Abstract—We consider a statistical model for finite-rank symmetric tensor factorization and prove a single-letter variational expression for its mutual information when the tensor is of even order. The proof uses the adaptive interpolation method, for which rank-one matrix factorization is one of the first problems it was successfully applied to. We show how to extend the adaptive interpolation to finite-rank symmetric tensors of even order, which requires new ideas with respect to the proof for the rank-one case. We also underline where the proof falls short when dealing with odd-order tensors.

I. INTRODUCTION

Tensor factorization is a generalization of principal component analysis to tensors, in which one wishes to exhibit the closest rank-$K$ approximation to a tensor. It has numerous applications in signal processing and machine learning, e.g., for compressing data while keeping as much information as possible, in data visualization, etc. [1].

An approach to explore computational and/or statistical limits of tensor factorization is to consider a statistical model, as done in [2]. The model is the following: draw $K$ column vectors, evaluate for each of them their $p$th tensor power and sum those $K$ symmetric order-$p$ tensors. For $p = 2$, and if no degeneracy occurs, this sum is exactly the eigendecomposition of a rank-$K$ positive semidefinite matrix. Tensor factorization can then be studied as an inference problem, namely, to estimate the initial $K$ vectors from noisy observations of the tensor and to determine information theoretic limits for this task. To do so, we focus on proving formulas for the asymptotic mutual information between the noisy observed tensor and the original $K$ vectors. Such formulas were first rigorously derived for $p = 2$ and $K = 1$, i.e., rank-one matrix factorization: see [3] for the case with a binary input vector, [4] for the restricted case in which no discontinuous phase transition occurs, [5] for a single-sided bound and, finally, [6] for the fully general case. The proof in [6] combines interpolation techniques with spatial coupling and an analysis of the Approximate Message-Passing (AMP) algorithm. Later, and still for $p = 2$, [7] went beyond rank-one by using a rigorous version of the cavity method. Reference [8] applied the heuristic replica method to conjecture a formula for any $p$ and finite $K$, which is then proved for $p \geq 2$ and $K = 1$. Reference [8] also details the AMP algorithm for tensor factorization and shows how the single-letter variational expression for the mutual information allows one to give guarantees on AMP’s performance. Afterwards, [9], [10] introduced the adaptive interpolation proof technique which they applied to the case $p \geq 2$, $K = 1$. Other proofs based on interpolations recently appeared, see [11] ($p = 2$, $K = 1$) and [12] ($p \geq 2$, $K = 1$).

In this work, we prove the conjectured replica formula for any finite-rank $K$ and any even order $p$ using the adaptive interpolation method. We also underline what is missing to extend the proof to odd orders.

The adaptive interpolation method was introduced in [9], [10] as a powerful improvement to the Guerra-Toninelli interpolation scheme [13]. Since then, it has been applied to many other inference problems in order to prove formulas for the mutual information, e.g., [14], [15]. While our proof outline is similar to [10], there are two important new ingredients. First, to establish the tight upper bound, we have to prove the regularity of a change of variable given by the solutions to an ordinary differential equation. This is non-trivial when the rank becomes greater than one. Second, the same bound requires one to prove the concentration of the overlap (a quantity that fully characterizes the system in the high-dimensional limit). When the rank is greater than one, this overlap is a matrix and a recent result [16] on the concentration of overlap matrices can be adapted to obtain the required concentration in our interpolation scheme.

II. LOW-RANK SYMMETRIC TENSOR FACTORIZATION

We study the following statistical model. Let $n$ be a positive integer. $X_1, \ldots, X_n$ are random column vectors in $\mathbb{R}^K$, independent and identically distributed (i.i.d.) with distribution $P_X$. These vectors are not directly observed. Instead, for each $p$-tuple $(i_1, \ldots, i_p) \in [n]^p$ with $i_1 \leq i_2 \leq \cdots \leq i_p$, one is given access to the noisy observation

