Wiener-SVD Approach to Data Unfolding

Hanyu Wei

LSU

Department of
Physics & Astronomy

CEWG meeting Dec 2, 2021

Data Unfolding with Wiener-SVD Method

W. Tang (Brookhaven), X. Li (SUNY, Stony Brook), X. Qian (Brookhaven), H. Wei (Brookhaven), C. Zhang (Brookhaven)

May 9, 2017

24 pages

Published in: *JINST* 12 (2017) 10, P10002

Published: Oct 4, 2017

e-Print: 1705.03568 [physics.data-an]

DOI: 10.1088/1748-0221/12/10/P10002

Report number: BNL-114378-2017-JA

View in: ADS Abstract Service

https://inspirehep.net/literature/1598758

Rising citations in the past year in the fields of neutrino cross section measurement, reactor neutrino flux unfolding, and collider physics.

Data unfolding

- **Solve an inverse problem (from result to cause)**
  - Unfolding, deconvolution, unsmearing, etc.
  - Discretized in general given the measurement in binned distribution (e.g. histograms)
- **In HEP community, to estimate the true signal based on a measurement with**
  - Detector response
  - Statistical fluctuation
  - Systematic (backgrounds, mis-modeling of detector response, imperfect calibration, etc.)
- **Necessity**
  - Close to the true signal without specific detector response
  - Comparison with theoretical predictions, e.g. cross section
  - Comparison between different experiments
  - Sometimes certain features may not be recognizable (intuitive) in the folded distribution
- **Not** necessary in many problems, e.g. hypothesis testing or parameter fit

- In general, the number of dimensions/bins in a measurement is more than that in true spectrum to be unfolded. The data unfolding problem is essentially a fitting problem.

$$\chi^2(s) = (\boldsymbol{m} - \boldsymbol{r} \cdot s)^T Cov^{-1} (\boldsymbol{m} - \boldsymbol{r} \cdot s)$$

  - $\boldsymbol{m}$: measured spectrum, *m*-dimensional vector; **central value background subtracted** and its uncertainty propagated to total covariance
  - $\boldsymbol{s}$: unknown spectrum, to be unfolded, *n*-dimensional vector
  - $\boldsymbol{r}$: smearing (response) matrix, *m X n* and $\boldsymbol{m} \geq \boldsymbol{n}$
  - *Cov* : total covariance matrix containing all statistical and systematic uncertainties

- Cholesky decomposition: $Cov^{-1} = Q^T Q$, Q is a lower triangular matrix

$$\chi^2(s) = (M - R \cdot s)^T \cdot (M - R \cdot s)$$

Pre-scaling

- $M \coloneqq Q \cdot \boldsymbol{m}$
- $R \coloneqq Q \cdot \boldsymbol{r}$

Solution (direct inversion)

