

Performance Limits of Matching Pursuit Algorithms

Yuzhe Jin

Department of Electrical and Computer Eng.
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0407, USA
Email: yujin@ucsd.edu

Bhaskar D. Rao

Department of Electrical and Computer Eng.
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0407, USA
Email: brao@ece.ucsd.edu

Abstract—In this paper, we examine the performance limits of the Orthogonal Matching Pursuit (OMP) algorithm, which has proven to be effective in solving for sparse solutions to inverse problem arising in overcomplete representations. To identify these limits, we exploit the connection between sparse solution problem and multiple access channel (MAC) in wireless communication domain. The forward selective nature of OMP helps it to be recognized as a successive interference cancellation (SIC) scheme that decodes non-zero entries one at a time in a specific order. We leverage this SIC decoding order and utilize the criterion for successful decoding to develop the information-theoretic performance limitation for OMP, which involves factors such as dictionary dimension, signal-to-noise-ratio, and importantly, the relative behavior of the non-zeros entries. Supported by computer simulations, our proposed criterion is demonstrated to be asymptotically effective in explaining the behavior of OMP.

I. INTRODUCTION

The problem of overcomplete representations and the computation of a sparse solution to the associated underdetermined linear system of equations has recently received much attention [1], [2], [3], [4]. It arises in many applications such as the biomagnetic inverse problems, image compression, bandlimited extrapolation and spectral estimation, channel equalization, sensor networks [5], [6], etc. The underlying sparse recovery problem is to represent a signal of interest by using the minimum number of vectors from the overcomplete dictionary. Solving the original problem involves an exhausted search over all possible solutions, which is NP-hard. To recover the sparse solution in practice, convex relaxation methods, typically represented by ℓ_1 -norm minimization, have proven to be effective in both noiseless [1], [7] and noisy settings [8], [2]. In contrast, greedy forward sequential selection methods, e.g. Matching Pursuit and Orthogonal Matching Pursuit, have also shown to be effective in solving the same problem [3] with lower computational burden [9]. Recently, Bayesian learning strategy [10] also emerges in the spotlight as a successful solver for sparse approximation problems.

Given various types of methods, it's natural and important to compare their abilities to reconstruct sparse solutions under various conditions. Since greedy sequential type methods are computationally simple and quite effective, our work focuses on the fundamental performance limitations associated with them, especially, by studying the behavior of a well-known representative—Orthogonal Matching Pursuit (OMP). Analysis in [3], [9] provides mathematical support for OMP. However, results in [1] also indicates OMP may not compete convex relaxation methods in challenging setting. To further the understanding of matching pursuit algorithms, we exploit the connection between sparse solution problems and wireless communication channels [11], namely that reconstructing sparse source can be modeled as decoding the competing users simultaneously transmitting information through a Multiple Access Channel (MAC). OMP can be viewed as a successive interference cancellation (SIC) method [12],

[13] due to its sequential selective nature. Based on this important analogy, starting at the necessary and sufficient conditions for an ideal successive fashion algorithm to succeed, we reach an important criterion, involving dictionary dimension as well as the nature of sources, that asymptotically explains the behavior of OMP, and naturally serves as the performance limit, upper-bounding the chance of success via OMP. This criterion also reveals the utility of highly overcomplete dictionaries. By further identifying different decoding possibilities for different sub-regions in the MAC capacity region, our approach provides information-theoretic insight into the potentially superior performance of joint decoding methods over successive type of methods.

II. BACKGROUND

A. Inverse problem with sparsity requirement

We consider the signal model with measurement noise as following,

$$\mathbf{b} = A\mathbf{x} + \mathbf{n} \quad (1)$$

where $A \in \mathcal{R}^{M \times N}$ with $M < N$, and, usually, $M \ll N$. A is often referred to as the overcomplete dictionary, with each column $\mathbf{a}_i$ as an atom. $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathcal{R}^N$, is the source signal to be recovered. We emphasize that the source is sparse, which means only K entries in $\mathbf{x}$ are nonzero, where $K \ll N$. $\mathbf{n} \in \mathcal{R}^M$, is measurement noise corrupting the measurements. In this work, each element of $\mathbf{n}$ is assumed to be i.i.d. Gaussian as $N(0, \sigma_n^2)$. $\mathbf{b} \in \mathcal{R}^M$, is the measurement vector. For notation simplicity, define integer set $\mathbb{N}_Q \triangleq \{1, 2, \dots, Q\}$.

