}) \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega} = \int_{\mathbb{R}^{k-1}} \tilde{\Lambda}_{\text{opt}}(\boldsymbol{\omega}) \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega} - \int\limits_{\boldsymbol{\omega}: \tilde{\Lambda}_{\text{opt}}(\boldsymbol{\omega}) < 0} \tilde{\Lambda}_{\text{opt}}(\boldsymbol{\omega}) \varepsilon(\boldsymbol{\omega}) \, d\boldsymbol{\omega} \tag{31}$$

The first integral is zero by (28), (29), and (30). When $\tilde{\Lambda}(\boldsymbol{\omega}) < 0$, $\Lambda_{\text{opt}}(\boldsymbol{\omega}) = 0$ and $\Lambda(\boldsymbol{\omega}) \geq 0$, so we must have $\varepsilon(\boldsymbol{\omega}) \geq 0$. Therefore the second integral is less than zero making the left hand side of (31) nonnegative. □

For $k = 2$, if we let $\beta = 1/5$, then $\alpha = \frac{3}{4\sqrt{5}}$ and (26) is the familiar Epanechnikov kernel [6]. This kernel is equivalent to the Bartlett-Priestly kernel in [13] and corresponds to the quadratic spectral lag-window in [1]. For $k = 3$, $\beta = 1/\pi^2$, and $\alpha = \sqrt{3}/\pi^3$, (26) is the Gabr-Rao optimum bispectral kernel [17].

The set of points that satisfies $\tilde{\Lambda}_{\text{opt}}(\boldsymbol{\omega}) > 0$ is equivalent to

$$\sum_{i=1}^{k-1} \omega_i^2 + \sum_{i<j} \omega_i \omega_j \leq \frac{1}{\beta}$$

where the left hand side is a positive quadratic form since

$$\sum_{i=1}^{k-1} \omega_i^2 + \sum_{i<j} \omega_i \omega_j = \frac{1}{2} \sum_{i=1}^{k-1} \omega_i^2 + \frac{1}{2} \left(\sum_{i=1}^{k-1} \omega_i \right)^2 \geq 0 \tag{32}$$

and equal to zero only if $\omega_j = 0$ for all j. Therefore there always exists some transformation of coordinates to transform the ellipsoid defined by $\tilde{\Lambda}_{\text{opt}}(\boldsymbol{\omega}) > 0$ to the unit sphere, i.e. a diagonalization of the quadratic form (32) to standard form with rank and signature $k - 1$. This transformation is easily computed for any given k [2], and the Jacobian of the transformation can be used to determine α for a given β and k. For instance for $k = 4$, the substitutions

$$\begin{aligned}
\omega_1 &\longrightarrow u_1 - \frac{u_2}{\sqrt{3}} - \frac{u_3}{\sqrt{6}} \\
\omega_2 &\longrightarrow -u_1 - \frac{u_2}{\sqrt{3}} - \frac{u_3}{\sqrt{6}} \\
\omega_3 &\longrightarrow \sqrt{\frac{3}{2}} u_3
\end{aligned}$$

transforms $\omega_1^2+\omega_2^2+\omega_3^2+\omega_1\omega_2+\omega_2\omega_3+\omega_1\omega_3$ into $u_1^2+u_2^2+u_3^2$. The Jacobian of this transformation is $\sqrt{2}$, and so α is computed to be

$$\alpha = \frac{15\beta^{3/2}}{8\sqrt{2}\pi}.$$

Polar coordinates on $\mathbb{R}^{k-1}$ can be used to exactly determine α for larger k [7].

6 Bivariate Examples

The bivariate optimal kernel with corresponding lag-window are plotted below.

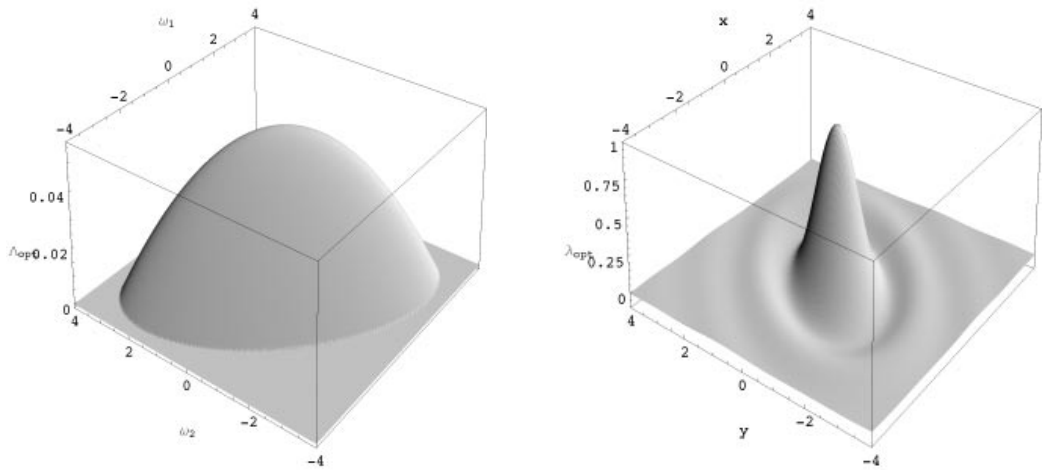

Figure 1: Plots of $\Lambda_{\text{opt}}(\omega_1,\omega_2)$ and $\lambda_{\text{opt}}(x,y)$.

