Estimation of Independent Component Analysis Systems

Vincent Starck[*][†]
LMU München

February 28, 2024

Abstract

I propose an approach to Independent Component Analysis (ICA) with square mixing matrix that does not require existence of higher-order moments or parametric restrictions, handles estimated sensors explicitely, and can achieve asymptotic efficiency. The estimator is shown to be consistent and asymptotically normal, with an asymptotic variance that can be consistently estimated. The optimal version of the estimator leverages results from the continuum Generalized Method of Moments of Carrasco and Florens (2000), which provides a global specification test which is valuable in many ICA applications. The method's effectiveness is illustrated through simulations, where the estimator outperforms efficient GMM and fastICA, and an application to the estimation of Structural Vector Autoregressions (SVAR), a popular model in the econometric time series literature.

Keywords: Independent Component Analysis, Structural VAR, characteristic function, continuum GMM.

1 Introduction

Independent Component Analysis (ICA; Comon (1994); Eriksson and Koivunen (2003*b*)) is a popular method which finds applications in fields as diverse as signal processing, machine learning, or structural Vector Autoregressions.

[*]Financial support from the European Research Council (Starting Grant No. 852332) is gratefully acknowledged.

[†]V.Starck@lmu.de

The common model posits that an observed vector (of 'sensors') at time t, η_t, is generated through $\eta_t = \Theta\varepsilon_t$ where ε_t has independent entries and Θ is a square, full-rank matrix. ε_t, sometimes referred as the 'sources', is a vector containing the latent factors that affects the system through the mixing matrix, Θ.

Various methods have been proposed to uncover the unmixing matrix, Θ^{-1}. Early methods typically attempted to maximize a measure of non-normality or use maximum entropy, often making use of third- and fourth-order moments. A popular and fast method based on non-Gaussianity is the fastICA algorithm (Oja and Yuan, 2006). A list of algorithms and applications can be found in in Hyvärinen, Karhunen and Oja (2001).

A broad estimation strategy is to rely on maximum likelihood or related methods. Many papers (Bach and Jordan, 2002; Samarov, Tsybakov et al., 2004; Chen, Bickel et al., 2006; Ilmonen, Paindaveine et al., 2011; Samworth, Yuan et al., 2012; Ablin, Cardoso and Gramfort, 2018) have assumed parametric, smooth, or log-concave densities to devise an estimation strategy for the unmixing matrix. Nevertheless, misspecification bias is a major concern as family of distribution is typically unknown and assumptions of smoothness, unimodality, or absence of atoms are not innocuous in applications. In addition, many of these approaches require a choice of tuning parameter or are not straightforward to implement.

I propose a nonparametric approach to estimate the unmixing matrix based on the empirical characteristic function that does not require existence of higher-order moments or restrictions on Θ's entries and can achieve asymptotic efficiency. I also explicitely allow η_t to be (consistently) estimated rather than directly observed to account for vanishing noise or an estimation step, as happens for instance in my application where the ICA system is derived through a first-step regression. Perhaps surprisingly, the approach avoids an explicit nonparametric estimation step and entails no bias-variance trade-off. The main drawback is computational cost, which may limit applicability to high-dimensional applications, at least with current capabilities. High-dimensional applications can be handled with sub-optimal but simpler weighting functions, or by turning to algorithms suited for large dimensions.

A related approach is Eriksson and Koivunen (2003*a*) (see also Chen and Bickel (2005) for the asymptotic properties), who start from the same identifying equation function to derive an estimator but work with the characteristic functions rather than its logarithm. Although this approach shares some of the benefits (such as the absence of parametric restrictions or higher-order moment assumptions), the approach developed in this paper derives straightforward asymptotics, allows for optimal weighting, and delivers a global test of the ICA system's validity.

2 Identification of Θ

Identification of systems featuring unknown linear combinations of unobserved independent variables has been extensively discussed in various places and extended to general setups (Reiersøl, 1950; Comon, 1994; Bonhomme and Robin, 2009; Ben-Moshe, 2016). Eriksson and Koivunen (2003*b*) provide a general proof of identification of unknown linear transformation of unobservables that applies to a broad range of ICA systems. Their Theorem 3.1 reads as follows:

If an observable vector X is generated through X = AS where A is a $p \times m$ constant matrix and S is a vector of independent real-valued variable, then A is identified if (i) no variable is S is normally distributed or (ii) A is of full column rank and at most one variable in S is normally distributed.

Eriksson and Koivunen (2003*b*) also establish that condition (ii) is sufficient to identify the distribution of S. In the 'square' ICA framework, Θ is a full rank-matrix so that, provided at most one source is normal, identification is achieved.

In this case, the result can also be derived by noting that two observationally equivalent systems (Θ, ε) and $(\Theta^*, \varepsilon^*)$ must satisfy $\Theta\varepsilon =^d \Theta^*\varepsilon^*$ and thus $\varepsilon =^d \Theta^{-1}\Theta^*\varepsilon^*$, where $=^d$ denotes equality in distribution. But if ε is to have independent entries, Darmois–Skitovich theorem (Darmois, 1953; Skitovitch, 1953) requires trivial linear combinations in the absence of normality. As a result, $\Theta^{-1}\Theta^*$ must be a (possibly scaled) permutation matrix.

Theorem 2.1 (Identification)**.** *Consider $\eta_t = \Theta\varepsilon_t$ where η_t is observed or consistently estimated. Θ is identified up to column scale and permutations if (i) it is invertible, and (ii) the vector ε_t is ergodic, strictly stationary, and contains independent random variables among which at most one is normal.*

3 Estimation

This section introduces the approach to the estimation of ICA systems.

I will make use of the following notation. I define $\vec{s}$ to be a $1 \times n$ row vector. φ_X denotes the characteristic function of the random vector X, *i.e.*

$\varphi_X(\vec{s}) \overset{\text{def}}{=} \mathbb{E}[e^{i\vec{s}x}]$. The j^{th} column of Θ is an $n \times 1$ vector denoted by $\Theta_{\cdot j}$; the j^{th} row of Θ is a $1 \times n$ vector denoted by $\Theta_{j\cdot}$.

3.1 Estimation based on empirical characteristic functions

By independence of sources, the observed variables' distribution is related to the distribution of their unobserved counterparts through

$$\varphi_{\eta_t}(\vec{s}) = \prod_{j=1}^{n} \varphi_{\varepsilon_{tj}}(\vec{s}\Theta_{\cdot j}) \tag{1}$$

while each source's characteristic function can be recovered from that of the sensors via

$$\varphi_{\varepsilon_{tj}}(s) = \varphi_{\eta_t}(s\Theta^{-1}_{j\cdot}) \tag{2}$$

where $\Theta^{-1}_{j\cdot}$ is the j-th row in Θ^{-1}.

A functional equation for the characteristic function of η in terms of the unknown Θ can be obtained using the last two expressions. First, define $P_j \overset{\text{def}}{=} \Theta_{\cdot j}\Theta^{-1}_{j\cdot}$ whose immediate properties are $P_j P_k = \mathbb{1}_{j=k} P_j \ \forall j, k$, $\sum_{j=1}^{n} P_j = I_n = \sum_{j=1}^{n} P_j'$, and $\text{rank}(P_j) = 1$.

In addition, the collection of P_j is isomorphic to Θ once the normalization on Θ is done (for instance, rows of Θ^{-1} are directly read in the corresponding row of each P_j when Θ has a unit diagonal). Next, substituting (2) into (1) yields an expression which implicitly relates the characteristic function of η to Θ without involving the distribution of the sources:

$$\varphi_{\eta_t}(\vec{s}) = \prod_{j=1}^{n} \varphi_{\eta_t}(\vec{s}P_j) \tag{3}$$

I propose to use the previous equation to form an estimator of Θ. As it appears from the consistency proof, it will be necessary to exclude some sequences toward degenerate matrices if η_t is to remain estimated. I make use of the following concept, which strengthens slightly the invertibility assumption by bounding the matrix an ϵ away from degeneracy.

Definition 3.1 (ϵ-invertibility)**.** Θ is ϵ-invertible if there exists $\epsilon > 0$ such that any two columns of Θ differ by an angle of at least ϵ.

Equivalently, one might bound the lowest (absolute) eigenvalue away from 0. Since identification is obtained only up to scale and column permutations, a normalization and an order is needed. From now on, I use a unit norm normalization for each column of Θ and denote the corresponding compact parameter space by $\overline{\Theta}$.

