JLAB-TN-07-030
Y. Chao

Concept for Extension to Steering Configuration Optimization Program for 12 GeV

Overview

This is motivated by the need for a more realistic estimate of the steering configuration performance of 12 GeV CEBAF, beyond what has been done so far[1]. The need arises on two fronts:

- Correlation between error inputs
  - BPM-Quad offset correlation
  - Long range correlation in alignment error (offset & roll)
  - Magnet field error correlation originating from common power supply
  - Injection position & angle correlation[2]
  - Other error correlations
- Non-Gaussian error distributions

Furthermore, it is desirable to gain insight into the following:

- Correlation between final corrected orbit errors and other derived statistical attributes

All these lead to the conceived extension of the existing algorithm. It is believed that, without foregoing adherence to a fully analytical framework, the above needs can be satisfied, provided the input error information can be augmented and made to closely model reality. A few scenarios of such input error models will be described. The context of correlated errors in fact makes for a stronger case for an analytical approach, as none of the typical simulation platforms (DIMAD, OPTIM, G4Beamline, etc.) can easily render error distributions with generic, user-specified correlation properties, especially when the dimensionality of errors runs to the hundreds or thousands.

The immediate goal for this extension is of course to better quantify the performance of the 12 GeV orbit correction system configuration, whence all the above special requirements arise. Some prototype error models can be tested once the new algorithm is implemented. The main question after that, as always, is how to construct the error input accurately reflecting reality. The prototype models described below may help in this respect.

---

[1] JLAB-TN-06-015, JLAB-TN-07-026. Also see [1].

[2] Injection position and angle can hardly be regarded as fully independent. They should tend to have normalized correlation dominated to certain extent by $\alpha/\sqrt{\beta\gamma}$ at the injection point.

Covariance Matrix for Errors

The existing algorithm assumes that all error inputs are independent and normally distributed. Thus the probability density around event

$$\left(x_1, x_2, \cdots x_k, \cdots x_N\right)$$

is (**A** is a normalizing factor)

$$\begin{aligned} \mathrm{P}\left(x_1, x_2, \cdots x_k, \cdots x_N\right) &= \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{x_k^2}{2\sigma_k^2}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^N \prod_{k=1}^{N} \sigma_k} e^{-\frac{1}{2}\boldsymbol{X}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{X}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^N \prod_{k=1}^{N} \sigma_k} e^{-\frac{1}{2}\boldsymbol{Y}^T \cdot \boldsymbol{M}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{M} \cdot \boldsymbol{Y}} \\ &= \mathrm{A} \cdot \mathrm{P}\left(y_1, y_2, \cdots y_j, \cdots y_M\right) \end{aligned} \tag{1.1}$$

where

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ \vdots \\ x_N \end{pmatrix} = \boldsymbol{M} \bullet \boldsymbol{Y} = \boldsymbol{M} \bullet \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_M \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \frac{1}{\sigma_1^2} & & & & & \\ & \frac{1}{\sigma_2^2} & & & & \\ & & \ddots & & & \\ & & & \frac{1}{\sigma_k^2} & & \\ & & & & \ddots & \\ & & & & & \frac{1}{\sigma_N^2} \end{pmatrix} \tag{1.2}$$

The alternate coordinates ***Y***, which we are interested in at the end, is related to ***X*** through ***M***, a potentially very complicated but linear matrix[3]. The existing algorithm exploits the fact that for originally independent and normally distributed errors $\boldsymbol{x_i}$, the new distribution in $\boldsymbol{y_j}$ is again normal, albeit correlated, with its RMS enhanced by the length of the j-th row of ***M***, and its height reduced accordingly.

Now instead of an RMS error for each independent value $\boldsymbol{x_i}$, the error space is defined via a covariance matrix

$$C^{ij} = \left\langle x_i x_j \right\rangle \tag{1.3}$$

[3] We make no distinction to the rank of ***M***, which can be over, under, or critically constrained. It is all correctly dealt with at the end as explained somewhere else.

