

Gini Index as Sparsity Measure for Signal Reconstruction from Compressive Samples

Dornoosh Zonoobi, *dornoosh@nus.edu.sg*

Ashraf A. Kassim*, *Member, IEEE, ashraf@nus.edu.sg*

Yedatore V. Venkatesh (former *Senior Member IEEE*), *yv.venkatesh@gmail.com*

Dept. of Electrical & Computer Engineering, National University of Singapore

Abstract

Sparsity is a fundamental concept in compressive sampling of signals/images, which is commonly measured using the ℓ_0 norm, even though, in practice, the ℓ_1 or the ℓ_p ($0 < p < 1$) (pseudo-) norm is preferred. In this paper, we explore the use of the *Gini index (GI)*, of a discrete signal, as a more effective measure of its sparsity for a significantly improved performance in its reconstruction from compressive samples. We also successfully incorporate the GI into a stochastic optimization algorithm for signal reconstruction from compressive samples and illustrate our approach with both synthetic and real signals/images.

I. INTRODUCTION

In many applications, it is often desired to reconstruct a discrete signal from a set of incomplete observations in the spatial/time domain or in one of its transform domains. This is not possible in general, unless we assume some constraints on the signal of interest in either of the two domains. In many natural signals, it is indeed the case that there is sparsity in a spatial/time (or transform domain) and the goal of many researchers is to exploit this sparsity by employing a smaller number (than normally required) of samples, called *compressive samples*, in order to reconstruct the original signal.

In most of the current literature on compressive sampling, sparsity is measured using the ℓ_0 norm of a vector. If the original signal \underline{f} needs N samples for its complete specification in the spatial domain, but has Z zeros in it, it is said to be K -sparse, where $K = N - Z$. When \underline{f} has no nonzero elements but has Z elements which have magnitudes which are small, then we can extend the above definition to “approximate K -sparsity”: if there exists a $\underline{f}^* \in R^N$ which is K -sparse, and $\inf \|(\underline{f} - \underline{f}^*)\|_p$ is small, where the subscript $p \leq 1$ denotes the ℓ_p (pseudo-) norm, then \underline{f} is approximately K -sparse.

The desire to use only $m \ll N$ compressive samples to reconstruct the original signal/image leads to the problem of solving an under-determined system of linear equations, which can be formulated as follows: Let \underline{f} be the K -sparse signal of interest and let A denote the sampling matrix of size $m \times N$. If $\underline{y} = A\underline{f}$ are the compressive samples, the problem is then to recover \underline{f} from \underline{y} , assuming that we only know that the signal of interest is sparse. Equivalently, we need to find a vector \underline{g} , which is an estimate of \underline{f} in the presence of noise, having $\|\underline{g}\|_{\ell_0}$ nonzero elements, as a solution to the following problem:

$$\min_{\underline{g} \in R^N} \|\underline{g}\|_{\ell_0} \quad \text{subject to} \quad \|A\underline{g} - \underline{y}\| \leq \epsilon \quad (1)$$

Solving the above minimization problem is known to be computationally unwieldy in view of its NP-hard nature. As a consequence, a majority of the literature on compressive sampling (e.g. [1], [2]) propose an alternative convex ℓ_1 minimization approach. It is hoped that, by minimizing $\|\underline{g}\|_{\ell_1}$ we can estimate \underline{g} which is (approximately) equivalent to the estimate of \underline{g} in (1). In fact, it has been shown that under some conditions, these solutions are the same [1]. The ℓ_p (pseudo-) norm for $0 < p < 1$, has been recently proposed ([3]–[5]) as a measure of sparsity which is a closer approximation of the ℓ_0 than the ℓ_1 norm. The convergence, stability analyses and the conditions under which the original signal can be recovered uniquely, are discussed in [6], [7].

As demonstrated in [8], and as shown by our experiments in the following sections, the ℓ_1 norm and ℓ_p (pseudo-) norm quantify sparsity in a way that runs counter to an intuitive understanding of sparsity. As an alternative measure of sparsity, we show in this paper that the Gini index (GI) not only overcomes the limitations of standard norm-based sparsity measures [8] but is also able to reconstruct signals from compressive samples more accurately and efficiently.

II. SPARSITY AND ITS MEASURES

In signal representation, we can define *practical* sparsity in many ways. For instance, a signal is sparse if the number of the non-zero coefficients in its representation is small compared to its dimension. But in the case of real signals, this definition may not be practical. Alternatively, a signal is sparse if its *energy* is concentrated in a small number of coefficients of its representation. As applied to an image, such a sparsity is scale invariant, which means that multiplying all the coefficients of the image representation by a constant does not affect its sparsity. Such a property, which is indeed desirable in the context of image reconstruction from sparse samples, is clearly not satisfied by ℓ_1 or ℓ_p , as a sparsity measure. In general, a sparsity measure should depend on the relative distribution of energy among the coefficients, as a fraction of the total energy, and not be calculated based solely on the absolute value of each coefficient.

