Modeling Vector and Axial Nucleon Form Factors with Bayesian Neural Networks

Krzysztof M. Graczyk

Institute for Theoretical Physics
University of Wrocław
Poland

June 16, 2020

Problem

Axial and Vector Form Factors (FF) of the nucleon:

- quasielastic charged current and elastic neutral current νN and ν-nucleus scattering cross-sections parameterized by vector and axial nucleon form factors.
- information about electroweak structure of the nucleon hidden in FF
- empirical information from:
  - elastic electron scattering data
  - quesielastic ν scattering data
- nucleon form factors: real-valued functions (scalars) depending on Q^2

Task

- obtain the G_{Ep} (electric), G_{Mp} magnetic protons FF
- obtain F_A axial FF

Neural Networks and Bayesian Statistics

Problems

- the choice of FF parametrization affects the results of the analysis: prediction of the uncertainties etc.
- bias-variance trade-off: **too simple models under-fit the data**, **too complex models tend to over-fit the data**

Method: Bayesian Neural Networks (BNN):

- Bayesian approach naturally embodies Occam's razor: **penalizes too complex models and favor simpler approaches** hence good generalization abilities
- Form Factors parametrized by Feed Forward Neural Networks in Multi Layer Perceptron (MLP) configuration
- Bayesian framework for Neural Networks: following MacKay's

$\rightarrow$ Estimate of the uncertainties of the model predictions

$\rightarrow$ validation data set is not required

$\rightarrow$ Quantitative comparison of different models

Bayesian model

i) Consider data $\mathcal{D}$ and Neural Network, $\mathcal{N}(\{w_i\})$

ii) two conditional probabilities: prior and likelihood

iii) from Bayes theorem **Posterior**

$$\underbrace{\mathbf{P}(\{\mathbf{w_i}\}|\mathcal{D},\mathcal{N})}_{posterior} = \frac{\overbrace{P(\mathcal{D}|\{w_i\},\mathcal{N})}^{likelihood}\,\overbrace{P(\{w_i\}|\mathcal{N})}^{prior}}{\underbrace{P(\mathcal{D}|\mathcal{N})}_{evidence}} \quad (1)$$

- **Evidence** for model $\mathcal{N}$

$$P(\mathcal{N}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{N})P(\mathcal{N})}{P(\mathcal{D})} \sim P(\mathcal{D}|\mathcal{N})P(\mathcal{N}) \sim P(\mathcal{D}|\mathcal{N}) \quad (2)$$

so evidence ranks models

- likelihood

$$-2\ln\mathcal{P}(\mathcal{D}|\{w_i\}),\mathcal{N})\sim\chi^2_{ex}(\mathcal{D},\mathcal{N},\{w_i\})\qquad(3)$$

- prior

$$\mathcal{P}(\{w_i\}|\alpha,\mathcal{N})\quad\sim\quad\exp\left(-\frac{\alpha}{2}E_w\right),\quad E_w=\sum_{k=1}^{W}w_k^2\qquad(4)$$

  α regularizer (hyperparameter)

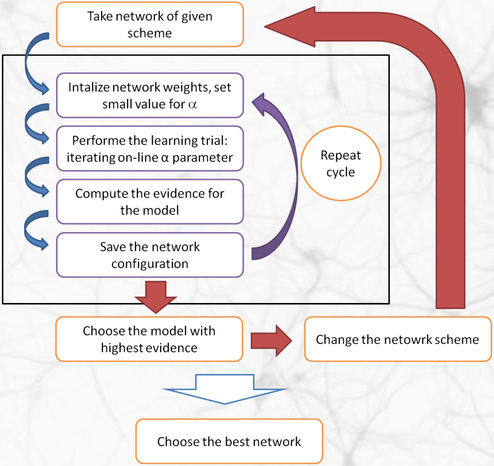
- posterior

$$-2\ln P(\{w_i\}|\mathcal{D},\mathcal{N},\alpha)\sim(\chi^2_{ex}(\mathcal{D},\{w_i\})+\alpha E_w)=\mathcal{E}(\mathcal{D},\{w_i\}),\qquad(5)$$

  $\mathcal{E}(\mathcal{D},\{w_j\})$ has a minimum at $\{w_j\}_{MP}$ and α_{MP}

- α_{MP} established during the training
- **1 step of inference:** the posterior is maximized
- **2 step of inference:** the evidence for each model is calculated to choose the best model!

