Application of Turchin's method of statistical regularization

Mikhail Zelenyi[1,2,⋆] and *Maria* Polyakova[1,2]

[1] *Institute for Nuclear Research RAS*
[2] *Moscow Institute of Physics and Technology*

Abstract. During analysis of experimental data, one usually needs to restore a signal after it has been convoluted with some kind of apparatus function. According to Hadamard's definition this problem is ill-posed and requires regularization to provide sensible results. In this article we describe an implementation of the Turchin's method of statistical regularization based on the Bayesian approach to the regularization strategy.

1 Introduction

1.1 Problem

Consider the usual situation: detector measures a signal ($\varphi(x)$) and produces observed data ($f(y)$) via convolution with its own apparatus function ($K(x,y)$). The resulting observed data $f(y)$ contains a random noise factor both from initial statistical uncertainty of $\varphi(x)$ and additional noise from measurement procedure. In order to reconstruct initial signal, one needs to solve Fredholm equation:

$$f(y) = \int dx\, K(x,y)\varphi(x), \tag{1}$$

but this equation is ill-possed: a small error in the measurement of $f(y)$ leads to big instability of $\varphi(x)$.

1.2 Solution

Solving such ill-possed problems requires additional operation called regularization. Regularization is a process of introducing additional information for transition from ill-possed problem to well-possed problem. The idea of statistical regularization [1, 2] is to look on the problem from the point of view of Bayessian statistics approach: unknown statistical value $\varphi(x)$ could be reconstructed using measured value $f(y)$, the model ($K(x,y)$) and some prior information about $\varphi(x)$ behavior. Main features of statistical regularization:

- Based on Bayesian approach and decision theory (choice theory).
- It's able to accommodate various forms of prior information, such as smoothness, constraints on boundary conditions, non-negativity, etc.
- It allows estimation of errors for reconstructed function ϕ

⋆e-mail: mihail.zelenyy@phystech.edu

- Gives clear data-driven procedure for choice of regularization strength.

While the idea of the regularization was proposed a long time ago, it was not published in the journal with sufficient visibility and therefore left unknown to the broad scientific auditory. The aim of this work is both to reconstruct the initial work (create a modern implementation for described algorithms), and study its performance under different conditions.

Since original work is very hard to find and is not accessible in languages other then Russia, we will present here the short description of the method.

2 Description of statistical regularization's method

2.1 Strategy

Consider eq. 1 in operator form:

$$\hat{K}\varphi = f. \tag{2}$$

According to statistical decision theory ([3]), one can define a reconstruction strategy $\hat{\varphi} = \hat{S}[f]$, which uses a prior information. Good strategy minimize an impact of prior information (which could be wrong). That could be achieved by introducing the loss-function:

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

For this loss-function:

$$\hat{S}[f] = E[\varphi|f] = \int \varphi P(\varphi|f) d\varphi \tag{3}$$

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

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

Errors of solution:

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

Let us consider what a prior information can we use?

2.2 A prior information

A general case of prior information is assumption about function smoothness. The applications of such prior information is done in three stages:

1. Minimize additional Shannon's information:

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

2. Normalize probability density:

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

3. Analyze information about smoothness of function $\varphi(x)$:

$$\int \langle \varphi, \hat{\Omega}\varphi \rangle P(\varphi) d\varphi = \omega, \tag{4}$$

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

The resulting prior probability density for the Gaussian random process:

$$P_\alpha(\vec{\varphi}) = \frac{\alpha^{Rg(\Omega)/2} \det \Omega^{1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}(\vec{\varphi}, \alpha\Omega\vec{\varphi})\right), \tag{5}$$

where $\alpha = \alpha(\omega)$ - parameter of smoothness. The prior information depends on α-smoothness parameter. This parameter could be estimated in a few different ways:

a) select manually using known smoothness, but this is rare probability;

b) use most probable parameter: $\alpha^* = \max P(\alpha|f)$;

c) use a prior information about smoothness:

$$P(\varphi) = \int P_\alpha(\varphi) P(\alpha)\, d\alpha; \tag{6}$$

d) use a posterior information about smoothness:

$$\hat{S}[f] = \int d\alpha \hat{S}_\alpha[f] P(\alpha|f). \tag{7}$$

In fact, methods (c) and (d) are equivalent.

