Correction of Cusps in Beam Location

M. Fischler

July 9, 2003

Contents

1 Overview and Problems **1**

2 Single-DFG Compensation **3**

3 Two-DFG Compensation **3**

4 Choosing Two DFG's **4**

5 Some 3-bump Techniques **4**

6 The Generic Minimization Approach **4**

7 Compensation For Non-Cusp (Distributed) Distortions **4**

8 Discrete Distortions **5**
8.1 Single Distortions 5
8.2 Special Cases of Multiple Distortions 6
8.2.1 Restricted-range 2-bumps 6
8.2.2 3-bumps 6
8.2.3 4-bumps 6

9 Distortion Due to a Distributed Dipole Field **7**

1 Overview and Problems

Say that in fitting the beam locations measured by BPM's, it is determined that there appears to be a cusp due to a single distortion in the field, of magnitutde A and at position μ_c.

The equation that describes both the effect of this unwanted dipole field distortion, and the effect of a corrector at position j, on the beam location at position i, is:

$$a_{ij} = \frac{\sqrt{\beta_i \beta_j}}{2\sin(\pi Q)} \cdot \cos(|\mu_i - \mu_j| - \pi Q) \tag{1}$$

where β_i (β_j) is the β-function and μ_i (μ_j) is the betatron phase at position i (j). For this document, the coefficient is moot, and we will re-write this as

$$a_{ij} = A\cos(|\mu_i - \mu_j| - \pi Q) \tag{2}$$

where Q is the beam tune.

If the cusp position lies precisely on top of a DFG (likely because the field distortion is being produced by misbehavior of the DFG) then (assuming that this DFG responds to corrections) it is easy to completely eliminate the cusp by setting the DFG to deliver a kick wich is negative of that which would produce that cusp. But in general, there is no DFG right on top of a given field distortion. The question addressed in this note is how best to assign corrections to other DFG's such that the cusp (and more importantly, the cosine curve distortion away from the cusp on either side) is compensated for as well as possible.

The question arises of what one means to compensate "as will as possible." We shall assume that there is some sort of penalty function which needs to be minimized; this function is in general an integral, over positions in betatron phase, of some function describing the beam loss due to incorrect location. This loss function is nearly zero for very small location distortions, and rises sharply when the distortion exceeds some radius. Although in principle the loss function can depend strongly on position along the beam–for example, there may be a known constriction of the pipe leadingto a position at which precise location is extremely important–we simplify by assuming the loss function is some position-independant norm-like function. Three attractive models to use are the square of the distortion (thus minimizing chi-squared for the beam), the value of the worst distortion (equivalent to an inf-norm) and the fourth power of the distortion (this 4-norm has properties which lie between the other two cases.)

One trap to avoid is the temptation to add up deviations only at the BPM positions. Clearly beam loss occurs all along the orbit, and a correction which does well at the 100-od BPM positions but leaves a severe distortion between two BPM's may not be a good correction.

The DFG's in the ring are spaced about 60 degrees apart in betatron phase, so finding a set of DFG's that will do a good job of compensating for a distortion at an arbitrary position can be difficult. Some approaches for correction are:

- To utilize the DFG nearest the cust to attempt to compensate.
- To use the two nearest DFG's on either side of the cusp.
- To select (based on criteria of how well they can compensate outside their range, and how much distortion they leave between them) one good DFG's on each side of the range and do 2-DFG correction based on those.
- In each of a sequence of ranges, create a 3-bump that best compensates for the distortion in that range (and which by the nature of 3-bumps does not affect the beam locations outsid ehtat range).
- To use some smooth-function minimization technique (for cases similar to the 2-norm or 4-norm penalty function) or linear programming (for the case where the penalty function to be minimized is the worst deviation in location) to compute settings for all DFG's at once, to compensate for all cusps at once.

At the end of this note, we will also discuss a separate problem: the case where it has been determined that a large-scale small distributed distortion is present. Unless the gerneric minimizaton approach is take, compensating for distributed distortions can require a different approach than compensating for individual cusps.

2 Single-DFG Compensation

The first remedy for a cusp distortion that comes to mind is to cancel it by giving the nearest DFG a distorting field of the same magnitude and opposite sign. Indeed, if the DFG is right at the position of the distortion, this prescription works perfectly.

