Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic ℓ_0-Minimization

Joshua Trzasko*, *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 oversampling above the theoretical minimum sampling rate to guarantee that exact reconstruction can be achieved. In this paper, we propose a generalization of the CS paradigm based on homotopic approximation of the ℓ_0 quasi-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 CS methods and the developed theoretical extensions, several example MRI reconstructions from highly undersampled K-space data are presented.**

Index Terms—**Compressed sensing, compressive sensing (CS), image reconstruction, magnetic resonance imaging (MRI), nonconvex optimization.**

I. INTRODUCTION

IN CONTEMPORARY clinical practice, magnetic resonance imaging (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 3-D 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, signal-to-noise ratio (SNR)].

The emerging theory of compressive (or compressed) sensing (CS) [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 (see Fig. 1 for an example). 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-E along with both its applied and theoretical implications. Moreover, a practical scheme is presented for addressing the ℓ_0 quasi-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.

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 successfully 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.
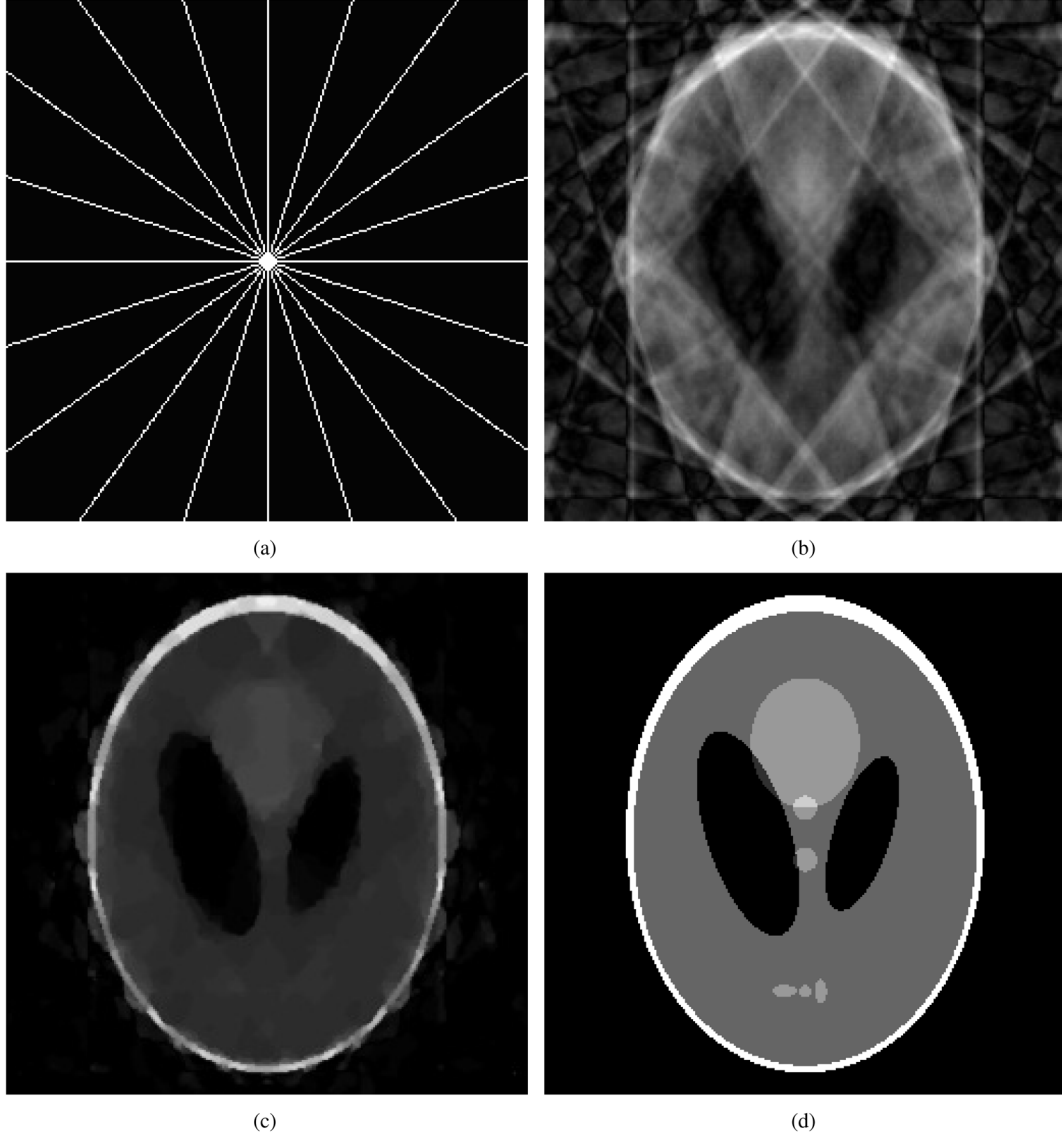
Fig. 1. A comparison of ℓ_1-minimization versus homotopic ℓ_0-minimization based reconstructions of the canonical Shepp–Logan phantom. Given a 96% undersampled K-space sampling mask containing only 10 radial lines (a) and its associated minimum energy or zero-filled reconstruction (b), ℓ_1-minimization (c) fails to recover the undersampled image while the proposed homotopic ℓ_0-minimization (d) achieves exact reconstruction.

An ideal sampler of a continuous signal is described by the periodic impulse train or Dirac comb

$$s_X(x) = \sum_{n=\infty}^{\infty} \delta(x - nX) = \frac{1}{X} \sum_{n=\infty}^{\infty} e^{\frac{-2\pi jnx}{X}} \quad (1)$$

where δ is the Dirac delta (generalized) function and X is the sampling period; the Fourier series expansion of the Dirac comb is also shown. Given (1), f can be sampled by simply taking the entrywise product between itself and the sampling vector, i.e., $f_X(x) = f(x) \cdot s_X(x)$. Computing the Fourier Transform of the sampled signal then yields

$$\hat{f}_X(k) = \frac{1}{X} \sum_{n=\infty}^{\infty} \hat{f}\left(k - \frac{n}{X}\right) \quad (2)$$

where k is a spectral index. Resultantly, the one-sided bandwidth of f must be less than $(1/2X)$ 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 $T_N = (1/2k_{max})$, 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

$$\Delta k = \frac{1}{N\Delta x} = \text{FOV}^{-1} \quad (3)$$

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 3-D 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. Let $\mathcal{F}$ denote the forward Fourier transform operator. A simple technique for choosing a solution, u, from this infinite set is to select the one with the minimum energy (ℓ_2) by solving

$$\min_u \|u\|_2^2 \quad \text{s.t. } \Phi u = \Phi f \tag{4}$$

where the measurement matrix $\Phi = \hat{\Phi}\mathcal{F}$ and $\hat{\Phi}$ is the characteristic function of the subset of measured *K*-space values. f is again the signal of interest. It is trivial to show that (4) is achieved by simply setting the complement of $\hat{\Phi}$ 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 SNR levels. Alternatively, if the signal of interest is assumed to be strictly real, Hermitian symmetry of *K*-space can be exploited 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 B_0 inhomogeneities, chemical shift, and magnetic susceptibility distortions 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 via 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 the Shepp–Logan phantom shown in Fig. 2. 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

$$\min_u \|\Psi u\|_0 \quad \text{s.t. } \Phi u = \Phi f \tag{5}$$

where, for some vector v, the zero quasi-norm, $\|v\|_0$, is a measure of functional cardinality defined as follows:

$$\|v\|_0 = \sum_{x \in \Omega} \mathbf{1}(|v(x)| > 0) \tag{6}$$

where Ω is the image domain and $\mathbf{1}$ is the indicator function. Although Ψ is often assigned to be a unitary operator (e.g., wavelet or discrete cosine transform), note that this is not a required condition. For instance, in variational models, Ψ may instead be a column vector of finite directional difference matrices analogous to the gradient operator. Redundant transformations and tight frames such as curvelets [6] have also recently been investigated [7].

Given knowledge of how to generate sparse or compressible representations for a specific signal class via some transform operator, the next logical question to ask is under what sampling conditions the signal can be accurately reconstructed Without loss of generality, consider the case where Ψ is equal to the identity operator, which can arise either from a signal of interest having inherent sparsity in its given domain or by shifting the problem formulation in (5) from analysis to synthesis form. The aforementioned question then resorts to identifying specific conditions on the operator Φ that assures the solution given by (5) is identically f.

