MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1619 September, 1997
C.B.C.L. Paper No. 156

Sparse Representation of Multiple Signals

Theodoros Evgeniou and Tomaso Poggio

This publication can be retrieved by anonymous ftp to publications.ai.mit.edu. The pathname for this publication is: ai-publications/1500-1999/AIM-xxx.ps.Z

Abstract

We discuss the problem of finding sparse representations of a class of signals. We formalize the problem and prove it is NP-complete both in the case of a single signal and that of multiple ones. Next we develop a simple approximation method to the problem and we show experimental results using artificially generated signals. Furthermore,we use our approximation method to find sparse representations of classes of real signals, specifically of images of pedestrians. We discuss the relation between our formulation of the sparsity problem and the problem of finding representations of objects that are compact and appropriate for detection and classification.

This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological and Computational Learning in the Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology. This research is sponsored by grants from MURI N00014-75-0600.

1 Introduction and Formulation of the Problem

In this note we discuss the problem of finding representations for reconstruction of a number of signals using "features" chosen from a large pool of possible ones. Specifically, we define the problem of finding sparse representations of a class of signals in terms of a small set of basis signals chosen from an overcomplete set of many basis vectors.

Finding sparse representations of signals has recently been an important topic of research in the vision community (ie see [4] [1],[3], [7] and references therein). In [1] the problem of finding a sparse representation of a single signal is defined and an approximation method is suggested. In [4] a sparsity criterion determines basis vectors to represent images of natural scenes that are similar to the receptive fields of neurons in primary visual cortex. In this case the basis functions "evolved" instead of being chosen from a predefined set of possible vectors.

In this paper we follow a different approach which, in a sense, is a combination of the work in [1] and [4]. Specifically, instead of trying to "evolve" (as in [4]) the basis functions used to represent signals (ie images), we try to find how an existing basis (neurons) can be used in order to sparsely represent input signals (images). Summarizing, the contributions of the paper are:

1. We formulate the problem of finding sparse representations of a family of signals.
2. We prove that both the sparsity problem in [1] as well as the one formulated here are NP-Complete.
3. We suggest approximation methods for the formulated problem.
4. We show preliminary experimental results using a simple approximation method.
5. We show how to use our formulation to find representations of classes of objects (such as images of pedestrians) that can be used for detection and classification.

2 Formulation of the Sparsity Problem

In the case of one signal the problem is as formulated in [1]: *given an N-dimensional signal S and a set of $M >> N$ vectors B_i ($i \in \{1, ...M\}$) that constitute an overcomplete basis for the N-dimensional space that the signal belongs to, choose the fewest possible basis vectors that reconstruct (or "best" approximate) the given signal S.* A number of approximation methods to this problem are presented in [1]. This paper discusses the extension of the single-signal case to the many-signals one. The problem now is: *Given a set of K N-dimensional signals S_j, ($j \in \{1, ...K\}$) and a set of $M >> N$ vectors B_i ($i \in \{1, ...M\}$) that* *constitute an overcomplete basis of the vector space that the signals lie in, find the smallest number of basis vectors that reconstruct (or best approximate)* **all** *the given signals.* The mathematical formulation of this problems is as follows.

$$\begin{gathered}\min_{\xi_i,\alpha_{ij}} \sum_{i=1}^{M} \xi_i \\ \text{subject to: } \sum_{i=1}^{M} \xi_i \alpha_{ij} B_i = S_j \text{ for every } j \in \{1, ...K\} \\ \text{where } \xi_i \in \{0,1\} \text{ and } \alpha_{ij} \in R\end{gathered}$$

Here the ξ_i are 0 when basis vector B_i is not used by any signal and 1 when it is used by at least one signal. Minimizing the sum of ξ_i means minimizing the number of basis vectors used by all signals.

Notice that this is an integer programming formulation with non-linear constraints, which is an indication that the sparsity problem is NP-complete. We present a formal proof of this below.

