DICTIONARY-FREE MRI PARAMETER ESTIMATION VIA KERNEL RIDGE REGRESSION

Gopal Nataraj[⋆] *Jon-Fredrik Nielsen*[†] *Jeffrey A. Fessler*[⋆]

[⋆]Dept. of Electrical Engineering and Computer Science, University of Michigan
[†]Dept. of Biomedical Engineering, University of Michigan

ABSTRACT

MRI parameter quantification has diverse applications, but likelihood-based methods typically require nonconvex optimization due to nonlinear signal models. To avoid expensive grid searches used in prior works, we propose to learn a nonlinear estimator from simulated training examples and (approximate) kernel ridge regression. As proof of concept, we apply kernel-based estimation to quantify six parameters per voxel describing the steady-state magnetization dynamics of two water compartments from simulated data. In relevant regions of fast-relaxing compartmental fraction estimates, kernel estimation achieves comparable mean-squared error as grid search, with dramatically reduced computation.

1. INTRODUCTION

Quantitative MRI describes a class of problems in experimental design and parameter estimation that seek to produce parameter "maps" describing physical processes of interest, *e.g.*, relaxation [1], diffusion [2], and myelin water content [3]. Motivated by such widespread applications, this paper describes a novel method for fast MRI parameter estimation.

Because MR signal models are often complicated nonlinear functions of both desired and nuisance parameters, likelihood-based estimation requires nonconvex optimization in general. To seek global optima, several recent works [4–7] approach estimation via grid search, which requires either storing or computing on-the-fly a "dictionary" of signal vectors. In these works, the number of estimated parameters is small, so grid search is suitable. However, for even moderately sized problems, the required number of dictionary elements renders grid search undesirable or even intractable, unless one assumes restrictive parameter constraints.

We observe that for many voxel-wise separable MRI parameter estimation problems, *e.g.* [4–7], training points are easily simulated. To address the challenges of likelihood-based nonlinear estimation, we thus propose to *learn* an estimator from simulated training examples. Specifically, this paper describes a fast MRI parameter estimation method based on kernel ridge regression (KRR). Though popular in the machine learning community, KRR has not (to our knowledge) been used before for MRI parameter mapping.

This paper is organized as follows. Section 2 introduces a general signal model, solves an associated functional optimization problem via KRR, and reduces the candidate estimator's computational requirements through an approximation. Section 3 applies kernel-based estimation to quantify six parameters arising from models describing the steady-state magnetization dynamics of two water compartments. Section 4 provides brief concluding remarks.

2. METHODS

2.1. MRI Parameter Estimation via Kernel Regression

After image reconstruction, many MRI acquisitions useful for parameter mapping produce at each voxel position a sequence of noisy voxel values $\mathbf{y} \in \mathbb{C}^D$, modeled here as

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

where $\mathbf{x} \in \mathbb{R}^L$ denotes L *latent* object parameters (*e.g.*, relaxation time constants); $\boldsymbol{\nu} \in \mathbb{R}^K$ denotes K *known* object parameters (*e.g.*, separately acquired and estimated field maps); $\mathbf{s} : \mathbb{R}^L \times \mathbb{R}^K \mapsto \mathbb{C}^D$ models the noiseless signals that arise from D datasets; and $\boldsymbol{\epsilon} \in \mathbb{C}^D$ is complex Gaussian noise, assumed to be distributed as $\mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Here we seek to estimate on a per-voxel basis each latent parameter $\mathbf{x}$ from corresponding data sequence $\mathbf{y}$ and known parameter $\boldsymbol{\nu}$.

To develop (or train) a non-iterative estimator $\widehat{\mathbf{x}}$, we simulate many instances of input-output relation (1) and use KRR to estimate a nonlinear inverse relation. We sample $\mathbb{R}^L \times \mathbb{R}^K \times \mathbb{C}^D$ and evaluate (1) N times to produce sets of training inputs $\{(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \ldots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)\}$ and data sequences $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. We seek a function $\widehat{\mathbf{h}} : \mathbb{R}^Q \mapsto \mathbb{R}^L$ for $Q := 2D + K$ and an offset $\widehat{\mathbf{b}} \in \mathbb{R}^L$ that together map each (real) $\mathbf{q}_n := [\mathrm{Re}(\mathbf{y}_n)^\mathsf{T}, \mathrm{Im}(\mathbf{y}_n)^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T}$, $n \in \{1, \ldots, N\}$, to an estimate $\widehat{\mathbf{x}}(\mathbf{q}_n) := \widehat{\mathbf{h}}(\mathbf{q}_n) + \widehat{\mathbf{b}}$ that is "close" to $\mathbf{x}_n$:

$$\left(\widehat{\mathbf{h}}, \widehat{\mathbf{b}}\right) \in \arg \min_{\substack{\mathbf{h} \in \mathbb{H}^L \\ \mathbf{b} \in \mathbb{R}^L}} \Psi\Big(\mathbf{h}, \mathbf{b}; \{(\mathbf{x}_n, \mathbf{q}_n)\}_1^N\Big), \text{ where} \tag{2}$$

Supported in part by NIH grant P01 CA87634 and in part by the University of Michigan MCubed funding program. Email: {gnataraj | jfnielse | fessler}@umich.edu.

$$\Psi(\dots) = \sum_{l=1}^{L} \Psi_l\Big(h_l, b_l; \{(x_{l,n}, \mathbf{q}_n)\}_1^N\Big); \tag{3}$$

$$\Psi_l(\dots) = \frac{1}{N}\sum_{n=1}^{N}(h_l(\mathbf{q}_n) + b_l - x_{l,n})^2 + \rho_l \|h_l\|_{\mathbb{H}}^2. \tag{4}$$

Here, each $h_l : \mathbb{R}^Q \mapsto \mathbb{R}$ is a scalar function that maps to the lth component of the output of $\mathbf{h}$; each $b_l, x_{l,n} \in \mathbb{R}$ are scalar components of $\mathbf{b}, \mathbf{x}$; $\mathbb{H}$ is a (reproducing kernel) Hilbert space (RKHS), whose norm $\|\cdot\|_{\mathbb{H}}$ is induced by inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}} : \mathbb{H}\times\mathbb{H}\mapsto\mathbb{R}$; each ρ_l controls for regularity in h_l; and $(\cdot)^{\mathsf{T}}$ denotes vector transpose.

Since (3) is separable in the components of $\mathbf{h}$, it suffices to consider optimizing each (h_l, b_l) separately via (4). Remarkably, a generalization of the Representer Theorem [8], restated as is relevant here for completeness, reduces (4) to a finite-dimensional optimization problem:

Theorem 1 (Generalized Representer, [8])**.** *Define $k : \mathbb{R}^Q \times \mathbb{R}^Q \mapsto \mathbb{R}$ to be the (symmetric positive definite) kernel function associated with RKHS $\mathbb{H}$, such that reproducing property $h_l(\mathbf{q}) = \langle h_l, k(\cdot,\mathbf{q})\rangle_{\mathbb{H}}$ holds for all $h_l \in \mathbb{H}$ and $\mathbf{q} \in \mathbb{R}^Q$. Then any minimizer $(\widehat{h}_l, \widehat{b}_l)$ of* (4) *over $\mathbb{H}\times\mathbb{R}$ admits a representation of $\widehat{h}_l$ of the form*

$$\widehat{h}_l(\cdot) \equiv \sum_{n=1}^{N} a_{l,n} k(\cdot, \mathbf{q}_n), \tag{5}$$

where each $a_{l,n} \in \mathbb{R}$ for $n \in \{1,\dots,N\}$.

Theorem 1 ensures that any solution to

$$(\widehat{\mathbf{a}}_l, \widehat{b}_l) \in \arg\min_{\substack{\mathbf{a}_l\in\mathbb{R}^N \\ b_l\in\mathbb{R}}} \rho_l \left\| \sum_{n'=1}^{N} a_{l,n'} k(\cdot, \mathbf{q}_{n'}) \right\|_{\mathbb{H}}^2 + \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{n'=1}^{N} a_{l,n'} k(\mathbf{q}_n, \mathbf{q}_{n'}) + b_l - x_{l,n}\right)^2 \tag{6}$$

corresponds via (5) to a minimizer of (4) over $\mathbb{H}\times\mathbb{R}$, where $\mathbf{a}_l := [a_{l,1},\dots,a_{l,N}]^{\mathsf{T}}$. Fortunately, a solution of (6) exists uniquely for $\rho_l > 0$ and can be expressed as

$$\widehat{\mathbf{a}}_l = (\mathbf{M}\mathbf{K} + N\rho_l \mathbf{I}_N)^{-1}\mathbf{M}\boldsymbol{x}_l; \tag{7}$$