The kernel $\Lambda_{\text{opt}}(\omega_1,\omega_2)$ can be derived from (10) (unlike (8) as proved by Theorem 1) by symmetrizing $f(x,y)=1-x^2-y^2$ with the symmetric function h being the arithmetic mean. Since the symmetrized f will be a quadratic form satisfying the appropriate symmetries, it must be of the form (25) for some α and β.

Now let $f(x,y)=(1-x^2-y^2)^+=\max(1-x^2-y^2,0)$. We produce three new lag-windows from the symmetric functions $h=\prod x_i$, $h=\max(x_i)$, and $h=\min(x_i)$.

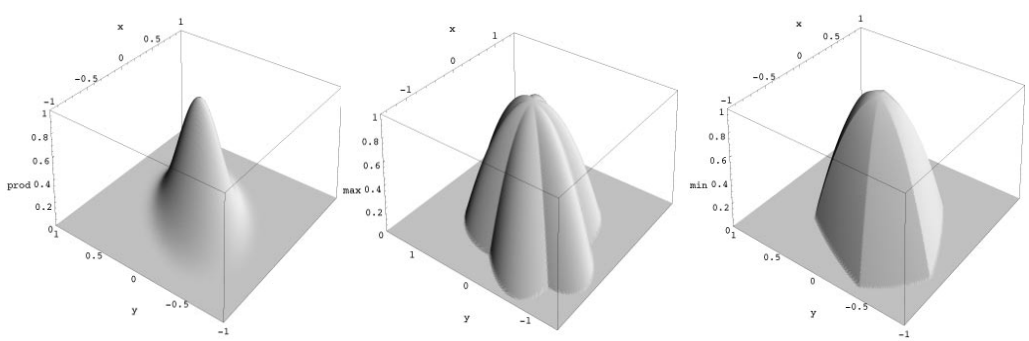

Figure 2: $\tilde{f}$ with $h=\prod x_i$, $h=\max(x_i)$, and $h=\min(x_i)$.

We conclude with two flat-top (infinite-order) lag-windows since estimation with these windows are asymptotically superior [3]. The first example is a right pyramidal frustum with the hexagonal base $|x| + |y| + |x - y| = 2$. We let $c \in (0, 1)$ be a scaling parameter that dictates when the frustum becomes flat, that is, the flat-top boundary is given by $|x| + |y| + |x - y| = 2c$. The equation of this lag-window is given by

$$\lambda_{\mathrm{rpf}}(\tau_1, \tau_2) = \frac{1}{1-c}\lambda_{\mathrm{rp}}(\tau_1, \tau_2) - \frac{c}{1-c}\lambda_{\mathrm{rp}}\left(\frac{\tau_1}{c}, \frac{\tau_2}{c}\right)$$

where λ_{p} is the equation of the right pyramid with base $|x| + |y| + |x - y| = 2$, i.e.,

$$\lambda_{\mathrm{rp}}(x, y) = \begin{cases} (1 - \max(|x|, |y|))^+, & -1 \le x, y \le 0 \text{ or } 0 \le x, y \le 1 \\ (1 - \max(|x + y|, |x - y|))^+, & \text{otherwise} \end{cases}$$

The second example is the right conical frustum with elliptical base $x^2 - xy + y^2 = 1$. Again $c \in (0, 1)$ is a scaling parameter, and the lag-window becomes flat in the ellipse $x^2 - xy + y^2 = c^2$. The equation of this lag-window is given by

$$\lambda_{\mathrm{rcf}}(\tau_1, \tau_2) = \frac{1}{1-c}\lambda_{\mathrm{rc}}(\tau_1, \tau_2) - \frac{c}{1-c}\lambda_{\mathrm{rc}}\left(\frac{\tau_1}{c}, \frac{\tau_2}{c}\right)$$

where λ_{rc} is the equation of the right cone with base $x^2 - xy + y^2 = 1$, i.e.,

$$\lambda_{\mathrm{rc}}(x, y) = (1 - \sqrt{x^2 - xy + y^2})^+$$

These two examples are plotted below with $c = 1/2$ in each case.

