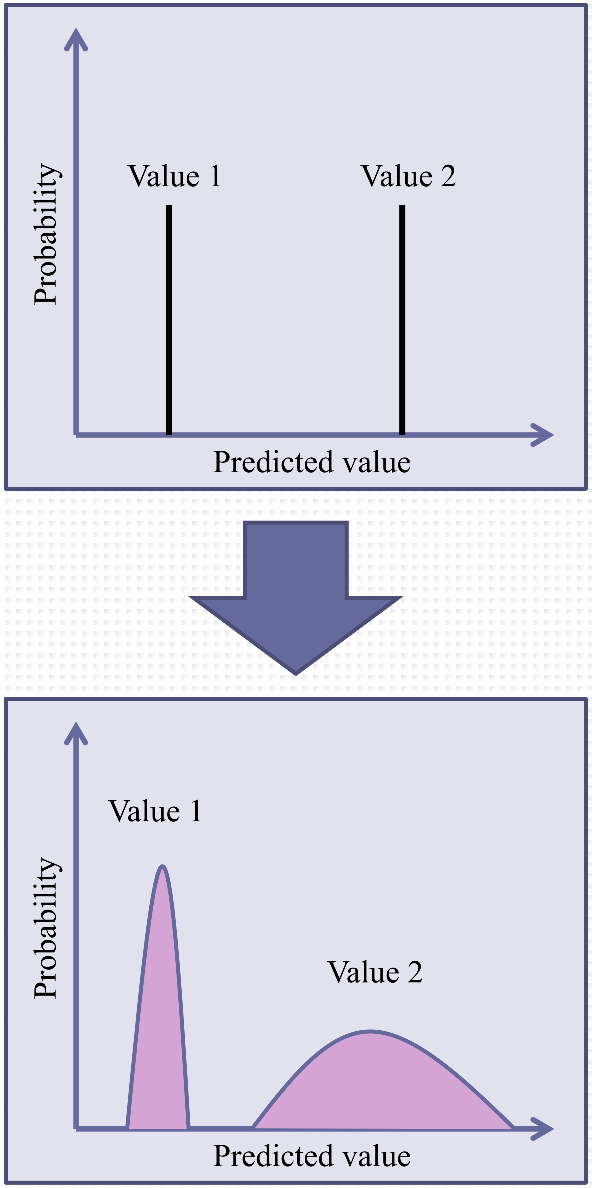
Bayesian Inference and nuclear structure models

Bartłomiej Szpak
H. Niewodniczański Institute of Nuclear Physics, Kraków, Poland

Physical motivation

We aim to:

- Provide statistical uncertainties for model parameters and predictions
- Investigate the predictive power of different nuclear structure models
- **Understand methods that could help to improve the predictive power**



With four parameters I can fit an elephant and with five I can make him wiggle his trunk (J. von Neumann)
Wei J. 1975. Least square fitting of an elephant. Chemtech 5: 128–129

Predictive power – a definition ?

There is no single measure of *predictive power*.

One cannot say that a Model A posses better *predictive power* then Model B.

Predictive power is always a function of:

- A set of predicted values (which remains unknown)
- A set of experimental data used to estimate the model parameters
- **A method used to fit the data**

However: every model can be characterized by the mean (expected) value of model parameters its covariance matrix.

Introductory statistics (linear regression)

Experimental data

$$d^{exp} = d^{th} + \delta(0,\sigma)$$

Random noise (inexact model)

$$d^{th} = Gm$$

G – model
m – parameters

Solution of a problem in a least-square sense:

Least-square solution

$$m^{L2} = (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G})^{-\mathbf{1}} G^T d^{exp}$$

Coviariance matrix of model parameters

$$cov(m^{L2}) = \sigma^2 (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G})^{-\mathbf{1}}$$

$$\sigma^2 = (d^{exp} - Gm^{L2})^T (d^{exp} - G m^{L2}) / dof$$

Nonlinear regression

The iterative solution can be obtained with e.g: the Gauss-Newton scheme or Levenberg-Marquardt algorithm:

$$\Delta m = (\boldsymbol{J}^{T}\boldsymbol{J})^{-1} J^{T}(Gm - d^{exp})$$

$$cov(m^{L2}) = \sigma^{2}(\boldsymbol{J}^{T}\boldsymbol{J})^{-1}$$

The statistical properties of a model are determined by the behaviour of a $(\boldsymbol{J}^{T}\boldsymbol{J})^{-1}$ term.

