Entropy-guided least squares: a robust approach to scattered data approximation

Felice Iavernaro (University of Bari, Italy)

FAIR Future Artificial Intelligence Research — 6 spoke

Symbiotic AI

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Spoke 5
Environment & Natural Disasters

PEOPLE INVOLVED

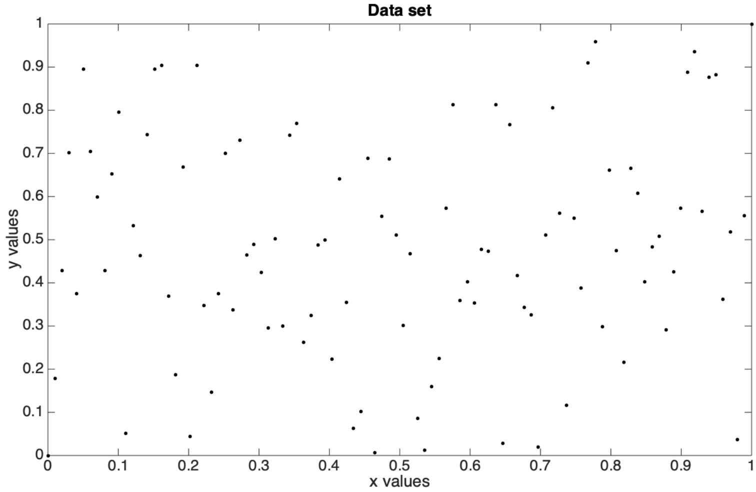
Path of Research: Exploit the meaning of entropy in statistical mechanics and information theory to increase the robustness of a given model fitting procedure.

- Pierluigi Amodio (University of Bari, Italy)
- Luigi Brugnano (University of Florence, Italy)
- Antonella Falini (University of Bari, Italy)
- Domenico Giordano (ESTEC (retired), European Space Agency, The Netherlands)
- Francesca Mazzia (University of Bari, Italy)

OUTLINE

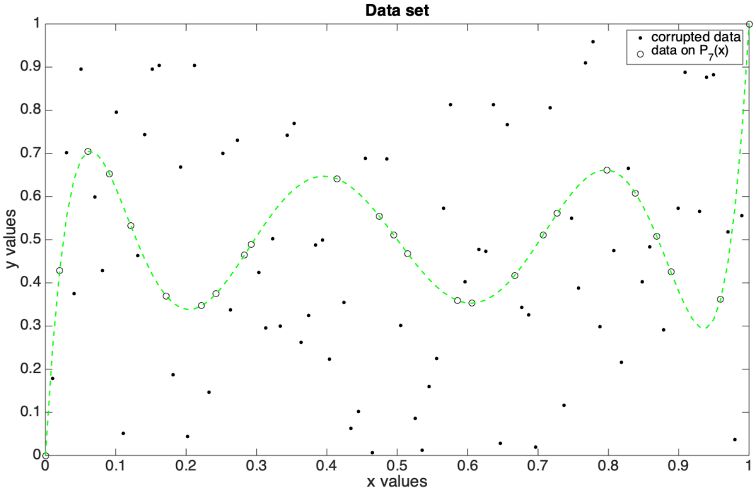
- The hidden polynomial problem
- Motivations
- The role of entropy in information theory and statistical mechanics
- Definition of the entropy-based methodology in a general setting
- Illustrations of the technique
- Applications of the maximum entropy principle in least squares approximation problems:
  - splines functions and curves
  - smoothing splines
  - bivariate splines
- Conclusions and future work

The hidden polynomial problem



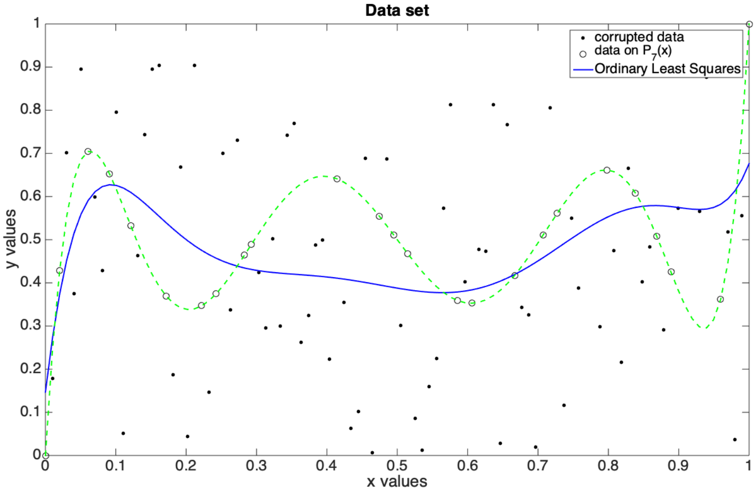
Total data set (dots) 100 points. *Find the polynomial of degree 7 which interpolates as many points as possible.*

The hidden polynomial problem



- 25 points lying on the Legendre polynomial of degree seven on $[0, 1]$;
- 75 points randomly chosen.

The hidden polynomial problem



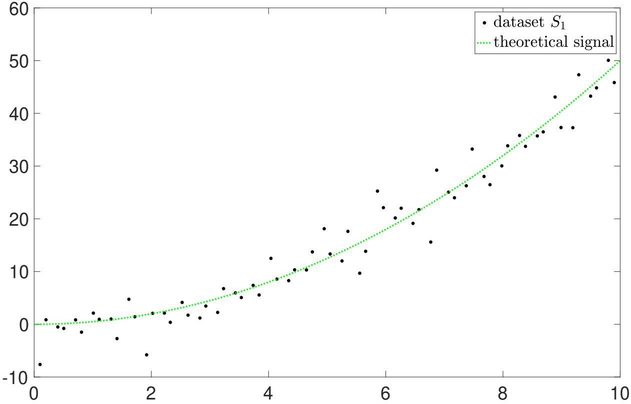
Ordinary Least Squares

Generalization to a linear algebra problem

"Solve" an overdetermined linear system in the following sense:

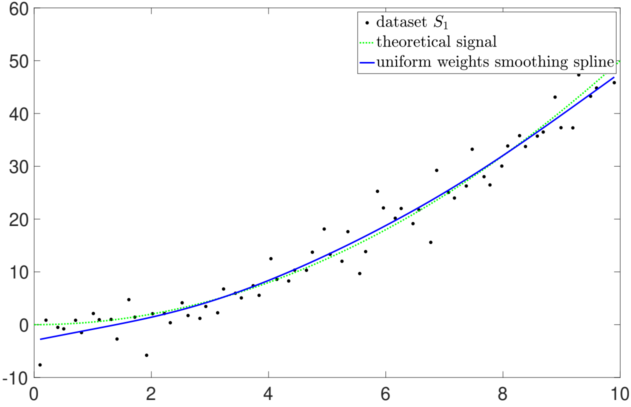
from an overdetermined (possibly inconsistent) linear system, remove the minimum number of equations to make the resulting system consistent.

Motivations



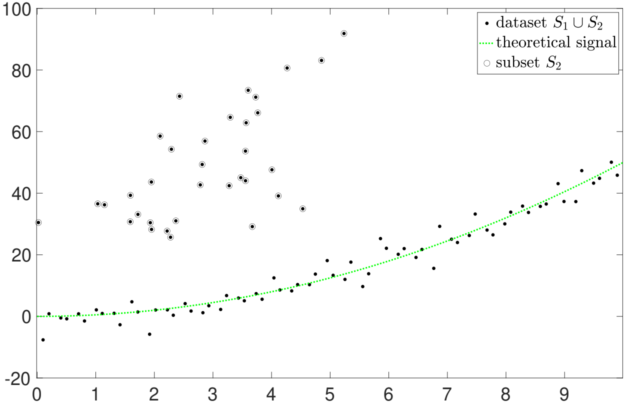
- Dataset S_1: 66 points aligned along the arc of parabola $y = 0.5 + x^2$ (green dotted line) but corrupted by Gaussian noise $\mathcal{N}(0, \sigma^2)$ with mean zero and variance $\sigma^2 = 9$.

Motivations



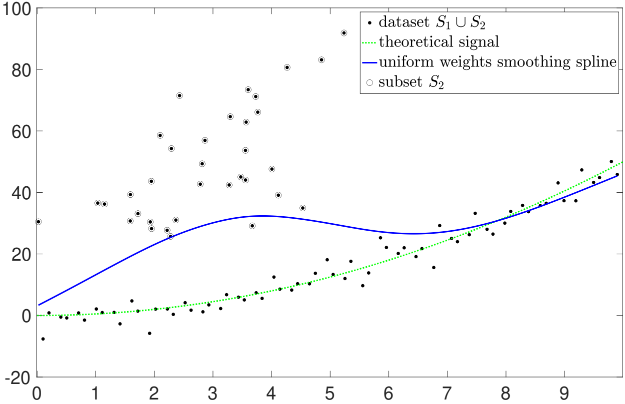
Ordinary least squares approximation with a smoothing spline, with smoothing parameter λ obtained through Leave-One-Out Cross Validation.

Motivations



- S_1: 66 points aligned along the arc of parabola $y = 0.5 + x^2$ (black dotted line) but are corrupted by Gaussian noise $\mathcal{N}(0, \sigma^2)$ with mean zero and variance $\sigma^2 = 9$.
- S_2: a random cloud of 34 points (black dots surrounded by circles) confined above the left part of the arc of parabola, departing from the pattern described by S_1.

Motivations



- Real data are often affected by massive uncertainties due to poor sampling or measurement, data entry or processing errors.
- In such cases the OLS (Ordinary Least Squares) approximation fails to capture the correct pattern.

Entropy in information theory: Shannon's theorem

Consider a collection of potential events with associated probabilities of occurrence $p_1, p_2, \dots, p_n$. Is it possible to quantify the level of uncertainty regarding the outcome? Denote by $H(p_1, \dots, p_n)$ such a measure.

Assumptions:

- continuity of $H(p_1, \dots, p_n)$ w.r.t. its arguments p_i
- monotonicity: if all the p_i are equal to $1/n$, H is a monotonic increasing function of n
- consistency: if we "split up" the probability space in different ways, the value of H should remain unchanged. For example: $H(1/2, 1/3, 1/6) = H(1/2, 1/2) + 1/2 \cdot H(2/3, 1/3)$

Up to a scaling positive constant, the only H satisfying the three above assumptions is of the form

$$H(p_1, \dots, p_m) = -\sum_{i=1}^{m} p_i \log p_i$$

C.E. Shannon, A Mathematical Theory of Communication. Bell System, Technical Journal 27(3), 379–423 (1948)

Jaynes' principle of maximum entropy

- When assigning probabilities to an event where no outcome appears more likely than any other, we traditionally distribute the probabilities equally among all possible outcomes. This approach is known as the principle of indifference, a concept rooted in the work of Laplace.

Jaynes' principle of maximum entropy

- When assigning probabilities to an event where no outcome appears more likely than any other, we traditionally distribute the probabilities equally among all possible outcomes. This approach is known as the principle of indifference, a concept rooted in the work of Laplace.
- However, when we happen to know or learn something about the non-uniformity of the outcomes, the method for assigning probabilities must be adjusted. This adjustment is guided by an extension of the principle of indifference, known as the principle of maximum entropy.

