NONPARAMETRIC LOW-RANK TENSOR IMPUTATION

Juan Andrés Bazerque, Gonzalo Mateos, and Georgios B. Giannakis

Dept. of ECE, Univ. of Minnesota, Minneapolis, MN 55455, USA

ABSTRACT

Completion or imputation of three-way data arrays with missing entries is a basic problem encountered in various areas, including bioinformatics, image processing, and preference analysis. If available, prior information about the data at hand should be incorporated to enhance performance of the imputation method adopted. This is the motivation behind the proposed low-rank tensor estimator which leverages the correlation across slices of the data cube in the form of reproducing kernels. The rank of the tensor estimate is controlled by a novel regularization on the factors of its PARAFAC decomposition. Such a regularization is inspired by a reformulation of the nuclear norm for matrices, which allows to bypass the challenge that rank and singular values of tensors are unrelated quantities. The proposed technique is tested on MRI data of the brain with 30% missing data, resulting in a recovery error of $-17dB$.

Index Terms— Tensor, low-rank, missing data, kernel methods.

1. INTRODUCTION

Recovering a low-rank matrix from a subset of its entries is a fundamental task that arises with localization of wireless sensors, Internet traffic as well as gene microarray data analysis, and preference modeling for recommender systems. A natural approach to the *low-rank matrix completion* problem is to minimize the rank of the target matrix, subject to a constraint on the error in fitting the observed entries. Since rank minimization is generally NP-hard, the nuclear norm has been advocated recently as a *convex* surrogate to the rank [5]. Beyond computational tractability, nuclear-norm minimization enjoys good performance both in theory as well as in practice [3, 4].

The focus here is on imputation of missing data from tensors (also known as multi-way arrays), which are high-order generalizations of matrices frequently encountered in chemometrics, medical imaging, and networking [7]. Leveraging the low-rank structure for tensor completion is challenging, since even computing the tensor rank is an NP-hard problem itself. Defining a nuclear norm surrogate is not obvious either, since singular values as defined by the Tucker decomposition are unrelated with the notion of rank. Traditional approaches to finding low-dimensional representations of tensors include unfolding the multi-way data and applying matrix factorizations such as the SVD, or, employing the parallel factor (PARAFAC) decomposition [12, 7]. In the context of tensor completion, an approach falling under the first category can be found in [6], while the PARAFAC decomposition was advocated in [2].

This paper introduces a kernel-based tensor completion method that is amenable to incorporating a priori known similarities (correlation) across slices. It builds on a novel tensor estimator presented in Section 3, inspired by the redefinition of the matrix nuclear norm in terms of its decomposition into low-rank factors.

Work was supported by an AFOSR MURI grant no. FA 9550-10-1-0567.

2. PRELIMINARIES

2.1. Nuclear-norm minimization for matrix completion

Low-rank approximation is a popular method for estimating missing values of a matrix $\mathbf{Z} \in \mathbb{R}^{N\times M}$, which capitalizes on "regularities" across the data [5]. For the imputation to be feasible, a binding assumption that relates the available entries with the missing ones is required. An alternative is to postulate that $\mathbf{Z}$ has low rank $R \ll \min(N, M)$. The problem of finding matrix $\hat{\mathbf{Z}}$ with rank not exceeding R, which approximates $\mathbf{Z}$ in the given entries specified by a binary matrix $\mathbf{W} \in \{0,1\}^{N\times M}$, can be formulated as ($\odot$ denotes Hadamard product, and subscript F the Frobenious norm)

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{X}} ||(\mathbf{Z}-\mathbf{X})\odot\mathbf{W}||_F^2 \quad \text{s. to } \operatorname{rank}(\mathbf{X}) \leq R\,. \tag{1}$$

The low-rank matrix $\mathbf{Z}$ implies that the vector $\mathbf{s}(\mathbf{Z})$ of its singular values is sparse. Hence, the rank constraint is equivalent to $\|\mathbf{s}(\mathbf{Z})\|_0 \leq R$, where the ℓ_0-(pseudo)norm $\|\cdot\|_0$ equals the number of nonzero entries of its vector argument.

