# FINDING THE CIRCULAR MAGNET APERTURE WHICH ENCLOSES AN ARBITRARY NUMBER OF MIDPLANE-CENTERED BEAM ELLIPSES*

J. Scott Berg, BNL, Upton, NY, USA

*Abstract*

In specifying the magnets for an accelerator, one must be able to determine the aperture required by the beam. In some machines, in particular FFAGs, there is a significant variation in the closed orbit and beta functions over the energy range of the machine. In addition, the closed orbit and beta functions may vary with the longitudinal position in the magnet. It is necessary to determine a magnet aperture which encloses the beam ellipses at all energies and all positions in the magnet. This paper describes a method of determining the smallest circular aperture enclosing an arbitrary number of midplane-centered ellipses.

## INTRODUCTION

The size of magnet apertures has a strong effect in determining the cost of an accelerator. Thus, it is generally important to determine the smallest magnet aperture which will still meet beam transmission requirements. For producing numerically optimized designs, it is necessary to have a systematic algorithm for computing the aperture.

A particular application of interest is fixed field alternating gradient (FFAG) accelerator design, where a very large energy range leads to a significant variation of the closed orbit with energy. In these cases, it is never a safe to assume that the beam ellipses are centered in the beam pipe. Thus, the algorithm for finding the beam pipe aperture must take into account both the axes of the ellipses and the position of the ellipse centers.

This paper presents an algorithm for computing the magnet aperture under the following assumptions:

- The magnet aperture is circular.
- The magnet must enclose a number of ellipses.
- The ellipse axes are horizontal and vertical.
- The ellipses are all centered vertically.

A circular magnet aperture is the most straightforward shape for a high-field superconducting magnet, so for some cases that is a good approximation. That assumption turns out to make the algorithm particularly simple. The assumption of upright, vertically centered ellipses translates into having no vertical dispersion, no coupling, and ignoring nonlinear effects which would distort the ellipses into other shapes. Many machine designs try to eliminate vertical dispersion and coupling, and mild nonlinear distortions generally leave the beam in a roughly elliptical shape anyhow.

* Work supported by US Department of Energy contract DE–AC02–98CH10886

## ALGORITHM

Imagine a series of ellipses, index by $k$, described by the equations

$$\frac{(x-c_k)^2}{a_k^2} + \frac{y^2}{b_k^2} = 1. \tag{1}$$

The goal of the algorithm is to find a circle which encloses all of these ellipses. For a given center of the circle, it is generally desirable to have the circle of smallest radius. Thus, the circle in question will be tangent to at least one of the ellipses.

The square of the radius of a circle, centered at $(z, 0)$, which is outside of ellipse $k$ and tangent to it, is $R_k^2(z)$, given by

$$R_k^2(z) = \begin{cases} R_{k-}^2(z) & z \leqslant z_{k-} \\ R_{k0}^2(z) & z_{k-} < z < z_{k+} \\ R_{k+}^2(z) & z_{k+} \leqslant z \end{cases} \tag{2}$$

$$R_{k-}^2(z) = (c_k + a_k - z)^2$$

$$R_{k0}^2(z) = b_k^2 \left[ \frac{(c_k - z)^2}{b_k^2 - a_k^2} + 1 \right]$$

$$R_{k+}^2(z) = (z - c_k + a_k)^2$$

$$z_{k-} = \min\left\{ c_k, c_k - \frac{b_k^2 - a_k^2}{a_k} \right\}$$

$$z_{k+} = \max\left\{ c_k, c_k + \frac{b_k^2 - a_k^2}{a_k} \right\}.$$

The goal of the algorithm is to find the function

$$R^2(z) = \max_k \{R_k^2(z)\}. \tag{3}$$

One can then minimize $R^2(z)$ with respect to $z$ if one wishes to find the smallest aperture, or minimize some cost function (e.g., a magnet cost) with respect to $z$.

Define $\mathcal{M}_k$ to be the set of values for $z$ (the horizontal coordinate of the circle's center) for which the circle with smallest radius that is outside all of the ellipses is tangent to ellipse $k$. If multiple ellipses are tangent to that circle, $z \in \mathcal{M}_k$ only if $k$ is the lowest index for which that holds:

$$\mathcal{M}_k = \{z : (\forall j)(R_k(z) > R_j(z) \text{ or } [R_k(z) = R_j(z) \text{ and } k < j])\}. \tag{4}$$

From the definition,

$$\mathcal{M}_k \cap \mathcal{M}_j = \emptyset \quad k \neq j \qquad \bigcup_k \mathcal{M}_k = \mathbb{R}. \tag{5}$$

In software, the $\mathcal{M}_k$ are each stored as a sequence of intervals. Each ellipse is stored as its triplet of values $(a_k, b_k, c_k)$ plus the list of pairs of points describing the intervals in $\mathcal{M}_k$.

