Fourier-based Systems Regression

James Rising

September 24, 2014

Abstract

An approach to solving completely endogenous systems of equations is presented, using Fourier representations.

1 Introduction

Linear regressions form the basis for much of the analysis of relationships in economics. While linear regressions produce easily intelligable results, the assumptions of linearity and exogeneity are generally difficult to support. This paper describes a alternative method of analyzing the relationships that exist within systems, drawing on techniques from signal processing.

There are a number of problems with these assumptions:

- Endogeneity is ubiquitious. Most analysis requires the identification of exogenous variables, but within an integrated system, processes are rarely independent.
- Measurement error is ubiquitous. Measurement error in otherwise exogenous variables result in a violation of the exogeneity assumption.
- The role of time and space must be constructed by the experimenter.
- Linearity
- Philosophically, as designers of the experiment, we are part of the system, and everything is correlated with our own error terms.

Systems approaches are very different from classical analyses. A system is a collection of elements, in which each element can influence every other element. This framework presents a completely different set of behaviors from the classical pysical sciences. As Von Bertalanffy notes,

> Concepts like those of organization, wholeness, directiveness, teleology, and differentiation are alien to conventional physics. However, they pop up everywhere in the biological, behavioral, and social sciences, and are, in fact, indispensable for dealing with living organisms or social groups. (34)

Finally, correlation and exogeneity are inappropriate concepts for systems.

The proposed method relies on a signal processing framework which relies heavily on Fourier analysis. This framework has many advantages:

OLS Compatibility Linear OLS models are natural to describe and to generalize as a signal processing system.

Explicit Time Time is given a unique role within the framework.[1]

Endogenous Frequencies Cycles are a tell-tale sign of cybernetic endogeneity, and Fourier analysis makes these explicit. To the extent to which different variables exhibit the same frequencies, they are likely to be related by feedback mechanisms.

1.1 Signal Processing Framework

Let $y\{t\}$ be a erratically sampled variable, or scattered function, with n data points. That is, for each time t_k (for $k = 1, 2, \ldots, n$), the value taken by the variable $y\{t\}$ is y_k. Moreover, allow $y\{t\}$ to not be well-defined except at the points $\{t_k\}_{k=1}^{N}$ (like a discrete-time function). Note that each scattered function has a unique corresponding range. Let $\vec{y}$ and $\vec{t}$ denote row vectors of these corresponding values. Formally, $y\{t\}\colon \{t_1, \ldots, t_n\} \to \{y_1, \ldots, y_n\}$.

Similarly, let $Y\{\omega\}$ be a scattered Fourier approximation to $y\{t\}$. That is, at a each ω_k in a collection of frequencies $\omega_k{}_{k=1}^{M}$, $Y\omega_k$ is defined as a complex value. The row vectors $\vec{Y}$ and $\vec{\omega}$ contain these values.

Below, when $y_i\{t\}$, $Y_i\{\omega\}$, $\vec{t_i}$, and so on, are written with subscripts, this denotes one of many such scattered functions. i alone will always refer to $\sqrt{-1}$, and as an index only when written as a subscript or summation variable. $x^\star$ is the complex conjugate of x, while x' is the transpose of x.

The following sections will describe the theory of system prediction from scattered data. First, the process for translating between scattered functions and scattered Fourier approximations is described below in section 1.2. Then the system model is presented in section 1.3. A simple analytic result of this model, linear system correlation, is presented in section 3. Finally, the first-order delay algorithm is sketched in full in section 7.

1.2 Scattered Fourier Approximation

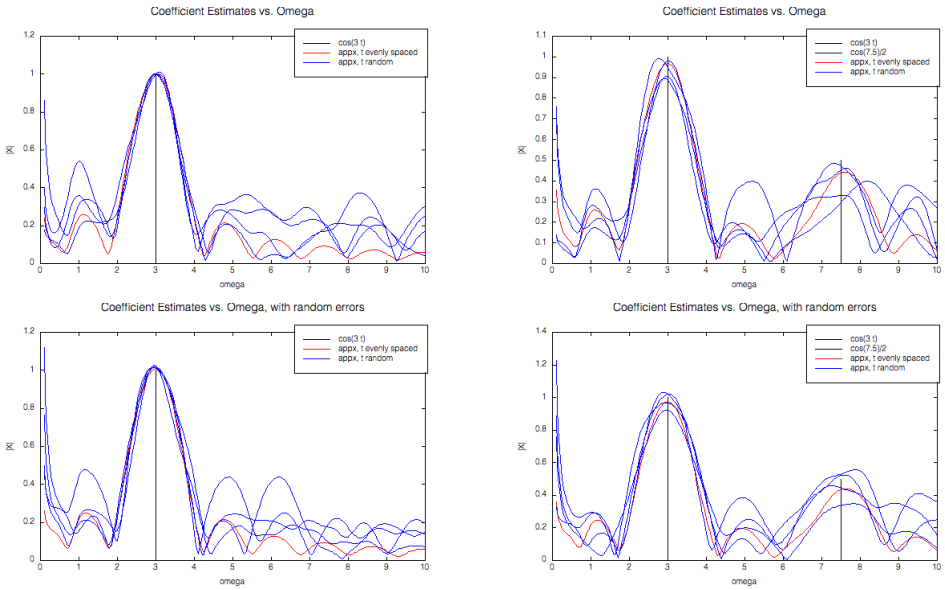
Throughout this paper, I use the unscaled Fourier transform,

