2018 5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)

Dictionary learning for multiplicative distortions with applications to SAR autofocus

Joachim H.G. Ender, Otmar Loffeld
Center for Sensor Systems ZESS, University of Siegen, Germany
Fabio Giovanneschi
Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR
Wachtberg, Germany

***Abstract*—When in the measurement model $\mathbf{Y} = \mathbf{DX} + \mathbf{N}$ with an unknown but sparse coefficient matrix $\mathbf{X}$ also the dictionary $\mathbf{D}$ is unknown, a pair $(\hat{\mathbf{D}}, \hat{\mathbf{X}})$ can be found minimizing the Euclidean distance between model and measurements. This procedure is known as *dictionary learning* and can be tackled e.g. with the K-SVD algorithm [1]. In our paper, we propose an algorithm for the case of linear constraints for the unknown dictionary, called *Linearly constrained dictionary learning (LCDL)*. The dictionary learning problem and the K-SVD algorithm can be modified to a model including an unknown multiplicative distortion to the measurements which can be evaluated as a linear combination of basis vectors enabling the application of LCDL. Further we propose a second algorithm based on *blind factorization* using ADMiRA (Atomic Decomposition for Minimum Rank Approximation) [2] as the most important component. Finally, we present applications for the removal of phases in SAR measurements induced by unknown motions of the carrier as an alternative to well-known autofocus algorithms like the phase gradient algorithm (PGA) [3] or the eigenvector method (EM) [4].**

I. Introduction

In this paper, we will address the problem to recover a sparse representation with respect to a dictionary $\mathbf{D}$ and the sparse coefficient matrix $\mathbf{X}$, observed by the measurements

$$\mathbf{Y} = \mathrm{diag}(\mathbf{e})\mathbf{DX} + \mathbf{N} \tag{1}$$

where $\mathbf{N}$ is the measurement noise and $\mathbf{e}$ is a multiplicative distortion vector, from which is only known that it is an element of a certain subspace E (for a better clarity, in this section we abstain from enumerating the dimensions).

The columns of $\mathbf{Y}$ may be regarded as 'training signals' for the estimation $\hat{\mathbf{e}}$ of $\mathbf{e}$, such that $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\mathbf{e}})\mathbf{D}$ can serve further for the simpler problem to recover future sparse coefficient matrices $\tilde{\mathbf{X}}$ from the model

$$\tilde{\mathbf{Y}} = \hat{\mathbf{D}}\tilde{\mathbf{X}} + \tilde{\mathbf{N}}. \tag{2}$$

The problem Eq. (1) may be slightly generalized by assuming that the measurements are performed via a matrix transformation:

$$\mathbf{Y} = \mathbf{G}\,\mathrm{diag}(\mathbf{e})\mathbf{DX} + \mathbf{N}. \tag{3}$$

The algorithm proposed in this paper will cover also this generalization.

Possible strategies to get a solution are either to minimize the error between measurement and model under the constraint of a bounded ℓ_0-norm of the coefficient vectors:

$$\begin{aligned} &\text{Minimize } \|\mathbf{Y} - \mathrm{diag}(\mathbf{e})\mathbf{DX}\|_F \\ &\text{subj. to } \|\mathbf{x}^{(q)}\|_0 \leq K \,\forall q \in [Q], \mathbf{e} \in E. \end{aligned} \tag{4}$$

Q here is the number of training signals, and $\mathbf{x}^{(q)}, q \in [Q]$ are the column vectors of $\mathbf{X}$, or to bound the measurement error and find an ℓ_p-minimum ($p = 0$ or $p = 1$):

$$\begin{aligned} &\text{Minimize } \|\mathbf{X}\|_p \\ &\text{subj. to } \|\mathbf{Y} - \mathrm{diag}(\mathbf{e})\mathbf{DX}\|_F \leq \sigma, \mathbf{e} \in E. \end{aligned} \tag{5}$$

We will concentrate to the first approach.

As mentioned, $\mathbf{e}$ is restricted to a subspace E, i.e. $\mathbf{e}$ can be expressed as a linear combination of basis vectors $\mathbf{b}_l, l \in [L]$:

$$\mathbf{e} = \sum_{l=1}^{L} \beta_l \mathbf{b}_l, \tag{6}$$

so

$$\tilde{\mathbf{D}} := \operatorname{diag}(\mathbf{e})\mathbf{D} \quad (7)$$

$$= \sum_{l=1}^{L} \beta_l \operatorname{diag}(\mathbf{b}_l)\mathbf{D} \quad (8)$$

$$= \sum_{l=1}^{L} \beta_l \mathbf{D}^{[l]}. \quad (9)$$

with $\mathbf{D}^{[l]} := \operatorname{diag}\mathbf{b}_l\mathbf{D}$. Since the coefficients β_l are unknown, the dictionary $\tilde{\mathbf{D}}$ is an unknown element of a linear subspace of dictionaries $\mathcal{D} = \operatorname{span}(\mathbf{D}^{[1]}, \ldots, \mathbf{D}^{[L]})$ where the spanning set of dictionaries $\mathbf{D}^{[l]}, l \in [L]$ are generated by the 'mother directory' $\mathbf{D}$. Consequently, our primary problem Eq. (1) can be formulated in the error-minimizing frame as a dictionary learning problem with a restricted set $\mathcal{D}$ of admissible dictionaries:

