Inverse problems

Regularization: Example Lecture 4

Nikolai Piskunov

2014

Now that we know the theory, let's try an application:

- Problem: Optimal filtering of 1D and 2D data
- Solution: formulate an inverse problem
- Refinement: add regularization
- See what happens

Problem

- In 1D we have a bunch of measured points on some grid
- We simplicity, let's assume an equispaced grid
- Thus we have a vector of measurements:

$$g_i \approx f_i$$

with associated uncertainties $\sigma_i \;\; (\omega_i = 1/\sigma_i^2)$

- Let's formulate an inverse problem:

$$\sum_i \omega_i \left[f_i - g_i\right]^2 = \min$$

Analysis

- This looks too trivial, but we know the solution may not be unique since we have measurement errors
- Let's look for the smoothest solution that still matches the error bars:

$$\Omega(f) \equiv \sum_{i=1,n} \omega_i \left[f_i - g_i\right]^2 + \Lambda \sum_{i=1,n-1} \left(f_{i+1} - f_i\right)^2$$

- We actually use Tikhonov regularization to impose extra constraint

Formal solution

- Now the solution becomes non-trivial
- Let's take the derivatives and set them to zero:

$$\frac{d\Omega}{df_i} = 0 = \begin{cases} \omega_i\,(f_i - g_i) + \Lambda(f_i - f_{i+1}) & \mathrm{i} = 1 \\ \omega_i\,(f_i - g_i) + \Lambda(2f_i - f_{i-1} + f_{i+1}) & 1 < \mathrm{i} < \mathrm{n} \\ \omega_i\,(f_i - g_i) + \Lambda(f_i - f_{i-1}) & \mathrm{i} = \mathrm{n} \end{cases}$$

- This is a system of n linear equations
- We can rewrite it in a more familiar form:

$$\begin{aligned} \omega_1 f_1 &\quad + \quad \Lambda(f_1 - f_2) &= \quad \omega_1 g_1 \\ \omega_i f_i &\quad + \quad \Lambda(2f_i - f_{i-1} - f_{i+1}) &= \quad \omega_i g_i \\ \omega_n f_n &\quad + \quad \Lambda(f_n - f_{n-1}) &= \quad \omega_n g_n \end{aligned}$$

Formal solution (cont'd)

- The matrix is tri-diagonal; other elements are 0
- Off-main-diagonal elements are all $-\Lambda$
- Main diagonal contains $\omega_i + \Lambda$ for the first and the last equation and $\omega_i + 2\Lambda$ elsewhere
- This is easy to solve – no iterations needed
- Uncertainty-based weights adjust the balance between measurements and regularization: precise measurement has higher weight and dominates local gradient control

Generalization

- What happens if we would like to apply higher order regularization?
- E.g. if we want to constrain also the amplitude and the 2nd derivative?

$$\Omega(f) \equiv \sum_{i=1,n} \omega_i \left[f_i - g_i\right]^2 +$$

$$+\Lambda_0 \sum_i f_i^2 + \Lambda_1 \sum_{i=1,n-1} (f_{i+1} - f_i)^2 +$$

$$+\Lambda_2 \sum_{i=2,n-1} (f_{i+1} + f_{i-1} - 2f_i)^2 = \min$$

- The matrix now has 5 non-zero diagonals!