$$Y_{i_1 \ldots i_p} = \sqrt{\lambda(p-1)! \over p^{p-1}} \sum_{k=1}^K X_{i_1,k}X_{i_2,k} \ldots X_{i_p,k} + Z_{i_1 \ldots i_p} \quad (1)$$

where $\lambda$ is a known signal-to-noise ratio (SNR) and the noise $Z_{i_1 \ldots i_p}$ is i.i.d. with respect to the standard normal distribution $\mathcal{N}(0, 1)$. Let $X$ be the $n \times K$ matrix whose $i$th row is given by $X_i$. All the observations (1) are combined into the symmetric order-$p$ tensor $Y = \sqrt{\lambda(p-1)! / p^{p-1}} \sum_{k=1}^K X_{i,k}^{\otimes p} + Z$.

Our main result is the proof of a formula for the mutual information in the limit $n \to +\infty$ while the rank $K$ is kept fixed. This formula is given as the optimization of a potential
over the cone of \( K \times K \) symmetric positive semi-definite matrices \( S_K^+ \). Let \( \tilde{Z} \sim \mathcal{N}(0, I_K) \) and \( X \sim P_X \). Define the convex (see [17], Appendix A) function
\[
\psi : S \in S_K^+ \mapsto \mathbb{E} \ln \int dP_X(x) e^{X^T S x + \tilde{Z}^T \sqrt{S} x - \frac{1}{2} x^T S x},
\]
and the potential
\[
\phi_p(S) \equiv \psi(\lambda S^{p(1) - 1}) - \frac{\lambda(p - 1)}{2p} \sum_{\ell, \ell' = 1}^K (S^{p})_{\ell \ell'}, \tag{2}
\]
where \( S^{p, k} \) is the \( k \)-th Hadamard power of \( S \). Note that, by the Schur Product Theorem [18], the Hadamard product of two matrices in \( S_K^+ \) is also in \( S_K^+ \). Introducing the second moment matrix \( \Sigma_X \equiv \mathbb{E} [X X^T] \in S_K^+ \), the conjectured replica formula [8] reads
\[
\lim_{n \to \infty} \frac{1}{n} f(X; Y) = \frac{1}{2p} \sum_{\ell, \ell' = 1}^K \left( \Sigma_X^{p, \ell} \right)_{\ell \ell'} - \sup_{S \in S_K^+} \phi_p(S). \tag{3}
\]
Remark: We can reduce the proof of (3) to the case \( \lambda = 1 \) by rescaling properly \( P_X \). From now on we set \( \lambda = 1 \).

Before proving (3), we introduce important information theoretic quantities, adopting the statistical mechanics terminology. Define the Hamiltonian for all \( x \in \mathbb{R}^{n \times K} \):
\[
\mathcal{H}_n(x; Y) \equiv \sum_{i \in \mathcal{I}} (p - 1) \left( \sum_{a=1}^p x_{i a} \right)^2 - \sum_{t, \epsilon} \sqrt{\frac{(p - 1)!}{n p - 1}} \sum_{i \in \mathcal{I}} \prod_{a=1}^p x_{i a},
\]
where \( \mathcal{I} = \{ i \in [n]^p : i_a \leq i_{a+1} \} \). Using Bayes’ rule, the posterior density written in Gibbs-Boltzmann form is
\[
dP_X(x|Y) = \frac{1}{Z_n(Y)} \left( \prod_{j=1}^n dP_X(x_j) \right) e^{-\mathcal{H}_n(x; Y)},
\]
with \( Z_n(Y) \equiv \int \prod_{j=1}^n dP_X(x_j) \exp\{-\mathcal{H}_n(x; Y)\} \) the normalization factor. Finally, we define the free entropy
\[
f_n = \frac{1}{n} \mathbb{E} \ln Z_n(Y), \tag{4}
\]
which is linked to the mutual information through the identity
\[
\frac{1}{n} I(X; Y) = \frac{1}{2p} \sum_{\ell, \ell' = 1}^K \left( \Sigma_X^{p, \ell} \right)_{\ell \ell'} - f_n + O(n^{-1}). \tag{5}
\]
In [5], \( O(n^{-1}) \) is a quantity such that \( nO(n^{-1}) \) is bounded uniformly in \( n \). Thanks to [5], the replica formula [3] will follow directly from the next two bounds on the asymptotic free entropy.

