Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic L_0-Minimization

Joshua Trzasko, *Student Member, IEEE,* and Armando Manduca, *Member, IEEE,*

***Abstract*— In clinical Magnetic Resonance Imaging (MRI), any reduction in scan time offers a number of potential benefits ranging from high-temporal-rate observation of physiological processes to improvements in patient comfort. Following recent developments in Compressive Sensing (CS) theory, several authors have demonstrated that certain classes of MR images which possess sparse representations in some transform domain can be accurately reconstructed from very highly undersampled K-space data by solving a convex L_1-minimization problem. Although L_1-based techniques are extremely powerful, they inherently require a degree of over-sampling above the theoretical minimum sampling rate to guarantee that exact reconstruction can be achieved. In this paper, we propose a generalization of the Compressive Sensing paradigm based on homotopic approximation of the L_0 semi-norm and show how MR image reconstruction can be pushed even further below the Nyquist limit and significantly closer to the theoretical bound. Following a brief review of standard Compressive Sensing methods and the developed theoretical extensions, several example MRI reconstructions from highly undersampled K-space data are presented.**

Keywords: Magnetic Resonance Imaging (MRI), Image Reconstruction, Compressive Sensing, Compressed Sensing, Nonconvex Optimization.

I. Introduction

In contemporary clinical practice, MRI is one of the most popular imaging modalities due to its excellent depiction of soft tissues, allowance of arbitrary vantage points, and inherent absence of emitted ionizing radiation. Despite its many advantages, a fundamental limitation of MRI is the linear relation between the number of measured data samples and net scan time. Increased scan duration presents a number of practical challenges in clinical imaging including higher susceptibility to physiological motion artifacts, diminished clinical throughput, and added patient discomfort. Recent trends towards large-scale applications such as 3D and time-resolved acquisitions generally require faster acquisition techniques to achieve clinical practicality. Unfortunately, such accelerations may result in a compromise of image quality (e.g. spatial and temporal resolution, SNR).

The emerging theory of Compressive Sensing [1], [2] has offered great insight into both when and how a signal may be recovered to high accuracy (or, in some instances, exactly) even when sampled at significantly below the Nyquist rate. To date, most CS applications, especially within medical imaging, have centered on the L_1-minimization problem. In this paper, we describe a method for reconstructing MR images at sampling rates even further below that which are achievable using L_1-based CS methods by directly attacking the ideal L_0-minimization problem. Following a review of Shannon sampling theory in the context of MRI and a discussion of L_1-based CS methods in Sections II-A and II-B, the L_0-minimization problem is described in Section II-C along with both its applied and theoretical implications. Moreover, a practical scheme is presented for addressing the L_0 semi-norm based on homotopic approximation using a wide class of deformable sparse priors, and an efficient semi-implicit numerical scheme for computation is described in Section III. Finally, several examples are presented in Section IV demonstrating the ability of our proposed technique to achieve accurate reconstructions beyond the capabilites of L_1-minimization and close to the true theoretical minimum sampling rate.

Manuscript received January XX, XXXX; revised January XX, XXXX.

Joshua Trzasko (trzasko.joshua@mayo.edu) and Armando Manduca (manduca@mayo.edu) are with the Center for Advanced Imaging Research, Mayo Clinic College of Medicine, 200 1st Street SW, Rochester, MN 55905, USA.

II. Theory

A. Shannon's Sampling Theory and MR Image Reconstruction

Shannon's sampling theorem has served as the dogma of signal processing theory for over half a century and successfuly guided the development of countless technologies ranging from telecommunication systems to MRI. While the reader is certainly familiar with Shannon's theorem, a brief review of this classical argument is given to offer a convenient juxtaposition against the more contemporary CS theory to be discussed later in this section.

Letting x be a continuous variable representing spatial position, suppose $f(x)$ is a signal of interest. For most practical applications, f is neither analytic nor finite and thus must be sampled into a discrete numerical sequence prior to any form of processing. The significance of Shannon's theorem is that, given a certain assumption about the spectral properties of f, conditions on the rate of sampling can be imposed such that the continuous signal can be recovered exactly from the discrete subset of sample measurements.

An ideal sampler is often described by the Dirac comb or impulse train,

$$s_X(x) = \sum_{n=\infty}^{\infty} \delta(x - nX), \tag{1}$$

where δ is the Dirac generalized function and X is the sampling period; as s_X is periodic, it can also be expressed as a Fourier series, namely

$$s_X(x) = \frac{1}{X} \sum_{n=\infty}^{\infty} \mathrm{e}^{\frac{-2\pi j n x}{X}}. \tag{2}$$

Given s_X, f can be sampled by simply taking the pointwise product between itself and the sampling vector, $f_X(x) = f(x) \cdot s_X(x)$. Computing the Fourier Transform of the sampled signal then yields

$$\hat{f}_X(k) = \frac{1}{X} \sum_{n=\infty}^{\infty} \hat{f}(k - \frac{n}{X}), \tag{3}$$

where k is a spectral index. Resultantly, the one-sided bandwidth of f must be less than $\frac{1}{2X}$ to ensure that aliasing in $\hat{f}_X$ does not arise from overlap of the periodic copies of $\hat{f}$. Conversely, any signal with *a priori* known maximum supported absolute frequency, k_{max}, must be sampled at a rate higher than $T_N = \frac{1}{2k_{max}}$, the so-called Nyquist limit, to ensure that the aforementioned overlap is avoided. When the Nyquist criterion is satisfied during the sampling process, the continuous signal, f, can be exactly recovered from f_X by simply filtering with an ideal low-pass signal of bandwidth equal to half the Nyquist rate.

In the discrete domain, the sampling period, X, and spatial resolution, Δx, of f_X are equivalent. If N measurements are acquired at intervals of Δk across the spectral band $B = [-k_{max}, k_{max}]$, then

$$\Delta k = \frac{1}{N \Delta x} = \mathrm{FOV}^{-1}, \tag{4}$$

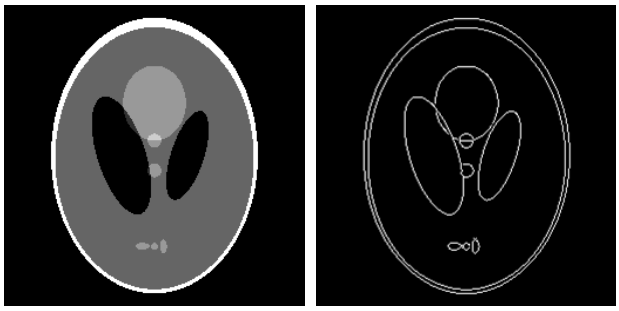
where FOV is the field-of-view of the signal space. Consequently, for a target image of fixed spatial resolution and FOV, absence of even a single measurement from the spectral grid changes the effective local value of Δk and leads to violation of Shannon's recovery condition. For higher-dimensional applications such as time-resolved 3D MRI, this rigid sampling constraint and its inherent dependence on dimensionality can be crippling.

In the event that the spectral or K-space measurement grid is only partially filled, the inversion problem becomes underdetermined and there are an infinite number of potential solutions to choose from. A simple technique for choosing a solution from this infinite set is to select the one with the minimum energy by solving

$$u = \arg\min_u \|u\|_2^2 \quad s.t. \quad \Phi u = \Phi f, \tag{5}$$

where the Fourier sampling operator $\Phi = \mathcal{F}^{-1} \hat{\Phi} \mathcal{F}$ and $\hat{\Phi}$ is the characteristic function of the subset of measured K-space values. It is quite trivial to show that (5) is achieved by simply setting $\hat{\Phi}^c$ to zero and directly inverting the spectral signal using standard Fourier methods. In practice, aliasing present within the minimum-energy solution arising from violation of Shannon's theorem typically results in an image of little diagnostic value.

