Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Instruments and Methods in Physics Research A 500 (2003) 391–405
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Combining correlated measurements of several different physical quantities

Andrea Valassi

CERN, IT Division, Genève 23 CH-1211, Switzerland

Received 25 June 2002; received in revised form 23 October 2002; accepted 9 December 2002

Abstract

Measurements of different physical quantities are often correlated when they are performed by the same experiment, using the same data or the same detector. Correlations may also exist between the results of different experiments, for instance if they rely on the use of the same theoretical models. All these correlations must be properly taken into account to provide the best combined estimate of each measured quantity. A procedure used to combine the correlated results of different high-energy physics experiments is reviewed in this paper.

PACS: 02.50.Sk; 13.38.−b

Keywords: Best linear unbiased estimate; Correlated measurements; LEPEWWG

1. Introduction

Most modern high-energy physics experiments are designed to perform simultaneous measurements of different physical quantities. The various results produced by a single experiment are often correlated because they are produced using the same detector, and in some cases the same data sample. It is also common for the same physical quantities to be measured by similar techniques at more than one experiment, in some cases using detectors installed on the same accelerator facility. The results produced by different experiments may thus be significantly correlated too, if they are based on the same theoretical models or they depend on the knowledge of a common accelerator parameter.

In order to provide the best estimate of each measured observable, it is clearly desirable to make use of all results from the different experiments. A well-known technique to combine many measurements of a single physical quantity is the best linear unbiased estimate method described in Ref. [1]. A simple approach could then be to perform as many separate combinations as there are different measured observables, using the above method in each case. In the absence of correlations between the measurements of different quantities, this would yield the "best" combined estimate of each of them, i.e. that of minimum variance. However, if the measurements of the various observables are correlated, a better approach, leading to smaller variances for the

E-mail address: andrea.valassi@cern.ch (A. Valassi).

0168-9002/03/$ - see front matter
doi:10.1016/S0168-9002(03)00329-2

combined results, consists in performing a joint combination of all measured quantities at the same time, making use of the complete covariance matrix of the input data and thus taking into account all correlations between them.

The purpose of this paper is to review this technique, which represents a straightforward extension to the case of multiple observables of the method described in Ref. [1] for a single physical quantity. A simple proof of the algebraic solution to the problem of finding best linear unbiased estimates for a set of correlated observables will be recalled in Section 2, in close analogy to Ref. [1]. The problem is equivalent to that of minimizing a sum of squared distances of the measurements from the combined estimates, as in the case of one observable; the minimum value of this sum, distributed as a χ^2, can be used to assess the consistency of all input measurements, taking properly into account their correlations. While the problem can also be solved by minimizing the χ^2 using numerical methods, it will be argued that the algebraic solution makes it easier to break down the errors on the combined estimates into individual components, representing the contributions of the different sources of uncertainty on the input measurements. In Section 3, this method will be compared to the approach of performing as many separate combinations as there are measured physical quantities, ignoring all correlations between the measurements of different observables, proving the above statement that it leads to smaller uncertainties on each combined estimate. Some features and consequences of this technique will then be discussed in Section 4 in the specific case of an ad hoc example using fictitious experimental data. Finally, conclusions will be drawn in Section 5.

The problems described above are often encountered in the context of the activities of the LEP ElectroWeak Working Group (LEPEWWG), which is responsible for combining the precision measurements of various electroweak observables performed by the LEP experiments ALEPH, DELPHI, L3 and OPAL. In several occasions, combinations of published and preliminary results of the four experiments have already been performed by the technique reviewed in this paper, using either numerical methods, or one of the many implementations of the algebraic solution developed within the LEPEWWG.[1] Examples include, amongst others: the published combinations of heavy flavour electroweak results [5] and of Z boson parameters [6] measured at LEP1; the preliminary combination of $q\bar{q}$, $\mu^+\mu^-$ and $\tau^+\tau^-$ cross-sections, and of $\mu^+\mu^-$ and $\tau^+\tau^-$ forward–backward asymmetries, at 12 different LEP energies between 130 and 207 GeV [2]; the combinations of published [3,7] and preliminary [4] W decay branching fractions to $e\nu_e$, $\mu\nu_\mu$ and $\tau\nu_\tau$ from W-pair events at LEP2; the preliminary combination of W-pair production cross-sections at eight different LEP2 energies between 183 and 207 GeV [4].

2. The method

The technique reviewed in this paper can be applied to obtain combined estimates of N distinct physical quantities, each of which has been measured by one or more experiments, under the hypothesis that all sources of errors are multivariate Gaussian distributed. It is also assumed that the total covariance matrix for the input data is positive definite, is known a priori and does not depend on the results of the measurements.

A more general discussion of this method, including its extension to the case where the input measurements are linear combinations of the observables to be determined, can be found in many textbooks on multivariate analysis (see, for example, Ref. [8]). The goal of this paper is to formulate this procedure in the language used for the combination of results produced by different high-energy physics experiments, pointing out its relevance and its properties in this context.

Estimating the input covariance matrix is one of the most delicate steps in the combination of results. This was discussed in Ref. [1] for a specific example where many measurements of a single

[1] Two such implementations are those independently developed by I. Tomalin and J. Holt for the LEP2 difermion group [2] and by the author of this paper for the LEP2 four-fermion group [3,4].

observable had to be combined. The case reviewed in this paper presents the extra complication that the correlations between measurements of different observables must be determined in addition to those between the many measurements of the same quantity and to the errors on the individual results. A general discussion of the techniques that may be used and of the practical difficulties that may arise in this task is, however, beyond the scope of this paper. Only the effect of the correlations between the input measurements on the combined results and on their covariance matrix will be discussed in Section 4, with the help of a few numerical examples.

2.1. Best linear unbiased estimates

Using Greek indices $(\alpha, \beta, \gamma, \ldots)$ and Roman indices $(i, j, k, \ldots)$ to denote observables and measurements, respectively, let $X_\alpha = \{X_1, \ldots, X_N\}$ be the true values of the N observables and $y_i = \{y_1, \ldots, y_n\}$ the n experimental results. Let there be $n_\alpha \geqslant 1$ measurements for each observable X_α, such that $n = (\sum_{\alpha=1}^{N} n_\alpha) \geqslant N$. The link between a measurement y_i and the observable X_α it refers to will be expressed using the $(n \times N)$ matrix $\mathscr{U}$, defined by

$$\mathscr{U}_{i\alpha} = \begin{cases} 1 & \text{if } y_i \text{ is a measurement of } X_\alpha, \\ 0 & \text{if } y_i \text{ is not a measurement of } X_\alpha. \end{cases} \quad (1)$$

Each of the n rows of $\mathscr{U}$ has one and only one element equal to 1, whereas the αth of the N columns of $\mathscr{U}$ has n_α elements equal to 1. As a consequence, left multiplication by the matrix $\mathscr{U}$ of the vector of observables X, by contraction over the index α, singles out for every i the observable $(\mathscr{U}X)_i = (\sum_{\alpha=1}^{N} \mathscr{U}_{i\alpha} X_\alpha)$ measured by a given result y_i, while $\sum_{\alpha=1}^{N} \mathscr{U}_{i\alpha} = 1$ for every i. Conversely, left multiplication of the vector of measurements y by the transpose matrix $\tilde{\mathscr{U}}$, by contraction over the index i, yields for every α the sum $(\tilde{\mathscr{U}}y)_\alpha = (\sum_{i=1}^{n} \mathscr{U}_{i\alpha} y_i)$ of the n_α measurements of a given observable X_α. In particular, the sum of any vector quantity V_i over the index i can be rewritten as $(\sum_{i=1}^{n} V_i) = \sum_{\alpha=1}^{N} (\tilde{\mathscr{U}}V)_\alpha$ to group together the terms relative to measurements of the same observable.

Given the measurements and their $(n \times n)$ covariance matrix

$$\mathscr{M}_{ij} = \mathrm{cov}(y_i, y_j) = \mathrm{cov}(y_j, y_i) = \mathscr{M}_{ji}, \quad (2)$$

symmetric and positive definite, the problem consists in finding the "Best Linear Unbiased Estimates" (BLUE) of the true values X_α of the N observables, i.e. their estimates $\hat{x}_\alpha$ such that:

(A) the $\hat{x}_\alpha$ are *linear* combinations of the input measurements y_i;
(B) the $\hat{x}_\alpha$ are *unbiased* estimates of the true values X_α of the observables;
(C) the $\hat{x}_\alpha$ are the *best* amongst all estimates of the observables compatible with points A and B, i.e. those of minimum variance.

