Paolo Fabbri

Task1

A) Write a short answer

In the context of an MEG experiment, how can the following sources of noise be most efficiently mitigated?
i) Earth's magnetic field
ii) Power line interference
iii) Intrinsic noise of SQUID sensors
iv) Background brain activity
Give a unique answer for each problem.

Responses:

i) A variety of techniques have been used to reduce the ambient magnetic fields for MEG measurements including:
(1) locating the MEG instrument and subject in a magnetically shielded room (MSR).
(2) using gradiometric magnetic field sensors coupled to SQUIDs.
(3) active compensation of magnetic noise with field coils.
(4) measuring or estimating the ambient noise fields and digitally subtracting the noise from sensor measuring brain activity.
(5) averaging MEG data from numerous stimuli. In addition to these hardware-based approaches, a broad array of additional post-processing software algorithms has been developed to reduce noise in MEG data such as simple filters (low-pass, high-pass, band-stop, etc).
(pdf)

ii) Most efficiently mitigated by Signal Space Separation (SSS), which mathematically separates external magnetic interference from internal brain-generated signals. Otherwise, as we know that the typycal frequency of the main power line devices is around 50-60 Hz, we can use a band-stop filter for these specific frequencies and their harmonics.

iii) The intrinsic noise of SQUID sensors is mainly mitigated by using high-sensitivity SQUIDs operated under cryogenic conditions and by optimizing Signal Space Projection (SSP), which projects out sensor-specific noise patterns and reduces noise originating from the sensors themselves.

iv) Since background brain activity originates from the brain itself, one of the primary techniques to mitigate it is signal averaging across multiple trials. By averaging numerous responses the random noise tends toward a mean of zero.

B) Calculate and explain

A MEG experiment is conducted to measure a stimulus-evoked response whose amplitude is known to be $100\ \mathrm{fT}$ (in one channel). The measurement system has sensors with intrinsic white noise with amplitude $4.0\ \mathrm{fT}/\sqrt{\mathrm{Hz}}$, and the background brain activity ("brain noise") is assumed to be zero-mean noise with power spectral density $S(f) = 3000/f\ \mathrm{fT}^2/\mathrm{Hz}$. The frequency band of interest is $1\text{--}40\ \mathrm{Hz}$. All noise sources can be assumed to be uncorrelated. Signal-to-noise ratio (SNR) is defined as the ratio between the signal mean and the noise standard deviation: $SNR = \mu/\sigma$.

(i) In this task, the standard deviation of the noise σ_{noise} equals the noise amplitude, i.e. its RMS value. Show why $\sigma_{noise} = RMS_{noise}$.

(ii) Averaging multiple trials can improve the signal-to-noise ratio. How many trials must be averaged to reach $SNR = 8.0$?

Hint: With repeated measurements, we sample multiple times from a distribution that describes the noise. What is the relationship between the standard error and the standard deviation?

(iii) Does the intrinsic sensor noise significantly reduce the signal-to-noise ratio in this case?

The evoked response can be assumed to be fully resilient to stimulus repetition, i.e. its amplitude remains constant across trials.

Responses:

(i) We know that noise has mean equal to 0 ($\mu = 0$):

$$\mu = E[X] = 0$$

Knowing this allows us to compute the value of the standard deviation and the MRS value.

$$\sigma = E[(X-\mu)2] >> \sigma = \sqrt{\mathbb{E}[X^2]}$$

Then, the mathematical formula for calculating the Root Mean Square is defined as the square root of the mean of the squares of a set of values.

$$RMS = E[x2]$$

So essentially we could assess:

$$\sigma = RMS$$

ii) While the signal amplitude remains constant across trials, the standard error of the noise decreases as more trials are averaged. It is calculated as

$$\sigma_{mean} = \frac{\sigma}{\sqrt{N}}$$

Now we could rewrite our SNR formula:

$$SNR = \mu/\sigma >> SNR = \mu/\frac{\sigma}{\sqrt{N}}$$

As we want to find how many trials must be averaged to reach a $SNR = 8$, we need to substitute the values in the formula and find our unknown variable N (number of trials).

