Generalized orthogonal matching pursuit for multiple measurements - A structural approach

Florian Bölßmann*

Abstract—Sparse data approximation has become a popular research topic in signal processing. However, in most cases only a single measurement vector (SMV) is considered. In applications, the multiple measurement vector (MMV) case is more usual, i.e., the sparse approximation problem has to be solved for several data vectors coming from closely related measurements. Thus, there is an unknown inter-vector correlation between the data vectors. Using SMV methods typically does not return the best approximation result as the correlation is ignored. In the past few years several algorithms for the MMV case have been designed to overcome this problem. Most of these techniques focus on the approximation quality while quite strong assumptions to the type of inter-vector correlation are made.

While we still want to find a sparse approximation, our focus lies on preserving (possibly complex) structures in the data. Structural knowledge is of interest in many applications. It can give information about e.g., type, form, number or size of objects of interest. This may even be more useful than information given by the non-zero amplitudes itself. Moreover, it allows efficient post processing of the data. We numerically compare our new approach with other techniques and demonstrate its benefits in two applications.

Index Terms—sparse approximation, multiple measurements, greedy algorithm, inter-signal correlation

I. INTRODUCTION

Sparse approximations of given data are of great interest in many different applications. They are used in image processing for e.g., denoising [1], [2], compression [3] or restoration [4]. Sparsity assumptions appear in face and speech recognition [5], [6], magnetic resonance imaging (MRI) and computer tomography (CT) [7], [8], as well as in non-destructive testing [9], [10] and seismic data processing [11], [12], [13], [14]. A detailed overview can also be found in [15], [16] and the references therein.

The sparse approximation itself can be stated as follows: Given a dictionary matrix $D \in \mathbb{C}^{N_1 \times N_3}$ and a measurement vector $b \in \mathbb{C}^{N_1}$, solve

$$\min_{x \in \mathbb{C}^{N_3}} \|x\|_0 \text{ s.t. } \|Dx - b\|_2 \leq \varepsilon \quad (1)$$

for a given $\varepsilon > 0$. Here $\|x\|_0$ is the $\ell_0$-quasi-norm, i.e., $\|x\|_0 := |\{k \mid x_k \neq 0\}|$. The vector $x$ is said to be $L$-sparse, if $\|x\|_0 \leq L$ for $L \in \mathbb{N}$. The matrix $D$ is constructed using a basis, frame or dictionary in which the given data vector $b$ is assumed to be sparse. Typical examples are Fourier [17] or Wavelet bases [18] as well as Curvelet [19] or Shearlet frames [20]. Dictionaries can be designed according to the underlying application, as e.g., the Gabor impulse in ultrasonic testing [9], [21].

The exact solution of (1) can in general only be found combinatorially, i.e., by considering all possible supports of $x$. Hence, finding the exact solution becomes NP-hard. There are two main strategies to find at least an approximate solution of (1): The first strategy is, replacing the $\ell_0$-quasi-norm by the $\ell_1$-norm what makes the problem convex. This approach is known as convex relaxation or basis pursuit [22], [23]. Greedy algorithms are another strategy to solve (1) approximatively. Those methods iteratively built up a global approximation by solving local subproblems [25]. Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP) [23] may be the most known algorithms in this context. In recent years, more advanced algorithms have been developed such as Stagewise OMP [26], Compressive Sampling MP [27], [28] and regularized OMP [27], [29]. An overview can be found in [30].

Eq. (1) is known as single measurement vector problem (SMV). It has been studied extensively over the last few years. However, there is an extension known as multiple measurement vector problem (MMV). Instead of only having one data vector $b \in \mathbb{C}^{N_1}$, several measurements $b := (b^1, \ldots, b^{N_2}) \in \mathbb{C}^{N_1 \times N_2}$ are given. The problem is stated similar as

$$\min_{X \in \mathbb{C}^{N_1 \times N_3}} \|X\|_{0,\infty} \text{ s.t. } \|DX - b\|_2 \leq \varepsilon \quad (2)$$

with $X = (x^1, \ldots, x^{N_2})$ and $\|X\|_{0,\infty} = \max_k \|x^k\|_0$, i.e., each vector $x^k$ is sparse. In fact this problem seems to be more common in many applications. It appears e.g., in non-destructive testing [9], [10], in seismic data [11], [12], [13], [14] or Magnetoencephalography (MEG) [24]. The MMV formulation can be used, whenever several measurements of the same or similar objects were made. One straight-forward approach to solve (2) is, to split it into $N_2$ SMV problems and solve these independently using the methods mentioned above. However, our intuition tells us, since the $N_2$ measurements were made in a quite similar set-up, also the obtained data vectors $b^1, \ldots, b^{N_2}$ should be correlated somehow. Simply solving $N_2$ SMV problems ignores this correlation and thus the solution quality might suffer.

Recently, new methods have been developed that consider inter-signal correlation. In [24] an extension for MP and the FOCal Underdetermined System Solver (FOCUSS) are presented. Bayesian methods are considered in [31], [32]. In [33], [34] the authors introduce Greedy pursuit and convex relaxation for the MMV problem. Theoretical results have been shown e.g., in [35]. All these methods force a common support in the reconstructed solution, i.e., the reconstructed matrix $X$ has only few nonzero rows or, in other words, the columns of $X$ have (nearly) the same support. In [36] two joint sparsity models (JSM) for compressed sensing are

*University of Passau, Mathematics / Digital Image Processing, Florian.Bossmann@uni-passau.de

arXiv:1705.08259v1 [math.NA] 23 May 2017
introduced. JSM-1 considers solutions where all columns $x^k$ can be written as the sum $x^k = x^c + x^{k,u}$ of a common sparse component $x^c$ that is equal for each column and another unique sparse vector $x^{k,u}$. JSM-2 is equal to the support constraint considered above. Another approach is presented in [37], [38] where correlated measurements are assumed to have sparse approximations that are close in the euclidean distance. This idea is related to dynamic compressed sensing [39], [40]. Here neighboring columns are assumed to have similar support. In both cases, the support is allowed to change slowly over different data vectors. In most cases the used methods penalize non-smooth rows in $X$.