Aiming to a convex relaxation of the NP-hard problem (1), one can leverage recent advances in compressive sampling [4, 5] and substitute the ℓ_0-norm with its ℓ_1-norm surrogate, which here equals the nuclear norm of $\mathbf{X}$ defined as $\|\mathbf{X}\|_* := \|\mathbf{s}(\mathbf{X})\|_1$; that is, the sum of its singular values. With this relaxation, the Lagrangian counterpart of (1) is

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{X}} ||(\mathbf{Z}-\mathbf{X})\odot\mathbf{W}||_F^2 + \mu\|\mathbf{X}\|_* \tag{2}$$

were $\mu \geq 0$ is rank-controlling parameter. Several iterative algorithms have been proposed to solve (2), and are effective in tackling low- to medium-size matrix completion problems; see e.g., [4, 10]. However, most algorithms require computation of singular values per iteration and become prohibitively expensive when dealing with high-dimensional data. To solve (2) efficiently, consider the following characterization of the nuclear norm [10]

$$\|\mathbf{X}\|_* = \min_{\{\mathbf{B},\mathbf{C}\}} \frac{1}{2}(\|\mathbf{C}\|_F^2 + \|\mathbf{B}\|_F^2) \quad \text{s. to } \mathbf{X} = \mathbf{C}\mathbf{B}^T. \tag{3}$$

For an arbitrary matrix $\mathbf{X}$ with singular value decomposition (SVD) $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the minimum in (3) is attained for $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}^{1/2}$ and $\mathbf{B} = \mathbf{V}\mathbf{\Sigma}^{1/2}$. The optimization in (3) is over all possible bilinear factorizations of $\mathbf{X}$, so that the number of columns of $\mathbf{B}$ and $\mathbf{C}$ is also a variable. Building on (3), one can arrive at the following equivalent reformulation of (2) [10]

$$\hat{\mathbf{Z}} = \arg\min_{\{\mathbf{X},\mathbf{B},\mathbf{C}\}} \|(\mathbf{Z}-\mathbf{X})\odot\mathbf{W}\|_F^2 + \frac{\mu}{2}(\|\mathbf{C}\|_F^2 + \|\mathbf{B}\|_F^2) \quad \text{s. to } \mathbf{X} = \mathbf{C}\mathbf{B}^T. \tag{4}$$

The equivalence implies that by finding the global minimum of (4), which could entail considerably less variables than (2), one can recover the optimal solution of (2). However, since (4) is *nonconvex*, it

978-1-4673-0183-1/12/$31.00 ©2012 IEEE

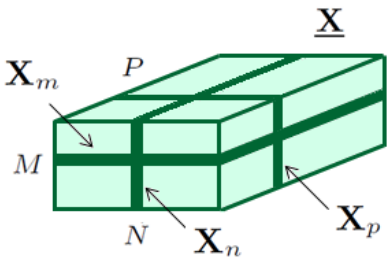

Fig. 1. Tensor slices along the row, column, and fiber dimensions.

may have stationary points which need not be globally optimum. Interestingly, the next proposition asserts that under mild assumptions every stationary point of (4) is globally optimum for (2) [9].

Proposition 1 *Let* $\bar{\mathbf{X}} = \bar{\mathbf{C}}\bar{\mathbf{B}}^T$ *be a stationary point of* (4) *. If* $\|\mathbf{Z} - \bar{\mathbf{X}}\|_2 < \mu/2$*, then* $\hat{\mathbf{Z}} := \bar{\mathbf{X}}$ *is the globally optimal solution of* (2)*.*

The Frobenious-norm regularization for controlling the rank in (4), will be useful to obtain tensor counterparts in Section 3.

2.2. PARAFAC decomposition

The PARAFAC decomposition of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{M\times N\times P}$ is at the heart of the proposed imputation method, since it offers the means to define the rank of $\underline{\mathbf{X}}$ [12, 7]. For given $R \in \mathbb{N}$, consider matrices $\mathbf{A} \in \mathbb{R}^{N\times R}$, $\mathbf{B} \in \mathbb{R}^{M\times R}$, and $\mathbf{C} \in \mathbb{R}^{P\times R}$, such that

$$\underline{\mathbf{X}}(m,n,p) = \sum_{r=1}^{R} \mathbf{A}(m,r)\mathbf{B}(n,r)\mathbf{C}(p,r). \tag{5}$$

The rank of $\underline{\mathbf{X}}$ is defined as the minimum value of R for which this decomposition is possible. For $R^* := \text{rank}(\underline{\mathbf{X}})$, the PARAFAC decomposition is given by the corresponding matrices $\mathbf{A} \in \mathbb{R}^{N\times R^*}$, $\mathbf{B} \in \mathbb{R}^{M\times R^*}$, and $\mathbf{C} \in \mathbb{R}^{P\times R^*}$ so that (5) holds.

