

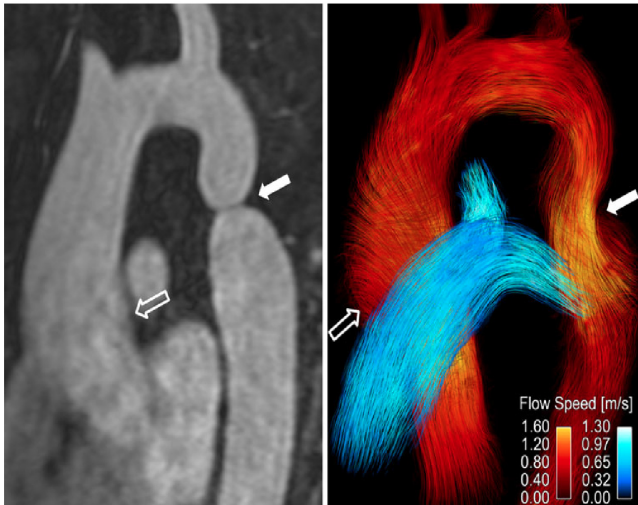
Advances in Quantitative MRI: Acquisition, Estimation, and Application

Gopal Nataraj

Dissertation Defense
March 23, 2018

Dept. of Electrical Engineering and Computer Science
University of Michigan

Example: flow imaging

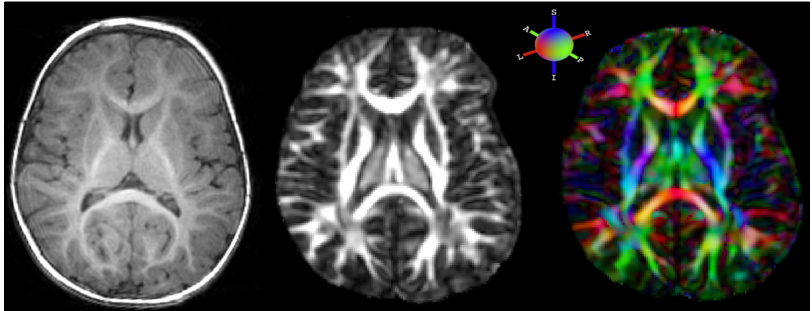


qualitative

quantitative¹

¹figure borrowed from [Hope et al., 2013]

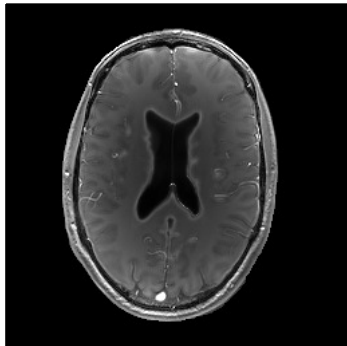
Example: diffusion imaging



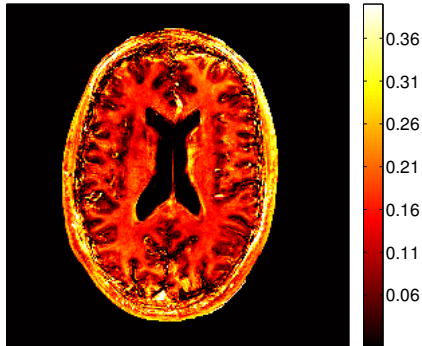
qualitative fractional anisotropy (FA) directional FA²

²figure borrowed from www.diffusion-imaging.com

Example: myelin water imaging



qualitative



fast-relaxing fraction³

³figure adapted from [Nataraj et al., 2017a]

Quantitative MRI (QMRI)

Goal: rapidly and accurately localize biomarkers from MR data

Quantitative MRI (QMRI)

Goal: rapidly and accurately localize biomarkers from MR data

- **biomarker** measurable tissue property (e.g., flow rate)
that indicates a biological process (e.g., blockage)
characteristic to specific disorders (e.g., stroke)

Quantitative MRI (QMRI)

Goal: rapidly and accurately localize biomarkers from MR data

- biomarker measurable tissue property (e.g., flow rate)
that indicates a biological process (e.g., blockage)
characteristic to specific disorders (e.g., stroke)
- localize produce quantitative MR images

Quantitative MRI (QMRI)

Goal: rapidly and accurately localize biomarkers from MR data

- biomarker measurable tissue property (e.g., flow rate)
that indicates a biological process (e.g., blockage)
characteristic to specific disorders (e.g., stroke)
- localize produce quantitative MR images
- accurately physically realistic signal models
- rapidly fast acquisition, fast estimation

Quantitative MRI (QMRI)

Goal: rapidly and accurately localize biomarkers from MR data

- biomarker measurable tissue property (e.g., flow rate) that indicates a biological process (e.g., blockage) characteristic to specific disorders (e.g., stroke)
- localize produce quantitative MR images
- accurately physically realistic signal models
- rapidly fast acquisition, fast estimation

Challenge: rapidly vs. accurately often competing goals

- more accurate models typically depend on more markers
- precisely estimating more markers usually requires longer scans and more computation

Advances in Quantitative MRI:

- **Acquisition**

[Ch. 4]

How can we assemble fast, informative collections of scans to enable precise biomarker quantification?

Advances in Quantitative MRI:

- **Acquisition** [Ch. 4]
How can we assemble fast, informative collections of scans to enable precise biomarker quantification?
- **Estimation** [Ch. 5]
Given accurate models and informative data, how can we rapidly quantify these biomarkers?

Advances in Quantitative MRI:

- **Acquisition** [Ch. 4]
How can we assemble fast, informative collections of scans to enable precise biomarker quantification?
- **Estimation** [Ch. 5]
Given accurate models and informative data, how can we rapidly quantify these biomarkers?
- **Application** [Ch. 6]
Using these tools, can we design a state-of-the-art biomarker?

Advances in Quantitative MRI:

- **Acquisition** [Ch. 4]
How can we assemble fast, informative collections of scans to enable precise biomarker quantification?
- **Estimation** [Ch. 5]
Given accurate models and informative data, how can we rapidly quantify these biomarkers?
- **Application** [Ch. 6]
Using these tools, can we design a state-of-the-art biomarker?

