Sequential Learning of Analysis Operators

Michael Sandbichler and Karin Schnass
Department of Mathematics, University of Innsbruck
Technikerstraße 13, A-6020 Innsbruck
Email: {michael.sandbichler, karin.schnass}@uibk.ac.at

I. Introduction

Many tasks in high dimension signal processing, such as denoising or reconstruction from incomplete information, can be efficiently solved if the data at hand is known to have intrinsic low dimension. One popular model with intrinsic low dimension is the union of subspace model, where every signal is assumed to lie in one of the low dimensional linear subspaces. However, as the number of subspaces increases, the model becomes more and more cumbersome to use unless the subspaces can be parametrised. Two examples of large unions of parametrised subspaces, that have been successfully employed, are sparsity in a dictionary and cosparsity in an analysis operator. In the sparse model the subspaces correspond to the linear span of just a few normalised columns from a $d \times K$ dictionary matrix, $\Phi = (\phi_1 \dots \phi_K)$ with $\|\phi_k\|_2 = 1$, meaning any data point y can be approximately represented as superposition of $S \ll d$ dictionary elements, $y \approx \sum_{j=1}^S \phi_{i_j} x_{i_j}$. In the cosparse model the subspaces correspond to the orthogonal complement of the span of some normalised rows from a $K \times d$ analysis operator $\Omega = (\omega_1^\star \dots \omega_K^\star)^\star$ with $\|\omega_k\|_2 = 1$, meaning any data point y is orthogonal to ℓ analysers, meaning the vector Ωy has ℓ zero entries.

However, before being able to exploit these models for a given data class it is necessary to identify the parametrising dictionary or analysis operator. This can be done either via a theoretical analysis or a learning approach. While dictionary learning is by now an established field, see [1] for an introductory survey, results in analysis operator learning are still relatively scarse, [2], [3], [4], [5], [6], [7]. In this work we will contribute to the development of the field by taking an optimisation approach, which will lead to an online algorithm for learning analysis operators. For further details, see [8].

II. The target function

Suppose we are given a batch of signals $y_1, \dots, y_N \in \mathbb{R}^d$ which are ℓ-cosparse with respect to some (unknown) operator $\Omega \in \mathbb{R}^{K \times d}$. Our goal is to learn the operator Ω from the data.

Define $\mathcal{A} = \{\Gamma \in \mathbb{R}^{K \times d} \colon \|\gamma_k\|_2 = 1\}$, where γ_k denotes the k-th row of Γ and $\mathcal{X}_\ell = \{(x_1, x_2, \dots, x_N) \in \mathbb{R}^{K \times N} \colon |\mathrm{supp}(x_n)| = K - \ell\}$. If we collect the given data $y_1 \dots, y_N$ in a matrix $Y = (y_1, \dots, y_N) \in \mathbb{R}^{d \times N}$, then the unknown operator Ω should satisfy $\|\Omega Y - X\|_F = 0$ for some $X \in \mathcal{X}_\ell$ since every column x_n can be used to zero out all non-zero entries in Ωy_n. This suggests to find Ω via the following opimisation program also suggested in [9], [6]

$$\hat{\Omega} = \operatorname*{argmin}_{\Gamma \in \mathcal{A}} \min_{X \in \mathcal{X}_\ell} \|\Gamma Y - X\|_F^2. \tag{1}$$

This can be rewritten as

$$\hat{\Omega} = \operatorname*{argmin}_{\Gamma \in \mathcal{A}} \sum_{n=1}^N \min_{J \subset [K] \colon |J| = \ell} \|\Gamma_J y_n\|_2^2, \tag{2}$$

where Γ_J denotes the submatrix of Γ consisting only of the rows of Γ indexed by J. So in order to find an estimate for Ω, we need to minimize the function f_N over $\mathcal{A}$, where

$$f_N(\Gamma) = \sum_{n=1}^N \min_{J \subset [K] \colon |J| = \ell} \|\Gamma_J y_n\|_2^2. \tag{3}$$

III. The ISAOL algorithm

Our next step consists in finding an efficient way to minimize the target function given in equation (3). In the spirit of online learning, we will employ a stochastic gradient descent algorithm in order to achieve this goal. The derivative of f_N with respect to a row γ_k of Γ is given by

$$\frac{\partial f_N}{\partial \gamma_k}(\Gamma) = \gamma_k \underbrace{\sum_{n \colon k \in J_n} 2 y_n y_n^\star}_{=: A_k(\Gamma)}, \tag{4}$$

where $J_n = \operatorname{argmin}_{|J| = \ell} \|\Gamma_J y_n\|_2^2$. Using this expression we now perform a semi-implicit gradient step followed by a projection onto the unit sphere for each of the rows of Γ, i.e. $\bar{\gamma}_k = \gamma_k - \alpha \bar{\gamma}_k A_k(\Gamma)$. Setting $\alpha = 1$, yields the Implicit Sequential Analysis Operator Learning (ISAOL) algorithm given in Table I. Note that this algorithm is sequential with respect to the data y_n, since every data point is used to perform the update of the matrices A_k, but not needed anymore later in the algorithm.

Figure 1 shows the performance of ISAOL on synthetic data. For a random initalization, the algorithm is not able to achieve full recovery due to multiple recovery of several rows. This can for example be tackled by introducing a replacement strategy, cf. [8].

Figure 2 (right) shows the result of performing analysis operator learning with the ISAOL algorithm for image patches taken from the Fabio image (Figure 2 left).

