Inverse problems

Finding regularized solution
Lecture 5

Nikolai Piskunov
2014

Finding minimum

Conjugate gradient search. In a simple gradient search we follow the direction of the local gradient on each iteration ignoring the information from previous steps:

$$\vec{f}^{n+1} = \vec{f}^{n} - \Delta_n \cdot \vec{\nabla}\Omega(\vec{f})$$

The algorithm is not very efficient as the linear approximation is no good for the search in multiple dimensions. One can notice that the consecutive steps are going in the right direction, so combining consecutive gradients may speed up convergence.

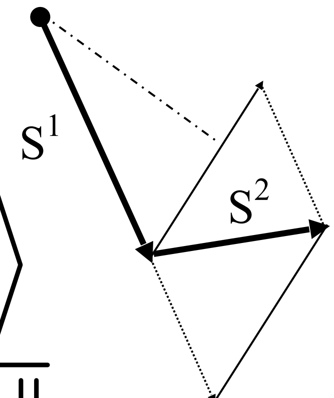
Instead of searching along the gradient, we will construct the search direction step-by-step:

$$\vec{f}^{n+1} = \vec{f}^{n} - \Delta_n \cdot \vec{S}_n$$

The simplest way is to linearly combine all the gradients with highest weight assign to the latest one:

$$\vec{S}^1 = \vec{\nabla}\Omega^1$$

$$\vec{S}^n = \vec{\nabla}\Omega^n + \vec{S}^{n-1} \frac{\left\langle \vec{S}^{n-1} \cdot \vec{\nabla}\Omega^n \right\rangle}{\left\| \vec{S}^{n-1} \right\| \cdot \left\| \vec{\nabla}\Omega^n \right\|}$$

