A general iterative imputation scheme with feedback control for tensor completion (IFCTC)

José Henrique DE MORAIS GOULART *,1,2, Gérard FAVIER[2],

[1]Univ. Grenoble Alpes, CNRS, Gipsa-Lab, F-38000 Grenoble

[2]Laboratoire I3S, CNRS, Univ. Côte d'Azur, 06900 Sophia Antipolis

jose-henrique.de-morais-goulart@gipsa-lab.fr, favier@i3s.unice.fr

Résumé – Les tenseurs et les décompositions tensorielles constituent des outils mathématiques très utiles pour représenter et analyser des données multidimensionnelles. Le problème de l'estimation de données manquantes dans un tenseur de mesures joue un rôle important dans de nombreuses applications. Dans cet article, nous proposons un schéma d'imputation général itératif incluant un mécanisme de rétroaction du premier ordre, avec l'objectif d'améliorer la performance de l'algorithme. Deux cas particuliers de ce schéma, faisant intervenir des opérateurs de seuillage doux et dur basés sur le modèle de Tucker, sont discutés. Puis, des résultats de simulations sont présentés pour illustrer leur performance.

Abstract – Tensors and tensor decompositions are very useful mathematical tools for representing and analyzing multidimensional data. The problem of estimating missing data in a tensor of measurements, named tensor completion, plays an important role in numerous applications. In this paper, to solve this problem, we propose a general iterative imputation scheme including a first-order feedback mechanism, aiming to improve algorithm performance. Two particularizations of this scheme, in which we apply soft and hard thresholding operators based on the Tucker model, are discussed. Then, simulation results are presented to illustrate their performance.

1 Introduction

The tensor completion (TC) problem constitutes an extension of matrix completion (MC) to higher-order tensors. During the last decade, this problem has received a growing attention from various scientific communities. Indeed, estimating missing data of a partially known tensor is of crucial importance in numerous applications like, for instance, biomedical signal processing, hyperspectral imaging, computer vision and graphics, road and network traffic analysis, etc. The missing data may result from sensor failures, from a compressive sensing scheme acquiring (or transmitting) only some tensor entries, or from the elimination of identified outliers. Promoting a low-rank solution is a standard and powerful approach for solving the MC and TC problems.

In the case of tensors, multiple notions of rank exist, the best known ones being the tensor and the multilinear rank (mrank). Also, several algorithms have been proposed for low-rank TC, such as those in [1, 5, 6, 9, 10, 11], most of them dealing with the mrank. In particular, [5, 6] have developed algorithms based on an iterative imputation scheme which, at each iteration, fills in the missing entries according to the current model estimate, and then re-estimates the model.

*During his PhD, J. H. de M. Goulart's work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, under the program Ciência sem Fronteiras. He is now supported by the European Research Council under the European Program FP7/2007-2013, Grant AdG-2013-320594 "DECODA."

In this paper, we present a general version of such an iterative imputation scheme including a first-order feedback control mechanism, aiming to improve algorithm performance. Two particularizations of this scheme, in which we apply soft and hard thresholding operators based on the Tucker model, are discussed. Then, simulation results with synthetic data are presented to illustrate their performance.

Notation : Scalars, column vectors, matrices, and tensors of order higher than two, are denoted by lower-case, bold lower-case, bold upper-case, and upper-case calligraphic letters, *e.g.*, $a, \mathbf{a}, \mathbf{A}, \boldsymbol{\mathcal{A}}$, respectively. The flat mode-p unfolding of a tensor $\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{T}} \triangleq \mathbb{R}^{N_1 \times \cdots \times N_P}$ is denoted by $\mathbf{X}_{\langle p \rangle} \in \mathbb{R}^{N_p \times \bar{N}_p}$, with $\bar{N}_p = N_1 \cdots N_{p-1} N_{p+1} \cdots N_P$. The mode-$p$ product of $\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{T}}$ with a matrix $\mathbf{A} \in \mathbb{R}^{I_p \times N_p}$ is denoted by $\boldsymbol{\mathcal{X}} \times_p \mathbf{A}$.

