Three proofs of the Makeenko–Migdal equation for Yang–Mills theory on the plane

Bruce K. Driver
University of California, San Diego
Department of Mathematics
La Jolla, CA 92093-0112 USA
bdriver@ucsd.edu

Brian C. Hall*
University of Notre Dame
Department of Mathematics
Notre Dame, IN 46556 USA
bhall@nd.edu

Todd Kemp†
University of California, San Diego
Department of Mathematics
La Jolla, CA 92093-0112 USA
tkemp@math.ucsd.edu

January 23, 2016

Abstract
We give three short proofs of the Makeenko–Migdal equation for the Yang–Mills measure on the plane, two using the edge variables and one using the loop or lasso variables. Our proofs are significantly simpler than the earlier pioneering rigorous proofs given by T. Lévy and by A. Dahlqvist. In particular, our proofs are “local” in nature, in that they involve only derivatives with respect to variables adjacent to the crossing in question.

1 Introduction

The (Euclidean) Yang–Mills field theory describes a random connection on a principal bundle for a compact Lie group $K$, known as the structure group. In two dimensions, the theory is tractable and has been studied extensively. In particular, for Yang–Mills theory on the plane, it is possible to use a gauge fixing to make the measure Gaussian, opening the door to rigorous calculations. This approach was developed simultaneously in two papers: [GKS] by L. Gross, C. King, and A. Sengupta; and [Dr] by B. Driver.

The typical objects of study in the theory are the Wilson loop functionals, given by

$$\mathbb{E}\{\text{trace}(\text{hol}(L))\},$$

(1.1)

where $\mathbb{E}$ denotes the expectation value with respect to the Yang–Mills measure, $\text{hol}(L)$ denotes the holonomy of the random connection around a loop $L$, and the trace is taken in some fixed representation of the structure group $K$. If $L$ is traced out on a graph in the plane, work of Driver [Dr] gives a formula [Dr, Theorem 6.4] for the Wilson loop functional in terms of the heat kernel measure on $K$. (See

---

*Supported in part by NSF grant DMS-1301534
†Supported in part by NSF CAREER award DMS-1254807
One noteworthy feature of the two-dimensional Yang–Mills theory is its invariance under area-preserving diffeomorphisms. This invariance is reflected in Driver’s formula: the expectation (1.1) may be expressed as a function (determined by the topology of the graph) of all the areas of the faces of the graph.

The Makeenko–Migdal equation relates variations of a Wilson loop functional in the neighborhood of a simple crossing to the associated Wilson loops on either side of the crossing, in the case $K = U(N)$. The original equations, in any dimension, were the subject of [MM]. In Section 0.6 of [Lévy2], T. Lévy shows that the so-called “keyboard-type” variation in Eq. (3) of [MM] can be interpreted in the planar case as the alternating sum of derivatives of the Wilson loop functional with respect to the areas of the faces surrounding a simple crossing. Lévy then provides a rigorous proof of the planar Makeenko–Migdal equation, using Driver’s formula. A different proof was subsequently given by A. Dahlqvist in [Dahl]. In this paper, we offer three new, short proofs of the equation.

We use the bi-invariant metric on $U(N)$ whose value on the Lie algebra $u(N) = T_c(U(N))$ is a scaled version of the Hilbert–Schmidt inner product:

$$\langle X, Y \rangle = N \text{trace}(X^*Y).$$

(1.2)

It is then convenient to express the Wilson loop functionals using the normalized trace,

$$\text{tr}(A) := \frac{1}{N} \text{trace}(A).$$

We now consider a loop $L$ with simple crossings, and we let $v$ be one such crossing. We label the four faces of $L$ adjacent to the crossing in cyclic order as $F_1$, $F_2$, $F_3$, and $F_4$, with $F_1$ denoting the face whose boundary contains the two outgoing edges of $L$. We choose the cyclic ordering of the faces so that the first edge traversed by $L$ lies between $F_1$ and $F_4$. We then let $t_1$, $t_2$, $t_3$, and $t_4$ denote the areas of these faces. (See Figure 1) We also let $L_1$ denote the loop from the beginning to the first return to $v$ and let $L_2$ denote the loop from the first return to the end. (See Figure 2) The Makeenko–Migdal equation then gives a formula for the alternating sum of the derivatives of the Wilson loop functional with respect to these areas.
Theorem 1.1 (Makeenko–Migdal equation for $U(N)$). Let $L$ be a closed curve with simple crossings and let $v$ be a crossing. Parameterize $L$ over the time interval $[0, 1]$ and with $L(0) = v$, and let $s_0$ be the unique time with $0 < s_0 < 1$ such that $L(s_0) = v$. Let $L_1$ be the restriction of $L$ to $[0, s_0]$ and let $L_2$ be the restriction of $L$ to $[s_0, 1]$. Then

$$
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \mathbb{E}\{\text{tr}(\text{hol}(L))\} = \mathbb{E}\{\text{tr}(\text{hol}(L_1))\text{tr}(\text{hol}(L_2))\}.
$$

(1.3)

We follow the convention that if any of the adjacent faces is the unbounded face, the corresponding derivative on the left-hand side of (1.3) is omitted. Note also that the faces $F_1, F_2, F_3$, and $F_4$ are not necessarily distinct, so that the same derivative may occur more than once on the left-hand side of (1.3).

The original argument of Makeenko and Migdal for the equation that bears their names was based on heuristic calculations with a path integral and is far from rigorous. Rigorous proofs have been given by Lévy [Lévy2] and Dahlqvist [Dahl]. The goal of the current paper is to provide three short proofs of the result, each of which is substantially simpler than the proofs in [Lévy2] and [Dahl].

The significance of (1.3) is that the two loops $L_1$ and $L_2$ on the right-hand side are simpler than the loop $L$. On the other hand, if one is attempting to compute Wilson loops expectations recursively, the right-hand side of (1.3) cannot be considered as a “known” quantity, because it involves the expectation of the product of traces, rather than the product of the expectations. Thus, Theorem 1.1 is not especially useful in computing Wilson loop expectations for a fixed rank $N$.

Nevertheless, it has been suggested at least since the work of ‘t Hooft [t Hooft] that quantum gauge theories with structure group $U(N)$ simplify in the $N \to \infty$ limit. Specifically, it has been suggested that in this limit, the Euclidean Yang–Mills path-integral concentrates onto a single connection (modulo gauge transformations), known as the master field. In the plane case, the structure of the master field has been described by I. M. Singer [Sing] and R. Gopakumar and D. Gross [GG]. Recently, rigorous analyses of the master field on the plane have been given by M. Anshelevich and A. N. Sengupta [AS] and T. Lévy [Lévy2]. Lévy, in particular, shows in detail that the Wilson loop functionals become deterministic in the large-$N$ limit.

In the large-$N$ limit, then, all variances and covariances go to zero, meaning that there is no difference between an expectation of a product and a product of the expectations. For the master field on the plane, the Makeenko–Migdal equation takes the form

$$
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \tau(\text{hol}(L)) = \tau(\text{hol}(L_1))\tau(\text{hol}(L_2)),
$$

(1.4)
where $\tau(\cdot)$ is the limiting value of $E\{\text{tr}(\cdot)\}$. Lévy shows (Section 9.4 of [Lévy2]) that by using the Makeenko–Migdal equation at each crossing of the loop (along with a simpler relation that we describe in Theorem 2.1), one can recover the derivative of a Wilson loop functional with respect to the area of any one face. This result leads to an effective procedure for computing, recursively, the Wilson loop functionals for the master field.

S. Chatterjee has given a rigorous version of the Makeenko–Migdal equation for lattice gauge theories in any dimension (Theorem 3.6 of [Chatt]). This equation takes a somewhat different form from the two-dimensional continuum result in Theorem 1.1.

We thank Franck Gabriel for comments on and corrections to an earlier version of this article.

2 Two proofs using edge variables

2.1 The set-up

Beginning at least from the work of Migdal, it has been apparent that the heat kernel on $K$ should play a role in the computation of expected traces of holonomies in the Yang–Mills field theory on the plane. (Equation (27) in [Mig], for example, can be understood as the expansion of the heat kernel on $K$ in terms of characters.) A rigorous approach to such computations was developed by L. Gross, C. King, and A. Sengupta in [GKS]. At the same time, B. Driver gave a formula [Dr, Theorem 6.4] for the Yang–Mills measure on a graph $G$ on the plane, under mild restrictions on the nature of the edges involved. Driver’s formula involves the heat kernel $\rho_t$ on the structure group $K$ with respect to a fixed bi-invariant metric. That is to say, $\rho_t$ satisfies the heat equation

$$\frac{\partial \rho_t}{\partial t} = \frac{1}{2} \Delta \rho_t,$$

(2.1)

where $\Delta$ is the Laplacian associated to the given metric, and $\rho_t$ concentrates to a $\delta$-function at the identity as $t$ tends to zero. It will be important in our computations to note that the heat kernel with respect to a bi-invariant metric on a compact Lie group is conjugation invariant:

$$\rho_t(uxu^{-1}) = \rho_t(x), \quad \forall x, u \in K.$$  

(2.2)

(This identity holds because the adjoint action of $K$ on itself is isometric and fixes the origin.)