The unit norm constraint can be explicited through polar coordinates. For instance, in the 2×2 case, $\Theta = \begin{pmatrix} \cos(\gamma_1) & \cos(\gamma_2) \\ \sin(\gamma_1) & \sin(\gamma_2) \end{pmatrix}$, where γ_1, γ_2 lie between 0 and π[1]. As a result, the properties can be discussed in terms of γ, with $\Theta = \tau(\gamma)$ for some function τ. In particular, the estimator is defined through an interior solution in the γ-space.

Consider a compact neighborhood Ω of the origin on which the characteristic function does not vanish. Then equation (3) can be written in log-form. Let $f_T(\vec{s}, \gamma) \overset{\text{def}}{=} \sum_{j=0}^{n} a_j \ln\left(\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}\right)$ with $a_j = (-1)^{\mathbb{1}_{j>0}}$ and $P_0 = I$, then consider the integrated quadratic form with weighting matrix W_T:

$$Q_T(\gamma) \overset{\text{def}}{=} \frac{1}{2}\int_{\Omega} \begin{pmatrix} \Re f_T(\vec{s}, \gamma) & \Im f_T(\vec{s}, \gamma) \end{pmatrix} W_T(\vec{s}) \begin{pmatrix} \Re f_T(\vec{s}, \gamma) \\ \Im f_T(\vec{s}, \gamma) \end{pmatrix} d\vec{s} \tag{4}$$

A simple choice of W is the identity matrix, delivering the integrated modulus of f_T as a sample criterion. Alternatively, simple weighting schemes can facilitate integration as in Eriksson and Koivunen (2003*a*). Nevertheless, more refined choices of W improve efficiency by deviating from the identity (as the real and imaginary part may be estimated with different accuracy) and dependence on $\vec{s}$ (as the characteristic function is more accurately estimated near the origin and estimates are correlated).

The proposed estimator is in all cases consistent, as summarized by the following Theorem which is proven in the appendix.

Theorem 3.1 (Consistency). *The estimator $\hat{\gamma} \overset{\text{def}}{=} \arg\min_{\gamma} Q_T(\gamma)$ is consistent for γ_0 if (i) $\Theta = \Theta(\gamma)$ is ϵ-invertible, (ii) the vector of sources ε_t is ergodic, strictly stationary, and contains independent random variables among which at most one is normal, (iii) there is a (root T) consistent estimator, $\hat{\eta}_t$, of η_t, (iv) Ω is a compact neighborhood of the origin which retains identification*[2]*, and (v) W_T converges uniformly to a positive definite matrix W.*

[1] Since Θ is only identified up to columns permutations, it is convenient to fix their order. Here, this can be done easily by setting, say, $\gamma_1 < \gamma_2$, with straightforward adaptations to higher-dimensional settings by lexicographic ordering.

[2] There is a possibility of losing point identification by not using the information from the characteristic function over the whole space. Notice however that the vicinity of the origin

3.2 Asymptotic Normality

I now turn to the derivation of the asymptotic distribution of the estimator.

In the following, the argument $\vec{s}$ is left implicit for notational convenience, though all integrals are computed with respect to it. I also use f to denote the population counterpart of f_T.

Noting that differentiability of characteristic functions is ensured by existence of moments, we have the following theorem.

Theorem 3.2 (Asymptotic Normality)**.** *$\hat{\gamma}$ is asymptotically normally distributed. Specifically, if (i) $\Theta = \Theta(\gamma)$ is ϵ-invertible, (ii) the vector of sources ε_t is iid and contains independent random variables with second moments among which at most one is normal, (iii) η_t is (root T) consistently estimated, (iv) Ω is a compact neighborhood of the origin which retains identification, and (v) W_T converges uniformly to a positive definite matrix W.*

Then, $\sqrt{T}(\hat{\gamma} - \gamma_0) \to^d N(0; BVB')$ where

$$B \stackrel{\mathsf{def}}{=} \left[\int_\Omega \begin{pmatrix} \Re\frac{\partial f}{\partial \gamma}(\gamma_0) & \Im(\frac{\partial f}{\partial \gamma}(\gamma_0)\end{pmatrix} W \begin{pmatrix} (\Re\frac{\partial f}{\partial \gamma}(\gamma_0))' \\ (\Im\frac{\partial f}{\partial \gamma}(\gamma_0))' \end{pmatrix}\right]^{-1} \tag{5}$$

and

$$V \stackrel{\mathsf{def}}{=} \int_\Omega \int_\Omega \begin{pmatrix} \Re\frac{\partial f}{\partial \gamma}(\gamma_0, \vec{u}) & \Im\frac{\partial f}{\partial \gamma}(\gamma_0, \vec{u})\end{pmatrix} W(\vec{u})K(\vec{u},\vec{v})W(\vec{v})' \begin{pmatrix} (\Re\frac{\partial f}{\partial \gamma}(\gamma_0, \vec{v}))' \\ (\Im\frac{\partial f}{\partial \gamma}(\gamma_0, \vec{v}))' \end{pmatrix} \tag{6}$$

where $K(\vec{u}, \vec{v})$ is the covariance function for the real and imaginary parts of the limiting process of f_T.

As usual, the iid assumption can be weakened to ergodicity and strict stationarity. This is done with a natural adaptation of the proof, noting that results about convergence of characteristic functions have generalizations to ergodic, stationary settings (Feuerverger, 1990). The main difference is that two corrections may apply to the asymptotic variance. First, the use of a long-run variance might be warranted since uncorrelatedness of sensors does not translate to that of their empirical characteristic functions. Second, estimation of η_t must be accounted for since its disappearance hinges on the vanishing

often contains crucial information, for instance knowledge of moments when they exist, and identification can be based on the first four. Hence, it is be reasonable to assume an Ω retaining identification of γ exists though marginal distributions are not always determined by the characteristic functions on a compact.

of the term $\frac{\partial f(\gamma_0,\eta_t(\overline{\beta}))}{\partial \beta}(\hat{\beta}-\beta)$ (following notation in the proof), which relied crucially on independence.

Finally, since Θ is usually the parametrization of interest, the Delta method can be applied and asymptotic normality ensues for the corresponding estimator with the same asymptotic variance scaled by $\partial_\gamma \tau$.

3.3 Estimation of asymptotic variance

Due to the asymptotic linear representation shown while establishing asymptotic normality, the variance can be approximated via bootstrap. Alternatively, the asymptotic variance can be consistently estimated. Indeed, most terms appearing in its expression have a natural estimator based on the use of $\hat{\gamma}$ in place of γ_0 and the use of the sample counterparts of population quantities. We have

$$\frac{\partial f}{\partial \gamma'} = -\sum_{j=1}^{n} \frac{\partial \ln(\mathbb{E}[e^{i\vec{s}P_j\eta_t}])}{\partial \gamma'} = -\sum_{j=1}^{n} \frac{\mathbb{E}[e^{i\vec{s}P_j\eta_t}(\eta_t^{'} \otimes \vec{s})\frac{\partial\,\mathrm{vec}(P_j)}{\partial \gamma'}]}{\mathbb{E}[e^{i\vec{s}P_j\eta_t}]} \tag{7}$$

As an illustration of the differentiated term, consider the two-dimensional case:

$$\Theta = \begin{pmatrix} \cos(\gamma_1) & \cos(\gamma_2) \\ \sin(\gamma_1) & \sin(\gamma_2) \end{pmatrix} \tag{8}$$

Tedious but straightforward algebra yields

$$\Theta^{-1} = \frac{1}{\sin(\gamma_2-\gamma_1)} \begin{pmatrix} \sin(\gamma_2) & -\cos(\gamma_2) \\ -\sin(\gamma_1) & \cos(\gamma_1) \end{pmatrix} \tag{9}$$

$$P_1 = \frac{1}{\sin(\gamma_2-\gamma_1)} \begin{pmatrix} \cos(\gamma_1)\sin(\gamma_2) & -\cos(\gamma_1)\cos(\gamma_2) \\ \sin(\gamma_1)\sin(\gamma_2) & -\sin(\gamma_1)\cos(\gamma_2) \end{pmatrix} \tag{10}$$

$$P_2 = \frac{1}{\sin(\gamma_2-\gamma_1)} \begin{pmatrix} -\sin(\gamma_1)\cos(\gamma_2) & \cos(\gamma_1)\cos(\gamma_2) \\ -\sin(\gamma_1)\sin(\gamma_2) & \cos(\gamma_1)\sin(\gamma_2) \end{pmatrix} \tag{11}$$

$$\frac{\partial\,\mathrm{vec}(P_1)}{\partial \gamma'} = \frac{1}{\sin^2(\gamma_2-\gamma_1)} \begin{pmatrix} \sin(\gamma_2)\cos(\gamma_2) & -\sin(\gamma_1)\cos(\gamma_1) \\ \sin^2(\gamma_2) & -\sin^2(\gamma_1) \\ -\cos^2(\gamma_2) & \cos^2(\gamma_1) \\ -\sin(\gamma_2)\cos(\gamma_2) & \sin(\gamma_1)\cos(\gamma_1) \end{pmatrix} \tag{12}$$

$$\frac{\partial\,\mathrm{vec}(P_2)}{\partial \gamma'} = \frac{\partial\,\mathrm{vec}(I-P_1)}{\partial \gamma'} = -\frac{\partial\,\mathrm{vec}(P_1)}{\partial \gamma'} \tag{13}$$

Hence, replacing expectations by sample averages and using consistent estimators in place of unknown parameters delivers a consistent estimate of $\frac{\partial f}{\partial \gamma'}$.