2 Problem statement

Consider a data tensor $\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{T}}$ and let Ω denote the set of tuples $(n_1, \ldots, n_P)$ such that $x_{n_1,\ldots,n_P}$ is observed. Define $\boldsymbol{\mathcal{T}}_\Omega \triangleq \{\boldsymbol{\mathcal{Y}} \in \boldsymbol{\mathcal{T}} : y_{n_1,\ldots,n_P} \neq 0 \text{ only if } (n_1, \ldots, n_P) \in \Omega\}$. Then, we can write $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_\Omega + \boldsymbol{\mathcal{X}}_{\overline{\Omega}}$, where $\boldsymbol{\mathcal{X}}_\Omega$ stands for the orthogonal projection of $\boldsymbol{\mathcal{X}}$ onto $\boldsymbol{\mathcal{T}}_\Omega$ and likewise for the complement of Ω, denoted by $\overline{\Omega}$. The TC problem generally consists in estimating $\boldsymbol{\mathcal{X}}_{\overline{\Omega}}$ from $\boldsymbol{\mathcal{X}}_\Omega$. Several approaches exist, depending on (1) the underlying tensor model (*e.g.*, canonical polyadic or Tucker decompositions) ; (2) the parsimony assumption (*e.g.*, low tensor rank or low mrank) ; (3) the minimized criterion,

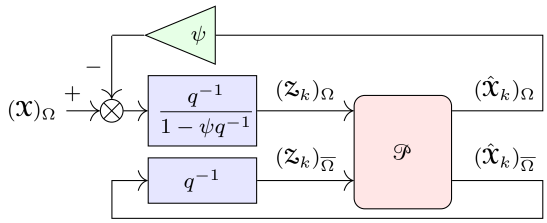

FIGURE 1 – Feedback control mechanism of IFCTC.

which often comprises a least-squares (LS) data fidelity term and a parsimony-inducing regularization term ; and (4) the optimization algorithm.

Here, we consider the general formulation

$$\min_{\hat{\mathcal{X}} \in \boldsymbol{\mathcal{T}}} \|\mathcal{X}_\Omega - \hat{\mathcal{X}}_\Omega\|_F^2 + \varphi(\hat{\mathcal{X}}), \tag{1}$$

where $\varphi(\cdot)$ is a parsimony-inducing penalization functional. In particular, $\varphi(\cdot)$ can be the indicator function of a set $\mathcal{S}$ of parsimonious tensors, such as low-rank or low-mrank tensors.

3 General imputation scheme

An iterative single imputation algorithm proceeds by performing the following steps at each iteration k :

1. construction of a corrected estimated model $\mathcal{Z}_k$ by imposing $\mathcal{Z}_k = \mathcal{X}_\Omega + (\hat{\mathcal{X}}_{k-1})_{\overline{\Omega}}$, where $\hat{\mathcal{X}}_{k-1}$ denotes the parsimonious tensor model estimated at iteration $k-1$;
2. computation of a new estimate $\hat{\mathcal{X}}_k = \mathscr{P}(\mathcal{Z}_k)$, where $\mathscr{P}$ is a parsimony-inducing operator depending on the choice of $\varphi(\cdot)$.

In [5], the above scheme is employed with $\varphi(\cdot)$ being the indicator function of the set of tensors having rank bounded by a given number R. The operator $\mathscr{P}$ then performs a single iteration of the alternating least-squares (ALS) algorithm in an attempt of approximating $\mathcal{Z}_k$ by a rank-R tensor. An interpretation of the resulting algorithm as an instance of the expectation minimization (EM) method is discussed in [5]. A similar approach is followed in [6], where $\mathcal{S}$ is the set of tensors with bounded mrank $\mathcal{L}_{\mathbf{r}} = \{\mathcal{Y} \in \boldsymbol{\mathcal{T}} : \operatorname{rank}(\mathbf{Y}_{\langle p \rangle}) \leq R_p\}$, for some target modal ranks $R_1, \dots, R_P$. So, $\mathscr{P}$ is in this case a hard thresholding operator which computes an approximate projection onto $\mathcal{L}_{\mathbf{r}}$ by truncating the higher-order singular value decomposition (HOSVD) of its argument [4].