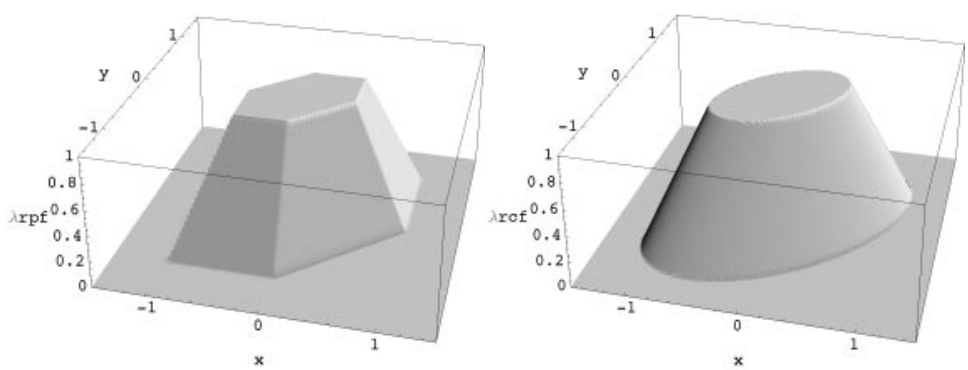

Figure 3: Lag-windows λ_{rpf} and λ_{rpf}.

7 Conclusions

This paper elicits a deep connection across permutations of S_k, symmetries of lag-windows and kernels, and invertible matrices over $\mathbb{R}$. The structure provided in group theory gives a clear understanding to many of the basic elements of higher-order spectral analysis. In particular, the symmetries of the auto-cumulant function are now well

understood, a general prescription for producing multivariate lag-windows and kernels is given, and the Gabr-Rao optimal kernel is easily generalized with the aid of basic properties of the symmetric group.

Acknowledgements. The author thanks John Hall for his helpful formulation of the representation produced in (12) and also thanks Jason Bandlow, Gregg Musiker, and Oded Yacobi for their helpful insights and discussions on the symmetric group and group representations.

References

[1] Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991. ISSN 0012-9682.

[2] Charles S. Beightler and Douglass J. Wilde. Diagonalization of quadratic forms by Gauss elimination. *Management Sci.*, 12:371–379, 1966. ISSN 0025-1909.

[3] A. Berg and D. Politis. Higher-order estimation of s-th order spectra with flat-top lag-windows. *currently unpublished.*

[4] David R. Brillinger. An introduction to polyspectra. *Ann. Math. Statist.*, 36: 1351–1374, 1965.

[5] Persi Diaconis. *Group representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988. ISBN 0-940600-14-5.

[6] V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications*, 14(1):153–158, 1969.

[7] Gerald B. Folland. *Real analysis.* Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-31716-0. Modern techniques and their applications, A Wiley-Interscience Publication.

[8] Melvin J. Hinich. Testing for Gaussianity and linearity of a stationary time series. *J. Time Ser. Anal.*, 3(3):169–176, 1982. ISSN 0143-9782.

[9] Mingquan Lu, Xianci Xiao, and Lemin Li. A general class of multidimensional lag windows for higher-order spectral estimations. In *3rd International Conference on Signal Processing*, volume 1, pages 576–579, 1996.

[10] C. L. Nikias and M. R. Raghuveer. Bispectrum estimation - A digital signal processing framework. *IEEE Proceedings*, 75:869–891, July 1987.

[11] Chrysostomos L. Nikias and Athina P. Petropulu. *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework.* PTR Prentice-Hall, 1993. ISBN 0-13-678210-8.

[12] C.L. Nikias and J.M. Mendel. Signal processing with higher-order spectra. *Signal Processing Magazine, IEEE*, 10(3):10–37, 1993.

[13] M. B. Priestley. *Spectral analysis and time series. Vol. 1.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981. ISBN 0-12-564901-0. Univariate series, Probability and Mathematical Statistics.

[14] Keiichi Saito and Tomoharu Tanaka. Exact analytic expression for gabr-rao's optimal bispectral two-dimensional lag window. *J. Nucl. Sci. Technol.*, 22(12): 1033–1035, 1985.

[15] K. Sasaki, T. Sato, and Y. Yamashita. Minimum bias windows for bispectral estimation. *Journal of Sound and Vibration*, 40:139–148, 1975.

[16] T. Subba Rao and M. M. Gabr. A test for linearity of stationary time series. *J. Time Ser. Anal.*, 1(2):145–158, 1980. ISSN 0143-9782.

[17] T. Subba Rao and M. M. Gabr. *An introduction to bispectral analysis and bilinear time series models*, volume 24 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1984. ISBN 0-387-96039-2.

[18] T. Subba Rao and Gy. Terdik. On the theory of discrete and continuous bilinear time series models. In *Stochastic processes: modelling and simulation*, volume 21 of *HandBOOK of Statist.*, pages 827–870. North-Holland, Amsterdam, 2003.

[19] Esin Turkbeyler. A three dimensional window function for the consistent estimation of the trispectrum. In *Proc. of the First Balkan Conference on Signal Processing, Communications, Circuits, and Systems*. Istanbul, Turkey, 2000.

[20] Wenbing Yao, Tianren Yao, and Han Tao. A novel method on designing bispectral windows. *Journal of Electronics (China)*, 19(9):192–198, 2001.