# Applying Sparse $\ell_1$-optimization to problems in robotics

Yasir Latif[1], Guoquan Huang[2], John J. Leonard[2] and José Neira[1]

***Abstract*— Sparse $\ell_1$-optimization techniques have received a lot of attention in the signal processing and computer vision communities, where they have been applied to problems such as denoising [1], deblurring [2], and face recognition [3]. Using $\ell_1$-objective to solve an optimization problem has been shown to induce sparsity. Moreover, the problem is convex allowing a global minimum solution. Well studied techniques and solvers exist that allow efficient solutions for the optimization problem by posing it as either a Linear Problem (LP) or taking advantage of the sparse nature of the problem, i.e., homotopy based methods. In this work, we provide an overview of this sparse $\ell_1$-formulation and apply it to various problems in robotics including loop closure detection, place categorization and topological SLAM.**

## I. INTRODUCTION

In many robotics applications, we are interested in solving a one-to-many data-association problem, that is, given the current observation, we are interested in finding one or a few among all the previous observations that are in some sense similar to the present observation. For example, given the current image, is there a matching image that the robot has seen before? Or, given an image, can we find which class (kitchen, hallway, etc) does the image come from? In these problems, we expect the solution to be sparse, that is, only a few (if any) of the previous images match the current image and only a small number of different instances of the observed classes explain the current observation. With this intuition in mind, we want to represent these problems in a more general formulation which can give us "sparse" solutions, i.e. solutions in which only a small number of previous observations interact to explain the current observation.

We begin by formulating this as: $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ where $\mathbf{A} \in \mathcal{R}^{n\times m}$ is a function of all the previous observations, $\boldsymbol{b} \in \mathcal{R}^n$ is the current observation and we want to find $\boldsymbol{x} \in \mathcal{R}^m$, which would indicate how the previous observations in $\mathbf{A}$ interact to generate $\boldsymbol{b}$. We call $\mathbf{A}$ the *dictionary*. Since there can be many more observations compared to the dimension of each observation ($m > n$), the problem is *under*-determined and infinitely many solutions exist (if $\mathbf{A}$ has full row-rank, that is, rank($\mathbf{A}$)=n)). We need to "regularize" i.e select a desirable solution based on our prior knowledge about it.

A first and commonly used approach is to look for the *least squares* solution, i.e. a solution that minimizes

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \quad \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \quad \Rightarrow \quad \boldsymbol{x}^\star = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\boldsymbol{b} \tag{1}$$

This allows a closed-form and unique solution but the solution is not-sparse in general. All the elements of $\boldsymbol{x}^*$ are non-zero indicating that all the columns of $\mathbf{A}$ are utilized in explaining $\boldsymbol{b}$. Instead, we want a solution that contains very few non-zero elements.

By quantifying the sparsity of a vector using the $\ell_0$-norm, which is defined as the number of non-zero elements in a vector $|\{x_i | x_i \neq 0\}|$, we look for a sparse solution under the constraint that this solution should explain our observation i.e.

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{2}$$

Solving for the $\ell_0$-norm is in general NP-hard [4] as all the possible solution have to enumerated and the constrained checked for fulfillment. Instead, we can relax the problem by solving for $\ell_1$-norm. For a vector $\boldsymbol{x}$, the $\ell_1$-norm is defined as the sum of absolute values of all the elements in the vector i.e $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$. (2) can now be relaxed to,

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \quad \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{3}$$

Note that the above formulation (3) assumes perfect reconstruction and does not cater for any noise. In order to deal with real-world problems, we also introduce a sparse noise term to explain our observations with both the original dictionary and the added noise, i.e.,

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \quad \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \mathbf{D}\boldsymbol{\alpha} = \boldsymbol{b} \tag{4}$$

where $\mathbf{D} = \begin{bmatrix} \mathbf{A} & \mathbf{I} \end{bmatrix}$ and $\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e} \end{bmatrix}$.Note that this still has the general form as (3).

