

# Decomposition of a covariance matrix into uncorrelated and correlated errors

Benno List

University of Hamburg

Alliance Workshop on Unfolding and Data Correction 27.5.2010

Introduction

Decomposition of the Covariance Matrix

Extended  $\chi^2$  Fit

Summary and Conclusions

## Motivation

Traditionally in HEP we have

- ▶ Statistical errors: uncorrelated between data points
- ▶ Systematic errors: correlated between data points

A popular way to incorporate correlated errors into a  $\chi^2$  fit:

$$\chi^2 = \sum_i \frac{(z_i + \sum_k \gamma_k^i b_k - f_i)^2}{\delta z_{i,\text{uncorr}}^2} + \sum_k b_k^2 \quad (1)$$

$$= ((\vec{z} + G\vec{b}) - \vec{f})^T \Sigma^{-1} ((\vec{z} + G\vec{b}) - \vec{f}) + \vec{b}^T \vec{b}. \quad (2)$$

Unfolding results in a fully occupied covariance matrix for the statistical errors.

- ▶ Difficult for visualization (no uncorrelated part)
- ▶ Current fit programs typically use form (2)

**Can an arbitrary covariance matrix be decomposed into an uncorrelated and a correlated part?**

## Introduction

The  $\chi^2$  in a fit is given by:

$$\chi^2 = (\vec{z} - \vec{f})^T \cdot C^{-1} \cdot (\vec{z} - \vec{f}) \quad (3)$$

$$= \sum_{i,j} (z_i - f_i) (C^{-1})_{i,j} (z_j - f_j). \quad (4)$$

In case of correlated errors, often this form is used:

$$\chi^2 = \sum_i \frac{(z_i + \sum_k \gamma_k^i b_k - f_i)^2}{\delta z_{i,\text{uncorr}}^2} + \sum_k b_k^2 \quad (5)$$

$$= ((\vec{z} + G\vec{b}) - \vec{f})^T \Sigma^{-1} ((\vec{z} + G\vec{b}) - \vec{f}) + \vec{b}^T \vec{b}. \quad (6)$$

- ▶ How are  $C$ ,  $\Sigma$  and  $G$  related?
- ▶ Can we transform any covariance matrix  $C$  into uncorrelated errors  $\Sigma$  and a coefficient matrix  $G$ ?
- ▶ Are the two  $\chi^2$  definitions equivalent?

## Notation

$$z_i = \hat{y}_i + \sum_k \gamma_k^i \hat{a}_k$$

- ▶  $z_i$ : The measured values, with  $\langle z_i \rangle = \langle y_i \rangle$  and uncorrelated (“statistical”) error  $\delta z_{i,\text{stat}}^2 = \delta y_i^2$ .
- ▶  $\hat{y}_i$ : Uncorrelated random variables with variance  $\delta y_i^2$  (“the measurement without systematics”)
- ▶  $\hat{a}_k$ : Uncorrelated random variables **with zero mean and unit variance**:  $\langle \hat{a}_k \rangle = 0$ ,  $\langle \hat{a}_k^2 - \langle \hat{a}_k \rangle^2 \rangle = 1$  (“the unknown systematics”)
- ▶  $\gamma_k^i$ : shift of measurement  $i$  due to systematic  $k$  (assumed to be known, e.g. from MC)

Note:  $\hat{a}_k$  represents the deviation from an **unknown, true value**, e.g. deviation from the true luminosity, the true energy scale etc.

The **observed** value of  $\hat{a}_k$  (e.g., the measured luminosity, the energy scale after some calibration) is generally nonzero and shifts all measurements.

## Notation, cont'd

For convenience, we define  $a_k$  such that the measured (or calibrated) value is zero:

$$a_k = \hat{a}_k + \Delta a_k \quad (7)$$

$$y_i = \hat{y}_i - \sum_k \gamma_k^i \Delta a_k \quad (8)$$

$$(9)$$

Then the measured values  $z_i$  are given by

$$z_i = y_i + \sum_k \gamma_k^i a_k \quad (10)$$

Define  $\Sigma = \text{diag}(\delta y_i^2)$  and  $G_{ik} = \gamma_k^i$ , then error propagation gives

$$C = \Sigma + GG^T$$

for the covariance matrix of the measurements  $z_i$ .