$$F(\omega) = \int_{-\inf}^{\inf} f(t) e^{-\omega t i}$$

First, a function represented by fourier coefficients $\vec{Y}$ at scattered frequencies $\vec{\omega}$ can be calculated efficiently at scattered points $\vec{t}$ using

$$\vec{y} = \Re\{\vec{Y} e^{\vec{\omega}' \vec{t} i}\}$$

[1] Both as the variable upon which processing happens, and as the variable with respect to which Fourier transformations are taken.



When approximating time-domain functions with Fourier coefficients, a unlimited number of frequency sets can reproduce the values of $y\{t\}$– in general any collection of $\frac{n}{2}$ unique frequencies can perfectly reproduce n values. It is therefore possible to select frequencies strategically or randomly, a combination of which will be done below.

Consider a scattered function $y\{t\}$, and let $\vec{\omega}$ be a set of frequencies that will compose the Fourier approximation. The problem of Fourier approximation is to minimize the errors:

$$\min_{\vec{Y}} \sum_{i=1}^{n} (y_i - \Re\{\vec{Y} e^{\vec{\omega}' t_i i}\})^2 =$$

$$\min_{\vec{Y}, \vec{Y^\star}} \left(\vec{y} - \frac{1}{2}\left(\vec{Y} e^{\vec{\omega}'\vec{t}i} + \vec{Y}^\star e^{-\vec{\omega}'\vec{t}i}\right)\right)\left(\vec{y} - \frac{1}{2}\left(\vec{Y} e^{\vec{\omega}'\vec{t}i} + \vec{Y}^\star e^{-\vec{\omega}'\vec{t}i}\right)\right)'$$

$$\implies 2\vec{y} e^{\vec{t}'\vec{\omega}i} = \vec{Y} e^{\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i} + \vec{Y^\star} e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i}$$

$$\implies \vec{Y^\star} = \left(2\vec{y} e^{\vec{t}'\vec{\omega}i} - \vec{Y} e^{\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i}\right)(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}$$

$$\implies 2\vec{y} e^{-\vec{t}'\vec{\omega}i} = \left(2\vec{y} e^{\vec{t}'\vec{\omega}i} - \vec{Y} e^{\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i}\right)(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}(e^{-\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i}) + \vec{Y}(e^{\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i})$$

$$\implies 2\vec{y}\left(e^{-\vec{t}'\vec{\omega}i} - e^{\vec{t}'\vec{\omega}i}(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}(e^{-\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i})\right) = \vec{Y}\left(e^{\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i} - e^{\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i}(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}(e^{-\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i})\right)$$

for $\omega \neq 0$.

Given a collection of frequencies $\vec{\omega}$, a scattered function $y\{t\}$ can be approximated as follows:

$$\vec{Y} = 2\vec{y}\left(e^{-\vec{t}'\vec{\omega}i} - e^{\vec{t}'\vec{\omega}i}(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}(e^{-\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i})\right)\left(e^{\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i} - e^{\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i}(e^{-\vec{\omega}'\vec{t}i} e^{\vec{t}'\vec{\omega}i})^{-1}(e^{-\vec{\omega}'\vec{t}i} e^{-\vec{t}'\vec{\omega}i})\right)^{-1}$$

