

# A measure of the information content of EIT data

Andy Adler<sup>1</sup>, Richard Youmaran<sup>2</sup>, William R.B. Lionheart<sup>3</sup>

<sup>1</sup>Systems and Computer Engineering, Carleton University, Ottawa, Canada

<sup>2</sup>School of Information Technology and Engineering, University of Ottawa, Canada

<sup>3</sup>School of Mathematics, University of Manchester, U.K.

E-mail: [adler@sce.carleton.ca](mailto:adler@sce.carleton.ca)

**Abstract.** We ask: how many bits of information (in the Shannon sense) do we get from a set of EIT measurements? Here, the term *information in measurements* (IM) is defined as: the decrease in uncertainty about the contents of a medium, due to a set of measurements. This decrease in uncertainty is quantified by the change from the the inter-class model,  $q$ , defined by the prior information, to the intra-class model,  $p$ , given by the measured data (corrupted by noise). IM is measured by the expected relative entropy (Kullback-Leibler divergence) between distributions  $q$  and  $p$ , and corresponds to the channel capacity in an analogous communications system. Based on a Gaussian model of the measurement noise,  $\Sigma_n$ , and a prior model of the image element covariances  $\Sigma_x$ , we calculate  $IM = \frac{1}{2} \sum \log_2([\text{SNR}]_i + 1)$ , where  $[\text{SNR}]_i$  is the signal to noise ratio for each independent measurement calculated from the prior and noise models. For an example, we consider saline tank measurements from a 16 electrode EIT system, with a 2 cm radius non-conductive target, and calculate  $IM = 179$  bits. Temporal sequences of frames are considered, and formulae for IM as a function of temporal image element correlations are derived. We suggest that this measure may allow novel insights into questions such as distinguishability limits, optimal measurement schemes and data fusion.

*Keywords:* Measurement Information, Kullback Leibler Divergence, Electrical Impedance Tomography

## 1. Introduction

Electrical Impedance Tomography (EIT) calculates internal conductivity from surface measurements; image reconstruction is most commonly formulated as an inverse problem using regularization techniques. Regularization adds “prior information” to address the solution ill-conditioning. In this paper we ask: 1) how much information (in the Shannon sense) is there in a set of measurements? and 2) how much information do we gain when we “add” prior information via regularization? To address these questions, we pose the regularized inverse problem in using an information-theoretic formulation. First, we define a term *information in measurements* (IM) to represent the information content of data, as follows:

*Information in Measurements:* the decrease in uncertainty about

the contents of a medium, due to a set of measurements.

This definition refers to two times: 1) *before the measurement*, where the prior information is known, which describes the medium as part of a general class,  $q$ ; and 2) *after the measurement*; where specific information on this medium,  $p$ , is known (although corrupted by noise). The claim that “regularization adds prior information” may be understood as a narrowing of the distribution of the inter-class distribution,  $q$ . In the most general case, conductivities are known to be positive; however, if the medium is known to be a body, then the conductivities will be in a physiological range and have spatial correlations.

In the following sections, we formulate an approach to measure IM in three cases: 1) for a single measurement, 2) for a set of measurements (frame) of a static medium, and 3) for a temporal series of frames on a changing medium. Results are shown for an EIT system, and we discuss the implications of this measure. This work represents an extension of our conference paper (Adler and Lionheart, 2007).

## 2. Methods

### 2.1. Information Theory

In Bayesian statistics the Kullback-Leibler divergence (KLD) can be used as a measure of the information gain in moving from a prior distribution,  $q(\mathbf{y})$ , to a posterior distribution,  $p(\mathbf{y})$ . It can also be understood to be the “extra bits” of information needed to represent  $p(\mathbf{y})$  with respect to  $q(\mathbf{y})$ . The KLD is defined as (Cover and Thomas, 1991):

$$D(p||q) = \int_{\mathbf{y}} p(\mathbf{y}) \log_2 \frac{p(\mathbf{y})}{q(\mathbf{y})} d\mathbf{y} \quad (1)$$

where the integral is over all measurement channels. A comment on notation: we use  $p$  to refer to both a specific realization of an EIT image, and the distribution of measurements from that image (including measurement noise), while  $q$  represents the *prior* distribution and the distribution of its features.

