Turchin's method of statistical regularization

Mikhail Zelenyi [1,2]   Maria Polyakova(Nelyubina)[1,2]

[1]Institute for Nuclear Research RAS

[2]Moscow Institute of Physics and Technology

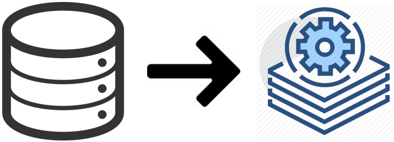
October 4, 2017

Scheme of an experiment

Experimental data = Observed value + Apparatus function + A bit of noise

$$f(y) = \varphi(x) * K(x,y) + \varepsilon_y$$

Processing: Solution of Fredholm integral equation



:

$$f(y) = \int dx\ K(x,y)\varphi(x)$$

$$\varphi(x) - ?$$

Solution of Fredholm equation (least squares)

In the integral form:

$$f(y) = \int dx \ K(x, y)\varphi(x)$$

For transition to matrix form:

$$\varphi(x) = \sum_n \varphi_n T_n(x),$$

where $\{T_n(x)\}$ is some function basis.

Then:

$$K_{mn} = \int K(x, y_m) T_n(x) \ dx$$

$$f_m = f(y_m)$$

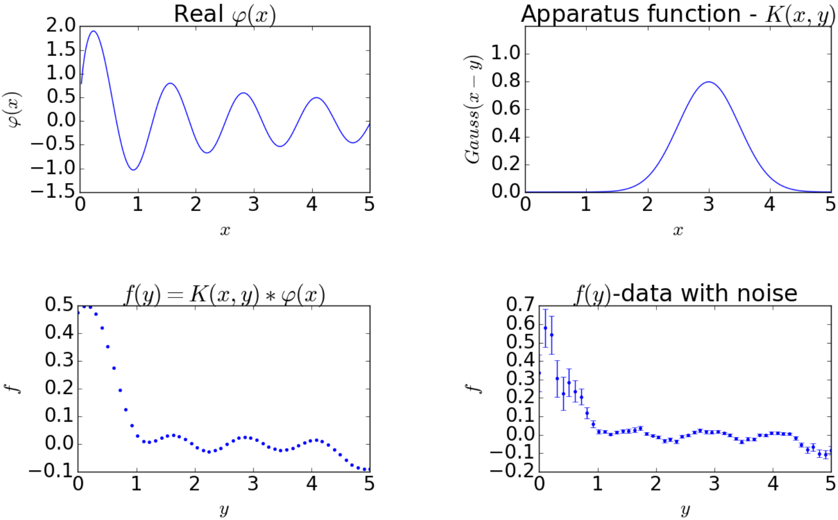
In the matrix form:

$$f_m = K_{mn}\varphi_n$$

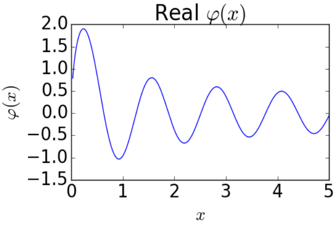
Solution with using method of least squares:

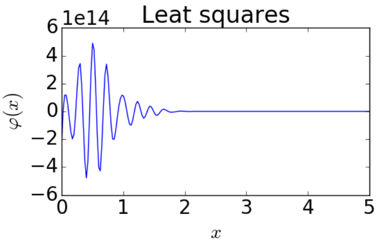
$$\varphi_n = (K_{mn}^T K_{mn})^{-1} K_{mn}^T f_m$$

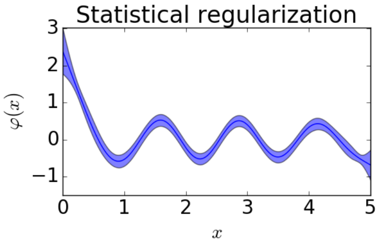
Numerical simulation: generation of data



Numerical simulation: comparison of two methods







Problem.