However, all methods have quite restricting support assumptions and hence cannot reconstruct simple geometries in the solution. As an example consider $X$ to be the identity matrix. There is no common support between all columns and the rows are non-smooth. Nevertheless, the matrix is still clearly structured. The linear structure can be described by a shift of 1 index per column. In the next section we introduce a generalized version of orthogonal matching pursuit for multiple measurements (GM-OMP) that takes complex structures in the data into account. Numerical evidence for the proposed method are shown in the third section of this work.

II. THE ALGORITHM

In this section we first introduce OMP and discuss its generalization to the MMV problem. GM-OMP increases the structural component of the reconstructed solution. For simplicity we assume that the columns of $D$ are normalized. Then the iterative scheme of OMP can be summarized as follows:

1) Set the residual $r = b$ and the support $I = \emptyset$.
2) Calculate $i = \arg \max |D^* r|_1$ and update $I \leftarrow I \cup \{i\}$.
3) Solve $x = \arg \min \|r - D y\|_2$ and set $r = r - D x$.
4) Iterate 2-3 until a stopping criterion holds.

Here $D^*$ is the transposed conjugate complex matrix. Hence, the algorithm chooses the column of $D$ that correlates most with the residual and adds its index to the support set in step 2. Step 3 calculates the best approximation according to the selected support. The algorithm may e.g., be stopped after $L$ iterations (the solution $x$ is L-sparse then), or when the residuum drops below a threshold, i.e., $\|x\|_2 \leq \varepsilon$.

Now, let us consider the MMV problem shown in (2). The idea of GM-OMP is surprisingly simple. We only adapt the second step of OMP, while all other steps stay the same. Therefore, note that OMP chooses one index $i$ and adds it to the support set $I$. Since we are now dealing with multiple measurements, GM-OMP is allowed to add not only one index $i$, but multiple indices to the support (e.g., one index per column of $X$). Let us denote the set of all indices added by $J$. This index set should be chosen from a set of feasible selections $J \in \mathcal{P} \{\{j,k \mid j \leq N_1, k \leq N_2\}\}$ where $\mathcal{P}(\cdot)$ denotes the power set. The second step of GM-OMP now reads as follows:

2) Choose $J \in \mathcal{J}$ and update $I \leftarrow I \cup J$.

The complete scheme of GM-OMP is shown in Alg. 1 where the maximum number of iterations $L$ and the minimal residuum norm $\varepsilon_R$ is included as stopping criterion. Of course we need to define the feasible set $\mathcal{J}$ and find a suitable choice $J \in \mathcal{J}$. This problem will be discussed in the next subsection. Let us first consider three examples to clarify the principle of GM-OMP and the feasible set $\mathcal{J}$.

Algorithm 1: GM-OMP

Data: $B; D$
Parameters: $\varepsilon_R, L$
Set $R = B$ and $I = \emptyset$;
for $l = 1, \ldots, L$ do
Choose $J \in \mathcal{J}$ and update $I \leftarrow I \cup J$;
Solve $X = \arg \min \|B - D Y\|_2$ and set $R = B - D X$;
if $\|R\|_2 \leq \varepsilon_R$ then stop;

For our first example, consider the set $\mathcal{J} = \mathcal{J}_p = \{\text{supp} M \mid \text{each column of } M \text{ is 1-sparse}\}$, (3) i.e., $\mathcal{J}_p$ contains all index sets that can be associated with the support of matrices $M$ having at most one non-zero element per column. Here $M$ is of same size as $X$. Choose $J \in \mathcal{J}_p$ such that

$(j,k) \in J \iff j = \arg \max_{l \leq J} |D^* l_k|, \quad \text{where } l_k \text{ is the } k\text{-th column of the residual matrix } R.$

This way, GM-OMP is equivalent to OMP parallelly applied to each column of $B$. In our next example define

$\mathcal{J} = \mathcal{J}_v = \{\text{supp } M \mid \|M\|_0 = 1\}$

as the set of all 1-sparse supports. It is easy to see that the best choice $J \in \mathcal{J}_v$ is $J = \{\arg \max |D^* R|\}$ the index set containing only the position of the maximum absolute value of $D^* R$. Now, GM-OMP is identical to using OMP on the vectorized formulation of $\tilde{X}$, i.e., rewrite $\tilde{X}$, $\tilde{B}$ as column vectors and $\tilde{D}$ becomes a block diagonal matrix. As our last example, consider

$\mathcal{J} = \mathcal{J}_s = \{\text{supp } M \mid M \text{ has at most one non-zero row }\}$

containing all support sets with constant row index. A possible choice $J \in \mathcal{J}_s$ is given by

$i = \arg \max_j \|\{D^* R\}_j\|_\lambda \quad \text{and } J = \{(i, j) \mid j \leq N_2\}.$

Here $\|\{D^* R\}_j\|_\lambda$ denotes the $\lambda$-norm of the $j$-th row of the matrix. For $\lambda = 1$ GM-OMP


becomes the simultaneous OMP (S-OMP) introduced in [33], the case $\lambda = 2$ is discussed in [24].

For the three demonstrated choices of $\mathcal{J}$, GM-OMP transforms into well known algorithms for the MMV problem. However, neither $\mathcal{J}_P$ nor $\mathcal{J}_V$ contain structured sets while $\mathcal{J}_S$ is bounded to row-sparsity of $X$ (see discussion in the introduction). Thus, we introduce a more general choice for $\mathcal{J}$ in the next subsection. Here, $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$ can be adapted by parameters. We will see that the three examples form the extreme cases of the parameter choice.

B. Feasible set and selection

Alg. 1 demonstrates the generalized scheme of OMP for multiple measurements. The set $\mathcal{J}$ represents the sparse sparsity patterns that can be chosen per iteration. However, since $\mathcal{J}$ is a subset of a power set, it can be of exponential size. Thus, it is not sufficient to leave it to the user as input data. Instead we will parametrize $\mathcal{J}$ and define a selection rule for $J \in \mathcal{J}$ based on the parameters. This way, the user only has to choose parameters that describe sparsity patterns suitable for his application.

**Remark 1:** Following, we describe a parametrization idea that, in the authors opinion, can be used in many applications. However, the reader might choose a different description of $\mathcal{J}$ and an according selection rule $J \in \mathcal{J}$ that is more suitable for the particular problem.

