Model Selection for Pion Photoproduction

R. Molina[1], J. Landay[1], C. Fernandez Ramirez[2], M. Doring[1,3] and B. Hu[1]

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

VNIVERSIDAD NACIONAL AVTONOMA DE MEXICO

Jefferson Lab

Contents

Analysis of photoproduction reactions

- HBChPT, RBChPT, low-energy neutral pion photoproduction, $E_\gamma \leq 170$ MeV. Bernard,Kaiser, Meißner, Scherer, Tiator
- Polynomial parametrizations which incorporate unitarity in the S-wave, $E_\gamma \sim 185$ MeV. Hornidge, C. Férnandez-Ramírez, Bernstein
- ChPT calculations including isosping breaking. Varu, Hanhart, Hoferichter, Nogga, Kubis,Nogga
- RBChPT with $\Delta(1232)$, $E_\gamma \sim 200$ MeV. Hiller Blin, Ledwig, Vicente Vacas
- Effective field theories, K-matrix parametrizations, Regge parametrizations, higher energies. Anisovich,Drechsel, Kamalov, Tiator, Haidenbauer, Krewald, Meißner
- Phenomenological parametrizations, basic principles of S matrix, unitarity, correct threshold behavior, Fermi-Watson Theorem, SAID approach. Workman, Paris, Briscoe, Strakovsky

Shrinkable methods: Ridge and Lasso

Linear model:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \epsilon \tag{1}$$

$$\beta_j = \frac{\partial Y}{\partial X_i}; \qquad \text{Penalty} \quad \lambda \sum_{j=1}^{p} f(\beta_j) \tag{2}$$

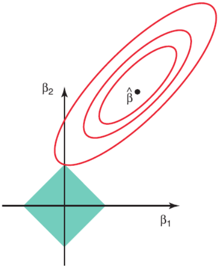
n observations, (x_{ij}, y_i). Find $\hat{\beta}_i^{\lambda}$ which minimize,

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} \beta_j^2 \qquad \text{RIDGE}$$

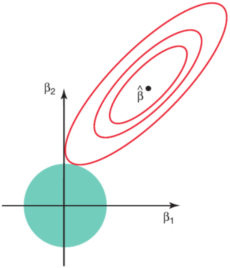
$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} |\beta| \qquad \text{LASSO}$$

Schrinkable methods: Ridge and Lasso

$$\text{minimize}_\beta \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \text{ subject to } \sum_{j=1}^{p} f(\beta_j) \leq s$$



LASSO



RIDGE

Bayesian Interpretation

Linear model:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \epsilon \tag{3}$$

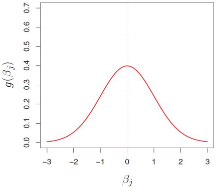
Assume

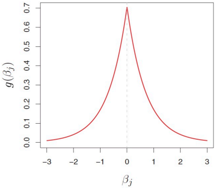
$\epsilon \sim N(\mu, \sigma)$, $p(\beta) = \Pi_{j=1}^{p} g(\beta_j)$, with $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$

$$p(\beta|X, Y) \propto f(Y|X, \beta)p(\beta|X) = f(Y|X, \beta)p(\beta) \tag{4}$$

It follows

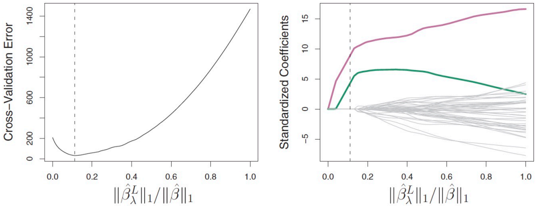
Ridge $g(\beta_j) \sim$ gaussian **LASSO** $g(\beta_j) \sim$ double exponential





How to fix λ?

1. Discretize λ
2. Split the data set into $k \sim 5$ (or 10) parts
3. In every iteration, set 4 parts to be the Training set, 1 part the Validation set.
4. Minimize χ^2_T
5. Evaluate Cross Validation Error (CV), $CV = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i)$ or χ^2_V for every i part and average it for every λ.



