## OPTIMIZATION OF HEAVY ION FOCUSSING SYSTEMS WITH LARGE NUMBERS OF MAGNETS

 John Steinhoff*
## Abstract

A method is presented for efficiently optimizing heavy ion focussing systems consisting of large numbers of magnets. The method is based on a technique for computing gradients of performance functions in systems where the transfer fuction at each point depends only on the state (trajectory) and control function (magnet strength) at that point. A computer program developed to implement this method is also described. The program can be used to minimize final spot size together with maximum bore radius for systems with both quadrupole and sextupole components. Geometric and chromatic abberations are included, with chromatic ones computed nonperturbatively. The results of some initial studies done for ideal quadrupole systems with large momentum spread ( $£ 1.5 \%$ ) are presented. Also, some results for systems that can simultaneously focus several beams, each with different momentum or charge, are presented.

## Introduction

Final focussing systems for ion beams can be designed perturbatively: Starting with a system of dipoles and quadrupoles designed to focus the beam to first order, geometric and chromatic abberations can be computed for the system and higher order multipole elements added as corrections. If these corrections are small, and if other constraints are not put on the system, then this approach can quickly lead to a focussing system. If, however, the abberations in the initial system are large, or many constraints such as maximum bore radius at various points in the system are imposed, then one set of corrective magnets may not be sufficient. For example, a large beam momentum spread would result in a number of terms proportional to $(\Delta P / P)^{2}$ and higher contributing significantly to the spot size at the target. In this case, the term-by-term corrective procedure may result in a prohibitively large number of additional magnets in the system. At some point it may be better to adjust the entire set of magnets simultaneously.

In this note we describe two optimization schemes that we have been working with that may provide systematic procedures for simultaneously computing the field strengths of a large number of magnets. Rather than attempting to drive a set of abberation terms to zero, these schemes optimize a "performance" function which may, for instance, be

[^0]the total number of particles reaching a 1 mm target, assuming a particular initial phase space distribution. Also, constraints such as beam radius in the magnets can easily be accommodated by adding a "penalty" function, related to the degree of violation of the constraint.

The usefulness of these algorithms in our case depends on whether they converge in a reasonable amount of computing time. Also, these schemes find local optima, not global ones, and their usefulness depends on whether or not there are a large number of local optimal configurations. As an initial attempt to answer some of these questions, we worked with a model system consisting only of ideal quadrupoles. No geometric abberations or space charge effects were included, but large $\Delta \rho / \rho$ with resulting large chromatic abberations was assumed. Numerical results of a series of optimization runs will be given for this case; comparisons of the two optimization schemes will be presented as well as other runs which give a measure of the uniqueness of the solution found and the ability to adjust higher order abherations.

In addition to quadrupoles, we have developed codes to treat systems with dipoles an sextupoles. We describe how these magnets can be included in the optimization schemes. At the time of writing, no new numerical results were available for the se systems.

## Equations of Motion

## 1. Quadrupoles

The equations describing the transverse coordinates ( $x, x^{\prime}, y, y^{\prime}$ ) of a particle after traversing an ideal quadrupole are ${ }^{(1)}$.

$$
\begin{aligned}
\binom{x_{j+1}}{x_{j+1}} & =M\left(u_{j}\right)\binom{x_{j}}{x_{j}^{\prime}} \\
\binom{y_{j+1}}{y_{j+1}^{\prime}} & =M\left(-u u_{j}\right)\binom{y_{j}}{y_{j}^{\prime}} \\
M(u) & =\left(\begin{array}{lr}
c(u) & z_{M} s(u) \\
-\frac{u}{z_{M}} s(u) & c(u)
\end{array}\right) \\
c(u) & =1-u / 2!+u^{2} / 4!-\cdots \\
s(u) & =1-u / 3!+u^{2} / 5!-\cdots \\
U & =G z_{M}^{2} Q / p
\end{aligned}
$$

where $Q / P$ is the charge/momentum, $G$ the field gradient and $\boldsymbol{Z}_{M}$ the quadrupole length.
2. Dipoles

