ACCELERATED AND VALIDATED MODEL BASED MR SIGNAL RECON

1. Background Reading

- MR signal model derivation [Haacke and Brown, 1999, Bernstein, 2004]. Equations for model based processing of the MR Signal is provided in Section 5.
- Signal equation from a digital signal processing perspective [Candy, 2005].
- Statistical Inverse Problem Framework [Tarantola, 2005]. Laplacian distribution [Eltoft et al., 2006, Babacan et al., 2010a].
- CS Background [Fessler, 2010, Donoho et al., 2006, Candes and Wakin, 2008].
  - Suppose we are given a signal $x \in \mathbb{R}^m$ and a basis $\theta_i =< \varphi_i, x > i = 1, ..., N$ in which the signal is sparse, ie that the 1-norm is bounded
  $$\|\theta\|_1 \leq R$$
  Given $n = N\ log\ m$ measurements, $\theta_1, ...\theta_n$, the approximation of the solution $\bar{x}$ computed as the solution to the L1 problem is of the same quality (with respect to the 2-norm) as using the N largest coefficients. ie
  $$\|x - \bar{x}\| \leq C \frac{1}{\sqrt{N}}$$
  ie there is little benefit, in terms of solution accuracy, if we were to wait longer to collect all measurement coefficients. This is similar to an iterative solver where a handful of matrix vector product product a solution accuracy $\mathcal{O}(10^{-6})$ precision with $\mathcal{O}(N)$ complexity and there is really no need to use a direct solver to get a $\mathcal{O}(10^{-16})$ precision with $\mathcal{O}(N^3)$ complexity.
  - As a big surprise, From [Donoho et al., 2006], "So adaptive information is of minimal helpdespite the quite natural expectation that an adaptive method ought to be able iteratively somehow "localize" and then "close in" on the "big coefficients."
  $$\text{error non-adaptive} \leq Const. \cdot \text{ error adaptive}$$
- Model Based CS [Baraniuk et al., 2010].
  - In order to get a good approximation, the measurement matrix $\Phi = [\varphi_1..\varphi_N]$ should satisfy the restricted isometry property (RIP)

(1) $$(1 - \delta_k)\|x\|_2 \leq \|\Phi x\| \leq (1 + \delta_k)\|x\|_2$$

  The RIP property ensures that submatrices are close to an isometry and are thus information preserving.
- Model based CS Recon Applications in MR [Cao et al., 2014, Wright et al., 2014, Assländer et al., 2013].
- GPU Code [Farber and Farber, 2011, Munshi, 2011, Hansen and Sø rensen, 2013].
- Computational Harmonic Analysis is an active field of research to determine basis functions with rapidly decaying coefficients [Feichtinger et al., 1989, Candes and Donoho, 2004].
- Dictionaries provide an alternative view for construction a basis for the sparse reconstruction [Candès et al., 2011, Aharon et al., 2006, Elad and Aharon, 2006].
  - From [Elad and Aharon, 2006] "with the growing realization that regular separable 1-D wavelets are inappropriate for handling images, several new tailored multiscale and directional redundant transforms have been introduced, including the curvelet [Candes and Donoho, 2002, Candes and Donoho, 2004], contourlet [Do and Vetterli, 2003a, Do and Vetterli, 2003b], wedgelet [Donoho et al., 1999], bandlet[Le Pennec and Mallat, 2005b, Le Pennec and Mallat, 20 , and the steerable wavelet [Freeman and Adelson, 1991, Simoncelli et al., 1992]."
- Markov Chain Monte Carlo (MCMC) was placed in the top 10 most important algorithms of the 20th century [1]. The basic idea behind MCMC is to construct a Markov chain on the state space whose stationary distribution is the target density of interest, Ch 24 [Murphy, 2012].

2. Mathematical Framework

The underlying modeling philosophy of the presentation is that the physics models are 1st order accurate or within 70-80% of the needed accuracy and the error "is within" Gaussian noise. Gaussian distributions provide analytical representations of the random variables of interest within the Bayesian setting and provide a crux for understanding. In particular, we say that a random variable η belongs to a multi-variate normal distribution of mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$

$$\eta \sim \mathcal{N}(\mu, \Sigma) \Rightarrow p(\eta) = \frac{1}{2\ \pi\ \det \Sigma} \exp\left(-\frac{1}{2}\|\mu - \eta\|_\Sigma^2\right)$$

[1] http://www.siam.org/pdf/news/637.pdf

1. We will consider multiple physics/sensor models, $\mathcal{G}_i(\vec{k},\theta) : \mathbb{R}^a \times \mathbb{R}^m \to \mathbb{R}^n$, $i = 1, ..., N_{\text{model}}$. Each model $\mathcal{G}_i$ maps deterministic acquisition parameters, $\vec{k} \in \mathbb{R}^a$, and uncertain parameters, $\theta \in \mathbb{R}^m$ to observables, $\vec{z} \in \mathbb{R}^n$ (or $\vec{z} \in \mathbb{C}^n$). To be explicit, will assume that the measurement models are corrupted by zero mean white noise noise of a **known** covariance matrix, $\Sigma_z \in \mathbb{R}^{n\times n}$

$$\vec{z} = \mathcal{G}_i(\vec{k};\theta) + \eta \qquad \eta \sim \mathcal{N}(0,\Sigma_z) \tag{2}$$

η may be interpreted as the measurement noise or the acquisition noise in the sensor model. For a deterministic measurement model $\mathcal{G}$, the conditional probablity distribution has an explicit analytical form and may be written as a **known** Gaussian distribution.

$$p(\vec{z}|\theta) = \mathcal{N}(\mathcal{G}_i(\vec{k};\theta),\Sigma_z) \qquad i = 1, ..., N_{\text{model}}$$

This model corresponds to the fifth sketch in Figure 1 ([Tarantola, 2005] Figure 3.2). Generally, this is a nonlinear model in which linear approximations are not appropriate.

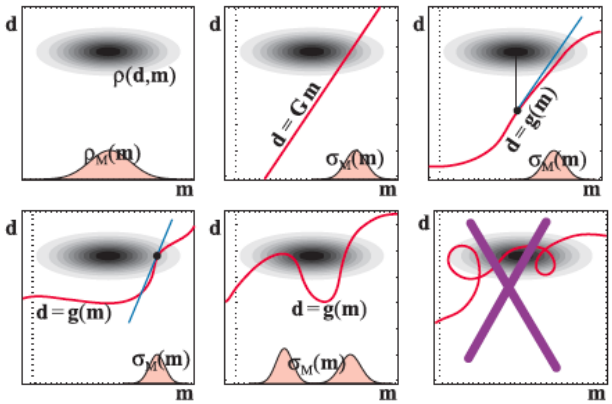

Figure 3.2. *The first sketch is a representation of the probability density* $\rho(\mathbf{d},\mathbf{m})$ *representing both the information we have on the observable parameters* $\mathbf{d}$ *(data) and the a priori information on the model parameters* $\mathbf{m}$ *(the marginal* $\rho_{\mathrm{M}}(\mathbf{m})$ *). In the second sketch, the forward equation* $\mathbf{d} = \mathbf{G}\,\mathbf{m}$ *is linear. The posterior probability density* $\sigma_{\mathrm{M}}(\mathbf{m})$ *is then also Gaussian. In the third sketch, the forward equation* $\mathbf{d} = \mathbf{g}(\mathbf{m})$ *can be linearized around* $\mathbf{m}_{\text{prior}}$ *, giving* $\mathbf{g}(\mathbf{m}) \simeq \mathbf{g}(\mathbf{m}_{\text{prior}}) + \mathbf{G}(\mathbf{m}-\mathbf{m}_{\text{prior}})$ *, where* $\mathbf{G}$ *represents the derivative matrix with elements* $G^i{}_\alpha = (\partial g^i / \partial m^\alpha)_{\mathbf{m}_{\text{prior}}}$ *. The posterior probability density* $\sigma_{\mathrm{M}}(\mathbf{m})$ *is approximately Gaussian. In the fourth sketch, the forward equation* $\mathbf{d} = \mathbf{g}(\mathbf{m})$ *can be linearized around the maximum likelihood point,* $\mathbf{m}_{\mathrm{ML}}$ *:* $\mathbf{g}(\mathbf{m}) \simeq \mathbf{g}(\mathbf{m}_{\mathrm{ML}}) + \mathbf{G}(\mathbf{m}-\mathbf{m}_{\mathrm{ML}})$ *, where now* $\mathbf{G}$ *represents the derivative operator with elements* $G^i{}_\alpha = (\partial g^i / \partial m^\alpha)_{\mathbf{m}_{\mathrm{ML}}}$ *. The point* $\mathbf{m}_{\mathrm{ML}}$ *has to be obtained by the nonquadratic minimization of* $S(\mathbf{m}) = \frac{1}{2}\big((\mathbf{g}(\mathbf{m})-\mathbf{d}_{\text{obs}})^t\,\mathbf{C}_{\mathrm{D}}^{-1}\,(\mathbf{g}(\mathbf{m})-\mathbf{d}_{\text{obs}}) + (\mathbf{m}-\mathbf{m}_{\text{prior}})^t\,\mathbf{C}_{\mathrm{D}}^{-1}\,(\mathbf{m}-\mathbf{m}_{\text{prior}})\big)$ *. In the fifth sketch, the forward equation* $\mathbf{d} = \mathbf{g}(\mathbf{m})$ *cannot be linearized, so the a posteriori probability density may be far from a Gaussian and special methods must be used (see text). In the last sketch, the nonlinearities between the parameters are so strong that the methods proposed in this elementary text cannot be used.*

FIGURE 1. [Tarantola, 2005] Figure 3.2. Different models include: single echo, multi-echo, multi-flip, motion correction, and sparsity. The models may be linear or nonlinear.

2. Additional **known** information is the prior probability distributions for the model parameters, $p(\theta)$.
   - Prior parameters may be a Gaussian distribution of **known** mean, $\hat{\theta}$ and covariance, Σ_θ

$$\theta \sim \mathcal{N}(\hat{\theta},\Sigma_\theta)$$

   - Sparsity in the model parameters may also be an appropriate assumption. Assume model parameters, θ are sparse under a given transformation Φ. A Laplacian distribution of **known** mean, $\hat{w}$ and covariance, Σ_w is defined in terms of $\|\cdot\|_1$ and is known to enforce sparsity [Aravkin et al., 2011a, Aravkin et al., 2011b].

$$w = \Phi\theta \qquad w \sim \mathcal{L}(\hat{w},\Sigma_w) \Rightarrow p(w) = \frac{1}{\sqrt{\det\,(2\,\Sigma_w)}} \exp\left(-\sqrt{2}\left\|\Sigma_w^{-1/2}\,(w-\hat{w})\right\|_1\right)$$

Within a compressed sensing setting, we typically assume the each component of the representation is **independent** with known covariance, λ and zero mean $\hat{w} = 0$.

$$p(w) = \prod_i \frac{1}{\sqrt{2\,\lambda}} \exp\left(-\frac{\sqrt{2}|w_i|}{\lambda}\right)$$

Intuitively, for a sparse representation, the probability that an individual independent component is zero is very high. However, the heavy tail promotes/allows outliers which is representative of the non-zero components. I.e. compared to a Gaussian the probability of an outlier is greater.

$$\exp(-x^2) < \exp(-|x|) \qquad x > 1$$

@dmitchell412 - TODO - validate the Laplacian assumption on various sparsity models. Given a transformation Φ_i,

$$w = \Phi_i \, \theta$$

each w_i is assumed independent. Statistical moments of the histograms of the w_i's should resemble a Laplacian distribution. Can the independent assumed be evaluated within this setting as well ?

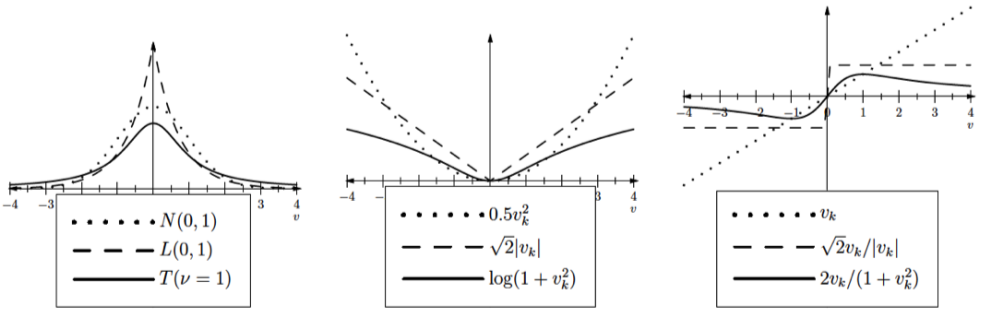

Fig. 1. Gaussian, Laplace, and Student's t Densities, Corresponding Negative Log Likelihoods, and Influence Functions (for scalar v_k).

FIGURE 2. [Aravkin et al., 2011a, Aravkin et al., 2011b]. Figure 1. Comparison of pdf.

3. For each model, $\mathcal{G}_i(\vec{k};\theta)$, determine the optimal measurement locations, $\vec{k}$. This may be determined by maximizing the mutual information between measurements, z, and parameters θ. The mutual information between parameters θ and z may be computed as the KL distance between the joint and independent probabilities.

$$I(\Theta, Z) = D_{\mathrm{KL}}(p(\theta, z)||p(\theta)p(z)) = \int_z \int_\theta p(\theta, z) \ln\left(\frac{p(\theta, z)}{p(\theta)p(z)}\right) \, d\theta \, dz$$

Motivation for this 'distance'/divergence measure may be understood from information theory [Pluim et al., 2003, Cover, 2006]. Given a probability space $(\Omega, \mathcal{F}, p)$ (probability maps from the sigma-algebra of possible events $p : \mathcal{F} \rightarrow [0, 1]$ sigma-algebra, $\mathcal{F}$, defined on set of 'outcomes' Ω [Durrett, 2010]), we will define information of an event as proportional to the inverse probability.

$$\text{information} \equiv \frac{1}{p(\theta)}$$

Intuitively, when a low probability event occurs this provides high information. The informational entropy is an *average* of the information content for a sigma algebra of events $\mathcal{F}$

$$H(\Theta) = \int_\theta p(\theta) \ln \frac{1}{p(\theta)}$$

Hence this entropy measure is an average of the information content for a given set of events, $\mathcal{F}$, and is proportional to the variance or uncertainty in which the set of events occur. This agrees with thermodynamic entropy; if the information containing events are completely spread out such as in a uniform distribution, the entropy is maxmized. The entropy is zero for a probability distribution in which only one event occurs. Zero information is gained when the same event always occurs ($0 \ln \frac{1}{0} = 0$). Entropy provides an intuitive interpretation of mutual information.

$$\begin{aligned}
I(\Theta, Z) &= \int_z \int_\theta p(\theta, z) \ln\left(\frac{p(\theta, z)}{p(\theta)p(z)}\right) \, d\theta \, dz = \int_z \int_\theta p(\theta, z) \ln\left(\frac{p(z|\theta)}{p(z)}\right) \, d\theta \, dz \qquad p(\theta, z) = p(z|\theta)p(\theta) \\
&= -\int_z \ln p(z) \int_\theta p(\theta, z) \, d\theta \, dz + \int_\theta p(\theta) \int_z p(z|\theta) \ln p(z|\theta) \, d\theta \, dz \\
&= -\int_z p(z) \ln p(z) \, dz + \int_\theta p(\theta) H(Z|\Theta = \theta) \qquad p(z) \equiv \int_\theta p(\theta, z) \, d\theta \\
&= H(Z) - H(Z|\Theta) \qquad \text{The conditional entropy, } H(Z|\Theta), \text{ is the average of } H(Z|\Theta = \theta).
\end{aligned}$$

With this expression the mutual information, $I(\Theta, Z)$, "translates to the amount of uncertainty about the measurements Z minus the uncertainty about the measurements Z when the model parameters Θ are known. In other words,

MI is the amount by which the uncertainy about the measurements Z decreases when the model parameters Θ are given: the amount of information the model parameters Θ contains about Z" [Pluim et al., 2003].

With this interpretation in mind, we want to find acquisition parameters, $\vec{k}$, for which the model parameters, Θ, contains the most information about the measurements Z. Similarly, find acquisition parameters, $\vec{k}$, that decrease the spread of measurement information given model prediction parameters Θ.

$$H(Z|\Theta) = \int_\theta p(\theta) \int_z p(z|\theta) \ln p(z|\theta) \; d\theta \; dz$$

Notice that the conditional entropy is an average of an average. The average measurment information given model parameter θ is computed for each model parameter θ. Then entropy for each θ is averaged. The decrease in the spread of information is evaluated with respect to the marginalized spread of measurement Z information, $H(Z)$.

Other measures of distances to maximize joint information are the Cauchy-Schwarz divergence [Kampa et al., 2011]

$$D_{\text{CS}}(p(\theta,z)||p(\theta)p(z)) = -\ln\left(\frac{\int_z\int_\theta p(\theta,z)p(\theta)p(z)\; d\theta\; dz}{\int_z\int_\theta p^2(\theta,z)\; d\theta\; dz \int_z\int_\theta p^2(\theta)p^2(z)\; d\theta\; dz}\right)$$

and the Jensen-Renyl divergence [Hamza and Krim, 2003]

$$D_{\text{JR}}(p(\theta,z)||p(\theta)p(z)) = ?$$

Bayes theorem is fundamental to the approach.

$$p(y|x)p(x) = p(x,y) = p(x|y)p(y)$$

@dmitchell412 - TODO - provide an intuitive interpretation of Bayes rule and an intuitive interpretation of conditional probability using Gaussian distributions

Bayes rule may be applied to simplify this expression to the expected value of the KL-distance

$$\begin{aligned} p(\theta,z) &= p(\theta|z)\; p(z) = p(z|\theta)\; p(\theta) \\ I(\theta,z) &= \int_z\int_\theta p(z|\theta)p(\theta)\ln\left(\frac{p(z|\theta)p(\theta)}{p(\theta)p(z)}\right)\; dz\; d\theta \\ &= \int_\theta p(\theta) \underbrace{\int_z p(z|\theta)\ln\left(\frac{p(z|\theta)}{p(z)}\right)\; dz}_{D_{KL}(p(z|\theta)||p(z))}\; d\theta \\ &= \mathbb{E}_\theta\left[D_{KL}(p(z|\theta)||p(z))\right] \\ &= \mathbb{E}_z\left[D_{KL}(p(\theta|z)||p(\theta))\right] \end{aligned} \tag{3}$$

The probability of the measurements $p(z)$ must be interprited in terms of the known information. The probability of the measurements may be derived from the marginalization of the joint probability and has the interpretation as the projection of the joint probability onto the measurement axis.

$$p(z) = \int_\theta p(\theta,z)\; d\theta = \int_\theta p(\theta|z)\; p(\theta)\; d\theta$$

Similarly, the probabilty of the model parameters given the data, $p(\theta|z)$ in understood from Bayes rule.

$$p(\theta|z) = \frac{p(z|\theta)\; p(\theta)}{p(z)}$$

In general, Markov Chain Monte Carlo (MCMC) [Murphy, 2012] are needed to compute the high dimensional integrals of the mutual information (3).