It remains to consider the central term in more detail. The (centered) log empirical characteristic function converges to a mean zero process with covariance function $\frac{\varphi(\vec{u}+\vec{v})}{\varphi(\vec{u})\varphi(\vec{v})}-1$ and, since $\overline{\ln(\varphi(\vec{s}))} = \ln(\varphi(-\vec{s}))$, the covariance functions is enough to characterize the complex process.

With $\varphi(\vec{s}) \stackrel{\text{def}}{=} \begin{pmatrix} \varphi_\eta(\vec{s}) \\ \varphi_\eta(P_1\vec{s}) \\ \dots \\ \varphi_\eta(P_n\vec{s}) \end{pmatrix}$ and $A \stackrel{\text{def}}{=} \begin{pmatrix} a_0 & a_1 & \dots & a_n \end{pmatrix}$, consider

$$\mathbb{C}ov\left(\begin{pmatrix} \Re f(\gamma_0, \vec{u}) \\ \Im f(\gamma_0, \vec{u}) \end{pmatrix}; \begin{pmatrix} \Re f(\gamma_0, \vec{v}) \\ \Im f(\gamma_0, , \vec{v}) \end{pmatrix}\right) = (I \otimes A)\mathbb{C}ov\left(\begin{pmatrix} \Re\varphi(\vec{u}) \\ \Im\varphi(\vec{u}) \end{pmatrix}; \begin{pmatrix} \Re\varphi(\vec{v}) \\ \Im\varphi(\vec{v}) \end{pmatrix}\right)(I \otimes A')$$

Letting $\vec{u}_j \stackrel{\text{def}}{=} P_j\vec{u}$, properties of the complex-normal distribution yield the relationships

$$\mathbb{C}ov(\Re\varphi(\vec{u}_j); \Re\varphi(\vec{v}_k)) = \frac{1}{2}\Re\left(\frac{\varphi(\vec{u}_j - \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(-\vec{v}_k)} + \frac{\varphi(\vec{u}_j + \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(\vec{v}_k)} - 2\right) \tag{14}$$

$$\mathbb{C}ov(\Re\varphi(\vec{u}_j); \Im\varphi(\vec{v}_k)) = \frac{1}{2}\Im\left(-\frac{\varphi(\vec{u}_j - \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(-\vec{v}_k)} + \frac{\varphi(\vec{u}_j + \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(\vec{v}_k)}\right) \tag{15}$$

$$\mathbb{C}ov(\Im\varphi(\vec{u}_j); \Re\varphi(\vec{v}_k)) = \frac{1}{2}\Im\left(\frac{\varphi(\vec{u}_j - \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(-\vec{v}_k)} + \frac{\varphi(\vec{u}_j + \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(\vec{v}_k)} - 2\right) \tag{16}$$

$$\mathbb{C}ov(\Im\varphi(\vec{u}_j); \Im\varphi(\vec{v}_k)) = \frac{1}{2}\Re\left(\frac{\varphi(\vec{u}_j - \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(-\vec{v}_k)} - \frac{\varphi(\vec{u}_j + \vec{v}_k)}{\varphi(\vec{u}_j)\varphi(\vec{v}_k)}\right) \tag{17}$$

Hence, the use of empirical characteristic functions as an estimate of their population counterparts allows the construction of consistent estimator of the central term of the asymptotic variance.

4 Efficient estimator

4.1 Optimal objective function

Finding the optimal weighting matrix is not straightforward and actually requires a more subtle general form of the objective function. In analogy with the formation of a quadratic form for estimating equations or Generalized Method of Moments (GMM; Hansen (1982)) in the continuous realm, consider

$$Q_T^E(\gamma) = \int_\Omega \int_\Omega \begin{pmatrix} \Re f_T(\vec{s},\gamma) & \Im f_T(\vec{s},\gamma) \end{pmatrix} W_T(\vec{s},\vec{r}) \begin{pmatrix} \Re f_T(\vec{r},\gamma) \\ \Im f_T(\vec{r},\gamma) \end{pmatrix} d\vec{r} d\vec{s} \tag{18}$$

Now $W(\vec{s},\vec{r})$ possesses the necessary additional degree of freedom to hope for cancellation between A and V. Solving for the optimal weighting scheme will later allow to consider the question of the optimal Ω.

This expression is reminiscent of the extension of GMM to a continuum of moment conditions developed by Carrasco and Florens (2000). While $\Re f_T(\vec{s},\gamma)$ are $\Im f_T(\vec{s},\gamma)$ are not *per se* sample moments, they do have a zero asymptotic counterpart and the results of Carrasco and Florens (2000) can be adapted to the present framework.

Specifically, denoting real and imaginary parts of f_T by $g_j, j = 1,2$, their objective function $\|B_n(f_T(\gamma))\|$ matches Q_T^E when B_n is an integral operator $(B_n g)(\vec{s}) = \left(\sum_{l=1}^2 \int_\Omega b^{jl}(\vec{s},\vec{r}) g_l(\vec{r}) d\vec{r}\right)_{j=1,2}$ that generates a weighting matrix through $W_T^{jk}(\vec{s},\vec{r}) = \sum_{l,l'=1,2} \int_\Omega b^{jl}(\vec{u},\vec{s}) b^{jl}(\vec{u},\vec{r}) d\vec{u}$.

As Carrasco and Florens (2000) establish, efficient estimation requires inverting a covariance operator $C : h \to \int K(\vec{t},\vec{s}) h(\vec{s}) d\vec{s}$ which is not possible on the whole reference space. They propose a regularized sample version $C_T^{\alpha_T}$ where α_T is a smoothing parameter that disturbs the eigenvalues of C. The choice of α_T has been discussed in subsequent papers, see Carrasco and Kotchoni (2017) and Amengual, Carrasco and Sentana (2020).

Eventually, the optimal estimator minimizes

$$\sum_{m=1}^T \frac{\mu_{m;T}}{\mu_{m;T}^2 + \alpha_T} \left(< \varphi_{m;T}, f_T >\right)^2 \tag{19}$$

where $\mu_{m;T}$ and $\varphi_{m;T}$ are the eigenvalues and eigenfunctions of C_T. Moreover, under the assumptions of Theorem 3.2 and provided $\alpha_T \to 0$ while $\alpha_T^3 T \to \infty$, the expected simplification of the asymptotic variance occurs so that the asymptotic distribution becomes

$$\sqrt{T}(\hat{\gamma} - \gamma_0) \to^d N\left(0; \left\|\frac{\partial f}{\partial \gamma}\right\|_C^{-2}\right) =^d N(0;B) \tag{20}$$

where the weighting matrix to compute B is now based on the inverted covariance operator.

Though the more sophisticated objective function implies a more computationally intensive procedure due to integration, the estimator is in practice obtained by minimizing equation (19) and the main computational burden arises from evaluating a matrix and computing its eigen-decomposition, which can be alleviated by the use of GPU and computational techniques (see, *e.g.*, Chen and Jiang (2017)). Furthermore, using the efficient form of the estimator carries significant benefits. It removes the need to specify an arbitrary form of the weighting matrix and furthers efficiency. In particular, Carrasco and Florens (2000) argue that using a continuum of moment conditions may allow to close the efficiency gap between GMM and MLE.