Jaynes' principle of maximum entropy

- When assigning probabilities to an event where no outcome appears more likely than any other, we traditionally distribute the probabilities equally among all possible outcomes. This approach is known as the principle of indifference, a concept rooted in the work of Laplace.
- However, when we happen to know or learn something about the non-uniformity of the outcomes, the method for assigning probabilities must be adjusted. This adjustment is guided by an extension of the principle of indifference, known as the principle of maximum entropy.

According to Jaynes:

> ...the distribution that maximizes H, subject to constraints which represent whatever information we have, provides the most honest description of what we know.

E.T. Jaynes, Foundations of probability theory and statistical mechanics, in: Mario B., ed. Delaware Seminar in the Foundations of Physics, Studies in the Foundations Methodology and Philosophy of Science 1, New York NY: Springer-Verlag (1967) 77–101.

Jaynes' principle of maximum entropy

- When assigning probabilities to an event where no outcome appears more likely than any other, we traditionally distribute the probabilities equally among all possible outcomes. This approach is known as the principle of indifference, a concept rooted in the work of Laplace.
- However, when we happen to know or learn something about the non-uniformity of the outcomes, the method for assigning probabilities must be adjusted. This adjustment is guided by an extension of the principle of indifference, known as the principle of maximum entropy.
- The principle of maximum entropy provides a systematic way to update probability distributions by incorporating new knowledge while maintaining the least bias possible.

The entropy-based approach in a general setting

Ingredients:

- a dataset $(y_i)_{i=1,\dots,m}$
- a proper model to fit the data $f(c_1,\dots,c_s)$
- a cost function C that may be written as a weighted mean over cost functions C_i for individual training points y_i:

$$C = \sum_{i=1}^{m} w_i C_i, \qquad \sum_{i=1}^{m} w_i = 1$$

The entropy-based approach in a general setting

Ingredients:

- a dataset $(y_i)_{i=1,\dots,m}$
- a proper model to fit the data $f(c_1,\dots,c_s)$
- a cost function C that may be written as a weighted mean over cost functions C_i for individual training points y_i:

$$C = \sum_{i=1}^{m} w_i C_i, \qquad \sum_{i=1}^{m} w_i = 1$$

ENTROPY. Due to the normalization condition, we can regard the weight distribution $\{w_i\}$ as a probability distribution and compute the associated entropy

$$H(w_1,\dots,w_m) = -\sum_{i=1}^{m} w_i \log w_i$$

This may be regarded as the expected value of the information content (or uncertainty).

The entropy-based approach in a general setting

Ingredients:

- a dataset $(y_i)_{i=1,\dots,m}$
- a proper model to fit the data $f(c_1,\dots,c_s)$
- a cost function C that may be written as a weighted mean over cost functions C_i for individual training points y_i:

$$C = \sum_{i=1}^{m} w_i C_i, \qquad \sum_{i=1}^{m} w_i = 1$$

Use the principle of maximum entropy and solve the constrained problem

maximize $\quad -\sum_{i=1}^{m} w_i \log w_i,$ (entropy)

subject to: $\quad \sum_{i=1}^{m} w_i = 1,$ (normalization condition)

C prescribed to a given value

Maximal entropy weighted least squares approximation

In this talk, I will consider:

- dataset:
  - a time series $(t_i, y_i)_{i=1,\ldots,m}$, $\quad y_i \in \mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$
  - a dataset representing an image or a point cloud
- model:
  - a polynomial $f(t, c_1, \ldots, c_s)$ to fit the data
  - a (univariate, bivariate) spline function, spline curve/surface,
  - a smoothing spline.
- cost function: the weighted mean squared error

$$\overline{\mathrm{E}^2} = \sum_{i=1}^{m} w_i ||f(t_i, c_1, \ldots, c_s) - y_i||_2^2,$$

or, for smoothing splines, the penalized weighted mean squared error:

$$\overline{\mathrm{E}^2}(g, \lambda, w) = \sum_{i=1}^{m} w_i (g(t_i) - y_i)^2 + \lambda \int_a^b (g''(t))^2 \mathrm{d}t.$$

OLS vs MEWLS

- In (weighted) Ordinary Least Squares, given a distribution of weights, we minimize the Mean Squared Error (MSE)

$$\overline{E^2} = \sum_{i=1}^{m} w_i ||f(t_i, c_1, \ldots, c_s) - y_i||_2^2,$$

which leads to a linear problem. $\overline{E^2}$ is an OUTPUT parameter

OLS vs MEWLS

- In (weighted) Ordinary Least Squares, given a distribution of weights, we minimize the Mean Squared Error (MSE)

$$\overline{\mathrm{E}^2} = \sum_{i=1}^{m} w_i ||f(t_i, c_1, \ldots, c_s) - y_i||_2^2,$$

which leads to a linear problem. $\overline{\mathrm{E}^2}$ is an OUTPUT parameter

- In Maximal Entropy Weighted Least Squares, we decide a priori what the Mean Squared Error should be, and the Maximum Entropy Principle guides us in the choice of the optimal weight distribution and the coefficients c_i:

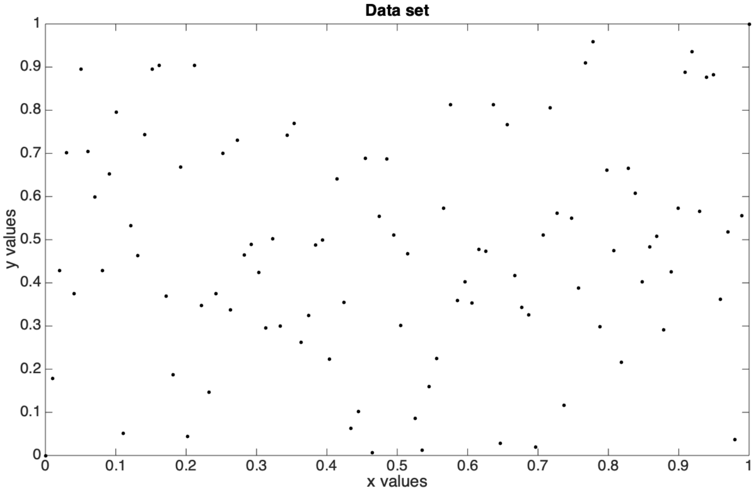
$$\text{maximize} \quad -\sum_{i=1}^{m} w_i \log w_i \qquad \text{(entropy)}$$

$$\text{subject to} \quad \sum_{i=1}^{m} w_i = 1 \qquad \text{(weights normalization condition)}$$

$$\sum_{i=1}^{m} w_i ||f(x_i, c) - y_i||^2 = \overline{\mathrm{E}^2} \qquad \text{(desired Mean Squared Error)}$$

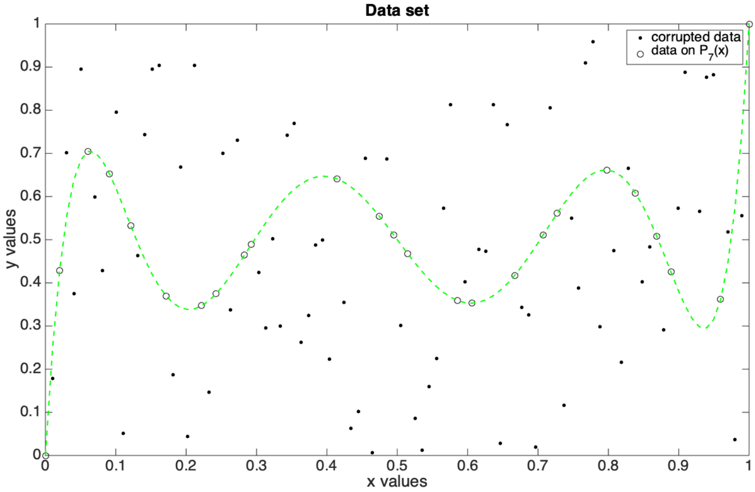
This leads to a nonlinear problem. $\overline{\mathrm{E}^2}$ is an INPUT parameter

The hidden polynomial problem solved by the MEWLS



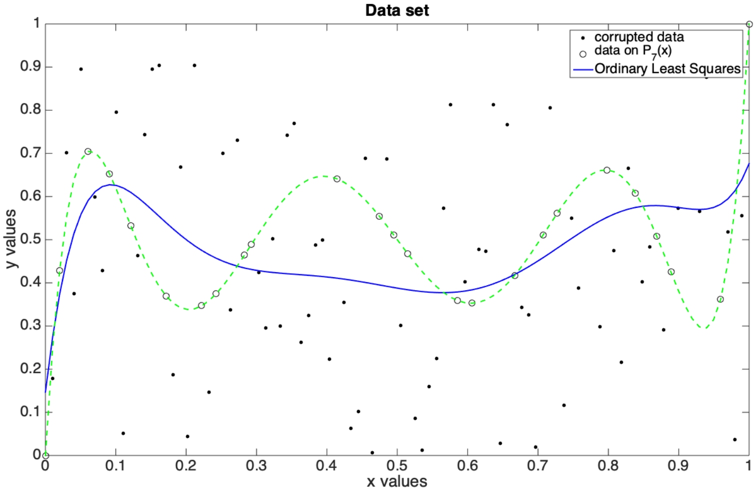
Total data set (dots) 100 points.

The hidden polynomial problem solved by the MEWLS



- 25 points lying on the Legendre polynomial of degree seven on $[0, 1]$;
- 75 points: corrupted data.

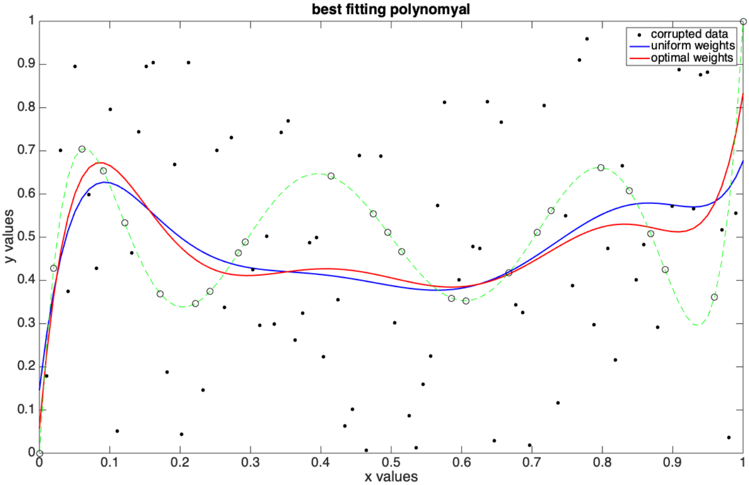
The hidden polynomial problem solved by the MEWLS