Adding dimensions

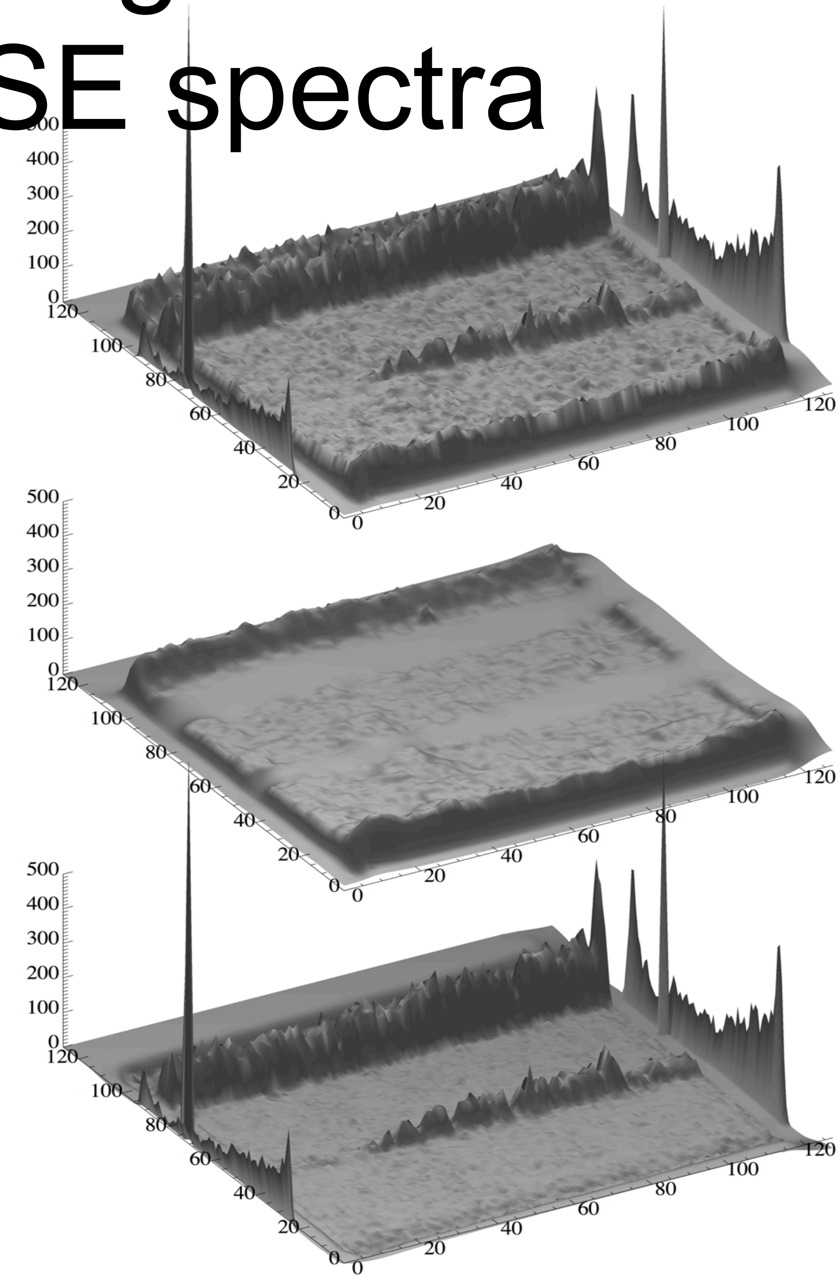
- What happens in 2D?

$$\begin{aligned}\Omega(f) &\equiv \textstyle\sum_{i,j} \omega_{i,j} \left[f_{i,j} - g_{i,j}\right]^2 + \\ &+ \Lambda_1 \textstyle\sum (f_{i+1,j} - f_{i,j})^2 + \\ &+ \Lambda_2 \textstyle\sum (f_{i,j+1} - f_{i,j})^2 = \min\end{aligned}$$

- Nothing really, just more book keeping
- If we decide to number grid points with i running faster ($0<i<n_i+1$) then sub-diagonals immediately below and above will contain Λ_1
- There will be two more sub-diagonals n_i with Λ_2 elements below and above the main diagonal

Subtracting background from the 2D FUSE spectra

- Observation
- Background model based on optimal filtering
- Clean spectrum



Questions

- What will happen if we add 2nd derivatives in 2D case?

- What happens in places where the weights are zero?

- Can you program this and try?

Relation to Singular Value Decomposition

- In SVD the matrix of a linear problem is represented by a product of three matrices.

$$\mathbf{A}f = \mathbf{U}\mathbf{B}\mathbf{V}^T f = g$$

- The middle one is diagonal and consists of eigenvalues of $\mathbf{A}$
- The SVD solution is found as:

$$f = \mathbf{V}\mathbf{D}\mathbf{U}^T g$$

where $\mathbf{D}$ is the inverse of $\mathbf{B}$ with all zero eigenvalues replaced by a constant

Relation to Singular Value Decomposition

- One have shown that replacing the original problem with a regularized one (Tikhonov):

$$\mathbf{A}f = g \rightarrow \mathbf{A}f + \Lambda\mathbf{R}f = g$$

results in modifying the diagonal elements of $\mathbf{D}$:

$$\frac{1}{\sigma_i} \rightarrow \frac{\sigma_i}{\sigma_i^2 + \Lambda^2}$$

- Thus the inversion of $\mathbf{B}$ will not lead to division by 0 and the problem will have a unique solution

Relation to optimal (Wiener) filtering

- The formulation for the optimal filtering is somewhat different. We assume that we measure a signal $g=f+w$ where w is a normally distributed noise. Noise is additive and uncorrelated with the signal!
- The problem is to find a filter $\mathbf{F}$ that damps the noise in the most probable sense. The expectation estimate is:

$$\mathbf{E}\|f - \mathbf{F}(f+w)\| = \max$$

Relation to optimal (Wiener) filtering

- The optimal filter function $\mathbf{F}$ is what we try to construct
- For discrete sampling the filter function can be represented as a matrix:

$$\mathbf{F}(g)\big|_j = \sum_i p_i u_j^i g_i$$

where matrix u is constructed based on the noise spectrum while p are the so-called Wiener factors. Using Tikhonov regularization sets them to:

$$p_i = \frac{\sigma_i^2}{\sigma_i^2 + \Lambda^2}$$

where sigma's are eigenvalues of matrix u making solution unique.

Next step

- We will look at more advanced optimization techniques

- Then we will start looking into very relevant astrophysical examples