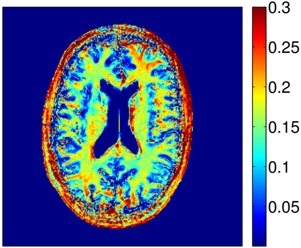
Dictionary-Free MRI Parameter Estimation via Kernel Ridge Regression

Gopal Nataraj*, Jon-Fredrik Nielsen†, & Jeffrey A. Fessler*

Int'l Symposium on Biomedical Imaging 2017

Departments of *EECS and †BME
University of Michigan



Quantitative MRI (QMRI) Parameter Estimation

Given: MR image sequence informative about a physical process

- flow
- diffusion
- multi-compartmental relaxation
- ...

Quantitative MRI (QMRI) Parameter Estimation

Given: MR image sequence informative about a physical process

- flow
- diffusion
- multi-compartmental relaxation
- . . .

Task: estimate MR tissue properties characterizing the process

- flow velocity
- diffusivity
- compartmental relaxivity
- . . .

QMRI Problem Statement

Given: at each voxel, image sequence $\mathbf{y} \in \mathbb{C}^D$ modeled as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}, \boldsymbol{\nu}) + \boldsymbol{\epsilon} \tag{1}$$

- $\mathbf{x} \in \mathbb{R}^L$ latent free parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ known parameters
- $\mathbf{s} : \mathbb{R}^{L+K} \mapsto \mathbb{C}^D$ signal model
- $\boldsymbol{\epsilon} \in \mathbb{C}^D$ noise $\sim \mathbb{C}\mathcal{N}(\mathbf{0}_D, \boldsymbol{\Sigma})$

QMRI Problem Statement

Given: at each voxel, image sequence $\mathbf{y} \in \mathbb{C}^D$ modeled as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}, \boldsymbol{\nu}) + \boldsymbol{\epsilon} \tag{1}$$

- $\mathbf{x} \in \mathbb{R}^L$ latent free parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ known parameters
- $\mathbf{s} : \mathbb{R}^{L+K} \mapsto \mathbb{C}^D$ signal model
- $\boldsymbol{\epsilon} \in \mathbb{C}^D$ noise $\sim \mathbb{C}\mathcal{N}(\mathbf{0}_D, \boldsymbol{\Sigma})$

Task: construct fast estimator $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Prior Work

Task: construct fast estimator $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Challenges:

- signal $\mathbf{s}$ often nonlinear in $\mathbf{x}$: non-convex inverse problems
- signal $\mathbf{s}$ might be difficult to write in closed form

Prior Work

Task: construct fast estimator $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Challenges:

- signal $\mathbf{s}$ often nonlinear in $\mathbf{x}$: non-convex inverse problems
- signal $\mathbf{s}$ might be difficult to write in closed form

Conventional Approaches:

- gradient-based local optimization
  - initialization-dependent solution
  - requires signal gradients

Prior Work

Task: construct fast estimator $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Challenges:

- signal $\mathbf{s}$ often nonlinear in $\mathbf{x}$: non-convex inverse problems
- signal $\mathbf{s}$ might be difficult to write in closed form

Conventional Approaches:

- gradient-based local optimization
    - initialization-dependent solution
    - requires signal gradients
- stochastic methods (*e.g.*, simulated annealing)
    - unclear convergence analysis [Bertsimas and Tsitsiklis, 1993]
    - several unintuitive tuning parameters

Prior Work

Task: construct fast estimator $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Challenges:

- signal $\mathbf{s}$ often nonlinear in $\mathbf{x}$: non-convex inverse problems
- signal $\mathbf{s}$ might be difficult to write in closed form

Conventional Approaches:

- gradient-based local optimization
  - initialization-dependent solution
  - requires signal gradients
- stochastic methods (*e.g.*, simulated annealing)
  - unclear convergence analysis [Bertsimas and Tsitsiklis, 1993]
  - several unintuitive tuning parameters
- grid search *e.g.*, for MR fingerprinting [Ma et al., 2013]