Since the covariance matrix $\boldsymbol{C}$ is real and symmetric, there always exists an orthonormal transform $\boldsymbol{O}$ that diagonalizes $\boldsymbol{C}$:

$$\left(O \bullet C \bullet O^{T}\right)_{kl} = \sum_{i,j} O^{ki} \left\langle x_i x_j \right\rangle O^{lj} = \sigma_k^2 \cdot \delta_{kl} \tag{1.4}$$

and thus we can use the orthonormal transform $\boldsymbol{O}$ to define new coordinates $\boldsymbol{z_i}$ with

$$z = O \bullet x, \quad x = O^{-1} \bullet z = O^{T} \bullet z \tag{1.5}$$

where orthonormality of $\boldsymbol{O}$ was used. If the errors $\boldsymbol{x_i}$ are intrinsically Gaussian, then $\boldsymbol{z_i}$'s are independent and normally distributed, taking on the form of overall distribution for independent and normally distributed $\boldsymbol{x_i}$'s in (1.1) with its own diagonal error matrix $\boldsymbol{\Sigma^z}$. Now the new distribution in the final errors $\boldsymbol{y_j}$ become simply

$$\begin{aligned} \mathrm{A} \cdot \mathrm{P}\left(y_1, y_2, \cdots y_j, \cdots y_M\right) &= \frac{1}{\left(\sqrt{2\pi}\right)^N \prod_{k=1}^{N} \sigma_k^z} e^{-\frac{1}{2} \boldsymbol{Z^T} \bullet \boldsymbol{\Sigma^z} \bullet \boldsymbol{Z}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^N \prod_{k=1}^{N} \sigma_k^z} e^{-\frac{1}{2} \boldsymbol{X^T} \bullet \boldsymbol{O^T} \bullet \boldsymbol{\Sigma^z} \bullet \boldsymbol{O} \bullet \boldsymbol{X}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^N \prod_{k=1}^{N} \sigma_k^z} e^{-\frac{1}{2} \boldsymbol{Y^T} \bullet (\boldsymbol{O} \bullet \boldsymbol{M})^T \bullet \boldsymbol{\Sigma^z} \bullet (\boldsymbol{O} \bullet \boldsymbol{M}) \bullet \boldsymbol{Y}} \end{aligned} \tag{1.6}$$

Thus by making the transition

$$\boldsymbol{M} \rightarrow \boldsymbol{O} \bullet \boldsymbol{M}$$
$$\sigma_k^x \rightarrow \sigma_k^z$$

the same formalism used for independent errors applies in the case of correlated errors. The matrix $\boldsymbol{O}$ always exists for physically consistent[4] error covariance matrices. The only question is whether numerical conditioning of a very large covariance matrix is always good enough to afford a numerical computation of $\boldsymbol{O}$.

[4] This can be simply stated as $\left\langle x_i x_j \right\rangle^2 \leq \left\langle x_i x_i \right\rangle \cdot \left\langle x_j x_j \right\rangle$ in order for all σ_k^z's to be positive definite.

More General Forms of Covariance

Here we investigate the evaluation of not only RMS, but also correlation between derived errors such as residual orbit after steering. In other words we are interested in the full error covariance of the derived quantities. In addition to being a worthwhile question itself, this question also has bearing on the next section on more general errors.

When we consider this problem, amounting to re-expressing an N-dimensional probability distribution in a different space through the map $\boldsymbol{y}=\boldsymbol{M}\cdot\boldsymbol{x}$, complication arises when the map $\boldsymbol{M}$ is not square, and thus does not have a well-defined inverse[5]. The geometry-based construction of the RMS of the derived quantities, with detail given in [1], is free of such complications caused by rank deficiency or surplus of the map $\boldsymbol{M}$. In a sense it is always constructed for an under-constrained map from one dimension back to the original N-dimensional space along the direction defined by a row in $\boldsymbol{M}$. To do the same for a more general covariance, one must find out how such probability distribution evolves under the map.

We resort to geometrical construction similar to but more involved than that in [1]. Firstly we must transform from the original error space to an error-normalized space

$$
\begin{aligned}
\mathrm{x} &\to \mathrm{C}\cdot\mathrm{x}, \\
\mathrm{M} &\to \mathrm{M}\cdot\mathrm{C}^{-1}, \\
\mathrm{y}=\mathrm{M}\cdot\mathrm{x} &\to \mathrm{M}\cdot\mathrm{C}^{-1}\cdot\mathrm{C}\cdot\mathrm{x}=\mathrm{y}, \\
\mathrm{C} &= \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1/\sigma_{\mathrm{N}} \end{pmatrix}.
\end{aligned}
\tag{1.7}
$$

such that the contour of constant probability density in this new space is an N-1 dimensional unit sphere. In the following we will use the symbol $\mathbf{M}$ for $\mathbf{M}\cdot\boldsymbol{C}^{-1}$, with the understanding that it is the map from the error-normalized space to the final physical space of errors.

