CLNS 05/1912

Simultaneous Least Squares Treatment of Statistical and Systematic Uncertainties

Werner M. Sun*

Cornell University, Ithaca, New York 14853, USA

(Dated: December 19, 2005)

Abstract

We present a least squares method for estimating parameters from measurements of event yields in the presence of background and crossfeed. We adopt a unified approach to incorporating the statistical and systematic uncertainties on the experimental measurements input to the fit. We demonstrate this method with a fit for absolute hadronic D meson branching fractions, measured in $e^+e^- \to \psi(3770) \to D\bar{D}$ transitions.

PACS numbers: 07.05.Kf; 29.85.+c
Keywords: least squares, nonlinear parameter estimation

*email: wsun@mail.lepp.cornell.edu

I. INTRODUCTION

Least squares fitting is a well-known and powerful method for combining information from a set of related experimental measurements to estimate the underlying theoretical parameters (see, for instance, Reference [1]). We discuss a specific implementation of this method for use in high-energy physics experiments, where the free parameters, denoted by the vector $\mathbf{m}$, are extracted from event yields for signal processes. Typically, these yields are subject to corrections for background, crossfeed, and efficiency. Because the sizes of these corrections depend on the values of the free parameters, we make all yield adjustments directly in the fit. Often, the uncertainties on these corrections are ignored during the fit and are propagated to the free parameters afterwards. However, if these uncertainties modify the relative weights of the measurements, then the above two-step procedure would bias both the fitted central values and the estimated uncertainties. Therefore, we build the χ^2 variable from a full description of the uncertainties, statistical and systematic, as well as their correlations, on both the yields and their corrections. Thus, the input measurements — event yields, signal efficiencies, parameters quantifying the background processes, and background efficiencies — and their uncertainties are all treated in a uniform fashion. In the χ^2 minimization, we account for the $\mathbf{m}$ dependence of the yield corrections.

II. FORMALISM

Below, we denote matrices by upper case bold letters and one-dimensional vectors by lower case bold letters. Let $\mathbf{n}$ represent a set of N event yield measurements, each for a different signal process. Each measurement may receive crossfeed contributions from other signal processes as well as backgrounds from non-signal sources. The background processes are described by $\mathbf{b}$, a vector of B estimated production yields, which can be functions of experimentally measured quantities, such as branching fractions, cross sections, and luminosities. In principle, the free parameters $\mathbf{m}$ can also appear in $\mathbf{b}$, although no additional degrees of freedom are introduced by $\mathbf{b}$. The rates at which these background processes contaminate the signal yields are given by the $N \times B$ background efficiency matrix, $\mathbf{F}$. Thus, the vector $\mathbf{s} \equiv \mathbf{n} - \mathbf{F}\mathbf{b}$ represents the background-subtracted yields.

We use an $N \times N$ signal efficiency matrix, $\mathbf{E}$, to describe simultaneously detection efficiencies (diagonal elements) and crossfeed probabilities (off-diagonal elements). The elements E_{ij} are defined to be the probabilities that an event of signal process j is reconstructed and counted in yield i. The corrected yields, denoted by $\mathbf{c}$, are obtained by acting on $\mathbf{s}$ with the inverse of $\mathbf{E}$:

$$\mathbf{c} = \mathbf{E}^{-1}\mathbf{s} = \mathbf{E}^{-1}(\mathbf{n} - \mathbf{F}\mathbf{b}). \tag{1}$$

Thus, $\mathbf{c}$ encapsulates all the experimental measurements. The variance matrix of $\mathbf{c}$, denoted by $\mathbf{V_c}$, receives contributions, both statistical and systematic, from each element of $\mathbf{n}$, $\mathbf{b}$, $\mathbf{E}$, and $\mathbf{F}$.

In the least squares fit, we define $\chi^2 \equiv (\mathbf{c} - \widetilde{\mathbf{c}})^T \mathbf{V}_{\mathbf{c}}^{-1} (\mathbf{c} - \widetilde{\mathbf{c}})$, where $\widetilde{\mathbf{c}}$ is the vector of predicted yields, which are also functions of $\mathbf{m}$. Because both $\widetilde{\mathbf{c}}$ and $\mathbf{c}$ (through $\mathbf{b}$) depend on $\mathbf{m}$, minimizing this χ^2 amounts to a nonlinear version of the total least squares method [2]. We solve this problem by extending the conventional least squares fit to include contributions from both $\widetilde{\mathbf{c}}$ and $\mathbf{c}$ in $\partial\chi^2/\partial\mathbf{m}$. Given a set of seed values, $\mathbf{m}_0$, the optimized estimate, $\widehat{\mathbf{m}}$, and its variance matrix, $\mathbf{V}_{\mathbf{m}}$, are