First, we calculate the total noise standard deviation, which is the result of combining the two noise sources, as the are uncorrelated.
$\sigma_{measurementsystem} + \sigma_{brain}$

$$\sigma^2 = \int_{f_1}^{f_2} S(f)\,df$$

When calculating the white noise contribution, we know the power spectrum is constant across the entire frequency band of interest.

$$\sigma^2_{measurementsystem} = S \cdot (f_2 - f_1) >> \sigma^2 = 16\ \mathrm{fT}^2/\mathrm{Hz} \cdot (40 - 1) = 624\ \mathrm{fT}^2$$

Standard deviation is:

$$\sigma_{measurementsystem} = \sqrt{624} \approx 25\ \mathrm{fT}$$

For the brain we have its spectrum so:

$$\sigma^2_{brain} = \int_1^{40} \frac{3000}{f}\,df = 3000 \cdot \ln(40/1) = 3000 \cdot \ln(40) = 11066.7\ \mathrm{fT}^2$$

Standard deviation is:

$$\sigma_{brain} \approx \sqrt{11066.7} \approx 105.2\ \mathrm{fT}$$

Finally:

$$\sigma^2 = \sigma^2_{measurementsystem} + \sigma^2_{brain} = 624 + 11066.7 \approx 11690.7$$

Standard deviation is:

$$\sigma \approx \sqrt{11690.7} \approx 108.1\ \mathrm{fT}$$

Now coming back to our main question:

$$8 = 100\ \mathrm{fT}/\frac{108.1 fT}{\sqrt{N}}$$

Then:

$$\sqrt{N} = \frac{8.0 \times 108.1 fT}{100 fT} = 8.6496 = \sqrt{8.6496} = 74.81$$

iii) As we can see from the calculations, to find the total standard deviation of the noise, we must add the contributions from the measurement system and the brain since they are uncorrelated.
Since the noise originating from the brain is significantly greater than the intrinsic sensor noise it dominates the total noise profile. In this context, the intrinsic sensor noise has not an significant effect on diminishing the SNR.

Task2

A) Explain

What happens if the dipole is radially directed, i.e. pointing in the same direction as □r0? What does
this mean for MEG and its sensitivity to different parts of the cortex (i.e. its sensitivity pattern)?

Responses:

A) If the dipole is radially oriented (Q□□r□0\vec Q \parallel \vec r_0Q□r0), this means that the primary current is directed along the line from the center of the head toward the cortex.
The volume currents generated by this dipole are distributed symmetrically around it.
As a consequence, the magnetic fields produced by these currents cancel out outside the head due to symmetry.

Task3

A) Derive the Minimum Norm Solution

One widely used way to find a source estimate $\tilde{x}$ for eq. 3 is to minimize the L2-norm of the estimate. Using Lagrange multipliers, find $\tilde{x}$ that minimizes

$$\tilde{x}^T \tilde{x}$$

subject to

$$y - L\tilde{x} = 0$$

Note that L is not a square matrix. λ has size $\mathbb{R}^{1 \times m}$.

B) Write a short answer

The solution derived in A) is usually not well behaving due to singularities and noise in the measurement. Tikhonov regularization is classical method of regularizing this kind of ill-posed problems.1 What does Tikhonov regularization do, and why is it used?

C) Write a short answer

In an overdetermined case the solution becomes of form

$$\tilde{x} = (L^T L)^{-1} L^T y$$

What is the difference between underdetermined and overdetermined problems?

Responses:

A)

We define the Lagrangian $\mathcal{L}$ by combining the objective function and the constraint:

$$\mathcal{L}(\tilde{x}, \lambda) = \tilde{x}^T \tilde{x} - \lambda^T (y - L\tilde{x})$$

- $\tilde{x}^T \tilde{x}$ the objective function
- $y - L\tilde{x} = 0$ is the constraint
- λ Lagrangian multiplier

Now we have to compute the partial derivatives and set them to zero.

Derivative with respect to $\tilde{x}$:

$$\frac{\partial \mathcal{L}}{\partial \tilde{x}} = 2\tilde{x} + L^T \lambda = 0 \implies \tilde{x} = -\frac{1}{2} L^T \lambda$$

Derivative of the Objective Term: $\frac{\partial}{\partial \tilde{x}}(\tilde{x}^T \tilde{x})$

This is the derivative of a quadratic form:

$$\tilde{x}^T \tilde{x} = [x_1, x_2, \ldots, x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2 = \sum_{i=1}^{n} x_i^2$$

So:

$$\frac{\partial}{\partial z}(z^T z) = 2z$$

- **Result:**

$$\frac{\partial}{\partial \tilde{x}}(\tilde{x}^T \tilde{x}) = 2\tilde{x}$$

Derivative of the Constraint Term: $\frac{\partial}{\partial \tilde{x}}(-\lambda^T (y - L\tilde{x}))$

$$-\lambda^T(y - L\tilde{x}) = -\lambda^T y + \lambda^T L\tilde{x}$$

Now we differentiate term by term with respect to $\tilde{x}$:

