A Greedy Blind Calibration Method for Compressed Sensing with Unknown Sensor Gains

Valerio Cambareri, Amirafshar Moshtaghpour, Laurent Jacques
Image and Signal Processing Group, ICTEAM/ELEN, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
E-mail: {valerio.cambareri, amirafshar.moshtaghpour, laurent.jacques}@uclouvain.be

Abstract—The realisation of sensing modalities based on the principles of compressed sensing is often hindered by discrepancies between the mathematical model of its sensing operator, which is necessary during signal recovery, and its actual physical implementation, which can amply differ from the assumed model. In this paper we tackle the bilinear inverse problem of recovering a sparse input signal and some unknown, unstructured multiplicative factors affecting the sensors that capture each compressive measurement. Our methodology relies on collecting a few snapshots under new draws of the sensing operator, and applying a greedy algorithm based on projected gradient descent and the principles of iterative hard thresholding. We explore empirically the sample complexity requirements of this algorithm by testing its phase transition, and show in a practically relevant instance of this problem for compressive imaging that the exact solution can be obtained with only a few snapshots.

Index Terms—Compressed Sensing, Blind Calibration, Iterative Hard Thresholding, Non-Convex Optimisation, Bilinear Inverse Problems

I. INTRODUCTION

The implementation of practical sensing schemes based on Compressed Sensing (CS) [1] often encounters physical non-idealities in realising the mathematical model of the sensing operator, whose accuracy is paramount to attaining a high-quality recovery of the observed signal [2]. Among such non-idealities, we here focus on the case in which each compressive measurement is affected by an unknown multiplicative factor or sensor gain, i.e.,

\[ \mathbf{y}_l = \text{diag}(\mathbf{g}) \mathbf{A}_l \mathbf{x}, \ l \in [p] = \{1, \ldots, p\}, \]  

(1)

where \( \mathbf{x} \in \mathbb{R}^n \) is the input signal, \( \mathbf{A}_l \in \mathbb{R}^{m \times n}, \ l \in [p] \) are independent and identically distributed (i.i.d.) random sensing matrices, and \( \mathbf{y}_l \in \mathbb{R}^m, \ l \in [p] \) are the respective snapshots of measurements obtained by applying each sensing matrix to \( \mathbf{x} \) (the reason why the acquisition is partitioned in snapshots will be cleared below). In this uncalibrated sensing model \( \mathbf{g} \in \mathbb{R}_+^p \) is an unknown set of positive-valued gains that remains identical throughout the snapshots, but whose value is unknown. Hence, this sensing model is bilinear in \( \mathbf{x} \) and \( \mathbf{g} \), and retrieving both quantities given the measurements is a non-trivial bilinear inverse problem (BIP). Note that (1) can be practically realised in compressive imaging schemes using snapshot (i.e., parallel) acquisition by convolving an input signal with one or more random masks, such as those detailed in [3]–[7]. When sensor gains are not calibrated, e.g., in the presence of fixed-pattern noise or strong pixel-response non-uniformity [8], taking a few snapshots allows for on-line blind calibration without missing any instance of the signal \( \mathbf{x} \) due to an off-line calibration process, as we showed in previous contributions [9], [10]. There, we proved that instances of (1) with sensing matrices having i.i.d. sub-Gaussian entries (for a rigorous definition, see [11]) and \( (\mathbf{x}, \mathbf{g}) \) being either unstructured or endowed with subspace models can be solved by a simple, suitably initialised projected gradient descent on a non-convex objective. The number of measurements ensuring the recovery of the exact solution was shown to be \( mp \gtrsim n + m \), i.e., a linear sample complexity in the dimensions of the unknowns (up to log factors, and referring to the findings in [10]).