One important general difficulty with direct information-theoretic measures is that of data availability. Distributions are difficult to estimate accurately, especially at the tails; and yet  $\log_2(p(\mathbf{y})/q(\mathbf{y}))$  will give large absolute values for small  $p(\mathbf{y})$  or  $q(\mathbf{y})$ . Instead, it is typical to fit data to a model with a small number of parameters. The Gaussian distribution is the most common model; it is often a good reflection of the real world distributions, and is analytically solvable in entropy integrals. Another important property of the Gaussian is that it gives the maximum entropy for a given standard deviation, allowing such models to be used to give an upper bound to entropy values. Thus, we model the distributions  $p$  and  $q$  as Gaussian with vector means  $\boldsymbol{\mu}_p, \boldsymbol{\mu}_q$  and covariances  $\boldsymbol{\Sigma}_p, \boldsymbol{\Sigma}_q$ . For Gaussian distributions the KLD is calculated:

$$D(p||q) = E_p \left[ (\mathbf{y} - \boldsymbol{\mu}_p)^t \boldsymbol{\Sigma}_p^{-1} (\mathbf{y} - \boldsymbol{\mu}_p) \right]$$

$$\begin{aligned}
& + (\mathbf{y} - \boldsymbol{\mu}_q)^t \boldsymbol{\Sigma}_q^{-1} (\mathbf{y} - \boldsymbol{\mu}_q) + \log_2 \sqrt{\frac{|2\pi \boldsymbol{\Sigma}_q|}{|2\pi \boldsymbol{\Sigma}_p|}} \\
& = \frac{1}{2} \log_2 |\boldsymbol{\Sigma}_q \boldsymbol{\Sigma}_p^{-1}| + \text{tr} (\boldsymbol{\Sigma}_p \boldsymbol{\Sigma}_q^{-1} - \mathbf{I}) + \mathbf{t}^T \boldsymbol{\Sigma}_q \mathbf{t}
\end{aligned} \tag{2}$$

where  $\mathbf{t} = \boldsymbol{\mu}_p - \boldsymbol{\mu}_q$ . As  $\mathbf{t}$  becomes large, the KLD increases dramatically as the current realization,  $p$ , moves to the tails of the prior,  $q$ .

Equation (2) represents the information for a specific  $p$ . Instead, we are interested in the average information from measurements, weighted over the prior probability of each measurement configuration. Thus, we define the *information in measurements* (IM) to be

$$\begin{aligned}
\text{IM} & = E_q [D(p||q)] = \frac{1}{2} \log_2 |\boldsymbol{\Sigma}_q \boldsymbol{\Sigma}_p^{-1}| + \text{tr} (\boldsymbol{\Sigma}_p \boldsymbol{\Sigma}_q^{-1} - \mathbf{I}) + E_q [\mathbf{t}^T \boldsymbol{\Sigma}_q \mathbf{t}] \\
& = \frac{1}{2} \log_2 |\boldsymbol{\Sigma}_q \boldsymbol{\Sigma}_p^{-1}| + \text{tr} (\boldsymbol{\Sigma}_p \boldsymbol{\Sigma}_q^{-1})
\end{aligned} \tag{3}$$

An analogy may be made between an EIT measurement system and a traditional communication system, in which IM corresponds to the channel information capacity (Cover and Thomas, 1991). The signal source transmits a symbol from an alphabet; this corresponds to one image configuration,  $p$ , from the set of possible images,  $q$ . The symbol is encoded into measurements (via the forward problem) and sent across a channel and is subject to channel noise. The communications system receiver detects the signal and must decide which symbol was sent (reconstruct the image). In this context,  $D(p||q)$ , is the differential information of a single symbol. The expected  $D(p||q)$  over all symbols,  $q$ , is the channel capacity (which we call IM, eqn 3).

Normally, the variability due to noise is much less than the prior class model, and  $|\boldsymbol{\Sigma}_p| \ll |\boldsymbol{\Sigma}_q|$ , giving

$$\text{IM} = \frac{1}{2} \log_2 (|\boldsymbol{\Sigma}_q| / |\boldsymbol{\Sigma}_p|) \tag{4}$$

The  $\text{tr}$  (trace) term will only have an effect in the very unusual case that the measurement distribution,  $p$ , is larger than that of the prior,  $q$ .