Expanding equation (6) in Brown ${ }^{(2)}$ in transverse coordinates (but not in $\delta a \operatorname{sp/p}$ ) :

$$
\begin{aligned}
& x^{\prime \prime}=\frac{h}{1+\delta}\left\{\delta-(1-\delta) h x-h^{2} x^{2}+\frac{x^{\prime}}{2}(1+2 \delta)+\frac{y^{\prime}}{2}\right\} \\
& y^{\prime \prime}=h x^{\prime} y^{\prime}
\end{aligned}
$$

where $h$ is the field strength of the dipole divided by $P / Q$. Instead of developing a perturbation expansion for the entire system, for our purposes it is convenient to develop an expansion for each individual magnet. In this way the higher order terms resulting from, say, second order terms of one magnet operating on second order terms of an upstream magnet are retained. Although both of our optimization schemes can be used with the exact equations of motion, the form given here is more convenient for our initial study. We then have

$$
\begin{aligned}
& x(s)=x_{A}(s)+x_{B}(s) \\
& y(s)=y_{A}(s)+y_{B}(s)
\end{aligned}
$$

where $S$ is the longitudinal coordinate and

$$
\left.\begin{array}{l}
x_{A}^{\prime \prime}=h \frac{\delta}{1+\delta}-\frac{(1-\delta)}{1+\delta} h^{2} x_{A} \\
y_{A}^{\prime \prime}=0
\end{array} \quad \begin{array}{rl}
x_{B}^{\prime \prime}=-\frac{(1-\delta)}{1+\delta} h^{2} x_{B}-h^{3} \frac{x_{A}^{2}}{1+\delta}+\frac{h(1+\delta \delta)}{2(1+\delta)} x_{A}^{\prime 2} \\
& +\frac{h}{2(1+\delta)} y_{A}^{\prime 2}
\end{array}\right\}
$$

The equations for $X_{A}$ and $y_{A}$ are solved first, and the result used in the solution for $X_{B}$ and $Y_{B}$. The result is

$$
\begin{aligned}
& x_{A}= x_{0}+h^{2} g C(s) \bar{x}_{0}+S(s) x_{0}^{\prime} \\
& x_{B}= h a C(s)+h^{2} f s s(s) \bar{x}_{0} \\
&+\left(f^{\prime}(S(s)-s)+h^{2} f s C(s)\right) x_{0}^{\prime} \\
&+h^{3}\left(c\left((s)+d S^{2}(s)\right) \bar{x}_{0}^{2}+h\left(c^{\prime}\left((s)+d^{\prime} S^{2}(s)\right) x_{0}^{3}\right.\right. \\
&+2 h^{3} d s(s) C(s) \bar{x}_{0} x_{0}^{\prime}+\frac{h C(s)}{2(1+\delta)} y_{0}^{\prime 2} \\
& y_{A}= y_{0}+y_{0}^{\prime} s \\
& y_{B}=h y_{0}^{\prime}\left((s(s)-s) \bar{x}_{0}+C(s) x_{0}^{\prime}\right) \\
& \bar{x}_{0}= x_{0}-\frac{s}{h(1-\delta)} \\
& S(s)= \sin (k s) / k \\
& C(s)=(1-\cos (k s)) / k^{2}
\end{aligned}
$$

$$
\begin{aligned}
& g=-(1-\delta) /(1+\delta) \\
& a=-\delta^{2} /\left((1+\delta)(1-\delta)^{2}\right) \\
& f=-\delta /(1-\delta 2) \\
& f^{\prime}=-\delta /(1-\delta)^{2} \\
& c=-2 \delta^{2} /\left(3(1+\delta)^{2}\right) \\
& d=-\left(3+3 \delta-2 \delta^{2}\right) /\left(6(1+\delta)^{2}\right) \\
& c^{\prime}=-\left(3+3 \delta+2 \delta^{2}\right) /\left(6\left(1-\delta^{2}\right)\right) \\
& d^{\prime}=\left(3+3 \delta-2 \delta^{2}\right) /\left(6\left(1-\delta^{2}\right)\right)
\end{aligned}
$$

where $k=(-g)^{\frac{1}{2}} h \quad$ and the coordinates at the beginning of the magnetare $x_{0}, x_{0}^{\prime}, y_{0}, y_{0}^{\prime \prime}$.

