Parameter selection in an Isobar model for $K^+\Sigma^-$ photoproduction. A *data-driven approach.*

Dimitrios Petrellis

Nuclear Physics Institute, Czech Academy of Sciences

Outline

- 1. Motivation
- 2. The Isobar model
  - features of the model
  - parameters and observables
- 3. Fitting procedure
  - problems with ordinary χ^2
  - regularized χ^2
  - information criteria
- 4. Numerical results

Photoproduction of Kaons and Hyperons off Nucleons

$$\gamma + p \to K^+ + \Lambda$$

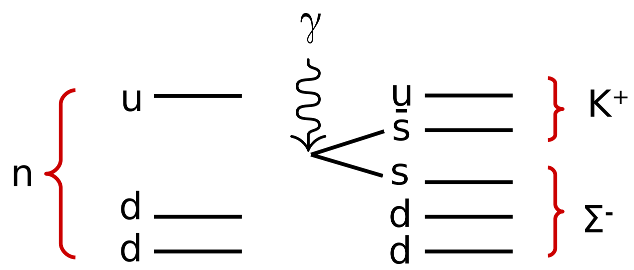
$$\gamma + p \to K^+ + \Sigma^0$$

$$\gamma + p \to K^0 + \Sigma^+$$

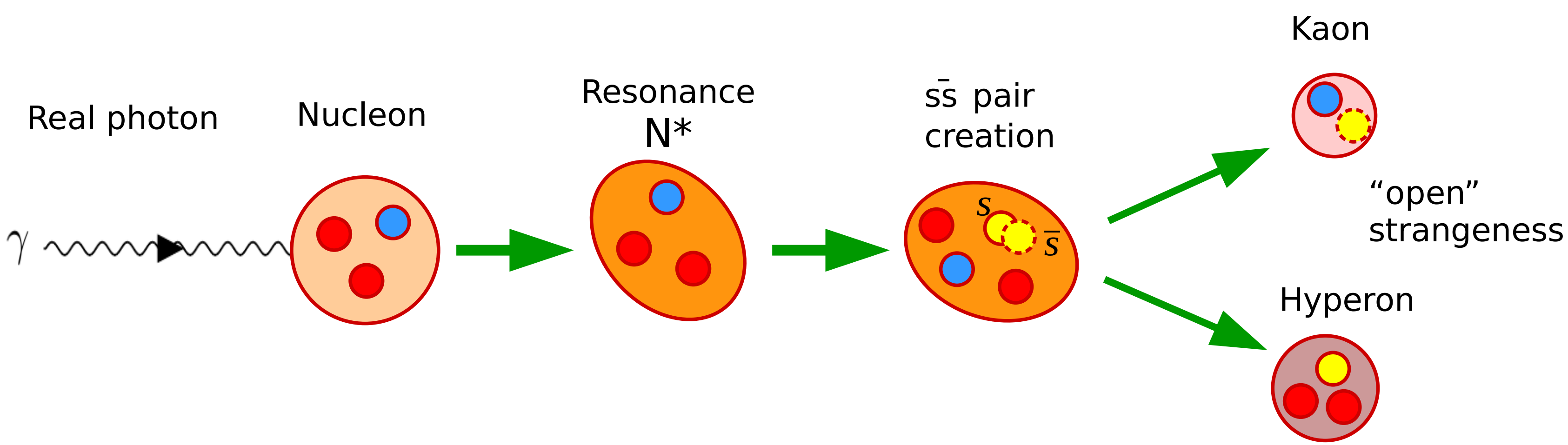
$$\gamma + n \to K^0 + \Lambda$$

$$\gamma + n \to K^0 + \Sigma^0$$

$$\gamma + n \to K^+ + \Sigma^-$$ [1, 2]



1. P. Bydzovsky et al., Phys. Rev. C 104, 065202 (2021), present work
2. N. Zachariou et al., Phys. Lett. B 827, 136985 (2022), new data from CLAS



3. Figure adapted from: L. De Cruz, PhD Thesis, Ghent University 2012.

1. Motivation

Motivation

- Why KΣ photoproduction?
  - Quark models predict more resonances than observed in π-N scattering experiments ("missing" resonance problem). Indications that these states may couple to KY (Y = Λ, Σ) channels
  - E/M interaction very well understood
  - Studied at several facilities: CEBAF, MAMI, ELSA, Spring-8, GRAAL
  - New data on photon beam asymmetries from CLAS
- Isobar model: phenomenological models useful in bridging the gap between fundamental theory and experiment

Motivation

- No single resonance dominates in $E^{lab}_{\gamma} \approx$ 1-2 GeV, but many (>20), broad and overlapping

  - extremely large number of possible combinations (**models**)

  - large number of parameters → ordinary χ^2 fitting: problematic similar minima, large *variations* in the parameter values

  - *Regularized* χ^2 fitting [4, 5] → penalty term constrains the number and magnitude of the parameters

    - improves the quality of the fits

    - + information criteria → selects the best subset of parameters (**model**)→ resonances evaluated as most "necessary" by the data

[4.] J. Landay et al., Phys. Rev. C 95, 015203 (2017)
[5.] J. Landay et al., Phys. Rev. D 99, 016001 (2019)

2. The Isobar model

General features of Isobar models

- interactions described by means of effective Lagrangians
    - effective degrees of freedom: hadrons

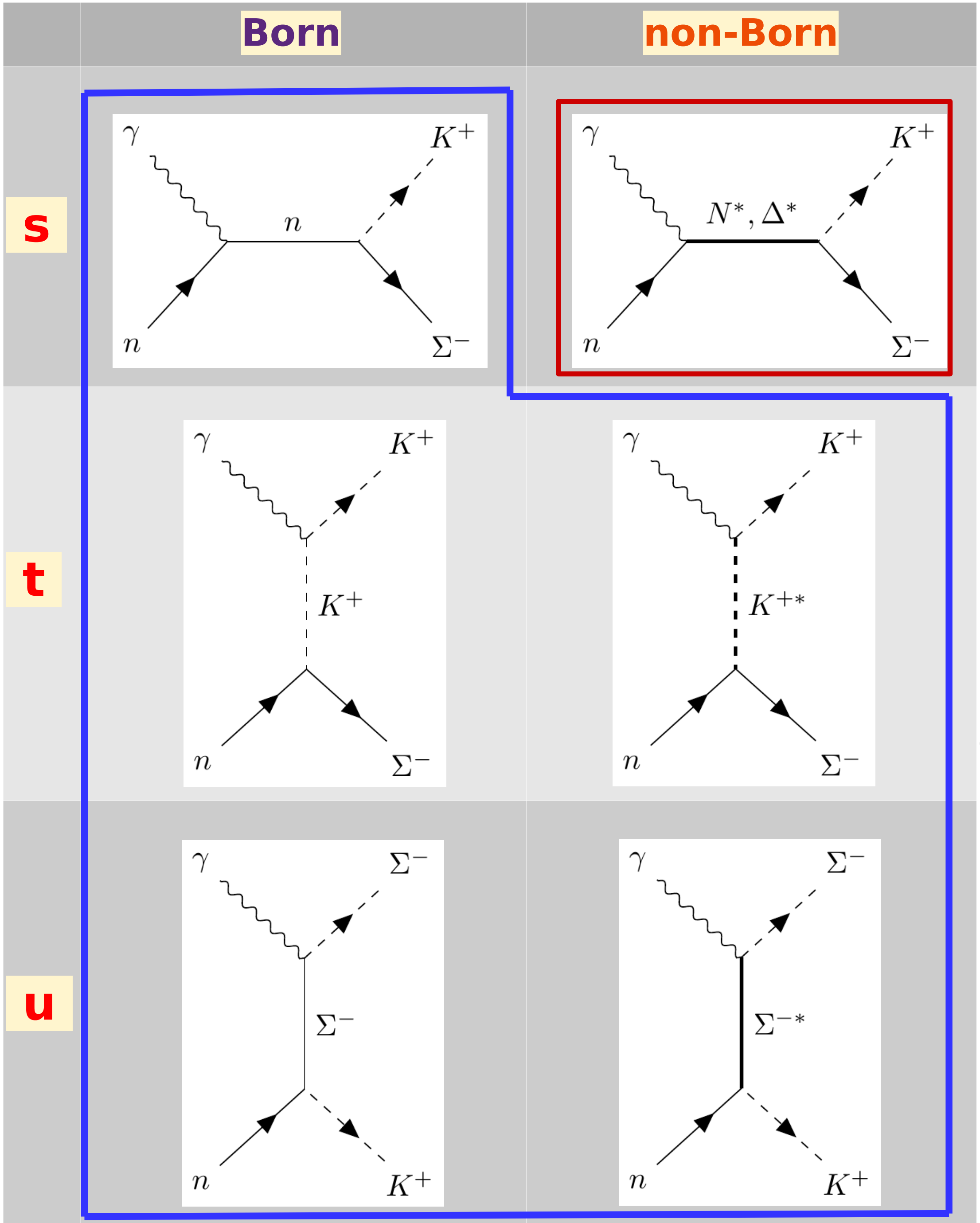
- amplitude = sum of tree-level Feynman diagrams
    - s-, t-, u- channels: exchange of nucleon, kaon, hyperon
    - intermediate state: ground state hadron (Born), resonance (non-Born)