$$\widehat{b}_l = \frac{1}{N}\mathbf{1}_N^{\mathsf{T}}(\boldsymbol{x}_l - \mathbf{K}\widehat{\mathbf{a}}_l), \tag{8}$$

where $\mathbf{K} \in \mathbb{R}^{N\times N}$ is the Gram matrix consisting of entries $k(\mathbf{q}_n, \mathbf{q}_{n'})$ for $n, n' \in \{1,\dots,N\}$; $\mathbf{M} := \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathsf{T}}$ is a de-meaning operator; $\boldsymbol{x}_l := [x_{l,1},\dots,x_{l,N}]^{\mathsf{T}}$; $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the identity matrix; and $\mathbf{1}_N \in \mathbb{R}^N$ is a vector of ones.

In the special case where each $\rho_l \leftarrow \rho$ for fixed $\rho > 0$, the scalar estimators $\left\{\widehat{x}_l(\cdot) := \widehat{h}_l(\cdot) + \widehat{b}_l\right\}_1^L$ that arise from plugging (7) into (5) can be concatenated as

$$\widehat{\mathbf{x}}(\cdot) \leftarrow \mathbf{X}\left(\frac{1}{N}\mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho\mathbf{I}_N)^{-1}\mathbf{k}(\cdot)\right), \tag{9}$$

where $\mathbf{k}(\cdot) := [k(\cdot,\mathbf{q}_1),\dots,k(\cdot,\mathbf{q}_N)]^{\mathsf{T}} - \frac{1}{N}\mathbf{K}\mathbf{1}_N : \mathbb{R}^Q \mapsto \mathbb{R}^N$ and $\mathbf{X} := [\mathbf{x}_1,\dots,\mathbf{x}_N] = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_L]^{\mathsf{T}} \in \mathbb{R}^{L\times N}$.

For $\rho > 0$, estimator (9) minimizes (3) over $(\mathbb{H}\times\mathbb{R})^L$. However, the utility of (9) depends on the choice of kernel k, which induces a choice on the RKHS $\mathbb{H}$ and thus the function space $(\mathbb{H}\times\mathbb{R})^L$ over which (2) optimizes. For example, if k was selected as the canonical dot product $k(\mathbf{q},\mathbf{q}') \leftarrow \langle\mathbf{q},\mathbf{q}'\rangle_{\mathbb{R}^Q} := \mathbf{q}^{\mathsf{T}}\mathbf{q}'$ (for which RKHS $\mathbb{H} \leftarrow \mathbb{R}^Q$), then (9) would reduce to affine ridge regression [9], which is optimal over $\left(\mathbb{R}^Q\times\mathbb{R}\right)^L$ but is unlikely to be useful when signal model $\mathbf{s}$ is nonlinear in $\mathbf{x}$.

Since we expect a useful estimate $\widehat{\mathbf{x}}(\mathbf{q})$ to depend nonlinearly (but smoothly) on $\mathbf{q}$ in general, we instead use a (symmetric, positive definite) k that is likewise nonlinear in its arguments and thus corresponds to a RKHS richer than $\mathbb{R}^Q$. Specifically, we use a Gaussian kernel

$$k(\mathbf{q},\mathbf{q}') \leftarrow \exp\left(-\frac{1}{2}\|\mathbf{q}-\mathbf{q}'\|_{\boldsymbol{\Lambda}^{-1/2}}^2\right), \tag{10}$$

where symmetric, positive definite $\boldsymbol{\Lambda} \in \mathbb{R}^{Q\times Q}$ controls the length scales in $\mathbf{q}$ over which the estimator $\widehat{\mathbf{x}}$ smooths.

Interestingly, the RKHS associated with Gaussian kernel (10) is infinite-dimensional. Thus, Gaussian KRR can be interpreted as first "lifting" via a nonlinear *feature map* $\mathbf{z} : \mathbb{R}^Q \mapsto \mathbb{H}$ each $\mathbf{q}$ into an infinite-dimensional *feature* $\mathbf{z}(\mathbf{q}) \in \mathbb{H}$, and then performing affine ridge regression on the features via dot products of the form $k(\mathbf{q},\mathbf{q}') \leftarrow \langle\mathbf{z}(\mathbf{q}),\mathbf{z}(\mathbf{q}')\rangle_{\mathbb{H}}$. From this perspective, the challenges of nonlinear estimation via likelihood models are avoided because we *select* (through the choice of kernel) characteristics of the nonlinear dependence that we wish to model and need only *estimate* via (6) the linear dependence of $\widehat{\mathbf{x}}$ on the corresponding features.

