Some new trends and old lessions in Geophysical Inversion

Malcolm Sambridge

Research School of Earth Sciences

Australian National University

Joint ICTP-IUGG Workshop on Data Assimilation and Inverse Problems in Geophysical Sciences

18 - 29 October 2021.

with contributions from Andrew Valentine, Matthias Scheiter and Buse Turunçtur

International Union of Geodesy and Geophysics IUGG UGGI Union Géodésique et Géophysique Internationale

ICTP The Abdus Salam International Centre for Theoretical Physics

UNESCO United Nations Educational, Scientific and Cultural Organization

EAIFR ICTP - East African Institute for Fundamental Research under the auspices of UNESCO

Emerging directions in geophysical Inversion

Malcolm Sambridge

Research School of Earth Sciences

Australian National University

Joint ICTP-IUGG Workshop on Data Assimilation and Inverse Problems in Geophysical Sciences

18 - 29 October 2021.

with contributions from Andrew Valentine, Matthias Scheiter and Buse Turunçtur

International Union of Geodesy and Geophysics IUGG UGGI Union Géodésique et Géophysique Internationale

ICTP The Abdus Salam International Centre for Theoretical Physics

UNESCO United Nations Educational, Scientific and Cultural Organization

EAIFR ICTP - East African Institute for Fundamental Research under the auspices of UNESCO

Review chapter

Emerging directions in geophysical inversion

Andrew P. Valentine[1] and Malcolm Sambridge[2]

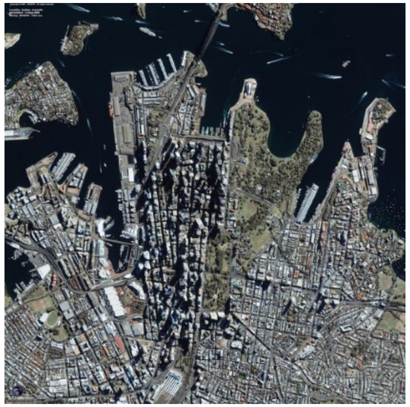
[1]Department of Earth Science, Durham University, South Road, Durham, DH1 3LE, UK.
Email: andrew.valentine@durham.ac.uk
[2]Research School of Earth Sciences, The Australian National University, 142 Mills Road, Acton ACT 2601, Australia.

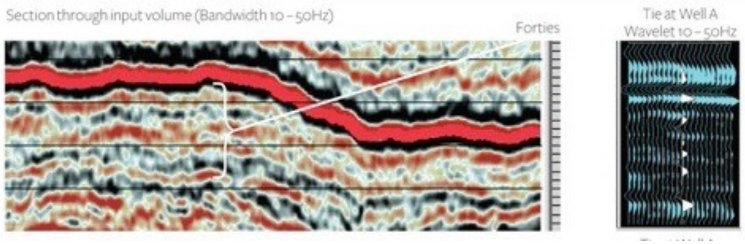
arxiv.org
Emerging Directions in Geophysical Inversion
In this chapter, we survey some recent developments in the field of geophysical inversion....

Sparsity constrained inversion; Optimal Transport; Ensemble based methods; Gaussian processes; Prior Sampling; Posterior Sampling; Variational Methods; Generative models; Surrogate Modelling; Physics informed Neural networks. Data-novel strategies.

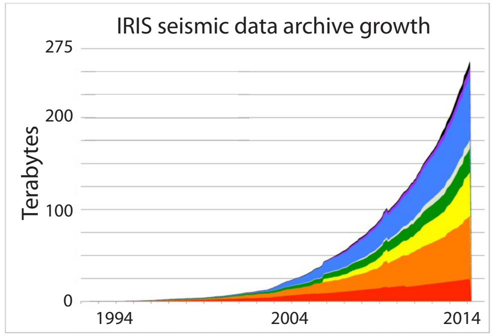
Draft available at https://arxiv.org/abs/2110.06017

Data, data everywhere





MEGATEM, Dash 7

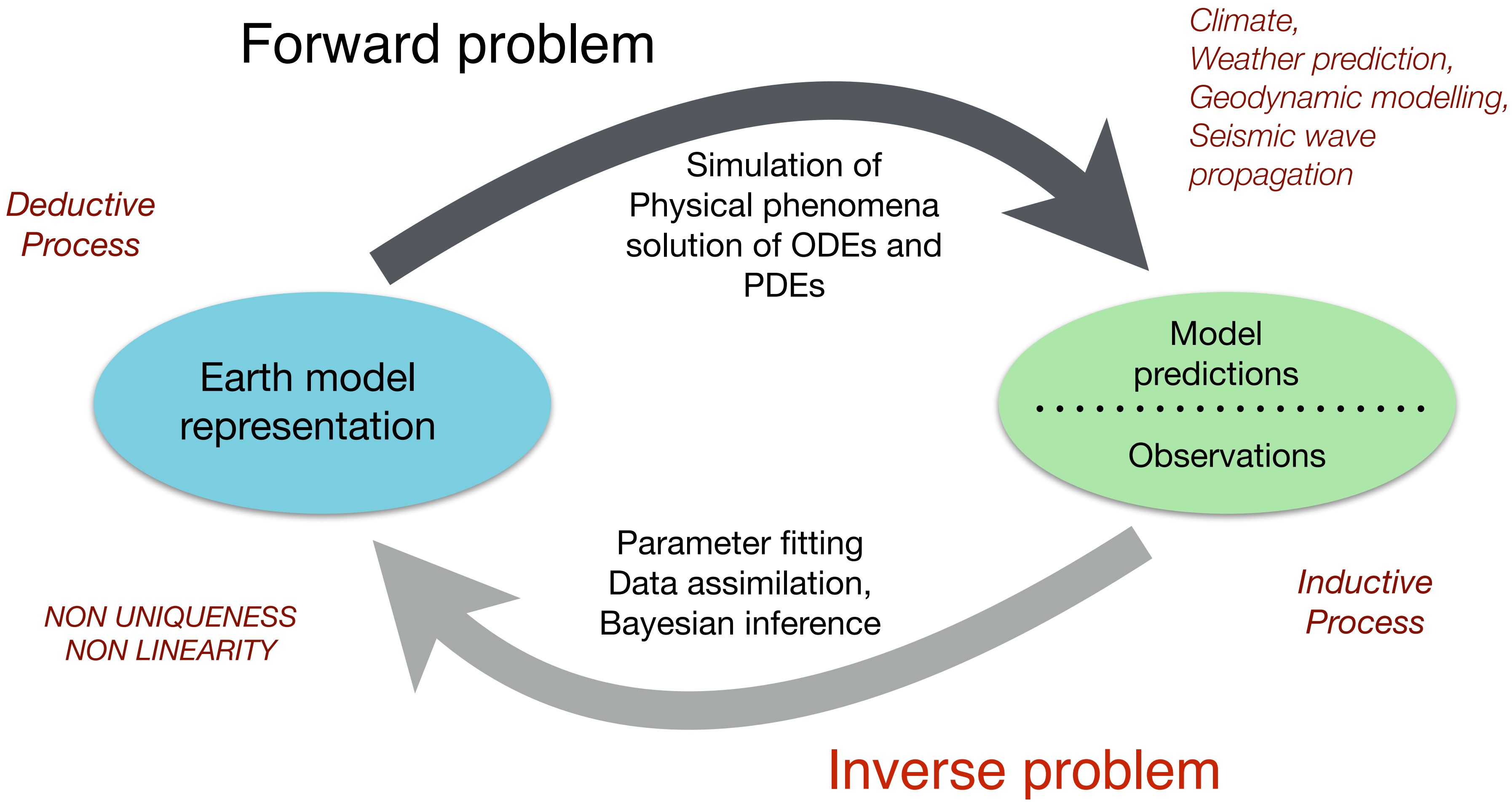


In 2010, global data collection rates from all sources was increasing at 58%, equal to 1.2 x10^{21} bytes p.a.
More than the estimated number of stars in the universe.