To appreciate why the aforementioned rank definition is natural, rewrite (5) as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \tag{6}$$

where $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ represent the r-th columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ respectively, and the outer products $\underline{\mathbf{P}}_\mathbf{r} := \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \in \mathbb{R}^{M\times N\times P}$ have entries $\underline{\mathbf{P}}_\mathbf{r}(m,n,p) := \mathbf{A}(m,r)\mathbf{B}(n,r)\mathbf{C}(p,r)$. The rank of a tensor is thus the minimum number of outer products (rank one factors) required to represent the tensor.

Let $\mathbf{X}_p,\ p = 1,\ldots,P$ denote the p-th slice of $\underline{\mathbf{X}}$ along its third (fiber) dimension, such that $\mathbf{X}_p(m,n) = \underline{\mathbf{X}}(m,n,p)$; see Fig. 1. The following compact form of the PARAFAC decomposition in terms of matrix factors will be used in the sequel

$$\mathbf{X}_p = \mathbf{A}\mathbf{Diag}\left[\boldsymbol{e}_p^T\mathbf{C}\right]\mathbf{B}^T, \quad p = 1,\ldots,P \tag{7}$$

where the diagonal matrix $\mathbf{D} = \mathbf{Diag}[\mathbf{u}]$ has the vector $\mathbf{u}$ on its diagonal, and $\boldsymbol{e}_p^T$ is the p-th row of the $P \times P$ identity matrix. The PARAFAC decomposition is symmetric [cf. (5)], and one can alternatively write $\mathbf{X}_m = \mathbf{B}\mathbf{Diag}\left[\mathbf{e}_m^T\mathbf{A}\right]\mathbf{C}^T$, or, $\mathbf{X}_n = \mathbf{C}\mathbf{Diag}\left[\mathbf{e}_n^T\mathbf{B}\right]\mathbf{A}^T$ in terms of slices along the first (row) and second (column) dimensions.

3. RANK REGULARIZATION FOR TENSORS

Generalizing the nuclear-norm regularization technique (2) from low-rank matrix to tensor completion is not straightforward, since singular values of a tensor (given by the Tucker decomposition) are not related to the rank [7]. The Frobenius-norm regularization outlined in Section 2.1 offers a viable option for tensor completion under the PARAFAC model however, by solving

$$\{\hat{\underline{\mathbf{Z}}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}\} := \arg \min_{\{\underline{\mathbf{X}},\mathbf{A},\mathbf{B},\mathbf{C}\}} \|(\underline{\mathbf{Z}} - \underline{\mathbf{X}}) \odot \underline{\mathbf{W}}\|_F^2 + \frac{\mu}{2}\left(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2\right) \tag{8}$$
$$\text{s. to } \mathbf{X}_p = \mathbf{A}\mathbf{Diag}\left[\boldsymbol{e}_p^T\mathbf{C}\right]\mathbf{B}^T,\ p = 1,\ldots,P$$

where the Frobenious norm of a tensor is defined as $\|\underline{\mathbf{X}}\|_F^2 := \sum_m^M \sum_n^N \sum_p^P \underline{\mathbf{X}}^2(m,n,p)$, and the Hadamard product as $(\underline{\mathbf{X}} \odot \underline{\mathbf{W}})(m,n,p) := \underline{\mathbf{X}}(m,n,p)\underline{\mathbf{W}}(m,n,p)$.

Different from the matrix case, it is in principle not clear whether the proposed regularization in (8) bears any relation with the tensor rank. Interestingly, analysis and simulations provided next corroborate the capability of (8) to produce a low-rank tensor $\hat{\underline{\mathbf{Z}}}$, for sufficiently large μ. The following proposition is in order (proofs are omitted due to space limitations).

