Manifold-regression to predict from MEG/EEG
brain signals without source modeling

D. Sabbagh, P. Ablin, G. Varoquaux, A. Gramfort, D. Engemann

NeurIPS 2019
Non-invasive measure of brain activity
Non-invasive measure of brain activity

EEG recordings

Equivalent Current Dipole

NEURAL CURRENT (POST SYNAPTIC)

MEG recordings

Equivalent Current Dipole

NEURAL CURRENT (POST SYNAPTIC)

First EEG recordings in 1929 by H. Berger

Hôpital La Timone Marseille, France
Objective: predict a variable from M/EEG brain signals
Objective: predict a variable from M/EEG brain signals

- We want to predict a continuous variable
- From M/EEG brain signals
Objective: predict a variable from M/EEG brain signals

- We want to predict a continuous variable
- From M/EEG brain signals

Sample MEG measurements

**EEG:**
- \(\approx 32\) to 100 sensors

**MEG:**
- \(\approx 150\) to 300 sensors

Data are multivariate time-series

Time frame: 10 seconds

\(\approx 1000\) samples / s
Model: generative model of M/EEG data

We measure M/EEG signal of subject \( i = 1 \ldots N \) on \( P \) channels:

\[ x_i(t) = A s_i(t) + n_i(t) \in \mathbb{R}^P \]

mixing matrix \( A = [a_1, \ldots, a_Q] \in \mathbb{R}^{P \times Q} \) fixed across subjects

source patterns \( a_j \in \mathbb{R}^P, j = 1 \ldots Q \) with \( Q < P \)

source vector \( s_i(t) \in \mathbb{R}^Q \)

noise \( n_i(t) \in \mathbb{R}^P \)

Under stationarity and Gaussianity assumptions, we can represent band-pass filtered signal by its second-order statistics

\[ C_i = E[x_i(t)x_i(t)^\top] \simeq X_i X_i^\top \in \mathbb{R}^{P \times P} \]
Model: generative model of M/EEG data

We measure M/EEG signal of subject \(i = 1 \ldots N\) on \(P\) channels:

\[
x_i(t) = A \, s_i(t) + n_i(t) \in \mathbb{R}^P \quad t = 1 \ldots T
\]

- *mixing matrix* \(A = [a_1, \ldots, a_Q] \in \mathbb{R}^{P \times Q}\) fixed across subjects
- *source patterns* \(a_j \in \mathbb{R}^P, j = 1 \ldots Q\) with \(Q < P\)
- *source vector* \(s_i(t) \in \mathbb{R}^Q\)
- *noise* \(n_i(t) \in \mathbb{R}^P\)

Under stationarity and gaussianity assumptions, we can represent band-pass filtered signal by its second-order statistics

\[
C_i = E \left[ x_i(t) x_i(t) \top \right] \approx X_i X_i \top \in \mathbb{R}^{P \times P}
\]

with \(X_i \in \mathbb{R}^{P \times T}\).
Model: generative model of M/EEG data

We measure M/EEG signal of subject $i = 1 \ldots N$ on $P$ channels:

$$\mathbf{x}_i(t) = \mathbf{A} \mathbf{s}_i(t) + \mathbf{n}_i(t) \in \mathbb{R}^P \quad t = 1 \ldots T$$

- mixing matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_Q] \in \mathbb{R}^{P \times Q}$ fixed across subjects
- source patterns $\mathbf{a}_j \in \mathbb{R}^P$, $j = 1 \ldots Q$ with $Q < P$
- source vector $\mathbf{s}_i(t) \in \mathbb{R}^Q$
- noise $\mathbf{n}_i(t) \in \mathbb{R}^P$

Under stationnarity and gaussianity assumptions, we can represent band-pass filtered signal by its second-order statistics

$$\mathbf{C}_i = \mathbb{E}[\mathbf{x}_i(t)\mathbf{x}_i(t)^\top] \simeq \frac{\mathbf{X}_i \mathbf{X}_i^\top}{T} \in \mathbb{R}^{P \times P} \quad \text{with} \quad \mathbf{X}_i \in \mathbb{R}^{P \times T}$$
Model: generative model of target variable

