Surprise talk

(in replacement of Tommaso Dorigo)

Unfolding in experimental particle physics

i.e., ill-posed linear inverse problems (for dummies)

Outline

- I: Unfolding: the basics
- II: *"...and then I accidentally divided by 0"*
- III: *"Let a hundred flowers bloom and a hundred schools of thought contend"*
- IV: *"Thou shalt not unfold"*
- V: In other fields (an example)

I: Basics

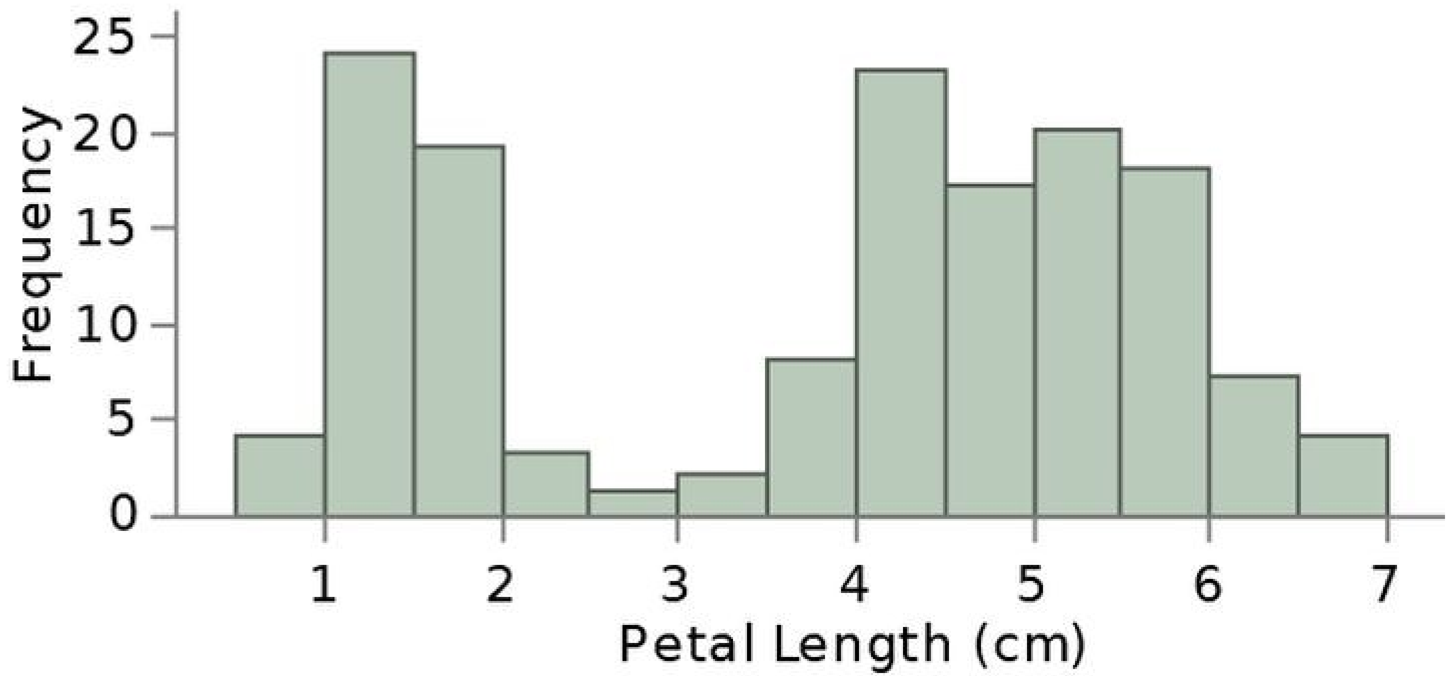
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

determinant

Picture from www.mathsisfun.com

This is the most complex math that you need to know to understand this talk

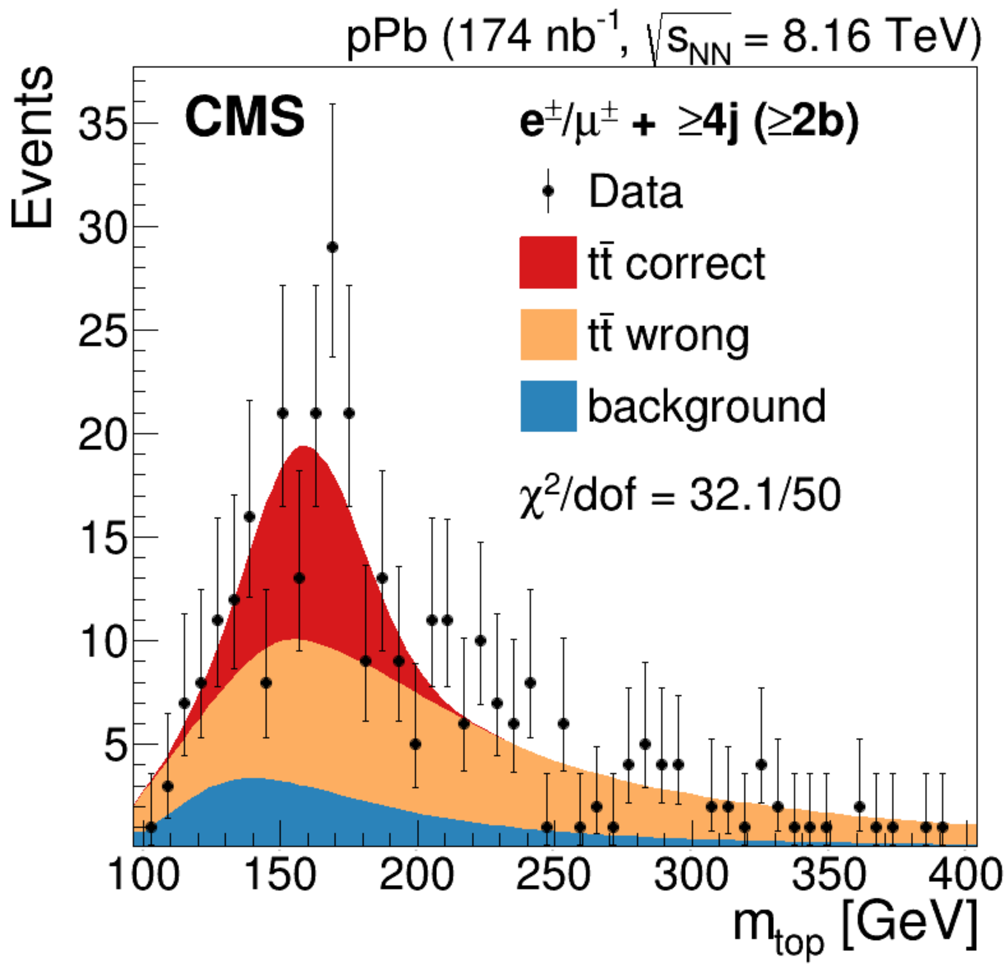
This is an histogram



Think of it as a vector

(it can be multi-dimensional, but the math is the same)

How we *usually* do analysis



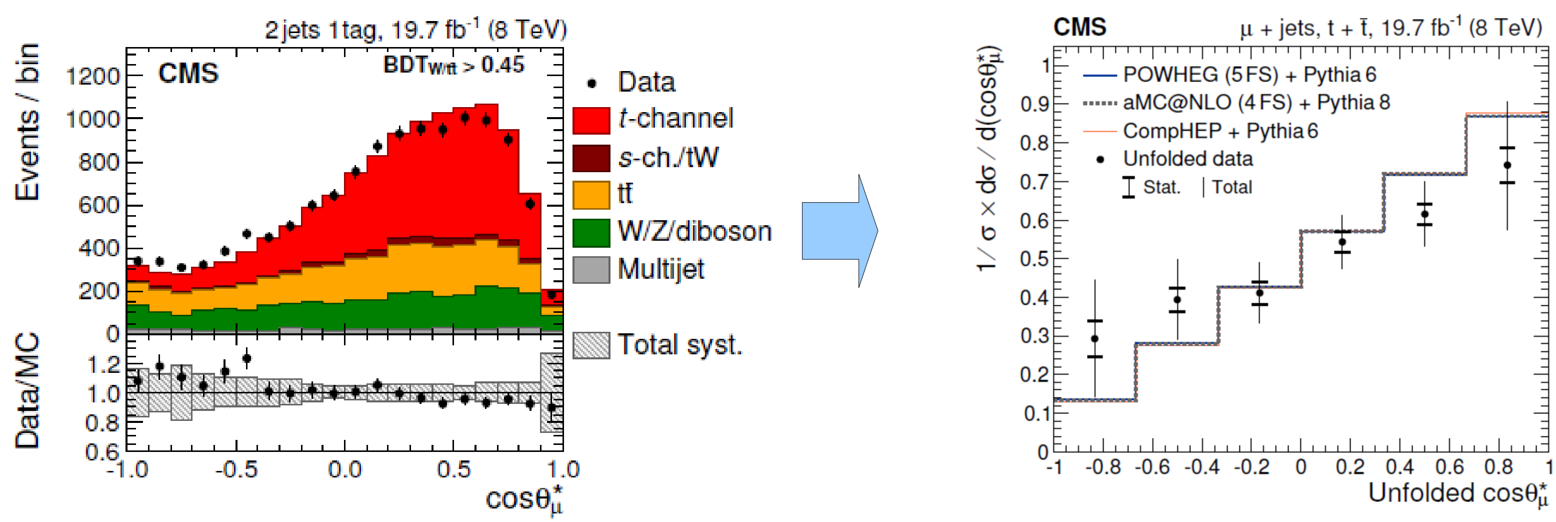
CMS Collaboration (G.Krintiras, AG, et ~2000 al.), arXiv:1709.07411 [nucl-ex], accepted by PRL

- These are *raw data*
  - This is not their "native" distribution because it is *smeared* by detector resolution, selection bias, backgrounds, etc.
- We compare data to *models*
  - Our models must include the same *smearing*
- From the comparison we want to understand stuff

How we *occasionally* do analysis

In some cases, useful to report some data distribution after correcting for the known sources of smearing:

CMS Collaboration (M.Komm, AG, et al.), JHEP 04 (2016) 073



This is what we call *unfolding*.

Why do we do that?

(Answer at the end, after some drama)

Unfolding = matrix equation

$$\boldsymbol{Ax} = \boldsymbol{y} \;\rightarrow\; \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y} \qquad \boldsymbol{V}_x = \boldsymbol{A}^{-1}\boldsymbol{V}_y(\boldsymbol{A}^{-1})^T$$

- $\boldsymbol{x}$ = n-histogram of true variable x
- $\boldsymbol{y}$ = m-histogram of measured variable y
- $\boldsymbol{A}$ = $m \times n$ response matrix
- $\boldsymbol{V}_x$ = error propagation for true variable x
- $\boldsymbol{V}_y$ = error of measured variable y

Imagine, e.g., that you want to infer ratio of spin-up vs spin-down particles via a measurement of some asymmetry. You count, e.g., the forward-going and the backward-going particles. Response matrix is 2x2:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

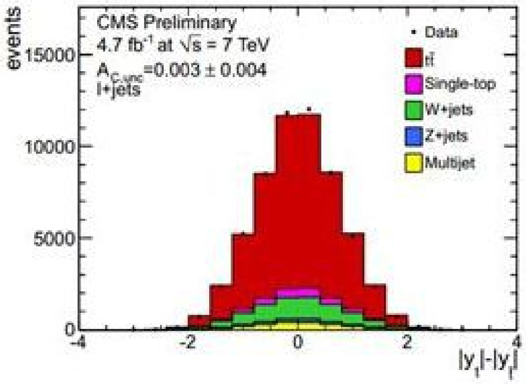
- a (d): how much of bin 1 (2) stays in bin 1 (2)
- b (c): how much of bin 1 (2) goes to bin 2 (1)

$$Ax = y \rightarrow x = A^{-1}y \qquad V_x = A^{-1}V_y(A^{-1})^T$$

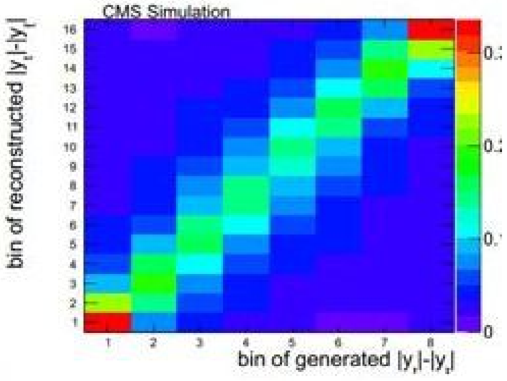
$\boldsymbol{x}$ = n-histogram of true variable x
$\boldsymbol{y}$ = m-histogram of measured variable y
$\boldsymbol{A}$ = $m \times n$ response matrix

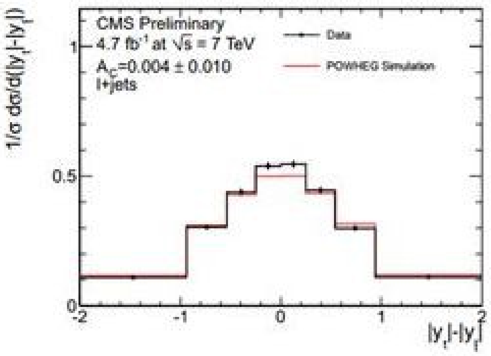
$\boldsymbol{V_x}$ = error propagation for true variable x
$\boldsymbol{V_y}$ = error of measured variable y