2.2. Practical Considerations: Kernel Approximations

In practical problems with even moderately large ambient dimension Q, the necessarily large number of training samples N complicates storage of (dense) Gram matrix $\mathbf{K}$. Using a kernel approximation can mitigate storage issues. Here we choose to sample Random (Fourier) Features [10], a recent method for approximating shift-invariants kernels having form $k(\mathbf{q},\mathbf{q}') \equiv k(\mathbf{q}-\mathbf{q}')$. This section briefly reviews [10] for the purpose of constructing an intuitive and computationally efficient approximation of (9).

The strategy of [10] is to construct independent probability distributions $p_{\boldsymbol{v}}$ and p_s associated with random $\boldsymbol{v} \in \mathbb{R}^Q$ and random $s \in \mathbb{R}$ as well as a random function (that is parameterized by $\mathbf{q}$) $\tilde{z}(\cdot,\cdot;\mathbf{q}) : \mathbb{R}^Q\times\mathbb{R}\times\mathbb{R}^Q \mapsto \mathbb{R}$, such that

$$\mathrm{E}_{\boldsymbol{v},s}(\tilde{z}(\boldsymbol{v},s;\mathbf{q})\tilde{z}(\boldsymbol{v},s;\mathbf{q}')) = k(\mathbf{q}-\mathbf{q}'), \tag{11}$$

where $\mathrm{E}_{\boldsymbol{v},s}(\cdot)$ denotes expectation with respect to $p_{\boldsymbol{v}}p_s$. If such a construction exists, one can build approximate feature maps $\tilde{\mathbf{z}}_Z$ by concatenating evaluations of $\tilde{z}_Z := \sqrt{2/Z}\tilde{z}$ on Z

samples $\{(\boldsymbol{v}_1, s_1), \ldots, \boldsymbol{v}_Z, s_Z\}$ of $(\boldsymbol{v}, s)$ (drawn jointly albeit independently), to produce approximate feature vectors

$$\tilde{\mathbf{z}}_Z(\mathbf{q}) := [\tilde{z}_Z(\boldsymbol{v}_1, s_1; \mathbf{q}), \ldots, \tilde{z}_Z(\boldsymbol{v}_Z, s_Z; \mathbf{q})]^\mathsf{T} \tag{12}$$

for any $\mathbf{q}$. Then by the strong law of large numbers,

$$\lim_{Z\to\infty} \langle \tilde{\mathbf{z}}_Z(\mathbf{q}), \tilde{\mathbf{z}}_Z(\mathbf{q}') \rangle_{\mathbb{R}^Z} \overset{a.s.}{\to} k(\mathbf{q}, \mathbf{q}') \qquad \forall \mathbf{q}, \mathbf{q}', \tag{13}$$

which, in conjunction with strong performance guarantees for finite Z [10], justifies the interpretation of $\tilde{\mathbf{z}}_Z$ as an approximate (and now finite-dimensional) feature map.

We use the Fourier construction of [10] that assigns $\tilde{z}(\boldsymbol{v}, s; \mathbf{q}) \leftarrow \cos\big(2\pi(\boldsymbol{v}^\mathsf{T}\mathbf{q} + s)\big)$. If $s \sim \mathrm{unif}(0, 1)$, then $\mathrm{E}_{\boldsymbol{v},s}(\tilde{z}(\boldsymbol{v}, s; \mathbf{q})\tilde{z}(\boldsymbol{v}, s; \mathbf{q}'))$ simplifies to

$$\int_{\mathbb{R}^Q} \cos\big(2\pi\boldsymbol{v}^\mathsf{T}(\mathbf{q} - \mathbf{q}')\big) p_{\boldsymbol{v}}(\boldsymbol{v}) \,\mathrm{d}\,\boldsymbol{v}. \tag{14}$$

For symmetric positive definite k, (14) exists [11] and is the Fourier transform of $p_{\boldsymbol{v}}$. Thus, for Gaussian kernel (10), choosing $\boldsymbol{v} \sim \mathcal{N}\Big(\mathbf{0}, \big(4\pi^2\boldsymbol{\Lambda}\big)^{-1}\Big)$ satisfies (11).