We consider this construction using the restricted isometry property (RIP) of Candès and Tao [8]. Suppose a signal, f, resides on a set of spatial indices denoted by Ω. An $N \times |\Omega|$
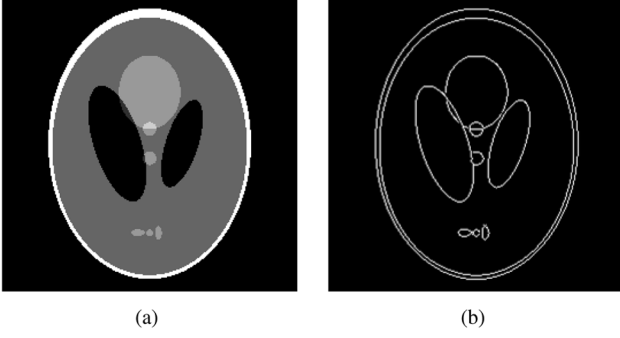
Fig. 2. The Shepp–Logan phantom (256×256) is a prime example of an image that posseses a sparse representation in a transform domain, here the magnitude of its gradient. While the image space representation (a) is clearly not sparse ($\|x\|_0 \approx 61\% \cdot N$), the transform representation of the image (b) exhibits very high sparsity ($\|\nabla x\|_0 \approx 3\% \cdot N$).

measurement matrix, Φ, is said to be an S-restricted isometry if, $\{\forall f \in \mathbb{R}^{|\Omega|} | \| f \|_0 \leq S\}$ and $N \geq S$

$$(1 - \delta_S) \| f \|_2^2 \leq \| \Phi f \|_2^2 \leq (1 + \delta_S) \| f \|_2^2 \qquad (7)$$

where $\delta_S \in (0, 1)$ is the restricted isometry constant. Satisfaction of (7) assures that no S-sparse signal is in the null space of Φ and thus inversion of the measurement process is feasible. Consider the $N \times |K|$ submatrix of Φ, Φ_K, where $K = \mathrm{supp}\{f\}$ and $|K| \leq S$. Complimentary to the null space argument, (7) provides a bound on the eigenvalues of the Gram matrix $\Phi_K^* \Phi_K$, where * denotes the complex conjugate operator. This in turn allows the condition number of the active system to be expressed by $\kappa(\Phi_K) = \sqrt{(1 + \delta_S / 1 - \delta_S)}$, demonstrating the benefit of utilizing measurement matrices with small restricted isometry constants. Furthermore, a simple proof by contradiction employing (7) asserts that if Φ satisfies $\delta_{2S} < 1$, (5) will have a unique minimizer [8] or, equivalently, unique recovery of an S-sparse signal via ℓ_0-minimization is possible even if as few as $2S$ measurements are acquired. Unlike the condition imposed by Shannon's theorem, the circumstances under which signal recovery by ℓ_0-minimization is possible are independent of dimensionality and instead linearly proportional to the underlying complexity of a signal.

While the general construction of matrices satisfying (7) is beyond the scope of this paper, the particular case of random Fourier measurement ensembles will be investigated in Section II-D. To date, RIP has been verified for several classes of random measurement ensembles (see [9] and [10] for a detailed discussion) and there is substantial active research (e.g., [11], [12]) into the construction of deterministic and structured matrices obeying (7). There is also compounding empirical evidence that suggests RIP holds for many practical deterministic measurement paradigms employed in medical imaging [1], [13], [14].

Despite the profound theoretical implications that arise for measurement matrices obeying (7), directly solving the ℓ_0-minimization problem to recover an S-sparse signal residing on Ω requires a combinatorial search through all $\binom{|\Omega|}{S}$ potential solutions and is thus intractable for most practical application. The

remarkable result of CS theory, pioneered by Candès *et al.* [1] and Donoho [2], is that if one replaces the ℓ_0 quasi-norm prior in (5) with the ℓ_1 norm, namely

$$\min_u \|\Psi u\|_1 \quad \text{s.t. } \Phi u = \Phi f \qquad (8)$$

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 (5) and (8) are equivalent. Moreover, (8) is a tractable convex optimization problem. In practice, the degree of oversampling required to achieve exact reconstruction of a signal of cardinality S via ℓ_1-minimization is roughly $3S - 5S$ [15], or about twice what is theoretical necessary for the analogous ℓ_0 case. Nonetheless, this approach offers the ability to sample drastically below the Nyquist rate and perform the signal reconstruction in a computationally practical manner.

C. Robust Compressive Sensing

In most medical imaging applications including MRI, signals are rarely noise-free and it is desirable to investigate how measurement error affects the stability of the reconstruction program. Suppose that the measured data is given by $f_n = f + n$, where n is a noise process such as additive white Gaussian noise (AWGN) found on each of the quadrature measurement channels in MRI. Addressing the potential for measurement error, (8) can be reformulated as

$$\min_u \|\Psi u\|_1 \quad \text{s.t. } \|\Phi u - \Phi f_n\|_2^2 \leq \epsilon \qquad (9)$$

where ϵ is a statistic describing the magnitude of the error. For AWGN processes observed in ℓ_2, ϵ is generally taken as the noise variance. Alternatively, ϵ may be subjectively defined as the maximum allowable error in the generated approximation. Again refering to the exemplary case $\Psi = I$, the solution to (9) will exhibit approximatory error on the order of $\|u - f\|_2 \leq C \cdot \epsilon$, where the multiplicative constant C depends only on the restricted isometry constant [16]. For a fixed measurement matrix, Φ, the accuracy of approximation for a signal reconstruction by ℓ_1-minimization thus degrades gracefully in the presence of increasing levels of noise and the reconstruction process is both robust and stable.

Additionally, most medical images do not possess truly sparse representations in any domain as nearly all transform coefficients are of nonzero value. Fortunately, many images are compressible in the sense that image information is strongly captured by a small subset of these coefficients. For example, the dominating morphological features of many anatomical MR images such as angiograms can be closely approximated by sparse models generated with transforms promoting piecewise smoothness, e.g., bounded variational models and many types of wavelets. For the case $\Psi = I$, let f_S denote the signal constructed from only the S-largest coefficients of f. In [16], it was shown that when a dense signal is reconstructed according to (8), the approximation error will be bounded as $\|u - f\|_1 \leq C \cdot \|f - f_S\|_1$, where C is again only dependent on the restricted isometry constant. Consequently, for a compressive signal, most of the information will nonetheless be

captured during the measurement process and the approximation will be on par with the best possible S-term approximation given full *a priori* specification of the signal.

Following the theoretical development of CS within the 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 3-D MR angiography (MRA) [13], [17], and time-resolved imaging [18]. More recently, several groups in the MRI community have proposed novel numerical techniques for attacking (9) with specific focus on image reconstruction including nonlinear conjugate gradient (CG) [13], interior point (IP) [19], Bregman iteration or inverse scale space [20], and iterative reweighted least squares or FOCUSS [14], [21] methods. As there has not been an extensive comparison of the computational performance of these techniques on large-scale problems to date, which numerical approach, if any, will best meet clinical demands is still an open question.

D. Extension to ℓ_p $(0 < p \leq 1)$

In Section II-B, it was stated that fewer measurements are necessary to exactly reconstruct a given signal via ℓ_0-minimization than by ℓ_1-minimization. While a direct solution of (5) 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 feasible. Following assertions made in [2] and that (assuming $0^0 = 0$) the zero quasi-norm is commonly defined as

$$\|v\|_0 = \lim_{p \to 0} \sum_{x \in \Omega} |v(x)|^p \qquad (10)$$

Chartrand [22] proposed use of the ℓ_p semi-norm $(0 < p \leq 1)$ class of sparsity priors within a basis pursuit reconstruction framework and showed asymptotic convergence of the ℓ_p sampling bounds towards ℓ_0 using RIP. Although Donoho [2] dismisses the practical use of ℓ_p semi-norm priors as numerical achievement of the global minimum can no longer be guaranteed as a consequence of their nonconvexity, these functionals (unlike the ℓ_0 prior) admit a nonzero gradient almost everywhere and 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. In [22], Chartrand offers several examples of computed "local" minima providing signal and image reconstructions that are exact to within numerical precision at sampling rates well below those achievable by ℓ_1-based methods.

Consider the ℓ_p recovery problem

$$\min_u \|\Psi u\|_p^p \quad \text{s.t. } \Phi u = \Phi f \qquad (11)$$

where $0 < p \leq 1$. Extending on [22] for $\Psi = I$, a sufficient condition for exact recovery of a signal of cardinality S via (11) is given by

$$\delta_{2M} < \left[1 + \sqrt{\frac{2M}{S}} \cdot \left(\frac{S}{M} \right)^{\frac{1}{p}} \right]^{-1} (0 < p \leq 1) \qquad (12)$$

for any $M \geq S$ (see Appendix for derivation). Next, suppose Φ is an $N \times |\Omega|$ matrix comprised of randomly selected rows (uniform probability) of the discrete Fourier transform (DFT) matrix defined on Ω; by definition, N denotes the number of measurements performed by applying Φ to a signal. For any $\gamma > 0$, Rudelson and Vershynin [10] showed that if

$$N \geq C \left(\frac{2M \log |\Omega|}{\gamma^2} \right) \log \left(\frac{2M \log |\Omega|}{\gamma^2} \right) \log^2 2M \qquad (13)$$

for some absolute constant C

$$\mathbb{E} \delta_{2M} \leq \gamma \qquad (14)$$

with high probability, where $\mathbb{E}$ denotes the expected value operator. When γ is equated to the right side of (12), it follows that decreasing p implies an analogous decrease in the required number of measurements, N, needed to achieve exact reconstruction. As anticipated, the use of diminishing values of p allows a signal to be exactly reconstructed from fewer and fewer measurements. Consider again the 256×256 Shepp–Logan phantom showed in Fig. 2, where $S \approx 2200$ for the gradient magnitude transformed signal. For this example case, Fig. 3 portrays the relaxation of the recovery conditions under diminishing values of p for a typically chosen partitioning $M = (3/2)S$ (e.g., [8], [15]).

Chartrand's work and the extension given in (12) provide an exact characterization of the performance of reconstructions employing the ℓ_p semi-norms $(0 < p \leq 1)$, but the results are not at all surprising—utilizing approximations that are closer to the ℓ_0 quasi-norm than ℓ_1 correspondingly lower the required number of measurements necessary to accurately reconstruct a signal. When considering p to be a static variable, as done in [22], an open problem exists on how to assign this variable. While relatively smaller values of p offer the potential for reconstruction of a signal from fewer measurements, this gain comes at the cost of increasing numerical stiffness. Which value of p offers the optimal tradeoff between these properties has yet to be determined.

