Subspace Estimation from Incomplete Observations: A Precise High-Dimensional Analysis

Chuang Wang[†], Yonina Eldar[‡] and Yue M. Lu[†]

[†]Harvard Paulson School of Engineering and Applied Sciences [‡]Technion, Israel Institute of Technology

***Abstract*—The problem of estimating and tracking low-rank subspaces from incomplete observations has received a lot of attention recently in the signal processing and learning communities. Popular algorithms, such as GROUSE [1] and PETRELS [2], are often very effective in practice, but their performance depends on the careful choice of algorithmic parameters. Important questions, such as the global convergence of these algorithms and how the noise level, subsampling ratio, and various other parameters affect the performance, are not fully understood. In this paper, we present a precise analysis of the performance of these algorithms in the asymptotic regime where the ambient dimension tends to infinity. Specifically, we show that the time-varying trajectories of estimation errors converge weakly to a deterministic function of time, which is characterized as the unique solution of a system of ordinary differential equations (ODEs.) Analyzing the limiting ODEs also reveals and characterizes sharp phase transition phenomena associated with these algorithms. Numerical simulations verify the accuracy of our asymptotic predictions, even for moderate signal dimensions.**

I. Introduction

Consider the problem of estimating a low-rank subspace using partial observations from a data stream. At any time k, an n-D sample vector $\boldsymbol{x}_k$ is generated as $\boldsymbol{x}_k = \boldsymbol{U}\boldsymbol{c}_k + \boldsymbol{z}_k$, where $\boldsymbol{U} \in \mathbb{R}^{n\times d}$ is an unknown matrix whose columns form an orthonormal basis of a d-D subspace, $\boldsymbol{c}_k \sim \mathcal{N}(0, \boldsymbol{I}_d)$ is a vector of expansion coefficients, and $\boldsymbol{z}_k \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$ denotes the noise. We assume that each coordinate of $\boldsymbol{x}_k$ is observed independently with probability α. Let $p_{k,i} = 1$ if the ith component of $\boldsymbol{x}_k$ is observed, and $p_{k,i} = 0$ otherwise. Introducing a diagonal matrix $\boldsymbol{P}_k = \text{diag}(p_{k,1}, p_{k,2}, \ldots, p_{k,n})$, we can then write our observation as $\boldsymbol{y}_k = \boldsymbol{P}_k \boldsymbol{x}_k$. Given the incomplete observations $\{\boldsymbol{y}_k, \boldsymbol{P}_k\}_k$ arriving in a stream, we aim to estimate the subspace spanned by the columns of $\boldsymbol{U}$.

GROUSE [1] and PETRELS [2] are two well-known methods in the literature for solving the above estimation problem. They are both online algorithms in the sense that they provide instantaneous, *on-the-fly* updates to their subspace estimates upon the arrival of a new data point $\{\boldsymbol{y}_k, \boldsymbol{P}_k\}$. The two differ in their update rules: GROUSE performs first-order incremental gradient descent on the Grassmannian, whereas PETRELS can be interpreted as a second-order stochastic gradient descent scheme. Both algorithms have been shown to be highly effective in practice, but their performance depends on the careful choice of algorithmic parameters such as the step size (for GROUSE) and the discount parameter (for PETRELS). Various convergence properties of these algorithms have been established in [2]–[5], but in general, the issue of global convergence with subsampled data is still open. Moreover, the important question of how the noise level σ, the subsampling ratio α, and various other algorithmic parameters affect the performance is not fully understood.

II. High Dimensional Analysis and Phase Transition

In this work, we provide an *exact* asymptotic performance analysis of GROUSE, PETRELS, and other related algorithms (*e.g.*, [6], [7]) in the large n limit. Due to space constraint, we present here the results for PETRELS when the subspace dimension $d = 1$, but our analysis can be extended to handle other algorithms and any finite d.

Let $\boldsymbol{D}_k \in \mathbb{R}^n$ be the estimate provided by the algorithm at step k. We measure the estimation performance via the squared *cosine similarity* $F_k^n \stackrel{\text{def}}{=} (\boldsymbol{D}_k^T \boldsymbol{U})^2 / (\|\boldsymbol{D}_k\|^2 \|\boldsymbol{U}\|^2)$, where the superscript n denotes the underlying ambient dimension. One can show that $\mathbf{E}(F_k^n - F_{k-1}^n) = \mathcal{O}(1/n)$. Thus, when n is large, it makes sense to "accelerate" the time by a factor n and study the algorithm at a rescaled time axis. To that end, we introduce the rescaled time as $t = k/n$, and define $f_n(t) \stackrel{\text{def}}{=} F_{\lfloor nt \rfloor}^n$. This way, we embed the original discrete-time process F_k^n into the space of continuous-time stochastic processes. As the main result of our work, we show that $f_n(t)$ will converge to a deterministic function $f(t)$ as $n \to \infty$. To establish this limit, we need to introduce an auxiliary parameter $G_k^n \stackrel{\text{def}}{=} n R_k^+ \|\boldsymbol{D}_k\|^{-2}$, where R_k^+ is the pseudo-inverses of the average of the correlation matrices defined in [2, Eq. (24)]. Similarly, we define the corresponding time-rescaled version as $g_n(t) = G_{\lfloor nt \rfloor}^n$.