$$\begin{gathered}\text{Find solutions } \hat{\mathbf{X}} \text{ and } \hat{\tilde{\mathbf{D}}} \in \mathcal{D} \\ \text{minimizing } \|\mathbf{Y} - \tilde{\mathbf{D}}\mathbf{X}\|_F \\ \text{subj. to } \|\mathbf{x}^{(q)}\|_0 \leq K \, \forall q \in [Q].\end{gathered} \quad (10)$$

This approach will be called in the following *Linearly constrained dictionary learning (LCDL)* while its application to multiplicative distortion correction is referred to as MDC.

Related problems known from literature are:

Dictionary learning:

$$\begin{gathered}\text{Find solutions } \hat{\mathbf{X}} \text{ and } \hat{\tilde{\mathbf{D}}} \\ \text{minimizing } \|\mathbf{Y} - \tilde{\mathbf{D}}\mathbf{X}\|_F \\ \text{subj. to } \|\mathbf{x}^{(q)}\|_0 \leq K \, \forall q \in [Q].\end{gathered} \quad (11)$$

This problem is solved by the well-known K-SVD algorithm. The difference to our problem Eq. (1) is that there are no constraints on the dictionary to be learned. Further, our algorithm differs in significant parts from the K-SVD algorithm.

A further optimization problem should also be mentioned in this context:

Blind factorization:

For the model

$$\mathbf{y} = \mathbf{e} \odot \mathbf{x} + \mathbf{n} \quad (12)$$

with where '$\odot$' stands for the element-wise multiplication.

$$\begin{gathered}\text{Find solutions } \hat{\mathbf{e}} \text{ and } \hat{\mathbf{x}} \\ \text{minimizing } \|\mathbf{y} - \mathbf{e} \odot \mathbf{x}\|_2 \\ \text{subj. to } \mathbf{e} \in E, \mathbf{x} \in X.\end{gathered} \quad (13)$$

E and X here are linear subspaces. This problem is the kernel of *blind deconvolution* [5], since the circular convolution in this paper is expressed in the Fourier domain as element-wise multiplication. The largest difference to our approach consists in making no use of sparsity of $\mathbf{x}$.

In section (II) we will introduce an algorithm for linearly constrained dictionary learning LCDL. The application to multiplicative distortion compensation MDC is treated in section (III).

The correction of motion errors for synthetic aperture radar (SAR) and inverse SAR (ISAR) is an old problem known as autofocus (AF). For a simple signal model for the corrupted SAR raw data - separability of range and azimuth processing, only lateral motion errors - well established AF techniques like the phase gradient algorithm (PGA) [3] or the eigenvector method (EM) [4] are available. These methods assume that there exist 'prominent scatterers' in the scene. Translated to the language of compressive sensing, this is comparable to the sparsity (or compressibility) of the SAR data. The AF problem can be directly formulated according to Eq. (1), where $\mathbf{e}$ is the vector of phasors $e_m = \exp\{j\phi_m\}, m \in M$ corrupting the measurements. This application is addressed in section (IV).

II. Linearly constrained dictionary learning

We now return to the LCDL problem and state the signal model more precisely:

A. Signal model

Let a set of training measurements $\mathbf{y}^{(q)} \in \mathbb{C}^M, q = 1, \ldots, Q$, collected in the $M \times Q$ matrix $\mathbf{Y} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(Q)})$ be given following

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{N} \quad (14)$$

where $\mathbf{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)})$ is the $N \times Q$ matrix of the unknown coefficient column vectors $\mathbf{x}^{(q)} \in \mathbb{C}^N, q = 1, \ldots, Q$, $\mathbf{D} \in \mathcal{D}$ is an unknown dictionary, constrained to be an element of a set of admissible dictionaries $\mathcal{D}$, and $\mathbf{N}$ is the measurement noise. Further it is assumed, that each $\mathbf{x}^{(q)}$ is K-sparse, i.e. $\|\mathbf{x}^{(q)}\|_0 \leq K \quad \forall q \in [Q] = \{1, \ldots, Q\}$. We denote by $\mathbf{\Omega}_K \subseteq \mathbb{C}^{N \times Q}$ the set of all $\mathbf{X}$ fulfilling the uniform K-sparsity.

