PART 5

Intro to data representation

Mauricio Sacchi
University of Alberta, Edmonton, AB, Canada

April 2014

Interpolation/Regularization: It is all about simplicity

- The Plane wave model in t-x, f-x and f-k
- Using windows to keep it simple

Interpolation/Regularization: It is all about simplicity

- The Plane wave model in t-x, f-x and f-k
  - Linear event in t-x
  - Linear event in f-x
  - Linear event in f-k

$$s(t,x) = a(t - px)$$

$$\Updownarrow$$

$$S(\omega,x) = A(\omega)\, e^{-i\omega px}$$

$$\Updownarrow$$

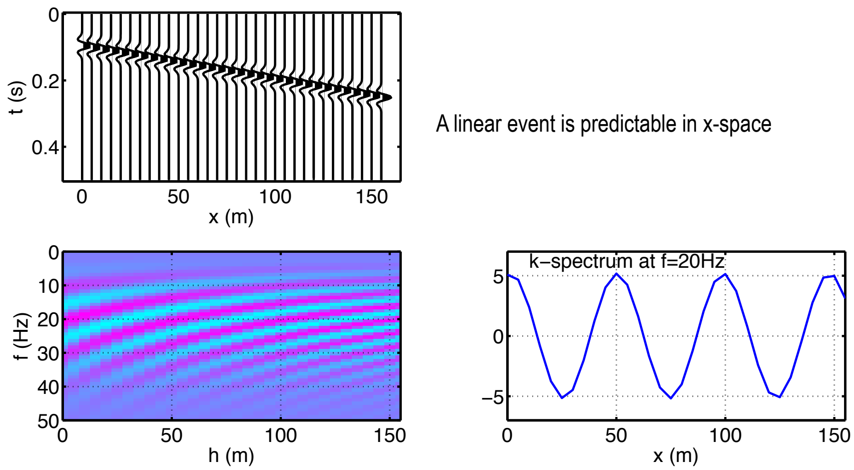
$$S(\omega,k) = A(k + p\omega)$$

$\omega = 2\pi f$

Interpolation/Regularization: It is all about simplicity

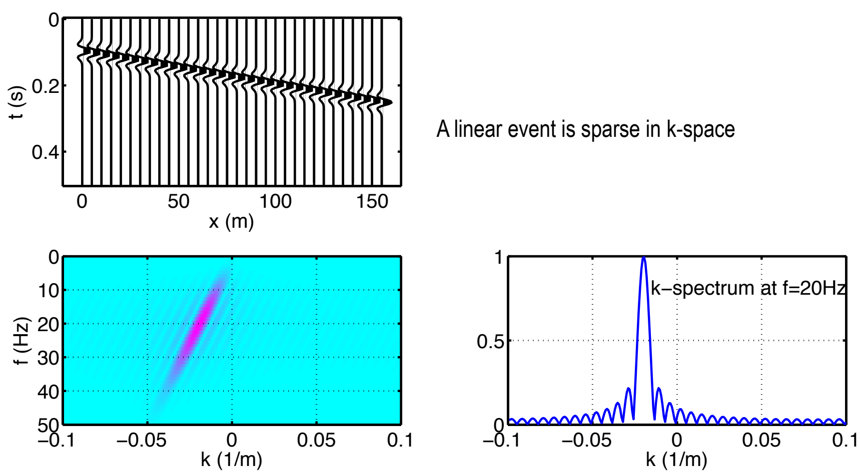
- The Plane wave model in t-x, f-x and f-k
  - Remarks: Consider one monochromatic temporal frequency f
    - The signal is predictable in x
      - Data are represented by complex exponentials
    - Sparsity in k
      - *One wavenumber per linear event* is needed to represent the spatial data. In other words, the k-spectrum is *sparse*

Interpolation/Regularization: It is all about simplicity



A linear event is predictable in x-space

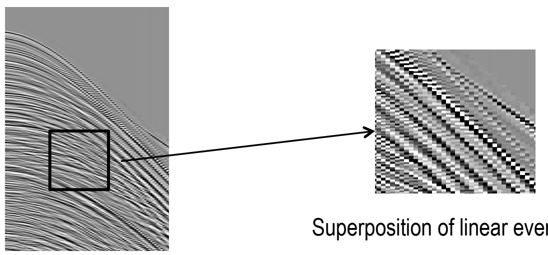
Interpolation/Regularization: It is all about simplicity



