

Relaxed Conditions for Sparse Signal Recovery with General Concave Priors

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

Abstract—The emerging theory of Compressive or Compressed Sensing challenges the convention of modern digital signal processing by establishing that exact signal reconstruction is possible for many problems where the sampling rate falls well below the Nyquist limit. Following the landmark works of Candès et al. and Donoho on the performance of ℓ_1 -minimization models for signal reconstruction, several authors demonstrated that certain nonconvex reconstruction models consistently outperform the convex ℓ_1 -model in practice at very low sampling rates despite the fact that no global minimum can be theoretically guaranteed. Nevertheless, there has been little theoretical investigation into the performance of these nonconvex models. In this work, a notion of weak signal recoverability is introduced and the performance of nonconvex reconstruction models employing general concave metric priors is investigated under this model. The sufficient conditions for establishing weak signal recoverability are shown to substantially relax as the prior functional is parameterized to more closely resemble the targeted ℓ_0 -model, offering new insight into the empirical performance of this general class of reconstruction methods. Examples of relaxation trends are shown for several different prior models.

Keywords: Compressive Sensing, Compressed Sensing, Restricted Isometry Property, Signal Recovery, Weak Recoverability.

I. INTRODUCTION

For over half a century the fundamental doctrine of signal processing has been Shannon's theorem, which states that any continuous signal sampled at a rate exceeding twice its maximal frequency will be devoid of aliasing. By construction, given fixed spatial or temporal resolution, increasing the number of elements of a digital signal requires an identical increase in the number of measurements taken, regardless of how much new information is actually introduced. Compressive or Compressed Sensing (CS) is a nascent sampling theorem asserting that the number of measurements truly required to accurately reconstruct a signal is, in fact, independent of dimensionality and instead solely proportional to the degree of underlying information. As a consequence, uncomplicated signals can be recovered from very few measurements even if the Nyquist criterion is never satisfied.

One approach to mathematically representing the notion of simplicity is sparsity. Consider a discrete, bounded signal $f \in \mathbb{R}^N$ such that $\Omega = \text{dom}\{f\}$ and $N = |\Omega|$. f is said to be S -sparse if $\|f\|_0 = S$, where the ℓ_0 -norm (technically not a

true norm due to lack of homogeneity)

$$\|f\|_0 = \sum_{x \in \Omega} \mathbf{1}(|f(x)| \neq 0). \quad (1)$$

If $S \ll N$, the pertinent information of the signal is concentrated on a very small (albeit potentially unknown) subset of the signal domain. While it is noted that many signals are sparse only in some transform domain (e.g. wavelet), without loss of generality only intrinsically sparse signals are considered in this work. For a detailed discussion on the relationship between intrinsic and transform sparsity models, see [1].

The process of linear sampling can be modeled as application of Φ , a $K \times N$ measurement matrix with bounded elements, to the target signal, f , resulting in a K -element measurement vector, $g = \Phi f$. Unlike direct sampling where Φ would simply be diagonal, CS applications typically employ dense measurement matrices such as range-limited discrete Fourier transforms (DFT) for Magnetic Resonance Imaging (MRI) [2], [3], [4] or Bernoulli/Rademacher [5] random matrices for single pixel digital cameras.

When $K < N$, the reconstruction problem is underdetermined and one must choose a solution from the infinitely large set of signals that could have potentially yielded g when sampled. If f happens to be sparse, one could simply discriminate a solution by selecting the sparsest signal from the infinite set. Formally, this search known as the ℓ_0 -minimization problem is defined as

$$\min_u \|u\|_0 \quad \text{s.t.} \quad \Phi u = g. \quad (2)$$

As will be discussed in Section II, under certain conditions on Φ , f can be recovered exactly as the solution of (2). Despite the profound implications of this claim, ℓ_0 -minimization is formally NP-hard [6] and solving this problem is computationally infeasible in practice.

As the gradient of (1) is zero almost everywhere (a.e.), even local minima of (2) are often evasive due to the inapplicability of descent-based methods. The catalyst of the CS revolution was the observation by Candès et al. [2] and Donoho [7] that, under more stringent conditions on Φ than required by (2), exact signal recovery can also be guaranteed under the classical ℓ_1 -minimization,

$$\min_u \|u\|_1 \quad \text{s.t.} \quad \Phi u = g. \quad (3)$$

As the ℓ_1 -problem is strictly convex its solution is provably feasible for many contemporary numerical techniques (e.g. interior point methods), transforming the signal reconstruction problem into a tractable search. Due to increased restrictions

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

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

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

on Φ , the practical cost of this tractability is that the ℓ_1 -problem requires more measurements than the ℓ_0 -problem for guaranteed recoverability. In other words, the sufficient conditions for exact signal recoverability are tighter for (3). Depending on the particular experiment, guaranteed tractability may outweigh the penalty of a higher required sampling rate.

In applications such as MRI where the cost (e.g. time) associated with performing measurements is very high, it may be desirable to trade the theoretical guarantee of tractability for the potential of reconstructing the target signal from even fewer samples than possible via ℓ_1 -minimization. With this goal in mind, several authors have relaxed the prior-term of the ℓ_0 -problem to nonconvex forms that more closely resemble (1) than does the ℓ_1 -norm. Despite differences in algorithmic implementation, each nonconvex reconstruction model effectively solves

$$\min_u \sum_{x \in \Omega} \rho_\sigma(|u(x)|) \quad \text{s.t.} \quad \Phi u = g \quad (4)$$

either directly or indirectly, where $\rho_\sigma(\cdot)$ is a concave metric (symmetric, positive-definite, and subadditive) whose gradient is a.e. non-zero. σ is a non-negative real scale parameter that coerces pointwise convergence of $\rho_\sigma(\cdot)$ to (1) as $\sigma \rightarrow 0$.

