LA-UR-02-1827

Cite Details

B. Wohlberg, “Noise Sensitivity of Sparse Signal Representations: Reconstruction Error Bounds for the Inverse Problem,” *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3053-3060, December 2003.

IEEE Copyright Notice

Noise Sensitivity of Sparse Signal Representations: Reconstruction Error Bounds for the Inverse Problem

Brendt Wohlberg

***Abstract*— Certain sparse signal reconstruction problems have been shown to have unique solutions when the signal is known to have an exact sparse representation. This result is extended to provide bounds on the reconstruction error when the signal has been corrupted by noise, or is not exactly sparse for some other reason. Uniqueness is found to be extremely unstable for a number of common dictionaries.**

***Index Terms*— dictionary, sparse representation, basis selection, adaptive decomposition, inverse problem, error bounds**

I. Introduction

In contrast to most traditional signal decompositions, such as Fourier and wavelet transforms, in which all signals are represented on the same basis, adaptive signal decompositions represent each signal using an optimal[1] subset of basis functions selected from a redundant *dictionary*. The representation of signal $\mathbf{s} \in \mathbb{C}^N$ using dictionary $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{M-1}\}$ may be expressed as $\Phi\boldsymbol{\alpha} = \mathbf{s}$, where *atoms* $\boldsymbol{\phi}_k \in \mathbb{C}^N$ are the columns of $N \times M$ matrix Φ, and the corresponding coefficients in the linear combination are the elements of $\boldsymbol{\alpha} \in \mathbb{C}^M$. One of the most common optimisation criteria is sparsity, where a linear combination is sought which represents the signal with the minimum possible number of non-zero coefficients. Such sparse representations have found a number of applications [1], including EEG (electroencephalography) and MEG (magnetoencephalography) estimation [2], time-frequency analysis [3], and spectrum estimation [4]. The most significant current decomposition algorithms are Matching Pursuit [3] and its variations [5], Basis Pursuit [6], and FOCUSS [2].

Sparse representations are of particular interest when one has reason, based on physics or other prior knowledge, to expect the signals in question to consist of a superposition of only a few fundamental functions, the coefficients of which are significant. In this case, it is useful to know when recovered coefficients may be expected to correspond to the original generating coefficients.

The author is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Phone: (505) 667 6886, Fax: (505) 665 5757, Email: *brendt@t7.lanl.gov*. Los Alamos National Laboratory is operated by the University of California for the U. S. Department of Energy under contract W-7405-ENG-36.

[1] A variety of optimality criteria, including minimum entropy and minimum l^1 norm, have been used.

II. Uniqueness Conditions and Error Bounds

Consider a dictionary Φ with the property that any N-cardinality subset of atoms selected from the dictionary is linearly independent. Call any set of coefficients $\boldsymbol{\alpha}$ with $N/2$ or fewer non-zero coefficients a *highly sparse* solution[2] with respect to dictionary Φ. It is easily shown that, if a highly sparse solution exists for signal $\mathbf{s}$, then it is the unique highly sparse solution: if two distinct solutions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ both have $N/2$ or fewer non-zero coefficients, then $\boldsymbol{\alpha} - \boldsymbol{\beta}$ has N or fewer non-zero coefficients, and lies in the null space of Φ, thus contradicting the linear independence assumption of N-cardinality subsets of the dictionary. Gorodnitsky and Rao first noted this uniqueness principle [2] [7] for general dictionaries, and Donoho and Huo recently applied the same principle in demonstrating uniqueness with respect to a specific dictionary constructed as a union of time and frequency dictionaries [8].

While instructive, this result is of little assistance when the signal is known to include a noise component, which is almost invariably the case. The sparse signal representation model is therefore extended so that the signal $\mathbf{s} = \Phi\boldsymbol{\alpha} + \boldsymbol{\eta}$ includes a residual component $\boldsymbol{\eta}$ without a sparse representation on dictionary Φ. This residual component will be referred to as the signal noise due to its role in the signal model, representing the actual signal noise when $\boldsymbol{\alpha}$ is the actual generating coefficient vector, and the hypothetical signal noise with respect to a specific reconstruction when $\boldsymbol{\alpha}$ is a solution to the inverse problem. Under these more realistic conditions, one would like to be able to bound the reconstruction error in terms of the signal noise magnitude.

Such a result would provide an indication of the significance of a particular solution $\boldsymbol{\alpha}$, perhaps obtained using one of the methods mentioned in Section I, by bounding the maximum $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$, for any alternative solution $\boldsymbol{\beta}$ of the same[3] or higher sparsity than $\boldsymbol{\alpha}$, in terms of the distance $\|\boldsymbol{\eta}\| = \|\mathbf{s} - \Phi\boldsymbol{\beta}\|$, considered to provide an indication of the relevant noise magnitude. (While the choice of norm is unconstrained at this level of generality, $\|\cdot\|$ should be considered to denote the l^2 norm when a specific choice of norm is necessary in the following sections.) If the bound is small for the primary solution $\boldsymbol{\alpha}$ and its corresponding noise component $\boldsymbol{\eta}$, then any other possible reconstruction is constrained to be similar

[2] The qualifier is required since Gorodnitsky and Rao [2] define *sparse* solutions as those with N or fewer non-zero coefficients.

[3] One might also consider trading slight decreases in sparsity for significant decreases in the magnitude of the non-sparse part of the solution, but this avenue opens a number of additional complications, and is not explored here.

to the primary solution, which is therefore likely to have special physical or other significance. Conversely, a large bound suggests the existence of alternative reconstructions which are not similar to the primary solution, which should therefore not be expected to have special significance.

III. Problem Geometry

Some notation is required in order to facilitate further exposition. Define

$$\Omega_{M,L} = \{(\omega_0, \omega_1, \ldots, \omega_{L-1}) \mid \omega_k \in \mathbb{N},\ 0 \leq \omega_k \leq M-1,\ \omega_k < \omega_{k+1}\},$$

so that $\Omega_{M,L}$ is the set of all

$$\begin{pmatrix} M \\ L \end{pmatrix} = \frac{M!}{L!(M-L)!}$$

distinct index subsets of size L for a dictionary of M atoms. For $\boldsymbol{\omega} \in \Omega_{M,L}$, define operator $P_{\boldsymbol{\omega}} : \mathbb{R}^M \rightarrow \mathbb{R}^L$ (where δ_k^l denotes the Kronecker delta)

$$P_{\boldsymbol{\omega}} = \begin{bmatrix} \delta_0^{\omega_0} & \delta_1^{\omega_0} & \ldots & \delta_{M-1}^{\omega_0} \\ \delta_0^{\omega_1} & \delta_1^{\omega_1} & \ldots & \delta_{M-1}^{\omega_1} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_0^{\omega_{L-1}} & \delta_1^{\omega_{L-1}} & \ldots & \delta_{M-1}^{\omega_{L-1}} \end{bmatrix},$$

which maps from the coefficient space of the full dictionary into the reduced coefficient space consisting of those components indexed by $\boldsymbol{\omega}$. The projection operator $Q_{\boldsymbol{\omega}} : \mathbb{R}^M \rightarrow \mathbb{R}^M$, projecting the full coefficient space into the subspace corresponding to the components indexed by $\boldsymbol{\omega}$, is defined as $Q_{\boldsymbol{\omega}} = P_{\boldsymbol{\omega}}^T P_{\boldsymbol{\omega}}$. Finally, define

$$\begin{aligned} \Gamma_{M,L} &= \{P_{\boldsymbol{\omega}}^T \boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbb{C}^L,\ \boldsymbol{\omega} \in \Omega_{M,L}\} \\ &= \{Q_{\boldsymbol{\omega}} \boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbb{C}^M,\ \boldsymbol{\omega} \in \Omega_{M,L}\} \end{aligned}$$

as the set (it is not a linear space) of all coefficient vectors in $\mathbb{C}^M$ with at most L non-zero coefficients.