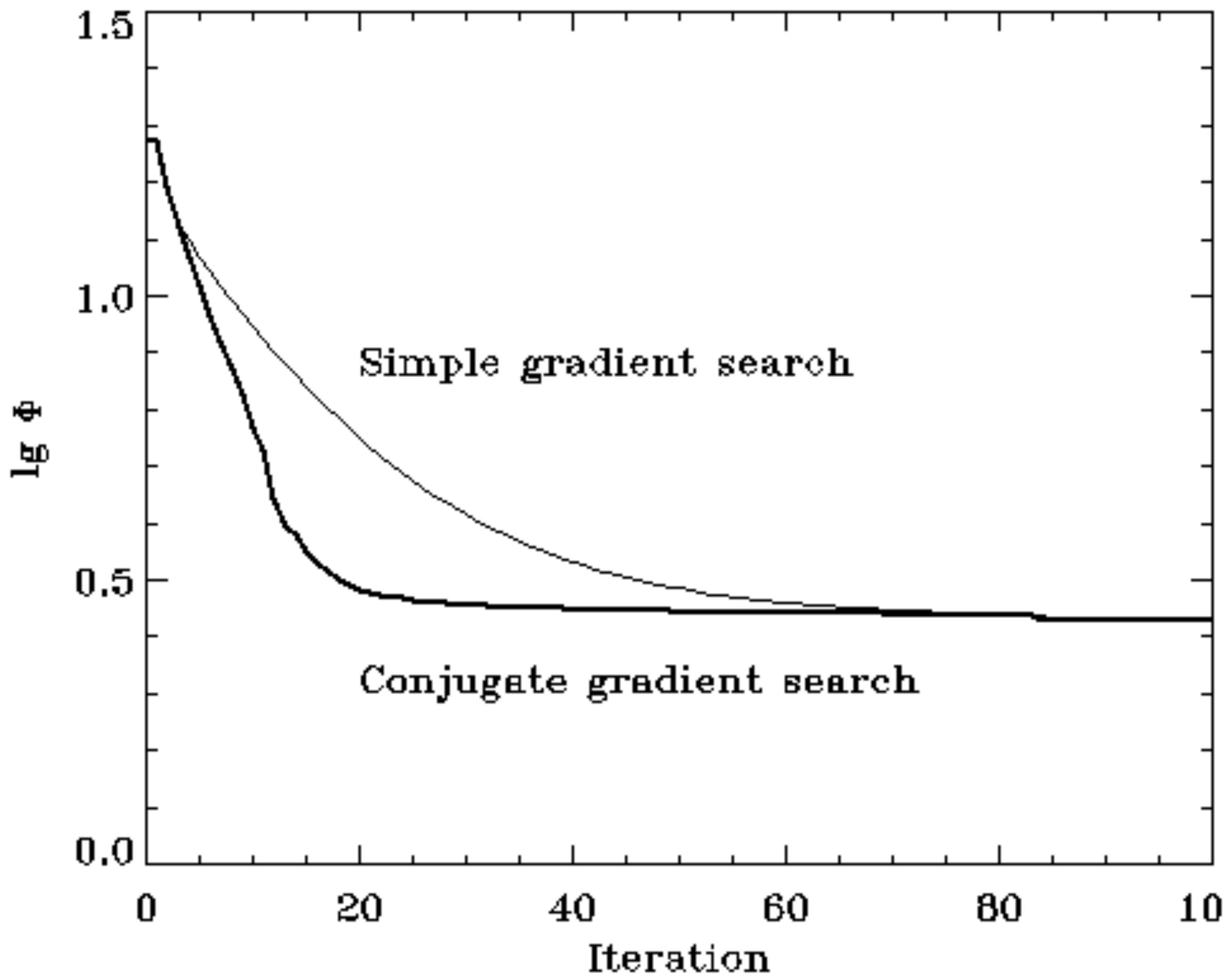


Other popular forms of linear combination were proposed by Fletcher & Reeves:

$$\vec{S}^n = -\vec{\nabla}\Omega^n + \vec{S}^{n-1}\frac{\|\vec{\nabla}\Omega^n\|^2}{\|\vec{\nabla}\Omega^{n-1}\|^2}$$

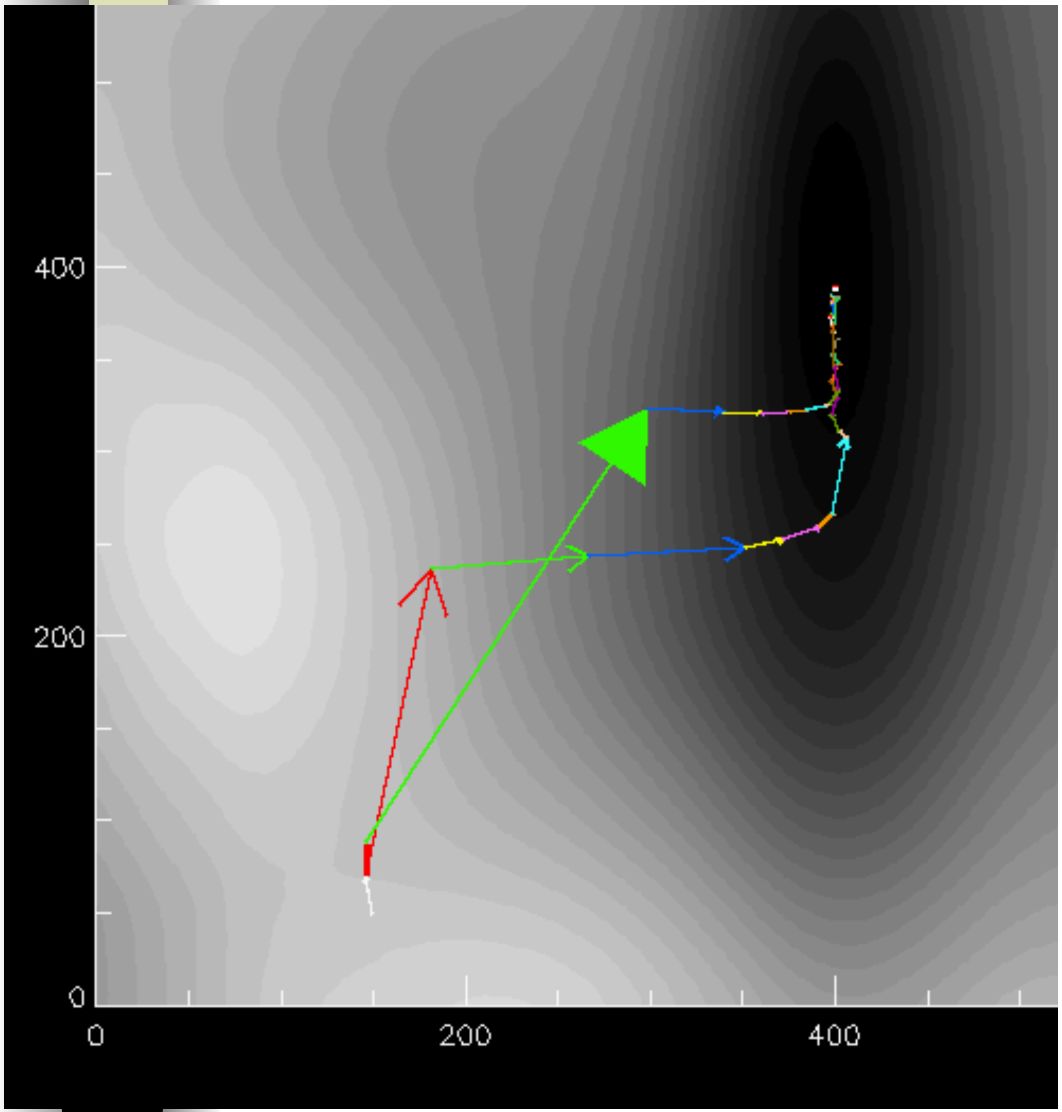
and Polak & Ribière:

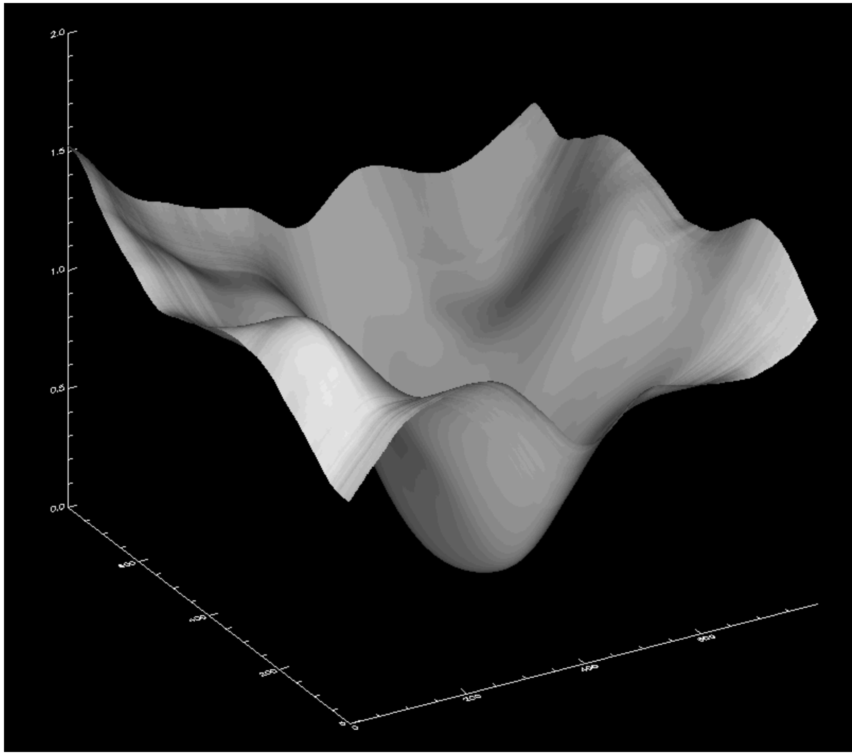
$$\vec{S}^n = -\vec{\nabla}\Omega^n + \vec{S}^{n-1}\frac{< \vec{\nabla}\Omega^n \cdot (\vec{\nabla}\Omega^n - \vec{\nabla}\Omega^{n-1}) >}{\|\vec{\nabla}\Omega^{n-1}\|^2}$$



1.5
1.0
0.5
0.0
lg Φ
Simple gradient search
Conjugate gradient search
0
20
40
60
80
10
Iteration

Example in 2D





Gradient search needs 30 iterations following complex zigzag path

Conjugate gradients reach the same goal in 19 iterations

You can find my IDL code [here]

- The theory predicts that the number of iterations equal to the number of variables is needed to achieve nearly quadratic approximation to the minimized function.
- After that it is worth resetting the direction to the pure gradient to avoid excessive accumulation of numerical errors.
- There are many different ways of combining gradients but the gain is marginal.

