

BLIND CALIBRATION FOR COMPRESSED SENSING BY CONVEX OPTIMIZATION

R. Gribonval[†], G. Chardon^{*} and L. Daudet^{*}

[†] INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France

^{*} Institut Langevin, CNRS UMR 7587, UPMC, Univ. Paris Diderot, ESPCI, 75005 Paris, France

ABSTRACT

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists in unknown gains on each measure. We focus on *blind* calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using ℓ_1 minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, we show that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrate the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observe that the success/failure of the approach seems to obey sharp phase transitions.

Index Terms— compressed sensing; calibration; dictionary learning; blind signal separation; sparse recovery.

1. INTRODUCTION

Linear inverse problems are ubiquitous in signal and image processing, where they are used to estimate an unknown signal $\mathbf{x}_0 \in \mathbb{R}^N$ or \mathbb{C}^N from noisy linear measurements:

$$\mathbf{y} := \mathbf{M}\mathbf{x}_0 + \mathbf{n} \in \mathbb{R}^m \text{ or } \mathbb{C}^m.$$

When $m < N$ this well-known under-determined problem admits infinitely many solutions, but if \mathbf{x}_0 is sparse enough it can be estimated accurately using sparse regularization. Among many other techniques, ℓ_1 regularization has recently become quite popular and consists in solving, e.g.,

$$\hat{\mathbf{x}}_1 := \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2 \leq \epsilon$$

for a well chosen ϵ . When the measurement matrix \mathbf{M} is perfectly known, this approach is known to perform well, and a well-established body of work characterizes its performance guarantees for the recovery of vectors \mathbf{x}_0 that are sufficiently well approximated by highly sparse vectors (see, e.g. [1, 2]). Such sparse linear regression problems occur in many practical scenarios where the measurement matrix \mathbf{M} is either dictated by the physics of the measurement system, or designed to have favorable properties with respect to the recovery of sparse vectors: this is the now famous compressed sensing (CS) scenario (see, e.g. [3, 4]), where \mathbf{M} *voluntarily* reduces dimension, exploiting the sparsity of \mathbf{x}_0 to capture it with fewer measurements than Nyquist sampling would require.

This work was partly funded by the Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006) and by the European Union, project SMALL (FP7-ICT-225913-SMALL). LD is on a joint affiliation between Univ. Paris Diderot and Institut Universitaire de France.

1.1. The decalibration issue

In practical situations, the true measurement system \mathbf{M} is not perfectly known: it may only have been modeled; or it may have been measured through a calibration process, but the physical conditions of the system (such as temperature) may have drifted since this calibration.

Exploiting sparse regularization with an inaccurate estimate $\widehat{\mathbf{M}}$ of the true measurement system \mathbf{M} is likely to hurt the reconstruction performance [5]. It is believed to be one of the reasons limiting the observed performance of several compressed sensing devices. To address this problem, the most standard existing approaches are:

1. **To ignore the problem.**
2. **To consider decalibration as noise [5]:** $\mathbf{y} \approx \widehat{\mathbf{M}}\mathbf{x}_0 + (\mathbf{M} - \widehat{\mathbf{M}})\mathbf{x}_0$. This leads to solving

$$\hat{\mathbf{x}}_1 := \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \widehat{\mathbf{M}}\mathbf{x}\|_2 \leq \epsilon + \eta$$

with η an estimate of the magnitude of this added noise.

3. **Supervised calibration:** using known training signals $\mathbf{x}_1, \dots, \mathbf{x}_L$ and the corresponding observations $\mathbf{y}_\ell = \mathbf{M}\mathbf{x}_\ell + \mathbf{n}_\ell$. Gathering all data in matrices, this takes the form $\mathbf{Y} = \mathbf{M}\mathbf{X} + \mathbf{N}$. The matrix \mathbf{M} is re-estimated, e.g., as

$$\widehat{\mathbf{M}} := \arg \min_{\mathbf{M}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F^2.$$

1.2. Constraints on calibration

It is sometimes useful to constrain the estimated calibration matrix $\widehat{\mathbf{M}}$ to belong to some family \mathcal{M} of matrices. For example, it is sometimes known/assumed [6] that the unknown \mathbf{M} is sparse in a given dictionary of measurement matrices $\{\mathbf{M}_k\}$: $\mathbf{M} \approx \sum_k \alpha_k \mathbf{M}_k$, $\|\alpha\|_0$ small. Supervised calibration can then be performed by solving, e.g., the convex relaxation

$$\min \|\alpha\|_1 \text{ s.t. } \|\mathbf{Y} - \sum_k \alpha_k \mathbf{M}_k \mathbf{X}\|_F \leq \epsilon.$$