-Gantz (2010)

Huge data volumes presents challenges for data custodianship,
but will also lead to new methods for drawing inferences .

Forward and Inverse problems



*'The purpose of models is **not** to fit the data, but to **sharpen** the questions.'*

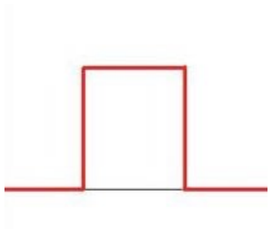
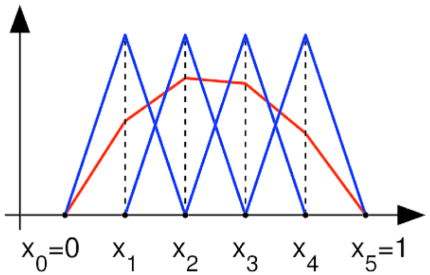
-S. Karlin, 1983

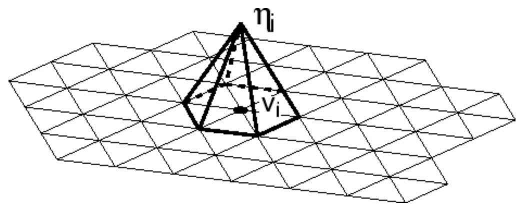
Discretizating a model.

Examples of *Basis functions*

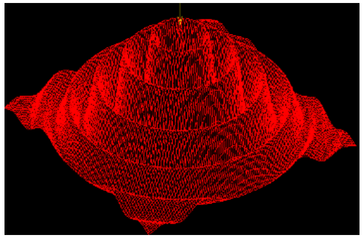
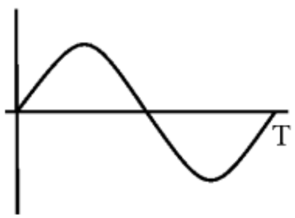
$$m(\mathbf{x}) = \sum_{j=1}^{M} m_j \phi_j(\mathbf{x})$$

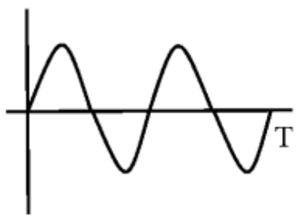
Local support

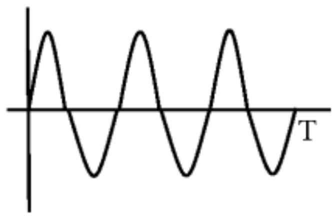



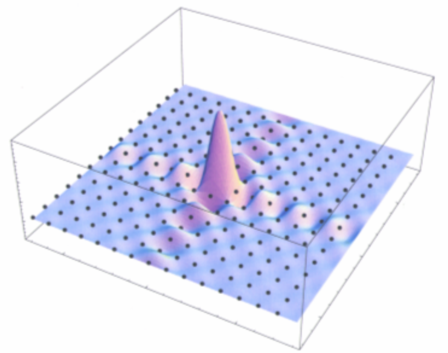


Global support







All inferences we can make about the continuous function will be influenced by the choice of basis functions. They must suit the physics of the forward problem. They bound the resolution of any model one gets out.

Classes of inverse problem

- Linear and discrete

$$\mathbf{d} = G\mathbf{m}$$

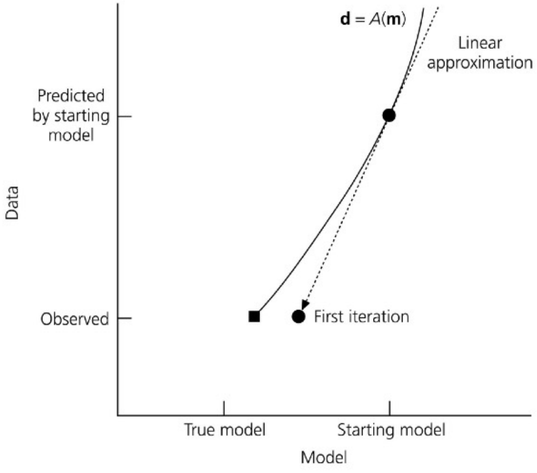
where $\mathbf{d}$ is a data vector, $\mathbf{m}$ is a vector of model parameters, and G is a constant matrix.

- Nonlinear and discrete

$$\mathbf{d} = g(\mathbf{m})$$

where $g(\mathbf{m})$ is some known function of the model parameters.

Figure 7.2-2: Illustration of the the effect of linearizing about an inverse problem starting model.



- Linearised and discrete

$$\delta\mathbf{d} = G\delta\mathbf{m}$$

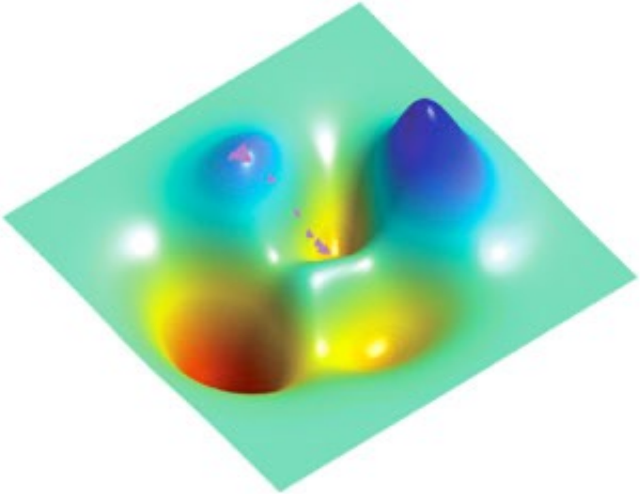
where $\delta\mathbf{m}$ is a perturbation in model parameter from some reference model
$\delta\mathbf{d}$ is a vector of differences between the observations and predictions from $\delta\mathbf{m}$

- Linear and Nonlinear continuous $\int_a^b g(s,x)m(x)dx = d(s)$ $\quad \int_a^b g(s,x,m(x))dx = d(s)$

Two common approaches

`What is the optimum model for my data?'

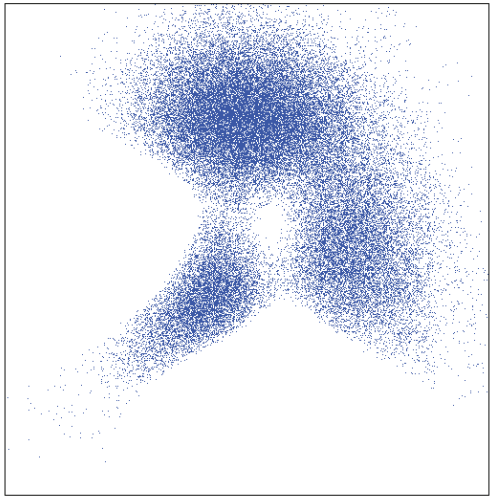
Minimising some misfit function…



$$\phi(\mathbf{m}) = \psi(\mathbf{d},\mathbf{m}) + \alpha\,\chi(\mathbf{m})$$

$$\phi(\mathbf{m}) = ||\mathbf{d} - g(\mathbf{m})||_2^2 + \alpha\,||\mathbf{m} - \mathbf{m}_o||_p^p$$

`What ensemble of possible models do my data support ?'



