Shallow Learning with Kernels for Dictionary-Free Magnetic Resonance Fingerprinting

Gopal Nataraj[1], Mingjie Gao[1], Jakob Assländer[2], Clayton Scott[1], and Jeffrey A. Fessler[1]
[1]*Electrical Engineering and Computer Science, University of Michigan (UM)*
[2]*Center for Biomedical Imaging, New York University (NYU) School of Medicine*

Target Audience
Researchers interested in dictionary-free magnetic resonance fingerprinting (MRF), especially in applications involving many unknown parameters per voxel.

Introduction and Purpose
As originally formulated [1], MRF requires solving a nonlinear least-squares problem at each voxel to estimate parameter maps from reconstructed images. The standard method to solve such problems involves simulating a dictionary of signals over a discretized parameter grid and exhaustively searching for the unknowns that yield the best fit with measurements. Such grid search scales poorly with the number of unknowns. As a more scalable alternative, researchers have recently proposed to use deep neural networks to learn estimators from simulated training data [2]. However, it is well known in the machine learning (ML) community that deep learning requires enormous numbers of training points to train many hyperparameters without overfitting, and its limited theoretical basis renders its practical use largely an art. Here, we develop a considerably simpler and theoretically well-founded "shallow" ML method for dictionary-free MRF.

Methods
After low-rank reconstruction, let $\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{\epsilon}$ model a low-rank MRF measurement sequence at a single voxel, where low-rank MRF signal model $\boldsymbol{s}$ relates unknowns $\boldsymbol{x}$ (e.g., (M_0, T_1, T_2)) to $\boldsymbol{y}$, barring noise $\boldsymbol{\epsilon}$. Following [3-4], we approach estimation by learning a nonlinear regression function from simulated (labeled) training data and evaluating this function on real (unlabeled) test data. We sample distributions on $\boldsymbol{x}, \boldsymbol{\epsilon}$ and use $\boldsymbol{s}$ to form N training points $\{(\boldsymbol{x}_1, \boldsymbol{y}_1), \dots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\}$. To design a nonlinear estimator for the lth component x_l of $\boldsymbol{x}$, we seek a function $\hat{g}_l$ and an offset $\hat{b}_l$ that nonlinearly map each simulated measurement $\boldsymbol{y}_n$ to an estimate $\hat{g}_l(\boldsymbol{y}_n) + \hat{b}_l$ that is "close" to the corresponding simulated true parameter $x_{l,n}$, for $n \in \{1, \dots, N\}$. Mathematically, we seek a solution to the following regularized function optimization problem:

$$(\hat{g}_l, \hat{b}_l) \in \underset{g_l \in \mathcal{G},\, b_l \in \mathbb{R}}{\arg\max} \frac{1}{N} \sum_{n=1}^{N} \left(g_l(\boldsymbol{y}_n) + b_l - x_{l,n} \right)^2 + \rho \left\| g_l \right\|_{\mathcal{G}}^2$$

where $\mathcal{G}$ is a reproducing kernel Hilbert space (RKHS) with function norm $\|.\|_{\mathcal{G}}$ and ρ is a regularization parameter [3]. This problem admits a unique solution with $\hat{g}_l$ taking the form $\hat{g}_l(\boldsymbol{y}) = \sum_{n=1}^{N} \hat{a}_{l,n}\, k(\boldsymbol{y}, \boldsymbol{y}_n)$ [5], where k is the kernel function associated with RKHS $\mathcal{G}$. Inserting this representation above yields an equivalent convex optimization problem over weights $a_{l,1}, \dots, a_{l,N}$ and offset b_l that we solve via standard methods. Choosing Gaussian kernel $k(\boldsymbol{y}, \boldsymbol{y}') \leftarrow \exp(-\|\boldsymbol{y} - \boldsymbol{y}'\|_{\boldsymbol{\Lambda}^{-2}}^2)$ with bandwidth matrix $\boldsymbol{\Lambda}$ induces optimization over a RKHS dense in the space of square-integrable continuous functions [6], so with enough samples $\hat{g}_l$ can approximate any continuous function over any bounded subset of the measurement space to arbitrary accuracy.

