Off-the-grid sparse estimation

Clarice Poon

University of Bath

Joint work with:

Nicolas Keriven, Gabriel Peyré, Mohammad Golbabaee

March 15, 2021

Outline

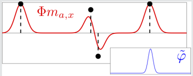
Problem description

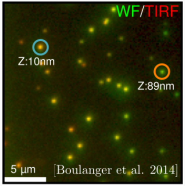
Sparse linear models

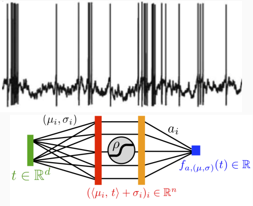
Unknown sparse measure: $\mathbf{m}_{a,\theta} = \sum_{i=1}^{n} a_i \delta_{\theta_i}$ where $a_i \in \mathbb{R}$, $\theta_i \in \Theta \subset \mathbb{R}^d$.

Observe linear model: Define $\varphi : \Theta \to \mathcal{H}$ continuous

$$\Phi : \mathcal{M}(\Theta) \to \mathcal{H}, \quad \Phi \mathbf{m} \stackrel{\text{def.}}{=} \int_\Theta \varphi(\theta) \mathrm{d}\mathbf{m}(\theta)$$







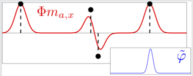
Problem description

Sparse linear models

Unknown sparse measure: $\mathbf{m}_{a,\theta} = \sum_{i=1}^{n} a_i \delta_{\theta_i}$ where $a_i \in \mathbb{R}$, $\theta_i \in \Theta \subset \mathbb{R}^d$.

Observe linear model: Define $\varphi : \Theta \to \mathcal{H}$ continuous

$$\Phi : \mathcal{M}(\Theta) \to \mathcal{H}, \; \Phi\mathbf{m} \overset{\text{def.}}{=} \int_{\Theta} \varphi(\theta) \mathrm{d}\mathbf{m}(\theta)$$



Fourier measurements: $\varphi(\theta) = (\exp(2\pi i \ell \theta))_{|\ell| \leqslant F} \in \mathbb{C}^{2F+1}$.
Then,

$$y = \Phi\mathbf{m} = \left(\sum_{i=1}^{n} a_i \exp(2\pi i \ell \theta_i) \right)_{|\ell| \leqslant F}.$$

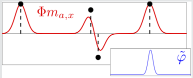
Problem description

Sparse linear models

Unknown sparse measure: $\mathbf{m}_{a,\theta} = \sum_{i=1}^{n} a_i \delta_{\theta_i}$ where $a_i \in \mathbb{R}$, $\theta_i \in \Theta \subset \mathbb{R}^d$.

Observe linear model: Define $\varphi : \Theta \to \mathcal{H}$ continuous

$$\Phi : \mathcal{M}(\Theta) \to \mathcal{H}, \quad \Phi\mathbf{m} \stackrel{\text{def.}}{=} \int_\Theta \varphi(\theta) \mathrm{d}\mathbf{m}(\theta)$$



Deconvolution: $\varphi(\theta) = \kappa(\cdot - \theta) \in L^2(\mathbb{R}^d)$. Then,

$$y = \Phi\mathbf{m} = \sum_{i=1}^{n} a_i \kappa(\cdot - \theta_i).$$

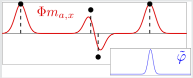
Problem description

Sparse linear models

Unknown sparse measure: $\mathbf{m}_{a,\theta} = \sum_{i=1}^{n} a_i \delta_{\theta_i}$ where $a_i \in \mathbb{R}$, $\theta_i \in \Theta \subset \mathbb{R}^d$.

Observe linear model: Define $\varphi : \Theta \to \mathcal{H}$ continuous

$$\Phi : \mathcal{M}(\Theta) \to \mathcal{H}, \quad \Phi\mathbf{m} \overset{\text{def.}}{=} \int_{\Theta} \varphi(\theta) \mathrm{d}\mathbf{m}(\theta)$$



Laplace: $\varphi(\theta) = (\exp(-\theta t_k))_{k=1}^{n} \in \mathbb{R}^m$. Then,

$$y = \Phi\mathbf{m} = \left(\sum_{i=1}^{n} a_i \exp(-\theta_i t_k) \right)_{k=1}^{m}.$$

Multicompartment effects in imaging

We observe at voxel some time series measurement $y \in \mathbb{R}^T$

$$y = \sum_{i=1}^{s} a_i \varphi(\theta_i)$$

where $\varphi(\theta) \in \mathbb{R}^T$ models the behaviour of tissue type θ over time.

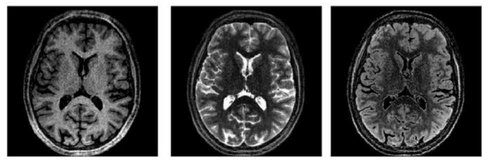


Figure 1: Contrast maps in quantitative MRI. Understanding multicompartment effects is important for accurate segmentation and studies of brain disorders.