Requires a Likelihood function

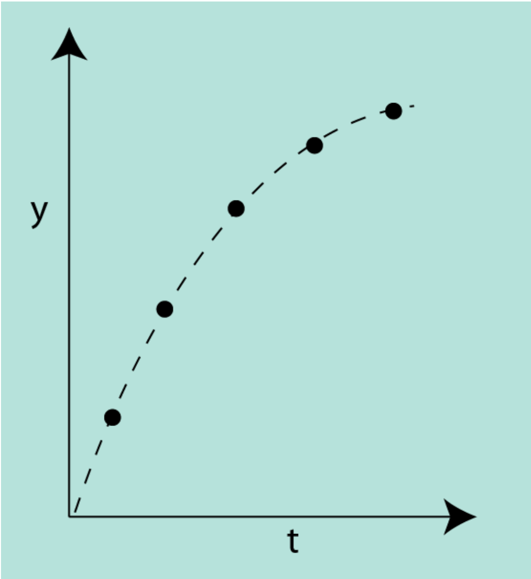
$$p(\mathbf{m}\,|\,\mathbf{d}) = \kappa \times p(\mathbf{d}\,|\,\mathbf{m})\;p(\mathbf{m})$$

$$\log p(\mathbf{m}\,|\,\mathbf{d}) = \log p(\mathbf{d}\,|\,\mathbf{m}) + \log p(\mathbf{m}) + \log\kappa$$

Discrete Linear inversion

Newton's laws of motion

$$y = m_1 + m_2 t - \frac{1}{2} m_3 t^2$$



Find the Initial height, velocity and gravitational constant

of a cannonball when data are height, y, versus time, t.

$$\phi(\mathbf{m}) = \frac{1}{2}(\mathbf{d} - G\mathbf{m})^T C_D^{-1} (\mathbf{d} - G\mathbf{m})$$

$$\mathbf{m}_{LS} = (G^T C_D^{-1} G)^{-1} G^T C_D^{-1} \mathbf{d}$$

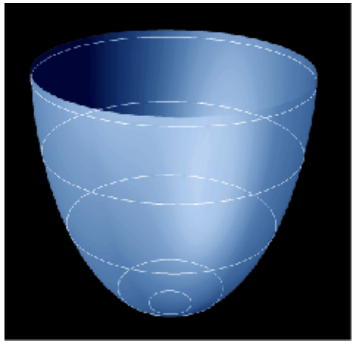
$$\mathbf{d} = G\mathbf{m}$$

$$\mathbf{d} = [y_1, y_2, \ldots, y_N]^T$$

$$\mathbf{m} = [m_1, m_2, m_3]^T$$

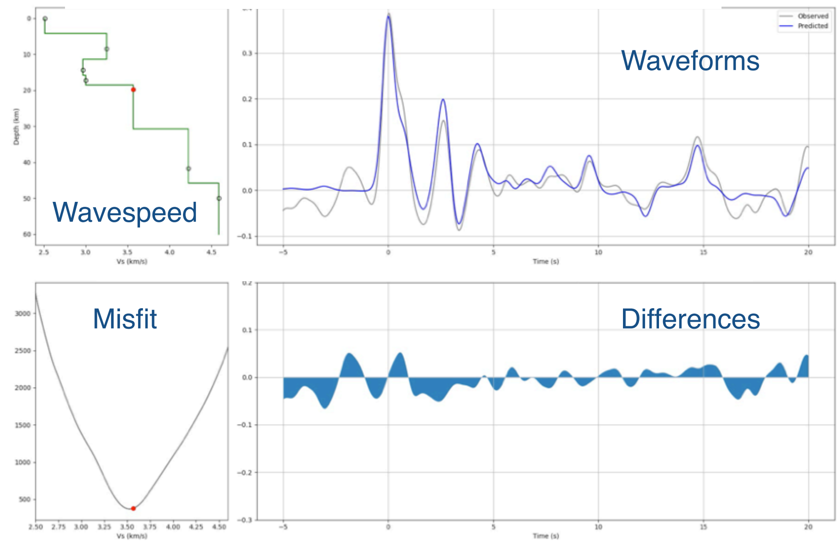
$$G = \begin{pmatrix} 1 & t_1 & -\frac{1}{2}t_1^2 \\ 1 & t_2 & -\frac{1}{2}t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_M & -\frac{1}{2}t_M^2 \end{pmatrix}$$

$\phi(\mathbf{m})$

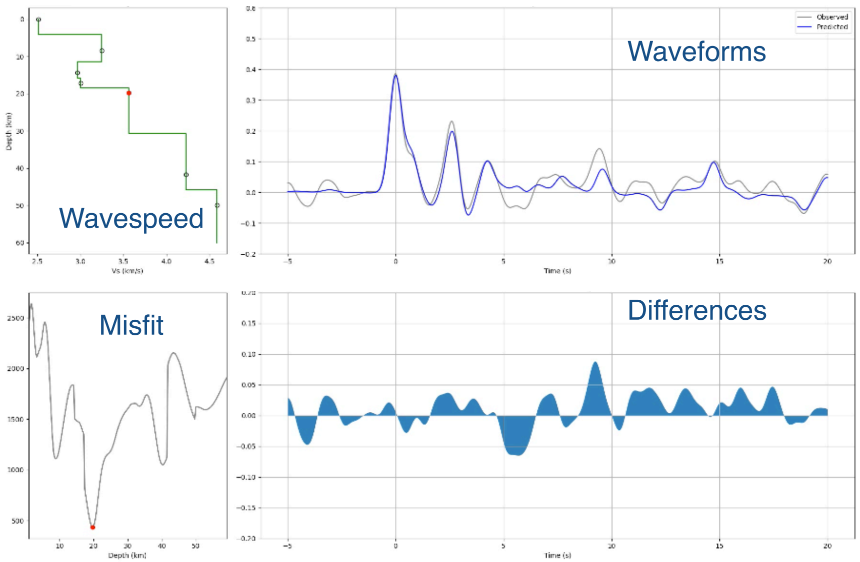


Discrete Nonlinear inversion

Changing the Shear wave speed in a single layer



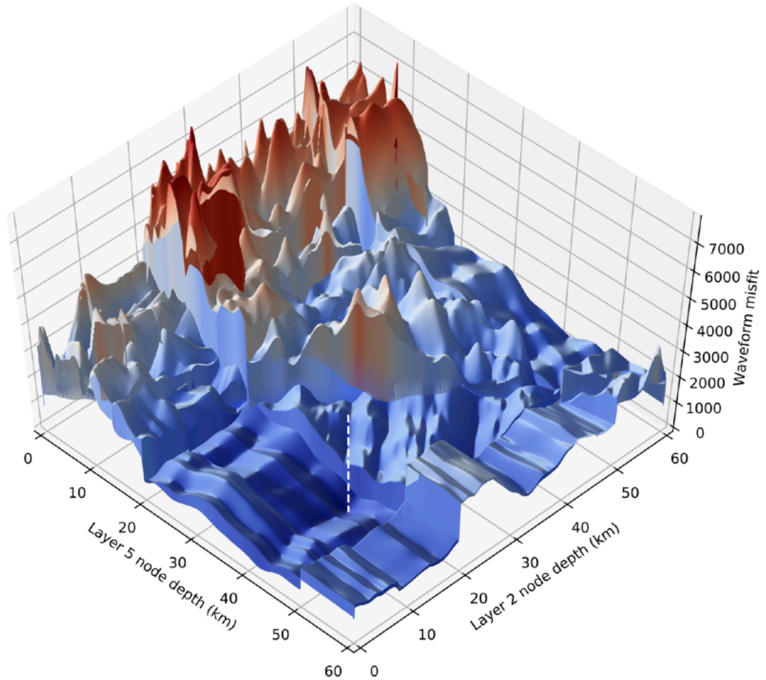
Changing the layer thicknesses



Model fitting can have simple near quadratic misfit functions or complex multi-modal distributions

…and sometimes both at the same time!

How do we explore 'good' fit and meaningful regions of a multi-dimensional parameter space?