- single-channel: intermediate channels (rescattering in final states) not taken into account

- Saclay-Lyon, MAID & Kaon-MAID, Gent, BS1,2,3[6,7] models

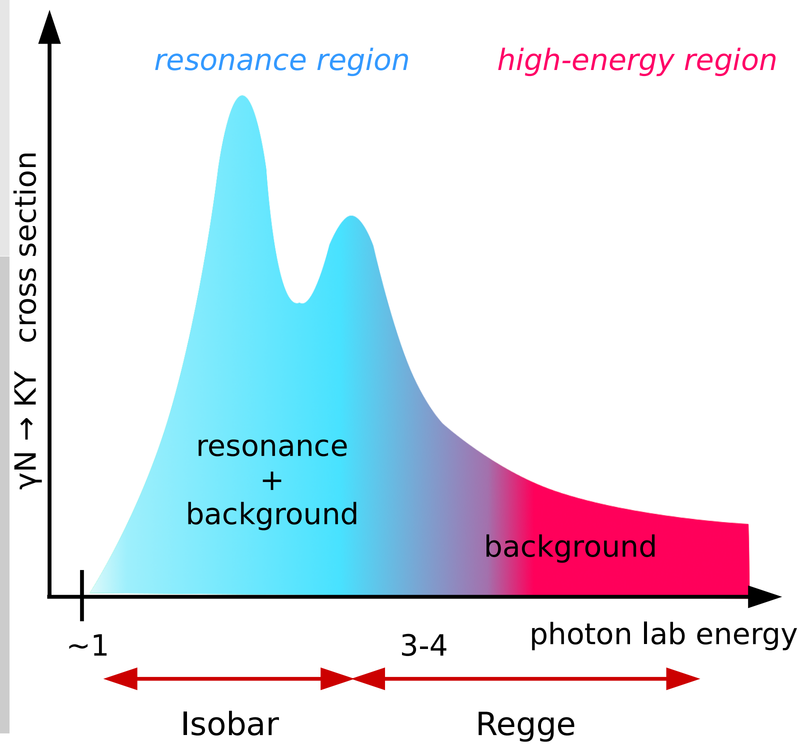
6. D. Skoupil and P. Bydzovsky, Phys. Rev. C 93, 025204 (2016)
7. D. Skoupil and P. Bydzovsky, Phys. Rev. D 97, 025202 (2018)

Tree-level contributions to $n(\gamma,K^+)\Sigma^-$



the only resonant diagrams

rest of diagrams
→ background



Specific features of Isobar models

Hadronic form factors

- hadron internal structure
- mitigate Born terms’ contribution to cross sections

$$F_d = \frac{\Lambda_h^4}{\Lambda_h^4 + \left(x - m_h^2\right)^2}$$

Λ_h	cutoff	of intermediate hadron h
x	4-momentum^2	
m_h	mass	

Decay widths

- finite lifetime of resonances
- decay widths introduced by hand in propagator denominators

$$\mathcal{P} \sim \frac{1}{q^2 - m^2} \qquad q^2 = s \qquad s - m_R^2 \rightarrow s - m_R^2 + i\, m^2\, \Gamma_R$$

Parameters and observables

Resonances

masses, widths: from PDG

Parameters to fit

($g_{K\Sigma n}$)

coupling constants of resonances

(= products of E/M and strong c.c.)

hadron form factor cutoffs

674 data points from: CLAS, LEPS

Observables

differential cross sections

photon beam asymmetries

Tag	Resonance	Mass (MeV)	Width (MeV)
K*	$K^*(892)$	891.7	50.8
K1	$K_1(1270)$	1270	90
N3	$N(1535)\ 1/2^-$	1530	150
N4	$N(1650)\ 1/2^-$	1650	125
N8	$N(1675)\ 5/2^-$	1675	145
N6	$N(1710)\ 1/2^+$	1710	140
N7	$N(1720)\ 3/2^+$	1720	250
P4	$N(1875)\ 3/2^-$	1875	200
P1	$N(1880)\ 1/2^+$	1880	300
Mx	$N(1895)\ 1/2^-$	1895	120
P2	$N(1900)\ 3/2^+$	1920	200
M4	$N(2060)\ 5/2^-$	2100	400
M1	$N(2120)\ 3/2^-$	2120	300
D1	$\Delta(1900)\ 1/2^-$	1860	250

Minimization with: MINUIT Library

Isobar code available at:

http://www.ujf.cas.cz/en/departments/department-of-theoretical-physics/isobar-model.html

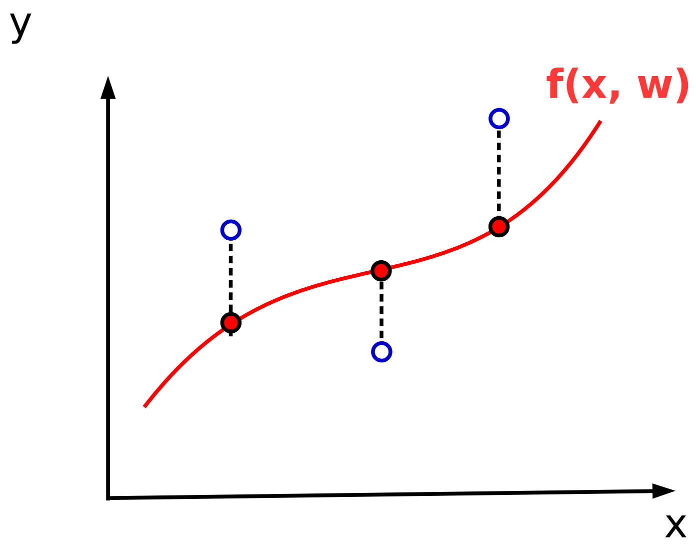
3. Fitting procedure

Ordinary Least Squares fitting

- set of data: pairs observations

$$D = \{(x_1, y_1), (x_2, y_2), \ldots (x_N, y_N)\}$$

e.g. $x = E, \quad y = d\sigma / d\Omega$



- Function: $f(x, \mathbf{w})$

- Parameters: $\mathbf{w} = (w_0, w_1, \ldots w_K)$

- Goal: determine values of the parameters **w*** that minimize some error function

$$E = \sum_{i=1}^{N} [y_i - f(x_i, \mathbf{w})]^2 \qquad \chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i, \mathbf{w})}{\sigma_i} \right]^2$$

[this approach,however, is problematic → we may overfit the data]