**Theorem 1:** (Lower bound) Assume \( p \) is even and \( P_X \) is such that its first \( 2p \) moments are finite. Then
\[
\lim_{n \to \infty} \inf f_n \geq \sup_{S \in S_K^+} \phi_p(S). \tag{6}
\]

**Theorem 2:** (Upper bound) Assume \( p \) is even and \( P_X \) is such that its first \( 4p - 4 \) moments are finite. Then
\[
\lim_{n \to \infty} \sup f_n \leq \sup_{S \in S_K^+} \phi_p(S). \tag{7}
\]

**III. ADAPTIVE PATH INTERPOLATION**

We introduce a “time” parameter \( t \in [0, 1] \). The adaptive interpolation interpolates from the original channel [11] at \( t = 0 \) to a decoupled channel at \( t = 1 \). In between, we follow an interpolation path \( R(t, \epsilon) : [0, 1] \to S_K^+ \), which is a continuously differentiable function parametrized by a “small perturbation” \( \epsilon \in S_K^+ \) and such that \( R(0, \epsilon) = \epsilon \). More precisely, for \( t \in [0, 1] \), we observe
\[
\begin{align*}
Y_t^{(i,t)} &= \sqrt{(1-t)(p-1)!} \sum_{k=1}^p X_{i k} + Z_i, \quad i \in \mathcal{I}; \\
\tilde{Y}_j^{(t,\epsilon)} &= R(t, \epsilon) X_j + \tilde{Z}_j, \quad j \in [n].
\end{align*}
\]
The noise \( \tilde{Z}_j \sim \mathcal{N}(0, I_K) \) is independent of both \( X \) and \( Z \). The associated interpolating Hamiltonian reads
\[
\mathcal{H}_{t,\epsilon}(x; Y^{(t)}, \tilde{Y}^{(t,\epsilon)}) \equiv \mathcal{H}_t(x; Y^{(t)}) + \mathcal{H}_{t,\epsilon}(x; \tilde{Y}^{(t,\epsilon)}), \tag{9}
\]
where
\[
\begin{align*}
\mathcal{H}_t(x; Y^{(t)}) &= \sum_{i \in \mathcal{I}} \frac{(1-t)(p-1)!}{2p-1} \left( \sum_{k=1}^p x_{i k} \right)^2 - \sum_{t, \epsilon} \sqrt{\frac{(p-1)!}{n p - 1}} Y_t^{(i,t)} \sum_{k=1}^p x_{i k}, \\
\mathcal{H}_{t,\epsilon}(x; \tilde{Y}^{(t,\epsilon)}) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{2} \mathbb{E} \ln Z_{t,\epsilon}(Y^{(t)}, \tilde{Y}^{(t,\epsilon)}). \tag{10}
\end{align*}
\]
Let
\[
Z_{t,\epsilon}(Y^{(t)}, \tilde{Y}^{(t,\epsilon)}) = \int \prod_{j=1}^n dP_X(x_j) e^{-\mathcal{H}_{t,\epsilon}(x; Y^{(t)}, \tilde{Y}^{(t,\epsilon)})},
\]
so that the posterior distribution of \( X \) given \( (Y^{(t)}, \tilde{Y}^{(t,\epsilon)}) \) is \( \prod_{j=1}^n dP_X(x_j) e^{-\mathcal{H}_{t,\epsilon}(x; Y^{(t)}, \tilde{Y}^{(t,\epsilon)})} / Z_{t,\epsilon}(Y^{(t)}, \tilde{Y}^{(t,\epsilon)}) \). The interpolating free entropy is similar to [3], i.e.,
\[
f_n(t, \epsilon) \equiv \frac{1}{n} \mathbb{E} \ln Z_{t,\epsilon}(Y^{(t)}, \tilde{Y}^{(t,\epsilon)}). \tag{11}
\]
Evaluating (10) at both extremes of the interpolation gives:
\[
\begin{align*}
f_n(0, \epsilon) &= f_n + O(\|\epsilon\|); \\
f_n(1, \epsilon) &= \psi(R(1, \epsilon)).
\end{align*}
\]
\( \|\cdot\| \) denotes the Frobenius norm and \( O(\|\epsilon\|) \) is a quantity such that \( |O(\|\epsilon\|)| \leq \frac{1}{2} \mathcal{O}(\|\epsilon\|) \|\epsilon\| / 2 \). It is useful, in order to deal with future computations, to introduce the Gibbs bracket \( \langle \cdot \rangle_{t, \epsilon} \) which denotes an expectation with respect to the posterior distribution, i.e.,
\[
\langle g(x) \rangle_{t, \epsilon} = \int g(x) \prod_{j=1}^n dP_X(x_j) e^{-\mathcal{H}_{t,\epsilon}(x; Y^{(t)}, \tilde{Y}^{(t,\epsilon)})} / Z_{t,\epsilon}(Y^{(t)}, \tilde{Y}^{(t,\epsilon)}). \tag{12}
\]
Combining (11) with the fundamental theorem of calculus
\[
f_n(0, \epsilon) = f_n(1, \epsilon) - \int_0^1 f_n'(t, \epsilon) dt,
\]
\( f_n'(\cdot, \epsilon) \) being the \( t \)-derivative of \( f_n(\cdot, \epsilon) \), we obtain the sum-rule of the adaptive interpolation.
Taking the liminf on both sides of this inequality, and bearing proof of Theorem 1.