E. Compressive Sensing With General Concave Priors

The clear advantage of migrating from ℓ_1 to ℓ_p reconstruction form suggests the following question: can other functionals be used in place of ℓ_1 to offer similar advantages as ℓ_p for the sparse reconstruction problem? Reconsider the limit definition of the ℓ_0 quasi-norm given in (10). While the ℓ_p functional is certainly the most popular choice for defining ℓ_0 as a sequence limit, there exist many possible functionals, generically denoted by ρ, that satisfy

$$\lim_{\sigma \to 0} \sum_{x \in \Omega} \rho(|u(x)|, \sigma) = \sum_{x \in \Omega} \mathbf{1}(|u(x)| > 0) \qquad (15)$$
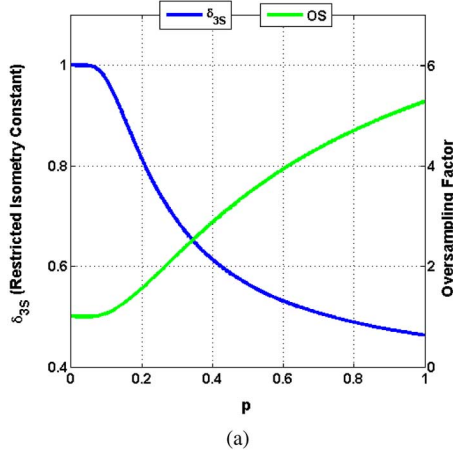
Fig. 3. As the ℓ_p sparsity prior is evolved towards ℓ_0, the restricted isometry constant, $\delta_{3S}(M = (3/2)S$, is progressively relaxed. Given (13) and defining the expected oversampling factor required for exact reconstruction by $(N(p)/N(0))$, a decrease in required number of samples towards the theoretical minimum sampling rate limit can be seen as $p \to 0$.
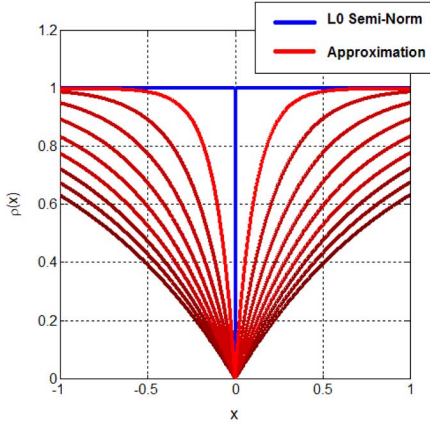

Fig. 4. Homotopic approximation of the ℓ_0 semi-norm using the Laplace error function, $1 - e^{(|u|/\sigma)}$, shown at diminishing values of σ.

where σ represents a sequence index or scale parameter. As will be discussed, only a small subclass of such functionals can be practically employed for the sparse reconstruction problem.

A key motivation for employing the ℓ_p limit sequence given in (10) is that every sequence element can be independently and successfully used as a prior for the sparse reconstruction problem and still yield a unique and exact solution. Consequently, there is a sense of stability for the entire solution path, not only at the sequence limit $\sigma = 0$. Consider the general class of semimetric functionals, i.e., positive definite and zero-symmetric, that are nondecreasing over $\mathbb{R}^+$. These functionals provide a measure of deviation of the function of interest from zero and as $\sigma \to 0$, by definition, they provide a progressively refined measure of signal sparsity that approaches that yielded by the ℓ_0 quasi-norm.

In the event that ρ is strictly concave over $\mathbb{R}^+$, the functional is inherently subadditive and thus a true metric. It can then be shown that, for any fixed value of σ, a unique and exact solution to the sparse reconstruction is possible (see Appendix), analogous to the results of Section II-D. For example, it can be shown

that many robust error norms and other redescending *M*-estimators [23] such as the Laplace function

$$\rho(|u(x)|, \sigma) = 1 - e^{-\frac{|u(x)|}{\sigma}} \quad (16)$$

shown in Fig. 4, the Geman–McClure function [24]

$$\rho(|u(x)|, \sigma) = \frac{|u(x)|}{|u(x)| + \sigma} \quad (17)$$

and the concave logarithmic penalty

$$\rho(|u(x)|, \sigma) = \log\left(\frac{|u(x)|}{\sigma} + 1\right) \quad (18)$$

all yield exact signal reconstructions for any $\sigma \in [0, \infty)$. In fact, if a signal can be exactly recovered by ℓ_1-minimization, it is guaranteed recoverable via any ρ that is nondecreasing and concave over $\mathbb{R}^+$. A proof of this claim can be found in the Appendix. Furthermore, for the above examples as well as many others, if ρ is scaled by a constant factor such that $\rho(1, \sigma) = 1$, $\sigma \to \infty$ admits $\rho(|u|, \sigma) \to |u|$. In other words, ρ interpolates the function space between ℓ_1 and ℓ_0 across $\sigma \in [0, \infty)$ in the same manner as does ℓ_p for $p \in [0, 1]$.

As anticipated and will be discussed in the proceeding sections, when σ is progressed towards zero, signals are observed to be recoverable from many fewer samples than via ℓ_1-based approaches when using priors such as (16)–(18); e.g., the result in Fig. 1(d) was achieved using the Laplace prior defined in (16). Although an exact formulation of the restricted isometry constants for general priors as a function of σ has yet to be determined, based on the interpolatory property of the described functionals and convincing empirical evidence (see Section IV), it is conjectured that a tight relation analogous to (12) given for the ℓ_p priors also exists.

Following the brief example in Fig. 1 where the Shepp–Logan phantom is shown to be recoverable from only 10 radial Fourier projections corresponding to roughly 4% of the data, Fig. 5 shows a comparison of results generated by homotopic ℓ_0 minimization and ℓ_1-minimization for a sequence of different sampling rates. The minimum-energy results obtained via zero-filling are also depicted. As expected, given a sufficient sampling rate (in this case, 18 or more Fourier projections), the results obtained by ℓ_1-minimization and homotopic ℓ_0 minimization are identical and both exact. As the sampling rate is further reduced, however, the ℓ_1-based reconstruction quickly begins to degrade whereas the homotopic ℓ_0 reconstructions continue to be exact all the way down to the case of only 10 Fourier projections. Beyond this point, neither the proposed or ℓ_1-based method are able to exactly reconstruct the phantom image although homotopic ℓ_0 continues to provide superior results.

For a tabulation of additional non-convex functionals that have potential application for sparse recovery, see [25]. We also note that Candès *et al.* [26] have recently introduced a reweighted ℓ_1 minimization scheme for CS which implicitly minimizes (18). Additionally, several authors have proposed the use of everywhere-differentiable quasi-convex priors satisfying (15) (e.g., the Tukey biweight and Gaussian error function) for
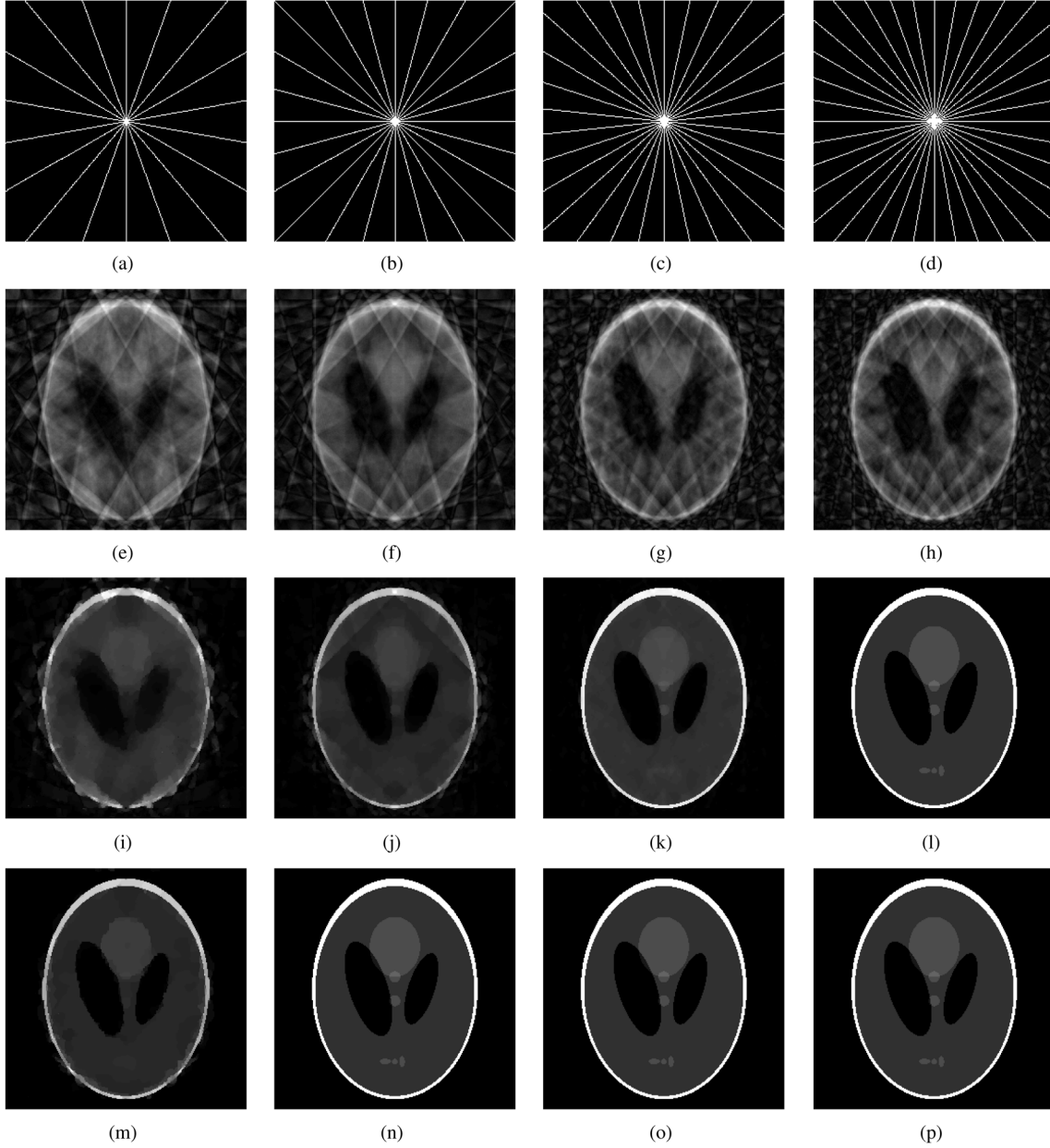
Fig. 5.   The practical benefit of homotopic ℓ_0 is elucidated when viewing a sequence of images recovered under progressively increasing sampling rates. Considering radial sampling using (a) 9, (b) 12, (c) 15, and (d) 18 equally spaced Fourier projections corresponding to approximate undersampling rates of 96%, 95%, 94%, and 93%, respectively. (e)–(h) The minimum energy reconstruction of (a)–(d) obtained by zero-filling. Similarly, (i)–(l) are the reconstructions obtained via ℓ_1-minimization and (m)–(p) are the reconstructions obtained via homotopic ℓ_0-minimization using the Laplace error function. While the ℓ_1 and homotopic ℓ_0 reconstructions are identical under 18 projections, ℓ_1-minimization fails to provide exact reconstructions from any fewer number of projections whereas homotopic ℓ_0 minimization achieves exact reconstructions even in the 12 projection example (and from as few as 10 projections, as shown in Fig. 1). Even in the case of nine projections where neither method yields exact signal recovery, the homotopic ℓ_0 result is clearly superior to its ℓ_1 analog.