By considering only solutions with at most L non-zero coefficients, one is effectively restricting one's attention to solutions for sub-dictionaries $\Phi P_{\boldsymbol{\omega}}^T$ with $\boldsymbol{\omega} \in \Omega_{M,L}$. All of the $\Phi P_{\boldsymbol{\omega}}^T$ are full rank if $L \leq N$, since only dictionaries with the linear independence condition discussed in Section II are considered. The behaviour of each of these sub-dictionaries is revealed by the Singular Value Decomposition (SVD) [9, pp. 70-73]. The SVD of $N \times L$ matrix A is

$$A = U \Sigma V^T,$$

where U is an $L \times L$ matrix, the columns $\mathbf{u}_k$ of which are the *left singular vectors*, V is an $N \times N$ matrix, the columns $\mathbf{v}_k$ of which are the *right singular vectors*, and Σ is a diagonal matrix of *singular values* $\sigma_k,\ 0 \leq k \leq \min\{N, L\} - 1$, ordered so that $\sigma_k \geq \sigma_{k+1}$. The maximum and minimum singular value of A are denoted as $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ respectively. Geometrically, the singular values are the lengths of the semi-axes of the hyperellipsoid constructed as the mapping by A of the unit hypersphere in the domain space of A. Conversely, the inverses of the singular values define a hyperellipsoid in the domain space of A as the pre-image of the unit sphere in its range space. The range space of each sub-dictionary $\Phi P_{\boldsymbol{\omega}}^T$ is a subspace of the range space of the full dictionary Φ.

IV. A Solution-Dependent Bound

Given a specific dictionary Φ with primary solution $\boldsymbol{\alpha}$, and maximum signal noise magnitude ϵ, the the maximum distance between $\boldsymbol{\alpha}$ and any other sparse solution with at most L non-zero coefficients may be expressed as[4]

$$\begin{aligned} \rho_{L,\boldsymbol{\alpha}}(\epsilon) &= \max_{\|\Phi\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\| \leq \epsilon,\ \boldsymbol{\beta} \in \Gamma_{M,L}} \|\boldsymbol{\beta} - \boldsymbol{\alpha}\| \\ &= \max_{\|\Phi Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\| \leq \epsilon,\ \boldsymbol{\beta} \in \mathbb{C}^M,\ \boldsymbol{\omega} \in \Omega_{M,L}} \|Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \boldsymbol{\alpha}\|, \end{aligned}$$

where the Φ and M subscripts of ρ, explicitly indicating dependence on these parameters, are suppressed for notational simplicity. The common assumption that all atoms in the dictionary have unit norm is adopted in order to avoid problems of dependence on dictionary scaling (an alternative approach is outlined in Appendix I). Computation of this value requires computation of

$$\rho_{L,\boldsymbol{\alpha},\boldsymbol{\omega}}(\epsilon) = \max_{\|\Phi Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\| \leq \epsilon,\ \boldsymbol{\beta} \in \mathbb{C}^M} \|Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \boldsymbol{\alpha}\|$$

for all $\boldsymbol{\omega} \in \Omega_{M,L}$, which is computationally infeasible except for very small values of M and L. It is also important to note that this bound is only valid for a specific primary solution $\boldsymbol{\alpha}$, and does not represent a general property of the dictionary for all signals and possible solutions. Nevertheless, computation of this value for very small problems is instructive, a method being described in Appendix II.

Using this method, the results plotted in Figures 1 and 2 were computed for two example dictionaries; one based on the Discrete Fourier Transform (DFT)

$$\phi_k(n) = \exp\left(-\frac{2\pi\imath}{M} kn\right) \quad n \in \{0, \ldots, N-1\},\ k \in \{0, \ldots, M-1\},$$

and the other on the Discrete Cosine Transform of Type II (DCT-II) [10, pp. 276-281]

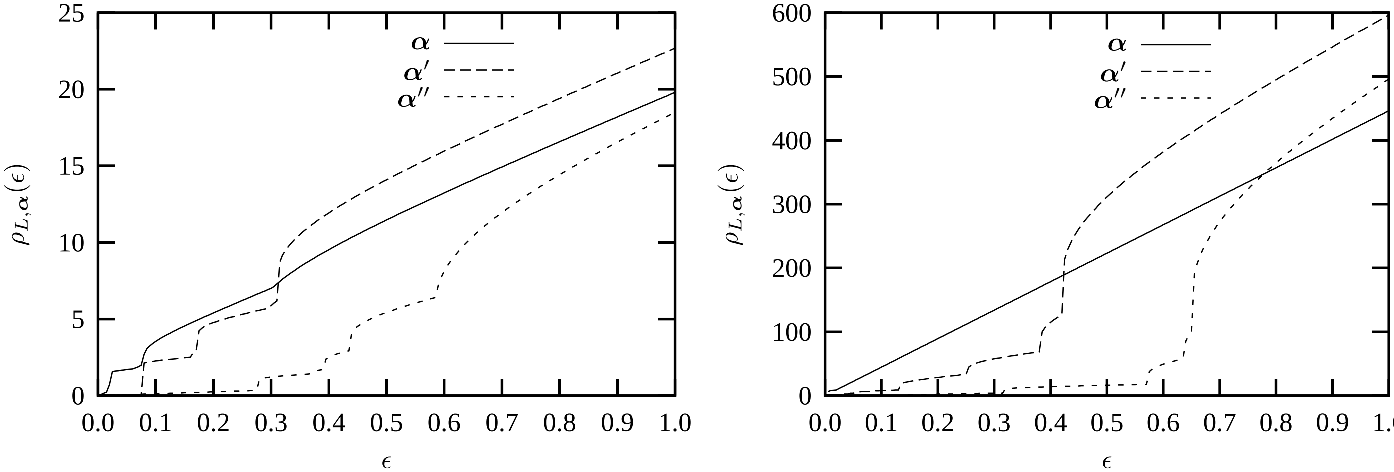
$$\phi_k(n) = \cos\left(\frac{\pi k(n + \frac{1}{2})}{M}\right) \quad n \in \{0, \ldots, N-1\},\ k \in \{0, \ldots, M-1\}$$

(the normalisations are omitted from these definitions for simplicity, but all results are presented for dictionaries with all atoms scaled to have unit norm). In all cases the primary solution $\boldsymbol{\alpha}$ has unit norm, so that the range $\epsilon \in [0, 1]$ corresponds to a signal to noise ratio range of infinity to 0 dB. Note the complex behaviour of the plots (for example, the bound for $\boldsymbol{\alpha}$ is larger than that for $\boldsymbol{\alpha}'$ for low noise, but becomes smaller for ϵ larger than about 0.3), the significant differences in stability between the DFT and DCT-II dictionaries, and the rapid decrease in reconstruction stability of both dictionaries with increasing L.