Subsequent sampling of $\boldsymbol{v}, s$ and construction via (12) of $\tilde{\mathbf{Z}}_Z := [\tilde{\mathbf{z}}_Z(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}_Z(\mathbf{q}_N)] \in \mathbb{R}^{Z\times N}$ produces for $Z \ll N$ a low-rank approximation $\tilde{\mathbf{Z}}_Z^\mathsf{T}\tilde{\mathbf{Z}}_Z$ of Gram matrix $\mathbf{K}$. Substituting this approximation into (9) and applying the matrix inversion lemma [12] yields

$$\widehat{\mathbf{x}}(\cdot) \leftarrow \mathbf{m}_{\mathbf{x}} + \mathbf{C}_{\mathbf{x}\tilde{\mathbf{z}}}(\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} + \rho\mathbf{I}_Z)^{-1}(\tilde{\mathbf{z}}_Z(\cdot) - \mathbf{m}_{\tilde{\mathbf{z}}}), \tag{15}$$

where $\mathbf{m}_{\mathbf{x}} := \frac{1}{N}\mathbf{X}\mathbf{1}_N$ and $\mathbf{m}_{\tilde{\mathbf{z}}} := \frac{1}{N}\tilde{\mathbf{Z}}_Z\mathbf{1}_N$ are sample mean vectors; and $\mathbf{C}_{\mathbf{x}\tilde{\mathbf{z}}} := \frac{1}{N}\mathbf{X}\mathbf{M}\tilde{\mathbf{Z}}_Z^\mathsf{T}$ and $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} := \frac{1}{N}\tilde{\mathbf{Z}}_Z\mathbf{M}\tilde{\mathbf{Z}}_Z^\mathsf{T}$ are sample covariance matrices. Estimator (15) is a regularized variation of the linear minimum mean-squared error estimator on the features, and illustrates that Gaussian KRR via estimator (9) is asymptotically (in Z) equivalent to affine ridge regression after nonlinear, high-dimensional feature mapping.

3. EXPERIMENT

As proof of concept, we apply kernel-based estimation to quantify parameters describing the magnetization dynamics of multiple water compartments, a challenging application of clinical interest, *e.g.*, for myelin water imaging [13]. Specifically, we use a simple model of two non-exchanging compartments and seek to estimate the associated latent parameters; more complex models would only add parameters and thereby increase the need for an alternative to grid search.

Such challenging estimation problems typically require multiple data acquisitions and thus long scans. To reduce scan times of classical methods [3], fast steady-state pulse sequences were recently proposed for two-compartment parameter estimation [14, 15]. We similarly take interest in steady-state sequences, but modify acquisition details to address possible concerns of insufficient estimation precision [16].

We begin with two-compartment models of the signals arising from Spoiled Gradient-Recalled Echo (SPGR) [17] and Dual-Echo Steady-State (DESS) [18] pulse sequences. We make appropriate assumptions to reduce model dependencies to seven free parameters per voxel: flip angle spatial variation (due to transmit field inhomogeneity) κ; fast-relaxing compartmental fraction f_F; (spin-lattice, spin-spin) relaxation time constants for the fast-relaxing $(T_{1,\mathrm{F}}, T_{2,\mathrm{F}})$ and slow-relaxing $(T_{1,\mathrm{S}}, T_{2,\mathrm{S}})$ compartments; and a complex proportionality constant[1] m_0. We assume prior knowledge of $\boldsymbol{\nu} \leftarrow \kappa$ (which in practice can be estimated from separate fast acquisitions, *e.g.* [19]) and collect the remaining $L \leftarrow 6$ latent parameters as $\mathbf{x} \leftarrow [f_\mathrm{F}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, m_0]^\mathsf{T}$.

In light of clinical need, we focus on estimating fast-relaxing fraction f_F and tailor our simulation accordingly. We optimize (by the method of [20]) the flip angles and repetition times of four SPGR and three DESS scans[2] for precise estimation of f_F in white matter (WM) and grey matter (GM) regions of the human brain. We consider the other five latent parameters to be nuisance parameters and thus do not evaluate the performance of their estimators in the following.