Note that $J \in \mathcal{J}$ is a set of two-dimensional elements (row and column indices). Our idea is, to use exactly two parameters $\alpha, \gamma$ to determine $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$. It is clear that we cannot cover all sets $\mathcal{J}$ with two parameters, since the number of possible choices for $\mathcal{J}$ grows exponential. Thus, we need a parametrization that generates suitable sets for applications. Analogous to the given examples (1)-(5), we identify an extreme cases of the parameter choice.

Analogous to the given examples (3)-(5), we identify an extreme cases of the parameter choice.

For the general setup of $\mathcal{J}$, GM-OMP transforms into well known algorithms for the MMV problem. However, neither $\mathcal{J}_P$ nor $\mathcal{J}_V$ contain structured sets while $\mathcal{J}_S$ is bounded to row-sparsity of $X$ (see discussion in the introduction). Thus, we introduce a more general choice for $\mathcal{J}$ in the next subsection. Here, $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$ can be adapted by parameters. We will see that the three examples form the extreme cases of the parameter choice.

**Remark 1:** Following, we describe a parametrization idea that, in the authors opinion, can be used in many applications. However, the reader might choose a different description of $\mathcal{J}$ and an according selection rule $J \in \mathcal{J}$ that is more suitable for the particular problem.

Note that $J \in \mathcal{J}$ is a set of two-dimensional elements (row and column indices). Our idea is, to use exactly two parameters $\alpha, \gamma$ to determine $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$. It is clear that we cannot cover all sets $\mathcal{J}$ with two parameters, since the number of possible choices for $\mathcal{J}$ grows exponential. Thus, we need a parametrization that generates suitable sets for applications. Analogous to the given examples (1)-(5), we identify an extreme cases of the parameter choice.

Analogous to the given examples (3)-(5), we identify an extreme cases of the parameter choice.

For the general setup of $\mathcal{J}$, GM-OMP transforms into well known algorithms for the MMV problem. However, neither $\mathcal{J}_P$ nor $\mathcal{J}_V$ contain structured sets while $\mathcal{J}_S$ is bounded to row-sparsity of $X$ (see discussion in the introduction). Thus, we introduce a more general choice for $\mathcal{J}$ in the next subsection. Here, $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$ can be adapted by parameters. We will see that the three examples form the extreme cases of the parameter choice.

**Remark 1:** Following, we describe a parametrization idea that, in the authors opinion, can be used in many applications. However, the reader might choose a different description of $\mathcal{J}$ and an according selection rule $J \in \mathcal{J}$ that is more suitable for the particular problem.

Note that $J \in \mathcal{J}$ is a set of two-dimensional elements (row and column indices). Our idea is, to use exactly two parameters $\alpha, \gamma$ to determine $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$. It is clear that we cannot cover all sets $\mathcal{J}$ with two parameters, since the number of possible choices for $\mathcal{J}$ grows exponential. Thus, we need a parametrization that generates suitable sets for applications. Analogous to the given examples (1)-(5), we identify an extreme cases of the parameter choice.

Analogous to the given examples (3)-(5), we identify an extreme cases of the parameter choice.

For the general setup of $\mathcal{J}$, GM-OMP transforms into well known algorithms for the MMV problem. However, neither $\mathcal{J}_P$ nor $\mathcal{J}_V$ contain structured sets while $\mathcal{J}_S$ is bounded to row-sparsity of $X$ (see discussion in the introduction). Thus, we introduce a more general choice for $\mathcal{J}$ in the next subsection. Here, $\mathcal{J} = \mathcal{J}(\alpha, \gamma)$ can be adapted by parameters. We will see that the three examples form the extreme cases of the parameter choice.


For implementation we recommend to replace (C)1, ﬁrstly set (C)1, ̸= 0 by (C)1, > 1 using a reasonable threshold 1 (see also the discussion in the theory part). Furthermore, if C = [D∗R] from start on has zero entries (or elements below the threshold), they will never be chosen by the algorithm. This assures that the support is not artificially enlarged, i.e., supp M ⊆ supp C always holds.

It is easy to see that Alg. 2 returns the exact solution for (α, γ) = (1, 1) and (α, γ) = (1, 1) (i.e., for Jp, Jp). Setting (α, γ) = (1, 1) (i.e., Jp) Alg. 2 solves (1) for λ = ∞.

C. Theoretical results

We first prove Theorem 1 that is the motivation for Alg. 2.

Proof of Theorem 1: We give a polynomial-time reduction of the coloring problem: Given a graph G and a number of colors C, assign a color to each vertex v ∈ G such that for each edge uv ∈ E G the vertices u, v have a different color.

Let G have M vertices, each vertex with at most n edges. For the reduction we need each vertex to have the same amount of edges. Thus, we add edges vj to each vertex vj until it has exactly n edges. Here vjr denotes an edge with no second vertex (or an “imaginary” vertex that will be considered for the coloring). Let the total amount of edges be N by where n ≤ N ≤ Mn. Now deﬁne mi = ei ∈ RN as the i-th unit vector, i.e., ‖mi − ωi‖∞ = 1 for i ̸= i. Furthermore, deﬁne χk ∈ RN, k = 1, . . . , M by

\[(\chi_k)_j = \begin{cases} 1 & \text{the j-th edge starts at } v_k, \\ 0 & \text{otherwise.} \end{cases} \]

Note that χk is exactly n-sparse. For the j-th unit vector e′ j in RN set χk = e′ j ⊗ χk ∈ RN where ⊗ is the Kronecker product. (For simplicity we keep the two-dimensional indexing of pjk instead of reordering it into a single index.) We obtain

\[\|p_{jk} − p_{j′,k′}\|_2 = \begin{cases} \sqrt{2n} & j ̸= j′, \\ \sqrt{2n} & j = j′ \text{ but } \forall k \in \mathbb{K}, v_k ∉ G, \\ \sqrt{2n+2} & j = j′ \text{ and } \forall k \in \mathbb{K}, v_k ∉ G, \end{cases} \]

i.e., for different colors or vertices which are not connected we obtain \(\sqrt{2n}\). Now choose α = ∞ and γ = ∞, then (8) only holds for pairs (j, k), (j′, k′) with either different colors \(j ≠ j′\) or not connected vertices \(v_k ∉ G\). Set D as identity matrix and \(R ∈ \mathbb{R}^{CM×M}\) to

\[(R)_{(j,k),j} = \begin{cases} 1 & k = j, \\ 0 & \text{otherwise.} \end{cases} \]

Then, there is a feasible coloring of G if and only if \(\|(D∗R)\|_1 ≤ M^{1/λ}\) for λ = 0 and \(J\) solution of (11). By construction of R the value \(M^{1/λ}\) can only be achieved with \(J = M\) and \(R)_{j,j} = 1\) for all \((j, j)\) which is the i-th vertex is assigned with the j-th color. Since \(|J| = M\) each vertex is colored. On the other hand, each feasible coloring deﬁnes an index set J ∈ (α, γ) with \(\|(D∗R)\|_1 ≤ M^{1/λ}\) what is maximal since |J| ≤ M for all J ∈ (α, γ).