The formulation in (4) can be posed as a linear program (LP), a class of convex optimization problems and solved using interior point methods [5], whose complexity is $O(m^3)$. This makes such methods computationally infeasible for long-term operation. Alternatively, homotopy methods [6], [7] are specifically designed to take advantage of the properties of $\ell_1$-minimization. Relaxing the equality constraint in (4), we have the following *constrained* problem:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \quad \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{D}\boldsymbol{\alpha} - \boldsymbol{b}\|_2 \leq \epsilon \tag{5}$$

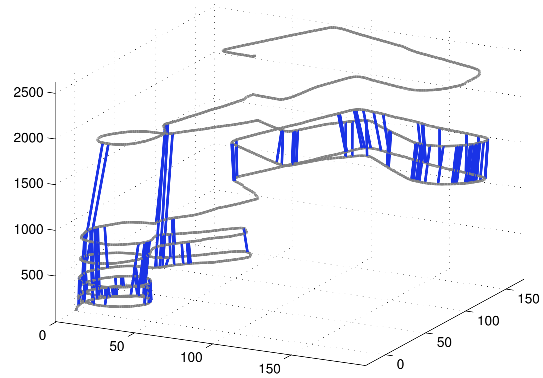
where $\epsilon > 0$ is a pre-determined noise level. This is termed the Basis Pursuit Denoising (BPDN) problem in compressive sensing. A variant of (5) is the *unconstrained* minimization:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \quad \lambda \|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\|\mathbf{D}\boldsymbol{\alpha} - \boldsymbol{b}\|_2^2 \tag{6}$$

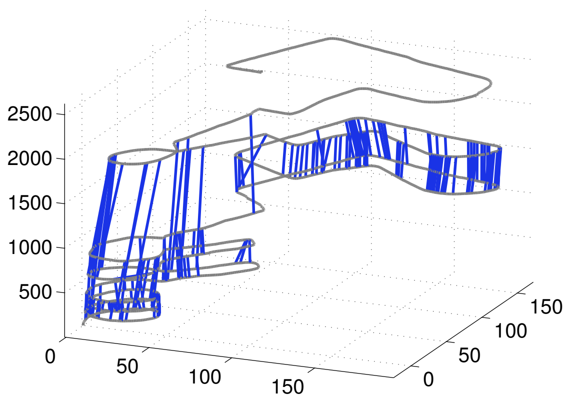
This work has been partially supported by the MINECO-FEDER project DPI2012-36070, and research grant BES-2010-033116.

[1]Y. Latif and J. Neira are with the *Departamento de Informática e Ingeniería de Sistemas, Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, Spain*, {ylatif, jneira}@unizar.es

[2]G. Huang and J. Leonard are with the *Marine Robotics Group, Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA* {ghuang, jleonard}@mit.edu




(a) Representation: Vectorized raw image down-sampled (size $8 \times 6$)




(b) Representation: GIST descriptor calculated over full-sized image

Fig. 1: Loop closures detected with the proposed approach using two different bases for the New College data set. In these plots, visual odometry (as provided with the dataset) is shown in gray and the loop closures in blue. The vertical axis represents time (in seconds) and the $x$ and $y$ axes represent horizontal position (in meters).

where $\lambda$ is a scalar weight parameter. The complexity of the homotopy method is $O(dn^2+dnm)$ for recovering a $d$-sparse signal in $d$ steps, but knowing that the solution is sparse, makes these methods more efficient compared to primal-dual methods.

Solving (6) gives us information about which basis (columns) of the dictionary are involved in explaining our current observation and well as their contribution towards making it. The noise component ($\boldsymbol{e}$) can be thought of as "innovation", it is information that can not be explained by any of the observations.

## II. Applications

In this sections, we show how the presented formulation can be applied to various problems in robotics.