Several definitions of $\rho_\sigma(\cdot)$ have been successfully used in practice. Initial work by Chartrand [8] demonstrated empirically that performance gains over ℓ_1 -minimization could be achieved using an allometric or power function prior,

$$\rho_\sigma(|\cdot|) = |\cdot|^\sigma \quad (0 < \sigma \leq 1). \quad (5)$$

In the reweighted ℓ_1 algorithm of Candès et al. [9], the symmetric log function,

$$\rho_\sigma(|\cdot|) = \ln \left(\frac{|\cdot|}{\sigma} + 1 \right), \quad (6)$$

is indirectly minimized through a sequence of modified ℓ_1 -problems. Trzasko and Manduca's homotopic ℓ_0 -minimization technique [4] attacks (2) asymptotically by chaining a series of nonconvex problems like (4) with progressively-decreasing values of σ , driving the prior term towards (1). In addition to (5) and (6), they advocate use of the Laplace error function,

$$\rho_\sigma(|\cdot|) = 1 - e^{-\frac{|\cdot|}{\sigma}}, \quad (7)$$

and the Geman function [10],

$$\rho_\sigma(|\cdot|) = \frac{|\cdot|}{|\cdot| + \sigma}. \quad (8)$$

The functional behavior of (5-8) relative to the zero-norm can be seen in Figure 1. While there is no theoretical guarantee that a global solution to any of these nonconvex forms is numerically feasible, in practice they all consistently yield superior reconstructions to ℓ_1 -minimization, especially at very low sampling rates (see Figure 2 for examples).

Despite compounding numerical evidence, there has been little theoretical analysis on the performance of nonconvex reconstruction problems. In [11] and [4] it was proven that an ℓ_1 -recoverable signal can also be reconstructed by any form of (4) with equal guarantees. Unfortunately, sheer equivalence

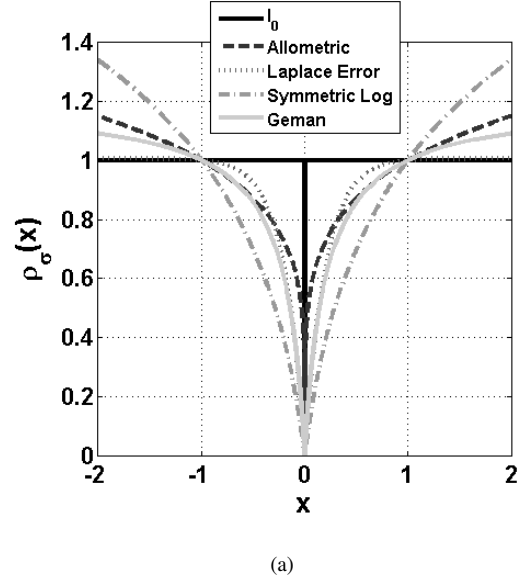


Fig. 1. Functional behavior of concave metric priors ($\sigma = 0.25$) defined in (5-8) relative to the zero-norm.

does not help identify why performance gains are observed when employing nonconvex priors. Noting the functional separability of $\rho_\sigma(|\cdot|) = |\cdot|^\sigma$ ($0 \leq \sigma \leq 1$), Chartrand showed for this specific case that the number of samples sufficient to guarantee exact signal reconstruction is continuously diminished as σ is relaxed from $1 \rightarrow 0$ [8]. While the existence of a similar relaxation for the general case given in (4) was conjectured in [4], to date there has been no theoretical exposition of such behavior.

In this work, the problem of reconstructing a sparse signal using the general concave prior model is investigated. Following [12] and [4], a bound governing the reconstruction residual for a solution of (4) is developed under the Restricted Isometry Property (RIP) hypothesis [13] in Section II. In Section III, a notion of weak recoverability is introduced for nonconvex sparse signal reconstruction models. It is shown that the sufficient conditions for establishing weak signal recoverability are continuously relaxed as $\sigma \rightarrow 0$, offering new insight on the observed empirical performance of nonconvex sparse signal reconstruction models that resemble the ℓ_0 metric more closely than does the ℓ_1 norm. Several examples of relaxation trends for different functional models are given in Section IV.

II. WHEN CAN A SIGNAL BE RECOVERED VIA COMPRESSIVE SENSING?

A. The Restricted Isometry Property

Let $T_0 = \text{supp}\{f\}$. The challenge of sparse signal reconstruction arises from the fact that T_0 is, in general, not known *a priori* although knowledge of $|T_0|$ may be assumed. This uncertainty can be accounted for when developing conditions for guaranteed signal recoverability by assuring that, for every S -sparse signal (recall $S = |T_0|$), the active submatrix of Φ is well-conditioned and thus the signal does not lie in $\text{null}\{\Phi\}$.

