Proceedings

Wilson Surfaces for Surface Knots

LMS/EPSRC Durham Symposium on Higher Structures in M-Theory

Roberto Zucchini[a, *]

Holonomy invariants in strict higher gauge theory have been studied in depth, aiming to applications to higher Chern–Simons theory. For a flat 2–connection, the holonomy of surface knots of arbitrary genus has been defined and its covariance properties under 1–gauge transformation and change of base data have been determined. Using quandle theory, a definition of trace over a crossed module has been given that yields surface knot invariants upon application to 2–holonomies.

1 Introduction

Knots are interesting in topology as well as in gauge theory [1].

Ordinary knots are embeddings of S^1 into a 3–dimensional manifold, say S^3 [2, 3]. Can one define higher dimensional knots generalizing this simple topological notion? In just one dimension higher there are at least two ways of doing that.

Since S^1 is the lowest dimensional non trivial sphere, one may define a 2–dimensional knot as an embedding S^2 into S^4. This yields the so called 2–knots. Since S^1 is also the lowest dimensional non trivial closed oriented manifold, one may define a 2–dimensional knot as an embedding of S_ℓ into S^4, where S_ℓ is a genus ℓ closed oriented surface. This leads to genus ℓ surface knots. Of course, 2–knots are just genus 0 surface knots. However, they have very special properties which make a separate study meaningful. 2- and surface knots are objects of intense investigation by topologists [4, 5].

Wilson lines [6] are relevant in the analysis of confinement in quantum chromodynamics, loop formulation of quantum gravity, symmetry breaking in string theory, condensed matter theory and knot topology. As shown in Witten's seminal work [7], one can study knot topology in Chern-Simons theory, an instance of gauge theory, relying on techniques of quantum field theory. With any knot ξ, one associates the Wilson line

$$W_R(\xi) = \mathrm{tr}_R\Big[\,\mathrm{Pexp}\Big(-\int_\xi A\Big)\Big]. \tag{1}$$

where R is a representation of the gauge group G. Chern–Simons correlators of Wilson line operators provide classic knot invariants.

Wilson surfaces [8, 9] may turn out to be relevant in the study of non perturbative aspects of higher gauge theory, brane theory, quantum gravity and higher knot topology. Following Witten's paradigm, one can presumably study 2– or surface knot topology computing correlators of knot Wilson surfaces in an appropriate higher version of Chern–Simons theory, an instance of higher gauge theory [10, 11], using again techniques of quantum field theory. To this end, one needs to associate with any surface knot Ξ a Wilson surface

$$W(\Xi) = ?\,, \tag{2}$$

whose expressions is at this point to be found. In this communication, we shall present a proposal for a definition of the Wilson surfaces $W(\Xi)$ in higher gauge theory based mainly on our work [12, 13].

The problem has two parts:

i) *define surface knot holonomy*;

ii) *define higher invariant traces*.

Parallel transport and holonomy are related but distinguished, holonomy being a special case of parallel transport.

Earlier endeavours on higher parallel transport includes the work of Caetano and Picken [14], Baez and Schreiber [15, 16] Schreiber and Waldorf [17–19], Faria Martins and Picken [20, 21], Chatterjee, Lahiri and Sengupta [22–24]

* Corresponding author e-mail: roberto.zucchini@unibo.it
a DIFA, University of Bologna, and INFN, Bologna, Italy

Soncini and Zucchini [25], Abbaspour and Wagemann [26] and Arias Abad and Schaetz [27,28]. Earlier results on higher holonomy were obtained by Cattaneo and Rossi [29] and Faria Martins and Picken [20, 21].

Following [12, 13], we shall present a framework for the construction of holonomy invariants of knots and surface knots. In a nutshell, our strategy rests on describing knots by parametrized curves and surface knots by parametrized surfaces. We outline it below, assuming that the reader is familiar with the basic ideas of strict higher gauge theory. In any case, those notion will be reviewed in greater detail in subsequent sections.

In a manifold M, a curve $\gamma : p_0 \to p_1$ is a parametrized path joining two points. A homotopy $h : \gamma_0 \Rightarrow \gamma_1$ of two curves is a parametrized path joining those curves. Curves can be composed by concatenation and inverted. The resulting operations make curves modulo homotopy a groupoid, the fundamental 1-groupoid $(M, P^0{}_1M)$ of M.

In ordinary gauge theory with gauge Lie group G, given a flat gauge field A one can construct a gauge covariant and homotopy invariant parallel transport functor

$$F_A : (M, P^0{}_1M) \to BG, \qquad (3)$$
$$\gamma \to F_A(\gamma),$$

where BG is the delooping of G, that is G seen as the morphism group of a one–object groupoid.

With a knot ξ based at p defined up to ambient isotopy one can associate a curve $\gamma_\xi : p \to p$ defined up to homotopy and with this the holonomy

$$F_A(\xi) = F_A(\gamma_\xi). \qquad (4)$$

One can check $F_A(\xi)$ is base point and isotopy invariant and gauge independent up to conjugation. Using invariant traces, one can extract an invariant from the holonomy $F_A(\xi)$.

A 'gentle' generalization of the above construction for surface knots is the following.

A curve $\gamma : p_0 \to p_1$ is a parametrized path joining two points. A surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$ is a parametrized path joining two curves in a manifold M. A thin homotopy $h : \gamma_0 \Rightarrow \gamma_1$ of two curves is a parametrized path joining those curves with degenerate (less than two–dimensional) range A homotopy $H : \Sigma_0 \Rrightarrow \Sigma_1$ of two surfaces is a parametrized path joining those surfaces.Curves can be composed by concatenation and inverted. Surfaces can be composed by concatenation and inverted in two distinct ways, usually called horizontal and vertical. The resulting operations make curves modulo thin homotopy and surfaces modulo homotopy a 2–groupoid, fundamental 2–groupoid $(M, P_1M, P^0{}_2M)$.

In strict higher gauge theory with gauge Lie crossed module (G, H), given a flat higher gauge field pair A, B one can construct a gauge covariant and (thin) homotopy invariant parallel transport 2–functor

$$F_{A,B} : (M, P_1M, P^0{}_2M) \to B(G, H), \qquad (5)$$
$$\gamma \to F_A(\gamma), \quad \Sigma \to F_{A,B}(\Sigma).$$

With a knot ξ based at p and a surface knot Ξ based at a genus dependent fundamental polygon τ stemming from cutting the image of Ξ along standard a and b cycles, both defined up to ambient isotopy, one can associate a curve $\gamma_\xi : p \to p$ and a surface $\Sigma_\Xi : \iota_p \Rightarrow \tau$ up to (thin) homotopy and from this the holonomy

$$F_A(\xi) = F_A(\gamma_\xi), \quad F_{A,B}(\Xi) = F_{A,B}(\Sigma_\Xi). \qquad (6)$$

One can check that $F_A(\xi)$ is base data and isotopy invariant and gauge independent up to the appropriate form of crossed module conjugation. Using higher invariant traces, one can extract invariants from the holonomy $F_A(\xi)$ and $F_{A,B}(\Xi)$.

There are open issues to be solved. It can be shown that surface knot holonomy necessarily lies in the kernel of the target map $H \xrightarrow{t} G$ of the Lie crossed module (G, H) and so is central. Thus, for many Lie crossed module this holonomy may turn out to be trivial. The existence of non trivial higher traces on (G, H) is also to be ascertained. This is an problem that can be formalized using higher quandle theory (see Crans [30] and Crans and Wagemann [31]).

From a quantum field theoretic point of view, the most delicate question remains obtaining surface knot invariants from a 4–dimensional higher Chern–Simons theory as proposed by Zucchini [32, 33] and Soncini and Zucchini [34]. There are problems with the definition of Wilson surface insertions in the quantum theory, which we shall point out in due course.

2 Curves, surfaces and homotopy

Closed curves and surfaces describe knots and surface knots in an ambient manifold M.

Curves and surfaces are smoothly parametrized subsets of M. They can be composed and inverted in various ways. In order to preserve smoothness, it is sufficient to require that their parametrization has sitting instants. A smooth map $f : S \times \mathbb{R} \to T$, where S and T are manifolds, has sitting instants if

$$f(-,x) = f(-,0) \quad \text{for } x < \epsilon, \tag{7}$$
$$f(-,x) = f(-,1) \quad \text{for } x > 1-\epsilon,$$

for some number ϵ such that $0 < \epsilon < 1/2$. In what follows, all maps will be tacitly assumed to have sitting instants for each factor $\mathbb{R}$ of their domains.

Formally, curves and surfaces are defined as follows.

For any two points $p_0, p_1 \in M$, a curve $\gamma : p_0 \to p_1$ in M is a map $\gamma : \mathbb{R} \to M$ such that

$$\gamma(0) = p_0, \quad \gamma(1) = p_1. \tag{8}$$

For any two points $p_0, p_1 \in M$ and any two curves $\gamma_0, \gamma_1 : p_0 \to p_1$ of M, a surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$ of M is a map $\Sigma : \mathbb{R}^2 \to M$ such that

$$\Sigma(0,y) = p_0, \quad \Sigma(1,y) = p_1, \tag{9}$$
$$\Sigma(x,0) = \gamma_0(x), \quad \Sigma(x,1) = \gamma_1(x)$$

Curve and surfaces can be combined through a set of natural operations based on the intuitive idea of concatenation. We begin by introducing the basic operations with curves.