$$\widehat{\mathbf{m}} = \mathbf{m}_0 + \left(\mathbf{D}\mathbf{V}_{\mathbf{c}}^{-1}\mathbf{D}^T\right)^{-1}\mathbf{D}\mathbf{V}_{\mathbf{c}}^{-1}\left[\mathbf{c}(\mathbf{m}_0) - \widetilde{\mathbf{c}}(\mathbf{m}_0)\right] \tag{2}$$

$$\mathbf{V}_{\mathbf{m}} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial\mathbf{m}\,\partial\mathbf{m}^T} = \left(\mathbf{D}\mathbf{V}_{\mathbf{c}}^{-1}\mathbf{D}^T\right)^{-1}, \tag{3}$$

where the $M \times N$ derivative matrix $\mathbf{D}$ is defined to be

$$\mathbf{D} \equiv \frac{\partial\widetilde{\mathbf{c}}}{\partial\mathbf{m}} - \frac{\partial\mathbf{c}}{\partial\mathbf{m}} = \frac{\partial\widetilde{\mathbf{c}}}{\partial\mathbf{m}} + \frac{\partial\mathbf{b}}{\partial\mathbf{m}}\mathbf{F}^T\left(\mathbf{E}^{-1}\right)^T. \tag{4}$$

In general, $\widetilde{\mathbf{c}}$ and $\mathbf{c}$ are nonlinear functions of $\mathbf{m}$, so the linearized solution $\widehat{\mathbf{m}}$ is approximate, and the above procedure is iterated until the χ^2 converges. Between iterations, all the fit inputs that depend on $\mathbf{m}$ are reevaluated with the updated values of $\widehat{\mathbf{m}}$.

Nonlinearities also occur when $\mathbf{V}_{\mathbf{c}}$ contains multiplicative or Poisson uncertainties that depend on the measurement values. With the least squares method, these nonlinearities result in biased estimators unless these variable uncertainties are evaluated using the predicted yields $\widetilde{\mathbf{c}}$ instead of the measured $\mathbf{c}$. Therefore, all three ingredients in the χ^2 — $\mathbf{c}$, $\widetilde{\mathbf{c}}$, and $\mathbf{V}_{\mathbf{c}}$ — are functions of $\mathbf{m}$. However, we do not include the derivatives $\partial\mathbf{V}_{\mathbf{c}}/\partial\mathbf{m}$ in $\mathbf{D}$ because doing so would generate biases in $\widehat{\mathbf{m}}$.

For a simple demonstration of the aforementioned biases, we consider two measured yields, c_1 and c_2, which are both estimators of a true yield $\bar{c}$. We assume that the uncertainties on c_1 and c_2 are uncorrelated, multiplicative, and of the same fractional size, λ. We construct an improved estimator, $\widehat{c}$, by minimizing $\chi^2 = (c_1 - c)^2/\sigma_{c_1}^2 + (c_2 - c)^2/\sigma_{c_2}^2$ with respect to c. If, following the prescription given above, we neglect the $\partial\sigma_{c_i}^2/\partial c$ terms in $\partial\chi^2/\partial c$ and assign (iteratively) the uncertainties $\sigma_{c_1} = \sigma_{c_2} = \lambda\widehat{c}$, then c_1 and c_2 are equally weighted, and $\widehat{c}$ is an unbiased estimate of $\bar{c}$:

$$\widehat{c}_{\text{unbiased}} = \frac{c_1 + c_2}{2} \tag{5}$$

$$\chi^2_{\text{unbiased}} = \frac{2}{\lambda^2}\left(\frac{c_1 - c_2}{c_1 + c_2}\right)^2. \tag{6}$$

On the other hand, including the $\partial\sigma_{c_i}^2/\partial c$ terms in $\partial\chi^2/\partial c$ results in an upward bias:

$$\widehat{c}_{\text{biased1}} = \frac{c_1^2 + c_2^2}{c_1 + c_2} = \widehat{c}_{\text{unbiased}}\left(1 + \frac{\lambda^2\chi^2_{\text{unbiased}}}{2}\right) \tag{7}$$

$$\chi^2_{\text{biased1}} = \frac{(c_1 - c_2)^2}{\lambda^2(c_1^2 + c_2^2)}. \tag{8}$$

Finally, if we assign uncertainties based on the measured yields, not the predicted yields, such that $\sigma_{c_1} = \lambda c_1$, $\sigma_{c_2} = \lambda c_2$, and $\partial\sigma_{c_i}^2/\partial c = 0$, then the resulting estimate is biased low:

$$\widehat{c}_{\text{biased2}} = \frac{c_1 c_2(c_1 + c_2)}{c_1^2 + c_2^2} = \widehat{c}_{\text{unbiased}}(1 - \lambda^2\chi^2_{\text{biased1}}) \tag{9}$$

$$\chi^2_{\text{biased2}} = \chi^2_{\text{biased1}}. \tag{10}$$

Thus, even though χ^2_{biased1} and χ^2_{biased2} are smaller than χ^2_{unbiased}, the corresponding estimators possess undesired properties.