We want to predict a continuous variable:

\[ y_i = \sum_{j=1}^{Q} \alpha_j f(p_{i,j}) \in \mathbb{R} \]

with \( p_{i,j} = \mathbb{E}_t[s_{i,j}^2(t)] \) band-power of sources
We want to predict a continuous variable:

$$y_i = \sum_{j=1}^{Q} \alpha_j f(p_{i,j}) \in \mathbb{R}$$

with $p_{i,j} = \mathbb{E}_t[s_{i,j}^2(t)]$ band-power of sources

- linear $y_i = \sum_{j=1}^{Q} \alpha_j p_{i,j}$
We want to predict a continuous variable:

$$y_i = \sum_{j=1}^{Q} \alpha_j f(p_{i,j}) \in \mathbb{R}$$

with $p_{i,j} = \mathbb{E}_t[s_{i,j}^2(t)]$ band-power of sources

- linear $y_i = \sum_{j=1}^{Q} \alpha_j p_{i,j}$

Euclidean vectorization leads to consistent model

$$y_i = \sum_{k \leq l} \Theta_{k,l} C_i(k, l)$$ i.e. $y_i$ is linear in coeff. of $\text{Upper}(C_i)$
Model: generative model of target variable

We want to predict a continuous variable:

\[ y_i = \sum_{j=1}^{Q} \alpha_j f(p_{i,j}) \in \mathbb{R} \]

with \( p_{i,j} = \mathbb{E}_t [s_{i,j}^2(t)] \) band-power of sources

- linear \( y_i = \sum_{j=1}^{Q} \alpha_j p_{i,j} \)

**Euclidean vectorization leads to consistent model**

\[ y_i = \sum_{k \leq l} \Theta_{k,l} \mathbf{C}_i(k,l) \]

i.e. \( y_i \) is linear in coeff. of Upper(\( \mathbf{C}_i \))

- log linear \( y_i = \sum_{j=1}^{Q} \alpha_j \log (p_{i,j}) \)
We want to predict a continuous variable:

\[ y_i = \sum_{j=1}^{Q} \alpha_j f(p_{i,j}) \in \mathbb{R} \]

with \( p_{i,j} = \mathbb{E}_t[s_{i,j}^2(t)] \) band-power of sources

- linear \( y_i = \sum_{j=1}^{Q} \alpha_j p_{i,j} \)

Euclidean vectorization leads to consistent model

\[ y_i = \sum_{k \leq l} \Theta_{k,l} C_i(k, l) \quad i.e. \ y_i \ is \ linear \ in \ coeff. \ of \ Upper(C_i) \]

- log linear \( y_i = \sum_{j=1}^{Q} \alpha_j \log (p_{i,j}) \)

\( C_i \) live on a Riemannian manifold so can’t be naively vectorized
Riemannian matrix manifolds (in a nutshell)
Riemannian matrix manifolds (in a nutshell)

Vectorization operator:

\[ \text{P}_M(\text{M}') = \varphi_M(\text{Log}_M(\text{M}')) \approx \text{Upper}(\text{Log}_M(\text{M}')) \]

\[ d(\text{M}_i, \text{M}_j) \approx \| \text{P}_M(\text{M}_i) - \text{P}_M(\text{M}_j) \|_2 \]

Vectorization operator key for ML

[Absil & al. Optimization algorithms on matrix manifolds. 2009]
Riemannian matrix manifolds (in a nutshell)

Vectorization operator:
\[ \mathcal{P}_M(M') = \phi_M(\text{Log}_M(M')) \simeq \text{Upper}(\text{Log}_M(M')) \]
\[ d(M, M') = \|\mathcal{P}_M(M')\|_2 + o(\|\mathcal{P}_M(M')\|_2) \]
\[ d(M_i, M_j) \simeq \|\mathcal{P}_M(M_i) - \mathcal{P}_M(M_j)\|_2 \]

Vectorization operator key for ML

[Absil & al. Optimization algorithms on matrix manifolds. 2009]
Regression on matrix manifolds

Given a training set of samples \( M_1, \ldots, M_N \in \mathcal{M} \) and target continuous variables \( y_1, \ldots, y_N \in \mathbb{R} \):
Given a training set of samples $M_1, \ldots, M_N \in \mathcal{M}$ and target continuous variables $y_1, \ldots, y_N \in \mathbb{R}$:

- compute the mean of the samples $\overline{M} = \text{Mean}_d(M_1, \ldots, M_N)$
Regression on matrix manifolds

Given a training set of samples $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}$ and target continuous variables $y_1, \ldots, y_N \in \mathbb{R}$:

- compute the mean of the samples $\overline{\mathbf{M}} = \text{Mean}_d(\mathbf{M}_1, \ldots, \mathbf{M}_N)$
- compute the vectorization of the samples w.r.t. this mean: $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathbb{R}^K$ as $\mathbf{v}_i = \mathcal{P}_{\overline{\mathbf{M}}} (\mathbf{M}_i)$
Regression on matrix manifolds

Given a training set of samples $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}$ and target continuous variables $y_1, \ldots, y_N \in \mathbb{R}$:

- compute the mean of the samples $\overline{\mathbf{M}} = \text{Mean}_d(\mathbf{M}_1, \ldots, \mathbf{M}_N)$
- compute the vectorization of the samples w.r.t. this mean: $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathbb{R}^K$ as $\mathbf{v}_i = \mathcal{P}_\overline{\mathbf{M}}(\mathbf{M}_i)$
- use those vectors as features in regularized linear regression algorithm (e.g. ridge regression) with parameters $\beta \in \mathbb{R}^K$
  assuming that $y_i \simeq \mathbf{v}_i^\top \beta$
Distance and invariance on positive matrix manifolds

Manifold of positive definite matrices:

\[ M_i \triangleq C_i \in \mathbb{S}^{++}_P \]

Geometric distance:

\[ d_G(S, S') = \| \log(S^{-1}S') \|_F = \left[ \sum_{P_i=1} \log^2 \lambda_k \right]^{1/2} \]

where \( \lambda_k \), \( k = 1 \ldots P \) are the real eigenvalues of \( S^{-1}S' \).

Tangent Space Projection:

\[ P_{S'}(S') = \text{Upper}(\log(S^{-1/2}S'S^{-1/2})) \]

Affine invariance property:

For invertible \( W \),

\[ d_G(W^{\top}SW, W^{\top}S'W) = d_G(S, S') \]

Affine invariance is key: working with \( C_i \) is then equivalent to working with covariance matrices of sources \( s_i \).

Distance and invariance on positive matrix manifolds

- Manifold of positive definite matrices: $M_i = C_i \in S^{++}_P$

$\text{Geometric distance: } d_G(S, S') = \left\| \log \left( S^{-1/2} S' S^{-1/2} \right) \right\|_F$

where $\lambda_k, k = 1, \ldots, P$ are the real eigenvalues of $S^{-1/2} S'$.

$\text{Tangent Space Projection: } P_S(S') = \text{Upper} \left( \log \left( S^{-1/2} S' S^{-1/2} \right) \right)$

$\text{Affine invariance property: }$ For $W$ invertible, $d_G(W^T S W, W^T S' W) = d_G(S, S')$

Affine invariance is key: working with $C_i$ is then equivalent to working with covariance matrices of sources $s_i$.

---

Manifold of positive definite matrices: $M_i = C_i \in S_{P}^{++}$

Geometric distance:

$$d_G(S, S') = \| \log(S^{-1}S')\|_F = \left[ \sum_{i=1}^{P} \log^2 \lambda_k \right]^{\frac{1}{2}}$$

where $\lambda_k, k = 1 \ldots P$ are the real eigenvalues of $S^{-1}S'$.

Tangent Space Projection:

$$\mathcal{P}_S(S') = \text{Upper}(\log(S^{-\frac{1}{2}} S' S^{-\frac{1}{2}}))$$

Distance and invariance on positive matrix manifolds

- **Manifold of positive definite matrices**: \( M_i = C_i \in S_{\mathbb{P}}^{++} \)
- **Geometric distance**:
  \[
  d_G(S, S') = \| \log(S^{-1}S') \|_F = \left[ \sum_{i=1}^{P} \log^2 \lambda_k \right]^{\frac{1}{2}}
  \]
  where \( \lambda_k, k = 1 \ldots P \) are the real eigenvalues of \( S^{-1}S' \).
- **Tangent Space Projection**:
  \( \mathcal{P}_S(S') = \text{Upper}(\log(S^{-\frac{1}{2}}S'S^{-\frac{1}{2}})) \)
- **Affine invariance property**:
  For \( W \) invertible, \( d_G(W^\top SW, W^\top S'W) = d_G(S, S') \)