Information Criteria

Nested models $\{M_i\}$ with i parameters, $M_1 \prec M_2 \prec .... \prec M_k$

Example

$$M_1 : y = \beta_0 + \beta_1 x_1 + \epsilon$$
$$M_2 : y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$M_1 \prec M_2$. How to *evaluate* the models given a set of data?

Training error, $\sum_{i=1}^{n}(y - \hat{y})^2$, χ^2_T? but.. size of test error?.

n, number of observations; k, number of parameters;

$$\textbf{AIC} = -2 \max \log(L(\hat{\theta}|data)) + 2k = \chi^2 + 2k$$
$$\textbf{AIC}_c = AIC + \frac{2k(k+1)}{n-k-1}$$
$$\textbf{BIC} = -2 \max \log(L(\hat{\theta}|data)) + 2log(n) = \chi^2 + k\, log(n)$$

All of them penalize overfitting

LASSO, cross validation, information criteria

Scan a large set of models and compare them.

Advantages

- Ability to recover the true model (least squares needs a large amount of data)
- Minimal predicted errors
- Good variance-bias trade-off
- Reduces the correlations between the parameters
- Select “simple” models

Parametrization for $\gamma p \to \pi^0 p$

For $\mathcal{M}$ the electric $E_{L\pm}$, and magnetic $M_{L\pm}$ multipoles,

$$\mathrm{Re}, \mathrm{Im}\, \mathcal{M}_{L\pm} = \frac{q_{\pi^0}^{l}}{m_{\pi^+}^{l+1}} \sum_{i=0}^{i_{\max}} \frac{a_i}{10^{-i}} \left(\frac{\omega_{\pi^0} - m_{\pi^0}}{m_{\pi^+}} \right)^i \qquad (5)$$

q_{π^0} c. m. momentum, ω_{π^0} energy, a_i fitting parameters

$$E_{1^+} = \frac{1}{6}(P_1 + P_2);\, M_{1^+} = \frac{1}{6}(P_1 - P_2 + 2\,P_3);\, M_{1^-} = \frac{1}{3}(P_3 + P_2 - P_1) \qquad (6)$$

Partial waves Includes P and D waves, and

$$\Delta E_{0^+} = i\,\frac{q_{\pi^+}}{m_{\pi^+}^2} \sum_{i=0}^{2} \frac{a_i}{10^{-i}} \left(\frac{q_{\pi^+}}{m_\pi^+} \right)^{2i} \qquad (7)$$

supplements the S wave multipole to take into account the $\pi^+ n$ threshold cusp. *Total 46 parameters.*

Parametrization for $\gamma p \to \pi^0 p$

Differential cross section and asymmetry

$$\frac{d\sigma}{d\Omega}(s,\theta) = \frac{q_{\pi^0}}{k_\gamma} W_T(s,\theta)$$

$$\Sigma(s,\theta) \equiv \frac{\sigma_\perp - \sigma_\parallel}{\sigma_\perp + \sigma_\parallel} = -\frac{W_S(s,\theta)}{W_T(s,\theta)} \sin^2\theta \qquad (8)$$

Electromagnetic responses, W_T and W_S:

$$W_T = T_0(s) + T_1(s)\mathcal{P}_1(\theta) + T_2(s)\mathcal{P}_2(\theta) + \ldots$$

$$W_S = S_0(s) + S_1(s)\mathcal{P}_1(\theta) + \ldots \qquad (9)$$

C. Fernández Ramírez and A. M. Bernstein, PLB724, PRC80; Legendre Polynomials $\mathcal{P}_j(\theta)$, and

$$T_n(s) = \sum_{ij} \text{Re}\left\{\mathcal{M}_i^*(s) T_n^{ij} \mathcal{M}_j(s)\right\} \qquad S_s(s) = \sum_{ij} \text{Re}\left\{\mathcal{M}_i^*(s) S_n^{ij} \mathcal{M}_j(s)\right\}$$

being $\mathcal{M}_j(s) = E_{0+}, E_{1+}, E_{2+}, E_{2-}, M_{1+}, M_{1-}, M_{2+}, M_{2-}$

LASSO

$$\chi_T^2(\lambda) = \chi^2(\lambda) + \lambda^4 \sum_{i=1}^{i_{\max}} |a_i| \qquad (10)$$