In the presence of measurements of different physical quantities, the most general way to express condition A is

$$\hat{x}_\alpha = \sum_{i=1}^{n} \lambda_{\alpha i} y_i = \sum_{\beta=1}^{N} \sum_{i=1}^{n} \lambda_{\alpha i} \mathscr{U}_{i\beta} y_i, \quad (3)$$

where the estimate $\hat{x}_\alpha$ of observable X_α is built as a linear combination of all input results y_i, irrespective of the observable which they measure.

Condition B requests that the expectation value of each $\hat{x}_\alpha$ be equal to the true value of the corresponding observable, $\mathrm{E}[\hat{x}_\alpha] = X_\alpha$. Assuming that each measurement is also unbiased, i.e. $\mathrm{E}[\mathscr{U}_{i\beta} y_i] = \mathscr{U}_{i\beta} X_\beta$ for every i and β, this translates into a normalization constraint for the linear weights of Eq. (3):

$$\sum_{i=1}^{n} \lambda_{\alpha i} \mathscr{U}_{i\beta} = \delta_{\alpha\beta} \quad \forall\alpha, \forall\beta, \quad (4)$$

where Kronecker's $\delta_{\alpha\beta}$ equals 1 if $\alpha = \beta$ and 0 otherwise. In other words, the n_α measurements of observable X_α contribute to $\hat{x}_\alpha$ with a total weight 1, while each set of n_β measurements of a different observable X_β, with $\beta \neq \alpha$, contributes to $\hat{x}_\alpha$ with a total weight 0.

From Eqs. (2) and (3), the $(N \times N)$ covariance matrix for the linear estimates $\hat{x}_\alpha$ is simply given by

$$\mathrm{cov}(\hat{x}_\alpha, \hat{x}_\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\beta j}. \quad (5)$$

The diagonal elements of this matrix represent the variances of the $\hat{x}_\alpha$:

$$\mathrm{var}(\hat{x}_\alpha) = \mathrm{cov}(\hat{x}_\alpha, \hat{x}_\alpha) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j}, \tag{6}$$

i.e. the squares of their total errors. Condition C requires that the linear weights $\lambda_{\alpha i}$ be determined as those that minimize all variances in Eq. (6) under the normalization constraints given in Eq. (4). Since the weights $\lambda_{\alpha i}$ and $\lambda_{\beta i}$ for two different estimates $\hat{x}_\alpha$ and $\hat{x}_\beta$, with $\alpha \neq \beta$, never appear together in either of Eqs. (4) and (6), the minimization problem can be solved separately for the N observables. For every α, the n linear weights $\lambda_{\alpha i}$ can be obtained by introducing N Lagrange multipliers $K_{\alpha\beta}$, as they must satisfy n differential equations under the N constraints of Eq. (4). Differentiating with respect to $K_{\alpha\beta}$ and $\lambda_{\alpha i}$ the sum $[\mathrm{var}(\hat{x}_\alpha) + 2\sum_{\gamma=1}^{N} K_{\alpha\gamma}(\delta_{\alpha\gamma} - \sum_{j=1}^{n} \lambda_{\alpha j} \mathscr{U}_{j\gamma})]$ yields

$$\begin{cases} \delta_{\alpha\beta} - \sum_{j=1}^{n} \lambda_{\alpha j} \mathscr{U}_{j\beta} = 0 & \forall\alpha,\ \forall\beta, \\ (\sum_{j=1}^{n} \mathscr{M}_{ij} \lambda_{\alpha j}) - (\sum_{\gamma=1}^{N} K_{\alpha\gamma} \mathscr{U}_{i\gamma}) = 0 & \forall\alpha,\ \forall i. \end{cases} \tag{7}$$

Using $\mathscr{M}^{-1}$ to denote the inverse of the covariance, i.e. the $(n \times n)$ symmetric matrix such that $\sum_{k=1}^{n} \mathscr{M}_{ik} \mathscr{M}^{-1}_{kj} = \sum_{k=1}^{n} \mathscr{M}^{-1}_{ik} \mathscr{M}_{kj} = \delta_{ij}$, it is easy to show that the solutions to the linear system of Eq. (7) are given by $K_{\alpha\gamma} = (\tilde{\mathscr{U}} \mathscr{M}^{-1} \mathscr{U})^{-1}_{\alpha\gamma}$ and

$$\lambda_{\alpha i} = \sum_{\beta=1}^{N} (\tilde{\mathscr{U}} \mathscr{M}^{-1} \mathscr{U})^{-1}_{\alpha\beta} (\tilde{\mathscr{U}} \mathscr{M}^{-1})_{\beta i}. \tag{8}$$

The coefficients $\lambda_{\alpha i}$ clearly satisfy the normalization condition in Eq. (4), since

$$\sum_{i=1}^{n} \lambda_{\alpha i} \mathscr{U}_{i\gamma} = \sum_{\beta=1}^{N} (\tilde{\mathscr{U}} \mathscr{M}^{-1} \mathscr{U})^{-1}_{\alpha\beta} (\tilde{\mathscr{U}} \mathscr{M}^{-1} \mathscr{U})_{\beta\gamma} = \delta_{\alpha\gamma}. \tag{9}$$

By substituting the weights of Eq. (8) into Eq. (5), the covariance between the best linear unbiased estimates $\hat{x}_\alpha$ is found to be

$$\mathrm{cov}(\hat{x}_\alpha, \hat{x}_\beta) = (\tilde{\mathscr{U}} \mathscr{M}^{-1} \mathscr{U})^{-1}_{\alpha\beta}. \tag{10}$$

In deriving Eq. (8), it has been assumed that the covariance matrix $\mathscr{M}$ can be inverted, i.e. that none of the n input measurements can be written as a linear combination of the others. If two of the input measurements were 100% correlated, for instance, $\mathscr{M}$ could not be inverted as the two measurements would be redundant: one of them would bring no additional information and should then be removed from the combination. As already pointed out in Ref. [1], one must in any case be very careful in the presence of large correlations between the input measurements, i.e. if the discriminant of the covariance matrix is close to zero: in that case, the weights of Eq. (8) may be very large and the results of the combination may become unstable with respect to slight biases in the input measurements or misassessments of the covariance matrix.

2.2. Linear combinations of observables

It is interesting to note that the $\hat{x}_\alpha$, determined above as the BLUE of the N measured observables X_α, also represent a linear base to obtain best unbiased linear estimates of any observable which can be built as a linear combination of the X_α. More precisely, if Z is the true value of one such observable,

$$Z = \sum_{\alpha=1}^{N} a_\alpha X_\alpha, \tag{11}$$

with coefficients a_α fixed a priori and not all equal to 0, the best unbiased linear estimate of Z determined from the n measurements y_i of $(\mathscr{U}X)_i$ is

$$\hat{z} = \sum_{\alpha=1}^{N} a_\alpha \hat{x}_\alpha, \tag{12}$$

where the $\hat{x}_\alpha$ are those given in Section 2.1. The variance of $\hat{z}$

$$\mathrm{var}(\hat{z}) = \sum_{\alpha=1}^{N} \sum_{\beta=1}^{N} a_\alpha a_\beta \,\mathrm{cov}(\hat{x}_\alpha, \hat{x}_\beta) \tag{13}$$

is, in fact, the minimum variance for any linear unbiased estimate of Z, because the covariance matrix for the $\hat{x}_\alpha$ is smaller than the covariance matrix of any other unbiased linear estimates of X_α, as discussed more in detail in Ref. [8]. Some examples of this statement will be given in Section 4.