Following, we prove reconstruction results for GM-OMP given exact data or data obtained with noised sparsity pattern. Therefore, we need results shown in [23]. We brieﬂy summarize Theorem 3.1, Corollary 3.2 and Theorem 3.5 of this work for the reader: Given the Babel function

\[μ_1(l) = \max_{a<\infty} \max_{\|\Omega\|_a\leq 1} \|DΩ\|_a.Ω,l\|, \]

OMP recovers the L-sparse solution of (1) in the noiseless case (1 = 0) whenever \(μ_1(L)+1 = μ(L−1)\). There exists a weak version of OMP that chooses an index i in each iteration such that \(\|(D∗p_i)\|_2 ≥ λ\) max \(\{(D∗p_l)\} l\leq l\) with a weak constant λ ≤ 1 holds. Weak OMP recovers an L-sparse solution whenever \(μ_1(L) = λ(1−μ(L−1))\).

Let \(R^I\) be the residual matrix after l iterations and \(J_l = J_l\) the greedy choice returned by Alg. 2. We calculate the weak parameter of GM-OMP by

\[\lambda = \min \left\{ \frac{\|\Omega\|_1}{\|\Omega\|_1}, (i, j) ∈ J_l, l = 1, . . . , L \right\}. \]

Note that λ can be very small depending on the range of amplitudes. Exemplary, consider

\[\frac{|\lambda|^{1/λ}}{\|\Omega\|_1} = \begin{pmatrix} 1001 & 999 & \ldots & 1 \\ 1 & 2 & \ldots & 1000 \end{pmatrix} \]

with α = ∞, γ = 0. The greedy choice would either select the first or the second row. In both cases we obtain \(λ = \min \{|1001−k|, k = 1, . . . , 1000\}\) and hence \(λ = 1/1000\). However, \(\lambda\) will be close to 1 for feasible sets where \(\max |\mathcal{X}| < |\mathcal{X}|\) for l < l hold. Before we state our ﬁrst theorem, we need the following deﬁnition.

Definition 1: For L sets \(J_1, . . . , J_L ∈ \mathbb{J}\) we say that the sets are (α, γ)-intersecting if there exists \((i, j) ∈ J_l, (i′, j′) ∈ J_{l′}, l ≠ l′\) such that (7) and (8) hold.

Let \(X\) be the unique L-column-sparse solution of \(B = DX\) and \(\mathcal{X} = \bigcup_{l=1}^{L} J_l\) with \(J_l ∈ \mathbb{J}\). We state

Theorem 2: GM-OMP recovers \(X\) and all elements \(J_1, . . . , J_L\) of the sparsity pattern in L iterations whenever \(J_1, . . . , J_L\) are (α, γ)-intersecting and \(μ_1(L) = λ(1−μ(L−1))\).

Proof: The reconstruction of the solution of \(X\) follows directly by the exactness of weak OMP. Since the sets are not (α, γ)-intersecting the indices selected by Alg. 2 belong to the same set \(J_l\). Suppose an index in \(J_l\) has not been selected. Then
the matrix \( C \) from Alg. 2 is not zero and Alg. 2 will not stop iterating. Thus, the complete set \( J_l \) is recovered.

Theorem 2 has two disadvantages. First, the conditions depend on the solution and hence cannot be checked beforehand. Second, as we have seen \( \lambda \) can be small and thus the conditions are quite restricting. To overcome the last problem, we define \( \beta = \mu_1(L)/(1 - \mu_1(L - 1)) \). Now, we can use an adaptive threshold in Alg. 2 by selecting only indices such that \( \lambda > \beta \) holds, i.e., in the \( l \)-th iteration only indices \((i, j)\) with

\[
\frac{|(D^s R^{l-1})_{i,j}|}{\| (D^s R^{l-1}) \|_m} > \beta
\]

will be selected. We can state

**Theorem 3:** Using this threshold strategy in Alg. 2 GM-OMP recovers the solution \( \tilde{X} \) in \( L \) iterations where \( L \leq L' \leq N_3 L \) whenever \( \beta \leq 1 \). Furthermore, let \( J_1', \ldots, J_l' \) be the feasible sets selected in each iteration and supp\( \tilde{X} = \bigcup_{l=1}^L J_l \). If \( J_1, \ldots, J_l \) are not \((\alpha, \gamma)\)-intersecting, then there exists a partition \( L_1', \ldots, L_l' \) of \( \{1, \ldots, L\} \) such that \( J_l = \bigcup_{l \in L_l'} J_l \).

**Proof:** Note that the first index selected in Alg. 2 is the maximum of \( C = |D^s R| \) and thus equivalent to an OMP choice, i.e., Alg. 2 selects at least one element per iteration. As \( \beta \leq 1 \) this choice is part of supp\( \tilde{X} \). The algorithm terminates after at most \( N_6 L \) iterations what is the number of non-zero entries in \( \tilde{X} \). If the sets \( J_1, \ldots, J_l \) are not \((\alpha, \gamma)\)-intersecting the indices selected by Alg. 2 belong to the same set \( J_l \). Since we use a threshold, we can no longer follow that \( J_l \) is found in one iteration. However, because supp\( \tilde{X} \) is recovered completely after \( L \) iterations, the existence of such a partition follows.