But attempting to correct a cusp by a single DFG works well only if that DFG is close in betatron phase to the position of the cusp, because two cosine curves can cancel only if they are in phase. Given that the DFG's are about 60° apart, the nearest DFG will be up to 30° away from the cusp. In such a case, the distortion curve far from the cusp and correction will reflect a 30° phase mismatch, and thus cannot have an amplitude which is less than a factor of $\sin 30 / \sin 75 = .517$ smaller than the original distortion. About half the time one can achieve reductions by a factor of .25 or less using a single DFG.

A side issue is the fact that this remaining distortion is not localized near the original distorting field, but manifests as a cosine curve around the ring. This can potentially be worse than a localized distortion of the same magnitude would be, because there is a much longer opportunity for beam loss. Certainly if the 2-norm penalty function is used, single DFG compensation would not cut the mustard.

For this technique, the value of Q makes little or no difference. A small value, say .1, would allow the distortion near the cusp position to be almost completely clattened, while the actual value of $Q = .585$ will cause the distortion near the original cusp to be almost as large as the fluctuations in the exterior curves. But it remains the latter fluctuations that dictate the quality of the correction.

3 Two-DFG Compensation

Using two DFG's which straddle the cusp distortion, one can always cause the distortion outside the range of those positions to be exactly zero. The easy way to compute this is to assign the DFG's at μ_1 and μ_3 kicks such that (μ_1, μ_2, μ_3) exactly form a 3-bump.

The fly in the ointment is that the fluctuations in the beam locations between μ_1 and μ_3 do not disappear when this is done. For example, at $Q = 2n + .3$ and with a cusp halfway between two DFG's separated by 60°, the correction which flat-lines the exterior distortions leaves a V-shaped local distortion at the position of the original cusp. This distortion has magnitude .467 times that of the original distortion. However, one is free to under-correct by reducing the amplitude computed for the compensating DFG's. In the case described, if you under-correct by a factor of .77, then the exterior curves are about .23 times their original amplitues, and the V-shaped distortion at the original cusp is down by a factor of .23 as well.

However, in the actual machine, $Q = 2n + .585$. This situation is worse for 2-DFG correcting. With that same cusp halfway between two DFG's separated by 60°, the correction which flat-lines the exterior distortions leaves a V-shaped local distortion of magnitude .56.

And this can only be reduced to .43 (by under-correcting by a factor of .57 in amplitudes of the corrector DFG's).

As before, there is room for debate about how much exterior fluctuation can be traded off to improve fluctation at and near the cusp. But assuming we want to go for minimum worst-case deviation, we see that 2-DFG compensation can always achieve a reduction by at least a factor of .43 if Q is fixed at a phase of .585.

When the cusp is more nearly in phase with one or the other of the DFG's we can do better than this; for example if the cusp is only $15°$ away from one of the correctors, we can achieve a reduction by a factor of .35. And matters improve quite rapidly if the separation is below $15°$. Still, at average, the best available reduction in worst-case deviation is by a factor of .3.

4 Choosing Two DFG's

Phase near the cusp

issues of other cusps

and wigles between the DFGs

Curve is terrible. Just does not help.

5 Some 3-bump Techniques

6 The Generic Minimization Approach

7 Compensation For Non-Cusp (Distributed) Distortions

Two-DFG Compensation

Consider a point i which is a small distance to the right of j, such that $\mu_i - \mu_j = \epsilon > 0$. Then (**??**) becomes

$$a_{ij} = A\cos(\epsilon - \pi Q) \tag{3}$$

Now consider moving point i along the ring, until ultimately it is at i', a small distance to the left of i, such that $\mu_j - \mu_i = \kappa > 0$. Then (**??**) becomes

$$a_{i'j} = A\cos(|(\mu_j - \kappa) - \mu_j| - \pi Q) = A\cos(\kappa - \pi Q) \tag{4}$$

But what if we had moved the point only about halfway around the ring, such that it is unclear whether to say that i' is to the right or to the left of j? Clearly, unless the two expressions (3) and (4) match, there is something fundamentally wrong with equation (**??**).

