Wave-Selection Techniques for Partial-Wave Analysis in Light-Meson Spectroscopy

Hadron 2019, Guilin, China

Florian Kaspar
Boris Grube, Fabian Krinner, Stephan Paul, Stefan Wallner

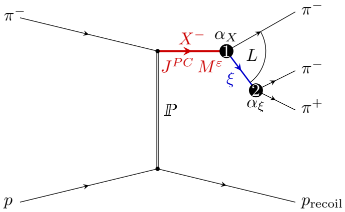
Saturday, August 17, 2019

Technische Universität München, Physics Department (E18)

Physics Motivation

Diffractive Dissociation

can produce light-meson resonances



diffractive dissociation into 3-pion final state [1]

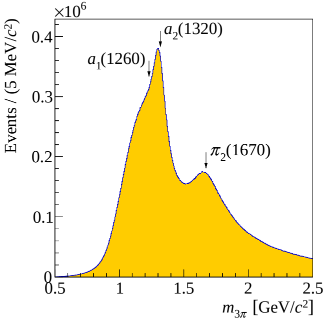
[1] C. Adolph et al. "Resonance production and $\pi\pi$ S-wave in $\pi^- + p \to \pi^-\pi^-\pi^+ + p_{\text{recoil}}$ at $190\ \text{GeV}/c$". In: *Phys. Rev. D* 95 (3 Feb. 2017), p. 032004. DOI: 10.1103/PhysRevD.95.032004. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.032004

Diffractive Dissociation

can produce light-meson resonances

Interfering Resonances

can be disentangled by partial-wave decomposition



m_X spectrum [1]

[1] C. Adolph et al. "Resonance production and $\pi\pi$ S-wave in $\pi^- + p \to \pi^-\pi^-\pi^+ + p_{\text{recoil}}$ at $190\ \text{GeV}/c$". In: *Phys. Rev. D* 95 (3 Feb. 2017), p. 032004. DOI: `10.1103/PhysRevD.95.032004`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.95.032004`

Diffractive Dissociation

can produce light-meson resonances

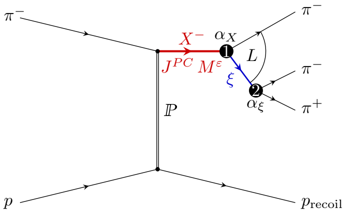
Interfering Resonances

can be disentangled by partial-wave decomposition

Extended Likelihood Fit

model with hundreds to thousands of parameters!

(actually ∞ number of partial waves)



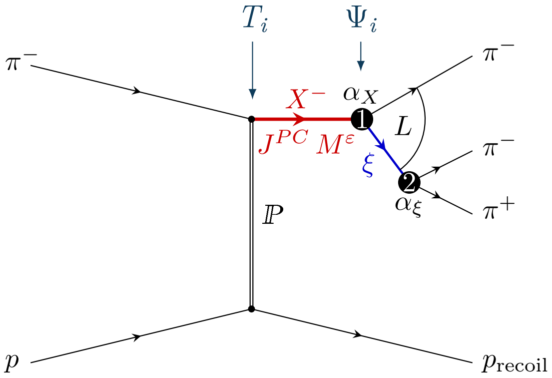
diffractive dissociation into 3-pion final state [1]

[1] C. Adolph et al. "Resonance production and $\pi\pi$ S-wave in $\pi^- + p \to \pi^-\pi^-\pi^+ + p_{\text{recoil}}$ at $190\ \mathrm{GeV}/c$". In: *Phys. Rev. D* 95 (3 Feb. 2017), p. 032004. DOI: 10.1103/PhysRevD.95.032004. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.032004

Intensity: (simplified)

$$\mathcal{I}(\tau; m_{3\pi}) = \Big| \sum_{i \in \{\text{waves}\}} T_i(m_{3\pi}) \Psi_i(\tau; m_{3\pi}) \Big|^2 + \Big| T_{\text{flat}}(m_{3\pi}) \Big|^2$$

$m_{3\pi}$ dependence unknown → narrow bins in $m_{3\pi}$



diffractive dissociation into three pions in the isobar model [2]

(extended) logLikelihood from intensity:

$$\log(\mathcal{L}) = \log\left(\frac{\bar{n}^n}{n!}\,e^{-\bar{n}}\prod_j^n \frac{\mathcal{I}(\tau_j)}{\int_\Omega \mathcal{I}(\tau)\,d\mathsf{LIPS}(\tau)}\right)$$

Maximize $\log\mathcal{L}$ to obtain estimate for complex transition amplitudes T_i (= Fit)

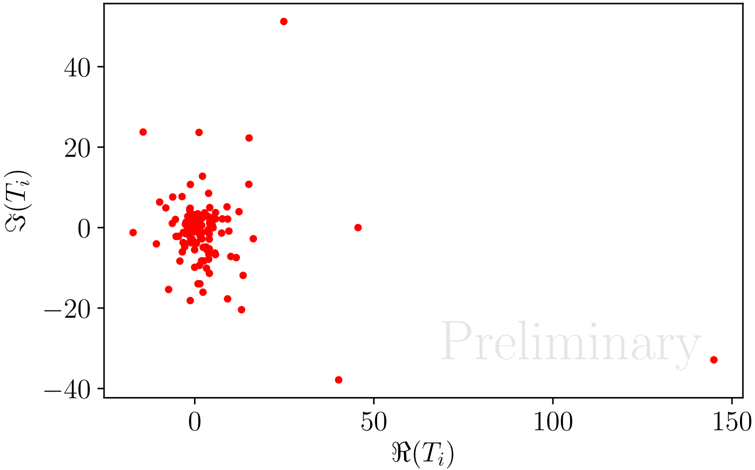
Transition amplitudes T_i:

Complex number: Phase and intensity ($|T_i|^2$)

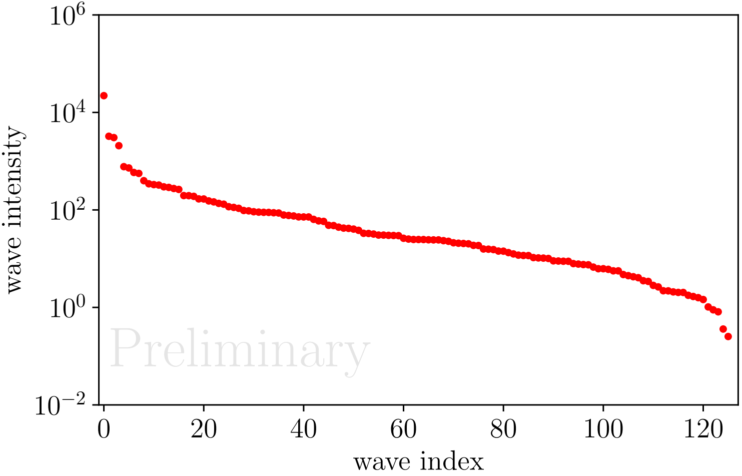
Two Monte Carlo Mock Datasets for 3π diffraction

- Generated mock-data sets with a few thousands events
- 3π-mass regions 1.00 GeV - 1.02 GeV with 19 waves and 1.80 GeV - 1.82 GeV with 126 waves (include only sensible isobars)

Every dot represents one amplitude T_i in the complex plane ...



… representation of their sorted intensities $|T_i|^2$ strong hierachy!



Model Selection

Different Fit Models:

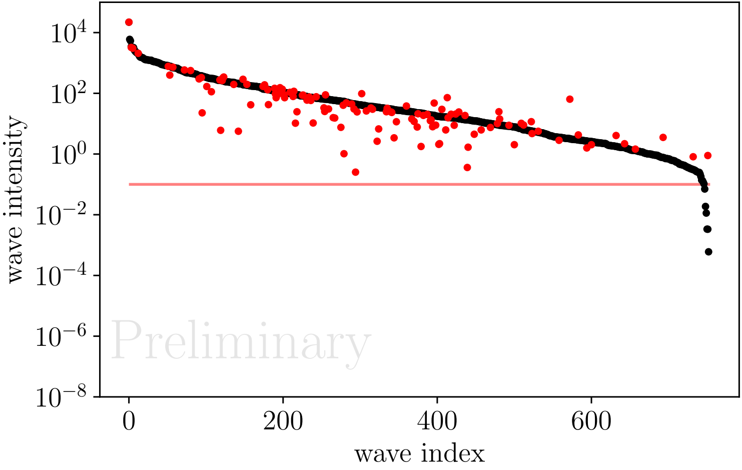
- selection of relevant subset of waves (we cannot fit a very large or even infinite number of waves) $\rightarrow$ model selection
- different parametrizations of decay amplitudes Ψ_i
- model for background
- ...

All aspects are interdependent $\rightarrow$ model selection required!

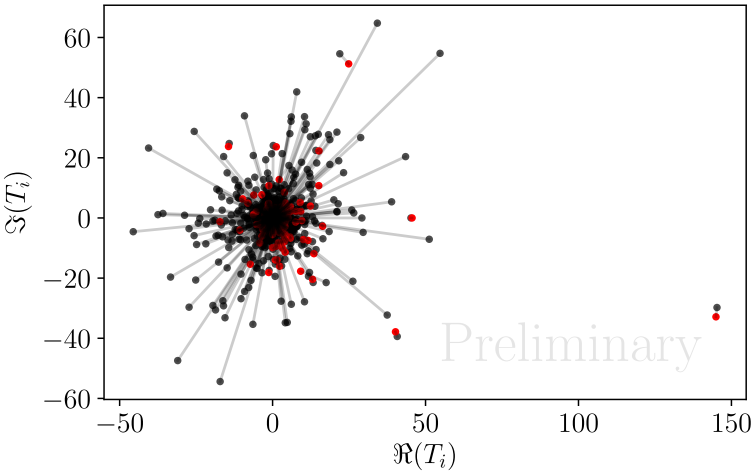
Model Selection:

- An objective way to select a subset of contributing waves
- Here:
  1. wave pool (= all waves up to certain QN & isobars)
  2. regularization

Fit of full 753 wave pool without regularization at 1.8 GeV. Black intensities of pool. Red intensities of reference fit. → Fails!



Fit of full 753 wave pool without regularization at 1.8 GeV. Black amplitudes of pool. Red amplitudes of reference fit.