In this paper we focus on the case in which the single input signal \( \mathbf{x} \) has a \( k \)-sparse representation in a known basis. To leverage this more involved model on \( \mathbf{x} \) we simply resort to a hard thresholding operator at each iterate of our former non-convex algorithm, turning it into a greedy scheme. The proposed greedy approach allows for blind calibration in actual CS schemes; the additional requirement of our methodology is a set of \( p \) snapshots that collects a sufficient amount of information on \( (\mathbf{x}, \mathbf{g}) \). Our emphasis is on assessing, at least empirically, how the sample complexity can be reduced in function of the signal-domain sparsity \( k \) (up to log factors). Hence, provided \( \mathbf{x} \) is sufficiently sparse, we will show empirically that the total amount of measurements \( mp \) can be lower than \( n \) while still recovering both \( (\mathbf{x}, \mathbf{g}) \).

A. Related Work

Blind calibration of sensor gains has been tackled in recent literature, starting from initial approaches for uncalibrated sensor networks in [12], [13], and more recently for radio-interferometry [14]. In the context of CS, some algorithms have been proposed to cope with such model errors [15]–[19]. Interestingly, most algorithms use sparse or known subspace models for several input signals, rather than random draws of the sensing operator itself (as typically feasible in optical CS schemes [6], [7]); moreover, these works do not attain sample complexity results that grant exact recovery. A first

\footnote{Hereafter, given two functions \( f, g \), \( f \gtrsim g \) indicates that \( f > Cg \) for some constant \( C > 0 \).}
work proposing such provable guarantees using a single sparse input signal was introduced by Ling and Strohmer [20] based on a lifting approach to the problem (as in [21]; improved guarantees were outlined in [22]). The main drawback of this approach is its computational complexity, given that it corresponds to very large-scale semidefinite programming.

Our former contributions [9], [10] then showed that a non-convex approach could provide exact recovery guarantees and computational advantages with respect to (w.r.t.) lifting approaches; these were inspired by the methodologies of Candès et al. [23], Sanghavi et al. [24], and Sun et al. [25] used for the closely related problem of phase retrieval.

For what concerns the related task of blind deconvolution, very recent approaches to this BIP apply similar non-convex schemes [26], [27] or alternating minimisation [28], [29], yet targeting a more general context than blind calibration and related schemes [26], [27] or alternating minimisation [28], [29], yet very recent approaches to this BIP adopt similar non-convex

B. Contributions and Outline

Our paper extends the non-convex algorithm devised in [9], [10] to account for a sparse model in the signal domain; this is a fundamental prior for CS, whereas sparse models on the gains $g$ could be inapplicable when these are drawn at random as each sensor is manufactured. Thus, we adopt a greedy algorithm to enforce signal-domain sparsity, and detail its empirical performances as a function of our BIP’s dimensions.

Our findings are presented as follows. In Sec. II we introduce the non-convex problem and propose a greedy algorithm based on hard thresholding. This algorithm is studied numerically in Sec. III, where we focus on the empirical phase transition as the problem dimensions vary. We then simulate a practical case of blind calibration for compressive imaging in Sec. IV. A conclusion is drawn afterwards.

II. A GREEDY AND NON-CONVEX APPROACH TO BLIND CALIBRATION

Our initial approach to the blind calibration problem involved defining a simple Euclidean loss,

$$f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^{p} \| \text{diag}(\gamma) A_l \xi - y_l \|_2^2,$$

and solving

$$(\hat{x}, \hat{y}) = \arg\min_{\xi \in \mathbb{R}^n; \gamma \in \Pi^+_m} f(\xi, \gamma)$$

where $\Pi^+_m := \{v \in \mathbb{R}^n : 1_m^T v = m \}$ is the scaled probability simplex and $1_m$ the vector of ones in $\mathbb{R}^m$. To begin with, up to a scaling all points in

$$\{(\xi, \gamma) \in \mathbb{R}^n \times \mathbb{R}^m : \xi = \alpha x, \gamma = \frac{y}{\alpha}, \alpha \neq 0\}$$

are minimisers of $f(\xi, \gamma)$ (i.e., the scaling of $(x, g)$ is anyway unrecoverable), so we adopted the constraint $\gamma \in \Pi^+_m$ which fixes one admitted solution for $\alpha = \frac{y}{\|y\|}$.