Motivation

Grid search computational costs

	L	$\sim$number dictionary atoms
(1-compartment) relaxivity	3	$\sim 100^2$

Motivation

Grid search computational costs

	L	$\sim$number dictionary atoms
(1-compartment) relaxivity	3	$\sim 100^2$
flow velocity	4	$\sim 100^3$
diffusivity tensor	7	$\sim 100^6$
2-3 compartment relaxivity	6-10	$\sim 100^5 - 100^9$

Motivation

Grid search computational costs

	L	$\sim$number dictionary atoms
(1-compartment) relaxivity	3	$\sim 100^2$
flow velocity	4	$\sim 100^3$
diffusivity tensor	7	$\sim 100^6$
2-3 compartment relaxivity	6-10	$\sim 100^5 - 100^9$

Can we scale computation with L more gracefully?

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \epsilon_1), \ldots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$ via signal model $\mathbf{s}$

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \dots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model $\mathbf{s}$
- design *nonlinear* functions $\widehat{x}_l(\cdot) := \widehat{h}_l(\cdot) + \widehat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\mathrm{Re}(\mathbf{y}_n)^\mathsf{T}, \mathrm{Im}(\mathbf{y}_n)^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T} \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \ldots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$ via signal model $\mathbf{s}$
- design *nonlinear* functions $\widehat{x}_l(\cdot) := \widehat{h}_l(\cdot) + \widehat{b}_l$ for $l \in \{1, \ldots, L\}$ that map each $\mathbf{q}_n := [\mathrm{Re}(\mathbf{y}_n)^\mathsf{T}, \mathrm{Im}(\mathbf{y}_n)^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T} \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg\min_{\substack{h_l \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 \right\}$$

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \dots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model $\mathbf{s}$
- design *nonlinear* functions $\widehat{x}_l(\cdot) := \widehat{h}_l(\cdot) + \widehat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\mathrm{Re}(\mathbf{y}_n)^\mathsf{T}, \mathrm{Im}(\mathbf{y}_n)^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T} \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg\min_{\substack{h_l \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 \right\}$$ ill-posed!

Machine Learning for QMRI Parameter Estimation

Idea: learn a *nonlinear* estimator from simulated training data

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \dots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model $\mathbf{s}$
- design *nonlinear* functions $\widehat{x}_l(\cdot) := \widehat{h}_l(\cdot) + \widehat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\mathrm{Re}(\mathbf{y}_n)^\mathsf{T}, \mathrm{Im}(\mathbf{y}_n)^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T} \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg \min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (2)$$

Solution: solve a *kernel ridge regression* (KRR) problem

- restrict function space over which we optimize
- include function regularization

A Function Space over which Optimization is Tractable

Hilbert space: complete inner product function space

A Function Space over which Optimization is Tractable

Hilbert space: complete inner product function space

Reproducing kernel Hilbert space (RKHS)

Hilbert space $\mathbb{H}$ over input space $\mathcal{Q}$ with *reproducing property*

$$\langle h, k(\cdot, \mathbf{q}) \rangle_{\mathbb{H}} = h(\mathbf{q}), \qquad \forall h \in \mathbb{H}, \mathbf{q} \in \mathcal{Q}$$

for some $k : \mathcal{Q}^2 \mapsto \mathbb{R}$ called a reproducing kernel (RK)

A Function Space over which Optimization is Tractable

Hilbert space: complete inner product function space

Reproducing kernel Hilbert space (RKHS)

Hilbert space $\mathbb{H}$ over input space $\mathcal{Q}$ with *reproducing property*

$$\langle h, k(\cdot, \mathbf{q}) \rangle_{\mathbb{H}} = h(\mathbf{q}), \qquad \forall h \in \mathbb{H}, \mathbf{q} \in \mathcal{Q}$$

for some $k : \mathcal{Q}^2 \mapsto \mathbb{R}$ called a reproducing kernel (RK)

Relevant facts

- Bijection between RKHS $\mathbb{H}$ and RK k [Aronszajn, 1950]
- Function $k(\cdot, \mathbf{q}) \in \mathbb{H}$ called a *feature mapping*