This depends on:

- Physical structure of a model
- Data used in the fit

Consider 5 x 5 Hilbert matrix:

$$H_{ij} = 1/(i+j-1)$$

and:

$$H'_{ij} = r_{ij} H_{ij}$$

where r is a random matrix with entries from (0.999-1.001). Then we can get:

$$\mathrm{H}^{-1}/\mathrm{H}'^{-1} = \begin{bmatrix} 6.7 & 20.8 & -1174.4 & -38.1 & -25.2 \\ 16.8 & 48.7 & -535.2 & -62.1 & -38.7 \\ 64.8 & 289.7 & -221.1 & -96.6 & -67.4 \\ -115.6 & -123.0 & -137.9 & -155.2 & -173.1 \\ -144.0 & -60.6 & -102.9 & -280.7 & 561.4 \end{bmatrix}$$

Condition number: ratio of maximal and minimal singular value of a matrix.

Ill-posed problems

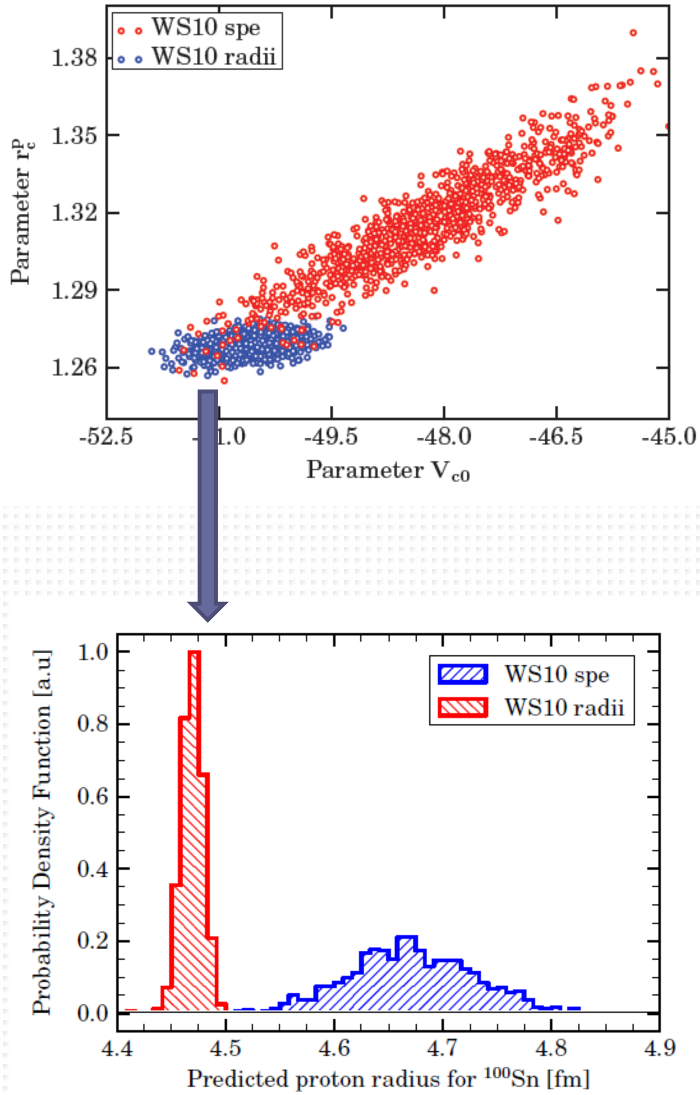
General conclusion: $(J^T J)^{-1}$ can be very unstable.