Since its inception, substantial effort has been made to decrease the required scan time in MRI. Early techniques such as echo-planar imaging (EPI) [3] traverse the entirety of K-space during a single relaxation cycle (TR), offering a dramatic increase in speed but at the expense of demanding hardware performance and significantly-lowered signal-to-noise (SNR) levels. Alternatively, if the signal of interest is assumed to be strictly real, Hermitian symmetry of K-space can be exploited



(a) (b)

Fig. 1. The Shepp-Logan phantom (256×256) is a prime example of an image that posseses a sparse representation in a transform domain, here the magnitude of its gradient. While the image space representation (a) is clearly not sparse ($\|x\|_0 \approx 61\% \cdot N$), the transform representation of the image (b) exhibits very high sparsity ($\|\nabla x\|_0 \approx 3\% \cdot N$).

such that it is only necessary to perform measurements on half of the spectral grid. In practice, however, the strict reality assumption is violated due to resonant frequency variations arising from thermal instabilities as well as artifacts from physiological motion and flow. As a result, errors in phase must be corrected prior to making the Hermitian assumption. If image phase is assumed to be smoothly-varying over space, the standard approach to phase correction involves measuring a symmetric low-frequency band of K-space and estimating the image space phase solely from this restricted measurement set. The solution magnitude is then derived from a moderately undersampled subset of K-space which also includes the small support used in the phase estimation step. A complex image is then formed by conjoining the image magnitude and phase estimates, with only the real portion of this generated image being retained as the solution. Due to the necessity of acquiring a symmetric, low-frequency spectral band for use in the phase estimation process, methods which rely on Hermitian symmetry such as POCS [4] and homodyne detection [5] can only decrease the number of required measurements by less than half of that delimited by Shannon's theorem.

B. Compressive Sensing and L_1-Minimization

For many applications, the signal of interest rarely exhibits true compact spectral support. Many images, such as those produced in high-resolution MRI, can essentially be modeled as piecewise-smooth functions containing a substantial number of jump discontinuities. As the Fourier basis is comprised of quadrature trigonometric functions, it is inherently inefficient at representing sharp spatial gradients and a large number of coefficients are required to provide sufficient suppression of Gibbs ringing. A similar phenomenon occurs in textured regions or signal areas containing significant high-frequency variation. Such limitation placed on the class of signals of interest naturally raises the question as to whether bandwidth, or effectively energy, is really the functional property that should serve as the basis for determining recoverability.

At first glance, signals are not typically characterized in terms of their bandwidth but rather anecdotally, such as, for example, "cartoon-like" for piecewise-constant images like

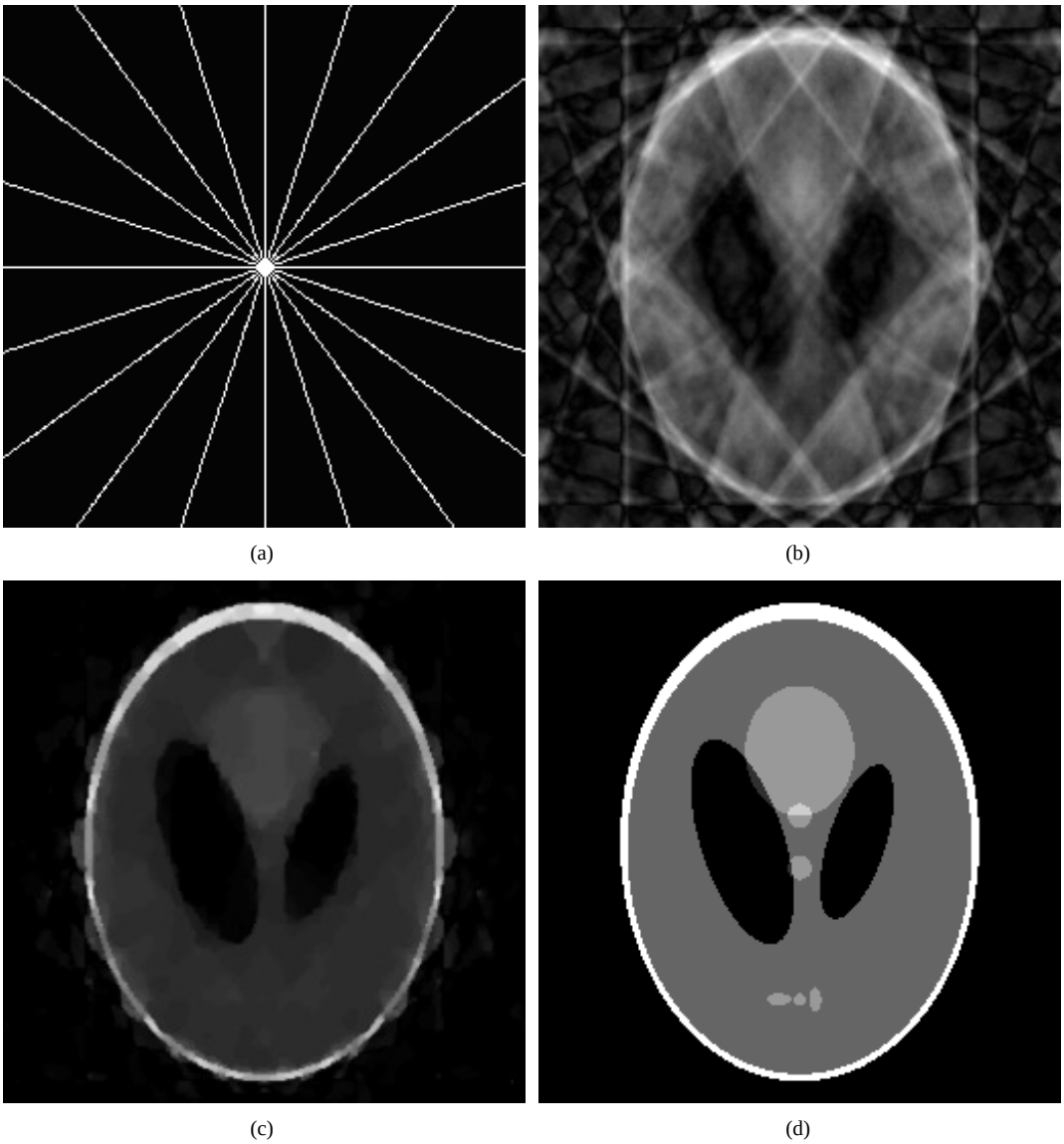


Fig. 2. A comparison of L_1-minimization versus Homotopic L_0-minimization based reconstructions of the canonical Shepp-Logan phantom shown in Figure 1a. Given a 96% undersampled K-space sampling mask containing only 10 radial lines (a) and its associated minimum energy reconstruction (b), L_1-minimization (c) fails to recover the undersampled image while Homotopic L_0-minimization (d) achieves exact reconstruction. For this example, the L_0 homotopy prior employed was $\rho(|\nabla u|, \sigma) = 1 - \mathrm{e}^{-\frac{|\nabla u|}{\sigma}}$.

the Shepp-Logan phantom shown in Figure 1. While this qualitative assessment may seem trivial at face value, such an intuitive descriptor is very powerful as it is, in essence, outlining a transform space in which the signal of interest possesses a sparse representation. For the aforementioned piecewise-constant scenario, the magnitude of the image gradient will accordingly be sparse. If such a sparsifying transform, Ψ, can be readily defined for a signal, one could ideally select the estimation with the sparsest representation in Ψ that still matches the limited observation set. Mathematically, this search consists of solving the equality-constrained L_0-minimization problem,

$$u = \arg\min_u \|\Psi u\|_0 \quad s.t. \quad \Phi u = \Phi f, \tag{6}$$

where the zero semi-norm, $\|\cdot\|_0$, is a measure of functional cardinality. Although Ψ is often assigned to be a unitary operator (e.g. wavelet, curvelet, etc...), note that this is not a required condition. For instance, in variational models, Ψ may instead be a vector collection of finite difference matrices analogous to the continuous gradient operator.

Without loss of generality, consider the case $\Psi = I$ which implies that signal of interest is intrinsically sparse in its given domain. Following the definition of Candès and Tao [6], Φ is an S-restricted isometry if, for every one of its $M \times K$ sub-matrices, $\Phi_K, \{\forall K \mid |K| < S\}$ and $M \geq K$, $\exists \delta_S \in (0, 1)$ such that

$$(1 - \delta_S)\|x\|_2^2 \leq \|\Phi_K x\|_2^2 \leq (1 + \delta_S)\|x\|_2^2 \tag{7}$$

for an arbitrary vector, x, of length K. Any measurement matrix which satisfies (7) acts as an approximate orthonormal system or isometry and the value of the constant, δ_S, provides a bound on eigenvalues of the Grammian matrix, $\Phi_K^* \Phi_K$, and

is thus proportional to the condition number of the inversion. From (7), it immediately follows that if Φ satisfies $\delta_{2S} < 1$, (6) will have a unique minimizer; conversely, this condition also asserts unique recovery of a signal with cardinality S cannot be guaranteed if less than $2S$ measurements are acquired. Exact determination of the sampling bounds generally requires probabilistic analysis of the sampling matrix as well as incorporation of the mutual incoherence between the sparsity and sampling bases; however, a detailed discussion of the techniques required to describe exact recovery bounds is beyond the scope of this paper and we refer the reader to [7], [8] for more information. One example for the particular case of random Fourier measurement ensembles will be given in Section II-C. Unlike that imposed by Shannon's theorem, this condition is independent of dimensionality and states that the number of measurements necessary to recover a signal is linearly proportional to the underlying complexity of a signal given *a priori* knowledge of a sparsifying transformation. Despite the remarkable theoretical implications of this result, directly solving the L_0-minimization problem to recover an S-sparse signal of length N requires a combinatorial search through all $\binom{N}{S}$ potential solutions and is thus intractable for any practical application.