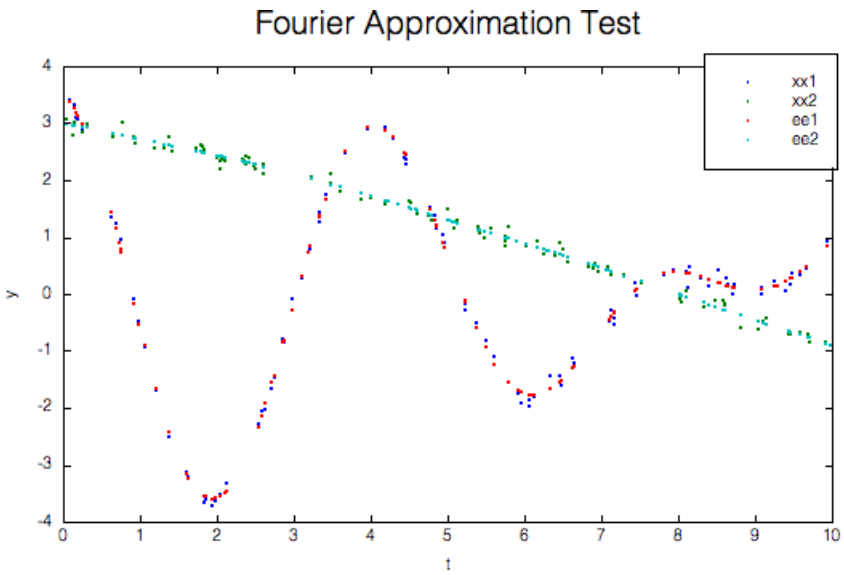


Figure 1: **Fourier Identification**: $xx1$ and $xx2$ are two simple functions with errors, each consisting of two frequencies. $ee1$ and $ee2$ are their recompositions, respectively, given all four frequencies.

If $\vec{\omega} = \omega$, a scalar, then the result is simpler. The vector of frequency responses to a vector of single frequencies (that is, each one considered on its own) is,

$$\vec{Y} = \vec{y}e^{-i\vec{t}'om\vec{e}ga}$$

1.3 The System Regression Model

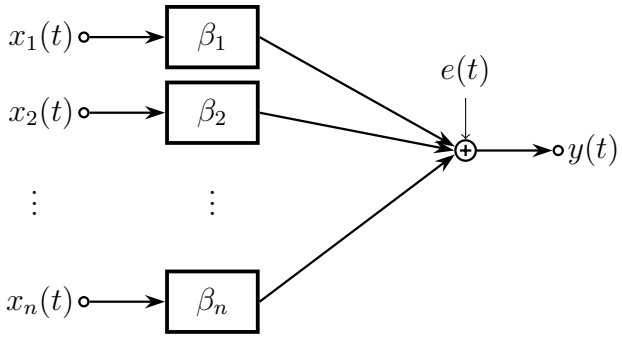
As a block diagram, the linear regression system could be described as follows:

Further Research I don't get the single-frequency result when I simplify my full method– which might mean that there's a further simplication that can be made to both!

$$\vec{Y} = 2\vec{y}\left(e^{-\vec{t}'\omega i} - e^{\vec{t}'\omega i}(e^{-\omega\vec{t}i}e^{\vec{t}'\omega i})^{-1}(e^{-\omega\vec{t}i}e^{-\vec{t}'\omega i})\right)\left(e^{\omega\vec{t}i}e^{-\vec{t}'\omega i} - e^{\omega\vec{t}i}e^{\vec{t}'\omega i}(e^{-\omega\vec{t}i}e^{\vec{t}'\omega i})^{-1}(e^{-\omega\vec{t}i}e^{-\vec{t}'\omega i})\right)^{-1}$$

$$\implies \vec{Y} = 2\vec{y}\left(e^{-\vec{t}'\omega i} - e^{\vec{t}'\omega i}(e^{\omega\vec{t}i}e^{\vec{t}'\omega i})^{\star}/n\right)\left(n - e^{\omega\vec{t}i}e^{\vec{t}'\omega i}(e^{\omega\vec{t}i}e^{\vec{t}'\omega i})^{\star}/n\right)^{-1}$$

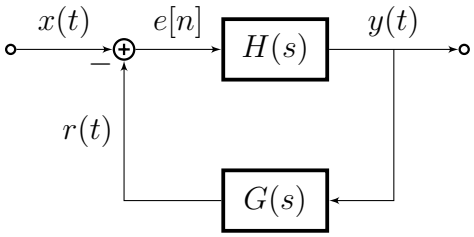
$$\implies \vec{Y} = 2\vec{y}\left(e^{-\vec{t}'\omega i} - e^{\vec{t}'\omega i}(e^{\omega\vec{t}i}e^{\vec{t}'\omega i})^{\star}/n\right)\left(n - |e^{\omega\vec{t}i}e^{\vec{t}'\omega i}|^2/n\right)^{-1}$$