If $n$ denotes the number of edges in $G$, Driver’s result then says that the expectation value of any gauge-invariant function of the parallel transport along the edges of $G$ may be computed as integration against a measure $\mu$ on $K^n$. To compute $\mu$, we associate an “edge variable” in $K$ to each edge of $G$, which we interpret as the parallel transport of a connection along the edge. Then $\mu$ is given by

$$d\mu = \left( \prod_{F_i} \rho_{|F_i|}(h_i) \right) dx,$$

(2.3)

where the product is over all the bounded faces $F_i$ of the graph, that is, over all the bounded components of the complement of the graph in the plane. Here $dx$ denotes the product of normalized Haar measures in all the edge variables, $|F_i|$ denotes the area of $F_i$, and $h_i$ denotes the “holonomy” around $F_i$, that is, the product of edge variables going around the boundary of $F_i$; in [MM], these discrete holonomies were referred to as plaquettes. (The value of $\rho_{|F_i|}(h_i)$ does not depend on where one starts on the boundary of $F_i$ or on the direction one proceeds.)

In particular, each Wilson loop functional as in (1.1) may be computed by means of a finite-dimensional integral over $K^n$, with respect to the measure $\mu$. 

4
2.2 A simple area-derivative formula

Before coming the Makeenko–Migdal equation itself, we record a simple result that can be proven much more easily. This result is a very special case of Corollary 6.5 of [Lévy2], but in this case, the proof simplifies dramatically. We include a proof here for completeness and to give an indication of the difficulties in computing area derivatives in general. In [Lévy2], Lévy shows that the master field (i.e. the large-$N$ limit of Yang–Mills for $U(N)$) is completely characterized by the limiting Makeenko–Migdal equation (1.4) and the large-$N$ limit of (2.5).

Theorem 2.1. Suppose $f$ is a smooth function of the edge variables associated to a graph $G$ and that $F$ is a bounded face of $G$ that shares an edge $e$ with the unbounded face. Let $a \in K$ denote the edge variable associated to the edge $e$ and let $t$ denote the area of $F$. Then we have

$$\frac{d}{dt} \int f \, d\mu = \frac{1}{2} \int \Delta^a f \, d\mu,$$

(2.4)

where $\Delta^a$ denotes the Laplacian with respect to $a$ with the other edge variables held constant.

In particular, suppose that $K = U(N)$, that $L$ is a loop traced out on $G$ in which the edge $e$ is traversed exactly once, and that $f = \text{tr}(\text{hol}(L))$. Then (2.4) reduces to

$$\frac{d}{dt} \mathbb{E}\{\text{tr}(\text{hol}(L))\} = -\frac{1}{2} \mathbb{E}\{\text{tr}(\text{hol}(L))\}.$$

(2.5)

In this case, if $L$ is a simple closed curve enclosing area $t$, we have $\mathbb{E}\{\text{tr}(\text{hol}(L))\} = e^{-t/2}$.

The key idea in the proof of (2.4) is that because the edge $e$ lies on the boundary of only one bounded face, the edge variable $a$ occurs in only one of the heat kernels in Driver’s formula. By contrast, a generic edge variable lies in two different heat kernels, which is a substantial complicating factor in the proof of the Makeenko–Migdal formula.

Proof. We may choose the orientation of the boundary of $F$ so that it contains the edge $e$ (as opposed to $e^{-1}$) exactly once. (For example, referring to Figure 5 below, we may take $F = F_5$ and $e = e_7$.) It is harmless to assume that $e$ is the first edge traversed, in which case, since parallel transport is order-reversing, the holonomy $h$ around $\partial F$ will have the form

$$h = \alpha a,$$
where \( \alpha \) is a word in edge variables other than \( a \). We then we note that \( (\Delta \rho_t)(h) \) may be computed as

\[
(\Delta \rho_t)(h) = \Delta^a(\rho_t(\alpha a)).
\]

Thus, using Driver’s formula and differentiating under the integral, we obtain

\[
\frac{d}{dt} \int f \, d\mu = \frac{1}{2} \int f [\Delta^a(\rho_t(\alpha a))] \prod_{F_i \neq F} \rho_t(F_i)(h_i).
\]

Now, since \( e \) lies between \( F \) and the unbounded face, the edge variable \( a \) does not occur in any other heat kernel besides \( \rho_t(\alpha a) \). Thus, if we integrate by parts the Laplacian does not hit any other heat kernel, but hits only \( f \), giving (2.4).

Meanwhile, suppose \( K = U(N) \) and \( f \) is the normalized trace of the holonomy of \( L \), where \( L \) traverses \( e \) exactly once. If \( L \) traverses \( e \) in the positive direction, then \( f \) will have the form

\[
f = \text{tr}(\beta a \gamma),
\]

where \( \beta \) and \( \gamma \) are words in edge variables distinct from \( a \). Then

\[
\Delta^a f = \sum_X \text{tr}(\beta a X^2\gamma),
\]

where \( X \) ranges over an orthonormal basis for the Lie algebra \( \mathfrak{k} = u(N) \). But a simple argument (e.g., Proposition 3.1 in [DHK]) shows that if the inner product on \( u(N) \) is normalized as in (1.2) we have

\[
\sum_X X^2 = -I,
\]

in which case, (2.4) reduces to (2.5). If \( L \) traverses \( e \) negatively, the argument is almost identical. Finally, if \( L \) has only one bounded face with area \( t \), Driver’s formula (2.3) tells us that at \( t = 0 \), the holonomy concentrates at the identity, so that the normalized trace of the holonomy is 1.

### 2.3 An abstract Makeenko–Migdal equation

Suppose now that \( G \) is a graph in the plane and \( v \) is a vertex of \( G \) with four incident edges. We assume for now that these edges are distinct; this assumption is removed in Section 4. We label the four (not necessarily distinct) faces of \( G \) adjacent to \( v \) in cyclic order as \( F_1, F_2, F_3, \) and \( F_4 \). We then let \( e_1, e_2, e_3, \) and \( e_4 \) be the outgoing edges at \( v \), labeled so that \( e_1 \) lies between \( F_4 \) and \( F_1 \) and \( e_2 \) lies between \( F_1 \) and \( F_2 \), etc. (See Figure 3.) We also let \( a_i \) denote the edge variable, with values in \( K \), associated to \( e_i \). We write the collection \( x \) of all edge variables in our graph as

\[
x = (a_1, a_2, a_3, a_4, b),
\]

where \( b \) is the tuple of all edge variables other than \( a_1, a_2, a_3, \) and \( a_4 \).

In [Lévy2], Lévy isolates a version of the Makeenko–Migdal equation that is valid for an arbitrary compact structure group \( K \), in which the function does not have to be the trace of a holonomy.

**Definition 2.2.** A function \( f(a_1, a_2, a_3, a_4, b) \) of the edge variables has extended gauge invariance if, for all \( x \in K \),

\[
f(a_1, a_2, a_3, a_4, b) = f(a_1x, a_2, a_3x, a_4, b) = f(a_1, a_2x, a_3, a_4x, b).
\]

(2.6)
Figure 4: Labeling of faces and edges adjacent to $v$

Suppose, for example, that $f$ is the trace of the holonomy around a loop $L$ traced out on $G$, as in Figure 4. We label things so that $L$ first traverses the edge between $F_4$ and $F_1$, namely $e_1$. The first return to $v$ must then be along the only incoming edge of $L$ at $v$ that is not “straight across” from $e_1$, namely $e_4^{-1}$. Thus, $L$ will have the form

$$L = e_1 A e_4^{-1} e_2 B e_3^{-1},$$

where $A$ and $B$ are sequences of edges not belonging to $\{e_1, e_2, e_3, e_4\}$. Since parallel transport is order-reversing, the trace of the holonomy around $L$ is then represented by a function of the form

$$f(a_1, a_2, a_3, a_4, b) := \text{tr}(a_3^{-1} \beta a_2 a_4^{-1} \alpha a_1),$$

(2.7)

where $\alpha$ and $\beta$ are words the $b$ variables. This function is easily seen to have extended gauge invariance.

Definition 2.3. If $f$ is a smooth function on $K$, the left-invariant gradient of $f$, denoted $\nabla f$, is the function with values in the Lie algebra $\mathfrak{k}$ of $K$ given by

$$(\nabla f)(x) = \sum_{X} \left( \frac{d}{ds} f(xe^{sX}) \bigg|_{s=0} \right) X,$$

where the sum is over any orthonormal basis of $\mathfrak{k}$. More generally, if $f$ is a smooth function of the edge variables and $a$ is one of the edge variables, we let $\nabla^a f$ denote the left-invariant gradient of $f$ with respect to $a$ with the other variables fixed.

If $f$ is smooth and has extended gauge invariance, then by differentiating (2.6), we obtain

$$\nabla^a_i f = -\nabla^{a_{i+2}} f,$$

where $i + 2$ is computed mod 4. Since, also, $\nabla^{a_i}$ commutes with $\nabla^{a_j}$, we have

$$\nabla^{a_i} \cdot \nabla^{a_j} f = -\nabla^{a_{i+2}} \cdot \nabla^{a_{j+2}} f = -\nabla^{a_{j+2}} \cdot \nabla^{a_{i+2}} f = \nabla^{a_{j+2}} \cdot \nabla^{a_{i+2}} f,$$

(2.8)

even though $\nabla^{a_j} f$ does not necessarily have extended gauge invariance. (To be clear: if $F$ is a $\mathfrak{k}$-valued smooth function, then

$$\nabla \cdot F(x) = \sum_{X} \left( \frac{d}{ds} F(xe^{sX}) \bigg|_{s=0} \right) X,$$
The derivative $\nabla^a_i F$ is defined similarly, holding variables other than $a_i$ fixed in $F$. Taking $F = \nabla^a_i f$ then gives the definition of the second derivative $\nabla^a_i \cdot \nabla^a_j f$ above.)