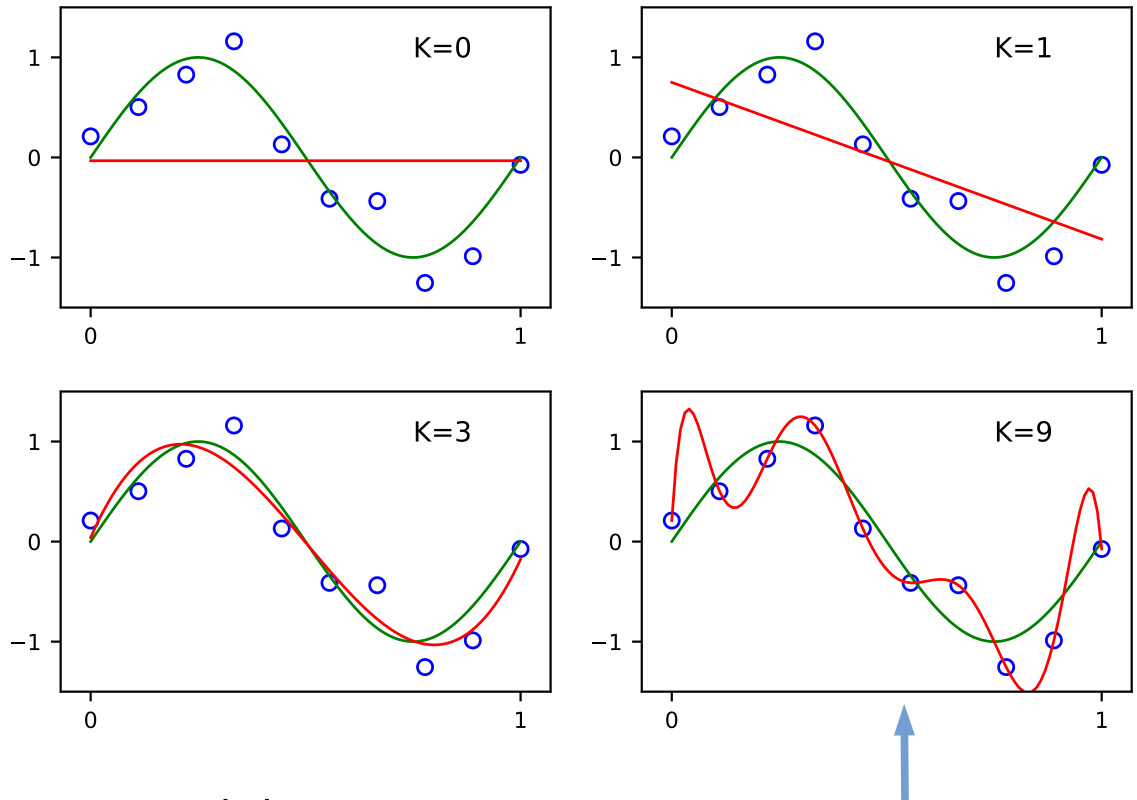
The problem of overfitting though an example

Create artificial data by adding Gaussian noise

$$y = \sin(x) + \epsilon$$

fit the data with a polynomial

$$f_K(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_K x^K$$



increasing order **K** of polynomial

- fits the data very well, but
- poor description of the function that generated them

Model fits the noise in the sample

Where do we stop?
What is the optimal complexity of our model?
Error minimization alone, does not guarantee the quality of the fitting

{ Occam's razor
Law of parsimony }

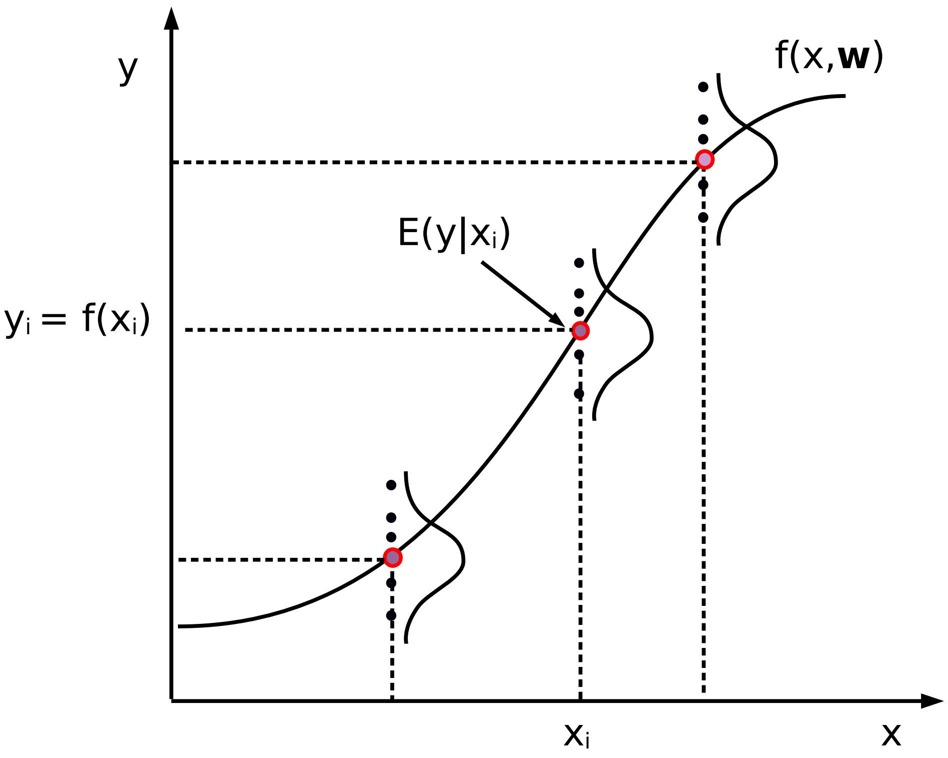
8. Figure adapted from:C. Bishop Pattern Recognition and Machine Learning, Springer 2006

Likelihood function

- each measurement is characterized by **uncertainty** σ_i

- assume: y has a Gaussian distribution around some **mean** μ

- we want our model f(x, w) to estimate that mean



$$p(y_i|x_i,\mathbf{w},\sigma_i)=\mathcal{N}(\mu=f(x_i,\mathbf{w}),\sigma_i^2)$$

- for N *independent, identically distributed* observations:

$$\boldsymbol{X}=\{x_1,x_2,\ldots x_N\},\quad \boldsymbol{Y}=\{y_1,y_2,\ldots y_N\}$$

probability of the whole set **Y** of observations:

$$p(Y|X,\mathbf{w})=\prod_{i=1}^{N}p(y_i|x_i,\mathbf{w},\sigma_i) \longrightarrow L(\mathbf{w})\equiv p(Y|X,\mathbf{w})$$

Likelihood function

Maximizing the log-likelihood

- under the Normality assumption*

$$L(\boldsymbol{w}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - f(x_i, \boldsymbol{w}))^2}{2\sigma_i^2}}$$

- since $\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i, \boldsymbol{w})}{\sigma_i} \right]^2$ $\Rightarrow$ Likelihood: $L(\boldsymbol{w}) \propto e^{-\chi^2}$

- taking the logarithm: $\ln(L(\boldsymbol{w})) \sim -\chi^2$

Maximizing the log-likelihood* is equivalent to minimizing χ^2
=> equally prone to overfitting

Regularization: a remedy for over-fitting

$$\textbf{min}\ \left(\chi^2 + \lambda \sum_{j=1}^{K} |w_j|^q\right)$$

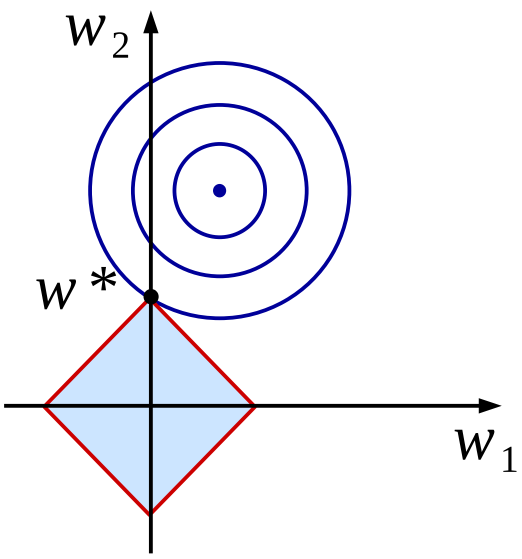
- introduction in χ^2 of a term that penalizes large values of the parameters w_j