Though the above described method is often effective, it may exhibit slow convergence or yield inaccurate estimates. Hence, we propose to improve it by changing step 1 so that $\mathcal{X}_\Omega$ is replaced by a modified update of the form

$$\mathcal{Z}_k = \mathcal{X}_\Omega + \psi[(\mathcal{Z}_{k-1})_\Omega - (\hat{\mathcal{X}}_{k-1})_\Omega] + (\hat{\mathcal{X}}_{k-1})_{\overline{\Omega}},$$

with $0 \leq \psi < 1$. By solving this recurrence relation, one can see that ψ acts as a *forgetting factor* of past modeling errors $(\mathcal{X})_\Omega - (\hat{\mathcal{X}}_{k-1})_\Omega$. Viewing the operator $\mathscr{P}$ applied on step 2 as a system with $\bar{N}$ inputs and $\bar{N}$ outputs, this modified update can be seen as the application of the first-order control rule

$$(\mathcal{Z}_k)_\Omega = \tfrac{1}{1-\psi q^{-1}} \left(\mathcal{X}_\Omega - \psi(\hat{\mathcal{X}}_{k-1})_\Omega\right) \tag{2}$$

Algorithm 1 Imputation scheme with feedback control for TC.

Inputs : Observed entries $\mathcal{X}_\Omega$ and parameters ψ, ϵ, K
Outputs : Estimate $\hat{\mathcal{X}}$ of $\mathcal{X}$

$k \leftarrow k+1$
repeat
  $\mathcal{Z}_k = \mathcal{X}_\Omega + \psi[(\mathcal{Z}_{k-1})_\Omega - (\hat{\mathcal{X}}_{k-1})_\Omega] + (\hat{\mathcal{X}}_{k-1})_{\overline{\Omega}}$
  $\hat{\mathcal{X}}_k = \mathscr{P}(\mathcal{Z}_k)$
until $\|\hat{\mathcal{X}}_k - \hat{\mathcal{X}}_{k-1}\|_F \, \|\hat{\mathcal{X}}_{k-1}\|_F^{-1} < \epsilon$ or $k = K$
  return $\hat{\mathcal{X}} = \mathcal{X}_\Omega + (\hat{\mathcal{X}}_k)_{\overline{\Omega}}$

to manipulate its inputs associated with the indices in Ω, with the goal of driving its outputs associated with indices in $\overline{\Omega}$ (*i.e.*, the missing data) more quickly to their true values (see Fig. 1).

The resulting scheme will be called imputation scheme with feedback control for TC (IFCTC). An explicit algorithm is given in Algorithm 1. As stopping criteria, we establish a maximum number of iterations K and check whether the inequality $\|\hat{\mathcal{X}}_k - \hat{\mathcal{X}}_{k-1}\|_F \, \|\hat{\mathcal{X}}_{k-1}\|_F^{-1} < \epsilon$ holds for a prescribed $\epsilon > 0$.

4 Tucker-based IFCTC algorithms

In this section, we present two particular cases of the IFCTC scheme based on the Tucker model. The first one seeks a solution of low mrank, by searching for it in $\mathcal{L}_{\mathbf{r}}$, for some target mrank $\mathbf{r} = (R_1, \cdots, R_P)$ chosen *a priori*. Its operator $\mathscr{P}$ truncates the HOSVD of its argument, thus corresponding to a hard thresholding. Therefore, this scheme, which we call imputation scheme with feedback and HOSVD hard thresholding (IFHHT), amounts to introducing the feedback control mechanism in the algorithm proposed by [6].

