Post-hoc regularisation

Smearing data for more accurate plots

Regularise by linear algebra

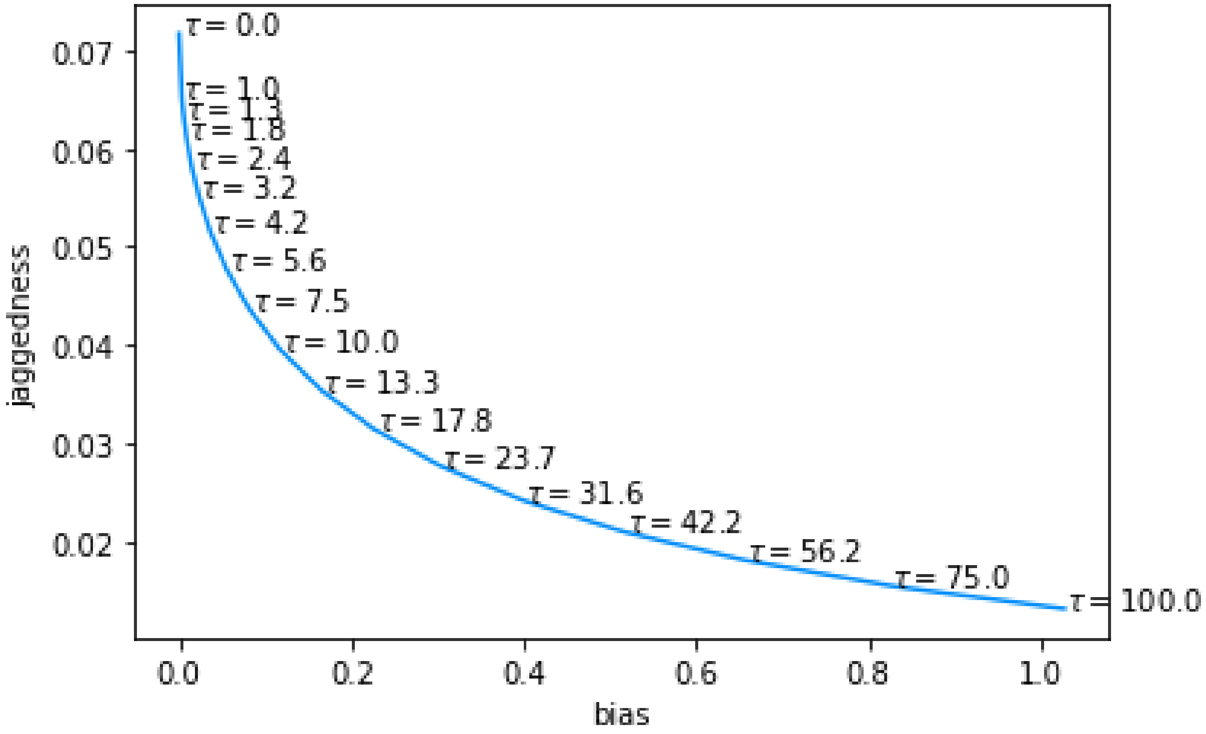
- New perspective: Regularisation as matrix multiplication
- Can construct matrix A equivalent to Tikhonov reg.

$$\chi^2 = -2\ln\left(L_{stat}(\theta)\right) - 2\ln\left(L_{prior}(\theta)\right)$$
$$\approx \left(\theta - \hat{\theta}\right)^T V^{-1}\left(\theta - \hat{\theta}\right) + const.$$
$$\chi^{2\prime} = -2\ln\left(L_{stat}(\theta)\right) - 2\ln\left(L_{prior}(\theta)\right) + P_{reg}(\theta)$$
$$\approx \left(\theta - \hat{\theta}\right)^T V^{-1}\left(\theta - \hat{\theta}\right) + \theta^T Q\, \theta + const.$$
$$= \left(\theta - \widehat{\theta'}\right)^T V'^{-1}\left(\theta - \widehat{\theta'}\right) + const.$$

- $\widehat{\theta'} = A\,\hat{\theta}$
- $A = (V^{-1} + Q)^{-1} V^{-1}$
- $V' = AVA^T$