(Sparse) Regularization:

- Use large pool of 'all possible' waves and add regularization term to $\log \mathcal{L}$ that pushes T_i towards zero during the fit
  $\rightarrow T_i = 0$ equivalent to exclusion
- Probably the most famous: LASSO[3]

$$\log \mathcal{L} - \frac{1}{\Gamma} \sum_i |T_i|$$

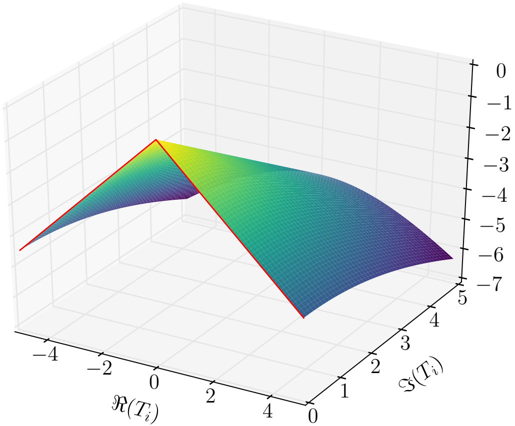
  Suggested for and applied to amplitude analysis by Guegan et al.[4] and for pion photo production by Landay et al.[5] also used for analysis of CLEO-c D-decay data (P. d'Argent et al.) and other analyses in the field!
  $\rightarrow$ established

[3] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246. URL: http://www.jstor.org/stable/2346178.

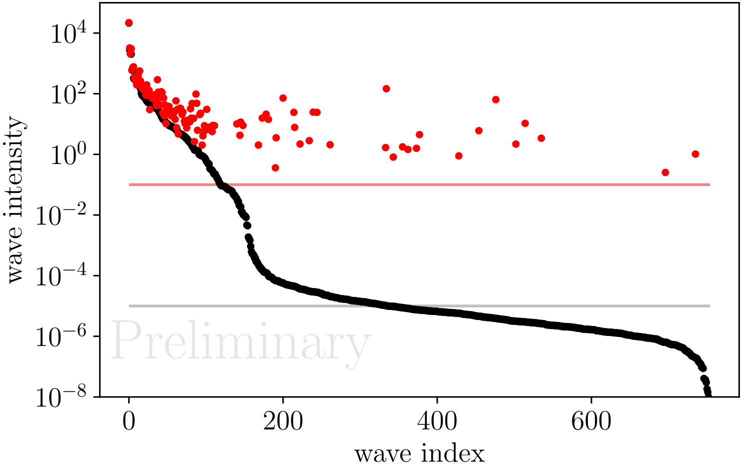
[4] Baptiste Guegan et al. "Model selection for amplitude analysis". In: *JINST* 10.09 (2015), P09002. DOI: 10.1088/1748-0221/10/09/P09002. arXiv: 1505.05133 [physics.data-an].

[5] J. Landay et al. "Model selection for pion photoproduction". In: *Physical Review C* 95.1 (Jan. 2017). DOI: 10.1103/physrevc.95.015203.

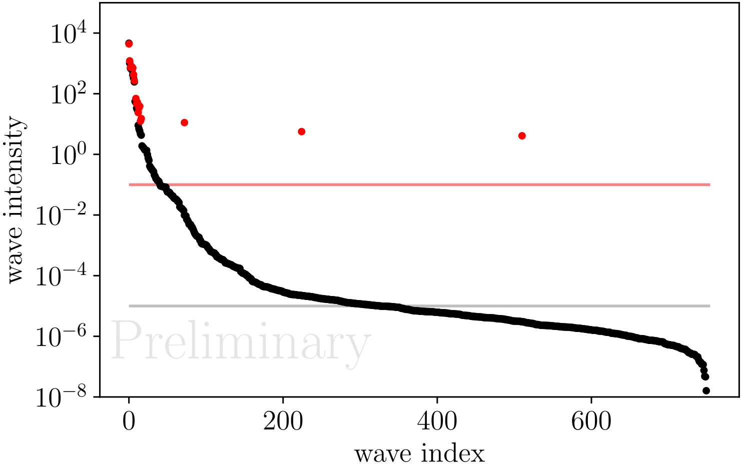
LASSO penalty is cone in the complex plane. For our fits: 'smooth' absolute value via $\sqrt{(x^2+\epsilon)}$ to remove non-differentiability at zero



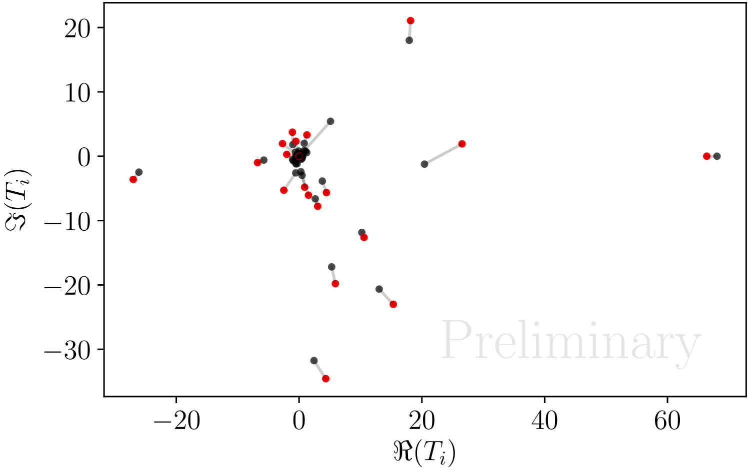
Sorted intensities of waves with LASSO regularization for $\Gamma_{\text{LASSO}} = 0.3$ and smoothing of $\epsilon = 10^{-5}$ for the high mass bin:



Sorted intensities of waves with LASSO regularization for $\Gamma_{\text{LASSO}} = 0.4$ and smoothing of $\epsilon = 10^{-5}$ for the low mass bin:



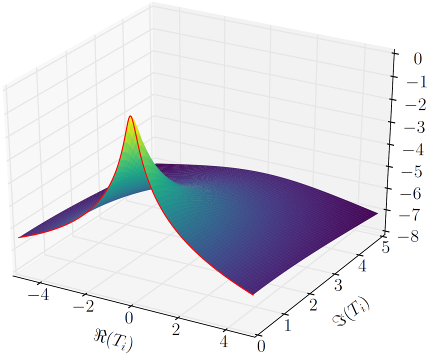
Amplitudes of waves with LASSO regularization for $\Gamma_{\text{LASSO}} = 0.4$ and smoothing of $\epsilon = 10^{-5}$ for the low mass bin:



BCM/Cauchy Regularization:

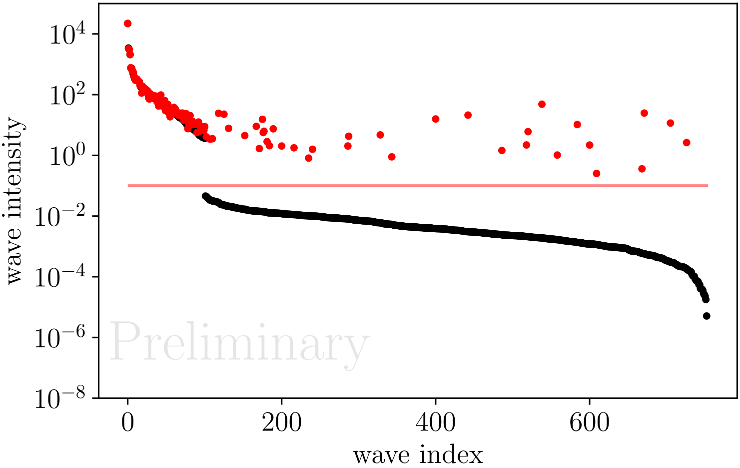
Model selection Ansatz by K. Bicker[6] 'Biggest Conceivable Model':

$$\log \mathcal{L} + \sum_i \log \left(1/(1 + |T_i|^2/\Gamma^2)\right)$$

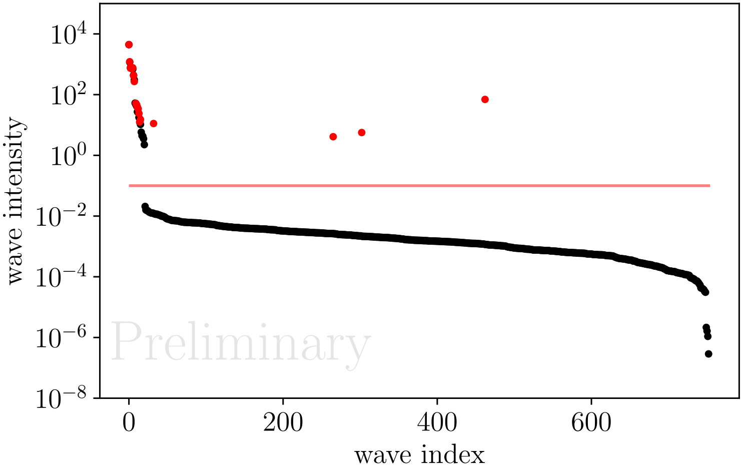


[6] Karl Bicker. "Model Selection for and Partial-Wave Analysis of a Five-Pion Final State at the COMPASS Experiment at CERN". PhD Thesis. Technische Universität München, Apr. 19, 2016, Oliver Drotleff. "Model Selection of $\pi^- + p \to \pi^-\pi^+\pi^- + p$ at the COMPASS Experiment at CERN". Diploma Thesis. Technische Universität München, Nov. 2, 2015.

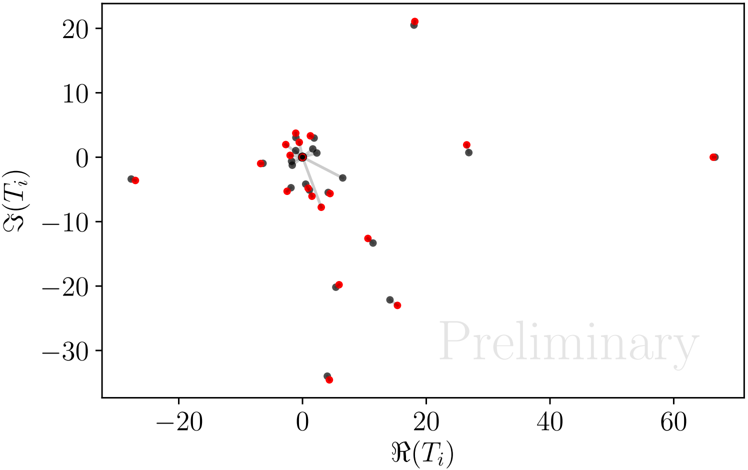
Sorted intensities of waves with BCM regularization for $\Gamma_{\text{BCM}} = 0.2$:



Sorted intensities of waves with BCM regularization for $\Gamma_{\text{BCM}} = 0.25$:



Amplitudes of waves with BCM regularization for $\Gamma_{\mathrm{BCM}} = 0.25$:



LASSO Properties:

- LASSO is convex $\rightarrow$ less problems with multimodality
- Pushes deselected wave intensities to zero (or smoothing scale ϵ)
- Stronger bias on large intensity waves

BCM Properties:

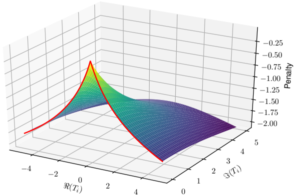
- BCM is NOT convex $\rightarrow$ many minima
- Does not push deselected waves all the way to zero
- Less bias on large intensity waves $\rightarrow$ bad at constraining destructive inference

$\rightarrow$ Invesitgate alternatives that combine advantages of both!