In the integral form:

$$f(y) = \int dx \; K(x, y)\varphi(x).$$

Fredholm equation is ill-possed.

In the matrix form:

$$f_m = K_{mn}\varphi_n$$

Matrix K is ill-conditioned

In summary

A small error when measuring $f(y)$ leads to big instability of $\varphi(x)$.

Alternative solution — to use regularization

Regularization is a process of introducing additional information for transition from ill-possed problem to well-possed problem.

Main features

- Based on Bayesian approach and decision theory (choice theory)
- Considered different a prior information: smoothness, non-negatives.
- Defined errors of obtained solution!
- Don't contend undefined parameter

Description

Choice of solution based on strategy $\hat{S}$, which used a prior information.

$$Optimal\ \varphi(x) = \hat{S}[f] = E[\varphi|f] = \int \varphi \frac{P(\varphi)P(f|\varphi)}{Norm} d\varphi$$

Error of solution:

$$D(x_1, x_2) = E[\varphi(x_1) - \hat{S}[f](x_1)][\varphi(x_2) - \hat{S}[f](x_2)]$$

Choice based on strategy $\hat{S}[f]$.
Good strategy minimize wrong from our ignorance.
Our ignorance is defined loss-function:

$$L(\varphi, \hat{S}[f]) = ||\varphi - \hat{S}[f])||_{L_2},$$

For this loss-function:

$$\hat{S}[f] = E[\varphi|f] = \int \varphi P(\varphi|f) d\varphi$$

Strategy depend on prior information $P(\varphi)$:

$$P(\varphi|f) = \frac{P(\varphi)P(f|\varphi)}{\int d\varphi P(\varphi)P(f|\varphi)}$$

Error of solution:

$$D(x_1, x_2) = E[\varphi(x_1) - \hat{S}[f](x_1)][\varphi(x_2) - \hat{S}[f](x_2)]$$

A prior information (smoothness)

Condition on prior information

Limit Shannon's information in prior probability:

$$I[P(\varphi)] = \int \ln P(\varphi) P(\varphi) d\varphi \rightarrow min$$

Normalize:

$$\int P(\varphi) d\varphi = 1$$

Choose more smoothness solutions:

$$\int \langle \varphi, \hat{\Omega}\varphi \rangle P(\varphi) d\varphi = \omega,$$

where ω - required level of smoothness, $\hat{\Omega}$ - operator of smoothness (for example $\hat{\Omega} = |\frac{d^2}{dx^2}\rangle\langle\frac{d^2}{dx^2}|$).

A prior information (smoothness)

Condition on prior information

$$I[P(\varphi)] = \int \ln P(\varphi) P(\varphi) d\varphi \to min$$

$$\int P(\varphi) d\varphi = 1$$

$$\int \langle \varphi, \hat{\Omega}\varphi \rangle P(\varphi) d\varphi = \omega,$$

In result: a prior probability density is Gauss random process

$$P_\alpha(\vec{\varphi}) = \frac{\alpha^{Rg(\Omega)/2} \det \Omega^{1/2}}{(2\pi)^{N/2}} \exp(-\frac{1}{2}(\vec{\varphi}, \alpha\Omega\vec{\varphi})),$$

where $\alpha = \alpha(\omega)$ - parameter of smoothness

A prior information

In result: $P(\varphi) = P_\alpha(\varphi)$ — prior information depend on parameter of smoothness

A prior information

In result: $P(\varphi) = P_{\alpha}(\varphi)$ — prior information depend on parameter of smoothness

Optimization of α

- Select manually using known smoothness. This is rare probability.

A prior information

In result: $P(\varphi) = P_\alpha(\varphi)$ — prior information depend on parameter of smoothness

Optimization of α

- Select manually using known smoothness. This is rare probability.
- Use most probable parameter: $\alpha^* = \max P(\alpha|f)$.