## 2.2. Information in a single channel

To clarify the measure IM and the selection of parameters, we initially consider an instrument which measures a single data channel: an impedance plethymography system, which measures a single impedance change measurement,  $\mathbf{y}$  (in voltage units), across the thorax, which varies with breathing, posture and heart activity.

In order to estimate the prior distribution, data acquisitions would be performed on a number of patients. From these data the inter-class mean ( $\mu_q$ ) and variance ( $\boldsymbol{\Sigma}_q = \sigma_q^2$ ) would be calculated. Since we are calculating difference measurements,  $\mu_q$  would be zero. For illustration purposes, consider that  $\sigma_q = 800 \text{ mV}$ .

When using the measurement system on a specific patient at a given time, the measurements,  $\mathbf{y}$ , will be due to the current physiological condition and the noise. Consider measurement noise  $\sigma_n = 10 \text{ mV}$ , and at a specific time, the difference

measurement is 1 V. The information via the KL divergence is

$$\log_2 \frac{\sigma_q}{\sigma_p} + \left( \frac{\mu_q - \mu_p}{\sigma_q} \right)^2 + \left( \frac{\sigma_p}{\sigma_q} \right)^2 - 1 \quad (5)$$

which is 6.9 bits. In order to calculate the *information in measurements*, we must average over measurements on the specific patient, giving

$$\text{IM} = \log_2 \frac{\sigma_q}{\sigma_p} + \left( \frac{\sigma_p}{\sigma_q} \right)^2 \approx \log_2 \frac{\sigma_q}{\sigma_p} \quad (6)$$

which is 6.3 bits.

### 2.3. Information in a data frame

Next, we consider a system which acquires a number of simultaneous measurements (a *data frame*) from a static medium. A naïve approximation would set IM to the sum of the information in each measurement channel considered separately. This is incorrect because of the correlation between measurements. For example, since height and weight are correlated, a measurement of height already gives some information about the likely weight. It is also possible for IM to be greater than the sum if channel measurements are negatively correlated, although this is rare for large data frames.

We consider a difference EIT system with  $n_E$  electrodes applied to a body. Using these electrodes,  $n_E$  current stimulation patterns are sequentially applied and  $n_V$  differential measurements are made for each stimulation. For an adjacent drive EIT system, voltages are typically not measured at driven electrodes, and  $n_V = n_E - 3$ . Each data frame measures a vector,  $\mathbf{v} \in \mathbb{R}^{n_M}$ , of  $n_M = n_E n_V$  data points (some of which are redundant if the medium is not changing). Difference EIT calculates difference data  $\mathbf{y} = \mathbf{v}_2 - \mathbf{v}_1$ . To improve its precision,  $\mathbf{v}_1$  is typically averaged over many data frames, at a time when the conductivity distribution may be assumed to be stable; we thus assume that  $\mathbf{v}_1$  is noise free.

The body under investigation is modelled using a finite element model (FEM) which discretizes the conductivity onto  $n_N$  piecewise smooth elements, represented by a vector  $\boldsymbol{\sigma} \in \mathbb{R}^{n_N}$ . (In this paragraph,  $\boldsymbol{\sigma}$  represents conductivity; elsewhere in this paper,  $\sigma$  is the standard deviation). Difference EIT calculates a vector of conductivity change,  $\mathbf{x} = \boldsymbol{\sigma}_2 - \boldsymbol{\sigma}_1$ , between the present conductivity distribution,  $\boldsymbol{\sigma}_2$ , and that at the reference measurement,  $\boldsymbol{\sigma}_1$ . For small variations around the reference conductivity  $\boldsymbol{\sigma}_1$ , the relationship between  $\mathbf{x}$  and  $\mathbf{y}$  can be linearized (giving the difference EIT forward model):

$$\mathbf{y} = \mathbf{J}\mathbf{x} + \mathbf{n} \quad (7)$$

where  $\mathbf{J} \in \mathbb{R}^{n_M \times n_N}$  is the Jacobian or sensitivity matrix and  $\mathbf{n} \in \mathbb{R}^{n_M}$  is the measurement noise which is assumed to be uncorrelated (white) Gaussian.  $\mathbf{J}$  is calculated from the FEM as  $\mathbf{J}_{ij} = \left. \frac{\partial \mathbf{y}_i}{\partial \mathbf{x}_j} \right|_{\boldsymbol{\sigma}_1}$ , and depends on the FEM, current injection patterns, the reference conductivity, and the electrode models. In many cases,  $n_N > n_M$ , making the system

ill-posed. The class statistics are calculated from the forward model (eqn. 7), as:

$$\begin{aligned}\boldsymbol{\mu}_y &= E[\mathbf{y}] = E[\mathbf{J}\mathbf{x} + \mathbf{n}] = \mathbf{J}\boldsymbol{\mu}_x + E[\mathbf{n}] \\ \boldsymbol{\Sigma}_y &= E[(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)^T] \\ &= \mathbf{J}E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T]\mathbf{J}^T + E[\mathbf{nn}^T]\end{aligned}\quad (8)$$