A list (henceforth "the list") of ellipses and their corresponding list of intervals is kept by the algorithm. If $\mathcal{M}_k$ becomes empty, it is removed from the list. The algorithm attempts to add a number of ellipses to the list. The first ellipse is added to the list with $\mathcal{M}_1 = \{(-\infty, \infty)\}$. Each subsequent ellipse, call its index $m$, is compared to each ellipse (index $k$) still in the list. Before going through the list of ellipses, $\mathcal{M}_m$ starts out empty. $R_k^2(z) = R_m^2(z)$ at two or fewer values of $z$ (henceforth called the breakpoints; see next section for proof). Furthermore, as $z \to -\infty$, $R_m^2(z) > R_k^2(z)$ if $c_m + a_m > c_k + a_k$, and as $z \to \infty$, $R_m^2(z) > R_k^2(z)$ if $c_m - a_m < c_k - a_k$.

If there are no breakpoints and $R_m^2(z) > R_k^2(z)$, then $\mathcal{M}_m = \mathcal{M}_m \cup \mathcal{M}_k$, and ellipse $k$ is removed from the list. If there are no breakpoints and $R_m^2(z) \leqslant R_k^2(z)$, we are finished processing ellipse $m$, and it is not added to the list. If there are one or more breakpoints, the intervals in $\mathcal{M}_k$ are examined in order, determining whether any of the breakpoints lie inside that interval. For this purpose, it is useful to keep the intervals sorted by their lower (or upper) bounds. If the breakpoint is outside the interval, the interval is completely removed from $\mathcal{M}_k$ and added to $\mathcal{M}_m$ if $R_m^2(z) > R_k^2(z)$ for $z$ in the interval. If one breakpoint lies in the interval, the interval is split in two, the part where $R_m^2(z) > R_k^2(z)$ is added to $\mathcal{M}_m$, and the other part replaces the original interval. If two breakpoints lie in the interval, the part(s) where $R_m^2(z) > R_k^2(z)$ is(are) added to $\mathcal{M}_m$, and the original interval is removed from $\mathcal{M}_k$ and is replaced by the remaining part(s). This procedure guarantees that the conditions in Eq. (5) continue to hold.

In practice, the list of intervals generally consists of fewer than 5 intervals, and so this algorithm is extremely fast. In addition, $R^2(z)$ can be evaluated rapidly by going through the intervals in $\mathcal{M}_k$ for each ellipse in the list, determining whether $z$ lies in that interval, and if it does returning $R_k^2(z)$. One could speed this somewhat by forming a list of intervals and the the ellipse that they correspond to. One could even do a binary search in that full list of intervals. The ability to evaluate $R^2(z)$ is useful for performing minimizations with respect to the placement of the center of the beam pipe.

In most cases, however, one is looking for the minimum value of $R^2(z)$. This occurs either at the boundary between intervals, or inside of a single interval. The minimum of $R_k^2(z)$ always occurs at $c_k$, so the minimum of $R^2(z)$ is found simply by searching determining whether $c_k \in \mathcal{M}_k$ for some $k$, and if not, searching through all the intervals and determining which interval boundary has the smallest value of $R^2(z)$. Note that the derivative of $R^2(z)$ is monotonically increasing (but not continuous), so if the intervals are examined in order, once the value of $R^2(z)$ increases, one knows that one has found that minimum. Figure 1 shows an example where the smallest circle enclosing a group of ellipses is found using this algorithm.

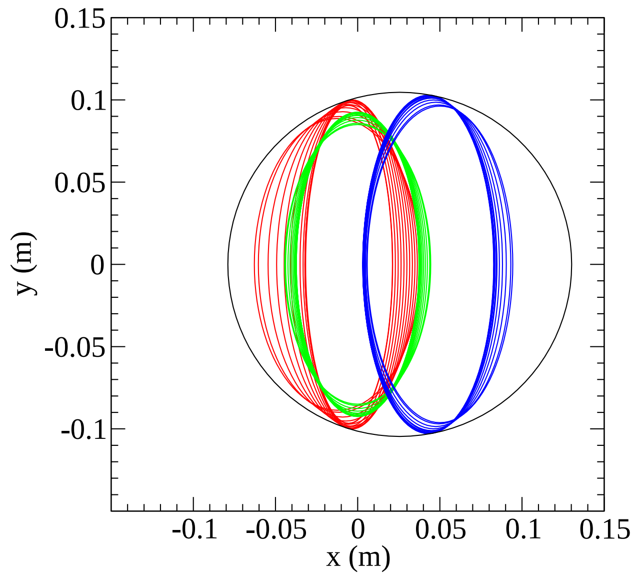



Figure 1: Results of the algorithm finding the smallest circle enclosing all ellipses.