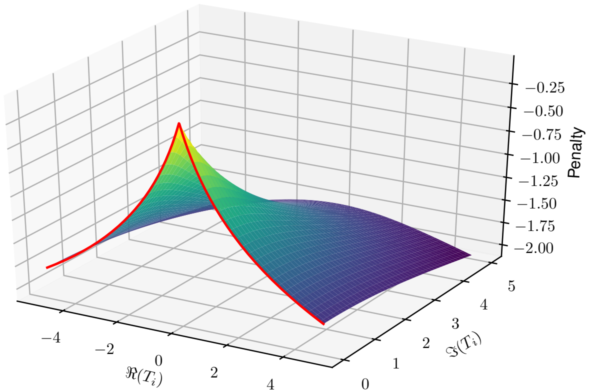
Generalized Pareto[7]:

- Logarithmic behavior for large intensities
- LASSO-like behavior for small intensities
- In the limit $\zeta \to 0$ this regularization falls back to the LASSO

$$\log \mathcal{L} - \frac{1}{\zeta} \sum_i \log\left(1 + \zeta \left| T_i \right| / \Gamma\right)$$



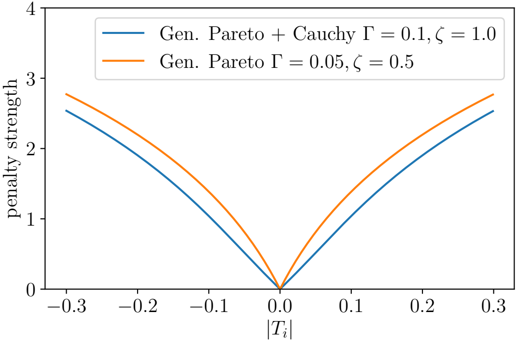
[7] Artin Armagan, David B. Dunson, and Jaeyong Lee. "Generalized double Pareto shrinkage". In: *Statistica Sinica* (2013). DOI: 10.5705/ss.2011.048.


−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
−2.00
Penalty
−4
−2
0
2
4
$\Re(T_i)$
0
1
2
3
4
5
$\Im(T_i)$

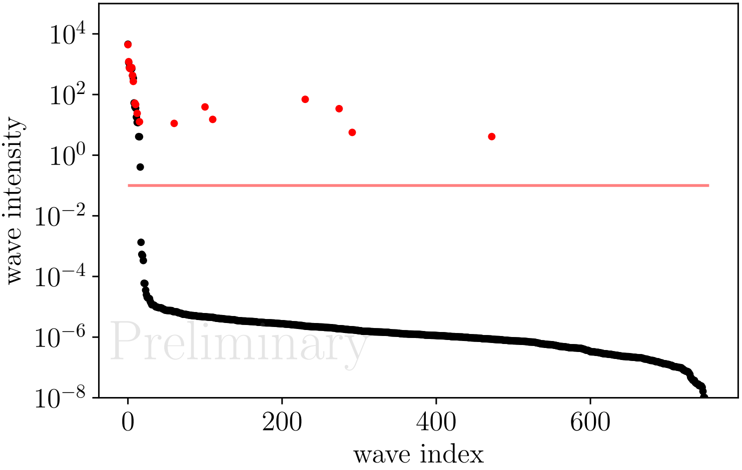
Generalized Pareto + Cauchy Smoothing:

Like generalized Pareto, but convex near minimum!

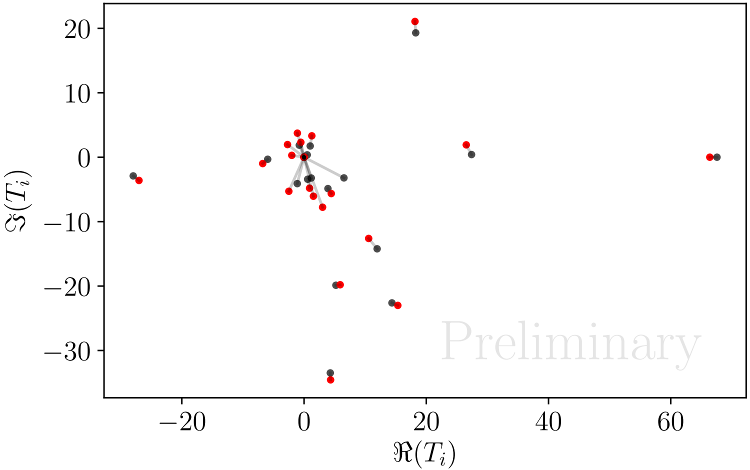
$$\log \mathcal{L} - \left(\sum_i \frac{1}{\zeta} \log \left(1 + \zeta \left| T_i \right| / \Gamma \right) + 0.5 \zeta \log \left(1 + \left| T_i \right|^2 / \Gamma^2 \right) \right)$$



Generalized Pareto + Cauchy smoothing in low mass bin.



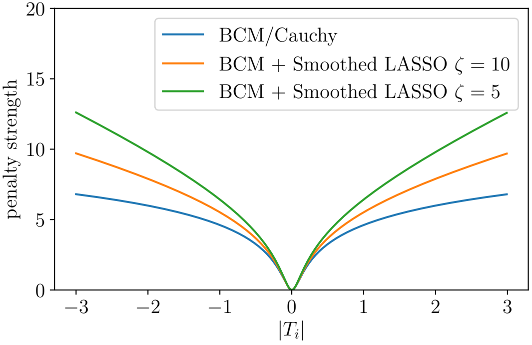
Generalized Pareto + Cauchy smoothing in low mass bin.