- ~ minimize χ^2,subject to constraint:

$$\sum_{j=1}^{K} |w_j|^q \leqslant \eta$$

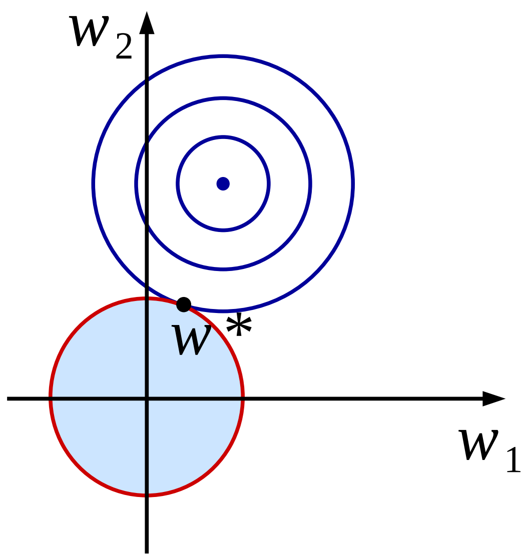
LASSO*

$q = 1 \rightarrow$ L1 norm



Ridge

$q = 2 \rightarrow$ L2 norm



w* = optimum value for **w** under the constraint

- for q = 1 (LASSO) $\rightarrow$ some parameters become zero ($w^*_1 = 0$)

* **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

LASSO for variable selection

- LASSO forces some of the parameters to zero → selects a subset

- λ, regularization parameter → strength of the penalty term
  smaller λ → more complex model

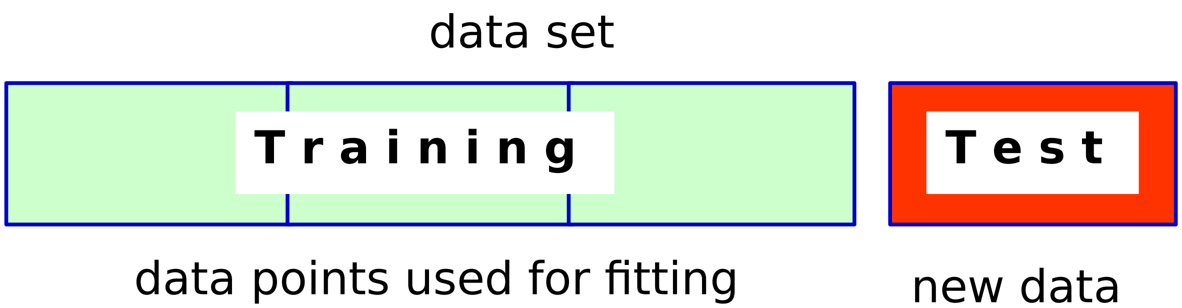
- λ: controls how many parameters are switched-off and how many remain
  → *λ practically selects a model*

- instead of taking a huge number of combinations of parameters,
  run LASSO with several λ_1, λ_2,... values and choose the optimal λ based on:

  either
  - **Validation**

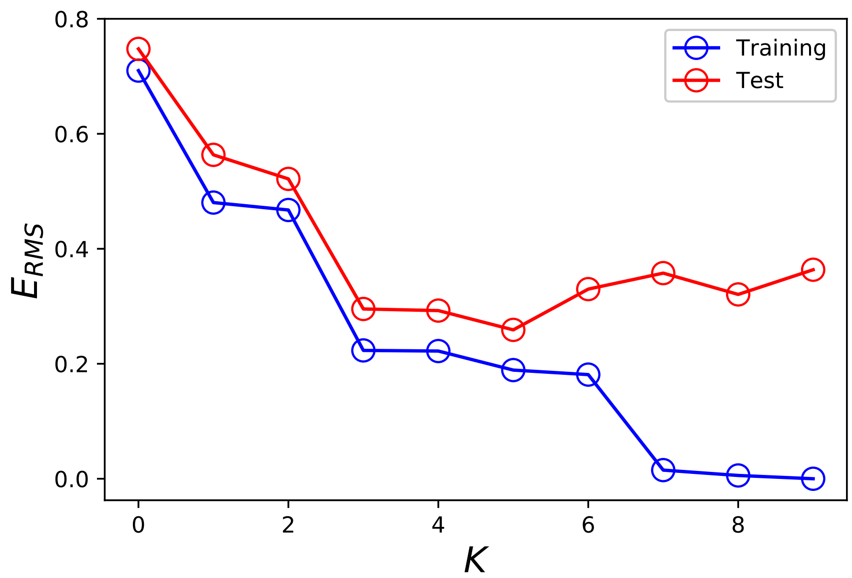
  or
  - **Information criteria**
    - Akaike Information Criterion (AIC)
    - Bayesian Information Criterion (BIC)

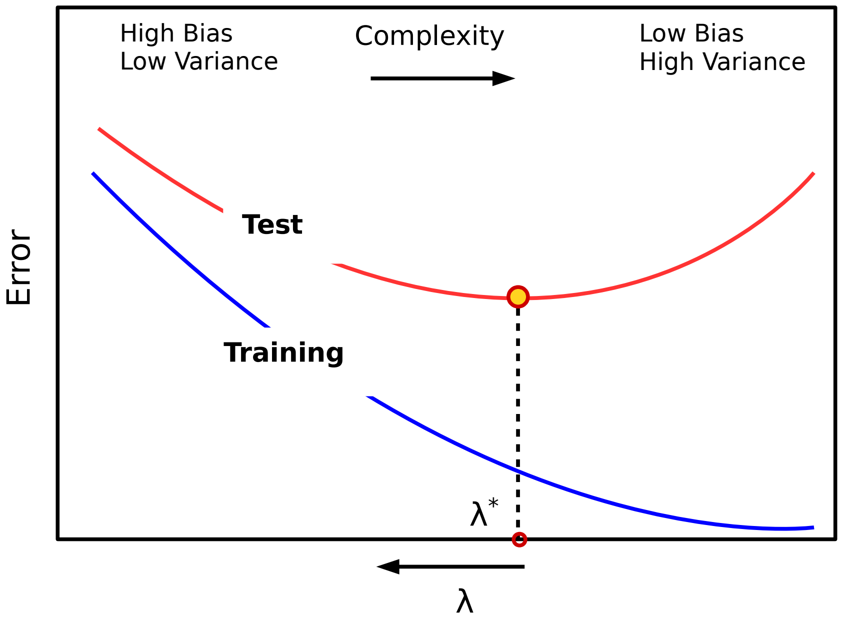
Validation: Training & Test set errors



- Fit model on the training set → Training Error
- Test the fitted model on the test set → Test Error
- Repeat while increasing complexity
  (Forward selection)