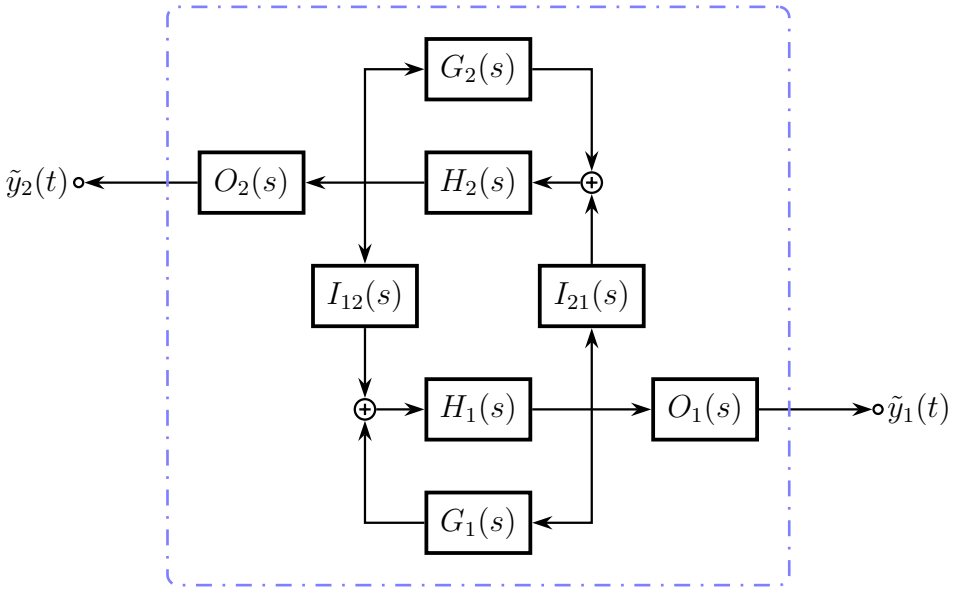
Suppose that there exist a collection of scattered data streams, with real values occuring at points in time, $d_i = [(t_{i0}, d_{i0}), (t_{i1}, d_{i1}), \ldots]$. Each data stream samples a distinct variable, f_i, which may be related to a collection of other variables. In other words, for each f_i, there are a set of indices, $\sigma = \{j | j \neq i\}$. The relationship between these variables is:

$$f_i(t) = \sum_{j \in \sigma} \alpha_i \int_{-\inf}^{t} f_j(t) e^{\frac{-t}{\tau}} + \epsilon_i(t)$$

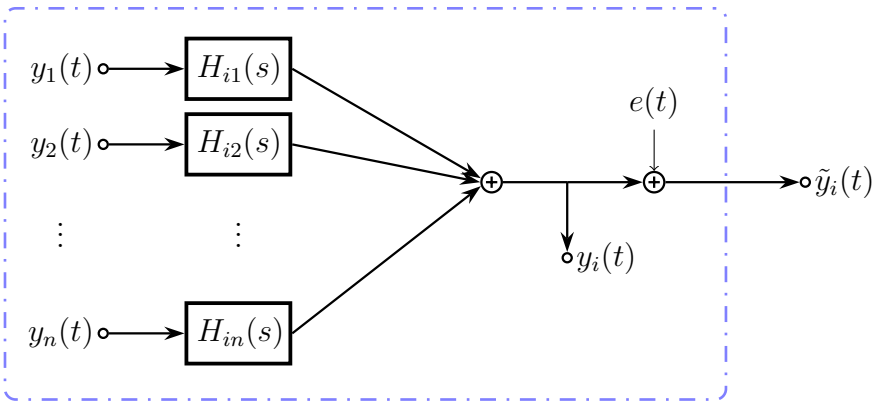
Basic feedback system:



Mutual feedback system:



System Neuron:



Note that the error is introduced outside the system. Without this, the behavior of the system is largely determined by the errors, and unbiased estimation is impossible.

2 Analytical Foundations

Let α_{ik} be fourier coefficient $\tilde{y}_i\{t\}$, measured at ω_k. Define γ_{ik} as the corresponding internal variable coefficient, so

$$\alpha_{ik} = \gamma_{ik} + \epsilon_{ik}$$

According to the diagram above, in the frequency domain,

$$\gamma_{ik} = \sum_j H_{ij}(\omega, \theta)\gamma_{jk}$$

where $H(\omega, \theta)$ has a supposed functional form, and θ is the parameter we want to estimate.

