

Iterative Estimation of Constrained Rank-One Matrices in Noise

Sundeep Rangan and Alyson K. Fletcher

Abstract—We consider the problem of estimating a rank-one matrix in Gaussian noise under a probabilistic model for the left and right factors of the matrix. The probabilistic model can impose constraints on the factors including sparsity and positivity that arise commonly in learning problems. We propose a simple iterative procedure that reduces the problem to a sequence of scalar estimation computations. The method is similar to approximate message passing techniques based on Gaussian approximations of loopy belief propagation that have been used recently in compressed sensing. Leveraging analysis methods by Bayati and Montanari, we show that the asymptotic behavior of the estimates from the proposed iterative procedure is described by a simple scalar equivalent model, where the distribution of the estimates is identical to certain scalar estimates of the variables in Gaussian noise. Moreover, the effective Gaussian noise level is described by a set of state evolution equations. The proposed method thus provides a computationally simple and general method for rank-one estimation problems with a precise analysis in certain high-dimensional settings.

I. INTRODUCTION

We consider the problem of estimating vectors $\mathbf{u}_0 \in \mathbb{R}^m$ and $\mathbf{v}_0 \in \mathbb{R}^n$ from a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ of the form

$$\mathbf{A} = \mathbf{u}_0 \mathbf{v}_0^T + \sqrt{m} \mathbf{W}, \quad (1)$$

where \mathbf{W} represents some unknown noise and \sqrt{m} is a normalization factor. The problem can be considered as a rank-one special case of finding a low-rank matrix in the presence of noise. Such low-rank estimation problems arise in a range of applications including blind channel estimation, antenna array processing, subspace system identification, and principal component or factor analysis.

When the noise term \mathbf{W} is zero, the vector pair $(\mathbf{u}_0, \mathbf{v}_0)$ can be recovered exactly, up to a scaling, from the maximal left and right singular vectors of \mathbf{A} [1]. However, in the presence of noise, the rank-one matrix can in general only be estimated approximately. In this case, *a priori* information or constraints on $(\mathbf{u}_0, \mathbf{v}_0)$ may improve the estimation. Such constraints arise, for example, in factor analysis in statistics, where one of the factors is often constrained to be either positive or sparse [2]. Similar sparsity constraints arise in the problem of dictionary learning [3]. In digital communications, one of the factors could come from a discrete QAM constellation.

Unfortunately, optimal estimation with constraints on \mathbf{u}_0 or \mathbf{v}_0 is often computationally intractable. The chief problem is the bilinear nature of the term $\mathbf{u}_0 \mathbf{v}_0^T$. However, the term is linear in \mathbf{u}_0 and \mathbf{v}_0 separately. Thus, many suboptimal estimation methods are performed in an iterative manner, alternately estimating \mathbf{u}_0 and \mathbf{v}_0 individually, while holding the estimate of the other factor constant.

This paper proposes a variant of such alternating optimization procedures that we call Iterative Factorization (IterFac), stated in detail in Algorithm 1 below. Through proper selection of relevant parameters, the IterFac algorithm can perform alternating minimizations to optimizations of the form

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}) := \arg \min_{\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n} F(\mathbf{u}, \mathbf{v}), \quad (2)$$

where the objective function is given by

$$F(\mathbf{u}, \mathbf{v}) = \frac{1}{2m} \|\mathbf{A} - \mathbf{u} \mathbf{v}^T\|_F^2 + c_U(\mathbf{u}) + c_V(\mathbf{v}). \quad (3)$$

Here, $\|\mathbf{X}\|_F$ is the Frobenius norm and $c_U(\mathbf{u})$ and $c_V(\mathbf{v})$ are cost or regularization functions on the left and right factors. In the case when the cost functions are separable, the IterFac algorithm reduces the vector optimization problem to a sequence of scalar optimization problems on the individual components of \mathbf{u} and \mathbf{v} ; it is thus computationally simple. The IterFac methodology is also general since the method can be applied to essentially arbitrary separable cost functions.