In fact, this worry can be seen even when i' is a small distance to the left of j, and at postion i which is a small distance to the right of j. Consider that i' has betatron phase of $\mu_j - \kappa$, but one could have taken i' to be a large distance to the right of j–in particular, to have a betatron phase of $\mu_j + (2\pi Q - \kappa)$. And we see in this case that in fact equation (**??**)

leads to no inconsistency because

$$\begin{aligned}\cos\left(|\mu_j - (\mu_j + (2\pi Q - \kappa))| - \pi Q\right) \\ = \cos(|\kappa - 2\pi Q| - \pi Q) \\ = \cos(2\pi Q - \kappa - \pi Q) \\ = \cos(\kappa - \pi Q)\end{aligned} \tag{5}$$

And now we can see that if the subtracted term in (??) were anything other than πQ, this equality would not hold, and the value for change in beam location would indeed be inconsistent. But since the correct value is subtracted, for any position around the ring, the change in beam location will be the same whether you consider the betatron phase offset to be positive or negative.

Finally, what if one were to subtract $4\pi Q$ from a positive offset, or add some multiple of $2\pi Q$? Then the answer gotten from equation (??) would change. Why is this not a problem? Because we know that the quadrupoles in the ring will correct the (small) distortion, so as to end up with a new (slightly displaced) orbit–that is, we know that the design orbit is stable. Equation (??) represents the change in that orbit; if you tried to take $\delta\mu$ to be multiple times around the ring, then you also would have to include multiple contributions from the stabilizing correction term, so (??) does not hold in the most naive manner.

8 Discrete Distortions

Obviously, a discrete distortion is only an approximation, saying that the distorting field is concentrated at one or several disrete points. As we will see, the mathematics of the cosine form of equation (??) make this approximation very solid.

8.1 Single Distortions

A field at j will always produce a distortion which, right at j, has the value $A\cos(\pi Q)$. When superposed on the existing location curve, this will appear as a cusp, with a change in slope of $-2A\sin(\pi Q)$ at $i = j$; this is simply $-\beta_j$. Here, A is the (integrated) dipole field strength, multiplied by $\sqrt{\beta_i\beta_j}$ which at $i = j$ is merely β_j. The salient feature is that this "kick" depends only on the field strength and the β-function at j, not on the betatron phase.

Because the β-function varies, the effect of a distortion is not a pure "cusped-cosine" curve. However, if we work with "rescaled distortion" $x(\mu) \equiv p(\mu)/\sqrt{\beta_\mu}$ then the form of $x(\mu)$ for a single distortion is, in fact, a pure "cusped-cosine" curve, with kick value of the (integrated) dipole field strength times $\sqrt{\beta_j}$. Henceforth we will work in terms of these more convenient rescaled distortions.

Because any distorting field at a single point leads to this form of beam location distortion, our fit to the beam location curves will always be of the form

$$A\cos(x + \delta) + k \tag{6}$$

That constant k represents an overall position drift, but also has another justification when we consider distorting fields which are constant over a region in betatron phase space.

8.2 Special Cases of Multiple Distortions

8.2.1 Restricted-range 2-bumps

If two distortions at μ_1 and $\mu_2 > \mu_1$ are separated by precisely $k\pi$, and their field strengths (times their β-functions) are related by $f_2 + (-1)^k f_1 = 0$ then for all phases outside of $\mu_1 < \mu < \mu_2$, the distortions caused exactly cancel. This is a "two-bump," and the concept is only of limited use in beam smoothing since in general, one has control of DFG's at fixed phases which generally don't differ by precise multiples of π.

At any rate, the shape of such a "two-bump" in rescaled postion measurement is that of k lobes of a sine curve, starting at μ_1 and ending at μ_2, and flat zero outside that range. And the peak magnitude is twice the peak magnitude of either single bump.

8.2.2 3-bumps

For three discrete distortions at three fixed positions, no matter what their relative betatron phases $\mu_1 < \mu_2 < \mu_3$ you can always adjust the magnitudes of the three kicks such that the sum of the distortions is zero outside the range $\mu_1 < \mu < \mu_3$. If we let the (fixed) phase differences be δ_{12}, δ_{13}, and δ_{23}, we find that for this to happen,

$$\frac{A_1}{A_2} = -\frac{\sin\delta_{23}}{\sin\delta_{13}}$$

$$\frac{A_1}{A_3} = -\frac{\sin\delta_{23}}{\sin\delta_{12}} \qquad (7)$$

$$\frac{A_2}{A_3} = -\frac{\sin\delta_{13}}{\sin\delta_{12}} \qquad (8)$$

$$(9)$$

These three equations are not linearly dependant; they represent only two constraints. Still, for a given set of three correctors at fixed positions, one has left only one free parameter. The shape of the bump is fixed (but depends of course on δ_{12}, δ_{13}, and δ_{23}); the magnitude (at some given point) can be varied.