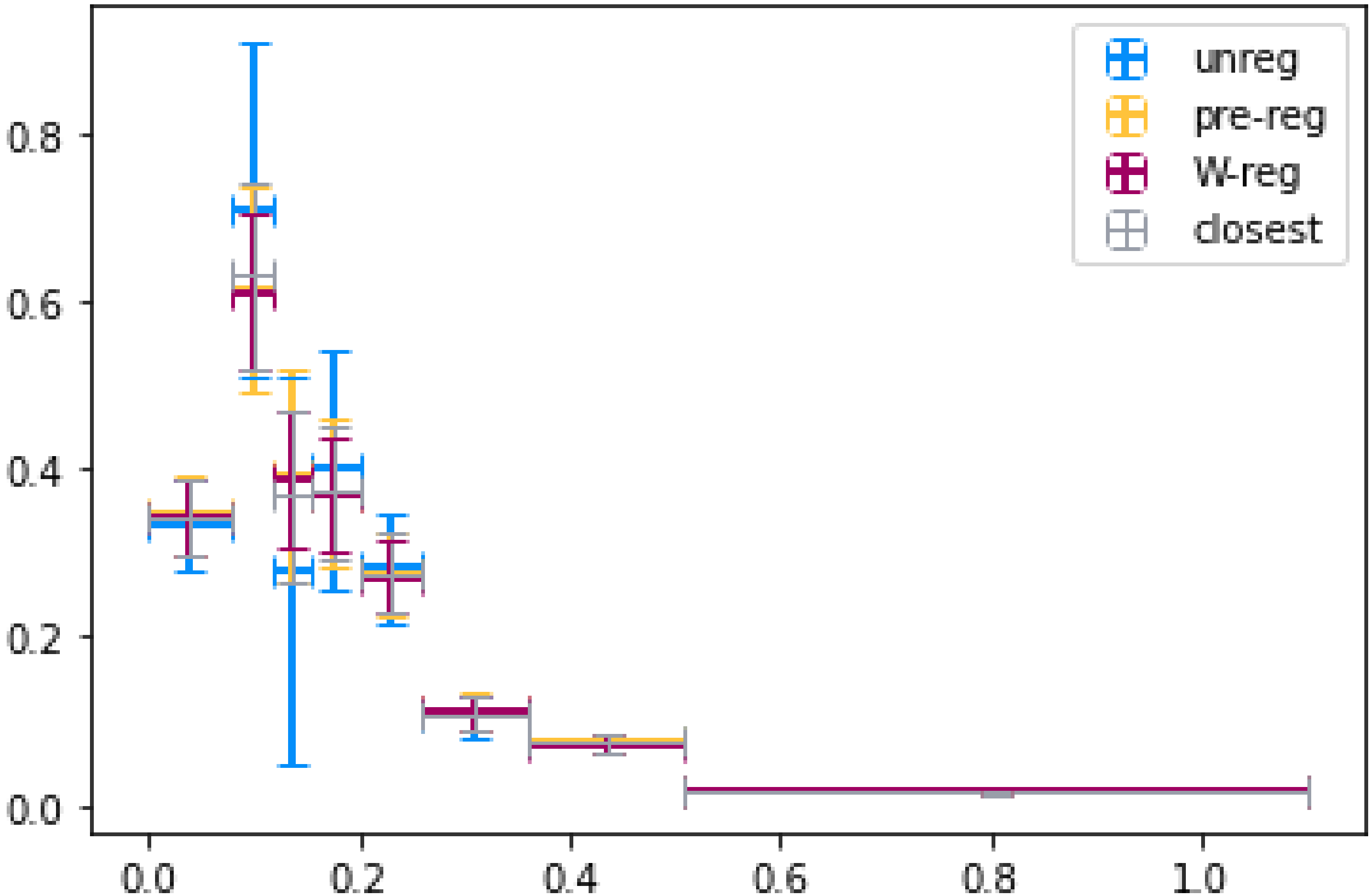
Computing efficiency

- Can introduce arbitrary regularisation after single fit
  - Potentially saving lots of time
  - Especially when done in XSEC space, rather than fit param.
- L-curve scan: Q → τQ
  - Ran on my laptop in < 15s



Works with any published result

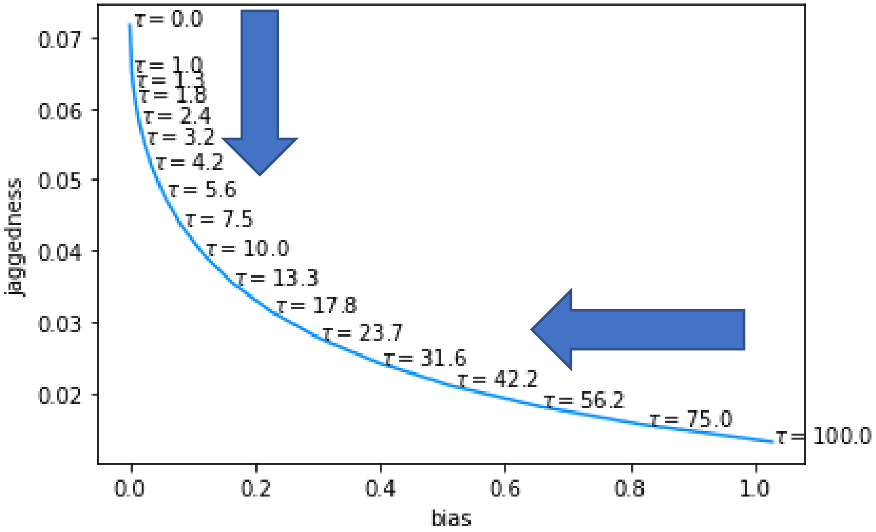
- No need to know details of extraction method
  - Just MLE and covariance
  - Example: T2K dpT measurement