Of course, such iterative algorithms are by no means new; they underlie many existing methods, including the classic alternating power method for finding maximal singular values [1] and some alternating methods in sparse or non-negative dictionary learning [3]–[7]. More recently, an alternating method has been proposed for estimation with matrix uncertainties [8], also using an approximate message passing technique. The IterFac iterations also have some similarities to the updates in gradient descent methods in sparse dictionary learning such as in [9].

Our main contribution is to show that, under a particular setting of a damping parameter, the IterFac algorithm admits an asymptotically-exact characterization when \mathbf{W} is i.i.d. Gaussian noise and the components of the true vectors \mathbf{u}_0 and \mathbf{v}_0 have limiting empirical distributions. In this scenario, we show that the empirical joint distribution of the components of \mathbf{u}_0 and the corresponding estimates from the IterFac algorithm are described by a simple *scalar equivalent model* where the IterFac component estimates are identically distributed to scalar estimates of the variables corrupted by Gaussian noise. Moreover, the effective Gaussian noise level in this model is described by a simple set of scalar *state evolution* (SE) equations. From the scalar equivalent model, one can compute the asymptotic value of almost any component-separable metric including mean-squared error or correlation. Thus, in addition to being computationally simple and general, the IterFac algorithm admits a precise analysis in the case of Gaussian noise. Moreover, since fixed points of the IterFac

algorithm correspond, under suitable circumstances, to local minima of objectives such as (3), the analysis can be used to characterize the behavior of such minima—even if alternate algorithms to IterFac are used.

The main analytical tool is a recently-developed technique by Bayati and Montanari [10] used in the analysis of approximate message passing (AMP) algorithms. AMP methods are Gaussian approximations of loopy belief propagation for estimation of vectors under large random linear measurements. The work [10] applied AMP techniques to compressed sensing, and proved that, in the limit for large Gaussian mixing matrices, the behavior of AMP estimates can be described by a scalar equivalent model with effective noise levels defined by certain scalar state evolution (SE) equations. Similar SE analyses have appeared in related contexts [11]–[16]. To prove the SE equations for the IterFac algorithm, we apply a key theorem from [10] with a simple change of variables and a slight modification to account for parameter adaptation. A complete version of this paper is available at [17], which contains all the proofs and more details.

II. ITERATIVE RANK-ONE FACTORIZATION

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, we consider the following iterative algorithm for estimating the rank-one factors \mathbf{u}_0 and \mathbf{v}_0 .

Algorithm 1 Iterative Factorization (IterFac)

Require: Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and factor selection functions

$$G_u(t, \mathbf{p}, \lambda_u) \text{ and } G_v(t, \mathbf{q}, \lambda_v).$$

1: $t \leftarrow 0$

2: Select initial values $\mathbf{u}(0), \mathbf{v}(0)$

3: **repeat**

4: {Update estimate of \mathbf{u} }

5: Select parameters $\lambda_u(t)$ and $\mu_u(t)$

6: $\mathbf{p}(t) \leftarrow (1/m)\mathbf{A}\mathbf{v}(t) + \mu_u(t)\mathbf{u}(t)$

7: $\mathbf{u}(t+1) \leftarrow G_u(t, \mathbf{p}(t), \lambda_u(t))$

8:

{Update estimate of \mathbf{v} }

9: Select parameters $\lambda_v(t)$ and $\mu_v(t)$

10: $\mathbf{q}(t) \leftarrow (1/m)\mathbf{A}^T\mathbf{u}(t+1) + \mu_v(t)\mathbf{v}(t)$

11: $\mathbf{v}(t+1) \leftarrow G_v(t, \mathbf{q}(t), \lambda_v(t))$

12: **until** Terminate

The output of the algorithm, $(\mathbf{u}(t), \mathbf{v}(t))$, $t = 0, 1, \dots$, is a sequence of estimates for $(\mathbf{u}_0, \mathbf{v}_0)$. The algorithm has several parameters including the initial conditions, the parameters in lines 5 and 9, the termination condition and, most importantly, the functions $G_u(\cdot)$ and $G_v(\cdot)$. In each iteration, the functions $G_u(\cdot)$ and $G_v(\cdot)$ are used to generate the estimates of the factors $\mathbf{u}(t)$ and $\mathbf{v}(t)$. $G_u(\cdot)$ and $G_v(\cdot)$ will thus be called the *factor selection functions*.