We now proceed to devise an algorithm solving (5) that accounts for the two constraints.

This ensures that the steps are taken on $1_m^\perp$. In theory, we would have to use the projection operator $\mathcal{P}_{\hat{G}_\rho}$ to ensure that a gradient step still belongs to this convex set; however, when we start from an initialisation $\gamma_0 := 1_m$, we have observed that the algorithm will remain inside $\hat{G}_\rho$ when convergent or, conversely, diverge independently of the presence of $\mathcal{P}_{\hat{G}_\rho}$. Thus, we will not practically use this projector, while it will be necessary for devising guarantees as in [9].

$$\mathcal{P}_{\hat{G}_\rho} := 1_m^\perp,$$

$$\hat{G}_\rho := 1_m^\perp \cap 1_m, i.e., in a subset \hat{G}_\rho \subset \Pi^+_m.$$
Secondly, as typically done in greedy algorithms, instead of adopting a proxy for sparsity such as the \( \ell_1 \)-norm we iteratively enforce it by evaluating the gradient

\[
\nabla_{\xi} f(\xi, \gamma) = \frac{1}{mp} \sum_{l=1}^{p} A_l^T \text{diag}(\gamma)(\text{diag}(\gamma)A_l \xi - y_l)
\]

and applying after each gradient step the hard thresholding operator \( H_k \), which sets all but the \( k \) largest-magnitude entries of the argument to 0. This operator is at the heart of Iterative Hard Thresholding (IHT, [35]) and allows us to enforce signal-domain sparsity. Finally, as in [9] we choose an initialisation by backprojection, i.e., \( \xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l) y_l \) that is an unbiased estimate of \( x \), i.e., as \( p \to \infty \) we have that \( \xi_0 \to \mathbb{E} x \). With all previous considerations, we approach our version of Blind Calibration with Iterative Hard Thresholding (BC-IHT), as summarised in Alg. 1. The line searches reported in step 3 can be computed in closed form, as they are crucial to accelerate the algorithm (albeit in a sub-optimal fashion). The step-size could be further optimised over the non-linear cost: this may yield faster convergence (see, e.g., [36]), but will be the subject of a future improvement of BC-IHT.

### III. Empirical Phase Transition

We here propose an extensive experimental assessment of the phase transition of BC-IHT. We explore the effect of the problem dimensions in (1) on the successful recovery of both the signal and the gains, by varying \( n = \{2^9, 2^{10}\}, k = \{2^5, 2^6, 2^7\}, p = \{2, 2^3, \ldots, 2^5\} \) and \( m = \{2, 2^4, \ldots, 2^5\} \cdot k \), while generating 144 random instances of the problem for each of the configurations. In detail, \( x \sim_{\text{i.i.d.}} \mathcal{N}(0, 1) \) is drawn as a standard Gaussian random vector; \( g \) is drawn uniformly at random on \( \mathbb{G}_p \) for \( \rho = \frac{1}{2} \); \( A_l \sim_{\text{i.i.d.}} \mathcal{N}^{m \times n}(0, 1) \) are drawn as i.i.d. Gaussian random matrices. We let the algorithm run given \( y_l \) and \( A_l, l \in \{p\} \) up to a relative change of \( 10^{-7} \) in the signal and gain updates. Then, we measure the probability of successful recovery

\[
P_\zeta(n, k, m, p) := \mathbb{P} \left[ \max \left\{ \frac{\|g-g\|_2}{\|g\|_2}, \frac{\|\hat{x}-x\|_2}{\|x\|_2} \right\} < \zeta \right]
\]

on the trials, with \( \zeta = -60 \text{ dB} \) (this corresponds to an early termination of the algorithm: when convergent, it will reach the exact solution, provided we let it run for a sufficient number

---

Figure 1. Empirical phase transition of Alg. 1 as \( n \) increases (top to bottom) and \( k \) increases (left to right), as a function of \( \frac{m}{n} \), \( p \) and fixing \( \rho = \frac{1}{2} \). We report the estimated contour levels of the probability of successful recovery, as it exceeds the value indicated above of each curve.
of iterations). The results are reported in Fig. 1 in terms of the contour levels of $P_z$, as a function of $\log_2 \frac{m}{k}$ and $\log_2 p$.