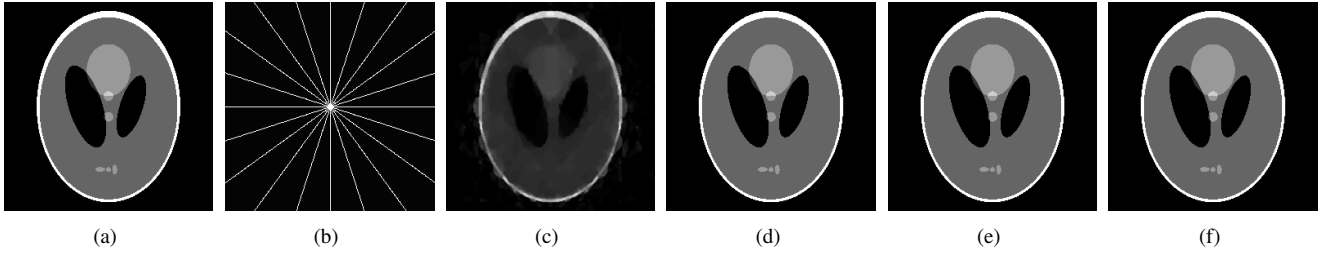


Fig. 2. A comparison of nonlinear reconstructions of the gradient-sparse 256×256 Shepp-Logan tomography phantom (a) from 10 Fourier projections (b) or 4% of the original data. While the convex ℓ_1 -minimization technique (c) fails to recover the image, several nonconvex methods including $\ell_{0.6}$ -minimization (d), reweighted ℓ_1 -minimization (e), and homotopic ℓ_0 -minimization (f) yield exact recovery even at this very low sampling rate.

Suppose $\exists \delta_S \in [0, 1)$ such that

$$(1 - \delta_S)\|u\|_2^2 \leq \|\Phi u\|_2^2 \leq (1 + \delta_S)\|u\|_2^2, \quad (9)$$

$\{\forall u \in \mathbb{R}^N : \|u\|_0 \leq S\}$. Any Φ satisfying (9) is said to possess the restricted isometry property (RIP) and will approximately preserve (Euclidean) signal norm under transformation [13]. Moreover, the frame bounds in (9) represent extremal singular values of the active submatrix of Φ and the effective system condition number is

$$\kappa(\Phi) = \sqrt{\frac{1 + \delta_S}{1 - \delta_S}}. \quad (10)$$

While Φ 's possessing small restricted isometry constants, δ_S , exhibit the best numerical conditioning, it can nonetheless be shown that even $\delta_{2S} < 1$ is a sufficient condition for exact signal recovery via (2) [13]. As will be seen in the remainder of this section and the next, assuring exact signal recovery under the general reconstruction model in (4) will require a tightening of the upper bound on the restricted isometry constant associated with Φ .

In typical CS and related sparse signal recovery applications, the target value for the restricted isometry constant is first identified and a measurement matrix or class of measurement matrices is then constructed such that (9) is satisfied under the given bound. While the design of measurement matrices possessing RIP is beyond the scope of this work, it is noted that the number of samples inherent to Φ typically holds an inverse relation to the targeted value of the restricted isometry constant. For example, it was shown in [14] that a $K \times N$ matrix whose elements are drawn from an independent and identically distributed (i.i.d) Gaussian process, $\Phi_{i,j} \sim \mathcal{N}(0, \frac{1}{K})$, will possess RIP with probability

$$\geq 1 - 2(12/\delta_S)^S e^{-C(\delta_S/2)K}, \quad (11)$$

where $C(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ is a monotonically increasing function. When the desired probability of satisfying the RIP hypothesis is fixed, K and δ_S must clearly be balanced. Similar results hold for other structures such as Bernoulli/Rademacher matrices [14], random Fourier ensembles [15], and noiselets [16]. Consequently, it is desirable to show that exact signal recovery is guaranteed under the least tight (i.e largest) possible bound on the restricted isometry constant. For more information on the design of measurement matrices satisfying RIP, the reader is referred to [2], [15], [14], [17], [18], [19].

B. Uniform Recoverability for Concave Prior Problems

It was stated in Section II-A that exact signal recovery can be guaranteed when the restricted isometry constant of a given measurement matrix, Φ , falls below some theoretically-defined threshold. Following the work of Candès in [12], Trzasko and Manduca showed in [4] that, for an arbitrary $M \geq S$, the residual of the minimizer of (4), $h = f - u$, will satisfy both

$$\sum_{x \in T_0} \rho_\sigma(|h(x)|) \geq \sum_{x \in \Omega \setminus T_0} \rho_\sigma(|h(x)|) \quad (12)$$

and

$$\sum_{x \in T_0} \left[\rho_\sigma(|h(x)|) - \frac{M}{S} \rho_\sigma \left(\frac{1 - \delta_{2M}}{\delta_{2M}} \sqrt{\frac{S}{2M}} |h(x)| \right) \right] \geq 0. \quad (13)$$

Trivially, if

$$\delta_{2M} < \left[1 + \sqrt{\frac{2M}{S}} \right]^{-1} \quad (14)$$

then (13) will hold only if $h(x) = 0, \forall x \in T_0$. (12) then implies that $h(x) = 0, \forall x \in \Omega$ and thus u will be identically f . For the ℓ_1 -based Compressive Sensing model [2], this bound can be relaxed to

$$\delta_{2M} < \left[1 + \sqrt{\frac{2S}{M}} \right]^{-1}. \quad (15)$$

Note that when $S = M$, (14) and (15) are equivalent. Consequently, $\delta_{2S} < \sqrt{2} - 1 \approx 0.4142$ [12] is a sufficient condition for exact recovery via both (3) and (4). A similar universal sufficient condition for all sparse signal reconstruction problems employing concave priors was also proven in [11] using mutual coherence arguments. In Section III, it will be shown that the upper bound on the restricted isometry constant can be relaxed by considering the specific functional definitions of $\rho_\sigma(|\cdot|)$ and allowing $\sigma \rightarrow 0$.