---

Distance and invariance on positive matrix manifolds

- **Manifold of positive definite matrices:** $M_i = C_i \in S_{++}^P$

- **Geometric distance:**
  \[
  d_G(S, S') = \| \log(S^{-1}S') \|_F = \left[ \sum_{i=1}^{P} \log^2 \lambda_k \right]^{\frac{1}{2}}
  \]
  where $\lambda_k, k = 1 \ldots P$ are the real eigenvalues of $S^{-1}S'$.

- **Tangent Space Projection:**
  \[
  P_S(S') = \text{Upper}(\log(S^{-\frac{1}{2}}S'S^{-\frac{1}{2}}))
  \]

- **Affine invariance property:**
  For $W$ invertible, $d_G(W^\top SW, W^\top S'W) = d_G(S, S')$

Affine invariance is key: working with $C_i$ is then equivalent to working with covariance matrices of sources $s_i$

---

Consistency of linear regression in tangent space of $S_{p}^{++}$
Consistency of linear regression in tangent space of $S^+_p$

**Geometric vectorization**

Assume $y_i = \sum_{j=1}^{Q} \alpha_j \log(p_{i,j})$. Denote $\overline{C} = \text{Mean}_G(C_1, \ldots, C_N)$ and $v_i = \mathcal{P}_C(C_i)$. Then, the relationship between $y_i$ and $v_i$ is linear.
Consistency of linear regression in tangent space of $S^{++}_P$

**Geometric vectorization**

Assume $y_i = \sum_{j=1}^{Q} \alpha_j \log(p_{i,j})$. Denote $\overline{C} = \text{Mean}_G(C_1, \ldots, C_N)$ and $v_i = \mathcal{P}_\overline{C}(C_i)$. Then, the relationship between $y_i$ and $v_i$ is linear.

We generate i.i.d. samples following the log linear generative model. $A = \exp(\mu B)$ with $B \in \mathbb{R}^{P \times P}$ random.
And in the real world?

We investigated 3 violations from previous idealized model:
And in the real world?

We investigated 3 violations from previous idealized model:

- Noise in target variable

\[ y_i = \sum_j \alpha_j \log(p_{ij}) + \varepsilon_i, \quad \text{with} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \]
And in the real world?

We investigated 3 violations from previous idealized model:

- **Noise in target variable**
  \[ y_i = \sum_j \alpha_j \log(p_{ij}) + \varepsilon_i , \quad \text{with} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \]

- **Subject-dependent mixing matrix**
  \[ A_i = A + E_i , \quad \text{with entries of} \quad E_i \sim \mathcal{N}(0, \sigma^2) \]
And in the real world?

We investigated 3 violations from previous idealized model:

- **Noise in target variable**
  \[
  y_i = \sum_j \alpha_j \log(p_{ij}) + \varepsilon_i , \quad \text{with} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)
  \]

- **Subject-dependent mixing matrix**
  \[
  A_i = A + E_i , \quad \text{with entries of} \quad E_i \sim \mathcal{N}(0, \sigma^2)
  \]
And in the real world?

- **Rank-deficient signals** (e.g. cleaning process): \( C_i \in S_{P,R}^+ \)
And in the real world?

- Rank-deficient signals (e.g. cleaning process): \( C_i \in S_{p,R}^+ \)

We can manipulate them in their native manifolds \( S_{p,R}^+ \):

- **Wasserstein distance:**
  \[
  d_W(S, S') = \left( \text{Tr}(S) + \text{Tr}(S') - 2\text{Tr}((S^{1/2}S'S^{1/2})^{1/2}) \right)^{1/2}
  \]

- **Tangent Space Projection:**
  \[
  P_{YY^\top}(Y'Y'^\top) = \text{vect}(Y'Q^* - Y) \in \mathbb{R}^{PR}
  \]
  where \( U\Sigma V^\top = Y^\top Y' \), \( Q^* = VU^\top \)

- **Orthogonal invariance property:**
  For \( W \) orthogonal, \( d_W(W^\top SW, W^\top S'W) = d_W(S, S') \)

---

And in the real world?