problems including noise removal [23], [27], [28] superresolution [29], [30] and recently medical image reconstruction [31]–[33]. As discussed in the Appendix, strict concavity over $\mathbb{R}^+$ of the prior functional is key to asserting solution uniqueness for the sparse reconstruction problem. Smooth quasi-convex functionals contain a small convex well about the origin which inherently violates the concavity condition and guaranteeing uniqueness under RIP is no longer straightforward. Although quality reconstructions are possible using quasiconvex regularizers, the migration to strictly concave

priors generally offers both improved practical performance as well as theoretical guarantees of recoverability.

F. Continuation Approach to ℓ_0-Minimization

In practice, computationally-challenging problems such as the ℓ_0-minimization problem are often attacked by a procedure known as homotopy or continuation. Under this scheme, a related but simplified problem is first solved and its solution is used as the warm-start initializer for a slightly harder problem that is closer to the target. This process is then repeated until

the target problem is reached or the solution becomes stationary [34]. While the optimality of this paradigm cannot, in general, be theoretically guaranteed, the observed practical gain is often substantial. In many cases, the solution to the simple problem lies relatively close in the function space to the solution of the hard problem. As the algorithm progresses, this extremal point is then tracked to the solution of the hard problem, avoiding many of the local extrema that would be encountered if the hard problem was attacked directly. If the exact solution of the hard problem is not achieved, the found local minimum is often far more satisfactory than that found without the continuation process.

We note that several authors have successfully employed everywhere-differentiable deformable approximators of non-convex target priors for image recovery and, in particular, limited-view tomography [31], [32]. Most of these approaches are based on the graduated nonconvexity (GNC) method developed by Blake and Zisserman [35] for solving the shape-from-shading problem in computer vision. We give a similar paradigm based on continuation of deformable concave metrics towards the ℓ_0 quasi-norm for undersampled MRI reconstruction within the CS setting.

For any particular concave metric ρ satisfying (15), consider the sequence of problems given by

$$u^{t+1} = \arg\min_{u^t} \sum_{x \in \Omega} \rho(|u^t(x)|, \sigma^t) \text{s.t.} \quad \|\Phi u^t - \Phi f\|_2^2 \leq \epsilon \tag{19}$$

where $\sigma^{t+1} < \sigma^t$. We call this process *homotopic ℓ_0-minimization* as the subproblems asymptotically approach the ℓ_0-minimization problem as $t \to \infty$ and thus $\sigma^t \to 0$. Practically, the solution to subproblem t is used as the initial estimate of the solution to subproblem $t+1$. An efficient, semi-implicit numerical solver for (19) will be presented in Section III and practical assignment of (σ_{t+1}/σ_t) will also be discussed.

Following discussions in Sections II-D and II-E, when $\sigma > 0$, ρ is continuous and possesses nonzero gradient on at least some subset of $\mathbb{R}$ and thus admits a weak Euler–Lagrange equation for which extrema can be readily computed using standard descent methods. It has also been shown in the previous sections that there exist conditions on Φ such that each subproblem of (19) is independently capable of sparse signal reconstructing and thus viable reconstructions are possible even under early termination of the sequence. For example, in Fig. 3 it was demonstrated that, for the ℓ_p prior case, the required number of samples is effectively stationary for $p < 0.1$. Consequently, it may occur that, as p is progressively reduced, the generated solution does not change under further reduction of p.

We briefly address the question of which functional prior, ρ, is optimal. As long as ρ satisfies the concavity and metric conditions described in Section II-E, (19) should, in theory, be invariant to the particular functional specification as all choices will asymptotically reach the ℓ_0 quasi-norm. In practice, we have consistently observed very similar results obtained by reconstructing a signal under different prior functionals and thus,

at this time, the optimal choice of ρ has not yet been investigated further. Several example reconstructions employing different sparsity priors can be seen in Figs. 5–8. In terms of computational efficiency, the optimal choice of ρ is still an open question that we hope to address in a future work.

G. Reconstructing Complex Sparse Signals

Multidimensional spatial localization in MRI relies on the provision of both frequency and phase atomic encoding, yielding complex signals which must be measured in quadrature. Following the discussion in Section II-A, if the signal of interest is dominantly real, a phase correction step such as homodyne detection [5] can be incorporated directly into the reconstruction process to avoid the need for complex signal analysis as well as offer the potential for improved performance [13]. In many scenarios such as velocity encoding of blood flow, phase information is crucial to image formation and the Hermitian assumption simply cannot be used. One must instead handle the data in its complex form.

In [20], 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., [36]). We also adopt this complex generalization of the sparsity prior, namely

$$\sum_{x \in \Omega} \rho(|\Psi u(x)|, \sigma) \to \sum_{x \in \Omega} \rho(|\Psi \mathcal{R} u(x)|, \sigma) + \rho(|\Psi \mathcal{I} u(x)|, \sigma) \tag{20}$$

where $\mathcal{R}$ and $\mathcal{I}$ denote the real and imaginary operators. We note that separate regularization of image-space magnitude and phase components is also possible [37] and will be investigated in a future work.

H. Brief Note on Multicoil MRI

A major advance in the MRI community came with the advent of parallel or multicoil imaging techniques. Multicoil methods such as SENSE [38], although offering accelerated acquisitions, can also be classified as "Nyquist-limited" as they are governed by Papoulis' generalized sampling expansion theorem [39] which essentially states that, given C orthogonal encoders, a signal can be exactly reconstructed if it is sampled at a rate above $(1/C)$ 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 coherence exists between sensitivity profiles, the degree of which tends to increase with the number of coils. Consequently, the anticipated $(1/C)$ speedup 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 [40].

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.* [38] formulates the reconstruction as an unregularized least-squares approximation, namely

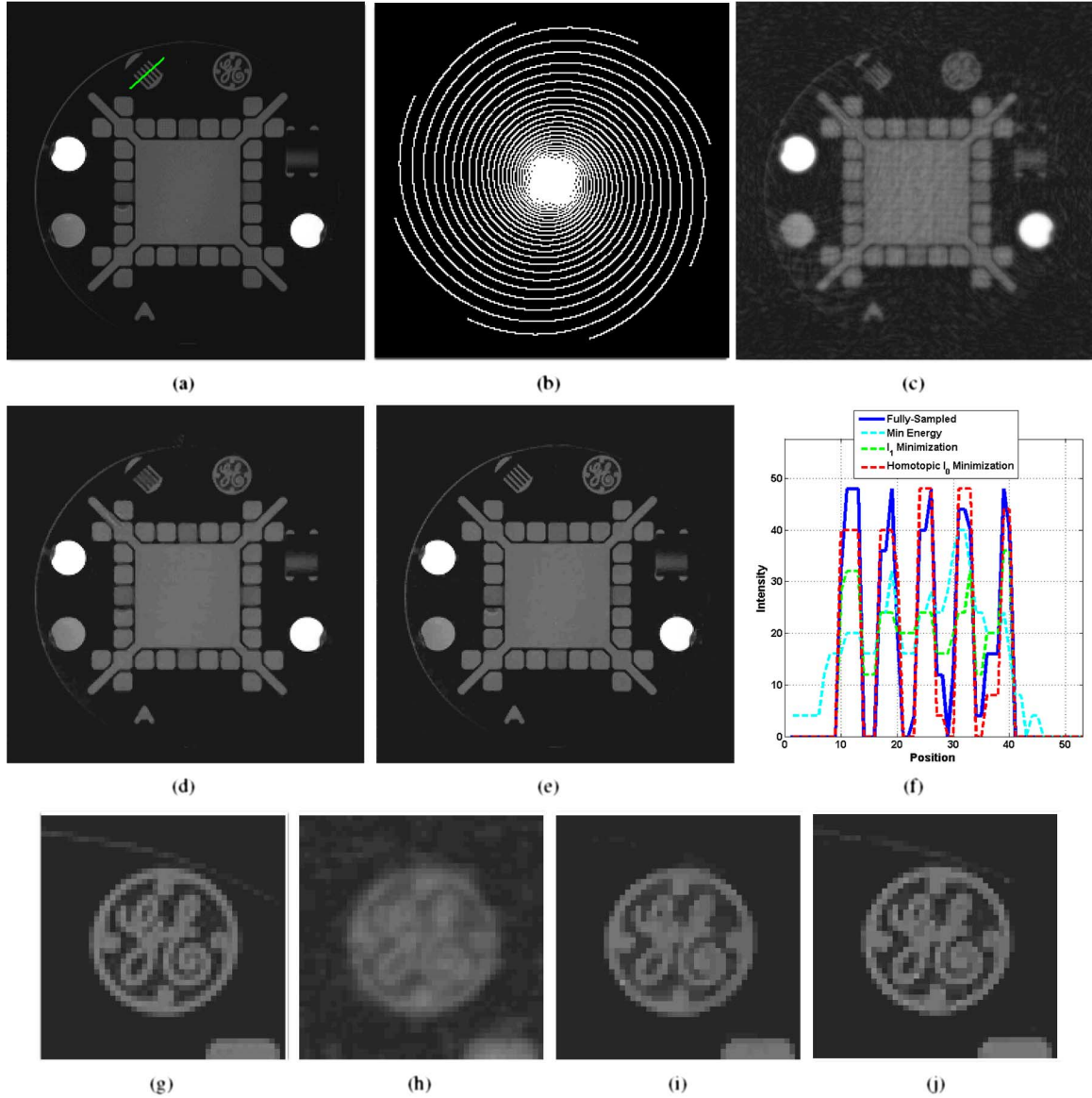$$\min_u \|\Gamma u - \Gamma f_n\|_2^2 \tag{21}$$

Fig. 6. Example: (a) standard General Electric resolution phantom, (b) simulated k-space trajectory (Cartesian approximation of multishot spiral, 86% 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.