Function Optimization over a RKHS

Choose: RK $k : \mathcal{Q}^2 \mapsto \mathbb{R}$, which induces choice of RKHS $\mathbb{H}$

Function Optimization over a RKHS

Choose: RK $k : \mathcal{Q}^2 \mapsto \mathbb{R}$, which induces choice of RKHS $\mathbb{H}$

- *Nonlinear* kernel corresponds to *nonlinear* estimation
- We use $k(\mathbf{q}, \mathbf{q}') \leftarrow \exp\left(-\frac{1}{2}\left\|\mathbf{\Lambda}^{-1}(\mathbf{q} - \mathbf{q}')\right\|_2^2\right)$

Function Optimization over a RKHS

Choose: RK $k : \mathcal{Q}^2 \mapsto \mathbb{R}$, which induces choice of RKHS $\mathbb{H}$

Solve: for each desired latent parameter $l \in \{1, \ldots, L\}$,

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg \min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (3)$$

Function Optimization over a RKHS

Choose: RK $k : \mathcal{Q}^2 \mapsto \mathbb{R}$, which induces choice of RKHS $\mathbb{H}$

Solve: for each desired latent parameter $l \in \{1, \ldots, L\}$,

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg \min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (3)$$

- Optimal $\widehat{h}_l$ over $\mathbb{H}$ takes form [Schölkopf et al., 2001]

$$\widehat{h}_l(\cdot) \equiv \sum_{n=1}^{N} \widehat{a}_{l,n} k(\cdot, \mathbf{q}_n) \quad (4)$$

Function Optimization over a RKHS

Choose: RK $k : \mathcal{Q}^2 \mapsto \mathbb{R}$, which induces choice of RKHS $\mathbb{H}$

Solve: for each desired latent parameter $l \in \{1, \ldots, L\}$,

$$\left(\widehat{h}_l, \widehat{b}_l\right) \in \left\{ \arg\min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^{N} \left(h_l(\mathbf{q}_n) + b_l - x_{l,n}\right)^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (3)$$

- Optimal $\widehat{h}_l$ over $\mathbb{H}$ takes form [Schölkopf et al., 2001]

$$\widehat{h}_l(\cdot) \equiv \sum_{n=1}^{N} \widehat{a}_{l,n} k(\cdot, \mathbf{q}_n) \quad (4)$$

- Plug (4) into (3); solve now instead for $\left(\widehat{a}_l, \widehat{b}_l\right)$; construct:

$$\widehat{x}_l(\cdot) = \sum_{n=1}^{N} \widehat{a}_{l,n} k(\cdot, \mathbf{q}_n) + \widehat{b}_l \quad (5)$$

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (6)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ training pt regressands

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M} (\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (6)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ training pt regressands
- $\mathbf{K} := \begin{bmatrix} k(\mathbf{q}_1, \mathbf{q}_1) & \cdots & k(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{q}_N, \mathbf{q}_1) & \cdots & k(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix}$ Gram matrix

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \tag{6}$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ — training pt regressands
- $\mathbf{K} := \begin{bmatrix} k(\mathbf{q}_1, \mathbf{q}_1) & \cdots & k(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{q}_N, \mathbf{q}_1) & \cdots & k(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix}$ — Gram matrix
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\mathsf{T}$ — de-meaning operator

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N}\mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N}\mathbf{K}\mathbf{1}_N \right) \right) \quad (6)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ — training pt regressands
- $\mathbf{K} := \begin{bmatrix} k(\mathbf{q}_1, \mathbf{q}_1) & \cdots & k(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{q}_N, \mathbf{q}_1) & \cdots & k(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix}$ — Gram matrix
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\mathsf{T}$ — de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^\mathsf{T}$ — nonlin kernel embedding

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (6)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ training pt regressands
- $\mathbf{K} := \begin{bmatrix} k(\mathbf{q}_1, \mathbf{q}_1) & \cdots & k(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{q}_N, \mathbf{q}_1) & \cdots & k(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix}$ Gram matrix
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\mathsf{T}$ de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^\mathsf{T}$ nonlin kernel embedding

Can we scale computation with L more gracefully?