- @madankan @dmitchell412 - TODO - need algorithm to evaluate (3).
- For the particular case of **linear** physics models: The mutual information integral may be simplified using an analytic expression for the KL distance obtained to compute the relative information between two Gaussians, see [Tarantola, 2005] Section 6.20

$$I(f_1;f_0) = \ln\left(\frac{det^{1/2}C_0}{det^{1/2}C_1}\right) + \frac{1}{2}\|x_1 - x_0\|^2_{C_0} + \frac{1}{2}\; trace(C_1C_0^{-1} - I) \tag{4}$$

- In general, measurements, $\vec{z}$, and model parameters, $\vec{\theta}$, are vector valued. Under the **strong assumption** that each is independent, same dimension $\vec{z},\vec{\theta} \in \mathbb{R}^N$, AND conditional probability independent, MI reduces to the component-wise sum. Note that each conditional probability, $p_i(z_i|\theta_i)$, depends ONLY on z_i and θ_i.

$$p(z) = \prod_i^N p_i(z_i) \qquad p(\theta) = \prod_i^N p_i(\theta_i) \qquad p(z|\theta) = \prod_i^N p_i(z_i|\theta_i)$$

$$
\begin{aligned}
I(\Theta, Z) &= H(Z) - H(Z|\Theta) = -\int_z p(z) \ln p(z)\ dz + \int_\theta p(\theta) \int_z p(z|\theta) \ln p(z|\theta)\ d\theta\ d\vec{z} \\
&= -\int_{z_1}\int_{z_2} \ldots \int_{z_N} \prod_i^N p_i(z_i) \ln \prod_j^N p_j(z_j)\ d\vec{z} + \int_{\vec{\theta}} p(\theta) \int_{z_1}\int_{z_2} \ldots \int_{z_N} \prod_i^N p_i(z_i|\theta) \ln \prod_j^N p_j(z_j|\theta)\ d\vec{\theta}\ d\vec{z} \\
&= -\sum_j \underbrace{\int_{z_1}\int_{z_2} \ldots \int_{z_N} \prod_i^N p_i(z_i) \ln p_j(z_j)\ d\vec{z}}_{\int_{z_i, i\neq j} p_i(z_i)=1} + \sum_j \int_{\vec{\theta}} p(\theta) \underbrace{\int_{z_1}\int_{z_2} \ldots \int_{z_N} \prod_i^N p_i(z_i|\theta) \ln p_j(z_j|\theta)\ d\vec{\theta}\ d\vec{z}}_{\int_{z_i, i\neq j} p_i(z_i|\theta)=1} \\
&= -\sum_j \int_{z_j} p_j(z_j) \ln p_j(z_j)\ dz_j + \sum_j \underbrace{\int_{\theta_1}\int_{\theta_2} \ldots \int_{\theta_N} \prod_k^N p_k(\theta_k)}_{\int_{\theta_k, k\neq j} p_k(\theta_k)=1} \int_{z_j} p_j(z_j|\theta) \ln p_j(z_j|\theta)\ d\vec{\theta}\ d\vec{z} \\
&= \sum_j \left(H(Z_j) - H(Z_j|\Theta_j)\right) = \sum_j I(\Theta_j, Z_j)
\end{aligned} \tag{5}
$$

- Yet another motivation for the use of Gaussian distributions is that the Entropy is analytic [Ahmed and Gokhale, 1989] for a multivariate Gaussian random variable, X.

$$
X \sim \mathcal{N}(\mu, \Sigma) \qquad \mu \in \mathbb{R}^n \qquad H(X) = \frac{1}{2} \ln\left((2\ \pi\ e)^n\ \det \Sigma\right) \tag{6}
$$

- For further intuition, consider the below toy problem in 1D with Gaussian distributions. Typically, we are given the prior, $p(\theta)$, and the conditional $p(z|\theta)$.

$$
p(\theta) \sim \mathcal{N}\left(\hat{\theta}, P_\theta\right) = \frac{1}{P_\theta \sqrt{2\ \pi}} \exp^{-\frac{(\theta-\hat{\theta})^2}{2\ P_\theta}} \qquad p(z|\theta) \sim \mathcal{N}\left(\frac{\theta+b}{a}, P_z\right) = \frac{1}{P_z \sqrt{2\ \pi}} \exp^{-\frac{(z-\frac{\theta+b}{a})^2}{2\ P_z}}
$$

Within this simple example the conditional entropy is analytic, (6).

$$
H(Z|\Theta) = -\int_\theta d\theta p(\theta) \underbrace{H(Z|\Theta=\theta)}_{\frac{1}{2}\ln((2\ \pi\ e)P_z)} = -\frac{1}{2}\ln\left((2\ \pi\ e)P_z\right) \underbrace{\int_\theta d\theta p(\theta)}_{=1} = \frac{1}{2}\ln\left((2\ \pi\ e)P_z\right)
$$

For this situation, the marginalizations of the measurements $p(z)$ is also Gaussian. Section 6.20, Convolution of Two Gaussians [Tarantola, 2005] in $\mathbb{R}^n$

$$
\begin{aligned}
I &= \int d\ \vec{d} \exp\left(-\frac{1}{2}\left((d-d_0)^\top C_d^{-1}(d-d_0) + (d-g(m))^\top C_T^{-1}(d-g(m))\right)\right) \\
&= (2\pi)^{n/2} \det(C_d^{-1} + C_T^{-1})^{-1/2} \exp\left(-\frac{1}{2}\left((d_0-g(m))^\top (C_d + C_T)^{-1}(d_0-g(m))\right)\right)
\end{aligned} \tag{7}
$$

For our case in $\mathbb{R}$

$$
\begin{aligned}
p(z) &= \int_\theta p(z|\theta)\ p(\theta)\ d\theta = \int_\theta \frac{1}{P_\theta\sqrt{2\ \pi}} \exp^{-\frac{(\theta-\hat{\theta})^2}{2\ P_\theta}}\ \frac{1}{P_z\sqrt{2\ \pi}} \exp^{-\frac{(z-\frac{\theta+b}{a})^2}{2\ P_z}}\ d\theta \\
&= \frac{1}{2\ \pi\ P_\theta\ P_z} \int_\theta d\theta \exp^{-\frac{1}{2}\left(\frac{(\theta-\hat{\theta})^2}{P_\theta} + \frac{(\theta-az+b)^2}{a^2\ P_z}\right)} \\
&= \frac{1}{2\ \pi\ P_\theta\ P_z} \left[\sqrt{\frac{2\ \pi}{\left(\frac{1}{P_\theta} + \frac{1}{a^2 P_z}\right)}} \exp^{-\frac{1}{2}\left(\frac{(\hat{\theta}-az+b)^2}{P_\theta + a^2\ P_z}\right)}\right] \underbrace{\Rightarrow}_{\text{verify!}} Z \sim \mathcal{N}\left(\frac{\hat{\theta}+b}{a}, P_\theta + a^2\ P_z\right)
\end{aligned} \tag{8}
$$

We will choose a particular representation of MI that exploits the Gaussian structure of our problem.

$$
I(\theta, z) = H(Z) - H(Z|\Theta) = \frac{1}{2}\ln\left((2\ \pi\ e)\left(P_\theta + a^2\ P_z\right)\right) - \frac{1}{2}\ln\left((2\ \pi\ e)P_z\right) = \frac{1}{2}\ln\left(\left(\frac{P_\theta}{P_z} + a^2\right)\right) \tag{9}
$$

The optimal acquistion parameters $\vec{k}$ maximize the mutual information

$$
\max_k I(\theta, z)
$$

As a more tractable alternative direction, a sensitivity analysis may be used to locate the measurements that provide the most variance in model predictions for uncertain parameters, θ

$$
\max_k \left\langle \mathcal{G}_i(\vec{k}; \theta) \right\rangle
$$

4. Given the optimal acquistion parameters $\hat{k}$, reconstuct the image from measurements z. The solution to the inverse problem is the probability of the model parameters given the data

$$p(\theta|z) = \frac{p(z|\theta)\ p(\theta)}{p(z)} \propto p(z|\theta)\ p(\theta)$$

- This should reduce the the misfit function S(m) of Tarantola, as seen in [Tarantola, 2005] Section 3.2
- *Deterministic regularized optimization is equivalent to finding the maximum likelihood point*

$$\max_{\theta} p(z|\theta)\ p(\theta)$$

For the case of sparse prior assumption under the tranformation Φ, the deterministic problem reduces to the form typically encountered in compressed sensing applications. (log likelihood)

$$p(z|\theta) = \mathcal{N}(\mathcal{G}_i(k;\theta), \Sigma_z) \qquad w \sim \mathcal{L}(\hat{w}, \Sigma_w) \qquad w = \Phi\theta$$

$$\min_{\theta} \left(\frac{1}{2} \left\| \mathcal{G}_i(k;\theta) - z \right\|_{\Sigma_z}^2 + \sqrt{2} \left\| \Sigma_w^{-1/2} \left(\Phi\theta - \hat{w} \right) \right\|_1 \right)$$

Under the assumption of zero mean, $\hat{w} = 0$, and uncorrelated parameters, $\Sigma_w^{-1/2} = \lambda I$, this reduces to the usual form

$$\min_{\theta} \left(\frac{1}{2} \left\| \mathcal{G}_i(k;\theta) - z \right\|_{\Sigma_z}^2 + \sqrt{2}\lambda \left\| \Phi\theta \right\|_1 \right)$$

- MCMC techniques are needed to find moments of quanitities of interest, $Q(\theta)$

$$\mathbb{E}[Q] = \int p(\theta) q(\theta)\ d\theta$$

$$p(\theta) = \int p(\theta, z)\ dz = p(\theta) \int p(z|\theta)\ dz \qquad \text{why is this true ?}$$

- Consider the L1 framework of [Tarantola, 2005] Section 4.4, $p(\theta)$ should be a Laplacian distribution [Eltoft et al., 2006, Babacan et al., 2010a].
- consider truncated Gaussian for Gaussian likelihood time uniform prior. `http://en.wikipedia.org/wiki/Truncated_no`

5. Validate the reconstruction and quantify the speedup.
6. Compute the model plausibility for each model, $\mathcal{G}_i$

3. Problem Definition

Consider the schematic of the pulse sequence for a multi-echo acquistion provided in Figure 3.

For a given tissue, $\Omega \subset \mathbb{R}^3$, with heterogenious intrinistic relaxation properties $T1(r)$ and $T2^*(r)$ we will assume a multi-echo signal, $s_l(t,n) \in \mathbb{C}$, from the l-th coil at the n-th echo of the form of a linear combination of damped exponentials. The signal strength is dependent on the coil sensitivity $c_l(r)$.

$$s_l(t,n) = \int_{\Omega} c_l(r) w[n] e^{-i\int_0^t \gamma \vec{G}(\tau)\cdot r d\tau}\ dr \qquad w[n] = \sum_j^{N_{\text{species}}} C_j \Lambda_j^n \qquad n = 0, 1, 2, \ldots, N_{\text{echo}} - 1$$

(10)

$$\Lambda_j^n = e^{-\left(\underbrace{i\ \Delta B_{0_j}(r)}_{\text{off-resonance}} + \frac{1}{T2_j^*} \right) (\text{TE} + n\ \text{ESP})} \qquad C_j = \frac{M_{0_j} \sin(\gamma\theta_N) \left(1 - e^{-TR/T1_j}\right)}{\left(1 - \cos(\gamma\theta_N)\, e^{-TR/T1_j}\right)} e^{-i\phi_j} \in \mathbb{C}$$

Each weighted exponential, $C_j\Lambda_j$, represents a distinct chemical species, ie water, fat, sodium hydroxide, etc. The complex amplitude, $C_j \in \mathbb{C}$, depends on (1) acquisition parameters - repetition time, TR, and flip angle, θ_N, as well as (2) tissue dependent properties - spin lattice relaxation, $T1_j$, proton density M_{0_j}, and initial phase offset, ϕ_j. Similarly, the complex exponential, Λ_j, depends on (1) acquisition parameters - echo time, TE, echo spacing , ESP, as well as (2) tissue dependent properties - spin spin relaxation, $T2_j^*$, and tissue depending off-resonsance arising from temperature change, motion, and/or susceptibility, $\Delta B_{0_j}(r)$.

$$\Delta B_{0_j}(r) = \begin{cases} 2\pi f_j = 2\pi \underbrace{\alpha B_0 \Delta T_j(r)}_{f_j} & \text{temperature} \\ \quad \ldots = & \text{susceptibility} \\ \quad \ldots = & \text{motion} \end{cases}$$

We will assume:

- off-resonance, ΔB_{0_j} does not depend on readout time

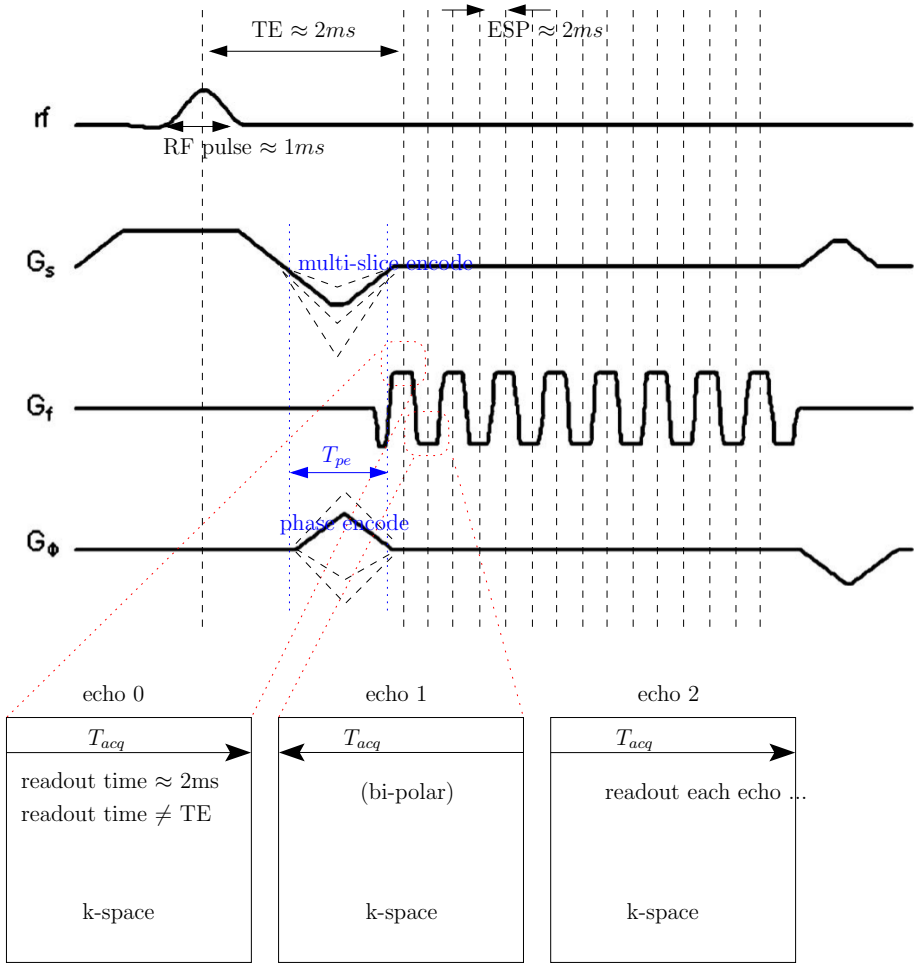

FIGURE 3. Diagram of multiecho sequence used for chemical shift imaging [Taylor, 2008]. In general the sequence plays out as (1) RF/slice select, multiple slices for stack of 2D images is possible. (2) Phase encode within the selected 2d slice. (3) Readout all echoes before moving on to next phase encode. A *bi-polar* k-space trajectory is shown: Readout alternates been left-to-right and right-to-left, is positive/negative. *Uni-polar* trajectories readout in a *single* direction.