Ordinary Least Squares

- $\overline{E^2}_{\mathrm{uw}} = 5.81 \cdot 10^{-2}$
- Entropy $= 4.60$

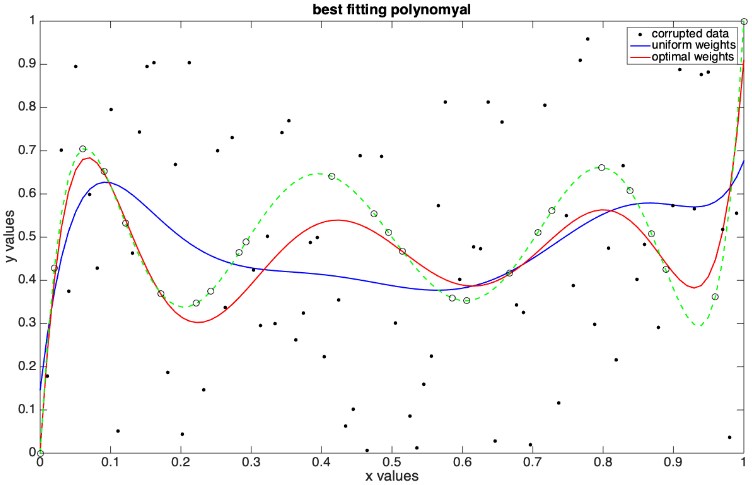
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw}/2 = 2.90 \cdot 10^{-2}$
- Entropy $= 4.49$

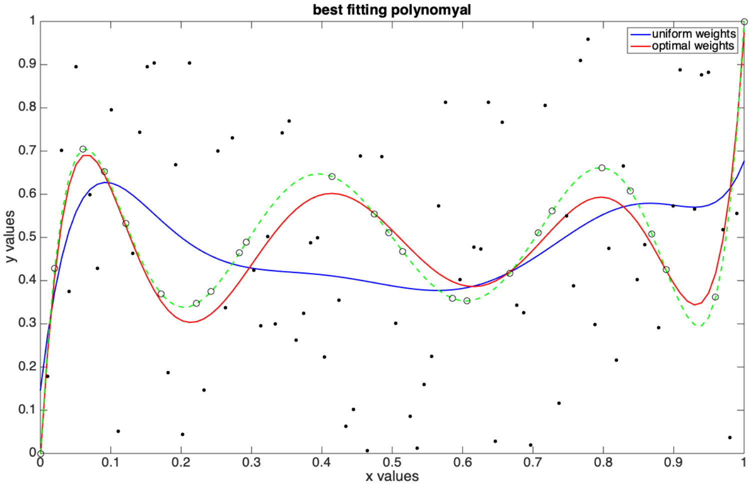
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 5 = 1.16 \cdot 10^{-2}$
- Entropy $= 4.23$

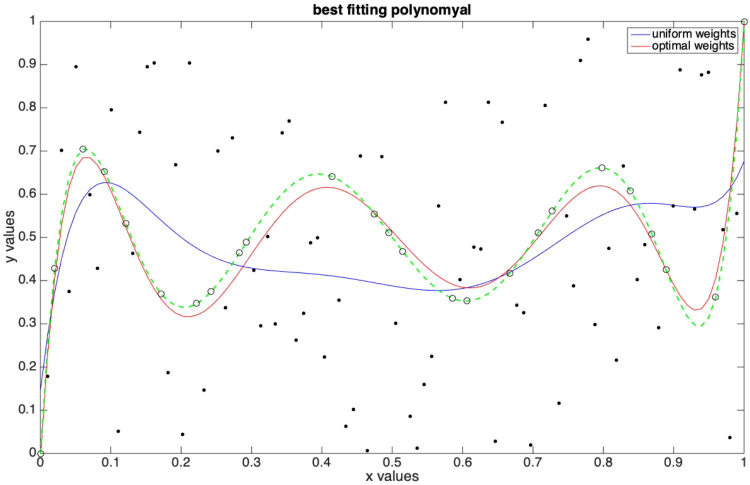
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 10 = 5.81 \cdot 10^{-3}$
- Entropy $= 4.04$

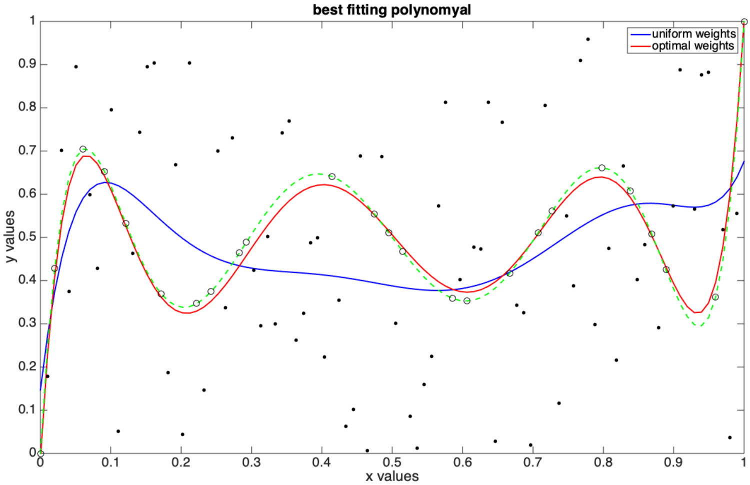
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 20 = 2.90 \cdot 10^{-3}$
- Entropy $= 3.86$

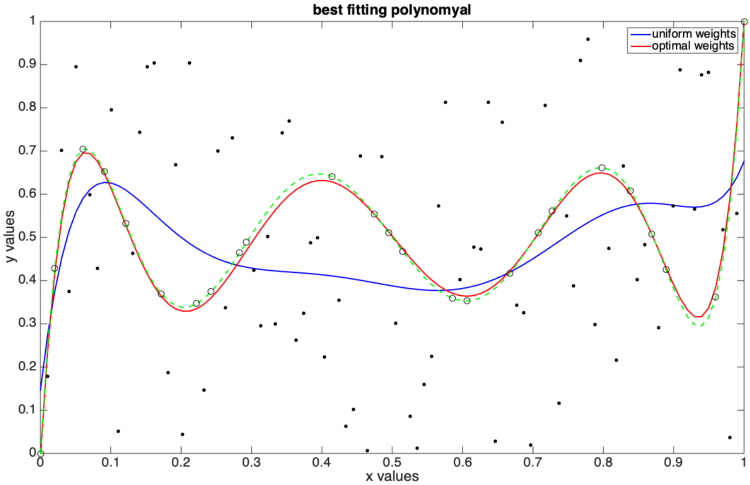
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 50 = 1.16 \cdot 10^{-3}$
- Entropy $= 3.67$

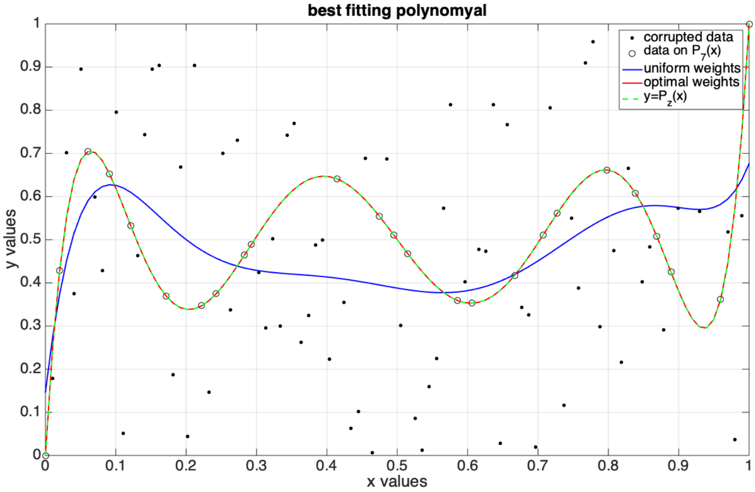
The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 100 = 5.80 \cdot 10^{-4}$
- Entropy $= 3.55$

The hidden polynomial problem solved by the MEWLS



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 1e6 = 5.81 \cdot 10^{-8}$
- Entropy $= 3.22 \simeq -\log(1/25)$

Solution of the MEWLS problem

Assume $y_i \in \mathbb{R}$ for simplicity. We build the Lagrangian...

$$\mathcal{L}(w, c, \alpha, \beta) = \sum_{i=1}^{m} w_i \log w_i + \alpha \left(\sum_{i=1}^{m} w_i - 1 \right) + \beta \left(\sum_{i=1}^{m} w_i (f(x_i, c) - y_i)^2 - \overline{\mathrm{E}^2} \right)$$

Solution of the MEWLS problem

Assume $y_i \in \mathbb{R}$ for simplicity. We build the Lagrangian...

$$\mathcal{L}(w, c, \alpha, \beta) = \sum_{i=1}^{m} w_i \log w_i + \alpha \left(\sum_{i=1}^{m} w_i - 1 \right) + \beta \left(\sum_{i=1}^{m} w_i (f(x_i, c) - y_i)^2 - \overline{\mathsf{E}^2} \right)$$

...and compute the stationary points

$$\frac{\partial \mathcal{L}}{\partial w_k} = 1 + \log w_k + \alpha + \beta (f(x_k, c) - y_k)^2 = 0, \qquad k = 1, \ldots, m$$

$$\frac{\partial \mathcal{L}}{\partial c_k} = 2 \sum_{i=1}^{n} w_i (f(x_i, c) - y_i) \frac{\partial f(x_i, c)}{\partial c_k} = 0, \qquad k = 1, \ldots, s$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{m} w_i - 1 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{m} w_i (f(x_i, c) - y_i)^2 - \overline{\mathsf{E}^2} = 0.$$

MWELS: the algebraic system to be solved

$$(a) \quad \sum_{i=1}^{m} w_i (f(x_i, c) - y_i) \frac{\partial f(x_i, c)}{\partial c_k} = 0, \qquad k = 1, \ldots, s$$

$$(b) \quad \sum_{i=1}^{m} (f(x_i, c) - y_i)^2 \exp(-\beta (f(x_i, c) - y_i)^2) = \sum_{i=1}^{m} \exp(-\beta (f(x_i, c) - y_i)^2) \cdot \overline{\mathrm{E}^2}$$

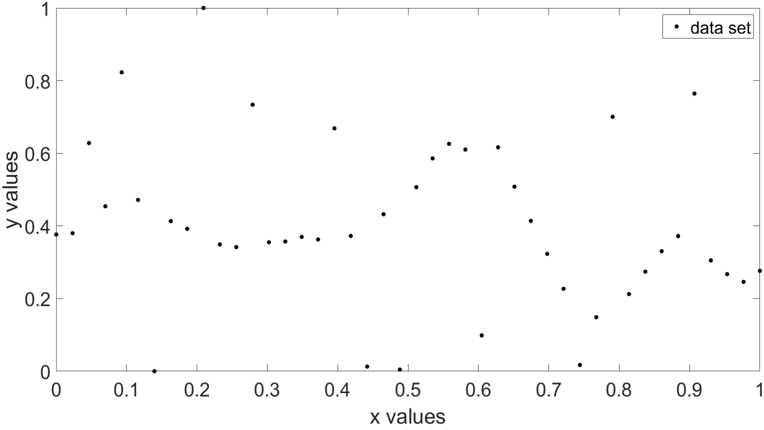
$$(c) \quad w_k = \frac{1}{\sum_{i=1}^{m} \exp(-\beta (f(x_i, c) - y_i)^2)} \cdot \exp(-\beta (f(x_i, c) - y_i)^2), \qquad k = 1, \ldots, m$$

ALGORITHM

1. Assign
   - an initial weight distribution (likely the uniform-weight distribution)
   - the desired minimal mean-squared error $\overline{\mathrm{E}^2}$
2. Evaluate the coefficients $c = (c_1, \ldots, c_s)$ from (a)
3. Update the Lagrangian multiplier β from (b)
4. Update the weights w_k from (c)
5. Repeat from step 2. until convergence is achieved.