The character of the misfit landscape is a function of how one measures fit.

Sparsity

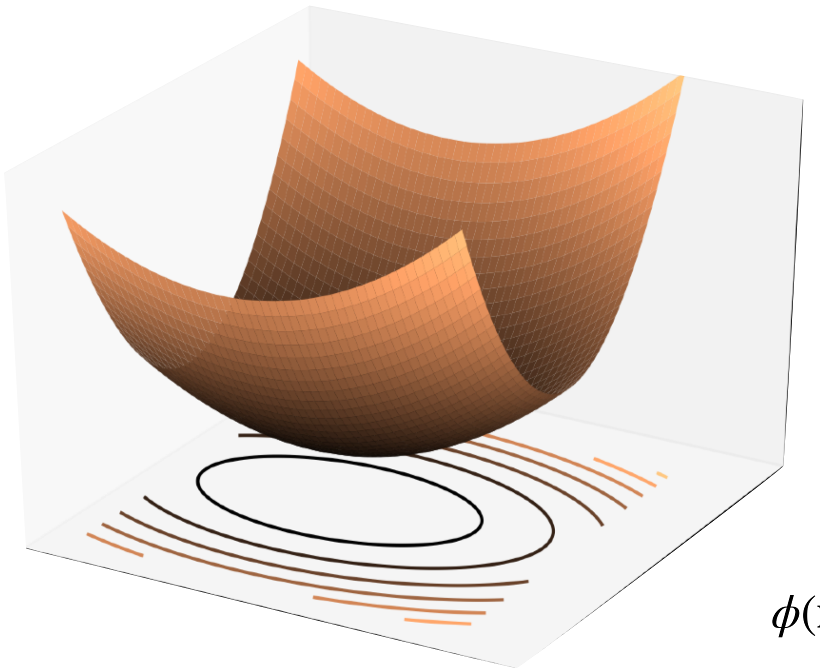
Applications to data collection, Seismic imaging, Seismic wavefield reconstruction

Under-determined problems

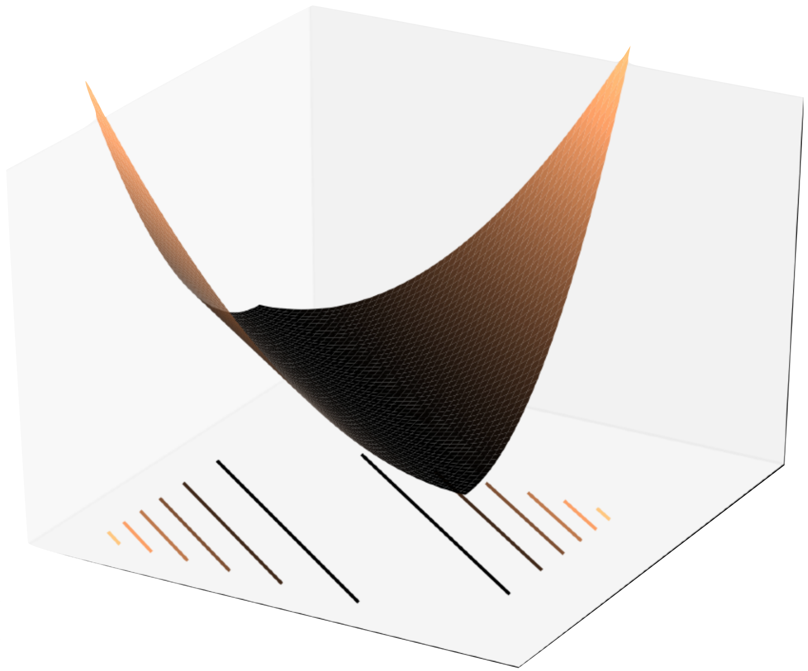
If the columns of matrix G are linearly dependent, then $G^T C_D^{-1} G$ is not full rank and cannot be inverted!

$$\mathbf{m}_{LS} = (G^T C_D^{-1} G)^{-1} G^T C_D^{-1} \mathbf{d}$$

Unique minimum in $\phi(\mathbf{m})$



Non-unique minimum $\phi(\mathbf{m})$



$$\phi(\mathbf{m}) = ||\mathbf{d} - G\mathbf{m}||_2^2$$

Sparsity

Looking for sparse solutions to linear and nonlinear parameter estimation

Reduce data misfit *Encourage sparsity*

$$\phi(\mathbf{m}) = ||\mathbf{d} - G\mathbf{m}||_2^2 + \alpha||\mathbf{m}||_1$$

$$||\mathbf{m}|| = \sum_j |m_j|$$

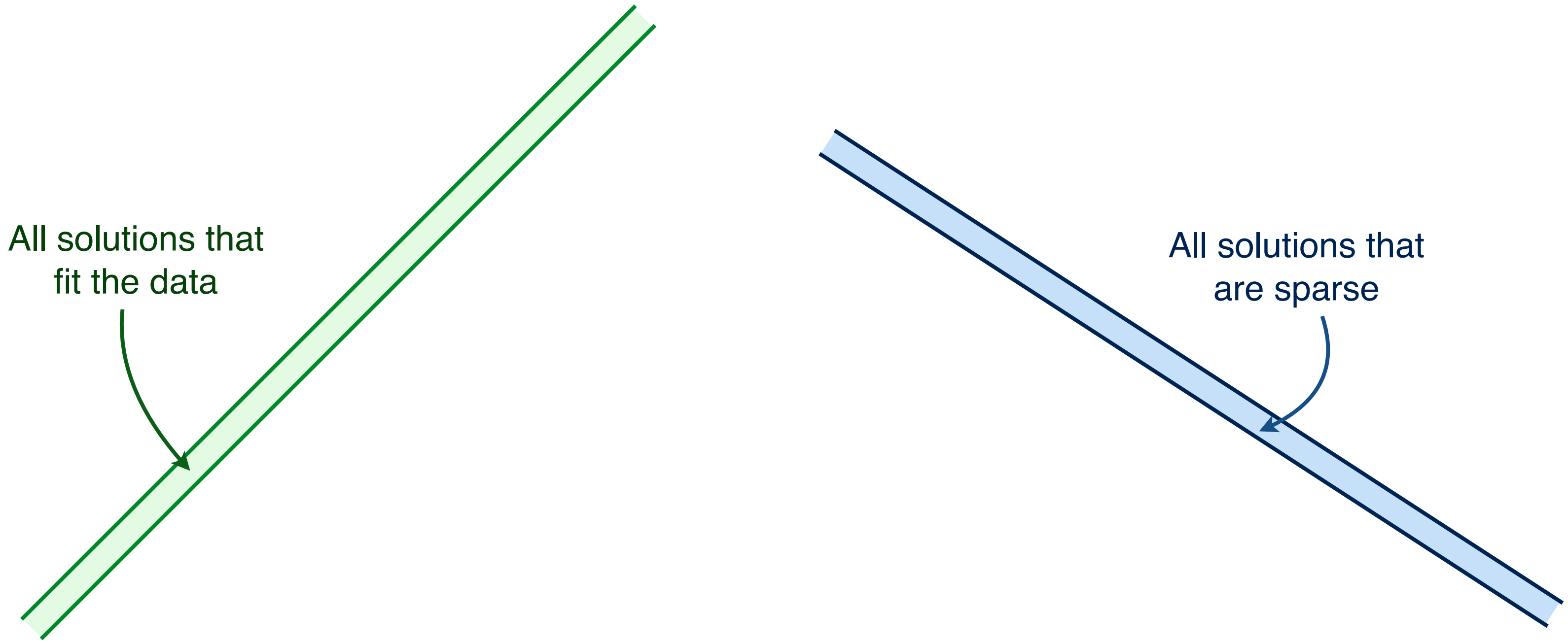
Use of L_1 norms in geophysical inversion dates back at least to Claerbout and Muir (1973), who applied it as a robust data misfit. Scales et al. (1998) used it as a regularization term in seismic tomography.