Reconstructed Δ|y|



Unfolding matrix

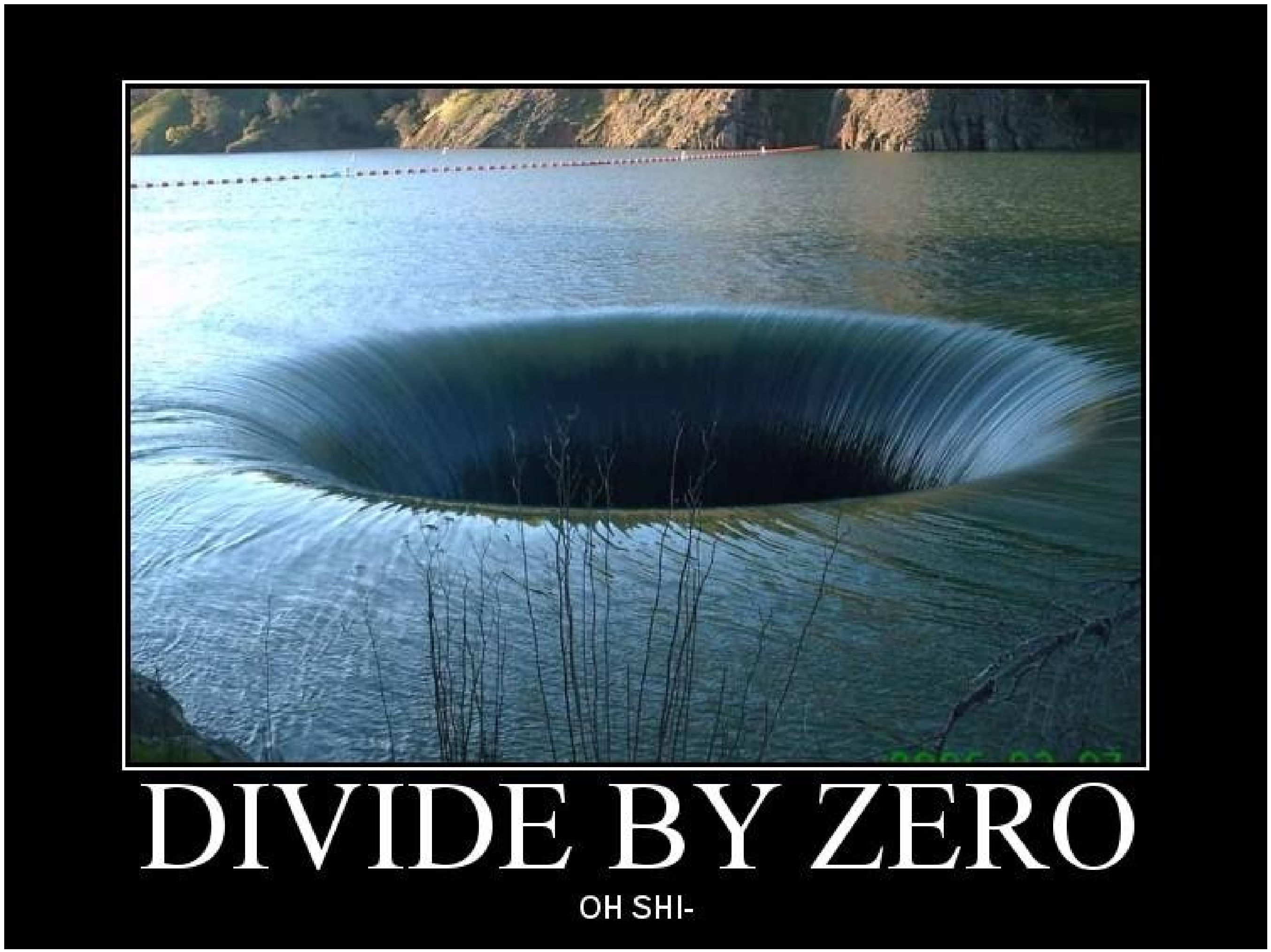




Unfolded Δ|y|
(SM MC truth)

It doesn't seem complicated.

II: "... *and then I accidentally divided by 0*"



Why the problem is said to be "ill-posed"

It is a trivial matrix equation (**y = Ax + b**), but *stochastic noise* affects y, A and b. In practice we usually maximize a likelihood (but that is conceptually equivalent to error propagation.)

No matter how you invert, anyway, nasty things can happen:

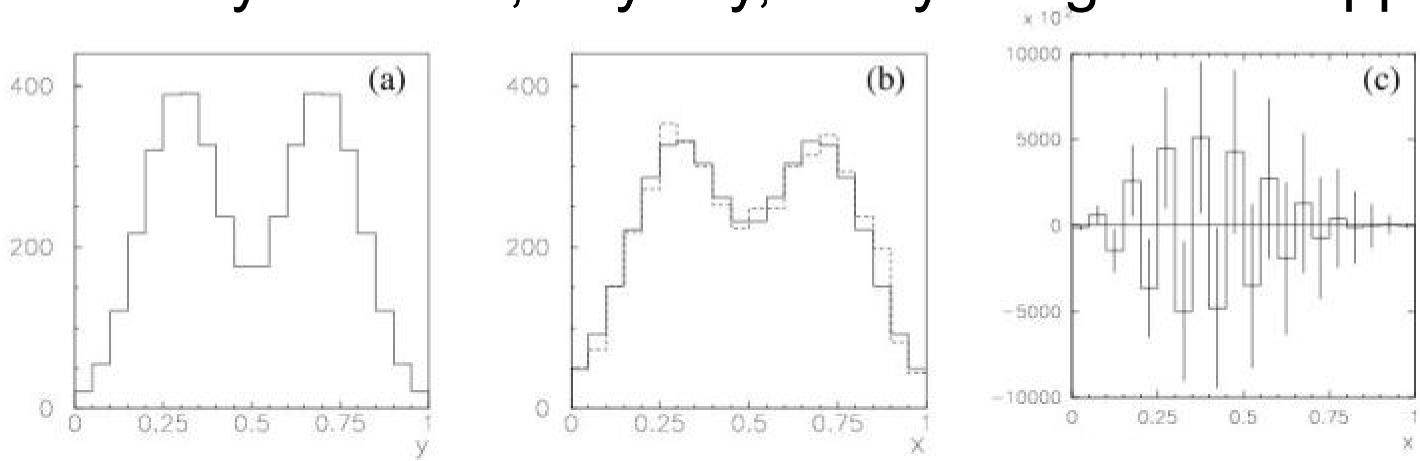


Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

(This is an extreme example, carefully designed for illustrative purposes; from G.Cowan, Conf.Proc. C0203181 (2002) 248-257)

Why the problem is said to be "ill-posed"

The last plot is supposed to be an unbiased estimate of the first.

Indeed, *it is* unbiased: if you run millions of simulations, the average in each bin does not deviate from the expected value...

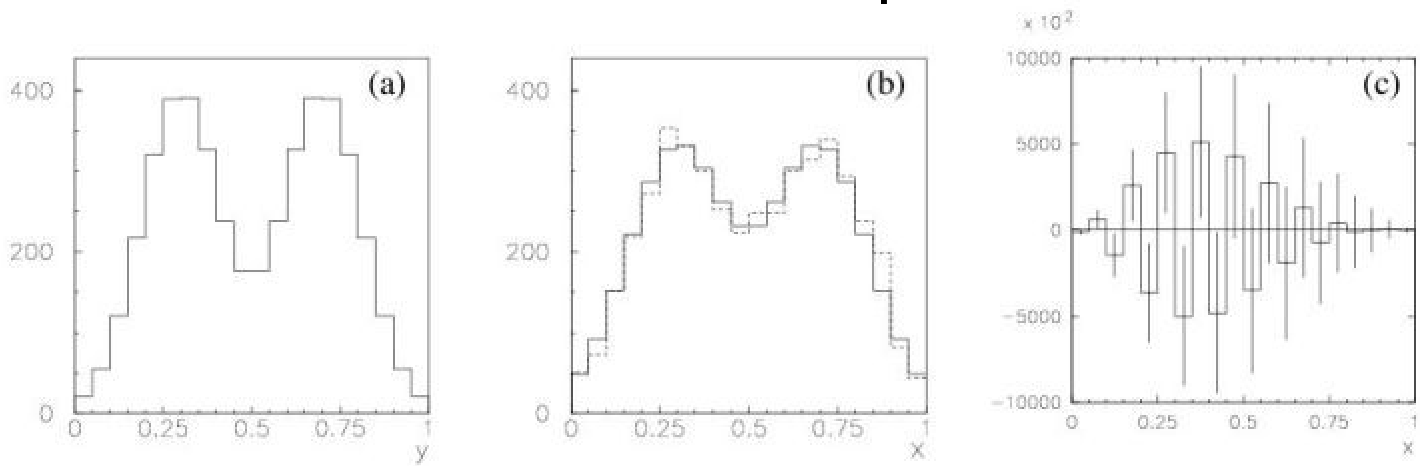


Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

But in each individual simulation, some huge high-frequency fluctuation is swamping the shape that we would like to observe.

Analogy from electronics: we are amplifying the noise

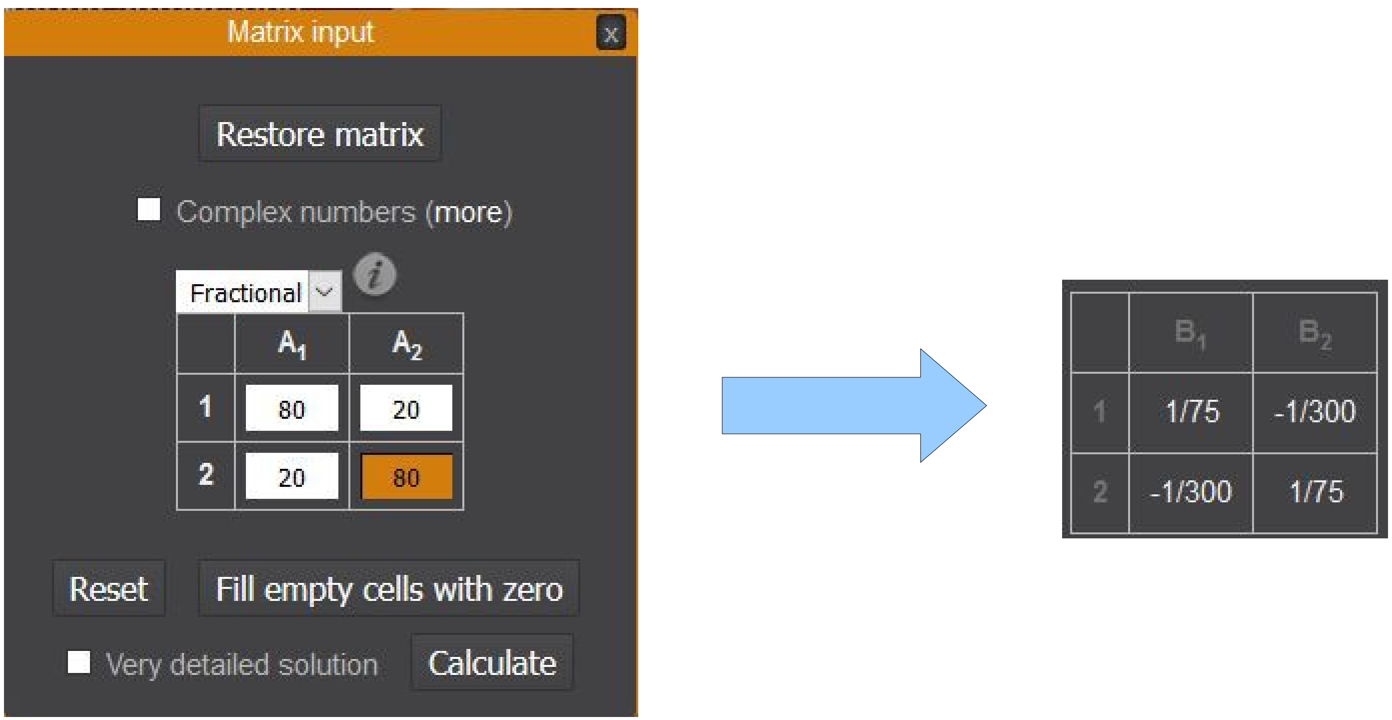
Back to the basics

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

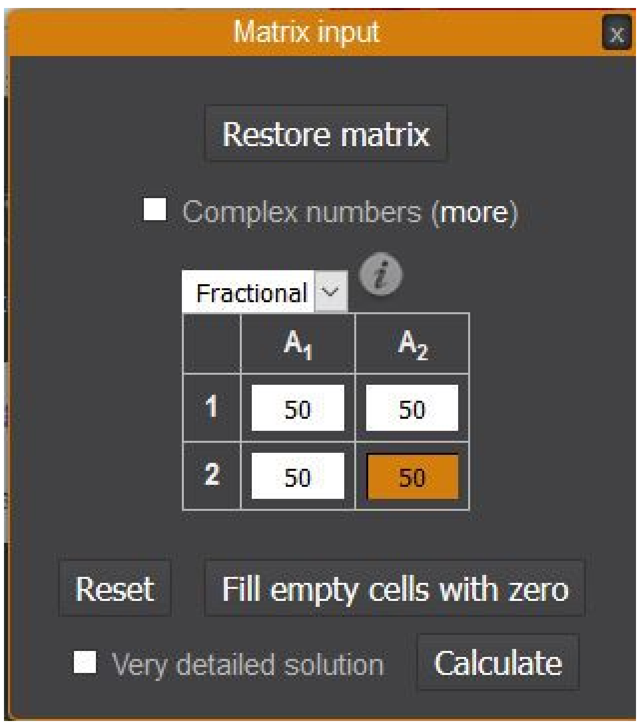
determinant

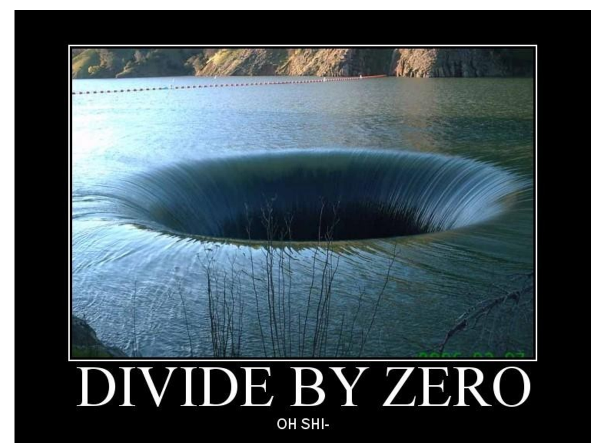
Back to the basics

Who needs to study math, when there are computers?
http://matrix.reshish.com/inverse.php



And then I accidentally...