Proposition 2 *Let* $\{\hat{\underline{\mathbf{Z}}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}\}$ *be the solution of* (8)*, and let* $\hat{\mathbf{a}}_r$*,* $\hat{\mathbf{b}}_r$*, and* $\hat{\mathbf{c}}_r$ *denote the* r*-th column of* $\hat{\mathbf{A}}$*,* $\hat{\mathbf{B}}$*, and* $\hat{\mathbf{C}}$*, respectively.*

a) If $\mu \geq \mu_0 := \|\underline{\mathbf{W}} \odot \underline{\mathbf{Z}}\|_F^{4/3}$*, then* $\hat{\underline{\mathbf{Z}}} = \mathbf{0}_{M\times N\times P}$*; and*

b) If for any $r \in \{1,\ldots,R\}$ *either* $\hat{\mathbf{a}}_r$*,* $\hat{\mathbf{b}}_r$*, or* $\hat{\mathbf{c}}_r$ *is null, then all three vectors* $\hat{\mathbf{a}}_r$*,* $\hat{\mathbf{b}}_r$*, and* $\hat{\mathbf{c}}_r$ *are null simultaneously.*

Proposition 2 a) asserts that if the penalty parameter is chosen large enough, the rank is reduced to the extreme case $\text{rank}(\hat{\underline{\mathbf{Z}}}) = 0$. To see why this is a non-trivial property, it is prudent to think of ridge-regression estimates where similar quadratic regularizers are adopted, but an analogous property does not hold. For instance, in ridge-regression one needs to let $\mu \to \infty$ in order to obtain an all-zero solution. The second part is relevant to the case $\mu < \mu_0$. Note first that low-rank factors $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$ are in general not sufficient to guarantee that the reconstructed tensor $\hat{\underline{\mathbf{Z}}}$ is low rank. The per-factor structure of columnwise linear dependencies must be shared across factors, to obtain a low-rank tensor. In particular, if a column of zeros appears in $\hat{\mathbf{A}}$, then the same column in $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ must be null in order to guarantee that the rank of $\hat{\underline{\mathbf{Z}}}$ is reduced by one. These simultaneously vanishing columns in the factor matrices are guaranteed by Proposition 2 b).

But Proposition 2 does not establish that some of these columns will indeed vanish, a desirable property that is demonstrated through simulations. Fig. 2 depicts the estimated rank of $\hat{\underline{\mathbf{Z}}}$ for different values of μ, when $\underline{\mathbf{Z}} \in \mathbb{R}^{4\times 5\times 6}$ is a tensor of rank three contaminated with additive Gaussian noise at $SNR = 20$dB. It is apparent from Fig. 2 that μ controls the rank of $\hat{\underline{\mathbf{Z}}}$. In addition, the existence of a finite $\mu = \mu_0$ such that $\text{rank}(\hat{\underline{\mathbf{Z}}}) = 0$ corroborates Proposition 2 a).

A limitation of (8) is that it does not allow for incorporating side information that could be available in addition to the given entries $\underline{\mathbf{W}} \odot \underline{\mathbf{Z}}$. In the context of recommender systems, a description of the users and/or products through attributes (e.g., gender, age) or measures of similarity, is typically available. It is thus meaningful to exploit both known preferences and descriptions to model the preferences of users [1]. In three-way (samples, genes, conditions) microarray data analysis, the relative position of single-nucleotide

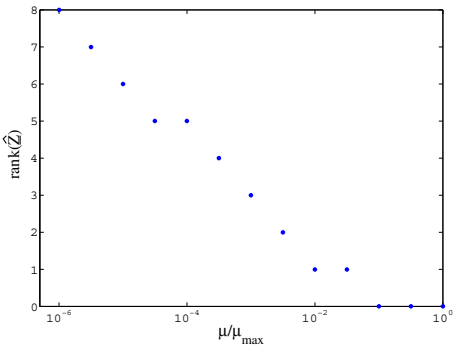

Fig. 2. Estimated rank vs. normalized μ.

polymorphisms in the DNA molecule implies degrees of correlation among genotypes [11]. These correlations could be available either through a prescribed model, or, through estimates obtained using a reference tensor $\underline{\mathbf{Z}}$. A kernel-based approach to tensor completion capable of incorporating such extra information is the next subject.

4. NONPARAMETRIC TENSOR APPROXIMATION

To incorporate a priori known similarities across its slices, it is instructive to look at a tensor $f : \mathcal{X}, \mathcal{Y}, \mathcal{T} \to \mathbb{R}$ as a function of three variables x, y, and t, living in measurable spaces $\mathcal{X}, \mathcal{Y}$, and $\mathcal{T}$, respectively. Identifying x, y, and t with the row, column and fiber indexes m, n, and p used so far, renders function $f(m, n, p)$ recognizable as a three-way array, but with x, y, and t expanded to incorporate the informative attributes. In the microarray data example, index n may be expanded to $y_n = (\text{gene id}, \text{gene position})$.