For the prior distribution  $q$ ,  $\boldsymbol{\mu}_x = \boldsymbol{\mu}_y = 0$ , since for difference EIT, positive and negative changes are equally likely. For the posterior distribution  $p$ ,  $\mathbf{x}$  is the realization of the image in the medium, and  $\boldsymbol{\mu}_y = \mathbf{J}\mathbf{x}$ . Assuming a Gaussian model, the distribution covariances may be calculated:

$$\begin{aligned}\boldsymbol{\Sigma}_q &= \mathbf{J}\boldsymbol{\Sigma}_x\mathbf{J}^T + \boldsymbol{\Sigma}_n \\ \boldsymbol{\Sigma}_p &= \boldsymbol{\Sigma}_n\end{aligned}\quad (9)$$

giving the *information in measurements*

$$\text{IM} = \frac{1}{2}\log_2|\boldsymbol{\Sigma}_q\boldsymbol{\Sigma}_p^{-1}| = \frac{1}{2}\log_2|\mathbf{J}\boldsymbol{\Sigma}_x\mathbf{J}^T\boldsymbol{\Sigma}_n^{-1} + \mathbf{I}|\quad (10)$$

In many cases, the measurement noise is independent and equal on all channels. However, for generality, we calculate a new set of independent measurements,  $\mathbf{y}'$ , with equal variance,  $\sigma_n^2$

$$\mathbf{y}' = \mathbf{D}_n^{-\frac{1}{2}}\mathbf{U}_n\mathbf{y}\quad (11)$$

using the singular value decomposition (SVD),  $\sigma_n^2\mathbf{U}_n\mathbf{D}_n\mathbf{U}_n^T = \boldsymbol{\Sigma}_n$ , where  $\sigma_n$  is the average measurement noise and is chosen so that  $\frac{1}{\sigma_n^2}\text{tr}\boldsymbol{\Sigma}_n = \text{tr}\mathbf{D}_n = n_M$ . Based on measurements,  $\mathbf{y}'$ , the forward model (7) becomes  $\mathbf{y}' = \mathbf{J}'\mathbf{x}$ , where  $\mathbf{J}' = \mathbf{D}_n^{-\frac{1}{2}}\mathbf{U}_n\mathbf{J}$ , and

$$\text{IM} = \frac{1}{2}\log_2\left|\frac{1}{\sigma_n^2}\mathbf{J}'\boldsymbol{\Sigma}_x\mathbf{J}'^T + \mathbf{I}\right|\quad (12)$$

In a similar way, a prior model can be constructed based on  $\boldsymbol{\Sigma}_x = \sigma_x^2\mathbf{R}^{-1}$ , where  $\mathbf{R}$  is the generalized Tikhonov regularization matrix. In this case, the linear regularized reconstructed image,  $\hat{\mathbf{x}}$ , will minimize the norm

$$\|\mathbf{y} - \mathbf{J}\hat{\mathbf{x}}\|_{\boldsymbol{\Sigma}_n}^2 + \|\hat{\mathbf{x}}\|_{\boldsymbol{\Sigma}_x}^2 = \|\mathbf{y}' - \mathbf{J}'\hat{\mathbf{x}}\|^2 + \lambda^2\|\hat{\mathbf{x}}\|_{\mathbf{R}}^2\quad (13)$$

where the regularization hyperparameter is  $\lambda = \frac{\sigma_n}{\sigma_x}$ . In order that  $\lambda^{-2} = (\frac{\sigma_x}{\sigma_n})^2$  represent the measurement signal to noise ratio (SNR),  $\mathbf{R}$  should be normalized so that

$$\text{tr}(\mathbf{J}'\mathbf{R}^{-1}\mathbf{J}'^T) = n_M\quad (14)$$

Using these values (10) becomes

$$\begin{aligned}\text{IM} &= \frac{1}{2}\log_2\left|\frac{\sigma_x^2}{\sigma_n^2}\mathbf{J}'\mathbf{R}^{-1}\mathbf{J}'^T + \mathbf{I}\right| \\ &= n_N\log_2\frac{\sigma_x}{\sigma_n} + \frac{1}{2}\log_2|\mathbf{J}'\mathbf{R}^{-1}\mathbf{J}'^T + \lambda^2\mathbf{I}|\end{aligned}\quad (15)$$