III. INPUT VARIANCE MATRIX

The uncertainties on the N elements of $\mathbf{n}$ and the B elements of $\mathbf{b}$ are characterized by the $N \times N$ matrix $\mathbf{V_n}$ and the $B \times B$ matrix $\mathbf{V_b}$, respectively. Usually, the elements of $\mathbf{E}$ and $\mathbf{F}$ share many common correlated systematic uncertainties, so we construct a joint variance matrix from the submatrices $\mathbf{V_E}$, $\mathbf{V_F}$, and $\mathbf{C_{EF}}$, where $\mathbf{V_E}$ ($N^2 \times N^2$) and $\mathbf{V_F}$ ($NB \times NB$) are the variance matrices for the elements of $\mathbf{E}$ and $\mathbf{F}$, respectively, and $\mathbf{C_{EF}}$ ($N^2 \times NB$) contains the correlations between $\mathbf{E}$ and $\mathbf{F}$. Below, we label each element of $\mathbf{E}$ or $\mathbf{F}$ by two indices (E_{ij} or F_{ij}), and the two dimensions of $\mathbf{E}$ or $\mathbf{F}$ are mapped onto one dimension of $\mathbf{V_E}$ or $\mathbf{V_F}$.

We form $\mathbf{V_c}$ by propagating the statistical and systematic uncertainties on $\mathbf{n}$, $\mathbf{b}$, $\mathbf{E}$, and $\mathbf{F}$ to $\mathbf{c}$ via

$$\mathbf{V_c} = \frac{\partial \mathbf{c}}{\partial \mathbf{n}}^T \mathbf{V_n} \frac{\partial \mathbf{c}}{\partial \mathbf{n}} + \frac{\partial \mathbf{c}}{\partial \mathbf{b}}^T \mathbf{V_b} \frac{\partial \mathbf{c}}{\partial \mathbf{b}} + \begin{pmatrix} (\partial \mathbf{c}/\partial \mathbf{E})^T & (\partial \mathbf{c}/\partial \mathbf{F})^T \end{pmatrix} \begin{pmatrix} \mathbf{V_E} & \mathbf{C_{EF}} \\ \mathbf{C_{EF}^T} & \mathbf{V_F} \end{pmatrix} \begin{pmatrix} \partial \mathbf{c}/\partial \mathbf{E} \\ \partial \mathbf{c}/\partial \mathbf{F} \end{pmatrix}. \quad (11)$$

Where appropriate, we substitute $\widetilde{\mathbf{c}}$ for $\mathbf{c}$, as discussed in Section II. The first term of Equation 11 is simply $\mathbf{E}^{-1}\mathbf{V_n}(\mathbf{E}^{-1})^T$, and the second term is $\mathbf{E}^{-1}\mathbf{F}\mathbf{V_b}\mathbf{F}^T(\mathbf{E}^{-1})^T$. For the third term, we evaluate the partial derivatives and find

$$\frac{\partial \mathbf{c}}{\partial \mathbf{E}} = \mathbf{s}^T \left(\frac{\partial \mathbf{E}^{-1}}{\partial \mathbf{E}}\right)^T = -\mathbf{s}^T \left(\mathbf{E}^{-1}\right)^T \left(\frac{\partial \mathbf{E}}{\partial \mathbf{E}}\right)^T \left(\mathbf{E}^{-1}\right)^T = -\mathbf{A}\left(\mathbf{E}^{-1}\right)^T \quad (12)$$

$$\frac{\partial \mathbf{c}}{\partial \mathbf{F}} = -\mathbf{B}\left(\mathbf{E}^{-1}\right)^T, \quad (13)$$

where $\mathbf{A} \equiv \mathbf{c}^T(\partial \mathbf{E}/\partial \mathbf{E})^T$ and $\mathbf{B} \equiv \mathbf{b}^T(\partial \mathbf{F}/\partial \mathbf{F})^T$, with elements given in terms of the Kronecker delta (δ_{ij}): $\partial E_{kl}/\partial E_{ij} = \partial F_{kl}/\partial F_{ij} = \delta_{ik}\delta_{jl}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ have rows labeled by two indices, which refer to the elements of $\mathbf{E}$ and $\mathbf{F}$, respectively, and columns labeled by one index, which refers to the elements of $\mathbf{c}$. In other words, the ij-th row of $\mathbf{A}$ is given by $\mathbf{c}^T(\partial \mathbf{E}/\partial E_{ij})^T$, where $(\partial \mathbf{E}/\partial E_{ij})_{kl} = \partial E_{kl}/\partial E_{ij}$. Therefore, the elements of $\mathbf{A}$ and $\mathbf{B}$ are $A_{ij,k} = \delta_{ik}\widetilde{c}_j$ and $B_{ij,k} = \delta_{ik} b_j$. For $N = B = 2$, these matrices are

$$\mathbf{A} = \begin{pmatrix} \widetilde{c}_1 & 0 \\ \widetilde{c}_2 & 0 \\ 0 & \widetilde{c}_1 \\ 0 & \widetilde{c}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_1 & 0 \\ b_2 & 0 \\ 0 & b_1 \\ 0 & b_2 \end{pmatrix}. \quad (14)$$