Generalizing (8) to this nonparametric framework, low-rank functions f are formally defined to belong to the following family

$$\mathcal{F}_R := \{f : \mathcal{X}\times\mathcal{Y}\times\mathcal{T}\to\mathbb{R} : \; f(x,y,t) = \sum_{r=1}^{R} a_r(x)b_r(y)c_r(t) \\ \text{such that } a_r(x) \in \mathcal{H}_{\mathcal{X}},\; b_r(y) \in \mathcal{H}_{\mathcal{Y}},\; c_r(t) \in \mathcal{H}_{\mathcal{T}}\}$$

where $\mathcal{H}_{\mathcal{X}}$, $\mathcal{H}_{\mathcal{Y}}$, and $\mathcal{H}_{\mathcal{T}}$ are Hilbert spaces constructed from specified kernels $k_{\mathcal{X}}$, $k_{\mathcal{Y}}$ and $k_{\mathcal{T}}$, in $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{T}$, respectively, while $R \geq R^*$ is an initial overestimate of the rank of f. The following nonparametric fitting criterion is adopted

$$\hat{f}_R := \arg\min_{f\in\mathcal{F}_R} \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{p=1}^{P} w_{mnp}(z_{mnp} - f(x_m, y_n, t_p))^2 \\ + \frac{\mu}{2}\sum_{r=1}^{R}\left(\|a_r\|^2_{\mathcal{H}_{\mathcal{X}}} + \|b_r\|^2_{\mathcal{H}_{\mathcal{Y}}} + \|c_r\|^2_{\mathcal{H}_{\mathcal{T}}}\right) . \qquad (9)$$

Recursive application of Representer's Theorem yields finite-dimensional representations of a_r, b_r, and c_r given by

$$\hat{a}_r(x) = \textstyle\sum_{m=1}^{M}\alpha_{rm}k_{\mathcal{X}}(x_m, x)$$
$$\hat{b}_r(y) = \textstyle\sum_{n=1}^{N}\beta_{rn}k_{\mathcal{Y}}(y_n, y)$$
$$\hat{c}_r(t) = \textstyle\sum_{p=1}^{P}\gamma_{rp}k_{\mathcal{T}}(t_p, t).$$

Defining vectors $\boldsymbol{k}^T_{\mathcal{X}}(x) := [k_{\mathcal{X}}(x_1, x), \ldots, k_{\mathcal{X}}(x_M, x)]$, and correspondingly $\boldsymbol{k}^T_{\mathcal{Y}}(y) := [k_{\mathcal{Y}}(y_1, y), \ldots, k_{\mathcal{Y}}(y_N, y)]$, and $\boldsymbol{k}^T_{\mathcal{T}}(t) := [k_{\mathcal{T}}(t_1, t), \ldots, k_{\mathcal{T}}(t_P, t)]$, along with matrices $\mathbf{A} \in \mathbb{R}^{M\times R}$: $A(m, r) := \alpha_{mr}$, $\mathbf{B} \in \mathbb{R}^{N\times R}$: $B(n, r) := \beta_{nr}$, and $\mathbf{C} \in \mathbb{R}^{P\times R}$: $C(p, r) := \gamma_{pr}$, it follows that

$$\hat{f}_R(x, y, t) = \sum_{r=1}^{R}\hat{a}_r(x)\hat{b}_r(y)\hat{c}_r(t) \\ = \boldsymbol{k}^T_{\mathcal{X}}(x)\mathbf{A}\mathbf{D}\mathbf{i}\mathbf{a}\mathbf{g}\left[\boldsymbol{k}^T_{\mathcal{T}}(x)\mathbf{C}\right]\mathbf{B}^T\boldsymbol{k}_{\mathcal{Y}}(y). \qquad (10)$$

Matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are further obtained by solving

$$\min_{\mathbf{A},\mathbf{C},\mathbf{B}} \sum_{p=1}^{P}\left\|\left(\mathbf{Z}_p - \mathbf{K}_{\mathcal{X}}\mathbf{A}\mathbf{D}\mathbf{i}\mathbf{a}\mathbf{g}\left[e_p^T\mathbf{K}_{\mathcal{T}}\mathbf{C}\right]\mathbf{B}^T\mathbf{K}_{\mathcal{Y}}\right)\odot\mathbf{W}_p\right\|_F^2 \\ + \frac{\mu}{2}\left(\mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{K}_{\mathcal{X}}\mathbf{A}) + \mathrm{trace}(\mathbf{B}^{\mathrm{T}}\mathbf{K}_{\mathcal{Y}}\mathbf{B}) + \mathrm{trace}(\mathbf{C}^{\mathrm{T}}\mathbf{K}_{\mathcal{T}}\mathbf{C})\right) \\ \text{s. to } \mathbf{A} \in \mathbb{R}^{M\times R},\; \mathbf{B} \in \mathbb{R}^{N\times R},\; \mathbf{C} \in \mathbb{R}^{P\times R} \qquad (11)$$