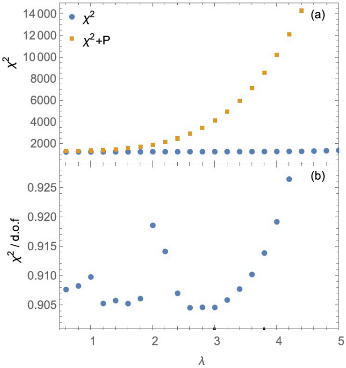
Test: Lasso in a benchmark model

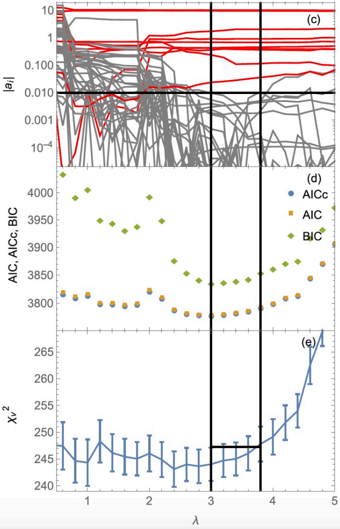
$\mathcal{B}_0$ model synthetic data

9 parameters, no D waves

D. Hornidge, PRL111 (2013)

Analysis with the 46 parameter model





Test: Lasso in a benchmark model

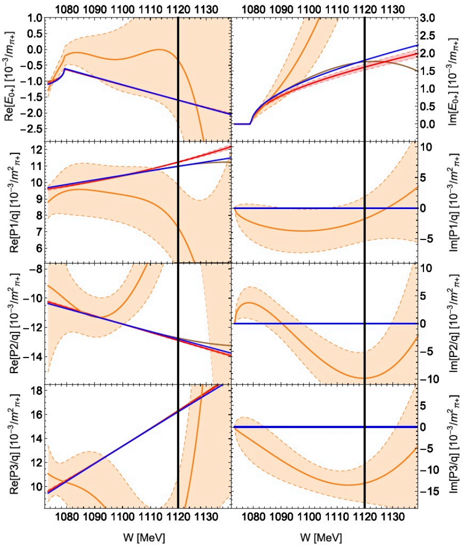
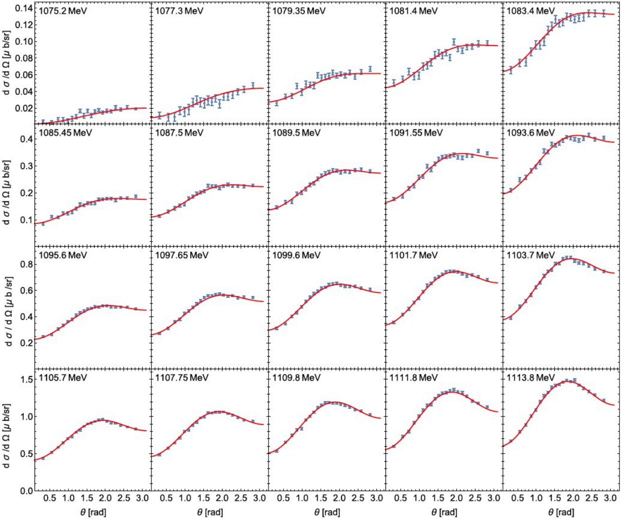


Figure. *S* and *P* partial waves. Red curves and bands indicates the Lasso solution. Blue lines stand for the benchmark solution. Orange curve and bands are the solution for $\lambda = 0$. The solution with $\lambda = 3$ is indicated with brown color.

Lasso in a $\mathcal{B}_0$ model: Differential cross section



Lasso in a $\mathcal{B}_0$ model: Differential cross section

Result:

The analysis with LASSO, cross validation and information criteria, recovers the true model with only one additional parameter, reducing the number of parameters from 46 to 10.