V. A Solution-Independent Bound

In addition to the computational expense of the bound $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ described in the previous section, it is valid only for a specific primary solution $\boldsymbol{\alpha}$. An alternative approach

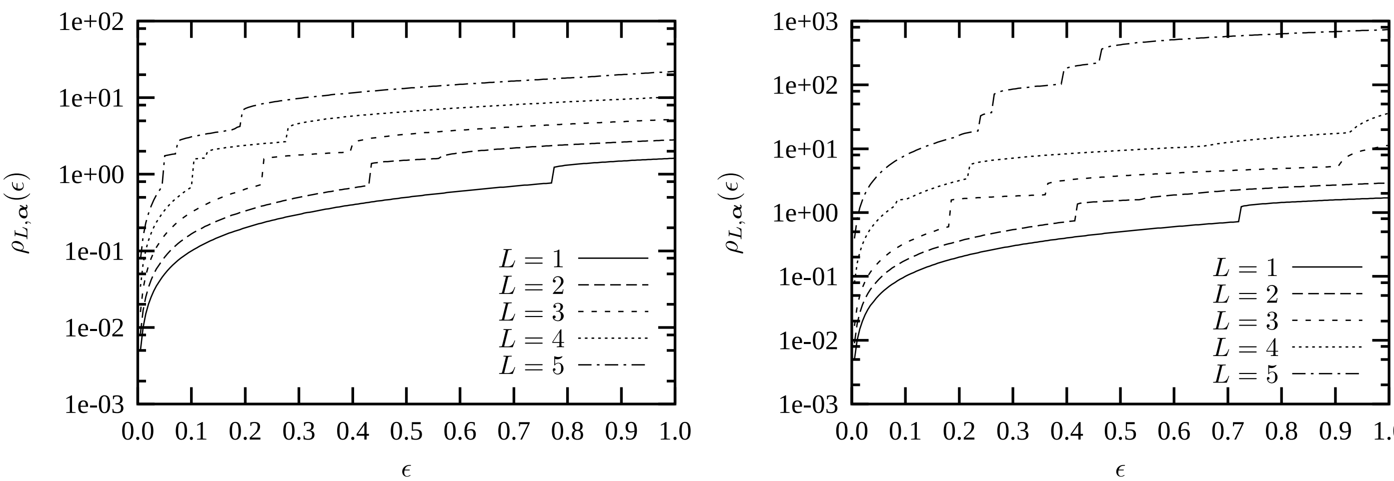
[4] Equivalently, the problem may also be expressed as the maximisation of $\|P_{\boldsymbol{\omega}}^T\boldsymbol{\beta} - \boldsymbol{\alpha}\|$ for $\boldsymbol{\beta} \in \mathbb{C}^L$, subject to the constraint $\|\Phi P_{\boldsymbol{\omega}}^T\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\| \leq \epsilon$.



(a) DFT dictionary, $N = 16$, $M = 32$.

(b) DCT-II dictionary, $N = 16$, $M = 32$.

Fig. 1. A comparison of $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ for DFT and DCT-II dictionaries for $L = 5$ and $\boldsymbol{\alpha} = (1,1,1,1,1,0,0,0,0,\ldots)/\sqrt{5}$, $\boldsymbol{\alpha}' = (1,0,1,0,1,0,1,0,1,0,0,0,0,\ldots)/\sqrt{5}$, and $\boldsymbol{\alpha}'' = (1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,0,0,\ldots)/\sqrt{5}$.



(a) DFT dictionary, $N = 16$, $M = 32$.

(b) DCT-II dictionary, $N = 16$, $M = 32$.

Fig. 2. A comparison of $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ for DFT and DCT-II dictionaries with $\boldsymbol{\alpha}$ such that $\alpha_7 = 1,\ \alpha_k = 0\ \forall k \neq 7$.

is to generalise the noise-free uniqueness result of Section II to obtain a bound that is independent of this variable. While the resulting bound is not as accurate as that of the previous section, it will be shown, by making a connection with $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$, to be the least upper bound independent of this variable (i.e. it is the smallest possible bound that does not depend on any specific primary solution).

For any $N \times L$ complex matrix A, define [11, pg. 216]

$$\mathrm{glb}(A) = \inf_{\mathbf{u}\neq 0} \frac{\|A\mathbf{u}\|}{\|\mathbf{u}\|} = \min_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|,$$

so that

$$\|A\mathbf{u}\| \geq \mathrm{glb}(A)\|\mathbf{u}\| \quad \forall \mathbf{u} \in \mathbb{C}^N.$$

Since $\|\mathbf{u}\| = \sqrt{\mathbf{u}^H\mathbf{u}}$ (in this section $\|\cdot\|$ denotes the l^2 norm), $\inf_{\mathbf{u}\neq 0} \frac{\|A\mathbf{u}\|}{\|\mathbf{u}\|}$ is equal to $\sqrt{\inf_{\mathbf{u}\neq 0} \frac{\mathbf{u}^H A^H A \mathbf{u}}{\mathbf{u}^H \mathbf{u}}}$, the square root of the minimum value taken on by the Rayleigh quotient of $A^H A$, which is equal to the smallest eigenvalue of $A^H A$ [11, pp. 108-109] and $\sigma_{\min}(A)$ in the SVD of A [9, pp. 70-73]. Note that $\mathrm{glb}(A) = 1/\|A^{-1}\|$ when the inverse exists (when $L = N$).

Define

$$\zeta_L = \min_{\boldsymbol{\omega}\in\Omega_{M,L}} \mathrm{glb}(\Phi P_{\boldsymbol{\omega}}^T)$$

for fixed $L \leq N$ ($\mathrm{glb}(A)$ is necessarily zero when $L > N$), providing the bound

$$\|\Phi\boldsymbol{\alpha}\| \geq \zeta_L \|\boldsymbol{\alpha}\|$$

for all $\boldsymbol{\alpha}$ with L or fewer non-zero coefficients. Once again, it is important to impose the dictionary normalisation requirement to avoid dependence on dictionary scaling (as before, an alternative approach is outlined in Appendix I). It should be emphasised that the bound is tight, since equality is attained by setting $\boldsymbol{\alpha}$ to the right singular vector corresponding to the minimum singular value defining ζ_L.

The value ζ_L is a measure of the stability of the linear independence of L-sized subsets of atoms of Φ. Given $\mathbf{s} = \Phi\boldsymbol{\alpha}$ and $\mathbf{s}' = \Phi\boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have maximum numbers of non-zero coefficients $L_{\boldsymbol{\alpha}}$ and $L_{\boldsymbol{\beta}}$ respectively, ζ_L for $L = L_{\boldsymbol{\alpha}} + L_{\boldsymbol{\beta}}$ provides the bound

$$\|\Delta\boldsymbol{\alpha}\| \leq \zeta_L^{-1}\|\Delta\mathbf{s}\|$$

on the difference $\Delta\boldsymbol{\alpha}$ between the two solutions in terms of the difference $\Delta\mathbf{s}$ between the two signals. (If the difference $\Delta\mathbf{s}$ between the two signals is known to be confined to some subspace[5] of the signal space, an improved bound may be obtained by restricting the minimisation in the computation of $\mathrm{glb}(A)$ to that subspace, as described in Appendix III.)

