‘Most-Informative’ Compressive Measurement Design for Classification and Reconstruction of Imagery Data

Miguel Rodrigues
Department of Electronic and Electrical Engineering
University College London
London WC1E 7JE
UNITED KINGDOM

m.rodrigues@ucl.ac.uk

Liming Wang, Robert Calderbank, and Lawrence Carin
Department of Electrical and Computer Engineering
Duke University
Durham, North Carolina 27708
UNITED STATES OF AMERICA

{liming.w,robert.calderbank,lcarin}@duke.edu

ABSTRACT

This paper overviews a framework that can be used to extract the set of ‘most-informative’ compressive measurements (or features) of a source, both for reconstruction and classification applications, using mutual information as the design criterion. The basis of the framework are connections between information theory and estimation theory, which enable one to express the gradient of the mutual information with respect to the design parameters in terms of estimation-theoretic quantities in closed-form. It is argued that the availability of such closed-form gradient expressions are twofold: First, it enables one to readily perform the measurement designs using gradient-descent algorithms. Second, it also often enables one to derive additional insight about the key operations performed by the ‘most-informative’ measurement procedure. The paper also overviews representative applications of the framework in the compression and reconstruction of imagery data.

1.0 INTRODUCTION

Compressive sensing is an emerging paradigm that offers the means to simultaneously sense and compress a signal without any loss of information. The sensing process is based on the projection of the signal of interest onto a set of vectors, which are typically constituted randomly, and the recovery process is based on the resolution of an inverse problem [1-4]. The result that has captured the imagination of the signal and information processing community is that it is possible to perfectly reconstruct an n-dimensional s-sparse signal with overwhelming probability with only $O(s \log(n/s))$ linear random measurements using tractable algorithms [1-4].

However, it has also been recognized even in early compressive sensing studies that it is possible to derive substantial performance gains, in terms of the minimum number of measurements necessary to achieve perfect or nearly perfect reconstruction, by using optimized measurement designs *in lieu* of the random ones [5]. Such performance gains derive from the adaptation of the measurements to the class of signals of interest that often exhibit additional structure beyond traditional sparsity, captured by models such as union of subspaces [6,7,8], wavelet trees [9], manifolds [10,11], or even by Bayesian models [5,12,13].

There are now various instances that portray the additional gains yielded by designed measurements over random ones, both in the compressive sensing literature (e.g. [14,15]) as well as in the machine learning literature under the rubric of supervised dimensionality reduction or feature extraction (e.g. [16,17,18]). However, the existing measurement design frameworks are not always linked to the optimization of quantities with operational relevance, or are only linked to approximations to quantities of operational relevance, or do not lead to insight about the optimal measurement process.

This paper instead introduces a principled framework, which capitalizes on intersections between information theory and estimation theory, to design a set of 'most-informative' compressive measurements for data reconstruction and classification applications. The proposed framework applies to non-adaptive settings, where the measurements are designed simultaneously offline, as well as to adaptive settings, where the measurements are designed sequentially online. The framework also applies to the Gaussian case, i.e. Gaussian noise contaminated data, and to the Poisson case, i.e. vector count data. The framework can also be generalized to scenarios where one wishes to perform nonlinear measurements in *lieu* of the standard linear projections or to scenarios where one may wish to strike a balance between the accuracy of data reconstruction and data classification.

This paper only provides a glimpse of the spirit and basic features of the framework, with some representative results and applications. A complete treatment covering both the details, generalizations and applications of our framework is reported elsewhere [19-23].

2.0 INFORMATION-THEORETIC BASED COMPRESSIVE MEASUREMENT DESIGN

2.1 Compressive Measurement Model

We consider the standard measurement model:

$$y = \Phi x + w \tag{1}$$

where $y \in \mathcal{R}^{m}$ represents the measurement vector, $x \in \mathcal{R}^{n}$ represents the source vector, $w \sim \mathcal{N}(0,R^{-1}) \in \mathcal{R}^{m}$ represents Gaussian noise with mean zero and precision matrix R, and $\Phi \in \mathcal{R}^{m \times n}$ represents the measurement matrix. It is assumed that $m<n$, in keeping with our interest in the compressive regime.