Proposition 1. *Assume that* $F_0^n \xrightarrow{n\to\infty} f(0) > 0$. *Then as* $n \to \infty$, *the stochastic processes* $\{f_n(t), g_n(t)\}_n$ *converge weakly to the unique solution of the following system of coupled ODEs:*

$$\frac{df(t)}{dt} = 2\alpha f(1-f)g - \sigma^2 f(\alpha f + \sigma^2) g^2$$
$$\frac{dg(t)}{dt} = -g^2(\sigma^2 g + 1)(\alpha f + \sigma^2) + \mu g,$$

where σ^2 *is the noise variance,* α *is the probability with which each coordinate of the sample vectors can be observed, and* $\mu > 0$ *is a constant such that the discount parameter* λ *in [2] is* $\lambda = 1 - \frac{\mu}{n}$.

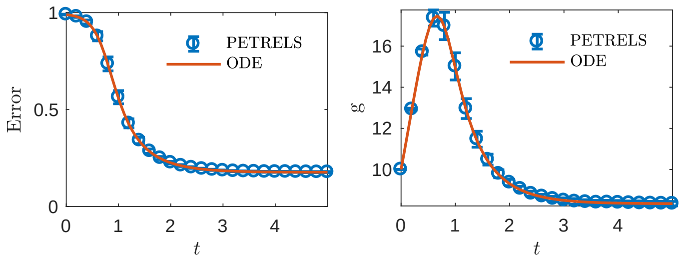
Using ODEs to analyze stochastic recursive algorithms has a long history [8], [9]. An ODE analysis of an early subspace tracking algorithm was given in [10], and this result was adapted to analyze PETRELS for the nonsubsampled case (*i.e.* $\alpha = 1$) [2]. Our results in Proposition 1 differ from previous analysis not only in that it can handle the more challenging case of incomplete observations. A more important distinction is as follows: The previous ODE analysis in [2], [10] keeps the ambient dimension n *fixed* and studies the asymptotic limit as the step size tends to 0. The resulting ODEs involve $\mathcal{O}(n)$ variables. In contrast, our analysis studies the limit as the dimension $n \to \infty$, and the resulting ODEs only involve 2 variables $f(t)$ and $g(t)$. This low-dimensional characterization makes our limiting results more practical to use, especially when the dimension is large.

Numerical verifications of our asymptotic results are shown in Figure 1. Figure 2 visualizes the solution trajectories of the ODEs starting from different initial conditions. Studying the stability of stationary points reveals the following sharp phase transition phenomenon:

Proposition 2 (Phase transition). $\lim_{t\to\infty} f(t) > 0$ *if only if*

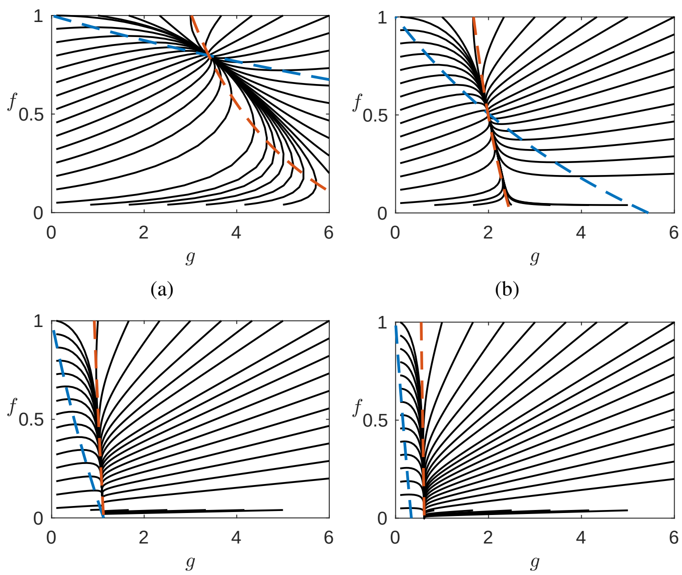
$$\mu < \left(2\alpha/\sigma^2 + 1/2\right)^2 - 1/4.$$

A "noninformative" solution corresponds to $f(t) = 0$, in which case the estimate $\boldsymbol{D}_{\lfloor nt \rfloor}$ and the underlying subspace $\boldsymbol{U}$ are uncorrelated. Proposition 2 predicts a critical choice of μ (as a function of σ and α) separating informative solutions from noninformative ones. This prediction is confirmed numerically in Figure 3. Similar asymptotic analysis can be carried out for GROUSE, as demonstrated in Figure 4.