In practice, such a hard-thresholding-based scheme may deliver unsatisfying results when applied to real-world data tensors. The reason is that, in practice, these tensors are mostly often characterized by modal unfoldings exhibiting a somewhat rapid decay of its singular values. If this decay is not sufficiently fast, then one needs to choose an mrank $\mathbf{r}$ having relatively high components, which implies the need of many measurements (*i.e.*, observed entries).

Hence, the second particularization of IFCTC that we present is rather aimed at exploiting the *compressibility* of the singular spectra of the tensor unfoldings, as measured by the sum of the corresponding singular values, *i.e.*, the nuclear norms of these unfoldings. In [7], it is shown that

$$\max_{p \in \langle P \rangle} \left\|\mathbf{X}_{\langle p \rangle}\right\|_* \leq \|\operatorname{vec}(\mathcal{S})\|_1 \leq \min_{p \in \langle P \rangle} \sqrt{\bar{N}_p} \, \left\|\mathbf{X}_{\langle p \rangle}\right\|_*, \tag{3}$$

where $\mathcal{S}$ denotes the core of the HOSVD of $\mathcal{X}$. These inequalities imply that the compressibility of the modal singular spectra of a tensor is connected to compressibility of its HOSVD core, in the ℓ_1-sense. Conceivably, other Tucker models of $\mathcal{X}$ could have even more compressible cores in that sense.

This property motivates choosing $\varphi(\hat{\mathcal{X}}) = \tau\| \operatorname{vec}(\mathcal{S})\|_1$, where $\mathcal{S}$ is the core of a Tucker model of $\hat{\mathcal{X}}$. A solution thus requires minimizing (1) with respect to $\mathcal{S}$ and to the factors of a Tucker model. Here, we consider instead a heuristic which seeks an approximate solution to this problem. Namely, we solve at step 2 of the imputation scheme the regularized problem

$$\min_{\mathcal{G}} \left\| \mathcal{Z}_k - \mathcal{G} \bigtimes_{p=1}^{P} \mathbf{U}_k^{(p)} \right\|_F^2 + \tau_k \| \operatorname{vec}(\mathcal{G}) \|_1, \tag{4}$$

where the Tucker factors $\mathbf{U}_k^{(p)}$ are those of the HOSVD of $\mathcal{Z}_k$. So, we aim to find an approximate Tucker model of $\mathcal{Z}_k$ having an even more compressible core than the HOSVD, by controlling the loss in modeling accuracy through the choice of τ_k. By computing the HOSVD $\mathcal{Z}_k = \mathcal{S}_k \bigtimes_{p=1}^{P} \mathbf{U}_k^{(p)}$, it is easy to show that the minimizer of (4) can be obtained via

$$\min_{\mathcal{G} \in \mathcal{T}} \|\mathcal{S}_k - \mathcal{G}\|_F^2 + \tau_k \|\operatorname{vec}(\mathcal{G})\|_1 , \tag{5}$$

since the matrices $\mathbf{U}_k^{(p)}$ are orthogonal. The above problem is solved by the l_1 proximity operator

$$\left[\operatorname{pr}_{\tau_k}^{\ell_1}(\mathcal{S}_k)\right]_{n_1,\dots,n_P} = \left(|[\mathcal{S}_k]_{n_1,\dots,n_P}| - \tau_k\right)_+ \operatorname{sign}([\mathcal{S}_k]_{n_1,\dots,n_P}),$$

where $(x)_+ \triangleq \max\{0, x\}$. Therefore, the new estimate is given by $\hat{\mathcal{X}}_k = \mathcal{P}(\mathcal{Z}_k) = \operatorname{pr}_{\tau_k}^{\ell_1}(\mathcal{S}_k) \bigtimes_{p=1}^{P} \mathbf{U}_k^{(p)}$. This algorithm is called imputation scheme with feedback and HOSVD soft thresholding (IFHST).