III. RELAXED CONDITIONS FOR RECOVERABILITY OF SPARSE SIGNALS

A. Exploiting the Separability of Allometric Priors

For the specific case of the allometric prior given in (5), $\rho_\sigma(|\cdot|) = |\cdot|^\sigma$ ($0 < \sigma \leq 1$), relaxation of the upper bound on the restricted isometry constant as $\sigma \rightarrow 0$ was first demonstrated by Chartand [8] and later extended by Trzasko

and Manduca [4] under (13). In the latter work, the general recovery condition for (4) resorts to

$$\left[1 - \frac{M}{S} \left(\frac{1 - \delta_{2M}}{\delta_{2M}} \sqrt{\frac{S}{2M}} \right)^\sigma \right] \sum_{x \in T_0} |h(x)|^\sigma \geq 0 \quad (16)$$

due to the natural separability of power functions and it immediately follows that, for any $0 < \sigma \leq 1$,

$$\delta_{2M} < \left[1 + \sqrt{\frac{2M}{S}} \cdot \left(\frac{S}{M} \right)^{\frac{1}{\sigma}} \right]^{-1} \quad (17)$$

is a sufficient condition for exact signal recovery via (4). If $M > S$, this upper bound on δ_{2M} approaches unity as $\sigma \rightarrow 0$; otherwise, $\delta_{2S} < \sqrt{2} - 1$ as seen in Section II-B.

B. Weak Recoverability Under Non-Separable Priors

Recall that $h \in \text{null}\{\Phi\}$. As both h and Φ are assumed to be of bounded magnitude, there exists a finite upper bound ϵ such that $\|h_{T_0}\|_\infty \leq \epsilon < \infty$, namely $h_{T_0} \in \Theta$ where

$$\Theta \triangleq \{\forall x \in T_0 : |h(x)| \in [0, \epsilon]\} \subset \mathbb{C}^S.$$

Determining signal recoverability essentially resorts to identifying the subset of Θ whose elements satisfy the residual condition in (13). Trivially, $h_{T_0} = \mathbf{0}^S$ satisfies (13). If it can be shown that $\Lambda = \{\emptyset\}$, where

$$\Lambda = \left\{ h_{T_0} \in \{\Theta \setminus \mathbf{0}^S\} : \sum_{x \in T_0} g(h(x)) \geq 0 \right\} \quad (18)$$

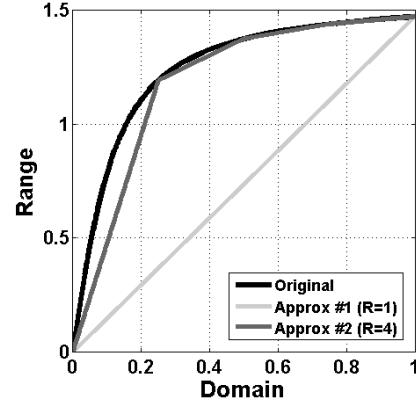
and

$$g(\cdot) = \rho_\sigma(|\cdot|) - \frac{M}{S} \rho_\sigma \left(\frac{1 - \delta_{2M}}{\delta_{2M}} \sqrt{\frac{S}{2M}} |\cdot| \right), \quad (19)$$

h_{T_0} must be the zero vector and exact recovery is guaranteed. When this level of assurance exists, a functional model can be said to possess strong recoverability. In conjunction with this logic, the separability of the allometric prior was exploited in Section III-A for development of an analytical upper bound on δ_{2M} that relaxes as $\sigma \rightarrow 0$.

Unfortunately, many commonly-employed priors like the Laplace error functional in (7) are not separable and functional-specific analytical bounds cannot be derived through simple algebraic manipulation. Only universal worst-case bounds on the restricted isometry constants like (14) are typically available. While universal bounds suffice to prove that exact signal recovery is always guaranteed (assuming numerical achievement of the global minimum) in some scenarios using (4), they do not permit investigation of the relative optimality of individual functionals with respect to one another akin to that possible for the allometric class.

Much can still be proclaimed about a prior functional model even if exact signal recovery cannot be universally guaranteed under a prescribed δ_{2M} . Consider the notion of weak recoverability, namely $\mu(\Lambda) = 0$ where μ is a measure (to be defined in Section III-C). Unlike strong recoverability, weak recoverability does not exclude the possibility that some non-zero residual signals may satisfy (13); it does, however,



(a)

Fig. 3. Piecewise linear approximation of the concave arctangent function. Whereas the single linear approximant ($R = 1$) is only close to the original function at the endpoints, the composite approximant ($R = 4$) provides a significantly more accurate approximation.

assert that (13) will be violated by almost all residual signals in $\{\Theta \setminus \mathbf{0}^S\}$. Moreover, exact recovery remains feasible when $\mu(\Lambda) = 0$ as $\mu(\emptyset) = 0$; if $\mu(\Lambda) > 0$ then Λ must contain an uncountable set of non-zero residual signals and $\Lambda = \{\emptyset\}$ is simply not achievable. It is this notion of weak recoverability that will guide the remainder of this work.

C. Bounding Measure of the Set of Viable Residual Signals

Unlike the exposition in Section III-A for $\rho_\sigma(|\cdot|) = |\cdot|^\sigma$ ($0 < \sigma \leq 1$), an analytical upper bound on δ_{2M} will not be developed for the general prior case. Instead, an upper bound on $\mu(\Lambda)$ will be derived for an arbitrary (σ, δ_{2M}) pair. It will then be shown that (σ, δ_{2M}) space exhibits nearly binary partitioning under this bound and an inverse relationship between σ and δ_{2M} characterizes the transition boundary.