Equations for information content, such as (15) often diverge because measurements are not independent and the covariance matrices are singular. To solve this problem,

we extract an independent set of measurements, using the SVD,  $\mathbf{U}\mathbf{D}\mathbf{U}^T = \mathbf{J}'\mathbf{R}^{-1}\mathbf{J}'^T$ . Using this decomposition,

$$\left| \frac{1}{\lambda^2} \mathbf{J}'\mathbf{R}^{-1}\mathbf{J}'^T + \mathbf{I} \right| = |\mathbf{U}| \left| \frac{1}{\lambda^2} \mathbf{D} + \mathbf{I} \right| |\mathbf{U}^T| = \prod_{i=1}^{n_M} \frac{[\mathbf{D}]_{i,i}}{\lambda^2} + 1 \quad (16)$$

Equation (15) is thus

$$\text{IM}_{\text{data frame}} = \frac{1}{2} \sum_{i=1}^{n_M} \log_2 ([\text{SNR}]_i + 1) \quad (17)$$

where  $[\text{SNR}]_i = \frac{\sigma_s^2}{\sigma_n^2} [\mathbf{D}]_{i,i}$  is the SNR of the  $i^{\text{th}}$  independent measurement in the SVD. For very small independent measurements,  $[\text{SNR}]_i < 1$ . Given the regularization (which contributes +1 in eqn 17) these small measurements have no effect on IM, without regularization, the contribution would be negative.

#### 2.4. Information in a temporal sequence of data frames

In EIT, the number of independent measurements, and hence the spatial resolution is low. However, the temporal resolution can be very high. For example, Wilkinson *et al* (2005) developed a system with frame rates up to 1000/s. It is interesting to be able to measure the information gathered by a temporal sequence of data frames. Similar to the previous section, a naïve assumption would be to use the sum of the IM in each frame. However, this value would overestimate the information content due to measurement correlations. For example, multiple frames measuring a medium known to be unchanging would only be useful to improve the measurement SNR by frame averaging.

Using the *temporal solver* model of Adler *et al* (2007), we consider  $n_D = 2d + 1$  measurement frames from  $t - d$  to  $t + d$  around the current frame,  $t$ . Given a vertically concatenated sequence of measurements frames  $\tilde{\mathbf{y}}_t$  and the corresponding concatenated images  $\tilde{\mathbf{x}}_t$ ,

$$\begin{aligned} \tilde{\mathbf{y}}_t &= [\mathbf{y}_{t-d}^T \cdots \mathbf{y}_t^T \cdots \mathbf{y}_{t+d}^T]^T \\ \tilde{\mathbf{x}}_t &= [\mathbf{x}_{t-d}^T \cdots \mathbf{x}_t^T \cdots \mathbf{x}_{t+d}^T]^T \end{aligned} \quad (18)$$

the direct temporal forward model (7) is rewritten as

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{J}}\tilde{\mathbf{x}}_t + \tilde{\mathbf{n}} \quad (19)$$

where  $\tilde{\mathbf{n}} = [\mathbf{n}_{t-d}^T \cdots \mathbf{n}_t^T \cdots \mathbf{n}_{t+d}^T]^T$ . We consider that the model structure is constant, and thus  $\mathbf{J}'$  does not vary with time, giving  $\tilde{\mathbf{J}} = \mathbf{I} \otimes \mathbf{J}'$ , where the identity matrix  $\mathbf{I} \in \mathbb{R}^{n_D \times n_D}$  and  $\otimes$  is the Kronecker product.