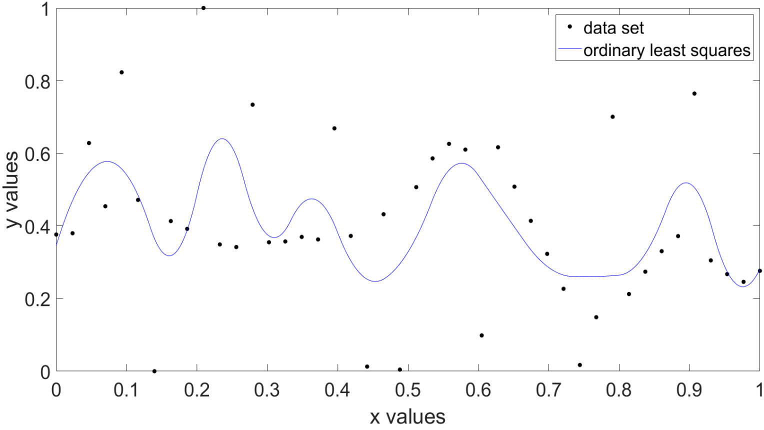
ILLUSTRATIONS

Spline function



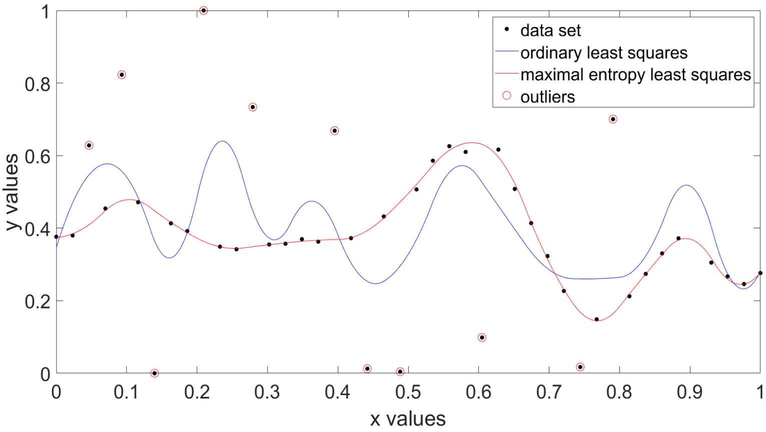
We consider a dataset consisting of 44 points, 32 of which closely follow a given profile while the remaining 12 consistently depart from it.
To fit the data we use a spline of degree $d = 2$ defined on a regular and uniform knot sequence composed by 20 nodes covering the interval $[0, 1]$.

Spline function



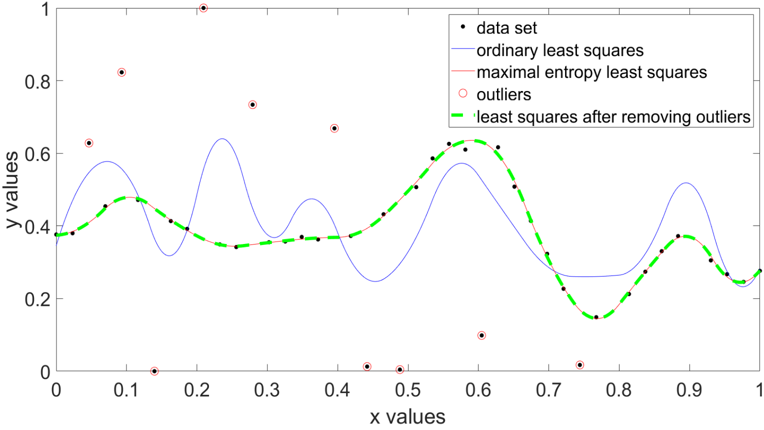
We see that the ordinary least squares approximation (blue line) is strongly influenced by the 12 anomalous data and, we want to improve it by reducing the mean squared error and exploiting the maximal-entropy argument to properly choose the weights.

Spline function



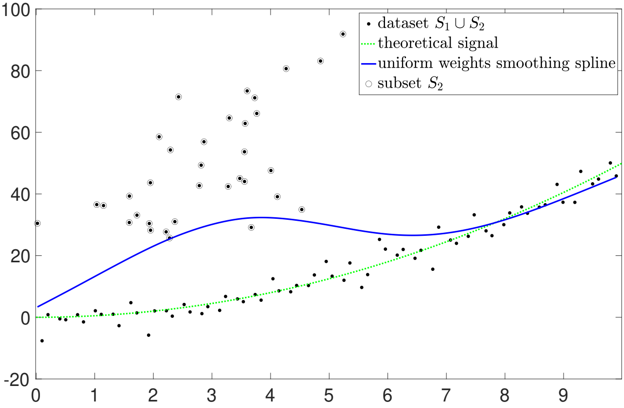
- Blue line: ordinary least squares approximation.
- Red line: maximal-entropy least squares approximation.
- Red circles: the outliers are detected by looking at the weights approaching zero.

Spline function



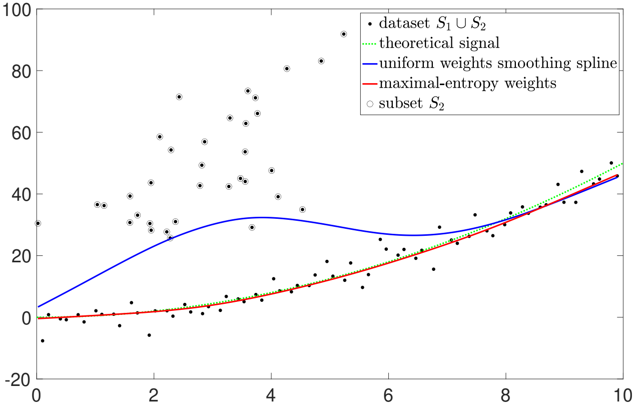
- Red line: maximal-entropy least squares approximation.
- Red circles: The outliers are detected by looking at the weights approaching zero.
- green dashed line: ordinary least squares approximation after removing outliers.

Smoothing spline



	$\overline{\mathrm{E}^2}_{\mathrm{uw}}$	λ
S_1	$2.89 \cdot 10^{-3}$	$1.78 \cdot 10^{-4}$
$S_1 \cup S_2$	$4.23 \cdot 10^{-2}$	$1.03 \cdot 10^{-4}$

Smoothing spline

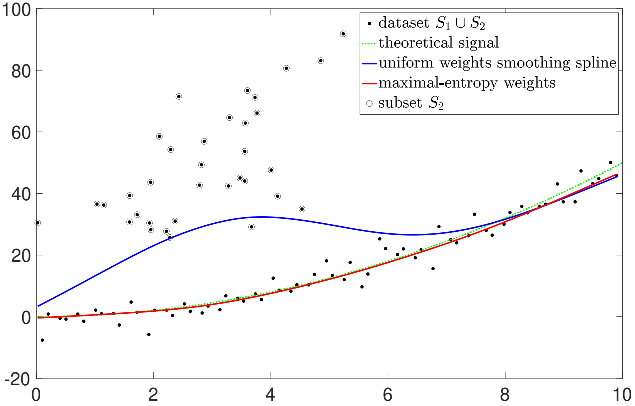


maximize $\quad -\sum_{i=1}^{m} w_i \log w_i, \qquad$ (entropy)

subject to: $\quad \sum_{i=1}^{m} w_i = 1, \qquad$ (normalization condition)

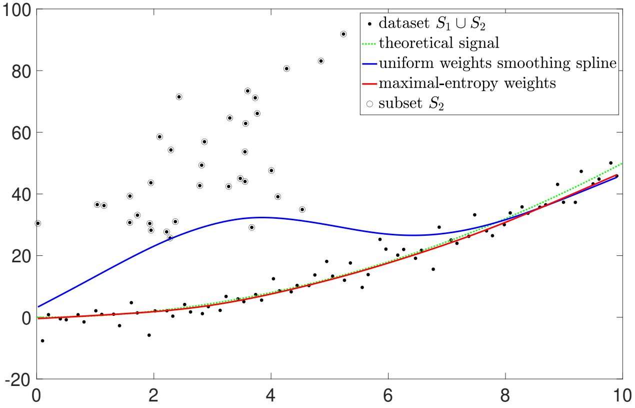
$$\sum_{i=1}^{m} w_i (f(c_1, \ldots, c_s, t_i) - y_i)^2 + \lambda \int_a^b (g''(t))^2 \mathrm{d}t = 2.89 \cdot 10^{-3}$$

Smoothing spline



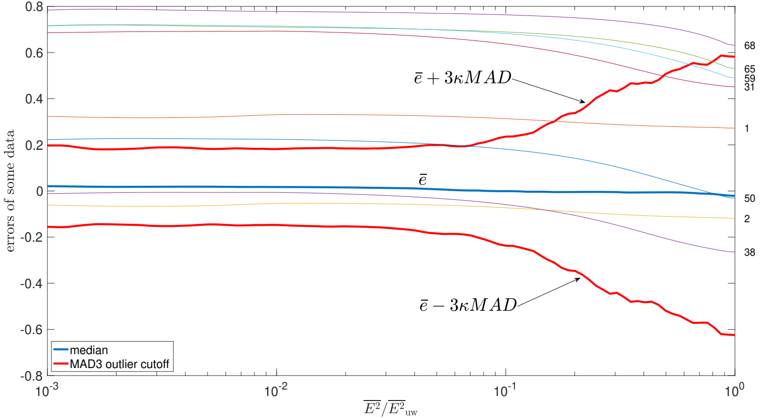
- The working assumption here is that the expected value of $\overline{E^2}$ is already determined within the context. BUT...

Smoothing spline



- The working assumption here is that the expected value of $\overline{E^2}$ is already determined within the context. BUT...
- What if we lack any prior information regarding the presence of outliers in the dataset?