Higher order methods

- Newton
- Levenberg-Marquardt

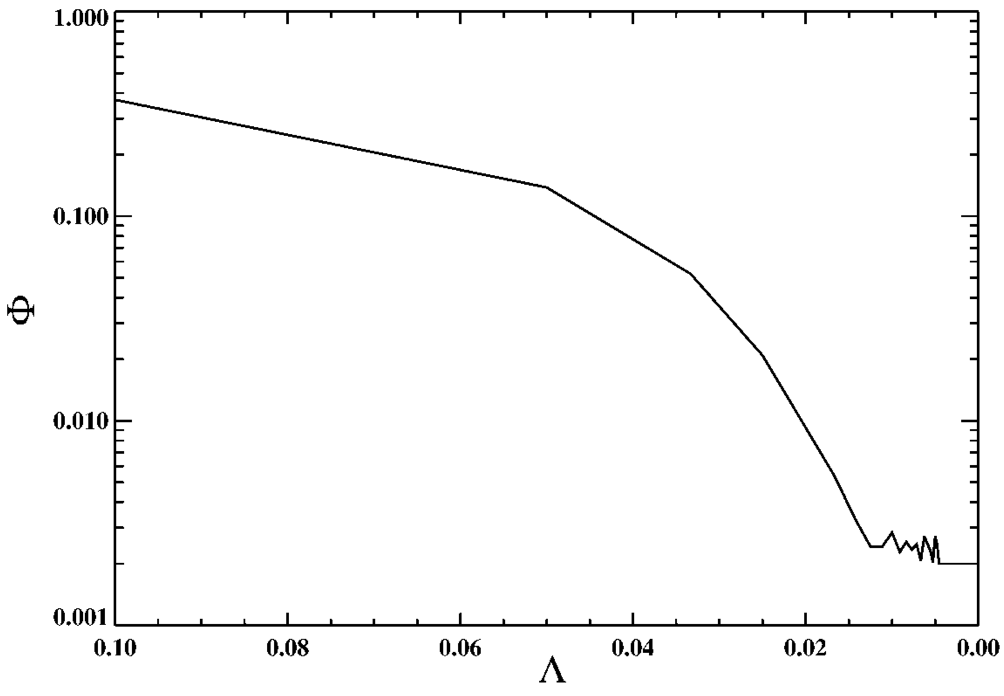
We will visit them later in the course

Selecting the regularization parameter

1. When the observation errors are known the regularization parameter Λ is selected so that $\Phi(f_{\min})$ matches the errors
2. Otherwise we should study a sequence of Λ. For each Λ we find minimum of Ω and the corresponding Φ. The key dependence is $\Phi(\Lambda)$.

Selecting the regularization parameter

We would expect Φ to decrease with decreasing Λ until certain level where it start behaving erratically:



Testing stability of the solution

Once the optimal Λ is selected on can run two tests:

1. Changing the initial guess
2. Perturbing the observations

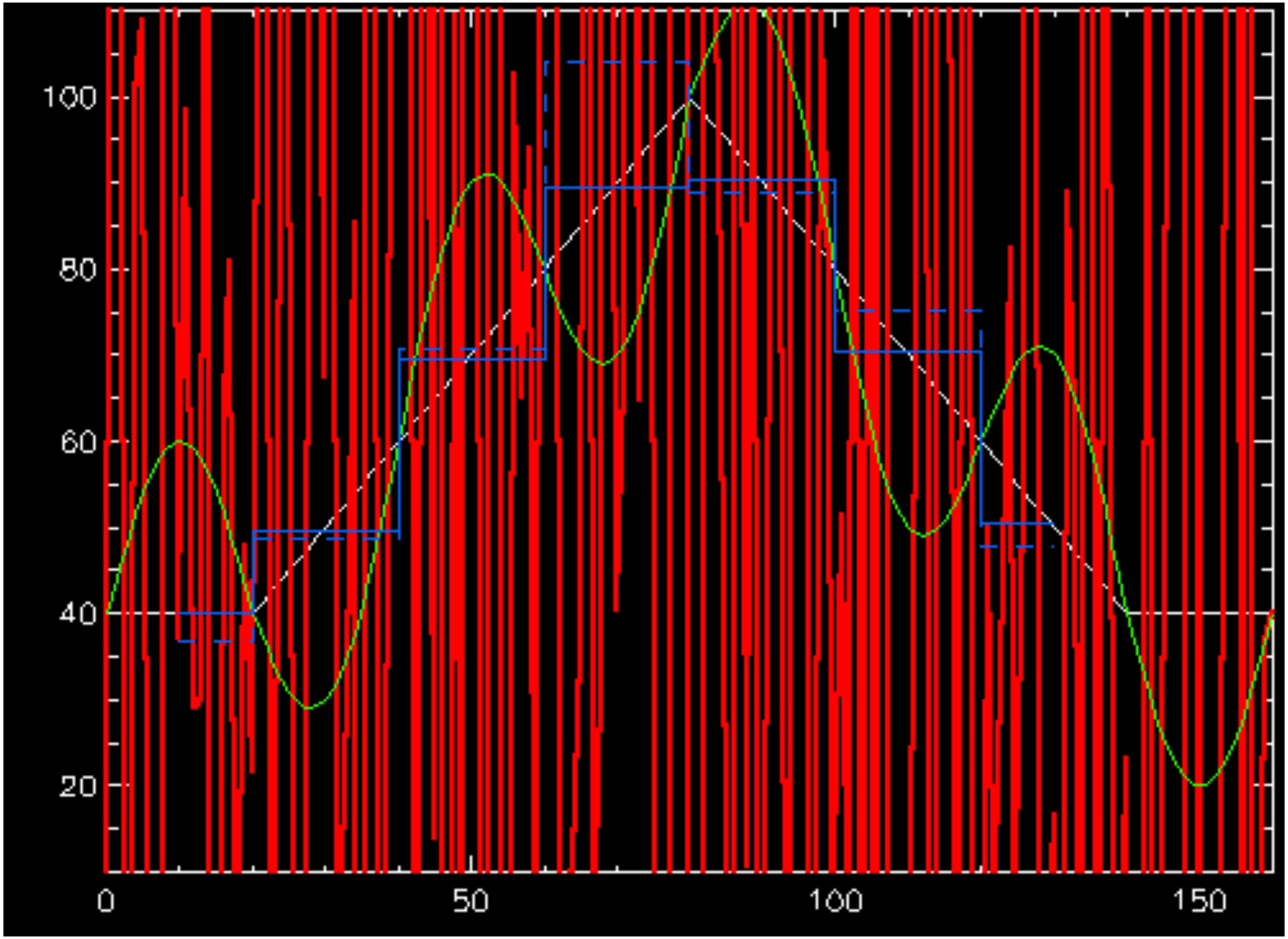
Testing stability against the observational errors:

In practice *g*'s are measured with certain accuracy. We can simulate that in our model by adding noise:

$$g' = g \cdot (1 + errors)$$

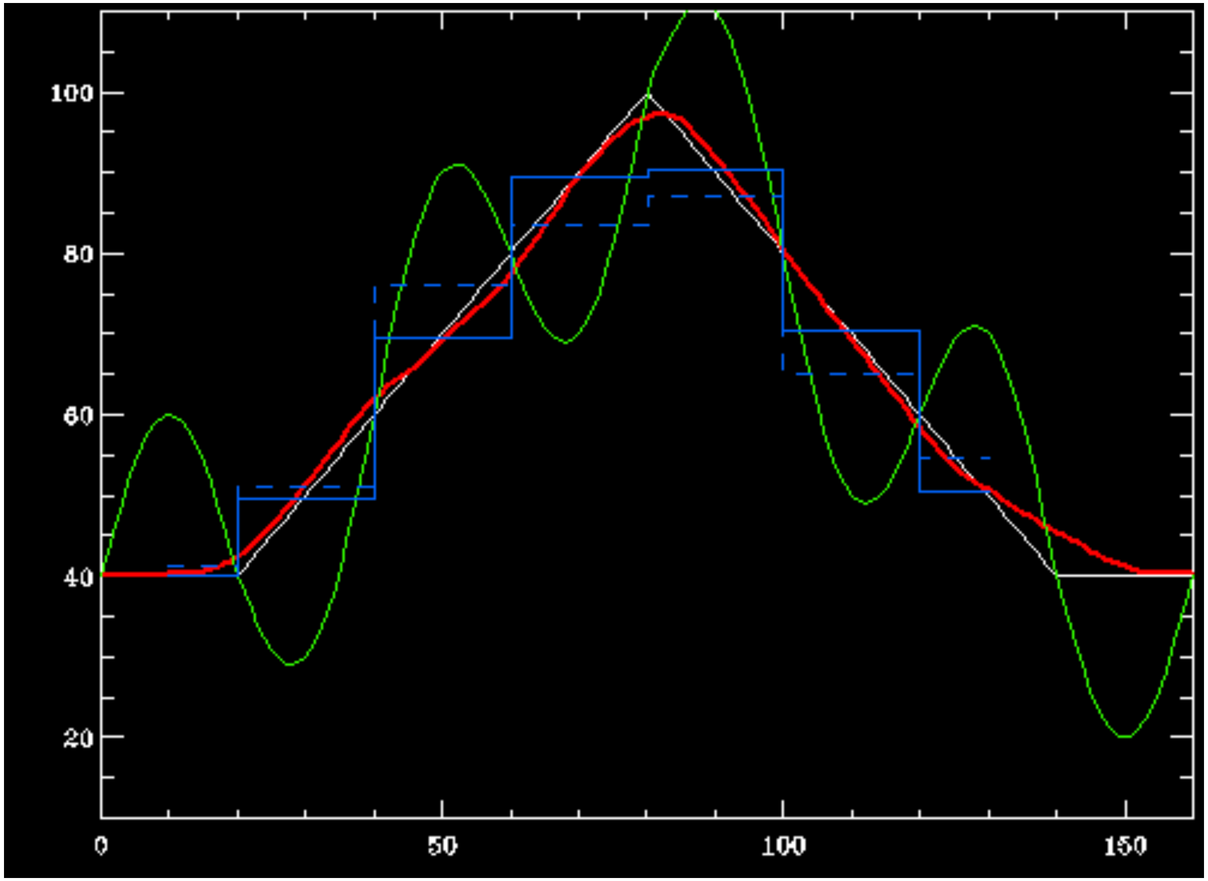
Without regularization the solution becomes unstable, that is for each realization of *errors* get very different solutions. The following example shows one realisation of *errors* corresponding to signal-to-noise ratio of 10.