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 (21). 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.* [41] employed nonconvex image priors similar to those discussed in Section II-E to the parallel MRI denoising problem using a formulation nearly identical to one subproblem of (19) but with σ fixed; a graph cut procedure was then used for minimization of the energy functional. More recently, several authors [13], [17] 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 nonlinear regularization in parallel MRI reconstruction, extension of the proposed reconstruction paradigm in Section II-E to this paradigm is straightforward. A more detailed discussion will be presented in a separate work.

III. METHODS

In this section, we discuss the practical implementation of the *homotopic ℓ_0-minimization* process given in Section II-F. An efficient fixed-point method for solving each complex subproblem of (19) is first derived and one approach to embedding the scheme within the continuation procedure is proposed. Algorithm parameter selection is also briefly discussed.
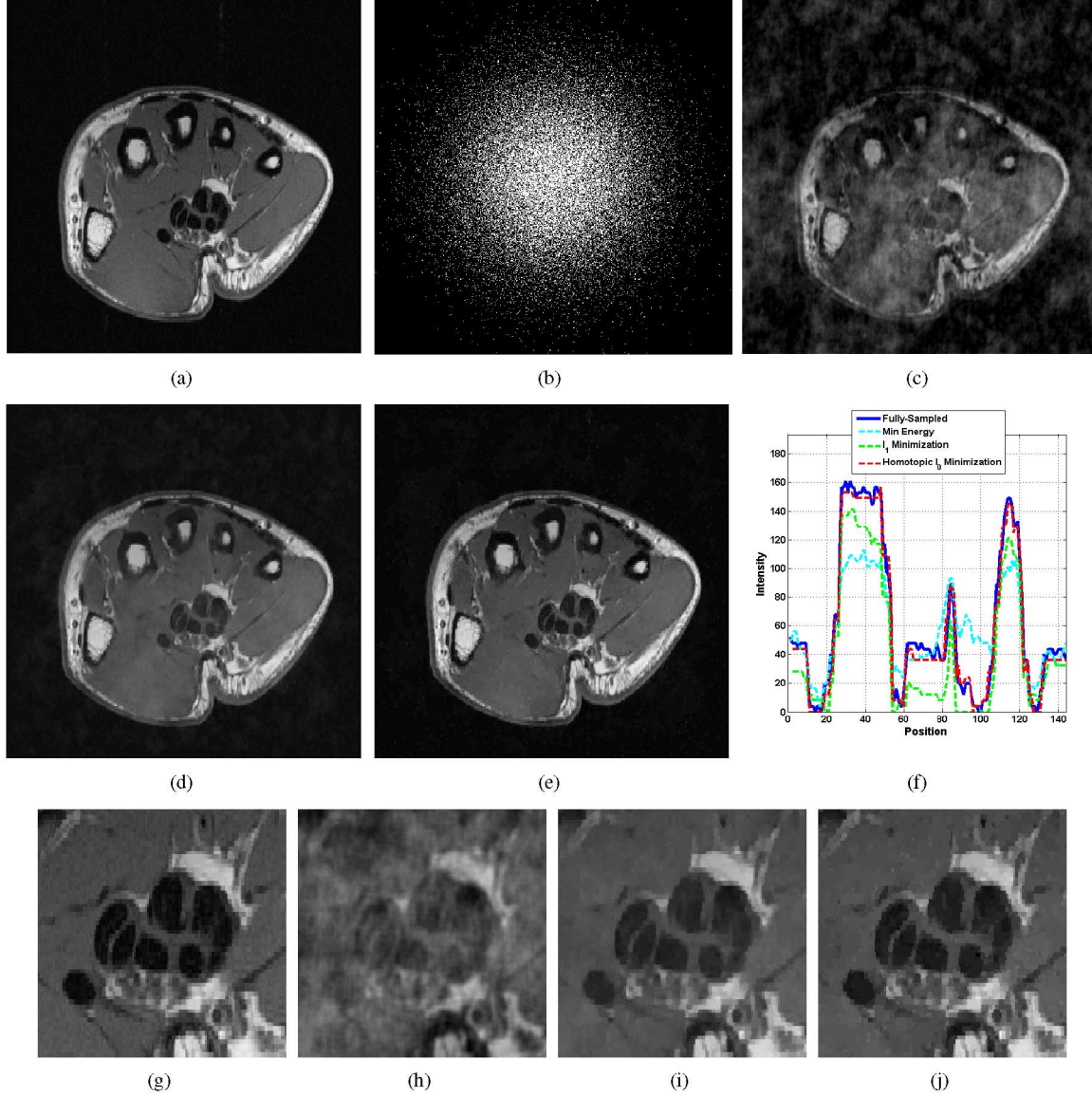
Fig. 7. Example: (a) axial T1-weighted image of the wrist, (b) simulated k-space trajectory (variable density random, 87% undersampling), (c) minimum-energy solution via zero-filling, (d) reconstruction by ℓ_1 minimization, (e) reconstruction by homotopic ℓ_0 minimization using $\rho(|\nabla u|, \sigma) = \log((|\nabla u|/\sigma + 1))$, (f) line profile across the second and third metacarpals, (g)–(j) enlargements of (a), (c)–(e), respectively.

A. Fixed-Point Method for Sparse Image Reconstruction

In many instances in optimization theory, constrained problems are converted to analogous unconstrained form prior to analysis to allow the use of more general and intuitive numerical solvers such as descent-based methods. It is this approach we will use to address the iterative reconstruction problem defined in (19). Consider the energy functional given by

$$E(u, \sigma, \lambda) = \sum_{x \in \Omega} \rho(|\Psi u(x)|, \sigma) + \frac{\lambda}{2} \|\Phi u - \Phi f\|_2^2 \qquad (22)$$

where λ controls the tradeoff between solution sparsity and data fidelity. For each subproblem of (19), an analogous unconstrained problem is given by

$$v = \arg \min_u E(u, \sigma, \lambda). \qquad (23)$$

When the prior functional ρ is convex, the Kuhn–Tucker theorem [42] asserts that there exists a parameter $\lambda \geq 0$ such that minimizers of (22) and its constrained analog are equivalent. Similarly, for the nonconvex case, λ can be chosen such that the so-called duality gap between the two problems is minimized and the solutions will be arbitrarily close to one another. 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 [43]; in this work, only the former approach is taken.

Given the functional definition in (22), minima of E are yielded as solutions of the associated Euler–Lagrange equation