The remarkable result of Compressive Sensing theory, pioneered by Candès et al. [1] and Donoho [2], is that if one replaces the L_0 semi-norm prior in (6) with the L_1 norm, namely

$$u = \arg\min_u \|\Psi u\|_1 \quad s.t. \quad \Phi u = \Phi f, \tag{8}$$

then exact signal recovery is still possibly albeit at the cost of a modest degree of oversampling. In other words, given a sufficient rate of sampling, the solutions to (6) and (8) are equivalent. Moreover, (8) is a tractable convex optimization problem. In practice, the requisite degree of oversampling necessary to achieve exact reconstruction of a signal of cardinality S is roughly $3S - 5S$ [9], where the multiplicative constant is proportional to the restricted isometry constant for a given Φ. Despite the fact that exact reconstruction via L_1-minimization cannot generally be achieved at the true minimum sampling rate associated with the L_0-minimization problem, this approach nonetheless offers the ability to sample drastically below the Nyquist rate and perform the recovery with a computationally tractable procedure.

In most medical imaging applications including MRI, signals of interest are rarely noise-free. Consequently, it is desirable to take into account the presence of noise during the reconstruction process such that measurement error is not propagated. Suppose that the measured data is given by $\hat{f}_n = \hat{\Phi} \circ \hat{f} + n$, where n is a stochastic noise process such as Additive White Gaussian Noise (AWGN) found on each of the quadrature measurement channels in MRI. Under this assumption, (8) can be reformulated as

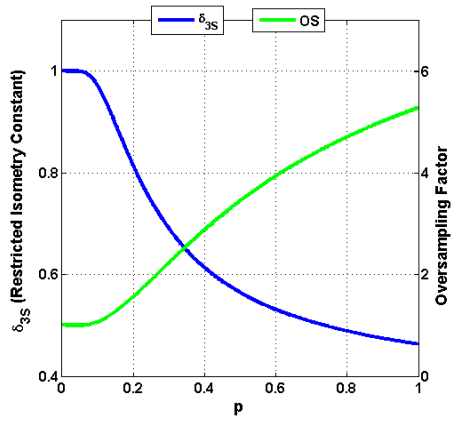
$$u = \arg\min_u \|\Psi u\|_1 \quad s.t. \quad \|\Phi u - \Phi f_n\|_2^2 \leq \epsilon, \tag{9}$$

where ϵ is a statistic of the noise process, n. It can be shown that the solution to (9) will exhibit reconstruction error on the order of $\|u - f\|_2 \leq C \cdot \epsilon$, where the constant C depends only on the restricted isometry constant [Candes]. The ability to recover an undersampled signal then degrades gracefully in the presence of noise. Moreover, medical images of interest are typically not truly sparse in any prescribed transform domain but do tend to exhibit high compressibility, i.e. their enumerated coefficients exhibit exponential decay in magnitude. Although the energy of the signal is concentrated within relatively few transform coefficients, the support of an exact representation of the signal may be dense. Nonetheless, when a compressible signal is undersampled and reconstructed via L_1-minimization, the approximation error will be on the order of $\|u - f\|_2 \leq C \cdot \|f - f_S\|_1$, where f_S consists of the S-largest coefficients of the sparse representation of Ψf. In other words, the solution generated by L_1-minimization using random sampling posseses error that is on the same order as the best possible S-term or compressive approximation [10].

Following the development of CS theory within the theoretical mathematics community, medical imaging and, in particular, MRI was immediately recognized as a strong candidate for a specific application where this new concept could have dramatic practical implications. In the past several years, the practical performance of CS theory and, specifically, L_1-minimization based techniques, has been successfully demonstrated for a large range of clinical applications including non-Cartesian and 3D MR angiography (MRA) [11], [12], and time-resolved or k-t SPARSE imaging [13]. More recently, several groups in the MRI community have proposed novel numerical techniques for attacking the L_1-minimization problem with specific focus on image reconstruction including Preconditioned Conjugate Gradient (PCG) [11], [14], Bregman iteration or inverse scale space [15], and reweighted Least Squares or FOCUSS [16], [17] methods. In essence, each of the aforementioned techniques solves the L_1-minimization problem either directly or asymptotically and the difference in quality of their produced results is anticipated to be minimal. As there has not been an extensive comparison of the computational performance of these techniques on large-scale problems, which method, if any, will best meet clinical demands is still an open question.

C. A Homotopy Method for L_0-Minimization

In Section II-B, it was stated that a signal or image can, in theory, be exactly reconstructed from fewer samples when the recovery task is cast using an L_0 semi-norm sparsity prior as opposed to the L_1 norm (see example in Figure 2). While a direct solution of (6) is infeasible for most practical applications, it is natural to ask whether there exists an alternative prior, or class of priors, that admit better sampling bounds than L_1 and are computationally achievable. Following assertions made in [2], Chartrand [18] proposed use of the L_p semi-norm $(0 < p < 1]$ class of sparsity priors within a basis pursuit reconstruction framework. While this approach generates a non-convex optimization problem and thus a global minima can no longer be guaranteed, Chartrand offered proof of asymptotic and continuous convergence of the L_p sampling bounds towards L_0 through assessment of the restricted isometry constants as well as several examples of



(a)

Fig. 3. As the L_p sparsity prior is evolved towards L_0, the restricted isometry constant, δ_{3S} ($M = \frac{3}{2}S$), is progressively relaxed. When considering the Shepp-Logan phantom in Figure 1, the cardinality of the gradient magnitude of the signal is $S \approx 2200$. Given the definition in (11) and defining the expected oversampling factor required for exact reconstruction by $\frac{K(p)}{K(0)}$, a decrease in required number of samples towards the theoretical minimum sampling rate limit can be seen as $p \rightarrow 0$.

local minima providing exact signal and image reconstructions to within numerical precision at sampling rates far below those achievable by L_1-based methods. Extending on [18], it can be shown that, for the class of L_p semi-norm priors,

$$\delta_{2M} < \left[1 + \sqrt{\frac{2M}{S}} \cdot \left(\frac{S}{M}\right)^{\frac{1}{p}}\right]^{-1} \qquad (0 < p < 1], \tag{10}$$

for some $M \geq S$ (see Appendix I for derivation). Let N denote the cardinality of the solution space and K be the required number of measurements necessary to reconstruct an S-sparse signal. When the measurement matrix is comprised of a random Fourier ensemble, $\mathbb{E}\delta_{2M} < \gamma$ whenever

$$K \geq C \left(\frac{2M \log N}{\gamma^2}\right) \log \left(\frac{2M \log N}{\gamma^2}\right) \log^2 2M, \tag{11}$$

for some absolute constant C [8]. Consequently, as $\gamma \propto \frac{1}{p}$, utilization of diminishing values of p allows a signal to be exactly reconstructed from fewer and fewer measurements (see Figure 3). Although Donoho [2] dismisses the practical use of the L_p semi-norms as a consequence of their intractability, energy functionals employing the L_p semi-norm class admit a non-zero gradient almost everywhere, unlike the L_0 problem. Thus at least local minima, many of which are more than sufficient in practice, can be found via standard descent methods as opposed to combinatorial search. When considering p to be a static variable, practical selection is non-trivial as there exists a tradeoff between reconstruction performance and problem stiffness and this is still an open problem.