Now we construct the contour of constant probability density in the mapped space, spanned by two vectors $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$. Eventually we want to arrive at the covariance $\langle \boldsymbol{y}_i \cdot \boldsymbol{y}_j \rangle$ for the mapped probability distribution. This is formulated as follows. Take the point

$$
\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{a}_i \\
\boldsymbol{y}_j &= \boldsymbol{a}_j ,
\end{aligned}
\tag{1.8}
$$

the probability for this to happen, up to a normalizing factor of the map $\mathbf{M}$, corresponds to the integrated probability back in the $\boldsymbol{x}$-space over all $\boldsymbol{x}$ satisfying

[5] Pseudo-inverse will not do.

$$\sum_{k,N} \boldsymbol{M}_{ik}\boldsymbol{x}_k = \boldsymbol{a}_i$$

(1.9)
$$\sum_{k,N} \boldsymbol{M}_{jk}\boldsymbol{x}_k = \boldsymbol{a}_j$$

$$\boldsymbol{M}\bullet\boldsymbol{X} = \boldsymbol{A}$$

where ***M***, ***X***, and ***A*** are vector representations of the N by 2 map from the ***x***-space to the 2D subspace of ***y***. The points ***x*** satisfying (1.9) are simply all the points connected by the null space vectors of ***M*** (an N-2 dim. subspace) to any point whose image by ***M*** gives ***A***, or more explicitly, those lying on the intersection of the two N-1 hyper-planes defined by the first two equations of (1.9), itself an N-2 hyper-plane. By integrating the probability density function over this entire N-2 hyper-plane one will get, up to the normalizing factor, the probability for (1.9) to be true. Thanks to the fact that we already normalized the original space such that the contours of constant probability density are spheres, the integral in question now only depends on its distance from the origin. To get this distance, we solve for the minimal distance from the origin of any point on (1.9) by differentiating the distance function $\boldsymbol{x}^T\cdot\boldsymbol{x}$ subject to constraints realized through Lagrange multipliers:

$$\frac{\partial F_{ij}}{\partial x_m} = 0, \qquad m = 1, 2, \ldots N$$

(1.10)
$$F_{ij} = \sum_{m=1}^{N} \boldsymbol{x}_m{}^2 - \lambda_1\left(\sum_{k=1}^{N} \boldsymbol{M}_{ik}\boldsymbol{x}_k - \boldsymbol{a}_i\right) - \lambda_2\left(\sum_{k=1}^{N} \boldsymbol{M}_{jk}\boldsymbol{x}_k - \boldsymbol{a}_j\right)$$

where dependence on the indices i & j is made explicit. Equation (1.10) leads to the following

$$\begin{pmatrix} 2 & 0 & \cdots & 0 & \boldsymbol{M}_{i1} & \boldsymbol{M}_{j1} \\ 0 & 2 & \cdots & 0 & \boldsymbol{M}_{i2} & \boldsymbol{M}_{j2} \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots \\ 0 & 0 & 0 & 2 & \boldsymbol{M}_{iN} & \boldsymbol{M}_{jN} \\ \boldsymbol{M}_{i1} & \boldsymbol{M}_{i2} & \cdots & \boldsymbol{M}_{iN} & 0 & 0 \\ \boldsymbol{M}_{j1} & \boldsymbol{M}_{j2} & \cdots & \boldsymbol{M}_{jN} & 0 & 0 \end{pmatrix} \bullet \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_1 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \boldsymbol{a}_i \\ \boldsymbol{a}_j \end{pmatrix}$$

(1.11)

or

$$\begin{pmatrix} 2\boldsymbol{I} & M^T \\ M & 0 \end{pmatrix} \bullet \begin{pmatrix} X \\ \Lambda \end{pmatrix} = \begin{pmatrix} 0 \\ A \end{pmatrix}$$

(1.12)

where ***M***, ***X***, and ***A*** are as defined in (1.9), ***I*** is the N-dim. Identity, and $\boldsymbol{\Lambda}$ is the vector formed by the Lagrange multipliers.