- Yes, in fact (6) separable in $l \in \{1, \dots, L\}$ by construction

MRI Parameter Estimation via KRR

Non-iterative closed-form solution, for $l \in \{1, \dots, L\}$:

$$\widehat{x}_l(\cdot) = \mathbf{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M} (\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (6)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$ training pt regressands
- $\mathbf{K} := \begin{bmatrix} k(\mathbf{q}_1, \mathbf{q}_1) & \cdots & k(\mathbf{q}_1, \mathbf{q}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{q}_N, \mathbf{q}_1) & \cdots & k(\mathbf{q}_N, \mathbf{q}_N) \end{bmatrix}$ Gram matrix
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\mathsf{T}$ de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^\mathsf{T}$ nonlin kernel embedding

Can we scale computation with L more gracefully?

- Yes, in fact (6) separable in $l \in \{1, \dots, L\}$ by construction
- However, explicitly computing $\mathbf{K}$ may be undesirable...

KRR as High-Dimensional Affine Regression

Suppose there exists "approximate feature mapping" $\tilde{\mathbf{z}} : \mathcal{Q} \mapsto \mathbb{R}^Z$ such that $\tilde{\mathbf{Z}} := [\tilde{\mathbf{z}}(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}(\mathbf{q}_N)]$ has for $\dim(\mathcal{Q}) \ll Z \ll N$

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^{\mathsf{T}} \tilde{\mathbf{Z}}. \tag{7}$$

KRR as High-Dimensional Affine Regression

Suppose there exists "approximate feature mapping" $\tilde{\mathbf{z}} : \mathcal{Q} \mapsto \mathbb{R}^Z$ such that $\tilde{\mathbf{Z}} := [\tilde{\mathbf{z}}(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}(\mathbf{q}_N)]$ has for $\dim(\mathcal{Q}) \ll Z \ll N$

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^\mathsf{T} \tilde{\mathbf{Z}}. \tag{7}$$

Plugging (7) into KRR solution (6) and rearranging gives

$$\widehat{x}_l(\cdot) \approx \frac{1}{N} \mathbf{x}_l^\mathsf{T} \mathbf{1}_N + \frac{1}{N} \mathbf{x}_l^\mathsf{T} \mathbf{M} \tilde{\mathbf{Z}}^\mathsf{T} \left(\frac{1}{N} \tilde{\mathbf{Z}} \mathbf{M} \tilde{\mathbf{Z}}^\mathsf{T} + \rho_l \mathbf{I}_Z \right)^{-1} \left(\tilde{\mathbf{z}}(\cdot) - \frac{1}{N} \tilde{\mathbf{Z}} \mathbf{1}_N \right)$$

KRR as High-Dimensional Affine Regression

Suppose there exists "approximate feature mapping" $\tilde{\mathbf{z}} : \mathcal{Q} \mapsto \mathbb{R}^Z$ such that $\tilde{\mathbf{Z}} := [\tilde{\mathbf{z}}(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}(\mathbf{q}_N)]$ has for $\dim(\mathcal{Q}) \ll Z \ll N$

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^\mathsf{T} \tilde{\mathbf{Z}}. \tag{7}$$

Plugging (7) into KRR solution (6) and rearranging gives

$$\widehat{x}_l(\cdot) \approx \widehat{m}_{x_l} + \widehat{\mathbf{c}}_{x_l \tilde{\mathbf{z}}}^\mathsf{T} \left(\widehat{\mathbf{C}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} + \rho_l \mathbf{I}_Z \right)^{-1} (\tilde{\mathbf{z}}(\cdot) - \widehat{\mathbf{m}}_{\tilde{\mathbf{z}}}) \tag{8}$$

which is regularized ("ridge") Z-dimensional affine regression!