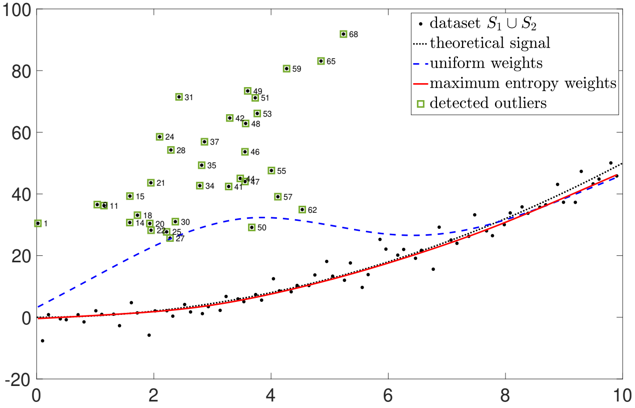
Smoothing spline



- $e_i = y_i - \hat{y}_i$: residuals ($\hat{y}_i$ is the value predicted by the smoothing spline at t_i)
- $\mathrm{MAD} = \mathrm{median}(|e_i - \bar{e}|)$: Median Absolute Deviation

$$(t_i, y_i) \text{ is an outlier} \quad \Longleftrightarrow \quad |e_i - \bar{e}| > 3\kappa \cdot \mathrm{MAD},$$

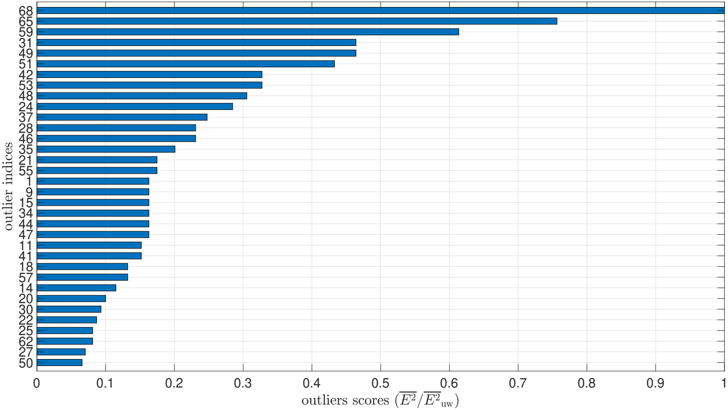
Smoothing spline



- $e_i = y_i - \hat{y}_i$: residuals ($\hat{y}_i$ is the value predicted by the smoothing spline at t_i)
- $\mathrm{MAD} = \mathrm{median}(|e_i - \bar{e}|)$: Median Absolute Deviation

$$(t_i, y_i) \text{ is an outlier} \quad \Longleftrightarrow \quad |e_i - \bar{e}| > 3\kappa \cdot \mathrm{MAD},$$

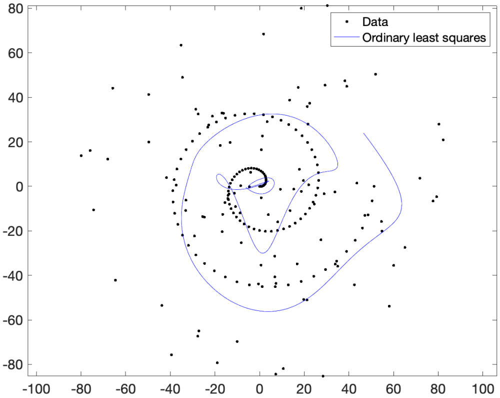
Smoothing spline



- $e_i = y_i - \hat{y}_i$: residuals ($\hat{y}_i$ is the value predicted by the smoothing spline at t_i)
- $\text{MAD} = \text{median}(|e_i - \bar{e}|)$: Median Absolute Deviation

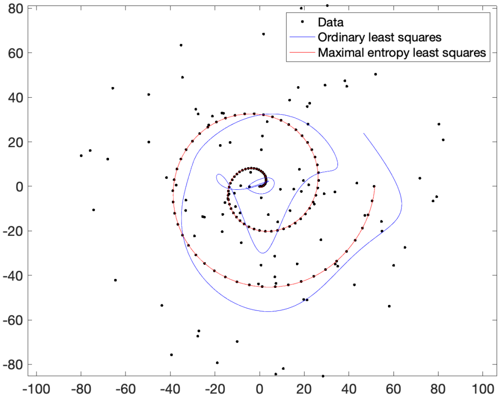
$$(t_i, y_i) \text{ is an outlier} \quad \Longleftrightarrow \quad |e_i - \bar{e}| > 3\kappa \cdot \text{MAD},$$

Archimedean spiral



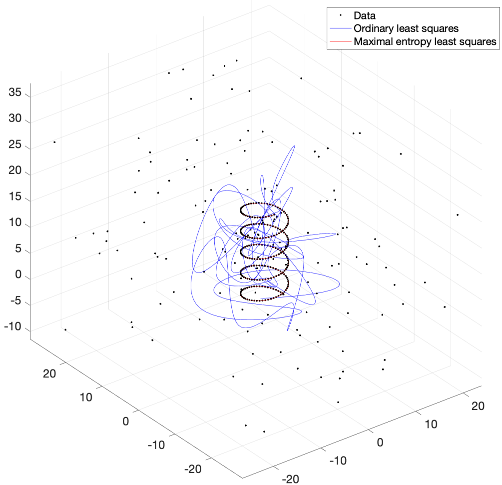
Data set: we define 200 points on the arithmetic spiral $y(t) = (1 + 4t)e^{it}$, $t \in [0, 4\pi]$ and introduce random noise over 100 of them (odd ones).

Archimedean spiral



- Blue line: ordinary least squares approximation.
- Red line: maximal-entropy least squares approximation.

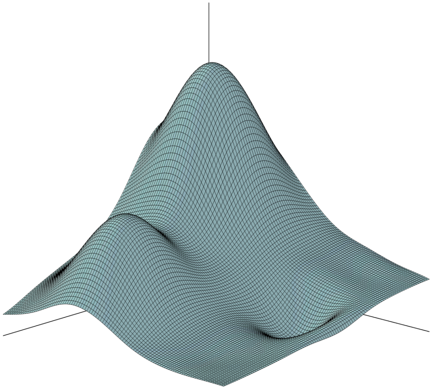
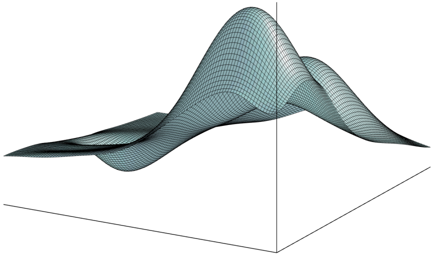
Helix



Data set: we start with 500 points on the helix $4[t,\ \cos(2\pi t),\ \sin(2\pi t)]$, and introduce random noise in 250 points (odd ones).

- Blue line: ordinary least squares approximation.
- Red line: maximal-entropy least squares approximation.

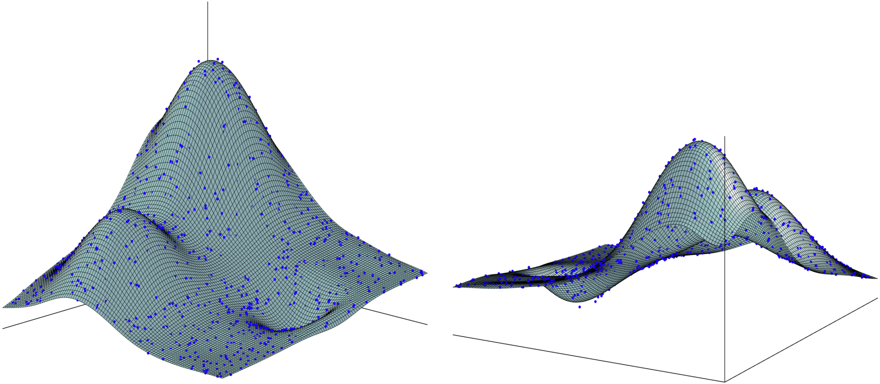
Surface approximation from a noisy point cloud

Franke's function has two Gaussian peaks of different heights, and a smaller dip, making it a common test function for interpolation problems.

- R. Franke, A critical comparison of some methods for interpolation of scattered data, Naval Postgraduate School, Tech. Rep. NPS-53-79-003 (1979).

Surface approximation from a noisy point cloud

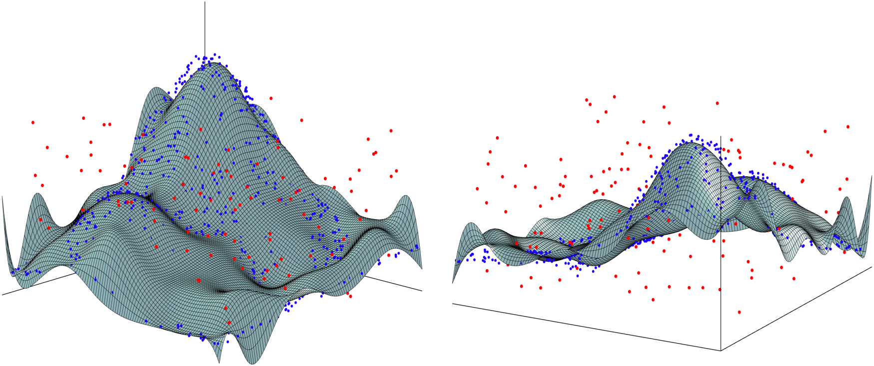


We have randomly sampled 10^3 dots form the function and perturbed their values with Gaussian noise with zero mean and variance $\sigma^2 = 10^{-3}$.

A cubic B-spline surface, defined on a 10×10 grid, has been used to reconstruct the surface. The knots are selected to ensure interpolation at the four corners.

In the absence of outliers, the reconstruction performs well!

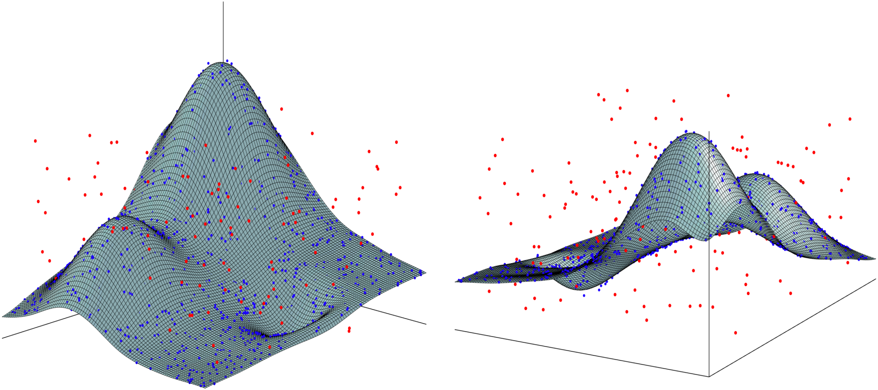
Surface approximation from a noisy point cloud



We have added 150 additional points, highlighted in red, randomly sampled from a uniform distribution to introduce outliers.

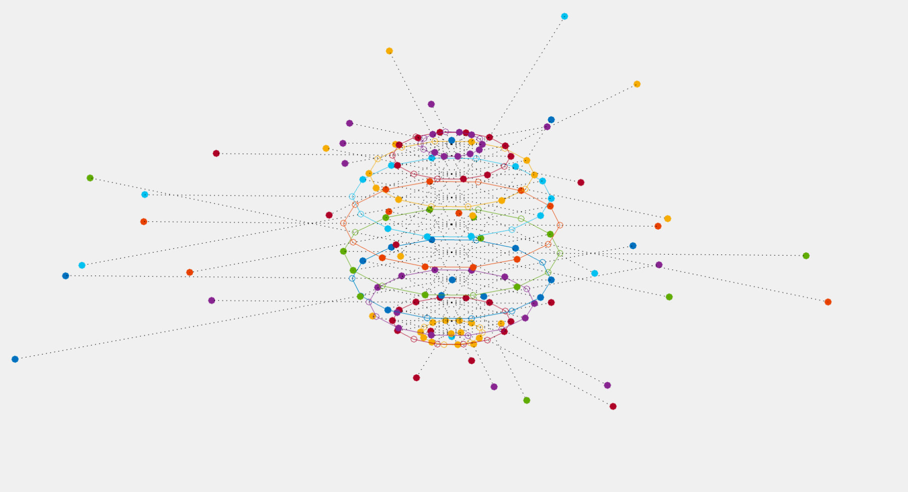
We see that the ordinary least squares approximation (OLS) is heavily affected by the presence of these anomalous data, leading to a poor reconstruction of Franke's function.