This can cause:

- Problems with convergence of minimization algorithm
- Unphysical solution with resonable σ^2
- Lack of predictive power for some groups of observables: see example

Solution:

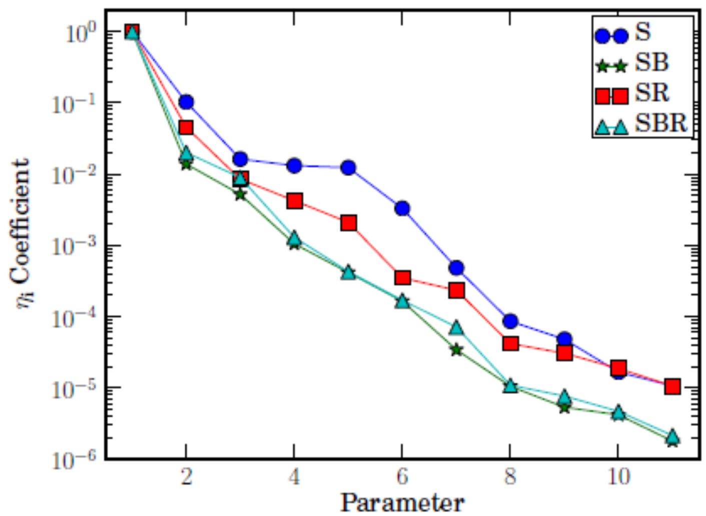
- Control the behaviour of a jacobian to check if a problem is ill-posed
- Change a model (???)
- Change a dataset (?)
- Apply regularization techniques (?)



Is this our problem ?

I consider typical spherical Skyrme-Hartree-Fock model with 11 parameters fitted to properties of doubly magic nuclei:
^{16}O, ^{40,48}Ca, ^{56}Ni, ^{90}Zr, ^{132}Sn, ^{208}Pb

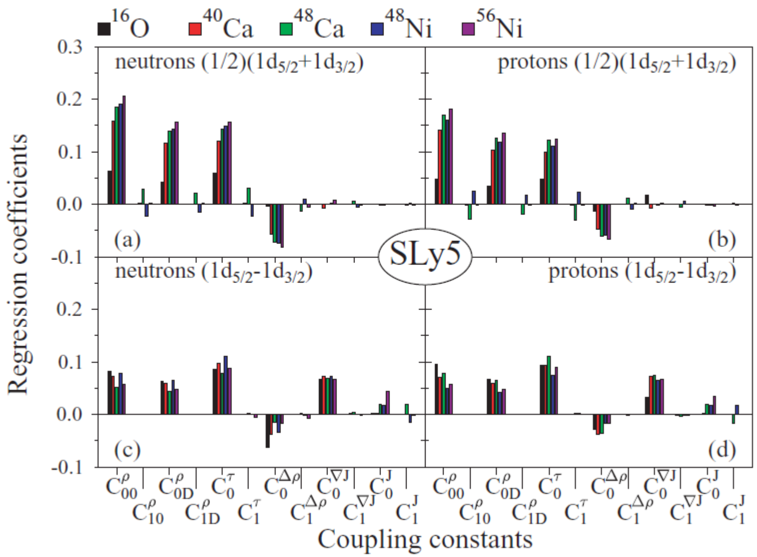
- binding energies
- proton radii
- single particle energies



Need to fix at least three model parameters

At least 240 parameterizations of a Skyrme force.

PRC **77**, 064307 (2008)
Dependence of single-particle energies on coupling constants of the nuclear energy density functional
M. Kortelainen, J. Dobaczewski, K. Mizuyama, and J. Toivanen



Is this our problem ?