We also consider a source characterized by the statistical mixture model:

$$p_x(x) = \sum_{m=1}^{M} p_c(c=m)\, p_{x|c}(x \mid c=m) \tag{2}$$

so that the source vector can be viewed to be drawn with probability $p_c(c)$ from the distribution $p_{x|c}(x|c)$, where c (c=1,…,M) corresponds to a latent class label. It is assumed that the class conditioned distribution $p_{x|c}(x|c)$ can be general (subject only to some minor regularity conditions [23]). It is also assumed that $p_c(c)$, c=1,…,M, and $p_{x|c}(x|c)$, c=1,…,M, are known (or can be estimated from training data).

The objective is to design a measurement matrix, subject to some appropriate design constraints, that extracts the 'most-informative' features from the source, either for reconstruction or classification applications.

2.2 Design Problems

Our design framework uses an information-theoretic metric – mutual information – as a proxy to extract the 'most-informative' measurement vector from the source vector. In particular, for reconstruction problems, where the objective is to infer the source vector x from the measurement vector y, the measurement matrix is designed as follows:

$$\Phi^* = \underset{\Phi:\, tr\left(\Phi\Phi^t\right)\leq m}{\operatorname{argmax}} \; I\left(x;\Phi x + w\right) \tag{3}$$

where $I(x;y=\Phi x+w)$ represents the mutual information between the source vector and the measurement vector. For classification problems, where the goal is to infer the underlying class label c from the measurement vector y, the measurement matrix is instead constructed as follows:

$$\Phi^* = \underset{\Phi:\, tr\left(\Phi\Phi^t\right)\leq m}{\operatorname{argmax}} \; I\left(c;\Phi x + w\right) \tag{4}$$

where $I(c;y=\Phi x+w)$ represents the mutual information between the underlying class label and the measurement vector. The use of mutual information as the basis of the design problems in (3) and (4) is also justified by the fact that quantities with operational relevance for classification and reconstruction problems, such as the Bayes classification error and the nonlinear minimum mean-squared error (MMSE), can be bounded by the mutual information between the class labels and the measurement vector [17] or the mutual information between source vector and the measurement vector [24]. The use of the trace constraint in the design problems in (3) and (4) regulates the energy of the sensing matrix design. Note, however, that other constraints, such as orthonormality constraints on the rows of the sensing matrix, are also possible [19-22].

The use of information-theoretic based metrics in feature design has also been attempted at in the literature [16,17,18]. However, earlier works often exploit approximations to mutual information in *lieu* of the exact quantity, such as quadratic mutual information (with quadratic Rényi entropy) [16], in view of the absence tractable closed-form mutual information expressions for a range of distributions. The key aspect in our framework that enables the use of mutual information as a proxy to effect the measurements design for data reconstruction and classification problems are connections between information theory and estimation theory.

2.3 Design Methodology

The resolution of the optimization problems embodied in (3) and (4) leverage closed-form expressions for the gradient of the mutual information with respect to the design parameters. Such closed-form expressions have been developed in [25] for settings where the measurements are contaminated by Gaussian noise and generalized in [23] for settings where the measurements obey a Poisson model – in fact, reference [23] unifies the gradient of the mutual information both for Gaussian and Poisson settings using the notion of a Bregman matrix.

In particular, for reconstruction scenarios the desired gradient is given by [25]:

$$\nabla_\Phi I\left(x;\Phi x + w\right) = R\Phi E \tag{5}$$

where $E=\mathrm{E}[(x-\mathrm{E}[x|y])(x-\mathrm{E}[x|y])^t]$ represents the MMSE matrix associated with the estimation of the source vector x from the measurement vector y. In turn, for classification scenarios the gradient is given by [20]:

$$\nabla_{\Phi} I\left(c;\Phi x+w\right)=R\Phi\left(E-\sum_{m=1}^{M}E_m\right) \tag{6}$$

where E also represents the MMSE matrix and E_m=E[(x−E[x|y,c=m])(x−E[x|y,c=m])^t|c=m] represents the class-conditioned MMSE matrix.