Surface approximation from a noisy point cloud



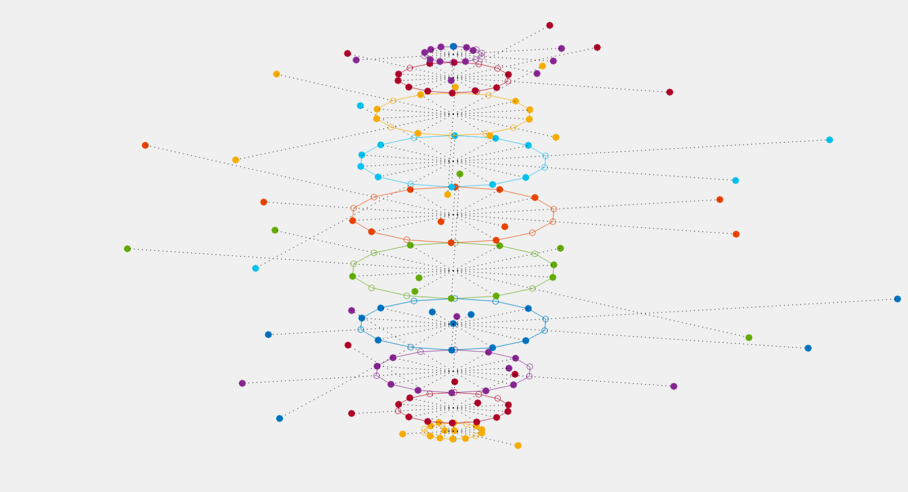
Exploiting the maximum entropy approach, we observe that a reduction of the $\overline{\mathsf{E}^2}_{\mathrm{uw}}$ of a factor 100 is enough to assign outliers a negligible weight and restore the original shape of the surface.

Sphere reconstruction



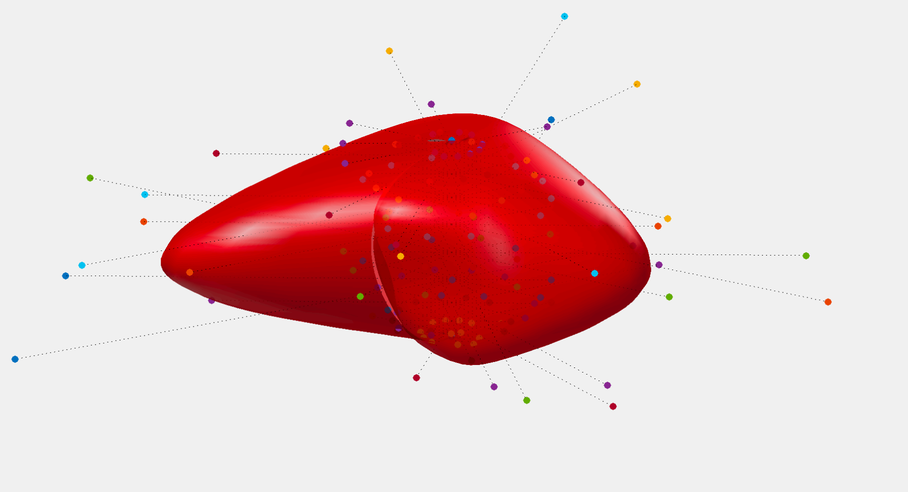
- 142 points lying on a sphere but half of them are perturbed.

Sphere reconstruction



To see how the experiment has been prepared, we stretch the sphere along the vertical axis for better visual clarity.

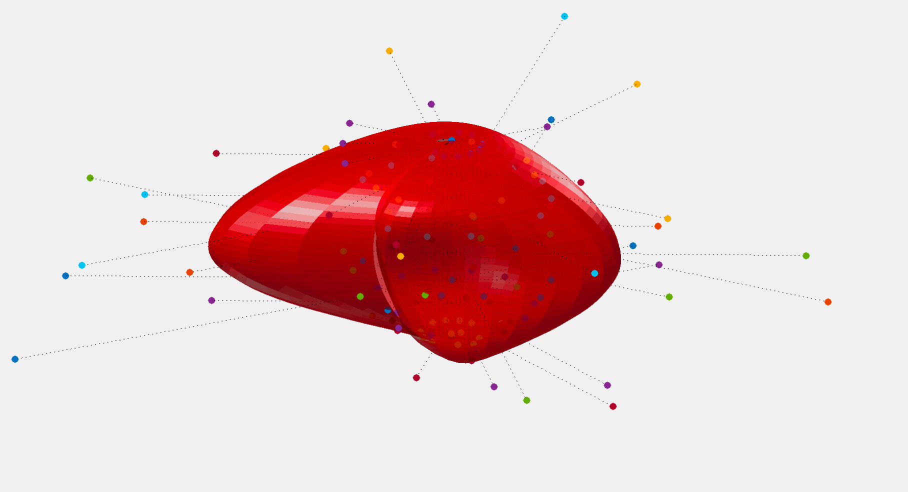
Sphere reconstruction



Ordinary Least Squares

- $\overline{E^2}_{\mathrm{uw}} = 0.465$

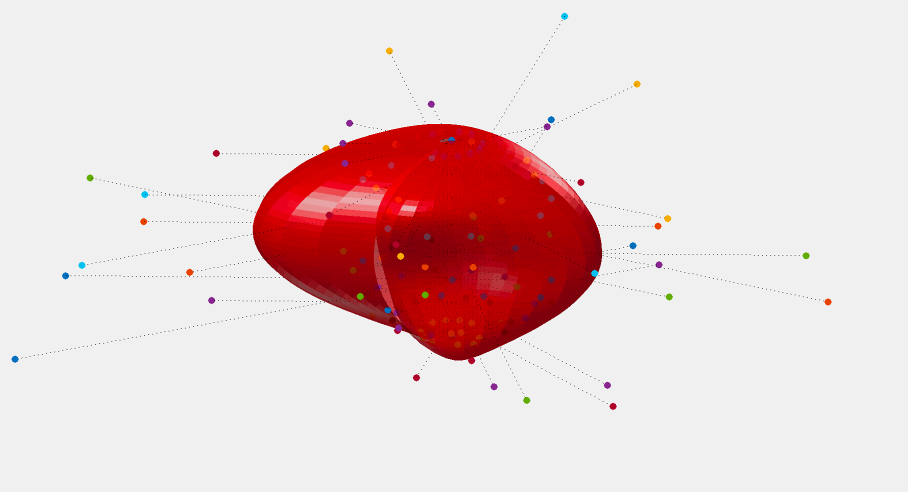
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}}/2$

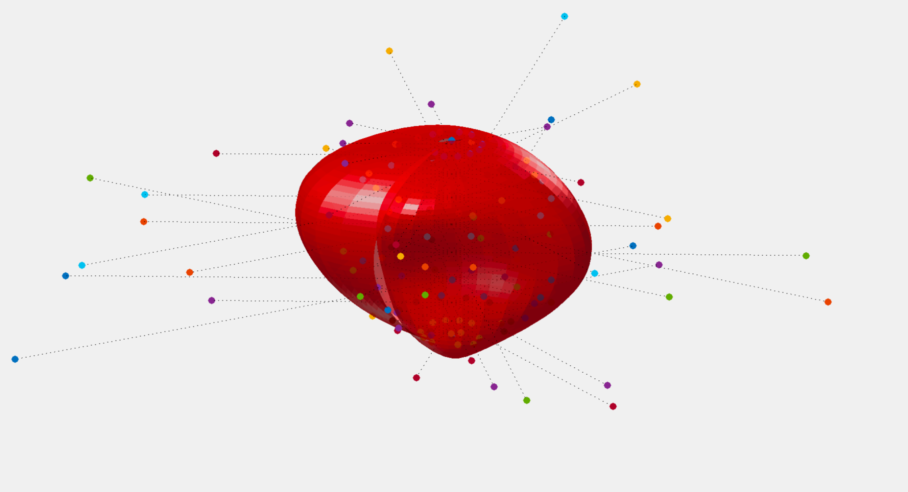
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 3$

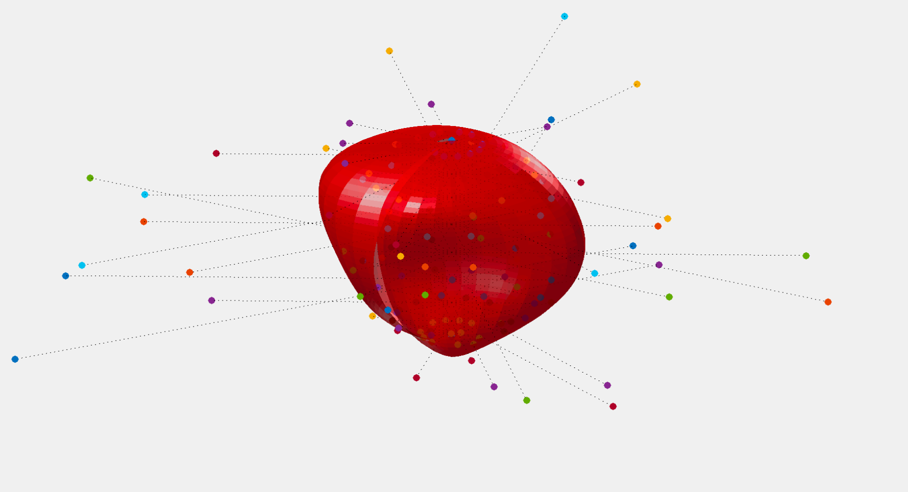
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}}/4$

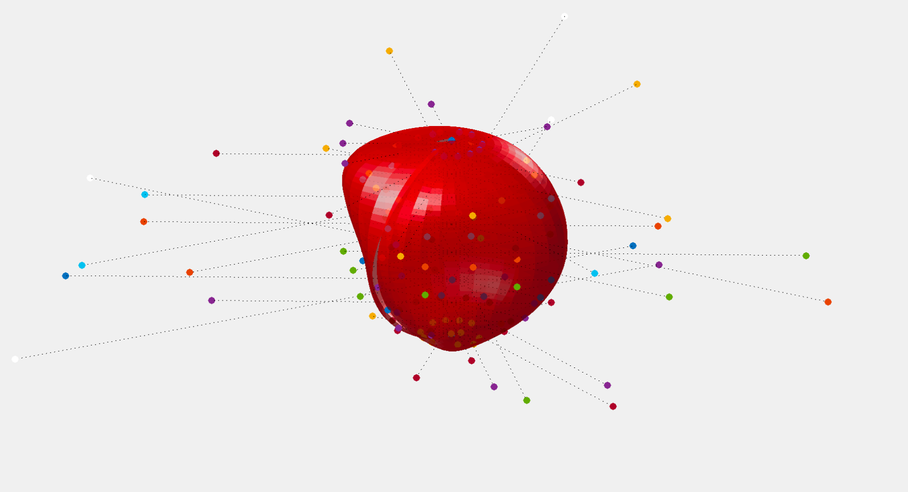
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 5$