UNEDFnb, 12 parameters

k	Par.	$\hat{\mathbf{x}}$	95% CI	% of Int.	σ
1	ρ_c	0.151046	[0.149,0.153]	10	0.001
2	E^{NM}/A	−16.0632	[−16.114,−16.013]	5	0.039
3	K^{NM}	337.878	[302.692,373.064]	70	26.842
4	$a_{\mathrm{sym}}^{\mathrm{NM}}$	32.455	[28.839,36.071]	72	2.759
5	$L_{\mathrm{sym}}^{\mathrm{NM}}$	70.2185	[11.108,129.329]	296	45.093
6	$1/M_s^*$	0.95728	[0.832,1.083]	21	0.096
7	$C_0^{\rho\Delta\rho}$	−49.5135	[−55.786,−43.241]	21	4.785
8	$C_1^{\rho\Delta\rho}$	33.5289	[−2.246,69.304]	36	27.292
9	V_0^n	−176.796	[−194.686,−158.906]	18	13.648
10	V_0^p	−203.255	[−217.477,−189.033]	14	10.850
11	$C_0^{\rho\nabla J}$	−78.4564	[−85.137,−71.775]	19	5.097
12	$C_1^{\rho\nabla J}$	63.9931	[23.460,104.526]	54	30.921

UNEDF0, 10 parameters,
K^{NM} and $1/M_s$ fixed

k	Par.	$\hat{\mathbf{x}}$	95% CI	% of Int.	σ
1	ρ_c	0.160526	[0.160,0.161]	10	0.001
2	E^{NM}/A	−16.0559	[−16.146,−15.965]	45	0.055
3	K^{NM}	230	–	–	–
4	$a_{\mathrm{sym}}^{\mathrm{NM}}$	30.5429	[25.513,35.573]	126	3.058
5	$L_{\mathrm{sym}}^{\mathrm{NM}}$	45.0804	[−20.766,110.927]	219	40.037
6	$1/M_s^*$	0.9	–	–	–
7	$C_0^{\rho\Delta\rho}$	−55.2606	[−58.051,−52.470]	9	1.697
8	$C_1^{\rho\Delta\rho}$	−55.6226	[−149.309,38.064]	94	56.965
9	V_0^n	−170.374	[−173.836,−166.913]	3	2.105
10	V_0^p	−199.202	[−204.713,−193.692]	6	3.351
11	$C_0^{\rho\nabla J}$	−79.5308	[−85.160,−73.901]	16	3.423
12	$C_1^{\rho\nabla J}$	45.6302	[−2.821,94.081]	65	29.460

- One of the parameters (isovector effective mass) has to be fixed a priori
- After other two parameters are fixed, some important model parameters lie outsite of the previously defined confidence intervals
- Increase of the standard errors when less parameters are used in the fit.

Bayesian approach

Bayesian interpretation of probability – probability measures a degree of belief

pdm - probability of observing data **d** for a specified model **m**

pmd - probability that model m is correct infered on a basis of d

$p(m)$ - prior knowledge of model parameters

$$pmd = pdm\, p(m)/c$$

In traditional regression one maximizes pdm. When there is no prior knowledge about model parameters two approaches are equivalent.

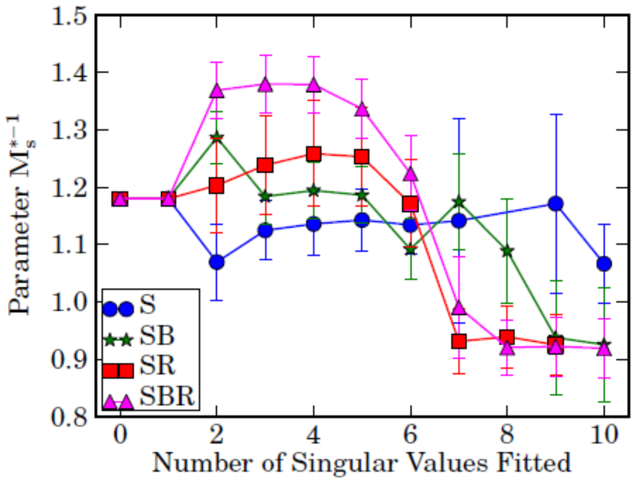
The probability of having rain when one observes the clouds and probability of having clouds when one observes the rain are certainly different.