Summary I

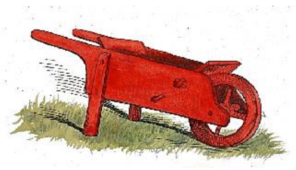
- New ways of thinking about regularisation
  - Equivalent to multiplying matrix A
    - Coordinate transformation → change of variable meaning
    - "Additional smearing" of result → same variable meaning
- Can apply regularisation after unregularized fit
  - Assumes likelihood is well described by covariance matrix
  - No re-fitting needed, just linear algebra
- Works on XSEC results directly, even after publication
  - Hence "post-hoc regularisation"

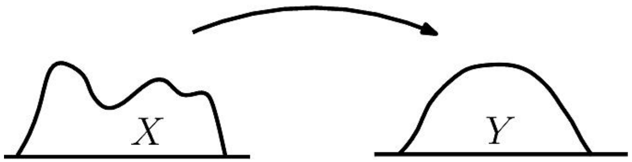
- Regularised result + A contain full information
  - Can be interpreted as transformation into other coord. syst.
  - Almost any A possible, but variable meanings change
    - E.g. A that switches bins, mirrors values, etc.
- Choice of Q and thus A is kind of arbitrary
  - Kink in L-curve method subjective
- More objective choice of A possible?
- Understand L-curve
  - Minimise “jaggedness” (Penalty term in likelihood)
  - Minimise shift of central value (squared Mahalanobis distance)
    - Interpret as change of result, NOT change of variable meaning



Better bias measure

- Result is not just central value!
  - Should take cov of regularised result into account
- How to measure difference between distributions rather than points?
- Earth mover's distance
  - Minimal total distance one has to shift probabilities to get from one distribution to the other





- Wasserstein metric
  - Neat formula for comparing multivariate normal distributions

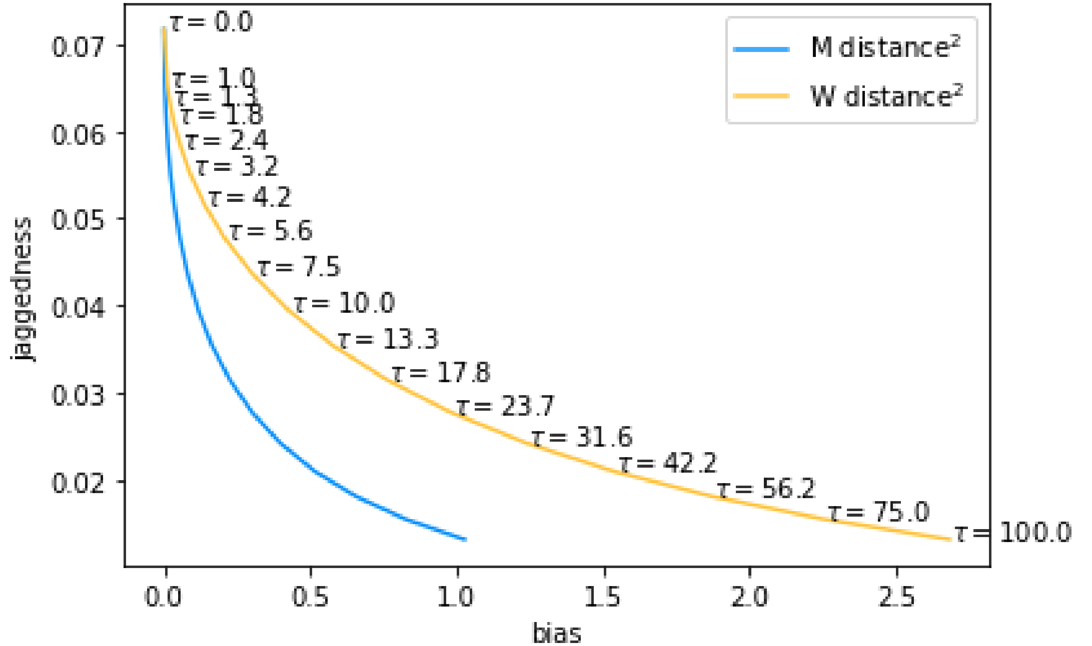
$$W^2 = \left|\widehat{\theta_1} - \widehat{\theta_2}\right|^2 + Tr\left(V_1 + V_2 - 2\left(V_2^{\frac{1}{2}} V_1 V_2^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)$$