The integral in (14) can then be split in two terms: one similar to the second summand in (2), and one that will vanish in the high-dimensional limit if the overlap concentrates. The next proposition states that indeed admits a solution, which at first sight is not clear as the Gibbs bracket (−) depends itself on R(·, ). Non-trivial properties required to show the upper bound (7) are also proved.

Proposition 2: For all ∈ SK+, there exists a unique global solution R(·, ) : [0, 1] → SK+ to the first-order ODE

This solution is continuously differentiable and bounded. If p is even then ∀ ∈ [0, 1], R(t, ·) is a C1-diffeomorphism from SK+ (the open cone of K × K symmetric positive definite matrices) into R(t, SK+) whose Jacobian determinant is greater than one, i.e.,

\[ ∀ ∈ SK^+ : \det J(R(\cdot, ))(\epsilon) ≥ 1 \]  

(17)

Here J(R(·, ·)) denotes the Jacobian matrix of R(·, ·).

Proof: We now rewrite (16) explicitly as an ODE. Let  be a matrix in SK+. Consider the problem of inferring X from the following observations:

\[ \begin{align*}
Y(t) &= \sqrt{\frac{(1-t)(p-1)}{n^p}} \sum_{k=1}^{p} X_i k + Z_i, \; i ∈ I; \\
\bar{Y}(t, R) &= √R_{j} X_j + Z_j, \quad j ∈ [n].
\end{align*} \]  

(18)

It is reminiscent of the interpolating problem (8). One can form a Hamiltonian similar to (9), where R(t, ·) is simply replaced by R, and (−)R, denotes the Gibbs bracket associated to the posterior of this model. One now defines the function

\[ F_n : [0, 1] × SK^+ \rightarrow SK^+ \]  

(19)

Note that E(Q)t, R is a symmetric positive semi-definite matrix. Indeed, from the Nishimori identity [10]:

\[ \E(Q)_{t, R} = \frac{1}{n} \E[\langle X \rangle^T_{t, R} X] = \frac{1}{n} \E[\langle X \rangle^T_{t, R} \langle X \rangle_{t, R}] . \]  