The shape of a 3-bump looks like a pair of sections of two different sine curves, joined at the middle position μ_2. Every such shape (with the proviso that the frequency of each sine curve is the universal frequency, 1 in our difintion of μ) can be achieved, but this represents little freedom since two curves of arbitrary magnitude won't necessarily meet at μ_2. In real life, one cannot adjust the positions of the DFG correctors, so only a very limited selection of 3-bump shapes can be attained.

8.2.3 4-bumps

For four discrete distortions, there is one additional degree of freedom, which can be used to adjust the magnitude and slope at a given point. This vastly increases the number of bump shapes which can be achieved with DFG's at fixed positions.

The question arises: Can every 4-bump be expressed as the superposition of two 3-bumps? The degrees-of-freedom counting suggests this, and a naive argument would imply this is so: When condition on the amplitudes of corrections for a 3-bump at (μ_1, μ_2, μ_3) , that ensures that the distortion is zero to the left of μ_1, is met, that same condition ensures that the distortion is zero to the right of μ_3. Similarly for (μ_2, μ_3, μ_4); so one might think that in order to be zero outside of (μ_1, μ_4) the 4-bump must be a sum of those two 3-bumps.

The hole in that argument is that one can imagine a "bad" 3-bump at (μ_1, μ_2, μ_3) which gives sine-curve distortions outside of (μ_1, μ_3), and a similarly flawed 3-bump at (μ_2, μ_3, μ_4) such that the outside distortion on the left and right sides precisely cancel with those of the first 3-bump. Can this possibility be used to grant additional freedom, to create a valid 4-bump whichis not the sum of two valid 3-bumps?

In general the answer is no. For any given three correctors, there is a relationship between the positions of the correctors, and the phase shifts of the leftover sine curves of the left and right sides. In general, this phase shift will be different for (μ_1, μ_2, μ_3) than for (μ_2, μ_3, μ_4), so that if the left side flaws cancel, the right side flaws cannot.

There are, however, possible placements of $(\mu_1, \mu_2, \mu_3, \mu_4)$ such that the phase shifts would be equal; in particular, uniform spacing has this property. In such a case, can one form a 4-bump which is not the sum of two valid 3-bumps? In that case, what becomes of the degrees-of-freedom counting?

The answer to the dilemma is that even for specialized spacing of the correctors, one cannot form a 4-bump which is not the sum of two valid 3-bumps. Yes, you can choose amplitudes (A_1, A_2, A_3) for the bump at (μ_1, μ_2, μ_3), and (B_2, B_3, B_4) for the bump at (μ_2, μ_3, μ_4), such that the left and right flaws cancel out. When you do this, however, you will always be able to transfer some amplitude from A_2 to B_2 and from A_3 to B_3 such that although you have the identical 4-bump, now each 3-bump is no longer flawed.

Thus, we come to the conclusion that *every 4-bump which is zero outside the range $\mu_1 < \mu < \mu_4$ can be expressed as the sum of a 3-bump which is zero outside the range $\mu_1 < \mu < \mu_3$ and another 3-bump which is zero outside the range $\mu_2 < \mu < \mu_4$.*

9 Distortion Due to a Distributed Dipole Field

Consider a dipole field which is not concentrated at a single point, but instead is distributed in betatron phase. Then many contributions from equation (**??**) must be integrated to get the net distortion.