The first term ($-\lambda^T y$) does not contain $\tilde{x}$, so its derivative is 0

For the second term ($\lambda^T L\tilde{x}$): This is a linear form in $\tilde{x}$

In our case, let $a^T = \lambda^T L$. Therefore, $a = (\lambda^T L)^T = L^T\lambda$

The reason we transform $\lambda^T L$ into $L^T\lambda$ is purely **dimensional**. When solving an optimization problem, we must sum all partial derivatives to satisfy the stationarity condition $\nabla\mathcal{L} = 0$. For this summation to be mathematically valid, all terms must have the same dimensions.

Thanks to this step, the final equation is perfectly consistent from a linear algebra perspective:

$$\underbrace{2\tilde{x}}_{(n\times1)} + \underbrace{L^T\lambda}_{(n\times1)} = \underbrace{\mathbf{0}}_{(n\times1)}$$

- Result:

$$\frac{\partial}{\partial\tilde{x}}(-\lambda^T y + \lambda^T L\tilde{x}) = 0 + L^T\lambda = L^T\lambda$$

Now we put together the two terms:

$$\frac{\partial\mathcal{L}}{\partial\tilde{x}} = 2\tilde{x} + L^T\lambda$$

To find the minimum, we set the gradient to zero:

$$2\tilde{x} + L^T\lambda = 0 \implies 2\tilde{x} = -L^T\lambda \implies \tilde{x} = -\frac{1}{2}L^T\lambda$$

Derivative with respect to λ:

We differentiate $\mathcal{L}$ with respect to the multiplier λ:

$$\mathcal{L}(\tilde{x}, \lambda) = \tilde{x}^T\tilde{x} - \lambda^T(y - L\tilde{x})$$

The term $-\lambda^T(y - L\tilde{x})$ is a linear form in λ

$$\frac{\partial}{\partial z}(z^T a) = a$$

- Result:

$$\frac{\partial\mathcal{L}}{\partial\lambda} = -(y - L\tilde{x})$$

  Setting this to zero recovers the constraint:

$$-(y - L\tilde{x}) = 0 \implies y = L\tilde{x}$$

Substituting $\tilde{x}$ into the Constraint to Find λ

The constraint is:

$$y = L\tilde{x}$$

And the X we foiund thanks to the derivative is:

$$\tilde{x} = -\frac{1}{2}L^T\lambda$$

So we need to substitute this value into the equation to find λ

$$y = L\left(-\frac{1}{2}L^T\lambda\right)$$

$$y = -\frac{1}{2}(LL^T)\lambda$$

Solving for λ

First, multiply both sides by -2 to eliminate the fraction:

$$-2y = (LL^T)\lambda$$

Next, to isolate λ, we multiply by the inverse of the matrix (LL^T). Assuming L has full row rank, the product (LL^T) is invertible:

$$(LL^T)^{-1}(-2y) = \lambda$$

Rearranging the terms:

$$\lambda = -2(LL^T)^{-1}y$$

Finding the Optimal $\tilde{x}$

Now that we have the value for the Lagrange multiplier λ, we can plug it back into our first optimality condition to find the solution for $\tilde{x}$:

$$\tilde{x} = -\frac{1}{2}L^T\left[-2(LL^T)^{-1}y\right]$$

The constants $-\frac{1}{2}$ and -2 cancel each other out:

$$\tilde{x} = L^T(LL^T)^{-1}y$$

B)

L2 Regularization (Tikhonov Regularization) adds a penalty proportional to the squared amplitudes of the sources to the cost function.

Regularized cost function:

$$\min_x \left(\|y - Lx\|^2 + \lambda \|x\|^2 \right)$$

What does it do?

- Penalizes large source amplitudes (x) by squaring them.
- Shrinks all estimated sources toward zero, improving stability
- Keeps all sources in the model, but reduces the effect of noisy or unstable components

C)

- **Overdetermined System**: This occurs when there are **more equations than unknowns**. It often leads to no solution or a unique solution, as the extra equations may impose conflicting constraints. More measurements than sources $\rightarrow$ generally no exact solution exists, but we seek the "best fit" (minimizing the error between the data and the prediction).
- **Underdetermined System**: This occurs when there are **fewer equations than unknowns**. It typically has infinitely many solutions, as there are not enough constraints to determine a unique solution. More sources than sensors $\rightarrow$ an infinite number of possible solutions.