By the Schur Product Theorem [18], the Hadamard power E(Q) t, R 0 p−1 also belongs to SK+, justifying that F_n takes values in the cone of symmetric positive semi-definite matrices. F_n is continuously differentiable on [0, 1] × SK+. Therefore, by the Cauchy-Lipschitz theorem, there exists a unique global solution R(·, ·) to the K(K + 1)/2-dimensional ODE:

\[ \forall t ∈ [0, 1] : \frac{dR(t)}{dt} = F_n(t, R(t)) , \; R(0) = ∈ SK^+ \]  

Each initial condition ∈ SK+ is tied to a unique solution R(·, ·). This implies that the function ∈ → R(t, ·) is injective. Its Jacobian determinant is given by Liouville’s formula [19]:

\[ \det J(R(·, ·))(\epsilon) = \exp \int_0^1 ds \sum_{1 ≤ i ≤ K} \left| \frac{∂(F_n)(t, R)}{∂R_i} \right|_{s, R(s, ·)} . \]  

(19)

1. The Nishimori identity is a direct consequence of the Bayes formula. In our setting, it states E(g(X, X))t, R = E(E(g(X, X))t, R | R = g(X, X))t, R where X, X are two samples drawn independently from the posterior distribution given Y(t), Y(t, R). Here g can also explicitly depend on Y(t), Y(t, R).
Thanks to (19), we can show that the Jacobian determinant is greater than (or equal to) one by proving that the divergence
\[
\sum_{t \leq t'} \frac{\partial (F_n t')}{\partial R t'} \bigg|_{t,R} \nonumber
\]
is nonnegative for all \((t, R) \in [0, 1] \times S^+_K\). A lengthy computation (see [17, Appendix C]) leads to the identity
\[
\sum_{t \leq t'} \frac{\partial (F_n t')}{\partial R t'} \bigg|_{t,R} = n(p - 1) \sum_{t \leq t'} E\left[\left(\frac{Q_{t', R} - Q_{t, R}}{t^*_{t, R}}\right)^{p-2}\right] \Delta t',
\]
where
\[
\Delta t' = E\left(\frac{Q_{t' + Q_{t'}}}{p} - \frac{Q_{t' + Q_{t'}}}{2}\right)^2 - \frac{E\left(\left(\frac{x^T \mathbf{r}_t + x^T \mathbf{r}}{n}\right)^2\right)}{n}.
\] (21)

If \(p\) is even then \(E\left[\left(\frac{Q_{t' + Q_{t'}}}{p} - \frac{Q_{t' + Q_{t'}}}{2}\right)^2\right] \) is nonnegative. We show next that the \(\Delta t'\)’s are nonnegative, thus ending the proof of (17). The second expectation on the right-hand side (r.h.s.) of (21) satisfies (we omit the subscripts of the Gibbs bracket):
\[
E\left(\frac{Q_{t' + Q_{t'}}}{p} - \frac{Q_{t' + Q_{t'}}}{2}\right)^2 - \frac{E\left(\left(\frac{x^T \mathbf{r}_t + x^T \mathbf{r}}{n}\right)^2\right)}{n}.
\]

The inequality is a simple application of Jensen’s inequality, while the equality that follows is an application of the Nishimori identity. The final upper bound is nothing but the first expectation on the r.h.s. of (21). Therefore \(\Delta t' \geq 0\). \[\]

2) Proof of Theorem 2

Let \(\epsilon\) be a symmetric positive definite matrix, i.e., \(\epsilon \in S^+_K\). We interpolate with the unique solution \(R(\cdot, \epsilon) : [0, 1] \rightarrow S^+_K\) to (19). Under this choice, the sum-rule (14) reads:
\[
f_n = \psi(R(1, \epsilon)) - \frac{p - 1}{2p} \sum_{t \leq t'} \int_0^1 dt E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p\right]
\]
\[
+ \frac{1}{2p} \int_0^1 dt \sum_{t \leq t'} E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p - E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p\right]\right]_{t', t}\nonumber
\]
\[
+ O(||\epsilon||) + O(n^{-1}) .
\] (22)