(a) (b)

Fig. 1. Monte Carlo simulations of the PETRELS algorithm v.s. asymptotic predictions obtained by the limiting ODEs given in Proposition 1. The error is defined as $1 - f(t)$. The signal dimension is $n = 10^4$. The error bars shown in the figure correspond to one standard deviation over 50 independent trials.



(a) (b)

(c) (d)

Fig. 2. Phase portraits of the nonlinear ODEs in Proposition 1: The black curves are trajectories of the solutions $(f(t), g(t))$ of the ODES starting from different initial values. The red and blue curves represent nontrivial solutions of the two stationary equations $\frac{df(t)}{dt} = 0$ and $\frac{dg(t)}{dt} = 0$. Their intersection point is the fixed point of the dynamical system. The fixed-points of the top two figures correspond to $f(\infty) > 0$, and thus the steady-state solutions in these two cases are informative. In contrast, the fixed-points of the bottom two figures are associated with noninformative steady-state solutions with $f(\infty) = 0$.

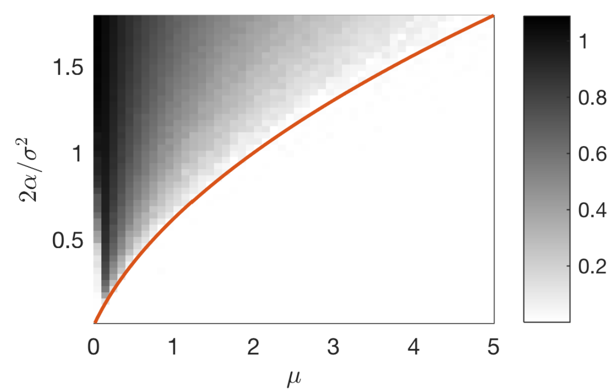


Fig. 3. The grayscale in the figure visualizes the steady-state errors of the PETRELS algorithm corresponding to different values of the noise variance σ^2, the subsampling ratio α, and the step-size parameter μ. The red curve is the theoretical prediction given in Proposition 2 of a phase transition boundary, below which no informative solution can be achieved by the algorithm. The theoretical prediction matches well with numerical results.

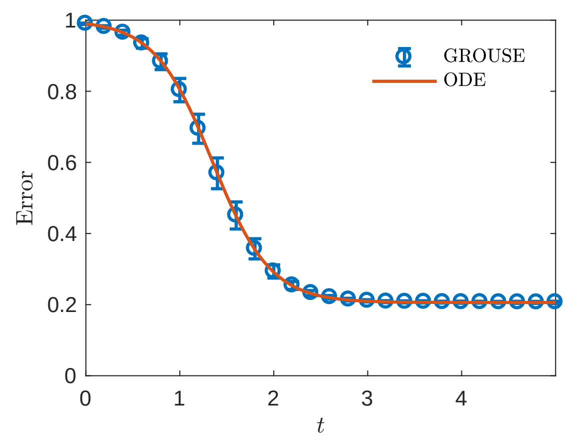


Fig. 4. Numerical simulations of the GROUSE algorithm [1] v.s. asymptotic predictions given by a limiting ODE: $\frac{df}{dt} = \tau(2\alpha - \tau\sigma^4)f - \alpha\tau(2 + \tau\sigma^2)f^2$, where $\tau > 0$ is a constant such that the step size parameter η_k defined in [1] is $\eta_k = \tau/n$. In the experiments, the error is defined as $1 - f(t)$, and the signal dimension is $n = 10^4$. The error bars shown in the figure correspond to one standard deviation over 50 independent trials.

References

[1] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Proc. Allerton Conference on Communication, Control and Computing.*, 2010.

[2] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5947–5959, 2013.

[3] L. Balzano and S. J. Wright, "Local convergence of an algorithm for subspace identification from partial data," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1279–1314, Oct. 2015.

[4] D. Zhang and L. Balzano, "Convergence of a Grassmannian gradient descent algorithm for subspace estimation from undersampled data," *arXiv:1610.00199*, 2016.

[5] A. Gonen, D. Rosenbaum, Y. C. Eldar, and S. Shalev-Shwartz, "Subspace learning with partial information," *Journal of Machine Learning Research*, vol. 17, no. 52, pp. 1–21, 2016.

[6] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *J. Math. Anal. Appl.*, vol. 106, no. 1, pp. 69–84, 1985.

[7] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[8] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *Journal of Applied Probability*, vol. 7, no. 1, p. 49, Apr. 1970.

[9] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 551–575, 1977.

[10] B. Yang, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Processing*, vol. 50, no. 1–2, pp. 123–136, Apr. 1996.