In fact, a good measure should be a weighted sum of coefficients of signal representation, based on the importance of a particular coefficient in the overall sparsity. As a consequence, any slight change in the value of a coefficient will affect sparsity only relative to the weight of that coefficient, which is a desirable property. More explicitly, large coefficients should have a smaller weight compared to the small ones so that they do not influence the sparsity measure in a way that does not respond to the changes of the smaller coefficients.

The authors of [8] examine and compare quantitatively, several commonly-used sparsity measures based on intuitive and desirable attributes which they call *Robin Hood*, *Scaling*, *Rising Tide*, *Cloning*, *Bill Gates*, and *Babies*. Their finding is that only the Gini index (GI), which is defined below, has all these attributes.

Gini Index (GI): Given a vector $\underline{f} = [f(1), \dots, f(N)]$, with its elements re-ordered and represented by $f_{[k]}$ for $k = 1, 2, \dots, N$, where $|f_{[1]}| \leq |f_{[2]}|, \dots, \leq |f_{[N]}|$, then

$$\text{GI}(\underline{f}) = 1 - 2 \sum_{k=1}^N \frac{|f_{[k]}|}{\|\underline{f}\|_1} \left(\frac{N - k + 1/2}{N} \right), \quad (2)$$

where $\|\underline{f}\|_1$ is the ℓ_1 norm of \underline{f} . We show in Appendix that $\text{GI}(\underline{f})$ is a *quasi-convex* function in $|\underline{f}|$.

An important advantage of GI over the conventional norm measures is that it is *normalized*, and assumes values between 0 and 1 for any vector. Further, it is 0 for the least sparse signal with all the coefficients having an equal amount of energy; and 1 for the most sparse one which has all the energy concentrated in just one coefficient. This gives us a meaningful measure, exhibiting the sparsity of the distribution. Moreover, unlike other norm measures, the value of GI is independent of the size of the vector, thereby enabling us to compare the sparsity of vectors of different sizes. As mentioned above, it is also scale-invariant and independent of the total energy of the signal. As a consequence, it is ideally suited for comparing the sparsity of a signal in different transform domains. When applied to the problem of signal reconstruction from compressive samples, GI facilitates the discovery of the sparsest domain of transform, if there is any. For example, consider the phantom image (Fig. 5) of size 512×512 . The GI of this image in different transform domains is presented in Table I, from which we can easily conclude that the phantom image is most sparse in the gradient domain and least sparse in the DFT domain. Moreover, with such a definition, it turns out that there is no need to define explicitly approximate sparsity measures: when GI is large (i.e., close to 1), then the signal has only a few values which are dominant; and when the GI is small, the signal has very few dominant values. For instance, the ℓ_0 norm does not take into account the energy distribution of the non-zero elements. Consider the two signals $\underline{X}_1 = [10, 1, 1, 0]$ and $\underline{X}_2 = [4, 4, 4, 0]$. Their sparsity according to the ℓ_0 norm is the same. However, intuitively, the sparsity

of \underline{X}_1 should be more than \underline{X}_2 , since most of the signal energy is concentrated in just one element. For other examples (of similar synthetic vectors) which demonstrate that GI is more general and consistent than K -sparsity, see [8].

The above observations serve as motivation for the use of GI as a sparsity measure for signal reconstruction from sparse samples. However, it should be noted that GI, which is neither a norm nor, even, a *pseudo-norm* (like the ℓ_p for $0 < p < 1$), has some undesirable characteristics, too. For instance, the GI of two signals cannot be algebraically manipulated to compute the GI of a composite signal. Further, we cannot analyze the sparsity of a signal by decomposing it into smaller segments, and computing the sparsity of each for summation later. Moreover, there seems to be no mathematical results that deal with conditions on the measurement matrix A for recovering uniquely the original signals from compressive samples when GI is invoked as a sparsity measure. The existing constraints of null space (NS) and restricted isometry properties (RIP) of A for unique reconstruction [9] of K -sparse signals from compressive samples are *not* applicable here. This is left as a challenging problem.

Based on the GI as a sparsity measure, we can formulate the basic problem of reconstruction as follows:

Problem: Given the measurements \underline{y} , find a vector \underline{g} (which is an estimate of \underline{f}), as a solution to the following optimization problem:

$$\max_{\underline{g} \in R^N} \text{GI}(\underline{g}) \text{ subject to } A\underline{g} = \underline{y}. \quad (3)$$

When the measurement is corrupted by noise, the signal reconstruction problem can be formulated as a corollary to the above.