We are now ready to state Lévy’s abstract form of the Makeenko–Migdal equation.

**Theorem 2.4** (T. Lévy). Suppose $G$ is a graph in the plane and $v$ is a vertex of $G$ with four distinct edges emanating from $v$. Label the four faces of $G$ adjacent to $v$ in cyclic order as $F_1, \ldots, F_4$ and label the outgoing edges in cyclic order as $e_1, \ldots, e_4$, with $e_1$ lying between $F_4$ and $F_1$. Then if $f$ is a smooth function of the edge variables of $G$ having extended gauge invariance, we have

$$\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int f \, d\mu = - \int \nabla^a_1 \cdot \nabla^a_2 f \, d\mu,$$

where $t_i$ is the area of $F_i$, $i = 1, \ldots, 4$. 

As usual, we set $\partial/\partial t_i$ equal to zero if $F_i$ is the unbounded face. A version of the theorem still holds even if the edges $e_1, \ldots, e_4$ are not distinct; see Section 4.

Theorem 2.4 is a special case of Proposition 6.22 in [Lévy2]. Specifically, since the Yang–Mills measure does not depend on the orientation of the plane, it is harmless to assume that the faces $F_1, F_2, F_3, F_4$ in our labeling scheme occur in counterclockwise order, as in Figure 1. We may take the set $I$ in Lévy’s Proposition 6.22 to be $\{e_1, e_3\}$, as in Figure 25 in [Lévy2]. Then the left-hand side of Proposition 6.22 is actually the negative of the usual alternating sum of area-derivatives. On the right-hand side of Proposition 6.22, meanwhile, there is only one term in the sum, namely $\int \Delta^{e_1:e_2} f \, d\mu$, which corresponds to $\int \nabla^a_1 \cdot \nabla^a_2 f \, d\mu$ in our notation.

Note that since $f$ is assumed to have extended gauge invariance, we have, as in (2.8),

$$\nabla^a_1 \cdot \nabla^a_2 f = -\nabla^a_2 \cdot \nabla^a_3 f = \nabla^a_3 \cdot \nabla^a_4 f = -\nabla^a_4 \cdot \nabla^a_1 f.$$

If we specialize Theorem 2.4 to the case in which $K = U(N)$ and $f$ is as in (2.7), we find that

$$\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int \text{tr}(a_3^{-1}ba_2a_4^{-1}a_1) \, d\mu$$

$$= - \sum_X \int \text{tr}(a_3^{-1}ba_2Xa_4^{-1}a_1) \, d\mu,$$

where the sum is over any orthonormal basis $\{X\}$ for $u(N)$. But an elementary argument (e.g. [DHK Proposition 3.1]) shows that if we normalize the inner product on $u(N)$ as in (1.2), then

$$\sum_X XCX = -\text{tr}(C)I$$

(2.12)

for any $N \times N$ matrix $C$. Thus, (2.11) reduces to

$$\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int \text{tr}(a_3^{-1}ba_2a_4^{-1}a_1) \, d\mu$$

$$= \int \text{tr}(a_4^{-1}a_1f(a_3^{-1}ba_2) \, d\mu,$$

which is—in light of Driver’s formula—just the Makeenko–Migdal equation for $U(N)$, as in Theorem 1.1.

The goal of this section is to give two short proofs of Theorem 2.4. In [Lévy2], Lévy develops a method of differentiating any function with respect to the area $t_i$ of some face $F_i$. Specifically, if $f$ is
any smooth function of the edge variables—which need not have any special invariance property—Lévy shows that

\[ \frac{\partial}{\partial t_i} \int f \, d\mu = \int Df \, d\mu, \]  

(2.13)

where \( D \) is a certain differential operator. (See Corollary 6.5 in \[Lévy2\].) The formula for \( D \) involves the choice of a maximal tree in \( G \) and a sum over a sequence of adjacent faces proceeding from \( F_i \) to the unbounded face. Thus, \( D \) contains, in general, derivatives involving edges far from the vertex in question.

Lévy then specializes his result to the case where \( f \) has extended gauge invariance and takes the alternating sum of derivatives around a vertex. At that point, a substantial cancellation occurs: all derivatives of \( f \) drop out, except for derivatives involving edges coming out of the crossing, and Lévy then obtains the abstract Makeenko–Migdal equation of Theorem 2.4. (See the proof of Proposition 6.22 in \[Lévy2\].)

Our strategy for a simplified proof of Theorem 2.4 is to think that if the cancellation described in the previous paragraph actually occurs, it should be possible to see the cancellation “locally,” that is, in such a way that derivatives involving far away edges never occur in the first place. Of course, Lévy’s formula (2.13) may be useful for various computations, but we do not use it in our proofs of the Makeenko–Migdal equation (1.3).

2.4 Two “local” proofs of the theorem

We consider at first the “generic” case, in which the faces \( F_1, F_2, F_3, \) and \( F_4 \) are distinct and bounded, and the outgoing edges \( e_1, e_2, e_3, \) and \( e_4 \) from \( v \) are distinct. (These assumptions are lifted in Section 4.) In that case, the boundary of \( F_i \) may be represented by a loop of the form

\[ \partial F_i = e_i A_i e_{i+1}^{-1}, \]  

(2.14)

where \( A_i \) is a sequence of edges not belonging to \( \{e_1, e_2, e_3, e_4\} \), where the index \( i \) is understood to be in \( \mathbb{Z}/4 \). Since parallel transport is order reversing, the holonomy \( h_i \) around \( \partial F_i \) is represented by an expression of the form

\[ h_i = a_i^{-1} \alpha_i a_i, \quad i = 1, 2, 3, 4, \]

where \( \alpha_i \) is a word in the \( b \) variables (i.e., the edge variables not belonging to \( \{a_1, a_2, a_3, a_4\} \)). Furthermore, none of the variables \( a_1, a_2, a_3, a_4 \) shows up in any holonomy other than ones associated to \( F_1, F_2, F_3, F_4 \). Thus, the Yang–Mills measure \( \mu \) takes the form

\[ d\mu = \rho_{t_1} (a_2^{-1} \alpha_1 a_1) \rho_{t_2} (a_3^{-1} \alpha_2 a_2) \rho_{t_3} (a_4^{-1} \alpha_3 a_3) \rho_{t_4} (a_1^{-1} \alpha_4 a_4) \nu(b) \, dx, \]  

(2.15)

where \( dx \) is the product of Haar measures in all the edge variables, and \( \nu(b) \) is a product of heat kernels in \( b \) variables.

Our first proof of Theorem 2.4 is a “local” version of Lévy’s proof, proceeding directly by computing the alternating sum of area-derivatives, and integrating by parts twice. Our second proof, which is even shorter, proceeds from the right-hand-side of (2.9) and relies on the decomposition of the density of \( \mu \) into the product of \( (t_1, t_2) \) heat kernels (both independent of edge variable \( a_1 \)) and \( (t_3, t_4) \) heat kernels (both independent of edge variable \( a_2 \)). Both proofs involve only derivatives with respect to the variables \( a_1, a_2, a_3, a_4 \).

Our third proof, based on the loop variables, is in Section 3.
2.4.1 First proof

Our strategy is to differentiate under the integral sign, use the heat equation satisfied by the heat kernel, and then integrate by parts. In this process, we will get “good terms” in which derivatives hit on the function \( f \), and “bad terms” in which derivatives hit on other heat kernels. In each of the two stages of integration by parts, we obtain a cancellation of the bad terms, allowing all of the derivatives to move off of the heat kernels and onto \( f \), at which point we easily obtain the abstract Makeenko–Migdal equation \((2.9)\). Our proof is entirely “local,” in the sense that the derivatives involved are with respect to the variables \( a_1, a_2, a_3, \) and \( a_4 \) only.

Since the heat kernel on \( K \) is invariant under conjugation, we can compute \((\Delta \rho_t)(a_i^{-1} \alpha_i a_i)\) by various combinations of derivatives with respect to \( a_i \) and derivatives with respect to \( a_{i+1} \). It turns out that the most convenient way to do the computation is as follows:

\[
(\Delta \rho_t)(a_i^{-1} \alpha_i a_i) = \frac{1}{4} (\nabla a_i - \nabla a_{i+1})^2 \left[ \rho_t(a_i^{-1} \alpha a_i) \right] = \frac{1}{4} \sum_x \left( \frac{\partial}{\partial s} - \frac{\partial}{\partial t} \right)^2 \rho_t(e^{sx}a_i^{-1} \alpha a_i e^{-tx}) \bigg|_{s=t=0}.
\]

Here, \( \nabla a_i \) is the left-invariant gradient of Definition 2.3. Thus, if we differentiate with respect to \( t_i \), we get

\[
\frac{\partial}{\partial t_i} \int f \, d\mu = \frac{1}{8} \int f[(\nabla a_i - \nabla a_{i+1})^2 \rho_t] \rho_{t_1} \cdots \hat{\rho}_{t_i} \cdots \rho_{t_4} \nu \, dx,
\]

where \( \hat{\rho}_{t_i} \) indicates that the given heat kernel is omitted. Here and below, we frequently omit the arguments of the heat kernels, since they are always the same.