L_1 regularization norms encouraging sparse solutions to under-determined problems, i.e. indicating that we prefer the model parameters to be zero.

L_0 regularization norms guarantee sparse solutions, but are too difficult to solve.

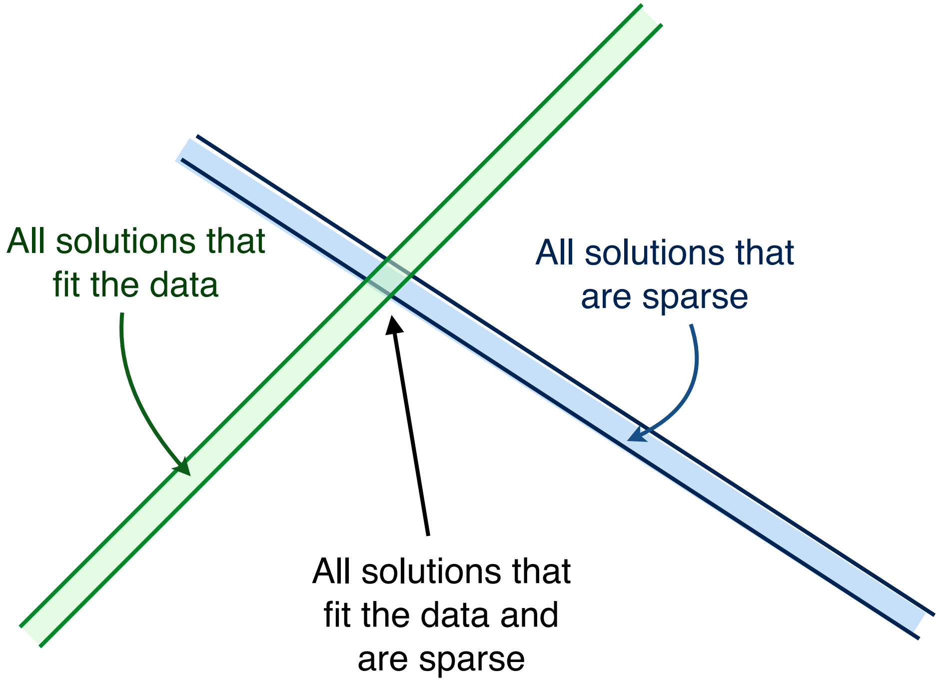
Why does sparsity maximisation work?

If sensor basis is incoherent with the model basis and the true model is sparse in a known basis



Why does sparsity maximisation work?

If sensor basis is incoherent with the model basis and the true model is sparse in a known basis



Compressive sensing in a nutshell

A novel technology for reconstructing time series from an irregularly sampled subset of data. It uses some novel mathematical principles that provide accurate reconstruction of complex signals from minimal observations.

Model parameter are coefficients of a model basis, $\phi_j(t)$

$$d_p(t) = \sum_j m_j \ \phi_j(t)$$

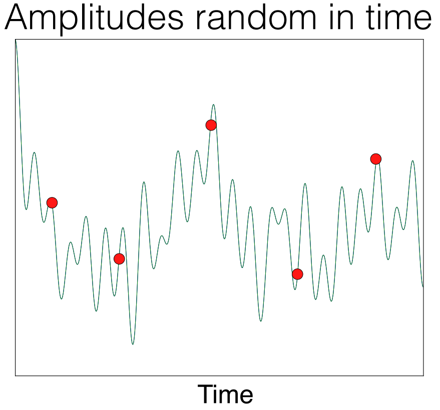
Data values are measured by a sensor basis,

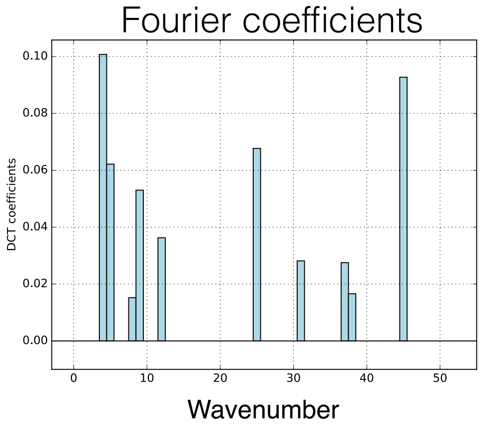
$$d_i^{obs} = \int s_i(t) \ d(t) \ dt$$

Provided there is *incoherence* between sensor and model basis and provided the model is *sparse* in its basis.

$$\phi(\mathbf{m}) = ||\mathbf{d} - G\mathbf{m}||_2^2 + \alpha||\mathbf{m}||_1$$

Sufficient random samples can yield exact solutions!





Candes et al. (2006), Candes & Tao (2006), Donoho (2006), Candes and Watin (2008), Hermann et al. (2008, 2009)

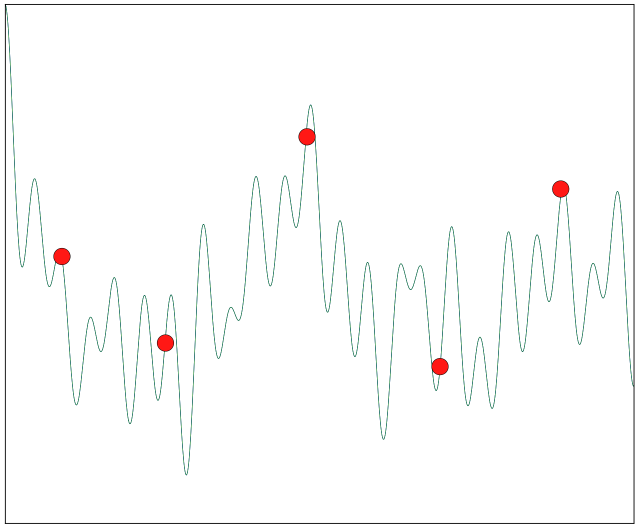
Compressive sensing example

Recovering a time series from random samples

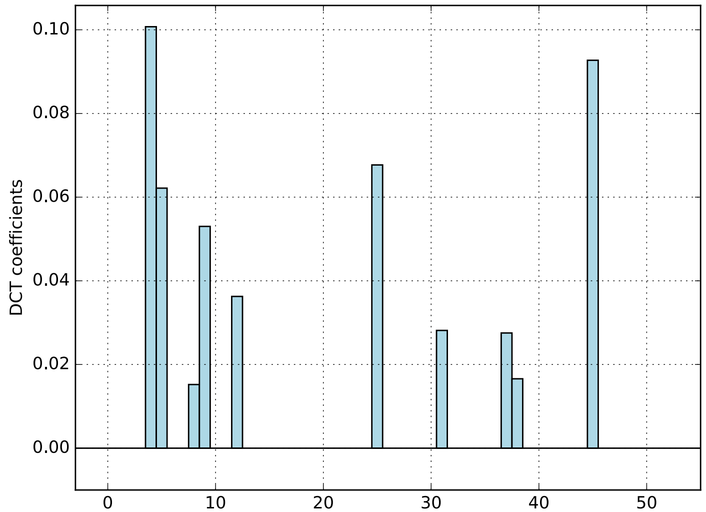
Sensor basis: $s_i(t) = \delta(t - t_i)$ Amplitudes at random times