2.1 Sparsity Problem is NP-Complete

It is known (see [2]) that the following problem is NP-complete:

Minimum Weight Solution for Linear Equations (MWSLE): *Given an $n \times m$ matrix A with integer entries, an $m \times 1$ vector b with integer entries and an integer $K <= m$, find whether there exists x with rational entries such that x has at most K non-zero entries and $Ax = b$.*

It is easy to see that the **Sparsity** problem is a "general" case of the **MWSLE** problem. So: if we assume that the **Sparsity** problem can be solved in polynomial time, then for a given instance (A, b, K) of **MSWLE** we could solve the **Sparsity** problem with basis A and signal b, and find a solution x_{sp} with the fewest non-zero entries (say L is the minimum number of non-zero entries, $L < m$). If x_{sp} has rational entries only, we are done with the **MWSLE** problem, since then: if $K > L$ then the answer to the problem is "yes", otherwise it is "no". So all we have to show now is that if we can find a solution with L non-zero entries for the *Sparsity* problem, then we can have a rational solution with at most L non-zero entries to the **MWSLE** problem.

In the case that the x_{sp} that we found solving the **Sparsity** problem is not rational, if we can show that there exists a rational x_{rat} with the same non-zero entries as x_{sp}, we are done. For this, consider the following problem: Construct $n \times L$ matrix A' which is matrix A with the columns corresponding to zero-entries of x_{sp} removed. Also, take x_{new} to be x_{sp} with all zero entries removed. Then we have a solution x_{new} to the system of equations:

$A'x = b$ with some of the entries of x_{new} being irrational numbers. If we can show that there exists also a solution x_{rnew} with rational entries such that $A'x_{rnew} = b$, we are done. For this we have the following lemma:

Lemma: If there is a solution x for the set of linear equations $Ax = b$ where A and b have integer entries and A is $n \times L$, then there is a solution x_{rat} to the same set of equations with all entries of x_{rat} being rational.

Proof: Since there is a solution to the set of equations, there is a solution that is given directly using determinants of matrices (if r is the rank of A, then there is an $r \times r$ square submatrix A' of A with rank r for which we get $x = A'^{-1}b$ which is a solution to our original system and clearly has rational entries since A' and b have integer entries. If $r = 0$, then clearly any x is a solution - since there exists at least one solution).

Therefore the *Sparsity* problem is also NP-Complete. Furthermore, the many signals problem can be shown to be NP-Complete trivially: the *Sparsity* problem can be trivially "reduced" to the many signals problem (since the first is a special case of the second).

3 Approximation Methods

In this section we first discuss ideas for how to approximate the many signals sparsity problem, and then we describe a simple approximation method that we also tested with artificial and real signals. A different approximation method is discussed in [3].

3.1 Iterative Approximation Methods

The layout of this family of approximation methods is as follows:
Given a set S_1 of K N-dimensional signals and a set B_1 of $M >> N$ basis vectors, the set G of "selected" basis vectors is initialized to "empty" and:

For $i = 1$ to N:

1. If all vectors in S_i are zero, return current G
2. For each of the non-zero signals S_{ik} in S_i find its sparsest representation in the sense of [1] using basis B_i. This is the solution of the LP problem:

$$\min_{\alpha_{kj}} \sum_{j=1}^{M} |\alpha_{kj}|_{L_1}$$
$$\text{subject to: } \sum_{j=1}^{M} \alpha_{kj} B_{ij} = S_{ik}$$

where $\alpha_{kj} \in R$ and $B_{ij} \in B_i$

3. For each basis vector in B_i: compute the sum of the absolute values of the coefficients (the α_{ij} found in 2) corresponding to this basis vector that the signals in S_i "use" - found in step (2).

4. Select the basis vector B_{ij} with the largest sum (as found in step (3)). For each non-zero signal S in S_i find its projection on the selected basis vector and subtract it from the signal. Delete this basis vector from the set of basis vectors and add it to set G. So now:
   $S = S - S \cdot B_{ij}$ for every non-zero signal $S \in S_i$ and
   $B_{i+1} = B_i - \{B_{ij}\}$
   $G_{i+1} = G_i \bigcup \{B_{ij}\}$
   Go back to step (1).