If we now integrate by parts a first time, we get some terms where \( \nabla a_i - \nabla a_{i+1} \) hits on \( f \) and some terms where \( \nabla a_i - \nabla a_{i+2} \) hits another heat kernel. But the only heat kernels besides \( \rho_t \) that contain the variables \( a_i \) or \( a_{i+1} \) are \( \rho_{t_{i-1}}(a_i^{-1} \alpha a_{i-1} a_{i-1}) \) and \( \rho_{t_{i+1}}(a_{i+1}^{-1} \alpha a_{i+1} a_{i+1}) \). Furthermore, using the conjugation invariance of \( \rho_t \) (cf. \((2.2)\)), we find that

\[
(\nabla a_i - \nabla a_{i+1})[\rho_{t_{i-1}}(a_i^{-1} \alpha a_{i-1})] = -(\nabla \rho_{t_{i-1}})(a_i^{-1} \alpha a_{i-1})
\]

\[
(\nabla a_i - \nabla a_{i+1})[\rho_{t_{i+1}}(a_{i+1}^{-1} \alpha a_{i+1})] = -(\nabla \rho_{t_{i+1}})(a_{i+1}^{-1} \alpha a_{i+1}).
\]

Each time we encounter a gradient of a heat kernel, we multiply and divide by \( \rho_t \) and use the identity

\[
\frac{\nabla \rho_t}{\rho_t} = \nabla \log \rho_t
\]

to write the answer as an integral against the Yang–Mills measure, giving

\[
\frac{\partial}{\partial t_i} \int f \, d\mu = -\frac{1}{8} \int [(\nabla a_i - \nabla a_{i+1}) f] \cdot [(\nabla a_i - \nabla a_{i+1}) \log \rho_t] \, d\mu
\]

\[
+ \frac{1}{4} \int f \nabla \log \rho_{t_i} \cdot \nabla \log \rho_{t_{i-1}} + \nabla \log \rho_{t_i} \cdot \nabla \log \rho_{t_{i+1}} \, d\mu.
\]

Upon taking the alternating sum, each term involving a product of two log-gradients of heat kernels will occur twice with opposite signs. Thus, even without assuming extended gauge invariance, we have a significant cancellation, yielding

\[
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int f \, d\mu
\]

\[
= -\frac{1}{8} \sum_i (-1)^{i+1} \int [(\nabla a_i - \nabla a_{i+1}) f] \cdot [(\nabla a_i - \nabla a_{i+1}) \rho_{t_i}] \rho_{t_1} \cdots \hat{\rho}_{t_i} \cdots \rho_{t_4} \nu \, dx.
\]
We now integrate by parts a second time, pushing the remaining derivatives off of $\rho_{t_i}$ and onto all the other functions involved. By the same argument as in the first integration by parts, this gives

\[
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int f \, d\mu = \frac{1}{8} \sum_i (-1)^i \int (\nabla^{a_i} - \nabla^{a_{i+1}})^2 f \, d\mu
\]

\[-\frac{1}{8} \sum_i (-1)^i \int \left[ (\nabla^{a_i} - \nabla^{a_{i+1}}) f(a) \cdot \left[ \nabla \log \rho_{t_{i-1}} + \nabla \log \rho_{t_{i+1}} \right] \right] d\mu.
\]

Now, since $f$ has extended gauge invariance, we have

\[
(\nabla^{a_i} - \nabla^{a_{i+1}}) f = -(\nabla^{a_{i+2}} - \nabla^{a_{i+3}}) f.
\]

Meanwhile, if we replace $i$ by $i + 2$, then $\nabla \rho_{t_{i-1}}$ becomes $\nabla \rho_{t_{i+1}}$ and $\nabla \rho_{t_{i+1}}$ becomes $\nabla \rho_{t_{i-1}}$, since $i$ is understood to be in $\mathbb{Z}/4$. Thus, all the derivatives of heat kernels cancel ($i = 1$ with $i = 3$ and $i = 2$ with $i = 4$), leaving us with

\[
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int f \, d\mu = \frac{1}{8} \sum_i (-1)^i \int (\nabla^{a_i} - \nabla^{a_{i+1}})^2 f \, d\mu.
\]  

(2.17)

We now note that

\[
(\nabla^{a_i} - \nabla^{a_{i+1}})^2 f = \nabla^{a_i} \cdot \nabla^{a_i} f - 2\nabla^{a_i} \cdot \nabla^{a_{i+1}} f + \nabla^{a_{i+1}} \cdot \nabla^{a_{i+1}} f.
\]

All the “Laplacian” terms will cancel in the alternating sum, so that (2.17) becomes

\[
\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int f \, d\mu = -\frac{1}{4} \sum_i \int (-1)^i \nabla^{a_i} \cdot \nabla^{a_{i+1}} f \, d\mu
\]

\[= -\int \nabla^{a_1} \cdot \nabla^{a_2} f \, d\mu,
\]

where we have used (2.10) in the second equality.

### 2.4.2 Second proof

In our second proof, which is likely to be about as short as possible, we begin by writing the density of $\mu$ as a product of two terms: those corresponding to $(t_1, t_2)$ and those to $(t_3, t_4)$:

\[
R_{12} = \rho_{t_1} \rho_{t_2}, \quad R_{34} = \rho_{t_3} \rho_{t_4}
\]

where, as above, we suppress the explicit variable dependences. For clarification, $R_{12}$ depends on $a_1, a_2, a_3$, while $R_{34}$ depends on $a_1, a_3, a_4$. Then

\[
d\mu = R_{12} R_{34} \nu \, dx,
\]

where $\nu$ is a product of heat kernels in variables not belonging to $\{a_1, a_2, a_3, a_4\}$ and $dx$ is the product of the Haar measures in the edge variables. For the remainder of the proof, we write integrals of functions $g$ against $dx$ simply as $\int g$.

Now, using (2.10), taking care to commute partial derivatives, we may write

\[
\nabla^{a_1} \cdot \nabla^{a_2} f = \frac{1}{2} (\nabla^{a_1} - \nabla^{a_3}) \cdot \nabla^{a_2} f.
\]
Then we integrate by parts once, and use the product rule.

\[-\int \nabla^{a_1} \cdot \nabla^{a_2} f \, d\mu = -\frac{1}{2} \int [(\nabla^{a_1} - \nabla^{a_3}) \cdot (\nabla^{a_2} f)] R_{12} R_{34} \nu = -\frac{1}{2} \int [\nabla^{a_2} f \cdot (\nabla^{a_1} - \nabla^{a_3})(R_{12} R_{34})] \nu = -\frac{1}{2} \int R_{34} [\nabla^{a_2} f \cdot (\nabla^{a_1} - \nabla^{a_3})(R_{12})] \nu + \frac{1}{2} \int R_{12} [\nabla^{a_2} f : (\nabla^{a_1} - \nabla^{a_3})(R_{34})] \nu.\]

We now use extended gauge invariance once more, in the second term, writing \(\nabla^{a_2} f = -\nabla^{a_4} f\), yielding

\[\frac{1}{2} \int R_{34} [\nabla^{a_2} f \cdot (\nabla^{a_1} - \nabla^{a_3})(R_{12})] \nu = -\frac{1}{2} \int R_{12} [\nabla^{a_4} f \cdot (\nabla^{a_1} - \nabla^{a_3})(R_{34})] \nu.\]

Since \(R_{12}\) does not depend on \(a_4\), and \(R_{34}\) does not depend on \(a_2\), we can integrate this by parts once more, and the \(\nabla^{a_2}\) and \(\nabla^{a_4}\) derivatives only hit the already differentiated factors. Thus,

\[-\int \nabla^{a_1} \cdot \nabla^{a_2} f \, d\mu = -\frac{1}{2} \int f R_{34} [\nabla^{a_2} \cdot (\nabla^{a_1} - \nabla^{a_3}) R_{12}] \nu + \frac{1}{2} \int f R_{12} [\nabla^{a_4} \cdot (\nabla^{a_1} - \nabla^{a_3}) R_{34}] \nu. \tag{2.18}\]

Finally, we compute the second derivatives. Recalling that \(R_{12} = \rho_t \rho_t\) and recalling the arguments of the heat kernels from (2.15), we have

\[(\nabla^{a_1} - \nabla^{a_3}) R_{12} = (\nabla^{a_1} - \nabla^{a_3})(\rho_t (a_2^{-1} a_1 a_1) \rho_t (a_3^{-1} a_2 a_2)) = \rho_t (a_3^{-1} a_2 a_2)(\nabla_t \rho_t)(a_2^{-1} a_1 a_1) + \rho_t (a_2^{-1} a_1 a_1)(\nabla_t \rho_t)(a_3^{-1} a_2 a_2).\]

Applying \(\nabla^{a_2}\) then yields

\[\nabla^{a_2} \cdot (\nabla^{a_1} - \nabla^{a_3}) R_{12} = \nabla_t \rho_t \cdot \nabla \rho_t - \rho_t \Delta \rho_t - \nabla \rho_t \cdot \nabla t \rho_t + \rho_t \Delta t \rho_t.\]

The first and third terms cancel, and we see that the first term on the right-hand side of (2.18) is equal to

\[-\frac{1}{2} \int f R_{34} [\nabla^{a_2} \cdot (\nabla^{a_1} - \nabla^{a_3}) R_{12}] \nu = -\frac{1}{2} \int f \rho_t \rho_t (-\rho_t \Delta \rho_t + \rho_t \Delta t \rho_t) = \left(\frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2}\right) \int f \, d\mu,\]

where we have used the heat equation (2.1) in the second equality.