In this paper, we concentrate on a different scenario where the measurement matrix is almost known, up to an unknown gain on each measure: that is to say, $\mathbf{M} = \mathbf{D}_0 \mathbf{M}_0$ where \mathbf{M}_0 is a perfectly known measurement matrix, and \mathbf{D}_0 is an unknown diagonal matrix which i -th entry is a (real or complex) gain d_i applied to the i -th measure of \mathbf{x}_0 associated to the i -th line of \mathbf{M}_0 . This leads to

$$\mathcal{M} := \{\mathbf{M} = \mathbf{D}\mathbf{M}_0, \mathbf{D} = \text{diag}(d_i), d_i \neq 0 \forall i\}.$$

Several practical scenarios can be associated with this assumption, for instance in the case of a microphone array where the frequency response of each microphone needs to be individually calibrated [7]. At each frequency, the calibration problem amounts to choosing unknown gains.

1.3. Unsupervised / blind sparse calibration

In this paper, we are interested in a *blind* calibration problem, where no *known* training signal can be used. While some training signals \mathbf{x}_ℓ have given rise to observed measures \mathbf{y}_ℓ , the training signals themselves are not known. Since one could not hope to calibrate the system without some form of knowledge of \mathbf{x}_ℓ , and the final scenario for using $\widehat{\mathbf{M}}$ will be sparse reconstruction, we assume that the unknown signals \mathbf{x}_ℓ are sparse and somehow statistically diverse. In matrix form, we have $\mathbf{Y} = \mathbf{D}_0 \mathbf{M}_0 \mathbf{X}_0$ where \mathbf{Y} is the $m \times L$ (known) observation matrix, \mathbf{D}_0 is the square and diagonal matrix of size m for the (unknown) calibration coefficients, \mathbf{M}_0 is the $m \times N$ (known) idealized measurement matrix, and \mathbf{X}_0 is the (unknown) $N \times L$ set of calibration signals upon which the only knowledge we have is that they are k -sparse.

The objective is to derive a calibration technique that exploits the knowledge of \mathbf{Y} and \mathbf{M}_0 , as well as the sparsity of the unknown \mathbf{X}_0 , to obtain estimates $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{X}}$, and finally $\widehat{\mathbf{M}} = \widehat{\mathbf{D}} \mathbf{M}_0$. It seems natural to consider the following optimization problem, where the ℓ_1 objective function is intended to promote the sparsity of \mathbf{X} :

$$\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{X}\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{D} \mathbf{M}_0 \mathbf{X}. \quad (1)$$

However, this would naively lead to two major issues:

1. Without further constraint on \mathbf{D} and/or \mathbf{X} , one can scale \mathbf{D} to infinity while letting \mathbf{X} go to zero, leading to a trivial but uninteresting solution; This is usually solved through a normalization constraint $\mathbf{D} \in \mathcal{D}$ where the set \mathcal{D} is bounded;
2. Even with appropriate normalizing constraints on \mathbf{D} , this seems at first a non-convex problem, because of the bilinear nature of the term $\mathbf{D} \mathbf{M}_0 \mathbf{X}$, which is linear / convex separately in \mathbf{D} and \mathbf{X} but not jointly.

1.4. Relation to previous work

The above issues are well known since they are encountered in *blind source separation* [8] and *dictionary learning* [9, 10]. In a way, the unsupervised calibration problem can be seen as a simplified instance of dictionary learning, where the general problem has been expressed, e.g. as [11, 12]

$$\min_{\Phi, \mathbf{X}} \|\mathbf{X}\|_1 \text{ s.t. } \mathbf{Y} = \Phi \mathbf{X}.$$

where the minimum is over $\Phi \in \mathcal{C}$ for a well-chosen bounded set \mathcal{C} (typically, the set of matrices with unit norm columns, or *oblique manifold*), but the dictionary Φ is not restricted to be diagonal.