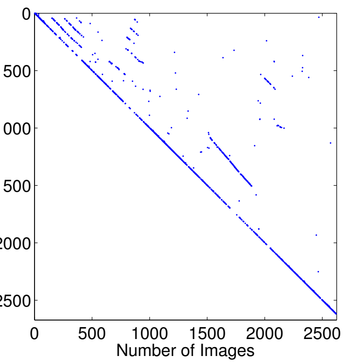



Fig. 2: Sparsity pattern induced by solving (4) for all the images. Each column $i$ of this matrix corresponds to the solution for the $i$-th image, and the non-zeros are the values in each column that are greater than $\tau = 0.99$. Note that the main diagonal occurs due to the current image being best explained by its neighbouring image, while the off-diagonal non-zero elements indicate the loop closures.

### *A. Loop Closure Detection*

Appearance based loop closing is the problem of finding images in the past that are similar to the image being observed, where similarity can be quantified as similarity in the image space or some descriptor space where each image is represented using a descriptor such as HOG [8] or GIST [9].

Loop closings occur sparsely and therefore it is natural to look for a sparse contribution from all the previous images. More detail can be found in [10] and a short summary is presented here. Until now, we have not placed any restriction on what the columns of $\mathbf{D}$ in (6) represent. For the task of appearance based loop closing, each column in the dictionary represents one of the previously observed images. In the simplest case, this can be the scaled-down image, reshaped into a vector in $R^n$ where $n = r \times c$, a product of the scaled rows and columns. An alternate approach would be to extract a descriptor such a HOG representing each image. The important aspect of this approach is that there is no prior learning involved, as is the case for state-of-the-art methods. Also, the formulation does not pose any restrictions on the type of representation, allowing a flexible way of defining "similarity".

Solving (6) then gives us the contribution of each previous image towards constructing the current image, most of which are zero. The relative contribution can then be obtained by normalizing the solution $\boldsymbol{\alpha}$ to obtain the unit vector $\hat{\boldsymbol{\alpha}}$ and applying a threshold ($\tau$) to find the most dominant contribution, hence detecting a loop closure.

Sample results are presented in Fig. 1 for the New College dataset [11]. Flexibility of representation allows us to use a scaled down image (size : $8 \times 6$) and well as a GIST descriptor for each image and be able to detect similar loops. The induced sparsity pattern for a threshold ($\tau = 0.99$) is given in Fig. 2 which shows that the solution is sparse when a loop can be detected because the image is mostly explained by a single image in the past, however in absence of any loops, the previous image is the best explanation for the current image, still keeping the solution sparse. The first three runs in the starting circular region can be observed in the top-left corner of Fig. 2. The problem can be solved efficiently both because of the inherent sparsity as well as the availability of solvers due to the maturity of convex-optimization techniques.

### B. Visual Place Categorization

A closely related problem to loop closing is that of place categorization in which we are interested in finding the class to which the current image belongs rather than a single image. Assuming we have annotated data that can provide us image-class correspondence, where the classes may be corridor, kitchen etc., we can collect images for each class in a contiguous submatrix $\mathbf{A}_i$ for $i = 1 \ldots n_c$ where $n_c$ is the number of classes. The dictionary in (6) will now have the form $\mathbf{D} = [\mathbf{A}_1\ \mathbf{A}_2\ \ldots\ \mathbf{A}_n\ \mathbf{I}]$.

We can then solve (6) and normalize the solution to have a unit norm as before. In this case, we are interested in finding the class with the maximum contribution, which can be obtained by summing up the contributions within each submatrix of the dictionary. $(\underset{i}{argmax} \sum_{j \in |\mathbf{A_i}|} \hat{\boldsymbol{\alpha_j}})$ where $|\mathbf{A_i}|$ are the indices of the columns in the dictionary ($\mathbf{D}$) corresponding to class $i$. Additionally, the columns corresponding to the Identity matrix form a *no-decision* class.

Using this approach along with a GIST based image representation, Carillo et. al [12] have demonstrated results that are comparable to state-of-the-art methods and perform well under different lighting conditions as well as minor changes in the environment over time.