After reconstruction, single voxel y_d in d th image modeled as

$$y_d = s_d(\mathbf{x}; \boldsymbol{\nu}, \mathbf{p}_d) + \epsilon_d \quad (1)$$

- $\mathbf{x} \in \mathbb{R}^L$ unknown parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ “known” parameters
- $\mathbf{p}_d \in \mathbb{R}^A$ acquisition parameters
- $s_d : \mathbb{R}^{L+K+A} \mapsto \mathbb{C}$ d th signal model
- $\epsilon_d \in \mathbb{C}$ noise $\sim \mathbb{CN}(0, \sigma_d^2)$

Signal Model

A *scan profile* is a set of D scans that produces at each voxel a measurement vector $\mathbf{y} := [y_1, \dots, y_D]^T$ modeled as

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

- $\mathbf{x} \in \mathbb{R}^L$ unknown parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ “known” parameters
- $\mathbf{P} := [\mathbf{p}_1, \dots, \mathbf{p}_D]$ acquisition parameter matrix
- $\mathbf{s} : \mathbb{R}^{L+K+AD} \mapsto \mathbb{C}^D$ vector signal model
- $\boldsymbol{\epsilon} \sim \mathbb{CN}(\mathbf{0}_D, \boldsymbol{\Sigma})$ noise, with $\boldsymbol{\Sigma} := \text{diag}(\sigma_1^2, \dots, \sigma_D^2)$

Signal Model

A *scan profile* is a set of D scans that produces at each voxel a measurement vector $\mathbf{y} := [y_1, \dots, y_D]^T$ modeled as

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

- $\mathbf{x} \in \mathbb{R}^L$ unknown parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ “known” parameters
- $\mathbf{P} := [\mathbf{p}_1, \dots, \mathbf{p}_D]$ acquisition parameter matrix
- $\mathbf{s} : \mathbb{R}^{L+K+AD} \mapsto \mathbb{C}^D$ vector signal model
- $\boldsymbol{\epsilon} \sim \mathbb{CN}(\mathbf{0}_D, \boldsymbol{\Sigma})$ noise, with $\boldsymbol{\Sigma} := \text{diag}(\sigma_1^2, \dots, \sigma_D^2)$

Task: design \mathbf{P} to enable precise unbiased estimation of \mathbf{x}

Towards an Objective Function

When \mathbf{s} is analytic in \mathbf{x} (as is typical),

Fisher information characterizes unbiased estimator precision:

$$\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) := (\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}))^H \boldsymbol{\Sigma}^{-1} \nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \quad (2)$$

Towards an Objective Function

When \mathbf{s} is analytic in \mathbf{x} (as is typical),

Fisher information characterizes unbiased estimator precision:

$$\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) := (\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}))^H \boldsymbol{\Sigma}^{-1} \nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \quad (2)$$

When \mathbf{F} is invertible, Cramér-Rao Bound (CRB) [Cramér, 1946] ensures covariance of unbiased estimates $\hat{\mathbf{x}}$ of \mathbf{x} satisfy

$$\text{cov } \hat{\mathbf{x}}; \boldsymbol{\nu}, \mathbf{P} \succeq \mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \quad (3)$$

Maximum-likelihood (ML) estimates achieve CRB asymptotically or (equivalently for Gaussian data) at sufficiently high SNR.

Towards an Objective Function

When \mathbf{s} is analytic in \mathbf{x} (as is typical),

Fisher information characterizes unbiased estimator precision:

$$\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) := (\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}))^H \boldsymbol{\Sigma}^{-1} \nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \quad (2)$$

When \mathbf{F} is invertible, Cramér-Rao Bound (CRB) [Cramér, 1946] ensures covariance of unbiased estimates $\hat{\mathbf{x}}$ of \mathbf{x} satisfy

$$\text{cov } \hat{\mathbf{x}}; \boldsymbol{\nu}, \mathbf{P} \succeq \mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \quad (3)$$

Maximum-likelihood (ML) estimates achieve CRB asymptotically or (equivalently for Gaussian data) at sufficiently high SNR.

Idea: choose \mathbf{P} such that imprecision matrix \mathbf{F}^{-1} “small”

Idea: choose \mathbf{P} to minimize the objective

$$\Psi(\mathbf{x}; \nu, \mathbf{P}) = \text{tr}(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \nu, \mathbf{P})\mathbf{W}^T), \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a pre-selected diagonal matrix of weights.

Idea: choose \mathbf{P} to minimize the objective

$$\Psi(\mathbf{x}; \nu, \mathbf{P}) = \text{tr}\left(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \nu, \mathbf{P})\mathbf{W}^T\right), \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a pre-selected diagonal matrix of weights.

Challenge: \mathbf{x}, ν vary spatially

Idea: choose \mathbf{P} to minimize the objective

$$\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) = \text{tr}(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})\mathbf{W}^T), \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a pre-selected diagonal matrix of weights.

Challenge: $\mathbf{x}, \boldsymbol{\nu}$ vary spatially

Two problems considered:

- min-max scan design [Nataraj et al., 2017b]

$$\check{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} \max_{\substack{\mathbf{x} \in \mathbb{X}^t \\ \boldsymbol{\nu} \in \mathbb{N}^t}} \Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) \right\} \quad (5)$$

where $\mathbb{X}^t \subseteq \mathbb{R}^L$ and $\mathbb{N}^t \subseteq \mathbb{R}^K$ are “tight” ranges of interest and \mathbb{P} is defined by acquisition/timing constraints

Idea: choose \mathbf{P} to minimize the objective

$$\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) = \text{tr}(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})\mathbf{W}^T), \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a pre-selected diagonal matrix of weights.

Challenge: $\mathbf{x}, \boldsymbol{\nu}$ vary spatially

Two problems considered:

- min-max scan design [Nataraj et al., 2017b]

$$\check{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} \max_{\substack{\mathbf{x} \in \mathbb{X}^t \\ \boldsymbol{\nu} \in \mathbb{N}^t}} \Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) \right\} \quad (5)$$

- Bayesian scan design

$$\check{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} E_{\mathbf{x}, \boldsymbol{\nu}}(\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})) \right\} \quad (6)$$

Idea: choose \mathbf{P} to minimize the objective

$$\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) = \text{tr}(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})\mathbf{W}^T), \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a pre-selected diagonal matrix of weights.