- Rank-deficient signals (e.g. cleaning process): \( C_i \in S_{P,R}^+ \)

We can manipulate them in their native manifolds \( S_{P,R}^+ \):

- **Wasserstein distance:**
  \[
  d_W(S, S') = \left[ \text{Tr}(S) + \text{Tr}(S') - 2 \text{Tr}((S_2^{-1} S' S_2^2)^{\frac{1}{2}}) \right]^{\frac{1}{2}}
  \]

- **Tangent Space Projection:**
  \[
  \mathcal{P}_{YY^\top}(Y' Y'^\top) = \text{vect}(Y' Q^* - Y) \in \mathbb{R}^{P \times R}
  \]
  where \( U \Sigma V^\top = Y^\top Y' \), \( Q^* = V U^\top \)

- **Orthogonal invariance property:**
  For \( W \) orthogonal, \( d_W(W^\top S W, W^\top S' W) = d_W(S, S') \)

---

**Wasserstein vectorization**

Assume \( y_i = \sum_{j=1}^Q \alpha_j \sqrt{p_{i,j}} \) and \( A \) orthogonal. Denote \( \bar{C} = \text{Mean}_W(C_1, \ldots, C_N) \) and \( v_i = \mathcal{P}_{\bar{C}}(C_i) \).

Then the relationship between \( y_i \) and \( v_i \) is linear.
Experiment: predict age from MEG data

Task-free MEG recordings from Cam-CAN dataset

Age is a dominant driver of cross-person variance in neuroscience data

Dimension:
\[ N = 595, \quad P = 102, \quad T \approx 520,000, \quad 65 \leq R_i \leq 73 \]

Signals is filtered into 9 frequency bands

mean absolute error (years)

log−diag Wasserstein geometric MNE
Experiment: predict age from MEG data

- Task-free MEG recordings from Cam-CAN dataset
Experiment: predict age from MEG data

- Task-free MEG recordings from Cam-CAN dataset
- Age is a dominant driver of cross-person variance in neuroscience data
Experiment: predict age from MEG data

- Task-free MEG recordings from Cam-CAN dataset
- Age is a dominant driver of cross-person variance in neuroscience data
- Dimension: $N = 595$, $P = 102$, $T \simeq 520,000$, $65 \leq R_i \leq 73$
Experiment: predict age from MEG data

- Task-free MEG recordings from Cam-CAN dataset
- Age is a dominant driver of cross-person variance in neuroscience data
- Dimension: $N = 595, P = 102, T \simeq 520,000, 65 \leq R_i \leq 73$
- Signals is filtered into 9 frequency bands
Experiment: predict age from MEG data

- Task-free MEG recordings from Cam-CAN dataset
- Age is a dominant driver of cross-person variance in neuroscience data
- Dimension: $N = 595$, $P = 102$, $T \approx 520,000$, $65 \leq R_i \leq 73$
- Signals is filtered into 9 frequency bands
Conclusion

- Proposed Riemannian method for regression from M/EEG data

[Sabbagh, Ablin, Varoquaux, Gramfort, Engemann (2019), Manifold-regression to predict from MEG/EEG brain signals without source modeling, Proc. NeurIPS 2019]

https://arxiv.org/abs/1906.02687
Conclusion

- Proposed Riemannian method for regression from M/EEG data
- With theoretical guarantees and empirical robustness

[Sabbagh, Ablin, Varoquaux, Gramfort, Engemann (2019), Manifold-regression to predict from MEG/EEG brain signals without source modeling, Proc. NeurIPS 2019]

https://arxiv.org/abs/1906.02687
Proposed Riemannian method for regression from M/EEG data
With theoretical guarantees and empirical robustness
Working to translate proposed method into the clinic
Conclusion

- Proposed Riemannian method for regression from M/EEG data
- With theoretical guarantees and empirical robustness
- Working to translate proposed method into the clinic


https://arxiv.org/abs/1906.02687