While a theoretical sample complexity result that grants provable convergence is still under study, we can already appreciate that the effect of increasing $n$ for fixed sparsity levels has a mild effect on the region in which $P_z > 0.95$, while it does sharpen the transition region as typically observed in standard CS. Moreover, we can appreciate the impact of increasing $k$ on the transition region while keeping the ratios $\frac{m}{n}$ fixed: for larger values of $k$, the region in which the algorithm fails to converge almost surely is rapidly reduced. Moreover, we reported in red the curve that matches $mp = C(k + m)$ (i.e., $\log_2 p = \log_2 C(1 + \frac{m}{k})$) for some $C > 0$, which roughly follows the contours’ shape in our experiments.

We highlight that all the empirical evidence collected in this context correctly suggests that $p > 1$: this agrees with our previous finding that $p \geq \log m$ (see [9, Proposition 2]), i.e., if no structure is leveraged on the gains $g$ more than one snapshot will always be needed for the algorithm to collect a sufficient amount of information on $g$.

Thus, by interpreting the results, we can expect that if $m \approx 5k$ (a widely used rule of thumb in CS), our blind calibration method will converge for most instances of (1) and $p < 1$, once we let $p > 4$ we will be able to recover both $(x, g)$. If furthermore $k$ is sufficiently low, the total undersampling factor $\frac{mp}{n}$ will be below 1.

IV. BLIND CALIBRATION FOR COMPRESSIVE IMAGING

We now proceed to apply BC-IHT in a practical case, in which we process a high-dimensional red-green-blue (RGB) image $x$ of dimension $n = 256 \times 256$ pixels, which is made sparse w.r.t. a Daubechies-4 orthonormal wavelet basis with only $k = 1800$ non-zero coefficients. Then $x$ is acquired by means of Gaussian random sensing matrices $A_l, l \in [p]$. This experiment could be carried out with other sub-Gaussian matrix ensembles such as Bernoulli sensing matrices, with the results being substantially unaltered. Since the sparsity level of the chosen test image is high, we can simulate its acquisition with a sensor array of $m = 103 \times 103$ pixels ($m \approx 6k$) and use $p = 5$ snapshots to meet the requirements of our method; thus $\frac{mp}{n} \approx 0.8$, and once the gains are retrieved this CS scheme could revert to $\frac{m}{n} \approx 0.16$ while benefiting from the improved model accuracy provided by blind calibration. As for the gains, we set $\rho = \frac{1}{2}$ and draw $g$ uniformly at random from $\mathcal{G}_p$.

We then run BC-IHT on each of the RGB channels separately, until the relative change in the signal and gain estimates falls below $10^{-7}$; the quality and data reported below are the worst case among the colour channels. This causes the algorithm to run for 884 iterations, achieving a high-quality estimate having $\text{RSNR}_{x,\hat{x}} = -20 \log_{10} \frac{\|x - \hat{x}\|}{\|x\|} = 153.16$ dB and $\text{RSNR}_{g,\hat{g}} = -20 \log_{10} \frac{\|g - \hat{g}\|}{\|g\|} = 122.76$ dB. The quality of the estimates can be observed in Fig. 2c and 2e.