Bias-Variance trade-off

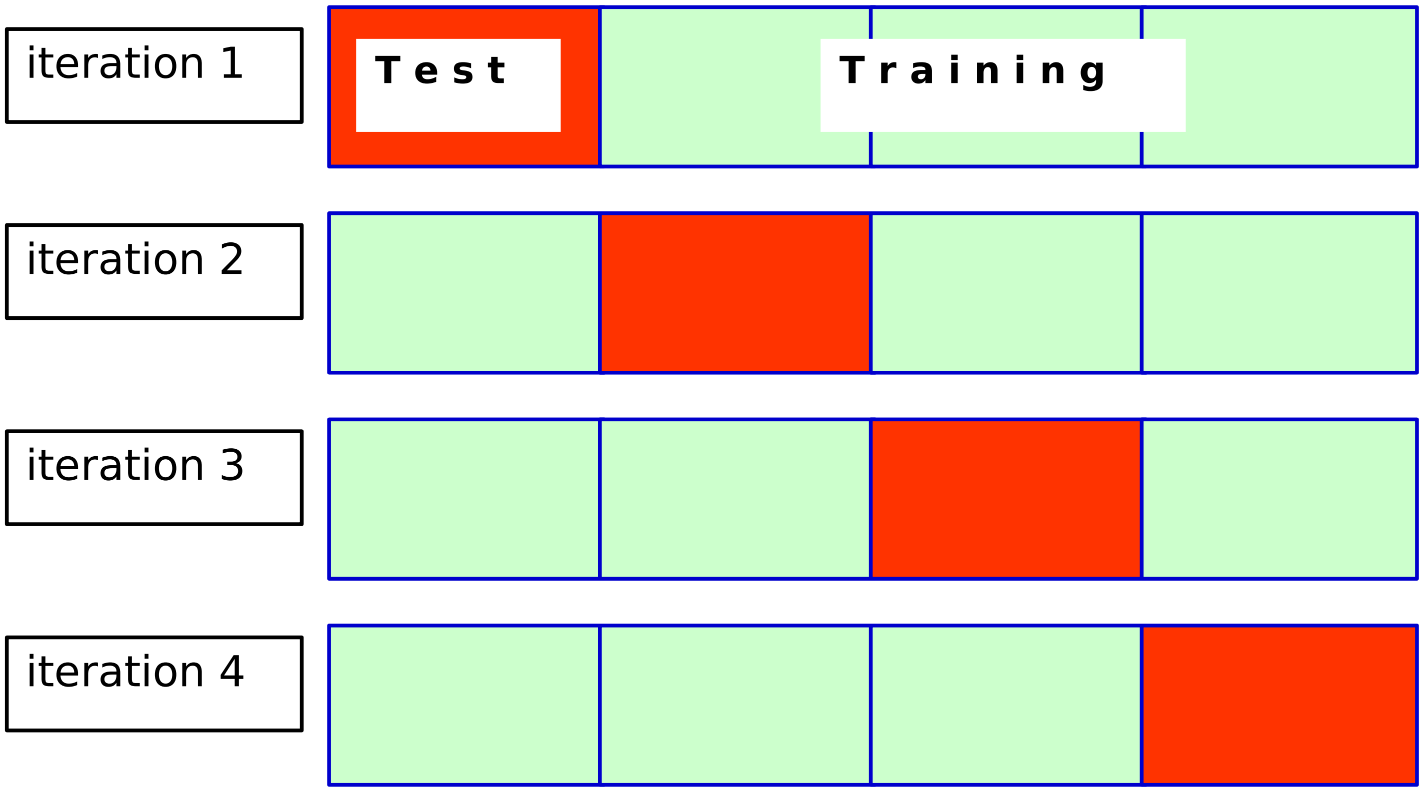




Cross-validation

- to avoid selection bias in the choice of Training / Test sets

e.g. 4-fold
cross-validation



in general: **n**-fold cross-validation → average over **n** runs

- drawback: it’s computationally costly

Information Criteria (IC)

Approach equivalent to validation

For a series of models *i* = 1, 2, ...m

Akaike IC[9]: $\text{AIC} = \chi^2_{\text{min}} + 2k_i$

Bayesian IC[10]: $\text{BIC} = \chi^2_{\text{min}} + k_i \ln(N)$

k_i : number of parameters corresponding to model i

N : number of data points

Choose the model with the minimum AIC, BIC

[both AIC and BIC give similar results, although BIC tends to penalize complexity more]

In the case of LASSO: model i → λ_i ⟹ Choose λ_i that results in the minimum IC

9. Akaike, IEEE Transactions on Automatic Control, 19 (6) 716 (1974)

10. G. Schwarz, Ann. Stat. 6(2), 461 (1978)

Bayesian approach: data fitting through posterior maximization

Likelihood   Prior

Posterior probability:
how probable **w** is,
given the data D

$$P(\mathbf{w}|D)=\frac{P(D|\mathbf{w})P(\mathbf{w})}{P(D)}$$

Evidence

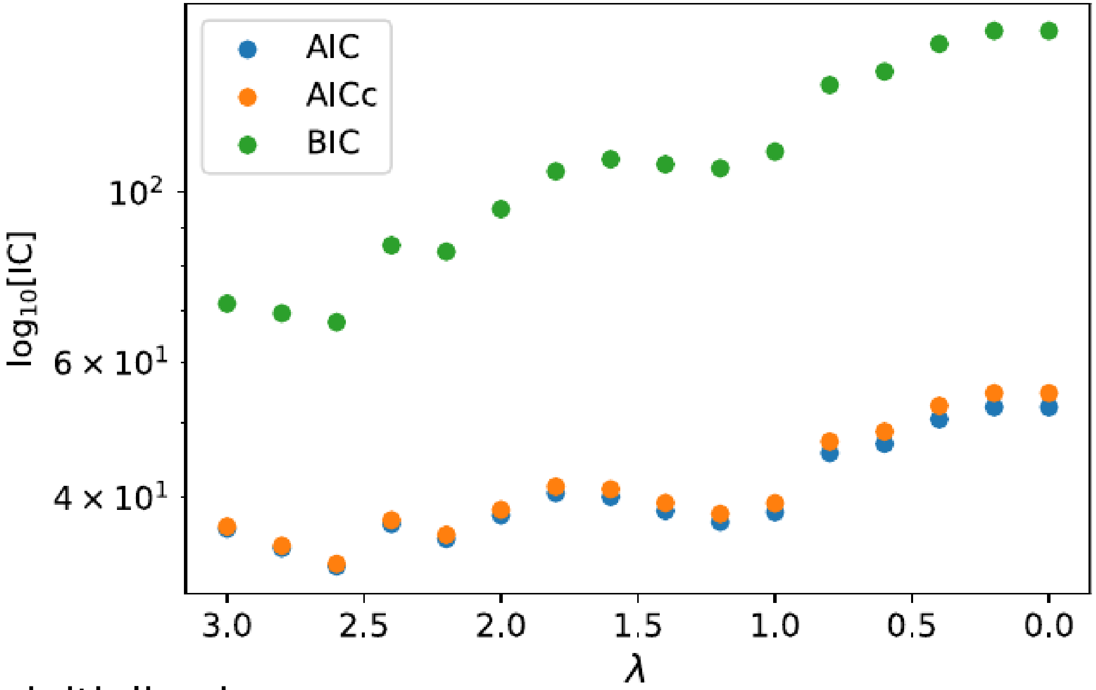
$$P(D)=\int P(D|\mathbf{w})P(\mathbf{w})d\mathbf{w}$$ ~ Normalization factor

Maximum Posterior

- in a fully Bayesian treatment we seek to maximize the Posterior P(**w**|D), instead of the Likelihood
- determine the most probable value of parameters $\mathbf{w}_{MP}$, given the data - no need for test runs
- is equivalent to minimizing regularized sum-of-squares error - Occam's principle automatically incorporated
- but computationally costly

4. Numerical results

Applying the Information Criteria



Forward selection:

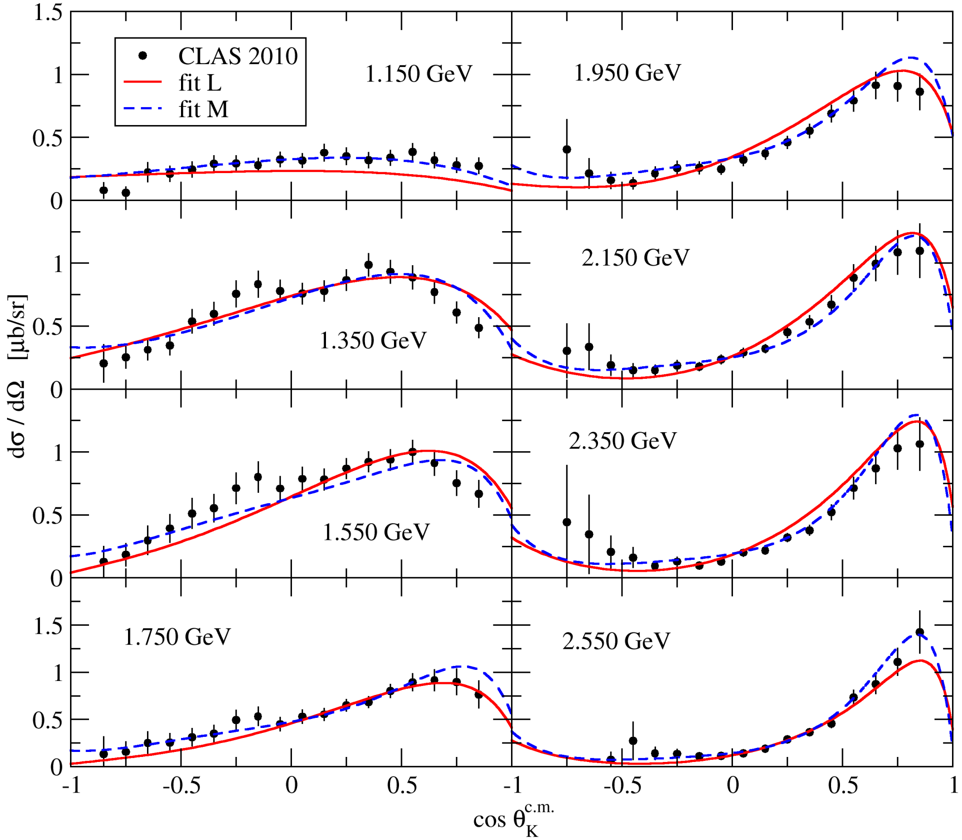
- start with the full model, all parameters initialized with random values and use some λ_{max}
- perform LASSO χ^2 minimization and compute AIC, BIC
- in each run progressively decrease λ and rerun LASSO using the fitted parameter values of the last run as starting values
- repeat until λ_{min} is reached
- optimal λ occurs at the minimum of BIC, AIC