Solution of (1.11) is straightforward through

(1.13)
$$\left.\begin{array}{l} X=\frac{1}{2}M^{T}\cdot\Lambda \\ M\cdot X=A \end{array}\right\}\rightarrow M\cdot X=\frac{1}{2}\left(MM^{T}\right)\cdot\Lambda=A$$
$$\rightarrow\begin{cases} X=M^{T}\cdot\left(MM^{T}\right)^{-1}\cdot A \\ \Lambda=2\left(MM^{T}\right)^{-1}\cdot A \end{cases}$$

Note that barring pathological maps for ***M***, the quantity

$$\left(MM^{T}\right)^{-1}$$

is well defined and always exists[6]. This is independent of whether the original **M** is over, under or critically constrained, thus averting the problem of finding the proper "inverse" map for **M** to correctly describe the mapping of the probability contours[7].

So the distance ***C*** from the origin to the N-2 hyper plane in ***x*** is from (1.13)

(1.14)
$$\begin{aligned} C=X^{T}\cdot X&=A^{T}\cdot\left[\left(MM^{T}\right)^{-1}\right]^{T}\cdot A \\ &=A^{T}\cdot\left(MM^{T}\right)^{-1}\cdot A \end{aligned}$$

The last equality comes from the fact that ***C*** is a scalar. Equation (1.14), up to the normalizing factor, basically defines the equation for the contour of constant probability density in the mapped i-j subspace, independent of whether the original **M** is over, under or critically constrained. All pairs (***a***$_i$, ***a***$_j$) of (1.8) satisfying the elliptic equation (1.14) have the same probability density ***C***:

(1.15)
$$\begin{aligned} P\left(A\right)=P\left(\boldsymbol{y}_i=\boldsymbol{a}_i,\boldsymbol{y}_j=\boldsymbol{a}_j\right)&=Q\cdot e^{-\frac{1}{2}A^{T}\cdot\left(MM^{T}\right)^{-1}\cdot A} \\ &=\frac{\sqrt{\det(N)}}{\pi}\cdot e^{-\frac{1}{2}A^{T}\cdot N\cdot A} \\ N&=\left(MM^{T}\right)^{-1} \end{aligned}$$

---

[6] The map from ***A*** back to ***X*** is closely related to the projection operator for an under constrained map ***M***

$$\Pi^{\mathrm{U}}=M^{T}\cdot\left(MM^{T}\right)^{-1}\cdot M$$

which projects out the subspace in ***X*** whose image encompasses the entire image space of ***M*** in a least-square sense back in the ***X*** space. This compares to the projection operator for an over constrained map projecting out the image subspace defined by ***M***:

$$\Pi^{\mathrm{O}}=M\cdot\left(M^{T}M\right)^{-1}\cdot M^{T}$$

Both are reduced to the identity transform when ***M*** becomes critically constrained, of course.

[7] In our notation remember **M** is the M by N (normalized) map from ***x*** to ***y***, while ***M*** is the 2 by N sub-map of **M** from ***x*** to the i-j subspace of ***y***,

where the normalizing factor can be easily determined. The covariance matrix for such a correlated bilinear Gaussian distribution can be readily worked out. Since $\boldsymbol{N}$ is real symmetric, it can always be diagonalized by an orthonormal transform $\boldsymbol{O}$ ($\boldsymbol{y}$ and $\boldsymbol{a}$ are used interchangeably):

$$\left(O \bullet N \bullet O^T\right)_{kl} = \sum_{i,j} O^{ki} \left\langle y_i y_j \right\rangle O^{lj} = \sigma_k^{-2} \cdot \delta_{kl} = \Sigma_{kl}$$
$$z = O \bullet y, \quad y = O^{-1} \bullet z = O^T \bullet z$$

Thus

$$\begin{aligned}
\left\langle \boldsymbol{a}_i \cdot \boldsymbol{a}_j \right\rangle &= \frac{\sqrt{\det(N)}}{\pi} \iint_{-\infty,\infty} d\boldsymbol{a}_i \, d\boldsymbol{a}_j e^{-\frac{1}{2} A^{\boldsymbol{T}} \cdot N \cdot A} \boldsymbol{a}_i \cdot \boldsymbol{a}_j \\
&= \frac{\sqrt{\det(\Sigma)}}{\pi} \sum_{m,n} O_{im}^{-1} O_{jn}^{-1} \iint_{-\infty,\infty} d\boldsymbol{z}_m d\boldsymbol{z}_n e^{-\frac{1}{2} \boldsymbol{z}^{\boldsymbol{T}} \cdot \Sigma \cdot \boldsymbol{z}} \boldsymbol{z}_m \cdot \boldsymbol{z}_n \\
&= O^{-1} \bullet \Sigma^{-1} \bullet O^{-1T} \\
&= N^{-1}
\end{aligned}$$