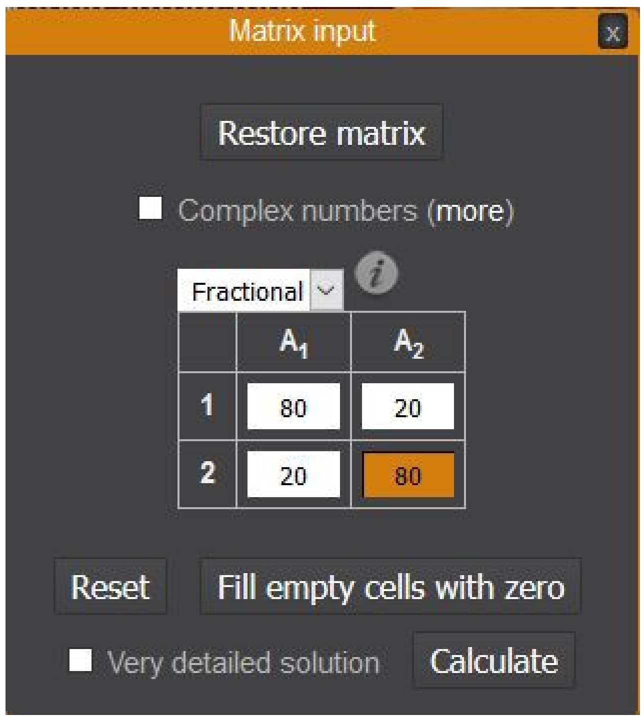


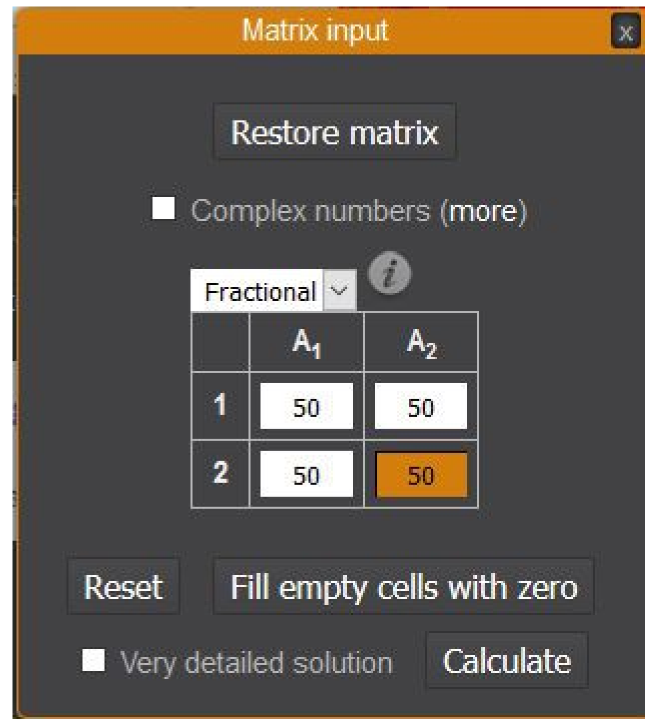
Why you still need to study math

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

determinant

And why you also need some physics sense





Going back to my example: you want to infer ratio of spin-up vs spin-down via a related angular asymmetry. You count the forward-going and the backward-going particles.

Left: you get the direction wrong 20% of the times. Fine.

Right: you get it wrong 50% of the times. Your detector, or your observable, has no sensitivity to the quantity of interest!

Approaching the singularity

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$$

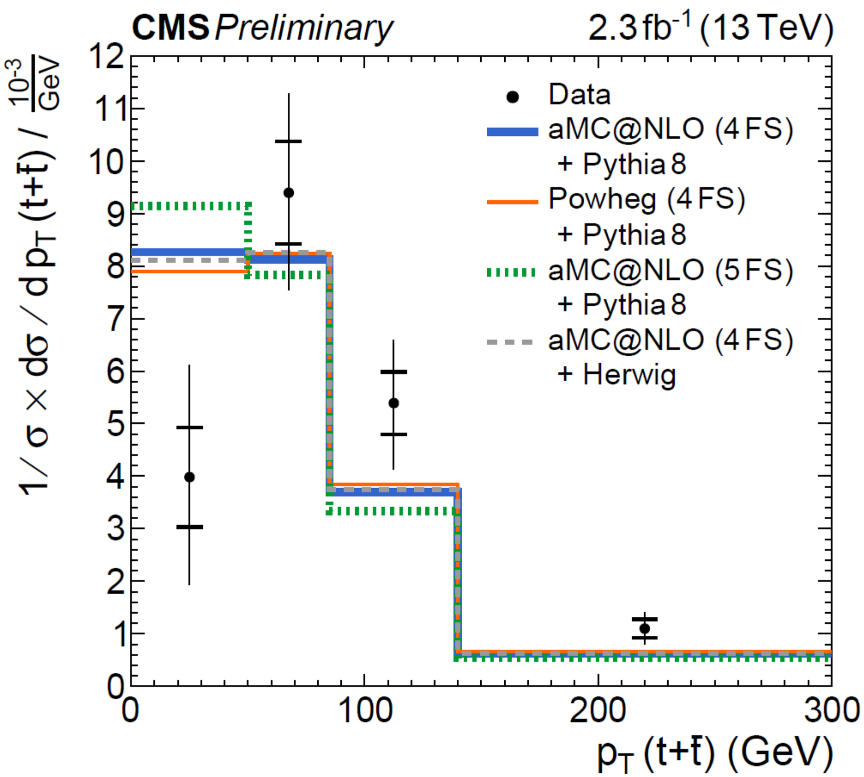
$$\begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \Rightarrow \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix} \Rightarrow \begin{pmatrix} 25.5 & -24.5 \\ -24.5 & 25.5 \end{pmatrix}$$

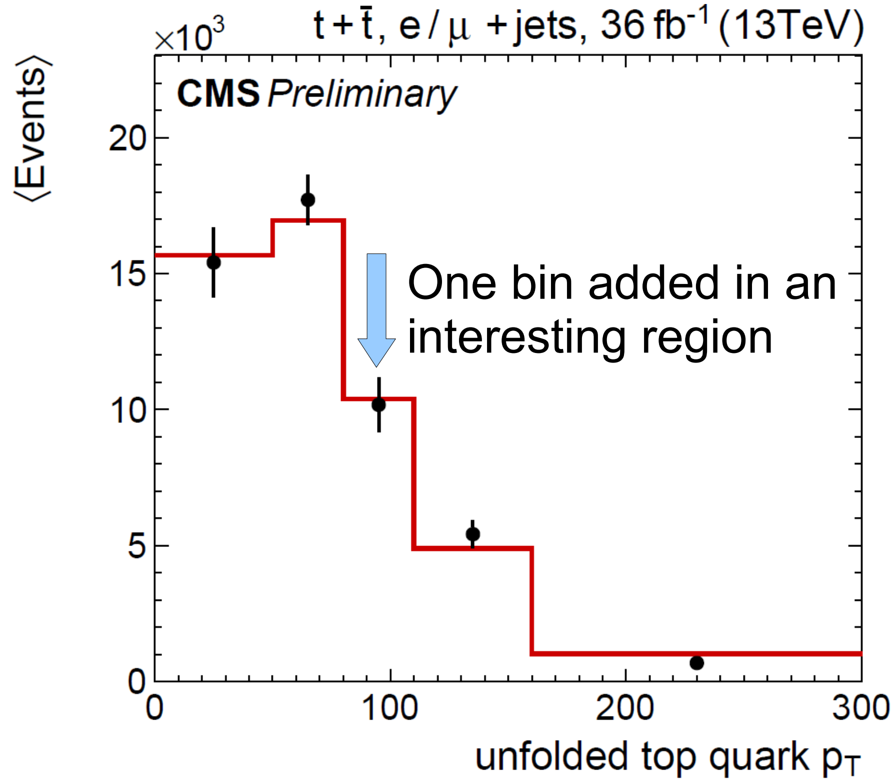
Suppose that all elements of the matrix have an uncertainty of ±0.01 (you estimate from MC samples; their statistics is finite)

Anyway, this was just for illustration. In 2x2 inversion, you get into trouble only when the resolution of your detector is so poor that you would not make the measurement anyway.
Let's now consider more bins.

Why we want to have many bins



CMS Coll. (M.Komm, AG, et al.), TOP-16-004



CMS Coll. (M.Komm, AG, et al.), in progress

The more data you have, the more you can afford to divide the sample in a larger number of bins, to achieve a more fine-grained understanding of the features in the spectrum

Doing math at a glance

$$M = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$$

This can be inverted

$$M = \begin{pmatrix} 0.80 & 0.20 & 0 & 0 & 0 \\ 0.20 & 0.80 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

This one too

$$M = \begin{pmatrix} 0.80 & 0.20 & 0 & 0 & 0 \\ 0.20 & 0.80 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

This is singular

$$M = \begin{pmatrix} 0.80 & 0.20 & 0 & 0 & 0 \\ 0.20 & 0.80 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \end{pmatrix}$$

This one too

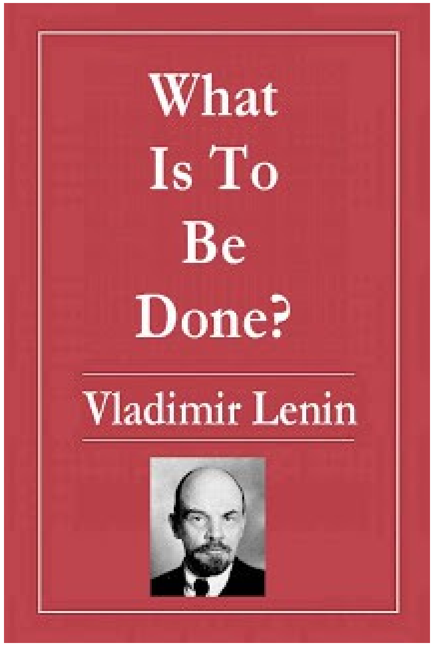
Not always so simple

$$M = \begin{pmatrix} 0.80 & 0.20 & 0 & 0 & 0 \\ 0.20 & 0.80 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.2 & 0.4 \\ 0 & 0 & 0.1 & 0.6 & 0.4 \\ 0 & 0 & 0.4 & 0.8 & 0.8 \end{pmatrix},$$

This is singular: last three columns (or rows) are in a linear relationship

But at a glance it is not so obvious

Imagine a very large matrix; consider that it is populated randomly; imagine how often you can get accidentally close to singularity in some of its sub-matrices


What
Is To
Be
Done?
Vladimir Lenin

III: *"Let a hundred flowers bloom, and a hundred schools of thought contend"*

100 Flowers Movement 1956-1957

Under the slogan "*Let a Hundred Flowers Bloom and a Hundred Schools of Thought Contend*," Mao actively encourages the opinions and criticisms from all walks of society, including intellectuals

(the name is a reference to the *Hundred Schools of Thought* period of the Warring States Period-Confucius, Lao Tze, etc).

Slide from: http://slideplayer.com/slide/9721247/

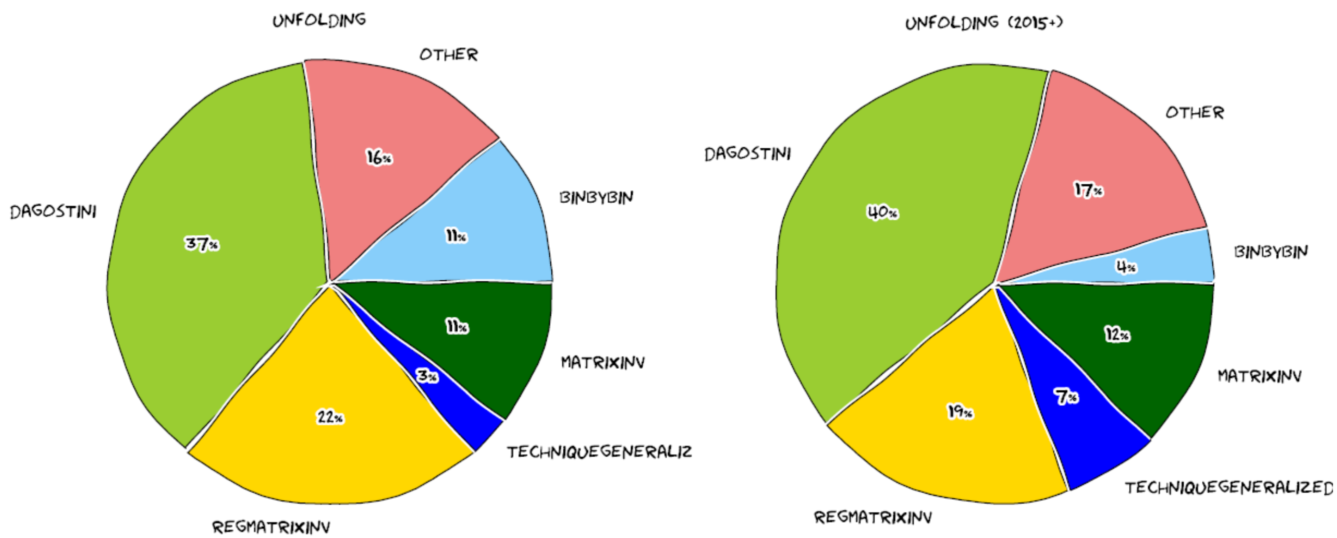
Problem is ill-posed → multiple solutions are possible

But we need to choose one! Matter of opinion?