Result difference

- Uses Euclidean distance in parameter space
  - E.g. how many cm^2/GeV you move the probability function
  - Not very informative, especially with abstract parameters
  - Transform into standard normal space of un-regularised result
  - W² = $\left(\hat{\theta} - \widehat{\theta'}\right)^T V^{-1}\left(\hat{\theta} - \widehat{\theta'}\right) + N + Tr\left(UV'U^T - 2(UV'U^T)^{\frac{1}{2}}\right)$

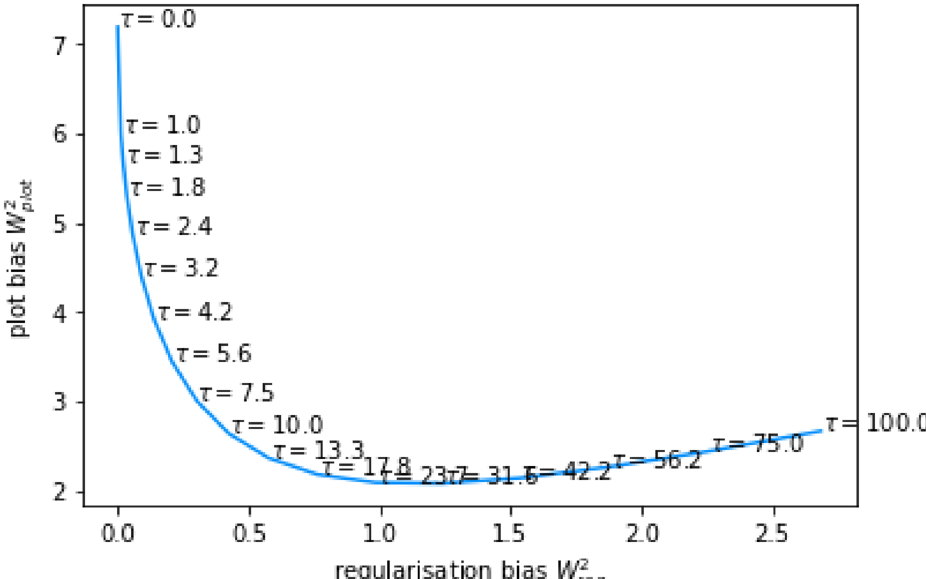
$$V^{-1} = U^T U$$

- For identical cov, value is same as M-distance!
- Metric in units of "z-score", "standard deviations", "chi-square"