$$L(u, \sigma, \lambda) = \Psi^* \Lambda(u) \Psi u + \lambda \Phi^* (\Phi u - \Phi f) = 0 \qquad (24)$$
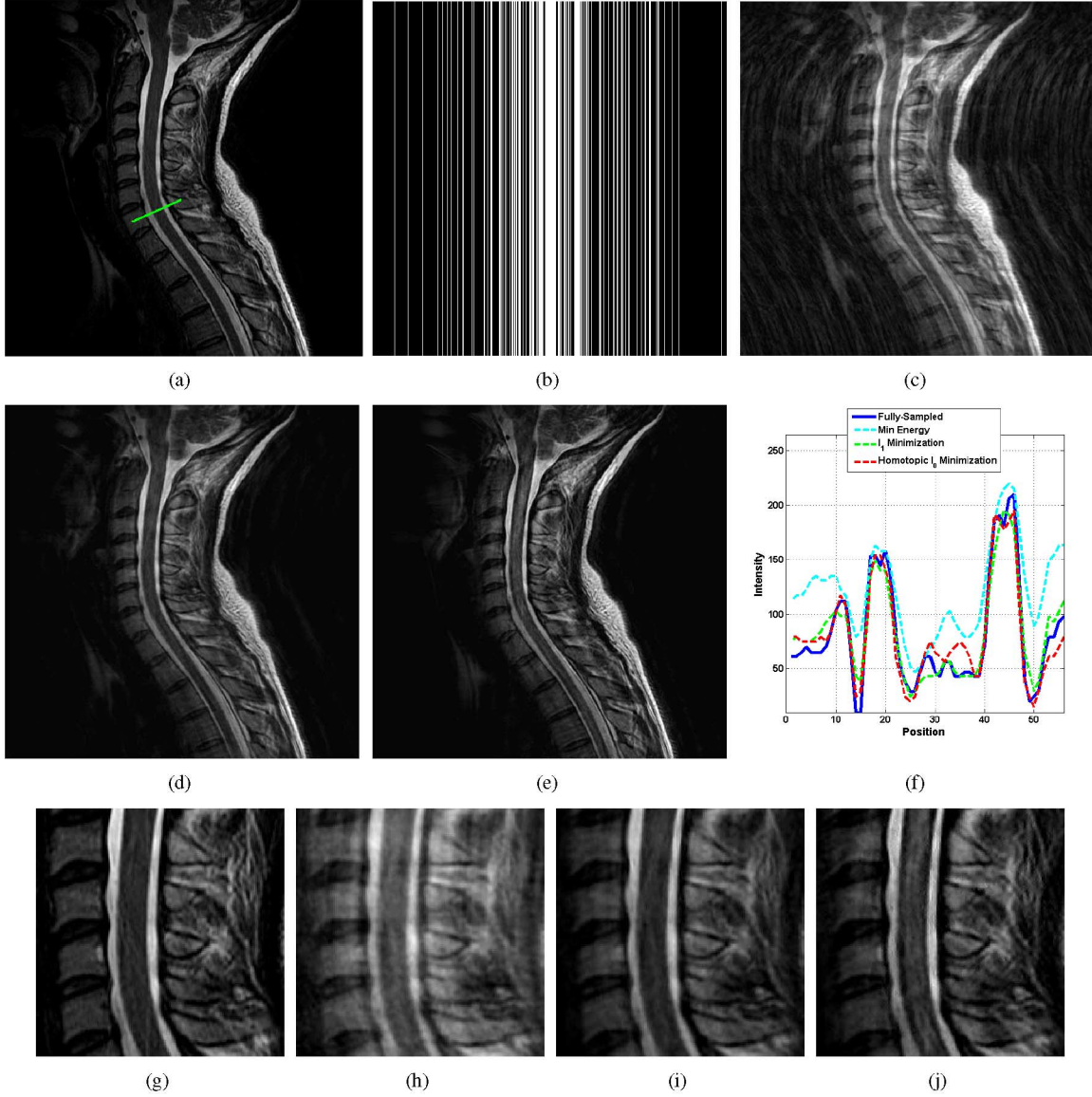
Fig. 8. Example: (a) Sagittal T2-weighted image of the spine, (b) simulated k-space trajectory (random phase encoding, 77% undersampling), (c) minimum-energy solution via zero-filling, (d) reconstruction by ℓ_1 minimization, (e) reconstruction by homotopic ℓ_0 minimization using $\rho(|\Psi u|, \sigma) = (|\Psi u|/|\Psi u| + \sigma)$ where Ψ is the discrete Curvelet transform, (f) line profile across C6, (g)–(j) enlargements of (a), (c)–(e), respectively.

where the diagonal matrix, $\Lambda(u)$, is defined by

$$\Lambda(u)_{x,x} = \frac{\rho'(|\Psi u(x)|, \sigma)}{|\Psi u(x)|}. \tag{25}$$

As ρ is, by construction, continuous but nondifferentiable at the origin, the derivative, ρ' is taken in the weak sense, e.g., a weak derivative of $\rho(|u|) = |u|$ is given by the signum function, $\rho'(|u|) = \text{sgn}(u)$. In practice, a small constant, ϵ is often added to the denominator of (25) for increased numerical stability, e.g., $|\Psi u(x)| \rightarrow \sqrt{|\Psi u(x)|^2 + \epsilon}$. When the sparsifying operator Ψ is a scalar transformation, e.g., a wavelet, its application within Λ is straightforward. For vectorial cases such as the 2-D discretized gradient operator given by

$$\Psi = \begin{bmatrix} \mathcal{D}_x \\ \mathcal{D}_y \end{bmatrix} \tag{26}$$

Ψ will transform u from $\Omega \times 1$ to $C\Omega \times 1$, where C is the dimensionality of the vector transform. For the example in (26), $C = 2$. Consequently, $\Lambda(u)$ must be $C\Omega \times C\Omega$. To rectify this discrepancy, $\Lambda(u)$ can be generalized as

$$\Lambda(u) \rightarrow I_C \otimes \Lambda(u) \tag{27}$$

where I_C is a $C \times C$ identity matrix and $\otimes$ is the Kronecker product.

Consolidation of the target variable u yields

$$[\Psi^* \Lambda(u)\Psi + \lambda\Phi]u = \lambda\Phi^*\Phi f. \tag{28}$$

Additionally, let

$$B(u) = \Psi^* \Lambda(u)\Psi + \lambda\Phi^*\Phi \tag{29}$$

denote the Hessian approximate. Following the Lagged Diffusivity approach given in [44] for TV minimization which is it-

self a special case of the half-quadratic regularization [24], [25], (28) can be cast into fixed-point form and iterated as

$$u^{t+1} = B^{-1}(u^t)\lambda\Phi^*\Phi f. \tag{30}$$

Moreover, when $\lambda\Phi f$ is defined by the recursion

$$\lambda\Phi^*\Phi f = B(u^t)u^t - L(u^t, \sigma, \lambda) \tag{31}$$

a robust quasi-Newton iteration

$$u^{t+1} = u^t + \Delta^t \tag{32}$$

is obtained for the computation of u, where

$$B(u^t)\Delta^t = -L(u^t, \sigma, \lambda). \tag{33}$$

In our particular implementation, inversion of the Hessian approximate implicit to (33) is performed via conjugate gradient (CG) iteration.

As discussed in Section II-G, reconstruction of undersampled MR images requires the analysis of complex data. Fortunately, such a generalization of the described fixed-point solver is straightforward. Following the discussion in Section II-G, a complex generalization of the unconstrained energy functional in (22) is given by

$$E(u, \sigma, \lambda) = \sum_{x \in \Omega} [\rho(|\Psi\mathcal{R}u(x)|, \sigma)$$
$$+ \rho(|\Psi\mathcal{I}u(x)|, \sigma)] + \frac{\lambda}{2}\|\Phi u - \Phi f\|_2^2 \tag{34}$$

where $\mathcal{R}$ and $\mathcal{I}$ again denote the real and imaginary operators. Differentiation of $E(u, \sigma, \lambda)$ with respect to u then yields the Cauchy–Riemann system

$$\begin{bmatrix} B(\mathcal{R}u)\mathcal{R} & i\lambda\Phi^*\Phi \\ \lambda\Phi^*\Phi & iB(\mathcal{I}u)\mathcal{I} \end{bmatrix} \begin{bmatrix} u \\ u \end{bmatrix} = \lambda \begin{bmatrix} \Phi^*\Phi f \\ \Phi^*\Phi f \end{bmatrix} \tag{35}$$

however, as this system is overdetermined, the least-squares approximate of the solution is computed. Denoting the complex Hessian approximate as

$$C(u) = B(\mathcal{R}u)\mathcal{R} + iB(\mathcal{I}u)\mathcal{I} + \lambda\Phi^*\Phi \tag{36}$$

and

$$\tilde{L}(u, \sigma, \lambda) = B(\mathcal{R}u)\mathcal{R}u + iB(\mathcal{I}u)\mathcal{I}u + \lambda\Phi^*\Phi u - 2\lambda\Phi^*\Phi f \tag{37}$$

a quasi-Newton iteration for the complex homotopic ℓ_0 minimization problem is given by

$$u^{t+1} = u^t + \Delta^t \tag{38}$$

where

$$C(u^t)\Delta^t = -\tilde{L}(u^t, \sigma, \lambda). \tag{39}$$

The iterative equation given in (38) will serve as the update step in the continuation algorithm that is given in the next section. In particular, (38) will be solved once per inner iteration of the homotopy continuation scheme described in (19).

B. Practical Employment of Homotopic ℓ_0-Minimization

In the previous section, an iterative numerical procedure for solving each complex subproblem of (19) was developed. We now show how this method can be embedded in a simple two-level algorithm to realize the continuation process described in Section II-F.

For some fixed value of σ, a minimum of the complex energy function can be found by placing (38) into a simple while-loop that is terminated only when a global measure of the difference between sequential iterations falls below some level $\text{tol}_{\text{inner}}$. When a (potentially local) minimum is found, σ can then be scaled by a factor $\beta \in (0, 1)$ and the process repeated until the solution becomes numerically stationary. Analogous to $\text{tol}_{\text{inner}}$, a threshold $\text{tol}_{\text{outer}}$ can also be defined such that the algorithm exits when some global measure of the difference between sequential solutions falls below this level. Denote u as the current solution estimate, u_{inner} as the last solution estimate obtained under the same value of σ, and u_{outer} the last solution estimate obtained under a different value of σ. The resulting algorithm is then given by

---

Algorithm 1. Pseudocode for Homotopic ℓ_0 Minimization

$u = \Phi^*\Phi f, \sigma \gg 0, \beta \in (0, 1)$

while $((\|u - u_{\text{outer}}\|_2/\|u_{\text{outer}}\|_2) \geq \text{tol}_{\text{outer}})$

    $u_{\text{outer}} = u, \sigma = \sigma \times \beta$

    while $((\|u - u_{\text{inner}}\|_2/\|u_{\text{inner}}\|_2) \geq \text{tol}_{\text{inner}})$

        $u_{\text{inner}} = u$

        Solve $C(u)\Delta = -\tilde{L}(u, \sigma, \lambda)$ for Δ

        via CG iteration.

        $u = u + \Delta$

    end

end

Future advances in numerical methods for the subproblems of (19) can be naturally incorporated into this scheme by modifying or replacing the inner loop off Algorithm 1, thus preserving the similar progression of diminishing σ values.

C. Some Remarks on Parameter Selection

Algorithm 1 contains several parameters which influence the solution path generated during the iterative procedure.

All parameter value ranges given will be under the assumption that the measured signal has been normalized such that $\max_\Omega\{|\Phi f(x)|\} = 1$.