where orthonormality of $\boldsymbol{O}$ was extensively used. Thus

$$\begin{aligned}
\left\langle \boldsymbol{a}_i \cdot \boldsymbol{a}_j \right\rangle &= N^{-1} = M \cdot M^{\boldsymbol{T}} \\
&= \sum_{m=1}^{N} M_{im} M_{jm} \\
&= \sum_{\substack{m=1 \\ n=1}}^{N} M_{im} \left\langle \boldsymbol{x}_m \cdot \boldsymbol{x}_n \right\rangle M_{jn} \qquad \left(\left\langle \boldsymbol{x}_m \cdot \boldsymbol{x}_n \right\rangle = \delta_{mn} \right)
\end{aligned} \tag{1.16}$$

where we used the fact that the $\boldsymbol{x}$-space is normalized.

It is useful to note that nowhere in the derivation from (1.7) to (1.15) did the use of 2 final coordinates ($\boldsymbol{a_i}$, $\boldsymbol{a_j}$), or an N by 2 map $\boldsymbol{M}$, take on special significance. The same algebra would have held for 3, 4 or more final coordinates. It should break down when the number of final coordinates exceeds N, the dimensionality of $\boldsymbol{X}$[8], when $\boldsymbol{M} \cdot \boldsymbol{M}^{\boldsymbol{T}}$ becomes rank-deficient and no longer supports an inverse. But we are safely far from that since the most interesting case corresponds to 2 final coordinates where the bilinear covariance is evaluated. For higher moments involving more than 2 final coordinates, only the final integral and (1.16) need be modified accordingly, but that is quite straightforward.

Equation (1.16) represents a quite intuitive result that the covariance of the final derived quantity can simply be inferred directly from the initial covariance and the map $\boldsymbol{M}$. This can conceivably be extended to cases where the initial probability densities are not even Gaussian, as discussed in the following section. However in such cases the probability in the final space would be difficult to interpret, unlike the current case, where the important property that the probability density P is a function only of the norm of the probability vector in the initial space,

[8] This is typically in the hundreds if not thousands.

$$P(a_1, a_1, \cdots, a_N) = P(\|a\|)$$

made intuitive geometrical picture possible.

Non Gaussian Errors

If the errors are no longer normally distributed, the nice persistent Gaussian behavior preserved by linear maps is lost. However, this will not keep us from propagating covariance matrices. It is only the interpretation of these matrices that suffers from the loss of Gaussian-ness.

Again consider error coordinates x_i and y_j as described in (1.2), with covariance of x_i as given in (1.3) by C_{ij}. It is straightforward to get the covariance of y_j as

(1.17)
$$\begin{aligned} C_Y^{ij} &= \langle y_i y_j \rangle \\ &= \langle M_{ik} x_k M_{jl} x_l \rangle \\ &= M_{ik} \langle x_k x_l \rangle M_{jl}, \\ C_Y &= M \bullet C_X \bullet M^T \end{aligned}$$

Note here we are not even assuming independent initial errors for x_i's. Of course it reduces to the equivalent form of (1.16) for independent x_i's, and supports a Gaussian interpretation if x_i's are normally distributed.

Therefore up to this point we have established the machinery for handling input errors that are neither independent[9] nor Gaussian. As a bonus we can also compute the covariance between any two final error coordinates, as indicated by (1.16) or (1.17).

[9] But correlated only in bilinear form, $\langle X_i X_j \rangle$ of course.

Constructing Realistic Covariance Matrices

We proceed to investigate how one can accurately realize the types of error correlations mentioned at the beginning through judicially constructed covariance matrices. The following gives prototypes for some of these examples.

BPM-Quad offset correlation

Define transverse offsets (in X) of a BPM and a quad sharing the same girder by ***Xb*** and ***Xq*** respectively. Define further their average ***Xa*** and difference ***Xd*** via

(1.18)
$$X_a = \frac{X_b + X_q}{2}$$
$$X_d = X_b - X_q$$

We take the logical view that it is ***Xa*** and ***Xd***, not ***Xb*** and ***Xq*** that are independent, with ***Xa*** considerably larger than ***Xd***. It is quite easy to see that the covariance matrix for ***Xb*** and ***Xq*** is

(1.19)
$$\left\langle X_b \cdot X_q \right\rangle = \begin{pmatrix} \sigma_a^2 + \frac{\sigma_d^2}{4} & \sigma_a^2 - \frac{\sigma_d^2}{4} \\ \sigma_a^2 - \frac{\sigma_d^2}{4} & \sigma_a^2 + \frac{\sigma_d^2}{4} \end{pmatrix}$$
$$\sigma_a^2 = \left\langle X_a \cdot X_a \right\rangle$$
$$\sigma_d^2 = \left\langle X_d \cdot X_d \right\rangle$$