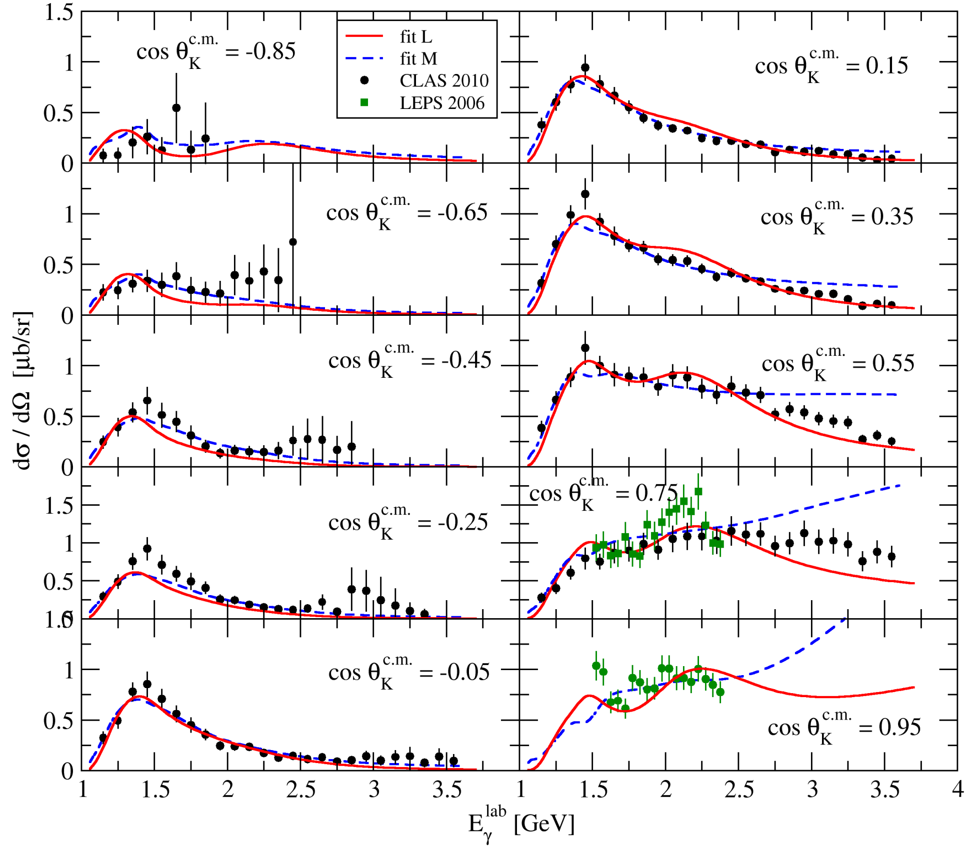
Reduction in the number pf parameters

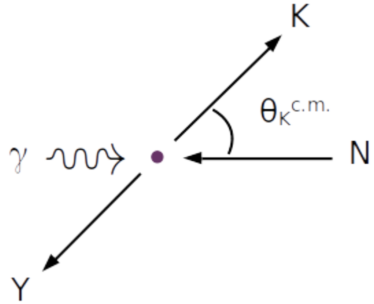
Tag	Resonance	Mass (MeV)	Width (MeV)	Branching ratio $K\Lambda$	Branching ratio $K\Sigma$	Fit M g_1	Fit M g_2	Fit L g_1	Fit L g_2
K*	$K^*(892)$	891.7	50.8			0.366 ± 0.024	1.103 ± 0.198	0.310 ± 0.019	
K1	$K_1(1270)$	1270	90			-1.448 ± 0.189	0.473 ± 0.156		
N3	$N(1535)\ 1/2^-$	1530	150			-0.709 ± 0.071			
N4	$N(1650)\ 1/2^-$	1650	125	0.07	0.00	0.314 ± 0.034		-0.085 ± 0.006	
N8	$N(1675)\ 5/2^-$	1675	145			-0.013 ± 0.001	0.022 ± 0.003	-0.010 ± 0.001	0.003 ± 0.002
N6	$N(1710)\ 1/2^+$	1710	140	0.15	0.01	-0.940 ± 0.093			
N7	$N(1720)\ 3/2^+$	1720	250	0.05	0.00	-0.098 ± 0.017	-0.082 ± 0.002	-0.187 ± 0.004	-0.126 ± 0.002
P4	$N(1875)\ 3/2^-$	1875	200	0.01	0.01	-0.220 ± 0.023	-0.223 ± 0.023	-0.042 ± 0.015	0.025 ± 0.013
P1	$N(1880)\ 1/2^+$	1880	300	0.16	0.14	-0.050 ± 0.064			
Mx	$N(1895)\ 1/2^-$	1895	120	0.18	0.13	-0.063 ± 0.005		0.019 ± 0.002	
P2	$N(1900)\ 3/2^+$	1920	200	0.11	0.05	-0.051 ± 0.005	-0.004 ± 0.001	0.027 ± 0.003	0.010 ± 0.001
M4	$N(2060)\ 5/2^-$	2100	400	0.01	0.03	-0.00001 ± 0.0001	0.003 ± 0.0003	-0.003 ± 0.0001	0.004 ± 0.0002
M1	$N(2120)\ 3/2^-$	2120	300			-0.034 ± 0.014	-0.010 ± 0.013	0.0003 ± 0.001	0.0 ± 0.0001
D1	$\Delta(1900)\ 1/2^-$	1860	250		0.01	0.298 ± 0.028			
D2	$\Delta(1930)\ 5/2^-$	1880	300						
D3	$\Delta(1920)\ 3/2^+$	1900	300						
D4	$\Delta(1940)\ 5/2^-$	1950	400						
S1	$\Sigma(1660)\ 1/2^+$	1660	100						
S2	$\Sigma(1750)\ 1/2^-$	1750	90						
S3	$\Sigma(1670)\ 3/2^-$	1670	60						
S4	$\Sigma(2010)\ 3/2^-$	1940	220						

