Profiling and Covariance Matrices

Say we are measuring 2 quantities B_1 and B_2. The measurement also has an uncertainty parameterized by a nuisance parameter θ, so what we actually measure is

$$B_1' = B_1 - \Delta_1\theta \tag{1}$$
$$B_2' = B_2 - \Delta_2\theta. \tag{2}$$

As usual, we assume an auxiliary measurement constrains θ by a normal distribution, so that the Δ_i correspond to $\pm 1\sigma$ uncertainties on the B_i. We also assume that the true values of the B_i are both 0, and write the observed values as $\hat{B}_1'$, $\hat{B}_2'$ and $\hat{\theta}$.

The probability distribution of the measurement process is then

$$P(\hat{B}_1', \hat{B}_2', \hat{\theta}) = K \exp\left(-\frac{1}{2}\hat{X}'^T C'^{-1} \hat{X}'\right) \tag{3}$$

where

$$\hat{X}' = \begin{pmatrix} \hat{B}_1' \\ \hat{B}_2' \\ \hat{\theta} \end{pmatrix} \tag{4}$$

and C' is the covariance matrix of the measurement,

$$C' = \left(\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{C_B}} & 0 \\ & & 0 \\ \hline 0 & 0 & \kappa^2 \end{array}\right) \tag{5}$$

where $C_B \equiv \langle \hat{B}_1' \ \hat{B}_2' \rangle$, and $\langle \hat{\theta}^2 \rangle = \kappa^2$. The θ auxiliary measurement is assumed to be independent of the B_i' measurement, so the matrix is block-diagonal. The auxiliary measurement alone would give $\langle \hat{\theta}^2 \rangle = 1$ by definition, but $\langle \hat{\theta}^2 \rangle$ can be lower if the main measurement constrains θ, leading to $\kappa < 1$ (but we assume that this still does not lead to correlation between θ and the B_i').

If we now represent the measurement in terms of the B_i which we want to measure, instead of the B_i', we have

$$P(\hat{B}_1, \hat{B}_2, \hat{\theta}) = K \exp\left(-\frac{1}{2}\hat{X}^T C^{-1} \hat{X}\right) \tag{6}$$

where

$$\hat{X} = \begin{pmatrix} \hat{B}_1 \\ \hat{B}_2 \\ \hat{\theta} \end{pmatrix} \tag{7}$$

and C is given by

$$C = \left(\begin{array}{c|c} C_B + \kappa^2 \Delta\Delta^T & \kappa^2\Delta \\ \hline \kappa^2\Delta^T & \kappa^2 \end{array}\right) \tag{8}$$

where $\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}$. This is since $\langle B_i\theta\rangle = \langle (B'_i + \Delta_i\theta)\theta\rangle = \kappa^2\Delta_i$ (again since θ is uncorrelated with the B'_i) and $\langle B_iB_j\rangle = \langle (B'_i + \Delta_i\theta)(B'_j + \Delta_j\theta)\rangle = C_{B,ij} + \kappa^2\Delta_i\Delta_j = (C_B + \kappa^2\Delta\Delta^T)_{ij}$.

Since we actually need the Hessian $H = C^{-1}$, we use the usual formula for block inversion,

$$\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) = \left(\begin{array}{c|c} \tilde{A}^{-1} & -\tilde{A}^{-1}BD^{-1} \\ \hline -D^{-1}C\tilde{A}^{-1} & D^{-1} + D^{-1}C\tilde{A}^{-1}BD^{-1} \end{array}\right) \tag{9}$$

where $\tilde{A} = A - BD^{-1}C$ is the Schur complement of D. So in our case,

$$H \equiv C^{-1} = \left(\begin{array}{c|c} C_B^{-1} & -C_B^{-1}\Delta \\ \hline -\Delta^T C_B^{-1} & \kappa^{-2} + \Delta^T C_B^{-1}\Delta \end{array}\right) = \left(\begin{array}{c|c} C_B^{-1} & -\Lambda \\ \hline -\Lambda^T & \kappa^{-2} + \Delta^T\Lambda \end{array}\right) \tag{10}$$

with $\Lambda = C_B^{-1}\Delta$.