The present issue is also related to the *basis mismatch* problem [13], where the signals are truly sparse in a basis that is slightly different from the chosen representation basis (for instance with a parametric dictionary using a grid of quantized parameters); or to the perturbed CS problem [5, 14], where multiplicative noise limits the effectiveness of CS. In both cases, the “noise” on the measurement matrix is different for every vector, whereas in the decalibration case considered here, this (unknown) basis is shared by all vectors.

1.5. Contributions

The main contribution of this paper is (Section 2) to show that, under proper parametrization and normalization, the considered unsupervised calibration problem (1) can actually be *exactly* expressed as a convex optimization problem. Numerically solving this problem is

straightforward using off-the-shelf algorithms. In numerical experiments (Section 3), we demonstrate the effectiveness of this approach even for highly uncalibrated measurements, whenever a sufficient number of (unknown but sparse) calibrating signals is provided. Remarkably, we observe that the success/failure of this problem obeys some sharp phase transitions, generalizing to the uncalibrated case the phase transitions studied by Donoho and Tanner [15].

2. PROPOSED APPROACH

As noticed above, the naive formulation of the problem is non-convex, even with a (convex) normalization constraint.

2.1. Convex formulation

To provide a convex formulation we propose to reparameterize the problem. Using the simple assumption that the unknown gains d_i are nonzero, we can write the constraint $\mathbf{Y} = \mathbf{D} \mathbf{M}_0 \mathbf{X}$ as $\Delta \mathbf{Y} = \mathbf{M}_0 \mathbf{X}$, where $\Delta = \mathbf{D}^{-1} = \text{diag}(\delta_i)$.

2.2. Normalization constraint

With this new parametrization, the constraint is convex in the pair (Δ, \mathbf{X}) . However, it is trivially satisfied for the pair $(0, 0)$. To avoid this trivial solution we introduce a convex normalization constraint $\text{Tr}(\Delta) = \sum_i \delta_i = m$. Therefore, the constraint set is $\mathcal{D} := \{\mathbf{D} = \text{diag}(d_i), \sum_i d_i^{-1} = m\}$. Note that many other alternatives could be considered such as $\delta_1 = 1$.

2.3. Proposed blind calibration approach

We end up proposing the following unsupervised calibration approach: given the collection of observed measures \mathbf{Y} and the model of the measurement system \mathbf{M}_0 , we estimate the inverse of the calibration gains $\widehat{\Delta}$ and the training signals $\widehat{\mathbf{X}}$ as:

$$(\widehat{\mathbf{X}}_{\text{cal}}, \widehat{\Delta}) := \arg \min_{\mathbf{X}, \Delta} \|\mathbf{X}\|_1 \text{ s.t. } \Delta \mathbf{Y} = \mathbf{M}_0 \mathbf{X}, \text{Tr}(\Delta) = m. \quad (2)$$

This is a convex problem. Note that, because of the arbitrary constraint on Δ , Δ and \mathbf{X} can only be estimated up to a scaling.

3. EXPERIMENTAL RESULTS

3.1. Considered techniques

We compare two techniques. The proposed approach (2), and an approach which ignores decalibration

$$\widehat{\mathbf{X}}_{\text{uncal}} := \arg \min_{\mathbf{X}} \|\mathbf{X}\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{M}_0 \mathbf{X}.$$

Both convex problems are solved using the `cvx` toolbox [16].

3.2. Data generation

The data was generated by drawing i.i.d. random k -sparse vectors \mathbf{x}_ℓ with a k -sparse support chosen uniformly at random and i.i.d. Gaussian nonzero entries. The idealized measurement matrix \mathbf{M}_0 was drawn from the Gaussian ensemble of size $m \times N$ with $N = 100$. Real positive decalibration coefficients were generated using $\mathbf{D}_0 = \text{diag}(\exp(\mathcal{N}(0, \sigma^2)))$, where σ is the parameter governing the amplitude of decalibration. Hence, a given value of σ