2.3. χ^2 minimization

As in the case of a single physical quantity, the problem of finding the linear unbiased estimates of minimum variance for the N observables X_α is equivalent to the problem of finding the estimates $\hat{x}_\alpha$ minimizing the sum

$$\begin{aligned} S &= -2\log\mathscr{L} = \sum_{i=1}^{n}\sum_{j=1}^{n}[y_i-(\mathscr{U}\hat{x})_i] \\ &\quad\times \mathscr{M}_{ij}^{-1}[y_j-(\mathscr{U}\hat{x})_j] \\ &= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\sum_{i=1}^{n}\sum_{j=1}^{n}[\mathscr{U}_{i\alpha}(y_i-\hat{x}_\alpha)] \\ &\quad\times \mathscr{M}_{ij}^{-1}[\mathscr{U}_{j\beta}(y_j-\hat{x}_\beta)], \end{aligned} \tag{14}$$

which measures the "distance" of the measurements y_i from the corresponding linear estimates $(\mathscr{U}\hat{x})_i$. Assuming that the experimental results y_i are multivariate Gaussian distributed, this corresponds to maximizing the likelihood $\mathscr{L}$ that each y_i is centred around $(\mathscr{U}\hat{x})_i$. By requiring that

$$\frac{\partial S}{\partial \hat{x}_\alpha} = 0 \quad \forall\alpha, \tag{15}$$

in fact, without any a priori assumptions that the combined estimates $\hat{x}_\alpha$ be linear or unbiased, it can easily be proved that Eqs. (3) and (8) must hold. In particular, it is interesting to note that an exact algebraic solution exists to Eq. (15), while this is often solved using numerical methods only.

The minimum value of S, i.e. that calculated for the best linear unbiased estimates $\hat{x}_\alpha$, is an interesting quantity in itself. Assuming that all errors are Gaussian, the minimum of S is distributed as a χ^2 with $(n-N)$ degrees of freedom and can be used to assess the extent to which the individual results y_i are consistent with the combined estimates $(\mathscr{U}\hat{x})_i$, i.e. with the hypothesis that they measure the same observables $(\mathscr{U}X)_i$.

2.4. Breakdown of error contributions

A good feature of the BLUE method is that it makes it very easy to break down the error matrix for the combined estimates into its individual components, such as those of statistical and systematic origin. Suppose, for instance, that U independent sources of uncertainty have been identified, i.e. that the total covariance matrix $\mathscr{M}_{ij}$ for the input measurements can be written as the sum

$$\begin{aligned} \mathscr{M}_{ij} &= \mathrm{cov}(y_i,y_j) = \sum_{u=1}^{U}\mathrm{cov}^{[u]}(y_i,y_j) \\ &= \sum_{u=1}^{U}\mathscr{M}_{ij}^{[u]}. \end{aligned} \tag{16}$$

The individual contributions to $\mathscr{M}$ can be separately propagated to the covariance matrix for the combined estimates $\hat{x}_\alpha$, yielding

$$\mathrm{cov}(\hat{x}_\alpha,\hat{x}_\beta) = \sum_{u=1}^{U}\mathrm{cov}^{[u]}(\hat{x}_\alpha,\hat{x}_\beta). \tag{17}$$

From Eq. (5), the contribution from the uth error source is simply given by

$$\mathrm{cov}^{[u]}(\hat{x}_\alpha,\hat{x}_\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{\alpha i}\mathscr{M}_{ij}^{[u]}\lambda_{\beta j}, \tag{18}$$

which can be computed exactly using the weights $\lambda_{\alpha i}$ of Eq. (8). The detailed breakdown of the error matrix is thus very simple to obtain if the problem is solved using the BLUE method: the weights $\lambda_{\alpha i}$ are already available for free, as they are needed to compute the central values and the total covariance matrix for the combined estimates $\hat{x}_\alpha$ according to Eqs. (3) and (5).

For comparison, the detailed breakdown of the error matrix can also be obtained if results are combined using a numerical approach instead of the BLUE method, although more minimizations are needed than that of the χ^2 in Eq. (15) that yields the central values and the total covariance of the $\hat{x}_\alpha$. This can be achieved, for instance, by the following technique, which was used [9] for the combination of LEP heavy-flavour results described in Ref. [5]. For every i, a new sum $S_{(\Delta y_i)}$ may be built, analogous to that in Eq. (14), assuming that the value of the ith input measurement is shifted by a fixed amount to $y_i+\Delta y_i$, while the values of all other y_j (with $j\neq i$) are unchanged. Since the problem is linear, the differences between the $\hat{x}_\alpha$ and the results of the numerical minimization of $S_{(\Delta y_i)}$ may be used to derive, for every i, the N derivatives $\partial\hat{x}_\alpha/\partial y_i$. From elementary error

propagation [7], the individual contributions to the covariance of $\hat{x}_\alpha$ and $\hat{x}_\beta$ may then be computed as

$$\begin{aligned}\mathrm{cov}^{[u]}(\hat{x}_\alpha, \hat{x}_\beta) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial \hat{x}_\alpha}{\partial y_i}\right)\{\mathrm{cov}^{[u]}(y_i, y_j)\}\left(\frac{\partial \hat{x}_\beta}{\partial y_j}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial \hat{x}_\alpha}{\partial y_i}\right)\mathscr{M}_{ij}^{[u]}\left(\frac{\partial \hat{x}_\beta}{\partial y_j}\right). \end{aligned} \tag{19}$$

This expression is strictly equivalent to Eq. (18), because Eq. (3) implies that

$$\frac{\partial \hat{x}_\alpha}{\partial y_i} = \lambda_{\alpha i} \quad \forall\alpha,\ \forall i. \tag{20}$$

In other words, the advantage of the BLUE technique over a numerical approach is that it is simpler and more elegant: the weights $\lambda_{\alpha i}$ are determined a priori from matrix algebra and are then used to compute both the combined estimates and all contributions to the error matrix, whereas, when the problem is solved by numerical methods, the derivatives $\partial\hat{x}_\alpha/\partial y_i$ must be determined a posteriori, from the results of more than one numerical minimizations, and are only used to compute the detailed breakdown of the error matrix. The advantage of the numerical approach, instead, is that it is more flexible, because it can be used also when the input measurements depend on one another or when the problem is not strictly linear or Gaussian, as pointed out in Ref. [5].

3. Effects of the correlations between different observables

In the absence of correlations between the measurements of different observables, it is easy to show that the method described in the previous section reduces to the simple approach of finding the best estimate of each observable from its measurements alone, using the technique of Ref. [1] in each case.

Let $\mathscr{M}'$ be the $(n \times n)$ covariance matrix to which $\mathscr{M}$ reduces when neglecting all correlations between different observables:

$$\mathscr{M}'_{ij} = \begin{cases} \mathscr{M}_{ij} & \text{if } y_i \text{ and } y_j \text{ measure the same observable,} \\ 0 & \text{if } y_i \text{ and } y_j \text{ measure different observables.} \end{cases} \tag{21}$$

Assuming for simplicity that the index i over the y_i is such that the n_α measurements of the same observable X_α are grouped together and ordered for increasing values of α, this means that $\mathscr{M}'$ is a block-diagonal matrix with N blocks, each of dimensions $(n_\alpha \times n_\alpha)$. Formally, Eq. (21) implies that

$$\mathscr{M}'_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\beta} = \mathscr{M}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\beta}\delta_{\alpha\beta} \quad \forall i,\ \forall j,\ \forall\alpha,\ \forall\beta, \tag{22}$$

so that $\mathscr{M}'$ can be rewritten as

$$\begin{aligned}\mathscr{M}'_{ij} &= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\mathscr{M}'_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\beta} = \sum_{\alpha=1}^{N}\mathscr{M}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\alpha} \\ &= \sum_{\alpha=1}^{N}(\mathscr{M}')_{ij}^{(\alpha)}, \end{aligned} \tag{23}$$

i.e. as the sum of N matrices $(\mathscr{M}')_{ij}^{(\alpha)} = \mathscr{M}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\alpha}$, each of which has all elements equal to zero except for one block of dimensions $(n_\alpha \times n_\alpha)$.