For a point p, the unit curve of p is the curve $\iota_p : p \to p$ defined by

$$\iota_p(x) = p. \tag{10}$$

For a curve $\gamma : p_0 \to p_1$, the inverse curve of γ is the curve $\gamma^{-1_\circ} : p_1 \to p_0$ given by

$$\gamma^{-1_\circ}(x) = \gamma(1-x). \tag{11}$$

For two curves $\gamma_1 : p_0 \to p_1$, $\gamma_2 : p_1 \to p_2$, the composition of γ_1, γ_2 is the curve $\gamma_2 \circ \gamma_1 : p_0 \to p_2$ piecewise given by

$$\gamma_2 \circ \gamma_1(x) = \gamma_1(2x) \qquad \text{for } x \le 1/2, \tag{12}$$

$$\gamma_2 \circ \gamma_1(x) = \gamma_2(2x-1) \qquad \text{for } x \ge 1/2.$$

We introduce next the basic operations with surfaces. These turn out to be of two types, called horizontal and vertical.

For a curve $\gamma : p_0 \to p_1$, the unit surface of γ is the surface $I_\gamma : \gamma \Rightarrow \gamma$ defined by

$$I_\gamma(x,y) = \gamma(x). \tag{13}$$

For a surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$, the vertical inverse of Σ is the surface $\Sigma^{-1_\bullet} : \gamma_1 \Rightarrow \gamma_0$ defined by

$$\Sigma^{-1_\bullet}(x,y) = \Sigma(x,1-y). \tag{14}$$

For two surfaces $\Sigma_1 : \gamma_0 \Rightarrow \gamma_1$, $\Sigma_2 : \gamma_1 \Rightarrow \gamma_2$, the vertical composition of Σ_1, Σ_2 is the surface $\Sigma_2 \bullet \Sigma_1 : \gamma_0 \Rightarrow \gamma_2$ given by

$$\Sigma_2 \bullet \Sigma_1(x,y) = \Sigma_1(x,2y) \qquad \textit{for } y \le 1/2, \tag{15}$$

$$\Sigma_2 \bullet \Sigma_1(x,y) = \Sigma_2(x,2y-1) \qquad \textit{for } y \ge 1/2.$$

For a surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$, the horizontal inverse of Σ is the surface $\Sigma^{-1_\circ} : {\gamma_0}^{-1_\circ} \Rightarrow {\gamma_1}^{-1_\circ}$ defined by

$$\Sigma^{-1_\circ}(x,y) = \Sigma(1-x,y). \tag{16}$$

For two surfaces $\Sigma_1 : \gamma_0 \Rightarrow \gamma_1$, $\Sigma_2 : \gamma_2 \Rightarrow \gamma_3$, the horizontal composition of Σ_1, Σ_2 is the surface $\Sigma_2 \circ \Sigma_1 : \gamma_2 \circ \gamma_0 \Rightarrow \gamma_3 \circ \gamma_1$ given by

$$\Sigma_2 \circ \Sigma_1(x,y) = \Sigma_1(2x,y) \qquad \text{for } x \le 1/2, \tag{17}$$

$$\Sigma_2 \circ \Sigma_1(x,y) = \Sigma_2(2x-1,y) \qquad \text{for } x \ge 1/2.$$

Unfortunately, these operations are not nice enough; associativity and invertibility fail to hold in general. The operation are in fact nice only up to homotopy.

A homotopy $h : \gamma_0 \Rightarrow \gamma_1$ of two curves $\gamma_0, \gamma_1 : p_0 \to p_1$ of M with the same end-points is a map $h : \mathbb{R}^2 \to M$ of M such that

$$h(0,y) = p_0, \quad h(1,y) = p_1, \tag{18}$$
$$h(x,0) = \gamma_0(x), \quad h(x,1) = \gamma_1(x).$$

The homotopy is thin if in addition $\operatorname{rank} dh(x,y) < 2$. *(Thin) homotopy of curves is an equivalence relation.*

A homotopy $H : \Sigma_0 \Rrightarrow \Sigma_1$ of two surfaces $\Sigma_0 : \gamma_0 \Rightarrow \gamma_1$, $\Sigma_1 : \gamma_2 \Rightarrow \gamma_3$, where $\gamma_0, \gamma_1, \gamma_2, \gamma_3 : p_0 \to p_1$ are four curves with the same end-points, is a map $H : \mathbb{R}^3 \to M$ such that $\operatorname{rank} dH(x,0,z)$, $\operatorname{rank} dH(x,1,z) \le 1$ *and*

$$H(0,y,z) = p_0, \quad H(1,y,z) = p_1, \tag{19}$$
$$H(x,y,0) = \Sigma_0(x,y), \quad H(x,y,1) = \Sigma_1(x,y).$$

The homotopy is thin if $\operatorname{rank} dH(x,y,z) < 3$. *(Thin) homotopy of surfaces is an equivalence relation.*

Let us denote by $\Pi_1 M, \Pi_2 M$ the sets of all curves and surfaces of M, by $P_1 M$ and $P^0{}_1 M$ the sets of thin homotopy and homotopy classes of curves and $P_2 M$ and $P^0{}_2 M$ the sets of thin homotopy and homotopy classes of surfaces, respectively. The following results are basic.

(M, P_1M) and $(M, P^0{}_1M)$ with the operations induced by those of $\Pi_1 M$ are groupoids, the path and fundamental groupoids of M.

(M, P_1M, P_2M) and $(M, P_1M, P^0{}_2M)$ with the operations induced by those of $\Pi_1 M$ and $\Pi_2 M$ are 2–groupoids, the path and fundamental 2–groupoids of M.

3 Higher parallel transport

In gauge theory, holonomy is a special case of parallel transport. Therefore, it is necessary to examine in some detail the definition and the properties of the latter. We begin be reviewing parallel transport in ordinary gauge theory and then we introduce and describe parallel transport in higher gauge theory.

Let G be a Lie group with Lie algebra $\mathfrak{g}$ and let M be a manifold. Consider an ordinary gauge theory on the trivial principal G–bundle $M \times G$.

A G–connection on M is a $\mathfrak{g}$–valued 1–form $\theta \in \Omega^1(M, \mathfrak{g})$. θ is flat if

$$d\theta + \frac{1}{2}[\theta, \theta] = 0. \tag{20}$$

Parallel transport requires a G–connection on M as input datum.

For a curve γ of M, the parallel transport along γ is the element $F_\theta(\gamma) \in G$ defined by

$$F_\theta(\gamma) = u(1), \tag{21}$$

where $u : \mathbb{R} \to G$ is the unique solution of the differential problem

$$d_x u(x) u(x)^{-1} = -\gamma^* \theta_x(x), \quad u(0) = 1_G. \tag{22}$$

The first relevant property of parallel transport is its consistency with the operations with curves defined in Section 2.

For any point p and any curves $\gamma, \gamma_1, \gamma_2$, one has

$$F_\theta(\iota_p) = 1_G, \tag{23}$$

$$F_\theta(\gamma^{-1_\circ}) = F_\theta(\gamma)^{-1}, \tag{24}$$

$$F_\theta(\gamma_2 \circ \gamma_1) = F_\theta(\gamma_2) F_\theta(\gamma_1). \tag{25}$$

whenever defined.

The second relevant property of parallel transport is its compatibility with homotopy of curves as defined in Section 2.

For any two thinly homotopic curves γ_0, γ_1, one has

$$F_\theta(\gamma_1) = F_\theta(\gamma_0). \tag{26}$$

When θ is flat, (26) holds also when γ_0, γ_1 are homotopic.

Parallel transport has an elegant categorical interpretation.

Parallel transport yields a functor $\bar{F}_\theta : (M, P_1M) \to BG$ from the path groupoid (M, P_1M) of M into BG. For flat θ, parallel transport yields a functor $\bar{F}^0{}_\theta : (M, P^0{}_1M) \to BG$ from the fundamental groupoid $(M, P^0{}_1M)$ of M into BG.

Any meaningful gauge theoretic construction should be gauge covariant in the appropriate sense. Parallel transport has also this property.

A G–gauge transformation is just a G–valued mapping $g \in \mathrm{Map}(M, G)$.

Gauge transformations act on connections in the well–known manner.

The gauge transform of the G–connection θ is the G–connection

$${}^g\theta = \mathrm{Ad}\, g(\theta) - dg g^{-1}. \tag{27}$$

If θ is flat, ${}^g\theta$ is flat, too.

Parallel transport has simple covariance properties under gauge transformation.

For any curve $\gamma : p_0 \to p_1$ of M,

$$F_{{}^g\theta}(\gamma) = g(p_1) F_\theta(\gamma) g(p_0)^{-1}. \tag{28}$$

Gauge transformation of parallel transport also has an elegant categorical interpretation.

A gauge transformation g encodes a natural transformation $\bar{F}_\theta \Rightarrow \bar{F}_{{}^g\theta}$ of parallel transport functors. When θ is flat, g encodes a natural transformation $\bar{F}^0{}_\theta \Rightarrow \bar{F}^0{}_{{}^g\theta}$ of flat parallel transport functors.

An appropriate form of parallel transport can be defined also in strict higher gauge theory. The intuitive idea of the construction is still simple, though the technical details are much more involved.