B. The optimization task

We regard the following optimization task:

Based on the signal model Eq. (14) estimate $\mathbf{D}$ and $\mathbf{X}$ as a solution of the optimization task

$$(\hat{\mathbf{X}}, \hat{\mathbf{D}}) = \operatorname{argmin} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F \quad (15)$$
$$\text{subj. to } \mathbf{D} \in \mathcal{D}, \mathbf{X} \in \mathbf{\Omega}_K.$$

If the set of admissible dictionaries is maximal: $\mathcal{D} = \mathbb{C}^{M \times N}$, the well-known K-SVD algorithm

[1] (see appendix) solves this optimization task under certain conditions. Let $\mathbf{D}^{[l]}, l \in [L]$ be a *"dictionary generation system"* with $M \times N$ matrix elements $\mathbf{D}^{[l]}$, and $\mathcal{D}$ the linear span generated by this system with additional restrictions:

$$\mathcal{D} = \{\sum_{l=1}^{L} \beta_l \mathbf{D}^{[l]} : \boldsymbol{\beta} = (\beta_l, \ldots, \beta_L)^T \in \mathcal{B}\}$$

with

$$\mathcal{B} = \{\boldsymbol{\beta} \in \mathbb{C}^L : \|\boldsymbol{\beta}\|_2 = 1, \arg(\sum_l \beta_l) = 0\}.$$

The additional restrictions are included to avoid scaling ambiguities between $\mathbf{X}$ and $\boldsymbol{\beta}$. Then the general task Eq. (15) now can be re-formulated as

$$(\hat{\mathbf{X}}, \hat{\boldsymbol{\beta}}) = \text{argmin}\|\mathbf{Y} - \sum_{l=1}^{L} \beta_l \mathbf{D}^{[l]}\mathbf{X}\|_F \quad \text{subj. to } \boldsymbol{\beta} \in \mathcal{B}, \mathbf{X} \in \boldsymbol{\Omega}_K. \quad (16)$$

Of course, a unique solution can only exist, if the generating elements $\mathbf{D}^{[l]}, l \in [L]$ are linearly independent.

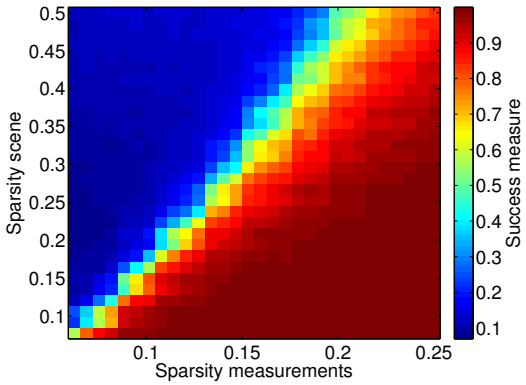

Fig. 1. *Phase transition diagram for LCDL. Measurement sparsity here is defined as the ratio between the number of measurements MQ and the number $NQ + (L-1)$ of freely selectable coefficients in $\mathbf{X}$ and $\boldsymbol{\beta}$, scene sparsity as the number KQ of non-zero coefficients divided by the number of coefficients NQ. As 'success measure' the mean value of the number of correct recoveries of the support indices relative to the total number KQ. All elements of the quantities $\mathbf{D}^{[l]}$, $\boldsymbol{\beta}$ and $\mathbf{X}_{\mathcal{K}}$ are chosen as i.i.d. complex Gaussian(0,1) random variables, and the support $\mathcal{K}$ is chosen as K-element index set by random for each $q \in [Q]$. The following quantities were fixed: $N = 50$, $L = 5$, $Q = 10$ while M and K were varied. The number of Monte Carlo simulations for each choice was 100.*

C. Proposed algorithm

K-SVD follows an approach alternating between finding a sparse representation $\hat{\mathbf{X}}$ and an update of the dictionary $\hat{\mathbf{D}}$ for step-wise decreasing the error term in Eq. (11).

Our algorithm will follow a similar way but with substantial differences. Let the tensor $\mathcal{T} \in \mathbb{C}^{M \times L \times N}$ be defined by

$$\mathcal{T}(., l, .) := \mathbf{D}^{[l]}, l \in [L]. \quad (17)$$

$\mathcal{T}$ comprises the dictionary generation system.

Algorithm 1: Linearly constrained dictionary learning (LCDL)

Input:
- Dictionary generation system $\mathcal{T}$
- Training vectors $\mathbf{Y}$
- Uniform sparsity constraint K
- Maximum number of iterations it_{max}

Initialization:
$it = 0$
(A) Perform modified sparse representation of order K for each $\mathbf{y}^{(q)}$
(B) Calculate common initial estimation $\hat{\boldsymbol{\beta}}$

Iteration: repeat
$it := it + 1$
(C) Update $\hat{\mathbf{D}}$
(D) Perform orthogonal matching pursuit (OMP) for each q
(E) Based on the new $\hat{\mathbf{X}}$ update $\hat{\boldsymbol{\beta}}$
until $it \geq it_{max}$.