Fully exploiting the distributional information implies using the full extent of zero conditions, *i.e.* asymptotically integrating over the whole space. This is backed up by the simplified form of the asymptotic variance, which decreases as Ω expands. As optimal weighting necessarily cancels infinite variances brought by the zeros of f, arbitrarily expanding the integration region becomes possible, though Carrasco and Florens (2000)'s assumption that $\mathbb{E}[\|f\|^4] < \infty$, required to derive the asymptotic distribution of the efficient estimator, is no longer a trivial assumption in presence of zeros[3] and unbounded space.

Lastly, the consistency proof suggests the rate of expansion of the diameter of Ω must be lower than root T. Investigating the optimal sequence of Ω and the trade-offs brought by finite sample considerations would require a deep analysis beyond the scope of this paper.

5 Implementation

Now, I derive the eigenfunctions and eigenvalues in details. Consider the integral operator $\hat{K}$ applied to $g : \mathbb{R}^n \to \mathbb{R}^2$

$$(Kg)(\vec{t}) = \int_\Omega \begin{pmatrix} \mathbb{C}ov(\Re f_T(\vec{s}), \Re f_T(\vec{t}))g_1(\vec{s}) + \mathbb{C}ov(\Im f_T(\vec{s}), \Re f_T(\vec{t}))g_2(\vec{s}) \\ \mathbb{C}ov(\Re f_T(\vec{s}), \Im f_T(\vec{t}))g_1(\vec{s}) + \mathbb{C}ov(\Im f_T(\vec{s}), \Im f_T(\vec{t}))g_1(\vec{s}) \end{pmatrix} d\vec{s}$$

[3]Also note that assumptions excluding zeros are common, *e.g.*, literature on nonparametric deconvolution (see Schennach (2004) and references therein).

Doing the algebra on its estimated counterpart yields

$$(\hat{K}g)(\vec{t}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \Re\sum_{k=0}^{n} a_k \left(\frac{e^{it_k\eta_\tau}}{\hat{\varphi}(t_k)} - 1\right) \\ \Im\sum_{k=0}^{n} a_k \left(\frac{e^{it_k\eta_\tau}}{\hat{\varphi}(t_k)} - 1\right)\end{pmatrix}$$
$$\times \left[\int_\Omega \Re\sum_{j=0}^{n} a_j \left(\frac{e^{i\vec{s}_j\eta_\tau}}{\hat{\varphi}(\vec{s}_j)} - 1\right) g_1(\vec{s})\ d\vec{s} + \Im\sum_{j=0}^{n} a_j \left(\frac{e^{i\vec{s}_j\eta_\tau}}{\hat{\varphi}(\vec{s}_j)} - 1\right) g_2(\vec{s})\ d\vec{s}\right]$$

This implies that the eigenfunctions g_1 and g_2 take the form

$$g_1(\vec{t}) = \frac{1}{T}\sum_{\tau=1}^{T} c_\tau \Re\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\hat{\varphi}(\vec{t}_k)} - 1\right) \tag{21}$$

and

$$g_2(\vec{t}) = \frac{1}{T}\sum_{\tau=1}^{T} c_\tau \Im\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\hat{\varphi}(\vec{t}_k)} - 1\right) \tag{22}$$

where the coefficients, $\{c_\tau^m, \tau = 1, \dots, T\}$ for $m = 1, \dots, T$, form the T eigenvectors of the matrix M with elements

$$M_{\tilde{\tau}\tau} = \Re\int_\Omega \sum_{j=0}^{n} a_j \left(\frac{e^{i\vec{s}_j\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}} - 1\right)\overline{\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{s}_k\eta_{\tilde{\tau}}}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1\right)}\ d\vec{s} \tag{23}$$

The associated eigenvalues correspond to $T\mu_{m;T}$.

The scalar products in computing the objective function then reads

$$< f_T, \varphi_{m;T} > = \int_\Omega \left(\Re\sum_{j=0}^{n} a_j \ln\left(\tfrac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}\right) \quad \Im\sum_{j=0}^{n} a_j \ln\left(\tfrac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}\right)\right)$$
$$\begin{pmatrix} \frac{1}{T}\sum_{\tau=1}^{T} c_\tau^m \Re\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1\right) \\ \frac{1}{T}\sum_{\tau=1}^{T} c_\tau^m \Im\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1\right)\end{pmatrix}\ d\vec{s}$$
$$= \frac{1}{T}\sum_{\tau=1}^{T} c_\tau^m\ \Re\int_\Omega \sum_{j=0}^{n} a_j \ln\left(\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}\right)\overline{\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1\right)}\ d\vec{s}$$

5.1 Tests

Asymptotic normality provides the basis for usual confidence intervals and tests. Moreover, an advantage of the analogy with GMM is the potential for a specification test, in the spirit of over-identifying restrictions. As detailed in Carrasco and Florens (2000), such a test can be constructed on the basis of

$$\frac{||\sqrt{T}f_T||^2_{C_T^{\alpha_T}} - \sum_{j=1}^T \frac{\mu^2_{j;T}}{\mu^2_{j;T}+\alpha_T}}{\sqrt{2\sum_{j=1}^T \frac{\mu^4_{j;T}}{(\mu^2_{j;T}+\alpha_T)^2}}} \to^d N(0;1) \tag{24}$$

provided $\alpha_T \sum_{j=1}^T \frac{\mu^4_{j;T}}{(\mu^2_{j;T}+\alpha_T)^2} \to \infty$ and the assumptions for asymptotic normality of the efficient estimator hold.

Such a test is valuable when working with ICA systems as it provides a feedback about the validity of the entire structure.

6 Simulations

I generate samples of η_t through equation $\eta_t = \Theta\epsilon_t$ with various distributions for the epsilons and a sample size of $T = 150$. I compare the performance in recovering the lag polynomial of the efficient estimator of Section 4 to that of efficient GMM based on moment conditions (*i.e.* deriving identifying equations implied by independence under the assumptions that moments up to order 4 exists, see *e.g.*, Guay and Normandin (2018)) and fastICA (Oja and Yuan, 2006). In the forthcoming tables, the corresponding estimators are denoted by log-cf, GMM, and fICA, respectively.

I consider the following distributions for the sources: student with 3 degrees of freedom, uniform on [-1;1], Binomial(20, 0.3), and Gamma(5, 1/7). All distributions are centered as to have mean zero. These distributions account for a variety of cases such as fat tails, skewness, or presence of atoms.

Consider first a student distribution with 3 degrees of freedom. In this case, the student distribution exhibits fat tails and moments higher than 2 do not exist, endangering identification strategies based on higher moments.

It appears the estimator based on log-empirical characteristic function considerably outperforms both efficient GMM and fastICA estimators when the sources are student distributed. The gains come mostly from a lower standard deviation, though there is some bias reduction especially compared to fastICA.

Now, I turn to uniform and binomial distributions. Both distribution have all their moments but one is continuous and symmetric while the other is

Table 1: Student distribution $\nu = 3$

	Bias			Standard deviation			RMSE		
$\Theta_{\cdot 1}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
0.71	−0.05	−0.02	−0.31	0.17	0.22	0.55	0.17	0.22	0.63
0.71	0.01	−0.07	0.01	0.13	0.28	0.17	0.14	0.29	0.17
$\Theta_{\cdot 2}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
−0.50	−0.03	0.03	0.11	0.17	0.28	0.55	0.18	0.28	0.56
0.87	−0.04	−0.05	−0.15	0.10	0.19	0.17	0.11	0.20	0.22

discrete and skewed. Both efficient GMM and the characteristic-function based estimator outperform fastICA for the uniform distribution. In the binomial case, the characteristic function based estimator again fares better than both efficient GMM and fastICA, with a considerable reduction in mean square error originating from lower standard deviations.

Table 2: Uniform distribution

	Bias			Standard deviation			RMSE		
$\Theta_{\cdot 1}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
0.71	−0.04	−0.10	−0.28	0.14	0.16	0.57	0.15	0.19	0.63
0.71	0.01	0.06	−0.01	0.12	0.11	0.12	0.12	0.13	0.12
$\Theta_{\cdot 2}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
−0.50	−0.01	−0.06	0.10	0.15	0.20	0.58	0.15	0.21	0.59
0.87	−0.02	−0.07	−0.17	0.08	0.11	0.12	0.08	0.13	0.21

Finally, the last tables show more contrasted results. In the case of a gamma distribution, log-cf and fICA estimators exhibit similar performance in terms of RMSE and tend to be outperformed by efficient GMM. The characteristic function based estimator occasionally displays a greater bias, which reduces its performance with these distributions, at least for some parameters.