Let K be a strict Lie 2 group with strict Lie 2–algebra $\mathfrak{k}$ and let M be a manifold. Consider a higher gauge theory on the trivial principal K–2–bundle $M \times K$. As it is natural and convenient, we shall view the Lie 2–group

K as a Lie crossed module $H \xrightarrow{t} G \xrightarrow{m} \mathrm{Aut}(H)$ and the Lie 2–algebra $\mathfrak{k}$ as the differential Lie crossed module $\mathfrak{h} \xrightarrow{\dot{t}} \mathfrak{g} \xrightarrow{\widehat{m}} \mathfrak{der}(\mathfrak{h})$ corresponding to it.

A (G,H)–2–connection on M is a pair formed by a $\mathfrak{g}$–valued 1–form $\theta \in \Omega^1(M,\mathfrak{g})$ and a $\mathfrak{h}$–valued 2–form $\Upsilon \in \Omega^2(M,\mathfrak{h})$ such that

$$\mathrm{d}\theta + \frac{1}{2}[\theta,\theta] - \dot{t}(\Upsilon) = 0. \tag{29}$$

(vanishing fake curvature condition). (θ,Υ) is flat if

$$\mathrm{d}\Upsilon + \widehat{m}(\theta,\Upsilon) = 0. \tag{30}$$

Analogously to the ordinary case, higher parallel transport requires a (G,H) connection on M as input datum.

For a curve γ of M, the parallel transport $F_\theta(\gamma) \in G$ is constructed as done earlier for the G–connection θ. For a surface Σ of M, the parallel transport along Σ is the element $F_{\theta,\Upsilon}(\Sigma) \in H$ defined by

$$F_{\theta,\Upsilon}(\Sigma) = E(0,1), \tag{31}$$

where $E : \mathbb{R}^2 \to H$ is the unique solution of the two step differential problem

$$\partial_x u(x,y)u(x,y)^{-1} = -\Sigma^*\theta_x(x,y), \quad u(1,y) = 1_G, \tag{32}$$

$$\partial_y v(x,y)v(x,y)^{-1} = -\Sigma^*\theta_y(x,y), \quad v(x,0) = 1_G, \tag{33}$$

$$\begin{aligned} &\partial_x(\partial_y E(x,y)E(x,y)^{-1}) = \\ &\qquad = -\dot{m}(v(1,y)^{-1}u(x,y)^{-1})(\Sigma^*\Upsilon_{xy}(x,y)) \ \text{ or} \\ &\partial_y(E(x,y)^{-1}\partial_x E(x,y)) = \\ &\qquad = -\dot{m}(u(x,0)^{-1}v(x,y)^{-1})(\Sigma^*\Upsilon_{xy}(x,y)), \\ &\qquad\qquad E(1,y) = E(x,0) = 1_H \end{aligned} \tag{34}$$

with $u, v : \mathbb{R}^2 \to G$.

The two forms of the differential problem for E are equivalent: any solution of one is automatically a solution of the other.

Higher parallel transport has several remarkable properties which extend those of the ordinary case. First, higher parallel transport along surfaces is compatible with that along their end-curves.

For a surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$ joining the curve γ_0 to the curve γ_1,

$$F_\theta(\gamma_1) = t(F_{\theta,\Upsilon}(\Sigma))F_\theta(\gamma_0). \tag{35}$$

Second, higher parallel transport is consistent with the operations with curves and surfaces defined in Section 2.

For any point p, any curves $\gamma, \gamma_1, \gamma_2$ and any surfaces $\Sigma, \Sigma_1, \Sigma_2$, relations (23)–(25) and the further relations

$$F_{\theta,\Upsilon}(I_\gamma) = 1_H, \tag{36}$$

$$F_{\theta,\Upsilon}(\Sigma^{-1\bullet}) = F_{\theta,\Upsilon}(\Sigma)^{-1}, \tag{37}$$

$$F_{\theta,\Upsilon}(\Sigma_2 \bullet \Sigma_1) = F_{\theta,\Upsilon}(\Sigma_2)F_{\theta,\Upsilon}(\Sigma_1), \tag{38}$$

$$F_{\theta,\Upsilon}(\Sigma^{-1\circ}) = m(F_\theta(\gamma_0)^{-1})(F_{\theta,\Upsilon}(\Sigma)^{-1}), \tag{39}$$

$$F_{\theta,\Upsilon}(\Sigma_2 \circ \Sigma_1) = F_{\theta,\Upsilon}(\Sigma_2)m(F_\theta(\gamma_2))(F_{\theta,\Upsilon}(\Sigma_1)), \tag{40}$$

hold whenever defined, where in the last two identities $\Sigma : \gamma_0 \Rightarrow \gamma_1$ and $\Sigma_2 : \gamma_2 \Rightarrow \gamma_3$.

Third, higher parallel transport is compatible with homotopy of curves and surfaces, as defined again in Section 2, in the following sense.

For any two thinly homotopic curves γ_0, γ_1

$$F_\theta(\gamma_1) = F_\theta(\gamma_0). \tag{41}$$

For any two thinly homotopic surfaces $\Sigma_0 : \gamma_{00} \Rightarrow \gamma_{01}$, $\Sigma_1 : \gamma_{10} \Rightarrow \gamma_{11}$,

$$F_\theta(\gamma_{10}) = F_\theta(\gamma_{00}), \tag{42}$$

$$F_\theta(\gamma_{11}) = F_\theta(\gamma_{01}), \tag{43}$$

$$F_{\theta,\Upsilon}(\Sigma_1) = F_{\theta,\Upsilon}(\Sigma_0). \tag{44}$$

The same relations hold if (θ,Υ) is flat and Σ_0, Σ_1 are homotopic.

Higher parallel transport has an elegant 2–categorical interpretation.

Higher parallel transport is equivalent to a strict 2–functor $\bar{F}_{\theta,\Upsilon} : (M, P_1M, P_2M) \to B(G,H)$ from the path 2–groupoid (M, P_1M, P_2M) of M into $B(G,H)$. For a flat (θ,Υ), higher parallel transport is likewise equivalent to a strict 2–functor $\bar{F}^0{}_{\theta,\Upsilon} : (M, P_1M, P^0{}_2M) \to B(G,H)$ from the fundamental 2–groupoid $(M, P_1M, P^0{}_2M)$ of M into $B(G,H)$.

Here, with an abuse of notation, $B(G,H)$ stands for the delooping of the strict Lie 2–group corresponding to the Lie crossed module (G,H).

Analogously to ordinary gauge theory, higher parallel transport is gauge covariant in the appropriate higher sense.

A (G,H)–1–gauge transformation is a pair of a G–valued map $g \in \mathrm{Map}(M,G)$ and an $\mathfrak{h}$–valued 1–form $J \in \Omega^1(M,\mathfrak{h})$.

1–gauge transformations act on 2 connections.

The 1–gauge transform of the (G,H)–2–connection (θ,Υ) is the 2–connection

$$ {}^{g,J}\theta = \mathrm{Ad}\, g(\theta) - dgg^{-1} - \dot{t}(J), \tag{45} $$

$$ {}^{g,J}\Upsilon = \dot{m}(g)(\Upsilon) - \mathrm{d}J - \frac{1}{2}[J,J] - \\ - \widehat{m}(\mathrm{Ad}\, g(\theta) - \mathrm{d}gg^{-1} - \dot{t}(J), J). \tag{46} $$

If (θ,Υ) is flat, $({}^{g,J}\theta, {}^{g,J}\Upsilon)$ is flat, too.

There exits a notion of parallel transport for 1–gauge transformations similarly to 2–connections.

For a curve γ of M, the gauge parallel transport along γ is the element $G_{g,J;\theta}(\gamma) \in H$ given by

$$ G_{g,J;\theta}(\gamma) = \Lambda(0), \tag{47} $$

where $\Lambda : \mathbb{R} \to H$ is the unique solution of the two–step differential problem

$$ \mathrm{d}_x u(x) u(x)^{-1} = -\gamma^* \theta_x(x), \quad u(1) = 1_G, \tag{48} $$

$$ \Lambda(x)^{-1}\mathrm{d}_x\Lambda(x) = -\dot{m}(u(x)^{-1}\gamma^* g(x)^{-1})(\gamma^* J_x(x)), \\ \Lambda(1) = 1_H. \tag{49} $$

As ordinary parallel transport, gauge parallel transport is consistent with the operations with curves defined in Section 2.

For any point p and any curves $\gamma, \gamma_1, \gamma_2$, one has

$$ G_{g,J;\theta}(\iota_p) = 1_H, \tag{50} $$

$$ G_{g,J;\theta}(\gamma^{-1_\circ}) = m(F_\theta(\gamma)^{-1})(G_{g,J;\theta}(\gamma)^{-1}), \tag{51} $$

$$ G_{g,J;\theta}(\gamma_2 \circ \gamma_1) = G_{g,J;\theta}(\gamma_2)\, m(F_\theta(\gamma_2))(G_{g,J;\theta}(\gamma_1)). \tag{52} $$

whenever defined.

Again as ordinary parallel transport, gauge parallel transport is compatible with homotopy of curves as defined in Section 2.

For any two thinly homotopic curves γ_0, γ_1, one has

$$ G_{g,J;\theta}(\gamma_1) = G_{g,J;\theta}(\gamma_0). \tag{53} $$

The reason why we introduced gauge parallel transport is that it enters in the 1–gauge covariance relation of higher parallel transport in a non trivial manner.