## PROOF OF TWO OR FEWER INTERSECTIONS

We want to prove that the function

$$d(z) = R_2^2(z) - R_1^2(z) \tag{6}$$

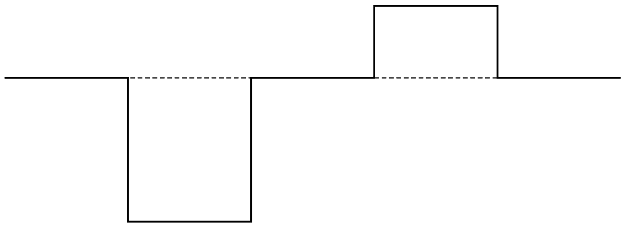
has two or fewer zeros in $z$. If $b_1 < a_1$ and $b_2 < a_2$, then $d(z)$ is piecewise linear, and can only be zero when either $z < c_1$ and $z > c_2$ or $z > c_1$ and $z < c_2$, and thus at only one point.

Let's now take the case where $b_1 > a_1$ and $b_2 > a_2$. It is easy to verify that the first derivative of $d$ is continuous and piecewise linear in this case. Without loss of generality, assume that $z_{1-} < z_{2-}$. If $z_{1-} > z_{2-}$, exchange the 1 and 2 subscripts and change the sign of $z$, and you have the case $z_{1-} < z_{2-}$, and $d(z)$ has the same zeros. $d''(z)$ is a piecewise constant function, and is zero for $z < z_{1-}$ and $z > \max\{z_{1+}, z_{2+}\}$. $d''(z)$ only changes value at the points $z_{1-}$, $z_{1+}$, $z_{2-}$, and $z_{2+}$. Furthermore, $d'(z) = 2(c_1 + a_1 - c_2 - a_2)$ for $z < z_{1-}$ and $d'(z) = 2(c_1 - a_1 - c_2 + a_2)$ for $z > \max\{z_{1+}, z_{2+}\}$.

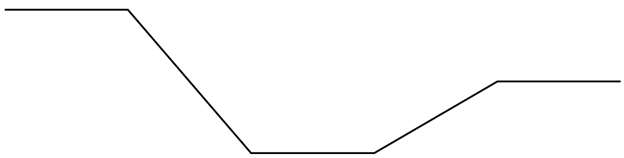
Since $z_{1-} < z_{1+}$, $z_{2-} < z_{2+}$, and $z_{1-} < z_{2-}$, there are only three possible sequences of the points $z_{1-}$, $z_{1+}$, $z_{2-}$, and $z_{2+}$:

1. $(z_{1-}, z_{1+}, z_{2-}, z_{2+})$
2. $(z_{1-}, z_{2-}, z_{1+}, z_{2+})$
3. $(z_{1-}, z_{2-}, z_{2+}, z_{1+})$

For the sequence $(z_{1-}, z_{1+}, z_{2-}, z_{2+})$, $d''(z)$ looks like

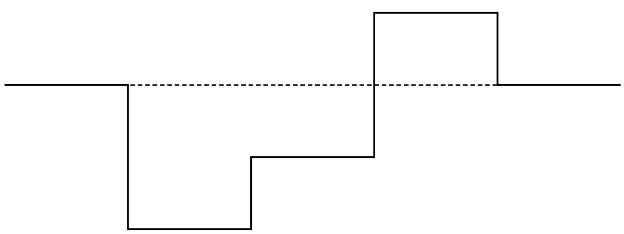


and $d'(z)$ looks like

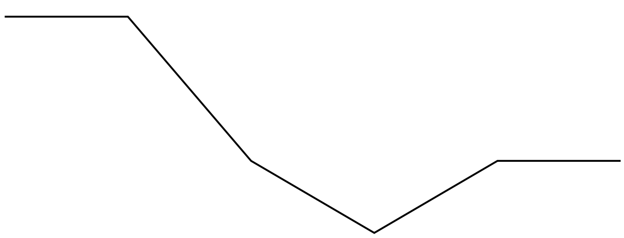


Note that the signs in $d''(z)$ and the signs of the slopes in $d'(z)$ must be as shown. For $d(z)$ to have three zeros, $d'(z)$ would need to have two zeros. This would require that $c_1+a_1-c_2-a_2>0$ and $c_1-a_1-c_2+a_2>0$. These two equations together imply that $c_1-c_2>|a_1-a_2|$. But since $z_{1-}<c_1<z_{1+}$ and $z_{2-}<c_2<z_{2+}$, we know that $c_1<c_2$. Thus, $d(z)$ must have two or fewer zeros in this case.