(And physicists are good at being *opinionated*)

Unfolding

- Which method is more popular for unfolding?
  - RegMatrix -> SVD
  - TechinqueGen -> TUnfold



Andrea Carlo Marini 28 Nov 2017 7

What is to be done

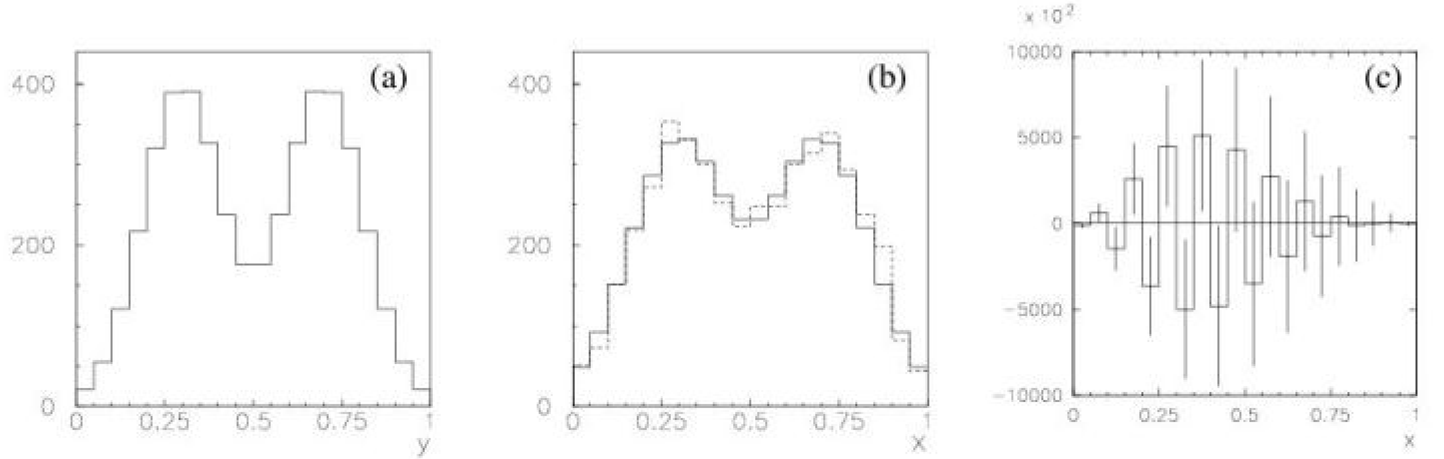


Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

Striking fact #1: in the third plot the variance of each bin is huge (orders of magnitude larger than the bin contents in the first and second plot), despite the fact that the author applied a maximum likelihood estimator, which is guaranteed to give the *smallest possible variance for an unbiased estimator* (see Kyle Cranmer's lectures last week).

What is to be done

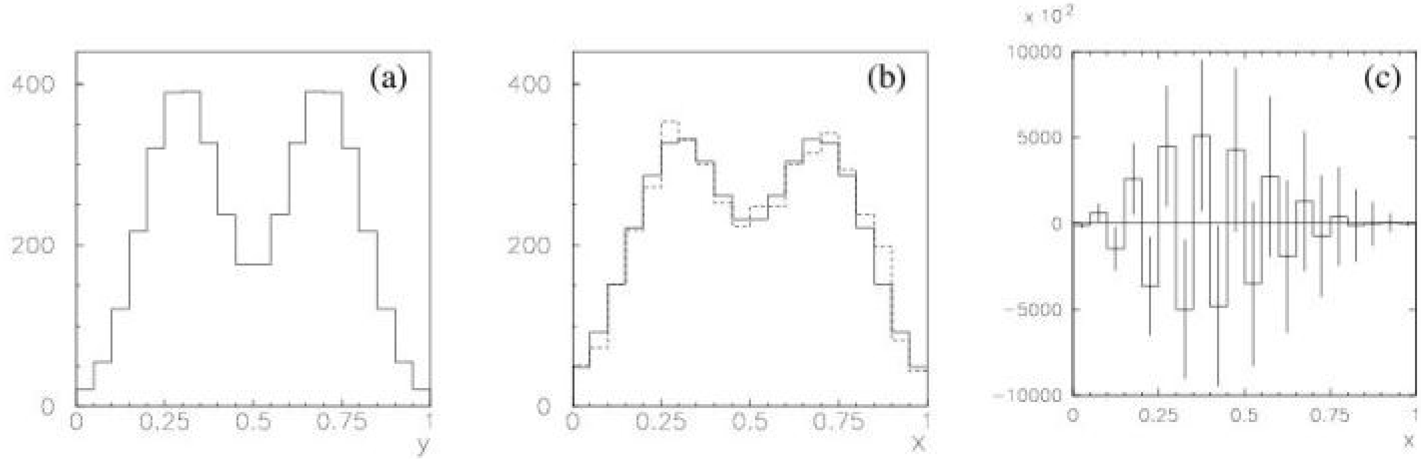


Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

Usually we strive to avoid (or at least minimise) bias.
But maybe, after all, bias is not so bad, if in the end the quadratic sum of bias and standard error is a reasonable number (and if you have ways to estimate the bias and account for that as an additional error component).

What is to be done

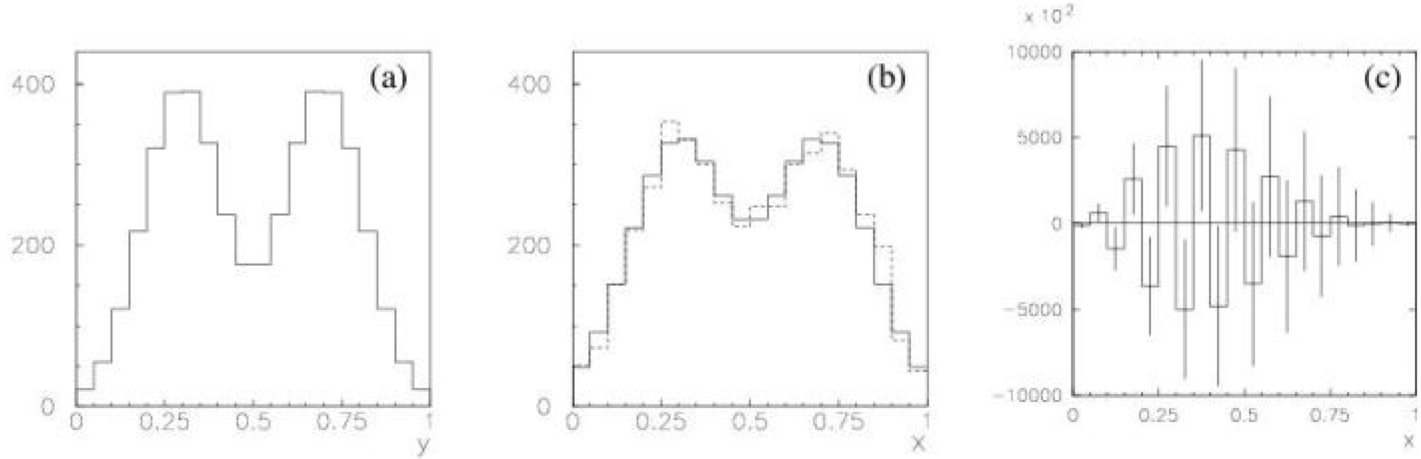
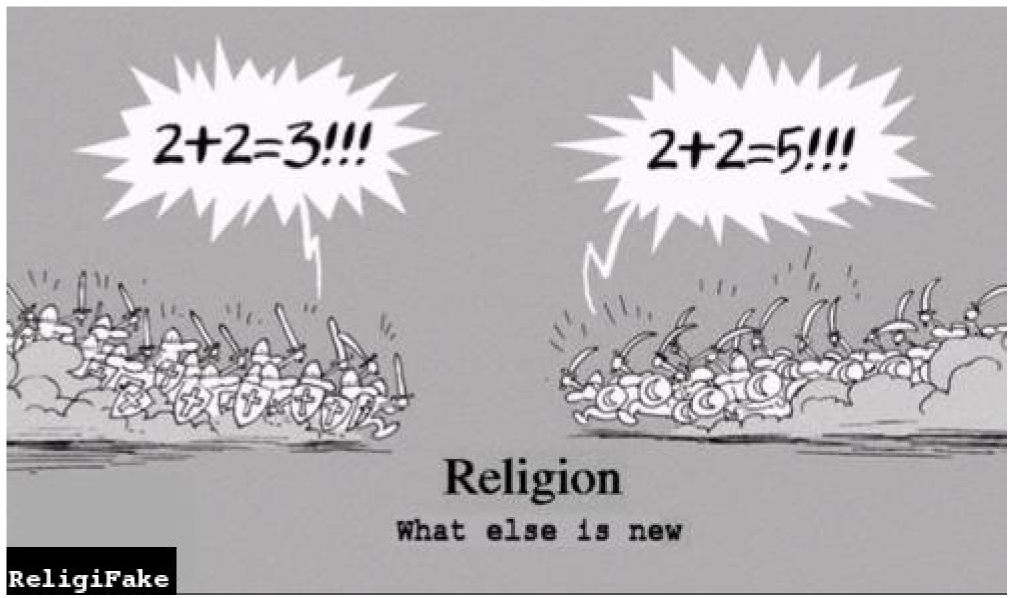


Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

Striking fact #2: the histogram is "oscillating", each bin seems to be anti-correlated with its neighbours. No matter how little you know about statistics, the first time you looked at the third plot you understood that "it was wrong", just because of this funny feature.
We may want to bias the unfolded shape by imposing our prejudice that it can not be so funny...

The two big Schools: regularization vs iteration



Tikhonov regularization

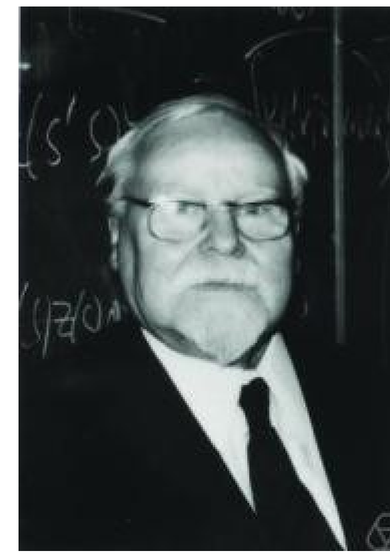
General: define a metric (‖*‖) in space of raw data

$$|| \, M \times \vec{t}_{unfold} - \vec{x} \, ||^2$$

Minimizing this distance as function of t_{unfold} is the same as solving the equation by matrix inversion

To solve our issue (the amplification of the bin uncertainties), introduce a damping term:

$$|| \, M \times \vec{t}_{unfold} - \vec{x} \, ||^2 + \tau \, || \, \Gamma \times \vec{t}_{unfold} \, ||^2$$



Tikhonov

Damping the high frequencies

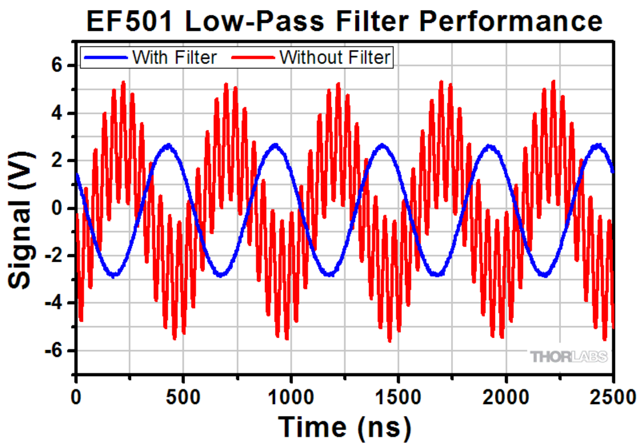
$$\| M \times \vec{t}_{unfold} - \vec{x} \|^2 + \tau \| \Gamma \times \vec{t}_{unfold} \|^2$$

$$\Gamma = \begin{pmatrix} +1 & -1 & 0 & 0 & \cdots & & \\ 0 & +1 & -1 & 0 & \cdots & & \\ 0 & 0 & +1 & -1 & \cdots & & \\ & \vdots & & & \ddots & & \\ & & & 0 & +1 & -1 & \end{pmatrix}$$

This matrix is a popular choice for regularization.
It addresses the *striking fact #2*: inserted in the penalty term of the distance above, it damps the high frequency features in the unfolded vector.

It is a sort of *first derivative* in a discretized space.

If you like electronics: it is a *low-pass filter*.