For any curve $\gamma : p_0 \to p_1$, one has

$$ F_{g,J_\theta}(\gamma) = g(p_1)\, t(G_{g,J;\theta}(\gamma)) F_\theta(\gamma) g(p_0)^{-1}. \tag{54} $$

For any two curves $\gamma_0, \gamma_1 : p_0 \to p_1$ and any surface $\Sigma : \gamma_0 \Rightarrow \gamma_1$, one has

$$ F_{g,J_{\theta},g,J_\Upsilon}(\Sigma) = \\ = m(g(p_1))\big(G_{g,J;\theta}(\gamma_1) F_{\theta,\Upsilon}(\Sigma) G_{g,J;\theta}(\gamma_0)^{-1}\big). \tag{55} $$

Gauge parallel transport has as expected a categorical interpretation.

Gauge parallel transport defines a pseudonatural transformation $\bar{G}_{g,J;\theta} : \bar{F}_{\theta,\Upsilon} \Rightarrow \bar{F}_{g,J_{\theta},g,J_\Upsilon}$ of parallel transport 2–functors. If (θ,Υ) is a flat 2–connection, gauge parallel transport defines a pseudonatural transformation $\bar{G}^0{}_{g,J;\theta} : \bar{F}^0{}_{\theta,\Upsilon} \Rightarrow \bar{F}^0{}_{g,J_{\theta},g,J_\Upsilon}$ of flat parallel transport 2–functors.

Higher gauge theory is characterized also by gauge for gauge symmetry.

A (G,H)–2–gauge transformation is just a mapping $\Omega \in$ $\mathrm{Map}(M,H)$.

(G,H)–2–gauge transformations describe gauge transformations of (G,H)–1– gauge transformations depending on an assigned (G,H)–2–connection (θ,Υ). They encode modifications $\bar{G}_{g,J;\theta} \Rrightarrow \bar{G}_{\bar{\Omega}_{g|\theta},\bar{\Omega}_{J|\theta};\theta}$ of gauge pseudonatural transformations of parallel transport functors. The apparently have no role in knot holonomy.

4 *C*– and *S*–knots

Knots are embeddings of a fixed closed model manifold into an ambient manifold M. Thus, knots are not simply subsets of M but mappings into M. Knots differing by an ambient isotopy are identified.

The simplest closed model manifold is the oriented circle C.

A C–marking of C is a pointing $p_C \in C$ of C.

A C–marking of an oriented manifold M is a pointing $p_M \in M$ of M.

C–knots are circles embedded in M.

A marked C–knot of M is embedding $\xi : C \to M$ of the circle C into M such that

$$ \xi(p_C) = p_M. \tag{56} $$

Ambient isotopy is the natural notion of mutual deformability of marked C–knots.

Two marked C–knots ξ_0, ξ_1 are ambient isotopic if there is a smooth family $F_z \in$ $\mathrm{Diff}_+(M)$, $z \in \mathbb{R}$, of orientation *preserving diffeomorphisms such that*

$$F_0 = \mathrm{id}_M, \quad \xi_1 = F_1 \circ \xi_0, \tag{57}$$

$$F_z(p_M) = p_M. \tag{58}$$

In order to compute C–knot holonomy, we need parametrized C–knots. This is achieved by assigning a curve to any marked C–knot as detailed next.

A compatible curve in C is a curve $\gamma_C : p_C \to p_C$ in C such that

i) $I_C = {\gamma_C}^{-1}(C \setminus p_C)$ *is an open interval in* $\mathbb{R}$*;*

ii) $\gamma_C|_{I_C} : I_C \to C \setminus p_C$ *is an orientation preserving diffeomorphism.*

Example. Let $C = S^1$ be the circle standardly embedded in $\mathbb{R}^2$ through

$$s_{S^1}(\vartheta) = (\cos\vartheta, \sin\vartheta), \tag{59}$$

where $\vartheta \in [0, 2\pi)$, with the C–marking $p_{S^1} = (1, 0)$. A compatible curve $\gamma_{S^1} : \mathbb{R} \to S^1$ is given by

$$\gamma_{S^1}(x) = s_{S^1}(2\pi\alpha(x)) \tag{60}$$

where $\alpha : \mathbb{R} \to [0, 1]$ is a function such that $d_x\alpha(x) \geq 0$ and $\alpha(x) = 0$ for $x < \epsilon$ and $\alpha(x) = 1$ for $x > 1 - \epsilon$.

A curve furnishing a natural parametrization of a given marked C–knot can now be constructed.

With every marked C–knot ξ there is associated a curve $\gamma_\xi : p_M \to p_M$ given by

$$\gamma_\xi = \xi \circ \gamma_C. \tag{61}$$

The curve γ_ξ has a number of nice properties.

γ_ξ is independent of the choice of the compatible curve γ_C up to thin homotopy.

Note that the C–marking p_C of C is fixed here.

Ambient isotopic marked C–knots have homotopic curves

If ξ_0, ξ_1 are ambient isotopic marked C–knots, then γ_{ξ_0}, γ_{ξ_1} are homotopic in the sense of Section 2.

It should be possible to alter marking data changing γ_ξ at most by a (thin) homotopy.

Two marked C–knots ξ_0, ξ_1 with respect to two distinct C–marking p_{M0}, p_{M1} of M are freely ambient isotopic if there is a smooth family $F_z \in \mathrm{Diff}_+(M)$, $z \in \mathbb{R}$, of orientation preserving diffeomorphisms such that

$$F_0 = \mathrm{id}_M, \quad \xi_1 = F_1 \circ \xi_0. \tag{62}$$

Again, the C–marking p_C of C is fixed.

Freely ambient isotopic marked C–knots have homotopic curves up to conjugation.

If ξ_0, ξ_1 two freely ambient isotopic marked C–knots, there is a curve $\gamma_1 : p_{M0} \to p_{M1}$ such that γ_{ξ_0}, ${\gamma_1}^{-1} \circ \gamma_{\xi_1} \circ \gamma_1$ are homotopic.

Notice that the "compose rightmost first" convention is used here and in the following for curve composition.

The same embedding of C into M can be a marked C–knot in more than one way. The corresponding curves are related in the expected manner.

If the embedding $\xi : C \to M$ is a marked C–knot with respect to two distinct C–markings p_{C0}, p_{M0} and p_{C1}, p_{M1} of C and M, there exists a curve $\gamma_1 : p_{M0} \to p_{M1}$ in $\xi(C)$ such that $\gamma_{\xi|0}$, ${\gamma_1}^{-1} \circ \gamma_{\xi|1} \circ \gamma_1$ are thinly homotopic.

The results just expounded are standard. Our aim is finding their generalization to surface knots. As we shall see, this task is not completely straightforward. Problems occur for higher genus knots. We shall propose a solution in due course. To this end, we need to introduce further notions.

To construct higher genus S–knot holonomy, it will be necessary to cut the model manifold S along its standard a– and b–cycles. The cuts are the images of spiky C–knots, generalized C–knots which are continuous but not smooth at the marked point.

A spiky C–knot is an embedding $\xi : C \to M$ that obeys

$$\xi(p_C) = p_M \tag{63}$$

and is smooth on $C \setminus p_C$ with finite derivatives and non zero first derivatives at both ends of $C \setminus p_C$.

Note that spiky C–knots are marked.

With any spiky C–knot ξ, one can associate a curve γ_ξ defined in the same way as above and smooth anyway.

For every spiky marked C–knot ξ, the curve γ_ξ is smooth.

We can now introduce S–knots. The next to simplest closed manifold is a genus ℓ_S closed oriented surface S.

An S–marking of M consists of the following elements:

i) *a C–marking p_M of M;*

ii) *a set of spiky C–knots ζ_{Mi} of M, $i = 1, \ldots, 2\ell_S$, such that:*

iii) *the $\zeta_{Mi}(C)$ intersect only at p_M;*

iv) *there is an embedding $\Phi : S \to M$ with the property that $\Phi(p_S) = p_M$, $\Phi \circ \zeta_{Si} = \zeta_{Mi}$.*

Note that the notion of S–marking of S is compatible with that of S–marking of M when $M = S$.

S–knots are surfaces embedded in M.

A marked S–knot of M an is embedding $\Xi : S \to M$ of the surface S into M such that

$$\Xi(p_S) = p_M, \tag{64}$$

$$\Xi \circ \zeta_{Si} = \zeta_{Mi}. \tag{65}$$

Ambient isotopy is the natural notion of mutual deformability also of marked S–knots.

Two marked S–knots Ξ_0, Ξ_1 are ambient isotopic if there is a smooth family $F_z \in \mathrm{Diff}_+(M)$, $z \in \mathbb{R}$, of orientation preserving diffeomorphisms such that

$$F_0 = \mathrm{id}_M, \quad \Xi_1 = F_1 \circ \Xi_0,$$

$$F_z(p_M) = p_M, \ \ F_z \circ \zeta_{Mi} = \zeta_{Mi}.$$

Analogously to C–knots, to compute S–knot holonomy we need parametrized S–knots. This is achieved by assigning a surface to any marked S–knot.

The fundamental polygon of S is the boundary of the simply connected open 2–fold that results cutting S along the standard a– and b–cycles. It plays a basic role in the subsequent constructions.