Let μ be a Lebesgue measure and recall

$$\mu(\Lambda) = \int_{\Theta \setminus \mathbf{0}^S} \mathbb{I}_\Lambda(h) \mu(dh) = \int_{\Theta} \mathbb{I}_\Lambda(h) \mu(dh), \quad (20)$$

where \mathbb{I}_Λ is the indicator function for Λ . Assuming Θ is uniformly dense, consider the normalized measure $\mu(dh) = |\Theta|^{-1} dh$ and note that

$$\mu(\Lambda) = \frac{1}{|\Theta|} \int_{\Theta} \mathbb{I}_\Lambda(h) dh = \frac{|\Lambda|}{|\Theta|}. \quad (21)$$

$\mu(\Lambda)$ is then an (unbiased) tally of the fraction of Θ that Λ occupies. Moreover, $\mu(\Lambda)$ is effectively a uniform probability. As in Monte Carlo integration [20], evaluation of (21) is technically equivalent to defining $\mu = \mathbb{P}$ and computing the probability that an independent and identically distributed (i.i.d) uniform random sequence satisfies (13). Standard probability inequalities can then be chained to derive an upper bound on $\mathbb{P}(\Lambda)$. The following Lemma is required to develop this bound.

Lemma 1: Let $k \sim \mathcal{U}(0, \epsilon)$. Additionally, let $\mathbb{E}[\cdot]$ denote the expectation operator. $\forall R \in \mathbb{Z}^+$,

$$\mathbb{E}[g(k)] \leq \rho_\sigma \left(\frac{\epsilon}{2} \right) - \xi(\sigma, \delta_{2M}),$$

where

$$\xi(\sigma, \delta_{2M}) = \frac{M}{2SR} \sum_{r=0}^{R-1} \left[\rho_\sigma \left(\epsilon \beta \frac{r}{R} \right) + \rho_\sigma \left(\epsilon \beta \frac{r+1}{R} \right) \right]$$

and

$$\beta = \frac{1 - \delta_{2M}}{\delta_{2M}} \sqrt{\frac{S}{2M}}.$$

Proof: First note that

$$\mathbb{E}[g(k)] = \mathbb{E}[\rho_\sigma(k)] - \frac{M}{S} \mathbb{E}[\rho_\sigma(\beta k)]$$

following (19). Bounding $\mathbb{E}[g(k)]$ requires the use of both upper and lower bounds on $\mathbb{E}[\rho_\sigma(k)]$. Invoking Jensen's inequality for concave functions [21],

$$\begin{aligned} \mathbb{E}[g(k)] &\leq \rho_\sigma(\mathbb{E}[k]) - \frac{M}{S} \mathbb{E}[\rho_\sigma(\beta k)] \\ &= \rho_\sigma\left(\frac{\epsilon}{2}\right) - \frac{M}{S} \mathbb{E}[\rho_\sigma(\beta k)]. \end{aligned}$$

Next, recall the composite trapezoid rule for definite integrals [22] (see Figure 3). For concave integrands, the expectation can be bounded as

$$\begin{aligned} \mathbb{E}[\rho_\sigma(\beta k)] &= \frac{1}{\epsilon} \int_0^\epsilon \rho_\sigma(\beta k') dk' \\ &= \frac{1}{\beta \epsilon} \int_0^{\beta \epsilon} \rho_\sigma(k') dk' \\ &\geq \frac{1}{R} \left[\frac{\rho_\sigma(\beta \epsilon)}{2} + \sum_{r=1}^{R-1} \rho_\sigma\left(\beta \epsilon \frac{r}{R}\right) \right] \\ &= \frac{1}{2R} \sum_{r=0}^{R-1} \left[\rho_\sigma\left(\epsilon \beta \frac{r}{R}\right) + \rho_\sigma\left(\epsilon \beta \frac{r+1}{R}\right) \right] \\ &= \frac{S}{M} \xi(\sigma, \delta_{2M}) \end{aligned}$$

via the Edmundson-Madansky inequality [23], [24]. As a result,

$$\mathbb{E}[g(k)] \leq \rho_\sigma\left(\frac{\epsilon}{2}\right) - \xi(\sigma, \delta_{2M}).$$

Theorem 1: Let $h(x) \sim \mathcal{U}(0, \epsilon)$. $\forall \tau > 0$ and $\forall R \geq 1$, an upper bound on the measure of the set viable residual signals for (4) is given by

$$\mathbb{P}(\Lambda) \leq \left(C_1 \cdot \left[\rho_\sigma\left(\frac{\epsilon}{2}\right) - \xi(\sigma, \delta_{2M}) \right] + C_2 \right)^S, \quad (22)$$

where

$$C_1 = \frac{e^{\tau b} - e^{\tau a}}{b - a} \quad C_2 = \frac{b e^{\tau a} - a e^{\tau b}}{b - a}$$

and

$$a = \inf_{\alpha \in [0, \epsilon]} \{g(\alpha)\} \quad b = \sup_{\alpha \in [0, \epsilon]} \{g(\alpha)\}.$$

Proof: Begin with application of the well-known Chernoff bound [25] for sums of i.i.d. random sequences. $\forall \tau > 0$,

$$\begin{aligned} \mathbb{P}(\Lambda) &= \mathbb{P}\left(\sum_{x \in T_0} g(h(x)) \geq 0\right) \\ &\leq \prod_{x \in T_0} \mathbb{E}\left[e^{\tau g(h(x))}\right] \\ &= \mathbb{E}\left[e^{\tau g(k)}\right]^S, \end{aligned}$$