This can be rewritten

$$\lambda_k \vec{\gamma_k} = H(\omega, \theta)\vec{\gamma_k} \, \forall k$$

where λ is added to allows different coefficients to be determined by the system up to a multiplicative constant; here $H(\omega, \theta)$ is $N \times K$, and $\vec{\gamma_k}$ is $N \times 1$.

This is an eigenequation, which has N solutions. That is,

$$\gamma_k = c_k v_{l(k)}(\omega_k, \theta)$$

where $l(k)$ maps every element k to a value from 1 to N.

In other words, the K vectors $\vec{\alpha_k}$ need to be partition into N collections. The estimation can proceed within each partition independently.

3 System Correlation Analysis

Discuss where I put the error, derivation of the eigen solution, etc.

Further Research [put results of problem.tex discussion right above here.] In general, there is no analytic expression for a given eigenvector... Is there an analytical solution to the n,nth eigenvector, if all lin. dep?

Further Research Let $\{x_i\}$ be a collection of objects and suppose that a distance measure, $d(x_1, x_2)$ is defined for any two objects. Also let there be a method, $f(\{x_i\})$ to construct a "representative object" from a collection of objects (in general not identitical to any of the objects in $\{x_i\}$).
The problem is to find the optimal partitioning of all the objects into n bins. Optimality is defined as the minimum of some function of the distance measures from each object to its corresponding bin representative object, e.g.,

$$\sum_i d(x_i, f(P_i))^2$$

where P_i is the collection of objects in the same bin as object i.

This may be solvable more efficiently by some kind of clustering algorithm, or by using the fact that each object is a complex vector and can be placed as a point in $\mathbb{C}^k$ space, but then I expect the particulars of the distance function (which is not at all like Euclidean distance) to matter a lot.

The number of ways to partitions of these values can get very large (called the Stirling numbers of the second kind and denoted $\left\{ {n \atop k} \right\}$).

n\k	2	3	4	5	6	7	8	9
2	1							
3	3	1						
4	7	6	1					
5	15	25	10	1				
6	31	90	65	15	1			
7	63	301	350	140	21	1		
8	127	966	1701	1050	266	28	1	
9	255	3025	7770	6951	2646	462	36	1
10	511	9330	34105	42525	22827	5880	750	45
11	1023	28501	145750	246730	179487	63987	11880	1155
12	2047	86526	611501	1379400	1323652	627396	159027	22275
13	4095	261625	2532530	7508501	9321312	5715424	1899612	359502
14	8191	788970	10391745	40075035	63436373	49329280	20912320	5135130
15	16383	2375101	42355950	210766920	420693273	408741333	216627840	67128490

Finding local optima of the partitioning problem is easy. However, finding the globally optimal partitioning is a difficult problem. The solution taken here is simulated annealing, which can be thought of as a random tunneling algorithm. At any point in the algorithm, there is a "current state", a partitioning. In each loop, random changes are proposed, and

Further Research It might also be fruitful to represent the complex numbers as $\rho\angle\theta$, where $\rho = ln(abs(c))$ and $\theta = angle(c)$, when solving for the components. Then both components of a complex number are add under multiplication. So the equation,

$$a = b * c \implies (\alpha_a + \beta_a i) = (\alpha_b + \beta_b i)(\alpha_c + \beta_c i) = (\alpha_b\alpha_c - \beta_b\beta_c) + (\alpha_b\beta_c + \alpha_c\beta_b)$$

becomes

$$\alpha = \beta * \gamma \implies \rho_a\angle\theta_a = (\rho_b\angle\theta_b)(\rho_c\angle\theta_c) = (\rho_b + \rho_c)\angle(\theta_b + \theta_c)$$

the alternative local optimum is identified. If that local optimum is better than the current state, it is taken. Sometimes, even if it isn't better, it might be taken– this is tunneling– as a potential path toward a better optimum.

4 Functional Forms for $H_{ij}(\omega, \theta)$

I have been exploring two functional forms for $H(\omega, \theta)$:

- $H_{ij}(\omega, \theta) = H_{ij}$ (that is, a constant). This is frequency correlation analysis, where $H = r\angle\phi$. r represents the magnitude of correlation; ϕ is the phase shift.
- $H_{ij}(\omega) = \frac{r}{-\tau + i\omega}$. This is a first-order delay. An infinite collection of first-order delays with certain values of τ can combine into a proper delay. With enough variables, first-order delays can represent any basic system dynamical system.