Corollary: Given the measurements $\underline{y} = A\underline{f} + \underline{\eta}$, where $\underline{\eta}$ is the noise vector, satisfying the inequality $\|\underline{\eta}\|_2 \leq \epsilon$, find a vector \underline{g} (which is an estimate of \underline{f}), as a solution to the following optimization problem:

$$\max_{\underline{g} \in R^N} \text{GI}(\underline{g}) \text{ subject to } \|A\underline{g} - \underline{y}\|_2 \leq \epsilon \quad (4)$$

III. PROPOSED APPROACH

Our goal is to find the sparsest solution, in the sense of GI, among the set of admissible signals. To this end, we make use of the “*Simultaneous Perturbation Stochastic Approximation*” (SPSA) [10] to find a solution to (3).

In the reconstruction of real images, any optimization algorithm must be able to contend with the high dimension of the problem and noisiness in the measurements of the chosen objective function. SPSA has been found to be efficient for this purpose by providing a satisfactory solution using a relatively small

number of measurements of the objective function. SPSA uses observations of the desired performance function (which, in our application, is the GI of the signal) without a *direct* reference to its gradient. It approximates the gradient using only two performance function observations per iteration, regardless of the dimension of the signal. The two observations are made by simultaneously varying randomly all the variables in the performance function. This essential feature of the SPSA makes it a powerful and relatively easy to use tool for difficult multivariate optimization problems.

Further, for global optimization in the face of multiple local optima, it has been established [11] that the SPSA converges under fairly general conditions including piecewise differentiability of the objective function and Lipschitz continuity of its gradient.

A. SPSA Algorithm

Assume that \underline{f} is the original sparse vector which is being measured using a matrix A , and \underline{y} is the set of observations, $\underline{y} = A\underline{f}$. Let $\phi(\underline{\theta})$ be the function to be maximized over a vector of parameters, $\underline{\theta}$. In our application, ϕ is the GI of the admissible signals. It is assumed that measurements of $\phi(\underline{\theta})$ are available at various values of $\underline{\theta}$. The standard SPSA invokes the Lagrangian multiplier method to include constraints. In contrast, we employ a different strategy to deal with constraints for both noise-less and noisy measurements: in each iteration, we project $\underline{\theta}$ onto the set of feasible solutions.

In the case of measurements free from noise, let the set of $\underline{\theta}$ that satisfy the observations be denoted by $\mathcal{O} = \{\underline{\theta} : A\underline{\theta} = \underline{y}\}$. Then, with A^\dagger denoting the pseudo-inverse of A , $P(\underline{\theta}) = (\underline{\theta} - A^\dagger(A\underline{\theta} - \underline{y}))$ is the nearest point to $\underline{\theta}$ on \mathcal{O} , or, in other words, $P(\underline{\theta})$ is the projection of $\underline{\theta}$ onto set \mathcal{O} .

Starting from the point $\underline{\theta}_0$, at each step, we simultaneously perturb all elements of $\underline{\theta}_k$ according to a distribution vector ($\underline{\delta}_k$) whose elements are usually (but not necessarily) generated by a Bernoulli distribution. We then evaluate $\phi(\underline{\theta}_k + \beta_k \underline{\delta}_k)$ and $\phi(\underline{\theta}_k - \beta_k \underline{\delta}_k)$, and update θ as follows:

$$\underline{\theta}_{k+1} = P \left(\underline{\theta}_k + \alpha_k \frac{\phi(\underline{\theta}_k + \beta_k \underline{\delta}_k) - \phi(\underline{\theta}_k - \beta_k \underline{\delta}_k)}{2\beta_k} \underline{\delta}_k^{-1} \right) \quad (5)$$

where the gain sequence $\alpha_k := \frac{\alpha}{(B+k+1)^\gamma}$, and the small positive time-varying constant $\beta_k := \frac{\beta}{(k+1)^\lambda}$. The inverse of $\underline{\delta}_k$ vector is defined to be an element-wise inverse, hence for the Bernoulli distribution $\underline{\delta}_k^{-1} = \underline{\delta}_k$. The algorithm terminates when it converges to a solution or when the maximum number of iterations is reached. In the case of measurements with noise, the set \mathcal{O} becomes $\mathcal{O}_\epsilon = \{\underline{\theta} : \|A\underline{\theta} - \underline{y}\|_r < \epsilon\}$, and $P(\underline{\theta}) = \arg \min_{\underline{x}} \{\|\underline{x} - \underline{\theta}\|_2 \text{ s.t. } \underline{x} \in \mathcal{O}_\epsilon\}$.