An entirely analogous computation shows that the second term on the right-hand side of (2.18) is equal to \((\frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4}) \int f \, d\mu\), and adding these up gives the left-hand-side of Theorem 2.4, concluding the proof.
3 A proof using loop variables

In Section 7 of [Dahl], Dahlqvist gave a proof of the Makeenko–Migdal equation for $U(N)$ (Theorem 1.1) using “loop” or “lasso” variables. This proof stands in contrast to the proof in [Lévy2] of the abstract Makeenko–Migdal equation (which implies Theorem 1.1) using “edge” variables. Like Lévy’s proof using edge variables, Dahlqvist’s proof is based on a formula for the derivative with respect to an individual time variable: a formula which contains a large number of terms that must cancel upon taking the alternating sum. As in our proof using edge variables, we will work with the alternating sum from the beginning and obtain the necessary cancellations without ever encountering all the terms arising in [Dahl]. We actually prove an abstract form of Makeenko–Migdal, similar to Theorem 2.4 based on loop variables. This result contains, as a special case, the Makeenko–Migdal equation for $U(N)$.

3.1 The loop variables

It is well known that the fundamental group of any graph is free. Given a planar graph $G$, a fixed vertex $v$ of $G$, and a maximal tree $T$ in $G$, Lévy gives a particular set of free generators for $\pi_1(G)$, which are in one-to-one correspondence with the bounded faces of $G$. We refer to these generators as loops or lassos. Each generator is obtained by starting at $v$, traveling along a certain path $p$ in $T$, then around the boundary of a particular bounded face, then back to $v$ along the inverse of $p$. For the details of this construction, we refer to Section 4.3 of [Lévy2]. (See especially Proposition 4.2.)

We now introduce the loop variables, which are simply the products of the edge variables associated to the edges in the just-defined loops. The loop variables are almost the same as the holonomy variables $h_i$ entering into Driver’s formula (2.3), except that they contain a “tail” representing the path $p$ in the previous paragraph. (Since $\rho_t$ is conjugation invariant, the tail may be omitted from the heat kernel.)

We choose as our basepoint $v$ the crossing involved in the Makeenko–Migdal equation. If $F_1$, $F_2$, $F_3$, and $F_4$ are the four faces surrounding $v$, then the associated loops just traverse the boundary of each face starting from $v$. Thus, as in (2.14), we have

$$L_i = \partial F_i = e_i A_i e_i^{-1}, \quad i = 1, 2, 3, 4,$$

where $A_i$ is a word in edges not belonging to $\{e_1, e_2, e_3, e_4\}$. Since parallel transport is order-reversing, the corresponding loop variables (with values in $K$) will have the form

$$\ell_i = a_i^{-1} \alpha_i a_i, \quad i = 1, 2, 3, 4,$$

(3.1)

where $\alpha_i$ is a word in the $b$ variables, that is, the edge variables not belonging to $\{a_1, a_2, a_3, a_4\}$. (See Figure 5) Properly speaking, (3.1) applies only when $F_i$ is a bounded face; if $F_i$ is unbounded, there is no loop variable associated to $F_i$.

Meanwhile, for the loop $L_j$ associated to any bounded face other than $F_1$, $F_2$, $F_3$, and $F_4$, the associated loop will have the form $L_j = e_{i_j} B_j e_{i_j}^{-1}$, where $e_{i_j} \in \{e_1, e_2, e_3, e_4\}$ is the first edge traversed by $L_j$ and where $B_j$ is a word in edge variable not in $\{e_1, e_2, e_3, e_4\}$. Thus, the corresponding loop variable will have the form

$$\ell_j = a_{i_j}^{-1} \beta_j a_{i_j}, \quad j \geq 5,$$

(3.2)

where $\beta$ is a word in the $b$ variables.

If $n$ is the number of edges and $m$ is the number of bounded faces, the assignments in (3.1) and (3.2) define a smooth map $\Psi : K^n \to K^m$, sending the edge variables to the loop variables. Typically,
there will be fewer loop variables than edge variables. Thus, not every function of the edge variables (whether or not the function has extended gauge invariance) will be expressible as a function of the loop variables. On the other hand, since the loop variables generate the fundamental group of $G$, the trace of the holonomy around any loop in $G$ will be expressible as a function of the loop variables. More generally, according to Lemma 2.1.5 of [Lévy1], every gauge-invariant function on $K^n$ can be expressed as a function of the loop variables.

Meanwhile, using Driver’s formula for the Yang–Mills measure for $G$, it is not difficult to show that the loop variables are independent and heat kernel distributed. (Compare Proposition 4.4 in [Lévy2].) That is to say: the push-forward of the measure $\mu$ under the map $\Psi$ is simply the product of heat kernel measures with time parameters equal to the areas of the bounded faces. If $\tilde{\mu}$ refers to this pushed forward measure, the measure-theoretic change of variables theorem says that if $f = g \circ \Psi$, then

$$\int_{K^n} f \, d\mu = \int_{K^n} g \, d\tilde{\mu}.$$  \hfill (3.3)

14
We now consider how changes in the four “adjacent” edge variables affect the loop variables. If we change, say, $a_1$ to $a_1x$, we can read off the corresponding change in the adjacent loop variables (as in (3.1)) as

$$(\ell_1, \ell_2, \ell_3, \ell_4) \mapsto (\ell_1 x, \ell_2, \ell_3, x^{-1} \ell_4).$$

Meanwhile, each nonadjacent loop variable $\ell_j$ with $j \geq 5$ (as in (3.2)) will either be conjugated by $x$ or unchanged, depending on whether $L_j$ goes out along $e_1$ or along $e_2, e_3$, or $e_4$. Similar transformation rules hold for changes in $a_2, a_3,$ and $a_4$.

In particular, if we make the substitution

$$(a_1, a_2, a_3, a_4) \mapsto (a_1x, a_2y, a_3x, a_4y)$$

we have the following substitutions for the adjacent loop variables:

$$(\ell_1, \ell_2, \ell_3, \ell_4) \mapsto (y^{-1} \ell_1 x, x^{-1} \ell_2 y, y^{-1} \ell_3 x, x^{-1} \ell_4 y),$$

whereas each $\ell_j$ with $j \geq 5$ loop changes either as

$$\ell_j \mapsto x^{-1} \ell_j x$$

or as

$$\ell_j \mapsto y^{-1} \ell_j y$$

depending on the first outgoing edge traversed by the loop $L_j$.

The above calculation motivates the following definition.

**Definition 3.1.** We say that a function on $K^m$ has extended gauge invariance if it is invariant under every transformation of the sort in (3.4), (3.5), and (3.6).

Since changes in the variables $a_1, a_2, a_3,$ and $a_4$ translate into simple changes in the loop variables, we can translate the differential operators $\nabla^a_1 \cdot \nabla^a_{i+1}$ on $K^n$ into differential operators on $K^m$. In a slight abuse of notation, we will continue to refer to the operators on $K^m$ as $\nabla^a_1 \cdot \nabla^a_{i+1}$.

**Definition 3.2.** Suppose $f : K^m \to \mathbb{C}$ is a smooth function of the loop variables. Then $\nabla^a_1 \cdot \nabla^a_j f : K^m \to \mathbb{C}$, with $i \neq j$ in $\{1, 2, 3, 4\}$, is the function computed as follows. Let $\ell = (\ell_i)_{i=1}^m$ be the loop variables defined in (3.1) and (3.2). Define a parametrized surface $\ell(s, t)$ in $K^m$ by replacing $a_i$ and $a_j$ with $a_i \mapsto a_i e^{tX}$ and $a_j \mapsto a_j e^{tX}$ in those two equations. We then set

$$(\nabla^a_1 \cdot \nabla^a_j f)(\ell) = \sum_X \left. \frac{\partial^2}{\partial s \partial t} f(\ell(s, t)) \right|_{s=t=0}.$$

We will compute these operators in the next subsection. We are now ready to state the abstract Makeenko–Migdal equation for the loop variables.

**Theorem 3.3.** Let $m$ be the number of loop variables and suppose $g : K^m \to \mathbb{C}$ is a smooth function having extended gauge invariance in the sense of Definition 3.1. Then

$$\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int_{K^m} g \, d\mu = - \int_{K^m} \nabla^a_1 \cdot \nabla^a_2 g \, d\mu.$$

In light of (3.3), Theorem 3.3 implies Theorem 2.4 for any function $f$ on $K^n$ that has extended gauge invariance and that can be expressed in the form $f = g \circ \Psi$ for some function $g$ on $K^m$. In particular, Theorem 3.3 implies the Makeenko–Migdal equation for $U(N)$, as in Theorem 1.1.
3.2 The proof

The advantage of working in the loop variables is that since each heat kernel is evaluated on a separate loop variable, when we differentiate and integrate by parts, none of the derivatives hits on any other heat kernel, but only the function $f$. Thus,

$$\left( \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} + \frac{\partial}{\partial t_3} - \frac{\partial}{\partial t_4} \right) \int_{K^m} f \, d\tilde{\mu} = \frac{1}{2} \int_{K^m} \left( \Delta_{t_1} - \Delta_{t_2} + \Delta_{t_3} - \Delta_{t_4} \right) f \, d\tilde{\mu}. \quad (3.7)$$

On the other hand, even if $f$ has extended gauge invariance, the integrand on the right-hand side of (3.7) will contain many other terms besides the one we want. We will have to show that these unwanted terms cancel out after integration.