A prior information

In result: $P(\varphi) = P_\alpha(\varphi)$ — prior information depend on parameter of smoothness

Optimization of α

- Select manually using known smoothness. This is rare probability.
- Use most probable parameter: $\alpha^* = \max P(\alpha|f)$.
- Use prior information about smoothness:

$$P(\varphi) = \int P_\alpha(\varphi) P(\alpha)\, d\alpha$$

A prior information

In result: $P(\varphi) = P_\alpha(\varphi)$ — prior information depend on parameter of smoothness

Optimization of α

- Select manually using known smoothness. This is rare probability.
- Use most probable parameter: $\alpha^* = \max P(\alpha|f)$.
- Use prior information about smoothness:

$$P(\varphi) = \int P_\alpha(\varphi) P(\alpha)\, d\alpha$$

- Use posterior information about smoothness:

$$\hat{S}[f] = \int d\alpha \hat{S}_\alpha[f] P(\alpha|f),$$

A prior information

In result: $P(\varphi) = P_\alpha(\varphi)$ — prior information depend on parameter of smoothness

Optimization of α

- Select manually using known smoothness. This is rare probability.
- Use most probable parameter: $\alpha^* = \max P(\alpha|f)$.
- Use prior information about smoothness:

$$P(\varphi) = \int P_\alpha(\varphi) P(\alpha)\, d\alpha$$

- Use posterior information about smoothness:

$$\hat{S}[f] = \int d\alpha \hat{S}_\alpha[f] P(\alpha|f),$$

- **Two last methods is equivalent!**

Solution for Gaussian noise

$$P(\vec{f}|\vec{\varphi}) = \frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\vec{f} - K\vec{\varphi})^T \Sigma^{-1} (\vec{f} - K\vec{\varphi})$$

Using most probable α:

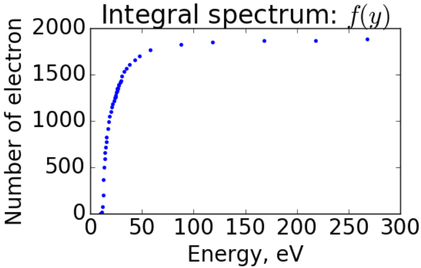
$$\vec{\varphi} = (K^T \Sigma^{-1} K + \alpha^* \Omega)^{-1} K^T \Sigma^{-1} \vec{f}$$

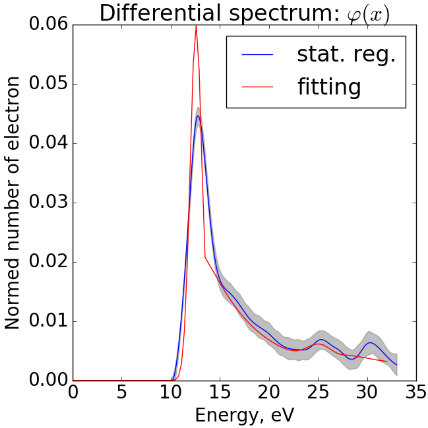
$$\Sigma_\varphi = (K^T \Sigma^{-1} K + \alpha^* \Omega)^{-1}$$

For comparison: method of least squares

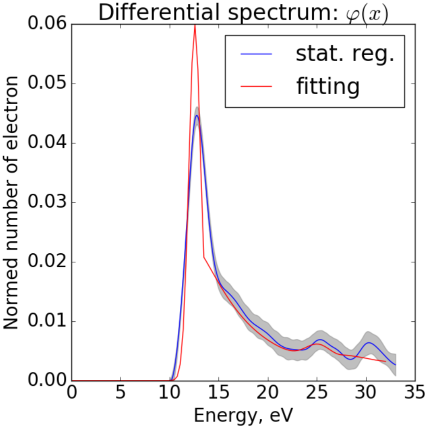
$$\vec{\varphi} = (K^T K)^{-1} K^T \vec{f}$$