Computation Scheme



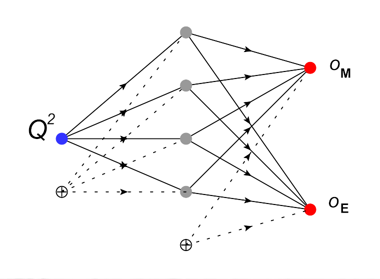
- every iteration step α's are changed (Hessian matrix is calculated and its eigenvalues)

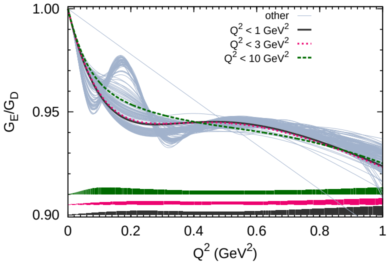
- C++ library developed by K.M. Graczyk and C. Juszczak
- user defined error function
- symbolic derivatives (but also numerical derivatives available)
- several optimization algorithms: gradient descent, QuickProp, RPROP (in several configurations, Adam, **Levenberg-Marquardt** (with approximate and exact Hessian)

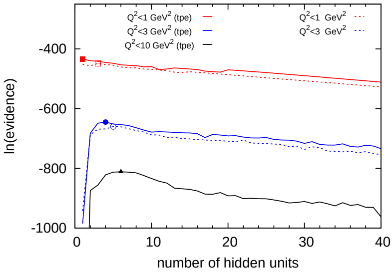
Electric and Magnetic Form Factors of the Proton: Graczyk and Juszczak, PRC90, 054334 (2014)

- 27 elastic ep scattering cross section data sets (27 normalization parameters in the fit)
- 15 polarization transfer data sets (ratios $\mu_p G_{Ep}/G_{Mp}$)

$$G_{Ep} = (1 - Q^2 o_E) G_D, \quad (6)$$
$$G_{Mp} = \mu_p (1 - Q^2 o_M) G_D, \quad (7)$$







Axial Form Factor

Alvarez-Ruso, Graczyk and Saul-Sala, PRC99 (2019), 025204

- **ANL** Bubble Chamber Data: $\nu_\mu + d \to \mu^- + p + p$ (PRD26, 537): distribution of events: $Q^2 \in (0.05, 2.5)$ GeV^2
- The least square-function:

$$\chi^2_{ex} = \chi^2_{\rm ANL} + \chi^2_{g_A} \tag{8}$$

where

$$\chi^2_{\rm ANL} = \sum_{i=k}^{n_{\rm ANL}} \frac{\left(N_i - pN_i^{th}\right)^2}{N_i} + \left(\frac{1-p}{\Delta p}\right)^2 \tag{9}$$

Δp – systematic uncertainty for $\# N$ of events

$$N_i^{th} = \int_0^\infty dE_\nu \frac{\frac{d\sigma}{dQ^2}(E_\nu, F_A)}{\sigma(E_\nu, F_A)} \frac{dN}{dE_\nu}, \quad \sigma(E_\nu, F_A) = \int_{min}^{max} \frac{d\sigma}{dQ^2}(E_\nu, F_A) dQ^2 \tag{10}$$

- $F_A(Q^2 = 0)$,

$$\chi^2_{g_A} = \left(\frac{F_A(0) - g_A}{\Delta g_A}\right)^2 \tag{11}$$

g_A and Δg_A from PDG

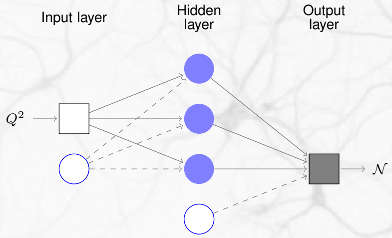
- Cross section modified by nuclear corrections for deuterium.