Model basis: $\phi_j(t) = e^{i2\pi f_j t}$, Fourier basis

Amplitudes random in time



Fourier coefficients



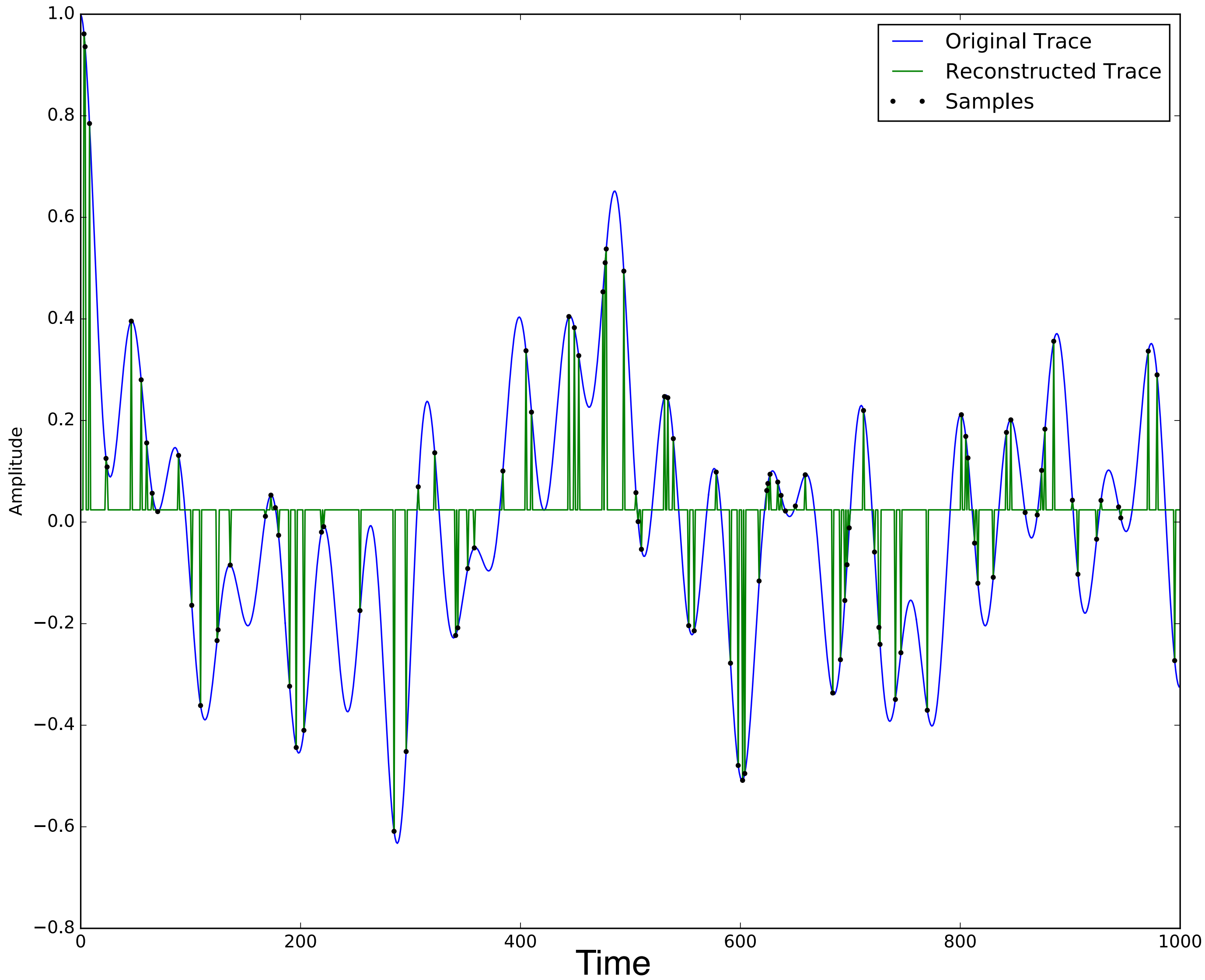
$$\phi(\mathbf{m}) = \sum_i (d_i^{obs} - d_p(t_i))^2 + \alpha \sum_j |m_j|^p$$

$p = 1$: Sparsity

$p = 2$: Least squares

Least squares reconstruction ($p = 2$)

Original signal (blue) and reconstruction (green).



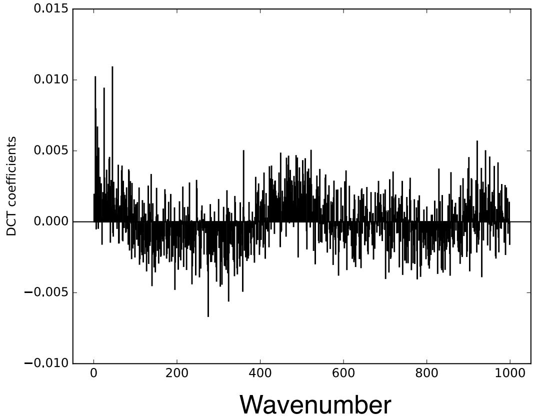
- data samples

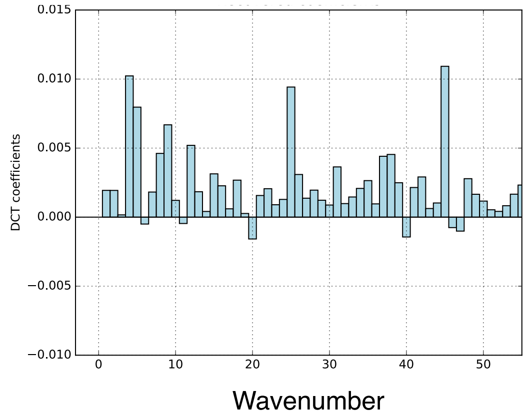
Least squares reconstruction ($p = 2$)

Recovering a time series from random samples

999 recovered finite coefficients

Recovered Fourier coefficients

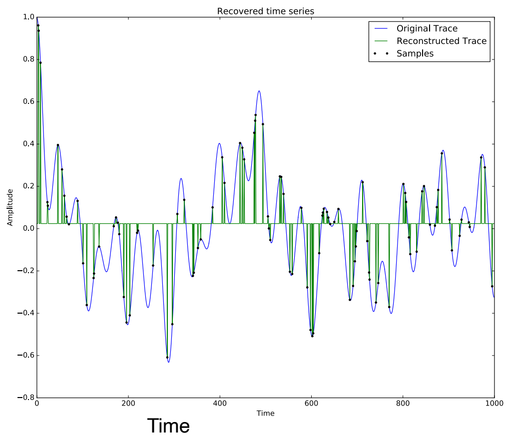




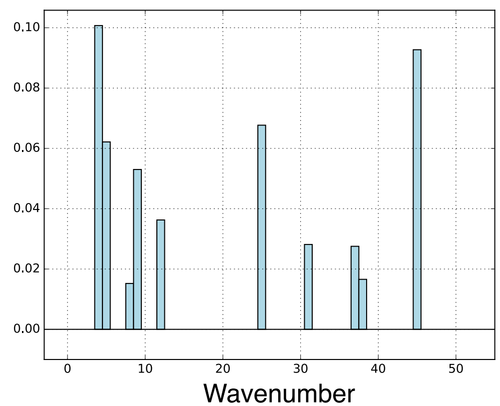
1000 potential wavenumber

Note poor amplitude recovery with many non zero coefficients

Original signal and reconstruction



True Fourier coefficients



10 True finite coefficients

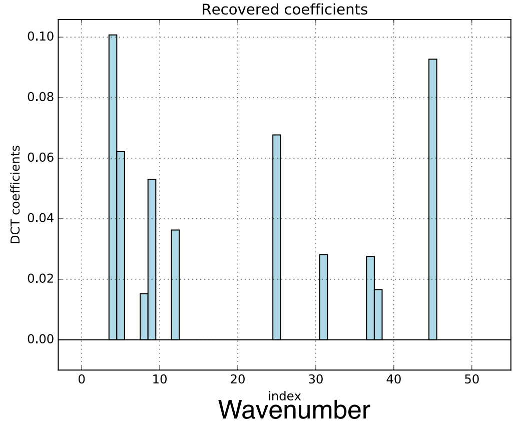
Compressed sensing reconstruction ($p = 1$)