View S as a C–marked manifold and let

$$\gamma_{Si} = \gamma_{\zeta_{Si}} \ \text{ that is } \ \alpha_{Sr} = \gamma_{\xi_{Sr}}, \ \beta_{Sr} = \gamma_{\eta_{Sr}}, \tag{66}$$

Then the fundamental polygon of S is the curve given by

$$\tau_S = \beta_{S\ell_S}{}^{-1\circ} \circ \alpha_{S\ell_S}{}^{-1\circ} \circ \beta_{S\ell_S} \circ \alpha_{S\ell_S} \circ \cdots \circ \beta_{S1}{}^{-1\circ} \circ \alpha_{S1}{}^{-1\circ} \circ \beta_{S1} \circ \alpha_{S1} \tag{67}$$

if $\ell_S = 0$, $\tau_S = \iota_{p_S}$.

As a compatible curve in C is required in order to associate a curve to each marked C–knot, a compatible surface in S is required in order to associate a surface to each marked C–knot.

A compatible surface in S is a surface $\Sigma_S : \iota_{p_S} \to \tau_S$ such that

i) *$D_S = \Sigma_S{}^{-1}(S \setminus \cup_i \zeta_{Si}(C))$ is an open simply connected domain in $\mathbb{R}^2$;*

ii) *$\Sigma_S|_{D_S} : D_S \to S \setminus \cup_i \zeta_{Si}(C)$ is an orientation preserving diffeomorphism.*

Example. Let $S = S^2$ be the sphere embedded in $\mathbb{R}^3$ as

$$S_{S^2}(\vartheta, \varphi) = (\cos\vartheta \sin\vartheta(1 - \cos\varphi), -\sin\vartheta \sin\varphi, 1 - \sin^2\vartheta(1 - \cos\varphi)), \tag{68}$$

where $\vartheta \in (0, \pi)$, $\varphi \in [0, 2\pi)$, with the S–marking $p_{S^2} = (0,0,1)$. A compatible surface $\Sigma_{S^2} : \mathbb{R}^2 \to S^2$ is given by

$$\Sigma_{S^2}(x, y) = S_{S^2}(\pi\alpha(y), 2\pi\alpha(x)), \tag{69}$$

where $\alpha : \mathbb{R} \to [0,1]$ is a function enjoying the properties listed below Equation (60).

The surface Σ_{S^2} describes a parametrized family of circles on S^2 which spring from the north pole on one side of it, sweep S^2 dilating, reaching the south pole and then contracting and finally converge to the north pole on the other side.

Example. Let $S = T^2$ be the torus embedded in $\mathbb{R}^3$ as

$$S_{T^2}(\vartheta_1, \vartheta_2) = (\cos\vartheta_1(1 + r\cos\vartheta_2), \sin\vartheta_1(1 + r\cos\vartheta_2), r\sin\vartheta_2), \tag{70}$$

where $r < 1$ is fixed and $\vartheta_1, \vartheta_2 \in [0, 2\pi)$, with the S–marking $p_{T^2} = (1 + r, 0, 0)$ and

$$\xi_{T^2}(\vartheta) = ((1 + r)\cos\vartheta, (1 + r)\sin\vartheta, 0),$$

$$\eta_{T^2}(\vartheta) = (1 + r\cos\vartheta, 0, r\sin\vartheta),$$

where $\vartheta \in [0, 2\pi)$. A compatible surface $\Sigma_{T^2} : \mathbb{R}^2 \to T^2$ is

$$\Sigma_{T^2}(x, y) = S_{T^2}(2\pi c_1(x, y)), 2\pi c_2(x, y)))$$

$$c_1(x, y) = \varrho(4\alpha(x), \alpha(y)) - \varrho(4\alpha(x) - 2, \alpha(y)),$$

$$c_2(x, y) = \varrho(4\alpha(x) - 1, \alpha(y)) - \varrho(4\alpha(x) - 3, \alpha(y)),$$

where $\alpha : \mathbb{R} \to [0,1]$ is a function with the same properties as before and $\varrho : \mathbb{R} \times [0,1] \to [0,1]$ is the function given by

$$\varrho(s, t) = t g_\beta\Big(\frac{1 - 2s}{(1 + s - t)(2 - s - t)}\Big), \tag{71}$$

where $g_\beta(w) = 1/(\exp(\beta w) + 1)$ with $\beta > 0$ is the Fermi–Dirac function.

Upon unfolding the torus T^2 into a square I^2 by cutting it along the a– and b–cycle, the surface Σ_{T^2} describes a parametrized family of closed curves on I^2 which spring from one corner of the square and sweep it all eventually approximating the square's boundary.

A surface furnishing a natural parametrization of a given marked S–knot can now be constructed. To this end, we need to identify a curve in M that matches the fundamental polygon of S.

View M as a C–marked manifold and let

$$\gamma_{Mi} = \gamma_{\zeta_{Mi}} \quad \text{that is} \quad \alpha_{Mr} = \gamma_{\xi_{Mr}},\ \beta_{Mr} = \gamma_{\eta_{Mr}}, \tag{72}$$

Then, the fundamental polygon of the marking of M is the curve

$$\tau_M = \beta_{M\ell_S}{}^{-1\circ} \circ \alpha_{M\ell_S}{}^{-1\circ} \circ \beta_{M\ell_S} \circ \alpha_{M\ell_S} \circ \cdots \circ \beta_{M1}{}^{-1\circ} \circ \alpha_{M1}{}^{-1\circ} \circ \beta_{M1} \circ \alpha_{M1} \tag{73}$$

if $\ell_S = 0$, $\tau_M = \iota_{p_M}$.

A surface furnishing a natural parametrization of a given marked S–knot can now be constructed.

With every marked S–knot Ξ, there is associated a surface $\Sigma_\Xi : \iota_{p_M} \Rightarrow \tau_M$ given by

$$\Sigma_\Xi = \Xi \circ \Sigma_S. \tag{74}$$

Note that $\tau_M = \Xi \circ \tau_S$.

For a marked C–knot ξ, the source and target of the associated curve $\gamma_\xi : p_M \to p_M$ are equal. In gauge theory, this ensures nice ambient isotopy and gauge covariance properties of C–knot holonomy. For a genus $\ell_S = 0$ marked S–knot Ξ, the source and target of the associated surface $\Sigma_\Xi : \iota_{p_M} \Rightarrow \iota_{p_M}$ are equal as well. In higher gauge theory, this also ensures nice ambient isotopy and gauge covariance properties of S–knot holonomy. However, for a genus $\ell_S > 0$ marked S–knot Ξ, the source and target of the associated surface $\Sigma_\Xi : \iota_{p_M} \Rightarrow \tau_M \neq \iota_{p_M}$ are different. This is likely to be a problem for ambient isotopy and gauge covariance properties of holonomy. We have a proposal for the solution of this difficulty.

For given ℓ_S and C–marking of M, pick a reference marked S–knot Δ_M (e. g. Hosokawa's and Kawauchi's surface unknots in S^4 [35]).

The normalized surface of a marked S–knot Ξ is the surface $\Sigma^\sharp{}_\Xi : \iota_{p_M} \Rightarrow \iota_{p_M}$ given by

$$\Sigma^\sharp{}_\Xi = \Sigma_M{}^{-1\bullet} \bullet \Sigma_\Xi, \tag{75}$$

where $\Sigma_M := \Sigma_{\Delta_M}$ and $\bullet$ denotes vertical surface composition (cf. Section 2).

An intuitive way of thinking of the normalized surface of Ξ is as a surface characterizing the S–knot "ratio" of Ξ to Δ_M, with Δ_M acting as a normalizing knot.

The normalized surface of a marked S–knot has nice properties.

$\Sigma^\sharp{}_\Xi$ is independent from the choice of Σ_S and γ_C up to thin homotopy.

Note that the markings p_C and (p_S, ζ_{Si}) are fixed.

Ambient isotopic reference S–knots yield homotopic normalized marked S–knot surfaces.

If the reference marked S–knots Δ_{M0}, Δ_{M1} are ambient isotopic, then for every marked S–knot Ξ the normalized surfaces $\Sigma^\sharp{}_{\Xi|0}$, $\Sigma^\sharp{}_{\Xi|1}$ are homotopic.

Ambient isotopic marked S–knots have homotopic normalized surfaces.

If Ξ_0, Ξ_1 are ambient isotopic marked S–knots, then the normalized surfaces $\Sigma^\sharp{}_{\Xi_0}$, $\Sigma^\sharp{}_{\Xi_1}$ are homotopic.

As for C–knots, it should be possible to alter the marking changing $\Sigma^\sharp{}_\Xi$ by a (thin) homotopy.

Two marked S–knots Ξ_0, Ξ_1 with respect to distinct S–markings (p_{M0}, ζ_{M0i}), (p_{M1}, ζ_{M1i}) of M are said to be freely ambient isotopic if there is a smooth family $F_z \in$ $\mathrm{Diff}_+(M)$, $z \in \mathbb{R}$, of orientation preserving diffeomorphisms such that

$$F_0 = \mathrm{id}_M, \quad \Xi_1 = F_1 \circ \Xi_0. \tag{76}$$

Notice that above the S–marking (p_S, ζ_{Si}) of S is kept fixed.

Two pairs Ξ_0, Ξ_1 and $\Xi_0{}'$, $\Xi_1{}'$ of freely ambient isotopic marked S–knots are called concordant if there exist ambient isotopies F_z of Ξ_0, Ξ_1 and $F'{}_z$ of $\Xi_0{}'$, $\Xi_1{}'$ s. t. $F_z(p_{M0}) = F'{}_z(p_{M0})$, $F_z \circ \zeta_{M0i} = F'{}_z \circ \zeta_{M0i}$.

Freely ambient isotopic marked S–knots have homotopic normalized surfaces up to conjugation under concordance with reference knots.