To understand the role of the factor selection functions, suppose that we wish to perform the optimization (2) for some regularization functions $c_U(\cdot)$ and $c_V(\cdot)$. Consider the factor

selections given by the minimization

$$G_u(t, \mathbf{p}, \lambda_u) := \arg \min_{\mathbf{u} \in \mathbb{R}^m} \left[-\mathbf{p}^T \mathbf{u} + c_U(\mathbf{u}) + \frac{\lambda_u}{2} \|\mathbf{u}\|^2 \right], \quad (4a)$$

$$G_v(t, \mathbf{q}, \lambda_v) := \arg \min_{\mathbf{v} \in \mathbb{R}^n} \left[-\mathbf{q}^T \mathbf{v} + c_V(\mathbf{v}) + \frac{\lambda_v}{2} \|\mathbf{v}\|^2 \right], \quad (4b)$$

where the parameters $\lambda_u(t)$ and $\lambda_v(t)$ are given by

$$\lambda_u(t) := \mu_u(t) + \|\mathbf{v}(t)\|^2/m, \quad (5a)$$

$$\lambda_v(t) := \mu_v(t) + \|\mathbf{u}(t+1)\|^2/m. \quad (5b)$$

With the factor selection functions (4) and non-negative values of the coefficients $\mu_u(t)$ and $\mu_v(t)$, it is shown in the full paper [17] that the objective function is monotonically decreasing,

$$F(\mathbf{u}(t+1), \mathbf{v}(t+1)) \leq F(\mathbf{u}(t), \mathbf{v}(t)), \quad (6)$$

where the coefficients $\mu_u(t)$ and $\mu_v(t)$ play a role in damping the steps.

III. ASYMPTOTIC ANALYSIS UNDER GAUSSIAN NOISE

A. Model and Assumptions

We analyze the algorithm under the following assumptions.

Assumption 1: Consider a sequence of random realizations of the estimation problem in Section I indexed by the dimension n . The matrix \mathbf{A} and the parameters in Algorithm 1 satisfy the following:

- (a) For each n , the output dimension $m = m(n)$ is deterministic and scales linearly with the input dimension in that

$$\lim_{n \rightarrow \infty} n/m(n) = \beta \quad (7)$$

for some $\beta > 0$.

- (b) The matrix \mathbf{A} has the form (1) where $\mathbf{u}_0 \in \mathbb{R}^m$ and $\mathbf{v}_0 \in \mathbb{R}^n$ represent “true” left and right factors of a rank one term, and $\mathbf{W} \in \mathbb{R}^{m \times n}$ is an i.i.d. Gaussian matrix with components $W_{ij} \sim \mathcal{N}(0, \tau_w)$ for some $\tau_w > 0$.

- (c) The factor selection functions $G_u(t, \mathbf{p}, \lambda_u)$ and $G_v(t, \mathbf{q}, \lambda_v)$ in lines 7 and 11 are componentwise separable in that for all component indices i and j ,

$$G_u(t, \mathbf{p}, \lambda_u)_i = G_u(t, p_i, \lambda_u), \quad (8a)$$

$$G_v(t, \mathbf{q}, \lambda_v)_j = G_v(t, q_j, \lambda_v), \quad (8b)$$

for some scalar functions $G_u(t, p, \lambda_u)$ and $G_v(t, q, \lambda_v)$ satisfying certain Lipschitz continuity and differentiability assumptions, see [17].

- (d) The damping factors are selected by the rules

$$\mu_u(t+1) = -\frac{\tau_w}{m} \sum_{j=1}^n \frac{\partial}{\partial q_j} G_v(t, q_j(t), \lambda_v(t)) \quad (9a)$$

$$\mu_v(t) = -\frac{\tau_w}{m} \sum_{i=1}^m \frac{\partial}{\partial p_i} G_u(t, p_i(t), \lambda_u(t)), \quad (9b)$$

with the initial damping factor $\mu_u(0) = 0$.