Challenge: $\mathbf{x}, \boldsymbol{\nu}$ vary spatially

Two problems considered:

- min-max scan design [Nataraj et al., 2017b]

$$\check{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} \max_{\substack{\mathbf{x} \in \mathbb{X}^t \\ \boldsymbol{\nu} \in \mathbb{N}^t}} \Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) \right\} \quad (5)$$

- Bayesian scan design

$$\check{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} E_{\mathbf{x}, \boldsymbol{\nu}}(\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})) \right\} \quad (6)$$

Detailed Example Study

Task: design fast acquisition for precise estimation of relaxation parameters T_1 , T_2 in white/gray matter (WM/GM) of brain

Detailed Example Study

Task: design fast acquisition for precise estimation of relaxation parameters T_1 , T_2 in white/gray matter (WM/GM) of brain

- Consider scan profiles consisting of two fast pulse sequences
 - Spoiled Gradient-Recalled Echo (SPGR) [Zur et al., 1991]
 - Dual-Echo Steady-State (DESS) [Redpath and Jones, 1988]

Detailed Example Study

Task: design fast acquisition for precise estimation of relaxation parameters T_1, T_2 in white/gray matter (WM/GM) of brain

- Consider scan profiles consisting of two fast pulse sequences
 - Spoiled Gradient-Recalled Echo (SPGR) [Zur et al., 1991]
 - Dual-Echo Steady-State (DESS) [Redpath and Jones, 1988]
- For each scan profile feasible under total time constraint:
 1. Let \mathbf{s} model corresponding single-component signal
 - $\mathbf{x} \leftarrow [m_0, T_1, T_2]^T$, where m_0 is a scale factor
 - $\nu \leftarrow$ flip angle variation
 - $\mathbf{P} \leftarrow$ nominal flip angles, repetition times
 2. Optimize \mathbf{P} subject to flip angle, sequence timing constraints
 - $\mathbf{W} \leftarrow \text{diag}(0, 0.1, 1)$ emphasizes T_1, T_2 est roughly equally
 - \mathbb{X}^t chosen to focus on WM/GM at 3T field strength
 - \mathbb{N}^t chosen to allow 10% flip angle variation

Scan Profile Comparison

(#SPGR, #DESS) Profiles	(2, 1)	(1, 1)	(0, 2)
SPGR nom. flip (deg)	(15, 5)	15	–
DESS nom. flip (deg)	30	10	(35, 10)
SPGR rep. times (ms)	(12.2, 12.2)	13.9	–
DESS rep. times (ms)	17.5	28.0	(24.4, 17.5)
optimal max cost	4.0	4.9	3.5

Scan Profile Comparison

(#SPGR, #DESS) Profiles	(2, 1)	(1, 1)	(0, 2)
SPGR nom. flip (deg)	(15, 5)	15	–
DESS nom. flip (deg)	30	10	(35, 10)
SPGR rep. times (ms)	(12.2, 12.2)	13.9	–
DESS rep. times (ms)	17.5	28.0	(24.4, 17.5)
optimal max cost	4.0	4.9	3.5

Main finding: 2 DESS sequences can yield T_1 , T_2 WM/GM estimates that are at least as precise as T_1 , T_2 estimates from SPGR/DESS scan profiles, under this competitive time constraint.

Experimental Setup

Candidate $(2, 1)$, $(1, 1)$, $(0, 2)$ SPGR/DESS scan profiles

- Prescribed optimized nominal flip angles, repetition times
- Used $256 \times 256 \times 8$ 3D matrix over $24 \times 24 \times 4$ cm FOV
- Required **1m37s** scan time for each profile

Experimental Setup

Candidate (2, 1), (1, 1), (0, 2) SPGR/DESS scan profiles

- Prescribed optimized nominal flip angles, repetition times
- Used $256 \times 256 \times 8$ 3D matrix over $24 \times 24 \times 4$ cm FOV
- Required **1m37s** scan time for each profile

Reference scan profile

- Four inversion recovery (IR) scans for T_1 estimation
- Four spin-echo (SE) scans for T_2 estimation
- 256×256 matrix over $24 \times 24 \times 0.5$ cm FOV
- Required **40m58s** scan time total

Experimental Setup

Candidate (2, 1), (1, 1), (0, 2) SPGR/DESS scan profiles

- Prescribed optimized nominal flip angles, repetition times
- Used $256 \times 256 \times 8$ 3D matrix over $24 \times 24 \times 4$ cm FOV
- Required **1m37s** scan time for each profile

Reference scan profile

- Four inversion recovery (IR) scans for T_1 estimation
- Four spin-echo (SE) scans for T_2 estimation
- 256×256 matrix over $24 \times 24 \times 0.5$ cm FOV
- Required **40m58s** scan time total

Bloch-Siegert (BS) acquisition for separate flip angle calibration

- Acquired 2 BS-shifted 3D SPGR scans in 1m40s total
- Used for T_1 , T_2 est from both candidate and reference profiles

Phantom Accuracy Results

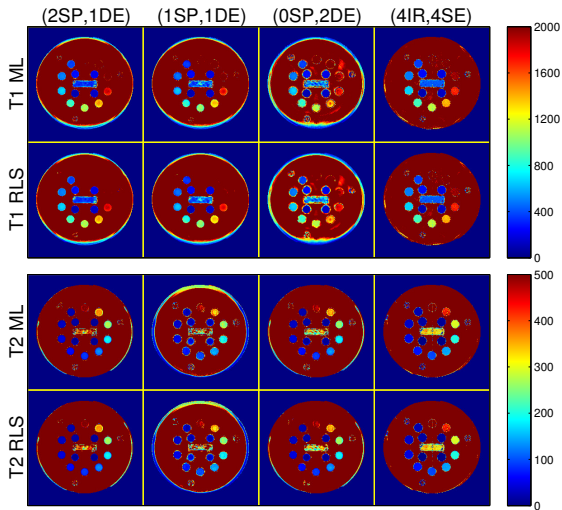
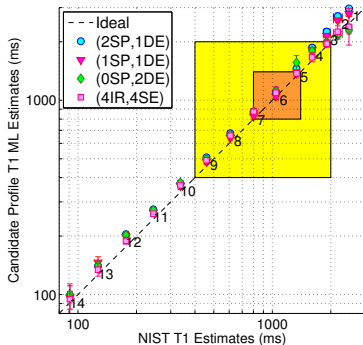
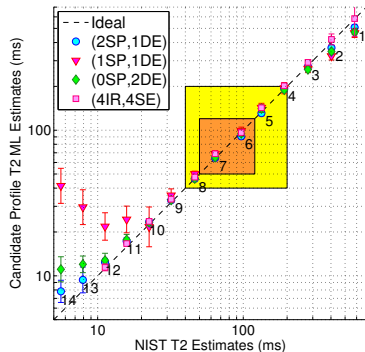


Figure 1: Colorbar ranges in ms.

Phantom Accuracy Results



(a) \hat{T}_1^{ML} Estimates



(b) \hat{T}_2^{ML} Estimates

Compared against NIST NMR measurements [Keenan et al., 2016]

Phantom Precision Results

- Repeated each profile 10 times
- Estimated T_1 , T_2 std dev of typical voxel across repetitions

Phantom Precision Results

	(2, 1)	(1, 1)	(0, 2)
V5 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	50 ± 12	$40 \pm 10.$	39 ± 9.4
V6 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	70 ± 18	60 ± 15	60 ± 16
V7 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	60 ± 13	50 ± 13	50 ± 13
V5 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	2.6 ± 0.63	6 ± 1.4	3.5 ± 0.84
V6 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	1.9 ± 0.46	5 ± 1.1	2.3 ± 0.54
V7 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	1.4 ± 0.34	3.4 ± 0.80	1.5 ± 0.35
$\sqrt{\text{opt max cost}}$ estimate	8.9 ± 1.8	11 ± 2.6	8.3 ± 2.1

Table 1: Pooled sample standard deviations \pm pooled standard errors of sample standard deviations (ms), from optimized SPGR/DESS profiles.

Phantom Precision Results

	(2, 1)	(1, 1)	(0, 2)
V5 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	50 ± 12	$40 \pm 10.$	39 ± 9.4
V6 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	70 ± 18	60 ± 15	60 ± 16
V7 $\hat{\sigma}_{\hat{T}_1^{\text{ML}}}$	60 ± 13	50 ± 13	50 ± 13
V5 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	2.6 ± 0.63	6 ± 1.4	3.5 ± 0.84
V6 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	1.9 ± 0.46	5 ± 1.1	2.3 ± 0.54
V7 $\hat{\sigma}_{\hat{T}_2^{\text{ML}}}$	1.4 ± 0.34	3.4 ± 0.80	1.5 ± 0.35
$\sqrt{\text{opt max cost}}$ estimate	8.9 ± 1.8	11 ± 2.6	8.3 ± 2.1

Table 1: Pooled sample standard deviations \pm pooled standard errors of sample standard deviations (ms), from optimized SPGR/DESS profiles.

Similar trends across profiles of empirical vs. theoretical std dev!

Contributions

- MR scan design method for precise parameter estimation
- Fast SPGR/DESS scan profile for T_1 , T_2 estimation in brain

Contributions

- MR scan design method for precise parameter estimation
- Fast SPGR/DESS scan profile for T_1 , T_2 estimation in brain
 - Phantom (and omitted simulation) results validate method as a predictor of unbiased estimation precision.

Summary

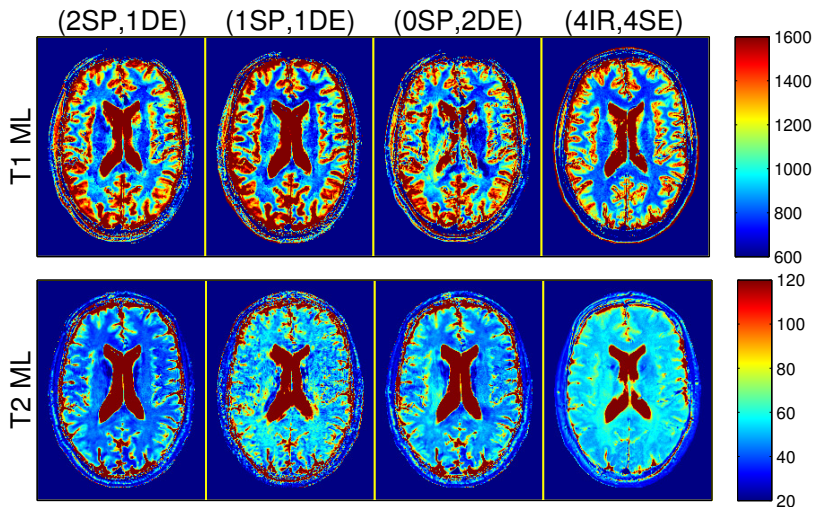


Figure 2: Colorbar ranges in ms.

Contributions

- MR scan design method for precise parameter estimation
- Fast SPGR/DESS scan profile for T_1 , T_2 estimation in brain
 - Phantom (and omitted simulation) results validate method as a predictor of unbiased estimation precision.
 - *In vivo* results reveal discrepancies (especially in T_2 estimates), suggesting T_1 , T_2 estimates sensitive to model mismatch.

Contributions

- MR scan design method for precise parameter estimation
- Fast SPGR/DESS scan profile for T_1 , T_2 estimation in brain
 - Phantom (and omitted simulation) results validate method as a predictor of unbiased estimation precision.
 - *In vivo* results reveal discrepancies (especially in T_2 estimates), suggesting T_1 , T_2 estimates sensitive to model mismatch.

How to address *in vivo* model mismatch?

- More accurate *in vivo* signal models
- More scalable parameter estimation

Advances in Quantitative MRI:

- **Acquisition** [Ch. 4]
How can we assemble fast, informative collections of scans to enable precise biomarker quantification?
- **Estimation** [Ch. 5]
Given accurate models and informative data, how can we rapidly quantify these biomarkers?
- **Application** [Ch. 6]
Using these tools, can we design a state-of-the-art biomarker?

Given: at every voxel, measurement vector $\mathbf{y} \in \mathbb{C}^D$ modeled as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}, \boldsymbol{\nu}) + \boldsymbol{\epsilon} \quad (7)$$

- $\mathbf{x} \in \mathbb{R}^L$ unknown parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ “known” parameters
- $\mathbf{s} : \mathbb{R}^{L+K} \mapsto \mathbb{C}^D$ vector signal model
- $\boldsymbol{\epsilon} \in \mathbb{C}^D$ noise $\sim \mathbb{CN}(\mathbf{0}_D, \boldsymbol{\Sigma})$

Given: at every voxel, measurement vector $\mathbf{y} \in \mathbb{C}^D$ modeled as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}, \boldsymbol{\nu}) + \boldsymbol{\epsilon} \quad (7)$$

- $\mathbf{x} \in \mathbb{R}^L$ unknown parameters
- $\boldsymbol{\nu} \in \mathbb{R}^K$ “known” parameters
- $\mathbf{s} : \mathbb{R}^{L+K} \mapsto \mathbb{C}^D$ vector signal model
- $\boldsymbol{\epsilon} \in \mathbb{C}^D$ noise $\sim \mathbb{CN}(\mathbf{0}_D, \boldsymbol{\Sigma})$

Task: design fast voxel-by-voxel estimator $\hat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Task: design fast voxel-by-voxel estimator $\hat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\nu})$

Challenges:

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

Task: design fast voxel-by-voxel estimator $\hat{\mathbf{x}}(\mathbf{y}, \nu)$

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

Task: design fast voxel-by-voxel estimator $\hat{\mathbf{x}}(\mathbf{y}, \nu)$

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

Task: design fast voxel-by-voxel estimator $\hat{\mathbf{x}}(\mathbf{y}, \nu)$

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
- dictionary-based grid search

Grid search computational costs

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

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	$\sim 100^5 - 100^9$

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 $(\mathbf{x}_1, \nu_1, \epsilon_1), \dots, (\mathbf{x}_N, \nu_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model \mathbf{s}

Machine Learning for QMRI Parameter Estimation

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

- sample $(\mathbf{x}_1, \boldsymbol{\nu}_1, \epsilon_1), \dots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model \mathbf{s}
- design *nonlinear* functions $\hat{x}_l(\cdot) := \hat{h}_l(\cdot) + \hat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\text{Re}(\mathbf{y}_n)^T, \text{Im}(\mathbf{y}_n)^T, \boldsymbol{\nu}_n^T]^T \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

Machine Learning for QMRI Parameter Estimation

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

- sample $(\mathbf{x}_1, \nu_1, \epsilon_1), \dots, (\mathbf{x}_N, \nu_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model \mathbf{s}
- design *nonlinear* functions $\hat{x}_l(\cdot) := \hat{h}_l(\cdot) + \hat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\text{Re}(\mathbf{y}_n)^\top, \text{Im}(\mathbf{y}_n)^\top, \nu_n^\top]^\top \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

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

Machine Learning for QMRI Parameter Estimation

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

- sample $(\mathbf{x}_1, \nu_1, \epsilon_1), \dots, (\mathbf{x}_N, \nu_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model \mathbf{s}
- design *nonlinear* functions $\hat{x}_l(\cdot) := \hat{h}_l(\cdot) + \hat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\text{Re}(\mathbf{y}_n)^\top, \text{Im}(\mathbf{y}_n)^\top, \nu_n^\top]^\top \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

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

Machine Learning for QMRI Parameter Estimation

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

- sample $(\mathbf{x}_1, \nu_1, \epsilon_1), \dots, (\mathbf{x}_N, \nu_N, \epsilon_N)$ from prior distributions
- simulate image data vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ via signal model \mathbf{s}
- design *nonlinear* functions $\hat{x}_l(\cdot) := \hat{h}_l(\cdot) + \hat{b}_l$ for $l \in \{1, \dots, L\}$ that map each $\mathbf{q}_n := [\text{Re}(\mathbf{y}_n)^T, \text{Im}(\mathbf{y}_n)^T, \nu_n^T]^T \in \mathcal{Q}$ to $x_{l,n} \in \mathbb{R}$

$$(\hat{h}_l, \hat{b}_l) \in \left\{ \arg \min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^N (h_l(\mathbf{q}_n) + b_l - x_{l,n})^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (8)$$

Solution: Param Estimation via Regression with **Kernels** (PERK)

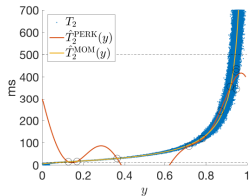
[Nataraj et al., 2018]

- restrict optimization to a **certain rich function space** \mathbb{H}
- optimal $\hat{h}_l \in \mathbb{H}$ takes form $\hat{h}_l = \sum_{n=1}^N \hat{a}_{l,n} k(\cdot, \mathbf{q}_n)$

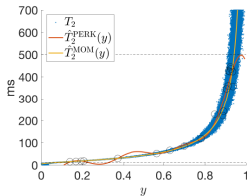
[Schölkopf et al., 2001]

PERK in a 1-D Toy Problem

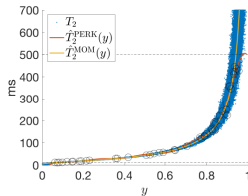
Task: estimate T_2 , given samples from $y = \exp(-T_E/T_2) + \epsilon$



(a) $N \leftarrow 10$



(b) $N \leftarrow 20$



(c) $N \leftarrow 50$

Compare: \hat{T}_2^{PERK} with method-of-moments (MOM) estimator

$$\hat{T}_2^{\text{MOM}}(\cdot) := -T_E / \log |\cdot|$$

(PERK more useful when good MOM estimator unavailable)

PERK Solution

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

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

PERK Solution

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

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

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^T$ 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
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$ de-meaning operator

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^T$ 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
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$ de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^T$ nonlin kernel embedding

PERK Solution

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^T$ 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
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$ de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^T$ nonlin kernel embedding

Can we scale computation with L more gracefully?

- Perhaps, since (9) separable in $l \in \{1, \dots, L\}$ by construction

PERK Solution

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

$$\hat{x}_l(\cdot) = \mathbf{x}_l^T \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}(\mathbf{K}\mathbf{M} + N\rho_l \mathbf{I}_N)^{-1} \left(\mathbf{k}(\cdot) - \frac{1}{N} \mathbf{K} \mathbf{1}_N \right) \right) \quad (9)$$

- $\mathbf{x}_l := [x_{l,1}, \dots, x_{l,N}]^T$ 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
- $\mathbf{M} := \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$ de-meaning operator
- $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^T$ nonlin kernel embedding

Can we scale computation with L more gracefully?

- Perhaps, since (9) separable in $l \in \{1, \dots, L\}$ by construction
- However, explicitly computing \mathbf{K} may be undesirable...

PERK as High-Dimensional Affine Regression

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

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}}. \quad (10)$$

PERK as High-Dimensional Affine Regression

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

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}}. \quad (10)$$

Plugging (10) into PERK solution (9) and rearranging gives

$$\hat{x}_I(\cdot) \approx \frac{1}{N} \mathbf{x}_I^\top \mathbf{1}_N + \frac{1}{N} \mathbf{x}_I^\top \mathbf{M} \tilde{\mathbf{Z}}^\top \left(\frac{1}{N} \tilde{\mathbf{Z}} \mathbf{M} \tilde{\mathbf{Z}}^\top + \rho_I \mathbf{I}_Z \right)^{-1} \left(\tilde{\mathbf{z}}(\cdot) - \frac{1}{N} \tilde{\mathbf{Z}} \mathbf{1}_N \right)$$

PERK as High-Dimensional Affine Regression

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

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}}. \quad (10)$$

Plugging (10) into PERK solution (9) and rearranging gives

$$\hat{x}_l(\cdot) \approx \hat{m}_{x_l} + \hat{\mathbf{c}}_{x_l \tilde{\mathbf{z}}}^T \left(\hat{\mathbf{C}}_{\tilde{\mathbf{z}} \tilde{\mathbf{z}}} + \rho_l \mathbf{I}_Z \right)^{-1} (\tilde{\mathbf{z}}(\cdot) - \hat{\mathbf{m}}_{\tilde{\mathbf{z}}}) \quad (11)$$

which is regularized Z -dimensional affine regression!

PERK as High-Dimensional Affine Regression

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

$$\mathbf{K} \approx \tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}}. \quad (10)$$

Plugging (10) into PERK solution (9) and rearranging gives

$$\hat{x}_I(\cdot) \approx \hat{m}_{x_I} + \hat{\mathbf{c}}_{x_I \tilde{\mathbf{z}}}^T \left(\hat{\mathbf{C}}_{\tilde{\mathbf{z}} \tilde{\mathbf{z}}} + \rho_I \mathbf{I}_Z \right)^{-1} (\tilde{\mathbf{z}}(\cdot) - \hat{\mathbf{m}}_{\tilde{\mathbf{z}}}) \quad (11)$$

which is regularized Z -dimensional affine regression!

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

- Yes, e.g. for Gaussian $k(\mathbf{q}, \mathbf{q}') \leftarrow \exp \left(-\frac{1}{2} \|\Lambda^{-1}(\mathbf{q} - \mathbf{q}')\|_2^2 \right)$
[Rahimi and Recht, 2007]
- In such cases, can reduce from $\sim N^2$ to $\sim NZ$ computations

Demonstrated PERK for T_1 , T_2 est from optim (2SP,1DE) scan

Demonstrated PERK for T_1, T_2 est from optim (2SP,1DE) scan

- PERK trained using $N \leftarrow 10^5$ samples from prior dist $p_{\mathbf{x},\nu}$
- To enable precise estimation, support of $p_{\mathbf{x},\nu}$ carefully chosen to coincide with min-max acquisition design support

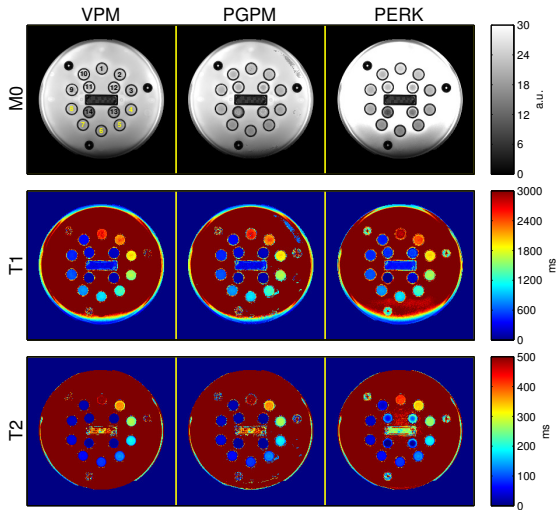
Demonstrated PERK for T_1 , T_2 est from optim (2SP,1DE) scan

- PERK trained using $N \leftarrow 10^5$ samples from prior dist $p_{x,\nu}$
- To enable precise estimation, support of $p_{x,\nu}$ carefully chosen to coincide with min-max acquisition design support

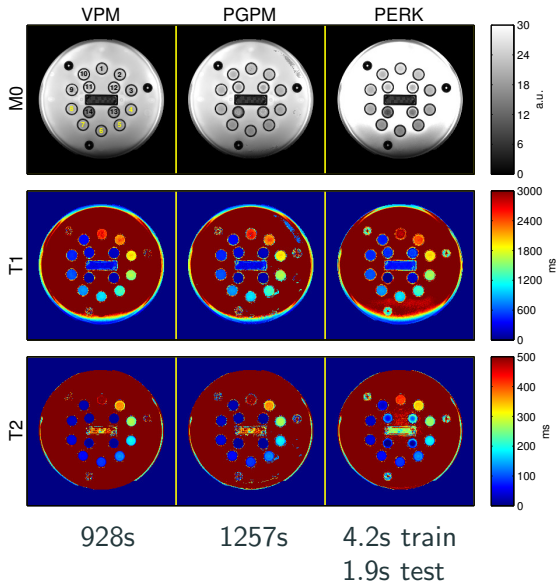
Compared PERK to two well-suited ML estimators:

- dictionary-based grid search estimator via variable projection method (VPM) [Golub and Pereyra, 2003]
- local optim estimator via preconditioned variant (PGPM) of classical gradient projection method [Rosen, 1960]

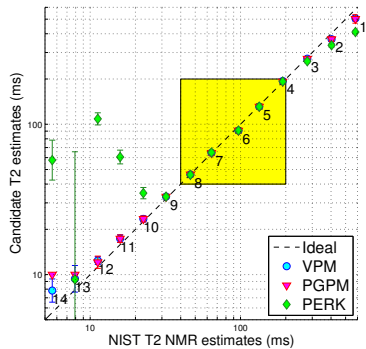
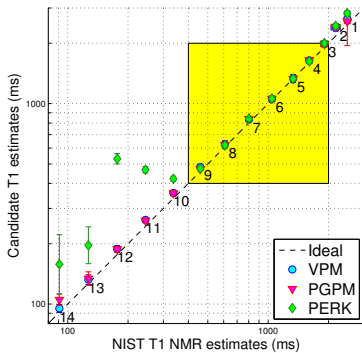
Phantom Results



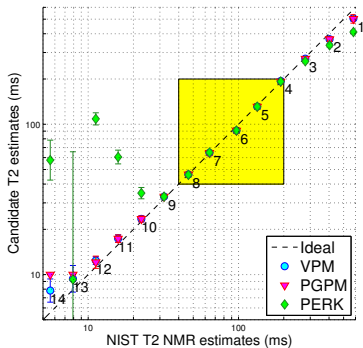
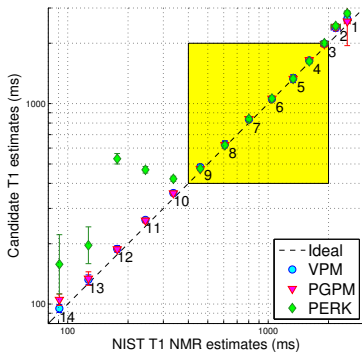
Phantom Results



Phantom Results

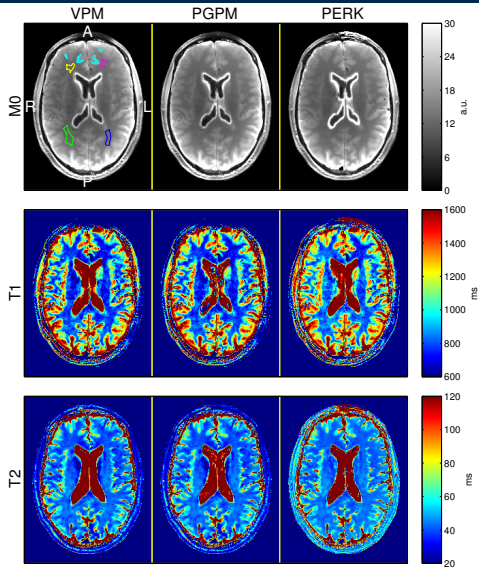


Phantom Results

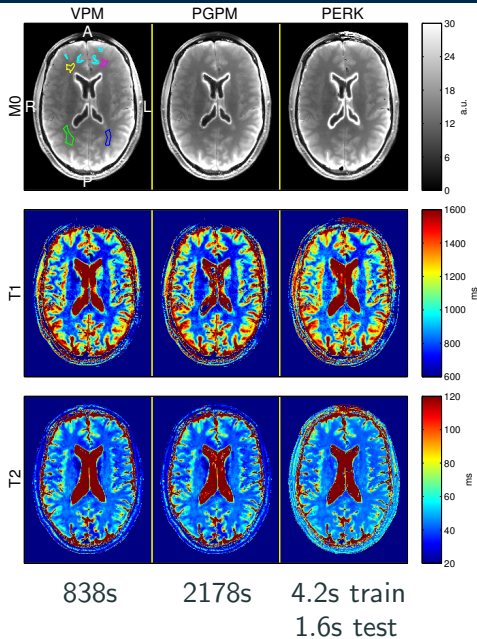


Within support of $p_{x,\nu}$,
PERK and grid search estimates agree excellently.

In vivo Results



In vivo Results



Contributions

- **PERK**: fast, dictionary-free estimator for QMRI

Contributions

- **PERK**: fast, dictionary-free estimator for QMRI
- demonstrated PERK for T_1 , T_2 estimation
 - Phantom (and omitted simulation) results show that PERK and ML estimators yield comparable accuracy/precision
 - *In vivo* PERK and ML estimates are comparable in WM/GM
 - **PERK is consistently at least 140x faster**

Contributions

- **PERK**: fast, dictionary-free estimator for QMRI
- demonstrated PERK for T_1 , T_2 estimation
 - Phantom (and omitted simulation) results show that PERK and ML estimators yield comparable accuracy/precision
 - *In vivo* PERK and ML estimates are comparable in WM/GM
 - **PERK is consistently at least 140x faster**

Can we exploit PERK's speed in a more compelling problem?

Advances in Quantitative MRI:

- **Acquisition** [Ch. 4]
How can we assemble fast, informative collections of scans to enable precise biomarker quantification?
- **Estimation** [Ch. 5]
Given accurate models and informative data, how can we rapidly quantify these biomarkers?
- **Application** [Ch. 6]
Using these tools, can we design a state-of-the-art biomarker?

Numerical Simulation: Acquisition Design

- Simulated many WM-like, GM-like voxel realizations
- Studied sample statistics of T_1, T_2 ML estimates $\hat{T}_1^{\text{ML}}, \hat{T}_2^{\text{ML}}$

Profile	(2, 1)	(1, 1)	(0, 2)	Truth
WM \hat{T}_1^{ML}	830 ± 17	830 ± 15	830 ± 14	832
GM \hat{T}_1^{ML}	$1330 \pm 30.$	1330 ± 24	1330 ± 24	1331
WM \hat{T}_2^{ML}	$80. \pm 1.0$	$80. \pm 2.1$	79.6 ± 0.94	79.6
GM \hat{T}_2^{ML}	$110. \pm 1.4$	$110. \pm 3.0$	$110. \pm 1.6$	110

Table 2: $\hat{T}_1^{\text{ML}}, \hat{T}_2^{\text{ML}}$ sample means \pm sample standard deviations

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*

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

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 $k : \mathcal{Q}^2 \mapsto \mathbb{R}$ that induces choice of RKHS \mathbb{H}

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

$$(\hat{h}_l, \hat{b}_l) \in \left\{ \arg \min_{\substack{h_l \in \mathbb{H} \\ b_l \in \mathbb{R}}} \frac{1}{N} \sum_{n=1}^N (h_l(\mathbf{q}_n) + b_l - x_{l,n})^2 + \rho_l \|h_l\|_{\mathbb{H}}^2 \right\} \quad (12)$$

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

$$\hat{h}_l(\cdot) \equiv \sum_{n=1}^N \hat{a}_{l,n} k(\cdot, \mathbf{q}_n) \quad (13)$$

- Plug (13) into (12); solve now instead for (\hat{a}_l, \hat{b}_l) ; construct:

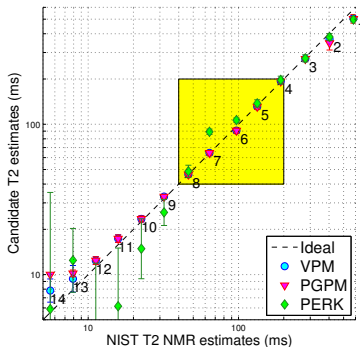
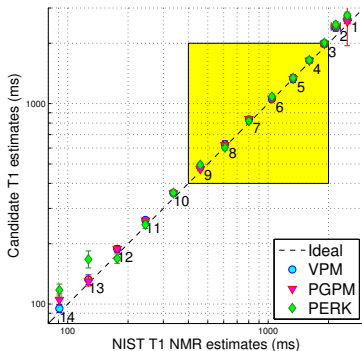
$$\hat{x}_l(\cdot) = \sum_{n=1}^N \hat{a}_{l,n} k(\cdot, \mathbf{q}_n) + \hat{b}_l \quad (14)$$

Numerical Simulation: PERK Estimation

	Truth	VPM	PGPM	PERK
WM T_1	832	832.1 ± 17.2 (17.2)	832.1 ± 16.2 (16.2)	833.0 ± 16.5 (16.5)
GM T_1	1331	1331.5 ± 31.1 (31.1)	1331.2 ± 29.7 (29.7)	1332.1 ± 30.4 (30.4)
WM T_2	79.6	79.61 ± 0.988 (0.988)	79.60 ± 0.952 (0.952)	79.46 ± 0.978 (0.989)
GM T_2	110.	110.02 ± 1.40 (1.40)	110.02 ± 1.35 (1.35)	109.91 ± 1.35 (1.35)

Table 3: Sample means \pm sample standard deviations (RMSEs) of VPM, PGPM, and PERK m_0 , T_1 , T_2 estimates, computed in simulation over 7810 WM-like and 9162 GM-like voxels.

Mismatch in Scan Design vs. Sampling Dist Support



Widening $\text{supp}(p_{x,\nu})$ degrades performance w/in scan design range.
Thus, scan design and param est should be considered in tandem.



Cramér, H. (1946).
Mathematical methods of statistics.

Princeton Univ. Press, Princeton.



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.



Golub, G. and Pereyra, V. (2003).
Separable nonlinear least squares: the variable projection method and its applications.

Inverse Prob., 19(2):R1–26.



Hope, M. D., Wrenn, S. J., and Dyverfeldt, P. (2013).

Clinical applications of aortic 4d flow imaging.

Curr. Cardiovasc. Imag. Rep., 6(2):128–39.



Keenan, K. E., Stupic, K. F., Boss, M. A., Russek, S. E., Chenevert, T. L., Prasad, P. V., Reddick, W. E., Cecil, K. M., Zheng, J., Hu, P., and Jackson, E. F. (2016).

Multi-site, multi-vendor comparison of T1 measurement using ISMRM/NIST system phantom.

In *Proc. Intl. Soc. Mag. Res. Med.*, page 3290.



Nataraj, G., Nielsen, J.-F., and Fessler, J. A. (2017a).

Myelin water fraction estimation from optimized steady-state sequences using kernel ridge regression.

In *Proc. Intl. Soc. Mag. Res. Med.*, page 5076.



Nataraj, G., Nielsen, J.-F., and Fessler, J. A. (2017b).

Optimizing MR scan design for model-based T1, T2 estimation from steady-state sequences.

IEEE Trans. Med. Imag., 36(2):467–77.



Nataraj, G., Nielsen, J.-F., Scott, C., and Fessler, J. A. (2018).

Dictionary-free MRI PERK: Parameter estimation via regression with kernels.

IEEE Trans. Med. Imag.

To appear.



Rahimi, A. and Recht, B. (2007).

Random features for large-scale kernel machines.

In *NIPS*.



Redpath, T. W. and Jones, R. A. (1988).

FADE-A new fast imaging sequence.

Mag. Res. Med., 6(2):224–34.



Rosen, J. B. (1960).

The gradient projection method for nonlinear programming, Part I: Linear constraints.

SIAM J. Appl. Math., 8(1):181–217.



Schölkopf, B., Herbrich, R., and Smola, A. J. (2001).

A generalized representer theorem.

In *Proc. Computational Learning Theory (COLT)*, pages 416–426.
LNCS 2111.



Zur, Y., Wood, M. L., and Neuringer, L. J. (1991).

Spoiling of transverse magnetization in steady-state sequences.

Mag. Res. Med., 21(2):251–63.