Regularization is too low

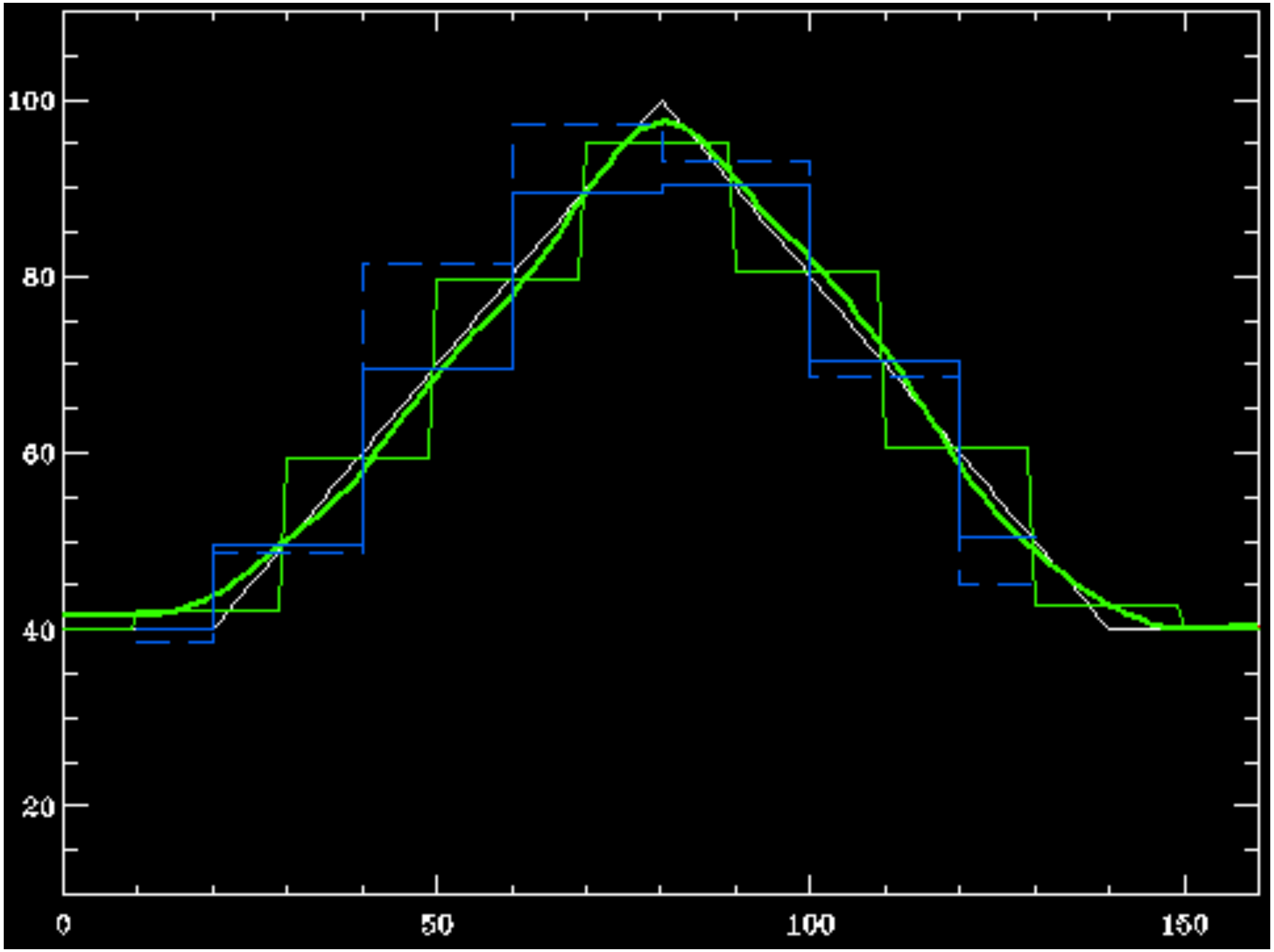


Regularization is optimal.

When regularization is added in the right proportion we get a stable solution reasonably close to the true one:



Now trying a different initial guess



Maximum Entropy Regularization

Regularization is given by:

$$\mathbf{R}^{MEM} = \sum_i p_i \ln p_i$$

where:

$$p_i = f_i / \bar{f}$$

$$\bar{f} = \tfrac{1}{N} \sum_k f_k$$

The gradient is then:

$$\frac{\partial \mathbf{R}}{\partial f_j} = \frac{\ln p_j - \mathbf{R}/N}{\bar{f}}$$

A Simplified form of Maximum Entropy Regularization

$$\mathbf{R}^{MEM} = \sum_i f_i \ln(f_i / \bar{f}); \quad \bar{f} = \tfrac{1}{N} \sum_k f_k$$