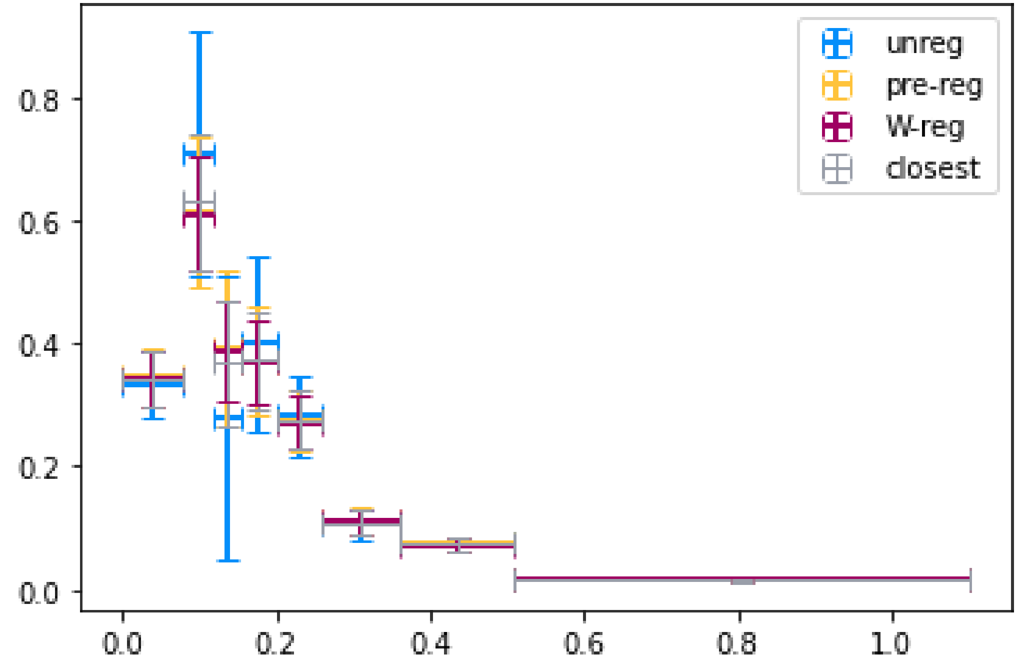
Plot bias

- Jaggedness not actually the problem
  - Data points fluctuate around true value, so what?
- Really want to make plots less misleading
  - Plots do not contain information about bin correlations
  - Chi-by-eye does not work
- Want to reduce difference between shown, implied uncorrelated distribution and actual correlated result
  - Can measure difference with W-metric!
  - Just set all off diagonals of regularised result to 0



Leads to similar regularisation!

- Minimise plot-bias: Wasserstein-Tikhonov-Regularisation



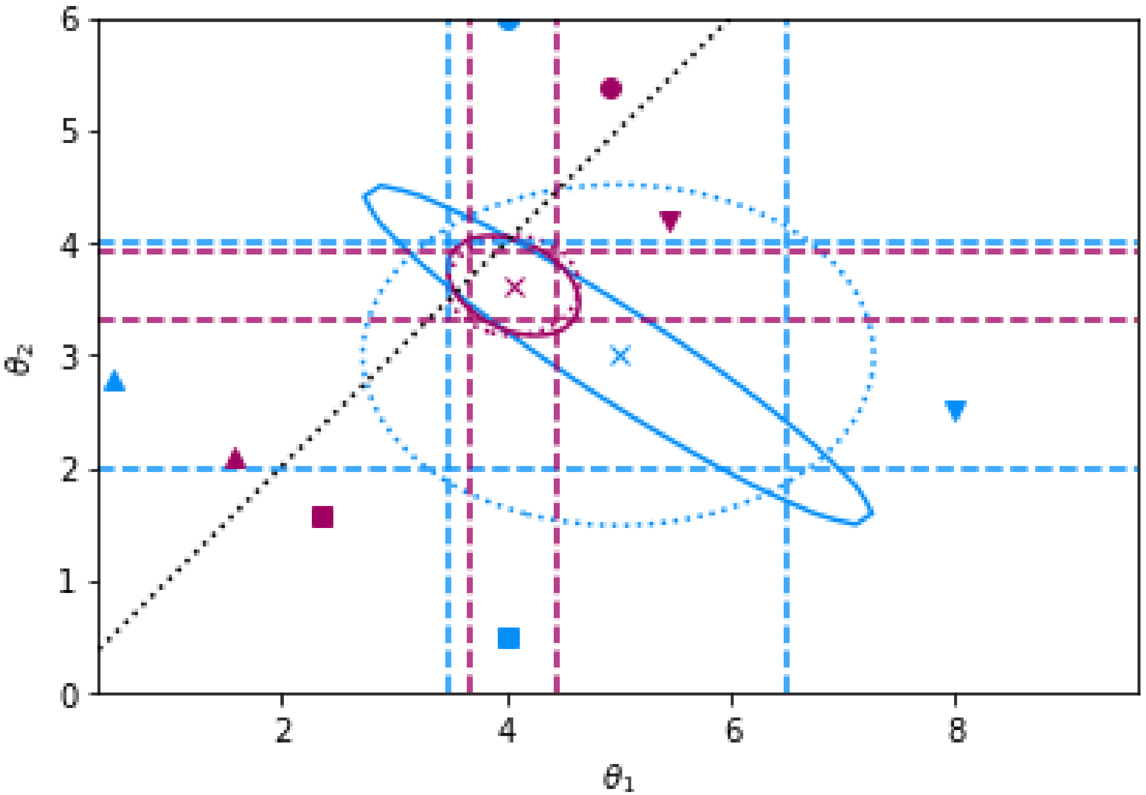
- Can probably ignore reg-bias!
  - Reg-bias shows difference between unreg. result and reg. result including covariance
  - Only meaningful if people use reg. cov. for model comparisons
  - But in that case could use unreg. result + cov. as well
    - Or equivalently reg. result + cov. + A matrix

Summary II

- Redefine aim of regularisation:
  - Not a statistical tool, but a data visualisation tool
    - Or maybe a bit of both
    - Uses prior assumptions to select subset of compatible results
  - Make plots less misleading
  - Reduce difference between implied uncorrelated distribution and actual, correlated, unregularised distribution
- Use W-metric as measure of that difference
  - “Objective” optimisation target