Suppose the reference marked S–knots Δ_{M0}, Δ_{M1} are freely ambient isotopic. If the marked S–knots Ξ_0, Ξ_1 are freely ambient isotopic concordantly with Δ_{M0}, Δ_{M1}, then there is curve a $\gamma_1 : p_{M0} \to p_{M1}$ such that $\Sigma^\sharp{}_{\Xi_0|0}$, $I_{\gamma_1}{}^{-1\circ} \circ \Sigma^\sharp{}_{\Xi_1|1} \circ I_{\gamma_1}$ are homotopic.

Before stating the next result, we recall the following property. For two S–markings (p_{S0}, ζ_{S0i}), (p_{S1}, ζ_{S1i}) of S, there is an orientation preserving ambient isotopy k_z of S such that $k_1(p_{S0}) = p_{S1}$, $k_1 \circ \zeta_{S0i} = \zeta_{S1i}$.

If the embeddings $\Delta_M, \Xi : S \to M$ are simultaneously the reference and considered marked S–knot with respect to two distinct S–markings (p_{S0}, ζ_{S0i}), (p_{M0}, ζ_{M0i}) and (p_{S1}, ζ_{S1i}), (p_{M1}, ζ_{M1i}) of S and M and there is an ambient isotopy k_z of S shifting $\{p_{S0}, \zeta_{S0i}\}$ to $\{p_{S1}, \zeta_{S1i}\}$ such that $\Xi \circ k_z(p_{S0}) = \Delta_M \circ k_z(p_{S0})$ and $\Xi \circ k_z \circ \zeta_{S0i} = \Delta_M \circ k_z \circ \zeta_{S0i}$, then there is a curve $\gamma_1 : p_{M0} \to p_{M1}$ lying in the image $\Xi(S)$ such that $\Sigma^\sharp{}_{\Xi|0}$, $I_{\gamma_1}{}^{-1\circ} \circ \Sigma^\sharp{}_{\Xi|1} \circ I_{\gamma_1}$ are thinly homotopic.

Relying on the above results, we can now tackle the task of constructing higher knot holonomy.

5 C– and S–knot holonomy

Our aim is constructing holonomy invariants of knots up to conjugation. We begin with reviewing how this is done for C–knots

We let G be a Lie group and θ be a flat G–connection on M. Further, we fix C–markings p_C and p_M of C and M, respectively.

The holonomy of a marked C–knot is built out of the curve associated to the knot.

The holonomy of a marked C–knot ξ is the element $F_\theta(\xi) \in G$ given by

$$F_\theta(\xi) = F_\theta(\gamma_\xi), \tag{77}$$

where $\gamma_\xi : p_M \to p_M$ curve of ξ (cf. Section 4) and F_θ is the parallel transport functor (cf. Section 3).

C–knot holonomy is independent from the choice of parametrization.

For any marked C–knot ξ, $F_\theta(\xi)$ is independent of the choice of the compatible curve γ_C of C.

C–knot holonomy is further invariant under ambient isotopy.

If ξ_0, ξ_1 are ambient isotopic marked C–knots of M, then

$$F_\theta(\xi_1) = F_\theta(\xi_0). \tag{78}$$

This property generalizes as follows. Fix the C–marking p_C of C but allow two distinct C–markings p_{M0}, p_{M1} of M.

If ξ_0, ξ_1 are freely ambient isotopic marked C–knots, then there exists a curve $\gamma_1 : p_{M0} \to p_{M1}$ of M such that

$$F_\theta(\xi_1) = F_\theta(\gamma_1)F_\theta(\xi_0)F_\theta(\gamma_1)^{-1}.$$

C–knot holonomy is independent of the way a given C–knot is marked up to conjugation.

If ξ is a marked C–knot with respect to two distinct C–markings p_{C0}, p_{M0} and p_{C1}, p_{M1} of C and M, then there is a curve $\gamma_1 : p_{M0} \to p_{M1}$ lying in $\xi(C)$ such that

$$F_{\theta|1}(\xi) = F_\theta(\gamma_1)F_{\theta|0}(\xi)F_\theta(\gamma_1)^{-1}.$$

C–knot holonomy is also gauge covariant as desired.

Let ξ be a marked C–knot of M. Then, for any G–gauge transformation g, one has

$$F_{{}^g\theta}(\xi) = g(p_M)F_\theta(\xi)g(p_M)^{-1}.$$

In summary, C–knot holonomy is C–marking and gauge independent and isotopy invariant up to G–conjugation.

Next, using the treatment of C–knot holonomy presented above as a model, we illustrate the construction of S–knot holonomy.

We let (G,H) be a Lie crossed module and (θ,Υ) be a flat (G,H)–2–connection pair on M. Furthermore, we fix S-markings (p_S,ζ_{Si}) and (p_M,ζ_{SMi}) of S and M, respectively.

The holonomy of a marked S–knot Ξ is the element $F_\theta(\Xi) \in H$ given by

$$F_{\theta,\Upsilon}(\Xi) = F_{\theta,\Upsilon}(\Sigma^\sharp{}_\Xi) = F_{\theta,\Upsilon}(\Sigma_M)^{-1}F_{\theta,\Upsilon}(\Sigma_\Xi), \tag{79}$$

where $\Sigma^\sharp{}_\Xi : \iota_{p_M} \Rightarrow \iota_{p_M}$ is the normalized surface of Ξ and $F_{\theta,\Upsilon}$ is the parallel transport 2–functor.

The fact that $\Sigma^\sharp{}_\Xi : \iota_{p_M} \Rightarrow \iota_{p_M}$ has the following crucial consequence.

For a marked S–knot Ξ,

$$t(F_{\theta,\Upsilon}(\Xi)) = 1_G. \tag{80}$$

Thus, $F_{\theta,\Upsilon}(\Xi) = 1_H$ unless $\ker t \neq \{1_H\}$. Further, $F_{\theta,\Upsilon}(\Xi) \in Z_H$.

Thus, unlike C–knot holonomy, S–knot holonomy is fundamentally Abelian and non trivial only for crossed modules whose target map has non trivial kernel.

S–knot holonomy is independent from the choice of parametrization.

For every marked S–knot Ξ, $F_{\theta,\Upsilon}(\Xi)$ is independent from the choice of the compatible surface Σ_S of S and curve γ_C of C.

S–knot holonomy is invariant under a change of the reference marked S–knots in the following sense.

If the reference marked S–knots Δ_{M0}, Δ_{M1} are ambient isotopic, then for any marked S–knot Ξ

$$F_{\theta,\Upsilon|0}(\Xi) = F_{\theta,\Upsilon|1}(\Xi). \tag{81}$$

S–knot holonomy is further invariant under ambient isotopy.

If Ξ_0, Ξ_1 are ambient isotopic marked S–knots, then

$$F_{\theta,\Upsilon}(\Xi_1) = F_{\theta,\Upsilon}(\Xi_0). \tag{82}$$

This property generalizes as follows. Fix the S-markings (p_S, ζ_{Si}) of S but allow distinct S–marking (p_{M0}, ζ_{M0i}), (p_{M1}, ζ_{M1i}) of M.

Suppose Δ_{M0}, Δ_{M1} are freely ambient isotopic reference marked S–knots. If the marked S–knots Ξ_0, Ξ_1 are freely ambient isotopic concordantly with Δ_{M0}, Δ_{M1}, then there is a curve $\gamma_1 : p_{M0} \to p_{M1}$ such that

$$F_{\theta,\Upsilon|1}(\Xi_1) = m(F_\theta(\gamma_1))(F_{\theta,\Upsilon|0}(\Xi_0)). \tag{83}$$

S–knot holonomy is independent of the way a given knot S–knot is marked up to conjugation.

If the embeddings $\Delta_M, \Xi : S \to M$ are simultaneously the reference and considered marked S–knot with respect to two distinct S–markings (p_{S0}, ζ_{S0i}), (p_{M0}, ζ_{M0i}) and (p_{S1}, ζ_{S1i}), (p_{M1}, ζ_{M1i}) of S and M and there is an ambient isotopy k_z of S shifting $\{p_{S0}, \zeta_{S0i}\}$ to $\{p_{S1}, \zeta_{S1i}\}$ such that $\Xi \circ k_z(p_{S0}) = \Delta_M \circ k_z(p_{S0})$ and $\Xi \circ k_z \circ \zeta_{S0i} = \Delta_M \circ k_z \circ \zeta_{S0i}$, then there is a curve $\gamma_1 : p_{M0} \to p_{M1}$ lying in the image $\Xi(S)$ such that

$$F_{\theta,\Upsilon|1}(\Xi) = m(F_\theta(\gamma_1))(F_{\theta,\Upsilon|0}(\Xi)). \tag{84}$$

In this higher gauge theoretic set–up, one can define also C–knot holonomy in the same way as before.

The holonomy of a marked C–knot ξ is the element $F_\theta(\xi) \in G$ given by

$$F_\theta(\xi) = F_\theta(\gamma_\xi) \tag{85}$$

is defined.

This C–knot holonomy has however weaker properties than in ordinary gauge theory.

C–knot holonomy is still independent of the choice of parametrization.

For any marked C–knot ξ, $F_\theta(\xi)$ is independent of the choice of the compatible curve γ_C of C.

Since, however, θ is not flat unless $\dot{t}(\Upsilon) = 0$, $F_\theta(\xi)$ is not ambient isotopy invariant.