B. Orthogonal Matching Pursuit (OMP)

Orthogonal matching pursuit is a greedy algorithm that utilizes sequential forward selection to determine the signal representation within certain number of steps. At iteration k , OMP selects atom $\mathbf{a}_{p_k}$ as part of signal representation such that

$$p_k = \arg \max_i \left| \mathbf{a}_i^T \mathbf{r}_{k-1} \right|, \quad i \in \mathbb{N}_N \quad (2)$$

where $\mathbf{r}_{k-1}$ is the residual signal from the $(k-1)$ th iteration. Note that (2) simply correlates the residual signal $\mathbf{r}_{k-1}$ with every atom in dictionary A , and then chooses the atom contributing the most in terms of signal strength.

To remove the contribution of $\mathbf{a}_{p_k}$, OMP compute the new residual signal $\mathbf{r}_k$ as follows,

$$\mathbf{r}_k = P_{S_k}^\perp \mathbf{r}_{k-1} \quad (3)$$

where S_k is the index set tracking the indices of all selected atoms up to iteration k , i.e. $S_k = \{p_1, p_2, \dots, p_k\}$. P_{S_k} is the orthogonal projection onto the subspace spanned by all atoms indexed by set

S_k . $P_{S_k}^\perp$ denotes the orthogonal complement projection of P_{S_k} . By definition, $P_{S_k}^\perp = I - P_{S_k}$.

By setting $\mathbf{r}_0 = \mathbf{b}$, iteratively operating on (2) and (3) until certain stopping criterion is satisfied will yield the sparse signal representation.

Using basic properties of orthogonal projections, it's straightforward to show that the residual signal $\mathbf{r}_k$ could be represented by

$$\mathbf{r}_k = \sum_{\substack{i=1, \\ i \notin S_k}}^N x_i \mathbf{a}_i - \sum_{\substack{i=1, \\ i \notin S_k}}^N x_i P_{S_k} \mathbf{a}_i + P_{S_k}^\perp \mathbf{n} \quad (4)$$

where the second term at the right-hand-side can be viewed as signal distortion introduced by orthogonal projections, due to correlations between atoms in dictionary A .

C. Gaussian multiple access channel (MAC)

The multiple access channel is illustrated in Fig. 1. For $i \in \mathbb{N}_{K_c}$, $\mathbf{a}_i$, with corresponding channel gains h_i , are K_c inputs to the channel. Only one receiver observes the noisy channel output $\mathbf{b}$ contaminated by white Gaussian noise $\mathbf{n}$, where $\mathbf{n} \sim N(0, \sigma_n^2)$. Note that the channel gains h_i could be either deterministic, or drawn according to certain probability distributions. In the latter case, Fig. 1 is usually termed as a slowly fading Gaussian MAC.

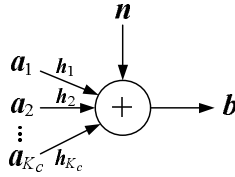


Fig. 1. Multiple access channel.

To transmit information through the channel, user j , $j \in \mathbb{N}_{K_c}$, will have access to a codebook $\mathcal{C}^{(j)} = \{\mathbf{c}_1^{(j)}, \dots, \mathbf{c}_{N_c^{(j)}}^{(j)}\}$, where $\mathbf{c}_i^{(j)}$, $i \in \mathbb{N}_{N_c^{(j)}}$, are codewords of length M_c . The Rate R_j of user j 's codebook is defined as

$$R_j = (\log N_c^{(j)}) / M_c \quad (5)$$

Each user will select a codeword from its codebook and send it through the channel, making M_c uses of the channel. To characterize the limited channel resource shared by competing users and the potential tradeoffs, a *capacity region* is usually computed to represent all possibilities of admissible user-rate allocations for successful transmission. Assume all users hold the same power constraint, i.e. $\text{var}(\mathbf{a}_i) \leq \sigma_a^2$, $i \in \mathbb{N}_{K_c}$. The capacity region of this MAC is then described by [14],

$$\sum_{i \in T} R_i \leq 0.5 \log \left(1 + \frac{\sigma_a^2}{\sigma_n^2} \sum_{j \in T} h_j^2 \right), \quad \forall T \subseteq \mathbb{N}_{K_c} \quad (6)$$

Note that Eq. (6) consists of $(2^{K_c} - 1)$ inequalities. Equalities hold when all channel inputs are i.i.d. Gaussian, i.e. $\mathbf{a}_i \sim N(0, \sigma_a^2)$, $i \in \mathbb{N}_{K_c}$.