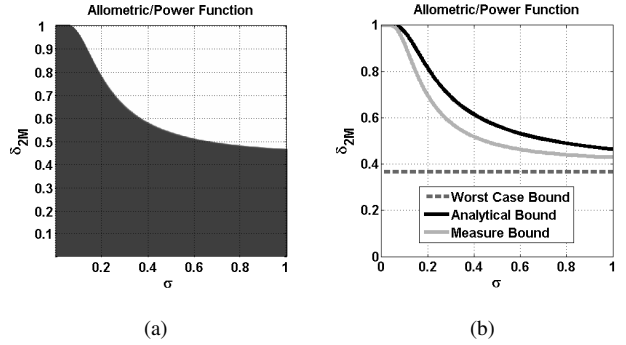


Fig. 4. (a) Investigation of the measure bound for 512×512 samples of the parameter space $\{[0, 1] \otimes [0, 1]\}$ under the allometric function prior with $S = 250$, $M = 1.5S$, and $R = 1000$. Dark gray corresponds to near-zero measure whereas white corresponds to full (unity) measure. (b) contrasts the $\mathbb{P}(\Lambda) = 1 \times 10^{-10}$ isoline of (a) with the known analytical upper bound on δ_{2M} .

where, again, $k \sim \mathcal{U}(0, \epsilon)$. As the exponential operator is convex with respect to $g(\cdot)$ when $\tau > 0$,

$$\begin{aligned} \mathbb{E}\left[e^{\tau g(k)}\right] &\leq \mathbb{E}\left[\frac{g(k) - a}{b - a} e^{\tau b} + \frac{b - g(k)}{b - a} e^{\tau a}\right] \\ &= C_1 \cdot \mathbb{E}[g(k)] + C_2 \end{aligned}$$

which is due to Hoeffding [26]; consequently,

$$\mathbb{P}(\Lambda) \leq (C_1 \cdot \mathbb{E}[g(k)] + C_2)^S.$$

Noting that $C_1 \geq 0$ due to $a \leq b$, Lemma 1 can be rightfully incorporated to yield

$$\mathbb{P}(\Lambda) \leq \left(C_1 \cdot \left[\rho_\sigma\left(\frac{\epsilon}{2}\right) - \xi(\sigma, \delta_{2M}) \right] + C_2 \right)^S.$$

Remark 1: By construction, $0 \leq \mathbb{P}(\Lambda) \leq 1$. Additionally, (14) asserts that $\mathbb{P}(\Lambda) = 0$ for $\delta_{2M} < \left[1 + \sqrt{2M/S}\right]^{-1}$.

Unlike the analytical bounds of Sections II-B and III-A which prescribe a range of values of δ_{2M} that guarantee exact signal reconstruction under a given σ , the upper bound on $\mathbb{P}(\Lambda)$ in (22) gives a bound on the measure of the viable residual signal set for a single arbitrary $(\sigma, \delta_{2M}) \in \{\mathbb{R}^+ \otimes [0, 1]\}$. When the upper bound in (22) is arbitrarily small, it can be conjectured that $\mu(\Lambda) = 0$ under the given (σ, δ_{2M}) parameterization and thus weak recoverability is established. Nonetheless, it is difficult to infer the interrelationship between the restricted isometry constant, σ , and the measure penalty solely by direct inspection of (22) as only the latter term is defined explicitly. In the next section, this measure bound will be evaluated numerically across a (finite) subset of the parameter space $\{\mathbb{R}^+ \otimes [0, 1]\}$ to illustrate a clear binary partitioning that suggests that, as $\sigma \rightarrow 0$, weak recoverability can be established for progressively larger values of δ_{2M} just as was shown for the allometric prior functional class.

IV. EXAMPLES OF RELAXED RECOVERABILITY TRENDS

Although the measure bound in (22) appears straightforward to investigate for prescribed (σ, δ_{2M}) pairs, modest numerical optimization is required to compute the values of a and b , the