The bound based on ζ_L may be shown to be the smallest possible solution-independent bound by examining the connection with the solution-specific bound $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ of the previous section. The obvious derivation from $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ of a solution-independent bound is the definition (the motivation for the $L' + L''$ subscript of $\rho_{L'+L''}(\epsilon)$ will become apparent shortly)

$$\begin{aligned}\rho_{L'+L''}(\epsilon) &= \max_{\boldsymbol{\alpha}\in\Gamma_{M,L''}} \rho_{L',\boldsymbol{\alpha}}(\epsilon) \\ &= \max_{\|\Phi(\boldsymbol{\beta}-\boldsymbol{\alpha})\|\leq\epsilon,\ \boldsymbol{\beta}\in\Gamma_{M,L'},\ \boldsymbol{\alpha}\in\Gamma_{M,L''}} \|\boldsymbol{\beta}-\boldsymbol{\alpha}\|,\end{aligned}$$

representing the maximum distance between any $\boldsymbol{\beta}$ with at most L' non-zero coefficients and any $\boldsymbol{\alpha}$ with at most L'' non-zero coefficients, when the maximum signal noise magnitude is ϵ. Noting that[6]

$$\{\boldsymbol{\beta}-\boldsymbol{\alpha} \mid \boldsymbol{\beta}\in\Gamma_{M,L'},\ \boldsymbol{\alpha}\in\Gamma_{M,L''}\} = \Gamma_{M,L'+L''},$$

which suggests the substitutions $L = L' + L''$ and $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\alpha}$ for $\boldsymbol{\gamma} \in \Gamma_{M,L}$, one may write[7]

$$\rho_L(\epsilon) = \max_{\|\Phi\boldsymbol{\gamma}\|\leq\epsilon,\ \boldsymbol{\gamma}\in\Gamma_{M,L}} \|\boldsymbol{\gamma}\| = \max_{\|\Phi P_{\boldsymbol{\omega}}^T\boldsymbol{\gamma}\|\leq\epsilon,\ \boldsymbol{\gamma}\in\mathbb{C}^L,\ \boldsymbol{\omega}\in\Omega_{M,L}} \|\boldsymbol{\gamma}\|,$$

from which it is clear that

$$\rho_L(\epsilon) = \zeta_L^{-1}\epsilon.$$

Computation of ζ_L is clearly intractable, in general, for large M and L. Certain dictionaries, may, however, exhibit sufficient structure to reduce the number of subsets to be considered to a manageable number. Consider, for example, a dictionary in which many distinct subset matrices $\Phi P_{\boldsymbol{\omega}}^T$ are unitary transforms of one another. Since $\mathrm{glb}(UA) = \mathrm{glb}(A)$ for unitary U, only one of these related subsets needs to be considered in the minimisation. When the number of subsets is intractable, an upper bound on ζ_L may obviously be obtained by consideration of as many subsets as possible (a random selection may be used, for example) under the prevailing computational constraints. Similarly, if a dictionary Φ consists of a union of the sets of atoms from dictionaries Φ_0 and Φ_1, then $\zeta_L(\Phi) \leq \min\{\zeta_L(\Phi_0), \zeta_L(\Phi_1)\}$.

Results for the example dictionaries defined in Section IV are presented in Figures 3(a), 3(b), 3(c), and 3(d). Note the rapid decay in stability with increasing L, and the significantly greater decay rate of the DCT-II dictionary.

More efficient computation of ζ_L for the DFT dictionary is possible by noting that any subset of L atoms with indices $(\omega_0, \omega_1, \ldots, \omega_{L-1})$ is a unitary transform of the set with indices $(\omega_0+k, \omega_1+k, \ldots, \omega_{L-1}+k)$ for $k \in \mathbb{Z}$ when indices are considered modulo M. In fact, empirical evidence obtained for a wide range of N, M, and L values supports the conjecture that the ζ_L for this dictionary may be obtained by considering only the single index subset $\boldsymbol{\omega} = (0, 1, \ldots, L-1)$. Results computed in this way for larger dictionaries are presented in Figures 4(a) and 4(b).

Bounds derived from ζ_L are compared with the more accurate $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ bounds in Figure 5. In each case $\boldsymbol{\alpha}$ is chosen to have a single non-zero coefficient, and the $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ bound is compared with the bound obtained from ζ_{L+1} (since the primary solutions $\boldsymbol{\alpha}$ are constrained to have a single non-zero coefficient, and are compared with all possible solutions with at most L non-zero coefficients, the difference between the solutions may have at most $L+1$ non-zero coefficients). Note that the ζ_L derived bounds in Figure 5(a) represent, for the chosen $\boldsymbol{\alpha}$, the tightest possible bounds linear in ϵ, while the bounds are somewhat looser for the DCT-II dictionary. It is interesting to note that, for the DFT dictionary (but not the DCT-II dictionary), the same results are obtained for any $\boldsymbol{\alpha}$ with a single non-zero coefficient - this phenomenon is likely to be related to the structural simplicity which allows the rapid computation of ζ_L for this dictionary.

When $\boldsymbol{\omega}$ is the index set on which $\boldsymbol{\alpha}$ has its non-zero coefficients, it is worth noting[8] that

$$\rho_{L,\boldsymbol{\alpha},\boldsymbol{\omega}}(\epsilon) = \left[\mathrm{glb}(\Phi P_{\boldsymbol{\omega}}^T)\right]^{-1}\epsilon,$$

so that the reconstruction error bound is linear in ϵ. This bound is relevant when the reconstruction error is sufficiently small that no equally sparse solutions exist in any other index set.

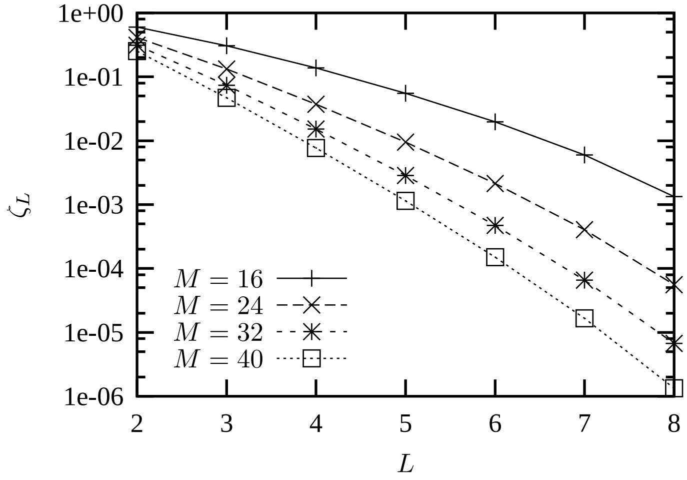
It is clear that none of the non-zero coefficients of primary solution $\boldsymbol{\alpha}$ (with L non-zero coefficients) may take on a zero value within the ball of radius $\min\{|\alpha_k| \mid 0 \leq k < M,\ \alpha_k \neq 0\}$ about $\boldsymbol{\alpha}$. Any alternative solution $\boldsymbol{\beta}$ with at most L non-zero coefficients, must, therefore, have its non-zero coefficients on the same index set $\boldsymbol{\omega}$ as $\boldsymbol{\alpha}$ if $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\| < \min\{|\alpha_k| \mid 0 \leq$

[5] Donoho and Huo [8], in contrast, restrict the noise so that a sparse representation is still possible on a dictionary combining signal and noise sub-dictionaries. Initial stability computations (for small N) for this combined dictionary, for which $M = 2N$, suggest that the representation is reasonably stable for L within the given uniqueness bounds.