The relevance of the closed-form mutual information gradient expressions in (5) and (6) (either for reconstruction or classification) is manifold:

1. It provides us with the means to effect the design of the measurement matrix using gradient-descent algorithms, in view of the fact that the MMSE matrix in (5) or the equivalent MMSE matrix in (6) can be readily estimated using Monte Carlo methods.

2. It can also provide us with insight about the operations performed by the optimal measurement matrix, as discussed next, despite the fact that the optimization problems in (3) and (4) are in general non-convex.

2.4 Measurement Matrix Designs: Interpretations

Our design framework can also reveal key operations performed by the optimal measurement matrix, both for reconstruction and classification problems. By capitalizing on the mutual information gradient expressions in (5) and (6) together with the Karush-Kuhn-Tucker optimality conditions associated with the optimization problems in (3) and (4), it is possible to show that [19,20]:

1. The optimal measurement matrix diagonalizes both the noise covariance matrix as well as the MMSE matrix (for reconstruction problems) [19] or the equivalent MMSE matrix (for classification problems) [20], creating effectively a set of parallel channels (in communications systems *parlance*) in order to convey the 'most informative' modes of the source. The diagonalization of the noise covariance matrix can be defined explicity but the diagonalization of the MMSE or the equivalent MMSE matrix can only be defined implicitly because these quantities are in general a function of the measurement matrix [19,20].

2. The optimal measurement matrix then aligns the modes of the source to the modes of the noise and also weighs such modes in a manner that often admits a waterfilling-like interpretation or generalizations thereof [26].

By performing such operations, one is guaranteed to capture maximal 'information' in the measurement vector about the source vector (for reconstruction) or the underlying class label (for classification).

3.0 REPRESENTATIVE RESULTS

Figure 1 presents some representative results associated with the reconstruction of an image from compressive measurements (20 compressive measurements per 8×8 patch). The exact experimental procedure is described in detail in [19], involving the decomposition of an image into several patches, and patch-based measurement and reconstruction. The experimental procedure also involves learning a Gaussian mixture model (GMM) for the image patches from training data. It is apparent that the quality of the image associated with designed measurements, based on our framework, is far superior to that based on random measurements. The designed measurements lead to a PSNR of 25.22 dB, whereas random measurements lead to a PSNR of 22.97 dB. Other representative applications of our measurement design framework, where we also portray the gains that designed measurements offer over random ones, appear in [20-23].

(a)

(b)

(c)

Figure 1: Example recovered images for the 'barbara' image:
(a) truth; (b) random measurements; (c) designed measurements.

4.0 CONCLUSIONS

This paper has overviewed a framework, which capitalizes on connections between information theory and estimation theory, in order to design the set of 'most-informative' measurements both for data and reconstruction applications. The paper has also shown that such a design methodology leads to state-of-the-art results in a representative compressive sensing application.

The focus has been on the design of linear projections for applications where the measurements are contaminated by standard Gaussian noise. However, the general methodology generalizes in various ways: it can be applied not only for Gaussian models [19,20] but also for Poisson models [21] that are relevant for various emerging applications; it can also be applied not only for linear but also nonlinear projections designs [22]; finally, it can also be generalized to scenarios where one may wish to strike a balance between classification and reconstruction accuracy.

We also note that the sensing, analysis and processing of high-dimensional data lies at the heart of the collaboration between our groups in UCL and Duke University. Other representative lines of research, which aim to unveil fundamental limits in high-dimensional signal reconstruction and classification from low-dimensional measurements, include [27,28,29].

REFERENCES

[1] E. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications in Pure Applied Mathematics, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.

[4] D. Donoho. Compressed sensing. IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. IEEE Transactions on Signal Processing, vol. 56, no. 6, pp. 2346–2356, 2008.