The inverse covariance $\mathscr{M}'^{-1}$ is also a block-diagonal matrix

$$\begin{aligned}\mathscr{M}'^{-1}_{ij} &= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\mathscr{M}'^{-1}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\beta} = \sum_{\alpha=1}^{N}\mathscr{M}'^{-1}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\alpha} \\ &= \sum_{\alpha=1}^{N}(\mathscr{M}'^{-1})_{ij}^{(\alpha)}, \end{aligned} \tag{24}$$

because it can be written as the sum of N matrices $(\mathscr{M}'^{-1})_{ij}^{(\alpha)} = \mathscr{M}'^{-1}_{ij}\mathscr{U}_{i\alpha}\mathscr{U}_{j\alpha}$, each of which, of dimensions $(n \times n)$, has all elements equal to zero except for one $(n_\alpha \times n_\alpha)$ block which is the inverse of the corresponding block of $(\mathscr{M}')^{(\alpha)}$:

$$\begin{aligned}\sum_{k=1}^{n}(\mathscr{M}'^{-1})_{ik}^{(\alpha)}(\mathscr{M}')_{kj}^{(\alpha)} &= \sum_{k=1}^{n}(\mathscr{M}')_{ik}^{(\alpha)}(\mathscr{M}'^{-1})_{kj}^{(\alpha)} \\ &= \mathscr{U}_{i\alpha}\mathscr{U}_{j\alpha}\delta_{ij}. \end{aligned} \tag{25}$$

In the absence of correlations between different observables, $(\tilde{\mathscr{U}}\mathscr{M}^{-1}\mathscr{U})$ in Eq. (8) thus reduces to the $(N \times N)$ diagonal matrix $(\tilde{\mathscr{U}}\mathscr{M}'^{-1}\mathscr{U})$, and the

best linear unbiased estimates of the N observables are simply given by

$$\hat{x}'_\alpha = \sum_{i=1}^{n} \lambda'_{\alpha i} y_i = \sum_{i=1}^{n} \mathscr{U}_{i\alpha} \lambda'_{\alpha i} y_i, \tag{26}$$

where the linear weight

$$\lambda'_{\alpha i} = \frac{[\tilde{\mathscr{U}}(\mathscr{M}'^{-1})^{(\alpha)}]_{\alpha i}}{[\tilde{\mathscr{U}}(\mathscr{M}'^{-1})^{(\alpha)}\mathscr{U}]_{\alpha\alpha}} = \frac{[\sum_{j=1}^{n} (\mathscr{M}'^{-1})^{(\alpha)}_{ji}]}{[\sum_{k=1}^{n} \sum_{j=1}^{n} (\mathscr{M}'^{-1})^{(\alpha)}_{jk}]}, \tag{27}$$

with which an experimental result y_i contributes to the BLUE of observable X_α, is 0 unless y_i is a measurement of X_α. Eqs. (26) and (27) above are equivalent to Eqs. (3) and (6) of Ref. [1], proving that the method presented reduces to that of Ref. [1] if the problem is that of finding best linear unbiased estimates for N uncorrelated measured observables.

Although the $\hat{x}_\alpha$ derived above only represent best linear unbiased estimates in the absence of correlations between different observables, it is interesting to consider them as "approximate" solutions of the BLUE problem when such correlations exist but are ignored, i.e. when the covariance matrix for the measurements is $\mathscr{M}$, but $\mathscr{M}'$ is used in Eq. (8). In particular, provided that the y_i are unbiased, the $\hat{x}'_\alpha$ still represent unbiased estimates of the true values X_α of the observables, although not those of minimum variance. In the following section, the $\hat{x}'_\alpha$ will be used to study how the properties of BLUE solutions change in the presence or in the absence of correlations between measurements of different observables. The minimum variances and χ^2 values obtained in the two cases will be compared in Sections 3.1 and 3.2, respectively.

3.1. Comparison of minimum variances

As already noted in Section 2.1, the problem of determining best linear unbiased estimates for a set of N observables always "diagonalizes" to that of finding the BLUE of each observable as the solution to an independent minimization problem, even when the measurements of different observables are correlated. For each X_α, in particular, the unbiased linear estimate of minimum variance $\hat{x}_\alpha$ is obtained by minimizing the quadratic form of Eq. (6) in the $(n-N)$-dimensional linear space spanned by the n weights $\lambda_{\alpha i}$, under the N normalization constraints of Eq. (4). The function to minimize

$$\begin{aligned} \mathrm{var}\left(\sum_{i=1}^{n} \lambda_{\alpha i} y_i\right) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j} \\ &= \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{n} \\ &\quad \times \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j} \mathscr{U}_{i\beta} \mathscr{U}_{j\gamma} \end{aligned} \tag{28}$$

can also be rewritten as

$$\begin{aligned} \mathrm{var}\left(\sum_{i=1}^{n} \lambda_{\alpha i} y_i\right) &= \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j} \mathscr{U}_{i\alpha} \mathscr{U}_{j\alpha}\right) \\ &+ \sum_{\beta(\beta\neq\alpha)}^{1,\ldots,N} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j} \mathscr{U}_{i\beta} \mathscr{U}_{j\beta}\right) \\ &+ \sum_{\beta,\gamma(\beta\neq\gamma)}^{1,\ldots,N} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{\alpha i} \mathscr{M}_{ij} \lambda_{\alpha j} \mathscr{U}_{i\beta} \mathscr{U}_{j\gamma}\right), \end{aligned} \tag{29}$$

where the $(n \times n)$ non-zero terms of Eq. (28) have been grouped into three lines containing, respectively, $(n_\alpha \times n_\alpha)$, $\sum_{\beta\ (\beta\neq\alpha)}^{1,\ldots,N}(n_\beta \times n_\beta)$ and $\sum_{\beta,\gamma\ (\beta\neq\gamma)}^{1,\ldots,N}(n_\beta \times n_\gamma)$ terms. The expression above represents the functional dependence, on the coefficients $\lambda_{\alpha i}$ and the matrix $\mathscr{M}_{ij}$, of the variance of a generic linear combination $(\sum_{i=1}^{n} \lambda_{\alpha i} y_i)$ of the measurements y_i, whether or not this function is calculated at its minimum. This illustrates that the variance of the unbiased estimate $\hat{x}'_\alpha$, defined by the coefficients $\lambda'_{\alpha i}$ of Eq. (27), is the same in the presence and in the absence of correlations between different observables, i.e. whether $\mathscr{M}$ or $\mathscr{M}'$ is used in Eq. (29). Since $\lambda'_{\alpha i} = 0$, if y_i is not a measurement of X_α, in fact, both the second and third lines are 0 and only the first remains:

$$\begin{aligned} \mathrm{var}^{(\mathscr{M})}(\hat{x}'_\alpha) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda'_{\alpha i} \mathscr{M}_{ij} \lambda'_{\alpha j} \mathscr{U}_{i\alpha} \mathscr{U}_{j\alpha} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda'_{\alpha i} \mathscr{M}'_{ij} \lambda'_{\alpha j} \mathscr{U}_{i\alpha} \mathscr{U}_{j\alpha} \\ &= \mathrm{var}^{(\mathscr{M}')}(\hat{x}'_\alpha). \end{aligned} \tag{30}$$

On the other hand, in the presence of correlations between different observables, $\hat{x}_\alpha$ is the unbiased linear estimate of minimum variance, so that $\mathrm{var}^{(\mathscr{M})}(\hat{x}_\alpha) \leqslant \mathrm{var}^{(\mathscr{M})}(\hat{x}'_\alpha)$. This implies that

$$\mathrm{var}^{(\mathscr{M})}(\hat{x}_\alpha) \leqslant \mathrm{var}^{(\mathscr{M}')}(\hat{x}'_\alpha). \quad (31)$$

In other words, in the presence of correlations between measurements of different observables, whether positive or negative, the error on each combined estimate is always lower than in their absence (or, at most, equal). Neglecting these correlations leads to a loss of information, as the combined estimate for each observable X_α is built only from the n_α measurements of X_α; conversely, useful information is added if the measurements of all other observables are also included. In fact, the reduction in variance is only possible thanks to the measurements of observables other than X_α, and more particularly to the third line of Eq. (29), whose contribution is generally negative.

The inequality in Eq. (31) can also be understood by realizing that $\mathrm{var}^{(\mathscr{M})}(\hat{x}_\alpha)$ is the absolute minimum of the variance of Eq. (29) in the full $(n-N)$-dimensional linear space spanned by weights $\lambda_{\alpha i}$, under the N normalization constraints of Eq. (4), while $\mathrm{var}^{(\mathscr{M}')}(\hat{x}'_\alpha)$ is the minimum of the same quadratic form in a smaller $(n_\alpha - 1)$-dimensional linear subspace defined by the additional constraint that $\lambda'_{i\alpha} = 0$ if y_i is not a measurement of X_α.