Damping the high frequencies

$$\| M \times \vec{t}_{unfold} - \vec{x} \|^2 + \tau \| \Gamma \times \vec{t}_{unfold} \|^2$$

$$\Gamma = \begin{pmatrix} -1 & +1 & 0 & 0 & \cdots & & \\ +1 & -2 & +1 & 0 & \cdots & & \\ 0 & +1 & -2 & +1 & \cdots & & \\ & \vdots & & & \ddots & & \\ & & & +1 & -2 & +1 \\ & & & 0 & +1 & -1 \end{pmatrix}$$

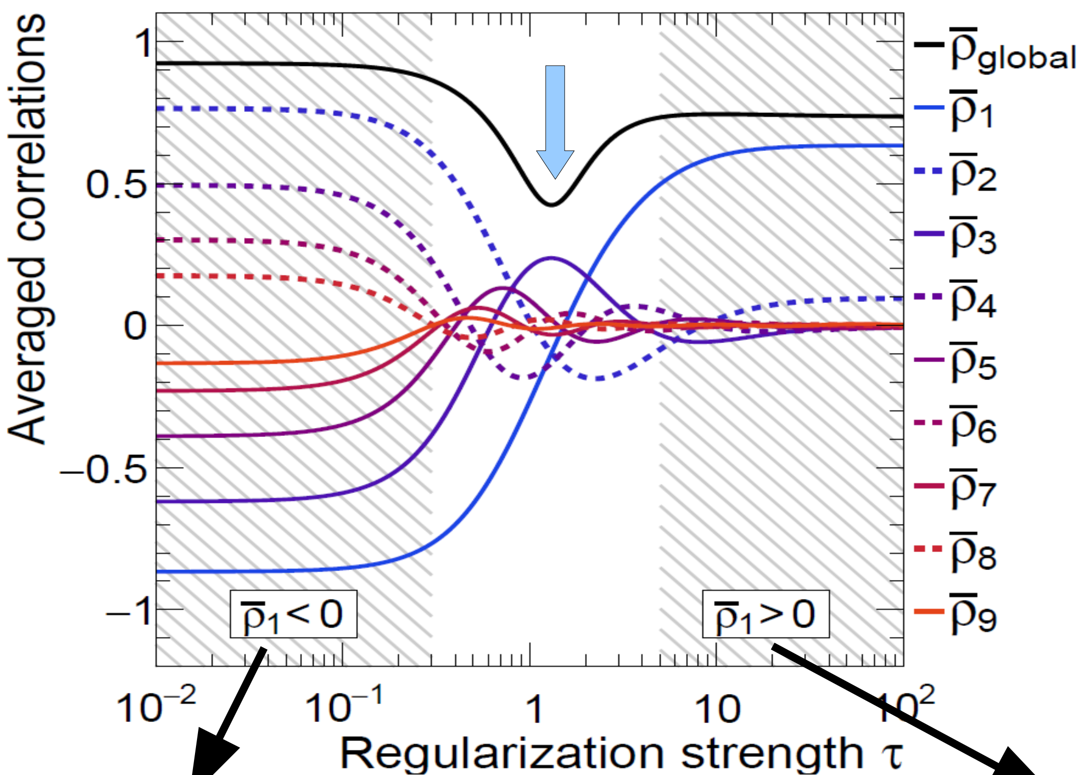
This matrix is also popular in unfolding.
It also addresses the *striking fact #2*, because it minimizes the local *curvature* of our space.

It is a sort of *second derivative* in a discretized space.

To address both the *striking facts #1* and *#2*, we need to play with parameter τ

How to choose the parameter τ

- "Subway plot" (Matthias Komm)



$$\rho_i^{\text{global}}[\tau] = \sqrt{1 - \frac{1}{\mathcal{C}_{x,ii}[\tau] \cdot (\mathcal{C}_x^{-1}[\tau])_{ii}}}$$

$$\bar{\rho}_j[\tau] \equiv \frac{1}{n-j-1} \sum_i^{n-j-1} \rho_{i,i+j}[\tau]$$

(e.g., $\rho_{i,1}$ is the correlation of i^{th} bin with its immediate neighbor, after unfolding)

negative bin-to-bin correlations
stat. fluctuations amplified
→ oscillation

positive bin-to-bin correlations
2^{nd} derivative dominates
→ solution pulled towards expectation

Iterative Bayesian unfolding
(or better, D'Agostini unfolding)

Also here we want to minimize this distance:

$$\| \, M \times \vec{t}_{unfold} - \vec{x} \, \|^2$$

D'Agostini (NIM A362 (1985) 487-498) reformulated the problem in terms of causes (true distribution, t) and effects (raw data, x):

To stay close to the application of interest, let us state Bayes' theorem in terms of several independent *causes* (C_i, $i = 1, 2, \cdots, n_C$) which can produce one *effect* (E). Let us assume we know the *initial probability* of the causes $P(C_i)$ and the conditional probability of the ith cause to produce the effect $P(E|C_i)$. The Bayes formula is then

$$P(C_i|E) = \frac{P(E|C_i)P(C_i)}{\sum_{l=1}^{n_C} P(E|C_l)P(C_l)}. \quad (1)$$

Bayes

(no pictures of D'Agostini)

Iterative Bayesian unfolding

(or better, D'Agostini unfolding)

For example, if we consider DIS events, the effect E can be the observation of an event in a cell of the measured quantities $\{\Delta Q^2_{meas}, \Delta x_{meas}\}$. The causes C_i are then all the possible cells of the true values $\{\Delta Q^2_{true}, \Delta x_{true}\}_i$.

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l)P_0(C_l)}.$$

If one observes $n(E)$ events with effect E, the expected number of events assignable to each of the causes is

$$\hat{n}(C_i) = n(E)P(C_i|E). \quad (2)$$

Bayes

Iterative Bayesian unfolding

(or better, D'Agostini unfolding)

1) choose the initial distribution of $P_0(C)$ from the best knowledge of the process under study, and hence the initial expected number of events $n_0(C_i) = P_0(C_i)N_{obs}$; in case of complete ignorance, $P_0(C)$ will be just a uniform distribution: $P_0(C_i) = 1/n_C$;

2) calculate $\hat{n}(C)$ and $\hat{P}(C)$;

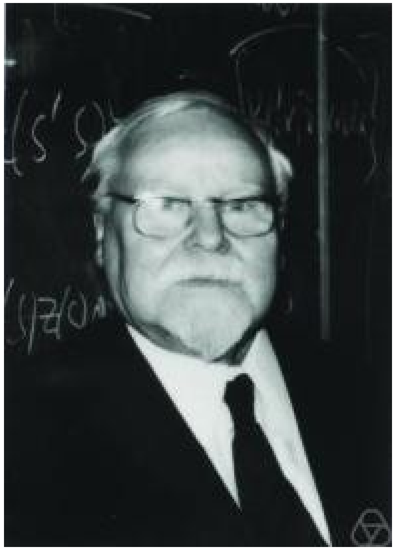
3) make a χ^2 comparison between $\hat{n}(C)$ and $n_0(C)$;

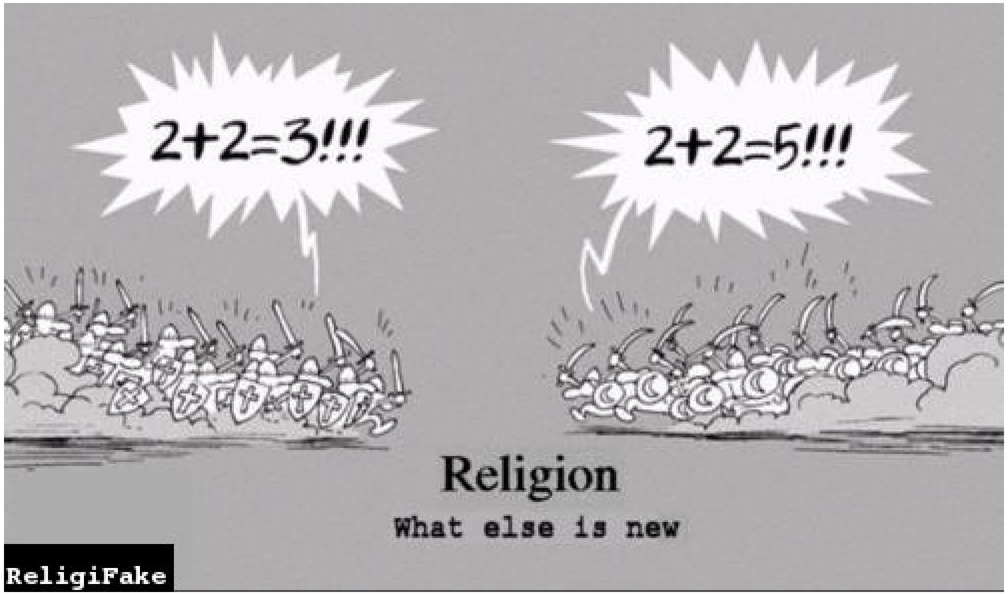
4) replace $P_0(C)$ by $\hat{P}(C)$, and $n_0(C)$ by $\hat{n}(C)$, and start again; if, after the second iteration the value of χ^2 is ''small enough'', stop the iteration; otherwise go to step 2. Some criteria about the optimum number of iterations will be discussed later.

Bayes

The number of iterations (N) in D'Agostini's method plays the same role as τ in Tikhonov regularization: N→∞ biases towards expectation, but N→0 is useless

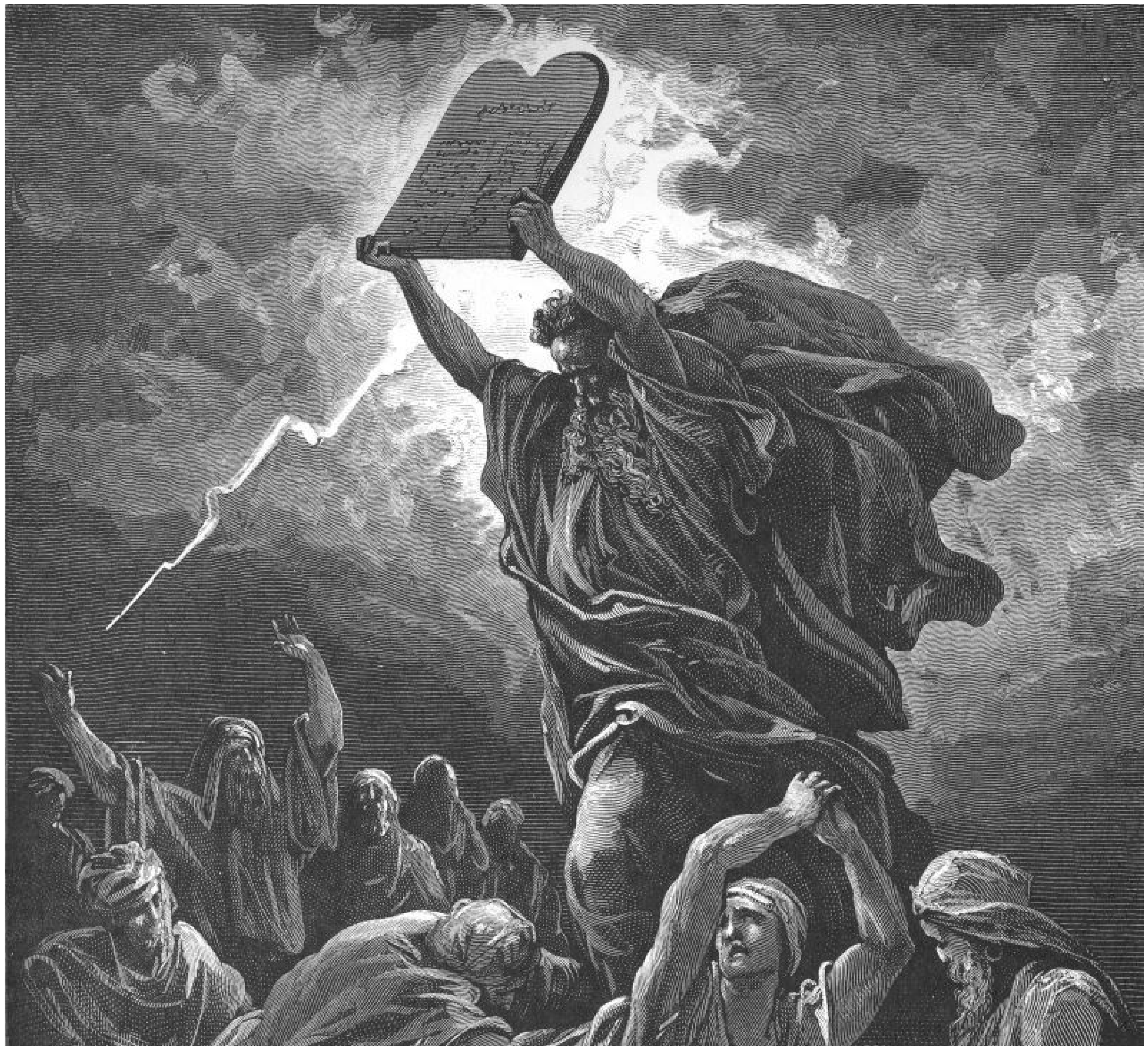
If hard-pressed to express an opinion, experts say:





Use whichever you like, but compare with the other as a cross-check.
If they agree: good. If they don't: you are in trouble.

IV: *"Thou shalt not unfold"*



The agenda...

Matrix unfolding techniques

14:00 **Introduction**
Speaker: Andreas Jung (Purdue University (US))

14:20 **Unfolding: Statistical issues**
Speaker: Mikael Kuusela (Ecole Polytechnique Federale de Lausann
Unfolding_May_201...

14:45 **Unfolding isses (TBC)**
Speaker: Igor Volobouev (Texas Tech University (US))

15:10 **Unfolding: SC recommendations put into practice**
Speaker: Olaf Behnke (Deutsches Elektronen-Synchrotron (DE))
obehnke160513.pdf

15:35 **Perspectives/Examples/other directions on unfolding**
Speaker: Juan Alcaraz Maestre (Centro de Investigaciones Energ. Medioambientales y Tecn. - (ES))

16:00 **Unfolding in the differential cross section measurement in the lepton+jets channel**
Speaker: Otto Heinz Hindrichs (University of Rochester (US))

16:40 **Toy model unfolding study for top mass**
Speaker: Fred Stober (Hamburg University (DE))

6:55 **Experience in TopModGen**
Speaker: Bugra Bilin (Middle East Technical University (TR))
bbilin_TopModGen_...