- (e) The parameters $\lambda_u(t)$ and $\lambda_v(t)$ are computed via an empirical mean of the functions of $(\mathbf{v}_0, \mathbf{v}(t))$ and $(\mathbf{u}_0, \mathbf{u}(t+1))$ – see [17] for details.
- (f) For each n and iteration number t , define the sets

$$\theta_u(t) = \{(u_{0i}, u_i(t)), i = 1, \dots, m\}, \quad (10a)$$

$$\theta_v(t) = \{(v_{0j}, v_j(t)), j = 1, \dots, n\}. \quad (10b)$$

For $t = 0$, as $n \rightarrow \infty$ the sets $\theta_u(0)$ and $\theta_v(t)$ empirically converge with bounded moments of order $2p - 2$ for some $p \geq 1$ to some random variable pairs $(U_0, U(0))$ and $(V_0, V(0))$. See [17] for a precise definition of the empirical convergence used here.

A complete discussion of the assumptions can be found in the full paper [17].

B. Main Result

Our main result, Theorem 1 below, will provide a characterization of the asymptotic empirical distribution for the sets $\theta_u(t)$ and $\theta_v(t)$ in (10). Specifically, we will show that, for all t , the sets have empirical limits of the form

$$\lim_{n \rightarrow \infty} \theta_u(t) \stackrel{d}{=} (U_0, U(t)), \quad (11a)$$

$$\lim_{n \rightarrow \infty} \theta_v(t) \stackrel{d}{=} (V_0, V(t)), \quad (11b)$$

for some random variable pairs $(U_0, U(t))$ and $(V_0, V(t))$. The distributions of the random variables $(U_0, U(t))$ and $(V_0, V(t))$ can be described recursively in t as follows: For $t = 0$, $(V_0, V(0))$ is the random variable pair in the initial condition Assumption 1(f). Then, for $t \geq 0$, $(U_0, U(t+1))$ is given by

$$U(t+1) = G_u(t, P(t), \bar{\lambda}_u(t)), \quad (12a)$$

$$P(t) = \beta \bar{\alpha}_{v1}(t) U_0 + Z_u(t), \quad (12b)$$

$$Z_u(t) \sim \mathcal{N}(0, \beta \tau_w \bar{\alpha}_{u0}(t)), \quad (12c)$$

where U_0 is the random variable in Assumption 1(f); $G_u(\cdot)$ is the scalar function in Assumption 1(d); and $Z_u(t)$ is Gaussian noise independent of U_0 . Thus, $U(t+1)$ is distributed identically to the output of the scalar function $G_u(t, P(t))$, where $P(t)$ is a scaled and Gaussian noise-corrupted version of the true variable U_0 . Similarly, for $t \geq 0$, the random variable $(V_0, V(t+1))$ in (11b) is given by

$$V(t+1) = G_v(t, Q(t), \bar{\lambda}_v(t)), \quad (13a)$$

$$Q(t) = \bar{\alpha}_{u1}(t+1) V_0 + Z_v(t), \quad (13b)$$

$$Z_v(t) \sim \mathcal{N}(0, \tau_w \bar{\alpha}_{u0}(t+1)), \quad (13c)$$

where V_0 is the empirical limit in Assumption 1(f), and $Z_v(t)$ is Gaussian noise independent of V_0 .

The constants $\bar{\alpha}_{u0}(t)$, $\bar{\alpha}_{u1}(t)$, $\bar{\alpha}_{v0}(t)$, $\bar{\alpha}_{v1}(t)$, $\bar{\lambda}_u(t)$, and $\bar{\lambda}_v(t)$ in (12) and (13) are deterministic sequences of scalars defined recursively by

$$\bar{\alpha}_{u0}(t) = \mathbb{E}[U(t)^2], \quad \bar{\alpha}_{u1}(t) = \mathbb{E}[U_0 U(t)], \quad (14a)$$

$$\bar{\alpha}_{v0}(t) = \mathbb{E}[V(t)^2], \quad \bar{\alpha}_{v1}(t) = \mathbb{E}[V_0 V(t)], \quad (14b)$$