First consider the tolerance parameter $\mathrm{tol_{inner}}$ used to determine when an estimate for one subproblem of (19) has reached numerical steady-state. As the sequence of problems utilizes progressively smaller values of σ and thus become numerically more stiff as the iterative process progresses, $\mathrm{tol_{inner}}$ should be set small enough that the algorithm advances only when no more substantial work can be done at a particular σ level. We typically employ $\mathrm{tol_{inner}} < 1 \times 10^{-2}$ such that the relative net image change is less than 1% and have found that this strict condition helps minimize propagation of error during the continuation process. Analogously, $\mathrm{tol_{outer}}$ should be set several orders of magnitude below $\mathrm{tol_{inner}}$ as too high a value will lead to premature termination of the recovery process. Typically, we set $\mathrm{tol_{outer}} \approx 10^{-3} \times \mathrm{tol_{inner}}$

As discussed in Section II-F, a large initial value of σ is crucial to the efficacy of Algorithm 1. As σ is roughly proportional to the width of the zero well of the sparsity prior, for unity-scaled images we generally employ $\sigma \gg 1$; however, for the $\ell_p(0 < p \le 1)$ case σ is always initially set to unity. Given a starting value of σ, β determines the rate of deformation of the prior term as exemplified in Fig. 4. Qualitatively, $\beta \in (0, 1)$ should balance rapid advancement of the algorithm while maintaining assurance that solutions of successive problems are adequately close. For the ℓ_p, we have found that $(\sigma_{t+1}/\sigma_t) \in (0.90, 0.95)$ works well for most examples. For all other cases where σ can range between 0 and ∞, we generally employ $(\sigma_{t+1}/\sigma_t) \approx (\sqrt{10}/10)$ or $(\sigma_{t+1}/\sigma_t) = 0.5$.

D. Computational Performance

Using parameterizations as outlined in the previous section, the presented numerical method is competitive with contemporary methods for CS signal reconstruction. In general, fewer than outer 40 iterations are required by Algorithm 1 to achieve sufficient results and, in many cases, far less are needed. We also note that, in all cases, the maximum number of internal CG iterations used for inversion of the Hessian approximate was limited to 250 with an early termination tolerance of 1×10^{-2}. On a standard PC with a 3 GHz Intel Pentium IV processor and 3 GB memory, a 256×256 complex MR image reconstructed using a variational sparsifying transformation runs on average at roughly 5–10 s per outer iteration or roughly 1–3 min total execution time using a Matlab implementation. Depending on the particular implementation, wavelet and related transforms may take longer to process. While still not necessarily optimal for widespread clinical usage, these execution times are comparable to the fastest existing methods known for large-scale ℓ_1-based reconstructions.

IV. EXAMPLES

Three example complex MR images reconstructed from undersampled k-space data using the presented homotopic ℓ_0 minimization scheme can be seen in Figs. 6–8. In each case, the fully sampled Cartesian k-space data (a) was undersampled retrospectively by the given binary mask (b) to allow a direct visual comparison of the generated reconstruction against the true image. Although only Cartesian approximations of the sampling paradigms such as radial and spiral are considered in this paper to demonstrate robustness to varying sampling patterns, the extension to true non-Cartesian k-space acquisition can be naturally embedded in this framework through the use of a regridding process or nonuniform Fourier transform and has been addressed in the context of ℓ_1 minimization in [13]. (c)–(e) shows 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 is shown in (f) 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.

In the three given examples, $\mathrm{tol_{inner}} = 1 \times 10^{-2}$ and $\mathrm{tol_{outer}} = 1 \times 10^{-4}$. Additionally, for the ℓ_p prior case in Fig. 6, $\beta = 0.9$, and for the other two examples, we used $\beta = (\sqrt{10}/10)$ as discussed in Section III-C. In each given example, the regularization parameter λ was hand-tuned to provide optimal performance for both the ℓ_1 and homotopic ℓ_0 examples although we note that the relative improvement of this fine-tuning step over simply setting λ statically to a large value was minimal as the performance of both algorithms appears to be fairly consistent for a wide range of large values of λ. For both the ℓ_1 and homotopic ℓ_0-minimization examples in all three given cases, $\lambda \in (1 \times 10^5, 1 \times 10^6)$. The automation of λ selection will be addressed in a future study.

For the standard General Electric resolution phantom (256×256) shown in Fig. 6, 86% undersampling (roughly a 7X reduction in acquisition time) by a Cartesian approximation of multishot (eight interleaves) variable-density spiral sampling 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 such as on the upper left-hand spoke of the center object as well as a slight blurring of the resolution bars (see Fig. 6(g)–(j) and GE logo). In contrast, the homotopic ℓ_0 reconstruction provides a crisper reconstruction of object edges here and, in general, provides greater intensity fidelity to the fully-sampled image.

Similarly, Fig. 7 shows a T1-weighted axial cross section of a wrist (512×512) undersampled by 87% using a variable density random mask akin to that described in [13]. As with the preceding example, note that homotopic ℓ_0-minimization is able to provide a more accurate reconstruction of image contrast and sharper anatomical depiction than ℓ_1-minimization as exemplified in the enlargement of the carpal tunnel region of the wrist [Fig. 7(g)–(j)]. In both of these cases, gradient-based priors were employed to emphasize the recovery of morphological information. As this transformation promotes sparse edge representations, fine-scale texture information is often lost and the images may appear somewhat piecewise constant or "blocky." When only morphology is of diagnostic interest, however, this characteristic is not a concern.

Fig. 8 depicts the reconstruction of a T2-weighted sagittal image of the spine (512×512). In this case, 1-D random phase

encoding along the anterior–posterior (AP) direction simulating 77% undersampling was utilized. The sparsifying operator was the fast discrete curvelet transform (FDCT) [6] whose semi-local behavior (as opposed to the strictly local gradient operator employed in the previous examples) offers the benefit of strong edge representation without blocking artifacts albeit at the cost of added computational complexity. In this example, both ℓ_1 and homotopic ℓ_0 reconstructions offer strong suppression of aliasing artifacts. However, the tendency of ℓ_1 minimization for uniform shrinkage of transform coefficients leads to a global smoothing of the image while coefficient hard thresholding associated with ℓ_0 (and thus homotopic ℓ_0 minimization) instead offers strong edge retention. This phenomena is especially evident in the enlargement of the cervical spine [Fig. 8(g)–(j)]. Additionally, the inherent over-completeness of curvelet frames minimizes the presence of spiking artifacts typically observed when thresholding orthogonal wavelet coefficients.

The ℓ_1-minimization procedure is generally very powerful and the relatively inferior quality of the examples presented using this approach is due to the fact that the degree of under-sampling was intentionally chosen to lie above the rate at which ℓ_1-minimization provided favorable results for the given image. The slower rate of degradation of the homotopic ℓ_0 method as a function of sampling rate was exemplified in Fig. 5. When the sampling rate is sufficiently raised, following [1], 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 minimum as does its convex ℓ_1 analog, the computed local minimum of the homotopic ℓ_0 minimization problem typically allows for accurate image reconstructions at higher undersampling rates than are achievable via ℓ_1-minimization. Moreover, we have demonstrated 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
Conditions for Exact Signal Recovery With General Concave Metric Priors

Consider the generalized signal recovery problem for the case $\Psi = I$, namely

$$\min_u \sum_{x \in \Omega} \rho(|u(x)|) \quad \text{s.t.} \ \Phi u = \Phi f \qquad (40)$$

where f and u are again the true and approximated signals defined on Ω. As described in Sections II-D and II-E, ρ is a metric that is concave and nondecreasing over $\mathbb{R}^+$. Before deriving the conditions under which exact signal recovery is possible, three lemmas are required.

Lemma 1: Let $T_0 = \text{supp}\{f\}$ and $T_0^c = \Omega \backslash T_0$. Additionally, define the reconstruction error $h = u - f$. A sufficient condition for the uniqueness a solution, u, of (40) is given by

$$\sum_{x \in T_0} \rho(|h(x)|) \geq \sum_{x \in T_0^c} \rho(|h(x)|) \quad \text{s.t.} \ \Phi h = 0. \qquad (41)$$

Proof: This proof essentially follows the approach taken in [1] for the ℓ_1 problem. A solution, u, of the minimization problem in (40) will be unique if

$$\sum_{x \in \Omega} \rho(|f(x)|) \geq \sum_{x \in \Omega} \rho(|u(x)|) \quad \text{s.t.} \ \Phi u = \Phi f.$$

Noting the subadditivity of concave functions over $\mathbb{R}^+$

$$\sum_{x \in T_0} \rho(|f(x)|) = \sum_{x \in \Omega} \rho(|f(x)|) \geq \sum_{x \in \Omega} \rho(|u(x)|)$$
$$= \sum_{x \in \Omega} \rho(|f(x) + h(x)|)$$
$$= \sum_{x \in T_0} \rho(|f(x) + h(x)|)$$
$$+ \sum_{x \in T_0^c} \rho(|h(x)|)$$
$$\geq \sum_{x \in T_0} \rho(|f(x)|) - \sum_{x \in T_0} \rho(|h(x)|)$$
$$+ \sum_{x \in T_0^c} \rho(|h(x)|).$$

(41) immediately follows. ∎

Lemma 2: With no effect on (41), reorder the elements of $h_{T_0^c}$ according to magnitude (descending). Next, partition $h_{T_0^c}$ into sequential subsets of size $M \geq K$ and denote these as h_{T_j}, where j is an index. By definition, $|h(s)| \geq |h(t)|, \forall s \in T_j, t \in T_{j+1}$. Also note that, without consequence, the last partitioned subset might be of cardinality smaller than M. Following enumeration of h by descending magnitude

$$\rho\left(M^{-\frac{1}{2}} \sum_{j \geq 2} \|h_{T_j}\|_2\right) \leq M^{-1} \sum_{x \in T_0^c} \rho(|h(x)|). \qquad (42)$$