The Beurling Lasso

Nonlinear least squares problem is nonconvex:

$$\min_{a,\theta} \frac{1}{2} \left\| \sum_i a_i \varphi(\theta_i) - y \right\|_2^2 + \lambda \left\| a \right\|_1$$

The Beurling Lasso

Nonlinear least squares problem is nonconvex:

$$\min_{a,\theta} \frac{1}{2} \left\| \sum_i a_i \varphi(\theta_i) - y \right\|_2^2 + \lambda \|a\|_1$$

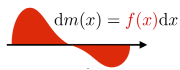
The Beurling Lasso

Minimisation over the space of measures is convex:

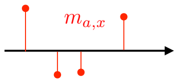
$$\min_{\mathbf{m} \in \mathcal{M}(\Theta)} \frac{1}{2} \|\Phi \mathbf{m} - y\|_2^2 + \lambda \|\mathbf{m}\|_{TV} . \qquad (\mathcal{P}_\lambda(y))$$

[Beurling, '38, De Castro & Gamboa, '12, Bredies & Pikkarainnen, '13]

$\|\mathbf{m}\|_{TV} \stackrel{\text{def.}}{=} \sup_{\{\mathcal{A}_i\} \subset \Theta} \sum_i |\mathbf{m}(\mathcal{A}_i)|.$



$\|\mathbf{m}\|_{TV} = \|f\|_{L^1}$



$\|\mathbf{m}_{a,x}\|_{TV} = \|a\|_1$

The Beurling Lasso

Nonlinear least squares problem is nonconvex:

$$\min_{a,\theta} \frac{1}{2} \left\| \sum_i a_i \varphi(\theta_i) - y \right\|_2^2 + \lambda \|a\|_1$$

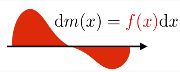
The Beurling Lasso

Minimisation over the space of measures is convex:

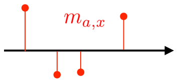
$$\min_{\mathbf{m} \in \mathcal{M}(\Theta)} \|\mathbf{m}\|_{TV} \quad \text{s.t.} \quad \Phi \mathbf{m} = y \qquad (\mathcal{P}_0(y))$$

[Beurling, '38, De Castro & Gamboa, '12, Bredies & Pikkarainnen, '13]

$\|\mathbf{m}\|_{TV} \overset{\text{def.}}{=} \sup_{\{\mathcal{A}_i\} \subset \Theta} \sum_i |\mathbf{m}(\mathcal{A}_i)|.$



$\|\mathbf{m}\|_{TV} = \|f\|_{L^1}$

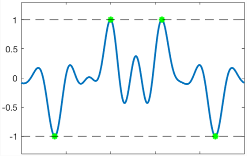


$\|\mathbf{m}_{a,x}\|_{TV} = \|a\|_1$

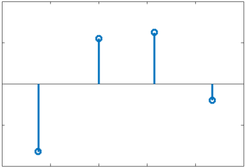
Key theoretical tool: dual certificates *

If you can find $\eta = \Phi^* p$ such that $|\eta(t)| < 1$ for all $t \notin \{\theta_i\}_i$ and $\eta(\theta_j) = \text{sign}(a_j)$, then

- Exact recovery of $\mathbf{m} = \sum_j a_j \delta_{\theta_j}$ from $y = \Phi\mathbf{m}$ by solving $\mathcal{P}_0(y)$.
- Stable recovery of $\mathbf{m} = \sum_j a_j \delta_{\theta_j}$ from $y = \Phi\mathbf{m} + w$ by solving $\mathcal{P}_\lambda(y)$, if in addition, $\text{sign}(a_j)\eta''(\theta_j) < 0$.



η



$\mathbf{m}$

*Most Blasso papers make use of this result...

Key theoretical tool: dual certificates *

If you can find $\eta = \Phi^* p$ such that $|\eta(t)| < 1$ for all $t \notin \{\theta_i\}_i$ and $\eta(\theta_j) = \text{sign}(a_j)$, then

- Exact recovery of $\mathbf{m} = \sum_j a_j \delta_{\theta_j}$ from $y = \Phi\mathbf{m}$ by solving $\mathcal{P}_0(y)$.
- Stable recovery of $\mathbf{m} = \sum_j a_j \delta_{\theta_j}$ from $y = \Phi\mathbf{m} + w$ by solving $\mathcal{P}_\lambda(y)$, if in addition, $\text{sign}(a_j)\eta''(\theta_j) < 0$.

Minimal norm certificate

Most of the time, we look at

$$\eta_0 \overset{\text{def.}}{=} \text{argmin}_{\eta=\Phi^* p} \|p\| \;\text{ s.t. }\; \begin{cases} \forall i,\; \eta(\theta_i) = \text{sign}(a_i) \\ \|\eta\|_\infty \leqslant 1. \end{cases}$$

*Most Blasso papers make use of this result...