This treatment of error propagation in matrix inversion agrees with that derived in Reference [3]. The above relations allow us to reexpress $\mathbf{V_c}$ as

$$\mathbf{V_c} = \mathbf{E}^{-1}\mathbf{V_{\Delta n}}\left(\mathbf{E}^{-1}\right)^T, \quad (15)$$

where $\mathbf{V_{\Delta n}} \equiv \mathbf{V_n} + \mathbf{F}\mathbf{V_b}\mathbf{F}^T + \mathbf{A}^T\mathbf{V_E}\mathbf{A} + \mathbf{B}^T\mathbf{V_F}\mathbf{B} + \mathbf{A}^T\mathbf{C_{EF}}\mathbf{B} + \mathbf{B}^T\mathbf{C_{EF}^T}\mathbf{A}$. As a result, we have $\chi^2 = \mathbf{\Delta n}^T\mathbf{V_{\Delta n}^{-1}}\mathbf{\Delta n}$, where $\mathbf{\Delta n} \equiv \mathbf{n} - \mathbf{E}\widetilde{\mathbf{c}} - \mathbf{F}\mathbf{b}$. Thus, the χ^2 minimization can be formulated equivalently in terms of $\mathbf{n}$ instead of $\mathbf{c}$: $\mathbf{V_m} = \left(\mathbf{D}'\mathbf{V_{\Delta n}^{-1}}\mathbf{D}'^T\right)^{-1}$ and $\widehat{\mathbf{m}} = \mathbf{m}_0 + \mathbf{V_m}\mathbf{D}'\mathbf{V_{\Delta n}^{-1}}\mathbf{\Delta n}$, where $\mathbf{D}' \equiv \mathbf{D}\mathbf{E}^T = (\partial\widetilde{\mathbf{c}}/\partial\mathbf{m})\mathbf{E}^T + (\partial\mathbf{b}/\partial\mathbf{m})\mathbf{F}^T$.

Systematic uncertainties on the efficiencies are often multiplicative and belong to one of three categories: those that depend only on the reconstructed mode (row-wise), those that depend only on the generated mode (column-wise), and those that are uncorrelated among elements of $\mathbf{E}$ and $\mathbf{F}$. For row-wise efficiency uncertainties, all the elements in any given row of $\mathbf{E}$ and $\mathbf{F}$ have the same fractional uncertainty, which we denote by $\lambda_i \equiv \sigma_{E_{ij}}/E_{ij} = \sigma_{F_{ij}}/F_{ij}$. The correlation coefficients between elements of different rows are $\lambda_{ij}^2/(\lambda_i\lambda_j)$, where λ_{ij} characterizes the uncertainties common to c_i and c_j. For instance, if λ_{track} is the fractional uncertainty associated with the charged particle tracking efficiency, then $\lambda_i = t_i\lambda_{\text{track}}$ and $\lambda_{ij}^2 = t_i t_j \lambda_{\text{track}}^2$, where t_i and t_j are the track multiplicities in modes i and j, respectively. Note that $\lambda_{ii} = \lambda_i$. Similarly, for column-wise uncertainties, we define the fractional uncertainties $\mu_j \equiv \sigma_{E_{ij}}/E_{ij} = \sigma_{F_{ij}}/F_{ij}$ and correlation coefficients $\mu_{ij}^2/(\mu_i\mu_j)$. We denote the uncorrelated fractional uncertainty on any element of $\mathbf{E}$ or $\mathbf{F}$ by $\nu_{ik,jl}$. Table I gives expressions for the elements of $\mathbf{V_E}$, $\mathbf{V_F}$, and $\mathbf{C_{EF}}$, as well as their contributions to $\mathbf{V_c}$ for row-wise, column-wise, and uncorrelated uncertainties.

TABLE I: Expressions for the elements of $\mathbf{V_E}$, $\mathbf{V_F}$, and $\mathbf{C_{EF}}$, as well as their contributions to $\mathbf{V_c}$. Repeated external indices are not summed over.