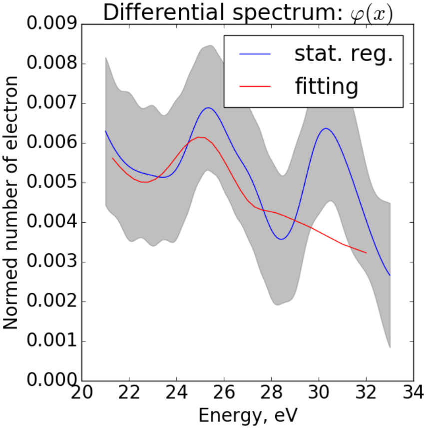
Experimental data: spectrum of electron scattering (Troitsk ν-mass data)





Experimental data: spectrum of electron scattering (Troitsk ν-mass data)





Procedure of fitting requires strong proposal about form of $\varphi(x)$.
Statistical regularization use less information about $\varphi(x)$.

Thank for you attention

Gaussian noise

$$P(\vec{f}|\vec{\varphi}) = \frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\vec{f} - K\vec{\varphi})^T \Sigma^{-1}(\vec{f} - K\vec{\varphi})$$

Define:

$$b = K^T \Sigma^{-1} \vec{f},\ B = K^T \Sigma^{-1} K$$

Than:

$$P(\vec{f}|\alpha) = \frac{\alpha^{Rg(\Omega)/2}|\Omega^{1/2}|}{(2\pi)^{(M)/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2} b^T B^{-1} b) \sqrt{|(B + \alpha\Omega)^{-1}|} \exp(\frac{1}{2} b^T (B + \alpha\Omega)^{-1} b)$$

Solution:

$$\vec{\varphi} = (K^T \Sigma^{-1} K + \alpha^* \Omega) K^T \Sigma^{-1} \vec{f}$$

$$\Sigma_\varphi = (K^T \Sigma^{-1} K + \alpha^* \Omega)^{-1}$$

Different methods of regularization

Tikhonov regularization

Find solution parametric approximate problem, which will trend to solution on accuracy problem for some value of parameter. For example, Fredholm equation can replaced by search minimum of next operator:

$$\Phi^{\alpha}[\varphi, f] = \int dy \left[f(y) - \int dx \; K(x,y)\varphi(x) \right]^2 + \underbrace{\alpha \left(\int r(x)\varphi^2 \; dx + \int q(x)(\varphi')^2(x) \right)}_{regularization \;\; operator},$$

where $r \geq 0, q \geq 0$, α - regularization parameter.

Disadvantages:

1. Correct α exist, but unknown,
2. Error of solution is unknown.

$$\int d\alpha \hat{S}_\alpha[f]P(\alpha|f) = \int d\alpha \left(\frac{\int \varphi P(f|\varphi)P(\varphi|\alpha)d\varphi}{\int P(f|\varphi)P(\varphi|\alpha)d\varphi} \right) * \frac{P(f|\alpha)P(\alpha)}{\int d\alpha P(f|\alpha)P(\alpha)} =$$

$$= \int d\alpha \left(\frac{\int \varphi P(f|\varphi)P(\varphi|\alpha)d\varphi}{\int P(f|\varphi)P(\varphi|\alpha)d\varphi} \right) * \frac{\left(\int d\varphi P(f|\varphi)P(\varphi|\alpha)\right) P(\alpha)}{\int d\alpha \int d\varphi P(f|\varphi)P(\varphi|\alpha)P(\alpha)} =$$

$$= \frac{\int d\alpha \left(\int \varphi P(f|\varphi)P(\varphi|\alpha)d\varphi\right) * P(\alpha)}{\int d\alpha \int d\varphi P(f|\varphi)P(\varphi|\alpha)P(\alpha)} = \frac{\int d\varphi \varphi P(f|\varphi) \int d\alpha P(\varphi|\alpha)P(\alpha)}{\int d\varphi P(f|\varphi) \int d\alpha P(\varphi|\alpha)P(\alpha)} =$$

$$= \frac{\int \varphi P(\varphi)P(f|\varphi)d\varphi}{\int P(\varphi)P(f|\varphi)d\varphi} = \hat{S}[f]$$