In the presence of correlations between measurements of different observables, even if the variance for the "approximate" solution $\hat{x}'_\alpha$ in Eq. (29) is the same whether it is computed using $\mathscr{M}$ or $\mathscr{M}'$, all such correlations must, of course, be taken into account if one needs to analyse several observables at the same time. In particular, all correlations must be propagated to the covariances between the $\hat{x}_\alpha$, according to Eq. (5): this is important, for instance, to estimate the variance of a quantity which depends on more than one $\hat{x}_\alpha$, or if measurements of different observables are used at the same time in a theoretical framework which predicts values for them all. As discussed in Section 3.2, it is also essential to take into account these correlations to estimate the χ^2 for the set of all input measurements, and to assess the extent to which these are compatible with one another. In all these situations, ignoring the correlations between the measurements of different observables leads to wrong results, just like it happens in the combination of measurements of one single observable if their correlations are ignored. Some examples will be presented in Section 4.

3.2. Comparison of minimum χ^2

Comparing the χ^2 values for the combination of results in the presence and in the absence of correlations between different observables is not as easy as comparing the variances obtained in the two cases. In the presence of correlations, best linear unbiased estimates are obtained by minimizing the sum S of Eq. (14). In the absence of correlations, $\mathscr{M}'$ must be used instead of $\mathscr{M}$ and S reduces to the sum $S' = \sum_{\alpha=1}^{N} (S')^{(\alpha)}$ of N independent terms:

$$(S')^{(\alpha)} = \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - \hat{x}_\alpha)(\mathscr{M}'^{-1})^{(\alpha)}_{ij}(y_j - \hat{x}_\alpha), \quad (32)$$

which can be separately minimized to yield the solutions $\hat{x}_\alpha$ given before.

Both sums S and S' have well-defined meanings, but only under different circumstances, as their functional dependence on the weights $\lambda_{\alpha i}$ is not the same. The minimum values of S and S', respectively, in the presence and in the absence of correlations between measurements of different observables (i.e. the values computed for the corresponding BLUE $\hat{x}_\alpha$ and $\hat{x}'_\alpha$), are both distributed as a χ^2 with $(n-N)$ degrees of freedom. It does not make sense, instead, to compute S for $\hat{x}'_\alpha$, or S' for $\hat{x}_\alpha$, as these values would not represent the minima of S and S'. In particular, there is no simple relationship between the χ^2 for the BLUE combination in the absence and in the presence of correlations between measurements of different observables (i.e., between the minima of S' and S): sometimes, such correlations may impose tighter constraints on the measurements and result in larger χ^2 values, whereas at other times the contrary may happen. While the variances of the BLUE are completely determined by the covariance matrix for the measurements (as shown for instance by

Eq. (10)), the comparison of χ^2 values is further complicated by the fact that these values depend on the central values of the input measurements. Simple examples of these statements will be presented in Section 4.

Finally, it is interesting to note that each of the $(S')^{(\alpha)}$ also has a well-defined meaning, as it is obtained using only the measurements of X_α and the corresponding covariance matrix. The minimum of $(S')^{(\alpha)}$ is that computed for the BLUE $\hat{x}'_\alpha$ and is distributed as a χ^2 with $(n_\alpha - 1)$ degrees of freedom. Whether or not correlations between measurements of different observables exist, this is the correct quantity to use if one wants to measure the internal consistency of the subset of all measurements of X_α alone, ignoring the measurements of all other observables.

4. Examples

For the purpose of illustration, the application of the method to a fictitious example is discussed in this section. The case of two experiments A and B measuring the branching fractions of the W boson in the two decay channels to electrons, B^e, and to taus, B^τ, is considered. The problem consists in finding the best linear unbiased estimates $\hat{B}^e$ and $\hat{B}^\tau$, given the four measurements

$$\begin{aligned}
\mathscr{B}^e_A &= (10.50 \pm 1.00)\%,\\
\mathscr{B}^e_B &= (13.50 \pm 3.00)\%,\\
\mathscr{B}^\tau_A &= (9.50 \pm 3.00)\%,\\
\mathscr{B}^\tau_B &= (14.00 \pm 3.00)\%,
\end{aligned} \tag{33}$$

and their covariance matrix. Various examples will be presented in the following section, differing only in the assumptions made about the correlations between the four measurements in Eq. (33).

All results presented below have been obtained using the formulas derived in Section 2; the central values and total errors on $\hat{B}^e$ and $\hat{B}^\tau$ have also been cross-checked by the numerical minimization of the χ^2 using Minuit [10]. Using the same vector notation adopted in Section 2, let $\mathscr{B}_i = \{\mathscr{B}^e_A, \mathscr{B}^e_B; \mathscr{B}^\tau_A, \mathscr{B}^\tau_B\}$ and $B_\alpha = \{B^e; B^\tau\}$ denote the four measurements and the true values of the two observables, respectively. The covariance matrix defined in Eq. (2) is then given explicitly by

$$\mathscr{M} = \begin{pmatrix}
\mathrm{cov}(\mathscr{B}^e_A, \mathscr{B}^e_A) & \mathrm{cov}(\mathscr{B}^e_A, \mathscr{B}^e_B) & \mathrm{cov}(\mathscr{B}^e_A, \mathscr{B}^\tau_A) & \mathrm{cov}(\mathscr{B}^e_A, \mathscr{B}^\tau_B)\\
\mathrm{cov}(\mathscr{B}^e_B, \mathscr{B}^e_A) & \mathrm{cov}(\mathscr{B}^e_B, \mathscr{B}^e_B) & \mathrm{cov}(\mathscr{B}^e_B, \mathscr{B}^\tau_A) & \mathrm{cov}(\mathscr{B}^e_B, \mathscr{B}^\tau_B)\\
\mathrm{cov}(\mathscr{B}^\tau_A, \mathscr{B}^e_A) & \mathrm{cov}(\mathscr{B}^\tau_A, \mathscr{B}^e_B) & \mathrm{cov}(\mathscr{B}^\tau_A, \mathscr{B}^\tau_A) & \mathrm{cov}(\mathscr{B}^\tau_A, \mathscr{B}^\tau_B)\\
\mathrm{cov}(\mathscr{B}^\tau_B, \mathscr{B}^e_A) & \mathrm{cov}(\mathscr{B}^\tau_B, \mathscr{B}^e_B) & \mathrm{cov}(\mathscr{B}^\tau_B, \mathscr{B}^\tau_A) & \mathrm{cov}(\mathscr{B}^\tau_B, \mathscr{B}^\tau_B)
\end{pmatrix} \tag{34}$$

and the matrix $\mathscr{U}$ of Eq. (1) by

$$\mathscr{U} = \begin{pmatrix} 1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1 \end{pmatrix}, \tag{35}$$

where $\mathscr{U}_{i\alpha} = 1$ if $\mathscr{B}_i$ is a measurement of observable B_α, 0 otherwise.

In the framework of the Standard Model, it is common [3,4] to measure the branching fractions of the W boson to leptons not only in the three separate decay channels to electrons, muons and taus, but also under the (so-called "lepton universality") assumption that the three are equal, $B^\ell = B^e = B^\mu = B^\tau$. In the examples considered in this section, a best linear unbiased estimate $\hat{B}^\ell$ will also be derived by combining the four experimental results for different channels and from different experiments, $\mathscr{B}^e_A, \mathscr{B}^e_B, \mathscr{B}^\tau_A, \mathscr{B}^\tau_B$, as if they were all measurements of the same observable B^ℓ. The same formulas needed for the combination without the assumption of lepton universality will be used, with the differences that the vector of observables has only one element $B_\alpha = \{B^\ell\}$ and that $\mathscr{U}$ reduces to a (4×1) matrix whose components are all 1.