and

$$\bar{\lambda}_u(t) = \mathbb{E}[\phi_{\lambda u}(V_0, V(t))], \quad (15a)$$

$$\bar{\lambda}_v(t) = \mathbb{E}[\phi_{\lambda v}(U_0, U(t+1))], \quad (15b)$$

which are initialized with the values $\bar{\alpha}_{v0}(0)$ and $\bar{\alpha}_{v1}(0)$ from the joint distribution of $(V_0, V(0))$ in Assumption 1(f). The subsequent values of $\bar{\alpha}_{u0}(t)$, $\bar{\alpha}_{u1}(t)$, $\bar{\alpha}_{v0}(t)$, $\bar{\alpha}_{v1}(t)$, $\bar{\lambda}_u(t)$ and $\bar{\lambda}_v(t)$ can then be computed recursively through (14) along with the definition of the variables in (12) and (13). In line with [10] upon which our analysis is based, we will call the set of recursive equations (12)–(15) the *state evolution* (SE) equations. From the solution to the SE equations, we get the exact descriptions of the limits in (11).

Theorem 1: Under Assumption 1, the sets $\theta_u(t)$ and $\theta_v(t)$ in (10) converge empirically with bounded moments of order p with the limits in (11).

The main contribution of Theorem 1 is that it provides a simple *scalar equivalent model* for the asymptotic behavior of the algorithm. The sets $\theta_u(t) = \{(u_{0i}, u_i(t))\}$ and $\theta_v(t) = \{(v_{0j}, v_j(t))\}$ in (10) are the components of true vectors \mathbf{u}_0 and \mathbf{v}_0 and their estimates $\mathbf{u}(t)$ and $\mathbf{v}(t)$. The theorem shows that empirical distribution of these components are asymptotically equivalent to simple random variable pairs $(U_0, U(t))$ and $(V_0, V(t))$ given by (12a) and (13a). Following [18], we can thus call the result a *single-letter* characterization.

From this single-letter characterization, one can exactly compute a large class of performance metrics of the algorithm. Specifically, the empirical convergence of $\theta_u(t)$ shows that for any pseudo-Lipschitz function $\phi(u_0, u)$ of order p , the following limit exists almost surely:

$$\lim_{n \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \phi(u_{0i}, u_i(t)) = \mathbb{E}[\phi(U_0, U(t))], \quad (16)$$

where the expectation on the right-hand side is over the variables $(U_0, U(t))$ with U_0 identical to the variable in the limit in Assumption 1(f) and $U(t)$ given by (12a). This expectation can be explicitly evaluated by a simple two-dimensional integral and consequently any component-separable performance metric based on a suitably continuous loss function $\phi(u_0, u)$, such as mean-squared error can be exactly computed. The full paper [17] shows, in particular, that one can evaluate the asymptotic correlations

$$\rho_u(t) := \lim_{n \rightarrow \infty} \frac{|\mathbf{u}(t)^T \mathbf{u}_0|^2}{\|\mathbf{u}(t)\|^2 \|\mathbf{u}_0\|^2} = \frac{\bar{\alpha}_{u1}^2(t)}{\bar{\alpha}_{u0}(t) \tau_u}, \quad (17a)$$

$$\rho_v(t) := \lim_{n \rightarrow \infty} \frac{|\mathbf{v}(t)^T \mathbf{v}_0|^2}{\|\mathbf{v}(t)\|^2 \|\mathbf{v}_0\|^2} = \frac{\bar{\alpha}_{v1}^2(t)}{\bar{\alpha}_{v0}(t) \tau_v}, \quad (17b)$$

which represent the cosines of the angles between the estimates $\mathbf{u}(t)$ and $\mathbf{v}(t)$ and the true vectors \mathbf{u}_0 and \mathbf{v}_0 .

IV. EXAMPLES

A. Linear Selection Functions

As a first simple application of the SE analysis, suppose we use linear selection functions of the form

$$G_u(t, \mathbf{p}, \lambda_u) = \lambda_u \mathbf{p}, \quad G_v(t, \mathbf{q}, \lambda_v) = \lambda_v \mathbf{q}, \quad (18)$$

where the parameters λ_u and λ_v allow for normalization or other scalings of the outputs. Linear selection functions of the form (18) arise when one selects $G_u(\cdot)$ and $G_v(\cdot)$ from (4) with zero cost functions, $c_U(\mathbf{u}) = c_V(\mathbf{v}) = 0$. With zero cost functions, the correct solution to the optimization (2) is for (\mathbf{u}, \mathbf{v}) to be the (appropriately scaled) left and right maximal singular vectors of \mathbf{A} . We will thus call the estimates $(\mathbf{u}(t), \mathbf{v}(t))$ of Algorithm 1 and linear selection functions (18) the *estimated* maximal singular vectors.