We now consider the right-hand side of the Makeenko–Migdal equation in Theorem 3.3. A key point will be to exploit the invariance of the measure $\tilde{\mu}$ under conjugation in each variable. We consider both a left-invariant gradient $\nabla^L$ and right-invariant gradient $\nabla^R$ in the loop variables (similar to Definition 2.3):

$$(\nabla^L f)(\ell) = \sum_X \left( \frac{d}{ds} f(e^{sX}) \right|_{s=0} ) X,$$

$$(\nabla^R f)(\ell) = \sum_X \left( \frac{d}{ds} f(e^{sX} \ell) \right|_{s=0} ) X.$$

**Lemma 3.4.** If $g$ is any smooth function on $K^m$, we have

$$\int_{K^m} \nabla^{\ell_j,L} \cdot \nabla^{\ell_k,L} g \, d\tilde{\mu} = \int_{K^m} \nabla^{\ell_j,L} \cdot \nabla^{\ell_k,R} g \, d\tilde{\mu} = \int_{K^m} \nabla^{\ell_j,R} \cdot \nabla^{\ell_k,L} g \, d\tilde{\mu} = \int_{K^m} \nabla^{\ell_j,R} \cdot \nabla^{\ell_k,R} g \, d\tilde{\mu},$$

where $\nabla^{\ell_j,L}$ and $\nabla^{\ell_j,R}$ denote the left-invariant and right-invariant gradients in the variable $\ell_j$, respectively.

**Proof.** The invariance of each heat kernel under conjugation tells us that for any $j$, we have

$$\int_{K^m} g(\ell_1, \ldots, \ell_j x, \ldots, \ell_m) \, d\tilde{\mu} = \int_{K^m} g(\ell_1, \ldots, x \ell_j, \ldots, \ell_m) \, d\tilde{\mu}.$$

If follows that

$$\frac{\partial^2}{\partial s \partial t} \left. \int_{K^m} g(\ell_1, \ldots, \ell_j e^{tX}, \ldots, \ell_k e^{sX}, \ldots, \ell_m) \, d\tilde{\mu} \right|_{s=t=0} = \frac{\partial^2}{\partial s \partial t} \left. \int_{K^m} g(\ell_1, \ldots, e^{tX} \ell_j, \ldots, e^{sX} \ell_k, \ldots, \ell_m) \, d\tilde{\mu} \right|_{s=t=0},$$

from which it follows that $\int_{K^m} \nabla^{\ell_j,L} \cdot \nabla^{\ell_k,L} g \, d\tilde{\mu}$ coincides with $\int_{K^m} \nabla^{\ell_j,R} \cdot \nabla^{\ell_k,R} g \, d\tilde{\mu}$. (The argument works equally well whether $j = k$ or $j \neq k$.) All the other claimed equalities follow by a similar argument. 

\[\Box\]
In what follows, we will make repeated use, usually without mention, of Lemma 3.4. It is convenient, in this context, to use the notation
\[ f \equiv g \]
to indicate that \( f \) and \( g \) have the same integral.

If we make the substitutions \( a_1 \mapsto a_1 e^{sX} \) and \( a_2 \mapsto a_2 e^{tX} \), the loop variables change in some computable way. Recall under such substitutions, each \( \ell_j \) with \( j \geq 5 \) merely gets conjugated (or not changed at all). Let \( m \) denote the tuple of variables \( \ell_5, \ldots, \ell_m \) and let \( m' \) denote the new value of these variables after changing \( a_1 \) and \( a_2 \) as above. Then, by the conjugation invariance of the measure (cf. (2.2)), we have
\[
\int_{K^m} \nabla^{a_1} \cdot \nabla^{a_2} g \, d\tilde{\mu} = \sum_X \frac{\partial^2}{\partial s \partial t} \int_{K^m} g(e^{-tX} \ell_1 e^{sX}, \ell_2 e^{tX}, \ell_3, e^{-sX} \ell_4, m) \, d\tilde{\mu} \bigg|_{s=t=0} = \sum_X \frac{\partial^2}{\partial s \partial t} \int_{K^m} g(e^{(s-t)X} \ell_1, e^{tX} \ell_2, \ell_3, e^{-sX} \ell_4, m) \, d\tilde{\mu} \bigg|_{s=t=0} = \int_{K^m} (\Delta \ell_1 + \nabla \ell_1 \cdot \nabla \ell_2 + \nabla \ell_1 \cdot \nabla \ell_4 - \nabla \ell_2 \cdot \nabla \ell_4) g \, d\tilde{\mu}.
\]
(Recall that \( m' \) differs from \( m \) only by conjugations in some of the variables, which does not affect the value of the integral.) Note that in light of Lemma 3.4, we do not have to specify whether the gradients are left-invariant or right-invariant.

After integrating and using the conjugation invariance of the measure, we are left with a “local” formula for \( \int_{K^m} \nabla^{a_1} \cdot \nabla^{a_2} g \, d\mu \), that is, one in which only derivatives in the variables \( \ell_1, \ell_2, \ell_3, \) and \( \ell_4 \) enter. Since (3.7) is also local in this sense, there is no need to consider derivatives in any variables not belonging to \( \{\ell_1, \ell_2, \ell_3, \ell_4\} \).

Using similar calculations for the other pairs of cyclically adjacent variables, we have
\[
\nabla^{a_1} \cdot \nabla^{a_2} g \equiv (\Delta \ell_1 + \nabla \ell_1 \cdot \nabla \ell_2 + \nabla \ell_1 \cdot \nabla \ell_4 - \nabla \ell_2 \cdot \nabla \ell_4) g
\]
\[
\nabla^{a_2} \cdot \nabla^{a_3} g \equiv (\Delta \ell_2 + \nabla \ell_2 \cdot \nabla \ell_3 + \nabla \ell_2 \cdot \nabla \ell_4 - \nabla \ell_3 \cdot \nabla \ell_4) g
\]
\[
\nabla^{a_3} \cdot \nabla^{a_4} g \equiv (\Delta \ell_3 + \nabla \ell_3 \cdot \nabla \ell_4 + \nabla \ell_3 \cdot \nabla \ell_4 - \nabla \ell_4 \cdot \nabla \ell_4) g
\]
\[
\nabla^{a_4} \cdot \nabla^{a_1} g \equiv (\Delta \ell_4 + \nabla \ell_4 \cdot \nabla \ell_1 + \nabla \ell_4 \cdot \nabla \ell_3 - \nabla \ell_1 \cdot \nabla \ell_3) g.
\]

Now, if \( g \) has extended gauge invariance, each of these terms reduces to one of \( \pm \nabla^{a_1} \cdot \nabla^{a_2} g \). Hence, we may take an alternating sum and divide by 4 to obtain
\[
\nabla^{a_1} \cdot \nabla^{a_2} g \equiv \left( -\frac{1}{2} \Delta \ell_1 + \frac{1}{2} \Delta \ell_2 - \frac{1}{2} \Delta \ell_3 + \frac{1}{2} \Delta \ell_4 \right) g
\]
\[
+ \left( \frac{1}{4} \Delta \ell_1 - \frac{1}{4} \Delta \ell_2 + \frac{1}{4} \Delta \ell_3 - \frac{1}{4} \Delta \ell_4 \right) g
\]
\[
+ \frac{1}{2} \left( \nabla \ell_1 \cdot \nabla \ell_3 - \nabla \ell_2 \cdot \nabla \ell_4 \right) g,
\]
where we have written the “correct” Laplacian term (as in (3.7)) on the first line. To establish the Makeenko–Migdal equation, we need to prove that the last two lines disappear after integration:
\[
\int_{K^m} \left( [\Delta \ell_1 - \Delta \ell_2 + \Delta \ell_3 - \Delta \ell_4 + 2\nabla \ell_1 \cdot \nabla \ell_3 - 2\nabla \ell_2 \cdot \nabla \ell_4] g \right) \, d\tilde{\mu} = 0, \quad (3.8)
\]
whenever \( g \) has extended gauge invariance.
To establish (3.8), we recall that extended gauge invariance means invariance under the transformations in (3.4), (3.5), and (3.6). Applying these transformations with $x = e^{tX}$ and $y = e$ and differentiating shows that

$$0 = (\nabla_{\ell_1} - \nabla_{\ell_2} + \nabla_{\ell_3} - \nabla_{\ell_4} g) + \sum_{j \in I} (\nabla_{\ell_j} - \nabla_{\ell_j}) g,$$

where $I$ refers to the set of indices $j \geq 5$ for which the loop goes out from the basepoint along $a_1$ or $a_3$. We now apply the operator $\nabla_{\ell_1} + \nabla_{\ell_2} + \nabla_{\ell_3} + \nabla_{\ell_4}$ to both sides of (3.9), integrate against $\tilde{\mu}$, and use Lemma 3.4. All terms involving derivatives with respect to $\ell_j$, $j \geq 5$, will drop out, and we do not have to specify whether the remaining derivatives are left-invariant or right-invariant, giving

$$\int_{K^m} \left[ (\nabla_{\ell_1} + \nabla_{\ell_2} + \nabla_{\ell_3} + \nabla_{\ell_4})(\nabla_{\ell_1} - \nabla_{\ell_2} + \nabla_{\ell_3} - \nabla_{\ell_4}) g \right] d\tilde{\mu} = 0. \quad (3.10)$$

If we expand out the product on the left-hand side of (3.10), we find that products of derivatives on cyclically adjacent variables (e.g., $\nabla_{\ell_1} \cdot \nabla_{\ell_2}$ or $\nabla_{\ell_4} \cdot \nabla_{\ell_1}$) cancel, while products of derivatives on “opposite” variables (i.e., $\nabla_{\ell_1} \cdot \nabla_{\ell_3}$ and $\nabla_{\ell_2} \cdot \nabla_{\ell_4}$) combine. Thus, (3.10) is precisely equivalent to the desired identity (3.8).