We simulate data to arise from two non-exchanging water pools with nominal fast $(T_{1,\mathrm{F}}, T_{2,\mathrm{F}}) \leftarrow (500, 20)$ms and slow $(T_{1,\mathrm{S}}, T_{2,\mathrm{S}}) \leftarrow (1000, 80)$ms relaxation time constants selected from prior measurements [3, 15]. We assign fast-compartment fractions $f_\mathrm{F} \leftarrow 0.15$ in WM and $f_\mathrm{F} \leftarrow 0.03$ in GM and constrain slow-compartment fractions as $1 - f_\mathrm{F}$. We prescribe these parameter values to the anatomy of the BrainWeb digital phantom [21] to produce ground truth parameter maps. Using optimized acquisition parameters and allowing κ to model $\pm 20\%$ flip angle variation, we apply two-compartment SPGR and DESS models to the 81st axial slice of the true parameter maps. We corrupt these (complex) noiseless signals with additive complex Gaussian noise whose covariance $\boldsymbol{\Sigma} \leftarrow \big(1.49 \times 10^{-7}\big)\mathbf{I}_{D\leftarrow 10}$ reflects measurements from normalized datasets [20]. This yields realistically noisy 217×181 (image-domain) datasets ranging from 24.3-48.8dB SNR in WM and 26.4-49.5dB SNR in GM.

Because the first and second DESS signals depend differently on phase accrual due to off-resonance effects [20], off-resonance related phase (unlike signal loss) cannot be collected into m_0. To avoid (separate or joint) estimation of an off-resonance field map, we elect to estimate parameters using magnitude SPGR and DESS image data and account for consequently Rician-distributed noise during training.

To sample training points, we assume prior distributions on latent parameters $\mathbf{x}$ (that distinguish the two compartments but are otherwise conservative) and directly measure known parameter $\boldsymbol{\nu}$ and noise $\boldsymbol{\epsilon}$ distributions from test data.

[1] We collect off-resonance effects in m_0 by approximating broadening distributions to be constant across compartments, as in prior works [14, 15]. We acknowledge this could lead to some bias in practice.

[2] Since SPGR (DESS) yields one (two) signal(s) per excitation, four SPGR and three DESS scans produce a total $D \leftarrow 10$ datasets.

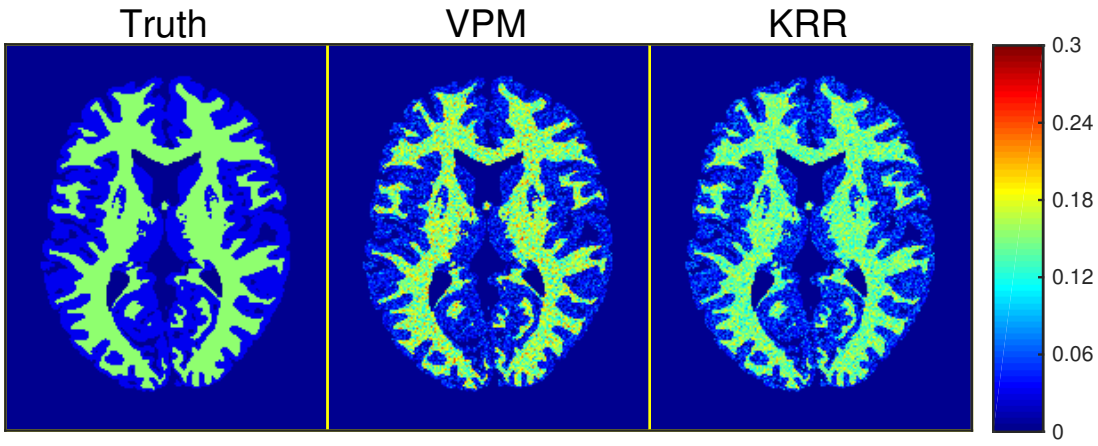

Fig. 1: True f_{F} (*left*) and estimated $\widehat{f}_{\mathrm{F}}$ fast-relaxing compartmental fraction maps, in simulation. Maximum-likelihood estimation via variable projection method and grid search (*center*) is accurate but is computationally expensive. In contrast, kernel ridge regression (*right*) is very fast and achieves comparable precision, at the expense of slightly increased bias (*cf.* Table 1). Voxels outside WM/GM regions are masked out in post-processing for visual clarity.