A linear event is sparse in k-space

The simplicity assumption

- Linear events in t-x
- Using windows to keep it simple



Superposition of linear events

Interpolation/Regularization: It is all about simplicity

- Predictability in *fx* (1 linear event)

$$s(t, n\,\Delta x) = a(t - p(n-1)\,\Delta x)$$

$$S(\omega, n\Delta x) = S_n(\omega) = A(\omega)\; e^{-i\omega p(n-1)\,\Delta x}$$

- Consider the complex amplitude at a given temporal frequency and spatial sample *n*

$$\begin{aligned} S_n(\omega) &= A\, e^{-i\alpha(n-1)} \\ &= e^{i\alpha} A\, e^{-i\alpha(n-2)} \\ &= a\, S_{n-1}(\omega) \end{aligned}$$

$$S_n(\omega) = a(\omega)\, S_{n-1}(\omega)$$

Interpolation/Regularization: It is all about simplicity

- Predictability in fx (*p* linear event)

$$s(t, n\Delta x) = \sum_k A_k w(t - p_k (n-1)\Delta x)$$

$$S(\omega, n\Delta x) = W(\omega) \sum_k A_k \, e^{-i\omega p_k (n-1)\Delta x}$$

- Consider the complex amplitude at a given temporal frequency and spatial sample *n*, one can show that

$$S_n(\omega) = a_1(\omega) S_{n-1}(\omega) + a_2(\omega) S_{n-2}(\omega) \cdots + a_p(\omega) S_{n-p}(\omega)$$

Interpolation/Regularization: It is all about simplicity

- Consider *p=2, n=1..N, N=6*

$$S_6(\omega) = a_1(\omega) S_5(\omega) + a_2(\omega) S_4(\omega)$$
$$S_5(\omega) = a_1(\omega) S_4(\omega) + a_2(\omega) S_3(\omega)$$
$$S_4(\omega) = a_1(\omega) S_3(\omega) + a_2(\omega) S_1(\omega)$$
$$S_3(\omega) = a_1(\omega) S_2(\omega) + a_2(\omega) S_1(\omega)$$

- Problem 1: Estimate prediction filter coefficients from data
- Problem 2: Estimate data from prediction filter

Data representation for interpolation

- Consider the expansion of a signal in terms of basis functions

$$d_n = s_n + \varepsilon_n$$

$$s_n = \sum_{j=1}^{M} c_j \phi_{n,j} \qquad n = 1 \cdots N$$

- Consider a regression in terms of an over-complete (underdetermined) expansion:

$$M > N$$

Data representation for interpolation

- We have an under-determined problem and therefore, an infinite number of solutions.

$$\vec{d} = \Phi \vec{c} + \vec{\varepsilon}$$

- Regularization methods are used to find a unique and stable solution. For this purpose we minimize the following cost function

$$J = \| \vec{d} - \Phi \vec{c} \|^2 + \mu^2 R(\vec{c})$$

Data representation for interpolation

$$J = \| \vec{d} - \Phi \vec{c} \|^2 + \mu^2 R(\vec{c})$$
$$= \text{Misfit} + \mu^2 \text{ Regularizer}$$

- The goal is to
  - find the coefficients of the expansion from the available data by minimizing the cost J
  - use the coefficients to synthesize unobserved data (new spatial and/or de-noised data)

Data representation for interpolation

$$J = \| \vec{d} - \Phi \vec{c} \|^2 + \mu^2 R(\vec{c})$$
$$= \text{Misfit} + \mu^2 \text{ Regularizer}$$

- In general, the coefficients of the expansion must provide physical information about the wave field (e.g. curvature, dip) – They need to have a physical label.

- In other words, they should model pieces of the wave field we wish to represent.