Some special cases are

- $\sigma_d^2 = 4\sigma_a^2$ BPM & quad offsets are completely independent.
- $\sigma_d^2 >> 4\sigma_a^2$ BPM & quad offsets are strongly anti-correlated.
- $\sigma_d^2 << 4\sigma_a^2$ BPM & quad offsets are strongly correlated, reaching 100% in the case $\sigma_d^2 = 0$.

Long range correlation in alignment error

Assume that the long-range alignment baseline, instead of conforming exactly to the theoretical central trajectory, oscillates in a slow sinusoidal fashion about this trajectory with characteristic amplitude and wavelength ***Wa*** and ***λa*** respectively. Remote elements are aligned with respect to this oscillating baseline with their own truly independent errors ***Xi*** for the i-th element, thereby developing a long-range oscillating correlation through the baseline. A priori there should be no preferred element or starting point of this oscillation to be singled out from the ensemble, thus all element offsets should depend equally in amplitude on this oscillation, which will be manifest only in inter-element correlations through phase advance between elements.

To realize this picture, we have

(1.20)
$$X^{i}_{E}(z) = X_{B}(z) + X^{i}_{F}$$

where the subscripts E, B and F denote "element", "(sinusoidal) baseline" and "fluctuation about baseline". The offset error of the element $\boldsymbol{X_E}$ is a function of both the distance z and the element index i. The errors $\boldsymbol{X_B}$ and $\boldsymbol{X_F}$ are clearly independent, as are different $\boldsymbol{X_F}$'s.

To evaluate the covariance of $\boldsymbol{X_E(z)}$, the dependence of its RMS on $\boldsymbol{X_B}$ will only be through the latter's RMS, or self-covariance

$$\sigma^{2}_{B} = \langle X_{B}(z) \cdot X_{B}(z) \rangle = W^{2}_{a}/2$$

without resorting to its detailed phase structure in z. This allows us to treat all elements on an equal basis with no undue preference. Note that by doing this we are still treating this oscillating baseline on a stochastic basis, with its offset distributed from 0 to a few σ's, rather than following a tight, well-defined sinusoidal line. The sinusoidal behavior of the baseline is manifested exclusively through the off-diagonal covariance matrix elements as

$$\langle X_{B}(z) \cdot X_{B}(z+\delta z) \rangle = \frac{W^{2}_{a}}{2} \cdot \cos\left(2\pi \frac{\delta_z}{\lambda_a}\right) \cdot e^{-\frac{\delta z}{\Delta}}$$

where a "damping term" has also been added with decay length Δ to account for diminishing long-range correlations. Apart from the damping this is fully consistent with the definition of the RMS of $\boldsymbol{X_B(z)}$ above.

Now the full covariance matrix for the offset of elements at position z with indices i, j,….N can be explicitly written

(1.21)
$$\langle X^{i}{}_{E}(z) \cdot X^{j}{}_{E}(z+\delta z) \rangle = \begin{cases} \frac{W^{2}_{a}}{2} + \sigma^{i}{}^{2}_{F} & i = j \\ \frac{W^{2}_{a}}{2} \cdot \cos\left(2\pi \frac{\delta_z}{\lambda_a}\right) \cdot e^{-\frac{\delta z}{\Delta}} & i \neq j \end{cases}$$

$$\sigma^{i}{}^{2}_{F} = \langle X^{i}{}_{F} \cdot X^{i}{}_{F} \rangle$$

It is interesting to note that this scheme automatically includes the previous case of BPM-quad offset covariance. If the BPM and quad in question are strongly correlated, they resemble two generic elements here, close in z with relatively small σ_F's compared with Wa. And their covariance matrix becomes (δ is a number <<1, from the oscillating and damping effects).

$$\begin{pmatrix} \frac{W_a^2}{2} + \sigma_B^{i\,2} & \frac{W_a^2}{2}(1-\delta) \\ \frac{W_a^2}{2}(1-\delta) & \frac{W_a^2}{2} + \sigma_Q^{i\,2} \end{pmatrix}$$

roughly equivalent to the strongly correlated case of (1.19). When σ_B, σ_Q, and δ are all 0 the correlation is 100%. In some sense this formalism is better as it allows independent assignments of fluctuating offset errors to different flavors of elements if one so chooses.