Seeking a compromise between convergence speed and estimation accuracy, we use a varying penalty parameter given by $\tau_k = \gamma \tau_{k-1}$, with $0 < \gamma < 1$. For a sufficiently large τ_0, this allows a fast progress at early iterations, without losing precision due to a "severe thresholding" near convergence. In practice, this leads to an interesting behavior, namely :

— the first iterates have very sparse cores, because τ_k is sufficiently large so that nearly all core components are zeroed by the soft thresholding operator for small k ;
— since τ_k decays exponentially, the core sparsity is gradually reduced, *i.e.*, more and more nonzero core entries are added to the model.

Thus, this yields a sequence of increasingly complex models. We have recently introduced the above described IFHST algorithm in [8].

5 Simulation results

We first highlight the improvement brought by the feedback mechanism in IFHHT. To this end, we generate several realizations of an mrank-$(4,4,4)$ $20 \times 20 \times 20$ tensor as $\mathcal{X}_0 = \mathcal{G} \bigtimes_{p=1}^{P} \mathbf{U}^{(p)}$, where $\mathcal{G} \in \mathbb{R}^{4\times4\times4}$ has zero-mean i.i.d. Gaussian entries and unit Frobenius norm and $\mathbf{U}^{(p)} \in \mathbb{R}^{20\times4}$ has orthonormal columns. Each generated tensor is completed from 15% randomly sampled entries. Before sampling, we add a rescaled Gaussian noise term $\mathcal{N}$ (also of unit Frobenius norm), yielding $\mathcal{X} = \mathcal{X}_0 + 10^{-5}\,\mathcal{N}$. The IFHHT algorithm is then applied for multiple values of ψ in $(0, 0.99)$. Note that when $\psi = 0$, IFHHT is equivalent to the algorithm of [6]. Fig. 2 shows the average normalized squared error NSE$=\|\mathcal{X} - \hat{\mathcal{X}}\|_F / \|\mathcal{X}\|_F$ achieved after k iterations, for 20 realizations. For $\psi = 0.99$,

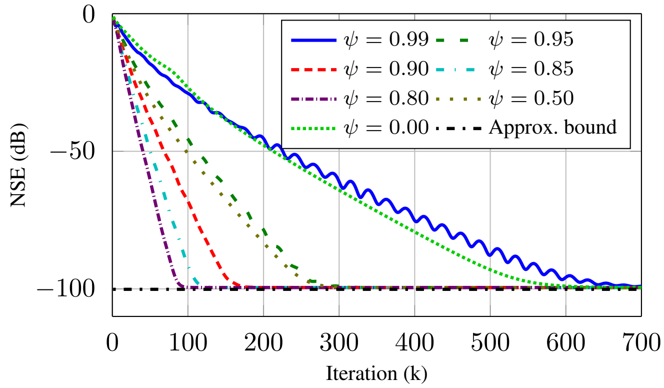

FIGURE 2 – IFHHT : acceleration of convergence due to first-order feedback control mechanism with time constant ψ.

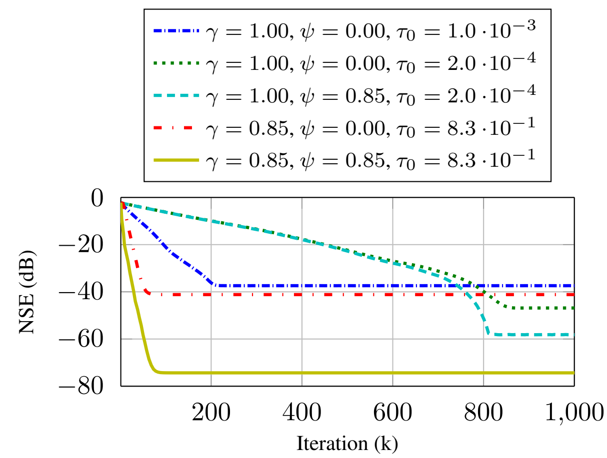

FIGURE 3 – IFHST : performance improvement brought by exponentially decaying threshold τ_k with rate γ and by first-order feedback control rule with time constant ψ.