	M “full” fit	L LASSO fit
no. of resonances	14	9
no. of parameters	25	17

Results: differential cross sections







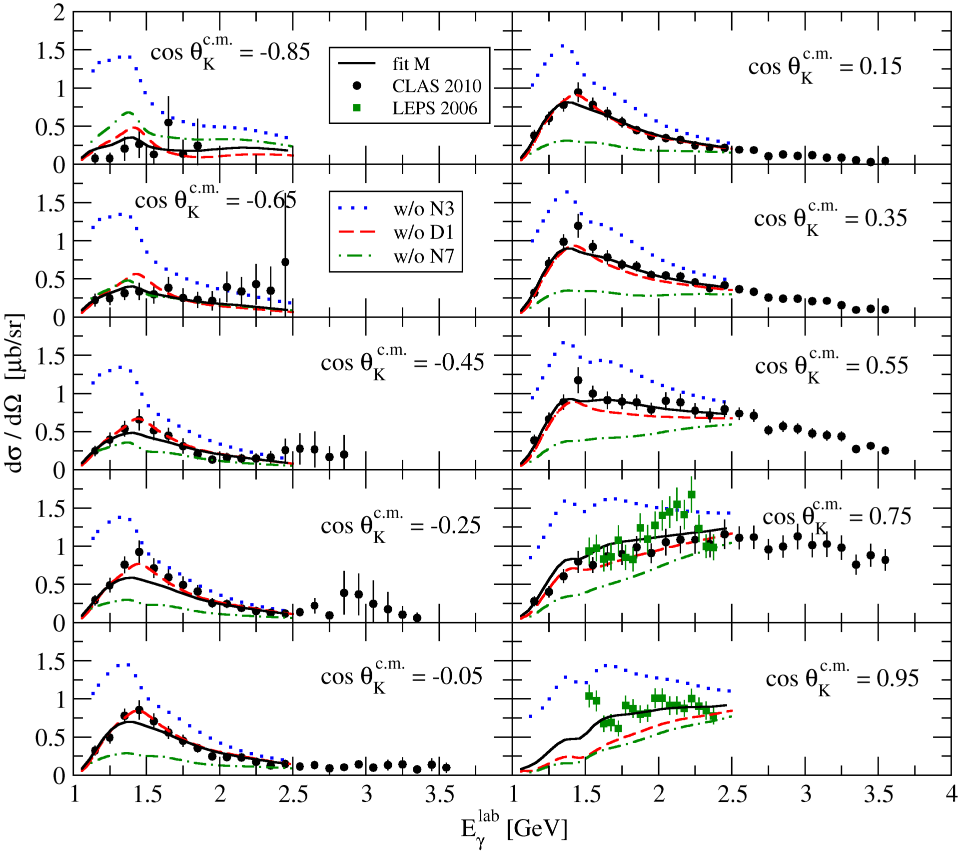
fit M: MINUIT

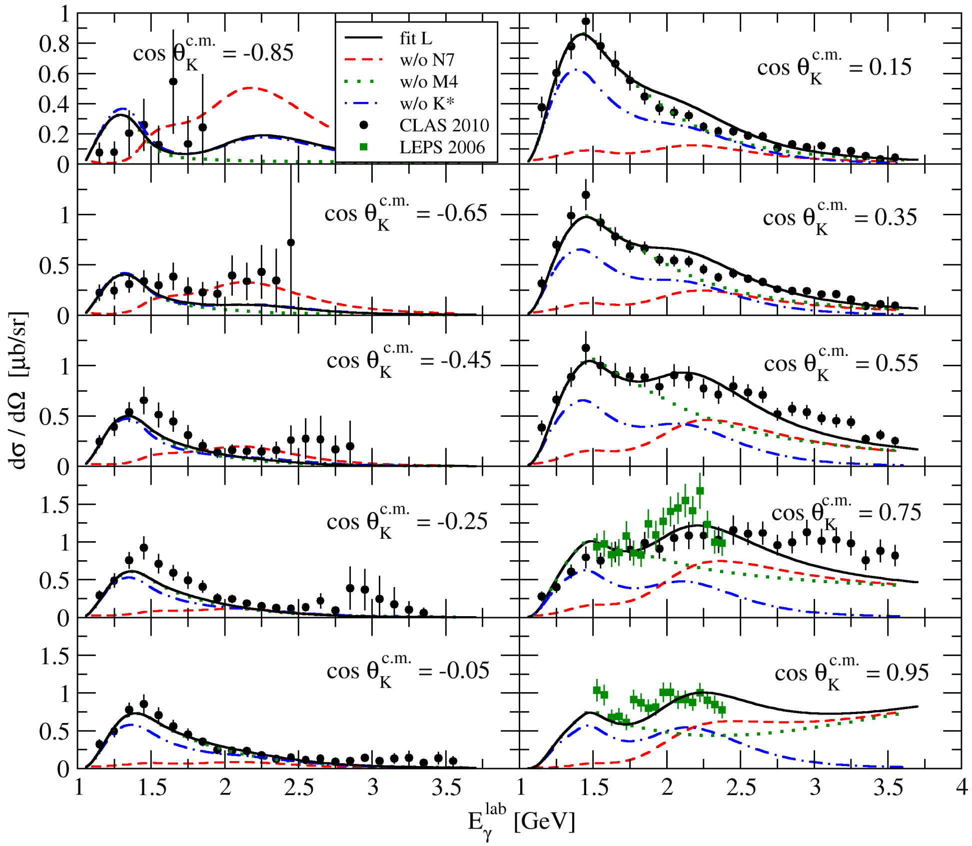
fit L: MINUIT + LASSO

$\theta_K^{c.m.}$: Kaon center-of-mass angle

E^{lab}_{γ}: incident photon energy

Results: differential cross sections

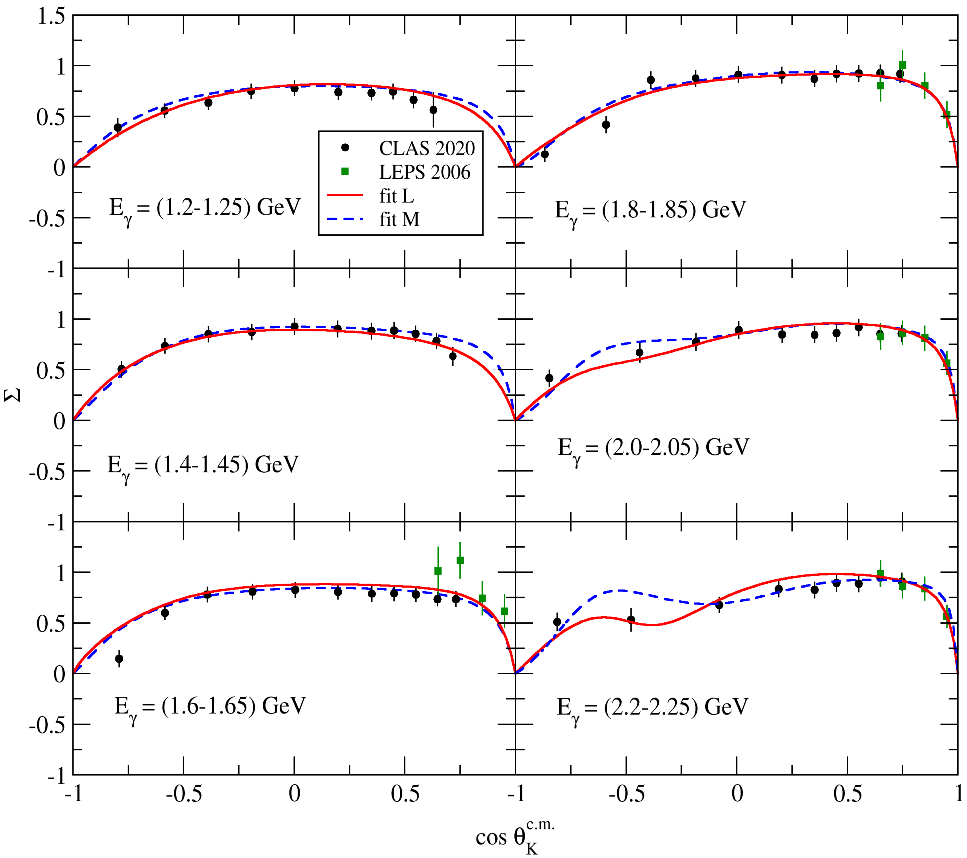


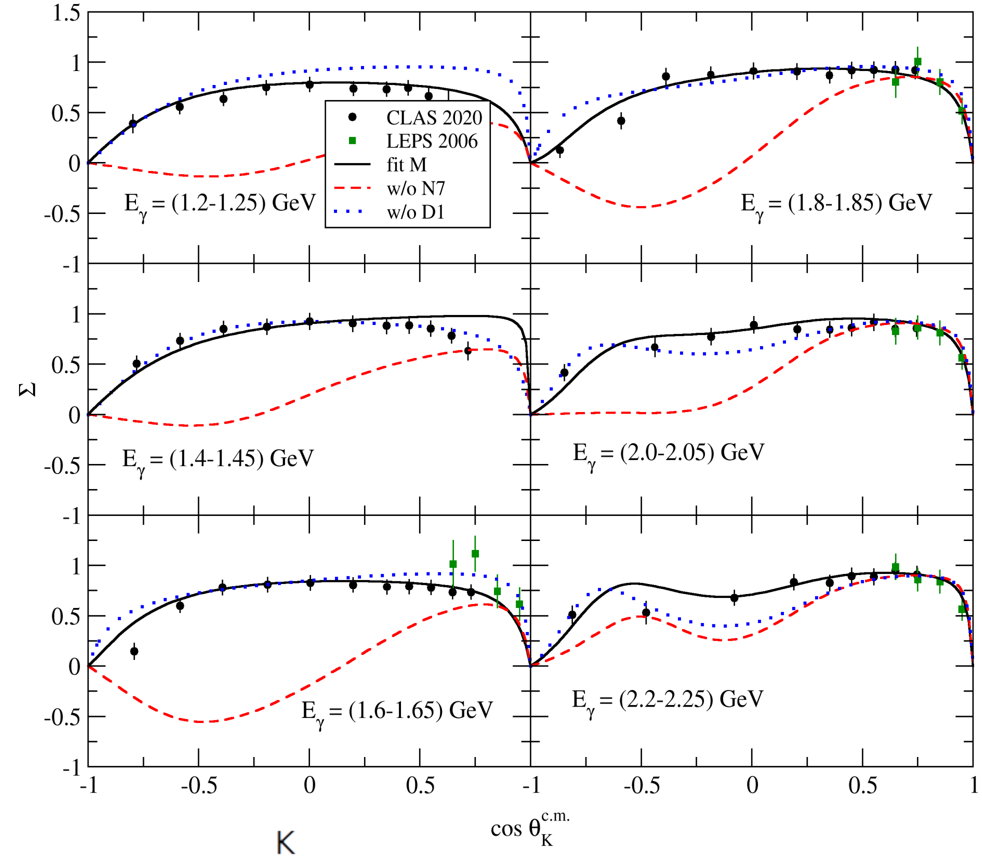


M: MINUIT fit
M fits w/o
N3 = N (1535) 1/2 −
D1 = Δ(1900) 1/2 −
N7 = N (1720) 3/2 +