The relationship between corresponding image elements between adjacent frames can be represented by an inter-frame correlation which has a value between 0 (independent) and 1 (fully dependent). The correlation could possibly be negative if subsequent frames have inverse correlation, although this scenario is physiologically unrealistic. As frames become separated in time, the inter-frame correlation decreases;

for a frame separation of  $\delta$ , the inter-frame correlation is  $\exp(-|\delta|/\gamma)$ , where  $\gamma$  is the temporal exponential decay factor in units of frames. Frames with large time difference,  $|\delta| > n_D$ , are considered independent. The value of  $\gamma$  may be estimated from the data using the approach of Dai *et al* (2007) by choosing  $\gamma$  to minimize

$$\left\| \boldsymbol{\Sigma}_{\tilde{y}} - \boldsymbol{\Sigma}_{\tilde{n}} - \tilde{\mathbf{J}} \boldsymbol{\Sigma}_{\tilde{x}} \tilde{\mathbf{J}}^T \right\|_F^2 \quad (20)$$

In order to calculate the covariance of the temporal model parameters, we assume that noise is independent between frames, and  $\boldsymbol{\Sigma}_{\tilde{n}} = E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T] = \mathbf{I} \otimes \boldsymbol{\Sigma}_n$ . On the other hand, the image elements are not independent, but are related by  $\gamma$ , so that  $\boldsymbol{\Sigma}_{\tilde{x}} = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = \boldsymbol{\Gamma} \otimes \boldsymbol{\Sigma}_x$ , where  $\boldsymbol{\Gamma} \in \mathbb{R}^{n_D \times n_D}$  and

$$[\boldsymbol{\Gamma}]_{i,j} = e^{-\frac{|i-j|}{\gamma}} \quad (21)$$

Based on this temporal formulation and (12),

$$\begin{aligned} \text{IM} &= \frac{1}{2} \log_2 \left| \frac{1}{\sigma_n^2} \tilde{\mathbf{J}}' \boldsymbol{\Sigma}_{\tilde{x}} \tilde{\mathbf{J}}'^T + \mathbf{I} \right| = \frac{1}{2} \log_2 \left| \frac{1}{\sigma_n^2} \boldsymbol{\Gamma} \otimes \sigma_x^2 \mathbf{J}' \mathbf{R}^{-1} \mathbf{J}'^T + \mathbf{I} \right| \\ &= \frac{1}{2} \log_2 \left| \boldsymbol{\Gamma} \otimes \mathbf{U} \frac{1}{\lambda^2} \mathbf{D} \mathbf{U}^T + \mathbf{I} \right| = \frac{1}{2} \log_2 \left| \boldsymbol{\Gamma} \otimes \left( \frac{1}{\lambda^2} \mathbf{D} + \mathbf{I} \right) + (\mathbf{I} - \boldsymbol{\Gamma}) \otimes \mathbf{I} \right| \\ &\approx \frac{1}{2} \log_2 |\boldsymbol{\Gamma} \otimes \left( \frac{1}{\lambda^2} \mathbf{D} + \mathbf{I} \right)| = \frac{1}{2} \log_2 |\boldsymbol{\Gamma}|^{n_M} \left| \frac{1}{\lambda^2} \mathbf{D} + \mathbf{I} \right|^{n_D} \\ &= \frac{1}{2} \log_2 (\eta^{n_D-1})^{n_M} \left| \frac{1}{\lambda^2} \mathbf{D} - \mathbf{I} \right|^{n_D} = (1 + (n_D - 1)\eta^{n_M}) \frac{1}{2} \log_2 \left| \frac{1}{\lambda^2} \mathbf{D} - \mathbf{I} \right| \\ &= (1 + (n_D - 1)\eta^{n_M}) \text{IM}_{\text{data frame}} \end{aligned} \quad (22)$$