- constant gradients

$$\begin{cases} k_x = \gamma G(t - TE) \\ k_y = \gamma G_y T_{pe} \\ k_z = \gamma G_z T_{pe} \end{cases} \qquad |t - TE| < T_{acq}/2 \qquad \vec{k} \equiv \frac{\gamma \, \vec{G} \, t}{2\pi}$$

The time in this signal model, t, represents the measurements during the readout and is less the the repetition time $t < TR \approx 500ms$.

$$t = iii \cdot \Delta t \quad iii = 0, ..., 255 \qquad \Delta t = \frac{T_{acq}}{256}$$

Note that the acquisition time for a single echo/single slice under this model may be estimated as the number of phase encodes times the repetition time.

$$\text{acquisition time} = \#\text{phase encodes} \ \cdot TR \leq 256 \cdot 500ms \approx 2\text{min}$$

- the phase induced temperature change is measured at the echo time

$$\int_0^{\text{TE}} f(\tau) d\tau \approx f(TE)TE$$

Under these assumptions, the measurements at the n-th echo $s(t_i, n)$, have the intuitive interpretation as the Fourier coefficient of the complex image, $c_l(r) \sum_j C_j \Lambda_j^n : \mathbb{R}^3 \to \mathbb{C}$.

$$(11)\qquad s_l(t,n) = \text{Fourier transform}\left\{\underbrace{c_l(r)\sum_j C_j\Lambda_j^n}_{\equiv f(r):\mathbb{R}^3\to\mathbb{C}}\right\} = \mathcal{F}(f(r)) = \int_\Omega \left(c_l(r)\sum_j C_j\Lambda_j^n\right) e^{-2\pi i\vec{k}\cdot r}\, dr = s_l(k_x,k_y,k_z,n)$$

TODO - @dmitchell412 - derive this model and verify

Within the context/notation of Eqn (2),

$$s_l(k_x,k_y,k_z,n) = \mathcal{G}\left(\underbrace{\vec{k}, TE, TR, ESP, N_{echo}, N_{species}, \theta_N}_{k}, \underbrace{T1, T2, \Delta B_0}_{\theta}\right)$$

3.1. **K-T Space, Single Echo, Single Species.** Under the simplifications, $c_l(r) = N_{echo} = N_{species} = 1$, the signal model reduces to the Fourier transform of a product of functions or the convolution of Fourier transform of each function.

$$s(\vec{k}) = \mathcal{F}\left\{\underbrace{Ce^{-\frac{\text{TE}}{T2^*}}e^{-i\ \text{TE}\ \Delta B_0(r)}}_{\text{complex valued signal: magnetization + phase}}\right\} = \mathcal{F}\{C\}\star\mathcal{F}\left\{e^{-\frac{\text{TE}}{T2^*}}\right\}\star\mathcal{F}\left\{e^{-i\ \text{TE}\ \Delta B_0(r)}\right\} \qquad C = \frac{M_0\sin(\gamma\theta_N)\left(1-e^{-TR/T1}\right)}{\left(1-\cos(\gamma\theta_N)\,e^{-TR/T1}\right)}e^{-}$$

In K-T space, we want to find acquistion parameters which provide the highest information content with respect to the rate of change of the signal for changing temperature. For example,

$$\frac{d}{dt}s(\vec{k}) = \mathcal{F}\{C\}\star\mathcal{F}\left\{e^{-\frac{\text{TE}}{T2^*}}\right\}\star\mathcal{F}\left\{e^{-i\ \text{TE}\ \Delta B_0(r)}\right\}\left(\frac{dR2^*}{du}\ \text{TE} + i\ \text{TE}\ \frac{d\Delta B_0(r)}{du}\ \text{TE}\right)\frac{du}{dt} \qquad \text{verify!}$$

Here the time rate of change of temperature is given by the bioheat equation

$$\rho c_p\frac{du}{dt} = k\Delta u + \omega(u-u_a) + q_{\text{source}}$$

3.2. **QOI based sensor model and k-space trajectory.** Consider the k-space trajectory across a readout line that is parametrized by the initial point, $\vec{k}_0$, and the direction, $\vec{s}$. The parameterization is chosen to reduce the search space for the optimization of the MI.

$$\vec{k}[n] = \vec{k}_0 + \vec{s}[n] \qquad n = 0,1,2,...,256$$

A corresponding measurement is taken at each point of the discretized k-space trajectory. For simplicity, consider only the signal magnitude of the off-resonance term.

$$s(\vec{k}[n],\vec{\theta}) = \left|\int_\Omega \left(e^{-i\ \text{TE}\ \Delta B_0(r,\vec{\theta})}\right)e^{-2\pi i\vec{k}[n]\cdot r}\, dr\right|$$

Assuming Gaussian noise (*terrible assumption ?*), the conditional probaility is a multi-variate distribution.

$$p(\vec{z}|\vec{\theta}) = \tilde{C}\exp^{\|\vec{z}-s(\vec{k},\vec{\theta})\|_\Sigma} \qquad s(\vec{k},\vec{\theta}) \equiv \begin{bmatrix} s(\vec{k}[0],\vec{\theta}) \\ s(\vec{k}[1],\vec{\theta}) \\ . \\ . \\ s(\vec{k}[256],\vec{\theta}) \end{bmatrix}$$

The conditional entropy for this model is analytic.

$$H(Z|\Theta) = -\int_\theta d\vec{\theta}p(\vec{\theta})\ \underbrace{H(Z|\Theta=\vec{\theta})}_{\frac{1}{2}\ln((2\ \pi\ e)^n\det\Sigma)} = -\frac{1}{2}\ln\left((2\ \pi\ e)^n\det\Sigma\right)\underbrace{\int_\theta d\vec{\theta}p(\vec{\theta})}_{=1} = \frac{1}{2}\ln\left((2\ \pi\ e)^n\det\Sigma\right)$$

Entropy integrals for the marginalized measurements will be prohibitive in general.

$$H(Z) = -\int_{z_1}\int_{z_2}...\int_{z_{256}} p(\vec{z})\ln p(\vec{z})\, d\vec{z} \qquad p(\vec{z}) = \int_\theta d\vec{\theta}p(\vec{z}|\vec{\theta})p(\vec{\theta})$$

In the spirit of compressed sensing [Fessler, 2010, Donoho et al., 2006, Candes and Wakin, 2008] or goal oriented error estimation [Prudhomme and Oden, 1999], a linear transform, Φ, of the measurments may be considered to reduce the measurement space

$$\Phi : \mathbb{R}^{256} \to \mathbb{R}^m \qquad m << 256$$

or

$$\text{QOI}(s(\vec{k},\vec{\theta})) \in \mathbb{R}$$

Thought exercise: Similar to our 1D example in which an analytic expresion for MI was derived in equation (9), suppose we have a reasonable linearization approximation of the signal model, $\vec{z} \approx L(\vec{k},\vec{\theta})\vec{\theta}$

$$p(\vec{z}|\vec{\theta}) = \tilde{C}\exp^{\|\vec{z}-s(\vec{k},\vec{\theta})\|_\Sigma} \underbrace{\approx}_{\text{jacobian...}} \tilde{C}\exp^{\|\vec{z}-L(\vec{k},\vec{\theta})\vec{\theta}\|_\Sigma} \qquad L:\mathbb{R}^M \to \mathbb{R}^N$$

Using the Convolution of Two Gaussians (Section 6.20 [Tarantola, 2005]) in $\mathbb{R}^n$ and the matrix identities (Section 6.30 [Tarantola, 2005] this is essentially the same posterior for the linear problem in (Section 3.2.2 [Tarantola, 2005])

$$\begin{aligned} p(z) &= \int d\,\vec{\theta}\exp\left(-\frac{1}{2}\left((\vec{\theta}-\vec{\theta}_0)^\top C_\theta^{-1}(\vec{\theta}-\vec{\theta}_0) + (L\vec{\theta}-\vec{z})^\top\Sigma^{-1}(L\vec{\theta}-\vec{z})\right)\right) \\ &= (2\pi)^{n/2}\det(C_\theta^{-1}+L^\top\Sigma^{-1}L)^{-1/2}\exp\left(-\frac{1}{2}\left((L\vec{\theta}_0-\vec{z})^\top(C_\theta^{-1}+L^\top\Sigma^{-1}L)^{-1}(L\vec{\theta}_0-\vec{z})\right)\right) \end{aligned}$$

TODO @madankan: double check algebra pls

Similar to Kalman [Maybeck, 1994]

$$(C_\theta^{-1}+L^\top\Sigma^{-1}L)^{-1} = C_\theta - C_\theta L^\top(\Sigma+LC_\theta L^\top)^{-1}LC_\theta$$

The marginalization is Gaussian and the MI is analytic.

$$I(\theta,z) = H(Z) - H(Z|\Theta) = \frac{1}{2}\ln\left((2\,\pi\,e)\frac{\det\left(C_\theta - C_\theta L^\top(\Sigma+LC_\theta L^\top)^{-1}LC_\theta\right)}{\det\Sigma}\right) = \frac{1}{2}\ln\left((2\,\pi\,e)\underbrace{\frac{\det\left((C_\theta^{-1}+L^\top\Sigma^{-1}L)^{-1}\right)}{\det\Sigma}}_{\text{lower dimension to compute}}\right)$$

For this interpretation the information of the conditional probability, $H(Z|\Theta)$, is essentially fixed. We are trying to find k-space parameters where the model predicts the most uncertainty in the measurements, $H(Z)$. This provides the most opportunity for the measurements to be informed by the model.

This kspace line method should be compared to a single pointwise solution.

4. Paper Outline

Accelerated implementions will reconstruct the $T1$, $T2$, f, etc maps from a sub-sampled k-space. For the dynamic applications, this may either be compressed sensing in space or time. Significant work is involved in choosing an appropriate subsampling trajectory and validating reconstruction of the image from the subsampled trajectory.

1. Initial implementations assume k-space is fully sampled. Data sets may be generated from:
   (a) In-silico using the signal model $s(t)$. *in-silico* verification of a mathematical phantom provides 'code verification' that the algorithm implementation is correct.
   (b) FFT of multi echo data acquired in an agar phantom may be used to simulate raw data signal.
   (c) Direct raw data acquired from scanner

   Compute $f(r)$ pointwise and use FFT! **Assume the following are given:**
   - $N_{echo} = 8$, $N_{species} = 2$
   - temperature change $\Delta T(r, \mu_{\text{eff}}(r))$
   - relaxation parameters, $T1 \sim \mathcal{U}(a,b)$, $T2^* \sim \mathcal{U}(a,b)$
   - coil sensitivity, $c_l(r) = 1$
   - flip angle, $\theta_N = 60^o$, echo time TE=25ms, repetition time TR =500ms
   - gyromagnetic ratio of water $\bar{\gamma} = 42.58$ MHz/T
   - temperature sensitivity, $\alpha = -0.0102 \pm 0.0005\left[\frac{\text{ppm}}{^oC}\right]$ (Note that units of ppm = 1e-6 are **dimensionless**), field strength B_0=1.5T
2. Analyze the PSF with respect to R2*. See temperature analysis of [Odéen et al., 2014].

   $$\Delta T(r) = \delta(r) \qquad \text{@dmitchell412}$$
3. create noise corrupted measurements z

   $$p(z|\theta) = \mathcal{N}(\mathcal{G}_i(k;\theta),\Sigma_z) \qquad z(t) = s(t,\theta)+\eta \qquad \eta \sim \mathcal{N}(0,\Sigma_z)$$
4. Find the k-space locations, $\vec{k}$, that maximizes the information gain in the model parameters $\mu_{\text{eff}}(r), T_1$, and T_2^*, i.e.

   $$\max_K I(k_x^1,k_y^1,\cdots,k_x^N,k_y^N) \tag{12}$$

   where, $K = \{(k_x^1,k_y^1),\cdots,(k_x^N,k_y^N)\}$ is a set of (k_x,k_y) pairs of the N observations $z_i = z(k_x^i,k_y^i)$ $(i = 1,2,\cdots,N)$, and $I(.)$ is the information content that we want to maximize. See (3).

   θ is the augmented parameter vector, composed of all uncertain parameters μ_{eff}, T_1, and T_2^*, etc.

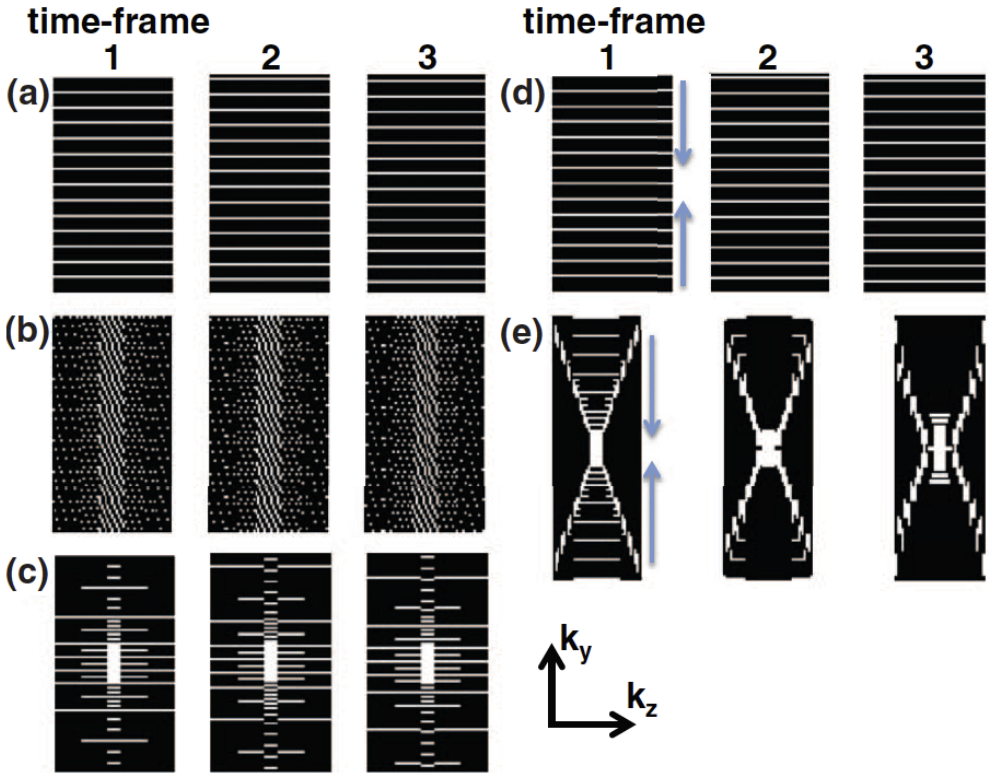

FIG. 1. k_y-k_z plane of k-space for three consecutive time-frames of the (a) ESS, (b) ESS-VD1, (c) NESS-VD2, (d) ESC, and (e) NESC2-VD2 schemes, respectively. White/black areas represent readout lines sampled/not sampled in the current time-frame, respectively. It can be noted that for the ESS and ESC schemes all of k-space have the same sampling density, whereas for the other three schemes the center of k_y-k_z-plane is more heavily sampled to varying degrees. For (d) and (e) arrows indicate centric sampling down the echo train.

FIGURE 4. @dmitchell412 look up references in [Odéen et al., 2014] introduction. How to these k-space locations compare to CURE, BRISK, TRICKS, centric sampling, Odeen [Odéen et al., 2014] trajectories?

- Note 1: Depending on the number of uncertain parameters and numbeer of quadrature points for measurement data z, evaluation of above integral can be computationally expensive. However, one can use parallelization techniques to alleviate this problem.
- Note 2: For evaluation of mutual information, one needs to perform uncertainty quantification for Pennes bioheat equation, given the uncertainty bounds for parameter Θ. In other words, one should first generate corresponding quadrature points for Θ, given $p(\Theta)$, and then Pennes biohear equation needs to be simulated for each of these quadrature realizations. This process is highly parallelizable and can be run through the GPU.
- Note3: Depending on the number of uncertain parameters, uncertainty quantification for Pennes bioheat equation exponentially increases. This *curse of dimensionality* can lead to an undesirable computational cost in presence of large number of uncertain parameters. For instance, given just 9 uncertain parameters, one needs to perform $5^9 \simeq 2 \times 10^6$ simulations for Pennes bioheat equation. However, the good news is that all of these simulations can be parallelized on GPU.
- Note 4: A much simpler alternative, but at the same time accurate option, is to make use of variance. In other word, one can find the points on k-space such that they maximize the variance of model predictions, instead of maximizing the mutual information between potential measurements and model predictions, i.e.

$$\max_K \langle \mathcal{G}_i(k;\theta) \rangle = Var\left(\mathcal{G}_i(k;\theta)\right)$$

The major benefit of using variance instead of mutual information is that its calculation is much simpler comparing to mutual information, and at the same time, it provides most of the information content given by $I(\Theta, z)$.

5. Explore different penalty terms and perform the optimization under L1 minimization [Eltoft et al., 2006, Babacan et al., 2010a] Evaluate the sensitivity of the optimal k-space locations with respect to the tissue parameters. ie change applicator location.

@dmitchell412 formulate this with respect to mixture model, wavelet, symlet, etc basis

6. Validate the reconstruction from the subsampled measurements. Significant validation work is needed to quantitatively assess the recovered model parameters, M_0, $T1$, $T2^*$, ϕ, f against the known parameters of the agar phantom experiment. Previous work includes [Taylor, 2008, Taylor, 2011, Taylor et al., 2012, Taylor, 2009].
7. 'validation' work includes:
   - Additional validation will be needed to compare the quantitative values from the accelerated reconstruction. Parameter reconstructions should be validated in both the fully sampled and subsampled datasets.
   - Validating the number of chemical species detected. This is a model selection problem. Mathematically applying 2-species or 3-species algorithm will find the best numerical fit, but how does this agree with the physical reality of what is being imaged.
   - A dictionary of physical calibration experiments may be created with varying species may be created to verify the recovery on the known physical case when applying the model to each.
     - (10% Species 1, 90% Species 2), (20% Species 1, 80% Species 2), (30% Species 1, 70% Species 2), ...
     - (10% Species 1, 10% Species 2, 80% Species 3), (20% Species 1, 20% Species 2, 60% Species 3), (30% Species 1, 30% Species 2, 40% Species 3), ...
   - A second dictionary of the simulated signal may be created using various concentrations of the species to create an *in-silico* dictionary.
8. Program the pulse sequence

@dmitchell412 for a given optimal trajectory, code the pulse sequence, validate the accuracy, and speedup

9. How do we extend this to multi flip angle signal model (13) ?

Compare and contrast this approach to [Odéen et al., 2014, Wright et al., 2014], Figure 4.

4.1. **Submission options.**

1. Apply existing state of the art algorithms [Hansen and Sø rensen, 2013] to existing data
   - Submit results to applied journal. SPIE, JMI, JCARS
2. Apply new techniques to existing data and compare to existing state of the art
   - Submit results to Med Phys, PMB, Neuroimage

5. Model Based MR Signal Processing

We collect N_{echo} echoes using the MFGRE sequence. Each pixel corresponds to a sequence of N_{echo} complex signal values $w[k]$, $k = 0, \ldots, N_{\text{echo}} - 1$, and want to approximate the sequence by the impulse response of a infinite impulse response (IIR) filter. The MR signal is of the form

$$w[k] = \sum_{j}^{N_{\text{species}}} C_j \Lambda_j^k \qquad k = 0, 1, 2, \ldots, N_{\text{echo}} - 1$$

$$\Lambda_j^k = e^{-\left(i2\pi f_j + \frac{1}{T2_j^*}\right)(\text{TE} + k\ \text{ESP})} \qquad C_j = \frac{M_{0_j} \sin(\gamma\theta_N)\left(1 - e^{-TR/T1_j}\right)}{\left(1 - \cos(\gamma\theta_N)\, e^{-TR/T1_j}\right)} e^{-i\phi_j} \in \mathbb{C} \tag{13}$$

Assume a linear temperature dependence of $T1$

$$T1_j(u) = T1_{0_j} + \psi_j\ \Delta u$$

- **Pulse Sequence Parameters Given:** θ_N, TR, TE, ESP
- **Find:** γ,ϕ_j, $T1_j$, $T2_j^*$, f_j, M_{0_j}

An important observation of Prony is that the signal model satisfies the difference equation

$$w[n] = \beta_1 w[n-1] + \beta_2 w[n-2] + \cdots + \beta_{N_{\text{species}}} w[n - N_{\text{species}}]$$

For any $N_{\text{species}}-$ order differential/difference equation, we would expect $N_{\text{species}}-1$ initial conditions. We will assume $N_{\text{species}}-1$ initial conditions.

$$\begin{aligned}
\alpha_0 &= w[0] \\
\alpha_1 &= w[1] + \beta_1 w[0] \\
\alpha_2 &= w[2] + \beta_1 w[1] + \beta_2 w[0] \\
&\cdot \\
&\cdot \\
\alpha_{N_{\text{species}}-1} &= w[N_{\text{species}}-1] + \beta_1 w[N_{\text{species}}-2] + \cdots + \beta_{N_{\text{species}}-1} w[0]
\end{aligned}$$

An equivalent formulation may interpreted within the context of an infinite impulse response (IIR) classical framework

$$\begin{aligned}
w[n] &= -\sum_{k=1}^{N_{\text{species}}} \beta_k w[n-k] + \sum_{k=0}^{N_{\text{species}}-1} \alpha_k x[n-k], \\
w[n] &= 0 \quad n<0 \qquad\qquad x[n] = \delta[n]
\end{aligned} \tag{14}$$

On a per voxel basis, an MR signal of N_{echo} is acquired and a set of α, β coefficients are processed. A causal relationship between w and x is assumed, $w[n]=0$ and $x[n]=0$ for all $n<0$. see, e.g. [**?**][2].

