

Semesterarbeit

Wilke Grosche
ETH Zurich

Supervisor: Solange Emmenegger

July 2020

Abstract

The controlled formation of magnetic fields is of relevance especially in the shielding of magnetic fields. This can be useful in experiments that require a highly homogeneous magnetic field for the purposes of heightened precision and in areas such as medicine, where shielding can be used in conjunction with MRIs. The motivation for this work is the field stabilisation for the neutron electric dipole measurement being conducted at the PSI. A refinement on the concept of Helmholtz coils allows for a controlled formation of a magnetic field within a large ($125m^3$) area to a field homogeneity of $\pm 1\%$. This potentially has applications in other experiments requiring predefined fields and also in medicine where it may be useful in conjunction with MRIs. The goal is to use this system as an active magnetic shield to reduce and homogenise fields from outside the system. In this work the focus is on optimising the points at which the field is measured before these values are fed back into the feedback algorithm. This allows the currents for field generation to be calculated. This optimisation is done via a gradient descent using the simultaneous perturbation stochastic approximation gradient descent. The results indicate a marginal performance increase over the work of Tim Rothlisberger done in 2018. Where his discretised Monte Carlo approach reached condition numbers of 1.7. With the randomly initialised gradient descent, condition numbers of 1.4 were reached. The model is altered in COMSOL to include a mu metal cube used in the experiment.

Contents

1	Introduction	3
1.1	nEDM Measurement	3
1.2	Magnetic Shielding	3
1.3	Feedback Algorithm	4
2	Method	4
2.1	Field Generation:	4
2.2	COMSOL Modelling	5
2.3	Optimisation:	5
2.4	Comparison of Optimisation Methods:	6
2.4.1	Nelder Mead and the Genetic Algorithm	6
2.4.2	Approximate Gradient Methods:	6
2.4.3	SPSA	6
2.5	Random Search	7
3	Results and Discussion	7
3.1	Convergence of the Algorithms	7
3.2	Number of Sensors	7
3.3	Refinements on Tim's Points	7
3.4	Refinements on Symmetrical Starting Points	7
3.5	Random Structure Search	8
3.6	Configuration Stability	8
3.7	Scaling	8
3.7.1	Matrix Scaling	8
3.8	PSI Measurement	8
3.9	Validity of the Final Positions	9
3.10	Stability of the Final Positions	9

1 Introduction

The standard model is still unable to explain the matter antimatter asymmetry in the universe. Whilst the established CP violations are capable of explaining the disparity to some extent (10^{-18}) in reality we observe 10^{-10} (??). This means that we require further sources of CP violation in order to explain the observed physics. One possible source of such violation is the neutron electric dipole moment (nEDM).

Since the Hamiltonian of a non-relativistic particle in an electromagnetic field is given by:

$$H = -\mu\vec{B} \cdot \frac{\vec{S}}{S} - d\vec{E} \cdot \frac{\vec{S}}{S} = -2(\mu\vec{B} + d\vec{E}) \cdot \vec{S} \quad (1)$$

Where μ and d are the magnetic moment and electric moment respectively. The second equality holds for spin $S = 1/2$ particles. A non-zero electric dipole moment d causes the Hamiltonian to be both P and T violating. This makes the nEDM a good probe for physics beyond the standard model. Here we will focus on the experimental setup of the nEDM measurement at the Paul Scherrer Institute (PSI). The work will build upon that of Michal Rawlik et al. (??).

1.1 nEDM Measurement

The measurement of the nEDM at the PSI relies on the Ramssey method. This works by using the differences in the aligned and unaligned Larmor precession frequencies. Looking at equation 1 we can see that the energy of the particle depends on the alignments of the E and B fields relative to the spin state. Larmor precession occurs when the magnetic field is not aligned with the particle spin. The frequency of this precession is given by the energy difference between the spin up/down states. From equation 1 we find that the energies are given by:

$$h\nu_{\uparrow\uparrow} = -2(\mu B_{\uparrow\uparrow} + d_n E_{\uparrow\uparrow}) \quad (2)$$