The advantage of using an implicit instead of an explicit step in the gradient descent is increased stability with respect to the choice of stepsize, cf. [8] and Figure 3. There ISAOL and SAOL (also presented in [8]) have been compared to the (incoherent) Analysis SimCO algorithms, which are similarly gradient descent based analysis operator learning algorithms.

In Table II the results for image denoising with the learned operators are presented again in comparison to the ASimCO algorithms. For the precise setups of the shown experiments, see [8].

IV. Conclusion

We presented a new sequential algorithm for learning analysis operators in an online setting along with numerical experiments showing the competitive performance of the algorithm in comparison to the state of the art algorithm ASimCO.

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) under Grant no. Y760. In addition the computational results presented have been achieved (in part) using the HPC infrastructure LEO of the University of Innsbruck.

ISAOL(Γ,ℓ,Y) - (one iteration)

- For all $n \in [N]$
  - Find $J_n = \operatorname{argmin}_{|J|=\ell} \|\Gamma_J y_n\|_2^2$
  - For all $k \in J_n$ update $A_k = A_k + y_n y_n^\star$
- Set $\bar{\gamma}_k = \gamma_k (I + A_k)^{-1}$
- Output $\bar{\Gamma} = (\frac{\bar{\gamma}_1}{\|\bar{\gamma}_1\|_2}, \dots, \frac{\bar{\gamma}_K}{\|\bar{\gamma}_K\|_2})^\star$

TABLE I
THE ISAOL ALGORITHM

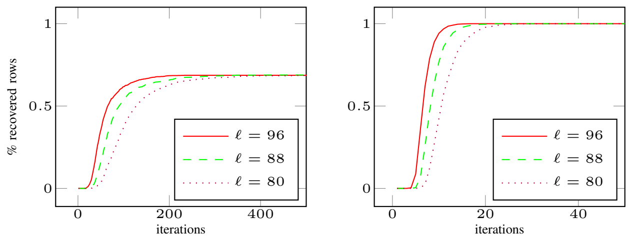

Fig. 1. Recovery rates of ISAOL from signals with various cosparsity levels ℓ in a noisy setting, using a random (left) and a closeby (right) initialisation. The saturation for the random initialization occurs due to multiple recovery of some of the rows of the operator. This can be for example tackled by introducing a replacement strategy.

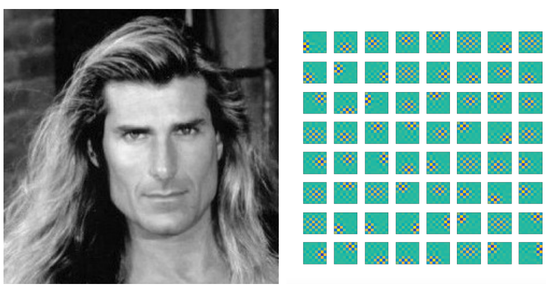

Fig. 2. The image, from which the analysis operator was learned (left) and the recovered operator for $\ell = 57$ (right).

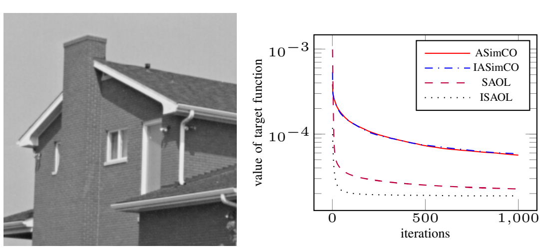

Fig. 3. Decay of the target function using (I)ASimCo and (I)SAOL (right) to learn a 128×64 operator for the House image (left).

REFERENCES

[1] R. Rubinstein, A. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[2] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[3] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Analysis operator learning for overcomplete cosparse representations," in *Signal Processing Conference, 2011 19th European*. IEEE, 2011, pp. 1470–1474.

[4] ——, "Constrained overcomplete analysis operator learning for cosparse signal modelling," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2341–2355, 2013.

[5] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2138–2150, 2013.

[6] R. Rubinstein, T. Peleg, and M. Elad, "Analysis k-svd: a dictionary-learning algorithm for the analysis sparse model," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 661–677, 2013.

[7] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-like algorithms for the cosparse analysis model," *Linear Algebra and its Applications*, vol. 441, pp. 22–60, 2014.

[8] M. Sandbichler and K. Schnass, "Sequential learning of analysis operators," *arXiv preprint arXiv:1704.00227*, 2017.

[9] J. Dong, W. Wang, W. Dai, M. Plumbley, Z. Han, and J. Chambers, "Analysis SimCO algorithms for sparse analysis model based dictionary learning," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 417–431, 2016.

Algorithm	σ(PSNR)	Pep	Cam	Hou	Bar	Man
ASimCO	12.8(25)	31.29	**30.35**	30.99	30.49	28.31
IASimCO		**31.32**	30.31	30.59	30.33	28.05
SAOL		31.26	30.27	32.67	**30.88**	**28.76**
ISAOL		31.08	30.24	**32.73**	30.85	28.62
ASimCO	45(15)	24.53	23.31	26.80	25.23	23.07
IASimCO		24.32	23.05	26.34	25.13	22.91
SAOL		24.50	23.24	26.53	25.12	23.30
ISAOL		**24.64**	**23.39**	**27.05**	**25.32**	**23.33**

TABLE II
PERFORMANCE OF (I)ASIMCO AND (I)SAOL FOR DENOISING FOR DIFFERENT PICTURES AND NOISE LEVELS. FOR EACH SETTING AND IMAGE, THE BEST ACHIEVED PSNR IS HIGHLIGHTED.