III. CONNECTION BETWEEN LINEAR INVERSE PROBLEM WITH SPARSITY CONSTRAINT AND MAC

Let's suppose throughout the paper that s_i , $i \in \mathbb{N}_K$, are the indices of the K non-zero entries in $\mathbf{x}$. To motivate our discussion, first note

that, by peeling off columns due to the zero entries in $\mathbf{x}$, the *effective* form of Eq. (1) is given by

$$\mathbf{b} = x_{s_1} \mathbf{a}_{s_1} + x_{s_2} \mathbf{a}_{s_2} + \dots + x_{s_K} \mathbf{a}_{s_K} + \mathbf{n} \quad (7)$$

Eq. (7) can be contrasted with a scalar Gaussian MAC depicted in Fig. 1 and mathematically determined by

$$b = h_1 a_1 + h_2 a_2 + \dots + h_K a_K + n \quad (8)$$

where $\mathbf{n} \sim N(0, \sigma_n^2)$. To bridge model (7) and (8), we now briefly discuss the key connections between the two domains [11].

1. Existence of one non-zero entry as one user to the MAC.

Direct comparison between (7) and (8) results in this observation.

2. $\mathbf{a}_{s_i}$ as codeword. We can treat the overcomplete dictionary A as a codebook with each column vector $\mathbf{a}_j$, $j \in \mathbb{N}_N$, as a codeword. Each element of $\mathbf{a}_{s_i}$ is fed one by one to the channel (8) as input a_i , resulting in M uses of the channel. Also, $\mathbf{a}_{s_i}$ can be viewed as stacking M transmissions of channel input a_i . Noise $\mathbf{n}$ and observation $\mathbf{b}$ can be related to channel noise $\mathbf{n}$ and channel output $\mathbf{b}$ in the same fashion.

3. Randomness in index s_i . For the inverse problem, we assume that the index s_i , $i \in \mathbb{N}_K$, of the non-zero entry can take any value equal-likely over the integer set $\mathbb{N}_N$. This, in the communication context, is akin to the user selecting randomly a codeword from the codebook $\mathcal{C}$ of size N , with each codeword of length M , for transmission.

4. x_{s_i} as channel gain. The channel gain h_i plays the same role as the value of source x_{s_i} . Assuming the location s_i of a non-zero entry in $\mathbf{x}$ is independent of the value of that source x_{s_i} translates into the fact that codeword selection is independent of the channel gains, which ensures a valid connection to MAC. Essentially, we can interpret the vector representation of (7) as M consecutive uses of the underlying Gaussian channel (8) with appropriate stacking of the inputs/outputs into vectors.

5. Similarity of objectives. To complete the analogy, we examine the goals in the two domains. For the sparse recovery problem, we aim at identifying the non-zero indices and source activities. In the communication context, the goal is successful decoding. Based on the above-mentioned connections, the reconstruction of sparse solutions in overcomplete representation problem is tantamount to identifying the correct codeword, i.e. successful decoding, in MAC communication.

IV. PERFORMANCE LIMIT FOR OMP

A. OMP as a successive interference cancellation scheme

In communication domain, when multiple users are present, successive interference cancellation (SIC) proves to be a valid candidate for decoding purpose. The basic characters of any SIC scheme can be summarized as follows.

- 1) Decode one user's information by treating all other undecoded users as interferences;
- 2) Remove the contribution of the newly decoded user's information from the residual signal.

By iterating over these two steps until certain stop criterion is satisfied, we will obtain the estimation of information sent by each user.

Clearly, based on the connection between linear inverse problem and multiple access channel established in section III, Orthogonal

Matching Pursuit, whose recipe is given in Eq. (2) and (3), can be viewed as a valid SIC decoding scheme. It first chooses a user whose codeword has the maximum correlation with the residual signal, and then removes its contribution by projecting the residual signal onto the orthogonal complement of the space spanned by previously decoded codewords. This connection will provide us with great opportunity to obtain insights into the fundamental performance limitation for OMP.