For the aligned spin. And:

$$h\nu_{\uparrow\downarrow} = -2(\mu B_{\uparrow\downarrow} - d_n E_{\uparrow\downarrow}) \quad (3)$$

For the anti-aligned spin. The energy difference between the two states is what matters for the dipole measurement. Subtracting equation ?? from ?? and solving for d_n gives the electric dipole moment in terms of the Larmor frequencies (equation 4).

$$d_n = \frac{h(\nu_{\uparrow\downarrow} - \nu_{\uparrow\uparrow}) - 2\mu(B_{\uparrow\uparrow} - B_{\uparrow\downarrow})}{2(E_{\uparrow\uparrow} + E_{\uparrow\downarrow})} \quad (4)$$

This measurement of the nEDM is often limited in its precision by the term $(B_{\uparrow\uparrow} - B_{\uparrow\downarrow})$. This is because this term is susceptible to fluctuations in the magnetic field. These fluctuations are harder to shield than those in the electric field.

1.2 Magnetic Shielding

The established way to shielding magnetic fields is through the use of passive magnetic shielding in the form of materials with a high magnetic permeability. For this purpose mu-metals were developed. These are nickel-iron soft ferromagnetic alloys with a permeability of order 10^4 . The high permeability allows the magnetic fields to follow a path within the material instead of penetrating into the enclosed area. However, for large fluctuations, fields can still penetrate into the shield. (maybe include COM-SOL sim image of field penetrating into the shield.) For this reason active magnetic shielding was proposed. Building on the established concept of the Helmholtz coil, the idea is to create a coil system capable of generating a desired field within an expanded region. Whilst the Helmholtz coil is capable of generating a desired field at given points the regional variation is large. A more complicated coil system proposed by Michal Rawlik (??) is able to create a stable field in an expanded region ($125m^3$ with a field homogeneity of $\pm 1\%$). The coil system being developed at the PSI is shown in figure ??.

The coil system is constructed from tiles of wires. The tiles together make up 3 current loops for each direction. 0.2A, 1A and 5A loops exist. These loops are shown in figure ???. This is done to simplify the complexity of the current decompositions whilst still being able to achieve the desired homogeneity of $\pm 1\%$ within the measurement zone.

The idea is to use this coil system in a manner similar to noise cancelling headphones, to generate a field that exactly compensates the background field at each point in the measurement zone. In order to have this happen in real time a feedback algorithm was developed [?].

1.3 Feedback Algorithm

In order to facilitate the real time field compensation, the field has to be measured at fixed points. Since the field generated by the coil system is given by the Biot Savart law, a static matrix can be calculated which links the field at each measurement point to the current required to generate a compensating field.

The matrix is calculated by running a linear regression over the field as a function of the current at the different measurement points in the measurement zone.

The field generated by the shield is given by $\vec{B} = \mathbb{M}\vec{I} + \vec{B}_{offset}$. Here the currents through the generating coils are gathered into a vector in $\mathbb{R}^{n_{coils}}$ where n_{coils} is the number of coils. The matrix \mathbb{M} is in $\mathbb{R}^{(n_{coils} \times 3n_{sensors})}$ and contains the gradients given by the linear regression of the field for given currents.

$$M_{i,j} = \begin{pmatrix} P1_{x,1} & P2_{x,1} & \cdots & P8_{x,1} \\ P1_{y,1} & P2_{y,1} & \cdots & P8_{y,1} \\ \vdots & \vdots & \ddots & \vdots \\ P1_{z,n} & P2_{z,n} & \cdots & P8_{z,n} \end{pmatrix} \quad (5)$$

This feedback matrix is static in the external fields as it only depends on the positions of the measurement points at which it is calculated. The danger with this matrix is that, with certain flux gate combinations, the matrix can become ill conditioned. This means that the matrix has a very high condition number. This should be avoided as a high condition number is an indicator for the fact that the matrix is coupling asymmetrically to the different coils.

Avoiding configurations with poor conditioning is the point of this work. The goal is to minimise the condition number of the feedback matrix for the sensor positions chosen.

2 Method

In order to determine the best positions for the sensors the system is be simulated and a gradient descent on the condition number is performed. For this to work the fields generated by the coil system have to be calculated.