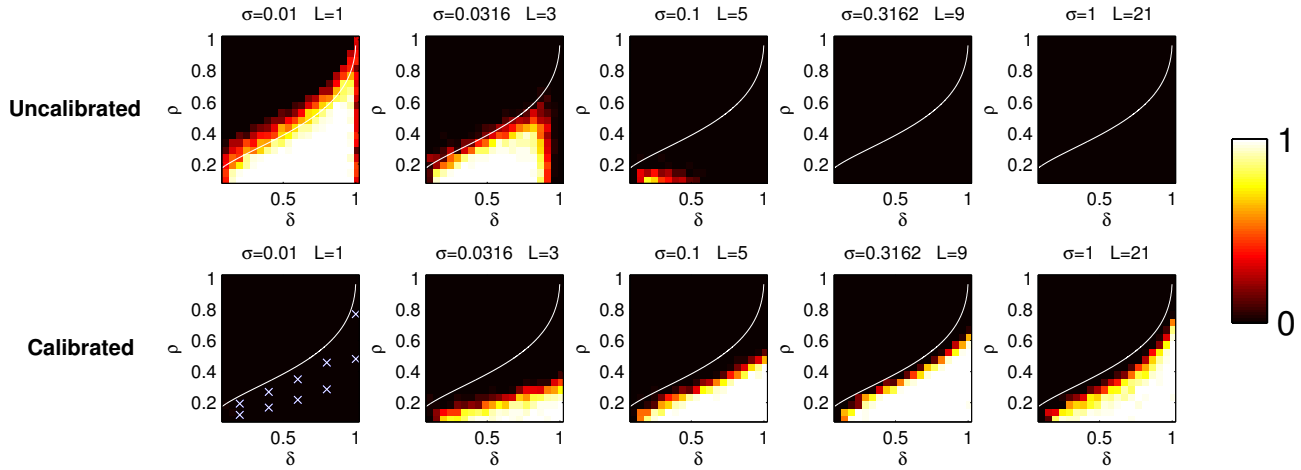


Fig. 1. Empirical phase-transitions with no calibration (top) and proposed calibration (bottom). Overwhelming success (white) or failure (black) is displayed as a function of dimension parameters $\delta := m/N$ (abscissa) and $\rho := k/m$ (ordinate). From left to right: increasing values of the number L of training signals and decalibration amplitude σ . The thin line is the asymptotic Donoho-Tanner phase transition curve for known Gaussian measurement matrix. Crosses on the bottom-left figure indicate choices of (δ, ρ) for the experiments on Fig. 2.

leads to a decalibration offset in decibels with zero mean and standard deviation $\pm 20\sigma/\ln(10) \approx \pm 8.7\sigma$ dB. The considered measures were $\mathbf{Y} = \mathbf{D}_0\mathbf{M}_0\mathbf{X}$, where \mathbf{X} was a concatenation of L i.i.d. k -sparse signals. Experiments were conducted for various configurations of $\delta := m/N$, $\rho := k/m$, L and σ to measure how accurately the signals \mathbf{X} were recovered with the different approaches. Recovery was considered successful if the normalized cross-correlation between original and estimated signals is above 99.5 % (note that, due to the global scaling invariance in Δ and \mathbf{X} mentioned in section 2.3, a distance measure between \mathbf{X} and $\hat{\mathbf{X}}$ would not be a relevant indication of success). All the empirical phase transition diagrams described below report the number of successful estimations, when the experiment was repeated over 50 random draws.

3.3. Can CS still succeed under calibration errors ?

The first study dealt with estimating how much classical CS was robust to calibration errors. To do this, we ran numerical experiments and plotted empirical Donoho-Tanner phase transition [15] (in the framework presented above). Different decalibration levels σ and number L of calibration signals have been tested.

The results can be seen on the top line of Fig.1: for small values of decalibration ($\sigma = .01$, $\sigma = .0316$), the transition curve is barely modified : CS seems indeed robust to small calibration errors. However, as the decalibration increases ($\sigma = .1$, i.e. the decalibration error was of the order ± 9 dB), then the region where CS succeeds drastically shrinks, and eventually disappears at even larger decalibration ($\sigma = .316$ and above).

The second line of fig. 1 shows the result of CS under the blind calibration procedure introduced above. For a large number of noise levels where the un-calibrated experiment failed, it now succeeds even with a relatively small number L of training samples (for instance, with $L = 21$ training samples, it still succeeds at $\sigma = 1$, i.e., a decalibration error of the order ± 8.7 dB). Interestingly, for very small values of decalibration ($\sigma = .01$ and below) and too few training samples, it is better not to perform the blind calibration, as introducing too many degrees of freedom is more detrimental to the calibration process than a slight decalibration is to classical CS.

3.4. Choosing the number of training samples

In this section we wish to determine how many training samples L are needed as a function of the “noise” level σ . We picked different values of (δ, ρ) where ideal CS works, and plotted on Fig.2 the rate of success as a function of decalibration level σ and number of training samples L , both in the un-calibrated and calibrated cases.