Though not the main focus of this paper, we believe it's inspiring to evaluate how close the performance of OMP is to an ideal successive cancellation scheme. Literature suggests, as M increases, at decoding step, the probability of p_k being equal to the true index converges to 1; and at removing step, the signal distortion term in (4) tends to introduce weaker disturbance, resulting in cleaner interference cancellation. Interested readers are referred to [9], [15] for detail analysis.

B. Criterion for OMP to success

To establish the fundamental performance limit of OMP, we employ the decoding criterion for SIC scheme to be successful. For ease of exposition, we first discuss the simple case where only two users ($K = 2$) access a MAC, whose capacity region is depicted in Fig. 2.

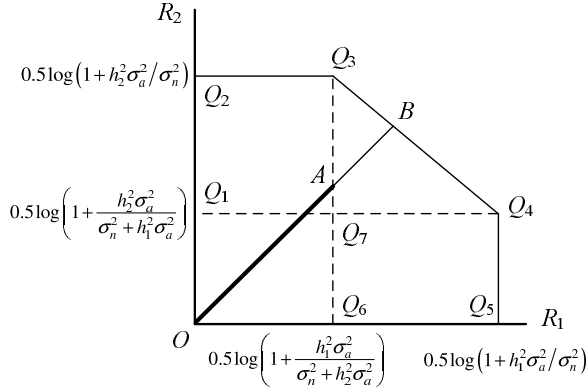


Fig. 2. Capacity region for AWGN MAC with two users.

The capacity region is given by the pentagon $OQ_2Q_3Q_4Q_5$. Note that in rectangular region $OQ_2Q_3Q_6$, where

$$R_1 \leq 0.5 \log \left[1 + \frac{h_1^2 \sigma_a^2}{\sigma_n^2 + h_2^2 \sigma_a^2} \right] \quad (9)$$

$$R_2 \leq 0.5 \log \left[1 + \frac{h_2^2 \sigma_a^2}{\sigma_n^2 + h_1^2 \sigma_a^2} \right] \quad (10)$$

user 1 can be decoded first by treating user 2's information as interference, which effectively leads to a raised noise level according to (9). After removing the decoded user 1's information from the received signal, (10) indicates user 2 should be successfully decoded as if it's the only single user accessing a Gaussian channel. Rate pairs in rectangular region $OQ_1Q_4Q_5$ can be decoded similarly but in the reverse order. The intersection $OQ_1Q_7Q_6$ can be decoded in an arbitrary order. However, no successive cancellation decoding fashion will be capable to successfully decode when the channel is operating in triangular region $Q_7Q_3Q_4$, which otherwise should be decoded by joint decoding schemes.

Hence, in this simple example, an ideal SIC scheme should be able to successfully decode both users if and only if there exists a user

sequence $\{i_{(1)}, i_{(2)}\}$, $i_{(1)}, i_{(2)} \in \mathbb{N}_2$, such that

$$R_{i_{(j)}} \leq 0.5 \log \left(1 + \frac{h_{i_{(j)}}^2 \sigma_a^2}{\sum_{u=j+1}^2 h_{i_{(u)}}^2 \sigma_a^2 + \sigma_n^2} \right), \forall j \in \mathbb{N}_2 \quad (11)$$

To generalize (11) for a K -user MAC, it can be readily shown that an ideal SIC scheme will successfully decode K users if and only if there exists a user sequence $\{i_{(1)}, i_{(2)}, \dots, i_{(K)}\}$ such that

$$R_{i_{(j)}} \leq 0.5 \log \left(1 + \frac{h_{i_{(j)}}^2 \sigma_a^2}{\sum_{u=j+1}^K h_{i_{(u)}}^2 \sigma_a^2 + \sigma_n^2} \right), \forall j \in \mathbb{N}_K \quad (12)$$

To leverage the important observation above to shed light into the behavior of OMP, we first note that by treating A as a codebook, its rate can be similarly defined as

$$R = (\log N)/M \quad (13)$$

It's worth pointing out that because different non-zero entries mirror different users, we implicitly assume different users won't select the same codeword, or they will collapse into one user otherwise. Meanwhile, all users have access to a common codebook A . Although these factors may have a negative impact on achievable capacity region, it is negligible and convenient to assume all users are operating at *equal* rate given in (13) when $K \ll N$.