Results and Discussion
We first illustrate shallow learning to estimate $\boldsymbol{x} \leftarrow (M_0, T_1, T_2)$ in a simulation that emulated the *in vivo* experiment described next. We sampled $N \leftarrow 10^4$ noisy training points assuming $M_0 \sim \text{unif}(0.9, 1.1)$, $T_1 \sim \text{logunif}(0.3, 4)$s, $T_2 \sim \text{logunif}(0.01, 0.25)$s, and $\boldsymbol{\epsilon} \sim \mathbb{CN}(\boldsymbol{0}, (0.01)^2 \boldsymbol{I})$. We trained the estimator using a Gaussian kernel whose bandwidth $\boldsymbol{\Lambda}$ is set save for a scale factor λ based on the scale of normalized test data ($\lambda \leftarrow 30$ and $\rho \leftarrow 2 \times 10^{-14}$ were separately chosen using a holdout criterion). We evaluated the estimator on a separate set of unlabeled test points. Figure 1 shows that within the training range, shallow learning achieves good estimation accuracy.

We next demonstrate shallow learning for MRF on *in vivo* data. An asymptomatic volunteer's brain was imaged on a 3T Skyra scanner (Siemens, Erlangen, Germany) according to a protocol approved by NYU's review board. Using a 16-channel head coil, 840 radial k-space spokes of data were acquired in about 3.8s using an MRF sequence optimized for T_1, T_2 estimation [7]. Fig. 2 displays low-rank images that were reconstructed from data by the method of [8] onto a 256x256x3 mm^3 field of view with 1x1x3mm^3 resolution. We simulated $N \leftarrow 10^4$ training points using a signal model that relates M_0, T_1, T_2 to a corresponding low-rank MRF measurement sequence. Using identical hyperparameters and priors as in simulations, we learned T_1, T_2 estimators and evaluated them voxelwise on low-rank MRF images to produce T_1, T_2 maps. On a 2.7GHz laptop with 8GB RAM running MATLAB® R2017a, training and testing took less than 30s total, far faster than deep learning methods [2].

Fig. 3 compares T_1, T_2 estimates computed via shallow learning versus via grid search (that took 117s using 174,225 dictionary elements). Table 1 compares sample statistics computed over white matter (WM) and gray matter (GM) regions of interest (ROIs) marked in Fig. 2. T_1, T_2 estimates in WM and GM ROIs are in excellent agreement.

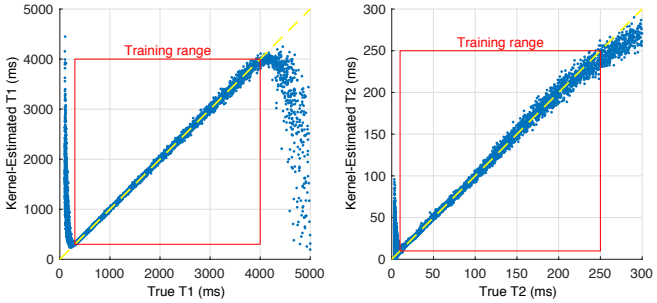

Figure 1: Shallow kernel learning T_1 and T_2 estimates, in simulation. Test points fix true T_2 values to 100ms as T_1 is varied (*left*) and true T_1 values to 1000ms as T_2 is varied (*right*). Within training ranges (indicated via red boxes), the learned estimator demonstrates good estimation accuracy.