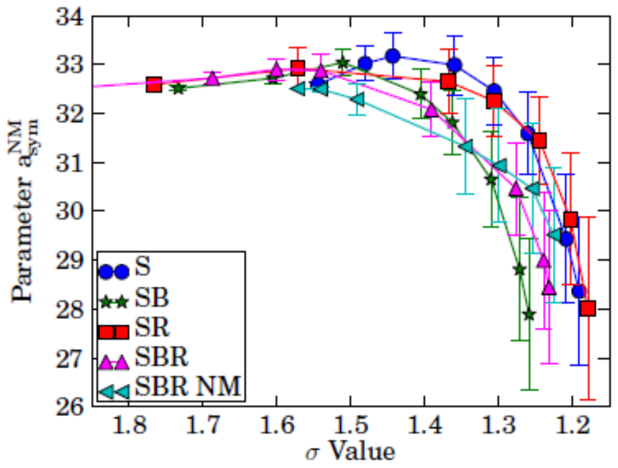
Relation to other regularization techniques

Regularization by fixing model parameters: priors with zero variance

In opposite to TSVD or Tikhonov method no need to choose regularization parameter: you believe in your prior.

For some cases equivalent to Tikhonov method for a special choice of regularization parameter.





Bayesian approach

Explicit inclusion of prior expectations about model parameters helps to regularize ill-posed problem.

Calculations much more complicated in general case. Analitical formulas only when one uses conjugated priors.

$$m^{L2} = (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G})^{-\mathbf{1}} G^{T} d^{exp}$$

$$m^{L2} = (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G} + \boldsymbol{\Sigma}^{-\mathbf{1}})^{-\mathbf{1}} (G^{T} d^{exp} + \boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{\mu})$$

$$cov(m^{L2}) = \sigma^{2} (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G})^{-\mathbf{1}}$$

$$cov(m^{L2}) = s^{2} (\boldsymbol{G}^{\boldsymbol{T}} \boldsymbol{G} + \boldsymbol{\Sigma}^{-\mathbf{1}})^{-\mathbf{1}}$$

We introduce the mean and variance of the a priori expectations:

μ **- mean**
Σ **- covariance matrix**

Applications to nuclear structure models

Advantages of the proposed extensions:

- Allows to naturally include expectations about the model parameters in fit
- No need to fix some of the model parameters
- Stabilization of the fit
- Method of choice when a small set of a new data need to be added to the fit (?)

However, the method is subjective – as every Bayesian approach is subjective.

Possible applications include:

- All nuclear structure models where a priori knowledge about model parameters can be defined
- Refitting of existing parameterizations (?)
- Application to models in which constrains of parameters can be put: $p_1>0$ etc.
- The only regularization technique with natural extension to non-linear models

Examples

We consider typical spherical Skyrme-Hartree-Fock model with 11 parameters fitted to properties of doubly magic nuclei: ^{16}O, ^{40,48}Ca, ^{56}Ni, ^{90}Zr, ^{132}Sn, ^{208}Pb

- binding energies
- proton radii
- single particle energies

We are interested in properties of ^{100}Sn (interpolation) and superheavy Z=114, N=162 nucleus (extrapolation)

Choice of a priori values based on:
M. Dutra *et. al.* PRC **85** 035201 (2012)
Skyrme interaction and nuclear matter constraints

$\rho_c = 0.16 \pm 0.05$

$E^{NM}/A = -16.0 \pm 0.2$

$K^{NM} = 230 \pm 30$

$a = 32.5 \pm 2.5$

$L = 58 \pm 18$

$1/M_s = 1.18 \pm 0.07$

$1/M_v = 1.36 \pm 0.29$

$C^{\Delta\rho}_0 = -63 \pm 14$

$C^{\Delta\rho}_1 = -27 \pm 52$

$C^{\nabla J}_0 = -92 \pm 10$

$C^{\nabla J}_1 = -16 \pm 33$

We assume no correlations between model parameters.