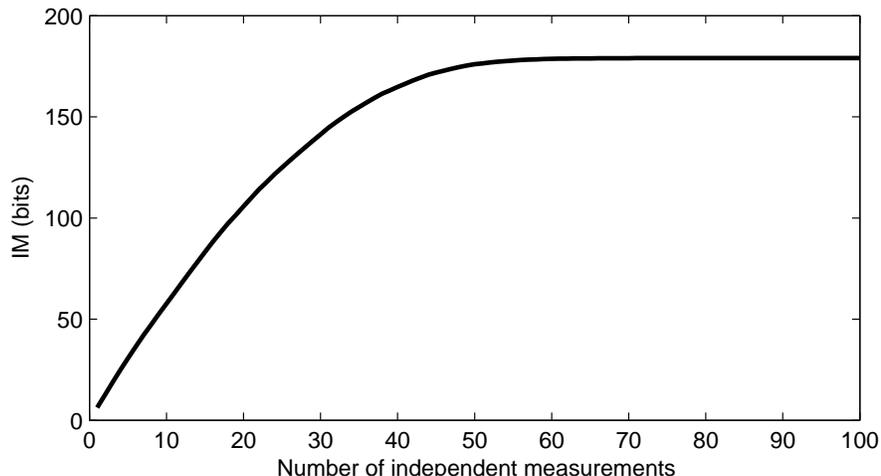
Since the factors in the Kronecker product are positive definite, we may take the product of determinants, and  $|\boldsymbol{\Gamma}| = \eta^{n_D-1}$ , where  $\eta = 1 - (e^{-\frac{1}{\gamma}})^2$ .

This expression may be interpreted as follows. When measuring multiple data frames in a temporal sequence, the first frame yields  $\text{IM}_{\text{data frame}}$  bits, and each subsequent frame adds  $\eta^{n_M}$  times this value. In this approximation, we have ignored a term  $(\mathbf{I} - \boldsymbol{\Gamma}) \otimes \mathbf{I}$ , which reflects the fact that the noise is not correlated between frames. This term is important if the medium is static, and  $\gamma = \infty$ . In this case measurements only differ by the noise, and frame data averaging may be used to increase the SNR.

### 3. Results

In order to illustrate these calculations, we present a simple example from EIT phantom measurements. Test data were measured from a 30 cm diameter and 30 cm tall plastic cylindrical phantom filled with 0.9% saline solution. Sixteen stainless steel electrodes were placed, equally spaced, around the circumference, and EIT data were acquired using the Goë-MF II EIT system (Viasys Healthcare, Höchberg, Germany) using an adjacent stimulation and measurement pattern. In order to measure the noise level, a sequence of 750 frames were measured from a homogeneous tank. To measure the signal level, a small non-conductive spherical object of 2 cm radius was introduced in the plane of the electrodes and moved into various positions on the horizontal plane. Note that this is a fairly small contrast; physiological changes are significantly larger.

To measure  $\sigma_n$ , the signal noise was calculated for each measurement channel by taking the standard deviation for the entire frame sequence. The noise ( $\sigma_n$ ) was assumed to be constant for all channels; it was reasonably consistent across measurement channels (std/mean = 0.71). We did not use the approach of (12) because we did not have enough data to accurately estimate the covariances. To measure  $\sigma_x$ , the average signal (difference from homogeneous) from different placements of the spherical object was calculated in the same way as the average noise. The regularization prior  $\mathbf{R}$  was set to the identity matrix,  $\mathbf{I}$ , scaled by the scheme of (14). The Jacobian,  $\mathbf{J}$  was calculated for a 2D 576 element mesh using the EIDORS software (Adler and Lionheart, 2006). Figure 1 shows the IM, calculated using (17), as a function of the number of independent measurements. It increases rapidly initially and reaches a plateau at approximately 50 measurements. The plateau value yields the calculation of IM= 179 bits.



**Figure 1.** *Information in measurements* (IM) (eqn 17) vs number of independent measurements ( $n_M$ ) for EIT measurements of a cylindrical phantom with a small (2 cm radius) non-conductive contrast

#### 4. Discussion

This paper has developed a method to measure the information content of a set of measurements. Three cases are considered: 1) for a single measurement channel, 2) for a data frame of multiple measurements, and 3) for a temporal sequence of data frames. A definition of IM, the *information of measurements* is introduced, and formulae to calculate its value based on the prior and posterior distributions are given. We acknowledge that the ideas we propose here are a relatively straightforward application of Bayesian regularization and information theory. We have therefore searched for previous examples of this analysis in the inverse problems literature, but have been unable to find any. Some authors have recognized the value of combining regularization and information theory (Idier *et al* , 1996, Godfrey, 1978) but have not pursued the analysis in the way presented here.