Minimum separation

Candès & Fernandez-Granda, CPAM 2012

Consider $\varphi(\theta) = \left(e^{2\pi i k\theta}\right)_{|k|\leqslant f_c}$. In dimension 1 and 2, η_0 is nondegenerate if

$$\Delta_\theta \stackrel{\text{def.}}{=} \min_{i\neq j} |\theta_i - \theta_j|_\infty \geqslant \frac{C}{f_c}$$

This result is **sharp**: If $\mathbf{m} = \delta_\theta - \delta_{\theta'}$ and $|\theta - \theta'| < \frac{1}{f_c}$, then no dual certificate exists (and this actually means that recovery is not possible).

Minimum separation

Candès & Fernandez-Granda, CPAM 2012

Consider $\varphi(\theta) = \left(e^{2\pi i k\theta}\right)_{|k|\leqslant f_c}$. In dimension 1 and 2, η_0 is nondegenerate if

$$\Delta_\theta \stackrel{\text{def.}}{=} \min_{i\neq j} |\theta_i - \theta_j|_\infty \geqslant \frac{C}{f_c}$$

This result is **sharp**: If $\mathbf{m} = \delta_\theta - \delta_{\theta'}$ and $|\theta - \theta'| < \frac{1}{f_c}$, then no dual certificate exists (and this actually means that recovery is not possible).

We first need to understand the minimum separation for arbitrary operators – need a metric to quantify what we mean by two spikes being close...

Fisher-Rao distance

For $\theta, \theta' \in \Theta \subset \mathbb{R}^d$, define $K(\theta, \theta') \overset{\text{def.}}{=} \langle \varphi(\theta), \varphi(\theta') \rangle$.

Fisher metric:

$$\mathfrak{g}_\theta = \nabla_1 \nabla_2 K(\theta, \theta) = [\nabla \varphi(\theta)][\nabla \varphi(\theta)]^\top \in \mathbb{R}^{d \times d}$$

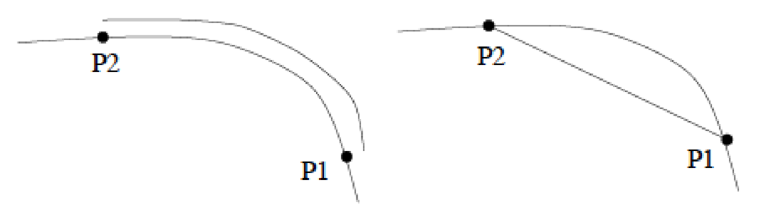
Fisher-Rao geodesic distance:

$$d_{\mathfrak{g}}(\theta, \theta') = \inf_{\gamma : \theta \to \theta'} \int_0^1 \sqrt{\langle \mathfrak{g}_{\gamma(t)} \gamma'(t), \gamma'(t) \rangle} \mathrm{d}t$$

Intuition

Statistical interpretation: If $\|\varphi(\theta)\| = 1$, then $\left(|\varphi(\theta)_i|^2\right)_i$ is a probability distribution.

Given $P_1 = \varphi(\theta)$ and $P_2 = \varphi(x')$:



The map $\theta \mapsto \varphi(\theta)$ embeds Θ into the sphere in $\mathcal{H}$ and

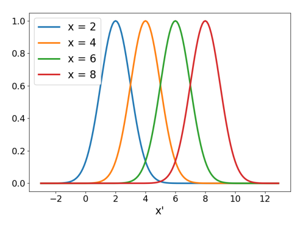
$$d_{\mathfrak{g}}(\theta, \theta') = \inf_{\gamma : \varphi(\theta) \to \varphi(\theta')} \int_0^1 \|\gamma'(t)\|_{\mathcal{H}} \, \mathrm{d}t.$$

Examples

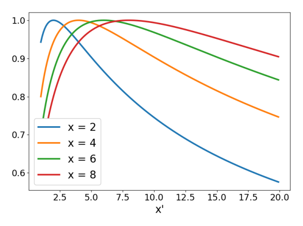
Fourier: $\Theta = \mathbb{T}^d$, $\varphi(\theta) = (e^{2\pi i k\theta})_{\|k\|_\infty \leqslant f_c}$ $\quad \mathfrak{g}_\theta = f_c^2 \mathrm{Id}$,
$d_{\mathfrak{g}}(\theta, \theta') = f_c \|\theta - \theta'\|_2$.

Gaussian convolution: $\Theta = \mathbb{R}^d$

$\varphi(\theta) \propto e^{\|\theta - \cdot\|_\Sigma^2}$, $\mathfrak{g}_\theta = \Sigma$,
$d_{\mathfrak{g}}(\theta, \theta') = \|\theta - \theta'\|_\Sigma$.