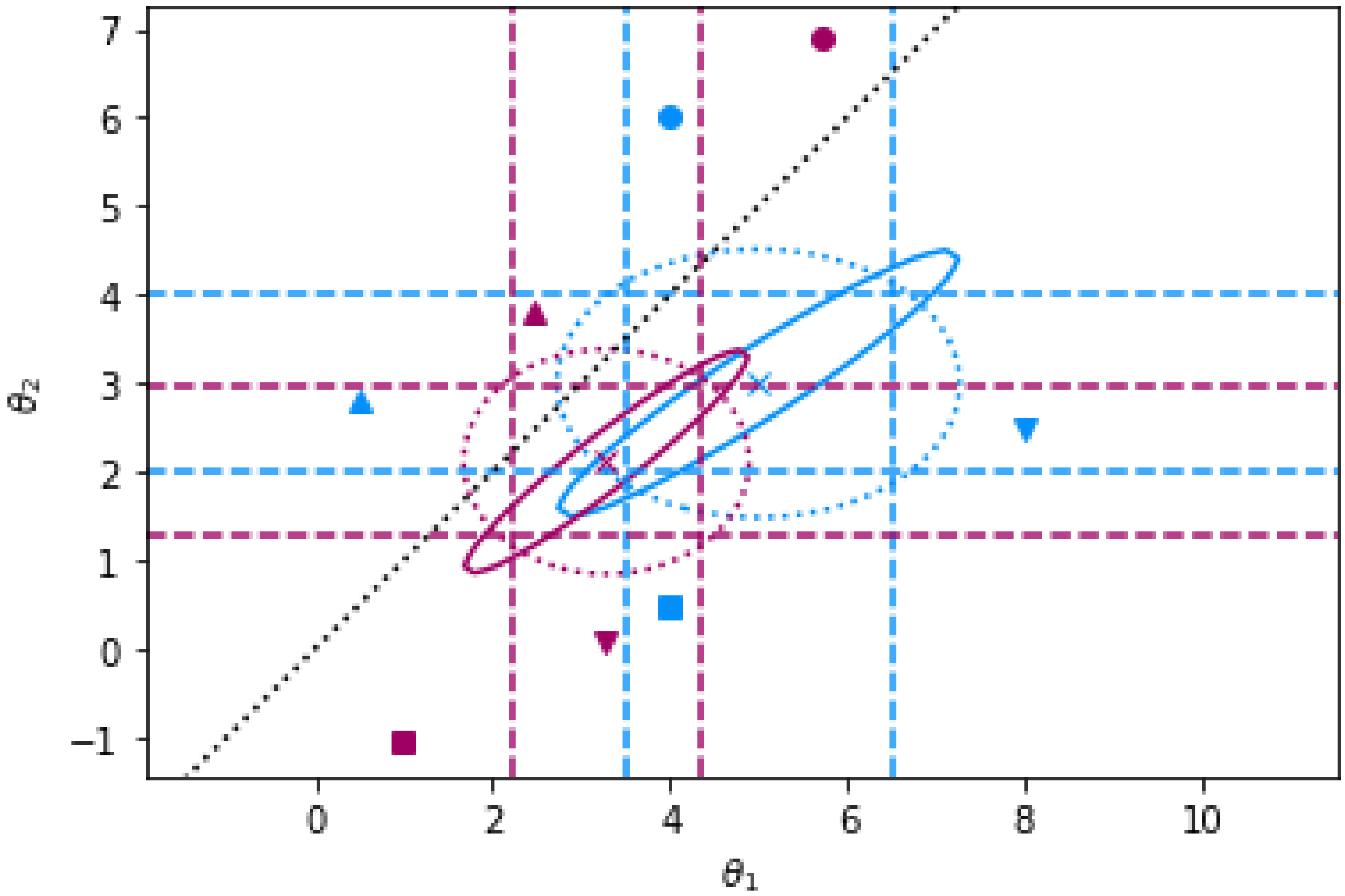
UNIVERSITY OF OXFORD

- Should use A to calculate chi2, but what about plots?
  - Show original models with regularised data?
  - Show models folded through A?
- Let's investigate 2D (= 2 bin) case
  - Regularisation pulls towards x=y
  - In this case, no big difference of conclusions