Now, let us discuss the reconstruction qualities of GM-OMP due to noised sparsity patterns. Instead of \( \tilde{B} = D \tilde{X} \) with supp\( \tilde{X} = \bigcup_{l=1}^L J_l \), the noised data \( \tilde{B} = D \tilde{X} \) with supp\( \tilde{X} = \bigcup_{l=1}^L J_l \) is given. Here \( J_l \) is the noised version of the pattern \( J_l \). We consider two different kinds and analyze in which case GM-OMP is able to reconstruct the sets \( J_l, l = 1, \ldots, L \). Given \( J_l \) we can try to recover \( J_l \) using a post processing strategy that will be introduced in the next subsection.

**Theorem 4:** Let \( J_l \) be corrupted by uniform noise \( \varepsilon_a > 0 \):

\[
(i, j) \in J_l \Rightarrow (i, j) \in J_l, d_{\varphi}(p_j, p_{j'}) \leq \varepsilon_a.
\]

Set \( m = \min_{i \neq j} d_{\varphi}(m_i, m_j) \). If \( \delta \) holds for \( J_l \) with parameter \( \gamma \), then \( \delta \) holds for \( J_l \) with Lipschitz parameter \( \overline{\gamma} = \gamma + 2\varepsilon_a/m \). If \( J_1, \ldots, J_l \) are not \((\alpha, \gamma + 4\varepsilon_a/m)\)-intersecting, then \( J_1', \ldots, J_l' \) are not \((\alpha, \gamma + 2\varepsilon_a/m)\)-intersecting.

**Proof:** For \((i_1, j_1), (i_2, j_2) \in J_l \) and \((i_1', j_1'), (i_2', j_2') \in J_{l'} \)

\[
d_{\varphi}(p_j, p_{j'}) \leq d_{\varphi}(p_{j'}, p_{j'}) + d_{\varphi}(p_{j'}, p_{j'}) + d_{\varphi}(p_{j'}, p_{j'}) \leq (\gamma + 2\varepsilon_a/m)d_{\varphi}(m_{i_1}, m_{i_2})
\]

holds. Equivalently, \((i_1, j_1) \in J_l, (i_2, j_2) \in J_l, (i_1', j_1') \in J_{l'}, (i_2', j_2') \in J_{l'} \) with \( l \neq l' \) using that \( J_1, J_{l'} \) are not \((\alpha, \gamma + 4\varepsilon_a/m)\)-intersecting and the inverse triangle inequality:

\[
d_{\varphi}(p_{j'}, p_{j'}) \geq d_{\varphi}(p_{j'}, p_{j'}) - 2\varepsilon_a > (\gamma + 2\varepsilon_a/m)d_{\varphi}(m_{i_1}, m_{i_2}).
\]

The noise assumed in Theorem 4 typically appears in applications where measurements may be corrupted due to shaking apertures. If an upper bound \( \varepsilon_a \) is known, the parameters of GM-OMP can be adapted.

**Theorem 5:** Let \( J_l \) be corrupted by Bernoulli distributed noise \( \varepsilon_B \in [0, 1] \), i.e.,

\[
\tilde{J}_l \subseteq J_l, \quad \Pr{(i, j) \notin \tilde{J}_l \mid (i, j) \in J_l} = \varepsilon_B
\]

where \( \Pr{(i, j) \notin \tilde{J}_l \mid (i, j) \in J_l} \) is the probability that an index \((i, j) \in J_l \) is not in the corrupted set \((i, j) \notin \tilde{J}_l \). Let \( \beta \) hold for \( J_l \) with parameter \( \alpha \) and \( \beta \) hold \( \Pr(\tilde{J}_l \in \mathbb{I}(ka, \gamma)) \) be the probability that \( \beta \) holds for \( \tilde{J}_l \) with parameter \( k \alpha, \gamma \). Then

\[
(1 - \varepsilon_B^k)^{N_2 - k + 1} \leq \Pr(\tilde{J}_l \in \mathbb{I}(ka, \gamma)).
\]

**Proof:** Note that \( \beta \) gives a connected graph. We search for a lower bound of the probability, that the graph is still connected when we remove points \( m_j \) with probability \( \varepsilon_B \) but add edges \( m_j m_{j+q} \) with \( q \leq k \). It follows that the graph will no longer be connected whenever \( k \) consecutive points vanish.

This problem is an application of success runs in Bernoulli trails [41]. In particular, \( \Pr(\tilde{J}_l \in \mathbb{I}(ka, \gamma)) \) is the probability that the longest success run is shorter than \( k \). This probability has an exact but rather complicated analytic expression. The simple lower bound \( (1 - \varepsilon_B^k)^{N_2 - k + 1} \) is shown in [42]. Other bounds and the exact analytic form can also be found in [41].

The noise assumed in Theorem 5 appears in applications e.g., whenever a single measurement is lost or a sensor fails. The parameter \( \alpha \) can be adapted according to Theorem 5.

Theorem 2 gives two conditions for exact recovery. The condition \( \mu_1(L) < \lambda(1 - \mu_1(L - 1)) \) ensures recovery of the right support set. It was deduced from OMP and was shown to be strict [23]. The \((\alpha, \gamma)\)-separation condition guarantees the separation of the support into \( L \) structures. This condition is not strict. The \( L \) feasible sets \( J_1, \ldots, J_L \) may still be reconstructed without having \((\alpha, \gamma)\)-separation depending on the amplitudes \((X)_{J_l}, l = 1, \ldots, L \). Theorem 3 gives reconstruction results if one or both of these conditions were penalized.

In Theorem 4 and 5 we discussed a noised sparsity pattern and how the parameters \( \alpha, \gamma \) should be adapted. In the next section, we present a post processing step to reconstruct the original pattern given. Beforehand, we give a statement on two other cases of noisy data. First, consider the case where \( B \) does not have an exact sparse representation \( X \) with \( DX = B \), but instead we search for a sparse approximation as in problem 2. This problem occurs e.g., when the data \( B \) is noised. We can easily obtain similar results to Theorem 2 and 3 by replacing the exact recovery condition of (weak) OMP with the optimal \( L \)-term approximation conditions given in [23].