KRR as High-Dimensional Affine Regression

Suppose there exists "approximate feature mapping" $\tilde{\mathbf{z}} : \mathcal{Q} \mapsto \mathbb{R}^Z$ such that $\tilde{\mathbf{Z}} := [\tilde{\mathbf{z}}(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}(\mathbf{q}_N)]$ has for $\dim(\mathcal{Q}) \ll Z \ll N$

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^{\mathsf{T}} \tilde{\mathbf{Z}}. \qquad (7)$$

Plugging (7) into KRR solution (6) and rearranging gives

$$\widehat{x}_l(\cdot) \approx \widehat{m}_{x_l} + \widehat{\mathbf{c}}_{x_l \tilde{\mathbf{z}}}^{\mathsf{T}} \Big(\widehat{\mathbf{C}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} + \rho_l \mathbf{I}_Z \Big)^{-1} (\tilde{\mathbf{z}}(\cdot) - \widehat{\mathbf{m}}_{\tilde{\mathbf{z}}}) \qquad (8)$$

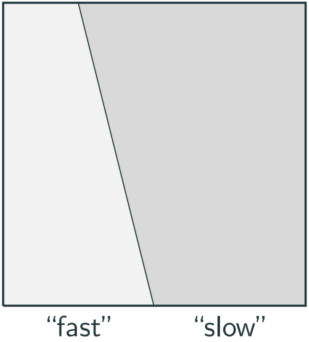
which is regularized ("ridge") Z-dimensional affine regression!

Does such a $\tilde{\mathbf{z}}$ exist and work well in practice?

- Yes, *e.g.* for "shift invariant" kernels (like our Gaussian) of form $k(\mathbf{q}, \mathbf{q}') \equiv k(\mathbf{q} - \mathbf{q}')$ [Rahimi and Recht, 2007]
- In such cases, can reduce from $\sim N^2$ to $\sim NZ$ computations

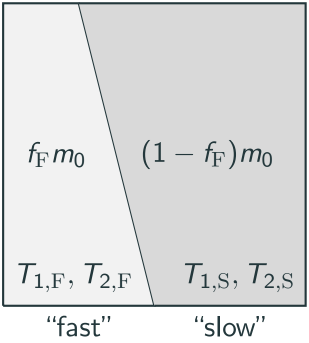
Application: Myelin Water Fraction (MWF) Imaging

simple two-compartment model



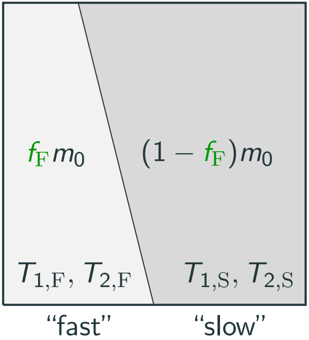
Application: Myelin Water Fraction (MWF) Imaging

simple two-compartment model



Application: Myelin Water Fraction (MWF) Imaging

simple two-compartment model



Goal: rapidly estimate f_{F} (proxy for MWF) in white matter (WM)

Application: MWF Imaging

Problem dimensions (per voxel)

- $\mathbf{x} \leftarrow [f_{\mathrm{F}}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, m_0]^{\mathsf{T}}$
- $\boldsymbol{\nu} \leftarrow$ flip angle variation
- $\mathbf{y} \leftarrow$ voxel values from 10 datasets [Nataraj et al., 2017]

Application: MWF Imaging

Problem dimensions (per voxel)

- $\mathbf{x} \leftarrow [f_{\mathrm{F}}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, m_0]^{\mathsf{T}}$
- $\nu \leftarrow$ flip angle variation
- $\mathbf{y} \leftarrow$ voxel values from 10 datasets [Nataraj et al., 2017]

Use KRR to estimate just f_{F}

- Separable prior on $\mathbf{x}$: f_{F}, m_0 uniform; others log-uniform
- $N \leftarrow 10^6$ training points
- $Z \leftarrow 10^3$ kernel approximation order

Application: MWF Imaging

Problem dimensions (per voxel)

- $\mathbf{x} \leftarrow [f_{\mathrm{F}}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, m_0]^{\mathsf{T}}$
- $\nu \leftarrow$ flip angle variation
- $\mathbf{y} \leftarrow$ voxel values from 10 datasets [Nataraj et al., 2017]