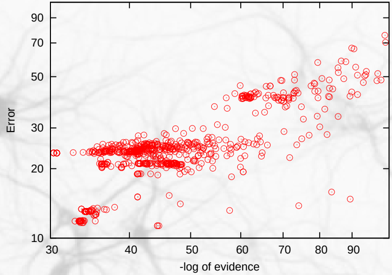
Analysis

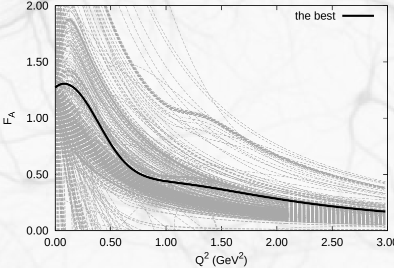
(i) BIN0: all ANL bins included

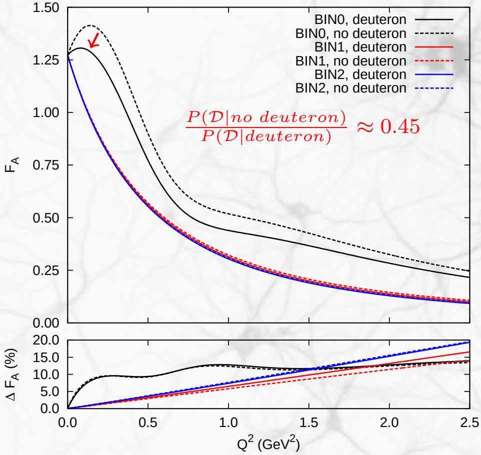
(ii) BINk: where $k = 1$ or $k = 2$: ANL bins without the first k bins

$$F_A(Q^2) = F_A^{\text{dipole}}(Q^2) \times \mathcal{N}_M(Q^2; \{w_i\}) \quad (12)$$



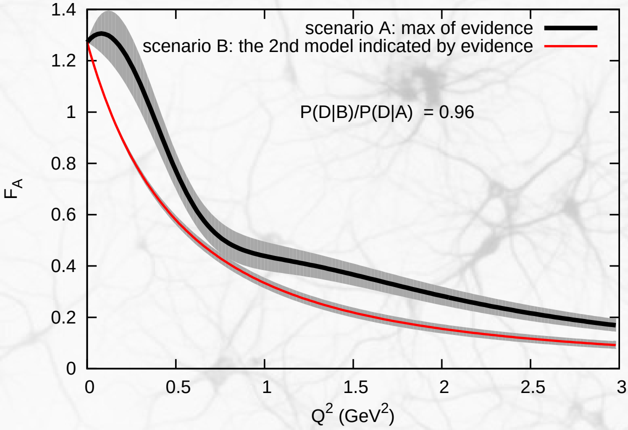






- For BIN0 The slope of F_A at $Q^2 = 0$ not consistent with other determinations...
- fits of BIN1 and BIN2 data consistent with original ANL analysis
- deuteron correction important for the first bin
- Possible explanation:
  - a low quality of the measurements at low-Q^2 due to low and not well understood efficiency
  - an improper description of the nuclear corrections
  - the actual value of the slope $dF_A(Q^2 = 0)/dQ^2$ might not be properly estimated because of the lack of very low-Q^2 data

two competitive scenarios: on the edge of dipole...



- small effects are important
- there is a tension between the first bin and constraint for F_A at $Q^2 = 0$
- assumption: $dF_A(Q^2 = 0)/dQ^2 < 0 \rightarrow$ the fit in agreement with dipole shape...

Final Remarks

- First Bayesian analyses of the electron-proton and neutrino-deuteron scattering data presented.

The method allows to:

- analyze small data sets
- reduce the model dependence of final results
- compare quantitatively different models

* Calculations done in WrocÂław Centre for Networking and Supercomputing, Grant No. 268 (http://www.wcss.wroc.pl)