B. Choice of SPSA parameters

Spall [12] provides general guidelines for choosing appropriate values for the parameters. Variables γ and λ are recommended to be set to 0.602 and 0.101, respectively; the constant B to 10% (or less) of the maximum number of expected iterations; and the parameter β to the standard deviation of the measurement noise. When the noise is small, β is chosen as 0.01. It is recommended to choose α such that the product $\alpha_0 \left(\frac{\phi(\underline{\theta}_0 + \beta_0 \underline{\delta}_0) - \phi(\underline{\theta}_0 - \beta_0 \underline{\delta}_0)}{2\beta_0 \delta_0} \right)$ is approximately equal to the smallest desired step size during the early iterations. Of all the SPSA parameters, selecting α requires the most effort, because it is often difficult to know ahead of time what a good initial step size should be. If the value of α is chosen to be too small, it can significantly increase the number of SPSA iterations required to reach the minima, while large values of α lead to unstable, diverging solutions. However, it has been shown through experimental evaluation that a wide range of α values will yield satisfactory results [13]. In the experiments presented here, α is set to a value in the range, 100 - 150. Finally, the vector $\underline{\delta}_k$ is so chosen as to have a Bernoulli distribution [14].

IV. EXPERIMENTAL RESULTS

In this section, we present experiments that evaluate the performance of GI as a sparsity measure in comparison with other conventional norm measures for 1D signals and 2D images. In all our experiments, we minimized (a) the ℓ_p -norm ($p = 0.1$) using the algorithm in [6]; and (b) the ℓ_1 -norm using NESTA [15].

A. One-Dimensional Signals

We first generate 100 random signals of size $N = 1000$ with $K = 100$, (i.e., 90% of the elements are zero) and then use m samples (where m is arbitrary) to measure the signals. For each choice of m , we select randomly the entries of the sampling matrices A of size $m \times N$, from a mean-zero Gaussian distribution. With the GI as the sparsity measure, we reconstruct the signals, and compare the reconstructions with those obtained by employing ℓ_1 and ℓ_p norms as the sparsity measures.

Figure 1 shows the success rate for perfect signal recovery versus the number of measurements for each sparsity measure. It is observed that our method outperforms the ℓ_1 and ℓ_p approaches in recovering the original signal.

In order to test the robustness of the GI as a measure of sparsity, we conducted two sets of experiments. In the first set, we performed the same experiment as discussed above, except that this time the measurements were corrupted with Gaussian noise (SNR=35). The equality constraints in all minimization

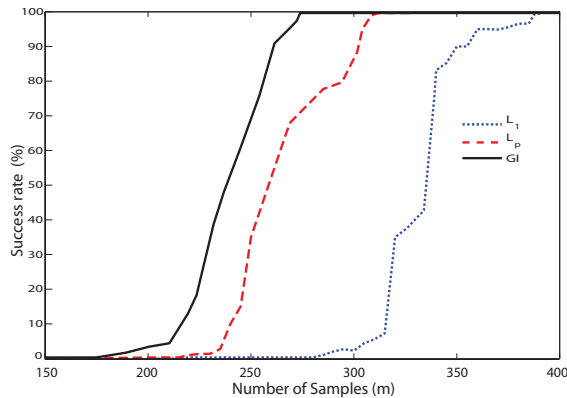


Fig. 1. Perfect signal recovery as a function of the number of samples.

problems are changed to inequality constraints with a same value of ϵ (proportional to the standard deviation of the noise). Figure 2 shows the mean square error (MSE) of the reconstructed signals versus the number of measurements. It is seen that our GI-based method significantly outperforms the ℓ_1 and ℓ_p norm-based approaches in recovering the original signal more accurately. In the second test, we (i) generate 100 random signals of size 1000 with 100 non-zero elements; (ii) make 350 measurements; and corrupt them with different levels of noise; and (iii) plot the MSE of the reconstructed signals versus SNR of the noise. Again the value of ϵ is the same for all methods. It is evident from Figure 3 that the GI-based method is superior to the ℓ_1 and ℓ_p norm-based approaches, in general, and more significantly when a high amount of noise is present.