The correlation defined in (1.21) does not need to be the only possible form of correlation either. Depending on realistic physical model for element offsets, it is clear that other forms of correlation can be constructed as well.

It is also useful to note that when the element-specific error σ_F's are negligible, and damping is ignored, the correlation between elements separated by distance z, driven purely by the oscillating baseline, can be diagonalized into <u>independent</u> errors

$$X_{\pm}(z,\delta z) = \frac{X_E(z) \pm X_E(z+\delta z)}{\sqrt{2}}$$

$$\left.\begin{matrix} \langle X_+(z,\delta z)\cdot X_+(z,\delta z)\rangle \\ \\ \langle X_-(z,\delta z)\cdot X_-(z,\delta z)\rangle \end{matrix}\right\} = \left\{\begin{matrix} W_a^2\cdot\cos^2\left(\pi\frac{\delta_z}{\lambda_a}\right) \\ \\ W_a^2\cdot\sin^2\left(\pi\frac{\delta_z}{\lambda_a}\right) \end{matrix}\right.$$

$$\langle X_+(z,\delta z)\cdot X_-(z,\delta z)\rangle = 0$$

where the choice between sin & cos depends on the quadrant in which the phase difference angle resides. Note also the trigonometric argument is <u>half</u> of the phase difference angle. Thus under this scheme we can always generate independent probability distributions for combinations of individual element offsets. The RMS's of these distributions give away the inter-element phase advances, although we never need to pinpoint a location where the oscillation starts, which would have been counter-intuitive.

Any long-range correlation in the dipole roll errors can conceivably be modeled in a similar fashion.

Magnet field error correlation originating from common power supply

Here we consider the correlation characterizing field errors in magnets powered by a common supply. We can take the field error for the i-th magnet to be

$$\delta B_i = F_i \cdot \delta I + \delta b_i$$

where ***I*** is the common power supply current, which is related to the field ***Bi*** through the *I-B* conversion factor ***Fi***. On top of this field driven by the main power supply, each magnet can take on a very small independent field fluctuation ***bi***. It is straightforward to show that the covariance matrix for these field errors is

$$\left\langle \delta B_i \cdot \delta B_j \right\rangle = \begin{cases} F_i^2 \cdot \sigma_I^2 + \sigma_{bi}^2 & i = j \\ F_i \cdot F_j \cdot \sigma_I^2 & i \neq j \end{cases}$$

(1.22)

$$\sigma_I^2 = \left\langle \delta I \cdot \delta I \right\rangle$$

$$\sigma_{bi}^2 = \left\langle \delta b_i \cdot \delta b_i \right\rangle$$

For small $\sigma_{\boldsymbol{b}}$'s this system is highly correlated, which shouldn't be surprising. The finite values of $\sigma_{\boldsymbol{b}}$'s keep the system from being 100% correlated, which can be unrealistic.

Injection position & angle

In most modeling & simulation programs, for lack of proper coding provision, injection position and angle of orbits are treated as independent quantities. This cannot be farther from the truth. Quick inspection of any BPM data[10] indicates that in reality these are usually highly correlated quantities because of the optics. Taking the view that for a transfer line[11] the initial position-angle correlation closely track the design beam ellipse, and that long range transport does not suffer too much from non-symplectic effects not included in the design model, one can readily assert that the correct covariance matrix to use for injection errors is simply

$$\begin{pmatrix} \langle x \cdot x \rangle & \langle x \cdot x' \rangle \\ \langle x \cdot x' \rangle & \langle x' \cdot x' \rangle \end{pmatrix} = \begin{pmatrix} E \cdot \beta_D & -E \cdot \alpha_D \\ -E \cdot \alpha_D & E \cdot \gamma_D \end{pmatrix} \tag{1.23}$$

where ***E*** is a measure of the size of the RMS of the single-particle emittance (or Courant Snyder invariant) of the trajectories, and α, β, and γ are the design Twiss parameters.

.

[10] Few are more dramatic than the FOPT data viewed in higher arcs when the launch starts from Arc 1.
[11] For storage rings this can be asserted with even higher confidence.