Table 3: Binomial distribution $n = 20, p = 0.3$

	Bias			Standard deviation			RMSE		
$\Theta_{\cdot 1}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
0.71	−0.02	−0.06	−0.30	0.13	0.27	0.51	0.13	0.28	0.60
0.71	0.01	−0.07	−0.01	0.14	0.30	0.31	0.14	0.31	0.31
$\Theta_{\cdot 2}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
−0.50	−0.01	0.00	0.12	0.18	0.32	0.53	0.18	0.32	0.54
0.87	−0.02	−0.11	−0.17	0.08	0.27	0.31	0.08	0.29	0.35

Table 4: Gamma distribution $\alpha = 5, \beta = 1/7$

	Bias			Standard deviation			RMSE		
$\Theta_{\cdot 1}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
0.71	−0.20	−0.14	−0.29	0.56	0.34	0.53	0.59	0.36	0.61
0.71	−0.15	−0.02	−0.01	0.35	0.30	0.25	0.39	0.31	0.25
$\Theta_{\cdot 2}$	log-cf	GMM	fICA	log-cf	GMM	fICA	log-cf	GMM	fICA
−0.50	0.02	−0.03	0.11	0.52	0.43	0.54	0.52	0.43	0.55
0.87	−0.23	−0.16	−0.19	0.32	0.36	0.25	0.40	0.42	0.30

7 Application to SVAR

7.1 Structural Vector Autoregression

Structural Vector Autoregressions (SVAR) have attracted a lot of interest in time series econometrics since the pioneering work of Sims (1980). The standard model postulates that some observed state of the economy characterized by a vector of n variables, Y_t, is related to unobserved (stationary) shocks (*e.g.*, monetary or oil shocks) through

$$Y_t = \Theta(L)\varepsilon_t \tag{25}$$

where $\Theta(L)$ is an unknown lag polynomial that represents the impulse response function[4]. $\Theta(L)$ describes the transmission mechanism of shocks to

[4]Similarly to ICA systems, shocks as well as their effects on the system are unobserved

the economy and a subset of its column typically constitutes parameters of interest.

The first step towards estimation of $\Theta(L)$ is usually to perform the vector autoregression $A(L)Y_t = \eta_t$ to recover estimates of the innovation vector, η_t. The fundamentalness assumption states that the span of the shocks and innovations are identical and thus $\eta_t = \Theta\varepsilon_t$ for some invertible matrix Θ. It is well-known in the literature (see for instance Forni, Gambetti and Sala (2019)) that Θ corresponds to the first term in the lag polynomial $\Theta(L)$.

Provided one can identify Θ, the whole lag polynomial is recovered as $A(L)^{-1}\Theta$. As a result, the problem is reduced to the system $\eta_t = \Theta\varepsilon_t$. While standard SVAR only assumes that entries in ε_t are uncorrelated, second moments – $\Sigma_\eta = \Theta\Sigma_\varepsilon\Theta'$ – bring too few equations to solve for Θ, even after a normalization. Various solutions have been proposed, among which short-run restrictions (Sims, 1980), long-run restrictions (King et al., 1987; Blanchard and Quah, 1989; Shapiro and Watson, 1988), identification by heteroskedasticity (Rigobon, 2003; Sentana and Fiorentini, 2001; Lewis, 2019), or sign restrictions (Uhlig, 2005). A good recent reference is Kilian and Lütkepohl (2017).

Although these restrictions solve the identification problem, assuming a priori knowledge of numerous ($n(n-1)$) shocks' effects is often an issue, as the transmission mechanism of shocks to the economy is primarily an empirical question. Thus, some authors (Siegfried et al., 2002; Gourieroux, Monfort et al., 2014) have pointed out that $\eta_t = \Theta\varepsilon_t$ can be identified by assuming that ε_t contains non-Gaussian independent variables, bypassing restrictions on Θ and introducing ICA methods to the SVAR literature. Many subsequent studies have followed that road, estimating the model using high-order moments (Guay and Normandin, 2018; Keweloh, 2019), or pseudo-maximum of likelihood Gouriéroux, Monfort and Renne (2017). See also related discussions and methods in Moneta et al. (2013); Herwartz (2015); Lanne, Meitz and Saikkonen (2017).

so that there is a scale indeterminacy: shocks can be arbitrarily re-scaled to get an observationally equivalent system in which the effect of shocks are inversely re-scaled. Hence a normalization is typically imposed, for instance the unit variance normalization (each shock has variance one) or the unit effect normalization ($\Theta_{jj} = 1\ \forall j$) are popular.

7.2 Application

I consider a standard SVAR system with monthly data Y_t on real economic, oil price, and stock market growths[5] as in Keweloh (2019). The first-step vector autoregression $A(L)Y_t = \eta_t$ is performed with four lags, as suggested by Akaike's information criterion.

The study is interesting to replicate for two reasons. First, as in many SVAR studies, there might be concerns about the fundamentalness assumption. This could for instance be caused by the presence of additional shocks (*e.g.*, due to measurement error). Though robustness results against non-fundamentalness exist (Sims and Zha, 2006; Sims, 2012; Feve and Jidoud, 2012; Beaudry et al., 2015; Forni, Gambetti and Sala, 2019), it is worthwhile to see if the test detects a problem about the validity of the ICA representation.

Second, shocks might have quite fat tails in practice. For instance, Keweloh (2019) obtains excess kurtosis for all shocks and find that the shock associated to economic activity has a kurtosis above 10. Thus an estimator robust to existence of moment and able to perform accurate estimation in presence of fat tails may be useful.

The object of interest is here the lag polynomial $\Theta(L)$, rather than solely the unmixing matrix. I report the estimated responses to shock in figure 1 and display bootstraped confidence intervals.

Shocks are here subject to the unit norm normalization, so they have the same overall variance over the system. Shocks 2 and 3 have similar variance of about 89, and affect strongly economic activity. The first shock accounts for less of the disturbances to the economic system (variance of 31) and has a lower contemporaneous effect on economic activity; it seems to affect the whole system negatively after a period, but the impact is imprecisely estimated.

The over-identification test' does not reject the null (p-value 0.21), so that there is no evidence against the validity of the ICA representation.

[5] Data comes from the following sources:
https://fred.stlouisfed.org/series/INDPRO (industrial production);
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=r0000····3&f=m (oil);
https://finance.yahoo.com/quote/%5EGSPC?p=%5EGSPC (SP);
https://fred.stlouisfed.org/series/CPIAUCSL (CPI)

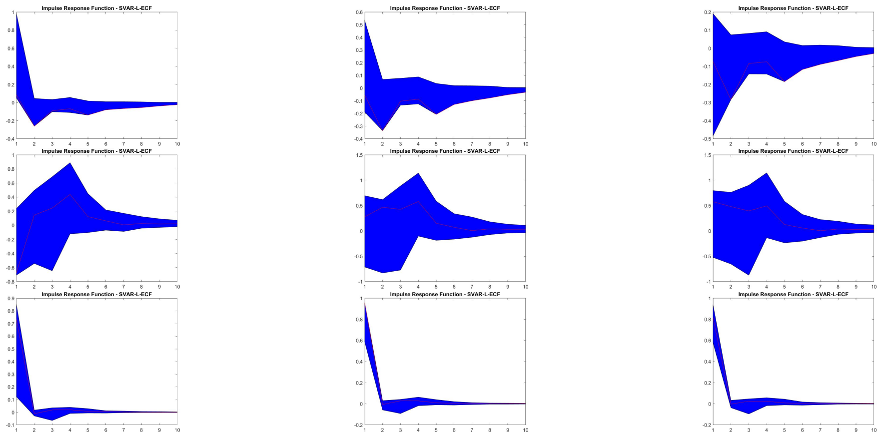

Figure 1: Plots of Impulse Responses Functions. Each column represents the 1-to-10-months impact of a shock on the S&P (first row), oil price (second row), and economic activity (third row). Shaded area depicts 90% bootstrap confidence interval.

Acknowledgements

I thank Susanne Schennach for many helpful comments. I am also grateful to participants at Brown university's lunch seminar for constructive discussions while presenting an early version of the paper.

References

Ablin, Pierre, Jean-François Cardoso, and Alexandre Gramfort. 2018. "Faster independent component analysis by preconditioning with Hessian approximations." *IEEE Transactions on Signal Processing*, 66(15): 4040–4049.