For the sequence $(z_{1-}, z_{2-}, z_{1+}, z_{2+})$, $d''(z)$ looks like



and $d'(z)$ looks like

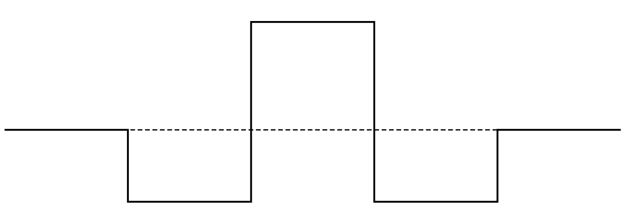


As before, having three zeros of $d(z)$ requires two zeros in $d'(z)$, which in turn requires that $c_1-c_2>|a_1-a_2|$. Since $z_{1-}<z_{2-}$ and $z_{1+}<z_{2+}$,

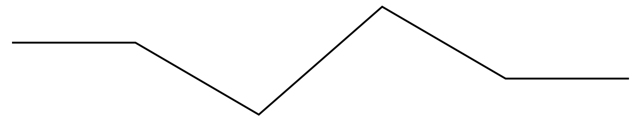
$$c_1-c_2<-\left|\frac{b_1^2-a_1^2}{a_1}-\frac{b_2^2-a_2^2}{a_2}\right|. \tag{7}$$

This is a contradiction, so again $d(z)$ has two or fewer zeros.

For the sequence $(z_{1-}, z_{2-}, z_{2+}, z_{1+})$, $d''(z)$ looks like



and $d'(z)$ looks like



The signs in $d''(z)$ must be as shown except for the central region. However, if $d''(z)$ is not positive in that central region, $d'(z)$ will clearly have at most one zero, and $d(z)$ will therefore have at most two zeros. It appears that $d'(z)$ as shown could have 0, 1, 2, or 3 zeros. If it has 2 or 3 zeros, any zeros occurring in the intervals $(z_{1-}, z_{2-})$ and $(z_{2+}, z_{1+})$ will be maxima. We can evaluate $d(z)$ at those maxima. In the interval $(z_{1-}, z_{2-})$, the zero of $d'(z)$ would occur when

$$z=\frac{b_1^2c_1}{a_1^2}-\frac{b_1^2-a_1^2}{a_1^2}(c_2+a_2), \tag{8}$$

and $d(z)$ takes on the value

$$\frac{b_1^2}{a_1^2}(c_2+a_2-c_1-a_1)(c_2+a_2-c_1+a_1) \tag{9}$$

at that point. If $z$ is in the interval $(z_{1-}, z_{2-})$, $c_2+a_2-c_1-a_1<0$ and $c_2+a_2-c_1+a_1>0$, and thus $d(z)<0$ at that maximum. Similarly, if there is a local maximum in the interval $(z_{2+}, z_{1+})$, $d(z)<0$ at that maximum as well. Since all the local maxima are negative, if $d'(z)$ has 2 zeros, $d(z)$ will have one zero, and if $d'(z)$ has 3 zeros, $d(z)$ will not have any zeros. If $d'(z)$ has 0 or 1 zero, $d(z)$ will have two of fewer zeros.

If $b_1<a_1$ and $b_2>a_2$, there are three cases:

1. $c_1<z_{2-}$: corresponds to $(z_{1-}, z_{1+}, z_{2-}, z_{2+})$.
2. $z_{2-}<c_1<z_{2+}$: corresponds to $(z_{1-}, z_{2-}, z_{2+}, z_{1+})$.
3. $c_1>z_{2+}$: corresponds to $(z_{1-}, z_{1+}, z_{2-}, z_{2+})$.

The proofs are nearly identical to those for the case where $b_1>a_1$ and $b_2>a_2$, and the correspondences given in the list indicate the appropriate case above to use for the proof. The case where $b_2<a_2$ and $b_1>a_1$ clearly is simply an exchange of indices, and thus since there are at most two zeros of $d(z)$ in this case as well.

## CONCLUSIONS

I have described a very efficient algorithm for finding a circular beam pipe which encloses a group of ellipses. The algorithm is used in practice for lattice design optimization, and consumes a negligible amount of computational resources compared to other parts of the optimization. It would be interesting to improve the algorithm to find other shapes of beam pipes (elliptical, for example), but the algorithm would likely be much more complex since it is unlikely that one could use the simple linear interval searches done here.