## 3. Sextupoles

The above procedure, when used for the second order sextupole equations

$$
\begin{aligned}
& x^{\prime \prime}=-H\left(x^{2}-y^{2}\right) \\
& y^{\prime \prime}=2 H x y
\end{aligned}
$$

results in

$$
\begin{aligned}
x_{A} & =x_{0}+s x_{0}^{\prime} \\
x_{B} & =-H\left(\frac{s^{2}}{2}\left(x_{0}^{2}-y_{0}^{2}\right)+\frac{s^{3}}{3}\left(x_{0} x_{0}^{\prime}-y_{0} y_{0}^{\prime}\right)+\frac{s^{4}}{4 \cdot 3}\left(x_{0}^{\prime}-y_{0}^{\prime}\right)\right) \\
y_{A} & =y_{0}+s y_{0}^{\prime} \\
y_{B} & =2 H\left(\frac{s^{2}}{2} x_{0} y_{0}+\frac{s^{3}}{3 \cdot 2}\left(x_{0} y_{0}^{\prime}+y_{0} x_{0}^{\prime}\right)+\frac{s^{4}}{4 \cdot 3}\left(x_{0}^{\prime} y_{0}^{\prime}\right)\right) \\
H & =(B \cdot Q) /\left(a_{0}^{2} \cdot p\right) \\
\text { where } \quad B & \text { is the field strength and } a_{p} \text { is the }
\end{aligned}
$$ aperture.

At each step (evaluation of the performance function) in our optimization run we propagate a set of particles with initially prescribed values of $x, x^{\prime}, y, y^{\prime}$ and $\delta$ through the system using these equations.

## Optimization Methods

## 1. Performance Function

The function to be minimized can consist of beam coordinates at the target or in the magnets. We have used

$$
P^{(k)}=\sum W_{i}\left\{\left(x_{i}(N+1)\right)^{K}+\left(y_{i}(N+1)\right)^{K}\right\}
$$

where the $\mathcal{W}_{i}$ are weighting functions representing the initial phase space density for particles near particle $i$, and $\chi_{i}(N+1)$ and $y_{i}(N+1)$ are the coordinates at the target. We have used values of 2 and 4 for $K$. For shaping the beam, we have added the function

$$
\underline{P}^{B}=\sum_{J} a_{s} \sum_{i} W_{i}\left\{\left(x_{i}(J)\right)^{2}+\left(y_{i}(J)\right)^{2}\right\}
$$

where $J$ is the magnet number. By adjusting the $a_{5}$, we can decrease the beam size at various points in the system (with somewhat of an increase in spot size, for a given magnet configuration).

[^1]
## 2. Random Search ${ }^{(3)}$

This optimization method involves first choosing an initial set of strengths $\left\{u_{j}\right\}$ and computing the coordinates $\left\{x_{i}(J), \gamma_{i}(J)\right\}$ and $p=P^{(k)}+p^{B}$. The $u_{j} ' s$ are then perturbed; $u_{j}^{n e w}=u_{j}+v a r_{j} \times R_{j}$, where the $R_{j}$ are random numbers with zero mean and unit variance and the varj are constants chosen initially to be $\sim 10^{-2}\left|u_{j}\right|$. A new value of $P$ is then computed. If the new $\mathcal{P}$ is smaller than the old, the old values of $P$ and $\left\{u_{j}\right\}$ are replaced by the new ones. Otherwise, the old values are kept. The calculation is then repeated.

As $P$ approaches a minimum and the $u_{j}$ approach their optimal values, we reach a point at which $\mid u_{j}-u_{j}$ opt $\mid \ll$ varj. At this point the probability of finding a smaller $P$ becomes very small and the varj must be decreased. We obtained good results by dividing the var; by 2 whenever there were 50 consecutive runs with no decrease in $P$.