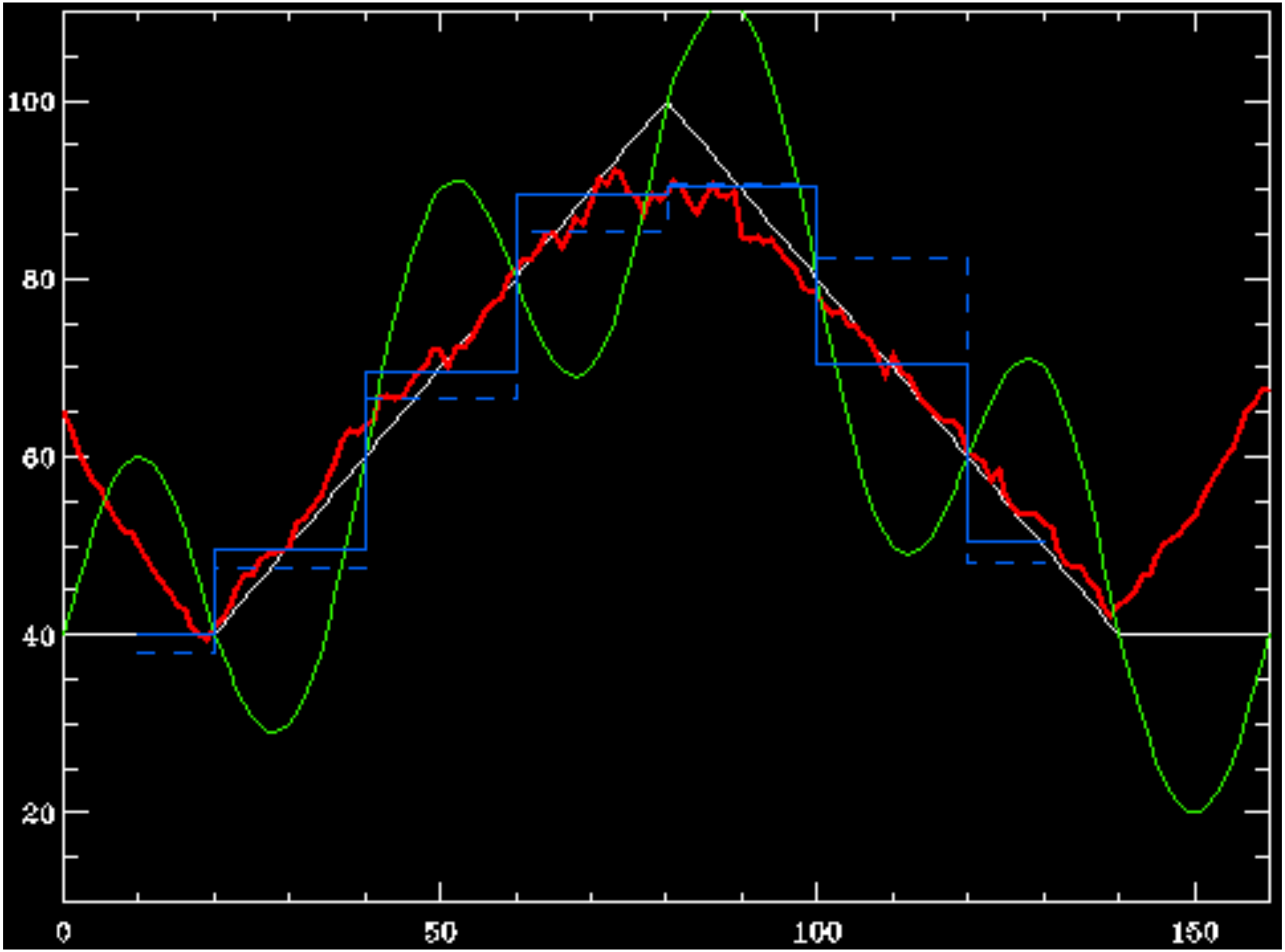
$$\frac{\partial \mathbf{R}}{\partial f_j} = \frac{\partial}{\partial f_j}\left[\sum_i f_i \left(\ln f_i - \ln \bar{f}\right)\right] =$$

$$= \ln f_j + 1 - \ln \bar{f} - \frac{\sum_i f_i}{\sum_k f_k} = \ln(f_j / \bar{f})$$