Recovering a time series from random samples

10 recovered finite coefficients

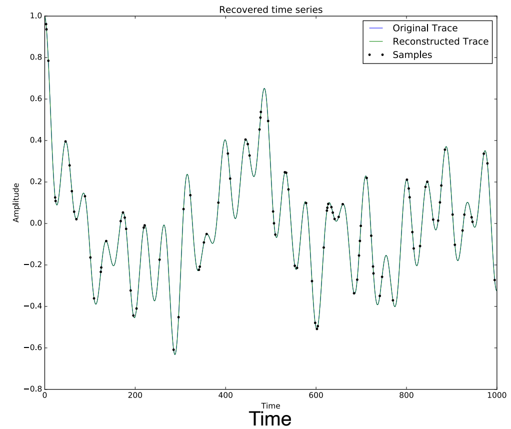
1000 possible coefficients

Recovered Fourier coefficients

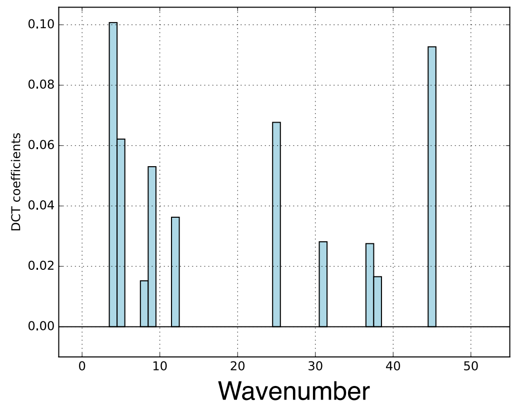


Exact amplitude and wavenumber recovery

Original signal and reconstruction



True Fourier coefficients

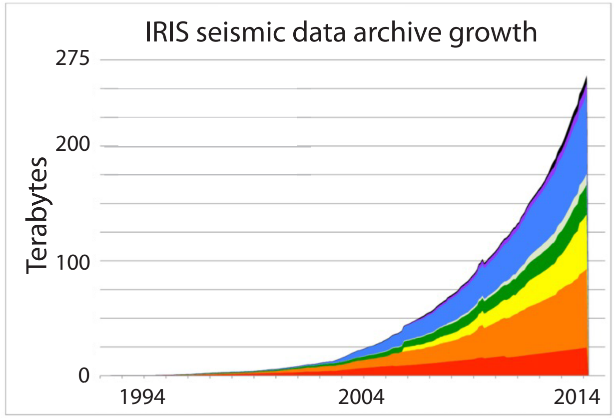


10 True finite coefficients

The age of big data

Looking for sparse solutions to linear and nonlinear parameter estimation

In seismology



The global rate of data collection is estimated to be increasing at 58% per year, which in 2010 alone amounted to 1250 billion gigabytes, more bytes than the estimated number of stars in the universe - (Gantz, 2010)

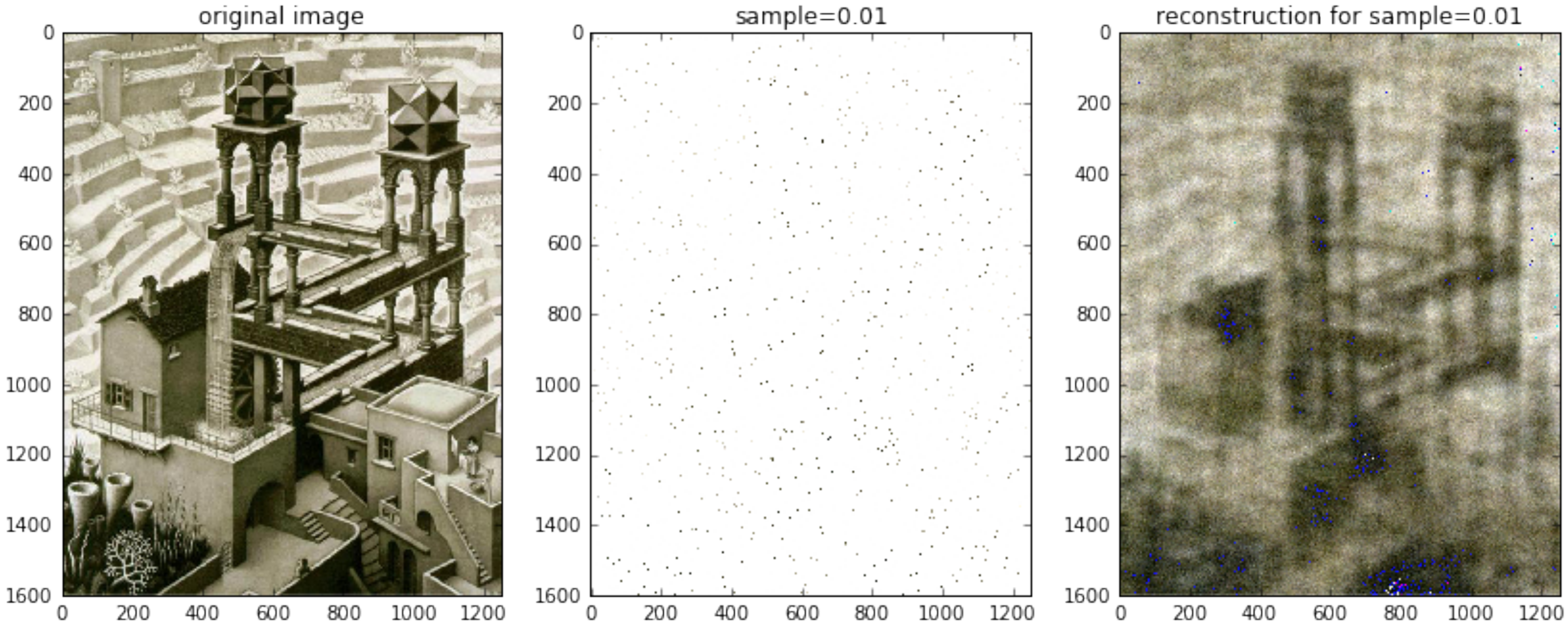
Since 2007 we have been generating more bits of data per year than can be stored in all of the world's storage devices (Gantz, 2010)

This creates challenges for storage and transmission of sensor recordings

Sparsity based image reconstruction

Compressive sensing concepts have applications in image reconstruction. Sparsity constrained optimization. Candes et al. (2006), Candes & Tao (2006), Donoho (2006), Candes and Watin (2008).

$$\min_{\mathbf{m}} \; ||\mathbf{d} - G\mathbf{m}||_2 + \lambda||\mathbf{m}||_1$$



Waterfall by M. C. Escher (1200 x1600)