Chartrand's work provides an exact characterization of the performance of reconstructions employing the L_p semi-norms $(0 < p < 1]$, but the results are not at all surprising – better approximations of the L_0 semi-norm simply yield sampling bounds that are closer to the theoretical minimum rate limit. Let the zero semi-norm of a signal, u, be defined as follows:

$$\|u\|_0 = \sum_{\Omega} \mathbf{1}\left(|u(n)| > 0\right), \tag{12}$$

where Ω is the image domain and $\mathbf{1}$ is the indicator function. Consider the general class of semimetric functionals, i.e. positive definite and symmetric about the origin, that are non-decreasing over $\mathbb{R}^+$. Intuitively, any semimetric functional, ρ, satisfying

$$\lim_{\sigma \rightarrow 0} \sum_{\Omega} \rho\left(|u(n)|, \sigma\right) = \sum_{\Omega} \mathbf{1}\left(|u(n)| > 0\right) \tag{13}$$

can be used as a sparsity prior as long as σ is chosen to be sufficiently small. The class of functionals satisfying (13) are said to be homotopic with the L_0 semi-norm in the sense that they can be continuously deformed into the desired measure. While the L_p semi-norm $(0 < p < 1]$ class clearly satisfies (13), so do many other functionals including robust error norms such as the Laplace function,

$$\rho\left(|u(n)|, \sigma\right) = 1 - \mathrm{e}^{-\frac{|u(n)|}{\sigma}}, \tag{14}$$

shown in Figure 4, the Geman-McClure function [19],

$$\rho\left(|u(n)|, \sigma\right) = \frac{|u(n)|}{|u(n)| + \sigma}, \tag{15}$$

and the concave logarithmic penalty

$$\rho\left(|u(n)|, \sigma\right) = \log\left(\frac{|u(n)|}{\sigma} + 1\right). \tag{16}$$

For a tabulation of additional non-convex functionals that are potentially homotopic with the L_0 semi-norm, see [20]. The authors also note that Candès et al. [21] have recently introduced a reweighted L_1 minimization scheme for CS which implicitly utilizes (16) with a fixed value of σ. When σ is assigned to be sufficienty small, as anticipated, results similar to those shown in Figure 2 are obtained. In effect, the reweighted L_1-minimization scheme can be viewed as a special instance of our generalized proposition.

While ρ will ideally be concave over $\mathbb{R}^+$ as this allows for straighforward generation of a uniqueness condition (see e.g. [22]), there are also many quasi-convex priors that are homotopic with the L_0 semi-norm such as the Tukey Biweight [23], Gaussian error function, and other redescending M-estimators [24], [25]. As a consequence of their uniform continuity, such functionals often contain a small convex well about the origin and are concave outside of this region. While successful reconstructions have been achieved using quasi-convex prior functions [26], it is conjectured that the presence of this convex well offers the potential for error accumulation during iterative solution generation and sub-optimal reconstructions are likely without extreme care being taken in numerical implementation. As will be demonstrated in Section IV, the practical ability to reconstruct a signal via homotopic L_0-minimization is fairly invariant to the particular employed ρ as long as the selected functional is concave. Given a reasonable schedule for the reduction of σ, it is conjectured that any such homotopy prior can provide asymptotically-equivalent results although the rate at which these are achieved is still open to theoretical investigation.

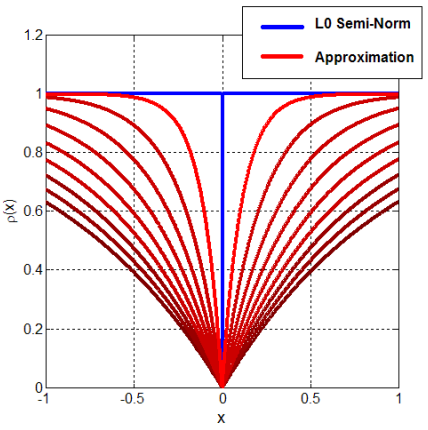


Fig. 4. Homotopic approximation of the L_0 semi-norm using the Laplace error function, $1 - e^{\frac{|u|}{\sigma}}$, shown at diminishing values of σ

Given any particular choice of ρ satisfying (13), a new reconstruction paradigm can then be defined by

$$\min_u \lim_{\sigma \to 0} \sum_{\Omega} \rho\left(|\Psi u(n)|, \sigma\right) \quad s.t. \quad \|\Phi u - \Phi f_n\|_2^2 \leq \epsilon, \quad (17)$$

where the condition $\sigma \to 0$ can be enforced through use of a standard continuation scheme such as that employed by [27] for total variation minimization (see Section III). Unlike the approach taken for L_p semi-norms, a continuation scheme allows the problem to be initially solved for large values of σ and the solution used as the initial condition of the next problem with diminished σ. Furthermore, p need not be fixed in advance and can simply be continually reduced during minimization. Relaxation of a non-convex measure is certainly not a new concept, with methods such as Graduated Non-Convexity (GNC) dating to early work on the shape-from-shading problem [28] as well as applications in Bayesian tomography [29], [30]; however, given recent insight on the conditions of when a signal can be recovered, the authors believe this to be the first application of this type of approach within the CS framework for attacking the L_0 minimization problem directly as well as the first application to magnetic resonance imaging.

D. Regularization and Complex Signals

Spatial localization in MRI relies on the provision of both frequency and phase encoding of atoms, necessitating the quadrature measurement of complex signals. If the signal of interest is dominantly real, as discussed in Section II-A, a phase correction step such as homodyne detection can be incorporated into the reconstruction process [11], offering strong additional *a priori* information as well as a potential for improved performance. In many scenarios such as velocity encoding of blood flow, phase information is crucial to image formation and the Hermitian assumption cannot be used and one must instead handle the data in its complex form. In [15], it was suggested that sparsity be enforced separately in the real and imaginary channels, an approach well known in the MR image denoising literature (e.g. [31]). The complex generalization of (17) is simply given by

$$\begin{aligned} \min_u \lim_{\sigma \to 0} \sum_{\Omega} \rho\left(|\Psi u_r(n)|, \sigma\right) + \rho\left(|\Psi u_i|(n)|, \sigma\right) \\ s.t. \quad \|\Phi u - \Phi f_n\|_2 \leq \epsilon, \end{aligned} \quad (18)$$

where u_r and u_i are the real and imaginary components of u, respectively, and it is this approach that is adopted here. Recently, the separate regularization of image-space magnitude and phase components has also been suggested [32], offering a number of potential benefits for improving the efficacy of reconstruction and a potential area of future investigation for the proposed reconstruction technique.

E. Extension to Multicoil MRI

A major advance in the MRI community came with the advent of parallel or multicoil imaging techniques. Multicoil methods such as SENSE [33], although offering accelerated acquisitions, can also be classified as "Nyquist-limited" as they are governed by Papoulis' Generalized Sampling Expansion theorem [34] which essentially states that, given C orthogonal encoders, a signal can be exactly reconstructed if it is sampled at a rate above $\frac{1}{C}$ of the Nyquist limit on each of the C channels. In practice, the individual elements of a coil array are never truly orthogonal and a certain amount of coherance exists between sensitivity profiles, the degree of which tends to increase with the number of coils. Consequently, the anticipated $\frac{1}{C}$ speed-up is rarely achievable, especially in the presence of noise. However, the loss in rate reduction can often be partially recovered though the use of the aforementioned Hermitian-driven partial K-space methods in conjunction with parallel MRI reconstruction [35].

When sampled at or above the Nyquist rate, the multicoil image reconstruction is inherently overdetermined and no solution to the problem exists. Consequently, the SENSE approach developed by Pruessmann et al. [33] formulates the reconstruction as an unregularized least-squares approximation, namely

$$\min_u \|\Gamma u - \Gamma f_n\|_2^2, \quad (19)$$

where the block operator, Γ, now incorporates both the Fourier measurement matrix as well as a vector of coil sensitivity profiles.

The amplification of noise during the reconstruction of parallel MR images is well-known and several works have addressed the regularization of (19). While the smoothing behavior of minimum-energy Tikhonov methods limits their clinical practicality for morphology-driven imaging such as MR angiograpy, *maximum a posteriori* (MAP) approaches employing edge-preserving priors have proven very promising. In particular, Raj et al. [36] employed non-convex image priors similar to those discussed in Section II-C to the parallel MRI denoising problem using a formulation nearly identical to (17) but with σ fixed; a graph cut procedure was then used for minimization of the energy functional. More recently, several authors [11], [12] have demonstrated the use of L_1-based regularization methods for undersampled parallel MRI

reconstruction analogous to that discussed in Section II-B for single receiver imaging. Given the well-established foundation of non-linear regularization in parallel MRI reconstruction, extension of the proposed reconstruction paradigm in Section II-C to this paradigm is straightforward. A more detailed discussion will be presented in a separate work.