7:10 **Experience in single top**
Speaker: Matthias Komm (Universite Catholique de Louvain (UCL) (BE))

7:25 **New unfolding methods: shape constraints**
Speaker: Mikael Kuusela (Ecole Polytechnique Federale de Lausanne (CH), Univ. of California Los Angeles (US))
Unfolding_May_201...

7:45 **Hot of the press summary**

A dedicated meeting of a large analysis group in CMS. Invited speakers: four of the best unfolding experts in the Collaboration, asked to give recommendations. All of them started with the same one...

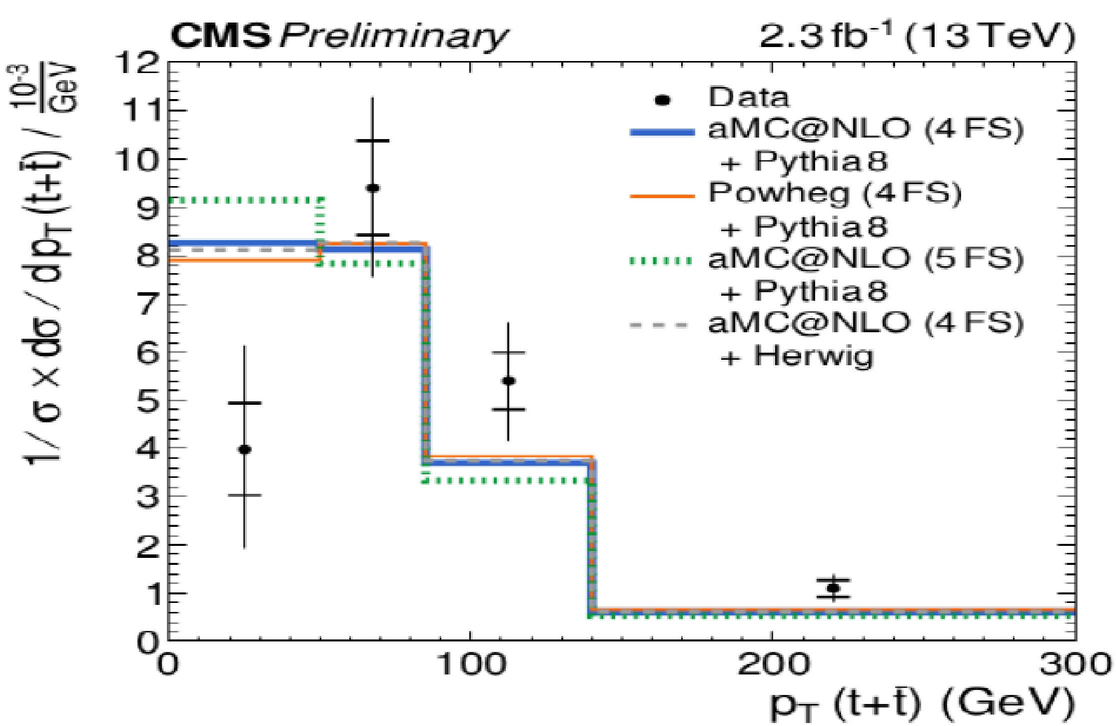
First advise on unfolding (probably useless)

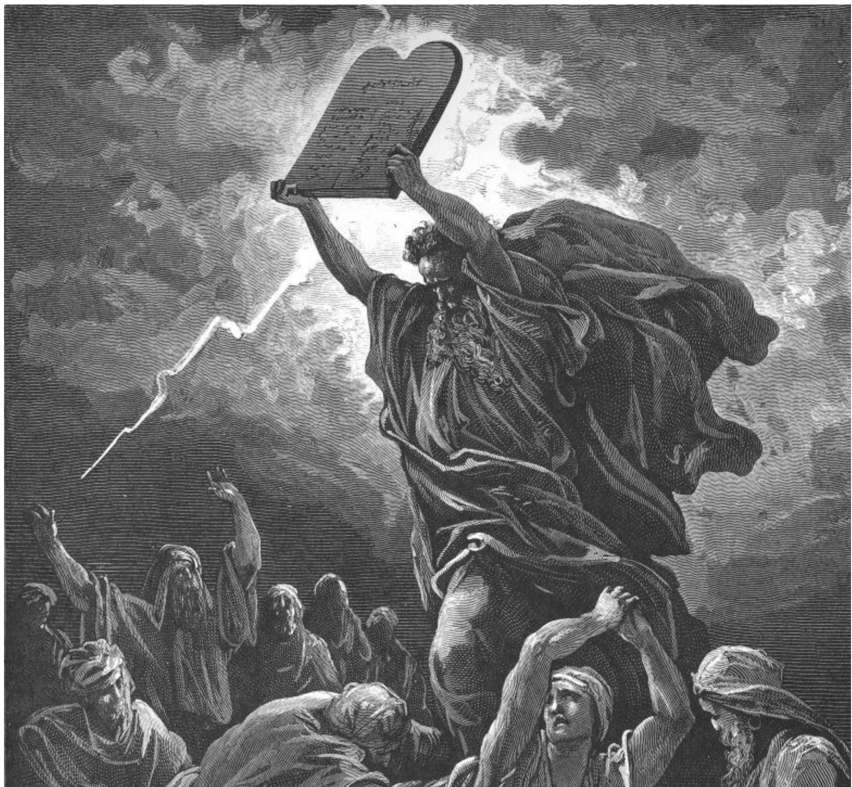
DO NOT DO IT :)

First advise on unfolding (probably useless)

DO NOT DO IT :)

If somebody needs to know the connection with the generator level, why do not you give the "response/migration matrix", from generator to reconstructed level?





Recommendations on Unfolding

Contents:

Unfolding How-to

This section describes our unfolding recommendations. They reflect our current best understanding and are likely to be updated as more experience is gained and new methods are developed. A part of the material presented here is based on the SMP Twiki pages [13].

We recommend to avoid unfolding when it is not deemed compulsory.

Why you should not unfold

- With respect to an histogram of the raw data, one in the "unfolded space" is:
  - Less sensitive to unexpected features (a discovery)
  - Inferior if the goal is a precise and accurate extraction of a parameter of the model

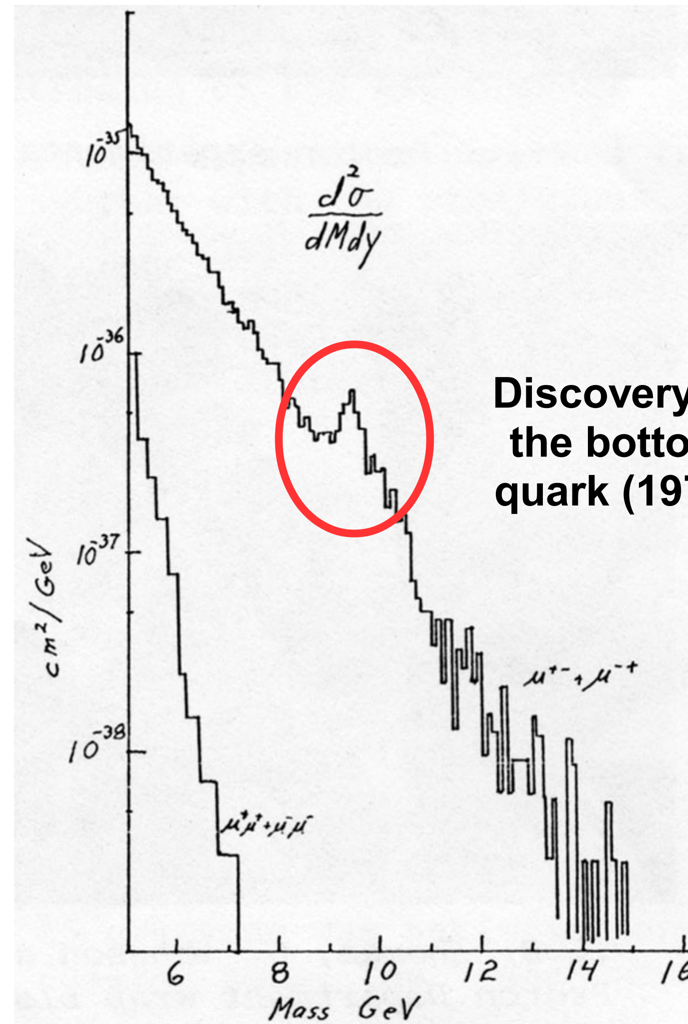
Unexpected features in data

Regularization / inversion / any method that cures the problem of high-frequency artifacts *has* to bias a bit towards expectation (= SM).

And new particles typically show up as high-frequency features:

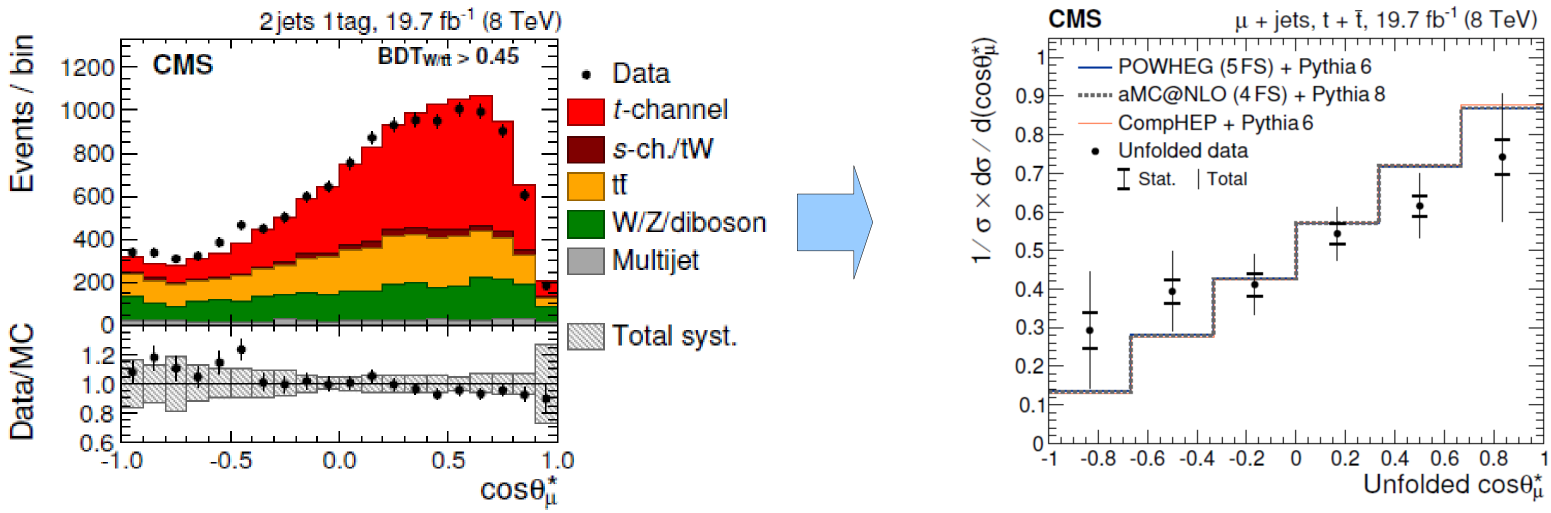
- Peaks (most frequent)
- Dips
- Peak-dips (oscillations!)

(But obviously we look at raw data way before looking at unfolded data, so that's never been a problem)



Discovery of the bottom quark (1977)

Parameter extraction



If the goal here is to extract the slope, a template fit to the raw data would be more precise: we could produce several MC samples with different values of the true slope, pass them through detector smearing, and check which slope agrees with the data better.