4.1. No correlations

In the absence of any correlations, the covariance matrix for the measurements listed in Eq. (33) is simply

$$\mathscr{M} = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 0 & 9.00 & 0 & 0 \\ 0 & 0 & 9.00 & 0 \\ 0 & 0 & 0 & 9.00 \end{pmatrix} \times 10^{-4}. \quad (36)$$

As intuitively expected, the best linear unbiased estimates for B^e and B^τ are

$$\begin{aligned} \hat{B}^e &= (0.900 \times \mathscr{B}^e_A) + (0.100 \times \mathscr{B}^e_B) \\ &\quad + (0.000 \times \mathscr{B}^\tau_A) + (0.000 \times \mathscr{B}^\tau_B), \\ \hat{B}^\tau &= (0.000 \times \mathscr{B}^e_A) + (0.000 \times \mathscr{B}^e_B) \\ &\quad + (0.500 \times \mathscr{B}^\tau_A) + (0.500 \times \mathscr{B}^\tau_B), \end{aligned} \quad (37)$$

where the two measurements of each branching fraction contribute only to the combined estimate for that decay channel, each with a linear weight inversely proportional to the square of its error. This leads to

$$\begin{aligned} \hat{B}^e &= (10.80 \pm 0.95)\%, \\ \hat{B}^\tau &= (11.75 \pm 2.12)\%, \end{aligned} \quad (38)$$

where the errors on B^e and B^τ are reduced with respect to the smaller of the errors on the input measurements, by $1/\sqrt{10/9} \simeq 1/\sqrt{1.1}$ and $1/\sqrt{2}$, respectively. The χ^2 for the combinations are 0.90 and 1.12, respectively, for 1 degree of freedom, summing up to 2.02 for 2 degrees of freedom for the joint combination of results. The covariance between $\hat{B}^e$ and $\hat{B}^\tau$ is, of course, 0.

Similarly, the combination under the assumption of lepton universality yields

$$\begin{aligned} \hat{B}^\ell &= (0.750 \times \mathscr{B}^e_A) + (0.083 \times \mathscr{B}^e_B) \\ &\quad + (0.083 \times \mathscr{B}^\tau_A) + (0.083 \times \mathscr{B}^\tau_B), \end{aligned} \quad (39)$$

where the four linear weights for $\hat{B}^\ell$ are inversely proportional to the squares of the errors on the four measurements. Numerically,

$$\hat{B}^\ell = (10.96 \pm 0.87)\%, \quad (40)$$

with a χ^2 of 2.19 for 3 degrees of freedom.

4.2. Correlations between measurements of the same observable

Assume now that a $+15\%$ correlation exists between the measurements of B^e performed by A and B, for the same errors given before. This could reflect a common systematic error affecting both experiments in the same direction, for instance a theoretical uncertainty on the production cross-section for a background process in the electron channel. The covariance matrix is then

$$\mathscr{M} = \begin{pmatrix} 1.00 & 0.45 & 0 & 0 \\ 0.45 & 9.00 & 0 & 0 \\ 0 & 0 & 9.00 & 0 \\ 0 & 0 & 0 & 9.00 \end{pmatrix} \times 10^{-4}. \quad (41)$$

The combination of B^τ, of course, remains unchanged, while the BLUE for B^e is now given by $\hat{B}^e = (0.940 \times \mathscr{B}^e_A) + (0.060 \times \mathscr{B}^e_B)$,

$$\begin{aligned} \hat{B}^e &= (10.68 \pm 0.98)\%, \\ \hat{B}^\tau &= (11.75 \pm 2.12)\%. \end{aligned} \quad (42)$$

The error on $\hat{B}^e$ is larger than the 0.95% found in the absence of correlations, because the measurements by A and B are not independent and their combination does not add as much information as required. As discussed in detail in Ref. [1], this error would be as large as that on the measurement with least variance, i.e. 1.00%, in the case that the correlation were equal to $\sigma_{\mathscr{B}^e_A}/\sigma_{\mathscr{B}^e_B} = 33\%$ (indicating a common contribution of 1.00% to both $\Delta\mathscr{B}^e_A$ and $\Delta\mathscr{B}^e_B$ from a fully correlated error source), while it would decrease again for larger correlations.

The χ^2 for combining B^e and B^τ are now 0.99 and 1.12, respectively, for 1 degree of freedom, summing up to 2.11 for 2 degrees of freedom for the joint combination. The increased χ^2 for the combination of B^e, with respect to 0.90 in the absence of correlations, reflects the fact that the $+15\%$ correlation makes the two measurements of B^e less compatible with each other: the expected error on $(\mathscr{B}^e_A - \mathscr{B}^e_B)$ is smaller, while the same measured central values are assumed. Of course, the covariance between $\hat{B}^e$ and $\hat{B}^\tau$ is 0 also in this case.

Although this is a less frequent case, $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$ can also be negatively correlated, for instance if the same source of systematic error affects the two measurements in opposite directions. For a -15% correlation, the BLUE for B^{e} is $\hat{B}^{e} = (0.867 \times \mathscr{B}^{e}_{A}) + (0.133 \times \mathscr{B}^{e}_{B})$, or, numerically, $\hat{B}^{e} = (10.90 \pm 0.90)\%$. The uncertainty on $\hat{B}^{e}$ is in this case smaller than the 0.95% observed in the absence of correlations, as the decrease in the error on $(\mathscr{B}^{e}_{A} + \mathscr{B}^{e}_{B})$ makes it possible to better constrain the value of B^{e}, around which both $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$ are distributed. The χ^2 for the combination of B^{e} is also reduced, to 0.83 for 1 degree of freedom, as the worse resolution on $(\mathscr{B}^{e}_{A} - \mathscr{B}^{e}_{B})$ allows the two measurements to disagree more with each other.

Similar trends to those observed for B^{e} can be noted for the combinations of B^{ℓ} performed under the assumption of lepton universality. For a $+15\%$ correlation between $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$, the error on

$$\hat{B}^{\ell} = (10.87 \pm 0.89)\%, \tag{43}$$

and the χ^2 value of 2.32 for 3 degrees of freedom are both larger than the 0.87% and 2.19 observed for no correlations. Similarly, for a -15% correlation between $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$, the error on $\hat{B}^{\ell} = (11.03 \pm 0.83)\%$ and the χ^2 of 2.09 for 3 degrees of freedom are smaller than those in the absence of correlations.

4.3. Correlations between measurements of different observables

4.3.1. Positive correlations

The effect of correlations between measurements of different observables is best understood by discussing a very extreme case: assume now that a $+99.5\%$ correlation exists between the measurements of B^{e} and B^{τ} performed by B, while the results of A and B are uncorrelated. This means that experiment B can measure the difference $(B^{e} - B^{\tau})$ with an extremely good resolution. This situation could be imagined, for instance, for an experiment with very small statistical errors, limited only by a systematic error affecting both channels in the same way. In this case, the covariance matrix is

$$\mathscr{M} = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 0 & 9.00 & 0 & 8.96 \\ 0 & 0 & 9.00 & 0 \\ 0 & 8.96 & 0 & 9.00 \end{pmatrix} \times 10^{-4}, \tag{44}$$

and the best linear unbiased estimates for B^{e} and B^{τ} are found to be

$$\begin{aligned} \hat{B}^{e} &= (0.820 \times \mathscr{B}^{e}_{A}) + (0.180 \times \mathscr{B}^{e}_{B}) \\ &\quad + (0.090 \times \mathscr{B}^{\tau}_{A}) + (-0.090 \times \mathscr{B}^{\tau}_{B}), \\ \hat{B}^{\tau} &= (0.808 \times \mathscr{B}^{e}_{A}) + (-0.808 \times \mathscr{B}^{e}_{B}) \\ &\quad + (0.098 \times \mathscr{B}^{\tau}_{A}) + (0.902 \times \mathscr{B}^{\tau}_{B}). \end{aligned} \tag{45}$$

This leads to

$$\begin{aligned} \hat{B}^{e} &= (10.64 \pm 0.91)\%, \\ \hat{B}^{\tau} &= (11.14 \pm 0.94)\%, \end{aligned} \tag{46}$$

where the errors on $\hat{B}^{e}$ and $\hat{B}^{\tau}$ are significantly smaller than the 0.95% and 2.12% observed in the absence of correlations. In both cases, error reduction is caused by the large negative contribution of the third line of Eq. (29).