Theorem 2: Consider the state evolution equations (12), (13), and (14) with the linear selection functions (18). Then:

- (a) The asymptotic correlation coefficients (17a) and (17b) satisfy the following recursive rules:

$$\rho_u(t+1) = \frac{\beta\tau_u\tau_v\rho_v(t)}{\beta\tau_u\tau_v\rho_v(t) + \tau_w}, \quad (19a)$$

$$\rho_v(t) = \frac{\tau_u\tau_v\rho_u(t)}{\tau_u\tau_v\rho_u(t) + \tau_w}. \quad (19b)$$

- (b) For any positive initial condition, $\rho_v(0) > 0$, the asymptotic correlation coefficients converge to the limits

$$\lim_{t \rightarrow \infty} \rho_u(t) = \rho_u^* := \frac{[\beta\tau_u^2\tau_v^2 - \tau_w^2]_+}{\tau_u\tau_v(\beta\tau_u\tau_v + \tau_w)}, \quad (20a)$$

$$\lim_{t \rightarrow \infty} \rho_v(t) = \rho_v^* := \frac{[\beta\tau_u^2\tau_v^2 - \tau_w^2]_+}{\beta\tau_u\tau_v(\tau_u\tau_v + \tau_w)}, \quad (20b)$$

where $[x]_+ = \max\{0, x\}$.

The theorem provides a set of recursive equations for the asymptotic correlation coefficients $\rho_u(t)$ and $\rho_v(t)$ along with simple expressions for the limiting values as $t \rightarrow \infty$. We thus obtain exactly how correlated the estimated maximal singular vectors of a matrix \mathbf{A} of the form (1) are to the rank one factors $(\mathbf{u}_0, \mathbf{v}_0)$. The proof of the theorem also provides expressions for the second-order statistics in (14) to be used in the scalar equivalent model.

The fixed point expressions (20) agree with the more general results in [19] that derive the correlations for ranks greater than one and low-rank recovery with missing entries. Similar results can also be found in [20]. An interesting consequence of the expressions in (20) is that unless $\sqrt{\beta}\tau_u\tau_v > \tau_w$, the asymptotic correlation coefficients are exactly zero. The ratio $\tau_u\tau_v/\tau_w$ can be interpreted as a scaled SNR.

B. Minimum Mean-Squared Error Estimation

Next suppose that the priors on U_0 and V_0 are known. In this case, given (12a) and (13a), a natural choice for the factor selection functions are

$$G_u(t, \mathbf{p}) = \mathbb{E}[U_0 | P(t)], \quad G_v(t, \mathbf{q}) = \mathbb{E}[V_0 | Q(t)], \quad (21)$$

which are the MMSE estimates of the variables. In this example, there are no parameters λ_u or λ_v . We can use the initial condition $v_j(0) = \mathbb{E}[V_0]$ for all j , so that the initial variable in Assumption 1(f) is $V(0) = \mathbb{E}[V_0]$. To analyze the algorithms define

$$\mathcal{E}_u(\eta_u) := \mathbf{var}(U_0 | Y = \sqrt{\eta_u}U_0 + D), \quad (22a)$$

$$\mathcal{E}_v(\eta_v) := \mathbf{var}(V_0 | Y = \sqrt{\eta_v}V_0 + D), \quad (22b)$$

where $D \sim \mathcal{N}(0, 1)$ is independent of U_0 and V_0 . That is, $\mathcal{E}_u(\eta_u)$ and $\mathcal{E}_v(\eta_v)$ are the mean-squared errors of estimating U_0 and V_0 from observations Y with SNRs of η_u and η_v . The functions $\mathcal{E}_u(\cdot)$ and $\mathcal{E}_v(\cdot)$ arise in a range of estimation problems and the analytic and functional properties of these functions can be found in [21], [22].