Quantity	Row-wise	Column-wise	Uncorrelated
$(\mathbf{V_E})_{ik,jl}$	$\lambda_{ij}^2 E_{ik}E_{jl}$	$\mu_{kl}^2 E_{ik}E_{jl}$	$\nu_{ik,jl}^2 E_{ik}E_{jl}\delta_{ij}\delta_{kl}$
$(\mathbf{V_F})_{ik,jl}$	$\lambda_{ij}^2 F_{ik}F_{jl}$	$\mu_{kl}^2 F_{ik}F_{jl}$	$\nu_{ik,jl}^2 F_{ik}F_{jl}\delta_{ij}\delta_{kl}$
$(\mathbf{C_{EF}})_{ik,jl}$	$\lambda_{ij}^2 E_{ik}F_{jl}$	$\mu_{kl}^2 E_{ik}F_{jl}$	0
$(\mathbf{A}^T\mathbf{V_E}\mathbf{A})_{ij}$	$\lambda_{ij}^2\widetilde{s}_i\widetilde{s}_j$	$\mu_{kl}^2 E_{ik}\widetilde{c}_k E_{jl}\widetilde{c}_l$	$\delta_{ij}\sigma_{E_{jk}}^2\widetilde{c}_k^2$
$(\mathbf{B}^T\mathbf{V_F}\mathbf{B})_{ij}$	$\lambda_{ij}^2 F_{ik}b_k F_{jl}b_l$	$\mu_{kl}^2 F_{ik}b_k F_{jl}b_l$	$\delta_{ij}\sigma_{F_{jk}}^2 b_k^2$
$(\mathbf{A}^T\mathbf{C_{EF}}\mathbf{B})_{ij}$	$\lambda_{ij}^2\widetilde{s}_i F_{jk}b_k$	$\mu_{kl}^2 E_{ik}\widetilde{c}_k F_{jl}b_l$	0

IV. EXAMPLE: HADRONIC D MESON BRANCHING FRACTIONS

The least squares method described in the previous sections has been employed by the CLEO-c collaboration [4] to measure absolute branching fractions for hadronic D meson decays. Using $D\bar{D}$ pairs produced through the $\psi(3770)$ resonance, the branching fraction for mode i, denoted by $\mathcal{B}_i$, is measured by comparing the number of events where a single $D \to i$ decay is reconstructed (called single tag, denoted by x_i) with the number of events where both D and $\bar{D}$ are reconstructed via $D \to i$ and $\bar{D} \to j$ (called double tag, denoted by y_{ij}). These yield measurements form the vector $\mathbf{n}$. The free parameters $\mathbf{m}$ are the $\mathcal{B}_i$ and the numbers of $D^0\bar{D}^0$ and D^+D^- pairs produced, denoted by $\mathcal{N}^{00}$ and $\mathcal{N}^{+-}$, respectively, and denoted generically by $\mathcal{N}$. Yields for charge conjugate modes are measured separately, so the predicted corrected yields $\widetilde{\mathbf{c}}$ are $\mathcal{N}\mathcal{B}_i$ for single tags and $\mathcal{N}\mathcal{B}_i\mathcal{B}_j$ for double tags. Thus, $\mathcal{B}_i$ and $\mathcal{N}$ can be extracted from various products and ratios of x_i, x_j, and y_{ij}: $\mathcal{B}_i \sim y_{ij}/x_j$, $\mathcal{N} \sim x_i x_j/y_{ij}$, up to corrections for efficiency, crossfeed, and background.

The matrix $\mathbf{V_n}$ describes the statistical uncertainties and correlations among the x_i and y_{ij}. The y_{ij} are uncorrelated, but because any given event can contain both single tag and double tag candidates, the x_i are correlated among themselves as well as with the y_{ij}. If the selection criteria for single and double tags are the same, then the events (signal and background) used to estimate y_{ij} are a proper subset of those for x_i and x_j. Thus, any single tag yield is a sum of exclusive single tags (x_i^{excl}) and double tags: $x_{\{i,j\}} = x_{\{i,j\}}^{\text{excl}} + y_{ij}$.

Propagating the uncertainties on the independent variables, $x_i^{\rm excl}$, $x_j^{\rm excl}$, and y_{ij}, gives the following elements for $\mathbf{V_n}$:

$$\begin{aligned}
\langle \Delta x_i \Delta x_j \rangle &= \delta_{ij}\sigma_{x_i}\sigma_{x_j} + (1-\delta_{ij})\sigma^2_{y_{ij}} && (16)\\
\langle \Delta y_{ij} \Delta y_{kl} \rangle &= \delta_{ik}\delta_{jl}\sigma_{y_{ij}}\sigma_{y_{kl}} && (17)\\
\langle \Delta x_i \Delta y_{jk} \rangle &= (\delta_{ij}+\delta_{ik})\sigma^2_{y_{jk}}, && (18)
\end{aligned}$$

where $\Delta x_i \equiv x_i - \langle x_i \rangle$, $\Delta y_{ij} \equiv y_{ij} - \langle y_{ij} \rangle$, and $\sigma^2_{x_{\{i,j\}}} = \sigma^2_{x^{\rm excl}_{\{i,j\}}} + \sigma^2_{y_{ij}}$. Thus, for any two single tag yields and the corresponding double tag yield, the three off-diagonal elements of $\mathbf{V_n}$ are all given by the uncertainty on the number of overlapping events. In addition to these statistical uncertainties, $\mathbf{V_n}$ can also receive contributions from additive systematic uncertainties.