In particular, say the kick from each point x is given by equation (**??**) times some $2\sin(\pi Q) \cdot \rho(x)/\sqrt{\beta(x)}$, where we divide out the square root of the β-function so that a given value of $\rho(x)$ represents the same strength of distortion at any point x. Then when you integrate over x, equation (**??**) takes the form

$$a_j = \int dx\, \rho(x) \cos(|\mu(x) - \mu_j| - \pi Q) \tag{10}$$

or, working with betatron phase μ instead of position x and assuming $\rho(\mu)$ is non-zero only in the range $a \le \mu \le b$,

$$a_j = \int_a^b \rho(\mu) \cos(|\mu - \mu_j| - \pi Q)\, d\mu \tag{11}$$

Outside the range $a \le \mu_j \le b$, the expression for which absolute value is taken in equation (11) is either strictly positive or strictly negative, and if $\rho(\mu)$ is some simple known function, the integral can be performed analytically. For example, if $\rho(\mu)$ is linear (or linear in each of N sub-regions) then the integral is easy to do, and works out to some factor times $\cos(\mu_j + \delta)$

where the factor and phase angle depend on the details of $\rho(\mu)$. An interesting case is to take a Gaussian shape: although this is strictly speaking not confined to a finite range, it dies off quickly and if we are far enough away from the "origin" the integral for $\rho(\mu) = e^{-c\mu^2}$ is $e^{-c}\cos(\mu + \pi Q)$. The general pattern persists, that outside the range, the effect of the distortion is always of the form $A\cos(\mu_j + \delta)$, which is equivalent to that of a single discrete distortion somewhere inside the range. This equivalence theorem can in fact be proven.

In fact, outside the range of an arbitrary distorting dipole field of finite extent, you can always mimic that field by a single discrete distortion, positioned somewhere within the range and of magnitude less than or equal to the integrated field. An easy way to visualize this is to pretend the distributed field were a large number of discrete distortions and express all distortions as a vector where the angle is the phase is the phase of the distortion. Then (outside the range) the net distortion is represented by the vector sum of the individual distortions. The properties asserted above immediately follow from that fact.

Inside the range $a \le \mu_j \le b$, the intgeral must be broken into two separate integrals, since the absolute value expression changes sign at $\mu = \mu_j$. This has two effects: The answer is no longer a simple cosine curve, and the integrals are a bit less likely to have simple answers. (For example, the Gaussian-shaped distortion gives bumps whose shape is a horrible combination of error functions and exponential integrals.) This motivates the consideration of only the simple linear case: $\rho(\mu) = m\rho + c$.

A priori, one would expect that within the range of a disributed distortion, the phase of the distortion would vary with x, and that for slowly varying fields, this could be expressed as a cosine curve with a slightly "incorrect" frequency which varies slowly. It would be tempting to fit BPM data to a function of this form, so take into account the possibility of a weak but widespread distortion. (This would probably not bear fruit, for the following reason: The combination of an underlying cosine curve plus another cosine curve of different frequency gives a total of at least 8 parameters to be determined: The 3 parameters for the underlying curve, plus the magnitude, frequency and phase of the second curve, plus starting and ending points for the distortion. Thus one needs at least 8 points to simply determine the fit, and since the whole idea is rejection of outlier measurements based on statistics of nearby measurements, probably at least 12 points to do anything practical. This amounts to more than a tenth of the ring; the assumption that the distorting field is almost constant over that large a region is shaky.)

But if you do the algebra for $\rho(\mu) = m\rho + c$, you find that the expression is:

$$\begin{aligned} m\cos(b - \pi Q - x) + (bM + c)\sin(b - \pi Q - x) \\ -m\cos(a + \pi Q - x) - (aM + c)\sin(a + \pi Q - x) \\ +2(mx + c)\sin(\pi Q) \end{aligned} \tag{12}$$

In this expression, all terms but the last can be absorbed into a cosine curve with frequency one and some phase and amplitude. There is no "incorrect frequency" term. The last term is merely a linear correction. In fact, if m were zero, the entire distrotion would not change the form–equation (6)–to which we were already fitting. The last term contributes to k in that equation.

This is fortuitous, because although the equivalence theorem mentioned above always allows you to substitute discrete set of distortions, one between each BPM, for continuous distortion fields, such a course would have rendered all analysis of BPM measurements (to reject outliers) useless. On the other hand, fitting to the constrained form of equation (6) gives an excellent framework in which to search for outliers.

The bottom line for distributed distorting dipole fields is:

- They don't alter the form $A\cos(x+\delta)+k$ to which we should fit the BPM measurements.
- When you cross the end or beginning or such a region, A, δ and k of the fit will change–but the same is true when you cross the location of a discrete distortion.
- To some extent, one can identify the k found in such a fit with an indication of a distorting field distributed in betatron phase.