## Decomposition of the Covariance Matrix

If only  $C$  is given, e.g. as result from an unfolding, can this calculation be reversed, i.e.:

Is it possible to decompose any given covariance matrix  $C$  into an uncorrelated part  $\Sigma$  and a correlated part  $G$  such that

$$C = \Sigma + GG^T?$$

Any covariance matrix  $C$  is

- ▶ symmetric:  $C = C^T$
- ▶ positive definite:  $\vec{x}^T C \vec{x} > 0$  for all  $\vec{x}$  with  $|\vec{x}| \neq 0$ .

Therefore:

- ▶ All eigenvalues  $\lambda_i$  of  $C$  are real and positive.
- ▶ An orthogonal matrix  $O$  of eigenvectors exists such that

$$C = O \Lambda O^T$$

with  $\Lambda = \text{diag}(\lambda_i)$  the diagonal matrix of the eigenvalues.

Setting  $F = O \text{diag}(\sqrt{\lambda_i})$ :

$$C = FF^T$$

This is not a solution to the problem, because there are **no** uncorrelated errors.  
But: **any** positive **semidefinite** matrix  $C$  can be written as  $FF^T$ .

## The Problem

The problem: Find a diagonal matrix  $\Sigma_{\text{unc}} = \text{diag}(\sigma_{\text{unc},1}^2, \dots, \sigma_{\text{unc},N}^2)$ , representing the uncorrelated errors, such that  $C - \Sigma_{\text{unc}}$  is positive semidefinite, i.e. all eigenvalues are  $\geq 0$ .

The uncorrelated errors should be **“as large as possible”**.



## The Solution

Ansatz:

$$\Sigma_{\text{unc}} = a\Sigma_{\text{tot}},$$

where  $\Sigma_{\text{tot}} = \text{diag}(C_{ii})$  is the diagonal part of  $C$ .

Define  $S = \text{diag}(\sqrt{C_{ii}})$ , i.e.  $\Sigma_{\text{tot}} = S^2$ , and the matrix of correlation coefficients

$$R = S^{-1}CS^{-1},$$

with eigenvalues  $\lambda_{R,i}$ .

Now

$$\tilde{C} = C - \Sigma_{\text{unc}} = S(R - aI)S$$

$\tilde{C}$  is positive semidefinite if  $\tilde{R} = R - aI$  is positive semidefinite.

The eigenvalues of  $\tilde{R}$  are given by  $\lambda_{R,i} - a$ .

Therefore: Choose

$$a = \min(\lambda_{R,i})$$

then at least one eigenvalue of  $\tilde{C}$  is zero, and all others are nonnegative.

## Properties of $G$

If  $O$  is the matrix of eigenvectors of  $R$ , then

$$G = SO \operatorname{diag}(\sqrt{\lambda_{R,i} - a})$$

and

$$C = a\Sigma_{\text{tot}} + GG^T$$

Elements of  $G$ :

$$\gamma_i^k = G_{ik} = \sigma_{\text{tot},i} O_{ik} \sqrt{\lambda_{R,k} - a}$$

Since  $O$  is orthogonal, for any column  $k$ :  $\sum_i O_{ik}^2 = 1$ , i.e. the  $O_{ik}$  are of order 1, and  $|O_{ik}| \leq 1$ .

If  $\lambda_{R,k} - a \ll 1$  it follows:  $\gamma_i^k \ll \sigma_{\text{tot},i}$

Therefore: Columns of  $G$  where  $\lambda_{R,k} - a = 0$  and  $\lambda_{R,k} - a \ll 1$  can be omitted, because they do not contribute noticeably to the overall error.