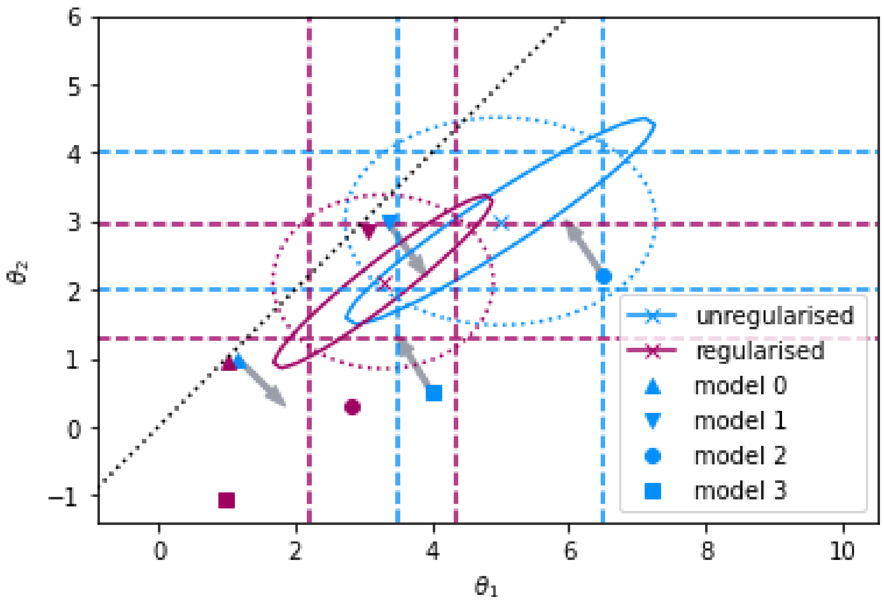
Now anticorrelated

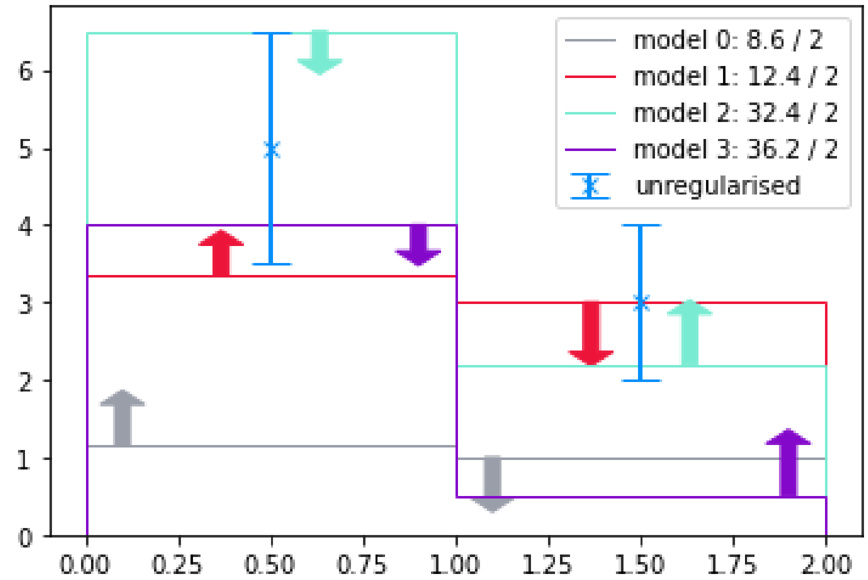
- Conclusions can be very different!
  - Seems to occur when model is less “regular” than data
  - Model gets shifted by A more than data itself



Add local gradient?

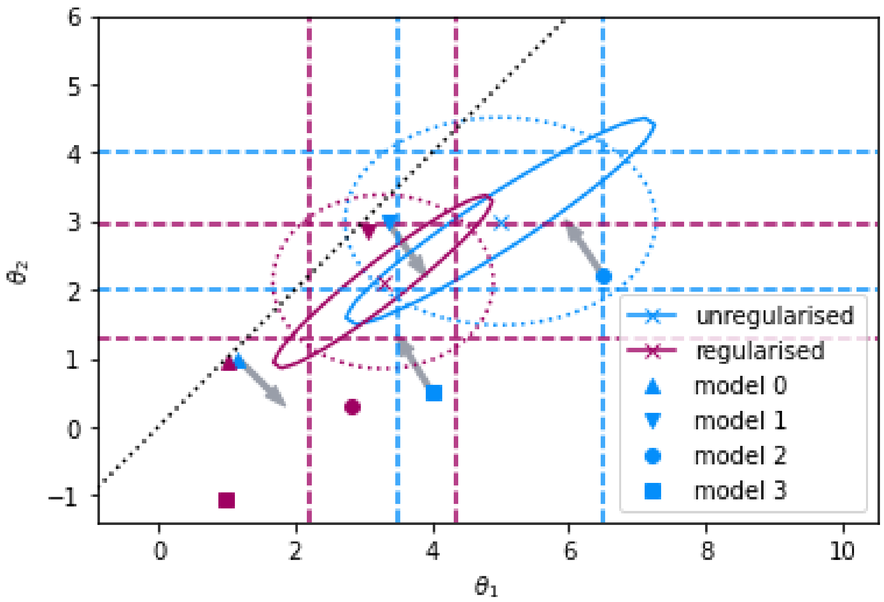
- Can add information about shape of likelihood surface
  - At least locally around the model
- Better conclusions when looking at plots of correlated data?