A Simple Example

Here is an example of how the global covariance matrix can be established based on the models described above. The entire error system is made up of injection errors (position and angle), 2 quad offsets, 3 BPM offsets, and 3 magnet errors from the same supply string. Roll errors in dipoles are omitted without losing the point of this example. Only upper half of the matrix is explicitly shown.

$$\begin{pmatrix}
E\cdot\beta_D & -E\cdot\alpha_D & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & E\cdot\gamma_D & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & W_a^2/2+{\sigma^1}_Q^2 & W_a^2/2.\cos\left(2\pi\delta z_{QQ}^{12}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{11}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{12}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{13}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & 0 & 0 & 0 \\
 & & & W_a^2/2+{\sigma^2}_Q^2 & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{21}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{22}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{QB}^{23}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & 0 & 0 & 0 \\
 & & & & W_a^2/2+{\sigma^1}_B^2 & W_a^2/2.\cos\left(2\pi\delta z_{BB}^{12}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & W_a^2/2.\cos\left(2\pi\delta z_{BB}^{13}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & 0 & 0 & 0 \\
 & & & & & W_a^2/2+{\sigma^2}_B^2 & W_a^2/2.\cos\left(2\pi\delta z_{BB}^{23}/\lambda a\right).e^{-\frac{\delta z}{\Delta}} & 0 & 0 & 0 \\
 & & & & & & W_a^2/2+{\sigma^3}_B^2 & 0 & 0 & 0 \\
 & & & & & & & F_1^2\cdot\sigma_I^2+\sigma_{b1}^2 & F_1\cdot F_2\cdot\sigma_I^2 & F_1\cdot F_3\cdot\sigma_I^2 \\
 & & & & & & & & F_2^2\cdot\sigma_I^2+\sigma_{b2}^2 & F_2\cdot F_3\cdot\sigma_I^2 \\
 & & & & & & & & & F_3^2\cdot\sigma_I^2+\sigma_{b3}^2
\end{pmatrix}$$

A few notes on nomenclature:

- δz_{PQ}^{mn} Distance in ***z*** between the m-th element of type P and the n-th element of type Q. As explained earlier, this frees one from the need to keep track of which quad and BPM should be paired together, but leaves all questions of modeling such correlations to evaluating the sinusoidal and exponential functions. The argument ***z*** in the exponent should carry the same indices m, n, P, Q as that of the cosine, but these indices are not made explicit for lack of space.
- ${\sigma^m}_P^2$ RMS-squared of the (independent) alignment error of the m-th element of type P relative to the (possibly fluctuating) baseline. By making allowance for these independent RMS, we maintain full flexibility to model any type-specific or even element-specific errors necessary.
- σ_{bm}^2 RMS-squared of the (independent) field ripple error of the m-th magnet relative to the common power supply. Again we maintain the flexibility to model element-specific errors.
- Other symbols are as described in previous sections.

Of course in a realistic beam line this matrix would be of dimensionality in the hundreds or thousands. Its construction however would be straightforward, through repetitive application of the few rules outlined above.

Discussion

- Extension of existing algorithm to systems with correlated error inputs is conceptually trivial, although retrofitting existing code to implement this extension may be complicated. One big question is input format.
- Defining form of correlation in terms of covariance matrix, on the other hand, requires some insight into physical properties of the system. Several physically justifiable implementations have been described.
- The new approach can also address non-Gaussian error inputs, as long as physically consistent covariance matrix can be established. No pretense is made in interpreting such results, yet.
- The new approach can also produce correlation between final corrected orbit errors and other derived statistical attributes. Extension to arbitrary-order moments has been described and appears trivial.
- The burden appears to shift toward generating the full error covariance matrix in this new scheme. This can be algorithmically defined following certain first-principle guidelines. A few examples were discussed for typical cases, including correlations between BPM-Quad offsets, between elements distributed over long range about a sinusoidal baseline, between magnets powered by a common supply, and between injection position and angle.
- Other correlations can be handled under the same framework as long as a covariance matrix can be specified.
- It is expected that the main advantage of the existing analytical algorithm over simulation, namely completeness, rigor, insight, and speed, will remain intact in the extension, since the formulation remains completely analytical.
- The case for an analytical approach seems even stronger in the presence of error correlations, since it appears that no available simulation platform can render error distributions conforming to user-specified high dimensional covariance with generality and ease.
- Although diagonalization of the covariance matrix was invoked in explaining the new algorithm, the end results actually do not depend on the execution of this procedure, as shown in (1.17). This is important in the context of numerical stability, as any need in diagonalizing a very large matrix is bypassed. Equation (1.17) suggests that only smooth, well-behaved numerical operations are required.

Reference

[1]. Y. Chao & V. Mertens, LHC Report 470.