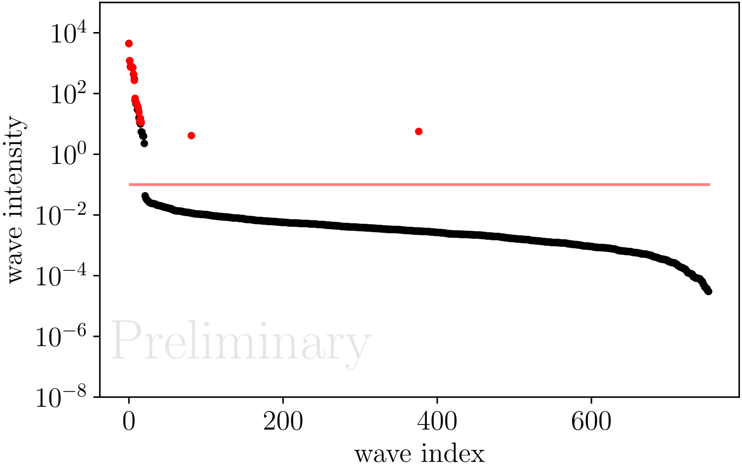
Cauchy + 'Smoothed LASSO':

ζ allows tuning of penalty tails towards LASSO

$$\log \mathcal{L} - \left(\sum_i \left(\sqrt{|T_i|^2 / \Gamma^2 + 1} \right) / \zeta + 2 \log \left(\sqrt{|T_i|^2 / \Gamma^2 + 1} \right) - 1/\zeta \right)$$



Sorted intensities of waves with BCM regularization for $\Gamma_{\text{BCM}} = 0.3$ and $\zeta = 100$:



PWA Fit: Start Value Generation

PWA fits oftentimes inherently multimodal: many different local optima

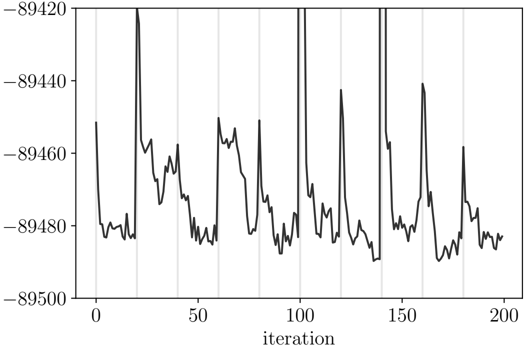
- Non-convex regularization cannot reduce fit to single optimum
- Situation similar to many Machine Learning problems: high-dimensional + many optima

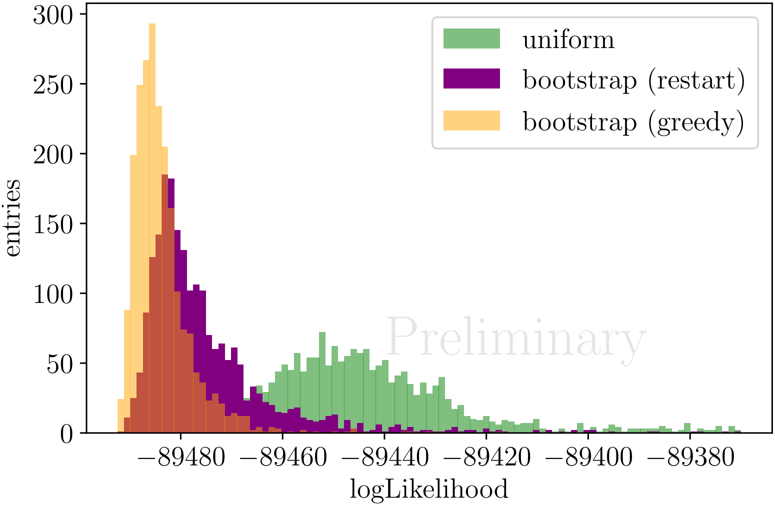
How to cope with multimodal and high-dimensional fits?

- Usual approach: Start with uniformly drawn parameters for each wave
- Alternative: Make use of boostrapping Ansatz[8]
- Fit on resampled data set to propose new start parameters → global structure should stay the same

[8] Simon N. Wood. “Minimizing Model Fitting Objectives That Contain Spurious Local Minima by Bootstrap Restarting”. In: *Biometrics* 57.1 (Mar. 2001), pp. 240–244. DOI: `10.1111/j.0006-341x.2001.00240.x`.

Boostrap restarting for 200 fit attempts with BCM regularization at 1.8 GeV: Potential to improve logLikelihood




300
250
200
150
100
50
0
entries
−89480
−89460
−89440
−89420
−89400
−89380
logLikelihood
uniform
bootstrap (restart)
bootstrap (greedy)
Preliminary

Model Comparison & Parameter Tuning

Regularization Parameters?

- Parameters of the regularization term influence number of selected waves: Every set of parameters yields a different model
- Compare these models to make a motivated choice of the parameter values!