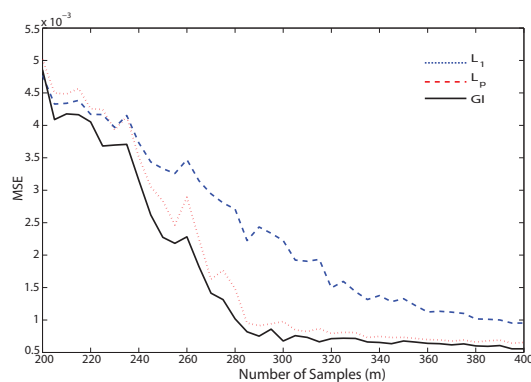


Fig. 2. MSE of the reconstructed signal versus number of samples taken for noisy signals.

B. Two-Dimensional Signals

Figures 4 and 5(a) show the images used in our experiments while Table I presents the GI values evaluated for these images in different representation domains. From Table I, it is evident that the wavelet

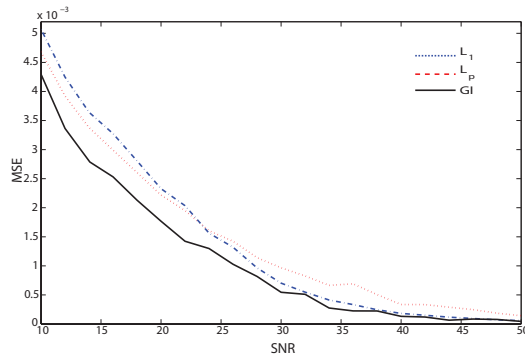
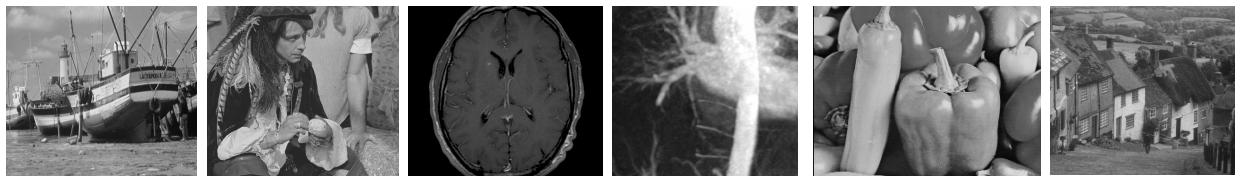


Fig. 3. MSE of the reconstructed signals vs. SNR of the noise.

domain is the sparsest for all chosen images except for the phantom image for which its gradient domain is the sparsest. The GI values in the table exhibit the extent of sparsity in the images more transparently than what can be discovered with the help of ℓ_p norm-based measures.



(a) Boat, 512×512 (b) Man, 512×512 (c) MRI, 512×512 (d) CT, 256×256 (e) Peppers, 256×256 (f) Hill, 512×512

Fig. 4. Original test images

Images/Domain	Spatial	DCT	DFT	Wavelets	Gradient
Boat	0.1922	0.6948	0.6508	0.8832	0.6282
Man	0.2450	0.6626	0.6099	0.8838	0.6099
MRI	0.5560	0.7132	0.6736	0.9389	0.8350
CT image	0.4534	0.8449	0.6098	0.9155	0.8165
Peppers	0.2467	0.6996	0.6552	0.8970	0.6771
Hill	0.2448	0.6570	0.6150	0.8781	0.6036
Phantom	0.6977	0.7600	0.5664	0.8887	0.9846

TABLE I

COMPARISON OF THE GI OF DIFFERENT TRANSFORM DOMAINS FOR THE TEST IMAGES.

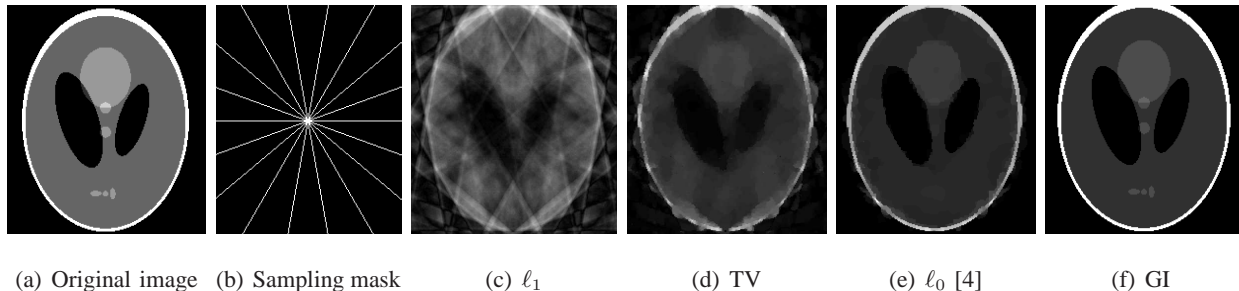


Fig. 5. Comparison of ℓ_1 -, TV -, ℓ_0 - and GI-based reconstruction of the phantom image.