As another scenario, consider a sampling \( B \) that is sparse in some dictionary \( D \), i.e., there exists a sparse solution \( X \) of \( DX = \tilde{B} \). Now, instead of \( D \) we have only given the dictionary \( \tilde{D} \). Exemplary, let \( B \) be sparse in Fourier domain but
not necessarily containing frequencies given by the discrete Fourier transform. Given the Fourier transform of \( \mathcal{R} \) is it possible to reconstruct the exact frequencies, i.e., given an approximation in \( \mathcal{D} \) is it possible to reconstruct the exact dictionary \( \mathcal{D} \)? This problem was analyzed under the keyword of super-resolution in [43] for the SMV problem. Only recently, the MMV problem with common support constraint was discussed in [44]. In both cases, an exact recovery is possible whenever the non-zero entries are separated by at least a distance depending on the super-resolution factor. An interesting question we consider for future work, is the connection between this separation and patterns that are not \((\alpha, \gamma)\)-intersecting. This connection may be used to design a super-resolution method for generalized patterns.

D. Post Processing

So far, we presented the GM-OMP algorithm and proved basic theoretical properties. Before we demonstrate the technique on numerical examples, we discuss how to use GM-OMP for powerful post processing of the data. Consider we reconstructed a solution \( \tilde{x} \) and its support \( \text{supp} \tilde{x} = I = J_1 \cup \ldots \cup J_L \), where \( J_i \subseteq \mathbb{I} \) is assumed to be a corrupted sparsity pattern.

While it is a common idea to denoise corrupted amplitude values of \( \tilde{x} \), the sparsity pattern has been of minor interest so far. Even though the pattern itself might be noised. Exemplary, in non-destructive testing external forces during the measurement can corrupt the probes positions what influences the geometry and thus the sparsity pattern [9], [10]. As another example, consider \( \mathcal{D} \) being the Fourier matrix. It only contains a fixed amount of Fourier frequencies. However, there are signals that are sparse in Fourier domain but only consist of frequencies not covered by the matrix. Then the reconstructed sparse approximation most likely rounds these frequencies up to the closest frequency of \( \mathcal{D} \), what can be interpreted as a corrupted sparsity pattern of \( \tilde{x} \). As last example, simply assume a failed measurement, i.e., a zero column in \( \mathcal{R} \). Surely the corresponding column in \( \tilde{x} \) will also be zero. To reconstruct the original signal, we can apply inpainting ideas on the sparsity pattern.

Remembering Fig. [1] i.e., \( J_i = \text{supp} M_i \) as a discrete sampling of a function, we can denoise the sparsity pattern for \( l = 1, \ldots, L \) by solving the problems

\[
    f_l = \arg \min_{f \in \mathcal{F}} \left\| (p_j - f(m_i))_{(i,j) \in J_l} \right\|_2 + \delta \left\| \text{supp} f \right\|,
\]

with a weight \( \delta > 0 \). Here, \( \mathcal{F} = \mathcal{F}(\mathcal{M}, \mathcal{P}) \) is a suitable function space (e.g., polynomials, splines, ...). Afterwards set the denoised pattern \( J_l \) to

\[
    J_l = \{(i,j) \mid m_i \in \text{supp} f_l, \quad p_j = [f_l(m_i)]\}
\]

where \([f_l(m_i)]\) is \( f_l(m_i) \) rounded to the closest of the elements \( p_1, \ldots, p_{N_F} \). For small \( \delta \) the support \( \text{supp} f \) may be large and hence \( |J_l| \) can increase. This gives an inpainting strategy to reconstruct missing structure elements.

A similar approach can be applied to denoise the amplitudes of \( \tilde{X} \). Given \( J_l \) and assume that \( J_l \cap J_{l'} = \emptyset \) for all \( l' \neq l \) we solve

\[
    g_l = \arg \min_{g \in \mathcal{G}} \left\| (\tilde{X}_{i,j} - g(m_i))_{(i,j) \in J_l} \right\|_2
\]

on a function space \( \mathcal{G} = \mathcal{G}(\mathcal{M}, \mathbb{C}) \) and update \( \tilde{X}_{i,j} = g_l(m_i) \), for all \( (i,j) \in J_l \).

III. NUMERICS

We demonstrate the advantages of our proposed algorithm in three examples. First, we compare the technique with other sparse approximation methods for the MMV problem. Afterwards we discuss two practical examples and illustrate the information given by the sparsity pattern.

A. Numerical comparison

We compare our method to three other techniques: OMP applied to each column separately, S-OMP [33] and MSBL, a technique presented in [45] based on sparse bayesian learning.

Let \( \mathcal{D} \) be a convolution matrix of a Gauss kernel with standard deviation \( \sqrt{2.5} \). We define the matrix \( \tilde{X} \in \mathbb{R}^{1000 \times 1000} \) by

\[
    (\tilde{X})_{i,j} = \begin{cases} 
    1 & i = [j \tan \xi] \\
    0 & \text{otherwise}
    \end{cases}
\]

for \( i, j \leq 1000 \), i.e., each column of \( \tilde{X} \) is 1-sparse. The matrix is clearly structured, it consists of one line with a slope of \( \xi \). For \( \xi = 45^\circ \) this becomes the identity matrix; \( \xi = 0^\circ \) gives a matrix with one non-zero row (which is the pattern that S-OMP and the MSBL assume). For \( \xi = 0^\circ, \ldots, 45^\circ \) we calculate \( \mathcal{B} = \mathcal{D} \mathcal{X} \) and use all methods to reconstruct \( \tilde{X} \). Therefore, we choose \( m_i = i, \quad p_j = j, \quad \alpha = 1 \) and \( \gamma = 1 \) (which corresponds to a maximal slope of \( 45^\circ \)). We use \( L = 1 \) iteration since \( \tilde{X} \) contains exactly one structure. In Fig. 2 the reconstruction error and the number of non-zero elements in the solution \( \tilde{X} \) are shown for all algorithms. Nearly all methods are able to find a good approximation. Only S-OMP forces row sparsity of \( \tilde{X} \) and thus produces stare casing effects which corrupt the solution. Both the MSBL and S-OMP assume that \( \tilde{X} \) is row-sparse, i.e., there are only a few non-zero rows. Once a row contains a non-zero element, the entire row is considered to be non-zero. This leads to an extreme overestimation of the support while OMP and GM-OMP can find the exact number of non-zero entries.