Model Selection Critera:

- Try condensing the choice in a single number
- Penalize model complexity → counter likelihood improvement

→ different assumptions lead to different criteria

Guegan et al. and Landay et al. make use of the AIC[9] and BIC[10] information criteria

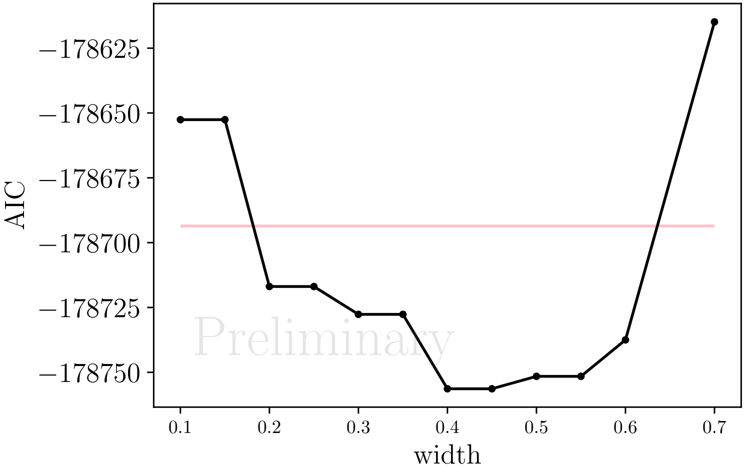
- AIC $= -2\log\mathcal{L} + 2k$
- BIC $= -2\log\mathcal{L} + k\log(n)$
- BIC is approximation of Bayesian evidence $\rightarrow$ Alternatively use Gaussian approximation to calculate evidence + prior defined by parameter volume with $\bar{n} = n$ [11]

[9] H. Akaike. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (Dec. 1974), pp. 716–723. ISSN: 0018-9286. DOI: `10.1109/TAC.1974.1100705`.

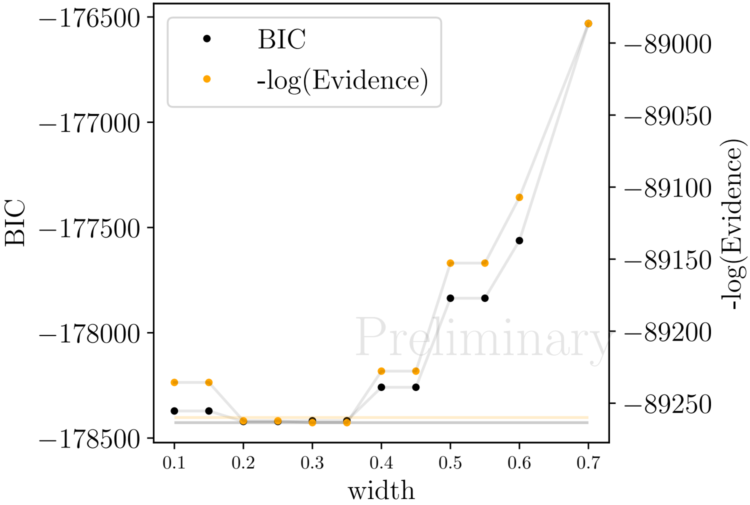
[10] Gideon Schwarz. "Estimating the Dimension of a Model". In: *Ann. Statist.* 6.2 (Mar. 1978), pp. 461–464. DOI: `10.1214/aos/1176344136`. URL: `http://dx.doi.org/10.1214/aos/1176344136`.

[11] Similar to what has been used for genetic model selection by [4]

AIC scan for BCM penalty at low mass bin



BIC and evidence scan for BCM penalty at low mass bin



Different Approach:

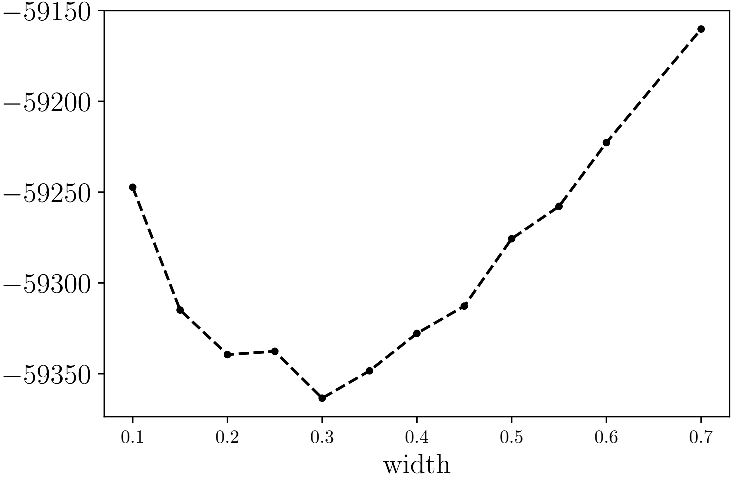
- AIC/BIC and the approximated Evidence are relatively easy to compute but require strong assumptions that might not be fulfilled on real data
- $\rightarrow$ Cross-Validation might be more robust
- Widely used in Machine Learning
- Used for example in Pion Photoproduction by Landay et al.[12]

K-Fold Cross-Validation:

- Split data in K sets, fit on K-1, evaluate 'loss' on the left out
- Repeat this procedure K times and add losses
- Here: Use $K = 20$

[12] J. Landay et al. "Model selection for pion photoproduction". In: *Physical Review C* 95.1 (Jan. 2017). DOI: 10.1103/physrevc.95.015203.

Parameter scan for BCM in the low mass bin. Summed loss of the 20 test sets:



Software

Framework Developments:

- ROOTPWA analysis framework currently undergoing changes towards Python implementation of PWA
- New interface makes use of the Autograd[13] package for automatic differentiation
- Autograd allows easy implementation of more complicated likelihoods/penalties: e.g. Penalty on fit fraction to reduce bias on total intensity

[13] Dougal Maclaurin et al. *Autograd*. URL: `https://github.com/HIPS/autograd`.