Now the single steps will be explained:

1) (A) Modified sparse representation of order K: For the model $\mathbf{y} = \mathbf{Dx} + \mathbf{n}$, $\|\mathbf{x}\|_0 \leq K$, this step is performed similar to the matching pursuit [6] by a greedy algorithm:

$\mathbf{r}^{(0)} = \mathbf{y}, \mathcal{K} = \emptyset$
For $\kappa := 1 \ldots K$

$$\begin{aligned} \hat{n} &= \underset{1 \leq n \leq N}{\text{argmax}} \|\mathbf{P}_n \mathbf{r}^{(\kappa-1)}\|_2 \quad (18) \\ \text{with } \mathbf{P}_n &:= \mathbf{D}_n(\mathbf{D}_n^H \mathbf{D}_n)^{-1}\mathbf{D}_n^H \\ \text{and } \mathbf{D}_n &:= \mathcal{T}(., ., n) \\ \text{Collect } \hat{\mathbf{w}}^{(\kappa)} &:= (\mathbf{D}_n^H \mathbf{D}_n)^{-1}\mathbf{D}_n^H \mathbf{r}^{(\kappa-1)} \\ \mathcal{K} &= \mathcal{K} \cup \{\hat{n}\} \\ \mathbf{r}^{(\kappa)} &= (\mathbf{I} - \mathbf{P}_{\hat{n}})\mathbf{r}^{(\kappa-1)} \end{aligned}$$

end κ.

At the end of the iteration, $\mathcal{K}$ is the estimated support of $\mathbf{x}$. Note that $\hat{\mathbf{w}}^{(\kappa)}$ is the result of

$$\hat{\mathbf{w}}^{(\kappa)} = \underset{\mathbf{w} \in \mathbb{C}^L}{\text{argmin}} \|\mathbf{r}^{(\kappa-1)} - \mathbf{D}_{\hat{n}}\mathbf{w}\|_2 \quad (19)$$

and $\mathbf{r}^{(\kappa)}$ is the remainder after the κth iteration.

2) (B) Common initial estimation $\hat{\boldsymbol{\beta}}$: After the step (A) we have collected QK versions of $\hat{\mathbf{w}}^{(\kappa,q)}$ which are arranged as a $L \times KQ$ dimensional matrix $\mathbf{W}$. We expect that each column of $\mathbf{W}$ is close to the vector $\boldsymbol{\beta}$, multiplied with the amplitude $x_n^{(q)}$ with the index n determined by the corresponding support element, i.e. $\mathbf{W}$ should be close to a dyade

$$\mathbf{W} = \boldsymbol{\beta}\boldsymbol{\xi} \quad (20)$$

with a $1 \times KQ$-vector $\boldsymbol{\xi}$ containing the complex amplitudes. To estimate $\boldsymbol{\beta}$, we perform a singular-value-decomposition:

$$[\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H] = svd(\mathbf{W}). \tag{21}$$

If the largest singular value σ_1 is placed at index 1, $\hat{\boldsymbol{\beta}}$ is taken as a scalar multiple c of the first column of $\mathbf{U}$, where c is determined by the normalization assumptions $\|\boldsymbol{\beta}\|_2 = 1, \arg(\sum_l \beta_l) = 0$.

3) (C) Update of $\hat{\mathbf{D}}$*:* The estimate of the dictionary is updated by

$$\hat{\mathbf{D}} = \sum_{l=1}^{L} \hat{\beta}_l \mathbf{D}^{[l]} \tag{22}$$

4) (D) OMP: The orthogonal matching pursuit of order K (OMP, [7]) is performed for each q. The result is a new sparse estimation $\hat{\mathbf{X}}$.

5) (E) Update of $\hat{\boldsymbol{\beta}}$*:* For each q, l calculate $\mathbf{Z}^{(q)}(:, l) = D^{[l]}\hat{\mathbf{x}}^{(q)}$. Then

$$\hat{\boldsymbol{\beta}} = \left(\sum_q \mathbf{Z}^{(q)H}\mathbf{Z}^{(q)}\right)^{-1} \left(\sum_q \mathbf{Z}^{(q)H}\mathbf{y}^{(q)}\right) \tag{23}$$

Afterwards, normalize $\hat{\boldsymbol{\beta}}$ with a scalar factor to fulfill $\|\boldsymbol{\beta}\|_2 = 1, \arg(\sum_l \beta_l) = 0$.

D. Discussion and simulation results

Unfortunately, we can not give any recovery guarantees in dependence on the choice of N, M, Q, L and the properties of the matrices. Just counting the number of unknowns and measurements, we get the necessary condition

$$MQ \geq KQ + L - 1. \tag{24}$$

Good results can be expected if $MQ \gg KQ + L - 1$, if the dictionaries $\mathbf{D}^{[l]}, l \in [L]$ are linearly independent and in a certain manner incoherent with respect to the spiky base.