Laplace transform $\Theta = \mathbb{R}_+^d$,
$\varphi(\theta) \propto e^{-\theta\cdot}$, $\mathfrak{g}_\theta = \mathrm{diag}(1/\theta_i)$,
$d_{\mathfrak{g}}(\theta, \theta') = \sqrt{\sum_i |\log(\theta_i/\theta_i')|^2}$.



Recovery under minimum separation

Theorem (P., Keriven, Peyré '19)

Let $s \in \mathbb{N}$ *and let* $(\theta_i)_{i=1}^s$ *be s.t.* $\min_{i \neq j} d_{\mathfrak{g}}(\theta_i, \theta_j) \geqslant \Delta_{s,K}$.

Then: η_0 *is nondegenerate.*

Recovery under minimum separation

Theorem (P., Keriven, Peyré '19)

Let $s \in \mathbb{N}$ *and let* $(\theta_i)_{i=1}^s$ *be s.t.* $\min_{i \neq j} d_{\mathfrak{g}}(\theta_i, \theta_j) \geqslant \Delta_{s,K}$.

Then: η_0 *is nondegenerate.*

Examples:

Fourier coefficients: $\Delta = \min\left(\sqrt{d\sqrt{s}}, 2^d\right)$.

Gaussian deconvolution: $\Delta = \sqrt{\log(s)}$.

Laplace transform: $\Delta = d + \log(ds)$.

The separation distance $\Delta_{s,K}$ *is independent of the problem parameters!*

Compressed sensing result

E.g. $\varphi(\theta) = \left(e^{-2\pi i \omega_k \theta}\right)_{k=1}^{m}$ where $|k| \leqslant f_c$ are drawn randomly.

Compressed sensing result

E.g. $\varphi(\theta) = \left(e^{-2\pi i \omega_k \theta}\right)_{k=1}^{m}$ where $|k| \leqslant f_c$ are drawn randomly.

Setting:

Let (Ω, Λ) be a probability space and let $\varphi(\theta) = (\varphi_{\omega_k}(\theta))_{k=1}^{m}$ where $\omega_k \overset{iid}{\sim} \Lambda$.

Consider recovery from $y = \Phi\left(\sum_{i=1}^{s} a_s \delta_{\theta_s}\right) + w$.

Compressed sensing result

E.g. $\varphi(\theta) = \left(e^{-2\pi i \omega_k \theta}\right)_{k=1}^m$ where $|k| \leqslant f_c$ are drawn randomly.

Setting:

Let (Ω, Λ) be a probability space and let $\varphi(\theta) = (\varphi_{\omega_k}(\theta))_{k=1}^m$ where $\omega_k \overset{iid}{\sim} \Lambda$.

Consider recovery from $y = \Phi\left(\sum_{i=1}^s a_s \delta_{\theta_s}\right) + w$.

Assumptions:

- Let $\theta \in \Theta^s$ be such that $\min_{j \neq k} d_{\mathfrak{g}}(\theta_j, \theta_k) \geqslant \Delta$.
- Let $\rho > 0$ and

  $$m \geqslant C_{\bar{L}} \cdot s \cdot (\log^2(s/\rho) + \log(N^d/\rho))$$

  where $C_{\bar{L}}$ and N depends on the derivatives of φ_ω and the domain diameter $\sup_{\theta,\theta' \in \Theta} d_{\mathfrak{g}}(\theta, \theta')$.

Compressed sensing result

Theorem (P., Keriven, Peyré '19)

Let $\lambda \sim \|w\| / \sqrt{s}$. With probability at least $1-\rho$, any solution $\mathbf{m}$ to $\mathcal{P}_\lambda(y)$ satisfies the following discrepancies to the true measures $\mathbf{m}_{a,\theta}$:

$$\max_{j=1}^{s} |a_j - \hat{a}_j| \lesssim s^{1/2} \|w\| . \quad \textit{and} \quad \mathcal{T}_{\mathfrak{g}}^2(|\mathbf{m}|, |\mathbf{m}_{a,\theta}|) \lesssim s^{3/2} \|w\| .$$

where

$$\mathcal{T}_{\mathfrak{g}}^2(\mu, \nu) \stackrel{\text{def.}}{=} \inf_{\hat{\mu},\hat{\nu}} W_{\mathfrak{g}}^2(\hat{\mu}, \hat{\nu}) + \|\mu - \hat{\mu}\|_{TV} + \|\nu - \hat{\nu}\|_{TV}.$$