If ξ_0, ξ_1 are two ambient isotopic marked C–knots of M, then there is a surface $\Sigma : \gamma_{\xi_0} \Rightarrow \gamma_{\xi_1}$ of M such that

$$F_\theta(\xi_1) = t(F_{\theta,\Upsilon}(\Sigma))F_\theta(\xi_0). \tag{86}$$

This property generalizes as follows. Fix the C–marking p_C of C but allow two distinct C–markings p_{M0}, p_{M1} of M.

If ξ_0, ξ_1 are freely ambient isotopic marked C–knots, then there exist a curve $\gamma_1 : p_{M0} \to p_{M1}$ and a surface $\Sigma : \gamma_{\xi_0} \Rightarrow {\gamma_1}^{-1} \circ \gamma_{\xi_1} \circ \gamma_1$ of M such that

$$F_\theta(\xi_1) = F_\theta(\gamma_1)\,t(F_{\theta,\Upsilon}(\Sigma))F_\theta(\xi_0)F_\theta(\gamma_1)^{-1}. \tag{87}$$

C–knot holonomy is again independent of the way a given C–knot is marked up to conjugation.

If ξ is a marked C–knot with respect to two distinct C–markings p_{C0}, p_{M0} and p_{C1}, p_{M1} of C and M, then there is $\gamma_1 : p_{M0} \to p_{M1}$ curve in $\xi(C)$ such that

$$F_{\theta|1}(\xi) = F_\theta(\gamma_1)F_{\theta|0}(\xi)F_\theta(\gamma_1)^{-1}. \tag{88}$$

Also in higher gauge theory, S– and C–knot holonomy is gauge covariant in the appropriate sense.

Let Ξ be a marked S–knot and ξ a marked C–knot. If (g, J) is a (G, H)–1–gauge transformation, then

$$F_{{}^{g,J}\theta,{}^{g,J}\Upsilon}(\Xi) = m(g(p_M))(F_{\theta,\Upsilon}(\Xi)) \tag{89}$$

and

$$F_{{}^{g,J}\theta}(\xi) = g(p_M)\,t(G_{g,J;\theta}(\gamma_\xi))F_\theta(\gamma)g(p_M)^{-1}, \tag{90}$$

where $G_{g,J;\theta}(\gamma_\xi)$ is the gauge parallel transport along γ_ξ defined in Section 3.

To summarize, C– and S–knot holonomy are C–marking and gauge independent and isotopy invariant up to the appropriate form of crossed module conjugation. We shall analyze this point in greater depth in the next section.

6 Invariant traces

Having applications to knot topology in mind, we aim at a construction of holonomy invariants. This requires working out invariant traces.

We let G again be a Lie group and θ be a flat G–connection on M. Further, we let C–markings p_C and p_M of C and M, respectively, be given.

We have seen in Section 5 that for a C–knot ξ, its holonomy $F_\theta(\xi)$ is C–marking and isotopy invariant and gauge independent up to G–conjugation, that is

$$F_\theta(\xi) \equiv aF_\theta(\xi)a^{-1} \tag{91}$$

for $a \in G$.

There is a well established way of extracting C–knot invariants from knot holonomy.

The Wilson line

$$W_{R,\theta}(\xi) = \mathrm{tr}_R(F_\theta(\xi)), \tag{92}$$

with R a representation of G provides a C–knot invariant.

Next, taking the procedure just reviewed to construct C–knot holonomy invariants as a model, we propose a systematic way to build S–knot holonomy invariants.

We let (G, H) be a Lie crossed module and (θ, Υ) be a flat (G, H)–2–connection pair on M. Furthermore, we let S-markings (p_S, ζ_{Si}) and (p_M, ζ_{SMi}) of S and M, respectively, be given.

In Section 5, we have also seen that for a C–knot ξ and an S–knot Ξ, the holonomy $F_\theta(\xi)$ and $F_{\theta,\Upsilon}(\Xi)$ is C– and S–marking and isotopy invariant and gauge independent up to (G, H)–conjugation

$$F_\theta(\xi) \equiv aF_\theta(\xi)a^{-1}t(A), \tag{93}$$

$$F_{\theta,\Upsilon}(\Xi) \equiv m(a)(F_{\theta,\Upsilon}(\Xi)) \tag{94}$$

with $(a, A) \in G \times H$. (G, H)–conjugation is defined by

$$u' = aua^{-1}t(A), \qquad U' = m(a)(U) \tag{95}$$

with $(u, U), (u', U'), (a, A) \in G \times H$ and is an equivalence relation.

To obtain knot invariants, one needs traces invariant under (G, H)–conjugation. To this end, one could proceed as follows.

Assume G, H are compact with bi-invariant Haar measures μ_G, μ_H. Pick representations R, S of G, H. Set

$$\mathrm{tr}_{R,S|b}(u) = \int_H d\mu_H(X)\,\mathrm{tr}_R(ut(X)), \tag{96}$$

$$\mathrm{tr}_{R,S|f}(U) = \int_G d\mu_G(x)\,\mathrm{tr}_S(m(x)(U)), \tag{97}$$

$(u, U) \in G \times H$.

The traces $\mathrm{tr}_{R,S|b}$, $\mathrm{tr}_{R,S|f}(U)$ are invariant under (G, H) conjugation,

$$\mathrm{tr}_{R,S|b}(aua^{-1}t(A)) = \mathrm{tr}_{R,S|b}(u), \tag{98}$$

$$\mathrm{tr}_{R,S|f}(m(a)(U)) = \mathrm{tr}_{R,S|f}(U) \tag{99}$$

for $(u, U), (a, A) \in G \times H$.

These invariant traces can be used to extract C– and S–knot invariants from knot holonomy as in the ordinary case.

The Wilson line and surface

$$W_{R,S,\theta|b}(\xi) = \mathrm{tr}_{R,S|b}(F_\theta(\xi)), \tag{100}$$

$$W_{R,S,\theta,\Upsilon|f}(\Xi) = \mathrm{tr}_{R,S|f}(F_{\theta,\Upsilon}(\Xi)). \tag{101}$$

provide a C– and S–knot invariant.

There is a problem with this way of proceeding. The traces may be trivial. For instance, if $t(H) = G$, $\mathrm{tr}_{R,Sb}(u)$ does not depend on u and $\mathrm{tr}_{R,Sf}(U) = \mathrm{tr}_S(U)$ for $U \in \ker t$ (the case of interest for surface knots).

In ordinary gauge theory with gauge group G, a trace is a map $\mathrm{tr}: G \to \mathbb{C}$ invariant under the action

$$a \triangleright u := aua^{-1} \tag{102}$$

with $a, u \in G$, that is

$$\mathrm{tr}(a \triangleright u) = \mathrm{tr}(u). \tag{103}$$

If G is a compact Lie group, then tr reduces to a linear combination of ordinary traces tr_R associated with the irreducible representations R of G.

What matters is not the group structure of G but its conjugation structure codified in the conjugation pointed quandle of G.

A pointed quandle is a set G with a binary operation $\triangleright : G \times G \to G$ and a distinguished element $1_G \in G$ such that

$$a \triangleright a = a, \tag{104}$$

$$a \triangleright (b \triangleright c) = (a \triangleright b) \triangleright (a \triangleright c) \tag{105}$$

with $a, b, c \in G$. Further, the map $a \triangleright \cdot : G \to G$ is invertible for any $a \in G$ and

$$a \triangleright 1_G = 1_G, \quad 1_G \triangleright a = a \tag{106}$$

for $a \in G$.

In higher gauge theory with gauge crossed module (G, H), a similar point of view is appropriate. A trace pair is a pair of maps $\mathrm{tr}_b : G \to \mathbb{C}$, $\mathrm{tr}_f : H \to \mathbb{C}$ invariant under the action

$$a \triangleright u := aua^{-1}, \tag{107}$$

$$A \succ u := ut(A), \tag{108}$$

$$a \triangleright U := m(a)(U) \tag{109}$$

with $a, u \in G$, $A, U \in H$, that is

$$\mathrm{tr}_b(a \triangleright u) = \mathrm{tr}_b(u), \tag{110}$$

$$\mathrm{tr}_b(A \succ u) = \mathrm{tr}_b(u), \tag{111}$$

$$\mathrm{tr}_f(a \triangleright U) = \mathrm{tr}_f(U). \tag{112}$$

What matters is not (G, H) itself but its conjugation augmented pointed quandle crossed module (G, H) [30, 31, 13]:

An augmented pointed quandle crossed module is a pair of sets G, H endowed with three operations $\triangleright : G\times G \to G$, $H\times H \to H$, $G\times H \to H$ and distinguished elements $1_G \in G$, $1_H \in H$ such that

i) G is a pointed quandle,

ii) H is a pointed quandle

and the following requirements are satisfied.

The relations

$$a \triangleright (b \triangleright A) = (a \triangleright b) \triangleright (a \triangleright A), \tag{113}$$

$$a \triangleright (A \triangleright B) = (a \triangleright A) \triangleright (a \triangleright B) \tag{114}$$

with $a, b \in G$, $A, B \in H$ hold.

For any $a \in G$, the map $a \triangleright \cdot : H \to H$ is invertible.

For $a \in G$, $A \in H$, the relations

$$1_G \triangleright A = A, \tag{115}$$

$$a \triangleright 1_H = 1_H \tag{116}$$

are fulfilled.