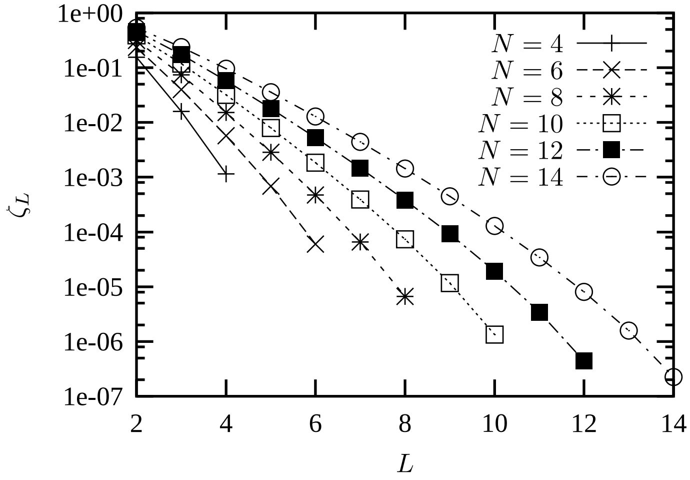
[6] This is easily shown; the difference $\boldsymbol{\beta} - \boldsymbol{\alpha}$ has at most $L' + L''$ non-zero coefficients and is therefore always in $\Gamma_{M,L'+L''}$, and any element of $\Gamma_{M,L'+L''}$ may be expressed as such a difference by choosing an appropriate partition of the indices on which it has non-zero coefficients.

[7] It is interesting to note that $\rho_L(\epsilon) = \rho_{L,\mathbf{0}}(\epsilon)$, the solution-dependent bound for the zero-vector, implying that the zero-vector is always the primary solution for which the solution-dependent bound is the largest.

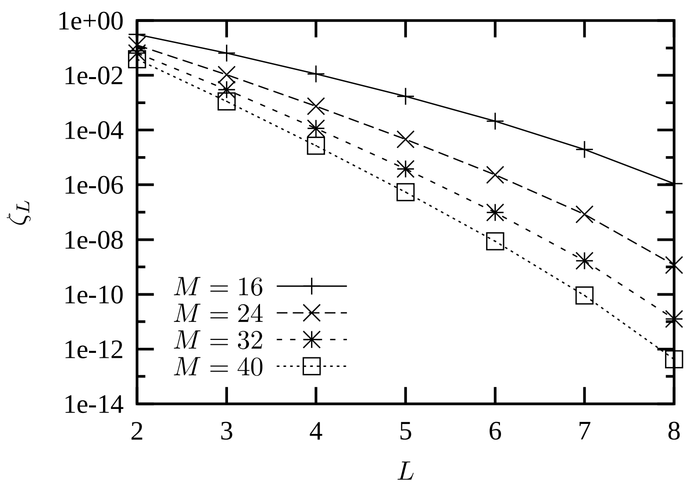
[8] This is easily shown by utilising the equivalent definition of $\rho_{L,\boldsymbol{\alpha},\boldsymbol{\omega}}(\epsilon)$ in terms of the operator $P_{\boldsymbol{\omega}}^T$ and observing that $\|P_{\boldsymbol{\omega}}^T\boldsymbol{\beta}\| = \|\boldsymbol{\beta}\|\ \forall\boldsymbol{\beta}\in\mathbb{C}^L$ and, in this case, $\boldsymbol{\alpha} = P_{\boldsymbol{\omega}}^T P_{\boldsymbol{\omega}}\boldsymbol{\alpha}$.



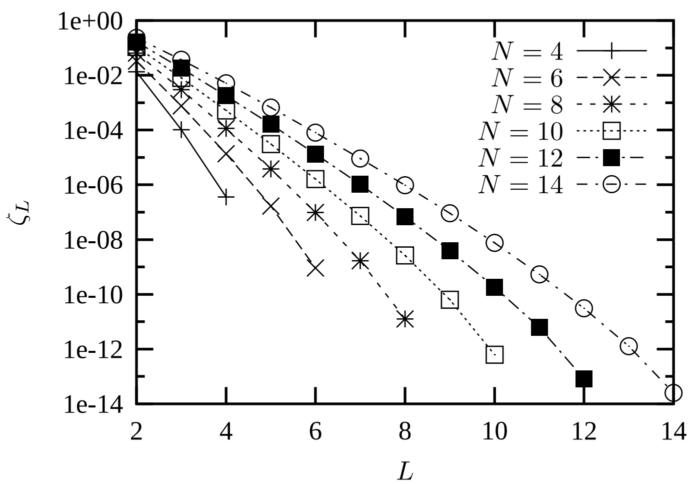
(a) DFT dictionaries with $N = 8$.



(b) DFT dictionaries with $M = 32$.



(c) DCT-II dictionaries with $N = 8$.



(d) DCT-II dictionaries with $M = 32$.

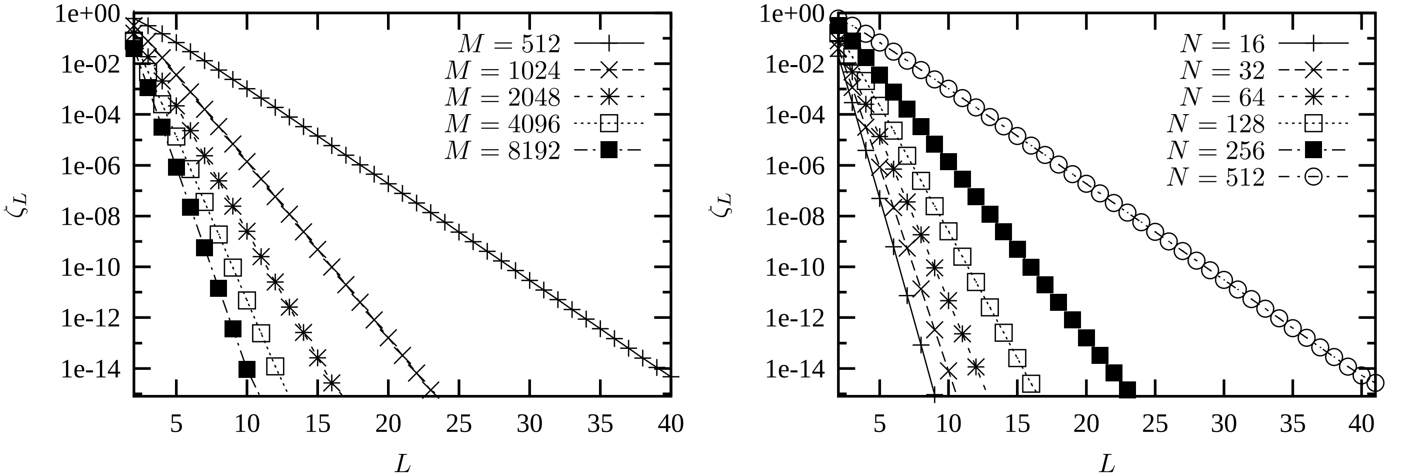
Fig. 3. Variation of ζ_L with L for example dictionaries.

$k < M,\ \alpha_k \neq 0\}$, since the sparsity restrictions require at least one of the non-zero coefficients in $\boldsymbol{\alpha}$ to become zero to allow a zero coefficient in $\boldsymbol{\alpha}$ to be non-zero in $\boldsymbol{\beta}$. This maximum reconstruction error may be guaranteed by imposing a signal noise bound of $\zeta_{2L} \min\{|\alpha_k| \mid 0 \leq k < M,\ \alpha_k \neq 0\}$. (Alternatively, $\boldsymbol{\beta}$ may be restricted to the same index set $\boldsymbol{\omega}$ as $\boldsymbol{\alpha}$ by imposing a signal noise bound smaller than the distance between $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}' = (\Phi Q_{\boldsymbol{\omega}'})^{+} \Phi Q_{\boldsymbol{\omega}} \boldsymbol{\alpha}$ for all other index sets $\boldsymbol{\omega}'$ of the same sparsity.)