Examples and remarks

Sampling Fourier coefficients with $\Theta = [0,1]^d$:

$$m \sim d^2 \cdot s \cdot \left(\log^2(s) + \log(f_c^d)\right)$$

Examples and remarks

Sampling Fourier coefficients with $\Theta = [0, 1]^d$:

$$m \sim d^2 \cdot s \cdot \left(\log^2(s) + \log(f_c^d)\right)$$

Sampling the Laplace transform with $\Theta = (0, 1]^d$ with $\Lambda(t) \propto e^{-\alpha t}$ and $\alpha_i \sim d$:

$$m \sim d^6 \cdot s \cdot \left(\log^2(m) \log^2(s) + \log^4(m) \log(\log(m)^d)\right)$$

Examples and remarks

Sampling Fourier coefficients with $\Theta = [0,1]^d$:

$$m \sim d^2 \cdot s \cdot \left(\log^2(s) + \log(f_c^d)\right)$$

Sampling the Laplace transform with $\Theta = (0,1]^d$ with $\Lambda(t) \propto e^{-\alpha t}$ and $\alpha_i \sim d$:

$$m \sim d^6 \cdot s \cdot \left(\log^2(m)\log^2(s) + \log^4(m)\log(\log(m)^d)\right)$$

Remark: Previous result by Tang et al. (2013) for sampling Fourier coefficients in 1D, **but** their result assumes that $\text{sign}(a_j)$ are distributed uniformly iid.

Outline

Introduction to the Blasso

Applying the Blasso to qMRI

Magnetic resonance imaging

MRI is one of the main applications of compressed sensing, this allows for subsampling

$$\min_{x \in \mathbb{R}^v} \lambda \|x\|_1 + \frac{1}{2} \|\mathcal{F}_\Omega x - y\|_2^2$$

Magnetic resonance imaging

MRI is one of the main applications of compressed sensing, this allows for subsampling

$$\min_{x \in \mathbb{R}^v} \lambda \|x\|_1 + \frac{1}{2} \|\mathcal{F}_\Omega x - y\|_2^2$$

Traditional MRI:

- The MR signal is obtained by applied the same radio frequency (RF) pulse repeatedly.
- $x \in \mathbb{R}^v$ is a gray-valued image, which captures the relative signal intensity changes between tissues, each image voxel is weighted by so-called T1,T2 values.

Quantitative MRI: measure T1,T2 values

Magnetic resonance fingerprinting [Ma et al '13, Nature] allowed this to be done in short clinically feasible scan times.

- Allow the RF pulse to vary over time.
- This results in a time-series magnetisation images (TSMI)

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_v \end{bmatrix} \in \mathbb{R}^{T \times v}$$

  with v voxels and T timeframes.
- The time dependent signal in each voxel is compared to a dictionary of fingerprints $\{\varphi(\theta_i)\}_i \subset \mathbb{R}^T$:
  - Precomputed by solving so-called **Bloch equations**,
  - Each fingerprint corresponds to $\theta = (T_1, T_2)$ values which depend on tissue type.

The quantitative MRI problem

Multicompartment effects: There can be more than one tissue type appearing in one image voxel.

TSMI with v voxels and T timeframes:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_v \end{bmatrix} \in \mathbb{R}^{T \times v}$$

For each $x = x_i$,

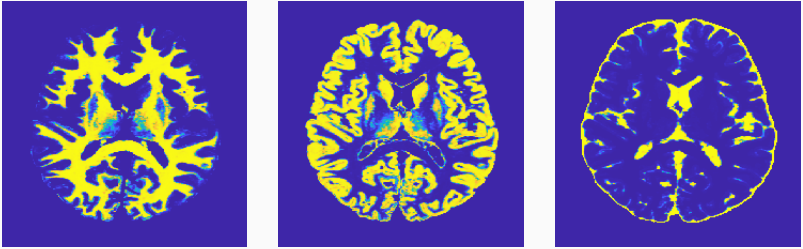
$$x = \sum_s c_s \varphi(\theta_s) \in \mathbb{R}^T$$

- $c_s \geqslant 0$ are mixture weights
- $\varphi : \Theta \to \mathbb{R}^T$ is the Block magnetisation response model.
- Θ is the domain of NMR properties.

Visualisation

TSMI:

Component maps:



$\theta_1 = (784, 77)$ $\theta_2 = (1216, 96)$ $\theta_3 = (4083, 1394)$

The quantitative MRI problem

Previous approaches:

- discretize the domain Θ, to $\{\theta_s\}_{s=1}^N$, form a dictionary

$$D_\theta \overset{\text{def.}}{=} \begin{bmatrix} \varphi(\theta_1) & \varphi(\theta_2) & \cdots & \varphi(\theta_N) \end{bmatrix}$$

- solve for $C \in \mathbb{R}^{v \times N}$,

$$D_\theta C^\top = X$$

  where each column $C_s \in \mathbb{R}^v$ correspond to the θ_s dependent mixture weights across all voxels.

Solve problem of form:

$$\min_C \frac{1}{2} \left\| D_\theta C^\top - X \right\|_F^2 + J(C)$$

Formulation as Blasso with vector-valued measures

Write $\mathbf{m} = \sum_{s=1}^{k} C_s^\top \delta_{\theta_s} \in \mathcal{M}(\Theta; \mathbb{R}^v)$.

