A matrix concentration inequality for products

Sina Baghal*

Abstract. We present a non-asymptotic concentration inequality for the random matrix product

$$Z_n = (I_d - \alpha X_n) (I_d - \alpha X_{n-1}) \cdots (I_d - \alpha X_1),$$

where \(\{X_k\}_{k=1}^{+\infty}\) is a sequence of bounded independent random positive semidefinite matrices with common expectation \(E[X_k] = \Sigma\). Under these assumptions, we show that, for small enough positive \(\alpha\), \(Z_n\) satisfies the concentration inequality

$$P \left(\|Z_n - E[Z_n]\| \geq t\right) \leq 2d^2 \cdot \exp \left(\frac{-t^2}{\alpha \sigma^2}\right) \text{ for all } t \geq 0,$$

where \(\sigma^2\) denotes a variance parameter.

1. Motivation

Products of random matrices appear as building blocks for many stochastic iterative algorithms, e.g. [5, 6]. While non-asymptotic bounds of averages of these matrices are well developed, e.g. [7, 8], the analogous bounds of their products are much harder to understand due to the non-commutative nature of matrix multiplication. As such, efforts to understand bounds of this type have become an active area of research e.g. [2, 3, 4].

2. Contribution

In this note, we provide a non-asymptotic concentration bound (2) for the random matrix product \(Z_n\) [1]. These instances appear, for example, in the stochastic gradient descent algorithm applied to the linear least squares problem. We remark that bound (2) will be of special interest when \(X_k\) is almost surely low rank for all \(k\). In this event, almost all eigenvalues of each factor in the matrix product \(Z_n\) are equal to 1 whereas \(E[Z_n]\) has an exponentially decaying operator norm. (Note: Without loss of generality, we can assume that \(\Sigma\) is positive

*Email: srezazadehbaghal@uwaterloo.ca
definite.) Hence, it is interesting to observe that \( Z_n \) concentrates around its mean with overwhelming probability as in (2), especially in the case where \( X_k \)'s are almost surely low rank matrices.

3. Related Work

In [3], using the uniform smoothness property of the Schatten \( p \)-norm, the authors have studied non-asymptotic bounds for the products of random matrices, in particular, random contractions [3, Theorem 7.1]. To apply their result to the matrix product (1), we will need to make some further assumptions. First, we need to assume some bound involving \( |I - \alpha X_k| \) since the Araki-Lieb-Thirring inequality [9, IX.2.11] is used in their analysis. Second, we need to assume a lower bound \( t^2 \geq c_\alpha \sum_{k=1}^{n} \|X_k - \Sigma\|^2 \) which may grow linearly in \( n \). This will be problematic particularly since we are only interested in the case where \( t \leq 1 \).

On the other hand, compared to our result, the bound in [3, Theorem 7.1] has a weaker dependency on the dimension \( d \) and, more importantly, it works in a broader variety of instances. For example, one can use their bound when in (1), instead of \( I - \alpha X_k \), we consider the factors \( I - \alpha_k X_k \) with \( \alpha_k \) decaying at a proper rate.

4. Concentration bound

In this section, we prove our result (2). The proof proceeds by constructing a martingale sequence satisfying bounded differences and then applying Azuma’s inequality. We assume that the positive semidefinite random matrices \( X_k \) in (1) are drawn independently and they satisfy \( \mathbb{E} [X_k] = \Sigma \) for all \( k \). In addition, we suppose that \( X_k \) are uniformly bounded in the operator norm, meaning that there exists \( r > 0 \) such that

\[
\|X_k\| \leq r \quad \text{almost surely.}
\]

Let \( u_1, \ldots, u_d \) denote the eigenvectors of \( \Sigma \) and \( \lambda_1 \geq \cdots \geq \lambda_d \geq 0 \) denote the corresponding eigenvalues. For each \( i \in [d] := \{1, \ldots, d\} \), define \( c_i \) to be the infimum over all positive reals for which

\[
\|(X_k - \lambda_i I) u_i\| \leq c_i \lambda_i \quad \text{almost surely.}
\]

Note that \( c_i < +\infty \) almost surely as \( c_i \leq 1 + \frac{r}{\lambda_i} \) and also, because \( X_k \) is positive semidefinite, \( c_i = 0 \) whenever \( \lambda_i = 0 \). We will use the following parameter to measure the amount of variation in \( X_k \)

\[
\sigma^2 := \frac{4d}{3} \sum_{i=1}^{d} c_i^2 \lambda_i.
\]
Theorem. Suppose that $\alpha \in (0, \frac{1}{27})$. Then the following concentration inequality holds.

$$
\mathbb{P} (\| Z_n - \mathbb{E} [Z_n] \| \geq t) \leq 2d^2 \cdot \exp \left( -\frac{t^2}{\alpha \sigma^2} \right). \tag{3}
$$

Proof. Without loss of generality, we can assume that $c_i, \lambda_i > 0$ for all $i \in [d]$. We will first show that, for any $i, j \in [d]$, the following bound holds for all $t \geq 0$:

$$
\mathbb{P} \left( \| u_i^T Z_n u_j - \mathbb{E} [u_i^T Z_n u_j] \| \geq t \sqrt{\frac{2\alpha}{3}} \cdot c_i \right) \leq 2 \exp \left( -\frac{t^2}{\alpha} \right). \tag{4}
$$