Data representation for interpolation

Seismic data can be represented with

- Fourier Bases (frequency / wavenumber)
- Gabor Bases (time-frequency / space-wavenumber)
- Wavelets (location and scale)
- Radon bases (intercept and dip or curvature)
- Local Radon bases (time-space-dip)
- Curvelets (time-space-scale-dip)

Data representation for interpolation

- One could also use data driven representations:
  - Cadzow / Multichannel Singular Spectrum Analysis
  - Tensor completion
  - Sparse coding

Data representation for interpolation

- **Sparsity**: General schemes for signal reconstruction based on the solution of an inverse problem can be found in numerous articles. Of particular interest to the discussion are contributions that considered *sparsity* constraints:
  - Thorson and Claerbout, GEO 1985
  - Sacchi and Ulrych, GEO 1995 and 1996
  - Sacchi, Ulrych and Walker, IEEE Trans. SP 1998
  - Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Inform. Theory 2004.
  - Zwartjes and Gisolf, GP 2007
  - Zwartjes and Sacchi, GEO 2007

Data representation for interpolation

Connection to Compressive Sensing

- The above references are early (*Non-mathematical but with good intuition*) attempts to what today is called *Compressive Sensing* (CS)
- CS arises from various contributions by Donoho, Candes, Romberg and Tao in articles starting around 2005
- CS provides conditions for reconstruction (from limited information) for signals that admit a sparse representation

Try to Google *Compressive Sensing*

Sparsity and data reconstruction

$$J = \| \vec{d} - \Phi \vec{c} \|^2 + \mu^2 R(\vec{c})$$

Min norm solution $R(c) = \| c \|^2$

Sparse solution $R(c) = | c |_1$

$R(c) = \sum_j \ln(1 + c_j^2 / \beta^2)$

MWNI $R(c) = c^H Q^{-1} c$

Assume Compressive solutions

Sparsity and data reconstruction

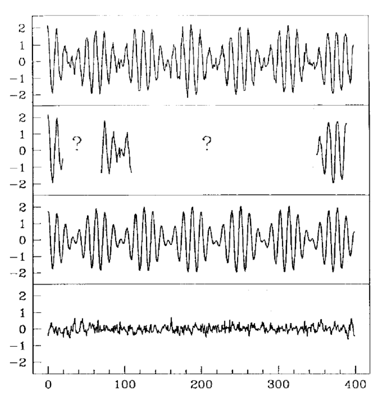
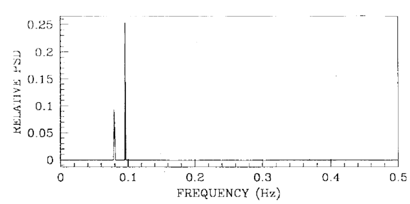


Fig. 9. (Top) Synthetic time series; the series contains two unit amplitude harmonics of frequencies 0.795 and 0.954 Hz, respectively; (second) Gapped time series; (third) Reconstructed time series using the Cauchy–Gauss DFT; (bottom) Error sequence obtained subtracting (third) from (top).



. 10. Power spectrum of the gapped time series computed using the ıchy–Gauss model.

Sacchi, M.D., Ulrych, T.J, and Walker, C., 1998, Interpolation and extrapolation using a high resolution discrete Fourier transform: IEEE Trans. on Signal Processing, 46, No. 1, 31-38.

Sparsity and data reconstruction

- Data are expanded as a sum of complex exponentials.
- The unknown coefficients are the complex Fourier amplitudes

$$d(x_j) = \sum_n c_n e^{ik_n x_j} + n(x_j)$$

$$\vec{d} = \Phi\vec{c} + \vec{n}$$

Sparsity and data reconstruction

- Use DFT kernel (Slow) and minimizing cost function

$$J = \| \vec{d} - \Phi\vec{c} \|^2 + \mu^2 R(\vec{c})$$

- Use estimator of Fourier coefficients to synthetize new data

$$\hat{d}((j-1)\Delta x) = \sum_n \hat{c}_n e^{ik_n (j-1)\Delta x}$$

- In Sacchi et al IEEE-SP-1998 sparsity is obtained by the Cauchy criterion

$$R(\vec{c}) = \sum_j \ln(1 + c_j^2 / \beta^2)$$

Sparsity and data reconstruction

- IRLS (Iterative reweighted least squares) solution

$$\min_{\vec{c}} \left\{ J = \| \vec{d} - \Phi \vec{c} \|^2 + \mu^2 R(\vec{c}) \right\}$$

$$\Rightarrow$$

$$\vec{c} = (\Phi^H \Phi + \mu^2 Q(\vec{c}))^{-1} \Phi^H \vec{d}$$

- With diagonal weights given by

$$Q_{jj} = \frac{2}{\beta^2 + | c_j |^2}$$

Sparsity and data reconstruction

- IRLS (Iterative reweighted least squares)

$$Q^0 = I$$

$$k = 1,2,3....$$

$$\vec{c}^k = (\Phi^H \Phi + \mu^2 Q(\vec{c}^{k-1}))^{-1} \Phi^H \vec{d}$$

We often replace the Inversion by a semi-iterative solver (CG)