Then, we have

$$\Phi\mathbf{m} = \int \varphi(\theta) \mathrm{d}\mathbf{m}(\theta) = \sum_s \varphi(\theta_s) C_s^\top.$$

Formulation as Blasso with vector-valued measures

Write $\mathbf{m} = \sum_{s=1}^{k} C_s^\top \delta_{\theta_s} \in \mathcal{M}(\Theta; \mathbb{R}^v)$.

Then, we have

$$\Phi\mathbf{m} = \int \varphi(\theta)\mathrm{d}\mathbf{m}(\theta) = \sum_s \varphi(\theta_s) C_s^\top .$$

Remark: Other models are possible, e.g.

$$\mathbf{m} = \sum_{s=1}^{k} C_s^\top g_\sigma(\theta_s - \cdot).$$

Then,

$$\Phi\mathbf{m} = \langle \varphi \star g_\sigma, \sum_s C_s^\top \delta_{\theta_s} \rangle$$

Formulation as Blasso with vector-valued measures

Write $\mathbf{m} = \sum_{s=1}^{k} C_s^\top \delta_{\theta_s} \in \mathcal{M}(\Theta; \mathbb{R}^v)$.

Then, we have

$$\Phi\mathbf{m} = \int \varphi(\theta)\mathrm{d}\mathbf{m}(\theta) = \sum_s \varphi(\theta_s) C_s^\top .$$

Total variation of vector valued measures

If a measure takes values in a normed space $\mathcal{V}$ endowed with norm $\|\cdot\|_{\mathcal{V}}$, then define

$$|\mathbf{m}|_{\mathcal{V}} = \sup_{\{\mathcal{A}_i\} \subset \mathcal{V}} \sum_{j=1}^{N} \|\mathbf{m}(\mathcal{A})\|_{\mathcal{V}} .$$

We need to choose $\|\cdot\|_{\mathcal{V}}$.

Sparse-group-Blasso

We consider regularisation with the following mixed norm:

$$\|\mathbf{m}\|_\beta \overset{\text{def.}}{=} (1-\beta)\,|\mathbf{m}|_1 + \beta\sqrt{v}\,|\mathbf{m}|_2\,.$$

So:

$$\min_{\mathbf{m}\in\mathcal{M}(\Theta;\mathbb{R}^v)} \lambda\,\|\mathbf{m}\|_\beta + \frac{1}{2}\,\|X - \Phi\mathbf{m}\|_F^2\,.$$

NB: $\|\sum_s C_s\delta_s\|_\beta = (1-\beta)\sum_s \|C_s\|_1 + \beta\sqrt{v}\sum_s \|C_s\|_2$.

- $\sum_s \|C_s\|_2$ enforces group sparsity.
- $\sum_s \|C_s\|_1$ enforces sparsity within each mixture map.

This is the continuous counterpart of the *sparse-group lasso* [Simon, Hastie & Tibshirani, JCGS, 2013].

Conditional gradient descent

Solve $\min_{x \in \mathcal{C}} f(x)$, $\mathcal{C}$ is a compact convex set in Banach space:

$$y_t \in \operatorname{argmin}_{y \in \mathcal{C}} \nabla f(x_t)^\top y$$
$$x_{t+1} = (1 - \gamma_t) x_t + \gamma_t y_t$$

Conditional gradient descent

Solve $\min_{x \in \mathcal{C}} f(x)$, $\mathcal{C}$ is a compact convex set in Banach space:

$$y_t \in \operatorname{argmin}_{y \in \mathcal{C}} \nabla f(x_t)^\top y$$
$$x_{t+1} = (1-\gamma_t)x_t + \gamma_t y_t$$

For our problem $\min_{\mathbf{m}} \lambda \|\mathbf{m}\|_{TV} + \frac{1}{2}\|\Phi\mathbf{m} - X\|_F^2$:

$$\lambda \|\mathbf{m}\|_\beta \leqslant \|0\|_{TV} + \frac{1}{2}\|\Phi 0 - X\|_F^2 = \|X\|_F^2/2.$$

Therefore, we solve

$$\min_{t, \mathbf{m} \in \mathcal{C}} f(t) = \lambda t + \frac{1}{2}\|X - \Phi\mathbf{m}\|_F^2$$

where $\mathcal{K} = \left\{(t, \mathbf{m}) \in \mathbb{R}_+ \times \mathcal{M} \;;\; \|\mathbf{m}\|_\beta \leqslant t \leqslant \|X\|_F^2/(2\lambda)\right\}$.
Convergence of objective is $\mathcal{O}(1/k)$ with k being iteration.