The main advantages of the random search are that it is trivial to program and gradients do not have to be computed. Also, "hard" constraints (inequalities) can easily be included.
3. Gradient Search ${ }^{(4)}$

To minimize $\boldsymbol{P}$, a sequence of one dimensional searches are done in the $N$ dimensional $u_{j}$ space. Each search is in a direction determined by the present and past gradient of $P$ :

$$
u_{j}^{\text {dew }}=u_{j}+\alpha\left(-\frac{\partial P}{\partial u_{j}}+\beta\left(u_{j}-u_{j}^{\text {o/d }}\right)\right)
$$

where

$$
\beta=\left(\sum_{1}^{N}\left(\frac{\partial P}{\partial u_{j}}\right)^{2}\right) /\left(\sum_{1}^{N}\left(\frac{\partial P}{\partial u_{j}}\right)^{\text {old }}\right)
$$

As a result of each search a value of $\alpha$ is determined that minimizes $\mathcal{P}\left(\left\{u_{3}\right\}\right)$ in one direction. The starting point for each search is the minimum point found in the previous search. In each step of the computation, the gradient terms must be computed, as well as a number of $P$ values needed to find a one dimensional minimum. For general functions of $N$ variables the gradient computation can take as much computing time as $N$ function ( $P$ ) evaluations and, for large $N$, the gradient part of the optimization can increase the total time by almost a factor of $N$. For "trajectory" equations, however, where $P$ has the above dependence on $\chi(J)$ and $Y(J)$, and where there is a recurrence relation which gives the coordinates at $J+1$ in terms of the coordinates at $J$ and $U_{J}$, we can compute the $N$ gradient terms in only several times the computing time required to compute $\mathcal{P}$, regardless of $N$.

For $p^{(k)}$, we define the vector

$\begin{aligned} & \text { where } \\ & \omega_{0}^{i}\end{aligned}=x_{i}, \omega_{1}^{i}=x_{i}^{\prime}, \omega_{2}^{i}=y_{i}, \omega_{3}^{i}=y_{i}^{\prime}$.
Then,

$$
\frac{\partial P^{(k)}}{\partial u_{J}}=\sum_{i} \sum_{v} P u_{v}^{i}(J+1) \frac{\partial \omega_{v}^{i}(J+1)}{\partial u_{J}}
$$

and $P u_{v}^{i}(J)$ obeys the recurrence relation

$$
P u_{v}^{i}(J)=\sum_{\mu} P u_{\mu}^{i}(J+1) \frac{\partial \omega_{\mu k}^{i}(J+1)}{\partial \omega_{j}^{i}(J)}
$$

Thus, after a forward sweep, in which the $\omega_{j}^{i}(J)$ 's are computed, the $N$ gradient terms can be developed in a single backward sweep.

This method is applied very easily to ideal $q$ adrupoles since the model is linear, as discussed in some detail in (1). For dipoles, if the field and length are kept fixed (as we are doing in our initial study), the model derived above can be used to compute $\partial \omega_{\mu}^{i}(J+1) / \partial \omega_{\nu}^{i}(\sigma)$ very easily since the coefficients in the equations can be prestored. The sextupole model can easily be used for both $\partial \omega_{\mu}^{i}(\tau+1) / \partial \omega_{v}^{i}(J)$ and $\partial \omega_{v}^{i}(J+1)$ $/ \partial u_{j}$ calculations (we are varying the sextupole strength).

Similar recurrence relations can be written for $\mathcal{P}^{B}$ : Define

$$
P B_{v}^{i}(J)=2 W_{i} \sum_{L=J}^{N}\left(\omega_{0}^{i}(L) \frac{\partial \omega_{0}^{i}(L)}{\partial \omega_{\nu}^{i}(J)}+\omega_{2}^{i}(L) \frac{\partial \omega_{\nu}^{i}(L)}{\partial \omega_{\nu}^{i}(J)}\right)
$$

then,

$$
\frac{\partial P^{B}}{\partial u_{J}}=\sum_{i} \sum_{V} P B_{V}^{i}(J+1) \frac{\partial \omega_{v}^{i}(J+1)}{\partial u_{J}}
$$



## 1. Convergence Speed

Optimization runs were made for an ideal quadrupole system with 16 magnets and large ( $\pm 1.5 \%$ ) momentum spread using both the random and gradient search algorithms. The magnets were .5 meters long with a 1 meter drift space in between, and the upstream beam was taken to be an ellipse in phase space with axes

$$
\begin{aligned}
& x_{i}(0)=y_{i}(0)=5 \mathrm{~cm} . \\
& x_{i}^{\prime}(0)=y_{i}^{\prime}(0)=.1 \mathrm{mr} .
\end{aligned}
$$

The system was started in an FoDO configuration (alternating fields), and the function $P^{(z)}$ was minimized. Spot size vs. computing time (for a Data General NOVA 800 minicomputer) is plotted in Fig. 1 for the two methods. It can be seen that the gradient method is about eight times faster than the random search method. Also, when the random search method appeared to converge, the spot size was still about $25 \%$ larger than the result of
the gradient scheme (which had converged to a point with very small gradients).