Examples

Mean values of model parameters and their variance.

	ρ_c	E/A	K^{NM}	a	L	$1/M_s$	$1/M_v$	$C_{\Delta\rho}^{0}$	$C_{\Delta\rho}^{1}$	$C_{\nabla J}^{0}$	$C_{\nabla J}^{1}$
LM	0.157	-15.900	230	31.35	55	1.18	1.36	-65.2	-52	-89.7	8
LM Var	0.024	0.295	**Const**	7.64	83	**Const**	**Const**	5.95	165	8.5	99
Bayes all	0.158	-16.005	264	31.47	57	1.08	1.47	-57.1	-46	-83	-6
Bayes all Var	0.002	0.065	18	1.36	14	0.05	0.23	3.03	45	3.6	26
Bayes BR	0.159	-16.032	253	31.86	57	1.19	1.35	-63.0	-44	-86.1	-10
Bayes BR Var	0.002	0.045	12	0.90	9	0.03	0.16	2.1	29	2.6	17

Root Mean Squere errors for groups of observables

Model	b.e	radii	spe
LM	0.669	0.05	1.449
Bayes all	0.637	0.02	1.343
Bayes BR	0.385	0.02	1.499

Examples

Interpolation to ^{100}Sn.

Property	LM	all	Bayes BR	Bayes SPE
Proton radius	4.436 (0.075)	4.400 (0.011)	4.395 (0.009)	+4.448 (0.030)
Neutron radius	4.364 (0.073)	4.330 (0.010)	4.324 (0.008)	+4.375 (0.029)
B. E	-828.88 (2.37)	-828.98 (1.42)	-828.83 (1.34)	-778.45 (16.62)
P1g9/2	-2.983 (0.193)	-2.872 (0.100)	-2.970 (0.092)	-2.445 (0.217)
P2d5/2	2.328 (0.284)	2.461 (0.132)	2.600 (0.139)	2.284 (0.196)
N1g9/2	-16.645 (0.239)	-16.674 (0.106)	-16.795 (0.103)	-16.074 (0.255)
N2d5/2	-11.291 (0.133)	-11.331 (0.124)	-11.159 (0.137)	-11.308 (0.153)

Examples

Z=114 N=162 Property	LM	Bayes all	Bayes BR	Bayes SPE
Proton radius	6.130 (0.103)	6.075 (0.015)	6.068 (0.012)	6.109 (0.038)
Neutron radius	6.224 (0.118)	6.168 (0.019)	6.161 (0.016)	6.213 (0.041)
B. E	-1971.90 (6.04)	-1975.71 (2.97)	-1973.12 (2.68)	-1858.81 (37.57)
P2f7/2	-2.560 (0.388)	-2.497 (0.179)	-2.581 (0.172)	-2.589 (0.247)
P2f5/2	-0.780 (0.348)	-0.766 (0.175)	-0.820 (0.164)	-0.829 (0.262)
N1j15/2	-9.043 (0.461)	-9.104 (0.164)	-9.151 (0.149)	-8.524 (0.307)
N2g7/2	-7.443 (0.214)	-7.268 (0.143)	-7.307 (0.140)	-7.030 (0.212)

Comparison of variances:
Bayes all model

Property	Superheavy	^{100}Sn
Proton radius	0.015	0.011
Neutron radius	0.019	0.010
B. E	2.97	1.42

Conclusions

I have presented the method which aim is not only to define the predictive power of a model but also to improve it.

It is possible to include a priori expectations about model parameters in quantitative way and that this helps to improve the ill-posedness of a problem.

The method is applicable to non-linear problems.

Further work necessary to obtain new parameterizations with improved predictive power (carefull definition of priors, extension of a data set for predictive power validation).

Open questions:

Can we use this method to update parameterizations when new data are available ?
Why some groups of observables seems to decrease the predictive power and how to interpret this ?