We take f_{F} to be uniformly distributed on $[-0.1, 0.4]$ and $T_{1,\mathrm{F}}$, $T_{2,\mathrm{F}}$, $T_{1,\mathrm{S}}$, $T_{2,\mathrm{S}}$ to be log-uniformly distributed on $[50, 700]$ms, $[5, 50]$ms, $[700, 2000]$ms, $[50, 300]$ms, respectively. We match the scaling of test data in training by taking m_0 to be uniformly distributed on $\left[2.22 \times 10^{-16}, u\right]$, where u is set as $5\times$ the maximum value of test data. We estimate the distribution of $\boldsymbol{\nu}$ via kernel density estimation [22]. We assume noise covariance $\boldsymbol{\Sigma}$ of form $\sigma^2 \mathbf{I}_{10}$ and compute noise variance estimate $\widehat{\sigma}^2$ from Rayleigh-distributed noise regions of magnitude test data, using estimators described in [23].

We sample $N \leftarrow 10^6$ training inputs from these distributions and use two-compartment SPGR/DESS signal models to evaluate corresponding (noisy, magnitude) responses. We set smoothing length scale $\boldsymbol{\Lambda}$ as a diagonal matrix, with diagonal entries set as squared sample means of test data and known κ. We sample $(\boldsymbol{v}, s)$ $Z \leftarrow 10^3$ times to construct approximate feature mapping $\tilde{\mathbf{z}}_Z$. We apply $\tilde{\mathbf{z}}_Z$ to training data and compute sample means $\mathbf{m}_{\mathbf{x}}$, $\mathbf{m}_{\tilde{\mathbf{z}}}$ and sample covariance matrices $\mathbf{C}_{\mathbf{x}\tilde{\mathbf{z}}}$, $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$. Lastly, we set $\rho \leftarrow \widehat{\sigma}^2$ and evaluate (15) using test datasets and κ on a per-voxel basis. On a 3.5GHz desktop computer with 32GB RAM running MATLAB® R2013a, training and estimating each took less than 40s and 2s.

Fig. 1 compares KRR estimates of fast-relaxing compartmental fraction f_{F} against not only ground truth maps but also maximum-likelihood estimates achieved via the "variable projection" method (VPM) [24] and grid search. As presented, the VPM estimate utilizes a dictionary of nearly 8×10^6 signal vectors computed using finely spaced samples on an unrealistically narrow feasible region consisting of a hypercube with boundaries set as $[-0.1, 0.4]$ in f_{F} and $\pm 20\%$ away from the truth in other dimensions. Using equal computational resources, estimation via VPM took nearly 4h.

Table 1 reports f_{F} sample statistics computed over 7810 WM-like and 9162 GM-like voxels. Overall, KRR and VPM achieve comparable estimation performance. In WM, KRR attains precision slightly higher than and accuracy similar to VPM. In GM, KRR attains precision comparable to and accuracy slightly lower than VPM. KRR attains root mean squared errors lower than VPM in WM (0.0229 versus 0.0295) and comparable to VPM in GM (0.0254 versus 0.0235).

	Truth	VPM	KRR
WM $\widehat{f}_{\mathrm{F}}$	0.15	0.1538 ± 0.0292	0.1440 ± 0.0221
GM $\widehat{f}_{\mathrm{F}}$	0.03	0.0336 ± 0.0232	0.0407 ± 0.0231

Table 1: Sample means $\pm$ sample standard deviations of fast-relaxing compartmental fraction estimates $\widehat{f}_{\mathrm{F}}$, computed over simulated WM- and GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, which is computed via formulas in [25].

4. SUMMARY

We introduced a fast and computationally efficient method for MRI parameter estimation from nonlinear models via KRR. As proof of concept, we applied KRR to quantify in simulation fast-relaxing compartmental fraction f_{F} maps (along with several nuisance parameters) using two-compartment signal models of realistic SPGR and DESS acquisitions. In $\widehat{f}_{\mathrm{F}}$ WM/GM regions, KRR achieved comparable estimation performance as VPM-accelerated grid search, with dramatically reduced computation. Due to its generality, KRR could potentially accelerate MRI parameter estimation in many other applications, particularly those involving multiple latent parameters and/or cumbersome if not altogether unavailable signal models (as in, *e.g.*, MR fingerprinting [5]).

5. REFERENCES

[1] N. Bloembergen, E. M. Purcell, and R. V. Pound, "Relaxation effects in nuclear magnetic resonance absorption," *Phys. Rev.*, vol. 73, no. 7, pp. 679–712, Apr. 1948.