3.3 An example

We now illustrate the preceding proof of the Makeenko–Migdal equation for the loop in Figure 7. We begin by identifying the four generating loops of this graph, each of which runs counter-clockwise around one of the bounded faces of the graph, as in Figure 8. It is straightforward to check that the loop in Figure 7 decomposes as

$$(L_1 L_2 L_3)(L_1^{-1} L_4^{-1} L_3^{-1}),$$

where the notation means that we first traverse $L_1$, then $L_2$, and so on. The expressions in parentheses indicate the component loops in the $U(N)$ version of the Makeenko–Migdal equation. (We will carry out the calculation for a general compact group $K$ and then indicate what happens when $K = U(N)$.) Since parallel transport is order-reversing, the holonomy around $L$ is expressed in terms of the loop variables as

$$\text{hol}(L) = \ell_3^{-1} \ell_4^{-1} \ell_1^{-1} \ell_3 \ell_2 \ell_1.$$
We start by computing $\nabla^{a_1} \cdot \nabla^{a_2} f$, with

$$f(\ell_1, \ell_2, \ell_3, \ell_4) = \text{tr}(\ell_3^{-1} \ell_4^{-1} \ell_1^{-1} \ell_2 \ell_1),$$

with $\text{tr}$ denoting the normalized trace in some representation of $K$. Recalling (3.1), we find that the substitutions $a_1 \mapsto a_1 e^{sX}$ and $a_2 \mapsto a_2 e^{tX}$ in the edge variables translates into the substitutions

$$(\ell_1, \ell_2, \ell_3, \ell_4) \mapsto (e^{-tX} \ell_1 e^{sX}, \ell_2 e^{tX}, \ell_3, e^{-sX} \ell_4)$$

in the loop variables. Thus,

$$-\nabla^{a_1} \cdot \nabla^{a_2} f = -\sum_X \frac{\partial^2}{\partial s \partial t} \text{tr}[\ell_3^{-1} (\ell_4^{-1} e^{sX}) (e^{-sX} \ell_1^{-1} e^{tX}) \ell_3 (\ell_2 e^{tX}) (e^{-tX} \ell_1 e^{sX})]_{s=t=0}$$

$$=-\sum_X \frac{\partial^2}{\partial s \partial t} \text{tr}[\ell_3^{-1} \ell_4^{-1} e^{tX} \ell_3 \ell_2 \ell_1 e^{sX}]_{s=t=0}$$

$$=-\sum_X \text{tr}[\ell_3^{-1} \ell_4^{-1} e^{tX} X \ell_3 \ell_2 \ell_1 X],$$

where in each term we sum $X$ over an orthonormal basis of the Lie algebra $\mathfrak{k}$ of $K$. In the $U(N)$ case, the last line of (3.11) simplifies to $\text{tr}(\ell_3^{-1} \ell_4^{-1} \ell_1^{-1} \ell_2) \text{tr}(\ell_3 \ell_2 \ell_1)$, by (2.12).

We now compute the alternating sum of time derivatives of $\int f \, d\mu$, using (3.7). By Lemma 3.4, we are free to evaluate the Laplacians using any combination of derivatives on the left and on the right. The computations work out most simply if we compute each Laplacian as a product of a gradient on the left and a gradient on the right. With this convention, we easily obtain

$$\Delta_{\ell_1} f = \sum_X (\text{tr}(\ell_3^{-1} \ell_4^{-1} X \ell_1^{-1} \ell_3 \ell_2 \ell_1) - \text{tr}(\ell_3^{-1} \ell_4^{-1} X \ell_1^{-1} \ell_3 \ell_2 X \ell_1)$$

$$- \text{tr}(\ell_3^{-1} \ell_4^{-1} \ell_1^{-1} X \ell_3 \ell_2 \ell_1 X) + \text{tr}(\ell_3^{-1} \ell_4^{-1} \ell_1^{-1} \ell_3 \ell_2 X \ell_1 X))$$

$$\Delta_{\ell_2} f = \sum_X \text{tr}(\ell_3^{-1} \ell_4^{-1} \ell_1^{-1} \ell_3 X \ell_2 X \ell_1)$$
\[ \Delta_{\ell_3} f = \sum_X \left( \text{tr}(X\ell_3^{-1}X\ell_4^{-1}\ell_1^{-1}\ell_3\ell_2\ell_1) - \text{tr}(X\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1) \right) \]
\[ \quad - \text{tr}(\ell_3^{-1}X\ell_4^{-1}\ell_1^{-1}\ell_3\ell_2\ell_1) + \text{tr}(\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1)) \]
\[ \Delta_{\ell_4} f = \sum_X \text{tr}(\ell_3^{-1}X\ell_4^{-1}X\ell_1^{-1}\ell_3\ell_2\ell_1), \]

where in each term, we sum \( X \) over an orthonormal basis for \( \ell \).

After taking half the alternating sum of these Laplacians, we obtain two “good” terms (the third term in \( \Delta_{\ell_1} \) and the second term in \( \Delta_{\ell_3} \)), namely
\[ -\frac{1}{2} \text{tr}(\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1) - \frac{1}{2} \text{tr}(X\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1). \]

(3.12)

After using the cyclic invariance of the trace, these terms reduce precisely to (3.11). We are left with eight “bad” terms in the alternating sum that must cancel out after integration.

To verify this cancellation directly, we compute that
\[ (\nabla_{\ell_1,L} \cdot \nabla_{\ell_2,L} - \nabla_{\ell_1,R} \cdot \nabla_{\ell_2,R}) f \]
\[ = \sum_X \left( - \text{tr}(\ell_3^{-1}\ell_4^{-1}X\ell_1^{-1}\ell_3\ell_2X\ell_1) + \text{tr}(\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}\ell_3\ell_2X\ell_1) \right) \]
\[ + \text{tr}(\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1) - \text{tr}(\ell_3^{-1}\ell_4^{-1}\ell_1^{-1}X\ell_3\ell_2\ell_1) \]  
(3.13)

and that
\[ (\nabla_{\ell_3,L} \cdot \nabla_{\ell_4,L} - \nabla_{\ell_3,R} \cdot \nabla_{\ell_4,R}) f \]
\[ = \sum_X \left( \text{tr}(X\ell_3^{-1}X\ell_4^{-1}\ell_1^{-1}\ell_3\ell_2\ell_1) - \text{tr}(\ell_3^{-1}X\ell_4^{-1}\ell_1^{-1}\ell_3\ell_2\ell_1) \right) \]
\[ - \text{tr}(\ell_3^{-1}X\ell_4^{-1}X\ell_1^{-1}\ell_3\ell_2\ell_1) + \text{tr}(\ell_3^{-1}\ell_4^{-1}X\ell_1^{-1}X\ell_3\ell_2\ell_1)) \].

(3.14)

One can easily check that the eight bad terms in the alternating sum of Laplacians are exactly the sum of the right-hand sides of (3.13) and (3.14). Thus, by Lemma 3.4, the bad terms integrate to zero.

4 Reduction to the generic case

In the preceding sections, we assumed that our graph was “generic” relative to the given vertex \( v \), meaning that the four adjacent faces are distinct and bounded and that the four edges emanating from \( v \) are distinct. (More precisely, the nongeneric situation is when one of the outgoing edges \( e_i \) emanating from \( v \) coincides with \( e_j^{-1} \) for some \( j \neq i \).) In this section, we show that Theorem 2.4 still holds even if the preceding assumptions are not satisfied.

Let us first describe the proper interpretation of the theorem in the non-generic case. If one of the adjacent faces \( F_i \) is the unbounded face, the corresponding time derivative \( \partial / \partial t_i \) should be interpreted as zero. If \( F_i = F_j \) for \( i \neq j \), we simply have the same area-derivative twice on the left-hand side of the Makeenko–Migdal equation.

Meanwhile, suppose an outgoing edge \( e_i \) coincides with \( e_j^{-1} \) for some \( j \neq i \). In that case, we do not consider the edge variable \( a_j \) as an independent variable. Rather, we simply regard \( a_j \) as another
In that case, we interpret $\nabla^{a_j}$ as $\nabla^{a_j^{-1}}$, where

$$\nabla^{a_j^{-1}} f(a_1, \ldots, a_i, \ldots, a_4, b) = \sum_X \left( \frac{d}{ds} f(a_1, \ldots, e^{-sX}a_i, \ldots, a_4, b) \bigg|_{s=0} \right) X.$$  \hspace{1cm} (4.1)

(Alternatively, we can think of $a_j$ as another name for $a_j^{-1}$; it makes no difference.)

In the notion of extended gauge invariance, we are supposed to insert an $x$ at the beginning of the outgoing edges $e_i$ and $e_{i+2}$. If it should happen that $e_i$ coincides with $e_{i+2}^{-1}$, then we have an $x$ at the beginning of $e_i$ and an $x^{-1}$ at the end of $e_i$. Thus, since parallel transport is order reversing, we should change $a_i$ to $x^{-1}a_i x$. In that case, we do not think of $a_{i+2}$ as an independent variable, and extended gauge invariance means invariance under $a_i \mapsto x^{-1}a_i x$. If, for example, $e_1 = e_3^{-1}$ but $e_2$ and $e_4^{-1}$ are distinct, extended invariance means that

$$f(a_1, a_2, a_4, b) = f(x^{-1}a_1 x, a_2, a_4, b) = f(a_1, a_2 x, a_4 x, b)$$

for all $x \in K$.