with kernel matrices $\mathbf{K}_{\mathcal{X}} := [\boldsymbol{k}_{\mathcal{X}}(x_1), \ldots, \boldsymbol{k}_{\mathcal{X}}(x_M)] \in \mathbb{R}^{M\times M}$, $\mathbf{K}_{\mathcal{Y}} := [\boldsymbol{k}_{\mathcal{Y}}(y_1), \ldots, \boldsymbol{k}_{\mathcal{Y}}(y_N)] \in \mathbb{R}^{N\times N}$, and correspondingly $\mathbf{K}_{\mathcal{T}} := [\boldsymbol{k}_{\mathcal{T}}(t_1), \ldots, \boldsymbol{k}_{\mathcal{T}}(t_P)] \in \mathbb{R}^{P\times P}$.

Problem (11) reduces to (8) when the side information is discarded by selecting $k_{\mathcal{X}}$, $k_{\mathcal{Y}}$ and $k_{\mathcal{T}}$ as Kronecker deltas. In the general case, (11) yields the sought nonlinear low-rank approximation method for $f(x, y, t)$ when combined with (10).

4.1. Block coordinate descent algorithm

An iterative algorithm is proposed for solving (11), by cyclically minimizing over the columns of $\mathbf{A}$ $\mathbf{B}$ and $\mathbf{C}$. In the first R steps of the cycle the objective function in (11) is minimized with respect to (w.r.t.) each of the R columns of $\mathbf{A}$; that is, for $r = 1, \ldots, R$,

$$\hat{\mathbf{a}}_r := \arg\min_{\mathbf{a}_r\in\mathbb{R}^M} \sum_{p=1}^{P} Q_p(\boldsymbol{a}_r) + \frac{\mu}{2}\|\boldsymbol{a}_r\|_2^2 \qquad (12)$$

with

$$Q_p(\boldsymbol{a}_r) := \left\|\left(\mathbf{Z}_p - \mathbf{K}_{\mathcal{X}}\left(\mathbf{A}^{(-r)} + \boldsymbol{a}_r\mathbf{e}_r^T\right)\mathbf{D}_p\mathbf{B}^T\mathbf{K}_{\mathcal{Y}}\right)\odot\mathbf{W}_p\right\|_F^2$$

where $\mathbf{A}^{(-r)} := \mathbf{A} - \mathbf{A}\boldsymbol{e}_r\boldsymbol{e}_r{}^T$ is the residual matrix formed by inserting M zeros in the rth column of $\mathbf{A}$, and $\mathbf{D}_p := \mathbf{Diag}\left[\mathbf{e}_p^T\mathbf{K}_{\mathcal{T}}\mathbf{C}\right]$.

Setting the gradient of (12) w.r.t. $\boldsymbol{a}_r$ equal to zero, and rearranging terms results in the closed-form minimizer

$$\hat{\mathbf{a}}_r = \left(\sum_{p=1}^{P}\mathbf{D}_{rp}\mathbf{K}_{\mathcal{X}} + \mu\mathbf{I}\right)^{-1}\sum_{p=1}^{P}\mathbf{Z}_{rp}\mathbf{K}_{\mathcal{Y}}\mathbf{B}\mathbf{D}_p\boldsymbol{e}_r \qquad (13)$$

where matrices $\mathbf{D}_{rp}$ and $\mathbf{Z}_{rp}$ are defined as

$$\mathbf{D}_{rp} := \mathbf{Diag}\left[\mathbf{W}_p\left((\mathbf{K}_{\mathcal{Y}}\mathbf{B}\mathbf{D}_p\mathbf{e}_r)\odot(\mathbf{K}_{\mathcal{Y}}\mathbf{B}\mathbf{D}_p\mathbf{e}_r)\right)\right]$$
$$\mathbf{Z}_{rp} := \mathbf{W}_p\odot\left(\mathbf{Z}_p - \mathbf{K}_{\mathcal{X}}\mathbf{A}^{(-r)}\mathbf{D}_p\mathbf{B}^T\mathbf{K}_{\mathcal{Y}}\right)$$