One can get several variations of this basic layout. For example one can change the criterion for selecting B_{ij} at each iteration. A possible criterion other than the one above is: "select the basis vector that is "used" (ie gives coefficients larger than a predefined threshold) by the largest number of signals". Other variations (ie changing step 2) can be developed.

3.2 Mathematical Programming Approximation Method

3.2.1 Two "naive" Approaches

One "naive" approach is to solve the many signals problem as formulated in section 2 after relaxing the constraints that $\xi_i \in \{0, 1\}$ - let ξ_i take any value between 0 and 1. However, although this relaxation would lead to a linear cost function, the constraints would still be non-linear. Solving the relaxed problem is still hard.

A simple approximation for the many signals sparsity problem is to solve the single signal problem for each of the input signals using the approximation method of [1] and define the final solution to be the union of all the basis vectors found. This could be achieved by solving the following linear programming problem:

$$\min_{\alpha_{ij}} \sum_{i=1,j=1}^{M,K} |\alpha_{ij}|_{L_1}$$
$$\text{subject to: } \sum_{i=1}^{M} \alpha_{ij} B_i = S_j \text{ for every } j \in \{1, ...K\}$$
$$\text{where } \alpha_{ij} \in R$$

The final solution consists of all basis vectors B_i for which at least one of the $|\alpha_{ij}|_{L_1}$ is non-zero (or greater than a threshold), $j \in \{1, ...K\}$ (notice that this linear programming problem can be decomposed to K smaller ones without changing the final solution). However such an approach is likely to give many basis vectors as a final solution since it does not try to find a set of basis vectors that are consistently used by all signals. Moreover, given that each of the individual signals is expected to have some "signal-specific" characteristics (ie each pedestrian has its own specific characteristics), such an approach will not give us only the "features" - basis vectors - that are consistently important for all signals in the class. It will also give basis vectors that are signal specific but not class specific.

3.2.2 A Simple Approximation Method

Alternatively we should search for an approximation method that tries to find a set of basis vectors **consistently** used by all signals. Furthermore the method should avoid finding characteristics that are specific to only some of the signals. Given these two goals we suggest the following method.

Given a set of K N-dimensional signals and a set B of $M >> N$ basis vectors:

1. Compute a small number, say 2, of different linear combinations of the K signals, say combinations C_1 and C_2.

2. Solve the following linear programming problem (which is a simple extension of the formulation in [1] from the one signal case to the two signals one):

$$\begin{gathered}\text{Min } (x_1)_{L_1} + (x_2)_{L_1} \\ \text{Subject to} \\ Bx_1 = C_1 \\ Bx_2 = C_2\end{gathered}$$

3. The final representation uses all basis vectors B_i for which $|x_{1i}|_{L_1}$ or $|x_{2i}|_{L_1}$ is non-zero or larger than a threshold. The number of basis vectors used can be restricted by altering this threshold.

For the second step we can alternatively use an approximation to the problem that also takes into account noise - it assumes the signals are noisy. In this case the problem is formulated as (again using the formulation in [1]):

$$\begin{gathered}\text{Min } (x_1)_{L_1} + (x_2)_{L_1} + \lambda((\epsilon_1)_{L_2} + (\epsilon_2)_{L_2}) \\ \text{Subject to} \\ Bx_1 + \epsilon_1 = C_1 \\ Bx_2 + \epsilon_2 = C_2\end{gathered}$$

In our experiments we use the noiseless formulation.

Before describing our experiments we explain the motivation behind this formulation. First, as mentioned above, we want to find a consistent set of basis vectors used by all signals and at the same time avoid picking vectors due to noise or due to "characteristics" specific to a signal. The formulation above is expected to satisfy both these requirements. By taking a linear combination of the signals we expect to eliminate noise and also "smooth out" the signal specific characteristics while enhancing the class specific ones. On the other hand, by taking only 2 (or maybe 3) linear combinations we make the problem tractable (we could potentially solve the problem using all the signals in our cost function, but such a formulation would quickly become intractable - as soon as the number of signals becomes significantly large). Moreover, solving the problem using all the signals instead of the linear combinations would not find a consistent solution among the signals. In a sense by taking the linear combination of the signals we "glue" them together so that only the basis vectors used by *all* of them is found. Finally, the reason we take 2 (or 3) linear combinations instead of just one is that taking only one linear combination may force some of the important "features" (basis vectors) to disappear (their coefficients to become zero). On the other hand we expect that a very small number of linear combinations (ie 2 or 3) is enough to avoid such a problem.