III. Methods

In abstraction, two levels of iteration are required to compute a local minima of the problem given in (18): an outer loop that progressively diminishes the homotopy parameter, σ, and an inner loop which updates the estimate of the solution for a fixed σ, generally by some degree of iteration. Let $maxIter$ be the maximum number of allowable outer iterations, tol be the threshold for significant relative change in the solution, and $\beta \in (0,1)$. Additionally, as demonstrated in Figure 3, actually reaching $\sigma = 0$ is often unnecessary and a minimum value for σ, σ_{target}, can be prescribed such that iteration terminates on reaching this parameterization with no loss in efficacy. An outline of the algorithm for homotopic L_0 minimization is given by

Algorithm 1. Pseudocode for Homotopic L_0 Minimization

$u^0 = f$, $\sigma >> 0$
while $(count < maxIter)$
  $u^{t+1} = \mathbf{update}(u^t, \sigma)$
  if $\left(\frac{\|u^{t+1}-u^t\|_2}{\|u^t\|_2} < tol\right)$
    $\sigma = \beta \times \sigma$;
  end
  if $(\sigma < \sigma_{target})$
    break
  else
    continue
  end
  $count = count + 1$
end

In our particular implementation and for all presented examples, a semi-implicit fixed point iteration scheme based on the Lagged Diffusivity model of Vogel and Oman [37] was used for the update step. The fixed-point iteration, which is a special case of the half-quadratic regularization [19], [37], [20] commonly employed for nonconvex regularization problems in image processing, is an efficient method with only one level of internal iteration associated with inversion of the Hessian approximate via Conjugate Gradient iteration. While the theoretical convergence of quasi-Newton forms is only linear, most often only a few tens of outer iterations are needed to achieve sufficiently accurate solutions. Specifically, the update step employed in our implementation is

$$u^{t+1} = u^t - C(u^t)^{-1}\tilde{L}(u^t, \sigma, \lambda), \tag{20}$$

where $C(v)$ is a Hessian approximate and L is the Lagrangian; λ is a regularization parameter that equates the constrained and unconstrained formulations of the energy functional implicit to (18). A detailed derivation and discussion of this update step is given in Appendix II.

Although the fixed-point method provides very accurate solutions with reasonable efficiency, the authors note that the computational performance of this approach is likely sub-optimal relative to the many recent proclaimed fast numerical techniques proposed for the L_1 form of the Compressive Sensing reconstruction problem such as GPSR [38], SPGL1 [39], $l_1\_l_s$ [14], and FPC [40]; however, as the paradigm in Algorithm 1 is generic, modification of any of these methods as well as future developments may be naturally embedded within our scheme.

IV. Examples

Several example MR images reconstructed from undersampled k-space data using the presented homotopic L_0 minimization scheme can be seen in Figures 5-7. In each case, the fully sampled Cartesian k-space data (a) was undersampled retrospectively by the given binary mask (b) to allow a direct comparison of the generated reconstruction against the true image. Although only Cartesian sampling examples are considered in this paper, the extension to non-Cartesian k-space acquisition can be naturally embedded in this framework through the use of a regridding process or non-uniform Fourier transform and has been addressed in the context of L_1 minimization in [11]. Subfigures (c-e) show the result of reconstructing the respective undersampled images via zero-filling, i.e. the minimum-energy solution, L_1-minimization, and homotopic L_0-minimization; the fixed-point numerical solver presented in Section III was used for the latter two reconstructions. A line profile of a pertinent section of each image and its respective L_0 reconstruction is shown in (e) and and enlargements of another section of the fully-sampled and undersampled reconstructions are shown in (g-j). The window and leveling of all images is uniform across all subimages of each example.

The parameterizations used for computation of each of the examples was fairly consistent. σ was initially set to be the maximum value of the magnitude image and $\lambda = 1 \times 10^5$. For the L_p semi-norm prior, $\beta = 0.9$ and $\sigma_{target} = 0.2$; for all other sparsity priors, $\beta = \frac{\sqrt{10}}{10}$ and $\sigma_{target} = 1 \times 10^{-8}$. The maximum number of internal Preconditioned Conjugate Gradient iterations using in the inversion of the Hessian approximate was limited to 250 with an early termination tolerance of 1×10^{-2}. tol was set to 1×10^{-3} for the results in Figure 5 and 1×10^{-4} for the results in Figures 6 and 7. In general, the number of outer iterations will generally increase as tol is reduced although the number of outer iterations was always less than 40 and typically only 10-20. On a standard PC with a 3GHz Intel© Pentium IV processor and 1GB memory, a Matlab© implementation of our reconstruction procedure runs on average at roughly 5-10s per outer iteration or about 1-3 minutes total execution time making it comparable with existing numerical methods for L_1-minimization in the MRI literature albeit still not optimal for widespread clinical usage.

For the sake of brevity, only results obtained under the gradient magnitude sparsity measure are presented. As the key

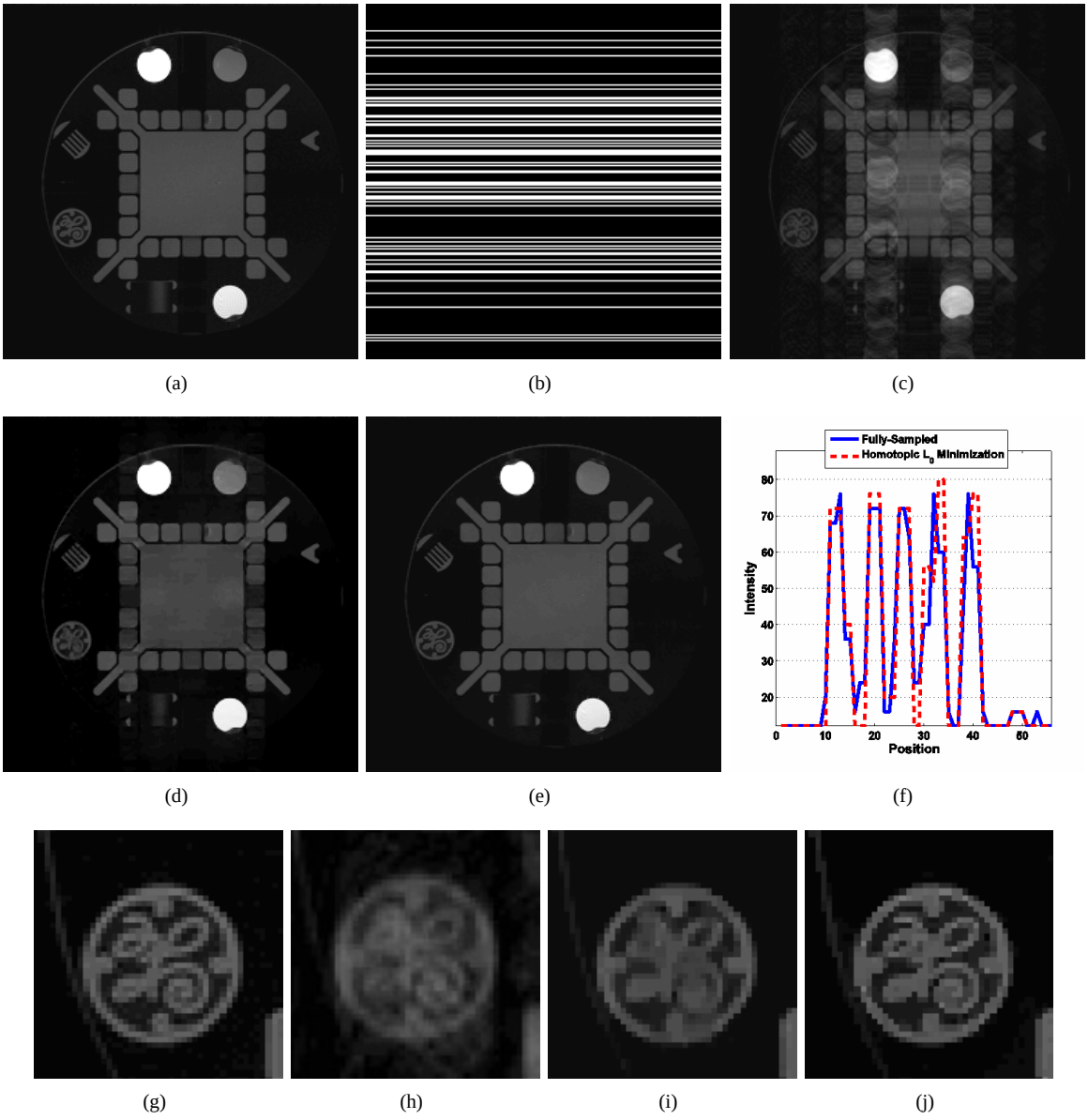


Fig. 5. Example: (a) standard General Electric© resolution phantom, (b) simulated k-space trajectory (random phase encoding, 78% undersampling), (c) minimum-energy solution via zero-filling, (d) reconstruction by L_1 minimization, (e) reconstruction by homotopic L_0 minimization using $\rho(|\nabla u|, p) = |\nabla u|^p$, (f) line profile across the phantom resolution bars, (g-j) enlargements of (a,c-e), respectively.

diagnostic information of many MR images is morphological, e.g. for assessment of stenosis in the spinal column in Figure 6, this choice is quite practical for many applications. Alternatively, wavelet sparsity measures would more easily model fine-scale detail such as anatomical texture albeit at a cost of diminished edge fidelity. A more thorough experimental comparison of results obtained using different sparsifying operators as well as the combination of operators [11], [15] will be presented in a future work.