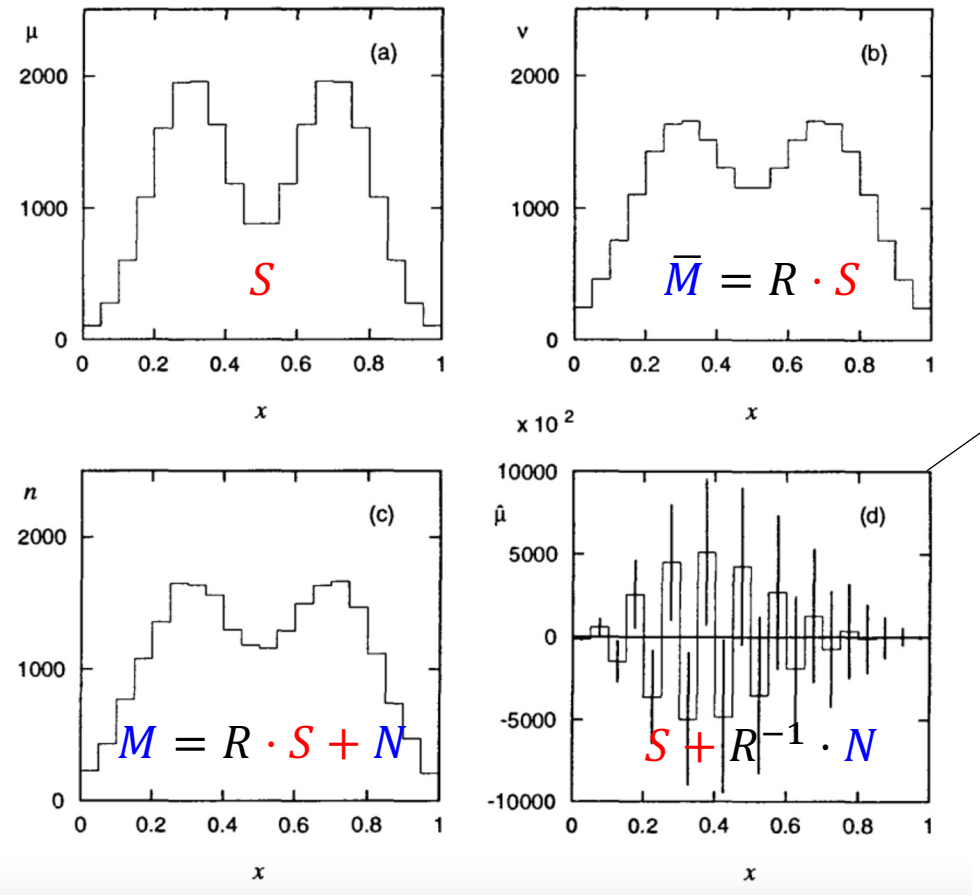
$$\hat{s} = (R^T R)^{-1} R^T M$$

$$M = R \cdot s_{true} + N(noise)$$

$$Noise \coloneqq total\ stat. + total\ sys.$$

The response matrix R is unnecessary to be a square matrix

Data unfolding



This is one unbiased solution (direct inversion, no regularization) to an unfolding problem. “True” information is buried in the catastrophic oscillations in the unfolded spectrum.

Practical solution: trade-off between **bias** and **variance** to suppress the “oscillation” --> **regularization [defines various unfolding methods]**

Regularization

- **Regularization is not needed** if
  - Bin-to-bin smearing is small
  - = Response matrix is well-conditioned
  - = Information is sufficient to solve the inverse problem without ambiguity
  - In an analysis: enlarge the bin size to achieve this if the granularity of the information to be unfolded is not that important

Note: too small binning is also meaningless considering the width of bin-to-bin smearing in the response (smearing) matrix [Nyquist theorem]

Simple traditional regularization

$$\chi^2 = (M - R \cdot s)^T \cdot (M - R \cdot s) + \tau \sum_i (C_k \cdot s)_i^{\,2}$$

Likelihood function

Regularization strength

← Regularization function

The unfolded result

$$\hat{s} = \left(R^T R + \tau C_k^T C_k\right)^{-1} \cdot R^T \cdot M$$

C_k is *k*-th order derivative matrix, e.g. k=0, identity matrix

Need to plug in C_k in the formula where $R = R \cdot C_k^{-1}$

$$C_0 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, C_1 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} -1+\epsilon & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2+\epsilon & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -2+\epsilon & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & -1+\epsilon \end{bmatrix},$$

Generalized regularization formalism

Generalization of regularization formulism is discussed in Wiener-SVD unfolding paper.

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M$$

$A_C = \left(R^T R + \tau C_k^T C_k\right)^{-1} \cdot (R^T R)$ Simple regularization for illustration

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot (R \cdot s_{true} + N)$$
$$= A_C \cdot s_{true} + A_C \cdot (R^T R)^{-1} \cdot R^T \cdot N$$

= *smeared truth (small bias)* + *suppressed noise (decent variance)*

- ✓ Additional smearing matrix A_C is fixed once a regularization method is chosen (the choice of derivative matrix C_k is arbitrary, C_2 is more common).
- ✓ A_C contains all information we need to calculate the bias introduced in the regularization.
  - ✓ A different model will have a different bias $(A_C - I) \cdot s_{model}$
  - ✓ $A_C \cdot s_{model}$ is to be compared to the unfolded result $\hat{s}$ $(= A_C \cdot s_{true} + suppressed\ noise)$
- ✓ Uncertainty of $\hat{s}$ can be directly calculated through standard propagation following the formula. A_C is part of the linear transformation.