Some of the sources of background we consider are non-signal D decays, $e^+e^- \to q\bar{q}$ events, and $e^+e^- \to \tau^+\tau^-$ events. If there are two non-signal D backgrounds with branching fractions $\mathcal{C}_1$ and $\mathcal{C}_2$, then the vector $\mathbf{b}$ is given by

$$\mathbf{b} = \begin{pmatrix} \mathcal{N}\mathcal{C}_1 \\ \mathcal{N}\mathcal{C}_2 \\ \mathcal{L}X_{q\bar{q}} \\ \mathcal{L}X_{\tau^+\tau^-} \end{pmatrix}, \qquad (19)$$

where $X_{q\bar{q}}$ and $X_{\tau^+\tau^-}$ are the cross sections for $q\bar{q}$ and $\tau^+\tau^-$ production, respectively, and $\mathcal{L}$ is the integrated luminosity of the data sample. Because of the non-signal D decays, the free parameter $\mathcal{N}$ appears in $\mathbf{b}$ but does not contribute any additional terms to the variance matrix $\mathbf{V_b}$, which takes the following block diagonal form:

$$\mathbf{V_b} = \begin{pmatrix} \mathcal{N}^2\sigma^2_{\mathcal{C}_1} & 0 & 0 & 0 \\ 0 & \mathcal{N}^2\sigma^2_{\mathcal{C}_2} & 0 & 0 \\ 0 & 0 & \mathcal{L}^2\sigma^2_{X_{q\bar{q}}} + X^2_{q\bar{q}}\sigma^2_{\mathcal{L}} & X_{q\bar{q}}X_{\tau^+\tau^-}\sigma^2_{\mathcal{L}} \\ 0 & 0 & X_{q\bar{q}}X_{\tau^+\tau^-}\sigma^2_{\mathcal{L}} & \mathcal{L}^2\sigma^2_{X_{\tau^+\tau^-}} + X^2_{\tau^+\tau^-}\sigma^2_{\mathcal{L}} \end{pmatrix}. \qquad (20)$$

Also, the matrix $\partial\mathbf{b}/\partial\mathbf{m}$ is nontrivial and is incorporated into the χ^2 minimization.

In the joint variance matrix for $\mathbf{E}$ and $\mathbf{F}$, uncertainties of all three types discussed in Section III are present. Row-wise effects arise from systematic uncertainties on simulated reconstruction efficiencies for charged tracks, $\pi^0 \to \gamma\gamma$ decays, $K^0_S \to \pi^+\pi^-$ decays, and particle identification (PID) for charged pions and kaons. Column-wise uncertainties reflect the poorly known resonant substructure in multi-body final states. Uncorrelated contributions come from statistical uncertainties due to the finite Monte Carlo (MC) simulated samples used to determine $\mathbf{E}$ and $\mathbf{F}$. Thus, for example, if mode i is $D^0 \to K^-\pi^+\pi^0$ and mode j is $D^+ \to K^0_S\pi^+$, then the row-wise uncertainties are given by

$$\begin{aligned}
\lambda_i^2 &= (2\lambda_{\rm track})^2 + \lambda^2_{\pi^0} + \lambda^2_{\pi^\pm{\rm PID}} + \lambda^2_{K^\pm{\rm PID}} && (21)\\
\lambda_j^2 &= (3\lambda_{\rm track})^2 + \lambda^2_{\pi^\pm{\rm PID}} && (22)\\
\lambda_{ij}^2 &= 6\lambda^2_{\rm track} + \lambda^2_{\pi^\pm{\rm PID}}. && (23)
\end{aligned}$$

Because these row-wise and column-wise uncertainties are completely correlated among the yields to which they pertain, they degrade the precision of $\mathcal{B}_i$ but not $\mathcal{N}$. Furthermore, they have no effect on the central values of $\widehat{\mathbf{m}}$ because the relative weight of each yield is unaltered by these uncertainties. However, they can introduce large systematic correlations among the fit parameters, even between statistically independent branching fractions of different charge.