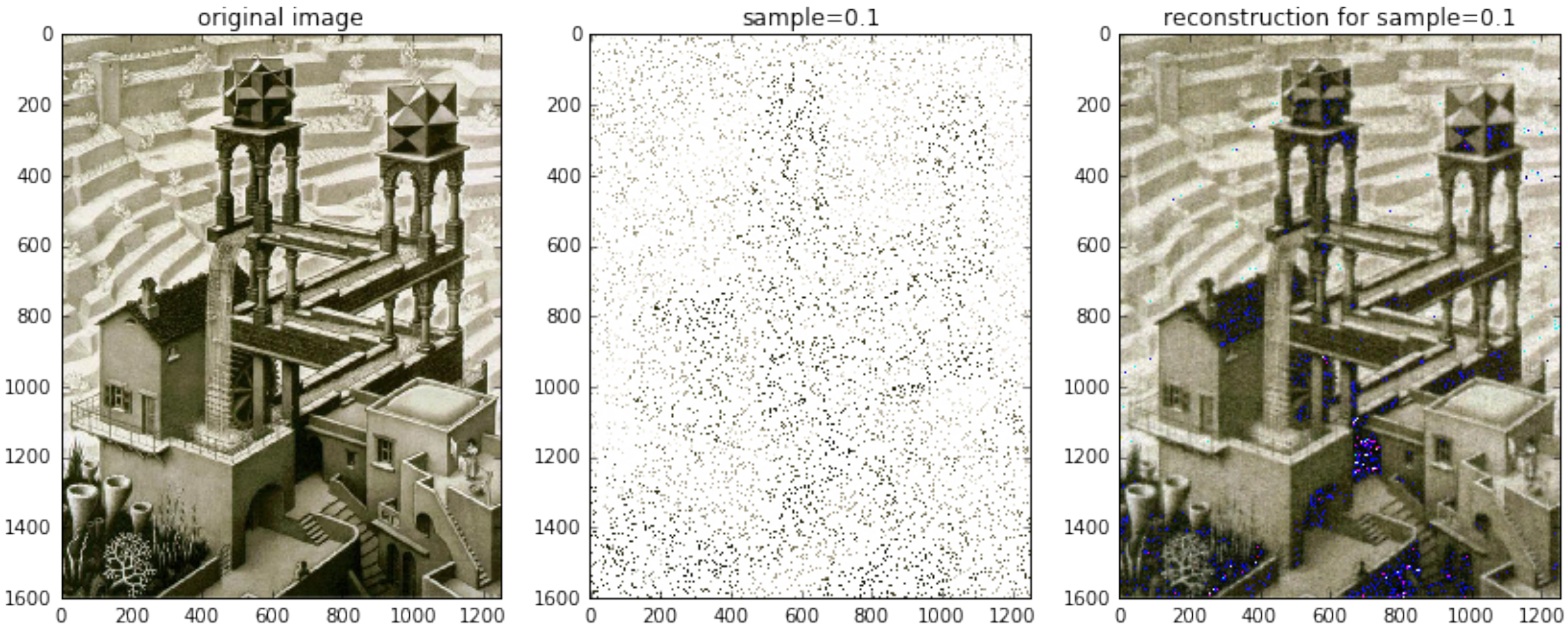
Discrete cosine basis

From Robert Taylor's pyrunner blog http://www.pyrunner.com/weblog/2016/05/26/compressed-sensing-python/

Sparsity based image reconstruction

Compressive sensing concepts have applications in image reconstruction. Sparsity constrained optimization. Candes et al. (2006), Candes & Tao (2006), Donoho (2006), Candes and Watin (2008).

$$\min_{\mathbf{m}} \; ||\mathbf{d} - G\mathbf{m}||_2 + \lambda||\mathbf{m}||_1$$



Waterfall by M. C. Escher (1200 x1600)

Discrete cosine basis

From Robert Taylor's pyrunner blog http://www.pyrunner.com/weblog/2016/05/26/compressed-sensing-python/

Use of L1 sparsity constraints has a long history in seismic imaging

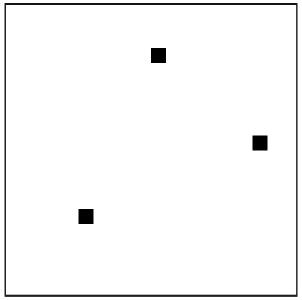
Scales et al. (1998); Hermann et al. (2008, 2009), Simons et al. (2011), Loris et al. (2012), Charléty et al. (2013).

But the Earth is not known to be sparse in any convenient basis.

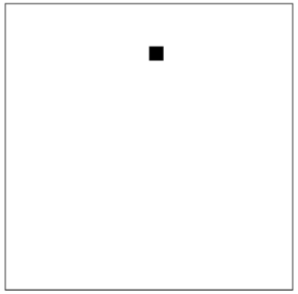
Overcomplete tomography

Sparse models $m(\mathbf{x}) = \sum_i^M \alpha_i \phi_i(\mathbf{x})$

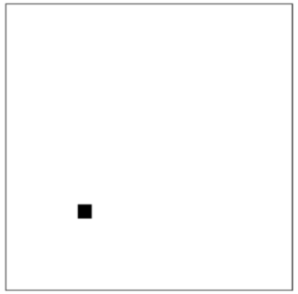
Pixel basis



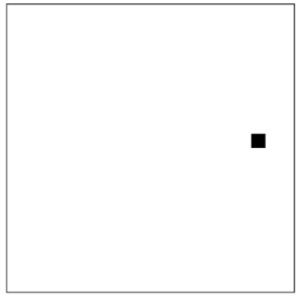
$=\ 1.0 \times$



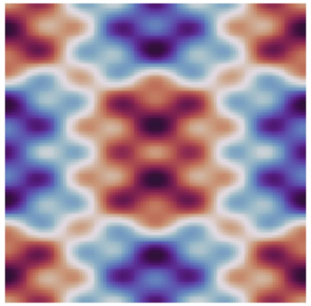
$+1.0 \times$



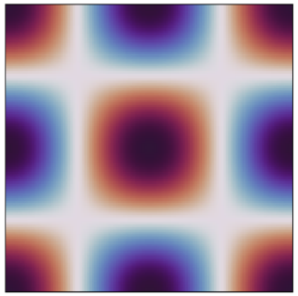
$-1.0 \times$



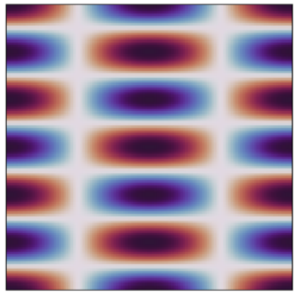
Cosine basis



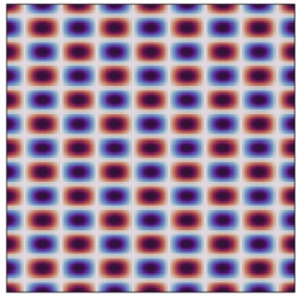
$=\ 0.9 \times$



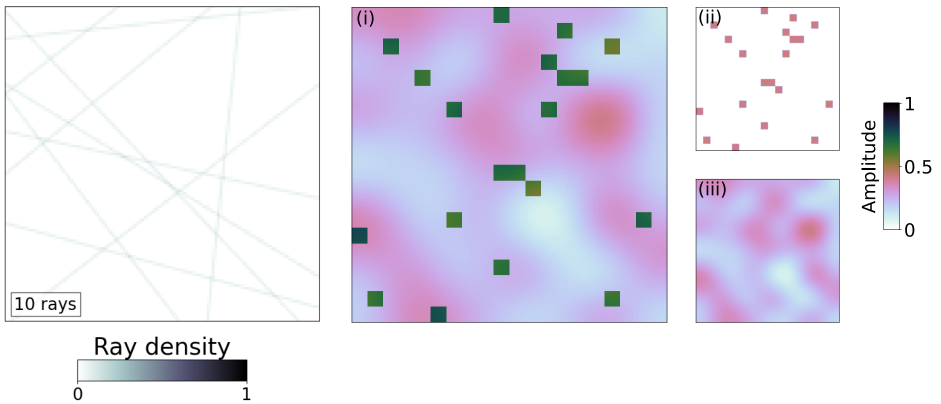
$-0.3 \times$



$-0.3 \times$



$$m(\mathbf{x}) = \sum_i^M \alpha_i \phi_i(\mathbf{x}) + \sum_j^M \beta_j \psi_j(\mathbf{x})$$



Courtesy Buse Turunçtur

Overcomplete tomography example