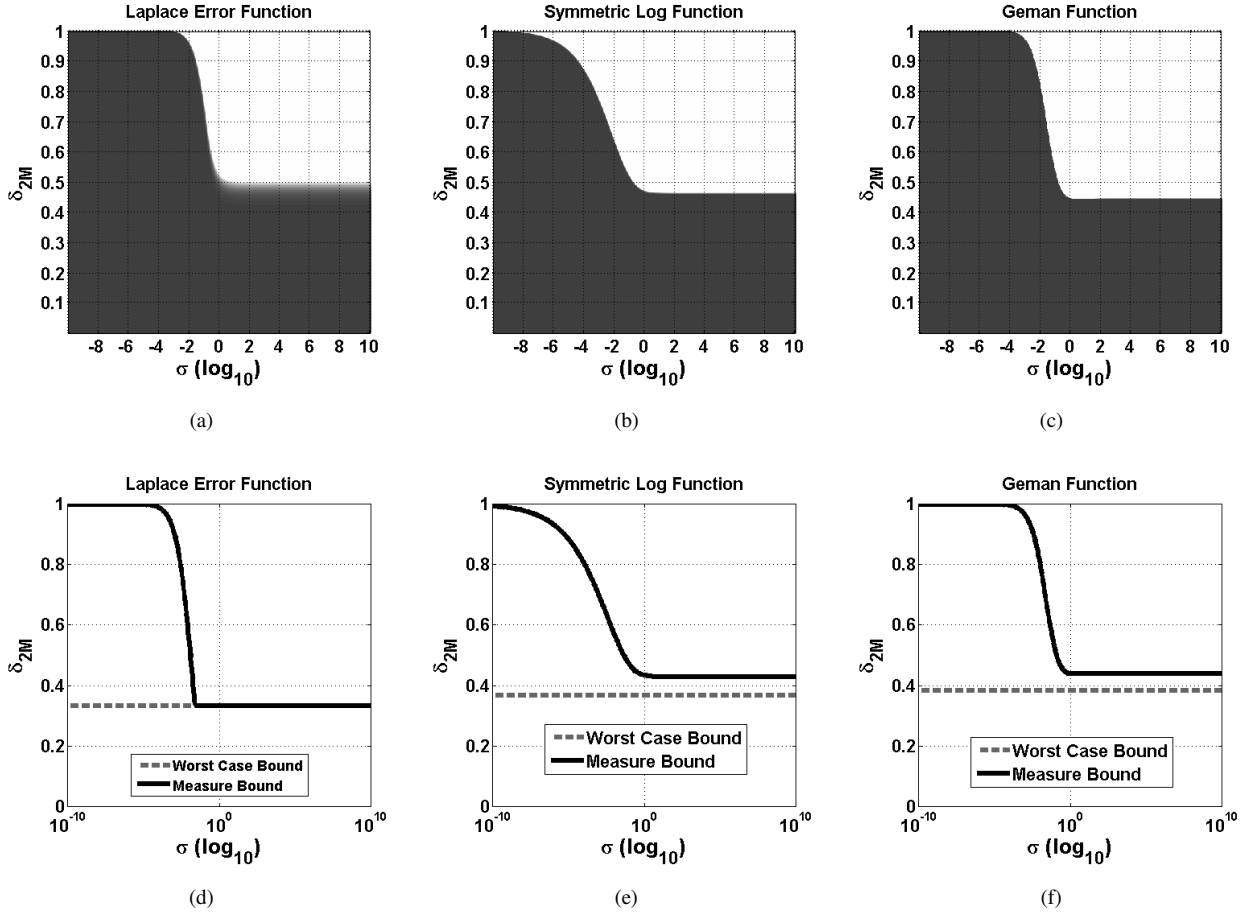


Fig. 5. Investigation of the measure bound for 512×512 samples of the parameter space $\{\mathbb{R}^+ \otimes [0, 1]\}$ under the (a) Laplace ($S = 40$, $M = 2S$, $R = 100$), (b) symmetric log ($S = 500$, $M = 1.5S$, $R = 20$), and (c) Geman ($S = 3000$, $M = 1.25S$, $R = 5$) function priors. Dark gray corresponds to near-zero measure whereas white corresponds to full (unity) measure. (d-f) represent the $\mathbb{P}(\Lambda) = 1 \times 10^{-10}$ isolines of (a)-(c), respectively.

respective lower and upper bounds on $g(\cdot)$. Fortunately, both variables can be determined to arbitrarily high precision using simple line search techniques. Recall that ρ_σ is monotone and note that $g(\cdot)$ is defined simply as the subtraction of a scaled version of this function from itself. As the difference of two monotone functions will be either be monotone or, at most, unimodal, bisection can be used to determine the extremal values over a prescribed range. For all presented examples, the classical Golden Section Search [22] ($\text{tol}=1 \times 10^{-10}$) was applied for determining a and b .

Practical investigation of the behavior of (22) under variable (σ, δ_{2M}) requisites selecting a finite subset of the parameter space $\{\mathbb{R}^+ \otimes [0, 1]\}$ to test. For the four proffered examples, δ_{2M} was chosen from 512 uniformly-spaced samples on $[0, 1]$; for the allometric prior case shown in Figure 4, σ was selected from 512 uniformly-spaced samples on $[0, 1]$ and for the three nonseparable examples in Figure 5 σ was selected from 512 log-uniformly-spaced samples over $[1 \times 10^{-10}, 1 \times 10^{10}]$. For the sake of exposition, τ and ϵ were arbitrarily fixed to unity in all four examples.

Figure 4a shows the evaluation of (22) with allometric priors and $S = 250$, $M = 1.5S$, and $R = 1000$, on the finite testing set of (σ, δ_{2M}) pairs. Note the marked

binary partitioning of the parameter space as being of either full (unity) or near-zero measure. In other words, for every distinct prior parameterization almost all or almost none of the elements of $\{\Theta \setminus \mathbf{0}^S\}$, the space of viable residual signals, satisfy the residual condition (13). If near-zero measure is assumed to establish weak recoverability under the prescribed (σ, δ_{2M}) parameterization, a sharp upper bound on δ_{2M} is visibly evident for every definition of σ . The exact locality of this bound can be made more concrete by focusing on a single arbitrarily small isoline of Figure 4a below which every (σ, δ_{2M}) parameterization yields equal or lesser measure. As the measure bound was derived by chaining several non-tight inequalities (see Section III-C), associated isolines are expected to be more conservative than their analytical counterpart. Figure 4b juxtaposes the $\mathbb{P} = 1 \times 10^{-10}$ isoline derived from the presented results with the analytical upper bound on δ_{2M} defined in (17). Under both models, this upper bound is clearly relaxed as $\sigma \rightarrow 0$ and always above the derived worst-case bound defined in (14). Following the discussion in Section II-A, this trend suggests that it may be possible to decrease the number of measurements sufficient for ensured signal recovery by considering prior parameterizations that more closely resemble the target ℓ_0 prior. Moreover, this

example substantiates the accuracy of the presented measure bound at characterizing the relationship between δ_{2M} and σ .