In the uncalibrated case, the columns of \mathbf{X} are estimated independently. This implies that the probability of recovering L vectors is the probability of recovering one vector raised to the power of L . This accounts for the fact that the noise threshold above which estimation fails remains constant as L raises, while, under this threshold, performances decreases.

Remarkably, the rates of success of calibrated CS obey sharp corner-like phase transitions. As long as the number of samples L is higher than a threshold that seems independent on the noise level and the ambient dimension N (results not shown here), calibration allows CS to work for σ increased by an order of magnitude compared to the un-calibrated case. Furthermore, in opposition to classical CS, for values of σ where estimation succeeds, the performance raises with L , as more samples are available for the calibration. Sharper transitions are observed for larger values of N . The number of samples necessary to perform calibrated CS thus seems to be dependent on ρ and δ only, i.e. the relative number of measurements and the relative sparsity of the vectors to be reconstructed.

4. CONCLUSION

In this study, we have experimentally observed that even mild decalibration on the measurement matrix can lead to a spectacular failure of standard CS recovery algorithms. This problem is highly relevant for engineering applications, as in many cases the observation matrix is not exactly known. The proposed blind calibration technique, formulated here as a convex optimization problem, is shown to offer a significantly improved robustness to decalibration: when a sufficient number of (unknown but sparse) calibration signals are provided, the algorithm can now succeed even for much larger calibration errors. Remarkably, the success / failure of this new calibrated CS exhibits

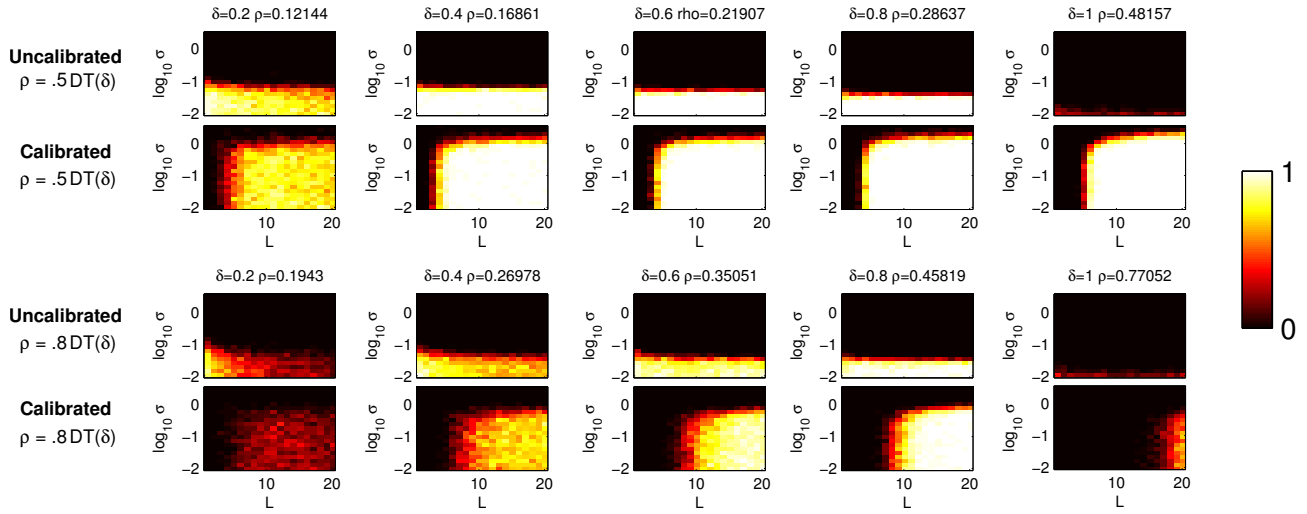


Fig. 2. Experiments at fixed (δ, ρ) , as a function of the number L of training signals and the decalibration amplitude σ . “Uncalibrated” and “calibrated” refer to the standard and proposed algorithms, respectively. White indicates that the algorithm succeeds in recovering the solution, and black indicates a failure.

sharp corner-like phase transitions. Being able to compute theoretically the asymptotic boundaries of these new transition diagrams is a promising extension of this work, together with the extension to more general decalibration cases. More immediate targets will be the design of efficient blind calibration algorithms to scale the approach and test in on real calibration problems arising in compressed sensing of acoustic fields [7].