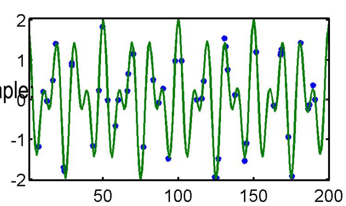
$$Q_{jj}^k = \frac{2}{\beta^2 + | c_j^k |^2}$$

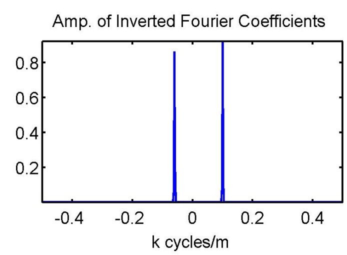
$$end$$

```
function [x, y_pred, mse] = cauchy_gauss(A,y,mu,beta,Max_Iter,r0);
%
% Solve Ax-y \approx 0
%
% with x sparse, using the minimization of the following cost:
%
%  J  =  ||Ax - y||_2^2 + mu^2 \sum_i \log (1 + (x(i)/beta)^2)
%

 [ny,nx] = size(A);

 x = r0;
     q = 1./(beta^2 + x.^2);
     q = q/max(q);
     Q = 0.001*diag(q) + 0.01*eye(nx);


 R = A'*A;
 g = A'*y;


   for k = 1:Max_Iter
      x = (R +  Q+0.001*eye(nx))\g;             % for Radon, I do this with CG
      xx = conv2(x.^2,hamming(3),'same');
      xx = xx/max(xx);
      b = 0.001*xxr;
      q = (mu^2 )./(b.^2 + x.^2);
      Q = diag(q);
   end

  y_pred = A*x;
  mse = (1/ny)*sum( (y-y_pred).^2);

  return
```

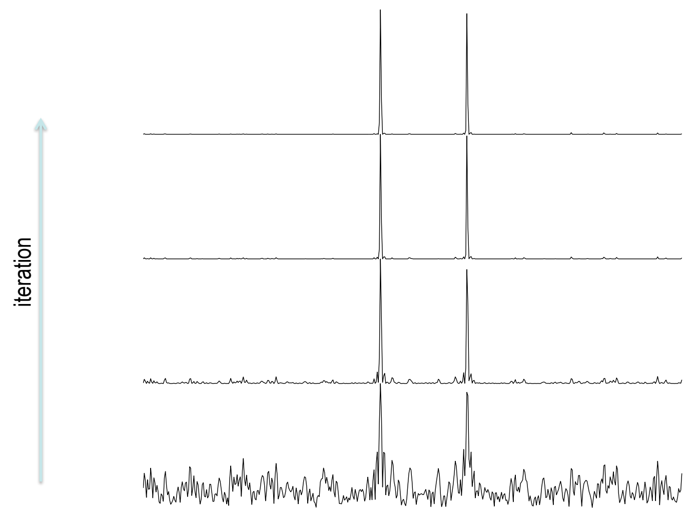

Sparsity and data reconstruction

Example





IRLS iteration



Irregular sampling and sparsity working together

Sparsity and Radon transforms

$$d(t,x) = \sum_p m(\tau = \varphi(t,x,p), p)$$

Linear RT $\varphi(t,x,p) = t - px$

Parabolic RT $\varphi(t,x,p) = t - px^2$

Hyperbolic RT $\varphi(t,x,p) = \sqrt{t - px^2}$

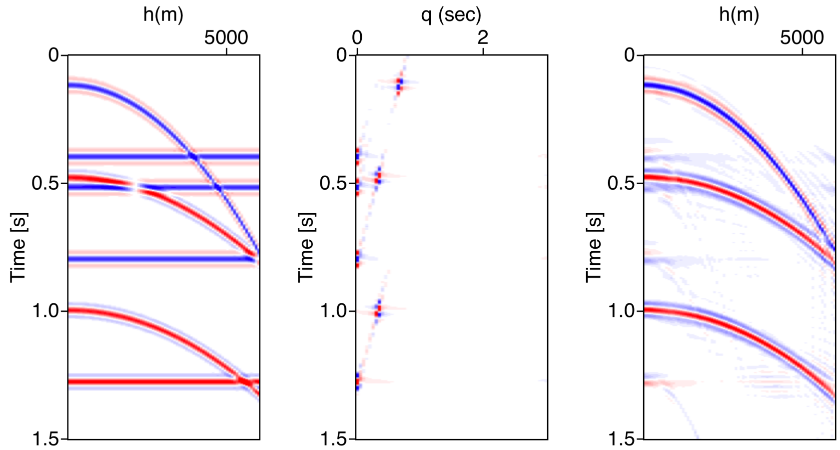
$$\min_{\vec{m}} J = \| \vec{d} - L\vec{m} \|_2^2 + \mu R(\vec{m})$$