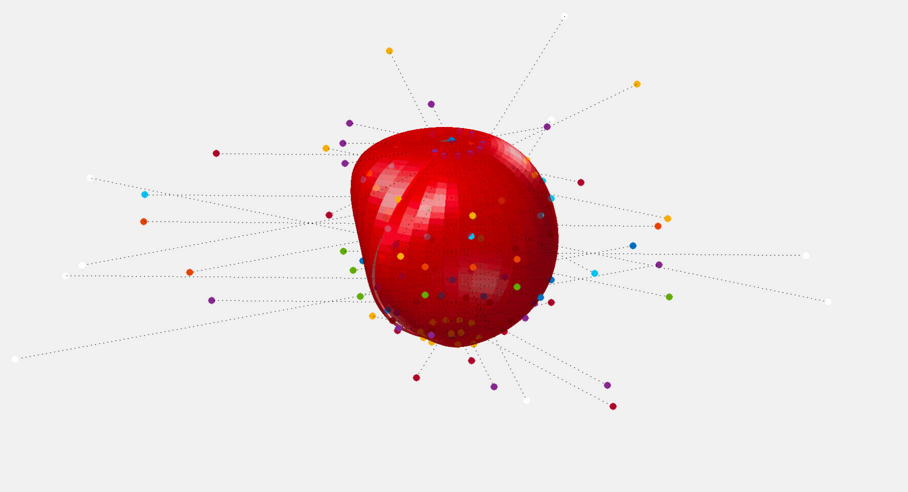
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 10$

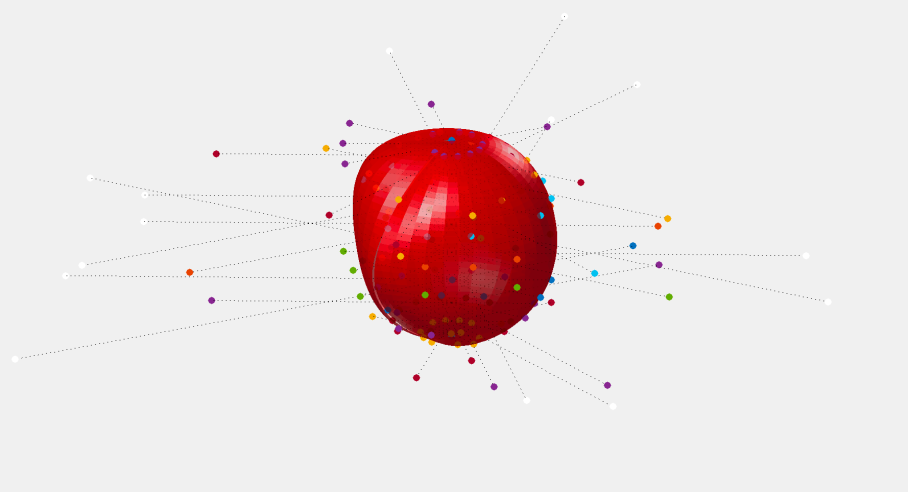
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 20$

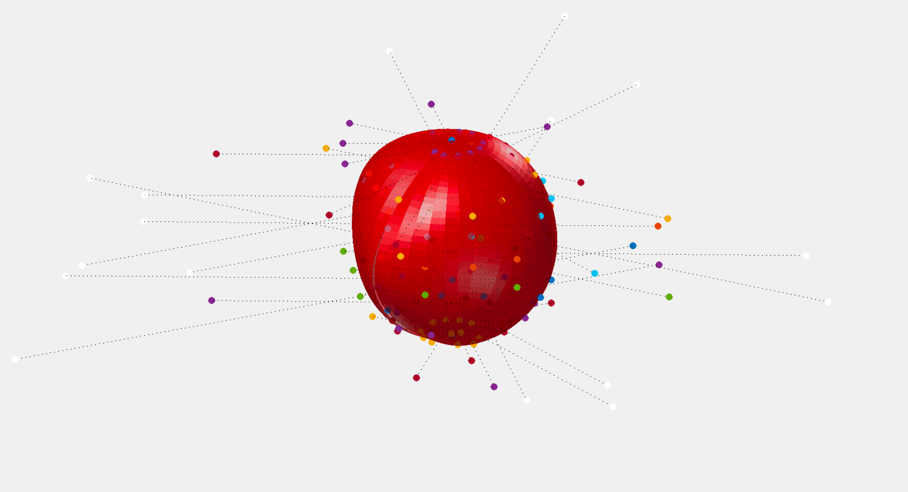
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathsf{E}^2} = \overline{\mathsf{E}^2}_{\mathrm{uw}} / 30$

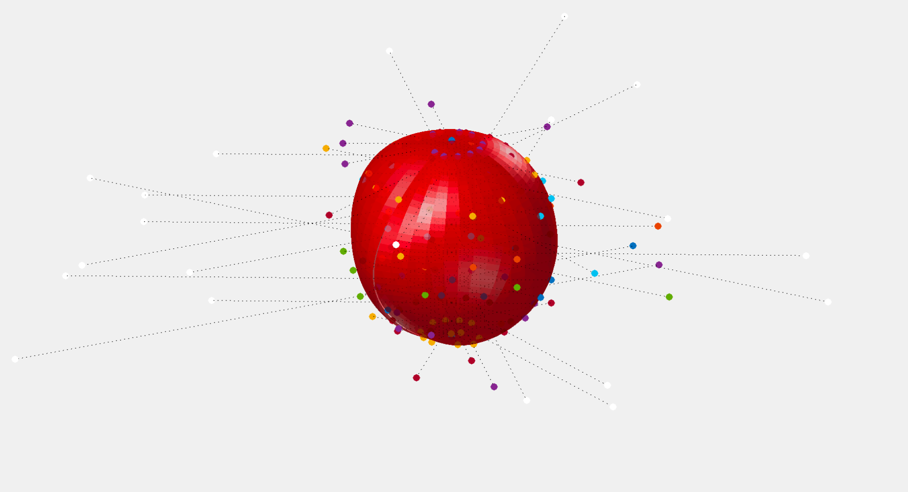
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{E^2} = \overline{E^2}_{uw} / 40$

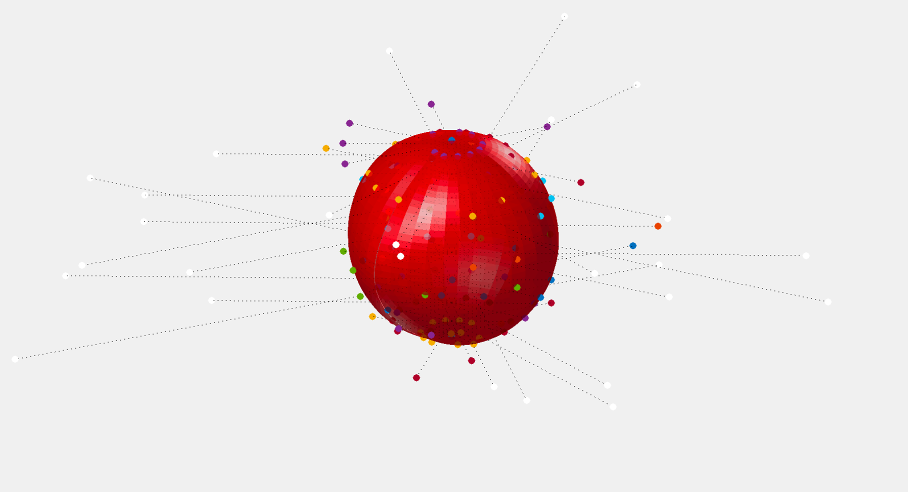
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 50$

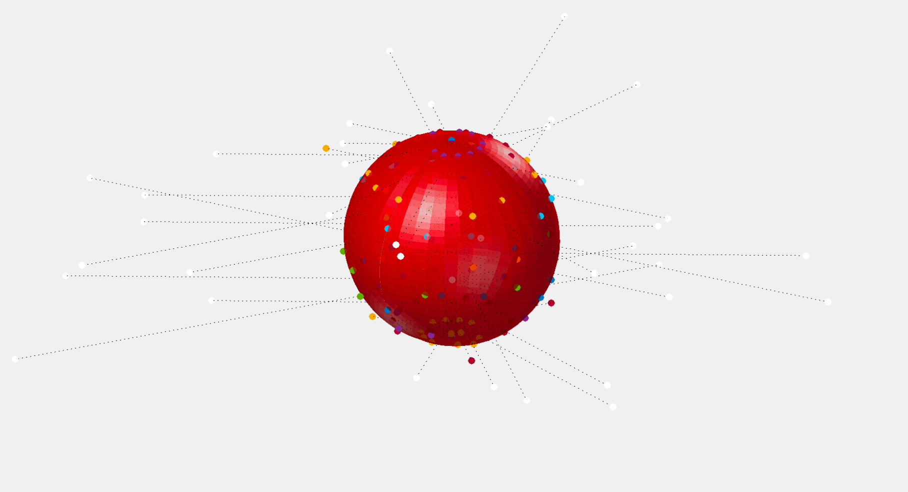
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 100$