2.1 Field Generation:

The fields generated by the SFC system can be modelled computationally via the use of the Biot Savart law:

$$B(\vec{r}) = \frac{\mu_0}{4\pi} \int_C \frac{Id\vec{l} \times \vec{r}'}{|\vec{r}'|^3} \quad (6)$$

For this purpose Michal Rawlik developed the Coils Package for Julia ([?]). This allows the generation of a system comprising line elements which carry a current. The coil systems can be read in and the currents required to generate a desired field are calculated. To decrease the risk of very high condition number matrices \mathbb{M} due to high order fields, the field is decomposed into Cartesian harmonic polynomials:

$$\vec{B}(\vec{r}) = \sum_n H_n \vec{P}_n(\vec{r}) \approx \sum_{n=1}^8 H_n \vec{P}_n \quad (7)$$

In order to achieve the desired homogeneity in the field of $\pm 1\%$ in the fiducial zone we only require fields up to those of linear order (??).

$$P_1(\vec{r}) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, P_2(\vec{r}) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, P_3(\vec{r}) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (8)$$

$$P_4(\vec{r}) = \begin{pmatrix} x \\ 0 \\ -z \end{pmatrix}, P_5(\vec{r}) = \begin{pmatrix} y \\ x \\ 0 \end{pmatrix}, P_6(\vec{r}) = \begin{pmatrix} 0 \\ y \\ -z \end{pmatrix}, P_7(\vec{r}) = \begin{pmatrix} z \\ 0 \\ x \end{pmatrix}, P_8(\vec{r}) = \begin{pmatrix} 0 \\ z \\ y \end{pmatrix} \quad (9)$$

Each field (P_1 to P_8) is generated independently and the matrix 5 can be determined.

2.2 COMSOL Modelling

The fields generated in Julia are for a system without a mu-metal cube. For this reason a further simulation step is required. The fields generated in Julia are read into COMSOL and linearly interpolated. The fields are set to 0 outside of the modelled domain. This can be done because the fields outside of the SFC have no bearing on the flux gate positions.

–image of the fields read into comsol x,y,z components of Bx –

The mu-metal shield is modelled in the form of a hollow cube surrounded by a spherical region filled with air. The boundary conditions are enforced via the use of an infinite element domain which sets the field as 0 at infinity.

–image of the modelled components –

The field generated by COMSOL for the maximal background field is read out at the Lagrange points determined by the program.

The resulting field configuration is read into Julia. In order for the field to be continuous the matrix is linearly interpolated using scipy's NDInterpolator in a PyCall wrapper[?]. The optimisation is applied to the continuous field in order to find the flux gate positions with minimal condition number in the coupling matrix.

2.3 Optimisation:

The feedback algorithm requires the magnetic field to be measured at given points in the region of interest in order to work. The danger with placing these measurement points randomly is that this may lead to a high condition number of the feedback matrix.

The condition number is a measure for how sensitive ([?]) the output of a function is to small changes in the input. In the case of a matrix equation the condition number of a matrix is given by:

$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \quad (10)$$

Where $\sigma_{max}(A)$ and $\sigma_{min}(A)$ are the maximal and minimal singular values of A respectively.

A high condition number in the feedback matrix corresponds to unequal coupling of the fields generated by the different coils at one or more of the points at which sensors are placed. This is a problem because it could result in the generated fields being dominated by the contribution due to one oversensitive measurement. This would lead to high instability in the generated field and an unreliable field compensation. As a result the condition number of the feedback matrix is to be minimised.

This optimisation takes place in a $24 = 3 \times n_{sensors}$ dimensional space. Since the calculated fields are read out on discrete points and the condition number itself is not easily differentiable a simple gradient descent is not possible. This lead to the consideration of the following techniques:

Genetic algorithm
Nelder Mead

2.4 Comparison of Optimisation Methods:

2.4.1 Nelder Mead and the Genetic Algorithm

Both the Nelder Mead optimisation method and the Genetic algorithm were discarded as options for this optimisation. The Nelder Mead algorithm was discarded due to its poor convergence in higher dimensions (d10)[?]. The genetic algorithm was discarded because it requires a 'genetic representation' of the solution domain in order for the evolution of the algorithm to take place. There were difficulties in defining this for this problem.