Next, the connection between sparse solution problem and MAC immediately reveals the following criterion which can be viewed as the counterpart of (12) for OMP:

It's possible for OMP to asymptotically reconstruct the sparse solution successfully if there exists a sequence comprised by all values of non-zero sources, say $\{x'_{(1)}, \dots, x'_{(K)}\}$, such that¹

$$\frac{\log N}{M} \leq 0.5 \log \left(1 + \frac{x'_{(i)}^2}{\sum_{u=i+1}^K x'_{(u)}^2 + \text{SNR}^{-1}} \right), \forall i \in \mathbb{N}_K \quad (14)$$

where $\text{SNR} \triangleq \sigma_a^2 / \sigma_n^2$, and σ_a^2 can be treated as the variance of each atom in A . Criterion (14) can be viewed as a straightforward interpretation of (12) into the sparse solution problem.

C. Insights into the performance limit of OMP

We now provide some observations that can help with the understanding of criterion (14).

1) *Asymptotical Guidance*: Our result (14) makes asymptotic sense, which is a direct observation from information-theoretic argument. Formally, define events:

$E = \{\text{There exists a sequence of } \{x'_{(1)}, \dots, x'_{(K)}\} \text{ s.t. (14) is satisfied.}\}$

$C = \{\text{OMP correctly decodes all the atoms in dictionary } A \text{ participating in true signal representation.}\}$

¹It should be noted that the capacity result (6) and its implication (12) are valid when receiver can track channel gains. In our linear inverse problem, the source activities, which in turns are the channel gains, are typically unknown and to be estimated. This actually brings a negative impact on the achievable capacity region. Hence, our analysis operates on an outer-bound. In practice, the outer-bound in (6) may be close to the true region.

Then, our results indicates, at best,

$$P(C|E) \rightarrow 1, \text{ as } M \rightarrow \infty \quad (15)$$

and for some $0 < \epsilon \leq 1$,

$$P(C|\bar{E}) \leq 1 - \epsilon, \text{ as } M \rightarrow \infty \quad (16)$$

where $\bar{E}$ means the complement event of E .

Here, we argue that (14) becomes informative when we operate in high signal dimensions, i.e. M and N are reasonably large. Meanwhile, it's suggested that dictionary A 's elements should be generated according to i.i.d. Gaussian distribution to maximize the performance.

2) *Insights into OMP's behavior*: Criterion (14) unveils the information-theoretic limits on performance of OMP, and provides novel explanation to its behavior, in terms of its success/failure determined by model parameters, such as degree of overcompleteness, i.e. M and N , signal-to-noise ratio, and the nature of the underlying sources, i.e. the strength of source activities. Especially, asymptotically perfect signal recovery is possible even if N grows at exponential order of M , i.e. $N = O(\beta^{RM})$, assuming all logarithms are base- β . This phenomenon coincides with important discoveries obtained via other analysis techniques in [9], which states the possibility of signal reconstruction with dictionary dimension at similar order.

Furthermore, the common codebook assumption requires the underlying MAC to work within the equal user-rate region, demonstrated by line $\overline{OB}$ in Fig. 2 for $K = 2$. OMP, because of its successive nature, is at best able to accomplish signal reconstruction only up to line $\overline{OA}$ (thick), which is a low-rate portion of $\overline{OB}$. We consider this comparison to be enlightening since (i) it shows the sub-optimality of OMP scheme, and (ii) it suggests that potential improvement could be enjoyed by using joint decoding methods. These observations echo the recent discoveries in [1], [2], where convex relaxation methods (which jointly estimate the sources via ℓ_1 -norm minimization) outperform OMP in challenging settings.

3) *Nature of Source matters*: From criterion (14) and its implications (15) and (16), it can be concluded that the nature of source intensities plays an important role in OMP's performance. The following two intensity profiles offer some insight.

Proposition 1: For sources with equal intensity, i.e. $|x_i| = c, \forall i$, the ability to decode all sources is equivalent to decoding any one of them at the first step.

Proposition 2: For exponentially decaying sources, i.e. $|x_i| = c^{i-1}$ with some constant $0 < c < 1$, then for any $k \in \mathbb{N}_{K-1}$, there exists c' s.t. when $c < c'$, OMP cannot recover the atoms with k smallest sources in magnitude.