Next, we demonstrate the power of the proposed denoising step. Therefore consider \( \mathcal{X} \) and its noised versions \( \mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{1000 \times 1000} \) with

\[
    (\mathcal{X}_1)_{i,j} = \begin{cases} 
    1 & i = 500 \\
    0 & \text{otherwise}
    \end{cases}, \quad (\mathcal{X}_1)_{i,j} = \begin{cases} 
    1 & i = 500 + [\varepsilon_u(j)] \\
    0 & \text{otherwise}
    \end{cases}
\]

where \([\varepsilon_u(j)] \leq \varepsilon_u \) is uniform noise and \( \varepsilon_B(j) \in \{0, 1\} \), \( \Pr(\varepsilon_B(j) = 0) = \varepsilon_B \) is Bernoulli distributed. Given \( \mathcal{B}_1 = \mathcal{D} \mathcal{X}_1 \) or \( \mathcal{B}_2 = \mathcal{D} \mathcal{X}_2 \) we want to reconstruct \( \tilde{X} \). The mean squared
error over 100 runs for both cases is plotted in Fig. 5. The noise level gives the values of \( \varepsilon_n \) respectively \( \varepsilon_b \).

For \( X_1 \) we adapted the parameters according to Theorem \( \mathbb{H} \) and set \( \gamma = 13 \geq 1 + 2\varepsilon_n/m \). In the second case, we choose \( \alpha = 6 \) for the reconstruction of \( X_2 \). Using Theorem \( \mathbb{M} \) this gives us a reconstruction probability of more than 78% even for \( \varepsilon_b = 0.25 \). In both cases we solve optimization problem \( \mathbb{R} \) with \( \delta = 0 \) and \( \mathcal{F} = \Pi_4 \) afterwards to denoise the pattern.

Note that the row-sparsity assumption of S-OMP and MSBL hold for \( X \). Nevertheless, only GM-OMP is able to find a good approximation in most cases. All other methods only reconstruct the noised matrices \( X_1, X_2 \). For a high probability of \( \sigma = 0 \) the MSE of GM-OMP increases, i.e., the parameter \( \alpha \) does no longer compensate the missing data. Interestingly, S-OMP profits from its stare casing effects when the pattern is uniformly noised and hence returns a slightly better solution.

**B. Application 1: Non-destructive testing**

As a first practical example we analyze ultrasonic data obtained from non-destructive testing of steel tubes. The original data shown in Fig. 4 was generated by the "Time-of-Flight Diffraction" (ToFD) method using an Olympus Omniscan iX system with 5MHz transducer, 6mm diameter and 70° angle of incidence. The tested tube was a large diameter pipe with outer diameter 1066mm and 23.3mm wall thickness. Each column of the data represents a measured signal at different positions on the tubes surface. The positions were equidistantly set on a straight line with a distance of 0.5mm, hence we set \( m_i = 0.5i \). The signals were measured in time with a sampling ratio of 0.01\( \mu \)s. Four major events can clearly been seen in the data. The topmost one is an ultrasonic impulse that directly travels through the surface from transducer to receiver - the lateral wave. The bottommost is an impulse that was reflected by the back wall - the back wall echo. The two events in between (recognizable as parabolas) indicate defects in the material. We use GM-OMP to recover and denoise these events.

Ultrasonic data is column-wise sparse when \( D \) is a convolution matrix based on the Gabor impulse \( (9), (11) \)

\[
g(t) = e^{-\theta t^2} \cos(\psi t + \theta)\]  

Here \( \theta \) is the bandwidth factor, \( \phi \) is the frequency and \( \psi \) is the phase. Thus, \( p_j = 0.01j \) is the shift of the \( j \)-th column in \( \mu \). \s{s}n. For the given data we choose \( p = 6.8486, \phi = 14.685 \) and \( \psi = -2.0836 \) (see \( \mathbb{N} \) for details about the parameter choice).

We define our feasible set \( \mathcal{J} \) using \( \alpha = 0.5 \) such that \( m_i - m_j \leq \alpha \) only for \( i = j - 1, i + 1 \). Note that \( \mathbb{J} \) compares distance in time (\( \mu \)) with a distance in space (\( \text{mm} \)). The ultrasonic speed in steel is about 5.9mm/\( \mu \), hence we chose \( \gamma = 0.1 \geq 5.9^{-1} \) what gives stability for noisy data.

After applying \( L = 4 \) iterations of GM-OMP to the data, we use the denoising strategies discussed in the last section and set \( \beta = 1 \). Since structures in pipe testing often behave linearly or quadratically, we use \( \mathcal{F} = \{ f \mid \chi_C / f \in \Pi_4, C \subseteq \mathcal{M} \text{ convex} \} \), i.e., polynomials up to degree four multiplied by a characteristic function. The characteristic function is needed for the support constraint in \( \mathbb{R} \). Moreover we choose \( \mathcal{G} = \Pi_0 \) as the space of all constant functions, i.e., the amplitudes of each structure are set to its mean value. This value can give a first hint about the underlying material in the pipe.

In Fig. 5 the four sparsity patterns found by GM-OMP are shown in data domain (i.e., multiplied by \( D \)). As we see, the algorithm is able to reconstruct all four structures of the original data. Due to the structural denoising, the pattern looks more smooth and effects caused by shaking probes are no longer visible. In Tab. \( \mathbb{T} \) the polynomial coefficients of each pattern are shown. As supposed, the lateral wave and back wall echo are mostly linear while the defects were approximated by a quadratic polynomial.

**C. Application 2: meteorologic data**

In this example we use GM-OMP without post-processing on meteorologic data provided by Deutscher Wetterdienst (DWD) \( \mathbb{D} \). We analyze the hourly precipitation in Germany from 25th to 28th of November 2008 where we use data of 932
Fig. 5. Reconstructed and denoised structures.