In each of the three presented cases, note the superior recovery of the image morphology and contrast by homotopic L_0 minimization as compared to the results of zero-filling and L_1-minimization. For the standard General Electric© resolution phantom shown in Figure 5, 78% undersampling (roughly a 5X reduction in acquisition time) by random phase encoding leaves significant artifacts in the minimum energy reconstruction. While both L_1 and L_0 minimization do an excellent job at suppressing aliasing, the L_1-minimization result suffers from several regions of intensity inaccuracies as well as a slightly blurred reconstruction of the GE logo. In contrast, the homotopic L_0 reconstruction provides a much crisper logo reconstruction and, in general, provides greater intensity fidelity to the fully-sampled image. Similar phenomena are observed in the T2-weighted sagittal image of the spine example shown in Figure 6. In this case, a Cartesian approximation of a uniform-density multishot spiral sampling trajectory simulating 83% undersampling is utilized. While L_1-minimization provides a good overall reconstruction, enlargement of the C6 vertebra demonstrates the superiority of L_0-minimization over the L_1-analog at recovering both contrast and morphology. Additionally, note the fidelity of edges in the image reconstructed with the proposed method to those

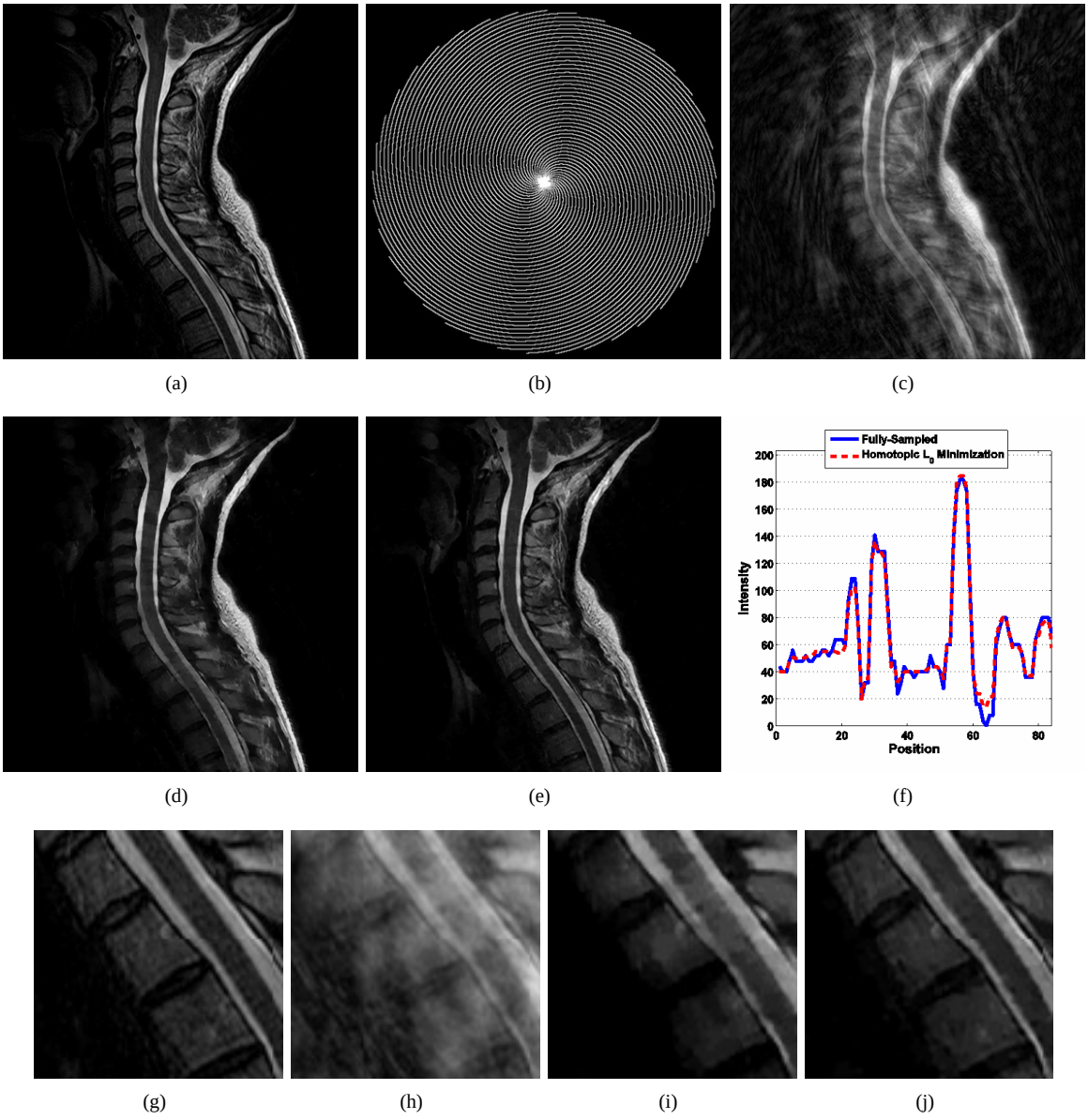


Fig. 6. Example: (a) Sagittal T2-weighted image of the spine, (b) simulated k-space trajectory (multishot Cartesian spiral, 83% undersampling), (c) minimum-energy solution via zero-filling, (d) reconstruction by L_1 minimization, (e) reconstruction by homotopic L_0 minimization using $\rho(|\nabla u|, \sigma) = \frac{|\nabla u|}{|\nabla u| + \sigma}$, (f) line profile across C6, (g-j) enlargements of (a,c-e), respectively.

of the fully-sampled data in the line profile. Figure 7 shows a T1-weighted axial cross-section of a wrist undersampled by 87% using a varaible density random mask akin to that described in [11] for multislice imaging. As with the preceding examples, again note that homotopic L_0-minimization is able to provide a substantially more accurate reconstruction than L_1-minimization as exemplified in the enlargement of the carpal tunnel region of the wrist.

The L_1-minimization procedure is generally very powerful and the relatively poor quality of the results obtained using this approach is due to the fact that the degree of undersampling was intentionally chosen to lie above the rate at which L_1-minimization is capable for the given image. When the sampling rate is sufficiently raised, following [Candes], the L_0 and L_1 minimization results are anticipated to converge; however, in this work, we are only interested in the case where the equivalence conditions are not satisfied and, in particular, homotopic L_0-minimization allows accurate reconstruction from undersampling rates greater than the level at which accurate L_1-based reconstruction is possible.

V. Conclusion

In this work, we have presented a novel extension of the Compressive Sensing paradigm for homotopic approximation of the L_0-minimization problem and shown its practical and successful application to the recovery of undersampled magnetic resonance images. Although the presented method has no guarantee of achieving a global minima as does its convex L_1 analog, the computed local minima of the homotopic L_0 minimization problem typically allow for accurate image reconstructions at higher undersampling rates than are achievable via L_1-minimization. Moreover, we have demonstrated