there is no gain with respect to $\psi = 0$. Also, the pronounced oscillations seen for this choice happen due to the "aggressive" tuning of the first-order control mechanism with a pole near the unit circle. However, by reducing ψ, convergence can be remarkably accelerated. In particular, with $\psi = 0.8$ IFHHT attains convergence at least five times faster than the algorithm with no feedback correction.

A similar evaluation is performed for IFHST. Here, we have generated 20 random tensors having fast decaying modal spectra and then applied IFHST for several values of ψ, γ and τ_0. Following [9], the target tensor is generated from the model $\mathcal{X} = \mathcal{G} \bigtimes_{p=1}^{3} \left(\mathbf{Q}_p \operatorname{Diag}(1, 2^{-\theta}, \dots, 20^{-\theta})\right)$, where $\mathcal{G} \in \mathbb{R}^{20\times20\times20}$ has standard normal entries, $\mathbf{Q}_p$ is a random orthogonal matrix and θ controls the rate of decay of the modal singular values, which we set as $\theta = 3$. The average NSE per iteration is shown in Fig. 3. In the first two curves, there is no feedback nor varying threshold ; the choice of τ_0 thus offers a compromise between convergence speed and accuracy. The third curve has the same convergence speed as the second one, but attains smaller error due to $\psi = 0.85$. Conversely, the

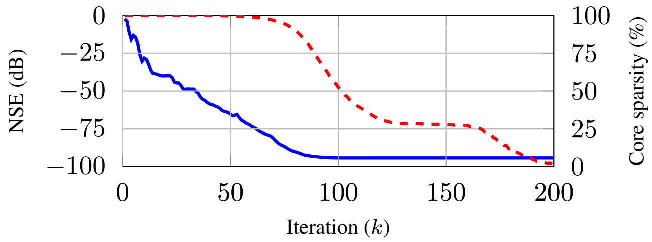

FIGURE 4 – Evolution of NSE (—) and sparsity of the softly thresholded HOSVD core (– –) along a run of the IFHST algorithm with $\gamma = \psi = 0.85$ and $\tau_0 = 1/1.2$.

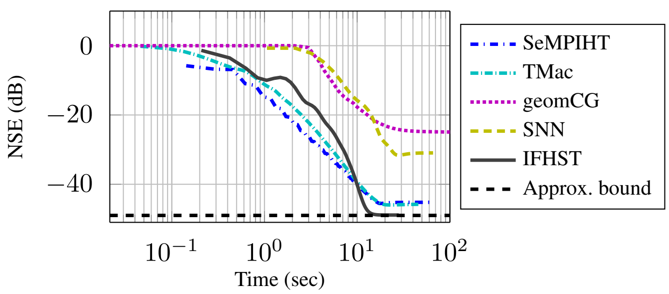

FIGURE 5 – Evolution of NSE for several TC algorithms when applied to recover $100 \times 100 \times 100$ tensors having fast decaying modal spectra ($\theta = 3/2$), with 15% sampled entries.

fourth curve converges faster than the first two due to the use of $\gamma = 0.85$, but there is no improvement with respect to the final error because $\psi = 0$. Finally, the last curve shows a remarkable improvement with respect to both aspects, due to the use of feedback and a varying penalty parameter.

The typical evolution of the sparsity of the thresholded Tucker core along the iterations of IFHST is illustrated by Fig. 4, which was generated with $\gamma = \psi = 0.85$ and $\tau_0 = 1/1.2$. It also shows that a highly sparse core can accurately model $\mathcal{X}$: at $k = 60$, the sparsity level is 98.7% and NSE = -73.5 dB.