Similarly, minimizing (11) w.r.t. the columns of $\mathbf{B}$ and $\mathbf{C}$ yields

$$\hat{\mathbf{b}}_r = \left(\sum_{m=1}^{M}\mathbf{D}_{rm}\mathbf{K}_{\mathcal{Y}} + \mu\mathbf{I}\right)^{-1}\sum_{m=1}^{M}\mathbf{Z}_{rm}\mathbf{K}_{\mathcal{T}}\mathbf{C}\mathbf{D}_m\boldsymbol{e}_r \qquad (14)$$

$$\hat{\mathbf{c}}_r = \left(\sum_{n=1}^{N}\mathbf{D}_{rn}\mathbf{K}_{\mathcal{T}} + \mu\mathbf{I}\right)^{-1}\sum_{n=1}^{N}\mathbf{Z}_{rn}\mathbf{K}_{\mathcal{X}}\mathbf{A}\mathbf{D}_n\boldsymbol{e}_r \qquad (15)$$

where

$$\mathbf{D}_{rm} := \mathbf{D}\Big[\mathbf{W}_m\left((\mathbf{K}_{\mathcal{T}}\mathbf{C}\mathbf{D}_m\mathbf{e}_r)\odot(\mathbf{K}_{\mathcal{T}}\mathbf{C}\mathbf{D}_m\mathbf{e}_r)\right)\Big]$$

$$\mathbf{Z}_{rm} := \mathbf{W}_m \odot \Big(\mathbf{Z}_m - \mathbf{K}_{\mathcal{Y}}\mathbf{B}^{(-r)}\mathbf{D}_m\mathbf{C}^T\mathbf{K}_{\mathcal{T}}\Big)$$

$$\mathbf{D}_{rn} := \mathbf{D}\Big[\mathbf{W}_n\left((\mathbf{K}_{\mathcal{X}}\mathbf{A}\mathbf{D}_n\mathbf{e}_r)\odot(\mathbf{K}_{\mathcal{X}}\mathbf{A}\mathbf{D}_n\mathbf{e}_r)\right)\Big]$$

$$\mathbf{Z}_{rn} := \mathbf{W}_n \odot \Big(\mathbf{Z}_n - \mathbf{K}_{\mathcal{T}}\mathbf{C}^{(-r)}\mathbf{D}_n\mathbf{A}^T\mathbf{K}_{\mathcal{X}}\Big)$$

with $\mathbf{D}_m := \mathbf{Diag}\left[\mathbf{e}_m^T\mathbf{K}_{\mathcal{Y}}\mathbf{A}\right]$, $\mathbf{D}_n := \mathbf{Diag}\left[\mathbf{e}_n^T\mathbf{K}_{\mathcal{Y}}\mathbf{B}\right]$, $\mathbf{B}^{(-r)} := \mathbf{B} - \mathbf{B}\boldsymbol{e}_r\boldsymbol{e}_r{}^T$, and $\mathbf{C}^{(-r)} := \mathbf{C} - \mathbf{C}\boldsymbol{e}_r\boldsymbol{e}_r{}^T$.

The limit points of the algorithm resulting by iteratively updating (13), (14), and (15), are stationary points of (11).

5. NUMERICAL TEST

Estimator (9) is tested against a corrupted version of the BRAINIX dataset in the OsiriX repository [8]. The tensor $\underline{\mathbf{Z}}$ to be estimated corresponds to a three-dimensional MRI scan of the brain comprising a set of $P = 50$ images of $M \times N = 64 \times 64$ pixels. Thirty percent of the data is removed uniformly at random together with the whole slice $\mathbf{Z}_n,\ n = 50$. Fig. 3 shows the results of applying estimator (9) to the remaining data, which yields a reconstruction error of -17dB. The original slice $\mathbf{Z}_p,\ p = 21$, together with its corrupted counterpart and the resulting estimate are shown on top and center left. Covariance matrices $\mathbf{K}_x$, $\mathbf{K}_y$ and $\mathbf{K}_z$ are assumed known, and the symmetry of the brain is reflected in the X-shaped pictorial view of $\mathbf{K}_y$ in Fig. 3 (center right), which shows relatively high correlation between slices $\mathbf{X}_n$ and $\mathbf{X}_{n'}$ at different sides of the head, as well as between contiguous slices. This correlation is the key enabler for the method to recover the missing slice (see Fig. 3 (bottom)) by interpolating its contiguous and symmetric counterparts.