Frank-Wolfe/Conditional gradient iterations[†]

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

[†] Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations†

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \mathrm{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$

† Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations[†]

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \operatorname{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$

$\theta^{i+\frac{1}{2}} = \theta^i \cup \{\theta\}$

[†] Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations†

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \mathrm{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$

$\theta^{i+\frac{1}{2}} = \theta^i \cup \{\theta\}$

$C^{i+\frac{1}{2}} \in \mathrm{argmin}_{C\in\mathbb{R}_+^{k\times v}} \frac{1}{2} \left\| X - D_{\theta^{i+\frac{1}{2}}} C \right\|_F^2 + \alpha \left\| C \right\|_\beta$

† Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations[†]

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \operatorname{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$
$\theta^{i+\frac{1}{2}} = \theta^i \cup \{\theta\}$
$C^{i+\frac{1}{2}} \in \operatorname{argmin}_{C\in\mathbb{R}_+^{k\times v}} \frac{1}{2} \left\| X - D_{\theta^{i+\frac{1}{2}}} C \right\|_F^2 + \alpha \|C\|_\beta$
Initialising with $C^{i+1/2}$ and $\theta^{i+1/2}$, solve

$$(C^{i+1}, \theta^{i+1}) \in \operatorname*{argmin}_{\theta, C} \frac{1}{2} \left\| X - D_\theta C^\top \right\|_F^2 + \alpha \|C\|_\beta$$

[†] Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations†

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \mathrm{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$

$\theta^{i+\frac{1}{2}} = \theta^i \cup \{\theta\}$

$C^{i+\frac{1}{2}} \in \mathrm{argmin}_{C\in\mathbb{R}_+^{k\times v}} \frac{1}{2}\left\|X - D_{\theta^{i+\frac{1}{2}}} C\right\|_F^2 + \alpha \|C\|_\beta$

Initialising with $C^{i+1/2}$ and $\theta^{i+1/2}$, solve

$$(C^{i+1}, \theta^{i+1}) \in \underset{\theta, C}{\mathrm{argmin}} \frac{1}{2}\left\|X - D_\theta C^\top\right\|_F^2 + \alpha \|C\|_\beta$$

Define $\eta^{i+1} = \frac{1}{\alpha}\Phi^*(X - D_{\theta^{i+1}}(C^{i+1})^\top)$

† Follows [Denoyelle et al, Inverse problems '19]

Frank-Wolfe/Conditional gradient iterations[†]

Inputs: TSMI X, Bloch model $\varphi(.)$, params $\alpha, \beta > 0$.
Outputs: NMR parameters θ, mixture weights C.
Initialise: $i = 0$, $\theta^0 = \{\}$, $C^0 = \{\}$, $\eta^0 = \frac{1}{\alpha}\Phi^* X$.
repeat

Let $\theta \in \text{argmax}_{\theta\in\Theta} \sum_{s=1}^{v} \left(\eta^i(\theta)_s - (1-\beta)\right)_+^2$

$\theta^{i+\frac{1}{2}} = \theta^i \cup \{\theta\}$

$C^{i+\frac{1}{2}} \in \text{argmin}_{C\in\mathbb{R}_+^{k\times v}} \frac{1}{2} \left\| X - D_{\theta^{i+\frac{1}{2}}} C \right\|_F^2 + \alpha \|C\|_\beta$

Initialising with $C^{i+1/2}$ and $\theta^{i+1/2}$, solve

$$(C^{i+1}, \theta^{i+1}) \in \underset{\theta, C}{\text{argmin}} \frac{1}{2} \left\| X - D_\theta C^\top \right\|_F^2 + \alpha \|C\|_\beta$$

Define $\eta^{i+1} = \frac{1}{\alpha}\Phi^*(X - D_{\theta^{i+1}}(C^{i+1})^\top)$

$i = i + 1$

until $\sup_{\theta\in\mathcal{T}} \sum_{s=1}^{v} (\eta_s^i(\theta) - (1-\beta))_+^2 \leqslant v\beta^2$

[†] Follows [Denoyelle et al, Inverse problems '19]

Setup for numerics

The MRF data came from a healthy volunteer's brain, a variable density spiral trajectory was used for *k*-space sampling.

- MRF excitation sequences with $T = 1000$ timepoints. That is $\varphi(\theta) \in \mathbb{R}^T$. Acquisition window around 10s.
- The number of image voxels per timeframe is 230x230.
- First recover the TSMI from *k*-space measurements using LRTV. This is standard compressed sensing with TV regularisation.
- We then apply SGB-Lasso to recover mixture maps.

qMRI tricks... Phase correction [‡]

Typically, TSMI is complex valued, however, it is often assume to have constant-valued phase which can be subtracted and removed.

Useful because positivity constraint helps in practice.