On the other hand, by unfolding, we can more easily verify that the relationship is truly linear, and not for example quadratic

Recommendations on Unfolding

Contents:

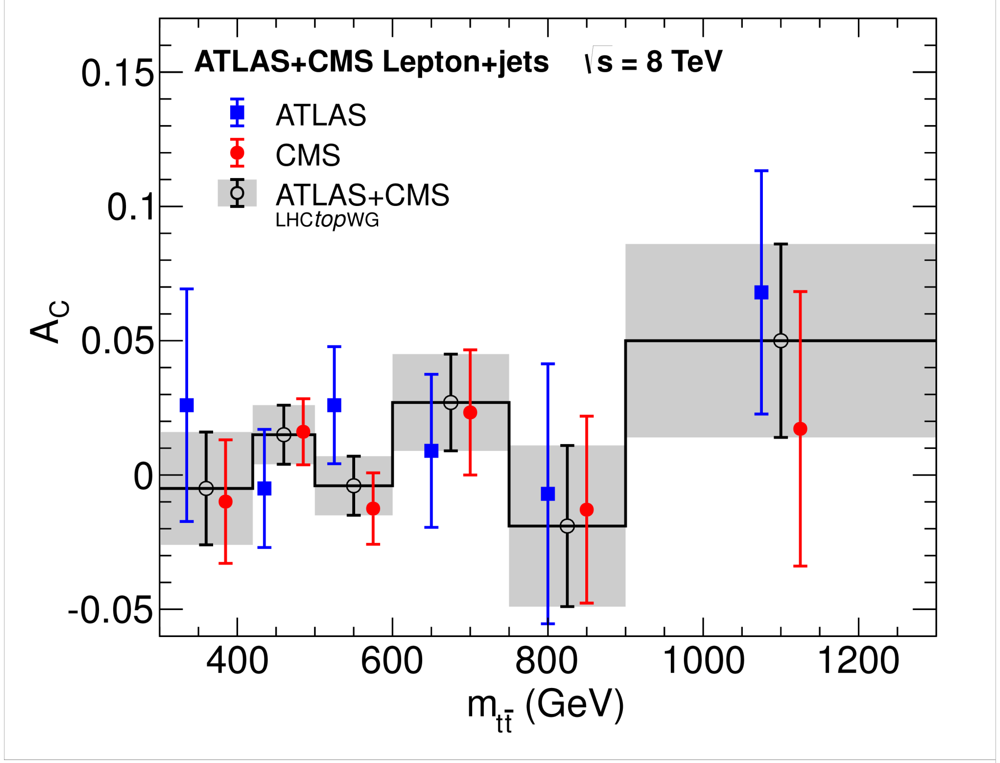
Unfolding How-to

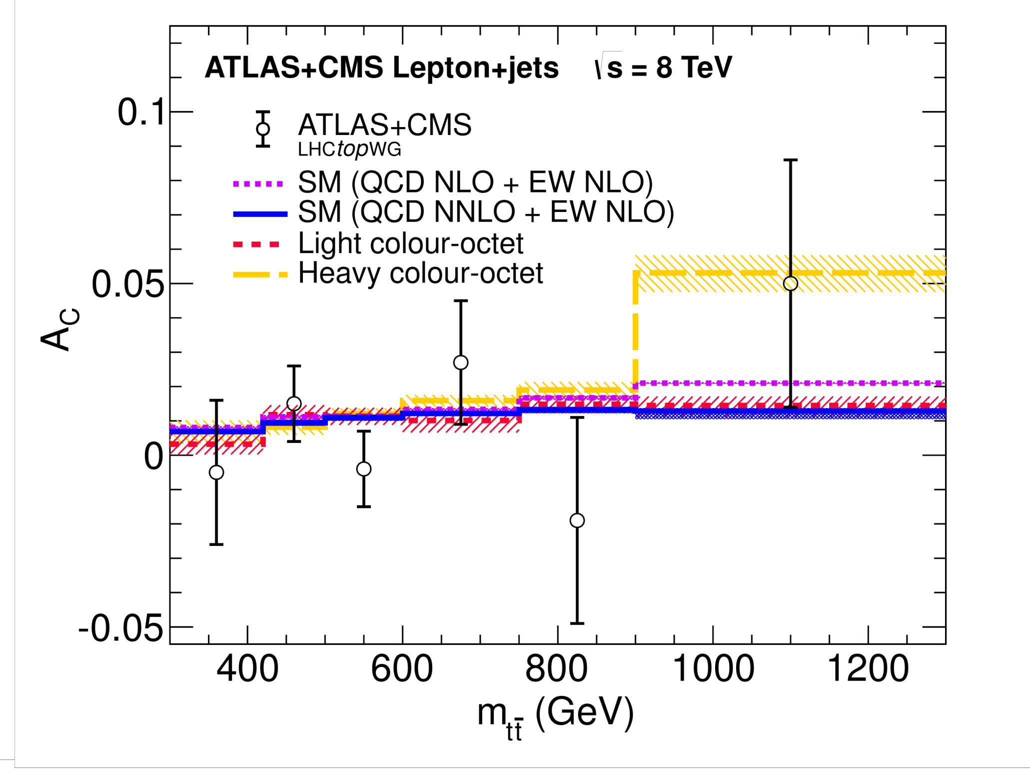
This section describes our unfolding recommendations. They reflect our current best understanding and are likely to be updated as more experience is gained and new methods are developed. A part of the material presented here is based on the SMP Twiki pages [13].

We recommend to avoid unfolding when it is not deemed compulsory.

To compare two experiments, and to combine them

CMS coll., arXiv:1709.05327 [hep-ex]



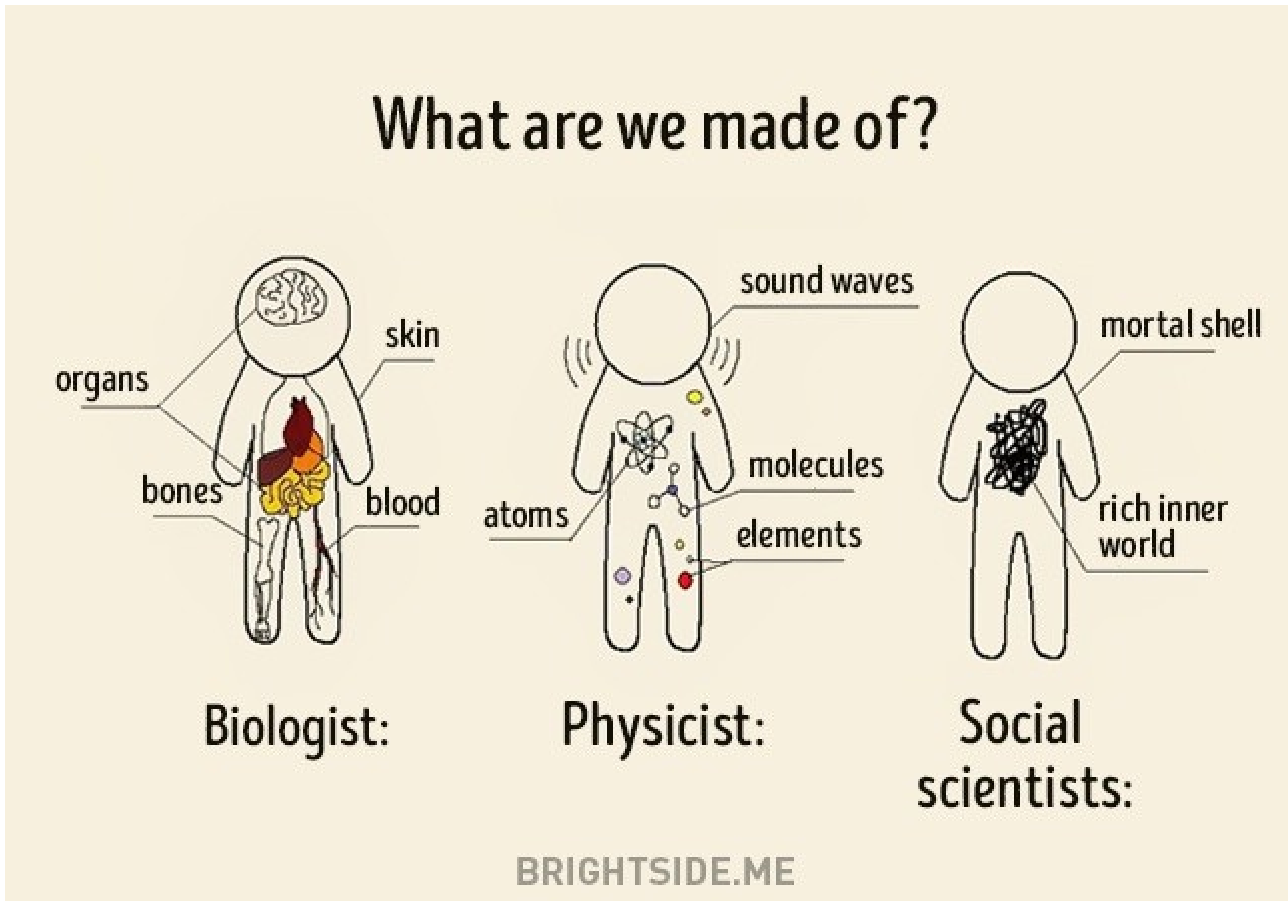


Very different detectors, also different selections, different reconstruction techniques, etc.: raw data are smeared very differently

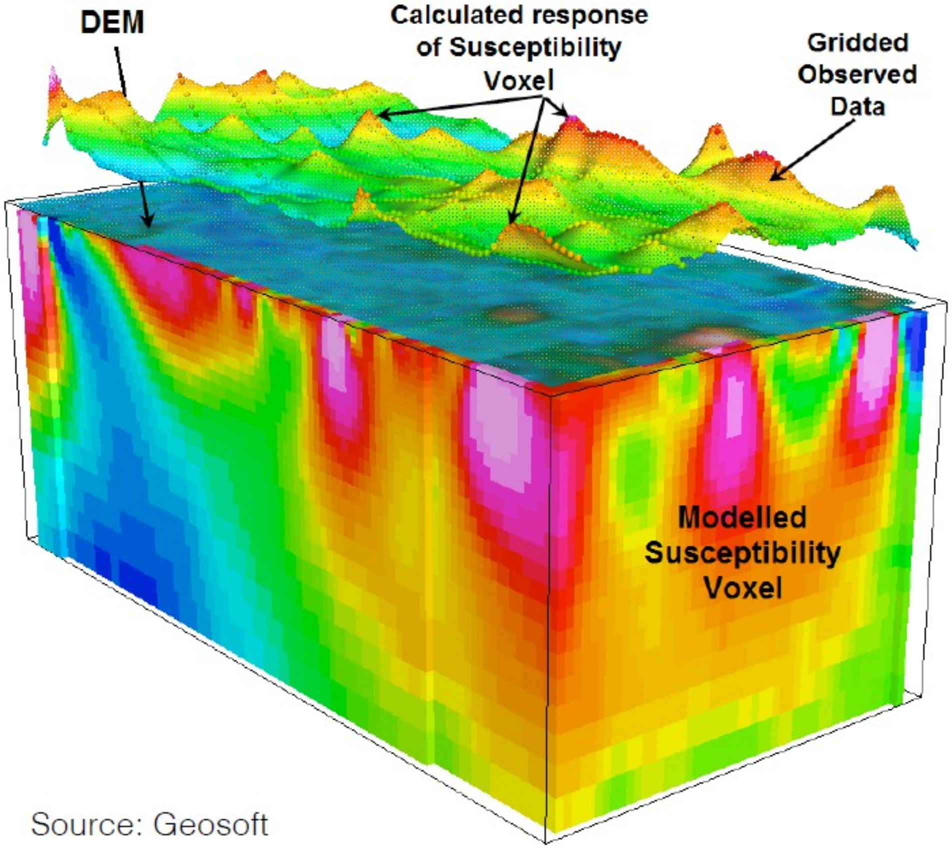
Convenient legacy

- Unfolding is an *easy* and *unexpensive* way of making your data useful to the relevant theory community, including the posterity
  - Theorists can come up with new promising models after the experiment stopped operating
  - PDF fitters need to combine distributions from several experiments, including old ones
- Note: growing trend to make *raw data* open (customary in astrophysics, novel for us)
  - But it is not trivial for external users to use raw data properly, and just making them accessible demands a lot of resources of the Collaboration

V: In other fields



Inversion in geophysics



Source: Geosoft

$$g(\boldsymbol{m}) = \boldsymbol{d}$$

Data, eg.:
- seismic wave travel times
- gravitational field of the Earth

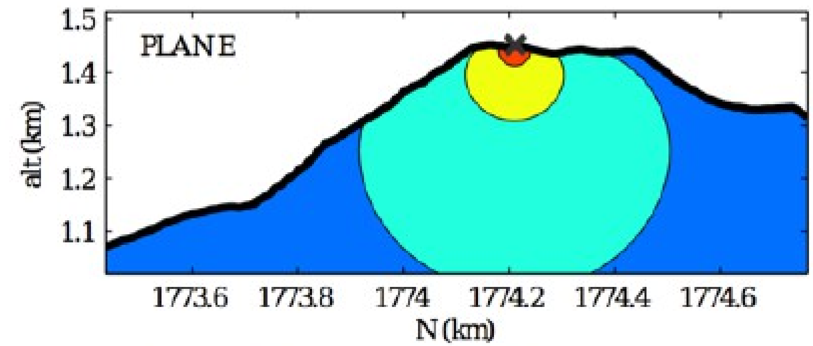
Forward modelling ↑ ↓ **Inversion**

Model, eg.:
- seismic velocity and density
- density

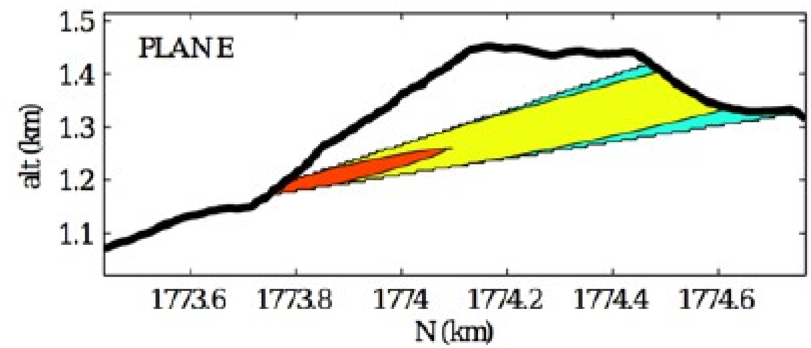
$$\mathbf{A} \quad \boldsymbol{\rho} \quad = \quad \boldsymbol{d}$$

$$\begin{bmatrix}\mathbf{G}\\ \mathbf{M}\end{bmatrix}\begin{bmatrix}\boldsymbol{\rho}\end{bmatrix} = \begin{bmatrix}\boldsymbol{g}\\ \boldsymbol{\varrho}\end{bmatrix}$$