A Monte Carlo simulations shows the success rate for some parameter variations, see Fig. 1

III. Multiplicative distortion compensation

A. Solution with LCDL

As explicated in Eq. (1) till Eq. (9), the multiplicative distortion problem can be traced back to LCDL, when the distortion $\mathbf{e}$ can be expressed as a linear combination of basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_L$. The algorithm Alg. 1 can directly be applied to the data, using $\mathbf{D}^{[l]} := \mathrm{diag}(\mathbf{b}_l)\mathbf{D}$ as a dictionary generation system. The estimation $\hat{\boldsymbol{\beta}}$ is used for the recovery of $\mathbf{e}$:

$$\hat{\mathbf{e}} = \sum_{l=1}^{L} \hat{\beta}_l \mathbf{b}_l \tag{25}$$

and the distortion may be corrected by

$$\mathbf{Y}_{cor} = \mathrm{diag}\left(\frac{1}{\hat{e}_1}, \ldots, \frac{1}{\hat{e}_M}\right)\mathbf{Y}. \tag{26}$$

B. Solution with blind factorization

In principle, also blind factorization could be used for the estimation of $\mathbf{e}$. Eq. (1) can be transformed to

$$\tilde{\mathbf{y}} = \mathrm{diag}(\tilde{\mathbf{e}})\tilde{\mathbf{D}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{27}$$

$$\text{with } \tilde{\mathbf{y}} := \begin{pmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(Q)} \end{pmatrix}, \tilde{\mathbf{x}} := \begin{pmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(Q)} \end{pmatrix},$$

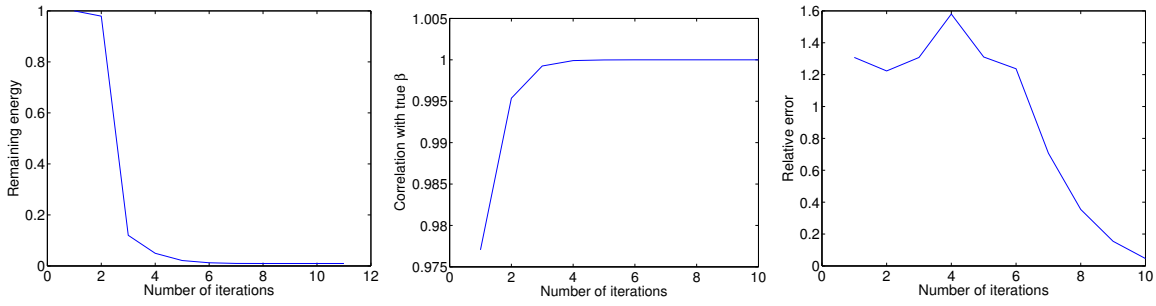

Fig. 2. Simulation result for the Multiplicative distortion compensation. The central dictionary is composed of independent complex Gaussian random variables with expectation 0 and variance 1. $M = 100$, $N = 150$; The basis for the dictionary generation system is an $M \times L$ partial Fourier matrix for the lowest frequencies, $L = 7$, and the linear coefficients β_l are also i.i.d. complex Gaussian random variables. The number of probes is $Q = 20$ and the sparsity is $K = 20$. The plot on the left shows the history of relative remaining energy over the iterations, the center plot the history of the correlation between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$, and the right plot the history of $\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2 / \|\boldsymbol{\beta}\|_2$.

$\tilde{\mathbf{D}} = \mathbf{I}_Q \otimes \mathbf{D}$ and $\tilde{\mathbf{e}} = \begin{pmatrix} \mathbf{e} \\ \vdots \\ \mathbf{e} \end{pmatrix} = \mathbf{1}_Q \otimes \mathbf{e}$ where $\mathbf{1}_Q$ is the column vector with Q coefficients equal to one.

If $\mathbf{B}$ is a basis for $\mathbf{e}$, then $\tilde{\mathbf{B}} := \mathbf{1}_Q \otimes \mathbf{B}$ is a basis for $\tilde{\mathbf{e}}$.

Suppose, $\tilde{\mathbf{e}} = \sum_{l=1}^{L} \beta_l \tilde{\mathbf{b}}_l$ and $\tilde{\mathbf{x}} = \sum_{p=1}^{P} \gamma_p \mathbf{g}_p$ where $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_P)$ is a basis for the subspace X introduced in Eq. (13).

Then Eq. (27) can be written as a function of products $\beta_l \gamma_p$:

$$\tilde{y}_\nu \quad = \quad \sum_{l,\mu,p} \tilde{b}_{\nu l} \tilde{D}_{\nu\mu} g_{\mu p} \, (\beta_l \gamma_p) + \tilde{n}_\nu. \quad (28)$$

$\beta_l \gamma_p$ is the l,pth coefficient of the rank 1 matrix $\mathbf{M} = \boldsymbol{\beta}\boldsymbol{\gamma}^T$, where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the column vectors built from $\beta_1, \ldots, \beta_L$ and $\gamma_1, \ldots, \gamma_P$.

So it is possible to write $\tilde{\mathbf{y}}$ as a linear function $\mathcal{A}$ of the matrix variable $\mathbf{M}$:

$$\tilde{\mathbf{y}} = \mathcal{A}(\mathbf{M}) + \tilde{\mathbf{n}} \quad (29)$$

This kind of transformation is known in the community as *'lifting'*.