To recover $H(\omega, \theta)$, note that for any matrix with eigen vectors $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) = V$, $H(\omega, \theta) = V\Lambda(\omega, \theta)V^{-1}$.

5 Estimating γ_k

Consider $H_{ij}(\omega, \theta) = H_{ij}$.

We minimize the least squares error to find the underlying

$$\epsilon_{ik} = \alpha_{ik} - c_k v_i \implies \mathcal{L} = \sum_i \sum_k (\alpha_{ik} - c_k v_i)(\alpha_{ik}^\star - c_k^\star v_i)$$

We define $c_k = a_k + ib_k$ and $v_i = x_i + iy_i$, and solve. The result is,

$$\frac{\vec{v}^{\star\prime}\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)\vec{v}}{\vec{v}'\vec{v}^\star}\vec{v} = \left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)\vec{v}$$

This is another eigenequation. The left coefficient of $\vec{v}$ is the Rayleigh quotient, equal to the eigenvalue of $\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)$. Furthermore, for small errors, the matrix $\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)$ will be nearly rank 1, and have only one eigensolution that is not near 0.

Further Research	Is there a closed-form solution to the last or largest eigenvalue of an eigenequation?

To my knowledge, there is no closed-form expression for this largest eigensolution. However, it can be solved easily, numerically.

Let that eigenvector be $\vec{v}$. Then $c_k = \frac{\vec{\alpha_k}'\vec{v}}{\vec{v}'\vec{v}^\star}$.

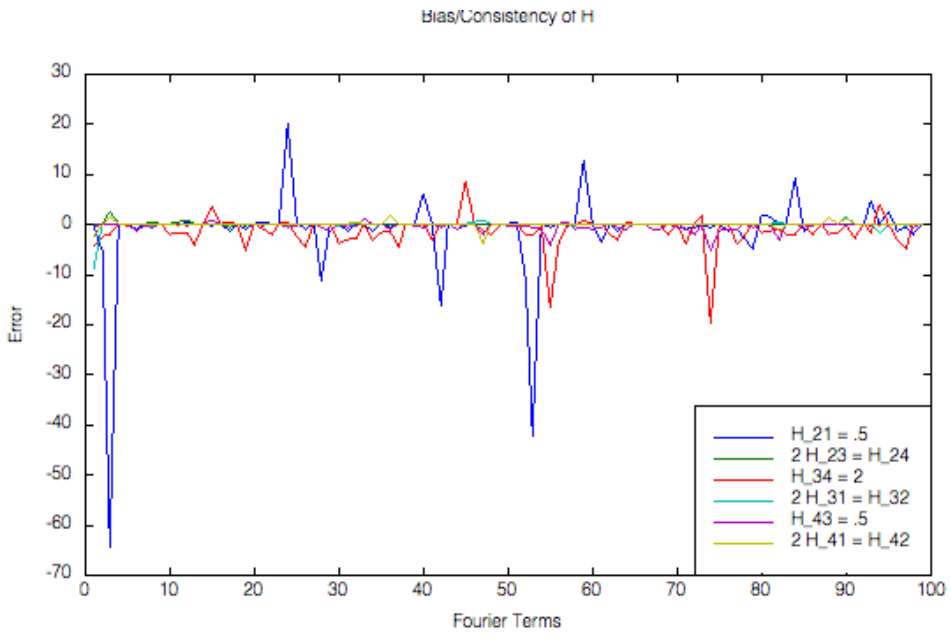
$$H_{ii} = \sum_k \left(\sum_l V_{il}\Lambda_{lk} \right) (V^{-1})_{ki} = \sum_k (V_{ik}\lambda_k)(V^{-1})_{ki} = \sum_k (V^{-1'})_{ik} V_{ik}\lambda_k = (V^{-1'} \times V)\vec{\lambda}$$

To get all solutions, set $H_{ii} = 1$ for a minimum number of values i to get solutions, and $H_{jj} = 0 \, \forall j \neq i$:

$$\vec{\lambda} = (V^{-1'} \times V)^{-1}(1, 0, 0, \ldots, 0)'$$

The result needs rescaling, but reliably works.