### C. Surprising images / Key-frame discovery

In many applications, we are interested in finding images that represent our lack of knowledge i.e. images that can contribute to the expansion of current model or understanding of the world. In these scenarios, given a image, we need to find a way of quantifying how well our current understanding of the world represents this image. That is exactly the question that $\ell_1$-optimization answers. More specifically, the solution $\boldsymbol{\alpha}$ is composed of two parts, one coming from all the previous observation $\boldsymbol{x}$ and one that comes from the noise (things that cannot be explained by the dictionary yet) part $\boldsymbol{e}$. In order to quantify how well our current image can be represented by all (or a subset of) the previous images, we can solve (6) given all the previous images and the current observation.

In order to find out the innovation (or noise), we find the unit vector $\hat{\alpha}$ and calculate the sum of the component corresponding the noise-term. This gives us an estimate of how well the current image is represented by the dictionary. If the contribution of the noise-term is above a certain threshold, the current image can be added to the dictionary as it represents information that we are currently lacking. Some results obtained using different image sizes and error threshold are give in Fig. 3, where the threshold $\tau$ is the maximum allowed contribution of error. If the error is greater than this threshold, the image in added to the dictionary.

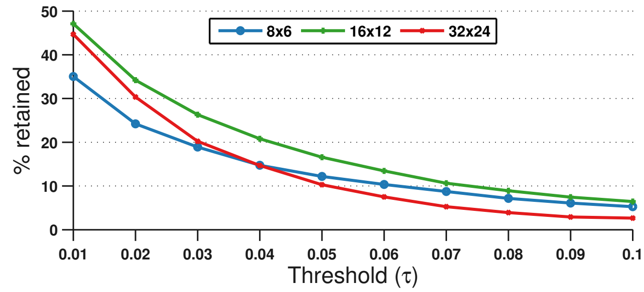



Fig. 3: Percentage of the total (2653) images retained using different thresholds ($\tau$) for the allowed error. Each line represents the size of the scaled image. All plots are for the New College dataset.

## III. Conclusion

We have presented a new formulation as an insight into mapping related problems. The formulation takes advantage of sparsity that is common and inherent to such problems. We have presented three problems to which the formulation can be applied.

## IV. Questions

(**1**) What other problems can be represented by this formulation? (**2**) For loop closure detection, specifically in topological SLAM, what strategies can be applied to ensure robustness against perceptual aliasing?

## References


[1] M. Elad, M. A. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.

[2] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.

[3] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang, “Learning With $\ell_1$-Graph for Image Analysis,” *IEEE transactions on image processing*, vol. 19, no. 4, pp. 858–866, 2010.

[4] E. Amaldi and V. Kann, “On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems,” *Theoretical Computer Science*, vol. 209, no. 12, pp. 237 – 260, 1998.

[5] A. Yang, Z. Zhou, A. Balasubramanian, S. Sastry, and Y. Ma, “Fast $\ell_1$-minimization algorithms for robust face recognition,” *Image Processing, IEEE Transactions on*, vol. 22, no. 8, pp. 3234–3246, Aug 2013.

[6] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Homotopy continuation for sparse signal representation,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*. IEEE, 2005.

[7] D. L. Donoho and Y. Tsaig, *Fast Solution of $\ell_1$-Norm Minimization Problems When the Solution May Be Sparse.* Department of Statistics, Stanford University, 2006.

[8] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 886–893.

[9] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[10] Y. Latif, G. Huang, J. Leonard, and J.Neira, “An Online Sparsity-Congnizant Loop-Closure Algorithm for Visual Navigation,” in *Robotics: Science and Systems*, 2014, (To Appear).

[11] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, “The new college vision and laser data set,” *The International Journal of Robotics Research*, vol. 28, no. 5, pp. 595–599, May 2009.

[12] H. Carrillo, Y. Latif, J. Neira, and J. Castellanos, “Place Categorization using Sparse and Redundant Representations,” in *Proc. IEEE/RJS Int. Conference on Intelligent Robots and Systems*, 2014, (To Appear).