$R(\vec{m}) = \| \vec{m} \|_2^2$ CG

$R(\vec{m}) = \| \vec{m} \|_1$ IRLS + CG

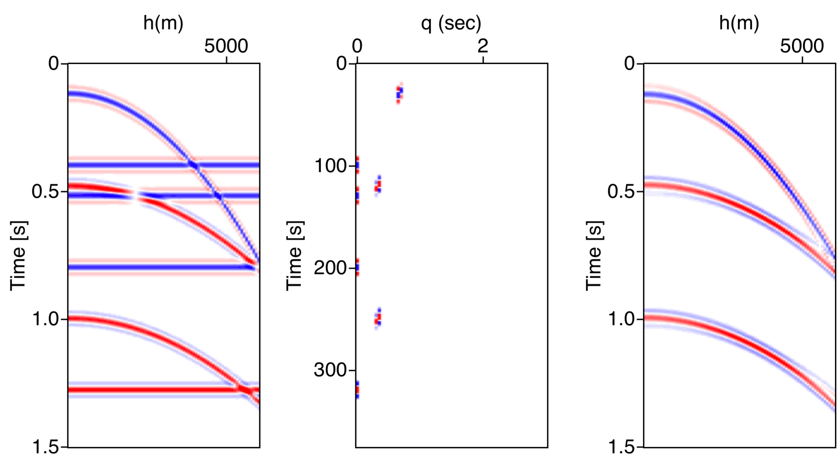
Thorson and Clearbout (GEOPHYSICS 1985)
Sacchi and Ulrych (GEOPHYSICS 1995)
Trad, Ulrych and Sacchi (GEOPHYSICS, 2003)