The algorithm ADMiRA (Atomic Decomposition for Minimum Rank Approximation) [2] solves the following optimization task:

Based on the signal model Eq. (29) estimate $\mathbf{M}$ as a solution of the optimization task

$$\hat{\mathbf{M}} \quad = \quad \text{argmin} \|\tilde{\mathbf{y}} - \mathcal{A}(\mathbf{M})\|_2 \quad (30)$$

subj, to $\text{rank}(\mathbf{M}) \leq r$.

In our case, ADMiRA may be used to find the rank-1 matrix $\hat{\mathbf{M}}$ with best fit Eq. (30). $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ can easily be extracted from $\hat{\mathbf{M}}$ except for a factor which can be determined by the normalization requirement for $\boldsymbol{\beta}$. All other quantities then can be deduced from $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$.

This approach to the solution of Eq. (4) needs a specification of the subspace X. The sparsity of $\mathbf{X}$, or $\tilde{\mathbf{x}}$, respectively, can be considered, if the support is known. Then as basis vectors $\mathbf{g}_p, p \in [P]$ the vectors with a one at the position of a support element, and zero, elsewhere, can be defined. The support is estimated with OMP or ℓ_1-minimization. We propose the following algorithm:

Algorithm 2: Blind factorization for distortion estimation under sparsity conditions
Input:
- Measurements $\tilde{\mathbf{y}}$
- Dictionary $\tilde{\mathbf{D}}$
- Error basis $\tilde{\mathbf{B}}$
- Initial guess for $\tilde{\mathbf{e}}$
- Maximum number of iterations it_{max}

Iteration: repeat
  $it := it + 1$
- Estimate support of $\tilde{\mathbf{x}}$ with any pursuit
- Build the elementary basis $\mathbf{G}$ for the support elements
- Fill the tensor representing $\mathcal{A}$ according to Eq. (28)
- Apply ADMiRA with $r = 1$
- Extract $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ from $\hat{\mathbf{M}}$
- Evaluate $\hat{\tilde{\mathbf{e}}}$ from $\hat{\boldsymbol{\beta}}$

until $it \geq it_{max}$.

Simulation results for this approach can be found in the next section.

IV. Application to SAR AF

SAR autofocusing with compressive sensing methods has been addressed in several papers, as [8], [9], [10], [11], [10], [12], [13]. The proposed approaches are different from that of this paper, so we continue with CDLS and blind factorization.

A. Model for SAR data

We regard range-compressed SAR data in the variables r (range) and t (slow time) under the assumption that the range curvature can be neglected. A single point scatterer at range r_0 and azimuth-time t_0 induces the approximate signal

$$s(t, r) = \delta(r - r0) c(t - t0, r) \quad (31)$$

with the azimuth chirp

$$c(t, r) = D(t, r) exp\{-j2k_0 \sqrt{r^2 + (vt)^2}\} \quad (32)$$

where k_0 is the center wave number, v is the platform velocity and $D(t, r)$ describes the influence of the two-way antenna characteristics. In a small strip around $r = r_0$ the dependence on r may be neglected:

$$c(t) \approx D(t, r_0) exp\{-j2k_0 \sqrt{r_0^2 + (vt)^2}\}. \quad (33)$$

After discretization a sensing matrix $\mathbf{A}$ can be defined:

$$a_{mn} = c(m\Delta T - n\Delta t), m \in [M], n \in [N] \quad (34)$$

where Δt defines the spacing of azimuth grid points in the image and ΔT the measurement grid spacing. In Fig. 3 a sensing matrix used for simulation is illustrated. For a fixed range bin the measurements now can be modeled as

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}. \quad (35)$$

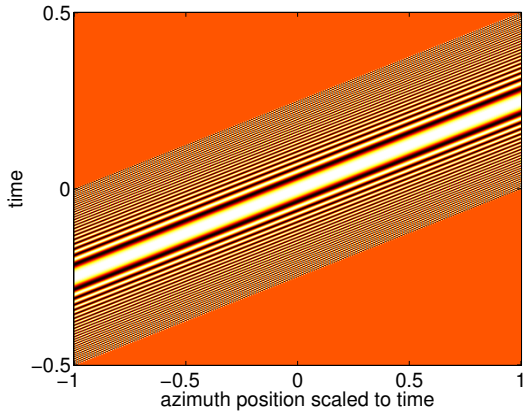

Fig. 3. *Real part of the sensing matrix.* $M = 400, N = 800$. *Length of chirp:* $N/2$. *Coherence of sensing matrix: 0.9.*

Because of unknown cross range motions of the carrier, phase distortions $\varphi_m, m = 1, \dots M$ are introduced leading to a multiplicative error term

$$\mathbf{e} = (\exp\{\varphi_1\}, \dots, \exp\{\varphi_M\})^t \tag{36}$$

and the distorted measurements now are

$$\mathbf{y} = \mathrm{diag}(\mathbf{e})\mathbf{A}\mathbf{x} + \mathbf{n}. \tag{37}$$

Similar as for the phase gradient algorithm, Q range cells are selected according to certain criteria as the *minimum variance principle* looking for range cells with minimum variance of the amplitude along azimuth which prefers dominant point-like scatterers. In contrary to the phase gradient algorithm we are not reliant to these isolated scatterers whereas we only demand sparsity.