Using the convexity of \(\psi\), we obtain by Jensen’s inequality:
\[
\psi(R(1, \epsilon)) = \psi\left(\epsilon + \int_0^1 dt E\left[\left(\frac{Q_{t'}}{t^*_{t, R}}\right)^p\right]\right)
\]
\[
= O(||\epsilon||) + \psi\left(\int_0^1 dt E\left[\left(\frac{Q_{t'}}{t^*_{t, R}}\right)^p\right]\right)
\]
\[
\leq O(||\epsilon||) + \int_0^1 dt \psi\left(\left(\int_0^1 dt E\left[\left(\frac{Q_{t'}}{t^*_{t, R}}\right)^p\right]\right)\right) .
\] (23)

Combining both (22) and (23) directly gives
\[
f_n \leq O(n^{-1}) + O(||\epsilon||) + \int_0^1 dt \phi_p\left(E\left[\left(\frac{Q_{t'}}{t^*_{t, R}}\right)^p\right]\right)
\]
\[
+ \frac{1}{2p} \int_0^1 dt \sum_{t \leq t'} E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p - E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p\right]\right]_{t', t}
\]
\[
\leq O(n^{-1}) + O(||\epsilon||) + \sup_{S \in S^+_K} \phi_p(S) (24)
\]

In order to end the proof of (7), we must show that the second line of the upper bound (24) vanishes when \(n\) goes to infinity. This will be the case if the overlap matrix \(Q\) concentrates on its expectation \(E(Q)_{t, t}\). Indeed, provided that the \((4p - 4)\)th-order moments of \(P_X\) are finite, there exists a constant \(C_X\) depending only on \(P_X\) such that
\[
\frac{1}{2p} \int_0^1 dt \sum_{t \leq t'} E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p - E\left[\left(\frac{Q_{t'} - Q_{t, R}}{t^*_{t, R}}\right)^p\right]\right]_{t', t}
\]
\[
\leq \frac{C_X}{2} \int_0^1 dt E\left[\left(\left\|Q - E(Q)_{t, t}\right\|\right)^2\right]_{t, t}^{1/2} .
\] (25)

However, proving that the r.h.s. of (25) vanishes is only possible after integrating on a well-chosen set of “perturbations” \(\epsilon\) that play the role of initial conditions in the ODE in Proposition 2. In essence, the integration over \(\epsilon\) smoothes the phase transitions that might appear for particular choices of \(\epsilon\) when \(n\) goes to infinity.

We now describe the set of perturbations on which to integrate. Let \((s_n) \in (0, 1)\) a sequence such that \(s_n\) goes to 0 and \(s_n^{K/(K + 1)}\) diverges to infinity when \(n \rightarrow +\infty\).

Define the following sequence of subsets:
\[
E_n = \left\{ \epsilon \in \mathbb{R}^{K \times K} \left\| \forall \ell \neq \ell' : \epsilon_{\ell, \ell'} = \epsilon_{\ell', \ell} \in [s_n, 2s_n] \right\} \right\}
\]

Those are subsets of symmetric strictly diagonally dominant matrices with positive diagonal entries, hence they are included in \(S^+_K\) (see (20) Corollary 7.2.3). The volume of \(E_n\) is
\[
V_{E_n} = s_n^{K(K + 1)/2} .
\]

Fix \(t \in [0, 1]\). First using the Cauchy-Schwarz inequality, and then making the change of variable \(\epsilon \rightarrow R \equiv R(t, \epsilon)\), which is justified because \(\epsilon \rightarrow R(t, \epsilon)\) is a \(C^1\)-diffeomorphism (see Proposition 2), one obtains
\[
\int_{E_n} d\epsilon E\left[\left\|Q - E(Q)_{t, t}\right\|^2\right]_{t, t}^{1/2}
\]
\[
\leq V_{E_n}^{1/2} \left( \int_{E_n} d\epsilon E\left[\left\|Q - E(Q)_{t, t}\right\|^2\right]_{t, t} \right)^{1/2}
\]
\[
= V_{E_n}^{1/2} \left( \int_{\mathbb{R}^{K \times K}} \frac{dR}{\left|\det J_R(t, \epsilon)\right|} E\left[\left\|Q - E(Q)_{t, R}\right\|^2\right]_{t, R} \right)^{1/2}
\]
\[
\leq V_{E_n}^{1/2} \left( \int_{\mathbb{R}^{K \times K}} dR E\left[\left\|Q - E(Q)_{t, R}\right\|^2\right]_{t, R} \right)^{1/2} ,
\] (26)
To conclude the proof, we integrate the inequality (24) over \( \epsilon \) that vanishes as \( O(\epsilon^2) \) (see [17 Appendix D]) uniformly in