It is particularly interesting to note that the uncertainty on $\hat{B}^{\tau}$ is now comparable to that on $\hat{B}^{e}$, and that the correlation between the two combined estimates, computed using Eq. (5), is equal to $+94.8\%$. This indicates that the good resolution on the measurement of $(B^{e} - B^{\tau})$ performed by experiment B is fully exploited in the combination. Indeed, the difference between $\hat{B}^{e}$ and $\hat{B}^{\tau}$ in Eq. (45) is essentially dictated by that between $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$, as

$$\begin{aligned} (\hat{B}^{e} - \hat{B}^{\tau}) &= (0.012 \times \mathscr{B}^{e}_{A}) + (0.988 \times \mathscr{B}^{e}_{B}) \\ &\quad + (-0.008 \times \mathscr{B}^{\tau}_{A}) + (-0.992 \times \mathscr{B}^{\tau}_{B}). \end{aligned} \tag{47}$$

The uncertainty on this difference is only 0.29%, much smaller than the uncertainties on either of $\hat{B}^{e}$ or $\hat{B}^{\tau}$, and essentially equal to the 0.30% resolution of experiment B on $(B^{e} - B^{\tau})$. As discussed in Section 2.2, it could actually be shown that the linear unbiased estimate of $(B^{e} - B^{\tau})$ given in Eq. (47), derived from the two individual BLUE for B^{e} and B^{τ}, is that of minimum variance.

The χ^2 for the combination, 1.23 for 2 degrees of freedom, is also smaller than the 2.02 obtained in the absence of correlations. As already observed in

Section 3.2, this does not depend only on the correlation between $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$, but also on the measured central values assumed in this example. In particular, while the large positive correlation between $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$ has the effect that the contribution to the χ^2 from their difference is numerically more relevant than that from their sum, this only results in a smaller χ^2 value than that obtained in the absence of correlations because, at the same time, the agreement between the branching ratios measured by experiments A and B is much better for their difference than for their sum. As a consequence, the contributions to the χ^2 from the two sets of measurements by experiments A and B, which can be calculated separately in this case as there are no correlations between them, are both smaller than those obtained in the absence of correlations between $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$. The contribution from A is smaller because $\hat{B}^{e}$ and $\hat{B}^{\tau}$ are closer to $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{\tau}_{A}$; the contribution from B decreases because the difference $(\hat{B}^{e} - \hat{B}^{\tau})$ remains very close to $(\mathscr{B}^{e}_{B} - \mathscr{B}^{\tau}_{B})$ even if the worse resolution on $(\mathscr{B}^{e}_{B} + \mathscr{B}^{\tau}_{B})$ allows $\hat{B}^{e}$ and $\hat{B}^{\tau}$ to move away from $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$.

Not surprisingly, the effects of the large positive correlation between $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$ on the results of the lepton universality fit are quite different, and no decrease in errors or in χ^2 values is observed. The uncertainty on

$$\hat{B}' = (10.71 \pm 0.90)\%, \tag{48}$$

and the χ^2 value of 4.01 for 3 degrees of freedom are, instead, both larger than the 0.87% and 2.19 observed in the absence of any correlations. Since the four results $\mathscr{B}^{e}_{A}, \mathscr{B}^{e}_{B}, \mathscr{B}^{\tau}_{A}$ and $\mathscr{B}^{\tau}_{B}$ are treated as if they were all measurements of the same observable B', the situation is similar to that described in Section 4.2 for a $+15\%$ correlation between $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$.

4.3.2. Negative correlations

It is important to realize that the presence of correlations between the measurements of different observables, whether positive or negative, may only decrease the error on the combined estimate of each observable, with respect to the case when such correlations are absent. This is quite different from the effects of correlations between measurements of the same observable, which tend to produce opposite results when they are of opposite sign.

In the example above, let the correlation between $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$ be equal to -99.5%. Negative correlations might, for instance, arise due to the mis-identification of an electron as a τ and vice versa, although never quite as large. While the central values of the combined estimates are of course quite different, the errors on the BLUE $\hat{B}^{e}$ and $\hat{B}^{\tau}$ are found to be exactly the same as those given in Eq. (46) for the case of a $+99.5\%$ correlation:

$$\hat{B}^{e} = (11.44 \pm 0.91)\%$$
$$\hat{B}^{\tau} = (15.98 \pm 0.94)\%. \tag{49}$$

Experiment B has in this case a good resolution on the sum of $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$, rather than on their difference, which results in a very small error of 0.29% on $(\hat{B}^{e} + \hat{B}^{\tau})$ and in a correlation of -94.8% between $\hat{B}^{e}$ and $\hat{B}^{\tau}$.

Even if the errors on the BLUE are the same as those observed for a $+99.5\%$ correlation, a different trend is observed for the χ^2 for the combination, which in that case was smaller than in the absence of correlations; as previously discussed, this is due not only to the correlations but also to the assumptions about the measured central values. For 3 degrees of freedom, the χ^2 rises now sharply to 6.07 from the 2.02 obtained in the absence of any correlations. The increase is due to the larger χ^2 contributions from $\mathscr{B}^{e}_{A}$ and especially $\mathscr{B}^{\tau}_{A}$, as $\hat{B}^{e}$ and $\hat{B}^{\tau}$ are forced to move closer to $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$, far away from $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{\tau}_{A}$.

While the effects of a $+99.5\%$ or -99.5% correlation between $\mathscr{B}^{e}_{A}$ and $\mathscr{B}^{e}_{B}$ on the error on the BLUE are strictly the same when two independent estimates $\hat{B}^{e}$ and $\hat{B}^{\tau}$ are determined, these two cases are quite different for the lepton universality fit, as these values come to represent correlations between measurements of the same observables. The uncertainty on

$$\hat{B}' = (13.67 \pm 0.15)\% \tag{50}$$

is now much smaller than the 0.87% observed in the absence of any correlations, and corresponds essentially to the resolution on $(\mathscr{B}^{e}_{B} + \mathscr{B}^{\tau}_{B})/2$. The χ^2 for the fit explodes instead to 12.3 for 3 degrees

of freedom, from the 2.19 observed for no correlations: this is due to the large increases in the χ^2 contributions from $\mathscr{B}^{\tau}_{A}$ and especially $\mathscr{B}^{e}_{A}$, already observed in the case of the combination without lepton universality.

4.3.3. Breakdown of error contributions

Although not very realistic, the two cases considered in Sections 4.3.1 and 4.3.2 are good examples to show how this method makes it easy to identify, in the errors on the BLUE, the contributions of the individual uncertainties on the input measurements. One can assume for instance that experiment B has very small statistical errors, having collected large samples of data, while it is limited by large systematic errors:

$$\begin{aligned}
\mathscr{B}^{e}_{A}/\% &= 10.50 \pm 1.00(\text{stat.}),\\
\mathscr{B}^{e}_{B}/\% &= 13.50 \pm 0.21(\text{stat.}) \pm 2.99(\text{syst.}),\\
\mathscr{B}^{\tau}_{A}/\% &= 9.50 \pm 3.00(\text{stat.}),\\
\mathscr{B}^{\tau}_{B}/\% &= 14.00 \pm 0.21(\text{stat.}) \pm 2.99(\text{syst.}).
\end{aligned} \tag{51}$$

One can further assume that all statistical errors are uncorrelated, while systematic errors on $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$ are 100% positively or negatively correlated, for the two cases considered in Sections 4.3.1 and 4.3.2, respectively. The covariance matrix of Eq. (44) can then be written as the sum of two contributions from statistical and systematic errors:

$$\frac{\mathscr{M}}{10^{-4}} = \begin{pmatrix} 1.00 & 0 & 0 & 0\\ 0 & 0.04 & 0 & 0\\ 0 & 0 & 9.00 & 0\\ 0 & 0 & 0 & 0.04 \end{pmatrix}_{(\text{stat.})} + \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 8.96 & 0 & \pm 8.96\\ 0 & 0 & 0 & 0\\ 0 & \pm 8.96 & 0 & 8.96 \end{pmatrix}_{(\text{syst.})}, \tag{52}$$

where the first matrix could be split up further as the sum of the uncorrelated contributions from the statistical errors on the four experiments. For comparison, individual error contributions will also be computed assuming that the systematic errors in Eq. (51) are uncorrelated, i.e. for the situation considered in Section 4.1: in that case, the two relevant covariance matrices are those given in Eq. (52), assuming that all non-diagonal terms are 0 in the second.