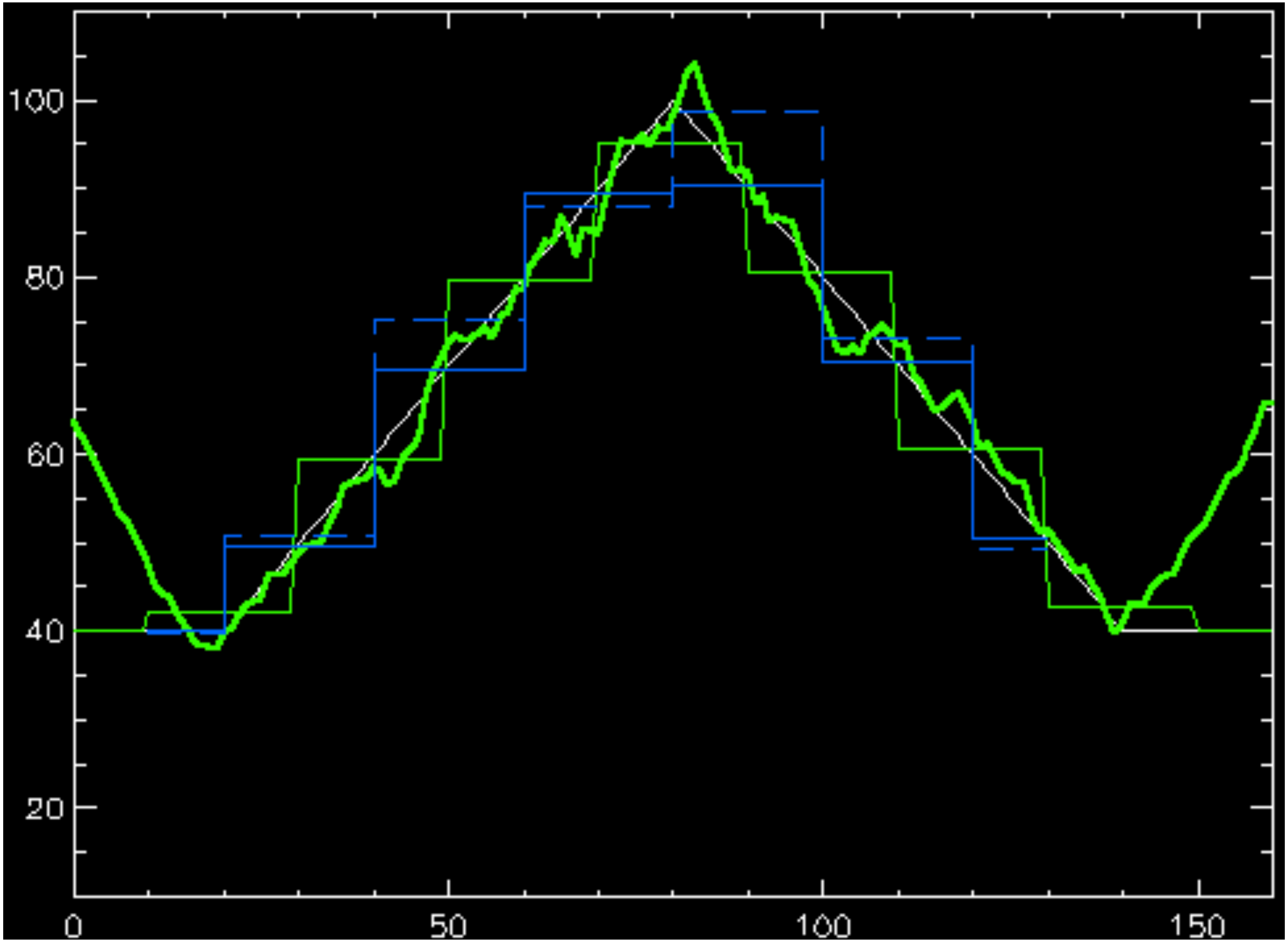
Maximum entropy regularization

- Since we have to "maximize" the entropy, the corresponding functional has to <u>subtracted</u> when constructing Ω
- The $\mathbf{R}^{MEM}$ "encourages" the solution to be close to the mean value, so a good test is to increase Λ and see if the solution comes closer to a horizontal line
- MEM regularization cannot handle non-positive values: apply an offset!

MEM regularization with different initial guesses



MEM regularization with different initial guesses



- Unfortunately MEM can create multiple local spurious minima (unlike Tikhonov regularization). The tactical approach is to start with large regularization parameter and decrease it as we are getting close to the solution. The procedure should stop when Φ reaches the level of observational errors.
- With Tikhonov regularization that is not necessary: we can use constant regularization parameter for the whole search.

Maximum likelihood

- For linear inverse problems the functional is represented by a rectangular matrix. In that case we would have to solve a system of linear equations where the number of equations is often larger than the number of unknowns.
- The method of choice is the Singular Value Decomposition (SVD, Num. Rec.).
- The statistical interpretation is that this method gives “the most likely” of all possible solutions (median of the distribution).

Next more serious example

- The PSF of a spectrometer is very important for detailed studies of line shapes and radial velocities. How to determine the PSF?
- Consider observations of a known absorption or emission spectrum. If the lines are unresolved (but not too dense) such spectrum can be used to derive the PSF.

The problem is a typical convolution type integral equation:

$$g = \int_{\lambda_1}^{\lambda_2} s(\lambda - x) f(x) dx$$

where S is the known spectrum (e.g. Iodine cell at +75° C measured with an Fourier Transform Spectrometer), f is the unknown PSF and the measured spectrum is g.

Home work for the next time

- Upgrade your minimization code to conjugate gradients search
- Try minimization with the noise in *g*'s and Tikhonov regularization
- Start thinking about application of inverse problem to your research