Proof: As ρ is monotone nondecreasing over $\mathbb{R}^+$, $\rho(|h(s)|) \geq \rho(|h(t)|), \forall s \in T_j, t \in T_{j+1}$. By construction

$$\min_{T_j}\{\rho(|h(x)|)\} \geq \max_{T_{j+1}}\{\rho(|h(x)|)\}$$

which in turn implies

$$M^{-1} \sum_{x \in T_j} \rho(|h(x)|) \geq \max_{T_{j+1}}\{\rho(|h(x)|)\} = \rho(\|h_{T_{j+1}}\|_\infty).$$

By the equivalence of norms over finite domains, $M^{-(1/2)}\|\cdot\|_2 \leq \|\cdot\|_\infty$ and thus, again by monotonicity

$$\rho(\|h_{T_{j+1}}\|_\infty) \geq \rho\left(M^{-\frac{1}{2}}\|h_{T_{j+1}}\|_2\right).$$

Summing over all subsets of T_0^c and utilizing the subadditivity of ρ yields

$$M^{-1} \sum_{x \in T_0^c} \rho(|h(x)|) \geq \rho\left(M^{-\frac{1}{2}} \sum_{j \geq 2} \|h_{T_j}\|_2\right).$$

∎

Lemma 3: Suppose Φ satisfies the RIP (7) with $\delta_{2M} \in [0, 1)$. The uniqueness condition in Lemma 1 then resorts to

$$\frac{M}{S} \sum_{x \in T_0} \rho\left(\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\sqrt{\frac{S}{M}}|h(x)|\right) \leq \sum_{x \in T_0} \rho(|h(x)|). \tag{43}$$

Proof: From [45], clever manipulation of the inner products inherent to the restricted isometry identity in (7) yields the relation

$$\|h_{T_0}\|_2 \leq \frac{\sqrt{2}\delta_{2M}}{1-\delta_{2M}} \sum_{j \geq 2} \|h_{T_j}\|_2 \tag{44}$$

under the assumption $\Phi h = 0$. Letting $S = |T_0|$, (44) resorts to

$$\sqrt{\frac{1}{SM}}\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\|h_{T_0}\|_1 \leq M^{-\frac{1}{2}} \sum_{j \geq 2} \|h_{T_j}\|_2$$

by simple algebraic manipulation. Incorporating Lemmas 1 and 2

$$\rho\left(\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\sqrt{\frac{S}{M}} \cdot S^{-1}\|h_{T_0}\|_1\right) \leq M^{-1} \sum_{x \in T_0} \rho(|h(x)|)$$

and by Jensen's inequality

$$\frac{M}{S} \sum_{x \in T_0} \rho\left(\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\sqrt{\frac{S}{M}}|h(x)|\right) \leq \sum_{x \in T_0} \rho(|h(x)|).$$

∎

Given Lemmas 1–3, conditions under which exact signal recovery is possible via (40) when using general concave metric priors will now be given. Additionally, a tight bound for the specific case of ℓ_p priors ($0 < p \leq 1$), given in (12) will be derived.

Theorem 1: Let the measurement operator Φ be a restricted isometry with

$$\delta_{2M} < \left(1 + \sqrt{\frac{2M}{S}}\right)^{-1}. \tag{45}$$

For any concave metric ρ, a signal can be recovered exactly via (40).

Proof: A solution, u, of (40) will be unique if it satisfies (43) in Lemma 3. Furthermore, u will be identically f only if $h = 0, \forall x \in \Omega$. By definition, $M \geq S$ and thus

$$\sum_{x \in T_0} \rho\left(\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\sqrt{\frac{S}{M}}|h(x)|\right) \leq \sum_{x \in T_0} \rho(|h(x)|).$$

As ρ is nondecreasing for positive arguments, $\rho(a|h|) \leq \rho(|h|)$ for any $a \in [0, 1]$. Conversely, if $a > 1$, this relation does not hold unless $h = 0$. If

$$\left(\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)\sqrt{\frac{S}{M}} > 1$$

(43) is not satisfied unless $h(x) = 0, \forall x \in \Omega$. Moreover

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

assures that exact recovery is feasible.   ∎

For the particular case of $S = M$, (45) resorts to $\delta_{2S} < \sqrt{2} - 1 \approx 0.4142$, which is exactly the condition for guaranteed recovery by ℓ_1-minimization given in [45]. Consequently, any signal that can be recovered by ℓ_1-minimization is also recoverable by (40). While this condition assures the feasibility of exact signal recovery via (40) and is as strong as what is known for ℓ_1-based methods, the bound is not tight and does not convey any benefit of utilizing smaller values of σ.

When considering the case of the ℓ_p ($0 < p \leq 1$) prior class, the bound in Theorem 1 can be analytically tightened by a considerable amount and clearly depicts the benefit of utilizing small values of p. This is exactly the relationship portrayed in Fig. 3. The proof for the ℓ_p case relies on the separability of power functions and involves straightforward manipulation of (43). For the generalized case, ρ is not necessarily separable and, to date, we have yet to derive an exact analytical bound that is variable with σ. Nonetheless, based on empirical evidence including the examples in Section IV as well as preliminary numerical simulation of (43), we conjecture that a similar σ-dependent bound can be derived for the generalized case.

Theorem 2: Let the measurement operator Φ be a restricted isometry with

$$\delta_{2M} < \left[1 + \sqrt{\frac{2M}{S}} \cdot \left(\frac{S}{M}\right)^{\frac{1}{p}}\right]^{-1} \tag{46}$$

for some $p \in (0, 1]$. For the given value of p, a signal can exactly be recovered via ℓ_p-minimization as posed in (11).

Proof: By the separability of $|\cdot|^p$ for any p, (43) trivially yields

$$\frac{M}{S}\left(\sqrt{\frac{S}{M}}\frac{1-\delta_{2M}}{\sqrt{2}\delta_{2M}}\right)^p \|h_{T_0}\|_p^p \leq \|h_{T_0}\|_p^p.$$

If δ_{2M} satisfies

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

the uniqueness condition can only be true if $h(x) = 0, \forall x \in \Omega$. Consequently, signal recovery is exact. The above condition is met only if δ_{2M} admits to (46). ∎

Unlike the general condition given in Theorem 1, for the case ℓ_p, δ is clearly relaxed towards unity as $p \to 0$. As discussed in Section II-D, such a relaxation allows for a reduction of the required number of measurements necessary to guarantee exact signal recovery.

ACKNOWLEDGMENT

REFERENCES

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

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

[3] P. Mansfield, "Multiplanar image formation using NMR spin echos," *J. Phys. 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. Magn. Reson.*, vol. 92, pp. 125–146, 1991.

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

[6] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[7] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2006.

[8] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[9] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[10] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, 2006, pp. 207–212.

[11] R. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, no. 4–6, pp. 918–925, Aug.–Dec. 2007.

[12] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE Workshop Stat. Signal Process.*, Aug. 2007, pp. 294–298.

[13] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 59, no. 6, pp. 1182–1195, Dec. 2007.

[14] J. Ye, S. Tak, Y. Han, and H. Park, "Projection reconstruction MR imaging using FOCUSS," *Magn. Reson. Med.*, vol. 57, pp. 764–775, 2007.

[15] E. Candès and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Int. Symp. Electron. Imag.: Computational Imag. III*, San Jose, CA, Jan. 2005, pp. 76–86.

[16] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.

[17] K. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint.," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007.

[18] M. Lustig, J. Santos, D. Donoho, and J. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. ISMRM*, Seattle, WA, May 2006, p. 2420.

[19] S. Kim, K. Koh, M. Lustig, and S. Boyd, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Selected Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[20] 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 Rep., 2006.

[21] H. Jung, J. Ye, and E. Kim, "Improved k-t blast and k-t SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, pp. 3201–3226, 2007.

[22] R. Chartrand, "Exact reconstruction of sparse signal via nonconvex minimization," *IEEE Signal Process. Lettt.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

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

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

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

[26] E. Candès, N. Braun, and M. Wakin, "Sparse signal and image recovery from compressive samples," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Arlington, VA, Apr. 2007, pp. 976–979.

[27] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

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

[29] Y. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. J. Comput. Vis.*, vol. 3, no. 1, pp. 73–102, 1989.

[30] N. Saito, "Superresolution of noisy band-limited data by data adaptive regularization and its application to seismic trace inversion," in *Proc. ICASSP*, Albuquerque, NM, Apr. 1990, pp. 1237–1240.

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

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

[33] J. Trzasko, A. Manduca, and E. Borisch, "Sparse MRI reconstruction via multiscale l0 continuation," in *Proc. IEEE Workshop Stat. Signal Process.*, Aug. 2007, pp. 176–180.

[34] J. Nocedal and S. Wright, *Numerical Optimization*. New York: Springer, 2006.

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

[36] J. Wood and K. Johnson, "Wavelet-packet denoising of magnetic resonance images: Importance of Rician statistics at low SNR," *Magn. Reson. Med.*, vol. 41, pp. 631–635, 1999.

[37] J. Fessler and D. Noll, "Iterative image reconstruction in MRI with separate magnitude and phase regularization," in *IEEE Int. Symp. Biomed. Imag.*, Apr. 2004, pp. 209 –212.

[38] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, pp. 952–962, 1999.

[39] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. 24, no. 11, pp. 652–654, Nov. 1977.

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

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

[42] D. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Boston, MA: Kluwer, 2003.

[43] P. Boufounos, M. Duarte, and R. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Proc. IEEE Workshop Stat. Signal Process.*, Aug. 2007, pp. 299–303.

[44] C. Vogel, *Computational methods for inverse problems*. Philadelphia, PA: SIAM, 2002.

[45] E. Candès, "The uniform uncertainty principle," presented at the IMA short course: Compressive sampling and frontiers in signal processing, Minneapolis, MN, Jun. 2007.