## 2. Uniqueness of Solution

In almost every case run with the gradient scheme, the final magnet configuration consisted of a triplet at the end preceeded by a 5-20 meter "drift space" where the strengths had been driven to zero. In Ref.(1) we present the final 8 magnet strengths for a 28 magnet system. The initial configuration was an FoDo lattice with $|G|$ equal to 20 Teslas/meter. The random search scheme did not find these configurations in a reasonable amount of time.


Fig. 1 Reduction of Focal Spot Size vs. Computing Time for Random and Gradient Search Algorithms.

## 3. Importance of Higher Order Abberations

The function $P$ was defined to be the sum of the squares of the spot sizes for particle sets with momenta $P=P_{0},(1 \pm .005) P_{0}$, ( $1 \pm .01$ ) $p_{0}$ and ( $1 \pm .015$ ) $p_{0}$. This function would be useful if the beam had a step function momentum dependence from (1-.015) $\rho_{0}$
to $(1+.015) p_{0}$. The target spot size for a 28 magnet system is presented in Fig. 2 as a function of $\Delta=P / \rho_{0}-1$. The importance of the higher order terms $\left((\Delta \rho / p)^{n}\right)$ in minimizing the spot size near the extreme momenta can be seen.

We also found another nonlinear effect: We were able to simultaneously focus, with the same system, two separate beams each with a $\pm .25 \%$ momentum spread but differing in $P_{0}$ by $20-40 \%$. The spot sizes found were about 3 mm . for both beams, as opposed to $1.5-2 \mathrm{~mm}$. for a single beam. This might be useful for "telescoping" beams (5) with different momenta, since both beams could have the same charge ( $\Phi / \rho_{0}$ is not constrained to be the same).


Fig. 2 Target Spot Size vs. Momentum Deviation: 28 Magnet System Optimized for $\pm 1.5 \%$ Momentum Spread.

## Conclusion

We have found that gradient optimization schemes can be efficient and versatile for optimizing spot sizes in simple ion focussing systems. Although higher order terms could be controlled, the effect of first order chromatic abberations, which cannot be made to vanish in a quadrupole system, contributed to a large spot size (3-5 mm) for large $\Delta p / p$ ( $1.5 \%$ ). Towards reducing these terms, we are currently working with a system with dipoles and sextupoles designed to cancel first order chromatic abberations.

## Acknowledgements

The author is grateful to Dr . S. Fenster and Dr. E. Co 1 ton of Argonne National Laboratory for many helpful discussions.

## Footnotes and References

1. J. Steinhoff, "Quadrupole Systems for Focussing Ion Beams With Large Momentum Spread," Proceedings of the 1978 Heavy Ion Fusion Workshop, Argonne National Laboratory, Sept. 19-26, 1978, to be published.
2. K. L. Brown, "A First- and Second-Order Matrix Theory for the Design of Beam Transport Systems and Charged Particle Spectrometers," SLAC-75, Rev. 3, Aug. 1972.
3. M. A. Schumer, K. Steiglitz, "Adaptive Step Size Random Search, " IEEE Transactions on Automatic Control, Vo1. AC-13, No. 3, p. 270-276, June 1968. Improvements to the basic method are also discussed.
4. A. E. Bryson Jr., Yu-Chi Ho, Applied Optimal Control, Blaisdell Publishing Co., Waltham, Mass., 1969.
5. R. J. Burke, "Telescoping Beams for Heavy Ion Fusion," to be published, this conference.

[^0]:    *Grumman Aerospace Corporation, Research Department, Bethpage, New York 11714

[^1]:    TIn the curvilinear system defined in Ref. (2).