4 Experimental Results

4.1 Synthetic Signals

We show the results of two experiments in tables 1 and 2. The signals used were 36 dimensional and were generated using some of the basis functions of an overcomplete dictionary with Gaussian noise added afterwards. Recovering the basis functions used was not always successful for each of the individual signals (especially when considerable noise was added to the signals), but it was possible most of the time for linear combinations of the signals.

In figure 1 a 4-fold cosine and sine overcomplete basis was used (146 basis vectors in total). 50 36-dimensional signals were generated using basis vectors 17 and 110. We added Gaussian noise to each of the signals, and then we solved the "sparsity" problem for a each of the individual signals (using the formulation of [1]). We also solved the sparsity problem using the approximation method described in the previous section. The first three lines of the table show the basis vectors chosen when the sparsity problem was solved for signals 2,3 and 4 respectively. Notice that for all 3 signals we fail to find the exact basis vectors used to construct them - due to noise. When we solved the problem using our simple approximation method we got the correct results shown in the last line of the table.

In figure 2 we used an overcomplete Haar wavelets basis (306 basis vectors plus one vector

Signal	Basis Vectors Used
2	1, 17, 110
3	17, 107
4	17, 108, 126
Weighted Averages solution	17, 110

Figure 1: *Using a 4-fold cosine and sine overcomplete basis.*

Signal	Basis Vectors Used
1	1, 35, 38, 40, 90, 93, 105, 140, 150, 151, 209, 220, 237, 277, 300
2	1, 39, 40, 45, 90, 123, 150, 175, 200, 209, 220, 285, 300
4	1, 10, 38, 40, 85, 132, 149, 150, 165, 200, 220, 276, 300
Weighted Averages solution	1, 10, 40, 90, 150, 220, 300

Figure 2: *Using an overcomplete Haar wavelets basis (306 basis plus one vector of ones).*

of ones - to capture the mean value of the signals). The signals were constructed using basis vectors 10, 40, 90, 150, 220 and 300. Noise was added as before. Again we show the solutions found for some individual signals as well well as the one found using our approximation method.

4.2 Application to the Representation of Pedestrians

An interesting application of the aforementioned ideas is finding representations of classes of objects. This idea is motivated from biology. It is well-known (ie see [8]) that the primary visual cortex has a set (overcomplete basis) of neurons with specific receptive fields (basis vectors). These "basis vectors" are used for the representation of all images. We expect that different classes of objects excite different neurons (basis vectors). Therefore, if we start with an overcomplete basis - similar to the receptive fields found in V1 - and examine which of the basis vectors are used by objects of the same class under the assumption that the representation should always be sparse, then the prediction is that a few basis vectors are commonly used by all objects of the same class. Each object individually may also use other basis vectors (due to noise or object-specific characteristics) but we expect to find a "small" set of basis functions used by all. Work in this direction can also be found in [5].
Having this in mind we conducted the following experiment. Given a number of aligned

Figure 3: Two typical images of pedestrians and an "average" pedestrian.

images of pedestrians (data used in [5]) and an overcomplete wavelet basis, we solved the "many signals" sparsity problem using the images as signals and the wavelet basis as our overcomplete basis. Following the simple approximation method described in the previous section, we first generated weighted averages of the images of pedestrians (the images were assumed to be aligned, so no correspondece was computed between them before averaging), and then we solved the problem for these averages. Figure 3 shows two typical images of pedestrians as well as a weighted average of 1000 such images. When the average is taken only the "significant" characteristics of the signals remain (ie the shape of a typical pedestrian). After solving (in the sense of [1]) the sparsity problem for this signal, we reconstructed the signal using only some of the found wavelet vectors (thresholding the computed coefficients). Figure 4 shows the reconstructed image using different number of basis vectors (different thresholds). Notice that only a few basis vectors are enough to yield sufficiently "good representation" of pedestrians (similar to the one found in [5]). Further tests need to be done to evaluate the quality of the found representation.