SVD regularization

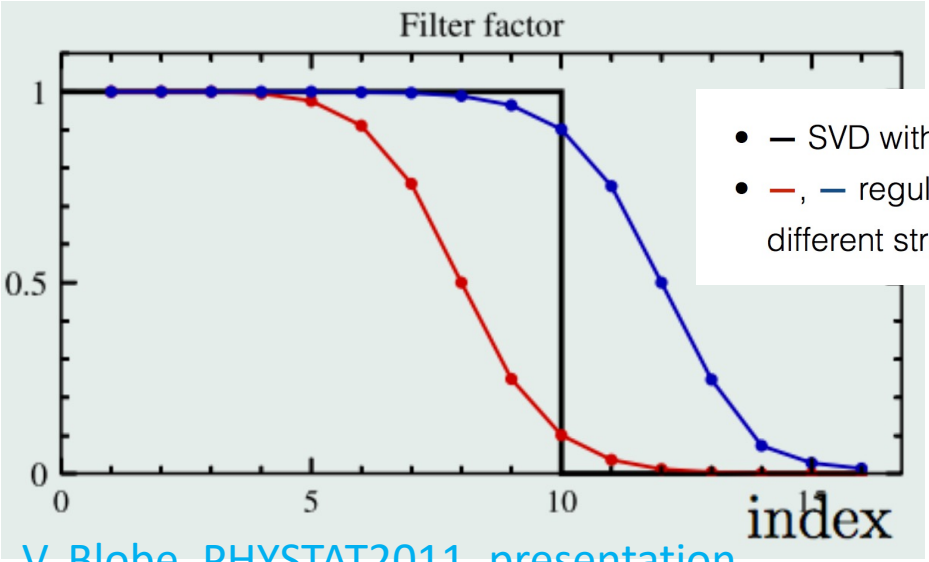
- Singular Value Decomposition (SVD) of response matrix

$$R = UDV^T$$

Geometrical Transformations

- Rotation → U: $m \times m$ orthogonal matrix
- Scaling → D: $m \times n$ diagonal matrix with non-negative real numbers on the diagonal; elements d_{ii} is in descending order
- Rotation → V: $n \times n$ orthogonal matrix

Columns of U, V are orthonormal bases



V. Blobe, PHYSTAT2011, presentation

Kernel

$$F_{ij} = \frac{d_i^2}{d_i^2 + \tau}\delta_{ij}$$

Regularization strength

SVD is a way to transfer the measurement domain to an "effective frequency" domain, therefore enabling the suppression of high "frequent" noise which in general corresponds to small d_{ii} (singular values).

SVD regularization

- Singular Value Decomposition (SVD) of response matrix

$$R = UDV^T$$

Geometrical Transformations

- Rotation → U: $m \times m$ orthogonal matrix
- Scaling → D: $m \times n$ diagonal matrix with non-negative real numbers on the diagonal; elements d_{ii} is in descending order
- Rotation → V: $n \times n$ orthogonal matrix

Columns of U, V are orthonormal bases

$$\hat{s} = V \cdot F \cdot D^{-1} \cdot U^T \cdot M$$

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M$$
$$A_C = V \cdot F \cdot V^T$$

Kernel

$$F_{ij} = \frac{d_i^2}{d_i^2 + \tau} \delta_{ij}$$

Regularization strength

SVD is a way to transfer the measurement domain to an "effective frequency" domain, therefore enabling the suppression of high "frequent" noise which in general corresponds to small d_{ii} (singular values).

Wiener-SVD Unfolding

A choice of additional smearing matrix A_C

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M \qquad A_C = C_k^{-1} \cdot V \cdot W \cdot V^T \cdot C_k$$

(SVD decomposition) $RC^{-1} = U \cdot D \cdot V^T$

(Wiener-SVD kernel) $$W_{ij} = \frac{d_i^2 \cdot \left(V^T \cdot C_k \cdot \bar{s}\right)_i^2}{d_i^2 \cdot \left(V^T \cdot C_k \cdot \bar{s}\right)_i^2 + 1} \delta_{ij}$$

- A_C is fixed once the expectation of signal $\bar{s}$ is chosen. **No strength parameter to tune with.**
- Unfolding is model dependent (any method is not?)
  - choose the "best" prediction of s_{true} or a weighted average from various models
  - Ensure the prediction from the best knowledge of s and the real measurement are well consistent within full uncertainties
  - (optional) Additional mock data study with varied models to further validate the choice of s
- What we care most:
  - Measurement uncertainties (model independent or largely independent) properly propagate to the unfolded results --> straightforward based on the unfolded result formula (math)
  - What is the model-dependent bias in this unfolding? Can be calculated using A_C.