Linearity and time shift properties of the Z-transform are used to transform the signal model to z-space (14)

- Linearity
$$\mathcal{Z}\left\{a_1 x_1[n] + a_2 x_2[n]\right\} = a_1 X_1(z) + a_2 X_2(z)$$
- Time Shift
$$\mathcal{Z}\left\{x_1[n-k]\right\} = z^{-k} X_1(z)$$
- Exponential
$$\mathcal{Z}\left\{e^{-\alpha n} u[n]\right\} = \frac{1}{1-e^{-\alpha} z^{-1}} \tag{15}$$

The output of such a filter can be expressed in the $z-$ domain by point wise multiplication of an input signal $X(z)$ with the transfer function $N(z)/D(z)$.

$$\mathcal{Z}\left\{\sum_{k=0}^{N_{\text{species}}} \beta_k w[n-k]\right\} = \mathcal{Z}\left\{\sum_{k=0}^{N_{\text{species}}-1} \alpha_k x[n-k]\right\} \quad \Leftrightarrow \quad \underbrace{\left(\sum_{k=0}^{N_{\text{species}}} \beta_k z^{-k}\right)}_{D(z)} W(z) = \underbrace{\left(\sum_{k=0}^{N_{\text{species}}-1} \alpha_k z^{-k}\right)}_{N(z)} X(z) \qquad \beta_0 = 1$$

$$W(z) = X(z)N(z)/D(z). \tag{16}$$

where

$$\begin{aligned}
N(z) &= \alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{N_{\text{species}}-1} z^{-(N_{\text{species}}-1)}, \\
D(z) &= 1 + \beta_1 z^{-1} + \cdots + \beta_{N_{\text{species}}} z^{-N_{\text{species}}}, \\
X(z) &= \sum_{n=-\infty}^{\infty} x[n] z^{-n} \qquad\qquad x[n] = \mathcal{Z}^{-1}\left\{X(z)\right\} \equiv \frac{1}{2\pi j} \oint X(z) z^{k-1} dz. \\
W(z) &= \sum_{n=-\infty}^{\infty} w[n] z^{-n}
\end{aligned}$$

Here, Parsevals theorem is used, and the contour integration is taken over the unit circle for the inverse transform. Note that if $D(z)=1$, i.e. $\beta_k = 0$, then (16) and (14) are convolutions in $z-$ domain and time domain respectively.

Within an ARMAX framework [Candy, 2005], the modeling error in the difference equation $\epsilon[n]$ may be interpreted as the difference of the signal model and the signal at time n.

$$\epsilon[n] = w[n] + \sum_{k=1}^{N_{\text{species}}} \beta_k w[n-k] - \sum_{k=0}^{N_{\text{species}}-1} \alpha_k x[n-k], \tag{17}$$

It is important to realize that the Z-domain provides an analysis framework and a new interpretion of the residual error. Calculations are NOT done in the Z-domain. Calculations will be done in the time domain. A duality exists in that

[2] In MATLAB the function *filter* computes the solution of the difference equation

the equivalent time domain representation of the Z-domain analysis is implemented on the computer. Consider the two representations of the time-domain residual error, $e[n]$ and $\epsilon[n]$ and the corresponding mean square error.

$$\epsilon[n] = \mathcal{Z}^{-1}\{X\ N - W\ D\} \quad \Rightarrow \quad \sum_{n=0}^{N_{\text{echo}}-1} \epsilon[n]^2 = \frac{1}{2\pi j}\oint |X\ N - W\ D|^2\,\frac{dz}{z}. \tag{18}$$

$$e[n] = \mathcal{Z}^{-1}\left\{X\ \frac{N}{D} - W\right\} \quad \Rightarrow \quad \sum_{n=0}^{N_{\text{echo}}-1} e[n]^2 = \frac{1}{2\pi j}\oint \left|X\frac{N}{D} - W\right|^2\,\frac{dz}{z}. \tag{19}$$

Here properties of contour integration on the complex plane simplify the expression for the mean square error. *Need to verify the complex plane integration simplification of the mean square error. Does an direct contour integration in the complex space result in a least square system of equations ??* Minimizing the residual error $\sum \epsilon^2[n]$ in (18) leads to a a linear least square approach to finding coefficients α_i and β_i. This is numerically attractive and straight forward to implement. However, the error ϵ does NOT have a direct physical interpretation and represents the residual error in the difference equation. Alternatively Steiglitz and McBride [**?**] proposed to minimize the modeling error directly such that the coefficients α_i and β_i are found, such that the modeling error $\sum e^2[n]$ (19) is minimized. It was suggested by Steiglitz to solve the problem by a fixed point iteration. During each iteration the objective function

$$\frac{1}{2\pi j}\oint \left|\frac{N_n}{D_{n-1}} - W_n\frac{D_n}{D_{n-1}}\right|^2\frac{dz}{z}, \qquad X(z) = \mathcal{Z}\{\delta[n]\} = 1 \tag{20}$$

is minimized. At the fixed point, where $D_{n-1} = D_n$, (19) and (20) are equivalent. where $\hat{X} = \frac{1}{D_{n-1}}$ and $\hat{W} = \frac{W_n}{D_{n-1}}$. Note, if we assume $D_0 := 1$, then the first iteration of (20) is

$$\min_{\alpha,\beta} \frac{1}{2\pi j}\oint |N_1(z) - W(z)D_1(z)|^2\,\frac{dz}{z}, \tag{21}$$

which is also called equation error and the formulation proposed by Kalman in [**?**].

5.1. **The Prony algorithm.** To derive the algorithm, we drop the subscript 1 from $D_1(z)$ and $N_1(z)$. Assume (21) is only minimized over β. If β is known, then α can be found through $N(z) = W(z)D(z)$, which can be written as a convolution $\alpha = \mathbf{w} * \beta$. In particular, a causal convolution is assumed:

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} = \begin{bmatrix} w_0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & & \\ w_2 & w_1 & w_0 & & \\ \vdots & & & & \vdots \\ w_M & & & & \end{bmatrix} \begin{bmatrix} 1 \\ \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}. \tag{22}$$

Note that from the definition of $N(z)$ it follows that $\alpha_{M+1} = 0, \cdots, \alpha_k = 0$, thus

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} w_{M+1} & w_M & \cdots & \\ w_{M+2} & w_{M+1} & & \\ \vdots & & & \\ w_k & w_{k-1} & \cdots & w_{k-N} \end{bmatrix} \begin{bmatrix} 1 \\ \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}, \tag{23}$$

or

$$\begin{bmatrix} -w_{M+1} \\ -w_{M+2} \\ \vdots \\ -w_k \end{bmatrix} = \begin{bmatrix} w_M & \cdots & \\ w_{M+1} & & \\ \vdots & & \\ w_{k-1} & \cdots & w_{k-N} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{bmatrix},$$

$$\mathbf{h_1} = H_2\mathbf{b} \tag{24}$$

If this system is over determined, a least squares approach is use, i.e. $H_2^T\mathbf{h_1} = H_2^T H_2\mathbf{b}$ is solved.

5.2. **The Steiglitz-McBride algorithm.** Instead of only minimizing (20) w.r.t. β we can minimize $\hat{X}_n(z)N_n(z) - \hat{W}_n(z)D_n(z)$ w.r.t. both α and β. First assume a non-over determined system, which can be solved directly. Similar to the Prony algorithm we write $\mathbf{0} = -\hat{\mathbf{x}} * \alpha + \hat{\mathbf{w}} * \beta$, or in matrix-vector form

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{w}_0 & 0 & \cdots & & -\hat{x}_0 & 0 & \cdots \\ \hat{w}_1 & \hat{w}_0 & & & -\hat{x}_1 & -\hat{x}_0 & \\ \vdots & & & & \vdots & & \\ \hat{w}_k & \hat{w}_{k-1} & \cdots & \hat{w}_{k-N} - \hat{x}_k & -\hat{x}_{k-1} & \cdots & -\hat{x}_{k-N} \end{bmatrix} \begin{bmatrix} 1 \\ \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \\ \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix}. \tag{25}$$

Similarly to the Prony algorithm the first column of the matrix is moved to the left hand side:

$$\begin{bmatrix} -\hat{w}_0 \\ -\hat{w}_1 \\ \vdots \\ -\hat{w}_k \end{bmatrix} = \begin{bmatrix} 0 & \cdots & & -\hat{x}_0 & 0 & \cdots \\ \hat{w}_0 & & & -\hat{x}_1 & -\hat{x}_0 & \\ \vdots & & & \vdots & & \\ \hat{w}_{k-1} & \cdots & \hat{w}_{k-N} - \hat{x}_k & -\hat{x}_{k-1} & \cdots & -\hat{x}_{k-N} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \\ \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix}. \tag{26}$$

If the system is overdetermined, it is solved using the normal equations. $\hat{w}$ and $\hat{x}$ are computed each iteration by using a difference equation similar to (14).

Filtered Signal Input. The filtered signals $\hat{W}$ and $\hat{X}$ appear within each steiglitz iteration. These may be interpreted in terms of the equivalent time domain representation.

$$\hat{Y}(z) = \frac{1}{D(z)}Y(z) \qquad \Leftrightarrow \qquad \hat{Y}(z) = \left.\frac{N(z)}{D(z)}\right|_{N(z)=1} Y(z) \qquad \Leftrightarrow \qquad \hat{y}[n] = -\sum_{k=1}^{N} \beta_k \hat{y}[n-k] + y[n]$$

Data: Echos from MFGRE sequence
Result: Coefficients α and β
Find initial set of coefficients for $D_1(z)$ using (24) or (26);
while *not converged* **do**
  Find $\hat{x}[n] = -\sum_{k=1}^{N} \beta_k \hat{x}[n-k] + \delta[n]$;
  Find $\hat{w}[n] = -\sum_{k=1}^{N} \beta_k \hat{w}[n-k] + w[n]$;
  POTENTIAL BUG- Notice that $w[n]$ and $\delta[n]$ are the ORIGINAL INPUT and DO NOT change with each iteration ;
  Find α_k, β_k from (26)
end

Algorithm 1: Steiglitz McBride

```
a_in = prony(x,0,p);
u_in = zeros(size(x));
u_in(1) = 1;          % make a unit impulse whose length is same as x
a = a_in;
N = length(x);
for i=1:niter
   u = filter( 1, a, x );
   v = filter( 1, a, u_in );
   C1 = convmtx(u(:),p+1);
   C2 = convmtx(v(:),q+1);
   T = [ -C1(1:N,:) C2(1:N,:) ];
   c = T(:,2:p+q+2)\(-T(:,1));   % move 1st column to RHS and do least-squares
   a = [1; c(1:p)];               % denominator coefficients
   b = c(p+1:p+q+1);              % numerator coefficients
end
```

```
a=a.';
b=b.';
```

Example. For the typical MFGRE dataset, each echo is one data point, $w[0], ..., w[15]$, $N_{\text{echo}} = 16$. Assuming a 2-peak model, $N = 2$, the data is modeled as a linear system of equations.

$$\underbrace{\begin{bmatrix}
w[0] = & & & & +\alpha_0 x[0] & \\
w[1] = & -\beta_1 w[0] & & & +\alpha_0 x[1] & +\alpha_1 x[0] \\
w[2] = & -\beta_1 w[1] & - & \beta_2 w[0] & +\alpha_0 x[2] & +\alpha_1 x[1] \\
w[3] = & -\beta_1 w[2] & - & \beta_2 w[1] & +\alpha_0 x[3] & +\alpha_1 x[2] \\
w[4] = & -\beta_1 w[3] & - & \beta_2 w[2] & +\alpha_0 x[4] & +\alpha_1 x[3] \\
w[5] = & -\beta_1 w[4] & - & \beta_2 w[3] & +\alpha_0 x[5] & +\alpha_1 x[4] \\
w[6] = & -\beta_1 w[5] & - & \beta_2 w[4] & +\alpha_0 x[6] & +\alpha_1 x[5] \\
w[7] = & -\beta_1 w[6] & - & \beta_2 w[5] & +\alpha_0 x[7] & +\alpha_1 x[6] \\
w[8] = & -\beta_1 w[7] & - & \beta_2 w[6] & +\alpha_0 x[8] & +\alpha_1 x[7] \\
w[9] = & -\beta_1 w[8] & - & \beta_2 w[7] & +\alpha_0 x[9] & +\alpha_1 x[8] \\
w[10] = & -\beta_1 w[9] & - & \beta_2 w[8] & +\alpha_0 x[10] & +\alpha_1 x[9] \\
w[11] = & -\beta_1 w[10] & - & \beta_2 w[9] & +\alpha_0 x[11] & +\alpha_1 x[10] \\
w[12] = & -\beta_1 w[11] & - & \beta_2 w[10] & +\alpha_0 x[12] & +\alpha_1 x[11] \\
w[13] = & -\beta_1 w[12] & - & \beta_2 w[11] & +\alpha_0 x[13] & +\alpha_1 x[12] \\
w[14] = & -\beta_1 w[13] & - & \beta_2 w[12] & +\alpha_0 x[14] & +\alpha_1 x[13] \\
w[15] = & -\beta_1 w[14] & - & \beta_2 w[13] & +\alpha_0 x[15] & +\alpha_1 x[14]
\end{bmatrix}}_{x[n]=\delta[n]} \quad \Leftrightarrow \quad \begin{bmatrix}
w[0] = & & & & & +\alpha_0\ 1 & \\
w[1] = & -\beta_1 w[0] & & & & & +\alpha_1\ 1 \\
w[2] = & -\beta_1 w[1] & - & \beta_2 w[0] & & & \\
w[3] = & -\beta_1 w[2] & - & \beta_2 w[1] & & & \\
w[4] = & -\beta_1 w[3] & - & \beta_2 w[2] & & & \\
w[5] = & -\beta_1 w[4] & - & \beta_2 w[3] & & & \\
w[6] = & -\beta_1 w[5] & - & \beta_2 w[4] & & & \\
w[7] = & -\beta_1 w[6] & - & \beta_2 w[5] & & & \\
w[8] = & -\beta_1 w[7] & - & \beta_2 w[6] & & & \\
w[9] = & -\beta_1 w[8] & - & \beta_2 w[7] & & & \\
w[10] = & -\beta_1 w[9] & - & \beta_2 w[8] & & & \\
w[11] = & -\beta_1 w[10] & - & \beta_2 w[9] & & & \\
w[12] = & -\beta_1 w[11] & - & \beta_2 w[10] & & & \\
w[13] = & -\beta_1 w[12] & - & \beta_2 w[11] & & & \\
w[14] = & -\beta_1 w[13] & - & \beta_2 w[12] & & & \\
w[15] = & -\beta_1 w[14] & - & \beta_2 w[13] & & &
\end{bmatrix}$$

Find initial set of coefficients for $D_0(z)$ using Prony

$$\begin{bmatrix} w[1] & w[0] \\ w[2] & w[1] \\ w[3] & w[2] \\ w[4] & w[3] \\ w[5] & w[4] \\ w[6] & w[5] \\ w[7] & w[6] \\ w[8] & w[7] \\ w[9] & w[8] \\ w[10] & w[9] \\ w[11] & w[10] \\ w[12] & w[11] \\ w[13] & w[12] \\ w[14] & w[13] \end{bmatrix}^\top \begin{bmatrix} -w[2] \\ -w[3] \\ -w[4] \\ -w[5] \\ -w[6] \\ -w[7] \\ -w[8] \\ -w[9] \\ -w[10] \\ -w[11] \\ -w[12] \\ -w[13] \\ -w[14] \\ -w[15] \end{bmatrix} = \begin{bmatrix} w[1] & w[0] \\ w[2] & w[1] \\ w[3] & w[2] \\ w[4] & w[3] \\ w[5] & w[4] \\ w[6] & w[5] \\ w[7] & w[6] \\ w[8] & w[7] \\ w[9] & w[8] \\ w[10] & w[9] \\ w[11] & w[10] \\ w[12] & w[11] \\ w[13] & w[12] \\ w[14] & w[13] \end{bmatrix}^\top \begin{bmatrix} w[1] & w[0] \\ w[2] & w[1] \\ w[3] & w[2] \\ w[4] & w[3] \\ w[5] & w[4] \\ w[6] & w[5] \\ w[7] & w[6] \\ w[8] & w[7] \\ w[9] & w[8] \\ w[10] & w[9] \\ w[11] & w[10] \\ w[12] & w[11] \\ w[13] & w[12] \\ w[14] & w[13] \end{bmatrix} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\vec{\beta}_0}$$

The steiglitz iteration requires a filter of the form

$$\underbrace{\left(1 + \beta_1 z^{-1} + \beta_2 z^{-2}\right)}_{D(z)} \hat{Y}(z) = Y(z) \quad \Leftrightarrow \quad \hat{y}[n] = -\beta_1 \hat{y}[n-1] - \beta_2 \hat{y}[n-2] + y[n] \quad n = 0, \ldots, 15$$

while *not converged* **do**

Find $\hat{x}[n] = -\beta_1 \hat{x}[n-1] - \beta_2 \hat{x}[n-2] + \delta[n] \qquad n = 0, \dots N_{\text{echo}} - 1$

Find $\hat{w}[n] = -\beta_1 \hat{w}[n-1] - \beta_2 \hat{w}[n-2] + w[n] \qquad n = 0, \dots N_{\text{echo}} - 1$