Our analysis is based on Gaussian models for the measurement noise and prior covariance. While such models have the advantage of permitting closed form solutions for entropy integrals, many feasible distributions, especially of the image prior, are not Gaussian. One useful approach may be to transform the image parameters – such as using the log conductivity – to enforce non-negativity. Another important consideration is that the prior information increases with each stimulation, if the prior distribution is calculated from the information gathered before a subsequent measurement. An example of the use of this strategy would be for optimal current patterns. In this paper, we consider the information content from the point of view of the measurements,  $\mathbf{y}$ . However, it would be conceptually interesting to develop this formulation considering the information about the image,  $\mathbf{x}$ . Distribution  $q$  would have mean,  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}_x$ , with distribution  $p$  based on the image reconstructed using Bayesian distribution,  $p(x||y)$ .

The developed formulae were implemented for a 16 electrode EIT system and a value of IM=179.0 bits calculated. Interestingly, a naïve guess for the IM for this EIT system would be the number of independent measurements  $\frac{1}{2}n_v n_e$  times  $\frac{1}{2}\log_2(\text{SNR})$ . This calculation yields  $104 \times 3.88 = 406.3$  bits. The actual IM is less because EIT measurements from the medium are highly correlated, and such correlated measures are less informative. Note that this test scenario uses a very small contrasting target. Most practical applications of EIT will give larger IM values.

The formula developed, (17), is intuitively appealing. The information in the measurements is proportional to the number of measurements and the log SNR, plus a term related to the Jacobian and the regularization matrix. This suggests that improvements to an EIT system design can thus be implemented by 1) improving the SNR, 2) adding measurement channels, and 3) designing better measurement schemes. That last function is the goal of the optimal current patterns, such as those of Demidenko *et al* (2005), Isaacson *et al* (1986), and Lionheart *et al* (2001).

We would like to suggest that this measure may allow novel insights into a number of questions regarding the performance of an EIT system, such as:

- Novel approaches to select the regularization hyperparameter. For example, the threshold for the truncated SVD may be selected at the knee of the IM vs independent measurements curve (figure 1).
- Inherent limits to the compressibility of measurement data. Measured data cannot be stored in less space than the calculated IM value.
- Distinguishability limits may be defined in terms of the IM content from small contrasts (Cheney and Isaacson, 1992).
- Fusion of EIT with other modalities. Measurements which are not independent will only add a small increment to the IM.
- Optimal current patterns may be defined in terms of maximizing IM

## Acknowledgments

This work was supported by a grant from NSERC Canada

## References

- Adler A Lionheart W R B 2006 “Uses and abuses of EIDORS: An extensible software base for EIT.” *Physiol. Meas.* 27:S25–S42
- Adler A Dai T Lionheart W R B 2007 “Temporal Image Reconstruction in Electrical Impedance Tomography” *Physiol. Meas.* 28 S1–S11
- Adler A Lionheart W R B 2007 “Information Content of EIT Measurements” *Proc. 13th Conf. Electrical Bioimpedance and 8th Conf Electrical Impedance Tomography II* 360–363 Graz, Austria, Aug 29-Sept 2
- Cheney M Isaacson D (1992) “Distinguishability in impedance imaging” *IEEE Trans Biomed Imag* 39 852–860
- Cover T M Thomas J A 1991 *Elements of Information Theory* New York: Wiley
- Dai T Soleimani M Adler A 2007 “Four Dimensional Regularization for Image Reconstruction in Electrical Impedance Tomography” *Proc. 13th Conf. Electrical Bioimpedance and 8th Conf Electrical Impedance Tomography II* 408–411 Graz, Austria, Aug 29-Sept 2
- Demidenko E Hartov A Soni N Paulsen K D 2005 “On optimal current patterns for electrical impedance tomography” *IEEE Trans Biomed Eng* 52 238–248
- Godfrey B 1978 “An Information Theory Approach to Deconvolution”, *Stanford Exploration Project: Report*
- Idier J Mohammad-Djafari A and Demoment G 1996 “Regularization Methods and Inverse Problems: An Information Theory Standpoint,” *2nd Int Conf Inverse Problems in Eng*, Le Croisic, France, June
- Isaacson D 1986 “Distinguishability of conductivities by electric current computed tomography” *IEEE Trans Med Imag* 5 91–95
- Lionheart W R B Kaipio J McLeod C N 2001 “Generalized optimal current patterns and electrical safety in EIT” *Physiol Meas* 22 85–90
- Wilkinson A J, Randall E W, Cilliers J J, Durrett D R, Naidoo T, Long T 2005 “A 1000-measurement frames/second ERT data capture system with real-time visualization” *IEEE Sensors Journal* 5 300–307