Figures 5a-5c display the evaluated measure bounds for the Laplace error function prior (7) under $S = 40$, $M = 2S$, and $R = 100$, the symmetric log prior in (6) under $S = 500$, $M = 1.5S$, and $R = 20$, and the Geman prior in (8) under $S = 3000$, $M = 1.25S$, and $R = 5$. In each of these cases, again note the marked binary partitioning of the parameter space as well as the apparent relaxation of the upper bound on δ_{2M} as $\sigma \rightarrow 0$. Following the example in Figure 4, the $\mathbb{P} = 1 \times 10^{-10}$ isolines for each of these three priors are also shown in Figures 5d-5f, respectively. Although no analogous analytical models exists for comparison in these examples, the relaxation phenomenon is again evident. Whereas the existence of this relaxation could previously only be conjectured following experimental observations, here it is explicitly demonstrated solely based on theoretical arguments.

The correspondence of these results to the number of samples sufficient to establish weak signal recoverability can be elucidated through a simple example. Consider the specific case of the Laplace prior (see Figures 5a and 5d) with $\sigma = 1 \times 10^{-3}$ and suppose it is desirable to satisfy the RIP hypothesis with failure probability less than 1×10^{-15} . By rearrangement of (11), for $S = 40$ it can be shown that roughly 34000 measurements are sufficient to achieve $\delta_{2S} < \sqrt{2} - 1$ and thus establish strong recoverability. To date, strong recoverability has not been established for larger values of δ_{2M} . From Figure 5d and following previous discussion, $\delta_{3S} < 0.8$ appears to generate $\mu(\Lambda) = 0$. Under this relaxed bound, weak recoverability can be established with as few as 12300 measurements. While the number of measurements for both recoverability forms is very large according to (11), the key observation here is that weak recoverability can be established with many fewer measurements than strong recoverability.

It is briefly noted that, in practice, ϵ will almost never be known *a priori* to reconstructing a signal and thus the choice of unity for exposition may appear trivial. For the priors employed in the three given examples as well as many other, inspection of the functional definitions reveals a fundamental coupling of σ and ϵ . In the above examples, this coupling is defined by an inverse multiplicative relationship between the two variables. As such, any scaling of ϵ away from unity can be wholly compensated for by an analogous scaling of σ to yield qualitatively similar relaxation trends.

V. SUMMARY

Despite an absence of theoretical guarantees regarding the numerical feasibility of global minimizers, nonconvex models for sparse signal recovery have proffered superior performance relative to their convex analogs in practical demonstration. In this work, these observed performance gains are corroborated by theoretical arguments. As the inherent nonseparability of most nonconvex priors has limited development of conditions for guaranteeing strong signal recoverability, a notion of weak recoverability was defined and sufficient conditions for its establishment were developed for the general concave metric sparsity prior class. Through several examples, it was demon-

strated that these conditions for establishing weak recoverability are continuously relaxed as the sparsity prior model is parameterized towards the ideal ℓ_0 measure. As such, fewer samples typically suffice to establish weak recoverability and if the measurement process is guided by satisfying this condition there is potential for an increase in the economy of the measurement process. Future work will hopefully reveal that the observed relaxation trends exist under strong recoverability as well.

REFERENCES

- [1] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [2] E. Candès, 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.
- [3] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [4] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0 -minimization," *IEEE Trans. Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.
- [5] M. Duarte, M. Davenport, D. Tahkar, J. Laska, S. Ting, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [6] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Scientific Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [7] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [8] R. Chartrand, "Exact reconstruction of sparse signal via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [9] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [10] D. Geman and G. Reynolds, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.
- [11] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, 2007.
- [12] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus de l'Académie des Sciences, Série I*, vol. 346, pp. 589–592, 2008.
- [13] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [14] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [15] 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.
- [16] J. Bobin, J. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE J. Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, 2008.
- [17] R. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, no. 4-6, pp. 918–925, 2007.
- [18] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. of the IEEE Workshop on Statistical Signal Processing*, 2007.
- [19] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.
- [20] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C, 2nd Edition*. Cambridge University Press, 1992.
- [21] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.
- [22] M. Heath, *Scientific computing: an introductory survey*. McGraw-Hill, 2002.
- [23] H. Edmundson, "Bounds on the expectation of a convex function of a random variable," *RAND Corporation Paper*, no. 982, 1956.

- [24] A. Madansky, "Bounds on the expectation of a convex function of a multivariate random variable," *Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 743–746, 1959.
- [25] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, 1952.
- [26] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.



Joshua Trzasko (S'99) received the B.S. degree in Electrical and Computer Engineering in 2003 from Cornell University (Ithaca, NY) and is working towards his Ph.D. in Biomedical Engineering at Mayo Graduate School (Rochester, MN). In the summer of 2002, he was an intern at General Electric's Global Research Center (Niskayuna, NY) in the MRI division. His research interests include limited and corrupted data recovery problems in medical imaging, sparse approximation theory, and large-scale optimization.



Armando Manduca received the B.S. degree in mathematics from the University of Connecticut in 1974 and the Ph.D. degree in astronomy from the University of Maryland in 1980. He was a Carnegie Fellow in astronomy in 1980-82, and then joined TRW in Redondo Beach, CA in the area of digital signal and image processing. He was manager of the Research and Analysis Staff of the System Development Division there from 1986 to 1989. He joined the Mayo Clinic in 1989, where he is currently a Professor of Biomedical Engineering and co-director of the Biomathematics Resource. His research interests are in image processing, image reconstruction, inverse problems and computer-aided radiology. He is an associate editor of the IEEE Transactions on Medical Imaging and serves on the editorial board of Medical Image Analysis.