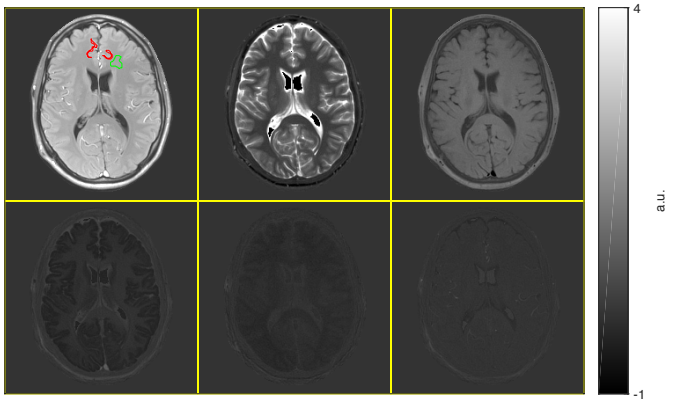

Figure 2: Rank-6 images from a low-rank reconstruction [8] of precision-optimized MRF data [7]. We learn separate T_1 and T_2 estimators that map on a per-voxel basis these six images into the estimates displayed in Fig. 3.

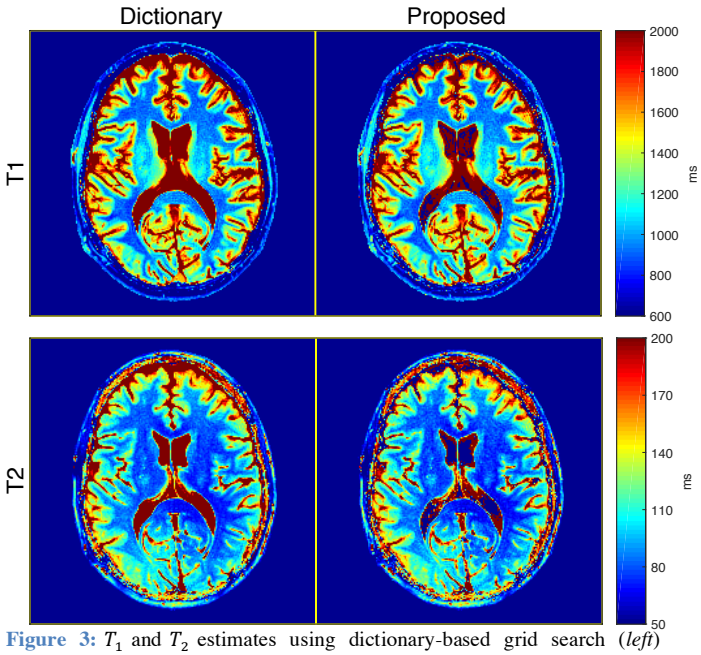

Figure 3: T_1 and T_2 estimates using dictionary-based grid search (*left*) versus dictionary-free shallow kernel learning (*right*). Within WM and GM ROIs, parameter estimates are in excellent agreement.

	Dictionary	Proposed
WM T_1	920 ± 35	930 ± 38
GM T_1	1500 ± 120	1500 ± 120
WM T_2	82 ± 4.0	82 ± 4.4
GM T_2	$120 \pm 10.$	120 ± 11

Table 1: T_1, T_2 sample means $\pm$ sample standard deviations, computed within pooled WM and GM ROIs (marked in Fig. 2) that respectively contain 142 and 111 voxels.

Conclusion
We have introduced a fast, dictionary-free shallow ML method for parameter estimation from MRF data. We demonstrated through simulations and *in vivo* experiments that the method can achieve T_1, T_2 estimates comparable in WM/GM ROIs to estimates via exhaustive grid search. The method will scale far better than grid search in applications requiring estimation of more unknowns per voxel, e.g. [4].

Acknowledgments
We thank UM for partial support via an "M-Cubed" seed grant and a predoctoral fellowship.

References
[1] Ma et al., *Nature*, 495(7440):187-92, 2013. [2] Cohen et al., *Proc. ISMRM*, 0688, 2017. [3] Nataraj et al. *Proc. IEEE ISBI*, 5-9, 2017. [4] Nataraj et al., *Proc. ISMRM*, 5076, 2017. [5] Scholkopf et al., *Proc. COLT*, 416-26, 2001. [6] Steinwert et al., *Support Vector Machines*, Springer, 1st ed., 2008. [7] Assländer et al., *arXiv* 1703.00481v1, 2017. [8] Assländer et al., *MRM*, 10.1002/mrm.26639, 2017.