POTENTIAL BUG- Notice that $w[n]$ and $\delta[n]$ are the ORIGINAL INPUT and DO NOT change with each iteration

$$\begin{bmatrix} 0 & 0 & \hat{x}[0] & 0 \\ -\hat{w}[0] & 0 & \hat{x}[1] & \hat{x}[0] \\ -\hat{w}[1] & -\hat{w}[0] & \hat{x}[2] & \hat{x}[1] \\ -\hat{w}[2] & -\hat{w}[1] & \hat{x}[3] & \hat{x}[2] \\ -\hat{w}[3] & -\hat{w}[2] & \hat{x}[4] & \hat{x}[3] \\ -\hat{w}[4] & -\hat{w}[3] & \hat{x}[5] & \hat{x}[4] \\ -\hat{w}[5] & -\hat{w}[4] & \hat{x}[6] & \hat{x}[5] \\ -\hat{w}[6] & -\hat{w}[5] & \hat{x}[7] & \hat{x}[6] \\ -\hat{w}[7] & -\hat{w}[6] & \hat{x}[8] & \hat{x}[7] \\ -\hat{w}[8] & -\hat{w}[7] & \hat{x}[9] & \hat{x}[8] \\ -\hat{w}[9] & -\hat{w}[8] & \hat{x}[10] & \hat{x}[9] \\ -\hat{w}[10] & -\hat{w}[9] & \hat{x}[11] & \hat{x}[10] \\ -\hat{w}[11] & -\hat{w}[10] & \hat{x}[12] & \hat{x}[11] \\ -\hat{w}[12] & -\hat{w}[11] & \hat{x}[13] & \hat{x}[12] \\ -\hat{w}[13] & -\hat{w}[12] & \hat{x}[14] & \hat{x}[13] \\ -\hat{w}[14] & -\hat{w}[13] & \hat{x}[15] & \hat{x}[14] \end{bmatrix}^{\top} \begin{bmatrix} \hat{w}[0] \\ \hat{w}[1] \\ \hat{w}[2] \\ \hat{w}[3] \\ \hat{w}[4] \\ \hat{w}[5] \\ \hat{w}[6] \\ \hat{w}[7] \\ \hat{w}[8] \\ \hat{w}[9] \\ \hat{w}[10] \\ \hat{w}[11] \\ \hat{w}[12] \\ \hat{w}[13] \\ \hat{w}[14] \\ \hat{w}[15] \end{bmatrix} = \begin{bmatrix} 0 & 0 & \hat{x}[0] & 0 \\ -\hat{w}[0] & 0 & \hat{x}[1] & \hat{x}[0] \\ -\hat{w}[1] & -\hat{w}[0] & \hat{x}[2] & \hat{x}[1] \\ -\hat{w}[2] & -\hat{w}[1] & \hat{x}[3] & \hat{x}[2] \\ -\hat{w}[3] & -\hat{w}[2] & \hat{x}[4] & \hat{x}[3] \\ -\hat{w}[4] & -\hat{w}[3] & \hat{x}[5] & \hat{x}[4] \\ -\hat{w}[5] & -\hat{w}[4] & \hat{x}[6] & \hat{x}[5] \\ -\hat{w}[6] & -\hat{w}[5] & \hat{x}[7] & \hat{x}[6] \\ -\hat{w}[7] & -\hat{w}[6] & \hat{x}[8] & \hat{x}[7] \\ -\hat{w}[8] & -\hat{w}[7] & \hat{x}[9] & \hat{x}[8] \\ -\hat{w}[9] & -\hat{w}[8] & \hat{x}[10] & \hat{x}[9] \\ -\hat{w}[10] & -\hat{w}[9] & \hat{x}[11] & \hat{x}[10] \\ -\hat{w}[11] & -\hat{w}[10] & \hat{x}[12] & \hat{x}[11] \\ -\hat{w}[12] & -\hat{w}[11] & \hat{x}[13] & \hat{x}[12] \\ -\hat{w}[13] & -\hat{w}[12] & \hat{x}[14] & \hat{x}[13] \\ -\hat{w}[14] & -\hat{w}[13] & \hat{x}[15] & \hat{x}[14] \end{bmatrix}^{\top} \begin{bmatrix} 0 & 0 & \hat{x}[0] & 0 \\ -\hat{w}[0] & 0 & \hat{x}[1] & \hat{x}[0] \\ -\hat{w}[1] & -\hat{w}[0] & \hat{x}[2] & \hat{x}[1] \\ -\hat{w}[2] & -\hat{w}[1] & \hat{x}[3] & \hat{x}[2] \\ -\hat{w}[3] & -\hat{w}[2] & \hat{x}[4] & \hat{x}[3] \\ -\hat{w}[4] & -\hat{w}[3] & \hat{x}[5] & \hat{x}[4] \\ -\hat{w}[5] & -\hat{w}[4] & \hat{x}[6] & \hat{x}[5] \\ -\hat{w}[6] & -\hat{w}[5] & \hat{x}[7] & \hat{x}[6] \\ -\hat{w}[7] & -\hat{w}[6] & \hat{x}[8] & \hat{x}[7] \\ -\hat{w}[8] & -\hat{w}[7] & \hat{x}[9] & \hat{x}[8] \\ -\hat{w}[9] & -\hat{w}[8] & \hat{x}[10] & \hat{x}[9] \\ -\hat{w}[10] & -\hat{w}[9] & \hat{x}[11] & \hat{x}[10] \\ -\hat{w}[11] & -\hat{w}[10] & \hat{x}[12] & \hat{x}[11] \\ -\hat{w}[12] & -\hat{w}[11] & \hat{x}[13] & \hat{x}[12] \\ -\hat{w}[13] & -\hat{w}[12] & \hat{x}[14] & \hat{x}[13] \\ -\hat{w}[14] & -\hat{w}[13] & \hat{x}[15] & \hat{x}[14] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \alpha_0 \\ \alpha_1 \end{bmatrix}$$

end

By direct substitution of the signal model

$$w[n] = C_1 \Lambda_1^n + C_2 \Lambda_2^n$$

into the difference equation, we get

$$c\Lambda^n = \beta_1 c\Lambda^{n-1} + \beta_2 c\Lambda^{n-2} \qquad \Rightarrow \qquad \Lambda^2 = \beta_1 \Lambda^1 + \beta_2 \qquad \Rightarrow \qquad (\Lambda_1, \Lambda_2) = \frac{\beta_1 \pm \sqrt{\beta_1^2 + 4\beta_2}}{2}$$

$$f_i = \frac{\Im[\ln \Lambda_i]}{2\ \pi\ \text{ESP}\ \ \gamma\ B_0} \qquad T_2^* = \frac{\text{ESP}}{\Re[\ln \Lambda_i]}$$

The complex amplitudes are found using complex analysis and a direct transformation of the exponential signal model to the z-domain, ie equation (15)

$$W(z) = \frac{N(z)}{D(z)} \underbrace{X(z)}_{=1} \qquad \Leftrightarrow \qquad \underbrace{\frac{C_1}{1 - \Lambda_1 z^{-1}} + \frac{C_2}{1 - \Lambda_2 z^{-1}}}_{W(z)} = \frac{\alpha_0 + \alpha_1\ z^{-1}}{(1 + \beta_1\ z^{-1} + \beta_2\ z^{-2})}$$

Using the residue theorem and the calculation of residues

$$\begin{aligned} \lim_{z \to \Lambda_i} \left(1 - \Lambda_i z^{-1}\right) W(z) =& C_i \\ &= \lim_{z \to \Lambda_i} \left(\frac{\left(1 - \Lambda_i z^{-1}\right) N(z)}{D(z)} \right) \\ &= \underbrace{\lim_{z \to \Lambda_i} \left(\frac{\left(1 - \Lambda_i z^{-1}\right) N'(z) + \Lambda_i z^{-2} N(z)}{D'(z)} \right)}_{\text{L'Hospital Rule, ie } 0/0} = \frac{N(\Lambda_i)}{\Lambda_i \frac{d}{dz} D(\Lambda_i)} \end{aligned}$$

(27)
$$C_1 = \frac{\alpha_0 + \alpha_1\ \Lambda_1}{\Lambda_1\ (\beta_1 + 2\ \beta_2\ \Lambda_1)} \qquad C_2 = \frac{\alpha_0 + \alpha_1\ \Lambda_2}{\Lambda_2\ (\beta_1 + 2\ \beta_2\ \Lambda_2)}$$

5.3. **T1 Calculations.** Applying STMCB processing to our general signal model, Eqn (10), we know that the complex amplitude, $C_j \in \mathbb{C}$, depends on (1) acquisition parameters - repetition time, TR, and flip angle, θ_N, as well as (2) tissue dependent properties - spin lattice relaxation, $T1_j$, proton density M_{0_j}, and initial phase offset, ϕ_j.

$$C_j = \frac{M_{0_j} \sin(\gamma\theta_N)\left(1 - e^{-TR/T1_j}\right)}{\left(1 - \cos(\gamma\theta_N)\, e^{-TR/T1_j}\right)} e^{-i\phi_j} \in \mathbb{C}$$

Multiple methods may be used to extract T1 information from the complex signal amplitude.

- The T1 may be found as the slope of the linearized signal

$$\frac{|C_j|}{\sin(\gamma\theta_N)} = e^{-TR/T1_j} \frac{|C_j|}{\tan(\gamma\theta_N)} + M_{0_j}\left(1 - e^{-TR/T1_j}\right)$$

  Given two flip angles, $\theta_N = \theta_1,\ \theta_2$, a B1 correction map γ, and the TR

$$e^{-TR/T1_j} = \frac{\frac{|C_j(\theta_2)|}{\sin(\gamma\theta_2)} - \frac{|C_j(\theta_1)|}{\sin(\gamma\theta_1)}}{\frac{|C_j(\theta_2)|}{\tan(\gamma\theta_2)} - \frac{|C_j(\theta_1)|}{\tan(\gamma\theta_1)}} \equiv m_j \in \mathbb{R} \quad \Rightarrow \quad T1_j = \frac{-TR}{\ln(m_j)}$$

- A low flip angle may be obtained prior to the scan.

$$\cos(\beta) \approx 1 \qquad \Rightarrow \qquad |C_j(\beta)| = \frac{M_{0_j} \sin(\beta)\left(1 - e^{-TR/T1_j}\right)}{\left(1 - \cos(\beta)\, e^{-TR/T1_j}\right)} \approx M_{0_j} \sin(\beta)$$

  The ratio of this low flip angle value to values acquired under normal condictions may be used to eliminate the M_{0_j} dependence.

$$\rho \equiv \frac{|C_j(\theta)|}{|C_j(\beta)|} = \frac{\sin(\theta)\left(1 - e^{-TR/T1_j}\right)}{\sin(\beta)\left(1 - \cos(\theta)\, e^{-TR/T1_j}\right)}$$

  Given β, θ, and TR, solve the nonlinear equation for T1.
  - Assuming a linear temperature dependence of the T1(u) [Rieke and Butts Pauly, 2008], the ratio of the amplitudes, ρ, is implicitly a temperature dependent measurement corrupted by noise $\nu \sim \mathcal{N}(0, \sigma)$

$$\rho(u) = \frac{\sin(\theta)\left(1 - e^{-TR/T1(u)}\right)}{\sin(\beta)\left(1 - \cos(\theta)\, e^{-TR/T1(u)}\right)} + \nu \qquad T1(u) = T1(u_{\text{ref}}) + m(u - u_{\text{ref}}) \qquad T1(u_{\text{ref}}) \sim \mathcal{U}(T1_{lb}, T1_{ub})$$

  In order to detect a temperature dependent signal change, the variance of the signal ratio must be greater than the noise variance.

$$\frac{\delta\rho(u)}{\delta u}\delta u > \sigma \tag{28}$$

TODO

- validate T1 values for both the double FA method and the ratio method against T1 values obtained from gold standard IR
- Verify temperature induced T1 measurements are viable. ie. when does equation (28) hold ? @madankan, is this condition satified in some sense under uncertainty in the reference T1 value ? $T1(u_{\text{ref}}) \sim \mathcal{U}(T1_{lb}, T1_{ub})$

5.4. **Root finding.** To find the roots of $N(z)$ and $D(z)$ the Companion matrix method can be used. It can be verified by direct computation that the so called Companion matrix

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & -c_{n-1} \end{pmatrix} \tag{29}$$

has the characteristic polynomial $p(t) = c_0 + c_1 t + \cdots + c_{n-1}t^{n-1} + t^n$. Thus the eigenvalues of (29) are the roots of $p(t)$. To find the eigenvalues the QR algorithm can be used. The algorithm performs iterations of the form.

$$A_{k+1} = R_k Q_k,$$

where Q_k is an orthogonal matrix and R_k an upper triangular matrix, such that $A_k = Q_k R_k$, i.e. the QR decomposition of A_k, and $A_0 = A$. It can be shown, that A_k has the same eigenvalues as A and that it converges to a triangular matrix, the *Schur form*. Thus, the eigenvalues of A and be read off the diagonal of A_k after convergence. Usually A is transformed into a upper Hessenberg matrix to reduce the costs of the QR decomposition during each iteration. However in our case, where $p(t)$ is of very small degree direct computation should be sufficient. One method for performing QR decomposition is Gram-Schmidt orthonormalization. This is a process for orthonormalizing a set of vectors in an inner product space, and

the algorithm facilitates QR decomposition when applied to the column vectors of a matrix. If projection is abbrevated such that $\text{proj}_u(v) = \frac{\langle u,v\rangle}{\langle u,u\rangle}u$, then the set of vectors v are transformed to the set of orthogonal vectors u by the following process:

$$\begin{aligned}
u_1 &= v_1 \\
u_2 &= v_2 - \text{proj}_{u_1}(v_2) \\
u_3 &= v_3 - \text{proj}_{u_1}(v_3) - \text{proj}_{u_2}(v_3) \\
&\vdots \\
u_k &= v_k - \textstyle\sum_{j=1}^{k-1} \text{proj}_{u_j}(v_k)
\end{aligned}$$

The orthogonal vectors u are normalized to the set of unit vectors $e_k = \frac{u_k}{\|u_k\|}$. When applying Gram-Schmidt orthonormalization to QR decomposition, the set of column vectors in A, such that $A = [a_1, \ldots, a_n]$, are orthogonalized.

$$\begin{aligned}
u_1 &= a_1 \\
u_2 &= a_2 - \text{proj}_{e_1}(a_2) \\
u_3 &= a_3 - \text{proj}_{e_1}(a_3) - \text{proj}_{e_2}(a_3) \\
&\vdots \\
u_k &= a_k - \textstyle\sum_{j=1}^{k-1} \text{proj}_{e_j}(a_k)
\end{aligned}$$

Once the set of vectors $[a_1, \ldots, a_n] = A = QR$ have been orthonormalized, the orthogonal matrix Q and upper triangular matrix R are reconstructed as follows:

$$Q = [e_1, \ldots, e_n]$$

$$R = \begin{pmatrix} \langle e_1, a_1\rangle & \langle e_1, a_1\rangle & \langle e_1, a_3\rangle & \cdots \\ 0 & \langle e_2, a_2\rangle & \langle e_2, a_3\rangle & \cdots \\ 0 & 0 & \langle e_3, a_3\rangle & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The classical Gram-Schmidt process is numerically unstable. A modified Gram-Schmidt algorithm corrects this instability by also orthogonalizing $u_k^{(i)}$ against rounding errors in $u_k^{(i-1)}$:

$$\begin{aligned}
u_k^{(1)} &= v_k - \text{proj}_{u_1}(v_k) \\
u_k^{(2)} &= u_k^{(1)} - \text{proj}_{u_2}(u_k^{(1)}) \\
&\vdots \\
u_k^{(k-2)} &= u_k^{(k-3)} - \text{proj}_{u_2}(u_k^{(k-3)}) \\
u_k^{(k-1)} &= u_k^{(k-2)} - \text{proj}_{u_2}(u_k^{(k-2)})
\end{aligned}$$

5.5. **Group Sparsity for Chemical Species.** Following [Deng et al., 2013], Consider a two species model in which we wish to enforce group total variation sparsity.

$$\min_{\alpha_0,\alpha_1,\beta_1,\beta_2} \int \sqrt{|\nabla\alpha_0|^2 + |\nabla\alpha_1|^2 + |\nabla\beta_1|^2 + |\nabla\beta_1|^2} \qquad |\nabla f| \equiv \sqrt{f_x^2 + f_y^2 + f_z^2}$$
$$\text{s.t.} \qquad W(z,x) = \frac{N(\alpha_0(x), \alpha_1(x), z)}{D(\beta_1(x), \beta_2(x), z)} X(z) \qquad \forall x$$

The equivalent optimization problem and Augmented lagrangian is of the form

$$\min_{z,y} \|z\|_{2,1} = \sum_{i=1}^{N} \|z_{g_i}\|_2 \qquad \Leftrightarrow \qquad \min_{z,y} \quad \|z\|_{2,1} - (\lambda_1, z - \Phi y) + \frac{\gamma_1}{2}\|z - \Phi y\|^2 - (\lambda_2, Ay - b) + \frac{\gamma_2}{2}\|Ay - b\|^2$$
$$z = \Phi y \qquad Ay = b$$

Where the auxillary variable, z, is define define through the total variation terms.

$$z = \begin{bmatrix} \Theta\vec{g}_1 \\ \Theta\vec{g}_2 \\ \cdot \\ \cdot \\ \cdot \\ \Theta\vec{g}_N \end{bmatrix} \equiv \Phi y \qquad \Theta\vec{g}_i \equiv \begin{bmatrix} \frac{\partial}{\partial x}\alpha_0(x_i) \\ \frac{\partial}{\partial y}\alpha_0(x_i) \\ \frac{\partial}{\partial z}\alpha_0(x_i) \\ \frac{\partial}{\partial x}\alpha_1(x_i) \\ \frac{\partial}{\partial y}\alpha_1(x_i) \\ \frac{\partial}{\partial z}\alpha_1(x_i) \\ \frac{\partial}{\partial x}\beta_1(x_i) \\ \frac{\partial}{\partial y}\beta_1(x_i) \\ \frac{\partial}{\partial z}\beta_1(x_i) \\ \frac{\partial}{\partial x}\beta_2(x_i) \\ \frac{\partial}{\partial y}\beta_2(x_i) \\ \frac{\partial}{\partial z}\beta_2(x_i) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & 0 & 0 & 0 \\ \frac{\partial}{\partial z} & 0 & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 & 0 \\ 0 & \frac{\partial}{\partial z} & 0 & 0 \\ 0 & 0 & \frac{\partial}{\partial x} & 0 \\ 0 & 0 & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} & 0 \\ 0 & 0 & 0 & \frac{\partial}{\partial x} \\ 0 & 0 & 0 & \frac{\partial}{\partial y} \\ 0 & 0 & 0 & \frac{\partial}{\partial z} \end{bmatrix}}_{\Theta} \begin{bmatrix} \alpha_0(x_i) \\ \alpha_1(x_i) \\ \beta_1(x_i) \\ \beta_2(x_i) \end{bmatrix} \qquad \|z_{g_i}\|_2 \equiv \sqrt{\begin{array}{l} (L_x\alpha_0(x_i))^2 + (L_y\alpha_0(x_i))^2 + (L_z\alpha_0(x_i))^2 \\ + (L_x\alpha_1(x_i))^2 + (L_y\alpha_1(x_i))^2 + (L_z\alpha_1(x_i))^2 \\ + (L_x\beta_1(x_i))^2 + (L_y\beta_1(x_i))^2 + (L_z\beta_1(x_i))^2 \\ + (L_x\beta_2(x_i))^2 + (L_y\beta_2(x_i))^2 + (L_z\beta_2(x_i))^2 \end{array}}$$

$$y \equiv \begin{bmatrix} \vec{g}_1 \\ \vec{g}_2 \\ \cdot \\ \cdot \\ \cdot \\ \vec{g}_N \end{bmatrix} \qquad \vec{g}_i = \begin{bmatrix} \alpha_0(x_i) \\ \alpha_1(x_i) \\ \beta_1(x_i) \\ \beta_2(x_i) \end{bmatrix}$$

The y-subproblem is

$$\min_y \quad (\lambda_1, \Phi y) + \frac{\gamma_1}{2}\|z - \Phi y\|^2 - (\lambda_2, Ay) + \frac{\gamma_2}{2}\|Ay - b\|^2 \quad \Leftrightarrow \quad \left(\gamma_1\Phi^T\Phi + \gamma_2 A^T A\right) y = \gamma_1\, z - \lambda_1 + \gamma_2 A^T b + A^T\lambda_2$$

Spatial derivatives in Θ eliminate the embarassingly parallel data fit and require communication from the finite different 5-point or 7-point stencil. perhaps solve Laplacian as convolution? ie FFT for heat transfer.

$$\left(\gamma_1\Theta^T\Theta + \gamma_2 A^T A\right)\vec{g}_i = \gamma_1\, z_{g_i} - (\lambda_1)_{g_i} + \gamma_2 A^T b_{g_i} + A^T(\lambda_2)_{g_i} \qquad i = 1, ..., N \qquad \Theta^T\Theta\vec{g}_i = \begin{bmatrix} \Delta\alpha_0(x_i) \\ \Delta\alpha_1(x_i) \\ \Delta\beta_1(x_i) \\ \Delta\beta_2(x_i) \end{bmatrix}$$

The z-subproblem is derived from the properties of the norm $\| \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \end{bmatrix} \|_2^2 = \|\vec{a_1}\|_2^2 + \|\vec{a_2}\|_2^2$

$$\min_z \quad \|z\|_{2,1} - (\lambda_1, z) + \frac{\gamma_1}{2}\|z - \Phi y\|^2 \underbrace{\qquad \Leftrightarrow \qquad}_{\text{defn of } \|\cdot\|_{2,1}\ \|\cdot\|_2} \min_z \sum_{i=1}^{N} \left\{ \|z_{g_i}\|_2 + \frac{\gamma_1}{2}\|z_{g_i} - \Theta\vec{g}_i - \frac{1}{\gamma_1}(\lambda_1)_{g_i}\|^2 \right\}$$

The motivation for the entire formulation is derived from the analytical solution to this form of the z-subproblem

$$z_{g_i} = \max\left\{ \|r_i\|_2 - \frac{1}{\gamma_1}, 0 \right\} \frac{r_i}{\|r_i\|_2} \qquad i = 1, ..., N \qquad r_i \equiv \Theta\vec{g}_i + \frac{1}{\gamma_1}(\lambda_1)_{g_i}$$

5.6. **Group Sparsity for Chemical Species (Alternative splitting).** Consider a two species model in which we wish to enforce group total variation sparsity.

$$\min_{\alpha_0,\alpha_1,\beta_1,\beta_2} \int \sqrt{|\nabla\alpha_0|^2 + |\nabla\alpha_1|^2 + |\nabla\beta_1|^2 + |\nabla\beta_1|^2}$$
$$\text{s.t.} \qquad W(z,x) = \frac{N(\alpha_0(x), \alpha_1(x), z)}{D(\beta_1(x), \beta_2(x), z)} X(z) \qquad \forall x$$
$$|\nabla f| \equiv \sqrt{f_x^2 + f_y^2 + f_z^2}$$

The equivalent optimization problem and Augmented lagrangian is of the form

$$\min_{p,y} \|\Phi p\|_{2,1} \equiv \sum_{i=1}^{N} \|\Theta p_{g_i}\|_2 \qquad \Leftrightarrow \qquad \min_{p,y} \quad \|\Phi p\|_{2,1} - (\lambda_1, p - y) + \frac{\gamma_1}{2}\|p - y\|^2 - (\lambda_2, Ay - b) + \frac{\gamma_2}{2}\|Ay - b\|^2$$
$$p = y \qquad Ay = b$$

Where the auxillary variable, p, is define define through the total variation terms.

$$p = \begin{bmatrix} \Theta\vec{g}_1 \\ \Theta\vec{g}_2 \\ . \\ . \\ . \\ \Theta\vec{g}_N \end{bmatrix} \equiv \Phi y \qquad \Theta\vec{g}_i \equiv \begin{bmatrix} \frac{\partial}{\partial x}\alpha_0(x_i) \\ \frac{\partial}{\partial y}\alpha_0(x_i) \\ \frac{\partial}{\partial z}\alpha_0(x_i) \\ \frac{\partial}{\partial x}\alpha_1(x_i) \\ \frac{\partial}{\partial y}\alpha_1(x_i) \\ \frac{\partial}{\partial z}\alpha_1(x_i) \\ \frac{\partial}{\partial x}\beta_1(x_i) \\ \frac{\partial}{\partial y}\beta_1(x_i) \\ \frac{\partial}{\partial z}\beta_1(x_i) \\ \frac{\partial}{\partial x}\beta_2(x_i) \\ \frac{\partial}{\partial y}\beta_2(x_i) \\ \frac{\partial}{\partial z}\beta_2(x_i) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & 0 & 0 & 0 \\ \frac{\partial}{\partial z} & 0 & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 & 0 \\ 0 & \frac{\partial}{\partial z} & 0 & 0 \\ 0 & 0 & \frac{\partial}{\partial x} & 0 \\ 0 & 0 & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} & 0 \\ 0 & 0 & 0 & \frac{\partial}{\partial x} \\ 0 & 0 & 0 & \frac{\partial}{\partial y} \\ 0 & 0 & 0 & \frac{\partial}{\partial z} \end{bmatrix}}_{\Theta} \begin{bmatrix} \alpha_0(x_i) \\ \alpha_1(x_i) \\ \beta_1(x_i) \\ \beta_2(x_i) \end{bmatrix}$$

$$\|p_{g_i}\|_2 \equiv \sqrt{\begin{array}{l} (L_x\alpha_0(x_i))^2 + (L_y\alpha_0(x_i))^2 + (L_z\alpha_0(x_i))^2 \\ + (L_x\alpha_1(x_i))^2 + (L_y\alpha_1(x_i))^2 + (L_z\alpha_1(x_i))^2 \\ + (L_x\beta_1(x_i))^2 + (L_y\beta_1(x_i))^2 + (L_z\beta_1(x_i))^2 \\ + (L_x\beta_2(x_i))^2 + (L_y\beta_2(x_i))^2 + (L_z\beta_2(x_i))^2 \end{array}}$$

$$y \equiv \begin{bmatrix} \vec{g}_1 \\ \vec{g}_2 \\ . \\ . \\ . \\ \vec{g}_N \end{bmatrix} \qquad \vec{g}_i = \begin{bmatrix} \alpha_0(x_i) \\ \alpha_1(x_i) \\ \beta_1(x_i) \\ \beta_2(x_i) \end{bmatrix}$$

The y-subproblem is a slightly different form of the original STMBC iteration.

$$\min_y \quad (\lambda_1, y) + \frac{\gamma_1}{2}\|p - y\|^2 - (\lambda_2, Ay) + \frac{\gamma_2}{2}\|Ay - b\|^2 \quad \Leftrightarrow \quad \left(\gamma_1\, I + \gamma_2 A^T A\right) y = \gamma_1\, p - \lambda_1 + \gamma_2 A^T b + A^T\lambda_2$$

The p-subproblem is derived from the properties of the norm $\| \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \end{bmatrix} \|_2^2 = \|\vec{a_1}\|_2^2 + \|\vec{a_2}\|_2^2$

$$\min_p \quad \|\Phi p\|_{2,1} - (\lambda_1, p) + \frac{\gamma_1}{2}\|p - y\|^2 \underbrace{\Leftrightarrow}_{\text{defn of } \|\cdot\|_{2,1}\ \|\cdot\|_2} \min_p \sum_{i=1}^{N} \left\{ \|\Theta p_{g_i}\|_2 + \frac{\gamma_1}{2}\|p_{g_i} - \vec{g}_i - \frac{1}{\gamma_1}(\lambda_1)_{g_i}\|^2 \right\}$$

A second splitting is applied to the p-subproblem.

$$\min_{z,p} \|z\|_{2,1} - (\lambda_1, p) + \frac{\gamma_1}{2}\|p - y\|^2 \quad z = \Phi p \qquad \Leftrightarrow \qquad \min_{z,p} \quad \|z\|_{2,1} - (\lambda_3, z - \Phi p) + \frac{\gamma_3}{2}\|z - \Phi p\|^2 - (\lambda_1, p) + \frac{\gamma_1}{2}\|p - y\|^2$$

The p-p-subproblem becomes a smoothing operation and the p-z-subproblem reduces to the group soft thresholding as before.

$$\min_p \quad (\lambda_3, \Phi p) + \frac{\gamma_3}{2}\|z - \Phi p\|^2 - (\lambda_1, p) + \frac{\gamma_1}{2}\|p - y\|^2 \qquad \min_z \quad \|z\|_{2,1} - (\lambda_3, z) + \frac{\gamma_3}{2}\|z - \Phi p\|^2$$

Spatial derivatives in Θ eliminate the embarassingly parallel data fit and require communication from the finite different 5-point or 7-point stencil. perhaps solve Laplacian as convolution? ie FFT for heat transfer.

$$\left(\gamma_3\Theta^T\Theta + \gamma_1 I\right) p_{\vec{g}_i} = \ldots \qquad i = 1, \ldots, N \qquad \Theta^T\Theta p_{\vec{g}_i} = \begin{bmatrix} \Delta\alpha_0(x_i) \\ \Delta\alpha_1(x_i) \\ \Delta\beta_1(x_i) \\ \Delta\beta_2(x_i) \end{bmatrix}$$

The motivation for the entire formulation is derived from the analytical solution to this form of the p-subproblem

$$z_{g_i} = \max\left\{ \|r_i\|_2 - \frac{1}{\gamma_1}, 0 \right\} \frac{r_i}{\|r_i\|_2} \quad i = 1, \ldots, N \qquad r_i \equiv \Theta p_{\vec{g}_i} + \frac{1}{\gamma_1}(\lambda_1)_{g_i}$$

6. Research Directions

- SPARSE ARTIFACTS: For a given fully sampled fourier dataset of $n = 256x256$ measurements, CS says that only a subset of the data $n = \mathcal{O}(K \log\ m)$ is needed to reconstruction and approximation of the Kth largest coefficients (with error decreasing as $\frac{1}{\sqrt{n}}$) and by taking multiple subsets of the fully sampled data we should be able to find multiple solutions with good accuracy. Are the artifacts sparse ? ie are the typical motion induced artifacts or applicator induced artifacts sparse and able to be detected and removed using the multiple good data approximations ? Similar idea to known-component artifact removal [Stayman et al., 2012]. How can we show that the artifacts are sparse ? Need example data set. Apply transform. subtract. Apply the transform to the residual. Evaluate sparsity with respect to the residual for a given a transform. Can the artifact be localized to the residual ?

– *In Silico* Simulate motion artifacts using motions models from Ch 23 Haacke [Haacke and Brown, 1999]. In this situation the solution x would be the concatenation of the intensities, I and voxel motion, v.

$$\min_x \left\| H \underbrace{\begin{bmatrix} I \\ v \end{bmatrix}}_{x} - z \right\|_{2,\Sigma} + \lambda \left\| \begin{bmatrix} \Phi_I \\ \Phi_v \end{bmatrix} \underbrace{\begin{bmatrix} I \\ v \end{bmatrix}}_{x} \right\|_{1,\Gamma}$$

Individual sparse transforms, Φ_I and Φ_v, would have to be considered for each component. Can Φ_v be used to sparsely represent ghosting/ blurring ? Look up similar work in literature. Potential artifacts of interest include: inter-/intra-scan motion, intravoxel lipid contamination, low SNR, tissue susceptibility changes, magnetic field drift, excessive heating, and modalitydependent applicator induced artifacts.

- Which is the "best basis" to perform the sparse recon ? Define multiple sparsity models.

$$\begin{array}{rcc} \|\Phi_1 x\|_1 & = & \|\nabla x\|_1 + \text{ stmcb} \\ \|\Phi_2 x\|_1 & = & \text{curvelet}(x) + \text{ stmcb} \\ \|\Phi_3 x\|_1 & = & \text{curvelet + motion} \\ \|\Phi_4 x\|_1 & = & ? \end{array} \tag{30}$$

ie take the transform of the exact solution and show that the signal coefficients are sparse. Compare the relative sparsity with another transform. For functional imaging that may be described by a PDE, a spectral decomposition of the PDE operator followed by a low rank approximation seems appropriate. Ie adaptive low rank approximation of the spectrum of the PDE operator. Can this be validated ? In addition to $TV(x) = \|\nabla x\|_1$, is curvelet [Candes and Donoho, 2002, Candes and Donoho, 2004], contourlet [Do and Vetterli, 2003a, Do and Vetterli, 2003b], wedgelet [Donoho et al., 1999], bandlet[Le Pennec and Mallat, 2005b, Le Pennec and Mallat, 2005a], or steerable wavelet [Freeman and Adelson, 1991, Simoncelli et al., 1992] appropriate ?

– Define multiple physics models.

$$\begin{array}{rcc} H_1 & = & \text{radial traj, FFT, regrid} \\ H_2 & = & \text{radial traj, NUFFT} \\ H_3 & = & \text{motion} \\ H_4 & = & ? \end{array} \tag{31}$$

Investigate the applicability of each model with respect to a clinical quantity of interest.

$$Q(x) \text{ for each permutation of } H_i\ \Phi_j$$

- As a baseline control, compute voxel wise maps of the $\alpha_i(x)$ and $\beta_i(x)$ coefficients discussed in [Taylor, 2008]. This processing uses the reconstructed complex valued image data (Cartesian sampling) at N_{echo} timepoints, $I(x, y, t_k)$, $k = 0, ..., N_{echo}$.
- compute the Fast Fourier transform of the data to approximate the raw data, $\tilde{d}(\omega_x, \omega_y, t_k)$, $k = 0, ..., N_{echo}$.

$$\tilde{d}(\omega_x, \omega_y, t_k) = \text{FFT}\left\{I(x, y, t_k)\right\}(\omega_x, \omega_y, t_k) \approx d(\omega_x, \omega_y, t_k)$$

- Subsample the Cartesian trajectory of the raw and synthetic raw datasets using a (1) spiral, (2) stack of spirals in the echo dimension, and (3) radial trajectory. Parametrize on the number of spokes, circles, etc. These will create a new data sets for the 'spiral', 'stack spiral' and 'radial' data sets respectively.

$$d_{\text{spiral}}(\omega_x, \omega_y, t_k) = \text{spiral}\left\{d(\omega_x, \omega_y, t_k)\right\} \qquad d_{\text{radial}}(\omega_x, \omega_y, t_k) = \text{radial}\left\{d(\omega_x, \omega_y, t_k)\right\}$$

$$d_{\text{stackspiral}}(\omega_x, \omega_y, t_k) = \text{stackspiral}\left\{d(\omega_x, \omega_y, t_k)\right\}$$

– Verify your synthetic subsample to the data computed with the nonuniform Fourier Transform (NUFFT). For example,

$$d_{\text{spiral}}(\omega_x, \omega_y, t_k) \quad ? = ? \quad \text{NUFFT}\left\{I(x, y, t_k)\right\}(\omega_x, \omega_y, t_k)$$

- Critically evaluate the performance gains of a Non-cartesian model based reconstruction [Wright et al., 2014] in solving for the α_i, β_i coefficients directly. ie What is the quantitative speed up if we reduce the number of echoes ? what if we reduce the number of spokes or spirals ? what is the accuracy loss ?

$$\min_{\alpha_0, \alpha_1, \beta_1, \beta_2} \int \sqrt{|\nabla\alpha_0|^2 + |\nabla\alpha_1|^2 + |\nabla\beta_1|^2 + |\nabla\beta_1|^2}$$

$$\text{s.t.} \qquad \sum_{k=0}^{N_{echo}-1} \|d_{\text{radial}}(\omega_x, \omega_y, t_k) - \text{NUFFT}\left\{w(\alpha_i, \beta_i, x, y, t_k)\right\}(\omega_x, \omega_y, t_k)\| < \epsilon$$

Here w is the signal model discussed in equation (13)

$$W(z, x) = \frac{N(\alpha_0(x), \alpha_1(x), z)}{D(\beta_1(x), \beta_2(x), z)} X(z) \qquad \forall x \qquad |\nabla f| \equiv \sqrt{f_x^2 + f_y^2 + f_z^2}$$

$$w(x,y,k,t) = \sum_{j}^{N_{\text{species}}} C_j \Lambda_j^k(x,y,t) \qquad \Lambda_j(x,y,t) = \exp\left\{-\left(2\pi i f(x,y,t) + R_2(x,y,t)\right)\text{ESP}\right\}$$

harmonics provide sparsity in echo dimension, still need to decide on sparsity of f and R_2. Need to look at the joint sparse recon

6.1. **Alternative Directions.**

- Build database/dictionary of models AND experiments, ie use phantom data experiments as prior probability distribution functions. Similar to [Ma et al., 2013], create a dictionary for exhaustive optimization from (13). This is a model selection problem! Why is the more expensive ode solver needed ? If you have a M x N dictionary matrix $M << N$, what is the rank of this matrix ? Can the information content of the matrix be related to the rank of the matrix ? Will a dictionary from our simplified signal model (13) be just as good ?
  - one way to test is: (1) given a representative signal, y, (2) decompose the signal into the corresponding basis for each dictionary, ie $y = D_1 x_1 = D_2 x_2$. The smallest 1-norm wins: $\min\{\|x_1\|_1, \|x_2\|_1\}$
  - Sparsity and Rank are not necessarily correlated. For example given a dictionary of the sines an cosines, the dictionary can have a low rank but still contain a basis vector that is an exact representation of a sinusoid data.
  - from [Candes and Tao, 2006], $\mathcal{O}(K\ log\ N)$, measurements are needed to recon a sparse signal f. accuracy compared to the true solution $\hat{f}$ is given by

$$\|f - \hat{f}\|_2 \leq C_{p,\alpha} \cdot R \cdot (K/logN)^{-r}$$

    * Given a typical MR image, $N = 256x256$, we typically sample $N = 256x256$ measurements of the Fourier coefficients to recon. how close does a radial recon approach the theory ? What is the percent increase in accuracy per measument ? ie

$$\frac{\text{accuracy increase}}{\#\text{ of measurements}} \text{ ??}$$

      how does this behave with increasing measurements ?
- Each sequence must be evaluted on a per application basis. How do we quantitatively quantify the information content with respect to a TCA procedure ?
- Hierarchical Gaussian Process modeling of data acquistion
  - GP regridding
  - Fast Gauss transform
- Model selection: what is the probability of the model given the data ?
  - How does the probability of the model given the data when training on LOOCV-style subsets of the data ?
- A probability distribution of the model parameters, T2, T1, freq. is well informed by the data if the probability distribution functions change substantially during calibration [Bauman et al., 2011], ie uniform prior becomes sharply peaked gaussian prior. Using entropy measures, how well is T1 T2 freq distributions informed by the data calibration ?
- Multi-band acquistion strategies ?
- Investigate preconditioners for the acquisition. IE we need a B1 map to quantitatively compute T1 from the multi-parametric sequence.
- given a dictionary from a set of normal and a set of tumor patients, does the trained dictionaries converge to the same basis ?
- Incorporate signal model information, T1, T2, etc into the segmentation data assimilation and random forest pipeline.

References

[Vas, 2008] (2008). Kalman Filtered Compressed Sensing. *Compute*, pages 893–896.

[Aelterman et al., 2011] Aelterman, J., Luong, H. Q., Goossens, B., Pižurica, A., and Philips, W. (2011). Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data. *Signal Processing*, 91(12):2731–2742.

[Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006). K -SVD : An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE trans. signal processing*, 54(11):4311–4322.

[Ahmed and Gokhale, 1989] Ahmed, N. A. and Gokhale, D. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*, 35(3):688–692.

[Aravkin et al., 2011a] Aravkin, A., Burke, V., and Pillonetto, G. (2011a). Robust and Trend-following Student t-Kalman Smoothers. (December):1–14.

[Aravkin et al., 2011b] Aravkin, A. Y., Bell, B. M., Burke, J. V., and Pillonetto, G. (2011b). Robust Kalman Smoother. *Trans Automatic Control*.

[Assländer et al., 2013] Assländer, J., Zahneisen, B., Hugger, T., Reisert, M., Lee, H.-L., LeVan, P., and Hennig, J. (2013). Single shot whole brain imaging using spherical stack of spirals trajectories. *NeuroImage*, 73:59–70.

[Babacan et al., 2010a] Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2010a). Bayesian compressive sensing using laplace priors. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 19(1):53–63.

[Babacan et al., 2010b] Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2010b). FAST BAYESIAN COMPRESSIVE SENSING USING LAPLACE PRIORS Department of Electrical Engineering Departamento de Ciencias and Computer Science de la Computaci on e I . A . pages 2873–2876.

[Baraniuk, 2007] Baraniuk, R. G. (2007). Compressive Sensing. *IEEE signal processing magazine*, (July):118–121.

[Baraniuk et al., 2010] Baraniuk, R. G., Cevher, V., and Duarte, M. F. (2010). Model-Based Compressive Sensing. 56(4):1982–2001.

[Bauman et al., 2011] Bauman, P. T., Jagodzinski, J., and Kirk, B. S. (2011). Statistical calibration of thermocouple gauges used for inferring heat flux. In *42nd AIAA Thermophysics Conference, AIAA*, volume 3779, page 2011.

[Beatty et al., 2005] Beatty, P. J., Nishimura, D. G., and Pauly, J. M. (2005). Rapid gridding reconstruction with a minimal oversampling ratio. *IEEE transactions on medical imaging*, 24(6):799–808.

[Bernstein, 2004] Bernstein (2004). *Hankbook of MRI PulseSequences*.

[Bilgic et al., 2011] Bilgic, B., Goyal, V. K., and Adalsteinsson, E. (2011). Multi-contrast reconstruction with Bayesian compressed sensing. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 000:1–15.

[Candes and Wakin, 2008] Candes and Wakin, M. B. (2008). An Introduction To Compressive Sampling [. *IEEE Signal Processing Magazine*, (March 2008):21–30.

[Candès, 2006] Candès, E. J. (2006). Compressive sampling. *Proc. Inter. Cong. Math.*

[Candes and Donoho, 2002] Candes, E. J. and Donoho, D. L. (2002). Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Annals of statistics*, pages 784–842.

[Candes and Donoho, 2004] Candes, E. J. and Donoho, D. L. (2004). New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. *Communications on pure and applied mathematics*, 57(2):219–266.

[Candès et al., 2011] Candès, E. J., Eldar, Y. C., Needell, D., and Randall, P. (2011). Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73.

[Candès et al., 2006] Candès, E. J., Romberg, J., and Tao, T. (2006). Robust Uncertainty Principles : Exact Signal Frequency Information. 52(2):489–509.

[Candes and Tao, 2006] Candes, E. J. and Tao, T. (2006). Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425.

[Candy, 2005] Candy, J. V. (2005). *Model-based signal processing*, volume 36. John Wiley & Sons.

[Cao et al., 2014] Cao, Z., Oh, S., Otazo, R., Sica, C. T., Griswold, M. a., and Collins, C. M. (2014). Complex difference constrained compressed sensing reconstruction for accelerated PRF thermometry with application to MRI-induced RF heating. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 00:1–12.

[Chan and Park, 2004] Chan, T. and Park, F. (2004). Recent Developments in Total Variation Image. pages 1–18.

[Cover, 2006] Cover, T. (2006). *Elements of Information Theory*.

[Cui et al., 2014] Cui, T., Marzouk, Y., and Willcox, K. (2014). Data Driven Model Reduction for the Bayesian Solution of Inverse Problems. *arxiv*, pages 1–24.

[Deka et al., 2013] Deka, B., Kumar, M., and Baruah, R. (2013). Single-Image Super-Resolution Using Compressive Sensing. 1(4):8–15.

[Deng et al., 2013] Deng, W., Yin, W., and Zhang, Y. (2013). Group sparse optimization by alternating direction method. In *SPIE Optical Engineering+ Applications*, pages 88580R–88580R. International Society for Optics and Photonics.

[Deshmane et al., 2012] Deshmane, A., Gulani, V., Griswold, M. a., and Seiberlich, N. (2012). Parallel MR imaging. *Journal of magnetic resonance imaging : JMRI*, 36(1):55–72.

[Do and Vetterli, 2003a] Do, M. N. and Vetterli, M. (2003a). Contourlets. beyond wavelets, j. stoeckler and gv welland.

[Do and Vetterli, 2003b] Do, M. N. and Vetterli, M. (2003b). Framing pyramids. *Signal Processing, IEEE Transactions on*, 51(9):2329–2342.

[Donoho and Elad, 2003] Donoho, D. L. and Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via l minimization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2197–202.

[Donoho et al., 1999] Donoho, D. L. et al. (1999). Wedgelets: Nearly minimax estimation of edges. *The Annals of Statistics*, 27(3):859–897.

[Donoho et al., 2006] Donoho, D. L., Santos, J. M., and Pauly, J. M. (2006). Compressed Sensing. *ieee trans. information theory*, (March 2008):72–82.

[Donoho and Tsaig, 2006] Donoho, D. L. and Tsaig, Y. (2006). Fast Solution of 1 -norm Minimization Problems When the Solution May be Sparse. pages 1–44.

[Duarte and Baraniuk, 2013] Duarte, M. F. and Baraniuk, R. G. (2013). Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129.

[Durrett, 2010] Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

[Elad and Aharon, 2006] Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 15(12):3736–45.

[Eltoft et al., 2006] Eltoft, T. r., Kim, T., Member, S., and Lee, T.-w. (2006). On the Multivariate Laplace Distribution. 13(5):300–303.

[Farber and Farber, 2011] Farber, R. and Farber (2011). *CUDA Application Design and Development*.

[Feichtinger et al., 1989] Feichtinger, H. G. et al. (1989). Atomic characterizations of modulation spaces through gabor-type representations. *Rocky Mountain J. Math*, 19(1):113–125.

[Fessler, 2002] Fessler, J. A. (2002). Iterative Tomographic Image Reconstruction Using Nonuniform Fast Fourier Transforms.

[Fessler, 2010] Fessler, J. A. (2010). Model Based Imaging Reconstruction for MRI. *IEEE Signal Processing Magazine*, 936726(July):81–89.

[Fessler et al., 2003] Fessler, J. A., Member, S., and Sutton, B. P. (2003). Nonuniform Fast Fourier Transforms Using Min-Max Interpolation. 51(2):560–574.

[Freeman and Adelson, 1991] Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence*, 13(9):891–906.

[Goldstein and Osher, 2009] Goldstein, T. and Osher, S. (2009). The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323.

[Greengard, 1997] Greengard (1997). A short course on fast multipole methods. *New York*.

[Haacke and Brown, 1999] Haacke, M. R. T. and Brown, R. W. (1999). *Magnetic resonance imaging: physical principles and sequence design*. Wiley-Liss.

[Hamza and Krim, 2003] Hamza, A. B. and Krim, H. (2003). Jensen-rényi divergence measure: theoretical and computational perspectives. In *Information Theory, 2003. Proceedings. IEEE International Symposium on*, pages 257–257. IEEE.

[Hansen and Sø rensen, 2013] Hansen, M. S. and Sø rensen, T. S. (2013). Gadgetron: an open source framework for medical image reconstruction. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 69(6):1768–76.

[He et al., 2012] He, C., Liu, L., Xu, L., Liu, M., and Liao, M. (2012). Learning Based Compressed Sensing for SAR Image. 5(4):1272–1281.

[Iqbal and Chen, 2012] Iqbal, M. and Chen, J. (2012). Unification of image fusion and super-resolution using jointly trained dictionaries and local information contents. *IET Image Processing*, 6(9):1299–1310.

[Jackson et al., 1991] Jackson, J. I., Meyer, C. H., Nishimura, D. G., and Macovski, A. (1991). Selection of a Convolution Function for Fourier Inversion Using Gridding. I(3).

[Ji et al., 2009] Ji, S., Dunson, D., and Carin, L. (2009). Multitask Compressive Sensing. 57(1):92–106.

[Jung et al., 2010] Jung, H., Park, J., Yoo, J., and Ye, J. C. (2010). Radial k-t FOCUSS for high-resolution cardiac cine MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 63(1):68–78.

[Kaiman, 1959] Kaiman, R. (1959). Design of self optimization control system. *Transaction of ASME*.

[Kampa et al., 2011] Kampa, K., Hasanbelliu, E., and Principe, J. C. (2011). Closed-form cauchy-schwarz PDF divergence for mixture of Gaussians. *Proceedings of the International Joint Conference on Neural Networks*, 2:2578–2585.

[Keiner et al., 2008] Keiner, J., Kunis, S., and Potts, D. (2008). Using NFFT 3 a software library for various nonequispaced fast Fourier transforms. V:1–23.

[Le Pennec and Mallat, 2005a] Le Pennec, E. and Mallat, S. (2005a). Bandelet image approximation and compression. *Multiscale Modeling & Simulation*, 4(3):992–1039.

[Le Pennec and Mallat, 2005b] Le Pennec, E. and Mallat, S. (2005b). Sparse geometric image representations with bandelets. *Image Processing, IEEE Transactions on*, 14(4):423–438.

[Lingala and Jacob, 2013] Lingala, S. G. and Jacob, M. (2013). Blind compressive sensing dynamic MRI. *IEEE transactions on medical imaging*, 32(6):1132–45.

[Lions, 1997] Lions, P.-L. (1997). Image Recovery via total variation minimization and related problems. *Numerische Mathematik*, pages 167–188.

[Liu, 2009] Liu, Y. (2009). *Fast Multipole Boundary Element Method*.

[Lustig et al., 2007] Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 58(6):1182–95.

[Ma et al., 2013] Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. a. (2013). Magnetic resonance fingerprinting. *Nature*, 495(7440):187–92.

[Madelin et al., 2012] Madelin, G., Chang, G., Otazo, R., Jerschow, A., and Regatte, R. R. (2012). Compressed sensing sodium MRI of cartilage at 7T: preliminary study. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 214(1):360–5.

[Majumdar and Ward, 2011] Majumdar, A. and Ward, R. K. (2011). Joint reconstruction of multiecho MR images using correlated sparsity. *Magnetic resonance imaging*, 29(7):899–906.

[Maleki et al.,] Maleki, A., Narayan, M., and Baraniuk, R. G. Anisotropic Nonlocal Means Denoising. pages 1–60.

[Marcia and Willett,] Marcia, R. F. and Willett, R. M. COMPRESSIVE CODED APERTURE SUPERRESOLUTION IMAGE RECONSTRUCTION Roummel F . Marcia and Rebecca M . Willett Department of Electrical and Computer Engineering.

[Martin and Ghattas, 2012] Martin, J. and Ghattas, O. (2012). A STOCHASTIC NEWTON MCMC METHOD FOR LARGE-SCALE STATISTICAL INVERSE PROBLEMS WITH APPLICATION TO SEISMIC INVERSION. *SIAM J. Sci. Comp.*, 34(3):1460–1487.

[Marzouk and Najm, 2009] Marzouk, Y. M. and Najm, H. N. (2009). Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902.

[Mathelin and Gallivan,] Mathelin, L. and Gallivan, K. A. A compressed sensing approach for partial differential equations with random input data. (November 2010):1–42.

[Maybeck, 1994] Maybeck, P. S. (1994). *Stochastic Models, Estimation, and Control*, volume 141 of *Mathematics in Science and Engineering*.

[Mead, 2008] Mead, J. L. (2008). A priori weighting for parameter estimation. *Journal of Inverse and Ill-posed Problems*, 16(2):175–193.

[Munshi, 2011] Munshi, A. (2011). *OpenCL Programming Guide*. Addison-Wesley.

[Murphy, 2012] Murphy, K. (2012). *Machine Learning A Probabilistic Perspective*.

[Murphy et al., 2012] Murphy, M., Alley, M., Demmel, J., Keutzer, K., Vasanawala, S., and Lustig, M. (2012). Fast l-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *IEEE transactions on medical imaging*, 31(6):1250–62.

[Nam et al., 2013] Nam, S., Akçakaya, M., Basha, T., Stehning, C., Manning, W. J., Tarokh, V., and Nezafat, R. (2013). Compressed sensing reconstruction for whole-heart imaging with 3D radial trajectories: a graphics processing unit implementation. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 69(1):91–102.

[Natarajan, 1995] Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234.

[Odéen et al., 2014] Odéen, H., Todd, N., Diakite, M., Minalga, E., Payne, A., and Parker, D. L. (2014). Sampling strategies for subsampled segmented EPI PRF thermometry in MR guided high intensity focused ultrasound. *Medical Physics*, 41(9):092301.

[Otazo et al., 2010] Otazo, R., Kim, D., Axel, L., and Sodickson, D. K. (2010). Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 64(3):767–76.

[Pan et al., 2013] Pan, Z., Yu, J., Huang, H., Hu, S., Zhang, A., Ma, H., and Sun, W. (2013). Super-Resolution Based on Compressive Sensing and Structural Self-Similarity for Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9):4864–4876.

[Park et al., 2003] Park, S. C., Park, M. K., and Kang, M. G. (2003). Super Resolution Image Reconstruction. *IEEE Signal Processing Magazine*, pages 21–36.

[Pluim et al., 2003] Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. a. (2003). Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004.

[Potts and Tasche,] Potts, D. and Tasche, M. Chapter 1 Fast Fourier transforms for nonequispaced data : A tutorial.

[Prudhomme and Oden, 1999] Prudhomme, S. and Oden, J. T. (1999). On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Computer Methods in Applied Mechanics and Engineering*, 176(1):313–331.

[Pruessmann, 2006] Pruessmann, K. P. (2006). Encoding and reconstruction in parallel MRI. *NMR in biomedicine*, 19(3):288–99.

[Pruessmann et al., 2001] Pruessmann, K. P., Weiger, M., Börnert, P., and Boesiger, P. (2001). Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 46(4):638–51.

[Rasmussen et al., 2006] Rasmussen, C. E., Williams, C. K. I., Processes, G., Press, M. I. T., and Jordan, M. I. (2006). *Gaussian Processes for Machine Learning*.

[Renaut et al., 2010] Renaut, R. a., Hnětynková, I., and Mead, J. (2010). Regularization parameter estimation for large-scale Tikhonov regularization using a priori information. *Computational Statistics & Data Analysis*, 54(12):3430–3445.

[Rieke and Butts Pauly, 2008] Rieke, V. and Butts Pauly, K. (2008). Mr thermometry. *Journal of Magnetic Resonance Imaging*, 27(2):376–390.

[Sankaranarayanan et al., 2013] Sankaranarayanan, A. C., Turaga, P. K., Chellappa, R., and Baraniuk, R. G. (2013). Compressive Acquisition of Linear Dynamical Systems. *SIAM Journal on Imaging Sciences*, 6(4):2109–2133.

[Sarvotham et al.,] Sarvotham, S., Baron, D., and Baraniuk, R. G. Measurements vs . Bits : Compressed Sensing meets Information Theory.

[Schiff, 1991] Schiff (1991). *The Laplace Transform: Theory and Applications*.
[Schomberg and Timmer, 1995] Schomberg, H. and Timmer, J. (1995). The gridding method for image reconstruction by Fourier transformation. *IEEE transactions on medical imaging*, 14(3):596–607.
[Seiberlich et al., 2007] Seiberlich, N., Breuer, F. A., Blaimer, M., Barkauskas, K., Jakob, P. M., and Griswold, M. A. (2007). Non-Cartesian Data Reconstruction Using GRAPPA Operator Gridding (GROG). 1265:1257–1265.
[Sen and Darabi, 2009] Sen, P. and Darabi, S. (2009). Compressive image super-resolution. *Forty-Third Asilomar Conference on Signals, Systems and Computers*, pages 1235–1242.
[Setzer, 2008] Setzer, S. (2008). Split Bregman Algorithm , Douglas-Rachford Splitting and Frame Shrinkage. *Science*, (December):1–13.
[Simoncelli et al., 1992] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2):587–607.
[Sorensen et al., 2009] Sorensen, T. S., Atkinson, D., Schaeffter, T., and Hansen, M. S. (2009). Real-time reconstruction of sensitivity encoded radial magnetic resonance imaging using a graphics processing unit. *IEEE transactions on medical imaging*, 28(12):1974–85.
[Sorensen et al., 2008] Sorensen, T. S., Schaeffter, T., Noe, K. O., and Hansen, M. S. (2008). Accelerating the nonequispaced fast Fourier transform on commodity graphics hardware. *IEEE transactions on medical imaging*, 27(4):538–47.
[Stayman et al., 2012] Stayman, J. W., Otake, Y., Prince, J. L., Khanna, A. J., and Siewerdsen, J. H. (2012). Model-based tomographic reconstruction of objects containing known components. *Medical Imaging, IEEE Transactions on*, 31(10):1837–1848.
[Studer and Baraniuk,] Studer, C. and Baraniuk, R. G. Approximately Sparse Signals.
[Sümbül et al., 2009] Sümbül, U., Santos, J. M., and Pauly, J. M. (2009). A practical acceleration algorithm for real-time imaging. *IEEE transactions on medical imaging*, 28(12):2042–51.
[Tarantola, 2005] Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.
[Taylor, 2008] Taylor, B. (2008). Dynamic chemical shift imaging for image-guided thermal therapy: Analysis of feasibility and potential. *Medical physics*, pages 793–803.
[Taylor, 2009] Taylor, B. (2009). Autoregressive moving average modeling for spectral parameter estimation from a multigradient echo chemical shift acquisition. *Medical physics*, pages 753–764.
[Taylor, 2011] Taylor, B. (2011). Correlation between the temperature dependence of intrinsic MR parameters and thermal dose measured by a rapid chemical shift imaging technique. *NMR in biomedicine*, (September 2010).
[Taylor et al., 2012] Taylor, B. a., Loeffler, R. B., Song, R., McCarville, M. B., Hankins, J. S., and Hillenbrand, C. M. (2012). Simultaneous field and R2 mapping to quantify liver iron content using autoregressive moving average modeling. *Journal of magnetic resonance imaging : JMRI*, 35(5):1125–32.
[Usman et al., 2013] Usman, M., Atkinson, D., Odille, F., Kolbitsch, C., Vaillant, G., Schaeffter, T., Batchelor, P. G., and Prieto, C. (2013). Motion corrected compressed sensing for free-breathing dynamic cardiac MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 70(2):504–16.
[Vasanawala et al., 2011] Vasanawala, S. S., Murphy, M. J., Alley, M. T., Keutzer, P. L. K., and Pauly, J. M. (2011). PRACTICAL PARALLEL IMAGING COMPRESSED SENSING MRI : SUMMARY OF TWO YEARS OF EXPERIENCE IN ACCELERATING BODY MRI OF PEDIATRIC PATIENTS . Electrical Engineering and Computer Science , University of California , Berkeley Radiology , Stanford University. pages 1039–1043.
[Vogel, 1996] Vogel, C. R. (1996). Iterative Methods for total variation denoising*. 17(1):227–238.
[Wang et al., 2007] Wang, Y., Yin, W., and Zhang, Y. (2007). CAAM Technical Report TR07-10 A Fast Algorithm for Image Deblurring with Total Variation Regularization. pages 1–19.
[Wright et al., 2014] Wright, K. L., Hamilton, J. I., Griswold, M. a., Gulani, V., and Seiberlich, N. (2014). Non-Cartesian parallel imaging reconstruction. *Journal of magnetic resonance imaging : JMRI*, 00.
[Yang et al., 2010] Yang, J., Wright, J., Huang, T., and Ma, Y. (2010). Image Super-Resolution via Sparse Representation. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 19(11):1–13.
[Yin et al., 2010] Yin, W., Morgan, S., Yang, J., and Zhang, Y. (2010). Practical Compressive Sensing with Toeplitz and Circulant Matrices. pages 77440K–77440K–10.
[Yin et al., 2008] Yin, W., Osher, S., Goldfarb, D., and Darbon, J. (2008). Bregman Iterative Algorithms for 1 -Minimization with Applications to Compressed Sensing . *Society*, 1(1):143–168.
[Zhang et al., 2012] Zhang, S., Zhan, Y., and Metaxas, D. N. (2012). Deformable segmentation via sparse representation and dictionary learning. *Medical Image Analysis*, 16(7):1385–1396.
[Zhao, 2011] Zhao, C.-h. (2011). An Improved Compressed Sensing Reconstruction Algorithm Based on Artificial Neural Network. *Distribution*, pages 1860–1863.