2.4.2 Approximate Gradient Methods:

The approximate gradient methods are of use in this situation because the loss function of the problem is not easily differentiable. This means that, in order to use gradient descent, the gradient must be estimated at each point. There are many established ways of doing this. The initial implementation of the gradient descent was with a static finite central differences approximation for the gradient. This was the most intuitive approach for estimating the gradient.

$$\nabla_i(\vec{x}) = \frac{f(|\vec{x}|(\hat{x} + \hat{x}_i\epsilon) - f(|\vec{x}|(\hat{x} - \hat{x}_i\epsilon))}{2|\vec{x}|\sqrt{\epsilon}} \quad (11)$$

Where \hat{x}_i is the unit vector in direction i and ϵ is the machine ϵ supplied by the system. This method is more accurate when it comes to approximating the gradient than forward or backwards differences gradient approximations [?].

This method is guaranteed to converge to a local minimum when coupled with the step size given by the Barzilai Borwein method [?]:

$$\gamma_n^{BB} = \frac{|(\vec{x}_n - \vec{x}_{n-1})^T [\nabla F(\vec{x}_n) - \nabla F(\vec{x}_{n-1})]|}{\|\nabla F(\vec{x}_n) - \nabla F(\vec{x}_{n-1})\|^2} \quad (12)$$

This quasi Newton method for calculating the step size offers guaranteed convergence without the increased computational requirement for small constant step sizes.

However, this approach has two major problems. Firstly, the algorithm is computationally expensive as it requires $2 \times d$ function evaluations to approximate the gradient, where d is the dimension of the problem. The second problem is the fact that, whilst this algorithm is guaranteed to converge, convergence is only guaranteed to a local minimum. It is impossible for the algorithm to escape local minima.

A more computationally efficient method of approximate gradient descent is the simultaneous perturbation stochastic approximation (SPSA) algorithm[?].

2.4.3 SPSA

In high dimensional optimisation problems the simultaneous perturbation stochastic approximation (SPSA) method can lead to large decreases in the computational cost of the descent. This is because, instead of calculating the gradient in each direction individually as in equation 11 the gradient is calculated in a random direction:

$$\nabla_{k,i}(\vec{x}_k) = \frac{f(\vec{x} + c_k \vec{\Delta}_k) - f(\vec{x} - c_k \vec{\Delta}_k)}{2c_k \Delta_{k,i}} \quad (13)$$

Where $\Delta_k =$

This leads to 2 instead of $2 \times n$ function calls for the gradient calculation. In our case this results in a decrease in the computation time of factor ≈ 20 . This was measured by benchmarking with the julia package BenchmarkTools [?].

	Gradient SPSA	Gradient BB	Step Size SPSA	Step size BB	SPSA Steps	BB Steps
Min Time	976.301 μ s	21.302 ms	140.349 ns	46.077 ms	1.451 ms	43.740 ms
Median Time	1.297 ms	26.591 ms	148.694 ns	49.123 ms	1.770 ms	50.147 ms
Mean Time	1.623 ms	31.850 ms	171.692 ns	55.272 ms	2.529 ms	57.252 ms
Max Time	63.520 ms	133.480 ms	2.318 μ s	108.894 ms	63.135 ms	131.147 ms
Memory Alloc.	76.81 KiB	1.77 MiB	48 B	3.54 MiB	131.17 KiB	3.54 MiB

Table 1: It should be known that this benchmarking is tied to computational performance and only the relative times are relevant. This data suggests that the SPSA descent is of the order 20 times faster than the BB method.

The reason for the significant reduction in calculation cost for the step size is that for the SPSA algorithm convergence is achieved reliably with a decreasing step size given by:

$$a_k = \frac{a}{(k + A)^\alpha} \quad (14)$$

SPSA has an additional benefit over the BBM. It is able to leave local minima with the correct choice of the perturbation coefficient c_k . Another problem with the optimisation is the question of what to do with runaway descents. If all fields couple equally weakly then the result may minimise the condition number whilst having no actual information about the fields inside the measurement zone. To fix this problem the search domain was constrained to a region within a fixed distance from the mu-metal shield. The constraint was applied by setting wandering coordinates to the edge of this domain.