Use KRR to estimate just f_{F}

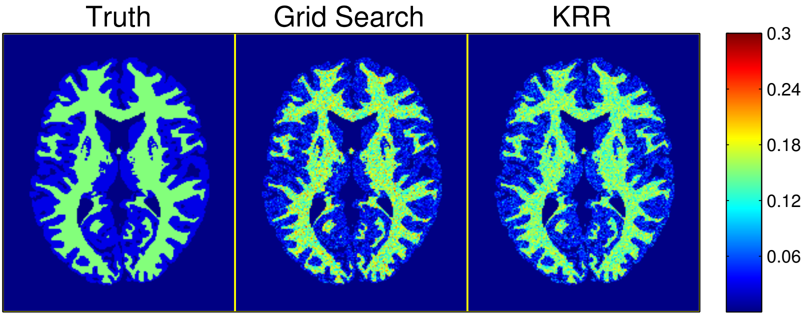
- Separable prior on $\mathbf{x}$: f_{F}, m_0 uniform; others log-uniform
- $N \leftarrow 10^6$ training points
- $Z \leftarrow 10^3$ kernel approximation order

Compare against grid search

- unconstrained search would require $\sim 100^5$ dictionary atoms
- we artificially constrain search here to limit computation

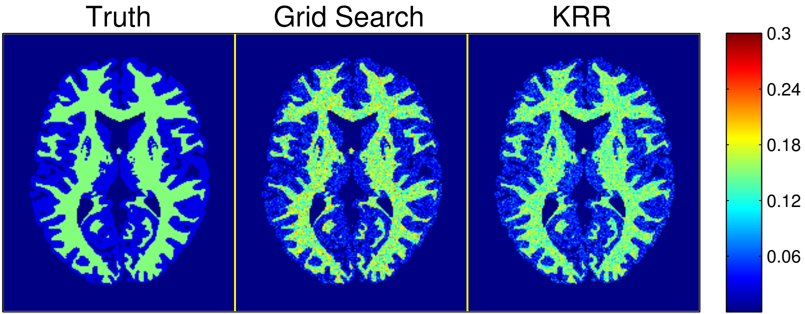
MWF Imaging: Simulation Result

Fast-fraction f_{F} estimates, in simulation:



MWF Imaging: Simulation Result

Fast-fraction f_{F} estimates, in simulation:



$\sim$4h 40s training, 2s testing

MWF Imaging: Proof-of-concept In Vivo Result

Fast-fraction f_{F} estimates, from 3D Cartesian data

MWF Imaging: Proof-of-concept In Vivo Result

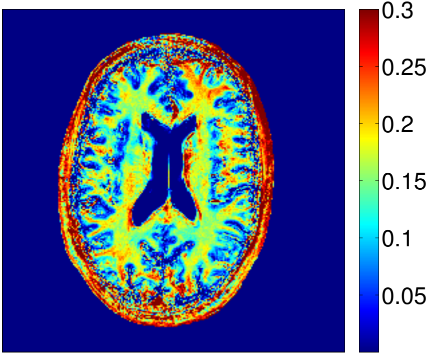
Fast-fraction f_{F} estimates, from 3D Cartesian data

- Full-scale grid search intractable on typical desktop

MWF Imaging: Proof-of-concept In Vivo Result

Fast-fraction f_{F} estimates, from 3D Cartesian data

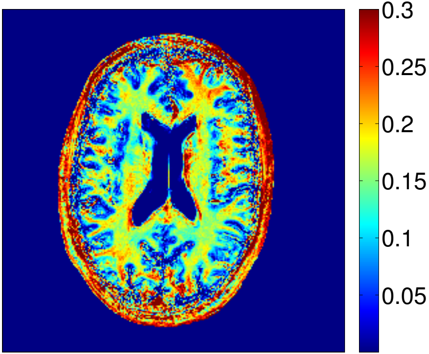
- Full-scale grid search intractable on typical desktop
- KRR estimates in single slice took about **70s**