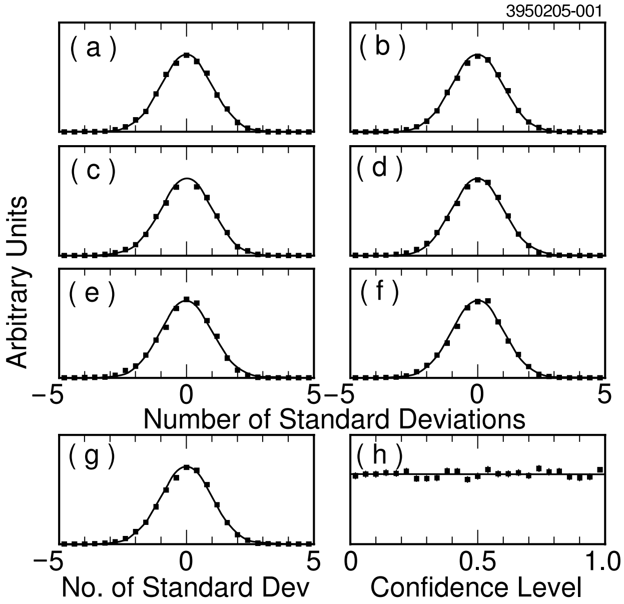

FIG. 1: Toy MC fit pull distributions for $\mathcal{N}^{00}$ (a), $\mathcal{B}(D^0 \to K^-\pi^+)$ (b), $\mathcal{B}(D^0 \to K^-\pi^+\pi^0)$ (c), $\mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+)$ (d), $\mathcal{N}^{+-}$ (e), $\mathcal{B}(D^+ \to K^-\pi^+\pi^+)$ (f), and $\mathcal{B}(D^+ \to K^0_S\pi^+)$ (g), overlaid with Gaussian curves with zero mean and unit width. The fit confidence level distribution (h) is overlaid with a line with zero slope.

A. Toy Monte Carlo Study

We test the method presented above using a toy MC simulation with Gaussian smearing of the fit inputs. We generate data for five decay modes, $D^0 \to K^-\pi^+$, $D^0 \to K^-\pi^+\pi^0$, $D^0 \to K^-\pi^+\pi^-\pi^+$, $D^+ \to K^-\pi^+\pi^+$, and $D^+ \to K^0_S\pi^+$ (charge conjugate particles are implied), for which there are ten single tag and thirteen double tag yields. The fit determines seven free parameters: $\mathcal{N}^{00}$, $\mathcal{N}^{+-}$, and five charge-averaged branching fractions. The input branching fractions are taken to be the world-average values given in Reference [1], and we use $\mathcal{N}^{00} = 2.0 \times 10^5$ and $\mathcal{N}^{+-} = 1.5 \times 10^5$. The efficiencies are mode-dependent: 30%–70% for single tags and 10%–50% for double tags, with fractional statistical uncertainties of 0.5%–1.0%. The yield uncertainties are specified to be close to the Poisson limit, and backgrounds correspond roughly to those expected in 60 pb^{-1} of e^+e^- collisions at the $\psi(3770)$. Also, we apply correlated systematic efficiency uncertainties of 1% for tracking, 2% for π^0 reconstruction, 2% for K^0_S reconstruction, and 1% for charged pion and kaon PID.

The fit reproduces the input parameters well. Figure 1 shows the pull distributions for the seven fit parameters and the fit confidence level for 10000 toy MC trials. All the pull distributions are unbiased and have widths consistent with unity. Also, the confidence level is flat. Table II gives the correlation coefficients among the fit parameters. Branching fractions tend to be positively correlated with each other and negatively correlated with $\mathcal{N}^{00}$ and $\mathcal{N}^{+-}$. In particular, the D^0 branching fractions are correlated with those for D^+. In the absence of correlated efficiency uncertainties, the D^0 and D^+ free parameters would essentially be independent.

Slight asymmetries can be observed in the pull distributions, especially in those for $\mathcal{N}^{00}$ and $\mathcal{N}^{+-}$. These asymmetries are caused by the nonlinear nature of the multiplicative

TABLE II: Correlation coefficients, including systematic uncertainties, for the free parameters determined by the fit to toy MC samples.