2.5 Random Search

As mentioned in section 2.4.3 this method guarantees convergence to a local minimum. Unfortunately, from the results it becomes clear that this problem is non convex and the local minima are not guaranteed to come close to the global minimum. To achieve a good approximation of the global minimum, an adapted method of the ab initio random structure search (AIRSS)[?], used to find the minimum energy solution of solid state systems, is used.

The optimisation algorithm is randomly initialised and the descent is performed separately for each optimisation. The lowest result is taken as an approximation of the global minimum.

3 Results and Discussion

3.1 Convergence of the Algorithms

3.2 Number of Sensors

In his work Tim determined the ideal number of sensors for the optimisation to be 8 [?]. This is verified by the optimisation run on the BabySFC simulation. As shown in table 2 the lowest condition numbers were achieved for 8 sensors in both the purely Julia simulated setup and that generated in COMSOL. These values were obtained by running the SPSA descent over 1000 iterations with 10 random initialisations and choosing the lowest result achieved.

3.3 Refinements on Tim’s Points

A descent was initialised using the points calculated by Tim Roethlisberger [?] to be the optimal configuration. These result in a condition number of — for the Julia setup and — for the COMSOL variant. This is a mild improvement on the condition number achieved by the initial points.

3.4 Refinements on Symmetrical Starting Points

Another option for initialisation is to start with points located at the corners of the search domain. This should represent the symmetries inherent in the setup. With this initialisation the condition

BabySFC	Lowest Condition	Achieved (10 Runs)
Number of Sensors	Julia	COMSOL
4	2.206	2.564
5	2.543	2.125
6	2.116	2.15
7	2.038	2.183
8	1.821	2.117
9	1.966	2.185
10	2.258	2.419
11	2.148	2.593
12	1.944	2.871
13	1.85	2.186
14	1.966	2.697
15	2.022	2.765
16	1.962	3.116
17	1.982	3.562
18	1.927	3.266
19	1.908	3.51
20	1.943	3.443

Table 2: The lowest condition numbers achieved over 10 runs of the SPSA algorithm for varying numbers of flux gates in the BabySFC configuration. In both the Julia and the COMSOL variant the lowest condition number was achieved for 8 flux gates.

number achieved is —...

3.5 Random Structure Search

The results obtained with the SPSA algorithm in conjunction with the AIRSS yielded the lowest condition numbers. These condition numbers did not compare favourably with those obtained through the Monte-Carlo Method suggested in —ref Tim SemesterProject-.

3.6 Configuration Stability

3.7 Scaling

One problem with the method arises when different goal field coupling constants are used for the homogeneous and gradient fields respectively. This unequal coupling results in large condition numbers (—) as the gradient fields couple far weaker (factor 10) than the homogeneous fields. In order to combat this, an attempt was made at scaling the relevant matrix components.

3.7.1 Matrix Scaling

Here the attempted fix was to scale the gradient components up by a factor of $\frac{m}{m_g}$. This yielded poor results when tested with the COMSOL fields for the BabySFC as shown in table ??

Table: scaling attempts:

Comparison: 50:50, 50:20 group scaled, 50:5 group scaled, 50:20 fine scaled, 50:5 fine scaled

For babysfc julia and comsol.

Comparison of the end points of a scaled descent from Tims Poi and the corner poi. Distances to the ideal positions found by the regular 50:50 descent.

3.8 PSI Measurement

The numbers resulting from the optimisation on the PSI setup are in the region of —. This increase over what is expected is likely due to the fact that the currents supplied to the optimisation are for

a system with $m = 50\mu\text{T}$ and $m_g = 20\mu\text{T}$. Since the scaling attempts detailed in section 3.7 did not solve the problem, other approaches are required. One possibility is to rerun the calculation of the currents for the PSI setup with the equal values for m and m_g and apply the standard descent. Another possibility is to find a method of scaling the components that results in the proper field values.

3.9 Validity of the Final Positions

3.10 Stability of the Final Positions