These two propositions extend the discussion in [16] to noisy setting. They can be viewed as the worst-case, where all sources are of equivalent scale, and best-case scenarios, where all sources are of different scales, for OMP, respectively. Moreover, [17] studies the effect of random source activities, which is, in parallel, related to outage analysis in wireless communication domain. Meanwhile, it's interesting to note that in any experimental study, one may want to take this factor into concern since the mechanics used to generate sources may have significant impact on the performance via certain decoding algorithm.

4) *Relation to other methods*: We point out that criterion (14) makes sense for a broad spectrum of successive type algorithms. Hence, similar analysis should hold for various greedy search algorithms, such as Basic Matching Pursuit (BMP), Recursive Orthogonal Matching Pursuit [5], etc. The difference lies in the gap between the optimal recoverability indicated by (14) and the ability of a specific forward selection algorithm. We choose OMP as a reasonable representative because of its superior performance over BMP and its relatively tractable analysis over much more sophisticated algorithms. Our result also advises caution, when proposing new algorithms, in balancing the computational complexity and the performance improvement, especially if existing algorithms already perform closely to the information-theoretic performance limit.

V. EXPERIMENTS

We present experiments to justify our theoretic analysis. All dictionaries have elements generated according to i.i.d. Gaussian distribution, and then with every column normalized to unit length. A large number of random realizations are processed for each experiment.

A. Correlation between SIC criterion and OMP's performance

We intend to verify that SIC criterion (14) exhibits practical guidance on OMP's behavior. To this end, we perform OMP in two different scenarios: one with underlying signal sources satisfying (14), where asymptotically perfect recovery is possible; one with sources failing the test in (14), which should result in incorrect reconstruction. To evaluate performance, we compute the percentage of OMP's results that correctly recover *all* non-zero locations in both scenarios, as shown in Fig. 3.

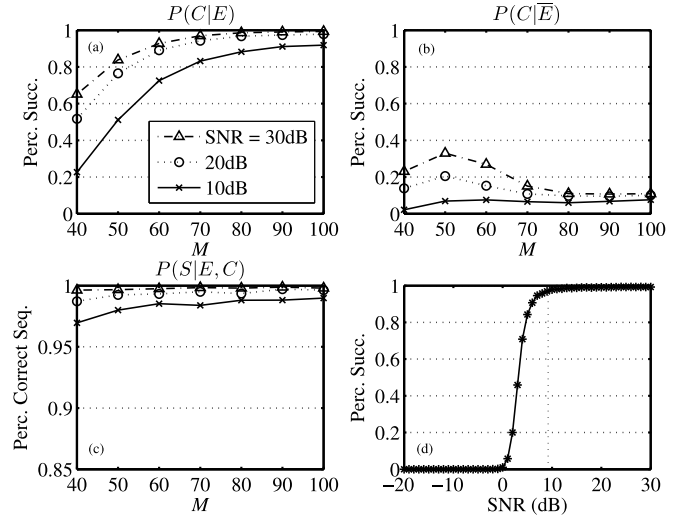


Fig. 3. OMP's performance in different scenarios. In (a) (b) (c): $N = 10^4, K = 7$. In (d): $M = 100, N = 1000, K = 7, x_i = 0.4, \forall i$.

Fig. 3 (a) and (b) illustrates the empirical probabilities of $P(C|E)$ and $P(C|\bar{E})$. The results clearly justify (15) and (16). As M increases, for data composed of sources passing the test (14), uptrend momentum propels OMP to approach near perfect performance. In contrast, for data synthesized by signal sources failing the test, OMP tends to be highly incapable to reconstruct them, resulting in a very low success rate.

To obtain a closer view of OMP's behavior, we define event:

$S = \{\text{The atom indices, ordered according to OMP's recovery stages, form a valid sequence that meets (14).}\}$

Fig. 3 (c) shows the finer scale scrutiny by examining $P(S|E, C)$, the probability that OMP proceeds at the correct order given that a sequence defined by (14) is available and OMP somehow identifies all of its ingredients. Note that this probability is very high in each case, showing when OMP can recover all atoms, it indeed prefers a sequence according to (14).