Amengual, Dante, Marine Carrasco, and Enrique Sentana. 2020. "Testing distributional assumptions using a continuum of moments." *Journal of Econometrics*, 218(2): 655–689.

Bach, Francis R, and Michael I Jordan. 2002. "Kernel independent component analysis." *Journal of machine learning research*, 3(Jul): 1–48.

Beaudry, Paul, Patrick Fève, Alain Guay, and Franck Portier. 2015. "When is nonfundamentalness in VARs a real problem? An application to news shocks." National Bureau of Economic Research.

Ben-Moshe, Dan. 2016. "Identification and estimation of coefficients in dependent factor models." Working paper, The Hebrew University of Jerusalem.

Blanchard, Olivier J, and Danny Quah. 1989. "The dynamic effects of aggregate demand and aggregate supply." *The American Economic Review*, 79(4): 655–673.

Bonhomme, Stéphane, and Jean-Marc Robin. 2009. "Consistent noisy independent component analysis." *Journal of Econometrics*, 149(1): 12–25.

Carrasco, Marine, and Jean-Pierre Florens. 2000. "Generalization of GMM to a continuum of moment conditions." *Econometric Theory*, 797–834.

Carrasco, Marine, and Rachidi Kotchoni. 2017. "Efficient estimation using the characteristic function." *Econometric Theory*, 33(2): 479–526.

Chen, Aiyou, and Peter J Bickel. 2005. "Consistent independent component analysis and prewhitening." *IEEE Transactions on Signal Processing*, 53(10): 3625–3632.

Chen, Aiyou, Peter J Bickel, et al. 2006. "Efficient independent component analysis." *The Annals of Statistics*, 34(6): 2825–2855.

Chen, Lu-Hung, and Ci-Ren Jiang. 2017. "Multi-dimensional functional principal component analysis." *Statistics and Computing*, 27(5): 1181–1192.

Comon, Pierre. 1994. "Independent component analysis, a new concept?" *Signal processing*, 36(3): 287–314.

Csorgo, Sandor. 1981. "Limit behaviour of the empirical characteristic function." *The Annals of Probability*, 130–144.

Darmois, George. 1953. "Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire." *Revue de l'Institut international de statistique*, 2–8.

Eriksson, Jan, and Visa Koivunen. 2003*a*. "Characteristic-function-based independent component analysis." *Signal Processing*, 83(10): 2195–2208.

Eriksson, Jan, and Visa Koivunen. 2003*b*. "Identifiability and separability of linear ICA models revisited." *Proc. of ICA*, 2003: 23–27.

Feuerverger, Andrey. 1990. "An efficiency result for the empirical characteristic function in stationary time-series models." *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 155–161.

Feuerverger, Andrey, Roman A Mureika, et al. 1977. "The empirical characteristic function and its applications." *The annals of Statistics*, 5(1): 88–97.

Feve, Patrick, and Ahmat Jidoud. 2012. "Identifying news shocks from SVARs." *Journal of Macroeconomics*, 34(4): 919–932.

Forni, Mario, Luca Gambetti, and Luca Sala. 2019. "Structural VARs and noninvertible macroeconomic models." *Journal of Applied Econometrics*, 34(2): 221–246.

Gouriéroux, Christian, Alain Monfort, and Jean-Paul Renne. 2017. "Statistical inference for independent component analysis: Application to structural VAR models." *Journal of Econometrics*, 196(1): 111–126.

Gourieroux, Christian, Alain Monfort, et al. 2014. "Revisiting identification and estimation in structural VARMA models."

Guay, Alain, and Michel Normandin. 2018. "Identification of Structural Vector Autoregresions Through Higher Unconditional Moments." *Unpublished Working Paper.[2, 11].*

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica: Journal of the Econometric Society*, 1029–1054.

Herwartz, H. 2015. "Structural VAR Modelling with Independent Innovations—An Analysis of Macroeconomic Dynamics in the Euro Area Based on a Novel Identification Scheme." *University of Göttingen, Working Paper.*

Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. 2001. "al. Independent component analysis. John Willey ans Sons." *Inc., New York.*

Ilmonen, Pauliina, Davy Paindaveine, et al. 2011. "Semiparametrically efficient inference based on signed ranks in symmetric independent component models." *the Annals of Statistics*, 39(5): 2448–2476.

Keweloh, Sascha Alexander. 2019. "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments." *Journal of Business & Economic Statistics*, 1–11.

Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural vector autoregressive analysis.* Cambridge University Press.

King, Robert G, Charles I Plosser, James H Stock, and Mark W Watson. 1987. "Stochastic trends and economic fluctuations." National Bureau of Economic Research.

Lanne, Markku, Mika Meitz, and Pentti Saikkonen. 2017. "Identification and estimation of non-Gaussian structural vector autoregressions." *Journal of Econometrics*, 196(2): 288–304.

Lewis, Daniel. 2019. "Identifying shocks via time-varying volatility." *FRB of New York Staff Report*, , (871).

Moneta, Alessio, Doris Entner, Patrik O Hoyer, and Alex Coad. 2013. "Causal inference by independent component analysis: Theory and applications." *Oxford Bulletin of Economics and Statistics*, 75(5): 705–730.

Newey, KW, and D McFadden. 1994. "Large sample estimation and hypothesis." *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, 2112–2245.

Oja, Erkki, and Zhijian Yuan. 2006. "The FastICA algorithm revisited: Convergence analysis." *IEEE Transactions on Neural Networks*, 17(6): 1370–1381.

Reiersøl, Olav. 1950. "Identifiability of a linear relation between variables which are subject to error." *Econometrica: Journal of the Econometric Society*, 375–389.

Rigobon, Roberto. 2003. "Identification through heteroskedasticity." *Review of Economics and Statistics*, 85(4): 777–792.

Samarov, Alexander, Alexandre Tsybakov, et al. 2004. "Nonparametric independent component analysis." *Bernoulli*, 10(4): 565–582.

Samworth, Richard J, Ming Yuan, et al. 2012. "Independent component analysis via nonparametric maximum likelihood estimation." *The Annals of Statistics*, 40(6): 2973–3002.

Schennach, Susanne M. 2004. "Estimation of nonlinear models with measurement error." *Econometrica*, 72(1): 33–75.

Sentana, Enrique, and Gabriele Fiorentini. 2001. "Identification, estimation and testing of conditionally heteroskedastic factor models." *Journal of econometrics*, 102(2): 143–164.

Shapiro, Matthew D, and Mark W Watson. 1988. "Sources of business cycle fluctuations." *NBER Macroeconomics annual*, 3: 111–148.

Siegfried, Nikolaus A, et al. 2002. "An information-theoretic extension to structural VAR modelling." Hamburg University, Department of Economics.

Sims, Christopher A. 1980. "Macroeconomics and reality." *Econometrica: journal of the Econometric Society*, 1–48.

Sims, Christopher A, and Tao Zha. 2006. "Does monetary policy generate recessions?" *Macroeconomic Dynamics*, 10(2): 231–272.

Sims, Eric R. 2012. "News, non-invertibility, and structural VARs." *Advances in Econometrics*, 28: 81.

Skitovitch, Viktor P. 1953. "On a property of the normal distribution." *DAN SSSR*, 89: 217–219.

Uhlig, Harald. 2005. "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics*, 52(2): 381–419.

8 Appendix A: proofs

8.1 Theorem 3.1 (Consistency)

Proof. Consistency follows by verifying assumptions of Theorem 2.1 in Newey and McFadden (1994). The parameter space is compact by construction, identification is been established under (i), (ii), (iii), and (iv), and the limiting objective function is continuous by inspection under compactness. It remains to show uniform convergence in probability.