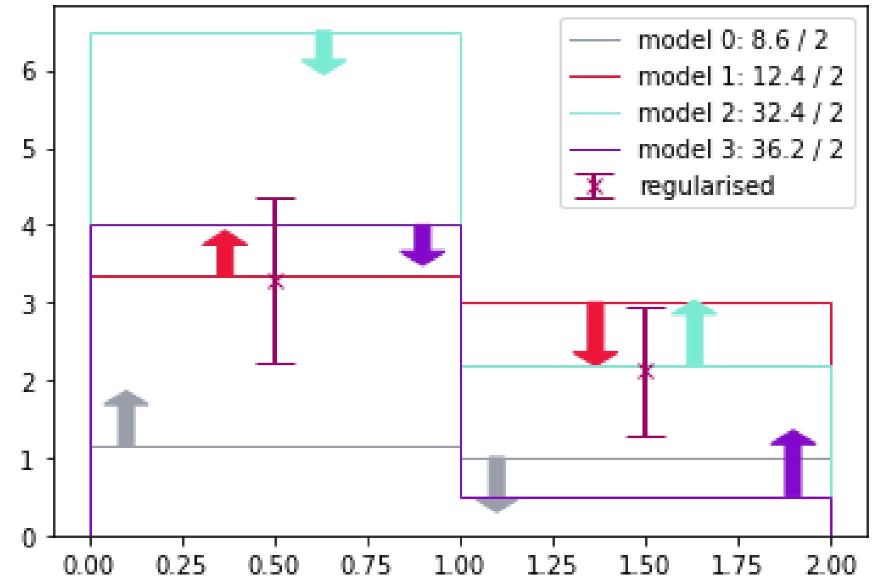




Add local gradient?

- Can add information about shape of likelihood surface
  - At least locally around the model
- Better conclusions when looking at plots of correlated data?
- Should always use gradient of unreg. chi2!





Summary III

- When plotting models, probably best to always plot original ones
  - "regularised" ones can become very strongly distorted
  - Regularisation designed to best describe original result in plot!
    - More visualisation tool than statistics tool
- Should still use A matrix to calculate correct chi2
  - And add the number to the plot!
- Local Likelihood gradient around models could add additional information
  - Calculate with unregularized data!
    - Or equivalently regularised data and A

$$\text{grad} \parallel \; A^T V'^{-1}(\hat{\theta}' - A\theta)$$

  - Can show together with either regularised or unreg. data

Summary summary

- Can achieve Tikhonov regularisation for any result after the unregularized fit/unfolding
  - No knowledge about unfolding procedure required
  - Fast linear algebra
- Regularisation should probably be seen as data visualisation tool
  - Define aim to make least misleading plots
  - Use Wasserstein distance to quantify difference between shown, implicitly uncorrelated distribution and unregularized result
- Plots should always show original models
  - Adding local gradient information can help interpreting differences between data and model

Backups

A new perspective

- New MicroBooNE pre-print XSEC measurements use "Wiener-SVD-unfolding"
- Method described in paper "Data Unfolding with Wiener-SVD Method" by W. Tang et al. https://doi.org/10.1088/1748-0221/12/10/P10002
- Unfolding by matrix inversion
  - Or rather pseudo inverse, applicable to non-square matrices
  - Leads to the usual sensitivity to stat. fluctuations and corr.
- Regularisation by applying an "additional smearing matrix", A, to the result
  - Reduces anticorrelations
- Regularised result + A = same information as unreg.
  - Coordinate transformation: truth space → pretty plot space

- Tikhonov matrix C: $Q = C^T C$
- C1

$$P = \sum_i (x_i - x_{i+1})^2$$

- C2

$$P = \sum_i (x_i - 2x_{i+1} + x_{i+2})^2$$

- “template scaling” on XSEC result

$$x_i \rightarrow x_i / m_i$$

  - m becomes part of Q

- Minimise plot-bias directly?
    - Does not move central values
    - Only scales errors