Now the measurements are modelled in the form of Eq. (1)

$$\mathbf{Y} = \mathrm{diag}(\mathbf{e})\mathbf{A}\mathbf{X} + \mathbf{N}. \tag{38}$$

Again, $\mathbf{e}$ is represented in a basis $\mathbf{B}$ and the quantities $\mathbf{D}^{[l]} = \mathrm{diag}(\mathbf{b}_l)\mathbf{A}$ build the dictionary generating system.

For our simulation we generated phases according to a random walk:

$\omega = 0, \varphi(1) = 0.$
for $m = 2 \dots M$
  $\omega := (1 - \mu)\omega + \lambda w(m).$
  $\varphi(m) = \varphi(m) + \omega.$
$w(m), m \in [M]$ i.i.d. $\mathcal{N}(0, 1)$ rd variables.
$\varphi := \varphi - \tilde{\varphi}$, where
  $\tilde{\varphi}$ is the best linear fit of φ.

A simulated phase random walk and the spectrum of the related phasors are depicted in Fig. 4.

B. Choice of the basis

Due to the limited dynamics of the cross range velocity it suggests itself to use a part of the Fourier matrix related to the small frequencies

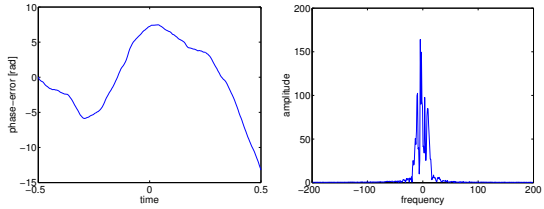

Fig. 4. *Left: example for a simulated random walk for phase error* $M = 400$, $\lambda = 0.02$, $\mu = 0.005$. *Right: spectrum of* $\mathbf{e}$

('low pass basis'); However, there is a problem which directly arises from the physics: a Doppler shift causes because of the chirp-nature of the azimuth signal an azimuth displacement without de-focusing. In our terms, if $\mathbf{B}$ denotes the low pass basis and $\mathbf{b}_\nu$ and $\mathbf{b}_\mu$ are two columns of $\mathbf{B}$ for each ξ not to close at the margins of $[1, N]$ a ζ can be found with

$$\mathrm{diag}(\mathbf{b}_\nu)\mathbf{a}^{(\xi)} \approx \mathrm{diag}(\mathbf{b}_\mu)\mathbf{a}^{(\zeta)} \tag{39}$$

in other words, the representation $\tilde{\mathbf{D}} = \sum_{l=1}^{L} \beta_l \mathrm{diag}(\mathbf{b}_\nu)\mathbf{A}$ is ambiguous in $\boldsymbol{\beta}$.

As a consequence, the function Eq. (18) is rather flat at its maximum, since in the vicinity of the maximum a shift in n can be compensated by another linear combination of basis elements.

We have chosen another way to a representative basis. According to the above phase generation scheme many realizations $\mathbf{e}_i, i = 1 : I$ were simulated and from the empirical covariance matrix $\mathbf{R} = \frac{1}{I}\sum_{i=1} I\mathbf{e}_i\mathbf{e}_i^H$ the eigenvectors to the L largest eigenvalues were taken as a basis. Fig. 5 shows a new generated error phase and its representation due to the so developed basis.

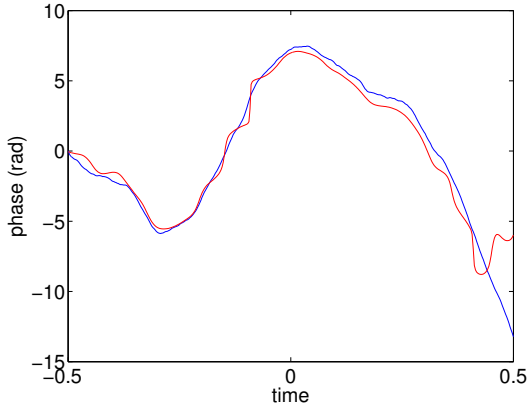

Fig. 5. *The error phase (blue) and its representation due to the basis* $\mathbf{B}$ *generated by random (red).* $M = 400$, $L = 19$

C. Simulation of LCDL applied to SAR AF

Now we introduce some simulation results for the linearly constrained dictionary learning. Fig. 6

shows the conventionally compressed azimuth signal (i.e. $\mathbf{w} = \mathbf{A}^H\mathbf{y}$ with and without multiplicative distortion.