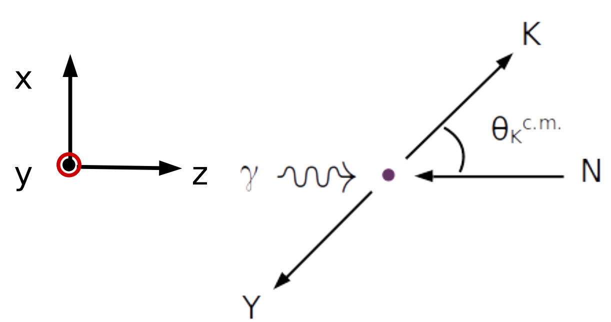
L: MINUIT + LASSO
L fits w/o
N7 = N (1720) 3/2 +
M4 = N (2060) 5/2 −
K ∗ (892)

Photon beam asymmetry Σ





Linearly polarized photons



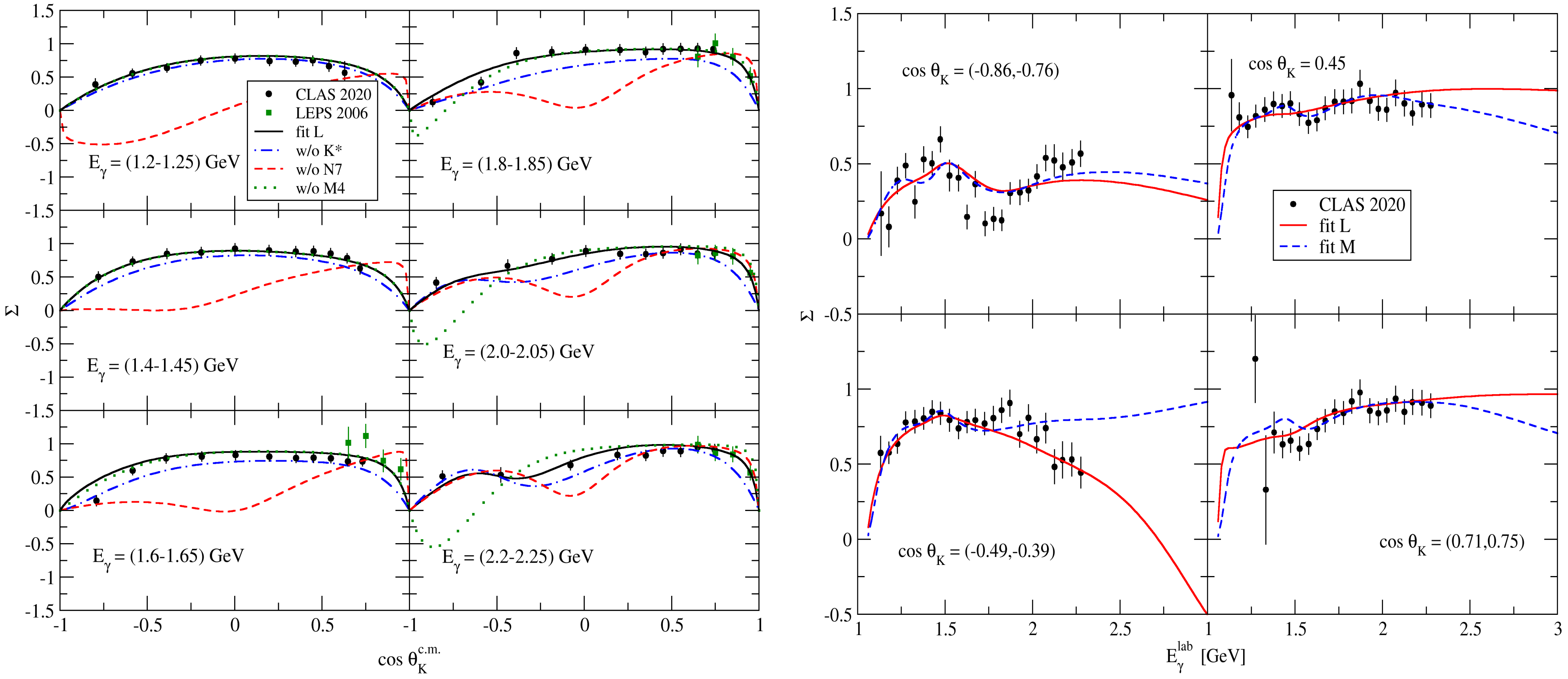
M: MINUIT fit
M fits w/o
D1 = Δ(1900) 1/2 −
N7 = N (1720) 3/2 +

$$\epsilon^{\lambda=x} \equiv \epsilon^{\parallel} = (0,1,0,0)$$

$$\epsilon^{\lambda=y} \equiv \epsilon^{\perp} = (0,0,1,0)$$

$$\Sigma = \frac{d\sigma^{\perp} - d\sigma^{\parallel}}{d\sigma^{\perp} + d\sigma^{\parallel}}$$

Photon beam asymmetry Σ



L: MINUIT + LASSO
L fits w/o
N7 = N (1720) 3/2 +
M4 = N (2060) 5/2 −
K ∗ (892)

Summary and outlook

- We modeled $K^+ \Sigma^-$ photoproduction with an Isobar model using regularization (LASSO) in combination with the Akaike and Bayesian information criteria.
- Regularization in a model with many parameters leads to more robust results, less prone to overfitting.
- The combination of the LASSO method with Information Criteria provides a method to choose the best subset of parameters (model).
- Future plans:
  - use Ridge regularization
  - fit simultaneously all 4 channels of $K\Sigma$ photoproduction → relate coupling constants by SU(2) (Isospin) symmetry

Thanks to:

P. Bydzovsky, A. Cieply, D. Skoupil and P. Vesely.

Bibliographic references

1). P. Bydzovsky, A. Cieply, D. Petrellis, D. Skoupil and N. Zachariou, Model selection for $K^+\Sigma^-$ photoproduction within an isobar model, Phys. Rev. C 104, 065202 (2021).

2). N. Zachariou et al., Beam-spin asymmetry Σ for Σ^- hyperon photoproduction off the neutron, Phys. Lett. B 827, 136985 (2022).

3). L. De Cruz, Bayesian model selection for electromagnetic kaon production in the Regge-plus-resonance framework, PhD Thesis, Ghent University (2012).

4). J. Landay, M. Döring, C. Fernández-Ramírez, B. Hu and R. Molina, Model selection for pion photoproduction, Phys. Rev. C 95, 015203 (2017).

5). J. Landay, M. Mai, M. Döring, H. Haberzettl and K. Nakayama, Towards the minimal spectrum of excited baryons. Phys. Rev. D 99, 016001 (2019).

6). D. Skoupil and P. Bydžovský, Photo- and electroproduction of K^+ with a unitarity-restored isobar model, Phys. Rev. C 97, 025202 (2018).

7). D. Skoupil and P. Bydžovský, Photoproduction of KΛ on the proton, Phys. Rev. C 93, 025204 (2016).

8). C. M. Bishop, Pattern Recognition and Machine Learning. Springer (2006).

9). H. Akaike, A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6): 716–723 (1974).

10). G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6(2), 461-464 (1978).