script for talk

Thanks for the introduction,
and thank you all
for coming to my thesis defense.

If your doctor ordered an MRI today,
chances are that it would look something like
the grayscale image on the left.
Such qualitative MR images
localize the MR signal,
a complicated function
of both interesting and uninteresting contrast mechanisms.
The types of MR images I'll focus on
will be quantitative
like this flow speed image on the right.
Quantitative MRI (QMRI) seeks
to be more informative than qualitative MRI
by imaging interesting contrast mechanisms more directly.
For instance,
this flow image very clearly delineates
the aorta from the pulmonary artery
and very clearly shows elevated flow speed
due to narrowing of the aorta.

Because MRI is a flexible imaging modality,
there are many types of QMRI applications,
such as diffusion imaging shown here.
On the left is a conventional t2−weighted image,
while the right two images quantify
the degree of diffusive anisotropy,
which for example is useful
for assessing the health
of white matter tracts.

And as a quick teaser for what is to come later,
we used qualitative images
like the one on the left
to create an image of fast−relaxing MR signal fraction,
which may correspond to intact myelin content
that is important in assessing demyelinating conditions
like multiple sclerosis.

In the and other QMRI applications,
the broad goal
is to rapidly and accurately localize biomarkers
(specifically, MR−imageable biomarkers)
from MR data.

For the purposes of this talk,
by biomarker
we mean some measurable tissue property
that indicates a biological process of interest
that is characteristic
to the onset and development of one or more specific disorders.
So for example,
flow rate could be indicative
of the development of a blockage,
which is characteristic to the onset and development
of an ischemic stroke.

By localize,
we mean that seek to quantify these biomarkers
at discretized positions in space.

By accurately,
we mean that we need to use signal models
that describe the underlying physics
in sufficient detail.

By rapidly,
we mean that we would like
to use both fast MR acquisitions
and fast estimation methods.

A key challenge of QMRI
is that 'rapidly' and 'accurately'
are often competing goals:
more accurate models typically depend
on both more biomarkers and nuisance markers,
but estimating more markers usually requires
longer scans and more computation.
I'll be speaking today
about some tools we recently developed
that help address these challenges.

Specifically,
I will be describing new answers
to three questions in this talk.
First,
how can we systematically assemble
fast, informative collections of scans
to enable precise biomarker quantification?

Second,
given data from an informative acquisition,
how can we rapidly quantify these biomarkers?

At the end of the talk,
I'll discuss our ongoing efforts
to apply these tools
for myelin imaging.

I will begin
with acquisition design.

After reconstruction,
each voxel within a conventional mr image
can be described
with this general nonlinear signal model.
Here,
s_d relates L unknown parameters x,
K known parameters nu,
and A acquisitions parameters p_d
to a single voxel y_d of the dth image,
barring complex-gaussian noise epsilon_d.

A scan profile contains capital D such measurements,
and is characterized
by a general vector signal model, bold s.
For example,
in t2 mapping
x might denote t2,
nu might denote a separately estimated b1 map,
y might denote a vector of measurements
at different echo times,
and
P might denote the echo times.

In acquisition design,
the task is to design acquisition parameter matrix P
to enable precise unbiased estimation
of parameters of interest,
which are some subset
of the elements of x.

To build an appropriate objective function,
we use the Fisher information matrix
to relate unbiased estimation precision

to the acquisition parameter matrix.
For complex Gaussian data,
the Fisher information can be interpretted
to mean that the signal provides high information in x
for acquisition parameters P
that yield a large signal gradient.

When the Fisher matrix exists,
the cramer−rao bound ensures
that the covariance of unbiased estimate x−hat
is bounded below
by the inverse of the Fisher matrix,
where this matrix inequality means
that the difference matrix
is positive semi−definite.

This inequality is asympotically tight
for maximum−likelihood estimators,
so for sufficiently high snr data,
the inverse fisher matrix
is a good proxy
for unbiased estimator covariance.

The idea then is to choose acquisition parameters P
such that the inverse fisher matrix
is 'small' in some matrix sense.

More concretely,
we seek P that minimizes an objective function
that is a weighted sum
of the diagonal entries
of the inverse Fisher matrix,
which characterize variances
of scalar entries of x.

This objective cannot be optimized directly
because of its dependence
on object parameters x and nu,
which vary spatially.

Instead,
we consider two alternate problems.
In min−max scan design,
we seek candidate scan parameters P−cup
that minimize the worst−case imprecision,
viewed over tight object parameter ranges set−X and set−N.

In Bayesian scan design,
we instead seek candidate scan parameters
that minimize the expected imprecision,
where the expectation is with respect to
some prior joint distribution on x and nu.
Comparing the two,
min−max design
takes milder distributional assumptions
but involves an objective
that is non−differentiable in P.

We study here the min−max design criterion
and will later return
to the Bayesian criterion
for a more challenging application.

As a demonstration of min−max scan design,
we design a fast acquisition
for precise estimation
of relaxation parameters t1,t2
in white matter and gray matter
of the human brain at 3−tesla field strength.

Specifically,

we consider scan profiles consisting
of spoiled gradient−recalled echo (spgr)
and dual−echo steady−state (dess),
two fast steady−state MR pulse sequeneces.

We choose a time constraint
that allows two spgr scans and one dess scan,
a somewhat conventional t1/t2 acquisition,
and then optimize all scan profiles
that are feasible under this time constraint.
Observe that we take flip angle variation
to be separately estimated and here assumed known.
Observe also
that we take weighting matrix W
to place no emphasis on scale factor m0 estimation
and to place roughly equal emphasis on t1,t2 estimation.

Under the previously mentioned time constraint,
three scan profiles are both feasible
and produce at least as many datasets as latent variables,
each respectively consisting
of (2,1), (1,1), or (0,2) spgr/dess scans.
Here,
we summarize their optimized flip angles,
repetition times,
and optimized cost function values.

Our main finding is
that under this time constraint,
2 optimized dess sequences alone
can produce t1,t2 estimates at least as precise
as spgr/dess scan profiles.

We compared our optimized acquisitions
through phantom experiments.
Over a 256x256x8 fully−sampled 3d matrix,
each of these three fast profiles
took less than 2 minutes.

To assess accuracy in vivo,
we also collected a slower reference scan profile
consisting of 4 inversion recovery
and 4 spin echo single−slice acquisitions.

Lastly,
we collected 2 Bloch−Siegert SPGR scans
for separate flip angle calibration.

Here are results in a quantitative MR phantom
consisting of 14 vials with different t1,t2 values.
Columns denote the three optimized profiles
and the reference profile.
The top images are t1 estimates
while the bottom images are t2 estimates.
Each pair of rows denotes two estimators,
but we will focus on the just the ML estimator for now.
Many of the candidate and reference estimates
appear visually similar,

but it is easier to assess accuracy with plots.
Here,
we plot sample means and sample standard deviations
pooled over 100s of voxels
within the 14 vials of interest.
We compare t1/t2 ml estimates
from the three candidate and one reference profile,
versus nist nmr measurements.
The orange region highlights
vials within the 'tight' parameter range.
Within this region,

t1,t2 estimates from all profiles
exhibit minimal bias.

We next report phantom precision results.
Here,
we repeated each profile 10 times
and estimated t1,t2 std dev
of typical voxels within each vial
across the repetitions.

Pooling these sample standard deviation estimates
within the orange−marked vials,
we can assess the performance
of min−max scan design
by examining the empirical worst−case precision
across scan profiles.

Comparing against the optimized costs,
we observe similar trends across profiles
of empirical vs. predicted standard deviations.

In summary,
we have introduced an MR scan design method
to enable precise parameter estimation,
and we have demonstrated the method
by designing three SPGR/DESS scan profiles
for t1,t2 estimation in the brain.

Phantom and simulation results validated the method
as predictive of unbiased estimation precision.
We did not attempt
to replicate these precision experiments in vivo
because of motion considerations,

but we did assess accuracy.
Again,
columns denote the three candidate profiles
and the slow reference profile;
rows denote t1 vs t2 estimates.
We used narrow colorbars
to distinguish the wm and gm boundary.
Overall,
the wm/gm boundaries are similarly distinguishable
across the t1 estimate
and the (1,1) t2 estimates exhibit much higher wm variation
than the other profiles,
as expected due to low relative precision.

However,
there are significant discrepencies across the profiles
(especially in t2 estimates),
which is suggestive of multi−compartmental relaxation.

To address this type of model mismatch,
we need to develop more complete in vivo signal models
and need to scalably estimate more parameters,

which brings me to the second portion of this talk.

Here,
we continue to use a general signal model,
except we now omit explicit dependence
on scan parameter matrix P,
as it is fixed during parameter estimation.

Our task then is to estimate
latent parameter x,
given image data y and known parameter nu,
on a voxel−by−voxel basis.

this can be a challenging problem
because as mentioned before
s is often a nonlinear function of x,
so inverse problems
based on standard likelihood functions
are non−convex.
furthermore,
signal s might be difficult
to write down analytically.

In some simpler applications
where good initializations
and signal gradients are available,
gradient−based local optimization
may be possible.

In more challenging applications,
researchers have tried
stochastic methods like simulated annealing.

In general though,
the most reliable and popular method
is based on discretizing the parameters
over a grid of possible values
and exhaustively searching
for the parameter
that produces the best fit
with the image data.
In fact,
we employed this method
in the first section.

but let's take a look
at the computational cost of grid search.
In simple t1,t2 estimation
there are 3 latent parameters.
discretizing over the 2 nonlinear parameters,
we need about 100 squared or 10,000 dictionary atoms,
which is manageable.
But let's recall
some of the more challenging
applications mentioned before.
These can have
4,7 and even up to 10 latent parameters,
and the corresponding numbers
of dictionary atoms quickly becomes unmanageable.

So our goal is
to find a method
that scales with L more gracefully.

We approach this problem
taking inspiration from machine learning.
The idea is to learn a nonlinear estimator
from simulated training data.

Specifically,
we sample N instances
of latent and known parameters
and noises
and simulate corresponding image data vectors.

We would like to then construct
nonlinear functions h_l and offsets b_l
that in some sense ''invert'' the signal model
by mapping image data vectors and known parameters
back to the lth latent parameter

Mathematically,

we would like to solve
this function optimization problem

but cannot because it is ill−posed:
there are infinitely many functions
that fit a finite N training points.

So we instead modify the problem
to restrict the function space
over which we optimize
and encourage regularity in the estimator
through function regularization.
It so happens
that the optimal estimator can be expressed
as a linear combination
of so−called kernel functions,
so solving the function optimization problem
is equivalent to solving
for the kernel weights and the offset,
which is a standard convex optimization problem
that we can solve exactly.

Before delving into details,
here is a demonstration of PERK
in a simple 1−dimensional toy problem.
The task is to estimate t2,
given N samples simulated
via simple monoexponential signal model.
The plots denote these samples as black circles
and evaluate the PERK estimator in red
at many test points denoted by blue dots.
PERK balances fitting the training points
and smoothly varying between training points.
Note that even with more training points,
PERK performance begins to degrade
near the edges of the sampling interval,
where poorer problem conditioning
causes for PERK to regress towards the sample mean.

When the dust settles,
the solution to the PERK training problem
can be written in one line.
Here bold x_l collects the training point regressands,

bold uppercase K is the NxN Gram matrix,

M is a demeaning operator
and itself introduces identity matrix In
and the Nx1 ones vector,

and bold lowercase k is a nonlinear kernel embedding operator.

So let us return to the original goal:
does this PERK estimator scale better with L?
Perhaps, since by construction the estimators
are constructed separably!

However,
in practice more complex problems
could require more training samples N
for good accuracy,
in which case it is undesirable
to explicitly compute, store, and/or invert
NxN Gram matrix K.
Fortunately,
for many useful kernels
there exist very accurate kernel approximations,
which as a plus can provide us
some further intuition about the PERK solution.

Suppose there exists a function z−tilde
of moderate dimension capital Z
such the Gram matrix admits this low−rank decomposition.
Here capital Z should be larger
than the dimension of the input space
(so as to still lift the inputs into higher dimension).

Plugging this low−rank decomposition
into the PERK solution
gives what appears to be a messy expression,
but recognizing some terms
as means and covariances,

we see that PERK appears (at least approximately)
to perform regular linear regression
but in a dimension higher
than the dimension of the input space!

But does such an approximate feature mapping
exist and work well?
Yes, at least for certain shift−invariant PSD kernels
like the Gaussian.
In this case,
if at least in practice
the product 'NZ' can be scaled less
than exponentially with L,
we have improved scalability.

We demonstrated PERK for t1,t2 estimation
from one of the previously optimized scan profiles.

To ensure we were learning a well−conditioned estimator,
we trained PERK using many samples drawn
from a prior distribution on object parameters x, nu
whose support was chosen to coincide
with the support over which we performed min−max scan design.

We then compared PERK estimates
to two well−suited ML estimators:
dictionary−based grid search estimates
via the variable projection method
as well as
preconditioned gradient projection method estimates
that were initialized
with a strongly biased method−of−moments estimate.

Here we compare vpm, pgpm, and perk estimates
of m0, t1, and t2.
This m0 image also indicates
which vials are within the training range.
The images appear visually similar,

but PERK is more than two orders of magnitude faster
including training time.
Here I've indicated training and testing time separately
because only the latter scales with the number of voxels,
so the acceleration factor would be closer
to three orders of magnitude
for a full 3D volume.

Here are t1 and t2 accuracy plots for the phantom.
The yellow boxes denote projections
of the sampling distribution's support
over which PERK was trained
and correspond to the yellow vials just shown.

We observe that within this support,
PERK and ML estimates agree excellently.

We then compared VPM, PGPM, and PERK in vivo.

The estimatoes agree reasonably
within the highlighted WM and GM ROIs,

but PERK is again more than two orders of magnitude faster,
including training time.

In summary,
we recently introduced PERK,
a fast, dictionary−free machine−learning inspired method
for QMRI parameter estimation.

We demonstrated PERK in a simple, easily−validated problem
in which it was consistently at least 140x faster
than dictionary−based grid search.

Recently however,
we have been interested
in whether we can exploit PERK's speed
for more challenging problems,

which brings me
to the final topic,
where we use acquisition design and PERK
for myelin water imaging.

Myelin is a lipid−rich substance
that in normal white matter wraps healthy axons,
thereby forming an electrically insulating layer
like rubber around a copper wire.
There is water trapped between these myelin layers,
and when myelin is damaged
in demyelinating diseases like multiple sclerosis,
this water is released
into the surrounding.

Myelin water fraction
denotes the proportion of mr signal
that arises from water trapped within the myelin layers
relative to the total water signal,

and has been shown
to correlate well with intact myelin content.

The gold−standard MW imaging acquisition remains
a multi−echo spin−echo sequence,
from which MWF is characterized
using multi−exponential or more complicated models
of the echo train decay.
Even with more recent MESE acceleration methods,
these experiments are typically speed−limited
by long repetition times.

More recently,
combinations of fast steady−state scans
using variable flip angles
(mcDESPOT)
were shown to produce whole−brain mwf images
in about a half−hour of imaging.
However,
myelin water fraction estimates
from mcDESPOT were shown to disagree with MESE estimates,
likely due to insufficient estimation precision.

So our goal here is to design an imaging workflow
that enables fast, accurate MW content quantification in WM.

Here is a simple voxel−scale model
of how myelin water influences mr signal.

Specifically,
we now distinguish signal
to arise from a 'fast'−relaxing
and a 'slow'−relaxing water compartment.

The physics of these compartments
can be described with six free parameters.

Of these,
we are interested in estimating the fast−relaxing fraction ff,
as a simple measure of myelin water content.

Since our previous single−compartment t1,t2 estimates
from spgr/dess scan profiles
demonstrated sensitivity to multi−compartmental relaxation,
we found it natural to study
2−compartment spgr/dess models.
We studied the assumptions
of a previously developed 2−compartment spgr model

and found
that their absorption of off−resonance effects
into m0 implies either assuming
that different compartments
have the same off−resonance broadening distribution
which may not be physically accurate,
or neglecting exchange
between excitation and readout
which is reasonable only for very short echo times.

Following their ideas,
we then derived a two−compartment dess signal model,
which interestingly required additional approximations
when including exchange
unlese we assumed that the difference
of compartmental off−resonance frequencies
remained constant over time.

Even with all the highlighted assumptions,
closed−form signal models still remain elusive,

and so for simplicity,
we neglect exchange in the ensuing studies.
This may be a reasonable assumption
when describing the interaction
of myelin water
and other water,
as interactions across the hydrophobic myelin membrane
are rather slow (~200ms) compared
to myelin water fraction t2 (~15−40ms).

To estimate six free latent parameters,
at minimum six datasets are required,
and so the scan parameter optimization
for these six datasets is higher−dimensional,
making grid search to optimize the min−max cost less desirable.
Instead,
we use gradient information
to locally optimize the bayesian scan design cost function.

Here x is six−dimensional,
nu again assumes known flip angle variation,
and P again contains nominal flip angles
and repetition times.

Weighting matrix W is fixed
to only place emphasis
on the fast−fraction ff,
and using the inverse mean
rescales the cost function
to be interpretable

as the expected coefficient of variation
of unbiased estimates of ff in wm.

Expectations are approximated
via sample means of samples drawn
from a separable prior,

and the constraint space ensures
reasonable flip angles
and enforces a total scan time constraint.
that is competitive with that of mcDESPOT.

Here are the optimized flip angles
and repetition times
of a scan profile
designed under a total time constraint
comparable to that of mcDESPOT.
We see that the predicted ff relative standard deviation
is 42.5 percent,
which is a significant improvement
over similar calculations
performed for the mcDESPOT acquisition.
Interestingly,
the optimized acquisition
consists entirely of DESS scans
with variable repetition times.
We observed this behavior
in other optimization instances
with different scan time constraints as well,
suggesting that DESS with variable flip angle and repetition times
can be sensitized to two−compartment relaxation.

We applied PERK for fast−fraction estimation
from the optimized DESS acquisition.
We trained PERK using a million samples
drawn from a prior distribution on parameters
that is similar to the Bayesian scan design distribution
but with finite support.

We compared PERK fast−fraction estimates in simulation
to ML fast−fraction estimates,
implemented via an unrealistically narrow grid search
around the ground truth.

We also compared PERK fast−fraction estimates
to unregularized and l2−norm regularized nnls estimates
that arise from two conventional estimators
from mese data.

We first studied these four estimators
in a simulation without any model mismatch.
We simulated data to arise
from two water compartments,
each with different t2 values
but the same t1 value.

Since there is no model mismatch,
ff and fm estimates are comparable.

Here is the two−compartment simulation result.
Along with ff and fm estimates
from the four estimators,
we also plot magnitude difference images
with respect to the ground truth.

We first observe
that PERK not only achieves lower WM RMSE than ML,

but does so in about 500x less time

including training time
for a single−slice experiment.

We next observe
that though NNLS and RNNLS RMSE are lower
than PERK RMSE in WM,
there is clear spatial variation in the MESE error maps
due to flip angle variation,
despite knowledge and compensation
for this flip angle variation.

We next repeated the previous study
in a simulation with model mismatch.
Here,
we simulated data
to perhaps more realistically arise
from three water compartments
each with different t2 and t1 values.
Now,
both DESS ff and MESE fm estimators incur bias,

and so ff and fm estimates need not be comparable.

Here are corresponding results.
PERK is again more than 500x faster than ML.
As expected,
the error maps are generally worse than before,

though PERK WM RMSE is the lowest.
Note that with model mismatch,
the spatial variation of MESE errors
is now apparent even in the fm maps.

We lastly report results from an in vivo experiment.
In a single long study,
we collected the optimized dess acquisition,
a conventional 32−echo mese acquisition,
a bloch−siegert acquisition
for separate flip angle calibration
(used for both DESS ff and MESE fm estimation)
and a variable−flip spgr acquisition
for separate bulk−t1 estimation
(used for MESE fm estimation only).

The mese acquisition was largely conventional,
except we repeated the acquisition twice
to increase effective snr through averaging
and used a somewhat short repetition time
to keep scan times reasonable.

We compared PERK ff estimates
to MESE NNLS and RNNLS fm estimates.
In previous studies,
we tried running a full−scale grid search
on a computing cluster
that ran for about 68 cpu−days
and still did not give reasonable results,
probably because the ML estimation problem
has multiple global minima here.
In any case,
since we acquired the data used here
in late january
and we didn't renew our cluster account,
I don't think I'd be defending today
if we had waited for similar ML results here!

Here are the associated in vivo results.
The elevated MESE fm estimates in medial wm
have been reported in several other MESE studies,

and have been attributed
to overlap of the myelin− and cellular water peaks
in internal capsule t2 distributions.
We also observe
that these patterns are visually similar
to those observed in simulation
due to flip angle variation.
The DESS PERK ff estimates are more homogeneous
across medial and lateral WM ROIs
and distinguish myelinated wm tracts
at least as well as mese fm estimates,

but the DESS ff estimates required 5x less scan time.

In summary
we have introduced a fast ss acquisition
for precise myelin water imaging.

Idealized simulations demonstrated
that PERK and ML ff estimates are comparable
but PERK is more than 500x faster.

More realistic simulations demonstrated
that the accuracy of both MESE fm and DESS ff estimates
may be sensitive to in vivo model mismatch.

Nevertheless,
in vivo experiments are the first
to demonstrate lateral wm myelin water content estimates
from a SS acquisition
that are at all similar
to conventional MESE MWF estimates.

As future work,
it will be interesting
to investigate DESS ff accuracy further
through ex vivo studies
and correlations with other myelin biomarkers.

It will also be interesting
to investigate whether MW imaging can be further improved
by exploiting compartmental off−resonance differences,
a topic that is an active area of research in our group.

More broadly,
it will be interesting to see
whether combining PERK with image reconstruction
can provide performance benefits,
particularly in cases of undersampling.

This concludes the main portion of my talk,

though there are few less mature but still novel topics
discussed in other parts of the thesis.

Finally,
I'd like to say thank you
to a small sample of the many people
who have made this work possible.
For this part,
I've prepared some words in advance,
for fear that I might get emotional.

First,
I thank my co−advisors Jeff and Jon.
Jeff realized that I might like research
before I even realized I might like research
all the way back in 2010
during an undergraduate UM visitation day,
when I was a sophomore at Cornell.

Jeff has since then helped me develop
my intuition and yet also rigor,
my persistence and yet also humility.
I hope to be half as productive as he is one day.
Asking Jon to be my co−advisor in 2014,
also per Jeff's recommendation by the way,
was perhaps the best decision I made during grad school.
Since our very first scan together,
Jon has treated me as a professional researcher
which was key to building my research confidence
in those early days.
Most practical things that I know about MRI today
are due to him.

Next,
I thank my thesis committee.
Clay was a key contributor and coauthor on the PERK paper.
Doug's medical imaging systems lectures are so good
that I visited his class two years ago,
even though I had already taken the class.
He has since then also contributed several times
in helping me get perhaps esoteric algorithms
working on real MR data with real MR nonidealities,
and understanding why things break when they do.
Scott was actually the first person who let me touch
an MR scanner back in 2013
during a medical imaging lab.
He has since then had several discussions with me
about myelin imaging.

Next,
I thank my collaborators at UM.
Mingjie adapted PERK to work on MR fingerprinting data
and ran the fast−fraction scan optimizations.
Steven has been
and I hope plans to continue working
on off−resonance−sensitive myelin imaging.

Next,
I thank the funding that made this work possible.
In particular,
the University of Michigan funded a large portion
of this proof−of−concept work
and deserves recognition
for supporting exploratory research.

Next,
I thank my many friends and colleagues here at UM,
old and new.
You have made what is ultimately a solitary journey
much more fun and fulfilling.
I am aware that I didn't work too often
in the LOJ, or lab of Jeff,
but that's only because I often found myself
socializing too much and working too little while there.

Next,
I thank my roommates, Adam and Trey.
All three of us moved here together
from Cornell back in 2012.
None of us had much facial hair back then.
Adam in particular has lived with me
in the same apartment
for the entire duration of our PhDs.
He has been with me
through every single one of the highs and lows.

Next,
I thank my family.
I don't think too many people can say
that three generations of family
attended their PhD defense.
I am humbled and am grateful

for their unconditional love and support.

And finally,
I thank Manisha.
Simply put,
she is the light of my life.
If you enjoyed the upma,
you have her to thank.
If you didn't enjoy it,
we'll say that I made it.
Manisha, I can't wait to get married
and start the next chapter of our lives together.

I hope that covers everyone here today,
but just in case I'll say again
thank you the audience
for attending my defense,
and I'd be happy to take questions.