To profile θ, we need to integrate it out of the probability distribution, which is given by

$$P(\hat{B}_1, \hat{B}_2, \hat{\theta}) = K \exp\left[-\frac{1}{2}\left(\hat{B}^T C_B^{-1}\hat{B} - 2\Lambda^T\hat{B}\hat{\theta} + (1 + \Delta^T\Lambda)\hat{\theta}^2\right)\right]. \tag{11}$$

We can rewrite it as

$$P(\hat{B}_1, \hat{B}_2, \hat{\theta}) = K \exp\left\{-\frac{1}{2}\left[\hat{B}^T\left(C_B^{-1} - \frac{1}{\kappa^{-2} + \Delta^T\Lambda}\Lambda\Lambda^T\right)\hat{B} + (\kappa^{-2} + \Delta^T\Lambda)\left(\hat{\theta} - \frac{\Lambda^T\hat{B}}{\kappa^{-2} + \Delta^T\Lambda}\right)^2\right]\right\}. \tag{12}$$

so that

$$P(\hat{B}_1, \hat{B}_2) = \int P(\hat{B}_1, \hat{B}_2, \hat{\theta})d\hat{\theta} = K \exp\left[-\frac{1}{2}\hat{B}^T\left(C_B^{-1} - \frac{1}{\kappa^{-2} + \Delta^T\Lambda}\Lambda\Lambda^T\right)\hat{B}\right]. \tag{13}$$

So ultimately, after profiling θ we get a covariance matrix for the B measurement equal to

$$C_B^{prof} = \left(C_B^{-1} - \frac{1}{\kappa^{-2} + \Delta^T\Lambda}\Lambda\Lambda^T\right)^{-1} = C_B(1 + \kappa^2\Delta^T\Lambda)\left[(1 + \kappa^2\Delta^T\Lambda)I - \kappa^2\Lambda\Delta^T\right]^{-1} \tag{14}$$

In the case of small systematics (relative to stat uncertainties), $\Delta \ll C_B$, we have

$$C_B^{prof} \approx C_B(1 + \kappa^2\Delta^T\Lambda)\left(I - \kappa^2\Delta^T\Lambda I + \kappa^2\Lambda\Delta^T\right) \approx C_B\left(I + \kappa^2\Lambda\Delta^T\right) = C_B + \kappa^2\Delta\Delta^T \tag{15}$$

and the effect of the systematic uncertainty amounts to an an extra term $\kappa^2\Delta\Delta^T$ in the covariance matrix, compared to the stat-only covariance matrix C_B. This is however only valid in the limit of small systematics, with the general case given by Eq. 14. It is actually easier to write this in terms of Hessians: in this case the stat-only case is given by the Hessian $H_B = C_B^{-1}$, and following Eq. 14 the systematics lead to the change

$$H_B \to H_B - \frac{1}{\kappa^{-2} + \Delta^T H_B\Delta} H_B\Delta\Delta^T H_B \tag{16}$$

which is exact even for large systematics.

So to summarize, we have the following covariance matrices:

- "Stat only" (ignoring θ) : C_B
- "First case" (profiling θ) : $C_B + \kappa^2 \Delta\Delta^T$, so $\kappa^2 \Delta\Delta^T$ is the extra covariance matrix accounting for the effect of systematics. This is however valid only for small systematics. In general one would need to either use the more general modification of Eq. 14, or express the change in the context of the Hessian, where systematics exactly manifest themselves as an extra term, given in Eq. 16.
- "Second case" (before profiling θ) :

$$\left(\begin{array}{c|c} C_B + \kappa^2 \Delta\Delta^T & \kappa^2 \Delta \\ \hline \kappa^2 \Delta^T & \kappa^2 \end{array}\right) \tag{17}$$