\[ t \in [0, 1] \quad \text{and} \quad R \in \bigcup_{k \geq 1} C(\mathcal{R}_{n,t}). \]

Such concentration of the free entropy is essential to guarantee the concentration of overlap matrices in a Bayesian inference framework. Then, we can adapt the proof of [16] Theorem 3) to show

\[
\int_{C(\mathcal{R}_{n,t})} dR \mathbb{E} \left( \left\| Q - \mathbb{E}[(Q)_{t,R}] \right\|_{t,R}^2 \right) \leq \frac{C_{p,K,P_X}}{s_n t^{1/6}}. \tag{27}
\]

Here \( C_{p,K,P_X} \) is a constant that depends only on \( p, K \) and \( P_X \). Note that the integral over the convex hull \( C(\mathcal{R}_{n,t}) \) is an upper bound on the integral over \( \mathcal{R}_{n,t} \). Combining (25), (26) and (27), one finally obtains:

\[
\left| \int_{\mathcal{E}_n} \frac{d\epsilon}{\mathcal{E}_n} \int_0^1 \frac{dt}{2t} \sum_{t',t} \mathbb{E} \mathbb{E}[(Q_{t't'})^{p-1} - \mathbb{E}[(Q_{t't'})^{p-1}]]_{t',t} \right| \leq C \sqrt{\frac{C_{p,K,P_X}}{s_n^{2/3} t^{1/6}}} = C \sqrt{\frac{C_{p,K,P_X}}{s_n^{2/3} (K+1)^{1/6}}}. \tag{28}
\]

To conclude the proof, we integrate the inequality (24) over \( \epsilon \) and, then, make use of (23) and

\[
\frac{1}{\mathcal{E}_n} \int_{\mathcal{E}_n} d\epsilon \mathbb{E}[(\epsilon)] \leq \mathcal{O}(1) \max_{\epsilon \in \mathcal{E}_n} \| \epsilon \| = \mathcal{O}(1) s_n = \mathcal{O}(n). \]

This gives the inequality

\[
f_n = \frac{1}{\mathcal{E}_n} \int_{\mathcal{E}_n} d\epsilon f_n \leq \sup_{S \in S_K^+} \phi_p(S) + \mathcal{O}(n),
\]

which directly implies the upper bound (7).

V. FUTURE WORK

We leave for future work the extension of both Theorems [1] and [2] to the odd-order case. For Theorem 1 it requires proving that the last summand on the r.h.s. of (15) is nonnegative. When \( K = 1 \), both \( \mathbb{E}[(Q)_{t,R}] \) and \( \mathbb{E}[(Q)_{t,R}] \) are nonnegative so that \( h_p(r,q') \)’s non-negativity for \( r, q \geq 0 \) suffices [8]. However, for \( K > 1 \), we can only say that \( \mathbb{E}[(Q)_{t,R}] \geq 0 \). Regarding Theorem 2 the whole proof directly applies to \( p \) odd if we can show that the divergence (20) is nonnegative, which is more difficult than for \( p \) even. Indeed, while the \( \Delta_{t'} \)’s are still \( \geq 0 \), it is not necessarily the case of \( \mathbb{E}[(Q_{t't'})^{p-2}] \) as \( p = 2 \) is odd.

ACKNOWLEDGMENT

C. L. acknowledges funding from the Swiss National Science Foundation, under grant no 200021-E-175541.