Summary

- Sparse regularization terms capable of selecting relevant partial waves (even for extremely many parameters)
- Can confirm usefulness of LASSO + parameter tuning
- LASSO experiences strong bias and includes additional (small) waves, but prevents destructive interference
- BCM/Cauchy reduces bias, but large destructive interference can overcome regularization
- Other approaches to combine desirable features of both penalties
- Use tools like bootstrap restart to find better optima …
- … and make cross validation feasible

Outlook

- Use more computational resources to study full CV scan of all the penalties
- Apply to real data

Thank you for your attention!

Acknowledgments

I would like to thank ...

Dr. Boris Grube, Dr. Fabian Krinner, Prof. Dr. Stephan Paul, Dr. Dmitri Ryabchikov and Stefan Wallner

This work was supported by the BMBF, the DFG Cluster of Excellence “Origin and Structure of the Universe” (Exc 153), and the Maier-Leibnitz-Laboratorium der Universtiät und der Technischen Universität München.

Questions?

C. Adolph et al. "Resonance production and $\pi\pi$ S-wave in $\pi^- + p \to \pi^-\pi^-\pi^+ + p_{\text{recoil}}$ at $190\ \mathrm{GeV}/c$". In: *Phys. Rev. D* 95 (3 Feb. 2017), p. 032004. DOI: 10.1103/PhysRevD.95.032004. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.032004.

H. Akaike. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (Dec. 1974), pp. 716–723. ISSN: 0018-9286. DOI: 10.1109/TAC.1974.1100705.

Artin Armagan, David B. Dunson, and Jaeyong Lee. "Generalized double Pareto shrinkage". In: *Statistica Sinica* (2013). DOI: 10.5705/ss.2011.048.

Karl Bicker. "Model Selection for and Partial-Wave Analysis of a Five-Pion Final State at the COMPASS Experiment at CERN". PhD Thesis. Technische Universität München, Apr. 19, 2016.

Michael Droettboom et al. *matplotlib: matplotlib v1.5.0*. Oct. 2015. DOI: 10.5281/zenodo.32914. URL: https://doi.org/10.5281/zenodo.32914.

Oliver Drotleff. "Model Selection of $\pi^- + p \to \pi^-\pi^+\pi^- + p$ at the COMPASS Experiment at CERN". Diploma Thesis. Technische Universität München, Nov. 2, 2015.

Baptiste Guegan et al. "Model selection for amplitude analysis". In: *JINST* 10.09 (2015), P09002. DOI: 10.1088/1748-0221/10/09/P09002. arXiv: 1505.05133 [physics.data-an].

J. Landay et al. "Model selection for pion photoproduction". In: *Physical Review C* 95.1 (Jan. 2017). DOI: 10.1103/physrevc.95.015203.

Dougal Maclaurin et al. *Autograd*. URL: `https://github.com/HIPS/autograd`.

Gideon Schwarz. "Estimating the Dimension of a Model". In: *Ann. Statist.* 6.2 (Mar. 1978), pp. 461–464. DOI: `10.1214/aos/1176344136`. URL: `http://dx.doi.org/10.1214/aos/1176344136`.

Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246. URL: `http://www.jstor.org/stable/2346178`.

Matthias Vogelgesang. *Metropolis Theme*. URL: `https://github.com/matze/mtheme`.

Simon N. Wood. "Minimizing Model Fitting Objectives That Contain Spurious Local Minima by Bootstrap Restarting". In: *Biometrics* 57.1 (Mar. 2001), pp. 240–244. DOI: `10.1111/j.0006-341x.2001.00240.x`.

Bayes' Theorem:

$$P(\theta|D,M) = \frac{P(D|\theta,M)\,P(\theta,M)}{P(D|M)} \tag{1}$$

Evidence: Probability of the data given the model can be used to compare models:

$$P(D|M) = \int P(D|\theta,M)\,P(\theta,M)\,d\theta \tag{2}$$

Select model with largest evidence

Reparametrization:

Reparametrization results in a Jacobian Determinant !!!

Problem:

- requires specification of prior and might be very sensitive to it
- computation prohibitively expensive for high-dimensional models $\rightarrow$ approximations needed!

Approximations:

- Laplace Approximation: $\log(\mathcal{L}) \approx$ Gaussian around maximum
  Bicker: prior is defined by parameter volume for which $\bar{n} = n$
  $\rightarrow$ defines Hyper-Ellipsoid
- BIC[14]: Asymptotic approximation $(n \rightarrow \infty) \rightarrow$ influence of prior negligible, results in "correction term" for $\log \mathcal{L}$ similar to AIC:
  $\text{BIC} = -2\log\mathcal{L} + k\log(n)$

[14] Gideon Schwarz. "Estimating the Dimension of a Model". In: *Ann. Statist.* 6.2 (Mar. 1978), pp. 461–464. DOI: 10.1214/aos/1176344136. URL: http://dx.doi.org/10.1214/aos/1176344136, Baptiste Guegan et al. "Model selection for amplitude analysis". In: *JINST* 10.09 (2015), P09002. DOI: 10.1088/1748-0221/10/09/P09002. arXiv: 1505.05133 [physics.data-an], J. Landay et al. "Model selection for pion photoproduction". In: *Physical Review C* 95.1 (Jan. 2017). DOI: 10.1103/physrevc.95.015203.

Amplitudes of waves with BCM regularization for $\Gamma_{\mathrm{BCM}} = 0.3$ and $\zeta = 100$: