Advanced Applications of Synthetic MR and MAGiC

June 22, 2016

1 Paper Outline Tasks

- @fuentesdt Code Eq. (8)
- @fuentesdt generate MI map of high information content regions, Figure 3
- @kenphwang Need 2 data sets. (1) phantom with known T1/T2 and (2) example data set in human. is this real/imaginary data?
- Recon all data sets with current TE and TD and optimal TE and TD. @kenphwang can we get the current code for this recon?
- @fuentesdt compare and report accuracy in the optimal vs current recon technique

2 Problem Statement

Consider a **complex** magnetization signal $M_{TD} \in \mathbb{C}$ that is defined as as function of **acquisition parameters** $\mathcal{K} = \{\vec{k}, T_R, T_D, \theta, T_E, \alpha\}$ and **tissue properties** $\mathcal{P} \equiv \{T_1, T_2, M_0\}$. Within the scope of this project, T_R=4sec, θ=120^o, and α=90^o are **fixed**. Delay time, T_D, echo time T_E, and $k-$space location, $\vec{k}$, are parameters under consideration for acquisition optimization.

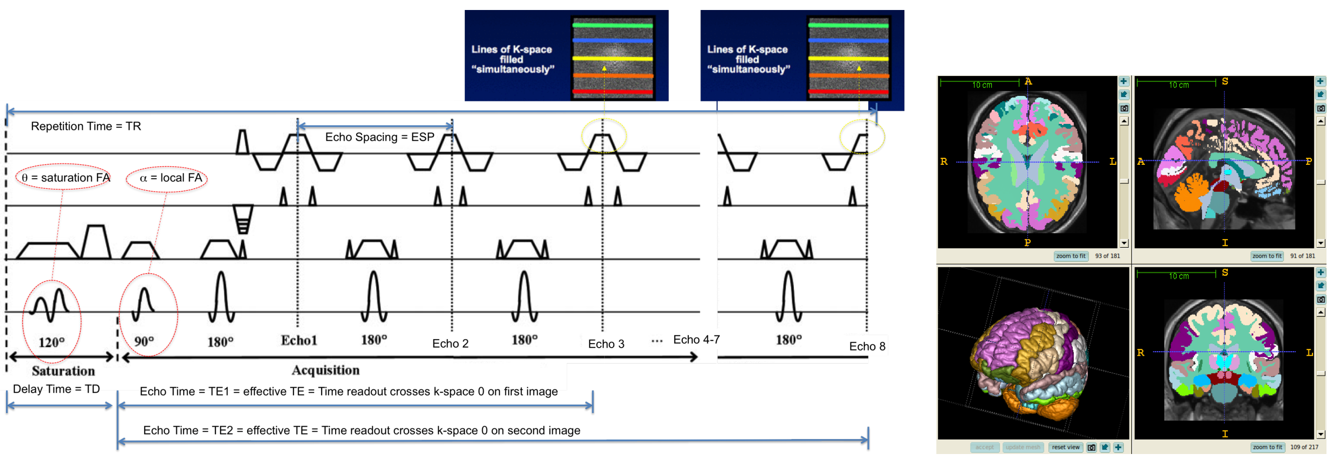


Figure 1: (a) Synthetic MR Pulse sequence. (b) Neuroimaging basis.

$$M_{TD}(T_R, T_D, \theta, T_E, \alpha, \mathcal{P}, \mathbf{x}) = M_0 e^{i\psi(\mathbf{x})} \left((\mathbf{x}) \frac{1 - (1-\cos\theta) e^{-\frac{T_D}{T_1(\mathbf{x})}} - \cos\theta e^{-\frac{T_D}{T_1(\mathbf{x})}}}{1 - \cos\theta e^{-\frac{T_R}{T_1(\mathbf{x})}} cos\alpha} \right) e^{-\frac{T_E}{T_2(\mathbf{x})}} \quad (1)$$

Here, M_0 is the unsaturated magnetization, $\psi(\mathbf{x})$ is the measured phase offset, θ represents the *saturation* flip angle, and T_R and T_E denote repetition time and echo time, respectively. Parameters T_1 and T_2 represents relaxation times, and α is the *local* excitation flip angle. In general, excitation pulse α is a function of flip angle, i.e. $\alpha = \alpha(\theta)$

(@kenphwang why is this?). Note that the unsaturated magnetization M_0, along with relaxation times T_1 and T_2, are a function of spatial coordination $\mathbf{x}$. Basis functions ϕ_i represent the neuroanatomy. For completeness, consider $\phi_1 = \phi_{\text{gm}}$, $\phi_2 = \phi_{\text{wm}}$, $\phi_3 = \phi_{\text{csf}}$, $\phi_4 = \phi_{\text{tumor}}$ as a simplified set of the regions illustrated in Figure 1(b).

$$T_1(\mathbf{x}) = \sum_{i=1}^{N=4} T1_i \phi_i(\mathbf{x}) \qquad T_2(\mathbf{x}) = \sum_{i=1}^{N=4} T2_i \phi_i(\mathbf{x}) \qquad M_0(\mathbf{x}) = \sum_{i=1}^{N=4} M0_i \phi_i(\mathbf{x})$$

$$\bigcup_{i=1}^{N=4} \Omega_i = \Omega \qquad \Omega_n \cap \Omega_m = \varnothing \qquad \phi_i(\mathbf{x}) = \begin{cases} 1 & x \in \Omega_i \\ 0 & \text{otherwise} \end{cases}$$

Under the assumptions:

- phase offset, $\psi(\mathbf{x})$, constant over acquisition time
- TE = effective echo time when readout crosses $k-$space center, i.e. ignore additional T2* decay of echo train
- 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 $k-$space measurements, $\mathcal{G} \in \mathbb{C}$, have the intuitive interpretation as the Fourier coefficients of the complex image weight by the coil sensitivities, $c(r) : \mathbb{R}^3 \to \mathbb{R}$.

$$\begin{aligned} \mathcal{G}(k_x, k_y, k_z, T_R, T_D, \theta, T_E, \alpha, \mathcal{P}) &= \text{Fourier transform} \left\{ \underbrace{c_l(r) M_{TD}(T_R, T_D, \theta, T_E, \alpha, \mathcal{P}, r)}_{\equiv f(r):\mathbb{R}^3 \to \mathbb{C}} \right\} = \mathcal{F}(f(r)) \\ &= \int_\Omega \left(M_{TD}(T_R, T_D, \theta, T_E, \alpha, \mathcal{P}, r) c_l(r) \right) e^{-2\pi i \vec{k} \cdot r}\, dr \end{aligned} \tag{2}$$

Assuming the real and imaginary components of the signal are independent Gaussian noise our signal model may equivalently be understood in $\mathbb{R}^2$ in terms of the real and imaginary components.

$$\begin{bmatrix} z_r(k_x, k_y, k_z, T_R, T_D, \theta, T_E, \alpha, \mathcal{P}) \\ z_i(k_x, k_y, k_z, T_R, T_D, \theta, T_E, \alpha, \mathcal{P}) \end{bmatrix} = \begin{bmatrix} \Re \left\{ \mathcal{G}(k_x, k_y, k_z, T_R, T_D, \theta, T_E, \alpha, \mathcal{P}) \right\} \\ \Im \left\{ \mathcal{G}(k_x, k_y, k_z, T_R, T_D, \theta, T_E, \alpha, \mathcal{P}) \right\} \end{bmatrix} + \begin{bmatrix} \nu_r \\ \nu_i \end{bmatrix}$$

For similicity in notation, vector notation will be assumed.

$$\boxed{z(\mathcal{K}, \mathcal{P}) = \mathcal{G}(\mathcal{K}, \mathcal{P}) + \nu \qquad \nu \sim N(0, \mathbf{R}) \qquad \mathbf{R} = \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_z \end{bmatrix} \qquad \sigma_z \in \mathbb{R}}$$

Note that the observation z is a function of control parameters $\mathcal{K}$ and parameters of interest $\mathcal{P}$. The ultimate goal is to provide accurate estimate of the parameters $\mathcal{P}$, given some measurements z. Precise estimation of parameters $\mathcal{P}$ crucially depends on the values of control parameters $\mathcal{K} = \{\vec{k}, T_D, T_E\}$ ($\{T_R, \theta, \alpha\}$ **fixed**). In other words, to ensure performance of the estimation algorithm, one needs to select the control parameters $\mathcal{K} = \{\vec{k}, T_D, T_E\}$ such that the observation z provides useful information about the parameters $\mathcal{P}$. This is achieved my maximizing the mutual information between the measurements z and parameters of interest $\mathcal{P}$.

2.1 Alternative Image Space Formulation

Assume that the signal model for M_{TD} (1) is our measurment model in **image space** and is polluted with a white noise ν (with mean zero and variance $\mathbf{R}$). Hence, Eq. (1) can be written as:

$$z(\mathcal{K},\mathcal{P}) = \underbrace{M_{TD}(\mathcal{K},\mathcal{P},\mathbf{x})}_{\mathcal{G}(\mathcal{K},\mathcal{P})} + \nu \qquad \nu \in \mathbb{R}^2 \quad \mathbf{R} = \begin{bmatrix} \sigma_r & 0 \\ 0 & \sigma_i \end{bmatrix} \tag{3}$$

3 Mathematical Framework

The underlying philosophy and assumptions within our approach is that the physics models are 1st order accurate or within 70-80% of the needed accuracy and the error is adequate within the assumed Gaussian noise. Gaussian distributions provide analytical representations of the random variables of interest (ie T1, T2) 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 $\Sigma \in \mathbb{R}^{n\times n}$

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

1. Our data acquistion model, $\mathcal{G}(\mathcal{K},\mathcal{P}) : \mathbb{R}^a \times \mathbb{R}^m \to \mathbb{R}^n$, maps deterministic acquisition parameters, $\mathcal{K} \in \mathbb{R}^a$, and uncertain parameters, $\mathcal{P} \in \mathbb{R}^m$ to observables, $\vec{z} \in \mathbb{R}^n$ (or $\vec{z} \in \mathbb{C}^n$). Explicitly, we will assume that the measurement models are corrupted by zero mean white noise noise of a **known** covariance matrix, $\Sigma_z \in \mathbb{R}^{n\times n}$

$$\vec{z} = \mathcal{G}(\mathcal{K};\mathcal{P}) + \nu \qquad \nu \sim \mathcal{N}(0,\Sigma_z) \tag{4}$$

   ν 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.

$$p(z|\mathcal{P}) = \mathcal{N}(\mathcal{G}(\mathcal{K};\mathcal{P}),\Sigma_z) = \frac{1}{2\ \pi\ \det \Sigma_z} \exp\left(-\frac{1}{2}\|\mathcal{G}(\mathcal{K};\mathcal{P}) - z\|^2_{\Sigma_z}\right)$$

2. Additional **known** information is the prior probability distributions for the model parameters, $p(\mathcal{P})$. For simplicity, assume that Prior parameters are Gaussian distributed of **known** mean, $\hat{\mathcal{P}}$ and covariance, $\Sigma_{\mathcal{P}}$

$$\mathcal{P} \sim \mathcal{N}(\hat{\mathcal{P}},\Sigma_{\mathcal{P}}) = \frac{1}{2\ \pi\ \det \Sigma_{\mathcal{P}}} \exp\left(-\frac{1}{2}\|\hat{\mathcal{P}} - \mathcal{P}\|^2_{\Sigma_{\mathcal{P}}}\right)$$

   We will consider the tissue properties to be normally distributed Gaussian parameters.

$$\mathrm{T1}_{\mathrm{WM}} = \mathcal{N}(100ms, 20ms) \qquad \mathrm{T1}_{\mathrm{GM}} = \mathcal{N}(120ms, 20ms) \qquad \mathrm{T1}_{\mathrm{CSF}} = \mathcal{N}(320ms, 20ms) \qquad \mathrm{T1}_{\mathrm{Tumor}} = \mathcal{N}(300ms, 20ms)$$

$$\mathrm{T2}_{\mathrm{WM}} = \mathcal{N}(100ms, 20ms) \qquad \mathrm{T2}_{\mathrm{GM}} = \mathcal{N}(120ms, 20ms) \qquad \mathrm{T2}_{\mathrm{CSF}} = \mathcal{N}(320ms, 20ms) \qquad \mathrm{T2}_{\mathrm{Tumor}} = \mathcal{N}(300ms, 20ms)$$

$$\mathrm{M0}_{\mathrm{WM}} = \mathcal{N}(100??, 20??) \qquad \mathrm{M0}_{\mathrm{GM}} = \mathcal{N}(120??, 20??) \qquad \mathrm{M0}_{\mathrm{CSF}} = \mathcal{N}(320??, 20??) \qquad \mathrm{M0}_{\mathrm{Tumor}} = \mathcal{N}(300??, 20??)$$

   (@kenphwang need exact numbers)

3. Bayes theorem is fundamental to the approach, Appendix A. 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.

$$p(z) = \int_{\mathcal{P}} p(\mathcal{P},z)\, d\mathcal{P} = \underbrace{\int_{\mathcal{P}} p(z|\mathcal{P})\, p(\mathcal{P})\, d\mathcal{P}}_{\int_{\mathcal{P}} = \int_{\mathrm{T1}_{\mathrm{WM}}}\int_{\mathrm{T1}_{\mathrm{GM}}}\int_{\mathrm{T1}_{\mathrm{CSF}}}\int_{\mathrm{T1}_{\mathrm{Tumor}}}\int_{\mathrm{T2}_{\mathrm{WM}}}\int_{\mathrm{T2}_{\mathrm{GM}}}\int_{\mathrm{T2}_{\mathrm{CSF}}}\int_{\mathrm{T2}_{\mathrm{Tumor}}}\int_{\mathrm{M0}_{\mathrm{WM}}}\int_{\mathrm{M0}_{\mathrm{GM}}}\int_{\mathrm{M0}_{\mathrm{CSF}}}\int_{\mathrm{M0}_{\mathrm{Tumor}}}}$$

$$= C\int_{\mathcal{P}} d\mathcal{P} \exp\left(-\frac{1}{2}\left(\|\mathcal{G}(\mathcal{K};\mathcal{P}) - z\|^2_{\Sigma_z} + \|\hat{\mathcal{P}} - \mathcal{P}\|^2_{\Sigma_{\mathcal{P}}}\right)\right) = C\int_{\mathcal{P}} d\mathcal{P} \exp\left(-\frac{1}{2}\left(\|\mathcal{F}M_{TD} - z\|^2_{\Sigma_z} + \|\hat{\mathcal{P}} - \mathcal{P}\|^2_{\Sigma_{\mathcal{P}}}\right)\right)$$

4. The concept of informational entropy [Madankan et al., 2015], $H(Z)$, provides a mathematically rigorous framework to look for measurement acquisition parameters, $\mathcal{K}$, with the high information content of the reconstruction, Figure 2. Given a probability space $(\Omega, \mathcal{F}, p)$ (probability maps from the sigma-algebra of possible events $p : \mathcal{F} \to [0, 1]$ 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.

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

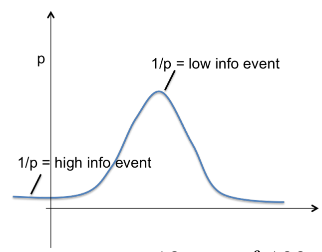
Intuitively, when a low probability event occurs this provides high information. The informational entropy is an *average* of the information content for a sigma algebra of events $\mathcal{F}$

$$H(Z) = \int_Z p(z) \ln \frac{1}{p(z)}\, dz \qquad p(z) = \int_{\mathcal{P}} p(z|\mathcal{P})\, p(\mathcal{P})\, d\mathcal{P}$$

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 ($0 \ln \frac{1}{0} = 0$).

Given a discrete probability function

$$p : \Omega \to \mathbb{R}^+ \qquad \Omega \equiv \left\{ \begin{array}{l} c_i = [a + i\ dx, a + (i+1)dx) \subset [a, b] \\ \qquad\qquad i = 0, ..., N = (b-a)/dx - 1 \end{array} \right\} \qquad \sum_{i=1}^{N} \underbrace{p(c_i)}_{\equiv p_i} = 1$$



Recall $\log \frac{x}{y} = \log x - \log y$, $\log x\ y = \log x + \log y$. The entropy is defined as:

$$H(p) = -\sum_i p_i \log(p_i) = +\sum_i p_i(\underbrace{\log 1}_{0} - \log(p_i)) = \sum_i p_i \log(1/p_i)$$

Here we are using a frequentist interpretation of probability. I.e. $p_0 = 1 \Rightarrow$ the event occurs 10 out of 100 times.

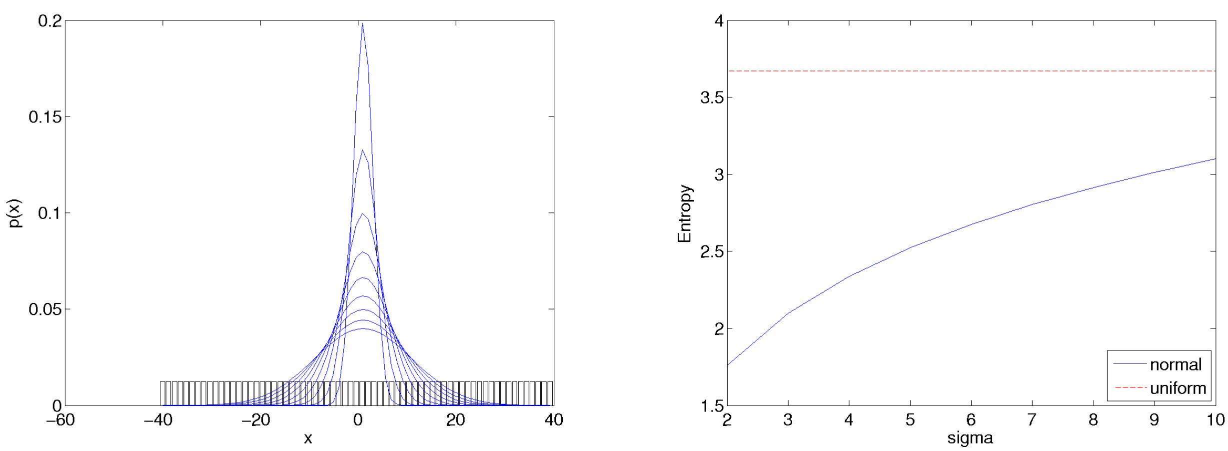


Figure 2: Entropy. Similar to thermodynamics, the entropy is a measure of the spread or uncertainty in a probability distribution. A uniform probability distribution has the largest entropy. This agrees within our intuition. If we have relatively little information about a model parameter/variable, then it can be anywhere uniformly within the interval. The more information we have the more 'peaked' the probability distribution. For example, suppose an object is located with uniform probability, $x \sim \mathcal{U}[a, b]$. The object is located between $[a, b]$ with equal probability and we are uncertain were it may actually be. Compared to $x \sim \mathcal{N}[(a + b)/2, 1]$, we are more certain that the object is likely to be located near the mid-point so we have relatively more information about the location of the object.

5. Performance of estimation process crucially depends on the value of control parameters $\mathcal{K}$. Hence, it is important to develop mathematical tools to identify the control parameters $\mathcal{K}$ such that they provide the best observation data for accurate estimation of parameter $\mathcal{P}$. This is equivalent with maximizing the mutual information between the observation data and parameters $\mathcal{P}$. Based on information theory, mutual information is defined as the reduction of uncertainty in one parameter due to knowledge of the other parameter.

$$I(\mathcal{P};z) = \int_z \int_p p(\mathcal{P},z) \ln\left(\frac{p(\mathcal{P},z)}{p(\mathcal{P})p(z)}\right) d\mathcal{P}dz \tag{5}$$

We make use of Bayes theorem to simplify the above equation. By substituting $p(\mathcal{P},z)$ with $p(z|\mathcal{P})p(\mathcal{P})$, Eq. (5) can be written as:

$$I(\mathcal{P};z) = \int_z \int_p p(z|\mathcal{P})p(\mathcal{P}) \ln\left(\frac{p(z|\mathcal{P})p(\mathcal{P})}{p(\mathcal{P})p(z)}\right) d\mathcal{P}dz \tag{6}$$

Or,

$$I(\mathcal{P};z) = \int_z \int_p p(z|\mathcal{P})p(\mathcal{P}) \ln\left[p(z|\mathcal{P})\right] d\mathcal{P}dz - \int_z p(z)\ln p(z)dz = H(Z) - H(Z|\mathcal{P}) \tag{7}$$

Note that due to dependence of observation data z on control parameters, the mutual information $I(\mathcal{P};z)$ is a function of control parameter $\mathcal{K}$. In order to maximize the reduction of uncertainty in parameter estimate (i.e. to have the most confident estimates of the parameter $\mathcal{P}$), one can simply maximize the mutual information between the observation data and parameters of interest:

$$\max_{\mathcal{K}\in\mathcal{F}} I(\mathcal{P};z) \qquad \mathcal{K} = \{\vec{k}, T_D, T_E\} \in \mathcal{F} = [\vec{k}_{\text{lb}}, \vec{k}_{\text{ub}}] \times (0, 4sec] \times (0, 140ms] \subset \mathbb{Z}^3 \times \mathbb{R}^2 \tag{8}$$

3.1 Mutual Information Approximation

@dmitchell412 need entropy approximations here

By assumption of Gaussian noise, $H(Z|\mathcal{P})$ is constant. Maximizing mutual information reduces to maximizing the entropy in the measurements. Intuitively, we want to find acquisition parameters, $\mathcal{K}$, for which the measurements are most uncertain Probilistic integrals may be computed from uncertainty quantification techniques using kernel density methods (KDM) about quadrature points [Walters-Williams and Li, 2009, Tobin and Houghton, 2013, Terejanu et al., 2012, Fahrenholtz et al., 2013].

$$\max_{\mathcal{K}} I(\mathcal{P};z) \propto \max_{\mathcal{K}} H(Z) \quad \Leftrightarrow \quad \max_{\mathcal{K}} \int_Z dz \underbrace{\int_{\mathcal{P}} d\mathcal{P}\; p(z|\mathcal{P})\; p(\mathcal{P})}_{p(z)} \underbrace{ln\left(\int_{\mathcal{P}} d\mathcal{P}\; p(z|\mathcal{P})\; p(\mathcal{P})\right)}_{ln\; p(z)}$$

$$p(z) = C\int_{\mathcal{P}} d\mathcal{P} \exp\left(-\frac{1}{2}\left(\|\mathcal{F}M_{TD} - z\|^2_{\Sigma_z} + \|\hat{\mathcal{P}} - \mathcal{P}\|^2_{\Sigma_{\mathcal{P}}}\right)\right)$$

To build intuition, Jensen inequality suggests that the expected value may be used to bound the entropy from below.

$$ln\;(E[Z]) \leq E[ln\; p(z)] = H(z)$$

$$E[Z] = \int_Z dz\; p(z) = \int_Z dz\; C\int_{\mathcal{P}} d\mathcal{P} \exp\left(-\frac{1}{2}\left(\|\mathcal{F}M_{TD} - z\|^2_{\Sigma_z} + \|\hat{\mathcal{P}} - \mathcal{P}\|^2_{\Sigma_{\mathcal{P}}}\right)\right)$$

Further, [Madankan et al., 2015] showed that the variance is a reasonable approximation to the entropy. Quadrature rules may be used to build intuition on expect maximal locations.

$$H(z) \propto H(\mathcal{F}M_{TD}) \approx E\left[\mathcal{F}M_{TD} - \mu\right]^2 \approx \sum_{p=1}^{Q} \omega_p \left(\mathcal{F}M_{TD}(\mathcal{P}_p)\right)^2$$

Resulting in

$$\max_{\vec{k},TE,TD} \sum_{p=1}^{Q} \omega_p \left(\int_\Omega M_{TD}(T_D, T_E, \mathcal{P}_p, r) e^{-2\pi i\vec{k}\cdot r}\; dr\right)^2$$

4 Optimal (T_D,T_E) Design

The above maximization results in *optimal* values of control parameter $\mathcal{K}$ for accurate estimation of parameter $\mathcal{P}$. The feasible set, $\mathcal{F}$, is determined by the pulse sequence acquisition physics. Note that in Eq. (7), $p(z|\mathcal{P})$ is defined as a Gaussian distribution with mean $h(\mathcal{K}, \mathcal{P})$ and variance $\mathbf{R}$. As well, $p(\mathcal{P})$ denotes the prior distribution of parameter $\mathcal{P}$, which for the ease of calculations, is considered to be a Gaussian distribution with some prior mean $\hat{\mathcal{P}}^-$ and prior covariance Σ^-, i.e. $p(\mathcal{P}) \sim \mathcal{N}(\hat{p}^-, \Sigma^-)$. Method of quadrature points can be used to evaluate Eq. (7).

We emphasize here that the mutual information will be the same on different pixels with the same tissue types. This is due to the similarities in statistics of $\mathcal{P}$ between the two different pixels with the same tissue properties. In other words, whenever two different pixels have the same tissue properties, then the distribution of parameter $\mathcal{P}$, denoted by $p(\mathcal{P})$, is the same and so is the value of mutual information. Hence, there is no need to evaluate the mutual information for each pixel in a region with the same tissue type. On the other hand, in a case that the tissue properties for each pixel are different from the other, then the mutual information needs to be evaluated for each and every pixel of interest.

The final plot over delay time and echo time is shown in Figure 3. The current acquistion utilizes 8 total combinations of (effective) echo times and delay times, 2 echo times and 4 delay times.

The information content of this acquisition scheme is calculated as the superposition

$$I^{\text{current}} = \sum_{i=1}^{8} I((T_D, T_E)_i) \qquad \text{(@kenphwang need exact numbers)}$$

$(T_D, T_E) = \{$(20ms, 15ms),(40ms, 15ms),(70ms, 15ms),(90ms, 15ms), (20ms, 35ms),(40ms, 35ms),(70ms, 35ms),(90ms, 35ms)$\}$

An optimal acquisition strategy either maintains or improves this information content with less samples to (1) improve time and (2) accuracy.

$$I^{\text{current}} < I^{\text{optimal}} = \sum_{i=1}^{M \leq 8} I((T_D^*, T_E^*)_i)$$

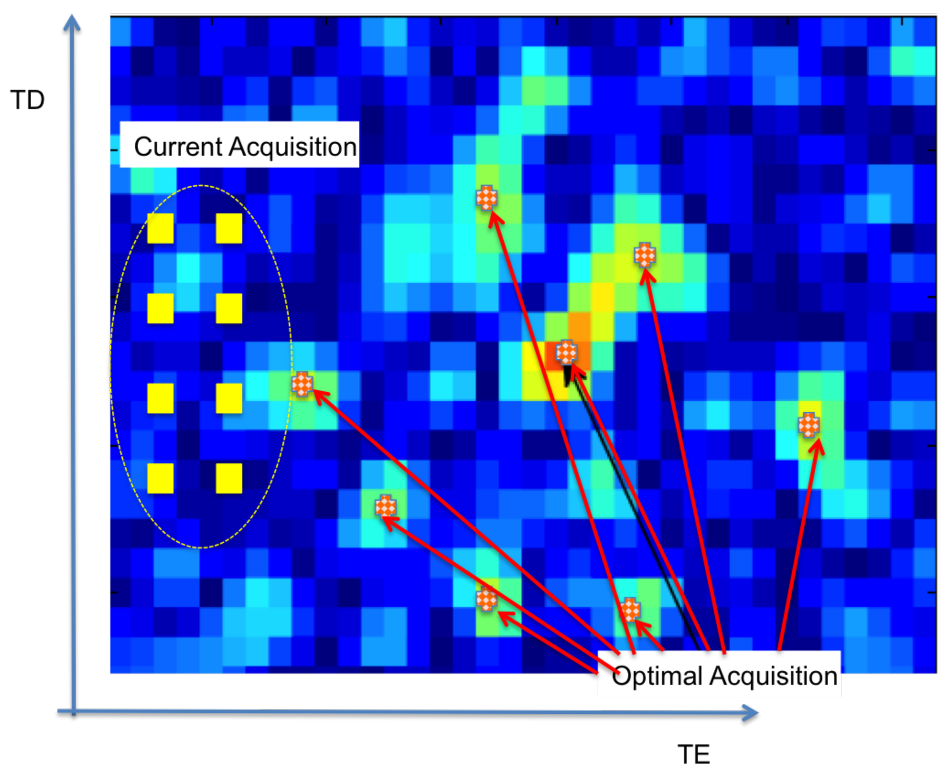


Figure 3: Information content as function of $(T_D, T_E) \in \mathcal{F} = (0, 4sec] \times (0, 140ms] \subset \mathbb{R}^2$. The feasible set, $\mathcal{F}$, is determined by the pulse sequence acquisition physics.

5 Bayesian Learning

The 'score' function from [Seeger et al., 2010] is the mutual information between the desired measurement $\tilde{y}$ and the image u conditioned on existing measurements y.

$$\mathcal{S} = I(U;\tilde{Y}|Y) = H(U|Y) - H(U|\tilde{Y},Y)$$

This has a very intuitive interpretation. The current measure of uncertainty in the parameter recovery (reconstruction) is $H(U|Y)$ for the given data, Y. The measure of uncertainty in the parameter recovery (reconstruction) for a new measurement, $\tilde{Y}$, is $H(U|Y,\tilde{Y})$. The measurement, $\tilde{Y}$, is 'optimal' when is reduction in uncertainty in the parameter recovery (reconstruction) is maximized.

We will assume that the joint probability between (measurement, acquisition) pairs is jointly normal. The joint probability of signal measurements as a function of design parameters is assumed to be indepdent.

$$p(y_1,y_2|u,k_1,k_2) = C\exp\begin{bmatrix} y_1-\mu_1 \\ y_2-\mu_2\end{bmatrix}^\top \begin{bmatrix} P_{11}^{-1} & 0 \\ 0 & P_{22}^{-1}\end{bmatrix}\begin{bmatrix} y_1-\mu_1 \\ y_2-\mu_2\end{bmatrix} = \underbrace{C_1\exp\|y_1(u,k_1)-\mu_1\|_{P_{11}}}_{p(y_1|u,k_1)}\underbrace{C_2\exp\|y_2(u,k_2)-\mu_2\|_{P_{22}}}_{p(y_2|u,k_2)}$$

For simplicity of notation, assume that two measures arise from two distinct acquisition parameters such that the dependence on k is assumed: $p(y_1,y_2|u) = p(y_1|u)p(y_2|u)$. The joint distribution for the measurements is generally correlated through the model.

$$p(y_1,y_2) = \int_U p(y_1,y_2,u)du = \int_U p(y_1,y_2|u)p(u)du = \int_U p(y_1|u)p(y_2|u)p(u)du \qquad p(u|y_i) = \frac{p(u,y_i)}{p(y_i)} = \frac{p(y_i|u)p(u)}{p(y_i)}$$

$$p(y_2|y_1) = \frac{p(y_1,y_2)}{p(y_1)} = \frac{1}{p(y_1)}\int_U p(y_1|u)p(y_2|u)p(u)du = \int_U \frac{p(y_1|u)}{p(y_1)}p(y_2|u)p(u)du = \int_U p(u|y_1)p(y_2|u)du$$

$$p(y_i) = \int_U p(y_i|u)p(u)du$$

Under these assumptions, the model for probability of parameters condition on the joint probability of the data reduces to the probabulity of the parametery recovery (reconstruction) with respect to the data multiplied by the probability of the new measurements given the parametery recovery (reconstruction) [Seeger et al., 2010].

$$p(u|\tilde{y},y) = \frac{p(u,\tilde{y},y)}{p(\tilde{y},y)} = \frac{p(\tilde{y},y|u)p(u)}{p(\tilde{y},y)} = \frac{p(\tilde{y}|u)p(y|u)p(u)}{p(\tilde{y},y)} = \frac{p(\tilde{y}|u)p(y|u)p(u)}{p(\tilde{y},y)} = p(\tilde{y}|u)p(u|y)\frac{p(y)}{p(\tilde{y},y)} \propto p(u|y)p(\tilde{y}|u)$$

The final information gain may be computed as:

$$\begin{aligned}
\mathcal{S} = I(U;\tilde{Y}|Y) &= H(U|Y) - H(U|\tilde{Y},Y) \\
&= -\int_y dy\ p(y)\int_u du\ p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\ \int_y dy\ p(\tilde{y},y)\int_u du\ p(u|\tilde{y},y)\log p(u|\tilde{y},y) \\
&= -\int_y dy\ p(y)\int_u du\ p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\ \int_y dy\ p(\tilde{y}|y)p(y)\int_u du\ p(u|\tilde{y},y)\log p(u|\tilde{y},y) \\
&= \int_y dy\ p(y)\left(-\int_u du\ p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\ p(\tilde{y}|y)\int_u du\ p(u|\tilde{y},y)\log p(u|\tilde{y},y)\right)
\end{aligned}$$

We have two options to compute the information gain $I(U;\tilde{Y}|Y)$ or scoring function $\mathcal{S}$ at this point:

1. Assume that the measurements are completely predicted by the model. ie

$$p(y) = \int_U p(y|u)p(u)du$$

2. Use Gaussian distribution around actual signal measurements y^* and ignore the model predicted measurements

$$p(y) = \mathcal{N}(y^*,\sigma) = \exp\left(\frac{y-y^*}{\sigma}\right) \neq \int_U p(y|u)p(u)du$$

@dmitchell412 how do we interpret this ? how does modeling errors affect this choice ? which is 'better' ? should we further separate into predicted measurement vs actual measurement ? ie Kalman Filter [Maybeck, 1979].

$$z = Hy + \nu \qquad \Rightarrow \qquad I(U;\tilde{Y}|Z)$$

6 Overall Picture

The following diagram illustrates the general work-flow of the process:

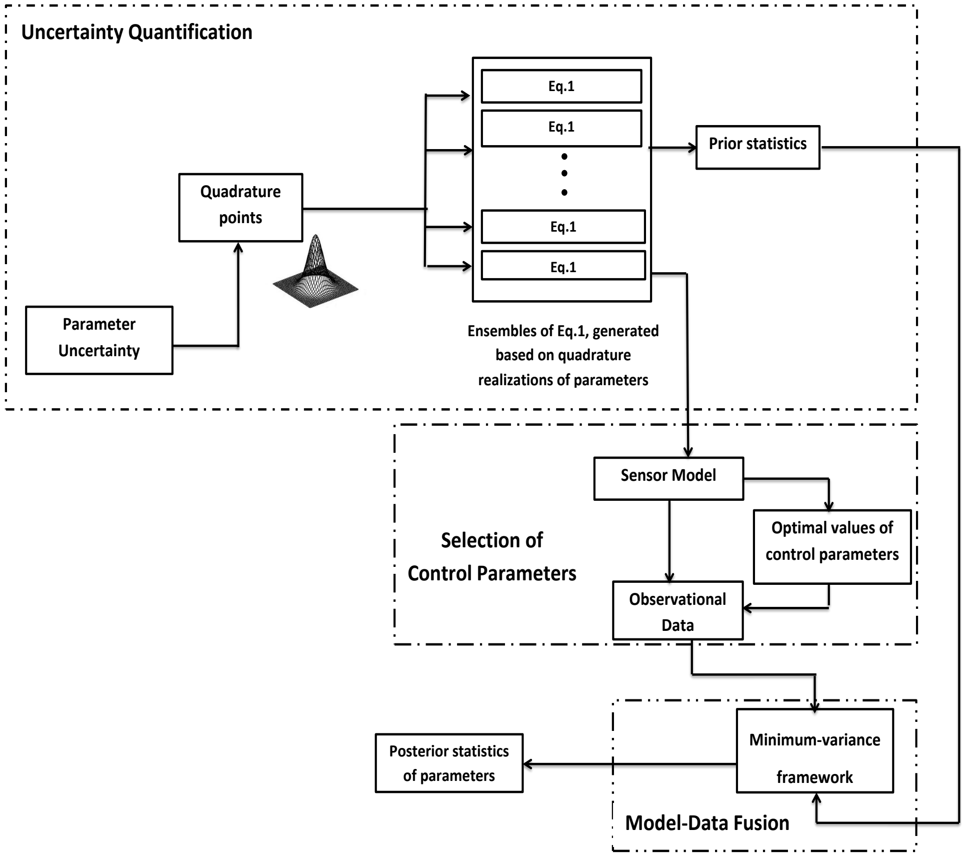


Figure 4: Schematic view of the estimation process

7 WIP - Inverse Problem Framework

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

$$p(z|\theta) = \exp\left(\|\vec{z} - \mathcal{G}(\theta)\|^2_{\Sigma_z}\right)$$

$$d\left(\vec{z}, \mathcal{G}(\theta^*)\right) = \min_{\theta \in \Omega} d\left(\vec{z}, \mathcal{G}(\theta)\right) \qquad \theta = (\mu_{\text{CSF}}, \mu_{\text{GM}}, \mu_{\text{WM}}, \mu_{\text{Tumor}})$$

8 WIP - T1, T2, M0 Reconstruction

Given the image space date for multiple acquistion parameters $\{M_{TD}(\mathcal{K}_1), M_{TD}(\mathcal{K}_2), M_{TD}(\mathcal{K}_2), ...\}$, $\mathcal{K}_i = \{T_{R_i}, T_{D_i}, \theta_i, T_{E_i}, \alpha_i\}$ (@kenphwang 2 delay times and 4 echoes correct?),

The reconstruction algorithms for T1, T2, M0 is as follows:

-
-
-

References

[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.

[Carp et al., 2004] Carp, S. A., Prahl, S. a., and Venugopalan, V. (2004). Radiative transport in the delta-P1 approximation: accuracy of fluence rate and optical penetration depth predictions in turbid semi-infinite media. *Journal of biomedical optics*, 9(3):632–47.

[Cover and Thomas, 2012] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

[Duck, 1990] Duck, F. (1990). Physical properties of tissue: a comprehensive reference book. *London, UK: Academic*.

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

[Fahrenholtz et al., 2015] Fahrenholtz, S., Moon, T., Franco, M., Medina, D., Hazle, J. D., Stafford, R. J., Maier, F., Danish, S., Gowda, A., Shetty, A., Warburton, T., and Fuentes, D. (2015). A Model Evaluation Study for Treatment Planning of Laser Induced Thermal Therapy. *International Journal of Hyperthermia*. in preparation.

[Fahrenholtz et al., 2013] Fahrenholtz, S., Stafford, R. J., Hazle, J., and Fuentes, D. (2013). Generalised polynomial chaos-based uncertainty quantification for planning MRgLITT procedures. *International Journal of Hyperthermia*, 29(4):324–335. PMC3924420.

[Fasano et al., 2010] Fasano, A., Hömberg, D., and Naumov, D. (2010). On a mathematical model for laser-induced thermotherapy. *Applied Mathematical Modelling*, 34(12):3831–3840.

[Fuentes et al., 2013] Fuentes, D., Elliott, A., Weinberg, J. S., Shetty, A., Hazle, J. D., and Stafford, R. J. (2013). An Inverse Problem Approach to Recovery of In-Vivo Nanoparticle Concentrations from Thermal Image Monitoring of MR-Guided Laser Induced Thermal Therapy. *Ann. BME.*, 41(1):100–111. PMC3524364.

[Fuentes et al., 2010] Fuentes, D., Feng, Y., Elliott, A., Shetty, A., McNichols, R. J., Oden, J. T., and Stafford, R. J. (2010). Adaptive Real-Time Bioheat Transfer Models for Computer Driven MR-guided Laser Induced Thermal Therapy. *IEEE Trans. Biomed. Eng.*, 57(5). Cover Page, PMC3857613.

[Fuentes et al., 2009] Fuentes, D., Oden, J. T., Diller, K. R., Hazle, J., Elliott, A., , Shetty, A., and Stafford, R. J. (2009). Computational Modeling and Real-Time Control of Patient-Specific Laser Treatment Cancer. *Ann. BME.*, 37(4):763. PMC4064943.

[Fuentes et al., 2011] Fuentes, D., Walker, C., Elliott, A., Shetty, A., Hazle, J., and Stafford, R. J. (2011). MR Temperature Imaging Validation of a Bioheat Transfer Model for LITT. *International Journal of Hyperthermia*, 27(5):453–464. Cover Page, PMC3930085.

[MacLellan et al., 2013] MacLellan, C. J., Fuentes, D., Elliott, A. M., Schwartz, J., Hazle, J. D., and Stafford, R. J. (2013). Estimating nanoparticle optical absorption with magnetic resonance temperature imaging and bioheat transfer simulation. *International Journal of Hyperthermia*, (0):1–9.

[Madankan et al., 2015] Madankan, R., Stefan, W., Fahrenholtz, S., MacLellan, C., Hazle, J., Stafford, J., Weinberg, J. S., Rao, G., and Fuentes, D. (2015). Accelerated magnetic resonance thermometry in presence of uncertainties. *arXiv preprint arXiv:1510.08875*.

[Martin et al., 2012] Martin, J., Wilcox, L. C., Burstedde, C., and Ghattas, O. (2012). A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487.

[Maybeck, 1979] Maybeck, P. S. (1979). *Stochastic models, estimation, and control*, volume 1. Academic press.

[Modest, 2013] Modest, M. F. (2013). *Radiative heat transfer*. Academic press.

[Needles et al., 2010] Needles, A., Heinmiller, A., Ephrat, P., Bilan-Tracey, C., Trujillo, A., Theodoropoulos, C., Hirson, D., and Foster, F. (2010). Development of a combined photoacoustic micro-ultrasound system for estimating blood oxygenation. In *Ultrasonics Symposium (IUS), 2010 IEEE*, pages 390–393. IEEE.

[Seeger et al., 2010] Seeger, M., Nickisch, H., Pohmann, R., and Schölkopf, B. (2010). Optimization of k-space trajectories for compressed sensing by bayesian experimental design. *Magnetic resonance in medicine*, 63(1):116–126.

[Stefan et al., 2015] Stefan, W., Fuentes, D., Yeniaras, E., Hwang, K., Hazle, J. D., and Stafford, R. J. (2015). Novel Method for Background Phase Removal on MRI Proton Resonance Frequency Measurements. *Trans. Medical Imaging*. in review.

[Terejanu et al., 2012] Terejanu, G., Upadhyay, R. R., and Miki, K. (2012). Bayesian experimental design for the active nitridation of graphite by atomic nitrogen. *Experimental Thermal and Fluid Science*, 36:178–193.

[Tobin and Houghton, 2013] Tobin, R. J. and Houghton, C. J. (2013). A kernel-based calculation of information on a metric space. *Entropy*, 15(10):4540–4552.

[Walters-Williams and Li, 2009] Walters-Williams, J. and Li, Y. (2009). Estimation of mutual information: A survey. In *Rough Sets and Knowledge Technology*, pages 389–396. Springer.

[Welch, 1984] Welch, A. J. (1984). The thermal response of laser irradiated tissue. *Quantum Electronics, IEEE Journal of*, 20(12):1471–1481.

[Welch and Van Gemert, 1995] Welch, A. J. and Van Gemert, M. J. (1995). *Optical-thermal response of laser-irradiated tissue*, volume 1. Springer.

[Wray et al., 1988] Wray, S., Cope, M., Delpy, D. T., Wyatt, J. S., and Reynolds, E. O. R. (1988). Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 933(1):184–192.

[Yung et al., 2015] Yung, J., Fuentes, D., MacLellan, C. J., Maier, F., Hazle, J. D., and Stafford, R. J. (2015). Referenceless Magnetic Resonance Temperature Imaging using Gaussian Process Modeling. *Medical Physics*. in review.

A Bayes - An intuitive example

Bayes theorem is fundamental to the approach and is immediately follows from the definition of conditional probability

$$\left.\begin{aligned} p(y|x) &\equiv \frac{p(x,y)}{p(x)} \\ p(x|y) &\equiv \frac{p(x,y)}{p(y)} \end{aligned}\right\} \Rightarrow p(y|x)p(x) = p(x,y) = p(x|y)p(y) \Rightarrow \qquad \begin{aligned} p(y|x) &= \frac{p(x|y)p(y)}{p(x)} \\ p(x|y) &= \frac{p(y|x)p(x)}{p(y)} \end{aligned}$$

As a concrete example, consider the explicit two dimensional joint Gaussian distribution as a medium for understanding. Here we have two random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ defined on the same probability space, Ω.

$$\mathbf{x}_i : \Omega \to \mathbb{R} \qquad P\left(\{\omega : \mathbf{x}_i(\omega) \in A\}\right) = \int_A p(\eta_i) d\eta_i$$

Intuitively, if we are **given** the joint distribution, $p(\eta_1, \eta_2)$, knowledge of the realization of one particular random variable provides information on the realization of the second random variable.

$$p(\eta_1, \eta_2) = \frac{1}{2\;\pi\;\sqrt{\det \Sigma}} \exp\left(\frac{1}{2} \begin{bmatrix} \eta_1 - \mu_1 \\ \eta_2 - \mu_2 \end{bmatrix}^\top \underbrace{\begin{bmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 \\ r_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}}_{\equiv \Sigma} \begin{bmatrix} \eta_1 - \mu_1 \\ \eta_2 - \mu_2 \end{bmatrix} \right)$$

See [Maybeck, 1979] (Sec 3.10), characteristic functions are used to show that individual marginal densities of joint Gaussian random variable is also Gaussian.

$$p(\eta_1) = \int_{\eta_2} p(\eta_1, \eta_2)\; d\eta_2 = \frac{1}{\sqrt{2\;\pi\;\sigma_2^2}} \exp\left(- \frac{(\eta_1 - \mu_1)^2}{2\sigma_1^2} \right)$$

$$p(\eta_2) = \int_{\eta_1} p(\eta_2, \eta_1)\; d\eta_1 = \frac{1}{\sqrt{2\;\pi\;\sigma_1^2}} \exp\left(-\frac{(\eta_2-\mu_2)^2}{2\sigma_2^2}\right)$$

Conditional probablity is *defined* through the algegraic reduction of the ratio of the joint and the marginal densities

$$p(\eta_1|\eta_2) = \frac{p(\eta_1,\eta_2)}{p(\eta_2)} = \frac{1}{\sqrt{2\;\pi\;\sigma_{1|2}^2}} \exp\left(-\frac{(\eta_1-\mu_{1|2})^2}{2\sigma_{1|2}^2}\right) = \frac{p(\eta_1)}{p(\eta_2)} \frac{1}{\sqrt{2\;\pi\;\sigma_{2|1}^2}} \exp\left(-\frac{(\eta_2-\mu_{2|1})^2}{2\sigma_{2|1}^2}\right)$$

$$\mu_{1|2} = \mu_1 - \frac{r_{12}\sigma_1\sigma_2}{\sigma_2^2}(\eta_2-\mu_2) \qquad \sigma_{1|2} = \sigma_1^2 - \frac{(r_{12}\sigma_1\sigma_2)^2}{\sigma_2^2}$$

$$\mu_{2|1} = \mu_2 - \frac{r_{12}\sigma_1\sigma_2}{\sigma_1^2}(\eta_1-\mu_1) \qquad \sigma_{2|1} = \sigma_2^2 - \frac{(r_{12}\sigma_1\sigma_2)^2}{\sigma_1^2}$$

B Information theory identities

Key ideas of active Bayesian learning follows from the definition of conditional mutual information [Cover and Thomas, 2012] and *repeeated* application of the definition of conditional probability and conditional entropy.

$$p(x,y|z) \equiv \frac{p(x,y,z)}{p(z)} = \frac{p(x|y,z)p(y,z)}{p(z)} = p(x|y,z)p(y|z)$$

$$H(Y|X) \equiv E_x\left[H(Y|X=x)\right] = \sum_x p(x) H(Y|X=x) = -\sum_x p(x) \sum_y p(y|x)\log p(y|x)$$

$$\begin{aligned}
I(X;Y|Z) &\equiv \sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)} = \sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x|y,z)p(y|z)}{p(x|z)p(y|z)} = \sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x|y,z)}{p(x|z)} \\
&= -\sum_x\sum_y\sum_z p(x,y,z)\log p(x|z) + \sum_x\sum_y\sum_z p(x,y,z)\log p(x|y,z) \\
&= -\sum_x\sum_z p(x,z)\log p(x|z) + \sum_x\sum_y\sum_z p(x|y,z)p(y,z)\log p(x|y,z) \\
&= -\sum_z p(z)\sum_x p(x|z)\log p(x|z) + \sum_y\sum_z p(y,z)\sum_x p(x|y,z)\log p(x|y,z) = H(X|Z) - H(X|Y,Z)
\end{aligned}$$