TABLE I

| L   | 1     | 2       | 3       | 4       |
|-----|-------|---------|---------|---------|
| $x^1$ | $1.05e-10$ | $1.57e-11$ | $1.48e-6$ | $5.65e-7$ |
| $x^2$ | $8.91e-08$ | $1.45e-08$ | $-3.02e-5$ | $4.22e-5$ |
| $x^3$ | $2.42e-05$ | $4.80e-06$ | $-0.000290$ | $0.001354$ |
| $x$  | $0.002202$ | $0.000576$ | $0.007798$ | $0.010080$ |
| const. | $2.8598$ | $0.770270$ | $1.6013$ | $1.6006$ |

stations shown in Fig. 5. Stations that were moved during this time or had too many missing values were neglected. In Fig. 7 the overall precipitation and data of two stations is plotted exemplary where 0 hour refers to Nov. 25th 2008, 0:00. We use a dictionary based on centered cardinal B-splines

$$B_i(t) = \chi_{[-0.5,0.5]}(t) \quad B_n(t) = \int_{-0.5}^{0.5} B_{n-1}(t-\tau)d\tau.$$  

We use the normalized versions of all B-splines with $n \leq 7$, i.e., $D$ contains all 96 shifts of $B_n$, $n \leq 7$. We have chosen a time period with low precipitation and thus the data can be sparsely approximated using splines. Note that a B-spline of order $n$ has a support of length $n$. Thus, the order directly correlates to the duration of the precipitation. Since a single precipitation event (e.g., rain) will be registered at several stations, we have an underlying structure in the data.

Choose $m_i$ to be the position coordinates of the $i$-th station and $p_j \in \mathbb{N}^2$ as the shift and order of the corresponding spline. We use $D_{\phi}$ as the geodetic distance and set $\alpha = 30$km. Let $d_{\phi} = ||\cdot||_{\infty}$ and $\gamma = 1/15$, i.e., neither the duration nor the time of occurrence should change more than 2 hours per 30km. We perform $L = 100$ iterations of GM-OMP.

Fig. 6 shows the time of occurrence of the largest precipitation event, i.e., the structure that includes the most stations (here 156). The mean duration is 1.28h (mean B-spline order) and one can clearly recognize the event moving from north to south caused e.g., by wind. In Fig. 9 we reconstructed the overall precipitation (see Fig. 7) using only 15 structures. In the left figure we choose the first 15 structure, i.e., those with the strongest precipitation; for the right figure the 15 largest events were used. While the strongest events contain the precipitation peaks, the largest events can better reconstruct the overall structure from Fig. 9.

Fig. 7. Left: Overall precipitation (log scale) [47]; right: data of two stations exemplary.

Fig. 8. Time of occurrence of the largest precipitation [47].

IV. Conclusion

We presented a generalized orthogonal matching pursuit for multiple measurements. The algorithm is able to recognize and reconstruct more general sparsity pattern in the solution as other algorithms for multiple measurements. Moreover, GM-OMP allows efficient post processing for each pattern. These patterns can provide crucial information in applications which was exemplary demonstrated in two practical examples.
Two parameters allow an adaption of the feasible patterns to the application and make the algorithm more flexible. The advantages of GM-OMP were shown in comparison to other techniques and confirmed by first theoretical results.

ACKNOWLEDGEMENTS

The author thanks Mannesmann Salzgitter GmbH for providing the ultrasonic data used in this paper. This work is supported by BMBF joined research project ZeMat (grant number: 05M13MGA).

REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering", IEEE Trans. Image Processing, vol. 16(8), pp. 2080-2095, 2007.

[2] M. Elad, M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", IEEE Trans. Image Processing, vol. 15(12), pp. 3736-3745, 2006.

[3] E. Le Pennec, S. Mallat, "Sparse geometric image representations with bandelets", IEEE Trans. Image Processing, vol. 14(4), pp. 423-438, 2005.

[4] J. Mairal, M. Elad, G. Sapiro, "Sparse representation for color image restoration", IEEE Trans. Image Processing, vol. 17(1), pp. 53-69, 2008.

[5] S. M. Katz, "Estimation of probabilities from sparse data for the language model component of a speech recognizer", IEEE Trans. Acoustics, Speech and Signal Processing, vol. 35(3), pp. 400-401, 1987.

[6] J. Wright, A. Y. Yang, A. Ganesh, Y. Ma, "Robust face recognition via sparse representation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31(2), pp. 210-227, 2009.

[7] M. Lustig, D. Donoho, J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging", Magnetic resonance in medicine, vol. 58(6), pp. 1182-1197, 2007.

[8] E. Y. Sidky, X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization", Physics in medicine and biology, vol. 53(17), pp. 4777-4807, 2008.

[9] F. Boëtmann, G. Plonka, T. Peter, O. Nemitz, T. Schmitte, M. Lustig, D. Donoho, "Sparse image and signal processing: wavelets, curvelets, morphological diversity", Cambridge university press, 2010.

[10] E. J. Candès, J. K. Romberg, T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", Comm. pure and applied mathematics, vol. 59(8), pp. 1207-1223, 2006.

[11] S. Mallat, A wavelet tour of signal processing: the sparse way, Academic press, 2008.

[12] E. J. Candès, D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges", Curves and Surfaces, Vanderbilt Univ. Press, 2000.

[13] X. Zhang, C. D. Frost, "Sparse multidimensional representation using shearlets", Proc. of the SPIE Conf. on Wavelets XI, San Diego, CA, 2005.

[14] S. Demirli, I. Sanie, "Model-based estimation of ultrasonic echoes. Part I: Analysis and algorithms", IEEE Trans. Ultrason. Ferroelect. Freq. Control, vol. 48(3), pp. 787-802, 2001.

[15] D. Labate, W. Lim, G. Kutyniok, G. Weiss, "Sparse multidimensional representation using shearlets", Proc. of SPIE Conf. on Wavelets XI, San Diego, CA, 2005.

[16] J. Mairal, M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", IEEE Trans. Image Processing, vol. 15(12), pp. 3736-3745, 2006.
[44] Z. Yang, L. Xie, "Exact Joint Sparse Frequency Recovery via Optimization Methods", IEEE Trans. Signal Processing, vol. 64(19), pp. 5145-5157, 2016.

[45] D.P. Wipf, B.D. Rao, "An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem", IEEE Trans. Signal Processing, vol. 55(7), pp. 3704-3716, 2007.

[46] Data provided by the Climate Data Center (CDC) of the Deutscher Wetterdienst (DWD) (www.dwd.de/DE/leistungen/cdcftp/cdcftp.html), downloaded Feb. 2016.

[47] Using m_map by R. Pawlowicz (www.eoas.ubc.ca/~rich/map.html) and coastline / boarder data by Natural Earth (www.naturalearthdata.com), downloaded Feb. 2016.