To see the beneficial effect of blind calibration, we use the accelerated version of IHT [36] given the exact sparsity level $k$, the snapshots $y_l$, and the corresponding sensing matrices, which form a standard CS model when concatenated vertically. Hence, accelerated IHT attempts to recover an estimate $\hat{x}$ while neglecting the model error. The algorithm converges in only 29 iterations to a local minimiser $\hat{x}$, whose $\text{RSNR}_{x,\hat{x}} = 17.83$ dB. Such modest performances can be seen directly in Fig. 2b. No comparison with other blind calibration algorithms

Figure 2. A numerical example of blind calibration for compressive imaging: the test image is a detail of “Tous les jours”, René Magritte, 1966, © Charly Herscovici, with his kind authorization - c/o SABAM-ADAGP, 2011. The artwork was retrieved at wikiart.org and is intended for fair use. A comparison of the original and retrieved signal and gains ($\rho = \frac{1}{2}$) is reported in a-c and d-e, respectively.
is here explored, since the choice of using a single sparse input and multiple snapshots is specific to our framework. Nevertheless, we note that (i) the computational complexity of our algorithm is competitively low, as it amounts to that of IHT plus an additional projected gradient step in the gain domain per iteration; (ii) just as a proof of convergence for IHT to a local minimiser has been devised, we expect to have provable convergence results in the same fashion, which will lead to a bound on the sample complexity that ensures the retrieval of the exact solution.

V. CONCLUSION

We proposed a novel approach to blind calibration based on the use of snapshots with multiple draws of the random sensing operator, and on a greedy algorithm which enforces sparsity on the steps resulting from gradient descent on a non-convex objective. Our approach is capable of achieving, within a few snapshots, perfect recovery of the signal and gains in computational efficiency. Hence, we conclude that when sensor calibration is a cause of concern in a sensing scheme, introducing a modality that follows (1) and using our method could be a viable option to cope with model errors.

We envision that our method may be used both for blind calibration of imaging sensors, as well as distributed sensor arrays or networks if suitably modified to allow for compressive sensing. While we presented empirical evidence on the phase transition of our algorithm, a more rigorous convergence guarantee is the subject of our current study and will be presented in a future communication.

REFERENCES

[1] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] M. A. Herman and T. Strohmer, “General deviations: An analysis of perturbations in compressed sensing,” IEEE J. Sel. Top. Signal Process., vol. 4, no. 2, pp. 342–349, 2010.
[3] J. Romberg, “Compressive sensing by random convolution,” SIAM J. Imag. Sci., vol. 2, no. 4, pp. 1098–1128, 2009.
[4] T. Björkland and E. Magli, “A parallel compressive imaging architecture for one-shot acquisition,” in 2013 IEEE Picture Coding Symposium (PCS), 2013, pp. 65–68.
[5] A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, “Imaging with nature: Compressive imaging using a multiply scattering medium,” Sci. Rep., vol. 4, 2014.
[6] S. Bahmani and J. Romberg, “Compressive deconvolution in random mask imaging,” IEEE Trans. Comput. Imaging, vol. 1, no. 4, pp. 236–246, 2015.
[7] J. P. Dumas, M. A. Lodhi, W. U. Bajwa, and M. C. Pierce, “Computational imaging with a highly parallel image-plane-coded architecture: challenges and solutions,” Opt. Express, vol. 24, no. 6, pp. 6145–6155, Mar. 2016.
[8] M. M. Hayat, S. N. Torres, E. Armstrong, S. C. Cain, and B. Yasuda, “Statistical algorithm for nonuniformity correction in focal-plane arrays,” Appl. Opt., vol. 38, no. 5, pp. 772–780, 1999.
[9] V. Cambalieri and L. Jacques, “A non-convex blind calibration method for randomised sensing strategies,” in 2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), Sep. 2016, pp. 16–20.
[10] ——, “Through the Haze: A Non-Convex Approach to Blind Calibration for Linear Random Sensing Models,” arXiv preprint arXiv:1610.09028, 2016, submitted to Information and Inference: A Journal of theIMA.
[11] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” in Compressed Sensing: Theory and Applications. Cambridge University Press, 2012, pp. 210–268.