Wiener-SVD Unfolding

- To automatically minimize the Mean Square Error (MSE) given a model S

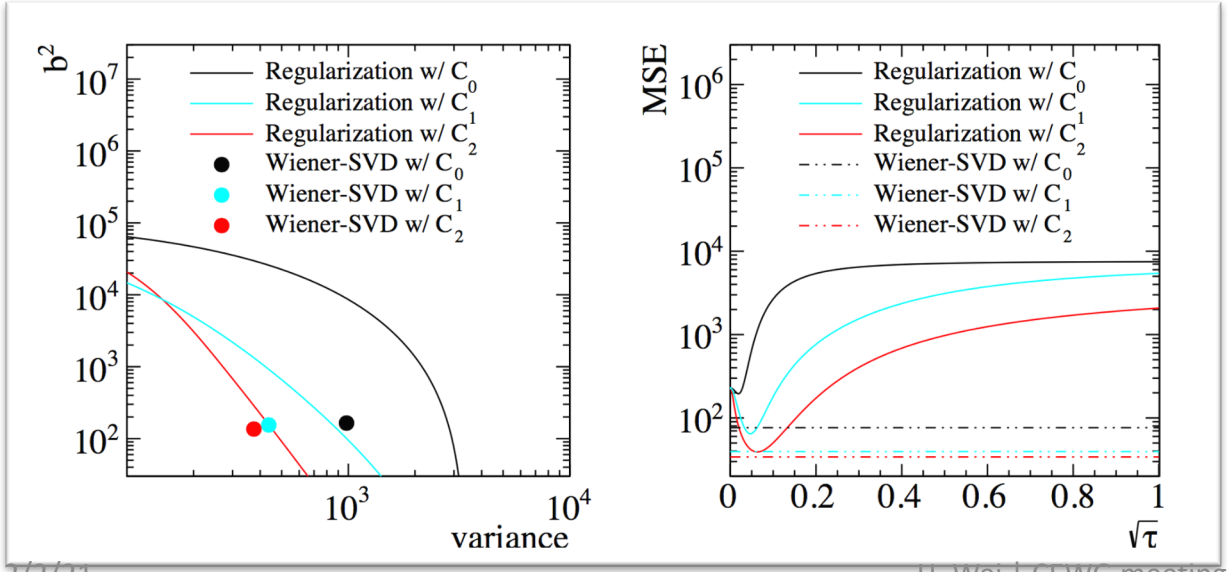
$$MSE = E\left[\left(\hat{S} - S\right)^2\right] = E\left[\left(F \cdot \frac{M}{R} - S\right)^2\right] = E\left[\left(F \cdot S + F \cdot \frac{N}{R} - S\right)^2\right]$$

$$= E\left[\left((\boldsymbol{F} - \boldsymbol{I}) \cdot \boldsymbol{S}\right)^2 + \left(\boldsymbol{F} \cdot \frac{\boldsymbol{N}}{\boldsymbol{R}}\right)^2\right]$$

Bias

Variance

F = additional smearing matrix = regularization



- Traditional regularization needs to "tune" a regularization strength parameter [curve in the left plot]
- Wiener-SVD regularization corresponds to a fixed point in the phase space of bias versus variance with **minimum MSE**

Uncertainty treatment

Pre-scaling

- $M := Q \cdot \boldsymbol{m}$
- $R := Q \cdot \boldsymbol{r}$

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M$$

Standard error propagation from a linear transformation

Best fit of regularized chi2

$$\hat{s} = K_{tot} \cdot m$$

$$Cov_s = K_{tot} \cdot \boldsymbol{Cov_m} \cdot K_{tot}^T$$

Cov_m is the covariance of predicited m considering deviation from the central value.

Predicted $m = r \cdot s + bkg$

- $\boldsymbol{s}$ is related to the "true parameter" to be unfolded and it has **NO** uncertainty (fixed at central value)
  - **Subtext: assume/ensure s_{CV} is consistent with s_{true} or at least they agree within uncertainties**
- r: response matrix derived from central value prediction and its variation is computed and reflected in the variation of m
- $\boldsymbol{bkg}$: central value prediction of background spectrum and its uncertainty is computed and reflected in the variation of m
- Statistical uncertainty of m

A case of extracting cross section

Recent MicroBooNE cross section measurement https://arxiv.org/abs/2110.14023 (A future talk from Wenqiang)

$$M(E_{rec}) = POT \cdot T \cdot \int F(E_\nu) \cdot \sigma(E_\nu) \cdot D(E_\nu, E_{rec}) \cdot \epsilon(E_\nu, E_{rec}) \cdot dE_\nu + B(E_{rec}),$$