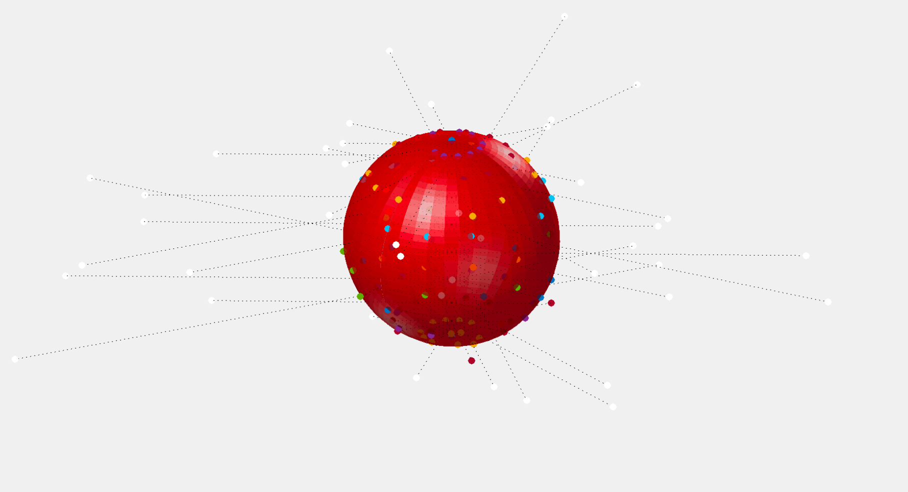
Sphere reconstruction



Maximal-entropy Least Squares

- $\overline{\mathsf{E}^2} = \overline{\mathsf{E}^2}_{\mathrm{uw}} / 500$

Sphere reconstruction

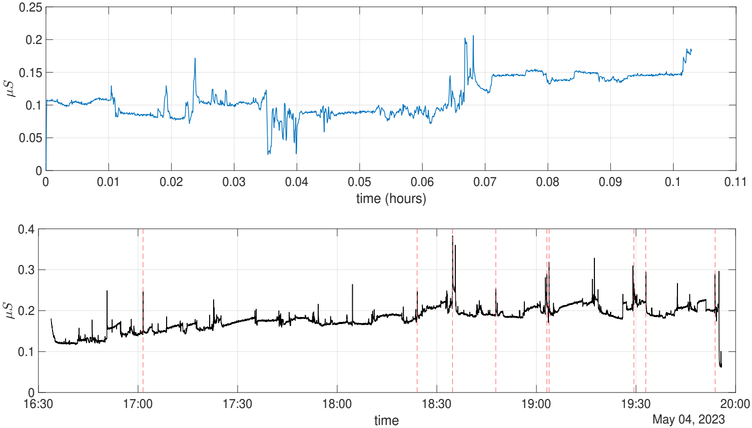


Maximal-entropy Least Squares

- $\overline{\mathrm{E}^2} = \overline{\mathrm{E}^2}_{\mathrm{uw}} / 1000$

SOME APPLICATIONS

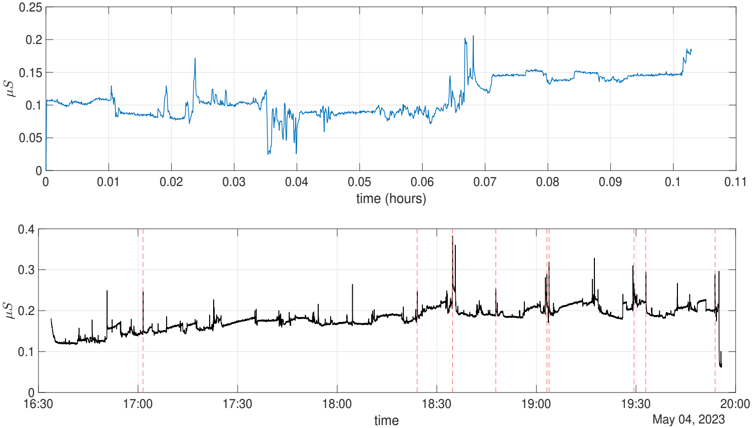
Peaks detection in electrodermal activity data



PROBLEM: identify peaks within a signal representing electrodermal activity (EDA) data of an individual.

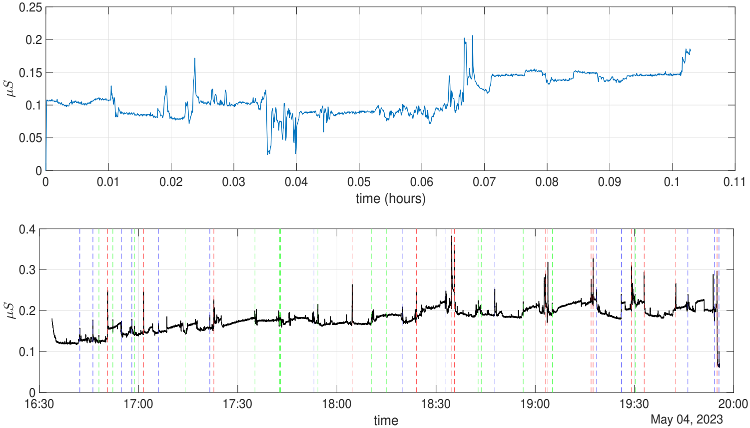
The location and amplitude of these peaks are correlated with the release of cortisol hormone, which occurs during stressful emotional activities.

Peaks detection in electrodermal activity data



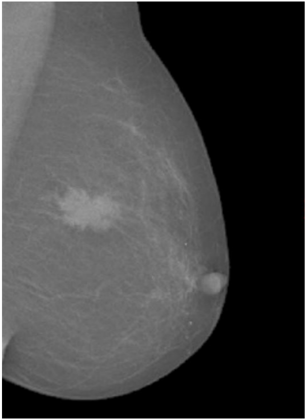
- TOP PICTURE: a baseline signal obtained through a six-minute relaxation session, which is assumed to be free of significant peaks, indicating a lack of stress.
- BOTTOM PICTURE: time series of conductance measurements (in microsiemens, μS) collected from an individual during daily activities from 4:30 PM to 8:00 PM. Peaks located by the MIT algorithm are highlighted as dashed vertical lines.

Peaks detection in electrodermal activity data



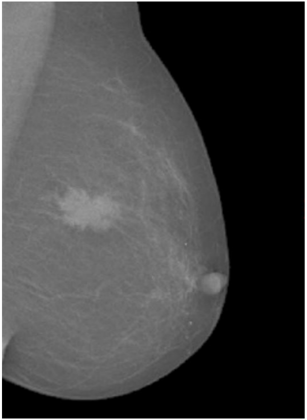
- A smoothing spline approximation with uniform weights applied to the baseline signal reveals a mean squared error of $\overline{E^2}_{uw} = 2.56 \times 10^{-5}$.
- In contrast, the mean squared error for the uniform-weight smoothing spline applied to the EDA signal is $\overline{E^2}_{uw} = 3.60 \times 10^{-4}$. The colors of the vertical bars, red, blue, and green) indicate stress severity levels in decreasing order.

Detecting masses in mammograms

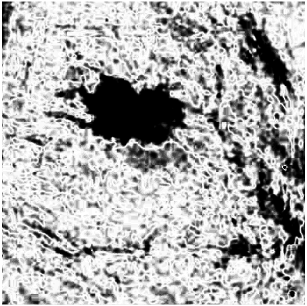


As a proof of concept, we have used MEWLS as a computer-aided diagnosis tool to identify Regions of Interest and assist radiologists in making accurate diagnoses of breast lesions.

Detecting masses in mammograms



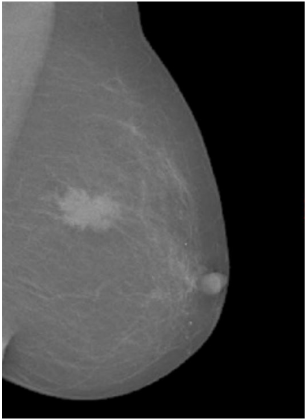
weight distribution

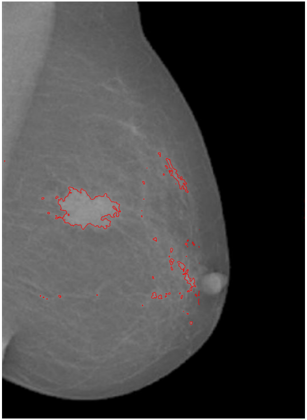


The original $\overline{E^2}_{uw}$ (uniform weights) has been reduced of a factor 10 to obtain the weight distribution shown in the picture

- black –> lowest weights
- white –> highest weights

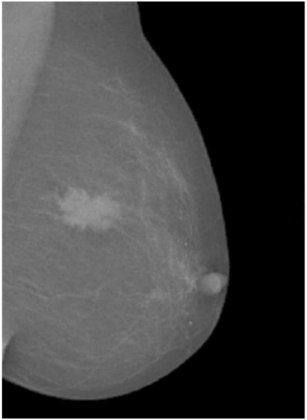
Detecting masses in mammograms

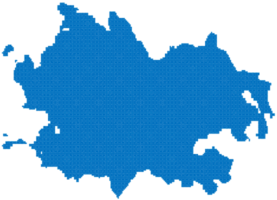




Superimposing the contour lines on the original image reveals the Regions of Interest (ROIs), highlighted as red lines in the picture.

Detecting masses in mammograms





weights close to zero

We can extract the portion of the weight distribution corresponding to the mass of interest: the Hausdorff dimension of its contour is $= 1.59$. This raises suspicion that the mass may be malignant.

- R.M. Rangayyan, T.M. Nguyen, Fractal Analysis of Contours of Breast Masses in Mammograms, Journal of Digital Imaging 20(3) 2007, 223–237.

Removing craquelure from an oil painting

Craquelure is a fine network of cracks that appear on the surface of oil paintings, particularly those on panel. It occurs naturally over time due to the aging of materials and environmental conditions.

Removing craquelure from an oil painting

In oil paintings on wooden panels, craquelure is often the result of the different rates of expansion and contraction between the layers of paint, varnish and the wooden support, as these materials respond to temperature and humidity changes in a different manner.

Removing craquelure from an oil painting

PROBLEM: Try to detect and reduce the visual impact of cracks by minimizing pixel-level interventions.

Removing craquelure from an oil painting

- We use a bivariate spline and exploit the MEWLS approximation to detect most of the pixel forming the pattern of cracks.
- Then we replace the original RGB values of the detected pixel by the ones predicted by the model.

Removing craquelure from an oil painting

ORIGINAL IMAGE

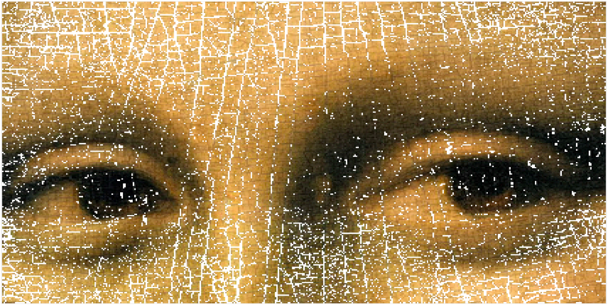
total pixels: 304725

Removing craquelure from an oil painting

ORIGINAL IMAGE

total pixels: 304725

DETECTED OUTLIERS: 13951 pixels



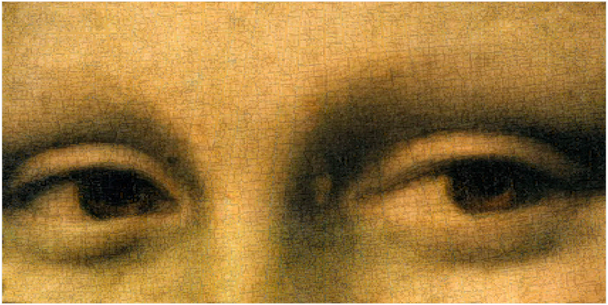
Removing craquelure from an oil painting

ORIGINAL IMAGE

total pixels: 304725

RESTORED IMAGE

(ONLY) 13.7% pixels replaced



Conclusions and Future Work

The following extensions will be the focus of future research:

- Nonlinear approximating functions
- Machine learning and deep learning models
- Stochastic differential equations with non-Gaussian noise

REFERENCES

- D. Giordano, F. Iavernaro, Maximal-entropy driven determination of weights in least-square approximation, *Math Meth Appl Sci.* **44** (2021), 6448–6461.

- L. Brugnano, D. Giordano, F. Iavernaro, G. Rubino, An entropy-based approach for robust least squares spline approximation, *J. Comput. Appl. Math.* **443**, 115773 (2024).

- M. De Giosa, A. Falini, F. Iavernaro, S. Losito, F. Mazzia, G. Rubino, A robust variant of cubic smoothing spline approximation, *Numer. Algor.*, (2025) `https://doi.org/10.1007/s11075-024-02003-7`

- L. Brugnano, F. Iavernaro, E.B. Weinmüller, Weighted least squares collocation methods, *Appl. Numer. Math.* **203**, 113128 (2024)

- P. Amodio et al., An Entropy-based Spline Approximation Technique for the Automatic Detection of Peaks in Electrodermal Activity Data, *AIP Conf. Proc.* (to appear)

- P. Amodio, L. Brugnano, F. Iavernaro, On the use of the principle of maximum entropy in bivariate splines least-squares approximation, (submitted)

THANK YOU FOR YOUR ATTENTION

P.S. Not every outlier is outlier :-)