[2] H. C. Torrey, "Bloch equations with diffusion terms," *Phys. Rev.*, vol. 104, pp. 563–5, 1956.

[3] A. Mackay, K. Whittall, J. Adler, D. Li, D. Paty, and D. Graeb, "In vivo visualization of myelin water in brain by magnetic resonance," *Mag. Res. Med.*, vol. 31, no. 6, pp. 673–7, June 1994.

[4] E. Staroswiecki, K. L. Granlund, M. T. Alley, G. E. Gold, and B. A. Hargreaves, "Simultaneous estimation of T2 and apparent diffusion coefficient in human articular cartilage in vivo with a modified three-dimensional double echo steady state (DESS) sequence at 3 T," *Mag. Res. Med.*, vol. 67, no. 4, pp. 1086–96, 2012.

[5] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 495, pp. 187–93, Mar. 2013.

[6] N. Ben-Eliezer, D. K. Sodickson, and K. T. Block, "Rapid and accurate T2 mapping from multi–spin-echo data using Bloch-simulation-based reconstruction," *Mag. Res. Med.*, vol. 73, no. 2, pp. 809–17, Feb. 2015.

[7] B. Zhao, K. Setsompop, H. Ye, S. Cauley, and L. L. Wald, "Maximum likelihood reconstruction for magnetic resonance fingerprinting," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1812–23, Aug. 2016.

[8] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. Computational Learning Theory (COLT)*, 2001, pp. 416–426, LNCS 2111.

[9] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[10] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *NIPS*, 2007.

[11] Z. Wu, "Generalized Bochner's theorem for radial function," *Approximation Theory and its Applications*, vol. 13, no. 3, pp. 47–57, 1997.

[12] M. A. Woodbury, "Inverting modified matrices," 1950, Tech. Report 42, Stat. Res. Group, Princeton Univ.

[13] E. Alonso-Ortiz, I. R. Levesque, and G. B. Pike, "MRI-based myelin water imaging: A technical review," *Mag. Res. Med.*, vol. 73, no. 1, pp. 70–81, Jan. 2015.

[14] S. C. L. Deoni, B. K. Rutt, T. Arun, C. Pierpaoli, and D. K. Jones, "Gleaning multicomponent T1 and T2 information from steady-state imaging data," *Mag. Res. Med.*, vol. 60, no. 6, pp. 1372–87, Dec. 2008.

[15] S. C. L. Deoni, "Correction of main and transmit magnetic field (B0 and B1) inhomogeneity effects in multicomponent-driven equilibrium single-pulse observation of T1 and T2," *Mag. Res. Med.*, vol. 65, no. 4, pp. 1021–35, Apr. 2011.

[16] C. L. Lankford and M. D. Does, "On the inherent precision of mcDESPOT," *Mag. Res. Med.*, vol. 69, no. 1, pp. 127–36, Jan. 2013.

[17] Y. Zur, M. L. Wood, and L. J. Neuringer, "Spoiling of transverse magnetization in steady-state sequences," *Mag. Res. Med.*, vol. 21, no. 2, pp. 251–63, Oct. 1991.

[18] H. Bruder, H. Fischer, R. Graumann, and M. Deimling, "A new steady-state imaging sequence for simultaneous acquisition of two MR images with clearly different contrasts," *Mag. Res. Med.*, vol. 7, no. 1, pp. 35–42, May 1988.

[19] L. I. Sacolick, F. Wiesinger, I. Hancu, and M. W. Vogel, "B1 mapping by Bloch-Siegert shift," *Mag. Res. Med.*, vol. 63, no. 5, pp. 1315–22, May 2010.

[20] G. Nataraj, J-F. Nielsen, and J. A. Fessler, "Optimizing MR scan design for model-based T1, T2 estimation from steady-state sequences," *IEEE Trans. Med. Imag.*, 2017, To appear.

[21] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–8, June 1998.

[22] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–76, Sept. 1962.

[23] M. M. Siddiqui, "Statistical inference for Rayleigh distributions," *RADIO SCIENCE Journal of Research NBS/USNC-URSI*, vol. 68D, no. 9, pp. 1005–10, Sept. 1964.

[24] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Prob.*, vol. 19, no. 2, pp. R1–26, Apr. 2003.

[25] S. Ahn and J. A. Fessler, "Standard errors of mean, variance, and standard deviation estimators," Tech. Rep. 413, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, July 2003.