MWF Imaging: Proof-of-concept In Vivo Result

Fast-fraction f_{F} estimates, from 3D Cartesian data

- Full-scale grid search intractable on typical desktop
- KRR estimates in single slice took about **70s**
- KRR MWF estimates in WM comparable to literature



Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation

Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation
    - Key insight: even with complicated MR signal models, can simulate training points “for free”

Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation
  - Key insight: even with complicated MR signal models, can simulate training points "for free"
  - Convert *nonlinear estimation* problem into *nonlinear regression* problem that we solve in closed-form with kernels

Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation
  - Key insight: even with complicated MR signal models, can simulate training points “for free”
  - Convert *nonlinear estimation* problem into *nonlinear regression* problem that we solve in closed-form with kernels
- Proof-of-concept *in vivo* application to MWF imaging

Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation
  - Key insight: even with complicated MR signal models, can simulate training points "for free"
  - Convert *nonlinear estimation* problem into *nonlinear regression* problem that we solve in closed-form with kernels
- Proof-of-concept *in vivo* application to MWF imaging

Ongoing work

- Conceptual: model selection, performance analysis
- Experimental: validation studies

Summary

Contributions

- Fast KRR method for nonlin MRI multiparameter estimation
  - Key insight: even with complicated MR signal models, can simulate training points "for free"
  - Convert *nonlinear estimation* problem into *nonlinear regression* problem that we solve in closed-form with kernels
- Proof-of-concept *in vivo* application to MWF imaging

Ongoing work

- Conceptual: model selection, performance analysis
- Experimental: validation studies

References i

Aronszajn, N. (1950).
Theory of reproducing kernels.
Trans. Amer. Math. Soc., 68(3):337–404.

Bertsimas, D. and Tsitsiklis, J. (1993).
Simulated annealing.
Statistical Science, 8(1):10–15.

Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. A. (2013).
Magnetic resonance fingerprinting.
Nature, 495:187–93.

References ii

Nataraj, G., Nielsen, J.-F., and Fessler, J. A. (2017).
Myelin water fraction estimation from optimized steady-state sequences using kernel ridge regression.
In *Proc. Intl. Soc. Mag. Res. Med.*
To appear.

Rahimi, A. and Recht, B. (2007).
Random features for large-scale kernel machines.
In *NIPS*.

Schölkopf, B., Herbrich, R., and Smola, A. J. (2001).
A generalized representer theorem.
In *Proc. Computational Learning Theory (COLT)*, pages 416–426.
LNCS 2111.

Backup: An Overview of Model Selection

Some model parameters **require manual selection**...

- Kernel shape $k(\mathbf{q}, \mathbf{q}') \leftarrow \exp\left(-\frac{1}{2}\left\|\boldsymbol{\Lambda}^{-1}(\mathbf{q} - \mathbf{q}')\right\|_2^2\right)$

Backup: An Overview of Model Selection

Some model parameters **require manual selection**...

- Kernel shape $k(\mathbf{q}, \mathbf{q}') \leftarrow \exp\left(-\frac{1}{2}\left\|\mathbf{\Lambda}^{-1}(\mathbf{q} - \mathbf{q}')\right\|_2^2\right)$
- Prior on $\mathbf{x}$ from tissue properties
- N, Z empirical methods

Backup: An Overview of Model Selection

Some model parameters **require manual selection**...

- Kernel shape $k(\mathbf{q}, \mathbf{q}') \leftarrow \exp\left(-\frac{1}{2}\left\|\boldsymbol{\Lambda}^{-1}(\mathbf{q} - \mathbf{q}')\right\|_2^2\right)$
- Prior on $\mathbf{x}$ from tissue properties
- N, Z empirical methods

...but others **tuned automatically**

- Kernel smoothing length-scale $\boldsymbol{\Lambda} \leftarrow \mathrm{diag}\left(\sum_{n=1}^{N} \mathbf{q}_n\right)$
- Regularization parameters $\rho_l \leftarrow \frac{1}{N^2}\mathbf{x}_l^\mathsf{T}\mathbf{M}\mathbf{x}_l$
- Prior on known $\boldsymbol{\nu}$ density estimation