Theorem 3: Consider the solutions to the SE equations (12), (13), and (14) under the MMSE selection functions (21) and initial condition $V(0) = \mathbb{E}[V_0]$. Then:

- (a) For all t , the asymptotic correlation coefficients (17a) and (17b) satisfy the recursive relationships

$$\rho_u(t+1) = 1 - \frac{1}{\tau_u} \mathcal{E}_u(\beta\tau_v\rho_v(t)/\tau_w), \quad (23a)$$

$$\rho_v(t) = 1 - \frac{1}{\tau_v} \mathcal{E}_v(\tau_u\rho_u(t)/\tau_w), \quad (23b)$$

with the initial condition $\rho_v(0) = (\mathbb{E}V_0)^2/\tau_w$.

- (b) If, in addition, $\mathcal{E}_u(\eta_u)$ and $\mathcal{E}_v(\eta_v)$ are continuous, then for any positive initial condition, $\rho_v(0) > 0$, as $t \rightarrow \infty$, the asymptotic correlation coefficients $(\rho_u(t), \rho_v(t))$ increase monotonically to fixed points (ρ_u^*, ρ_v^*) of (23) with $\rho_v^* > 0$.

Again, we see that we can obtain simple, explicit recursions for the asymptotic correlations.

V. NUMERICAL SIMULATION

To validate the SE analysis, we consider a simple case where the left factor $\mathbf{u}_0 \in \mathbb{R}^m$ is i.i.d. Gaussian, zero mean and $\mathbf{v}_0 \in \mathbb{R}^n$ has Bernoulli-Exponential components:

$$v_{0j} \sim \begin{cases} 0 & \text{with prob } 1 - \lambda, \\ \text{Exp}(1) & \text{with prob } \lambda, \end{cases} \quad (24)$$

which provides a simple model for a sparse, positive vector. The parameter λ is the fraction of nonzero components and is set in this simulation to $\lambda = 0.1$. The dimensions are $(m, n) = (1000, 500)$, and the noise level τ_w is set according to the scaled SNR defined as

$$\text{SNR} = 10 \log_{10}(\tau_u\tau_v/\tau_w). \quad (25)$$

Estimating the vector \mathbf{v}_0 in this set-up is related to finding sparse principal vectors of the matrix $\mathbf{A}^T\mathbf{A}$ [2], [23]–[27].

The results of the simulation are shown in Fig. 1, which shows the simulated and SE-predicted performance of the IterFac algorithm with both the linear and MMSE selection functions for the priors on \mathbf{u} and \mathbf{v} . The algorithm is run for $t = 10$ iterations and the plot shows the median of the final correlation coefficient $\rho_v(t)$ over 50 Monte Carlo trials at each value of SNR. It can be seen that the performance of the IterFac algorithm for both the linear and MMSE estimates are in excellent agreement with the SE predictions. The correlation coefficient of the linear estimator also matches the correlation of the estimates produced from the maximal singular vectors of \mathbf{A} . This is not surprising since, with linear selection functions, the IterFac algorithm is essentially an iterative method to find the maximal singular vectors. The figure also shows the benefit of exploiting the prior on \mathbf{v}_0 , which is evident from the

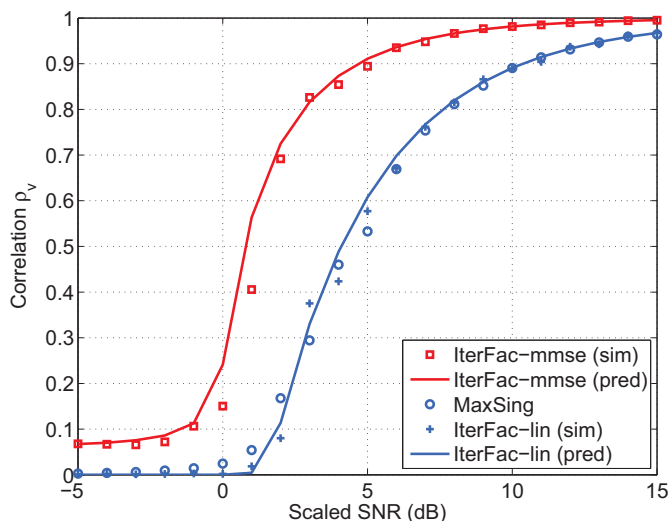


Fig. 1. Simulation of the IterFac algorithm for both the linear selection functions in Section IV-A (labeled iter-lin) and MMSE selection functions in Section IV-B (labeled iter-mmse). Plotted are the correlation values after 10 iterations. The simulated values are compared against the SE predictions. Also shown is the simple estimate from the maximal singular vectors of \mathbf{A} .