Set $Z_0 = I_d$. Then we note that for all $k \geq 0$,

$$
\mathbb{E} [u_i^T Z_k u_j] = u_i^T \mathbb{E} [Z_k] u_j = u_i^T (I - \alpha \Sigma)^k u_j = (1 - \alpha \lambda_i)^k \cdot \delta_{i,j},
$$

where $\delta_{i,j}$ stands for Kronecker delta. For notational convenience, let us denote $z_k := u_i^T Z_k u_j$. We have that

$$
\mathbb{E} [z_k | X_{k-1}, \cdots, X_1] = u_i^T \mathbb{E} [I - \alpha X_k] Z_{k-1} u_j = (1 - \alpha \lambda_i) z_{k-1}. \tag{5}
$$

Denote $q_i := 1 - \alpha \lambda_i$ and define the random variable $Y_k := q_i^{-k} \cdot z_k$. Dividing both sides of (5) by $q_i^k$, we obtain that $\mathbb{E} [Y_k | X_{k-1}, \cdots, X_1] = Y_{k-1}$. Thus, $(Y_k)_{k=1}^\infty$ is a martingale with respect to $(X_k)_{k=1}^\infty$. We observe that for all $k \geq 1$

$$
q_i^k \cdot |Y_k - Y_{k-1}| = |z_k - q_i \cdot z_{k-1}| = \alpha |u_i^T (X_k - \lambda_i I) Z_{k-1} u_j| \leq \alpha c_i \lambda_i,
$$

where the assumption $\alpha r \leq \frac{1}{27}$ yielded the bound $\| Z_{k-1} \| \leq 1$ a.s. Thus, by Azuma’s inequality, see e.g. [4], we have that for any $\epsilon \geq 0$

$$
\mathbb{P} (|z_n - \mathbb{E} [z_n]| \geq \epsilon) = \mathbb{P} (|Y_n - Y_0| \geq \epsilon \cdot q_i^{n-1}) \leq 2 \exp \left( \frac{-\epsilon^2}{2 \alpha^2 \lambda_i^2 / \sigma^2 \sum_{k=0}^{n-1} q_i^{2k}} \right). \tag{6}
$$

Note that by Jensen’s inequality

$$
\lambda_i \leq \| \Sigma \| = \| \mathbb{E} [X_k] \| \leq \mathbb{E} [\| X_k \|] \leq r. \tag{7}
$$

Therefore, by (7) and since $\alpha \in (0, \frac{1}{27})$, we obtain that $\sum_{k=0}^{n-1} q_i^{2k} \leq \frac{2}{\alpha (1 - q_i)}$. Plugging this bound into the right-hand side of (6) and letting $\epsilon = t \sqrt{\frac{2\alpha}{3}} \cdot c_i$, we will obtain (4). Finally, in order to see (3), we observe that by (4), with probability exceeding $1 - 2d^2 \cdot \exp \left( -\frac{t^2}{\alpha} \right)$, it holds that

$$
\| Z_n - \mathbb{E} [Z_n] \|_F^2 \leq t^2 \cdot \frac{4d}{3} \sum_{i=1}^{d} c_i^2 \lambda_i,
$$

where $\| Z_n - \mathbb{E} [Z_n] \|_F$ is the Frobenius norm. Therefore,

$$
\mathbb{P} (\| Z_n - \mathbb{E} [Z_n] \|_F \geq t \cdot \sigma) \leq 2d^2 \cdot \exp \left( -\frac{t^2}{\alpha} \right).
$$

The result immediately follows since $\| Z_n - \mathbb{E} [Z_n] \|_F \leq \| Z_n - \mathbb{E} [Z_n] \|_F$. \qed
 References

[1] Alon, Noga, and Joel H. Spencer. The probabilistic method. John Wiley & Sons, 2004.

[2] Henriksen, Amelia, and Rachel Ward. "Concentration inequalities for random matrix products." Linear Algebra and its Applications 594 (2020): 81-94.

[3] Huang, De, Jonathan Niles-Weed, Joel A. Tropp, and Rachel Ward. "Matrix Concentration for Products." arXiv preprint arXiv:2003.05437 (2020).

[4] Kathuria, Tarun, Satyaki Mukherjee, and Nikhil Srivastava. "On Concentration Inequalities for Random Matrix Products." arXiv preprint arXiv:2003.06319 (2020).

[5] Oja, Erkki. "Simplified neuron model as a principal component analyzer." Journal of mathematical biology 15, no. 3 (1982): 267-273.

[6] Strohmer, Thomas, and Roman Vershynin. "A randomized Kaczmarz algorithm with exponential convergence." Journal of Fourier Analysis and Applications 15, no. 2 (2009): 262.

[7] Tropp, Joel A. "User-friendly tail bounds for sums of random matrices." Foundations of computational mathematics 12, no. 4 (2012): 389-434.

[8] Wainwright, Martin J. High-dimensional statistics: A non-asymptotic viewpoint. Vol. 48. Cambridge University Press, 2019.

[9] Bhatia, Rajendra. Matrix analysis. Vol. 169. Springer Science & Business Media, 2013.