Integrated beam luminosity, target mass

Flux

Cross section

Detector response (smearing and efficiency)

Background

$$M_i = \sum_j \widetilde{S}_{ij} + B_i,$$

$$\begin{aligned}\widetilde{S}_{ij} &= \frac{POT \cdot T \cdot \int_j F(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot D(E_{\nu\ j}, E_{rec\ i}) \cdot \epsilon(E_{\nu\ j}, E_{rec\ i}) \cdot dE_{\nu\ j}}{POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}} \\ &\cdot \left(POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j} \right) \cdot \frac{\int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}}{\int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j}} \\ &= \widetilde{\Delta}_{ij} \cdot \widetilde{F}_j \cdot S_j,\end{aligned}$$

Cross section term cannot be cleanly extracted given the finite size of binning and energy-dependent flux

A case of extracting cross section

$$\widetilde{\Delta}_{ij} = \frac{\text{(Selected no. of events in reco. energy bin i from true energy bin j after event weights)}}{\text{(Generated no. of events in true energy bin j after event weights)}}$$

$$M_i = \sum_j \widetilde{S}_{ij} + B_i,$$

Response matrix

$$\widetilde{S}_{ij} = \frac{POT \cdot T \cdot \int_j F(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot D(E_{\nu\ j}, E_{rec\ i}) \cdot \epsilon(E_{\nu\ j}, E_{rec\ i}) \cdot dE_{\nu\ j}}{POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}}$$

$$\cdot \left(POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j} \right) \cdot \frac{\int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}}{\int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j}}$$

$$= \widetilde{\Delta}_{ij} \cdot \widetilde{F}_j \cdot S_j,$$

CV flux weighted cross section **TO UNFOLD**

Constant including luminosity, target mass, and central value (CV) flux

Wiener-SVD Unfolding

Uncertainty treatment

$$\hat{s} = K_{tot} \cdot m$$

$$Cov_s = K_{tot} \cdot \boldsymbol{Cov_m} \cdot K_{tot}^T$$

$$M_i = \sum_j \widetilde{S}_{ij} + B_i,$$

Response matrix

Flux (only numerator), cross section, and detector systematics

$$\widetilde{S}_{ij} = \frac{POT \cdot T \cdot \int_j F(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot D(E_{\nu\ j}, E_{rec\ i}) \cdot \epsilon(E_{\nu\ j}, E_{rec\ i}) \cdot dE_{\nu\ j}}{POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}}$$

$$\cdot \left(POT \cdot T \cdot \int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j} \right) \cdot \frac{\int_j \overline{F}(E_{\nu\ j}) \cdot \sigma(E_{\nu\ j}) \cdot dE_{\nu\ j}}{\int_j \overline{F}(E_{\nu\ j}) \cdot dE_{\nu\ j}}$$

$$= \widetilde{\Delta}_{ij} \cdot \widetilde{F}_j \cdot S_j,$$

CV flux weighted cross section

(TO UNFOLD) No flux nor cross section uncertainty to be considered

Constant including luminosity, target mass, and central value (CV) flux

Uncertainty of beam intensity and target mass

What to report

- Unfolded spectrum $\hat{s}$
- Covariance matrix of $\hat{s}$

- Additional smearing matrix A_C
- Central value (nominal) flux $\bar{F}$

} The things are to be applied to any other models for comparisons

Summary

- Wiener-SVD unfolding
  - Generalized formulism, additional smearing matrix Ac
  - Minimal MSE
  - No regularization parameter
  - Uncertainty standard propagration
- Cross section uncertainty
  - Flux uncertainty treatment (nominal flux weighted cross section)

Backup slides

Effective frequency domain

$$\hat{s} = V \cdot F \cdot D^{-1} \cdot U^T \cdot M$$

- ✓ "Time" → "Energy", response matrix is **not energy-invariant** & unclear physics of Fourier transform on the response matrix
- ✓ Fourier transform doesn't work

$$(V^T \hat{s})_i = F_{ii} \cdot \frac{(U^T M)_i}{D_{ii}}$$

Three numbers + filter given the i-th row

- ✓ "i" corresponds to "frequency"
- ✓ Refer to $M_U = U^T \cdot M$ as the measurement in 'effective frequency domain' and $R_U = U^T \cdot R$, $N_U = U^T \cdot N$ (still independent normal distribution),
- ✓ U^T is a rotation (transformation). A 'scaling' D connects M_U and S_V, so they can be treated as in the same domain.