superior performance of the MMSE estimate over the linear reconstruction.

CONCLUSIONS AND FUTURE WORK

We have presented a computationally-efficient method for estimating rank-one matrices in noise. The estimation problem is reduced to a sequence of scalar AWGN estimation problems which can be performed easily for a large class of priors or regularization functions on the coefficients. In the case of Gaussian noise, the asymptotic performance of the algorithm is exactly characterized by a set of scalar state evolution equations which appear to match the performance at moderate dimensions well. Thus, the methodology is computationally simple, general and admits a precise analysis in certain asymptotic, random settings. However, to make the algorithm practical, we need to extend the method to higher ranks and handle the case where the priors are not known. The full paper [17] suggests some methods for addressing these issues in the future.

ACKNOWLEDGMENTS

The authors thank Vivek Goyal, Ulugbek Kamilov, Andrea Montanari and Phil Schniter for detailed comments on an earlier draft.

REFERENCES

- [1] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985, reprinted with corrections 1987.
- [2] J. Cadima and I. T. Jolliffe, "Loadings and correlations in the interpretation of principal components," *J. Appl. Stat.*, vol. 22, pp. 208–214, 1995.
- [3] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Network Comp. Neural Syst.*, vol. 7, no. 2, pp. 333–339, 1996.
- [4] —, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *J. Vis. Res.*, vol. 37, pp. 3311–3325, 1997.

- [5] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Neural Information Process. Syst.*, Vancouver, Canada, Dec. 2001.
- [6] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comp.*, vol. 12, pp. 337–365, 2000.
- [7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [8] J. T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *Conf. Rec. Asilomar Conf. on Signals, Syst. & Computers*, Pacific Grove, CA, Nov. 2011.
- [9] J. Mairal, F. Bach, G. Sapiro, and J. Ponce, "Online dictionary learning for sparse coding," in *Proc. ICML*, Montréal, Canada, Jun. 2009.
- [10] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [11] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.
- [12] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," arXiv:cs/0602028v1 [cs.IT], Feb. 2006.
- [13] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.
- [14] —, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 946–950.
- [15] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," arXiv:1001.2228v1 [cs.IT], Jan. 2010.
- [16] —, "Generalized approximate message passing for estimation with random linear mixing," arXiv:1010.5141v1 [cs.IT], Oct. 2010.
- [17] S. Rangan and A. K. Fletcher, "Iterative Estimation of Constrained Rank-One Matrices in Noise," arXiv preprint, Feb. 2011.
- [18] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep.–Oct. 2009.
- [19] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.
- [20] M. Capitaine, C. Donati-Martin, and D. Féral, "The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations," *Ann. Probability*, vol. 37, no. 1, pp. 1–47, Jan. 2009.
- [21] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
- [22] Y. Wu and S. Verdú, "Functional Properties of Minimum Mean-Square Error and Mutual Information," *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
- [23] I. T. Jolliffe, N. Trendafilov, and M. Uddin, "A modified principal component technique based on the LASSO," *J. Comput. Graphical Stat.*, vol. 12, pp. 531–547, 2003.
- [24] Z. Zhang, H. Zha, and H. Simon, "Low rank approximations with sparse factors I: Basic algorithms and error analysis," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 3, pp. 706–727, 2002.
- [25] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graphical Stat.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.
- [26] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Rev.*, vol. 49, no. 3, pp. 434–448, Mar. 2007.
- [27] H. Shena and J. Z. Huang, "Sparse principal component analysis," *J. Multivariate Anal.*, vol. 99, no. 6, pp. 1015–1034, Jul. 2008.