All in all, the experiment evidences the merits of low-rank PARAFAC decomposition for modeling a tensor, the ability of the kernel estimator (9) in recovering missing data, and the usefulness of incorporating correlations as side information.

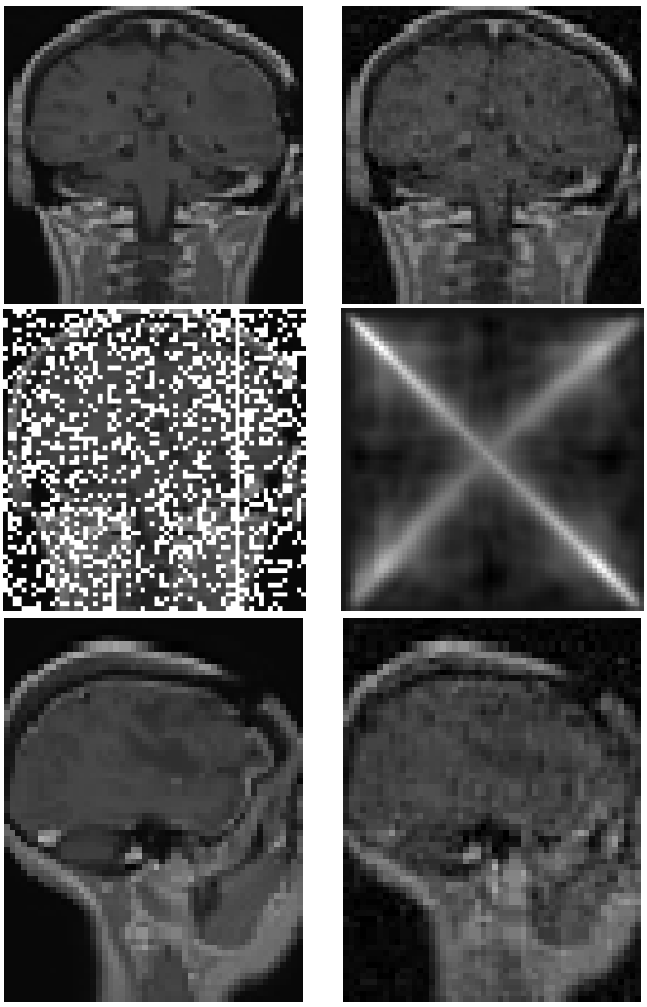

Fig. 3. Results of applying (9) to the BRAINIX dataset. (top) original and recovered fibers $\mathbf{Z}_p$ and $\hat{\mathbf{Z}}_p$ for $p = 21$. (center) input fiber $\mathbf{Z}_p,\ p = 21$ with missing data, and covariance matrix $\mathbf{K}_y$. (bottom) original and recovered columns $\mathbf{Z}_n$ and $\hat{\mathbf{Z}}_n$ for the position $n = 50$ in which the whole input slice is missing)

6. REFERENCES

[1] J. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert, "Low-rank matrix factorization with attributes," *Technical Report N24/06/MM,* Ecole des Mines de Paris, 2006; see also *J. of Machine Learning Res.,* vol. 10, pp. 803-826, Mar. 2009.

[2] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mrup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems,* vol. 106, no. 1, pp. 41-56, 2011.

[3] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, pp. 925–936, June 2010.

[4] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics,* vol. 9, pp 717-772, 2009.

[5] M. Fazel, "Matrix rank minimization with applications" *PhD Thesis,* Electrical Engineering Dept., Stanford University, vol. 54, pp. 1-130, 2002.

[6] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems,* vol. 27, no. 2, pp. 1-19, 2011.

[7] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455-500, 2009.

[8] OsiriX, "DICOM sample image sets repository," *http://pubimage.hcuge.ch:8080/.*

[9] M. Mardani, G. Mateos, and G. B. Giannakis, "In-network sparsity-regularized rank minimization: Algorithms and applications," *IEEE Trans. Signal Process.*, Mar. 2012 (submitted), see also arXiv:1203.1507v1 [cs.MA].

[10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,"*SIAM Review*, vol. 52, no. 3, pp. 471-501, 2010.

[11] P. Scheet, and M. Stephens, "A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase," *American Journal of Human Genetics*, vol. 78, pp. 629-644, 2006.

[12] J. M. F. ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP/PARAFAC," *Psychometrika,* vol. 67, no. 3, pp. 399-409, 2002.