F test

Compare two models

- Model 1, k parameters
- Model 2, m+k parameters

If the true values of the m extra parameters vanish,

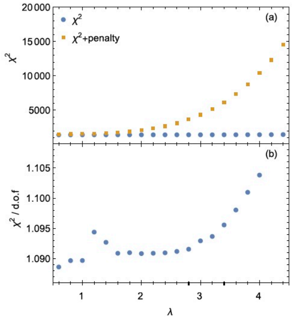
$$y = \frac{(\chi_1^2 - \chi_2^2)/k}{\chi_2^2/(n-m-k)} \tag{11}$$

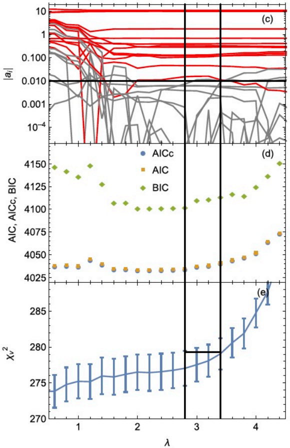
is $F(k, n-m-k)$ distributed. A value of y beyond a chosen CL limit indicates that the more complex model 2 is significantly better than 1. **We find $y = 1.64$, below the 90% CL interval ending at $y = 2.63$, the overfit is not significantly better than the simplest fit.**

Real data for the $\gamma p \to \pi^0 p$ reaction

MAMI $d\sigma/d\Omega$, Σ, $d\sigma_T/d\Omega$.

D. Hornidge et al., PRL111,062004 (2013), S. Schumann et al., PLB750, 252 (2015)





$\gamma p \to \pi^0 p$: partial waves, beam asymmetry

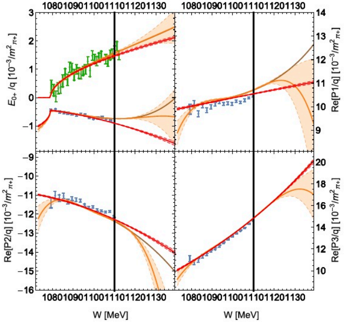


Figure. *S* and *P* partial waves. Red curve with bands indicates the Lasso solution with bootstrap.

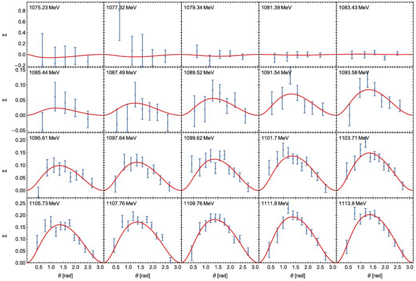
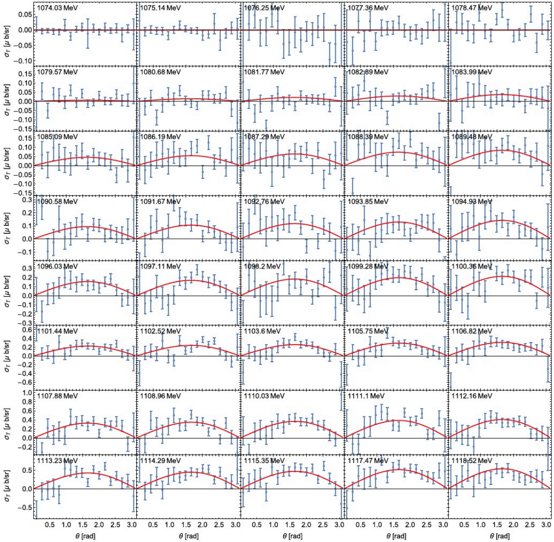


Figure. Beam Asymmetry

$\gamma p \to \pi^0 p$: Polarized differential cross section $\sigma_T = d\sigma/d\Omega T$



β_0, Cusp parameter at threshold

$$\beta_0 = (2.41 \pm 0.05) \times 10^{-3} m_{\pi^+}^{-1}$$

Agrees with the value from, S. Schumann et al., PLB750,252,
$\beta_0 = (2.2 \pm 0.2[\textit{stat.}] \pm 0.6[\textit{syst.}]) \times 10^{-3} m_{\pi^+}^{-1}$.
(A2 Collaboration, 2015)

Conclusions

- LASSO, in combination with cross validation and criteria from information theory provides a tool to scan large classes of models, selecting the *simplest* model with minimum number of parameters and prediction error
- It has a wide range of applicability, being a promissing tool for the analysis of data of the excited baryon and meson spectra in future experiments