Appendix A. Alternative Derivations

A.1. **Signal Model Filter.** The augmented signal $\hat{X}$ appears within each steiglitz iteration.

$$\hat{X}(z) = X(z)\frac{1}{D(z)} \qquad \Rightarrow \qquad \hat{x}[n] = x[z] \star \mathcal{Z}^{-1}\left\{\frac{1}{D(z)}\right\}$$

For an N-peak model, partial fractions is needed to expand the rational polynomial about the roots, λ_i.

$$\frac{1}{D(z)} = \frac{1}{1+\beta_1 z^{-1}+\beta_2 z^{-2}+...+\beta_N z^{-N}} = \sum_{i=i}^{N}\frac{a_i}{1+\lambda_i z^{-1}}$$

$$\Rightarrow \qquad \mathcal{Z}^{-1}\left\{\frac{1}{D(z)}\right\}[n] = \sum_{i=i}^{N} a_i \lambda_i^n u[n]$$

Here a_i are coefficients of the expansion.

Example. Given β_i, for a two peak model, partial fraction expansion leads to a 2x2 system of equations in terms of the roots of the polynomial.

$$\vec{\lambda} = \text{roots}([1 \quad \vec{\beta}])$$

$$\frac{1}{1+\beta_1 z^{-1}+\beta_2 z^{-2}} = \frac{a_1}{1+\lambda_1 z^{-1}} + \frac{a_2}{1+\lambda_2 z^{-1}} \Rightarrow \qquad \begin{bmatrix} \lambda_2 & \lambda_1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\hat{x}[n] = x[z] \star (a_1 \lambda_1^n u[n] + a_2 \lambda_2^n u[n]) \qquad \hat{w}[n] = w[z] \star (a_1 \lambda_1^n u[n] + a_2 \lambda_2^n u[n])$$

A.2. **Complex Amplitude Deriviation.** In derivation of the complex amplitudes, (27), Is the complex amplitude calculated using direct substitution of the initial conditions equivalent ?? -or- use least square on all 16 equations ??

$$\begin{array}{l} w[0] = \alpha_0 \\ w[1] = -\beta_1 w[0] + \alpha_1 \end{array} \Leftrightarrow \begin{array}{c} (C_1 + C_2) = \alpha_0 \\ (C_1\Lambda_1 + C_2\Lambda_2) = -\beta_1 (C_1 + C_2) + \alpha_1 \end{array} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ (\Lambda_1 + \beta_1) & (\Lambda_2 + \beta_1) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$$

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \frac{1}{(\Lambda_2 + \beta_1) - (\Lambda_1 + \beta_1)} \begin{bmatrix} (\Lambda_2 + \beta_1) & -1 \\ -(\Lambda_1 + \beta_1) & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$$

A.3. **Spectral Compressed Sensing.** Consider the 10 species signal model

$$w[k] = \sum_{j}^{N_{\text{species}}=10} C_j \Lambda_j^k$$

Will group sparsity on the α, β coefficients distinguish between a 1, 2, 3 peak model ? Ie. do the L1 minimization act as a model selection ?

$$\min_{\alpha,\beta} \sum_{n=0}^{N_{\text{echo}}-1} e[n]^2 + \left\| \begin{matrix} \alpha_0^2 + \beta_1^2 \\ . \\ . \\ . \\ \alpha_9^2 + \beta_{10}^2 \end{matrix} \right\|_1 \qquad \sum_{n=0}^{N_{\text{echo}}-1} e[n]^2 = \frac{1}{2\pi j} \oint \left| X \frac{N}{D} - W \right|^2 \frac{dz}{z}.$$

- Run simulated data w/ 1,2,3 peaks
- Apply L1 min of stmcb and verify recovery

Appendix B. FFT nuFFT

for non integer valued k-space trajectories. define when nuFFT is needed. what datasets is nuFFT appropriate for ? why is integer valued data sufficient for FFT?

Appendix C. Separate Measurement Model and Physics Model WIP

As an alternative framework to Section 2, consider a separation of the physics and measurement models.

In general, we assume that we are given a possibly nonlinear measurement model, $\mathcal{H} : \mathbb{R}^d \to \mathbb{R}^n$ that maps state space variables, $y \in \mathbb{R}^d$, to observables, $z \in \mathbb{R}^n$. To be explicit, will assume that the measurement models are corrupted by zero mean white noise noise of a **known** covariance matrix, $\Sigma_z \mathbb{R}^{n \times n}$

$$z = \mathcal{H}(y) + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma_z)$$

η may be interpreted as the measurement noise or the acquisition noise in the sensor model. For a deterministic measurement model $\mathcal{H}$, the conditional probablity distribution has an explicit analytical form

$$p(z|y) = \mathcal{N}(\mathcal{H}(y), \Sigma_z)$$

A general nonlinear Physics model, $\mathcal{G} : \mathbb{R}^m \to \mathbb{R}^d$ is a mapping from parameter space to state space. Modeling error due to improper assumptions in the model derivations are characterized by ζ. Again to be explicit, we will assume that the physics model is corrupted by zero mean white noise noise of a **known** covariance matrix, $\Sigma_y \mathbb{R}^{n \times n}$

$$y = \mathcal{G}(m) + \zeta \qquad \eta \sim \mathcal{N}(0, \Sigma_y)$$

$$p(y|m) = \mathcal{N}(\mathcal{G}(m), \Sigma_y)$$

The **Goal** is to find the probability distribution of the model parameters given all available measurement data.

$$p(m|z) = ?$$

Within our assumptions, a particular realization of ζ for a given set of model parameters m is assumed to represent the true state y. Within this context, the probability of the measurement given the state may be interepreted as the probability of the measurement given the model parameters and the modeling error

$$p(z|y) = p(z|y(m,\eta)) = \mathcal{N}(\mathcal{H}(y(m,\eta)), \Sigma_z)$$

Invoke Bayes theorem for multiple variables

$$p(m,\eta|z) = \frac{p(z|m,\eta)p(m,\eta)}{p(z)} = \underbrace{\frac{p(z|m,\eta)p(m)p(\eta)}{p(z)}}_{\text{independence}}$$

TODO - verify

$$\begin{aligned} p(m|z) = \int d\eta\, p(m,\eta|z) &= \int d\eta\, p(z|m,\eta)p(\eta) \\ &= \frac{p(m)}{p(z)} \underbrace{\int dy\, p(z|y)p(y|m)}_{\text{change of var } \eta=y-g(m) \quad d\eta=dy} \\ &= \frac{p(m)}{p(z)} \frac{1}{2\,\pi\,\det \Sigma_y + \Sigma_z} \exp\left(\frac{1}{2}\|z-\mathcal{G}(m)\|^2_{\Sigma_y+\Sigma_z}\right) \end{aligned}$$

- **Known**: Σ_y Σ_z
- **Find**:

Difficult to distinguish model error from measurement error formally.

$$\begin{aligned} z &= \mathcal{H}(\mathcal{G}(m)+\zeta)+\eta \\ &\approx g(m)+\epsilon \end{aligned}$$

Appendix D. L1 minimization within the Bayesian framework

Suppose that we have prior belief that the solution to our reconstruction is sparse. We can write this mathematically by stating that our solution, x, is a multivariate Laplacian [Eltoft et al., 2006] distribution characterized by the mean, covariance, and λ, ie $x \sim \text{ML}(\mu, \Gamma, \Lambda)$ with probability distribution function given by:

$$p(x) = \text{... TODO - copy explicity form of pdf from [Eltoft et al., 2006]}$$

e are interested in finding a solution field $x \in \mathbb{R}^n$ for measurement $z \in \mathbb{R}^m$ under noise corruption v

$$z = Hx + v \qquad v \sim \mathcal{N}(0,\Sigma)$$

Here, H is a physics based measurement model. Does the measurment matrix, H, satisfy the RIP (1) or model based RIP property of CS ? This is an NP hard questions [Natarajan, 1995]. Can you show this ? Can this noise corrupted measurement model be shown to satisfy the RIP condition of CS [Donoho et al., 2006, Baraniuk et al., 2010]? Significant success has been demonstrated in recovering a sparse solution from the sparse transform Φ. Here Φ represents a sparsifying transform or dictionary of basis function in which the solution x is sparse. This is essentially a laplacian prior [Eltoft et al., 2006, Babacan et al., 2010a] and may be thought of as a statistical inverse problem for $\Phi x \sim ML(\lambda, \mu, \Gamma)$. Withing this setting the problem formulation of interest is related (TODO - need to derive) to the usual L1 formulation

$$\max_{x,a} \text{Const.}\ \exp\left\{\frac{1}{2}\|H\ Dx - z\|_{2,\Sigma}\right\} \cdot \exp\left\{\lambda\|\Phi x\|_{1,\Gamma}\right\} \quad \Rightarrow \quad \min_x \|Hx-z\|_{2,\Sigma} + \lambda\|\Phi x\|_{1,\Gamma}$$

Alternative philosphies: (1) x is the solution to the statistical inverse problem, hence μ, Γ AND λ define the solution. IE as opposed to traditional solution schemes, λ should be optimized as well. (2) Given λ, the mean, μ, defines the solution for this deterministic model, the λ parameter should be CALIBRATED from previous training data. Within this statistical formulation, λ is a property of the assumed laplacian prior on the state x and the 'best' value must be learned/calibrated from the data and/or the dictionary/transform used to enforce sparsity.

- The data, z, is provided by a Fourier transform coefficients. System noise, Σ, of MR can be characterized/measured.
- Calibrate λ on previous training data.

$$\min_x \|Hx-z\|_{2,\Sigma} + \lambda\|\Phi x\|_{1,\Gamma}$$

The solution $\bar{x}$ is the best approximation of K largest coeffients of Φ_j transform of the true solution

$$\|x-\bar{x}\| \leq C\frac{1}{\sqrt{K}}$$

Investigate how the approximation error varies with the sparsifying transform Φ_j ? Does it agree with theory ? This will be very application dependent. A low rank decomposition of the spectrum of the forward projection operator may be a good sparse transform for instance. How does the physics model, H_i, affect the recovery of the true solution ? Ie model selection, on a gold standard data set.

The goal of the reconstruction is to find the maximum likelihood or maximum probability of the state given the model parameters. Typically we consider the model parameters to be the physics equation represente by the measurement model H, however within this probability setting we must expand our notion of model parameters to contain the parameterizations of the assumed probability distributions functions.

$$\max_x f_{x|z,a}(\xi|\zeta, a) \qquad a = (H, \lambda, D)$$

Given a training data set $\{z_1, z_2, ...\}$ we make also consider calibration of the model parameters a.

$$\max_{x,z} f_{x|z,a}(\xi|\zeta, a) = \frac{f_{x,z|a}(\xi, \zeta|a)}{f_{z|a}(\zeta|a)}$$

To be mathmetically precise we need to explicitly write down the joint Gaussian/Laplacian distribution

$$f_{x,z|a} = \text{ TODO - derive from [Maybeck, 1994, Eltoft et al., 2006]}$$

as well as understand the linear transformation of multi-variate laplacians. See [Eltoft et al., 2006] ie is Y is ML then $Z = AY + b$ is also ML with mean, covariance, lambda defined as

$$\mu = ... \qquad \lambda = ... \qquad \Gamma = ... \text{ TODO - copy explicity form of pdf from [Eltoft et al., 2006]}$$

Depending on the splitting, the final optimization problem could resemble the another form of the L1 min.

$$\max_{x,a} \text{Const. } \exp\left\{\frac{1}{2}\|H\ Dx - z\|_{2,\Sigma}\right\} \cdot \exp\{\lambda\|x\|_{1,\Gamma}\} \quad \Rightarrow \quad \min_{x,a} \frac{1}{2}\|HDx - z\|_{2,\Sigma} + \lambda\|x\|_{1,\Gamma} \qquad a = (H, \lambda, D)$$

ie λ is a property of the assumed laplacian prior on the state x and the 'best' value must be learned/calibrated from the data and/or the dictionary/transform used to enforce sparsity.

A MCMC solution process should be explored. A Karhunen-Loeve decompostion of the solution field of interest has great practical benifit for reducing the solution space [Marzouk and Najm, 2009]. A polynomial chaos (PC) expansion of the physics model also provides a promising surrogate for the stochastic forward problem within the Bayesian inverse setting. However, in [Marzouk and Najm, 2009], the PC solution assumes input parameters are Gaussian and projects from this input Gaussian space to the solution space. it is non Gaussian in the solution space by the PDE mapping. It is not clear how the input probability distribution affects the solution.

Posterior exploration and model reduction may also be pursued simultaneously [Cui et al., 2014]. The *computational efficiency* of the MCMC algorithm is defined by the effective sample size for a given budget of CPU time. There are two ways to improve this: (1) Improve the statistical efficiency. Stochastic Newton [Martin and Ghattas, 2012] use local derivative information to construct efficient proposals that adapt to the local geometry of the posterior distribution. (2) Increase the number of MCMC steps for a given amount of CPU time, ie use GPU or surrogates for more reduced sampling time or forward evaluatiions.

Appendix E. Measurement Operator, H, Forward Projection Operator

Adopting the notation of Fessler [Fessler, 2010], "a typical goal in MR imaging is to form an image of the transverse magnetization of an object, $f(r)$, immediately post excitation". Anticipating a compressed sensing formulation with an $N_{channel}$ receive system, let $c_l(r, \eta)$ denote a random coil sensitivity field at of channel l. Here it is assumed that the number of channels is less than the number number of coils, $N_{channel} < N_{coil}$. Random variable η is used to denote a particular realization of a coil sensitivity pattern, c_l, formed by the random multiplexing of several coils, Figure 5. A typical model of

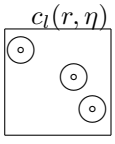

Figure 5. Sensitivity map, $c_l(r, \eta)$, at receive channel l is formed by a random multiplexing from several coils.

the signal received at coil l, *during read-out*, is of the form

$$s_l(t) = \int c_l(r,\eta) f(r) e^{\frac{-t}{T_2^*(r)}} e^{-i\phi(r,t)}\, dr \qquad \phi(r,t) = \int_0^t \left(\gamma B_z(r,\tau) - \omega_0 + \underbrace{\text{velocity term}}_{\text{optional}}\right) d\tau \tag{32}$$

The velocity terms should be obtained from Ch 23 [Haacke and Brown, 1999]. A Galerkin expansion in the image domain is assumed.

$$f = \sum_j f_j \psi_j(r)$$

Here a basis is formed by $\psi_j(r)\ j = 1, .., N$. In matrix notation, the signal model is of the form

$$\begin{bmatrix} s_l(t_1) \\ s_l(t_2) \\ . \\ s_l(t_i) \\ . \\ s_l(t_{N_{RO}}) \end{bmatrix} = \begin{bmatrix} . & . & . \\ . & \int \psi_j(r) c_l(r,\eta) e^{\frac{-t_i}{T_2^*(r)}} e^{-i\phi(r,t_i)}\, dr & . \\ . & . & . \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ . \\ f_j \\ . \\ f_N \end{bmatrix} \equiv A_l f$$

Notice, matrix A_l should not be stored; the matrix-vector multiply operation should be coded instead. Consider using a fast multipole method [Liu, 2009, Greengard, 1997] for the matrix vector product. For a particular phase-encode $\phi(r,t)$, the total measurements, $M = N_{RO} \cdot N_{channel}$, acquired at each channel during read-out is of the form characteristic to the compressing sensing literature [**?**], $y = \Phi\Psi x$. Here the random matrix, Φ, is formed by a particular multiplexing realization of all coils and all channels; $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ is the transformation from a sparse representation, x, to the image domain, f.

$$y = \begin{bmatrix} s_1(t_1) \\ s_1(t_2) \\ . \\ s_1(t_i) \\ . \\ s_1(t_{N_{RO}}) \\ . \\ . \\ s_{N_{channel}}(t_1) \\ s_{N_{channel}}(t_2) \\ . \\ s_{N_{channel}}(t_i) \\ . \\ s_{N_{channel}}(t_{N_{RO}}) \end{bmatrix}_{M\times 1} = \begin{bmatrix} A_1 \\ A_2 \\ . \\ . \\ A_{N_{channel}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ . \\ f_j \\ . \\ f_N \end{bmatrix} = \Phi\Psi x \qquad \Phi \equiv \begin{bmatrix} A_1 \\ A_2 \\ . \\ . \\ A_{N_{channel}} \end{bmatrix}$$