Parabolic Radon Transform



$$R(\vec{m}) = |\vec{m}|_2^2$$

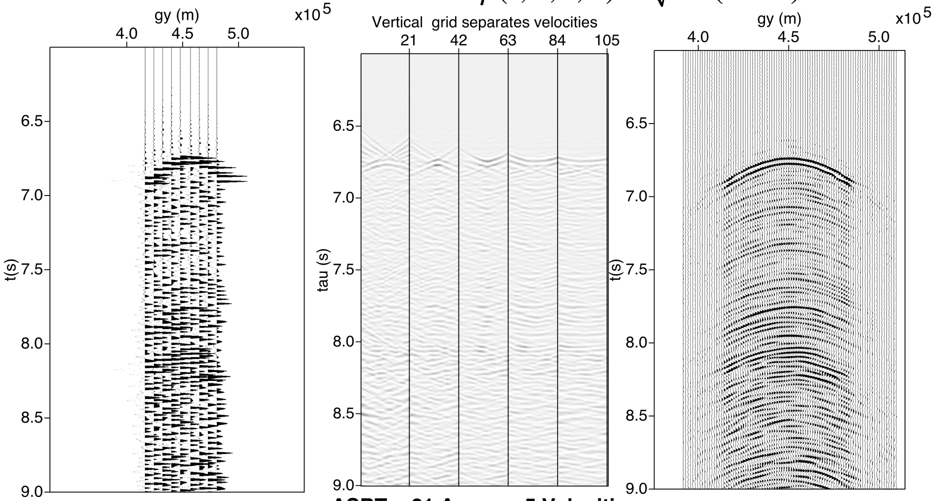
Sparse Parabolic Radon Transform



$$R(\vec{m}) = |\vec{m}|_1$$

Hyperbolic Radon transform

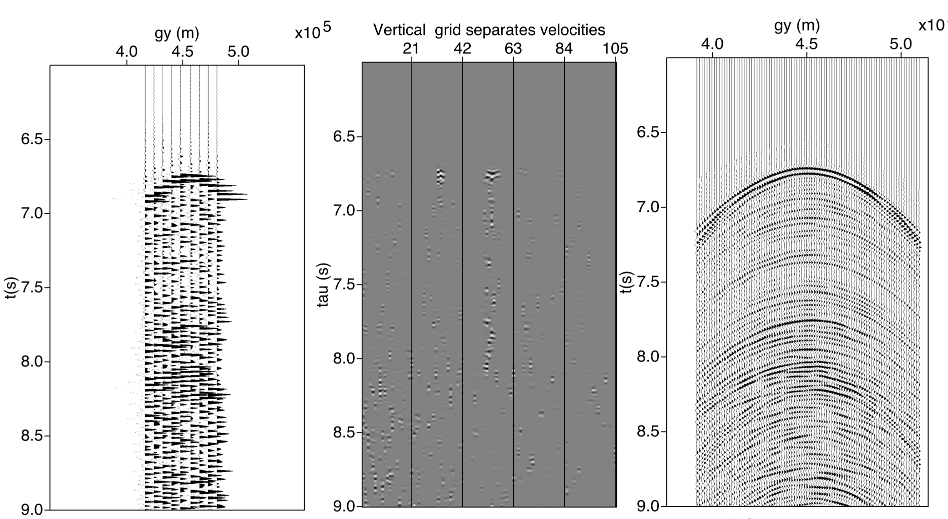
$$\varphi(t,x,v,a)=\sqrt{t-(x-a)^2/v^2}$$



DATA | ASRT - 21 Apexes, 5 Velocities | RECOVERED DATA

$$R(\vec{m}) = \| \vec{m} \|_2^2$$

Sparse Hyperbolic Radon Transform



DATA | ASRT - 21 Apexes, 5 Velocities | RECOVERED DATA

$$R(\vec{m}) = \| \vec{m} \|_1$$

The simplicity assumption and connection to rank (Singular Spectrum Analysis (SSA)/Embedding)

$$S_n = aS_{n-1} \Rightarrow$$

$$\mathbf{M} = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ s_2 & s_3 & s_4 & s_5 \\ s_3 & s_4 & s_5 & s_6 \\ s_4 & s_5 & s_6 & s_7 \end{pmatrix} = \begin{pmatrix} s_1 & as_1 & a^2\, s_1 & a^3 s_1 \\ s_2 & as_2 & a^2 s_2 & a^3 s_2 \\ s_3 & as_3 & a^2 s_3 & a^3 s_3 \\ s_4 & as_4 & a^2 s_4 & a^3 s_4 \end{pmatrix}$$

$$rank(\mathbf{M}) = 1$$

Cadzow, J.A. Signal Enhancement, A Composite Property Mapping Algorithm. IEEE Trans. on Acoustics, Speech and Signal Processing 36 (1988)

Ghil, M., M.R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. Advance Spectral Methods for Climatic time series. Reviews of Geophysics 40 (2002)

- Predictability in *fx* (*p* linear events)

$$s(t, n\,\Delta x) = \sum_k A_k w(t - p_k (n-1)\,\Delta x)$$

$$S(\omega, n\Delta x) = W(\omega) \sum_k A_k \; e^{-i\omega p_k (n-1)\,\Delta x}$$

$$S_n(\omega) = a_1(\omega) S_{n-1}(\omega) + a_2(\omega) S_{n-2}(\omega) \cdots + a_p(\omega) S_{n-p}(\omega)$$

$$rank(\mathbf{M}) = p$$

Yang, H.H. and Y. Hua. On rank of block Hankel matrix for 2-D frequency detection and Estimation. IEEE Transactions on Signal Processing 44 (1996)

Hua, Y. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. IEEE Transactions on Signal Processing 40 (1992)

- Sparsity, Predictability and Rank are connected for the Fourier synthesis model
    - Methods to interpolated data based on simplicity or sparsity in Fourier domain (MP, ALFT, MWNI, POCS) work all under similar assumptions
        - Wakefield can be synthesized by a superposition of plane waves
        - Simplicity (or sparsity) in the distribution of Fourier coefficients
        - The idea very similar to ideas used today in CS (Compressive Sensing)
    - *Rank is the new sparsity but it is not so new if one thinks the connection that exists between predictability (SPITZ, 91), sparsity (SACCHI et al 98) and rank (Cazdow 88)… as someone said: All goes back to Gauss or Laplace.*