If a graph is not generic relative to a given vertex $v$, we construct a new graph by adding four new vertices and connecting them in a circular pattern as in Figures 9.

For each outgoing edge $e_i$ emanating from $v$, we make a substitution either as

$$e_i \mapsto \tilde{e}_i e_i$$

(if $e_i$ is distinct from each $e_j^{-1}$ with $i \neq j$) or as

$$e_i \mapsto \tilde{e}_i e_i e_j^{-1}$$
(if \(e_i\) coincides with \(e_j^{-1}\) for some \(i \neq j\)). (See Figure 10.) We then make a corresponding substitution of the associated edge variables: let \(\phi_i\) denote the edge variable corresponding to the edge \(e_i\). Since the edge variables are interpreted as parallel transport and since parallel transport is order-reversing, these substitutions take the form

\[ a_i \mapsto \phi_i \tilde{a}_i \]  

or

\[ a_i \mapsto \tilde{a}_j^{-1} b_i \tilde{a}_i. \]

Here \(a_i\) is the edge variable associated to \(e_i\) in the original graph, and \(\tilde{a}_i\) and \(\phi_i\) are the edge variables associated to \(\tilde{e}_i\) and \(e_i\), respectively, in the new graph.

Suppose now \(f\) is a function of the edge variables of the original graph. We may then associate to \(f\) a function \(\tilde{f}\) of the edge variables in the new graph by means of the preceding substitutions. To be a little more explicit, we divide the edge variables in the new graph into four groups, the variables \(\tilde{a}_i\), the variables \(\phi_i\), the variables \(c_i\) associated to the four “circular” edges in the new graph, and all the remaining edge variables \(b\), which simply correspond to the edge variables \(b\) in the original graph. Then

\[ \tilde{f}(\tilde{a}, \phi, c, b) = f(a, b), \]

where \(a\) depends on \(\tilde{a}\) and \(\phi\) as in (4.2) and (4.3). Note that \(\tilde{f}\) is independent of the \(c\) variables.

We now let \(\mu\) denote the Yang–Mills measure for the original graph and \(\tilde{\mu}\) the Yang–Mills measure for the new graph. The key point is the consistency of the Yang–Mills measures, namely that

\[ \int f \, d\mu = \int \tilde{f} \, d\tilde{\mu}. \]  

(4.4)

From the point of the continuum theory of \([\text{GKS}, \text{Dr}]\), this consistency holds because the stochastic parallel transport is anti-multiplicative. One can also verify consistency directly from the formulas for \(\mu\) and \(\tilde{\mu}\); it amounts to the fact that the heat kernel forms a convolution semigroup, \(\rho_{s+t} = \rho_s * \rho_t\).

Since the new graph is generic relative to \(v\), all of our proofs of the Makeenko–Migdal equation apply to \(\int \tilde{f} \, d\tilde{\mu}\). It now remains only to see that the Makeenko–Migdal equation for \(\int \tilde{f} \, d\tilde{\mu}\) reduces to the Makeenko–Migdal equation for \(\int f \, d\mu\). We begin with the time derivatives and we consider first the possibility that one of the adjacent faces \(F_i\) in the original graph is the unbounded face. In that case, the face \(\tilde{F}_i\) in the new graph will share a “circular” edge with the unbounded face. Since the circular edge lies between \(\tilde{F}_i\) and the unbounded face, the corresponding edge variable \(c_i\) will occur in only one of the heat kernels in the definition of \(\tilde{\mu}\). Thus, the density of \(\tilde{\mu}\) will take the form

\[ \rho_{\tilde{t}_i}(c_i \gamma) \delta, \]

where \(\gamma\) is a word in edge variables other than \(c_i\) and where \(\delta\) is a product of heat kernels evaluated on edge variables other than \(c_i\). Thus,

\[ \frac{\partial}{\partial \tilde{t}_i} \int \tilde{f} \, d\tilde{\mu} = \frac{1}{2} \int \tilde{f} \Delta c_i [\rho_{\tilde{t}_i}(c_i \gamma)] \delta \, d\text{Haar} \]

\[ = \frac{1}{2} \int (\Delta c_i \tilde{f}) \, d\tilde{\mu} \]

\[ = 0, \]

since \(\tilde{f}\) is independent of \(c_i\).
We next consider the possibility that for some \( i \neq j \), two bounded faces \( F_i \) and \( F_j \) in the original graph coincide. In that case, the face \( F_i = F_j \) is divided into three faces in the new graph, \( \tilde{F}_i, \tilde{F}_j, \) and one other face \( G \). Thus,
\[
t_i = \tilde{t}_i + \tilde{t}_j + s,
\]
where \( s \) is the area of \( G \), which means that varying \( t_i \) has the same effect as varying \( t_i \). It follows from this observation and (4.4) that
\[
\frac{\partial}{\partial t_i} \int \tilde{f} \, d\tilde{\mu} = \frac{\partial}{\partial t_i} \int f \, d\mu. \tag{4.5}
\]
If three or more bounded faces in the original graph coincide, a very similar argument shows that (4.5) still holds.

We may summarize the results of the two previous paragraphs as follows: Each time derivative applied to \( \int \tilde{f} \, d\tilde{\mu} \) coincides with the corresponding time derivative applied to \( \int f \, d\mu \), where in the latter case we follow the convention that a derivative with respect to the area of the unbounded face is zero. Thus, the left-hand side of the abstract Makeenko–Migdal equation for \( \int \tilde{f} \, d\tilde{\mu} \) reduces to the left-hand side of the abstract Makeenko–Migdal equation for \( \int f \, d\mu \).

A similar analysis applies to the right-hand side of the abstract Makeenko–Migdal equation. If we compute \( \nabla \tilde{a}_i \tilde{f} \), then using (4.2) or (4.3), we can easily see that
\[
\nabla \tilde{a}_i \tilde{f} = \nabla \tilde{a}_i \tilde{f}. \tag{4.6}
\]
(If \( e_i \) coincides with \( e_j^{-1} \), we may think of \( a_i \) as the independent variable, or we may think of \( a_j \) as the independent variable and interpret \( \nabla \tilde{a}_i \) as \( \nabla \tilde{a}_j^{-1} \) as in (4.1). Either way, (4.6) is easily seen to hold.) Thus, by (4.4),
\[
\int \nabla \tilde{a}_i \tilde{f} \, d\tilde{\mu} = \int \nabla \tilde{a}_i \tilde{f} \, d\mu. \tag{4.7}
\]

Finally, it is easy to check that if \( f \) has extended gauge invariance, \( \tilde{f} \) also has extended gauge invariance. Since the new graph is generic relative to \( v \), the proofs of the Makeenko–Migdal equation from Section 2.4 apply to \( \tilde{f} \). Using (4.5) and (4.7), it follows that the Makeenko–Migdal equation also holds for \( f \).

References

[AS] M. Anshelevich and A. N. Sengupta, Quantum free Yang-Mills on the plane. J. Geom. Phys. 62 (2012), 330–343.

[Chatt] S. Chatterjee, Rigorous solution of strongly coupled \( SO(N) \) lattice gauge theory in the large \( N \) limit. preprint: arXiv:1502.07719v3 [math.PR].

[Dahl] A. Dahlqvist, Free energies and fluctuations for the unitary Brownian motion, preprint: arXiv:1409.7793v1 [math.PR].

[Dr] B. K. Driver, YM\(_2\): continuum expectations, lattice convergence, and lassos. Comm. Math. Phys. 123 (1989), 575–616.

[DHK] B. K. Driver, B. C. Hall, and T. Kemp, The large-\( N \) limit of the Segal–Bargmann transform on \( U_N \), J. Funct. Anal. 265 (2013), 2585–2644.

[Gop] R. Gopakumar, The master field in generalised QCD\(_2\), Nuclear Phys. B 471 (1996), 246–260.
[GG] R. Gopakumar and D. Gross, Mastering the master field, *Nuclear Phys. B* 451 (1995), 379–415.

[GKS] L. Gross, C. King, and A. N. Sengupta, Two-dimensional Yang-Mills theory via stochastic differential equations, *Ann. Physics* 194 (1989), 65–112.

[Lévy1] T. Lévy, Two-dimensional Markovian holonomy fields. Astrisque No. 329 (2010), 172 pp.

[Lévy2] T. Lévy, The master field on the plane, preprint: arXiv:1112.2452v2 [math-ph].

[MM] Y. M. Makeenko and A. A. Migdal, Exact equation for the loop average in multicolor QCD, *Physics Letters* 88B (1979), 135-137.

[Mig] A. A. Migdal, Recursion equations in gauge field theories, *Sov. Phys. JETP* 42 (1975), 413-418.

[Sing] I. M. Singer, On the master field in two dimensions, *In: Functional analysis on the eve of the 21st century, Vol. 1* (New Brunswick, NJ, 1993), 263–281, *Progr. Math.*, 131, Birkhäuser Boston, Boston, MA, 1995.

[’t Hooft] G. ’t Hooft, A planar diagram theory for strong interactions, *Nuclear Physics B* 72 (1974), 461-473.