VI. Discussion

The tools introduced above allow quantification of the noise sensitivity of sparse reconstruction problems, and provide bounds on the reconstruction error when the signal noise magnitude is known. Except at very low noise levels, very high degrees of sparsity, or small overcompleteness factors M/N, these results indicate very high noise sensitivities for the common DFT and DCT-II dictionaries. In superresolution applications using overcomplete sinusoidal dictionaries [4], for example, these results allow an explicit quantification of the tradeoff between spectral resolution (depending on the degree of overcompleteness of the dictionary) and noise sensitivity of the result, and also suggest that the DFT dictionary is a better choice for superresolution than the DCT-II dictionary due to the significantly lower noise sensitivity of the former dictionary.

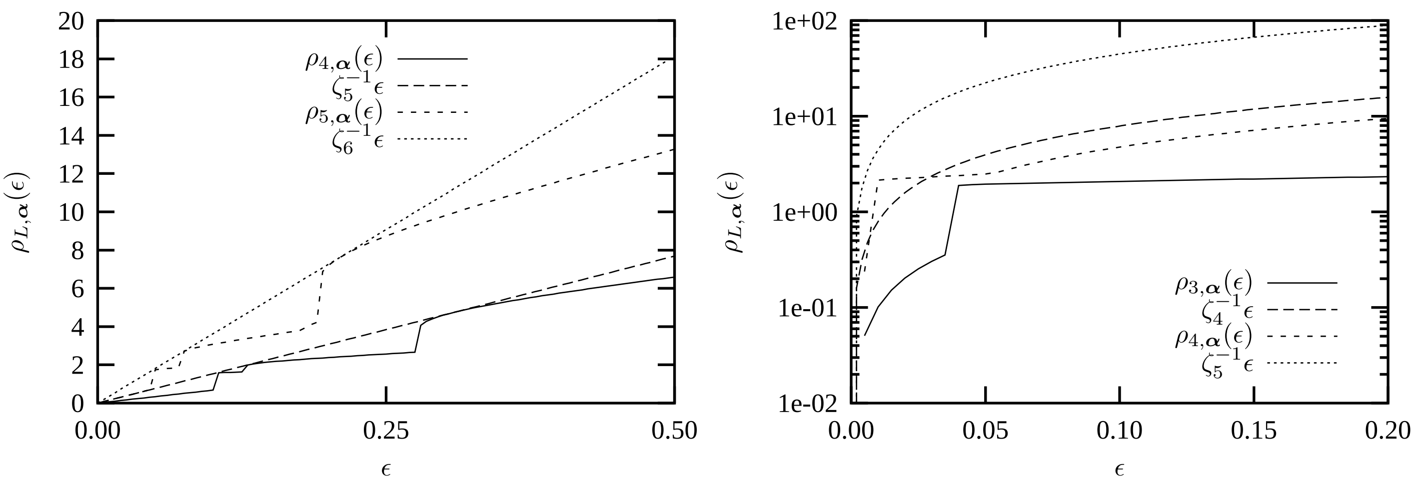
While the bound based on ζ_L is less informative than the more accurate $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ bound, it does appear to provide a useful indication of the relative noise sensitivities of different dictionaries, as well as of the increase in reconstruction error with decreasing sparsity (increasing L). Given the significant differences in the stabilities of the DFT and DCT-II dictionaries, an upper bound on the stability of any dictionary of a given size would be valuable, but is difficult to obtain. The $N \times M$ dictionary with the largest possible ζ_L is related to the optimum packing in the complex Grassmannian space $G(N, L)$



(a) DFT dictionaries with $N = 256$.

(b) DFT dictionaries with $M = 1024$.

Fig. 4. Variation of ζ_L with L for large DFT dictionaries. The vertical axis has been restricted to avoid display of values which are inaccurate due to limited numerical precision in the computation of ζ_L.



(a) DFT dictionary, $N = 16,\ M = 32,\ \alpha_7 = 1,\ \alpha_k = 0\ \forall k \neq 7$.

(b) DCT-II dictionary, $N = 16,\ M = 32,\ \alpha_{31} = 1,\ \alpha_k = 0\ \forall k \neq 31$.

Fig. 5. A comparison of $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ and $\rho_{L+1}(\epsilon)$ bounds for DFT and DCT-II dictionaries

of the $\begin{pmatrix} M \\ L \end{pmatrix}$ L-dimensional subspaces associated with that dictionary, but existing results for packings in Grassmannian spaces [12] are not applicable since the distance measure used is not appropriate for this application.

Appendix I
Alternative Bounds Independent of Dictionary Scaling

The bounds defined in Sections IV and V require restrictions to be placed on dictionary normalisation to avoid dependence on dictionary scaling. Similar bounds may also be defined which are, by their construction, independent of dictionary scaling (although the modified $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ is expected to present greater computational difficulties).

A version of $\rho_{L,\boldsymbol{\alpha}}(\epsilon)$ that is independent of the scaling of Φ may be defined as

$$\rho'_{L,\boldsymbol{\alpha}}(\epsilon) = \max_{\|\Phi Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\|/\|\Phi\boldsymbol{\alpha}\| \leq \epsilon,\ \boldsymbol{\beta}\in\mathbb{C}^M,\ \boldsymbol{\omega}\in\Omega_{M,L}} \frac{\|Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \boldsymbol{\alpha}\|}{\|\boldsymbol{\alpha}\|}.$$

Therefore, for all $\boldsymbol{\beta}$ with at most L non-zero coefficients,

$$\frac{\|s - \Phi\boldsymbol{\beta}\|}{\|s\|} < \epsilon \Rightarrow \frac{\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|}{\|\boldsymbol{\alpha}\|} < \rho'_{L,\boldsymbol{\alpha}}(\epsilon).$$

Defining signal to noise ratios

$$\mathrm{SNR}_{\boldsymbol{\alpha}} = -20\log_{10}\frac{\|\Delta\boldsymbol{\alpha}\|}{\|\boldsymbol{\alpha}\|} \qquad \mathrm{SNR}_{\mathbf{s}} = -20\log_{10}\frac{\|\Delta\mathbf{s}\|}{\|\mathbf{s}\|}$$

in the coefficient and signal spaces respectively, this may be expressed as

$$\mathrm{SNR}_{\mathbf{s}} > -20\log_{10}\epsilon \ \Rightarrow\ \mathrm{SNR}_{\boldsymbol{\alpha}} > -20\log_{10}\rho'_{L,\boldsymbol{\alpha}}(\epsilon).$$

An alternative version of ζ_L that is invariant to scaling of Φ may be defined as

$$\zeta'_L = \min_{\boldsymbol{\omega}\in\Omega_{M,L}} \frac{\mathrm{glb}(\Phi P_{\boldsymbol{\omega}}^T)}{\|\Phi P_{\boldsymbol{\omega}}^T\|} = \min_{\boldsymbol{\omega}\in\Omega_{M,L}} \frac{\sigma_{\min}(\Phi P_{\boldsymbol{\omega}}^T)}{\sigma_{\max}(\Phi P_{\boldsymbol{\omega}}^T)},$$

from which the bound

$$\zeta'_L \frac{\|\Delta\boldsymbol{\alpha}\|}{\|\boldsymbol{\alpha}\|} \leq \frac{\|\Delta \boldsymbol{s}\|}{\|\boldsymbol{s}\|}$$

may be obtained. Using the signal to noise ratios defined above, this may be expressed as

$$\mathrm{SNR}_{\boldsymbol{\alpha}} \geq \mathrm{SNR}_{\mathbf{s}} + 20\log_{10}\zeta'_L.$$