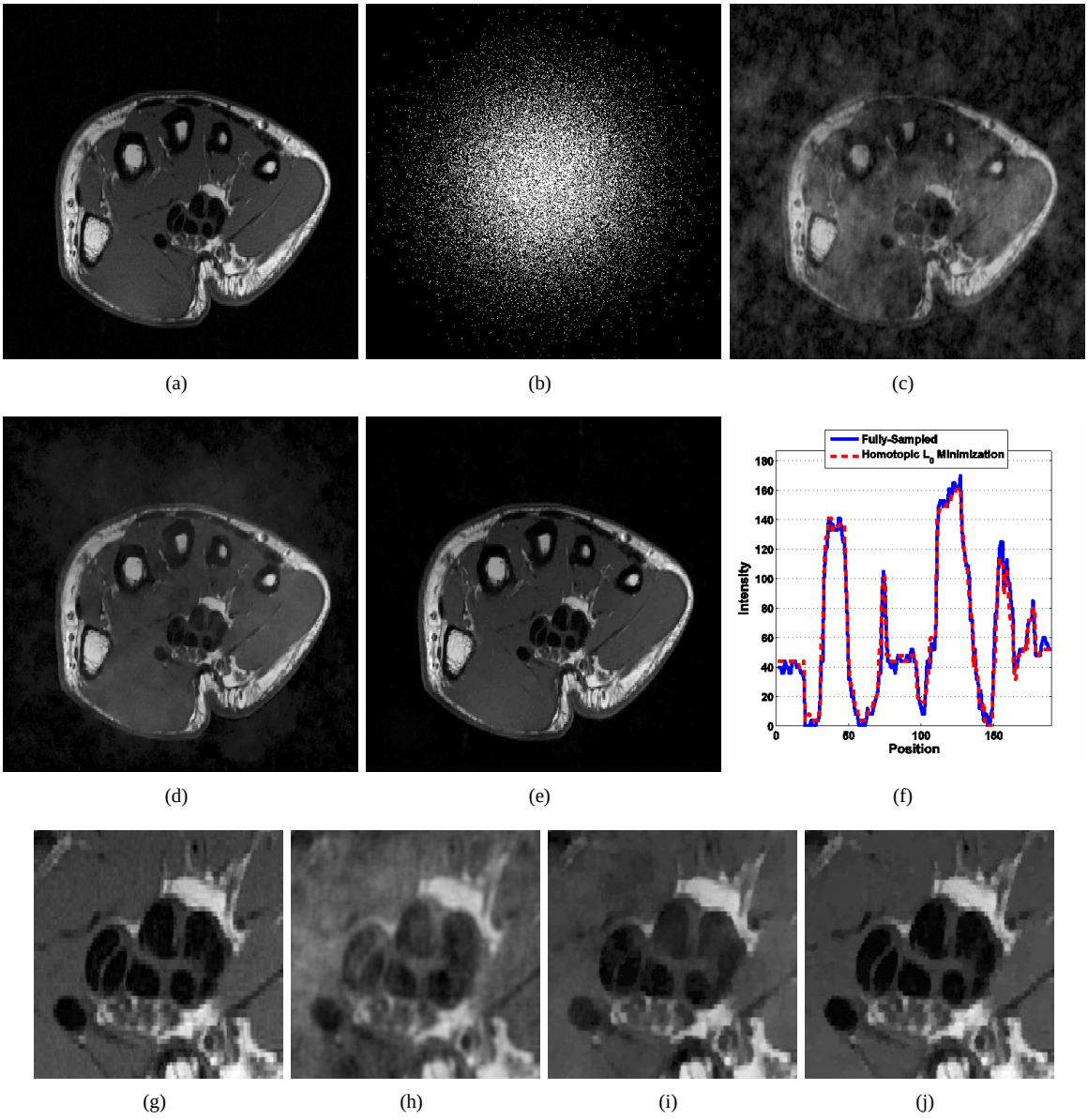


Fig. 7. Example: (a) Axial T1-weighted image of the wrist, (b) simulated k-space trajectory (variable density random, 87% undersampling), (c) minimum-energy solution via zero-filling, (d) reconstruction by L_1 minimization, (e) reconstruction by homotopic L_0 minimization using $\rho(|\nabla u|, \sigma) = \log\left(\frac{|\nabla u|}{\sigma} + 1\right)$, (f) line profile across the 2^{nd} and 3^{rd} metacarpals, (g-j) enlargements of (a,c-e), respectively.

that such local minima can be achieved in a computationally practical manner even under the usage of several different homotopic sparsity priors. While the specific application of this technique within this work is for complex MRI recovery, the extension to other medical imaging modalities such as x-ray CT is quite natural and will be investigated in a future publication.

Appendix I: Derivation of Restricted Isometry Constants for L_p Semi-Norm Priors $(0 < p < 1]$

Consider the special case of (17) where $\epsilon = 0$, $\Psi = I$, and $\rho(|u|, p) = |u|^p$, namely

$$\min_u \lim_{p \to 0} \sum_{\Omega} |u(n)|^p \quad s.t. \quad \Phi u = \Phi f,$$

where f and u are again the true and approximated signals, respectively. Let the reconstruction error $h = u - f$. For a fixed value of p, a sufficient condition for the uniqueness of a solution to the equality-constrained L_p-minimization problem is given by

$$\sum_{T_0} |h(n)|^p \geq \sum_{T_0^c} |h(n)|^p,$$

where $T_0 = \text{supp}\{f\}$ and $p \in (0, 1]$ [18]. Next, suppose that $h_{T_0^c}$ is reordered by magnitude (descending) into sets of size $M \geq K$, each denoted as h_{T_j}, namely $h(t) \leq h(s), \forall s \in T_j, t \in T_{j+1}$. Without consequence, note that the last subset might be of cardinality smaller than M. Invoking the monotonicity of the L_p semi-norms over $\mathbb{R}^+$, it immediately follows

that

$$\left(\|h_{T_{j+1}}\|_\infty\right)^p \leq M^{-1}\sum_{T_j}|h(n)|^p.$$

From [41], the relation

$$\|h_{T_0}\|_2 \leq \frac{\sqrt{2}\delta_{2M}}{1-\delta_{2M}}\sum_{j\geq 2}\|h_{T_j}\|_2$$

can also be derived through clever manipulation of the inner products inherent to the restriced isometry identity in (7). Noting the equivalence of norms over finite domains, simple algebraic manipulation of the above relation yields

$$S^{-1}\|h_{T_0}\|_1 \leq \sqrt{\frac{2M}{S}}\frac{\delta_{2M}}{1-\delta_{2M}}M^{-\frac{1}{2}}\sum_{j\geq 2}\|h_{T_j}\|_2.$$

Again exploiting finite norm equivalence and subsequently summing over all subsets $T_j \subset T_0^c$,

$$\left(M^{-\frac{1}{2}}\sum_{j\geq 2}\|h_{T_j}\|_2\right)^p \leq M^{-1}\sum_{T_0^c}|h(n)|^p,$$

where the subadditivity of concave functions over $\mathbb{R}^+$ is utilized. Consequently,

$$\sum_{T_0}|h(n)|^p \leq \frac{S}{M}\left(\sqrt{\frac{2M}{S}}\frac{\delta_{2M}}{1-\delta_{2M}}\right)^p\sum_{T_0^c}|h(n)|^p.$$

On imposing the uniqueness condition, if

$$\frac{S}{M}\left(\sqrt{\frac{2M}{S}}\frac{\delta_{2M}}{1-\delta_{2M}}\right)^p < 1,$$

$\sum_{T_0}|h(n)|^p = 0$ and the exact recovery condition is implictly fulfilled. Resultantly, this condition is met whenever the upper bound on the restricted isometry constant given in (10) is satisfied.

Appendix II: Fixed Point Solver for Homotopic L_0 Minimization

Consider the energy functional

$$E(u,\sigma,\lambda) = \sum_\Omega \rho\left(|\Psi u(n)|,\sigma\right) + \frac{\lambda}{2}\|\Phi u - \Phi f_n\|_2^2,$$

where λ is a regularization parameter controlling the tradeoff between the sparsity and fidelity measures. It is well-known in optimization theory that the Karush-Kuhn-Tucker (KKT) conditions guarantee $\exists\lambda \geq 0$ such that the solutions to

$$\min_u E(u,\sigma,\lambda)$$

and the constrained problem given in (17) are equivalent. In general, analytic determination of λ as a function of ϵ is quite challenging and this parameter is often either assigned manually or determined iteratively using estimation techniques such as cross -validation [42]; in this work, only the former approach is taken.

Let C equal the number of distinct sparsifying operators implicit in Ψ. For most applications such as when Ψ is a basis transformation operator (e.g. wavelet), $C = 1$; however, it is possible that $C > 1$ such as with variational models that require use of multiple finite difference operators. When the sparsifying operators in Ψ are arranged to yield a $C\Omega \times \Omega$ column vector, the magnitude operator can be written in matrix form as

$$|\Psi u| = \theta\left(J_C\left[\Psi u \circ \Psi u\right]\right)$$

where θ is an element-wise square-root operator, J_C is a row vector of C identity matrices, and $\circ$ denotes the Hadamard product. Although $E(u,\sigma,\lambda)$ is not necessarily convex, for fixed σ and λ, local minima of the energy functional are given by

$$\begin{aligned} L(u,\sigma,\lambda) &= \frac{dE}{du} = 0 \\ &= \Psi^*\left(I_C \otimes \Lambda(u)\right)\Psi u + \lambda\Phi^*\left(\Phi u - \Phi f\right), \end{aligned}$$

where I_C denotes the $C \times C$ identity matrix, $\otimes$ is the Kronecker product operator, and $\Lambda(u)$ is a diagonal matrix whose elements are given by

$$\Lambda(u)_{n,n} = \frac{\rho'(|\Psi u(n)|,\sigma)}{|\Psi u(n)|}.$$

In practice, a small constant is generally added to the denominator of $\Lambda(u)_{n,n}$ to circumvent the need for consideration of subdifferentials. While continuation on this variable is also possible, we simply assign it statically to be $< 1 \times 10^{-4}$. Noting that Φ is both Hermitian and idempotent, consolidation of the target variable u yields

$$\left[\Psi^*\left(I_C \otimes \Lambda(u)\right)\Psi + \lambda\Phi\right]u = \lambda\Phi f_n.$$