In Fig. 5, the IFHST algorithm is compared with the algorithms SNN [1], TMac [9], geomCG [10] and SeMPIHT [11]. We generate $100 \times 100 \times 100$ tensors having fast decaying modal spectra, with decay parameter $\theta = 3/2$, and sample 15% of their entries. Both TMac and geomCG were run with mrank-increasing strategies starting from $\mathbf{r}_0 = (1, 1, 1)$. The target mrank of SeMPIHT, TMac and geomCG was $\mathbf{r} = (40, 40, 40)$, and unit mrank increments were applied in all three algorithms. The tolerance controlling mrank increase was set as 0.01 for TMac and geomCG. In SeMPIHT, mrank components were incremented every two iterations. IFHST was run with $\psi = \gamma = 0.85$ and $\tau_0 = 1/12$. From this figure, we can conclude that the proposed IFHST algorithm outperforms the other algorithms with regard to convergence speed and to estimation accuracy.

6 Conclusion

We have proposed a general iterative imputation scheme with feedback correction for addressing the tensor completion problem, describing two concrete particularizations based on the Tucker model, IFHHT and IFHST. Our simulation results show the improvement brought by incorporating the feedback correction mechanism. Furthermore, they evidenced the advantage of employing IFHST in comparison with other existing solutions in a scenario where the generated tensors have rapidly decaying modal spectra, which often happens for real-world data.

Future developments include studying other ways of exploiting Tucker core compressibility, and also designing new IFCTC algorithms based on hierarchical tensor models. Finally, though we observe that IFHHT and IFHST generally provide a reasonable approximation of $\mathcal{X}$ (provided enough entries are observed), being competitive with other TC algorithms, proving their convergence remains an open problem.

References

[1] R. Tomioka, K. Hayashi, H. Kashima. *Estimation of low-rank tensors via convex optimization*. arXiv :1010.0789v2, 2011.

[2] S. Gandy, B. Recht, I. Yamada. *Tensor completion and low-n-rank tensor recovery via convex optimization.* Inv. Prob., 27(2), 025010, 2011.

[3] J. Liu, P. Musialski, P. Wonka, J. Ye. *Tensor completion for estimating missing values in visual data*. IEEE Trans. Pattern Anal. Mach. Intell., 35(1), 208-220, 2013.

[4] L. De Lathauwer, B. De Moor, J. Vandewalle. *A multilinear singular value decomposition*. SIAM J. Matrix Anal. Appl., 21(4), 1253-1278, 2000.

[5] G. Tomasi, R. Bro. *PARAFAC and missing values*. Chemometrics and Intel. Lab. Syst., 75(2), 163-180, 2005.

[6] N. Kreimer, M. D. Sacchi. *A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation.* Geophysics, 77(3), V113-V122, 2012.

[7] J. H. de M. Goulart. *Estimation of structured tensor models and recovery of low-rank tensors*. Ph. D. thesis, Université de Nice Sophia-Antipolis, France, 2015.

[8] J. H. de M. Goulart, A. Y. Kibangou, G. Favier. *Traffic data imputation via tensor completion based on soft thresholding of Tucker core*. Transportation Research Part C : Emerging Technologies, 2017 (submitted).

[9] Y. Xu, R. Rao, W. Yin, Z. Su. *Parallel matrix factorization for low-rank tensor completion*. Inverse Problems & Imaging, 9(2), 601-624, 2015.

[10] D. Kressner, M. Steinlechner, B. Vandereycken. *Low-rank tensor completion by Riemaniann optimization.* BIT Numer. Math., 54, 447-468, 2014.

[11] J. H. de M. Goulart, G. Favier. *Low-rank tensor recovery using sequentially optimal modal projections in iterative hard thresholding (SeMPIHT)*. SIAM J. Sci. Comput., 39(3), A860-A889, 2017.