Appendix II
Computation of the maximum reconstruction error

The computation of $\rho_{L,\boldsymbol{\alpha},\boldsymbol{\omega}}(\epsilon)$ as defined in Section IV involves the maximisation of $\|Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \boldsymbol{\alpha}\|$, subject to the constraint $\|\Phi Q_{\boldsymbol{\omega}}\boldsymbol{\beta} - \Phi\boldsymbol{\alpha}\| \leq \epsilon$, for $\boldsymbol{\beta} \in \mathbb{C}^M$ (the equivalent formulation in terms of $\Phi P_{\boldsymbol{\omega}}^T$ leads to a derivation similar to that which follows). This is closely related to *least squares with quadratic constraint* [13][9, Ch. 12] ($\|\cdot\|$ denotes the l^2 norm in this section), but does not conform to all of the restrictions on which the standard approaches are based.

A simpler solution may be derived by consideration of the geometry of the problem in signal space. As illustrated in Figure 6, the feasible region represents the intersection of the ϵ-ball about $\mathbf{s} = \Phi\boldsymbol{\alpha}$ and the range space of $\Phi Q_{\boldsymbol{\omega}}$. The closest point to $\mathbf{s}$ in this range space is its orthogonal projection into that space (A^+ denotes the pseudo-inverse of A)

$$\mathbf{s}' = \Phi Q_{\boldsymbol{\omega}}(\Phi Q_{\boldsymbol{\omega}})^+\mathbf{s},$$

and there is no feasible point when $\|\mathbf{s} - \mathbf{s}'\| > \epsilon$. The feasible region is itself a hypersphere. Since the projection is orthogonal, $\mathbf{s}' - \mathbf{p} \perp \mathbf{s}' - \mathbf{s}$ for any $\mathbf{p}$ in the feasible region, and the radius of this hypersphere is

$$\epsilon' = \sqrt{\epsilon^2 - \|\mathbf{s} - \mathbf{s}'\|^2}.$$

Now, define $\boldsymbol{\alpha}' = (\Phi Q_{\boldsymbol{\omega}})^+\mathbf{s}$ so that $\mathbf{s}' = \Phi Q_{\boldsymbol{\omega}}\boldsymbol{\alpha}'$. The original problem may be expressed as the maximisation of $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|$ subject to the constraint that $\boldsymbol{\beta} = Q_{\boldsymbol{\omega}}\boldsymbol{\beta}$ (i.e. $\boldsymbol{\beta}$ is in the subspace defined by $Q_{\boldsymbol{\omega}}$) and $\|\Phi Q_{\boldsymbol{\omega}}(\boldsymbol{\beta} - \boldsymbol{\alpha}')\| \leq \epsilon'$. This feasible region represents an ϵ'-ball about $\mathbf{s}'$ in signal space, with a corresponding feasible region in coefficient space consisting of a hyperellipsoid about $\boldsymbol{\alpha}'$, defined as the preimage under $\Phi Q_{\boldsymbol{\omega}}$ of this ϵ'-ball about $\mathbf{s}'$. Under the change of coordinates $\boldsymbol{\beta}' = \boldsymbol{\beta} - \boldsymbol{\alpha}'$, this becomes the maximisation of $\|\boldsymbol{\beta}' - (\boldsymbol{\alpha} - \boldsymbol{\alpha}')\|$ with the constraint $\|\Phi Q_{\boldsymbol{\omega}}\boldsymbol{\beta}'\| \leq \epsilon'$, transforming the feasible region into a hyperellipsoid about the origin.

After introducing simplified notation[9], the problem is to find, for transform A (which has rank R), point $\mathbf{p}$, and radius r, the vector $\mathbf{q}$ maximising the distance $\|\mathbf{q} - \mathbf{p}\|$ such that $\|A\mathbf{q}\| \leq r$ and $\mathbf{q}$ is orthogonal to the null space of A (this final requirement ensuring that $\mathbf{q}$ respects the subspace constraint). Using the SVD $A = U\Sigma V^T$, express $\mathbf{q}$ as a linear combination

$$\mathbf{q} = \sum_{k=0}^{R-1} c_k \mathbf{v}_k$$

of the right singular vectors $\mathbf{v}_k$, so that $\mathbf{q}$ is orthogonal to the null space of A and

$$A\mathbf{q} = \sum_{k=0}^{R-1} c_k A\mathbf{v}_k = \sum_{k=0}^{R-1} c_k\sigma_k\mathbf{u}_k.$$

Since the left singular vectors $\mathbf{u}_k$ are mutually orthonormal, $\|A\mathbf{q}\|^2 = \sum_{k=0}^{R-1}\sigma_k^2|c_k|^2$, and the problem may be posed as maximising $\|\sum_{k=0}^{R-1} c_k\mathbf{v}_k - \mathbf{p}\|$ subject to the constraint $\sum_{k=0}^{R-1}\sigma_k^2|c_k|^2 \leq r^2$.

Now, the optimal $\mathbf{q}$ must lie on the boundary of the feasible region since any interior $\mathbf{q}$ may be moved a finite distance in a direction which increases its distance from $\mathbf{p}$ (the distance function does not have local maxima within the feasible region). The constraint may therefore be expressed as $\sum_{k=0}^{R-1}\sigma_k^2|c_k|^2 = r^2$, allowing the Lagrange multiplier approach

$$L(\mathbf{c},\lambda) = \sum_{k=0}^{R-1}|c_k|^2 - 2\,\mathrm{Re}\left(\sum_{k=0}^{R-1} c_k\langle\mathbf{v}_k,\mathbf{p}\rangle\right) + \|\mathbf{p}\|^2 - \lambda\left(\sum_{k=0}^{R-1}\sigma_k^2|c_k|^2 - r^2\right),$$

with the solution

$$c_k = \frac{\langle\mathbf{v}_k,\mathbf{p}\rangle}{1-\lambda\sigma_k^2} \quad \text{and} \quad \sum_{k=0}^{R-1}\sigma_k^2\left(\frac{\langle\mathbf{v}_k,\mathbf{p}\rangle}{1-\lambda\sigma_k^2}\right)^2 = r^2.$$

The constraint equation on the right is solved numerically for λ, and the corresponding $\mathbf{q}$ is obtained via the c_k. Each λ corresponds to a stationary point of the distance function; the solution to the problem is provided by the λ for which the corresponding $\mathbf{q}$ has the greatest distance from $\mathbf{p}$.

Two special cases of the problem require different treatment:

The vector $\mathbf{p}$ *is at the origin* If $\mathbf{p} = 0$ one obtains the equations $(1-\lambda\sigma_k^2)c_k = 0$ so that, for each k, either $c_k = 0$ or $\lambda = \sigma_k^{-2}$. In the general case in which all of the σ_k are distinct, the choice $\lambda = \sigma_k^{-2}$ may only be made for a single k, imposing $c_k = 0$ for all other $k \in \{0, 1, \ldots, R-1\}$. The desired solution is obtained by choosing the c_k corresponding to the smallest σ_k to be non-zero, so that $c_{R-1} = \pm r\sigma_{R-1}^{-1}$ and $\mathbf{q} = r\sigma_{R-1}^{-1}\mathbf{v}_{R-1}$. The geometrical interpretation of this solution is that the maximum distance from the centre of the hyperellipsoid to a point on its boundary is along the direction of the longest semi-axis.