## Short Summary

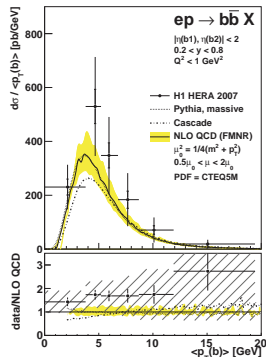
For any covariance matrix  $C$ , one can define an uncorrelated part  $\Sigma = a\Sigma_{\text{tot}}$  and a coefficient matrix  $G$  such that

$$C = \Sigma + GG^T$$

The uncorrelated errors are a fixed fraction  $\sqrt{a}$  of the total errors for all data points.

Caveat: For  $N$  data points, the full covariance matrix has  $N \cdot (N + 1)/2$  elements, while  $\Sigma$  and  $G$  have up to  $N + N \cdot (N - 1) = N^2$  elements.

The number of independent coefficients decreases only if at least  $N/2$  eigenvalues of  $\tilde{R}$  are so small that the corresponding columns of  $G$  can be omitted.



M.

Sauter, PhD thesis ETH Zurich 2009,  
DESY-THESIS-2009-047.

## Extended $\chi^2$ Fit

In an extended  $\chi^2$  fit, shifts and scale values are determined for all sources of systematic errors.

Sometimes the errors are significantly reduced (“cross calibration”), sometime the result is quite unexpexted (e.g. large shifts of the systematics).

Would that also happen if the “pure” covariance matrix were used?

**Are a  $\chi^2$  fit with a full covariance matrix and an extended  $\chi^2$  fit with parameters for each systematic error equivalent?**

A formal proof is presented for the equivalence.

Define:

Vector of measured values:

$$\vec{z}' = (z_1 \dots z_N, a_1 \dots a_M)^T = (z_1 \dots z_N, 0 \dots 0)^T$$

Vector of independent random variables

$$\vec{y}' = (y_1 \dots y_N, a_1 \dots a_M)^T$$

$\vec{z}'$  and  $\vec{y}'$  are related by a matrix

$$M = \begin{pmatrix} I & G \\ 0 & I \end{pmatrix} \quad \text{and} \quad M^{-1} = \begin{pmatrix} I & -G \\ 0 & I \end{pmatrix}.$$

such that

$$\vec{z}' = M \cdot \vec{y}'.$$

The covariance matrix of the elements of  $\vec{y}'$  is given by

$$\Sigma' = \text{diag}(\delta z_{1,\text{uncorr}}^2 \dots \delta z_{N,\text{uncorr}}^2, 1 \dots 1).$$

From error propagation, the covariance matrix  $C'$  of the extended vector  $z'$  is given by

$$C' = M \Sigma' M^T \quad (11)$$

$$= \begin{pmatrix} \Sigma + G^T G & G \\ G^T & I \end{pmatrix}. \quad (12)$$

and we note that

$$C'^{-1} = (M^T)^{-1} \Sigma'^{-1} M^{-1} = \begin{pmatrix} \Sigma^{-1} & -\Sigma^{-1} G \\ -G^T \Sigma^{-1} & 1 + G^T \Sigma^{-1} G \end{pmatrix}.$$

Introduce  $K$  additional fit parameters  $b_k$  representing the best fit values of  $a_k$ , i.e. our best estimate of the true mean of the  $a_k$ .

Form a vector

$$\vec{f}' = (f_1 \dots f_N, b_1 \dots b_M)^T,$$

Use

$$C'^{-1} = (M^T)^{-1} \Sigma'^{-1} M^{-1} = \begin{pmatrix} \Sigma^{-1} & -\Sigma^{-1}G \\ -G^T \Sigma^{-1} & 1 + G^T \Sigma^{-1}G \end{pmatrix}.$$

Extended  $\chi^2$ :

$$\tilde{\chi}^2 = (\vec{z}' - \vec{f}')^T C'^{-1} (\vec{z}' - \vec{f}') \quad (13)$$

$$= (\vec{z} - \vec{f})^T \Sigma^{-1} (\vec{z} - \vec{f}) + (\vec{z} - \vec{f})^T \Sigma^{-1} G \vec{b} \quad (14)$$

$$+ \vec{b}^T G^T \Sigma^{-1} (\vec{z} - \vec{f}) + \vec{b}^T \vec{b} + \vec{b}^T G^T \Sigma^{-1} G \vec{b} \\ = (\vec{z} + G \vec{b} - \vec{f})^T \Sigma^{-1} (\vec{z} + G \vec{b} - \vec{f}) + \vec{b}^T \vec{b}. \quad (15)$$

By definition  $\vec{b} - \vec{a} = \vec{b}$ , because the measured values  $a_k$  all vanish.