6 Bias and Consistency



We have,

$$\hat{\gamma} = \hat{H}\hat{\gamma}$$
$$\alpha = \hat{\gamma} + \hat{\epsilon}$$

These can be rearranged to give

$$\hat{H}\alpha = H\alpha + (\hat{H} - I_n)\hat{\epsilon} - (H - I_n)\epsilon \implies$$

$$\hat{H} = H + \left((\hat{H} - I_n)\hat{\epsilon}\alpha' - (H - I_n)\epsilon\alpha' \right) (\alpha\alpha')^{-1}$$

The second form is conducive to proving consistency. So, $\hat{H}$ is unbiased and consistent if $E(\hat{H} - I_n)\hat{\epsilon} = (H - I_n)\epsilon$. Or, equivalently, if

$$E\hat{\epsilon}(\hat{\gamma} + \hat{\epsilon})' = E\epsilon(\gamma + \epsilon)'$$

By the model, $E\epsilon\gamma' = 0$, and if it seems plausible that

$$E\hat{\epsilon}\hat{\epsilon}' = E\epsilon\epsilon'$$

(that is, that σ_ϵ can be consistently estimated).

So, the remaining criteria is that

$$E(\hat{H} - I_n)\hat{\epsilon}\hat{\gamma} = 0$$

Rather than selecting λ arbitrarily as above, we can select λ to minimize this value, thereby minimizing the bias.

7 Sketch of First-Order Delay Algorithm

Let there be a collection of n endogenous scattered variables, $y_i\{t\}$, for $i \in I = 1, \ldots, n$. For each variable, define a Fourier approximation, $Y_i\{\omega\}$. Initially, this approximation may be empty, e.g., $\omega = \{\}$. Also, for each variable

$< \{y_i\{t\}, Y_i\{\omega\}, \{\alpha_{ij}, \tau_{ij}\}_{j \in J_i}\}_{i=1}^n >$

Initially, $\omega_i = \{\}$, $m = 0$.

7.1 Phase 1: Fourier Improvements

■ Is bestspec sensible? The idea is that I can approximate a function many ways: I can specdata with all the omegas, or I can just specdata with some, take out those elements, and specdata with others. What's sensible?

In comfour, I can specdata each with common + kk new, and just strip out the kk best. Then I'll have potentially many solutions. So start eliminating the weakest. ■

[2]

Require: $J \subset I$, $N > 0$

$pool \leftarrow \{: \omega_i : \ random(\{-\pi \ldots \pi\}, 2N) \ : \omega_{J_1} : \cdots : \omega_{J_J} :\}$

$\omega \leftarrow random(pool, N)$

[2]Notation: $\{\cdot\}$ denotes a (unordered) set, $[\cdot]$ denotes a (row) vector, and $< \cdot >$ denotes a tuple. Within any of these, $: x :$ denotes all of the elements within x, that is, $x_1, x_2, \ldots, x_n$. $x[i]$ is the ith element of x, but xt refers to the entireity of a scattered function. x_i denotes the entire variable x associated with input i.

$\hat{\omega} \leftarrow [\,]$
$\hat{\alpha} \leftarrow [\,]$
$\hat{y}\{t\} \leftarrow \{0\}_{\forall t}$
for $i = 1$ to N **do**
  $\alpha \leftarrow fourappx(y\{t\} - \hat{y}\{t\}, \omega)$
  $k \leftarrow$ index of largest element of $|\alpha|$
  $\hat{\omega} \leftarrow [: \hat{\omega} : \ \omega[k]]$
  $\hat{\alpha} \leftarrow [: \hat{\alpha} : \ \alpha[k]]$
  $\hat{y}\{t\} \leftarrow \hat{y}\{t\} + \Re\{\alpha[k]e^{\omega[k]t}\}$
  $\omega[k] \leftarrow random(pool)$
end for

1. Randomly select a few other variables, $J \subset I$, and take the union of their ω's: $\omega = \cup_{j \in J} \omega_j$. Let $\hat{\omega}$ and $\hat{\alpha}$ denote the progressive Fourier appoximation, initially both empty sets. Define $\hat{y}\{t\}$, the approximation, at the same values of t as $y\{t\}$; initial $\hat{y}\{t\}$ is everywhere 0.
2. Produce a scattered Fourier approximation of $y\{t\} - \hat{y}\{t\}$, using ω, producing the corresponding coefficient set α.
3. Identify the largest component α, α_k at ω_k.
4. Add α_k and ω_k to $\hat{\alpha}$ and $\hat{\omega}$ respectively. Replace ω_k in ω with a new (random) frequency.
5. Add the contribution of α_k to $\hat{y}\{t\}$. That is, let $\hat{y}\{t\} = ...$

Acknowledgements

Many thanks to Serene Ng, for endless help in econometrics, and Raju Krishnamoorthy.