The constraint region is a hypersphere If $\sigma_k = \sigma_l \quad \forall 0 \leq k, l < R$, one may drop

[9] The original problem is addressed by making the substitutions

$$\mathbf{p} = \boldsymbol{\alpha} - \boldsymbol{\alpha}' \qquad \mathbf{q} = \boldsymbol{\beta}' \qquad A = \Phi Q_{\boldsymbol{\omega}} \qquad r = \epsilon'.$$

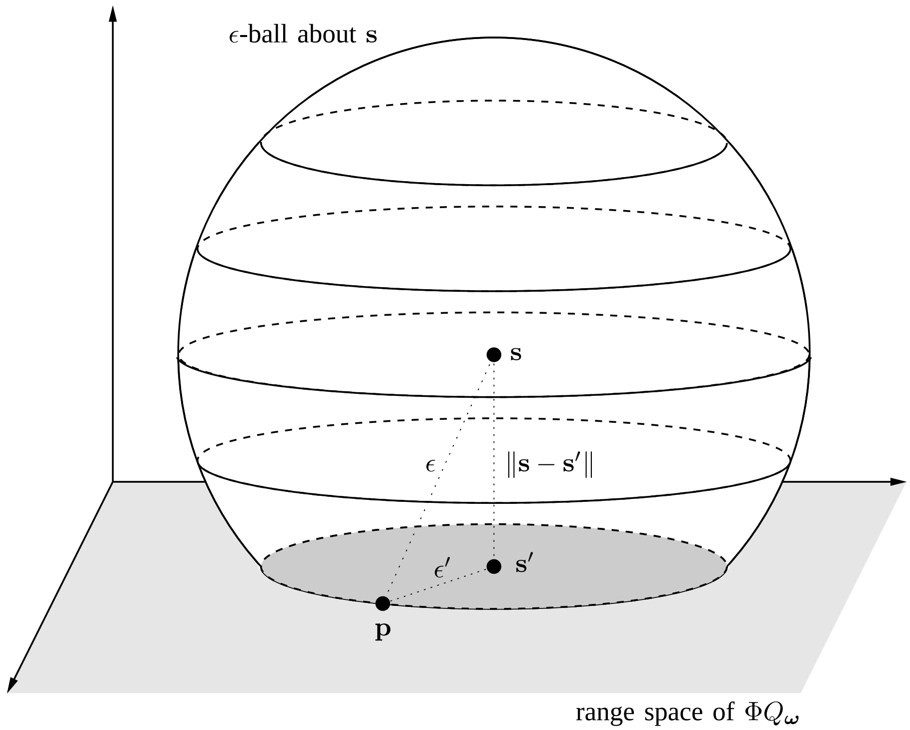


Fig. 6. Computation of $\rho_{L,\boldsymbol{\alpha},\boldsymbol{\omega}}(\epsilon)$: illustration of the geometry of the feasible region in signal space (for $N=3$).

the subscript on σ and write the constraint as $\sigma^2 \sum_{k=0}^{R-1} |c_k|^2 = r^2$, obtaining the expression

$$\lambda = \sigma^{-2}\left(1 \pm \sqrt{\sum_{k=0}^{R-1} \langle \mathbf{v}_k, \mathbf{p}\rangle^2}\right),$$

so that

$$c_k = \frac{-\langle \mathbf{v}_k, \mathbf{p}\rangle}{\frac{\sigma}{r}\sqrt{\sum_{k=0}^{R-1} \langle \mathbf{v}_k, \mathbf{p}\rangle^2}}.$$

The geometrical interpretation of this solution is that, since the feasible region is a hypersphere, the furthest point from $\mathbf{p}$ is the point directly opposite the origin from $\mathbf{p}$.

Appendix III
Restriction of the Greatest Lower Bound to a Subspace

The greatest lower bound of A restricted to the subspace spanned by the columns of B may be expressed as

$$\mathrm{glb}_B(A) = \min_{\|\mathbf{u}\|=1,\ A\mathbf{u}\in \mathrm{ran}(B)} \|A\mathbf{u}\|,$$

where $\mathrm{ran}(B)$ denotes the range space of B. The minimisation is restricted to all $\mathbf{u}$ for which $A\mathbf{u} = B\mathbf{v}$ for some $\mathbf{v}$. Rewriting as

$$\begin{pmatrix} A & -B \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = 0,$$

it is apparent that, if the columns of F are a basis for the null space of $\begin{pmatrix} A & -B \end{pmatrix}$, all valid $\mathbf{u}, \mathbf{v}$ pairs may be generated as

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = F\mathbf{w} = \begin{pmatrix} F_A \\ F_B \end{pmatrix} \mathbf{w}.$$

Therefore $\mathbf{u} = F_A\mathbf{w} = Q\mathbf{w}$, where the columns of Q (which may be computed using the complex QR factorisation [9, pg. 233] of F_A) form an orthonormal basis for $\mathrm{ran}(F_A)$, and

$$\begin{aligned} \mathrm{glb}_B(A) &= \sqrt{\min_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^H Q^H A^H A Q \mathbf{w}}{\mathbf{w}^H Q^H Q \mathbf{w}}} \\ &= \sqrt{\min_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^H Q^H A^H A Q \mathbf{w}}{\mathbf{w}^H \mathbf{w}}} \\ &= \mathrm{glb}(AQ). \end{aligned}$$

Acknowledgment

The majority of this research was performed while the author was with the Center for Nonlinear Studies at Los Alamos National Laboratory. The author wishes to thank Markus Berndt and Kevin Vixie for valuable discussion and comments.

References

[1] B. D. Rao, "Signal processing with the sparseness constraint," in *Proceedings ICASSP-98 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, vol. 3, Seattle, WA, USA, May 1998, pp. 1861–1864.

[2] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[3] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[4] S. Chen and D. Donoho, "Application of basis pursuit in spectrum estimation," in *Proceedings ICASSP-98 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, vol. 3, Seattle, WA, USA, May 1998, pp. 1865–1868.

[5] J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Comparison of basis selection methods," in *30th Asilomar Conference on Signals, Systems & Computers*, vol. 1, Pacific Grove, CA, USA, Nov. 1996, pp. 252–257.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[7] I. F. Gorodnitsky, "Can compact neural currents be uniquely determined?" in *Proceedings 10th International Conference on Biomagnetics*, Santa Fe, NM, USA, Feb. 1996.

[8] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins University Press, 1996.

[10] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA, USA: Wellesley-Cambridge Press, 1996.

[11] P. Lancaster, *Theory of Matrices*. New York, NY, USA: Academic Press, 1969.

[12] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[13] W. Gander, "Least squares with a quadratic constraint," *Numerische Mathematik*, vol. 36, pp. 291–307, 1981.

PLACE PHOTO HERE

Brendt Wohlberg received the BSc(Hons) degree in applied mathematics, and the MSc(Applied Science) and PhD degrees in electrical engineering from the University of Cape Town, South Africa, in 1990, 1993 and 1996 respectively. He is currently a technical staff member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include image coding, pattern recognition, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.