The empirical characteristic function using a consistent estimator of η_t converges uniformly in probability:

$$
\begin{aligned}
\sup_{\Theta\in\overline{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\hat{\eta}_t}-\mathbb{E}[e^{i\vec{s}P_j\eta_t}]\right| &= \sup_{\Theta\in\overline{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\hat{\eta}_t}-\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\eta_t}+\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\eta_t}-\mathbb{E}[e^{i\vec{s}P_j\eta_t}]\right| \\
&\leq \sup_{\Theta\in\overline{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\hat{\eta}_t}-\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\eta_t}\right|+\sup_{\Theta\in\overline{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\eta_t}-\mathbb{E}[e^{i\vec{s}P_j\eta_t}]\right| \\
&\leq \sup_{\Theta\in\overline{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}e^{i\vec{s}P_j\eta_t}(e^{i\vec{s}P_j(\hat{\eta}_t-\eta_t)}-1)\right|+o_p(1) \\
&\leq \sup_{\Theta\in\overline{\Theta}}\frac{1}{T}\sum_{t=1}^{T}\left|e^{i\vec{s}P_j(\hat{\eta}_t-\eta_t)}-1\right|+o_p(1) \\
&\leq \sup_{\Theta\in\overline{\Theta}}\frac{1}{T}\sum_{t=1}^{T}|\vec{s}P_j(\hat{\eta}_t-\eta_t)|+o_p(1) \\
&\leq \sup_{\Theta\in\overline{\Theta}}\frac{1}{T}\sum_{t=1}^{T}\text{diameter}(\Omega)\frac{1}{\cos(\frac{\pi}{2}-\epsilon)}|\hat{\eta}_t-\eta_t|+o_p(1) \\
&\to^p 0
\end{aligned}
$$

where the second inequality follows from the uniform law of large numbers, the fourth from $|e^{ix}-1|\leq|x|$, the fifth from ϵ-invertibility, and the convergence is implied by compactness of Ω and consistency of $\hat{\eta}_t$.

Hence, by Theorem 2.1 in Newey and McFadden (1994), $\hat{\gamma}\to^p\gamma_0$. □

8.2 Theorem 3.2 (Asymptotic Normality)

Proof. Derivation of asymptotic normality follows the approach in Newey and McFadden (1994). By dominated convergence, first-order conditions read

$$\int_\Omega \begin{pmatrix}\Re\frac{\partial f_T}{\partial\gamma}(\hat\gamma) & \Im(\frac{\partial f_T}{\partial\gamma}(\hat\gamma))\end{pmatrix} W_T \begin{pmatrix}\Re f_T(\hat\gamma)\\ \Im f_T(\hat\gamma)\end{pmatrix} d\vec{s} = 0 \tag{26}$$

Applying the mean-value theorem around the true value γ_0 yields

$$\int_\Omega \begin{pmatrix}\Re\frac{\partial f_T}{\partial\gamma}(\hat\gamma) & \Im(\frac{\partial f_T}{\partial\gamma}(\hat\gamma))\end{pmatrix} W_T \begin{pmatrix}\Re f_T(\gamma_0) + (\Re\frac{\partial f_T}{\partial\gamma}(\tilde\gamma_R))'(\hat\gamma-\gamma_0)\\ \Im f_T(\gamma_0) + (\Im\frac{\partial f_T}{\partial\gamma}(\tilde\gamma_I))'(\hat\gamma-\gamma_0)\end{pmatrix} d\vec{s} = 0$$

so that, rearranging

$$\sqrt{T}(\hat\gamma-\gamma_0) = -\left[\int_\Omega \begin{pmatrix}\Re\frac{\partial f_T}{\partial\gamma}(\hat\gamma) & \Im(\frac{\partial f_T}{\partial\gamma}(\hat\gamma))\end{pmatrix} W_T \begin{pmatrix}(\Re\frac{\partial f_T}{\partial\gamma}(\tilde\gamma_R))'\\ (\Im\frac{\partial f_T}{\partial\gamma}(\tilde\gamma_I))'\end{pmatrix} d\vec{s}\right]^{-1}$$
$$\sqrt{T}\int_\Omega \begin{pmatrix}\Re\frac{\partial f_T}{\partial\gamma}(\hat\gamma) & \Im(\frac{\partial f_T}{\partial\gamma}(\hat\gamma))\end{pmatrix} W_T \begin{pmatrix}\Re f_T(\gamma_0)\\ \Im f_T(\gamma_0)\end{pmatrix} d\vec{s}$$

Consider the right-hand side. The first term can be proven to converge uniformly in probability, proceeding as in the uniform convergence step in proving consistency. For the second term, the empirical log-characteristic function converges to a complex normal stochastic process. This follows from convergence of finite dimensional distributions (by the multivariate central limit theorem and the delta method) and tightness (tightness was proven for the empirical characteristic function by Feuerverger, Mureika et al. (1977)[6]. Since they also proved almost sure convergence of $\sup_{-K\leq s\leq K}|c_n(s)-c(s)|$, where c denotes the empirical characteristic function and c_n its empirical counterpart, the empirical characteristic function is almost surely bounded away from zero on Ω for T large enough, and tightness follows).

Then convergence of $\sqrt{T}(\ln(\frac{1}{T}\sum_{i=1}^T e^{i\vec{v}\eta_t}) - \ln(\mathbb{E}[e^{i\vec{v}\eta_t}]))$ to a complex normal stochastic process together with the continuous mapping theorem and the condition $\sum_{j=0}^n a_j \ln(\mathbb{E}[e^{i\vec{s}P_j\eta_t}])) = 0$ deliver asymptotic normality. If η_t is known, we directly obtain

$$\sqrt{T}(\hat\gamma-\gamma_0) \to^d N(0; BVB')$$

Interestingly, estimation of η_t does not affect the asymptotic variance thanks to assumption (ii). Indeed, estimation of η_t can be accounted for by expanding $f(\gamma_0,\hat\eta_t)$ into $f(\gamma_0,\eta_t(\beta_0)) + \frac{\partial f(\gamma_0,\eta_t(\overline{\beta}))}{\partial\beta}(\hat\beta-\beta)$ by the mean-value theorem, where β is the underlying parameter vector in estimating η_t. The first term

[6] As pointed out by Csorgo (1981), the result requires slightly stronger conditions than initially thought. Existence of moments larger than 1 suffices.

corresponds to the case where η_t is observed. If η_t is an error term independent from its regressors w_t, as in the case of the VAR since $\eta_t = \Theta\varepsilon_t$ is independent of the past under assumption (ii), then, by the law of iterated expectations and properties of the P_j, $\frac{\partial f(\gamma_0,\overline{\beta})}{\partial \beta} = \sum_{j=0}^{n} a_j \frac{\mathbb{E}[e^{i\vec{s}P_j\eta_t}(-w_t)P_j'\vec{s}']}{\mathbb{E}[e^{i\vec{s}P_j\eta_t}]} = -\sum_{j=0}^{n} a_j \frac{\mathbb{E}[\mathbb{E}[e^{i\vec{s}P_j\eta_t}|w_t]w_tP_j'\vec{s}']}{\mathbb{E}[e^{i\vec{s}P_j\eta_t}]} = -\mathbb{E}[w_t](\sum_{j=0}^{n} a_jP_j')\vec{s}' = 0$.

□

Appendix B: Additional derivations

I derive the eigenfunctions and eigenvalues in details. Consider the estimated counterpart of the integral operator K:

$$
\begin{aligned}
(\hat{K}g)(\vec{t}) &= \int_\Omega \begin{pmatrix} \widehat{\mathbb{C}ov}(\Re f_T(\vec{s}), \Re f_T(\vec{t}))g_1(\vec{s}) + \widehat{\mathbb{C}ov}(\Im f_T(\vec{s}), \Re f_T(\vec{t}))g_2(\vec{s}) \\ \widehat{\mathbb{C}ov}(\Re f_T(\vec{s}), \Im f_T(\vec{t}))g_1(\vec{s}) + \widehat{\mathbb{C}ov}(\Im f_T(\vec{s}), \Im f_T(\vec{t}))g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} \frac{a_k a_j}{2} \\
&\times \int_\Omega \begin{pmatrix} \Re\left(\frac{\hat\varphi(\vec{s}_j-\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{\hat\varphi(\vec{s}_j+\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)} - 2\right) g_1(\vec{s}) + \Im\left(\frac{\hat\varphi(\vec{s}_j-\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{\hat\varphi(\vec{s}_j+\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_2(\vec{s}) \\ \Im\left(-\frac{\hat\varphi(\vec{s}_j-\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{\hat\varphi(\vec{s}_j+\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_1(\vec{s}) + \Re\left(\frac{\hat\varphi(\vec{s}_j-\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} - \frac{\hat\varphi(\vec{s}_j+\vec{t}_k)}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
&= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{k=0}^{n}\sum_{j=0}^{n} \frac{a_k a_j}{2} \\
&\times \int_\Omega \begin{pmatrix} \Re\left(\frac{e^{i(\vec{s}_j-\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{e^{i(\vec{s}_j+\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)} - 2\right) g_1(\vec{s}) + \Im\left(\frac{e^{i(\vec{s}_j-\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{e^{i(\vec{s}_j+\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_2(\vec{s}) \\ \Im\left(-\frac{e^{i(\vec{s}_j-\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} + \frac{e^{i(\vec{s}_j+\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_1(\vec{s}) + \Re\left(\frac{e^{i(\vec{s}_j-\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(-\vec{t}_k)} - \frac{e^{i(\vec{s}_j+\vec{t}_k)\eta_\tau}}{\hat\varphi(\vec{s}_j)\hat\varphi(\vec{t}_k)}\right) g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
=& \frac{1}{T}\sum_{\tau=1}^{T}\sum_{k=0}^{n}\sum_{j=0}^{n} a_k a_j \\
&\times \int_\Omega \begin{pmatrix} \Re\left(\left(\frac{e^{-i\vec{t}_k\eta_\tau}}{\hat\varphi(-\vec{t}_k)} + \frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\right)\frac{e^{i\vec{s}_j\eta_\tau}}{2\hat\varphi(\vec{s}_j)} - 1\right) g_1(\vec{s}) + \Im\left(\left(\frac{e^{-i\vec{t}_k\eta_\tau}}{\hat\varphi(-\vec{t}_k)} + \frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\right)\frac{e^{i\vec{s}_j\eta_\tau}}{2\hat\varphi(\vec{s}_j)} g_2(\vec{s})\right) \\ \Im\left(-\frac{e^{-i\vec{t}_k\eta_\tau}}{\hat\varphi(-\vec{t}_k)} + \frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\right)\frac{e^{i\vec{s}_j\eta_\tau}}{2\hat\varphi(\vec{s}_j)} g_1(\vec{s}) + \Re\left(\left(\frac{e^{-i\vec{t}_k\eta_\tau}}{\hat\varphi(-\vec{t}_k)} - \frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\right)\frac{e^{i\vec{s}_j\eta_\tau}}{2\hat\varphi(\vec{s}_j)}\right) g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
=& \frac{1}{T}\sum_{\tau=1}^{T}\sum_{k=0}^{n}\sum_{j=0}^{n} a_k a_j \int_\Omega \begin{pmatrix} \left(\Re\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\Re\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_1(\vec{s}) + \Re\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\Im\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} g_2(\vec{s}) \\ \Im\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\Im i\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} g_1(\vec{s}) + \Im\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)}\Re(-i)\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
=& \frac{1}{T}\sum_{\tau=1}^{T}\sum_{k=0}^{n}\sum_{j=0}^{n} a_k a_j \\
&\times \int_\Omega \begin{pmatrix} \left(\Re\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)} - 1\right)\left(\Re\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_1(\vec{s}) + \left(\Re\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)} - 1\right)\left(\Im\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_2(\vec{s}) \\ \left(\Im\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)} - 1\right)\left(\Re\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_1(\vec{s}) + \left(\Im\frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)} - 1\right)\left(\Im\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_2(\vec{s}) \end{pmatrix} \ d\vec{s} \\
&= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \Re\sum_{k=0}^{n} a_k\left(\frac{e^{it_k\eta_\tau}}{\hat\varphi(t_k)} - 1\right) \\ \Im\sum_{k=0}^{n} a_k\left(\frac{e^{it_k\eta_\tau}}{\hat\varphi(t_k)} - 1\right) \end{pmatrix} \\
&\times \left[\int_\Omega \Re\sum_{j=0}^{n} a_j\left(\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_1(\vec{s}) \ d\vec{s} + \Im\sum_{j=0}^{n} a_j\left(\frac{e^{i\vec{s}_j\eta_\tau}}{\hat\varphi(\vec{s}_j)} - 1\right) g_2(\vec{s}) \ d\vec{s}\right]
\end{aligned}
$$

noting that $\frac{1}{T}\sum_\tau \frac{e^{i\vec{t}_k\eta_\tau}}{\hat\varphi(\vec{t}_k)} = 1$.

This implies that the eigenfunctions g_1 and g_2 take the form

$$g_1(\vec{t}) = \frac{1}{T}\sum_{\tau=1}^{T} c_\tau \Re \sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\hat{\varphi}(\vec{t}_k)} - 1 \right) \tag{27}$$

and

$$g_2(\vec{t}) = \frac{1}{T}\sum_{\tau=1}^{T} c_\tau \Im \sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{t}_k\eta_\tau}}{\hat{\varphi}(\vec{t}_k)} - 1 \right) \tag{28}$$

Substituting these in the system $(\hat{K}g)(\vec{t}) = \mu g(\vec{t})$, it follows that the coefficients $\{c_\tau^m, \tau = 1, \dots, T\}$ for $m = 1, \dots, T$, form the T eigenvectors of the matrix M with elements

$$M_{\tilde{\tau}\tau} = \Re \int_\Omega \sum_{j=0}^{n} a_j \left(\frac{e^{i\vec{s}_j\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}} - 1 \right) \overline{\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{s}_k\eta_{\tilde{\tau}}}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1 \right)}\, d\vec{s}$$

The associated eigenvalues correspond to $T\mu_{m;T}$.

If one desires to integrate on $\mathbb{R}^n$, then we can consider integral with respect to another measure than Lebesgue's. A leading possibility, as in Carrasco and Kotchoni (2017), is to use a density function π as weight.

One then needs

$$M_{\tilde{\tau}\tau} = \Re \int_{\mathbb{R}^n} \sum_{j=0}^{n} a_j \left(\frac{e^{i\vec{s}_j\eta_\tau}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_j\eta_t}} - 1 \right) \overline{\sum_{k=0}^{n} a_k \left(\frac{e^{i\vec{s}_k\eta_{\tilde{\tau}}}}{\frac{1}{T}\sum_{t=1}^{T} e^{i\vec{s}P_k\eta_t}} - 1 \right)}\, \pi(\vec{s})\, d\vec{s}$$

and, to avoid integrating on $\mathbb{R}^n$, one can use a change of variable such as $s_i = \tan(\tilde{s}_i)$ coordinate-wise.

Note:

- M/T has eigenvalues μ, eigenvectors E.

$\hat{\phi} = \frac{1}{T}\sum_{\tau=1}^{T} c_\tau (\sum_{k=0}^{n} a_k (\frac{e^{i\vec{s}P_k\eta_\tau}}{T^{-1}\sum_\tau^T e^{i\vec{s}P_k\eta_\tau}} - 1))/\|\phi\| = E'B/T/\|\phi\|$ is a a normalized eigenfunction, using E as the eigenvector. B is the base, i.e., the collection of $(\sum_{k=0}^{n} a_k (\frac{e^{i\vec{s}P_k\eta_\tau}}{T^{-1}\sum_\tau^T e^{i\vec{s}P_k\eta_\tau}} - 1))$.

- $\|\phi\| = \sqrt{E'ME/T} = \sqrt{\mu}$
- $< \hat{\phi}, f >= \int f E'B/T/\sqrt{\mu}\Pi(d\vec{s})$
- $Q = 1/T^2 \sum_\tau \frac{1}{\mu_\tau^2 + \alpha} |E'_\tau \int f \overline{B}|^2$.

Complex-valued derivation

$$
\begin{aligned}
(Kg)(\vec{t}) &= \int \sum_j \sum_k a_j a_k \sum_\tau \left(\frac{e^{-\vec{t}P_k\eta_\tau} e^{i\vec{s}P_j\eta_\tau}}{T\hat{\varphi}(-\vec{t}_k)\hat{\varphi}(\vec{s}_j)} - 1 \right) g(\vec{s})\ d\vec{s} \\
&= \frac{1}{T} \sum_\tau \sum_k a_k \left(\frac{e^{-\vec{t}P_k\eta_\tau}}{\frac{1}{T}\sum_\tau e^{-\vec{t}P_k\eta_\tau}} - 1 \right) \int_\Omega \sum_j a_j \left(\frac{e^{i\vec{t}P_j\eta_\tau}}{\frac{1}{T}\sum_\tau e^{i\vec{t}P_j\eta_\tau}} - 1 \right) g(\vec{s})\ d\vec{s}
\end{aligned}
\tag{29}
$$

so eigenfunctions are linear combinations of the $\sum_{k=1}^n a_k(\frac{e^{-i\vec{t}P_k\eta_\tau}}{\frac{1}{T}\sum_{\tau=1}^T e^{-i\vec{t}P_k\eta_\tau}} - 1)$.