Jourde *et al.* 2015



(1) gravimetry acquisition kernel, $\mathcal{G}$



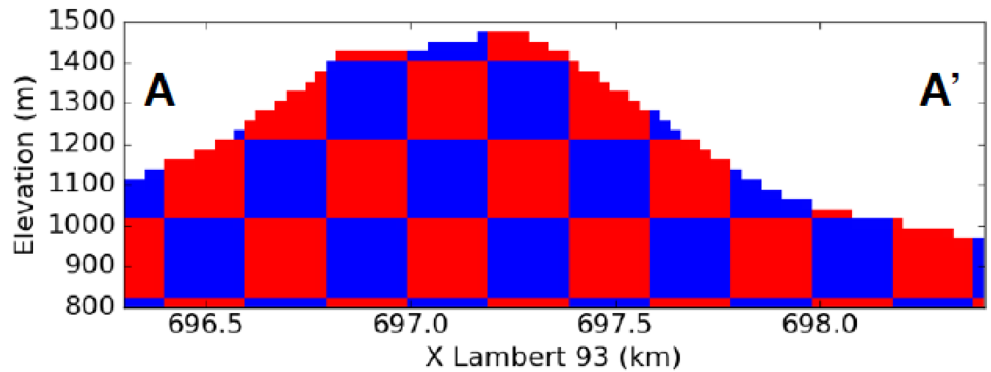
(2) tomography acquisition kernel, $\mathcal{M}$

Most geoprospecting methods are non-linear inversion problems: solutions wildly degenerate, need strong constraints to converge, different assumptions lead to qualitatively different results

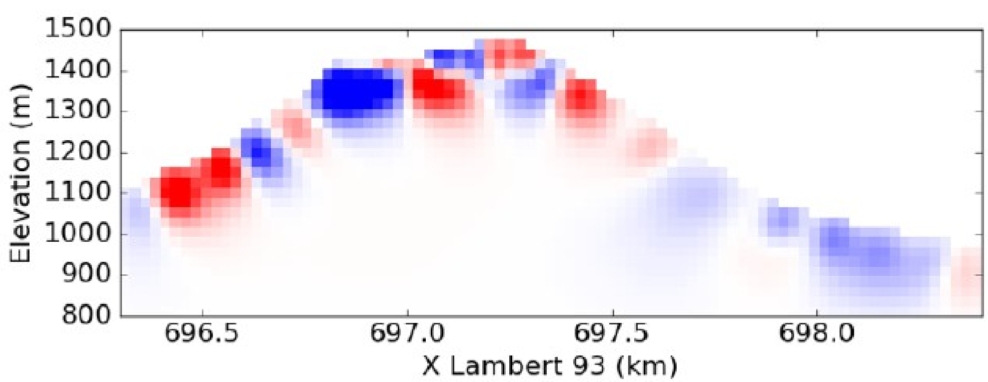
New method based on cosmic-ray detectors (muography): statistics-limited, but linear

Checkerboard test

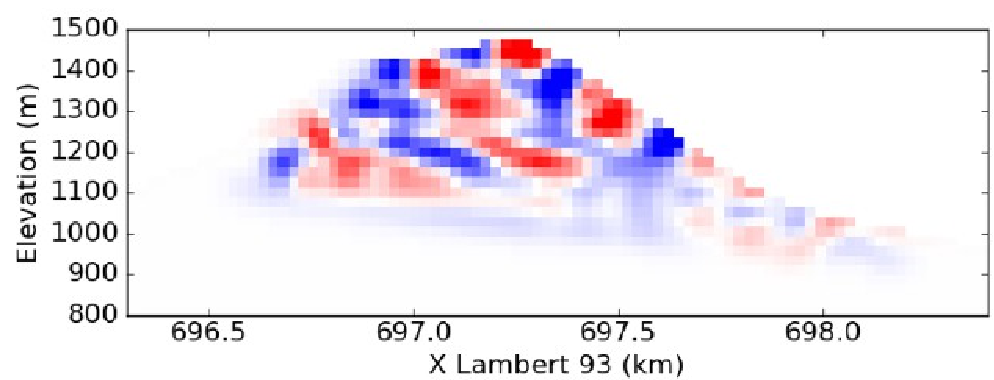
Simulated density pattern:



Red: high density
Blue: low density



Seen from gravimetric inversion

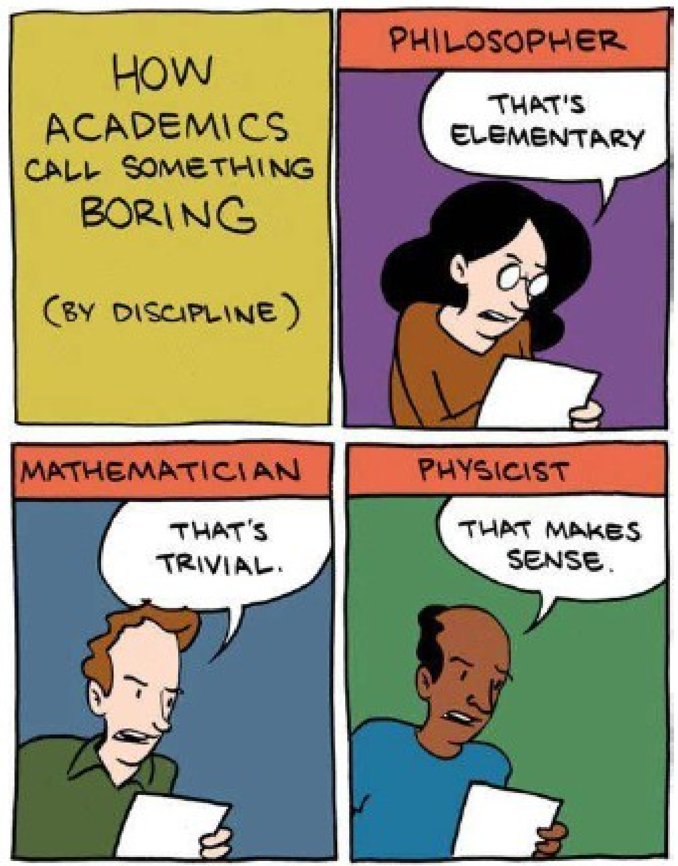


Seen from muographic inversion

Summary

- "Unfolding" is about how to invert a matrix that you should not invert

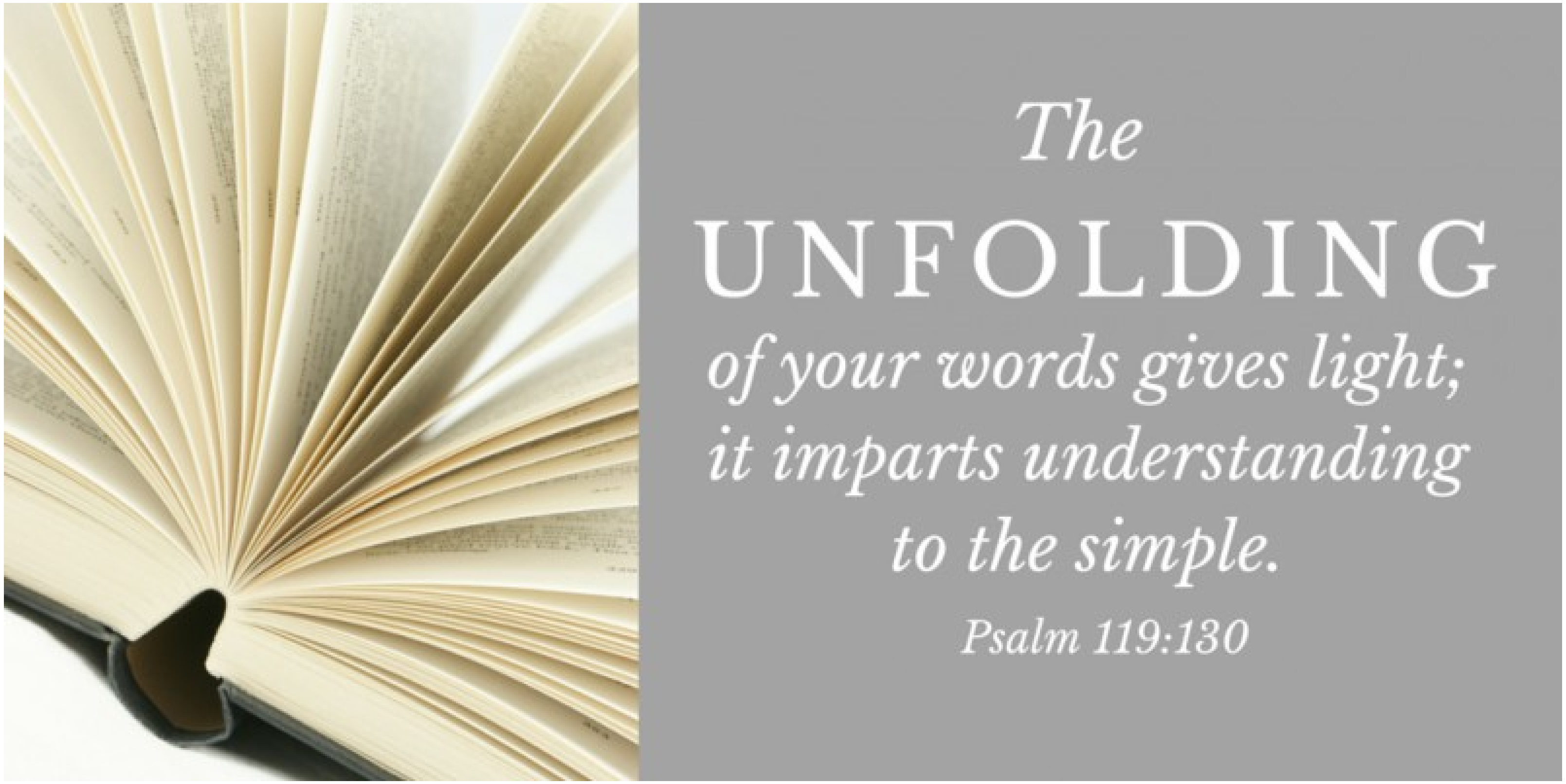
Thanks for your attention



From http://www.smbc-comics.com

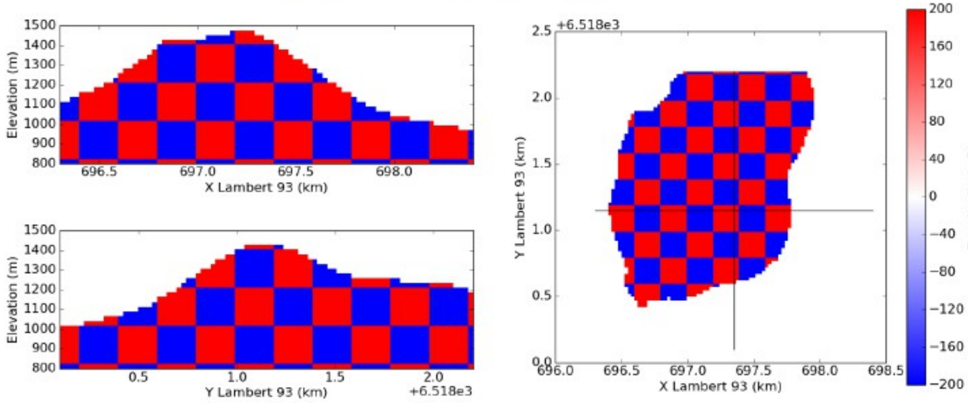
Some material stolen from:
Matthias Komm, Andreas Jung, Juan Alcaraz Maestre, Anne Barnoud, Andrea Marini

Extra slides

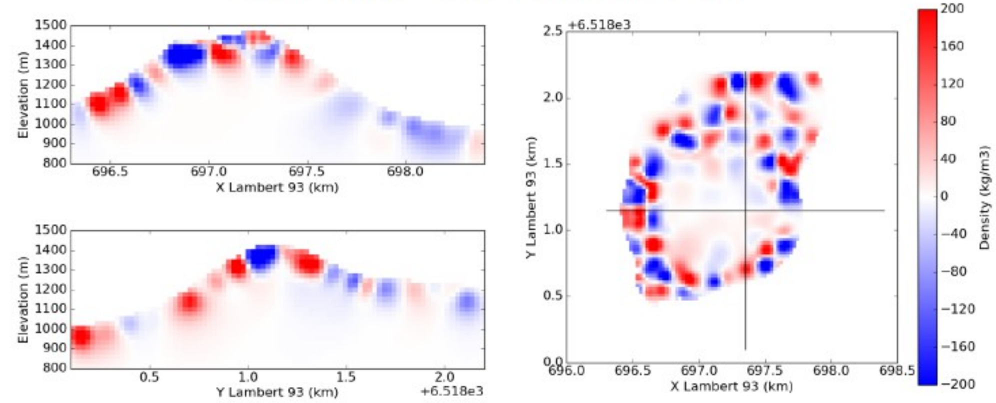


Comparison: Densities

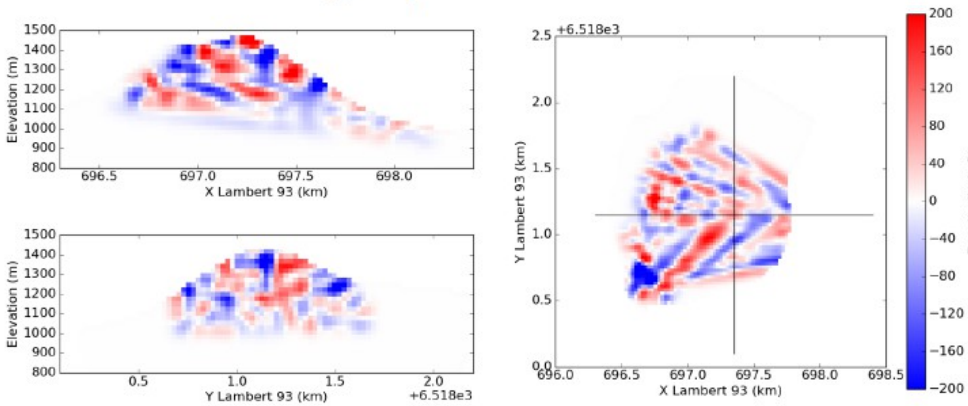
« True » densities



Gravimetric inversion



Muographic inversion



Joint inversion