2.3 Additional prior information

In eq. 4 one can use not only smoothness operator, but also different or even multiple different conditions:

$$P_{\alpha\beta}(\varphi) \sim \det|\alpha\Omega_1 + \beta\Omega_2| \exp\left(-\frac{1}{2}(\vec{\varphi}, (\alpha\Omega_1 + \beta\Omega_2)\vec{\varphi})\right)$$

This features allow to combine different types of parametric prior information. Also one can use non-parametric prior. For example, information about non-negativity:

$$P(\varphi) \sim P_\alpha(\varphi) P(\varphi(x) > 0)$$

2.4 Discretization

In order to find numerical solution, one need to go from continuous functions in eq. 2 to discrete matrix form:

$$\begin{gathered}
\varphi(x) = \sum_n \varphi_n T_n(x), \\
K_{mn} = (\hat{K} T_n(x))(y_m), \\
\Omega_{ij} = \int_a^b \left(\frac{d^p T_i(x)}{dx}\right)\left(\frac{d^p T_j(x)}{dx}\right) dx, \\
D[\varphi(x)] = D[\sum_n \varphi_n T_n(x)] = \sum_{i,j} \varphi_i \varphi_j cov(T_i(x), T_j(x)), \\
\vec{f} = K\vec{\varphi},
\end{gathered} \tag{8}$$

where $T_n(x)$ - some basis in function space. For example, cubic spline, Fourier series and Legendre polynomials.

2.5 Solution for Gaussian distributed noise

The most common experimental case of noise distribution is normal (or Gaussian) distribution:

$$P\left(\vec{f}|\vec{\varphi}\right) = \frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\vec{f} - K\vec{\varphi})^T\Sigma^{-1}(\vec{f} - K\vec{\varphi})\right) \tag{9}$$

Using the most probable α, one can get the best solution in a simple form.

$$\hat{S}_\alpha[\vec{f}] = (K^T\Sigma^{-1}K + \alpha\Omega)^{-1}K^T\Sigma^{-1T}\vec{f}, \tag{10}$$

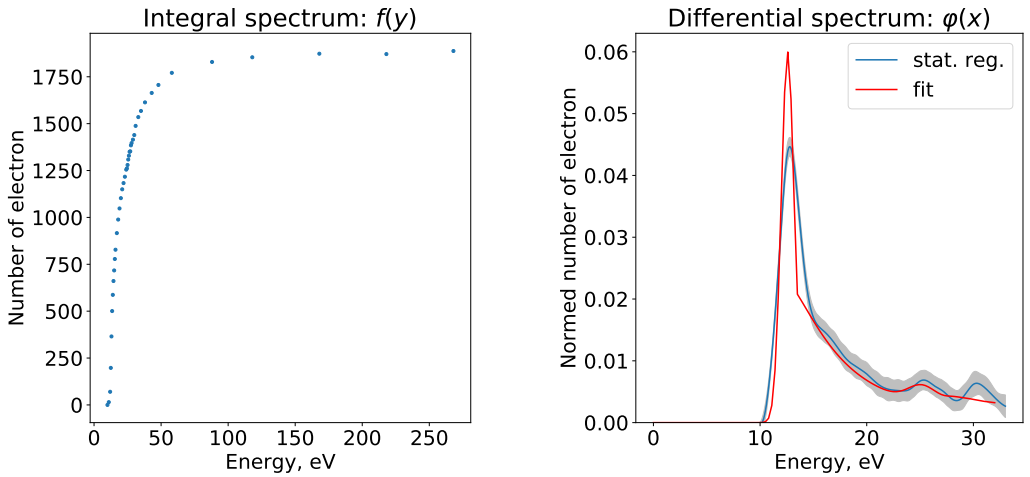
$$\mathrm{cov}_\alpha(\varphi_m, \varphi_n) = \|(K^T\Sigma^{-1}K + \alpha^*\Omega)^{-1}\|_{mn} \tag{11}$$