For the reconstruction of the MRI-phantom image of size 256×256 , we use the samples along only 9 radial lines (see Fig. 5(b)) in its Fourier domain, and maximize the GI of its gradient magnitude. From Fig. 5, which also contains the results of ℓ_1 - and ℓ_0 -based [4] reconstructions, it can be seen that our (GI-based) method outperforms the others in reconstructing the image from only 9 radial lines.

When applying our algorithm to various images, we maximized the GI of the wavelet coefficients, subject to random observations in the Fourier domain. These results were then compared with those obtained by *minimizing* (a) Total Variation (TV)-norm, (b) ℓ_1 , and (c) $\ell_p, p = 0.1$. In the case of noisy measurements, the same value of ϵ has been used for all algorithms. From Table II, which presents the PSNR of the reconstructed images using these different approaches, it is clear that the GI-based reconstruction is superior to reconstruction using the TV, ℓ_1 and ℓ_p norm based minimizations for most of the test images.

From Table II, our method appears to be inferior to TV minimization in terms of PSNR for the CT and Pepper test images. However it is evident from Fig. 6 that images reconstructed using the GI-based method are perceptually better with less loss of details. Moreover, TV minimization blurs the output, thereby affecting the visual quality, which, in medical images, may lead to loss of important minutiae, such as small vessels.

V. CONCLUSIONS

For the reconstruction of signals/images from compressive samples, we have explored the use of the Gini index (GI) as a sparsity measure. The GI, is a more reliable and robust alternative to the currently popular ℓ_p (pseudo-) norm-based (for $0 < p \leq 1$) sparsity measure [8]. Furthermore, our proposed GI-based stochastic optimization method to reconstruct signals and images has been shown to be superior over other commonly used norm based minimization methods. An interesting challenge would be to

Input Data		Noise-less images				Noisy images				
Image name		PSNR								
% of Samples	TV min.	ℓ_1 min.	ℓ_p min.	GI max.	Noisy image	TV min.	ℓ_1 min.	ℓ_p min.	GI max.	
Boat	50%	31.77	32.23	33.12	34.71	25.36	26.92	24.91	27.06	29.81
Man	38%	30.35	30.55	30.36	34.64	26.44	25.72	27.79	28.11	29.01
MRI	27%	32.66	34.11	39.27	42.98	27.11	30.42	28.01	29.04	32.28
CT image	27%	36.07	33.61	32.81	35.81	26.73	28.91	26.12	27.43	29.10
Peppers	39%	32.90	29.66	28.52	31.82	25.55	25.81	23.64	26.09	25.83
Hill	39%	31.51	25.52	30.90	33.12	26.12	27.82	26.54	26.78	27.65

TABLE II
PSNR OF THE RECONSTRUCTED IMAGES IN NOISE-LESS AND NOISY SETTINGS.

establish the theoretical conditions under which exact reconstruction is possible with GI as a signal sparsity measure.

REFERENCES

- [1] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21 – 30, 2008.
- [2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406 – 25, 2006.
- [3] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [4] J.Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0 -minimization," *IEEE Trans. on Medical Imaging*, vol. 28, no. 1, 2009.
- [5] Y. V. Venkatesh, A. A. Kassim, and D. Zonoobi, "Medical image reconstruction from sparse samples using simultaneous perturbation stochastic optimization," in *Proc. of the ICIP Conference, Hong Kong*, September 26-29, 2010, pp. 3369–3372.
- [6] S. Foucart and M. jun Lai, "Sparsest solutions of underdetermined linear systems via l_q minimization for $0 \leq q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, May 2009.
- [7] R. Saab and O. Yilmaz, "Sparse recovery by nonconvex optimization - instance optimality," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 30–48, 2010.
- [8] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. on Information Theory*, vol. 55, no. 10, pp. 4723 – 4741, 2009.
- [9] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.