	$\mathcal{N}^{00}$	$K^-\pi^+$	$K^-\pi^+\pi^0$	$K^-\pi^+\pi^-\pi^+$	$\mathcal{N}^{+-}$	$K^-\pi^+\pi^+$	$K^0_S\pi^+$
$\mathcal{N}^{00}$	1	-0.63	-0.52	-0.38	-0.01	-0.01	-0.01
$K^-\pi^+$		1	0.79	0.87	-0.01	0.40	0.29
$K^-\pi^+\pi^0$			1	0.77	-0.01	0.37	0.27
$K^-\pi^+\pi^-\pi^+$				1	-0.01	0.53	0.39
$\mathcal{N}^{+-}$					1	-0.82	-0.77
$K^-\pi^+\pi^+$						1	0.87
$K^0_S\pi^+$							1

efficiency uncertainties and of the functions $\tilde{\mathbf{c}}(\mathbf{m})$. Because the fit parameters are effectively estimated from ratios of the input yields, Gaussian fluctuations in the denominators produce non-Gaussian fluctuations in the ratios, which are most visible in $\mathcal{N}^{00}$ and $\mathcal{N}^{+-}$, where the uncertainties in the denominators are dominant. Similarly, multiplicative uncertainties, which affect only the branching fractions, scale with the fitted values and, therefore, give rise to asymmetric $\mathcal{B}$ pulls. In both cases, larger fractional uncertainties would heighten the asymmetries.

If we form the matrix $\mathbf{A}$ in Equation 14 using the measured yields $\mathbf{c}$ rather than the predicted yields $\tilde{\mathbf{c}}$, then the variance matrix $\mathbf{V_c}$ need not be reevaluated after each fit iteration. However, in this case, the pull distributions become significantly biased, as shown in Figure 2. Thus, obtaining unbiased fit results and the correct uncertainties requires proper handling of the efficiency variance matrices $\mathbf{V_E}$ and $\mathbf{V_F}$.

V. SUMMARY

We have developed a least squares fit that simultaneously incorporates statistical and systematic uncertainties, as well as their correlations, on all the input experimental measurements. Biases from nonlinearities are reduced by introducing fit parameter dependence in the input variance matrix. This fitting method is used to measure absolute branching fractions of hadronic D meson decays, and toy Monte Carlo studies validate the performance of the fitter. By including all known sources of measurement uncertainty in the χ^2, we obtain unbiased fit parameters with correct estimated uncertainties.

Acknowledgments

We wish to thank Roy Briere, David Cassel, Lawrence Gibbons, Wolfgang Rolke, Anders Ryd, and Ian Shipsey for many helpful discussions. This work was supported in part by the

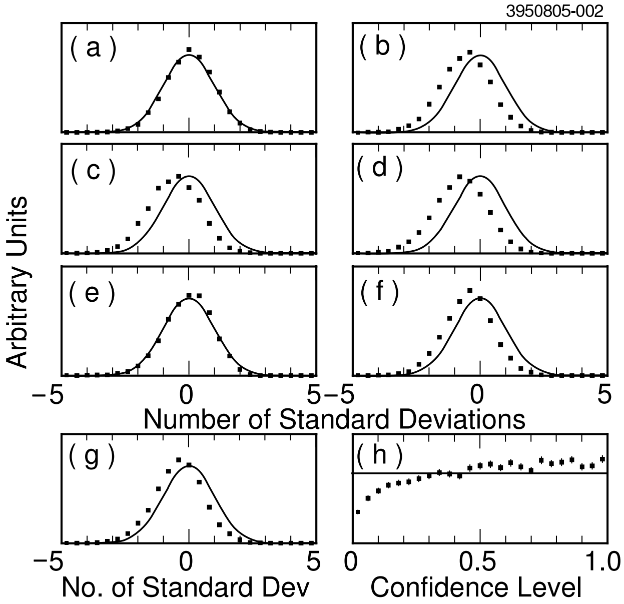

FIG. 2: Toy MC fit pull distributions, with $\mathbf{V_c}$ calculated using $\mathbf{c}$ instead of $\tilde{\mathbf{c}}$, for $\mathcal{N}^{00}$ (a), $\mathcal{B}(D^0 \to K^-\pi^+)$ (b), $\mathcal{B}(D^0 \to K^-\pi^+\pi^0)$ (c), $\mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+)$ (d), $\mathcal{N}^{+-}$ (e), $\mathcal{B}(D^+ \to K^-\pi^+\pi^+)$ (f), and $\mathcal{B}(D^+ \to K^0_S\pi^+)$ (g), overlaid with Gaussian curves with zero mean and unit width. The fit confidence level distribution (h) is overlaid with a line with zero slope.

National Science Foundation under Grant No. PHY-0202078.

[1] Particle Data Group, S. Eidelman *et al.*, Phys. Lett. B **592**, 1 (2004).
[2] G. H. Golub and C. F. Van Loan, SIAM J. Numer. Anal. **17** 883–893 (1980).
[3] M. Lefebvre, R. K. Keeler, R. Sobie and J. White, Nucl. Instrum. Meth. A **451**, 520 (2000) [arXiv:hep-ex/9909031].
[4] B. I. Eisenstein *et al.*, CLEO CONF 04-10, ICHEP04 ABS11-0775, conference submission presented at the 32$^{\text{nd}}$ International Conference on High Energy Physics, Aug. 16-22, 2004, Beijing, China.