Fig. 3 (d) simulates a smooth transition from an undecodable to a decodable model by varying a model parameter. We hope to see the performance evolution of OMP as condition changes. The vertical dashed line shows the boundary ($\text{SNR} = 9.2\text{dB}$) between two regions. Clearly, at low SNR, OMP is very likely to fail. At high SNR, OMP can succeed with very high probability. Before crossing the watershed, OMP starts to be able to decode with increasing probability. Note that the intersection occurs at overwhelming probability of success. This fact strongly agrees with our theoretic analysis.

The practical guidance of (14) is exploited by these discoveries.

B. Stable reconstruction with highly overcomplete dictionary

The scale of overcompleteness that may lead to successful signal recovery has been pursued as a practically important issue. In [2], [18], [19], convex relaxation methods, e.g. Basis Pursuit, prove to be effective to recovery the sparse source with signal dimension N at exponential order of number of measurements M when dictionary atoms are generated from random ensembles. [9] starts from error probability analysis and reaches the similar order for OMP to succeed. To echo this discussion, we realize that criterion (14) uncovers the fact that OMP would possibly be able to accomplish the reconstruction as long as $N = O(\beta^{RM})$. This suggests great opportunity for recovery that increasing the number of measurements by 1 may greatly boost the scale of the underlying problem. Fig. 4 justifies our consideration.

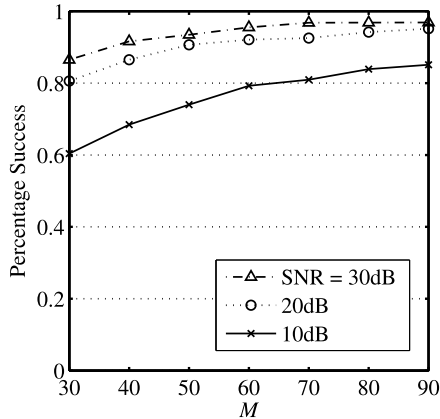


Fig. 4. OMP is able to reconstruct signal using highly overcomplete dictionaries. ($N = \lfloor 2^{0.2M} \rfloor$, $K = 7$)

In this experiment, all data processed by OMP pass the test (14). Clearly, OMP shows sufficient ability to reconstruct *all* non-zero entries as M and N maintain the exponential relationship. Smaller error probability is obtained for larger M , which is in accordance with (15). Hence, our result provides an information-theoretic perspective of interpreting parallel discovery in previous literatures.

VI. SUMMARY

We established the connection between sparse solution problem and channel coding problem. OMP can be then viewed as a successive interference cancellation scheme. By leveraging the decoding criterion for SIC schemes, we derive the underlying criterion that determines OMP's behavior, particularly fundamental performance limits and an explanation to the performance discrepancy with joint decoding methods. Together with computer simulations, we demonstrate the efficiency of this criterion, and the utility of highly overcomplete dictionaries.

ACKNOWLEDGMENT

This research was supported by NSF grant DGE-0333451, the RESCUE project NSF grant 0331690, and MICRO grant #07-034.

REFERENCES

- [1] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution," *Preprint*, 2004.
- [2] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, 2006.
- [3] J. A. Tropp, "Greedy is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theo.*, 2004.
- [4] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," *Proc. Int'l. Symp. Info. Theo.*, June 2007.
- [5] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, 2005.
- [6] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, 2006.
- [7] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [8] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 -norm near-solution is also the sparsest near-solution," *Preprint*, 2004.
- [9] J. A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theo. preprint*, 2007.
- [10] D. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704–3716, 2007.
- [11] Y. Jin and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," *ICASSP*, 2008.
- [12] A. C. Gilbert and J. A. Tropp, "Applications of sparse approximations in communications," *ISIT*, 2005.
- [13] G. Karabulut, T. Kurt, and A. Yongaçoglu, "Blind estimation and multiuser detection for wireless cdma systems," *IEEE ICC*, pp. 4953–4958, 2006.
- [14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.
- [15] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *preprint*, 2006.
- [16] D. Wipf, "Bayesian methods for finding sparse representations," *Ph.D. dissertation, Univ. Calif., San Diego*, 2006.
- [17] D. Wipf and B. D. Rao, "Comparing the effects of different weight distributions on finding sparse representations," *NIPS*, 2006.
- [18] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [19] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.