Computing weights according to Eq. (8), and variances according to Eq. (18), the results previously given in Eqs. (46), (38) and (49) can then be rewritten as

$$\begin{aligned}
\hat{B}^{e}/\% &= 10.64 \pm 0.91 = 10.64 \pm 0.86(\text{stat.})\\
&\quad \pm 0.27(\text{syst.}),\\
\hat{B}^{\tau}/\% &= 11.14 \pm 0.94 = 11.14 \pm 0.90(\text{stat.})\\
&\quad \pm 0.28(\text{syst.})
\end{aligned} \tag{53}$$

if the systematic errors on $\mathscr{B}^{e}_{B}$ and $\mathscr{B}^{\tau}_{B}$ are 100% correlated,

$$\begin{aligned}
\hat{B}^{e}/\% &= 10.80 \pm 0.95 = 10.80 \pm 0.90(\text{stat.})\\
&\quad \pm 0.30(\text{syst.}),\\
\hat{B}^{\tau}/\% &= 11.75 \pm 2.12 = 11.75 \pm 1.50(\text{stat.})\\
&\quad \pm 1.50(\text{syst.})
\end{aligned} \tag{54}$$

if they are uncorrelated, and

$$\begin{aligned}
\hat{B}^{e}/\% &= 11.44 \pm 0.91 = 11.44 \pm 0.86(\text{stat.})\\
&\quad \pm 0.27(\text{syst.}),\\
\hat{B}^{\tau}/\% &= 15.98 \pm 0.94 = 15.98 \pm 0.90(\text{stat.})\\
&\quad \pm 0.28(\text{syst.})
\end{aligned} \tag{55}$$

if the correlation between them is -100%, respectively. The statistical errors on $\hat{B}^{e}$ and $\hat{B}^{\tau}$ in Eq. (53) are highly correlated as they are dominated by the statistical error on $\mathscr{B}^{e}_{A}$, which contributes with a large positive weight to the two BLUE. The same is true in Eq. (55), with the only difference that $\mathscr{B}^{e}_{A}$ contributes with large weights of opposite sign to $\hat{B}^{e}$ and $\hat{B}^{\tau}$, so that the correlation between their statistical errors is negative.

The comparison of Eq. (54) with Eqs. (53) and (55) shows that, in the absence of correlations between the measurements of different observables, the individual contributions to the errors on the BLUE tend to be larger as they must add up to larger total uncertainties. In particular, the statistical and systematic error contributions computed by neglecting these correlations, for instance using the formulas of Ref. [1] for the combination of one

physical quantity, cannot be used to estimate the corresponding error contributions in the presence of correlations, as they would generally result in overestimated values.

It is also interesting to compare these results to those obtained by assuming that only statistical errors are present in Eq. (51), while the systematic errors are all equal to 0. In this case, the combination for $\hat{B}^e$ and $\hat{B}^\tau$ yields

$$\hat{B}^e/\% = 13.37 \pm 0.21 = 13.37 \pm 0.21(\text{stat.}) \pm 0.00(\text{syst.}),$$

$$\hat{B}^\tau/\% = 13.98 \pm 0.21 = 13.98 \pm 0.21(\text{stat.}) \pm 0.00(\text{syst.}), \quad (56)$$

with a χ^2 of 10.85 for 2 degrees of freedom. The resulting statistical errors on $\hat{B}^e$ and $\hat{B}^\tau$ are much smaller than the statistical contributions to the uncertainties on $\hat{B}^e$ and $\hat{B}^\tau$ in the presence of $\pm 100\%$ correlated systematic errors, given in Eqs. (53) and (55). In other words, it is dangerous to approximate the statistical contributions to the errors on the BLUE in the presence of correlated systematic uncertainties by the "statistical errors in the absence of systematics", as this may lead to largely underestimated values for them.

As for the results of the lepton universality fits, previously given in Eqs. (48), (40) and (50), they can now be rewritten as

$$\hat{B}^\ell/\% = 10.71 \pm 0.90 = 10.71 \pm 0.86(\text{stat.}) \pm 0.27(\text{syst.}) \quad (57)$$

if the systematic errors on $\mathscr{B}_B^e$ and $\mathscr{B}_B^\tau$ are 100% correlated,

$$\hat{B}^\ell/\% = 10.96 \pm 0.87 = 10.96 \pm 0.79(\text{stat.}) \pm 0.35(\text{syst.}) \quad (58)$$

if they are uncorrelated, and

$$\hat{B}^\ell/\% = 13.67 \pm 0.15 = 13.67 \pm 0.15(\text{stat.}) \pm 0.00(\text{syst.}) \quad (59)$$

if the correlation between them is -100%, respectively. Under the assumption that systematic errors are equal to zero, the lepton universality fit yields instead

$$\hat{B}^\ell/\% = 13.67 \pm 0.15 = 13.67 \pm 0.15(\text{stat.}) \pm 0.00(\text{syst.}), \quad (60)$$

with a χ^2 of 15.04 for 3 degrees of freedom. It is not surprising that this result is essentially the same as that obtained in Eq. (59) under the assumption that the correlation between the systematic errors on $\mathscr{B}_B^e$ and $\mathscr{B}_B^\tau$ is -100%: in both situations, in fact, $\mathscr{B}_B^e$ and $\mathscr{B}_B^\tau$ contribute to the BLUE with equal weights, so that the total systematic error contribution from the two measurements is in any case equal to zero. The linear weights for $(\mathscr{B}_B^e + \mathscr{B}_B^\tau)/2$, $\mathscr{B}_A^e$ and $\mathscr{B}_A^\tau$ are thus exactly the same in Eqs. (59) and (60).

5. Conclusions

A "best linear unbiased estimate" method to combine correlated measurements of many different physical quantities has been reviewed. Various properties of this technique have been discussed in the general case and for the specific example of two fictitious experiments. It has been observed, in particular, that any correlations between the measurements of different observables, whether positive or negative, always result in decreasing the combined errors on each observable, with respect to the case where these are not taken into account.

The BLUE method is an analytical solution to the problem of minimizing the χ^2 for the combination of the measurements. As such, it provides a better means of understanding the results of the combination than a numerical approach, and a simpler means of correctly estimating the individual contributions to the errors on the combined results.

Acknowledgements

The contents of this paper largely reflect common knowledge and expertise shared by the members of the LEP ElectroWeak Working Group, many of whom have helped me with their remarks and suggestions. I especially wish to

thank Martin Grünewald, Atul Gurtu, John Holt, Klaus Mönig and Ian Tomalin for many useful discussions. I am also grateful to Günther Dissertori for several interesting comments.

References

[1] L. Lyons, D. Gibaut, P. Clifford, Nucl. Instr. and Meth. A 270 (1988) 110.
[2] J. Holt, Fermion pair production above the Z^0 resonance, JHEP Proceedings, PrHEP hep2001/115, Proceedings of the International Europhysics Conference on High Energy Physics, EPS HEP 2001, Budapest, July 2001.
[3] The ALEPH, DELPHI, L3 and OPAL Collaborations and the LEP WW Working Group, LEP W Branching Fraction Results for the 2001 Update of the PDG Review of Particle Physics, LEPEWWG/XSEC/2001–02, unpublished, available on the LEP WW Working Group WWW pages (URL: http://lepewwg.web.cern.ch/LEPEWWG/lepww/4f/PDG01/).
[4] R. Chierici, W-pair and single W production at LEP, JHEP Proceedings, PrHEP hep2001/117, Proceedings of the International Europhysics Conference on High Energy Physics, EPS HEP 2001, Budapest, July 2001.
[5] The ALEPH, DELPHI, L3 and OPAL Collaborations, Nucl. Instr. and Meth. A 378 (1996) 101.
[6] The ALEPH, DELPHI, L3 and OPAL Collaborations and the Lineshape Subgroup of the LEPEWWG, Combination procedure for the precise determination of Z boson parameters from results of the LEP experiments, CERN-EP-2000-153, hep-ex/0101027; http://arxiv.org/pdf/hep-ex/0101027.
[7] D.E. Groom et al. (The Particle Data Group), Eur. Phys. J. C 15 (2000) 1;
D.E. Groom et al. (The Particle Data Group), 2001 off-year partial update for the 2002 edition available on the PDG WWW pages (URL: http://pdg.lbl.gov/).
[8] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Section 6.6.2, Academic Press, New York, 1979.
[9] K. Mönig, private communication.
[10] F. James, M. Roos, Comput. Phys. Commun. 10 (1975) 343.