References

Von Bertalanffy, L. (1956). General system theory. *General systems*, 1(1):11–17.

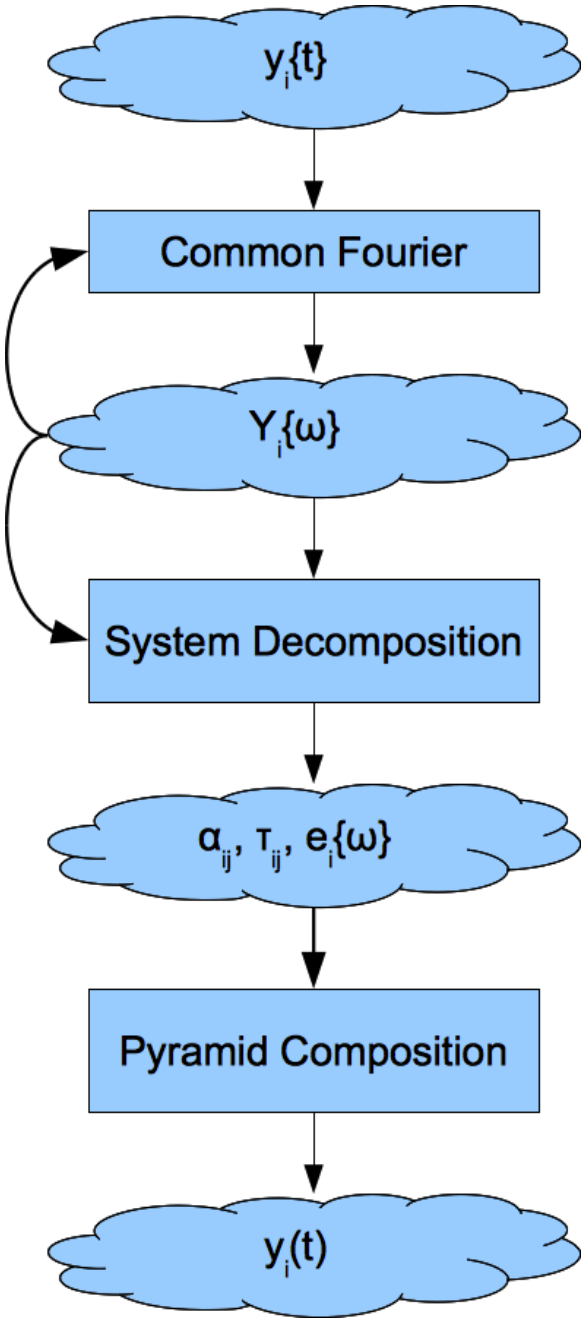


Figure 2: *FOD Algorithm Stages:* A rough sketch of the algorithm.

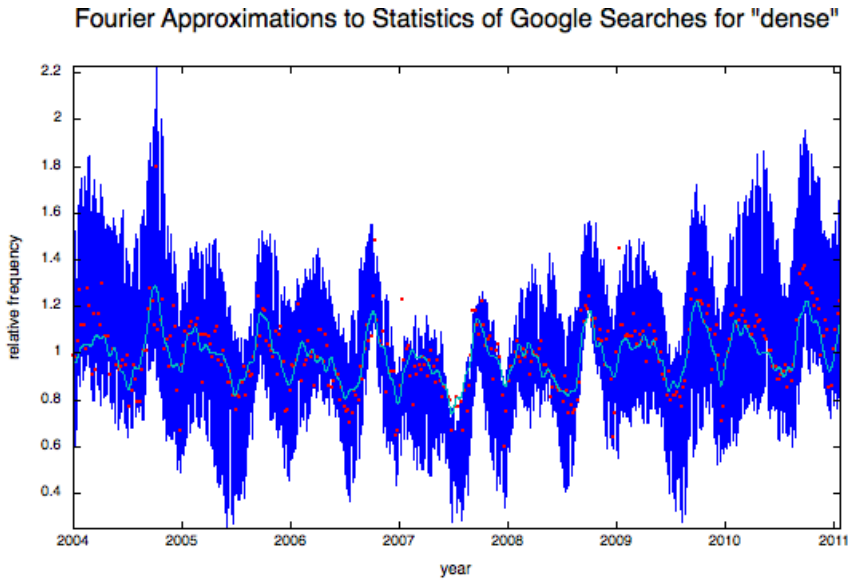


Figure 3: *Fourier Approximations:* Blue shows an over-saturated approximation; note that it produces oscillations not supported by the data. The cyan curve includes only the largest 19 components.