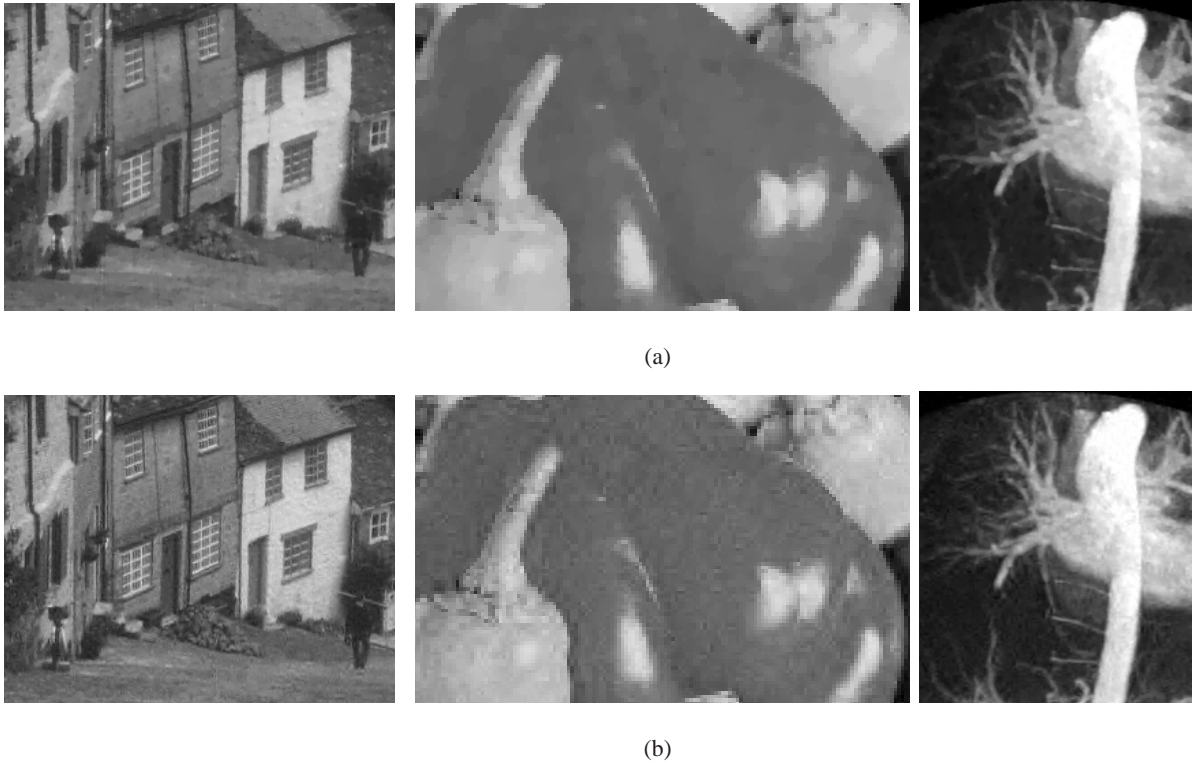


Fig. 6. (a) TV vs. (b) GI-based reconstruction of some test images (zoomed in).

- [10] J. C. Spall, "Stochastic optimization and the simultaneous perturbation method," in *Proc. of WSC 99, USA, 1999*, pp. 101–109.
- [11] J. Maryak and D. Chin, "Global random optimization by simultaneous perturbation stochastic approximation," *IEEE Trans. on Automatic Control*, vol. 53, no. 3, pp. 780 – 3, 2008.
- [12] P. Sadegh and J. Spall, "Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation," in *Proc. of 1997 ACC*, vol. vol.6, USA, 1997, pp. 3582 – 6.
- [13] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. on Automatic Control*, vol. 37, no. 3, pp. 332 – 41, 1992.
- [14] C. Renjifo, D. Barsic, C. Carmen, K. Norman, and G. S. Peacock, "Improving radial basis function kernel classification through incremental learning and automatic parameter selection," *Neurocomputing*, vol. 72, no. 1, pp. 3 – 14, 2008.
- [15] S. Becker, J. Bobin, and E. J. Candes, "NESTA: A fast and accurate first-order method for sparse recovery," <http://www.acm.caltech.edu/nesta/nesta.html>, 2009.
- [16] M. Bloem and B. Sridhar, "Optimally and equitably distributing delays with the aggregate flow model," *27th Digital Avionics Systems Conf., October 26-30, 2008*, p. AF20091, 2008.

APPENDIX

GI, which has been defined in (2), can also be expressed as follows:

$$GI(\underline{f}) = \frac{\sum_{i=1}^N \sum_{j=1}^N \left| |f(i)| - |f(j)| \right|}{2N \|\underline{f}\|_1}, \quad (6)$$

Proof: From (6), we have

$$GI(\underline{f}) = \frac{2 \sum_{i=1}^N \sum_{j=i+1}^N \left| |f(i)| - |f(j)| \right|}{2N \|\underline{f}\|_1} = \frac{2 \left(\left| |f(1)| - |f(2)| \right| + \cdots + \left| |f(N-1)| - |f(N)| \right| \right)}{2N \|\underline{f}\|_1}$$