[6] T. Blumensath and M. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. IEEE Transactions on Information Theory, vol. 55, no. 4, pp. 1872–1882, 2009.

[7] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. IEEE Transactions on Information Theory, vol. 55, no. 11, pp. 5302–5316, 2009.

[8] Y. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. IEEE Transactions on Signal Processing, vol. 58, no. 6, pp. 3042–3054, 2010.

[9] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hedge. Model-based compressive sensing. IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1982– 2001, 2010.

[10] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. Foundations of Computational Mathematics, vol. 9, no. 1, pp. 51–77, 2009.

[11] M. Chen, J. Silva, J. Paisley, D. Dunson, and L. Carin. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. IEEE Transactions on Signal Processing, vol. 58, no. 12, pp. 6140–6155, 2010.

[12] G. Yu and G. Sapiro. Statistical compressed sensing of Gaussian mixture models. IEEE Transactions on Signal Processing, vol. 59, no. 12, pp. 5842– 5858, 2011.

[13] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. IEEE Transactions on Signal Processing vol. 21, no. 5, pp. 2481–2499, 2012.

[14] M. Elad. Optimized projections for compressed sensing. IEEE Transactions on Signal Processing, vol. 55, no. 12, pp. 5695 –5702, 2007.

[15] J. Duarte-Carvajalino and G. Sapiro. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. IEEE Transactions on Image Processing, vol. 18, no. 7, pp. 1395 – 1408, 2009.

[16] K. Hild, D. Erdogmus, K. Torkkola, and J. Principe. Feature extraction using information-theoretic learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pp. 1385–1392, 2006.

[17] Z. Nenadic. Information discriminant analysis: Feature extraction with an information-theoretic objective. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1394–1407, 2007.

[18] K. Torkkola. Feature extraction by non-parametric mutual information maximization. Journal of Machine Learning Research, vol. 3, pp. 1415–1438, 2003.

[19] W. R. Carson, M. Chen, M. R. D. Rodrigues, A. R. Calderbank, and L. Carin. Communications-inspired projection design with application to compressive sensing. SIAM Journal on Imaging Sciences, vol. 5, pp. 1185–1212, 2012.

[20] M. Chen, W. R. Carson, M. R. D. Rodrigues, A. R. Calderbank, and L. Carin. Communications inspired linear discriminant analysis. International Conference on Machine Learning (ICML), Edinburgh, UK, June 2012.

[21] L. Wang, D. Carlson, M. R. D. Rodrigues, D. Wilcox, A. R. Calderbank and L. Carin. Designed measurements for vector count data. Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada, USA, December 2013.

[22] L. Wang, A. Razi, M. Rodrigues, R. Calderbank, and L. Carin. Nonlinear information-theoretic compressive measurement design. International Conference in Machine Learning (ICML), Beijing, Chine, June 2014.

[23] L. Wang, D. Carlson, M. R. D. Rodrigues, A. R. Calderbank and L. Carin. A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels. IEEE Transactions on Information Theory, to appear.

[24] S. Prasad. Certain relations between mutual information and fidelity of statistical estimation. Available online at: http://arxiv.org/pdf/1010.1508v1.pdf.

[25] D. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. IEEE Transactions on Information Theory, vol. 52, no. 1, pp. 141–154, 2006.

[26] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú. MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. IEEE Transactions on Information Theory, vol. 56, no. 3, pp. 1070–1084, 2010.

[27] F. Renna, A. R. Calderbank, L. Carin and M. R. D. Rodrigues. Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements. IEEE Transactions on Signal Processing, to appear.

[28] H. Reboredo, F. Renna, A. R. Calderbank and M. R. D. Rodrigues. Compressive classification of a mixture of Gaussians: Analysis, designs and geometrical interpretation. Available online at: http://arxiv.org/abs/1401.6962.

[29] M. Nokely, M. R. D. Rodrigues, A. R. Calderbank. Discrimination on the Grassmann manifold: Fundamental limits of subspace classifiers. Available online at: http://arxiv.org/abs/1404.5187.