If we want to get more accurate solution, we should average solution over a posterior probability, which is calculated using Bayes' rule:

$$P(\alpha|\vec{f}) = C\alpha^{\frac{R_\theta(\Omega)}{2}}\sqrt{|(B + \alpha\Omega)^{-1}|}\exp(-\frac{1}{2}b^TB^{-1}b)\exp(\frac{1}{2}b^T(B + \alpha\Omega)^{-1}b), \tag{12}$$

where norming factor C doesn't depend on α. This formula allows to take integrals (7) and (6) numerically.

3 Application



(a) (b)

Figure 1: The spectrum of electron scattering (Troitsk ν-mass data)

The basic concepts of the described method (alongside some improvements) were implemented in a new prototype program written in the Python language. Then the program was used to reconstruct the differential cross-sections of electron scattering on hydrogen isotopes measured in Troitsk nu-mass experiment. In this experiment gaseous hydrogen (as well as deuterium and tritium) was irradiated by electrons with energies of about 20 keV from electron gun. Troitsk nu-mass spectrometer on the other side of gas volume registers the integral electron spectrum (like at fig 1a) with relative resolution of about 10^{-5}. In previous works ([4] and [5]) the differential cross-section of scattering electrons on $H2$ molecules was reconstructed using fit procedure against 5 or 6 parameters shape. That analysis had three major flaws:

- The parametric shape for fit does not have sufficient physics basis and therefore result is subjective in respect to the parametrization.
- The chosen parametrization has too many parameters and produces very tight correlations between these parameters, making the fit results instable.
- There is not clean way to determine uncertainties for any given point of reconstructed function.

Turchin's regularization allows to avoid all these problems and produces clean model-independent result with uncertainties. The result of regularization and previously published fit are presented at Fig. 1b. The regularization reconstruction is in good agreement with fitted curve. Also there are two very small peaks picture shows two additional peaks at 25 and 30 eV.The peak on 25 eV is known to be produced by double scattering events. Peak on 30 eV is previously unknown. It could be either statistical anomaly (errors are quite large in this point), or possibly point to existence of additional dissociative scattering.

4 Conclusion

Turchin's statistical regularization is a very powerful tool to solve ill-posed inversed problems. It provides a flexible way to introduce almost any kind of prior knowledge into reconstruction problem. Also it opens a new way to solve some problems which previously were solved only by direct approach. In this work, we created an implementation of Turchin's regularization using modern language and successfully applied it to the Troitsk nu-mass data for electron scattering on hydrogen isotopes. The work was preceded by a lot of testing of the algorithm with different artificial functions and discretization options. The results of those tests are beyond the scope of this article.

This work is supported by the Ministry of Education and Science of the Russian Federation under the contract No. 3.3008.2017/PP.

References

[1] V.F. Turchin, V.Z. Nozik, in *Physics of atmosphere and ocean* (Nauka, 1969), v. 5, `https://books.google.ru/books?id=N3RUAAAAMAAJ`

[2] L.S. Turovceva, Ph.D. thesis (1973)

[3] F. James, *Statistical Methods in Experimental Physics* (World Scientific, 2006), ISBN 9789812705273, `https://books.google.ru/books?id=S8N\_QgAACAAJ`

[4] V. Aseev, A. Belesev, A. Berlev, E. Geraskin, O. Kazachenko, Y. Kuznetsov, V. Lobashev, R. Ostroumov, N. Titov, S. Zadorozhny et al., The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics **10**, 39 (2000)

[5] D.N. Abdurashitov et al., Physics of Particles and Nuclei Letters **14**, 892 (2017), `1603.04243`