Effective frequency domain

$$M_U = U^T \cdot M$$

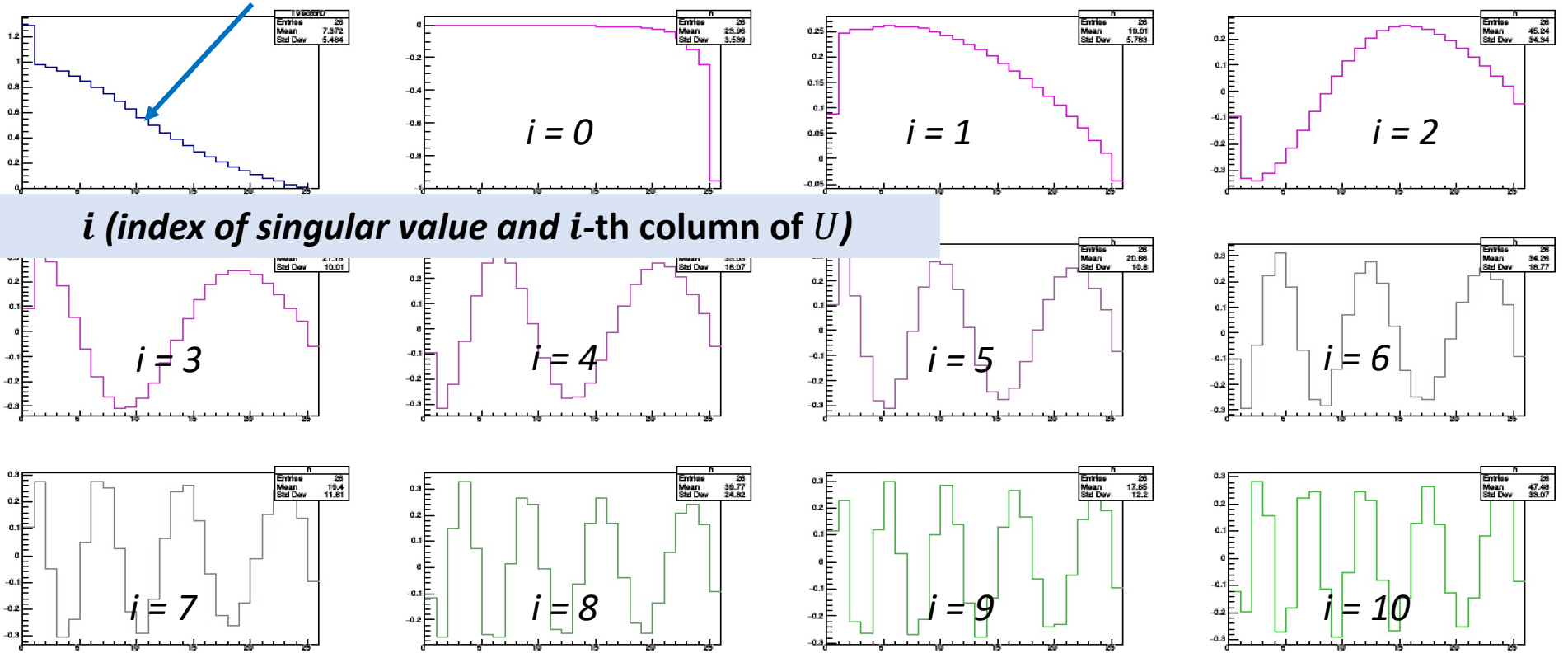
- Consider M measured in the "energy" domain
  - Bases: (1, 0, 0, …), (0, 1, 0, …), etc
- U^T transform (rotate) M into the "effective frequency" domain
- **The bases in "effective frequency" domain are the columns(rows) of orthogonal matrix $U(U^T)$ if represented by "energy"** (*hint: representation theory in matrix mechanics*)
- The higher "frequency" corresponds to smaller singular value (wild oscillation in the direct inverse solution).

Effective frequency domain

- SVD decomposition ($R = UDV^T$) of a realistic response matrix (Daya Bay experiment, neutrino spectrum to measured spectrum)

Singular values (d_i) in descending order

The rest: first 11 columns of matrix U



***i* (index of singular value and *i*-th column of U)**

X-axis: energy index, corresponds to M, measured in energy domain