[‡] Jiang et al., MRI, 2015; Nagtegaal et al., Magnetic resonance in medicine, 2020.

qMRI tricks... Low rank approximations [§]

It is observed that Block responses have low rank approximations

$$\varphi(\theta) \approx VV^\top \varphi(\theta)$$

where $V \in \mathbb{R}^{T\times\tau}$ with $\tau \ll T$ (we took $\tau = 10$) and the columns of V form an orthonormal system.

This V comes from PCA of a large simulated dictionary.

So, instead, work with $\tilde{\varphi} = V^\top \varphi(\theta) \in \mathbb{R}^\tau$ and $\tilde{X} = V^\top X$.

[§] McGivney et al. IEEE TMI (2014). Cline et al. MRI (2017)

qMRI tricks... Neural network approximations. ¶

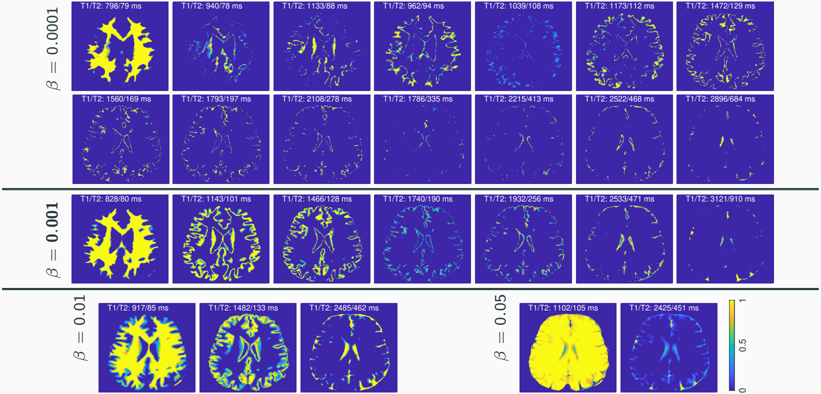
Instead of working with $\tilde{\varphi} = V^\top \varphi(\theta) \in \mathbb{R}^\tau$, train a 2 layer neural network

$$\mathcal{N} : \theta \in \Theta \mapsto \tilde{\varphi}(\theta).$$

This means that $\tilde{\varphi}$ and its Jacobian can be evaluated efficiently.

¶ Chen et al, MICCAI (2020); Gómez et al, Scientific reports (2020)

Effects of β



Comparison against baseline methods:

- PVMRF ‖. Estimate dictionary using k-means
- SPIJN ** Group sparsity regularization.
- BayesianMRF †† Enforces sparsity.

T1 (ms)

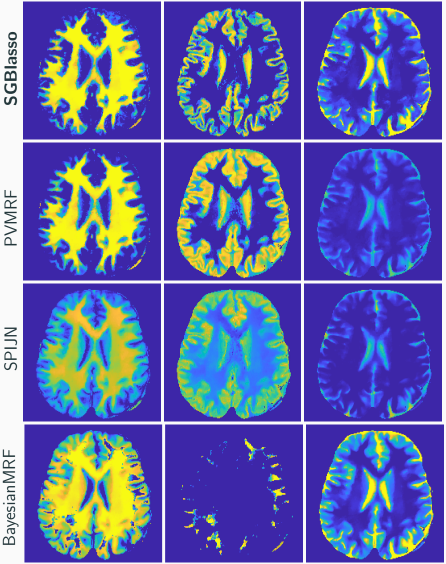
Tissue	Literature	**SGBlasso**	PVMRF	SPIJN	BayesianMRF
WM	694 – 862	829	806	699	821
GM	1074 – 1174	1114	1165	1483	874

T2 (ms)

Tissue	Literature	**SGBlasso**	PVMRF	SPIJN	BayesianMRF
WM	68 – 87	81	80	51	77
GM	87 – 103	102	105	164	82

‖Deshmane et al, NMR in Biomedicine, 2018
**Nagtegaal et al, Magnetic resonance in medicine 2020
††McGivney et al, Magnetic resonance in medicine 2018

Comparison with existing methods



Summary

- Introduction of the Fisher metric, which offers a way of imposing the separation condition. This provides a unified way of approaching nontranslational invariant problems.
- The Blasso framework gives promising results for the problem of multi-compartment analysis in MRF.

Papers:

- *The geometry of off-the-grid compressed sensing, P., Keriven & Peyré, arXiv:1802.08464*
- *An off-the-grid approach to multi-compartment magnetic resonance fingerprinting, Golbabaee & P., arXiv:2011.11193*

Summary

- Introduction of the Fisher metric, which offers a way of imposing the separation condition. This provides a unified way of approaching nontranslational invariant problems.
- The Blasso framework gives promising results for the problem of multi-compartment analysis in MRF.

Papers:

- *The geometry of off-the-grid compressed sensing, P., Keriven & Peyré, arXiv:1802.08464*
- *An off-the-grid approach to multi-compartment magnetic resonance fingerprinting, Golbabaee & P., arXiv:2011.11193*

Thanks for listening