For the application of LCDL of course the additional information that the amplitudes of the coefficients of $\mathbf{e}$ are always equal to one, should be used by normalizing the coefficients of $\hat{\mathbf{e}}$ to one at each iteration step.

In Fig. 7 all simulated training lines after conventional compression are shown, left without error correction, right after phase correction via LCDL.

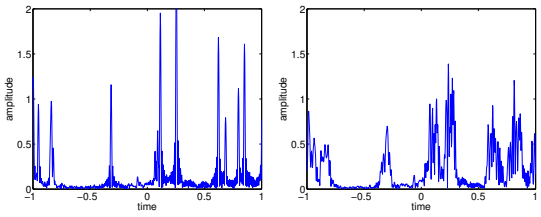

Fig. 6. *Left: example for a simulated conventionally compressed azimuth signal without error, $K = 15$. Right: the same with the above shown phase error*

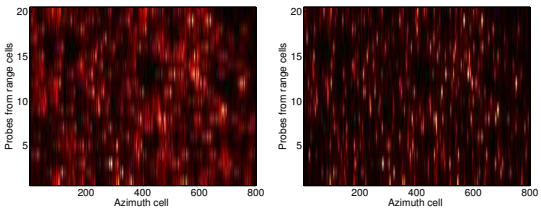

Fig. 7. *All training azimuth lines after conventional compression. Left: Without error correction. Right: After phase compensation via LCDL*

D. Reduction of complexity by matrix transformation

As indicated in refEqeq:030, the LCDL method can be applied also if the measurements are linearly transformed, i.e. pre-processed:

$$\mathbf{Z} = \mathbf{GY} \tag{40}$$

A possible pre-processing may consist of imaging by the matched filter: $\mathbf{G} = \mathbf{A}^H$. Because of the band-structure of $\mathbf{A}$ the response to a scatterer will be concentrated to an interval around its position, even if a phase error is present. The probes then could be truncated around the maxima leading to a computational mitigation.

E. SAR AF with blind factorization

Also the algorithm 2 (blind factorization) was tested for SAR AF. Representative results are illustrated in Fig. 8 and Fig. 9.

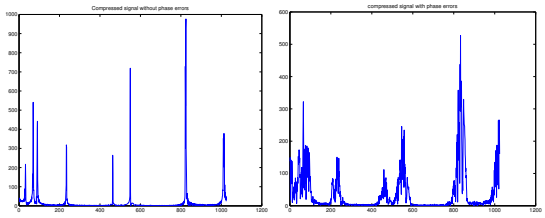

Fig. 8. *Left: Compressed signal without phase error. Right: The same with phase error*

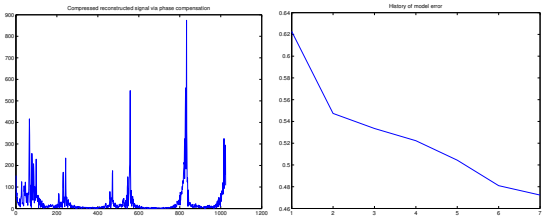

Fig. 9. *Left: After phase correction with blind factorization. Right: Decrease of the reconstruction error during the iterations*

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, "rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[2] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4402–4416, Sept 2010.

[3] D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, "Phase gradient autofocus-a robust tool for high resolution sar phase correction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 827–835, 1994.

[4] J. Charles V. Jakowatz and D. E. Wahl, "Eigenvector method for maximum-likelihood estimation of phase errors in synthetic-aperture-radar imagery," *J. Opt. Soc. Am. A*, vol. 10, no. 12, pp. 2539–2546, Jul 1993.

[5] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, March 2014.

[6] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44 vol.1.

[8] L. Xia, D. Zhu, and R. Jiang, "A research on sar autofocusing algorithms for compressive sensing technique," in *Radar Conference 2013, IET International*, 2013, pp. 1–6.

[9] Z. Zhe, Z. Yao, J. Chenglong, Z. Bingchen, H. Wen, and W. Yirong, "Autofocus of sparse microwave imaging radar based on phase recovery," in *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on*, 2013, pp. 1–5.

[10] T. Jihua, S. Jinping, H. Xiao, and Z. Bingchen, "Motion compensation for compressive sensing sar imaging with autofocus," in *2011 6th IEEE Conference on Industrial Electronics and Applications*, June 2011, pp. 1564–1567.

[11] M. Cetin, I. Stojanovic, O. Onhon, K. Varshney, S. Samadi, W. C. Karl, and A. S. Willsky, "Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 27–40, July 2014.

[12] S. I. Kelly, M. Yaghoobi, and M. E. Davies, "Auto-focus for under-sampled synthetic aperture radar," in *Sensor Signal Processing for Defence (SSPD 2012)*, Sept 2012, pp. 1–5.

[13] S. Ugur and O. Arikan, "Sar image reconstruction and autofocus by compressed sensing," *Digit. Signal Process.*, vol. 22, no. 6, pp. 923–932, Dec. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.dsp.2012.07.011