Suppose that $|f(a)|$ is the k -th element in the sorted vector then $|f(a)| = |f_{[k]}|$. In the numerator $|f(a)|$ is compared with the other $(N-1)$ elements. It is clear that if $|f(a)| = |f_{[k]}|$, then $(k-1)$ of \underline{f} elements are smaller than $|f(a)|$, while $(N-k)$ elements of \underline{f} are greater. Therefore, we have

$$\left| |f(1)| - |f(2)| \right| + \cdots + \left| |f(1)| - |f(N)| \right| + \cdots + \left| |f(N-1)| - |f(N)| \right| = \sum_{k=1}^N (k-1)|f_{[k]}| - (N-k)|f_{[k]}|.$$

Then it follows that

$$\begin{aligned} GI(\underline{f}) &= \frac{2 \sum_{i=1}^N \sum_{j=i+1}^N \left(\left| |f(i)| - |f(j)| \right| \right)}{2N \|\underline{f}\|_1} = \frac{2 \sum_{k=1}^N \left((k-1)|f_{[k]}| - (N-k)|f_{[k]}| \right)}{2N \|\underline{f}\|_1} \\ &= \frac{2 \sum_{k=1}^N (k|f_{[k]}|) - \|\underline{f}\|_1 - N \|\underline{f}\|_1}{N \|\underline{f}\|_1} = \frac{N \|\underline{f}\|_1 + 2 \sum_{k=1}^N (k|f_{[k]}|) - \|\underline{f}\|_1 - 2N \|\underline{f}\|_1}{N \|\underline{f}\|_1} \\ &= 1 + \frac{2 \sum_{k=1}^N (k|f_{[k]}|) - \|\underline{f}\|_1 - 2N \|\underline{f}\|_1}{N \|\underline{f}\|_1} \end{aligned}$$

which is the same as (2).

To show that $GI(\underline{f})$ is quasi-convex in $|\underline{f}|$, it is sufficient to show that the sub-level sets of Eqn. (6) are convex sets [16]. The c -sublevel set can be written as:

$$\frac{\sum_{i=1}^N \sum_{j=1}^N \left| |f(i)| - |f(j)| \right|}{2N \|\underline{f}\|_1} \leq c,$$

We can rewrite this as :

$$\sum_{i=1}^N \sum_{j=1}^N \left| |f(i)| - |f(j)| \right| - 2cN \|\underline{f}\|_1 \leq 0, \quad (7)$$

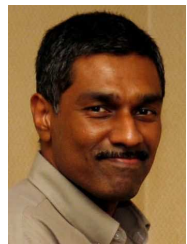
Since the first term on the left hand side of (7) can be rewritten as a point-wise maximum of linear expressions, it is convex. The second term is linear. Therefore, the above expression is convex over $|\underline{f}|$.

ACKNOWLEDGMENT

The work of the last author was partially supported by project grant NRF2007IDM-IDM002-069 on Life Spaces from the IDM Project Office, Media Development Authority of Singapore. He wishes to express grateful thanks to Professor Lawrence Wong (Director, IDMI) for his keen interest and encouragement.



Dornoosh Zonoobi received the BSc degree with honors from Shiraz University, Iran, in 2005, and the MEng degree in computer engineering from the National University of Singapore (NUS) in 2008, where she is currently pursuing the PhD degree. Her research interests include compressive signal/image processing, medical image analysis, and image coding.



Ashraf A. Kassim received his BEng (First Class Honors) in Electrical Engineering from the National University of Singapore (NUS) and had worked on machine vision systems at Texas Instruments before proceeding to obtain his PhD in Electrical & Computer Engineering from Carnegie Mellon University in 1993. He has been with the Electrical & Computer Engineering Department at NUS since 1993 and is currently Vice-Dean of the engineering school. Dr Kassim's research interests include image analysis, machine vision, and video/image processing, and compression.



Y. V. Venkatesh (SM-IEEE91) received the Ph.D. degree from the Indian Institute of Science (IISc), Bangalore. He was an Alexander von Humboldt Fellow at the Universities of Karlsruhe, Freiburg, and Erlangen, Germany; a National Research Council Fellow (USA) at the Goddard Space Flight Center, Greenbelt, MD; and a Visiting Professor at the Institut fuer Mathematik, Technische Universitat Graz, Austria, Institut fuer Signalverarbeitung, Kaiserslautern, Germany, National University of Singapore, Singapore and others. His research interests include 3D computer and robotic vision; signal processing; pattern recognition; biometrics; hyperspectral image analysis; and neural networks. As a Professor at IISc, he was also the Dean of Engineering Faculty and, earlier, the Chairman of the Electrical Sciences Division. Dr. Venkatesh is a Fellow of the Indian Academy of Sciences, the Indian National Science Academy, and the Indian Academy of Engineering. He is on the editorial board of the International Journal of Information Fusion.