Best value for  $\vec{b}$ , i.e. minimum of  $\tilde{\chi}^2$ :

$$0 = \frac{\partial \tilde{\chi}^2}{\partial \vec{b}} \quad (16)$$

$$= 2G^T \Sigma^{-1} (\vec{z} + G\vec{b} - \vec{f}) + 2b \quad (17)$$

Solution:

$$\vec{b} = (1 + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{f} - \vec{z}) \quad (18)$$



## Rewrite

$$\tilde{\chi}^2 = (\vec{z} + G\vec{b} - \vec{f})^T \Sigma^{-1} (\vec{z} + G\vec{b} - \vec{f}) + \vec{b}^T \vec{b} \quad (19)$$

$$\begin{aligned} &= (\vec{z} - \vec{f})^T \Sigma^{-1} (\vec{z} - \vec{f}) \\ &\quad - \vec{b}^T G^T \Sigma^{-1} (\vec{z} - \vec{f}) - (\vec{z} - \vec{f})^T \Sigma^{-1} G \vec{b} + \vec{b}^T G^T \Sigma^{-1} G \vec{b} + \vec{b}^T \vec{b}. \end{aligned} \quad (20)$$

Insert  $\vec{b} = -(1 + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f})$ :

$$\tilde{\chi}^2 = (\vec{z} - \vec{f})^T \Sigma^{-1} (\vec{z} - \vec{f}) \quad (21)$$

$$\begin{aligned} &- (\vec{z} - \vec{f})^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f}) \\ &- (\vec{z} - \vec{f})^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f}). \\ &+ (\vec{z} - \vec{f})^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f}) \\ &+ (\vec{z} - \vec{f})^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f}). \end{aligned}$$

after a bit of calculation...

$$\begin{aligned} &= (\vec{z} - \vec{f})^T \Sigma^{-1} (\vec{z} - \vec{f}) \\ &\quad - (\vec{z} - \vec{f})^T \Sigma^{-1} G (I + G^T \Sigma^{-1} G)^{-1} G^T \Sigma^{-1} (\vec{z} - \vec{f}) \end{aligned} \quad (22)$$

$$C = \Sigma + GG^T$$

is an rank- $M$  update of  $\Sigma$ .

Apply the the Sherman–Morrison–Woodbury formula  
 $(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I - V^T A^{-1}U)^{-1}V^T A^{-1}$   
 to calculate inverse of  $C$ :

$$C^{-1} = \Sigma^{-1} - \Sigma^{-1}G \left( I + G^T \Sigma^{-1}G \right)^{-1} G^T \Sigma^{-1}.$$

Insert into  $\chi^2$  expression:

$$\chi^2 = (\vec{z} - \vec{f})^T \cdot C^{-1} \cdot (\vec{z} - \vec{f}) \quad (23)$$

$$= (\vec{z} - \vec{f})^T \cdot \Sigma^{-1} \cdot (\vec{z} - \vec{f}) \quad (24)$$

$$- (\vec{z} - \vec{f})^T \cdot \Sigma^{-1}G \left( I + G^T \Sigma^{-1}G \right)^{-1} G^T \Sigma^{-1} \cdot (\vec{z} - \vec{f})$$

Result is identical to  $\tilde{\chi}^2$  at minimum w.r.t.  $\vec{b}$ .

Conclusion:

**The extended  $\chi^2$  fit is exactly equivalent to a fit with the full covariance matrix.**

## Summary and Conclusions

- ▶ Any covariance matrix, e.g. from a unfolding problem, can be decomposed into an uncorrelated part and a correlated part, with a coefficient matrix for a number of sources of correlation.
- ▶ The extended  $\chi^2$  fit, where shifts and scale factors for all all sources of systematics are determined, is exactly equivalent to a fit with only the full correlation matrix of the data points.