5 Conclusion and Future Directions

We proved that the problem of finding sparse representations starting from an overcomplete basis is NP-Complete. Given this, finding approximation methods is the only feasible approach. In this paper we suggest a simple approximation method to the "sparsity" problem in the many signals case. Preliminary experimental results using artificially generated signals were promising. Furthermore we applied our formulation of the sparsity problem and our approximation method to the problem of finding representations of classes of objects such as images of pedestrians. The results are promising but further tests need to be done

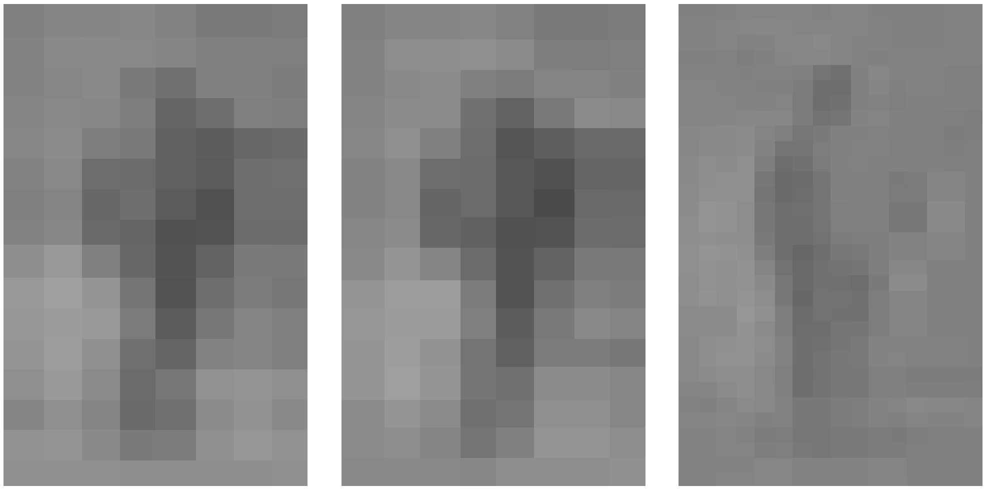

Figure 4: Reconstructed "average pedestrian" using 41, 25 and 19 of the basis vectors. In the first two images we used a basis of Haar wavelets with resolutions 4 and 8. For the third image we used resolutions 2, 4 and 8. The images were 32x64.

to better evaluate the performance of our approximation methods.

References

[1] S. Chen and and D. Donoho and M. Saunders Atomic Decomposition by Basis Pursuit Technical Report 479 , Department of Statistics, Stanford University, May 1995

[2] Michael R. Garey and David S. Johnson, 1979 Computers and Intractability: A Guide to the Theory of NP-Completeness

[3] F. Girosi, An Equivalence Between Sparse Approximation and Support Vector Machines, MIT Artificial Intelligence Laboratory Memo 1606, 1997

[4] B.A. Olshausen and D.J. Field Emergence of simple-cell receptive field properties by learning a sparse code for natural images, In *Nature, 381, 607-609, 1996*

[5] M. Oren and C. Papageorgiou and P. Sinha and E. Osuna and T. Poggio Pedestrian Detection Using Wavelet Templates In *Proc. Computer Vision and Pattern Recognition, Puerto Rico, June 16–20,1997*

[6] E. Osuna and R. Freund and F. Girosi Support Vector Machines: Training and Applications MIT A. I. Lab. Memo 1602, 1997,

[7] T. Poggio and F. Girosi Notes on PCA, Regularization, Support Vector Machines and Sparsity, MIT Artificial Intelligence Laboratory Memo (in press), 1997

[8] Young, M.P. and Yamane, S. Sparse Population coding of faces in the inferotemporal cortex In *Science, 256, 1327-1331, 1992*