Following the Lagged Diffusivity approach given in [37] for total variation minimization, the above can also be cast into fixed-point form and iterated to admit

$$u^{t+1} = B(u^t)^{-1}\lambda\Phi f_n,$$

where the Hessian approximate is given by

$$B(v) = \Psi^*\left(I_C \otimes \Lambda(v)\right)\Psi + \lambda\Phi.$$

When $\lambda\Phi f_n$ is defined by the recursion

$$\lambda\Phi f_n = B(u^t)u^t - L(u^t,\sigma,\lambda),$$

a robust quasi-Newton iteration for the computation of u,

$$u^{t+1} = u^t - B(u^t)^{-1}L(u^t,\sigma,\lambda),$$

is obtained. In our particular implementation, inversion of the Hessian approximate is performed via conjugate gradient (CG) iteration with a simple Jacobi preconditioner.

As discussed in Section II-D, reconstruction of undersampled MR images requires the analysis of complex data and a generalization of the described fixed-point solver is straightforward. For the problem described in (17), the associated unconstrained energy functional is given by

$$\begin{aligned} E(u,\sigma,\lambda) = &\sum_\Omega\left[\rho\left(|\Psi u_r(n)|,\sigma\right) + \rho\left(|\Psi u_i(n)|,\sigma\right)\right] \\ &+ \frac{\lambda}{2}\|\Phi u - \Phi f_n\|_2^2, \end{aligned}$$

where u_r and u_i again denote the real and imaginary components of u, respectively. Differentiation of $E(u, \sigma, \lambda)$ w.r.t u then yields the Cauchy-Riemann system

$$\begin{pmatrix} B(u_r) & i\lambda\Phi \\ \lambda\Phi & iB(u_i) \end{pmatrix} \begin{pmatrix} u_r \\ u_i \end{pmatrix} = \begin{pmatrix} I \\ I \end{pmatrix} \lambda\Phi f;$$

however, as this system is not holomorphic, no solution exists and only an approximation may be obtained. The least-squares approximate of the solution to the Cauchy-Riemann system is given by

$$\tilde{L}(u, \sigma, \lambda) = B(u_r)u_r + iB(u_i)u_i + \lambda\Phi u - 2\lambda\Phi f = 0.$$

Defining the complex Hessian approximate by

$$C(v) = B(\mathcal{R}v)\mathcal{R} + iB(\mathcal{I}v)\mathcal{I} + \lambda\Phi,$$

where $\mathcal{R}$ and $\mathcal{I}$ denote the real and imaginary operators, a quasi-Newton iteration for the complex homotopic L_0 minimization problem,

$$u^{t+1} = u^t - C(u^t)^{-1}\tilde{L}(u^t, \sigma, \lambda),$$

is obtained.

REFERENCES

[1] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] P. Mansfield, "Multiplanar image formation using nmr spin echos," *J. Physics C*, vol. 10, pp. 349–352, 1977.

[4] E. Haacke, E. Lindskog, and W. Lin, "A fast, iterative, partial-fourier technique capable of local phase recovery," *J. Magnetic Resonance*, vol. 92, pp. 125–146, 1991.

[5] D. Noll, D. Nishimura, and A. Macovski, "Homodyne detection in magnetic resonance imaging," *IEEE Trans. Medical Imaging*, vol. 10, no. 2, pp. 154–163, 1991.

[6] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[7] ——, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[8] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. of the 40th Annual Conference on Information Sciences and Systems*, 2006.

[9] E. Candés and J. Romberg, "Practical signal recovery from random projections," in *Proc. of the SPIE International Symposium on Electronic Imaging: Computational Imaging III*, 2005.

[10] E. Candés, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications in Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[11] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, 2007, in press.

[12] K. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint." *Magnetic Resonance in Medicine*, vol. 57, pp. 1086–1098, 2007.

[13] M. Lustig, J. Santos, D. Donoho, and J. Pauly, "k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. of the International Society for Magnetic Resonance in Medicine*, 2006.

[14] S. Kim, K. Koh, M. Lustig, and S. Boyd, "An interior-point method for large-scale l1-regularized least squares," in *Proc. of the IEEE Workshop on Statistical Signal Processing*, 2007.

[15] L. He, T. Chang, S. Osher, T. Fang, and P. Speier, "MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods," *UCLA CAM Reports*, vol. 06-35, 2006.

[16] j. Ye, S. Tak, Y. Han, and H. Park, "Projection reconstruction MR imaging using FOCUSS," *Magnetic Resonance in Medicine*, vol. 57, pp. 764–775, 2007.

[17] H. Jung, J. Ye, and E. Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Physics in Medicine and Biology*, vol. 52, pp. 3201–3226, 2007.

[18] R. Chartrand, "Exact reconstruction of sparse signal via nonconvex minimization," *IEEE Signal Processing Lettters*, vol. 14, no. 10, pp. 707–710, 2007.

[19] D. Geman and G. Reynolds, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[20] M. Nikolova and R. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Processing*, vol. 16, no. 6, pp. 1623–1627, 2007.

[21] E. Candés, N. Braun, and M. Wakin, "Sparse signal and image recovery from compressive samples," in *Proc. of the IEEE International Symposium on Biomedical Imaging*, 2007.

[22] R. Gribnoval and M. Nielsen, "On the strong uniqueness of highly sparse expansions from redundant dictionaries," in *Proc. of the International Conference on Independent Component Analysis*, 2004.

[23] M. Black, G. Sapiro, D. Marimon, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[24] P. Huber, *Robust Statistics*. Wiley, New York, 1981.

[25] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[26] J. Trzasko, A. Manduca, and E. Borisch, "Sparse MRI reconstruction via multiscale l0 continuation," in *Proc. of the IEEE Workshop on Statistical Signal Processing*, 2007.

[27] T. Chan, H. Zhou, and R. Chan, "Continuation method for total variation denoising," *UCLA CAM Reports*, vol. 95-28, 1995.

[28] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.

[29] A. Delaney and Y. Bresler, "Globally convergent edge-preserving reguarlized reconstruction: an application to limited-angle tomography," *IEEE Trans. Image Processing*, vol. 7, no. 2, pp. 204–231, 1998.

[30] D. Yu and J. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *IEEE Trans. Medical Imaging*, vol. 21, no. 2, pp. 159–173, 2002.

[31] J. Wood and K. Johnson, "Wavelet-packet denoising of magnetic resonance images: importance of Rician statistics at low SNR," *Magnetic Resonance in Medicine*, vol. 41, pp. 631–635, 1999.

[32] J. Fessler and D. Noll, "Iterative image reconstruction in MRI with separate magnitude and phase regularization," in *IEEE International Symposium on Biomedical Imaging*, 2004.

[33] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, pp. 952–962, 1999.

[34] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits and System*, vol. 24, no. 11, pp. 652–654, 1977.

[35] H. Hu, A. Madhuranthakam, D. Kruger, J. Glockner, and S. Riederer, "Combination of 2D sensitivity encoding and 2D partial Fourier techniques for improved acceleration in 3D contrast-enhanced MR angiorgaphy," *Magnetic Resonance in Medicine*, vol. 55, pp. 16–22, 2005.

[36] A. Raj, G. Singh, R. Zabih, B. Kressler, Y. Wang, N. Schuff, and M. Weiner, "Bayesian parallel imaging with edge-preserving priors," *Magnetic Resonance in Medicine*, vol. 57, pp. 8–21, 2007.

[37] C. Vogel, *Computational Methods for Inverse Problems*. SIAM, 2002.

[38] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE J. Selected Topics in Signal Processing*, 2007, in press.

[39] E. van den Berg and M. Friedlander, "In pursuit of a root," *UBC Computer Science Technical Reports*, vol. TR-2007-16, 2007.

[40] E. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *Rice University CAAM Technical Reports*, vol. TR07-07, 2007.

[41] E. Candés, "The uniform uncertainty principle," in *IMA Short Course: Compressive Sampling and Frontiers in Signal Processing*, 2007.

[42] P. Boufounos, M. Duarte, and R. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Proc. of the IEEE Workshop on Statistical Signal Processing*, 2007.