Further, a quandle morphism $\alpha : H \to G$ (a map respecting $\triangleright$ and 1) is given such that

$$\alpha(a \triangleright A) = a \triangleright \alpha(A)\ , \tag{117}$$

$$\alpha(A) \triangleright B = A \triangleright B \tag{118}$$

with $a \in G$, $A, B \in H$.

Finally, an augmentation map $\succ : H\times G \to G$ is given with the following properties.

For $a, b \in G$, $A \in H$,

$$a \triangleright (A \succ b) = (a \triangleright A) \succ (a \triangleright b)\ . \tag{119}$$

For $A \in H$, $A \succ \cdot : G \to G$ is invertible.

For $a \in G$, $A \in H$

$$A \succ 1_G = \alpha(A), \tag{120}$$

$$1_H \succ a = a. \tag{121}$$

Notice that α is the quandle crossed module analog of the group crossed module target morphisms. In particular (118) is the quandle counterpart of the Peiffer identity.

The following question is still open. If G, H are compact, does a trace pair tr_b, tr_f reduce to linear combinations of traces $\mathrm{tr}_{R,S|b}$, $\mathrm{tr}_{R,S|f}$ of the form (96), (97) with R, S irreducible representations of G, H, respectively?

7 Higher Chern–Simons theory

To compute knot invariants in quantum field theory, one needs Chern–Simons theory. This has been known for a long time since Witten's 1988 paper [7].

Chern–Simons theory is a Schwarz type topological gauge theory on a closed 3-dimensional manifold M_3. Suppose that G is the gauge group and $\mathfrak{g}$ is its Lie algebra. Suppose further that $\mathfrak{g}$ is equipped with a properly normalized invariant non singular bilinear form $(\cdot,\cdot) : \mathfrak{g}\times\mathfrak{g} \to \mathfrak{g}$ so that

$$([z,x],y) + x,[z,y]) = 0 \tag{122}$$

with $x, y, z \in \mathfrak{g}$.

The Chern–Simons action is given by

$$\mathrm{CS}(\theta) = \frac{k}{4\pi}\int_{M_3}\Big(\theta, \mathrm{d}\theta + \frac{1}{3}[\theta,\theta]\Big) \tag{123}$$

with θ a G–connection. The coefficient k is called level.

The Chern–Simons field equations are equivalent to the flatness condition of θ (cf. Equation (20)):

$$\mathrm{d}\theta + \frac{1}{2}[\theta,\theta] = 0. \tag{124}$$

The Chern–Simons action is invariant under a G–gauge transformations g only modulo $2\pi\mathbb{Z}$:

$$\mathrm{CS}({}^g\theta) = \mathrm{CS}(\theta) - 2\pi k\cdot \mathrm{wn}(g), \tag{125}$$

where $\mathrm{wn}(g)$ *is the winding number of g.*

Quantum gauge invariance holds if the level k is integer.

Chern–Simons correlators of Wilson loop $W_{R,\theta}(\xi)$ yield knot invariants, for instance:

$G = \mathrm{SU}(2)$, $R = F \Rightarrow$ Jones polynomial;

$G = \mathrm{SU}(n)$, $R = F \Rightarrow$ HOMFLY polynomial;

$G = \mathrm{SO}(n)$, $R = F \Rightarrow$ Kauffman polynomial...

In the Chern–Simons path integral, θ is not flat and consequently $W_{R,\theta}(\xi)$ is not ambient isotopy invariant. However, the theory somehow localizes on the moduli space of flat connections even though it is not a cohomological topological field theory. This has been proven by Beasley and Witten for M_3 Seifert, e. g. $S^1\times S^2$, S^3, ... Therefore, Chern–Simons Wilson loop correlators $W_{R,\theta}(\xi)$ furnish genuine knot invariants.

In order to compute surface knots invariants in quantum field theory, one needs a higher version of Chern–Simons theory, 2-Chern–Simons theory. We have a proposal for such a model. There are however unsolved problems to be discussed below.

2–Chern–Simons theory is a Schwarz type topological gauge theory on a closed 4-dimensional manifold M_4. Assume that $H \xrightarrow{t} G \xrightarrow{m} \mathrm{Aut}(H)$ is the gauge Lie crossed module and that $\mathfrak{h} \xrightarrow{\dot{t}} \mathfrak{g} \xrightarrow{\widehat{m}} \mathfrak{Der}(\mathfrak{h})$ is its differential Lie crossed module. Assume further $(\mathfrak{g}, \mathfrak{h})$ is equipped with a properly normalized invariant non singular bilinear pairing $(\cdot,\cdot) : \mathfrak{g} \times \mathfrak{h} \to \mathbb{R}$ such that

$$(\dot{t}(X), Y) - (\dot{t}(Y), X) = 0,$$

$$([y, x], X) + (x, \widehat{m}(y)(X)) = 0$$

with $x, y \in \mathfrak{g}$, $X, Y \in \mathfrak{h}$. Note that this requires that the crossed module is balanced, that is $\dim \mathfrak{g} = \dim \mathfrak{h}$.

The 2–Chern–Simons action is given by

$$\mathrm{CS}_2(\theta, \Upsilon) = \kappa_2 \int_{M_4} \Big(\mathrm{d}\theta + \frac{1}{2}[\theta, \theta] - \frac{1}{2}\dot{t}(\Upsilon), \Upsilon\Big), \tag{126}$$

with (θ, Υ) a (G, H) 2–preconnection [34, 33]. κ_2 is a coefficient analog to level.

A (G, H) 2–preconnection is just a pair $(\theta, \Upsilon) \in \Omega^1(M_4, \mathfrak{g}) \times \Omega^2(M_4, \mathfrak{h})$. (θ, Υ) a (G, H)–2–connection if in addition it satisfies the vanishing fake curvature condition (29).

The 2–Chern–Simons field equations are equivalent to (θ, Υ) being a flat (G, H)–2–connection (cf. Equation (29)), (30)):

$$\mathrm{d}\theta + \frac{1}{2}[\theta, \theta] - \dot{t}(\Upsilon) = 0, \tag{127}$$

$$\mathrm{d}\Upsilon + \widehat{m}(\theta, \Upsilon) = 0. \tag{128}$$

Thus, at once, (θ, Υ) satisfies the vanishing fake curvature condition, which makes it a genuine (G, H)–2–connection, and the vanishing curvature condition, which characterizes as a flat one.

This is quite nice, but it signals a potential problem for the construction of 2–Chern–Simons theory as a full quantum field theory. Since (θ, Υ) does not obey the zero fake curvature condition in the 2–Chern–Simons path integral, the insertion of Wilson surfaces $W_{R,S,\theta,\Upsilon}(\Xi)$ of surface knots Ξ in the path integral is problematic, as the definition of the $W_{R,S,\theta,\Upsilon}(\Xi)$ requires that condition in a basic way.

Another unexpected feature of the model concerns 1–gauge invariance.

The 2-Chern–Simons action is invariant under (G, H)–1–gauge transformation (g, J),

$$\mathrm{CS}_2({}^{g,J}\theta, {}^{g,J}\theta\Upsilon) = \mathrm{CS}_2(\theta, \Upsilon). \tag{129}$$

In 2–Chern–Simons theory, there is no shift by some kind of higher winding number such to cause level quantization as in the ordinary Chern–Simons model. This surprising and somewhat disappointing finding can be explained by hypothesising that either all (G, H)–1–gauge transformations (g, J) are small unlike ordinary gauge transformation or that we are missing all the topologically non trivial (G, H)–1–gauge transformations. This is still an open problem.

In spite of these open issues, the possibility of obtaining surface knot invariants as correlators of Wilson surface insertion in 2–Chern–Simons theory remains an intriguing possibility. Here are further reasons for this.

Studying pull–backs of knots may be interesting.

All orientation preserving diffeomorphisms $f \in \mathrm{Diff}_+(C)$ of the circle C are homotopic to id_C. Consequently, for a C–knot ξ, the curves γ_ξ, $\gamma_{f^*\xi}$ are thinly homotopic and the C–knots ξ and $f^*\xi$ have the same holonomy.

Conversely, for a higher genus surface S, not all orientation preserving diffeomorphisms $f \in \mathrm{Diff}_+(S)$ of S are homotopic to id_S. Consequently, for a S–knot Ξ, the normalized surfaces $\Sigma^\sharp{}_\Xi$, $\Sigma^\sharp{}_{f^*\Xi}$ are not thinly homotopic and the S–knots Ξ and $f^*\Xi$ do not have the same holonomy in general

This suggests that S–knot invariants computed using higher gauge theory may have interesting covariance properties under the mapping class group

$$\mathrm{MCG}_+(S) = \mathrm{Diff}_+(S) / \mathrm{Diff}_0(S), \tag{130}$$

about which there exists a well–developed mathematical theory.

Acknowledgements. We thank R. Picken for his interest in the subject and for many useful discussions.

Key words. Higher gauge theory, higher Chern-Simons theory, higher knots, A field theoretic route to higher knots

References

[1] L. H. Kauffman, *Knots and physics,* World Scientific, Singapore, 1991.

[2] C. C. Adams, *The knot book: an elementary introduction to the mathematical theory of knots,* American Mathematical Society, Providence, 2004.

[3] W. R. Lickorish, *An introduction to knot theory,* Springer, 1997.

[4] J. S. Carter and M. Saito, *Knotted surfaces and their diagrams*, American Mathematical Society, Providence, 1998.
[5] S. Kamada, *Braid and knot theory in dimension four*, American Mathematical Society, Providence, 2002.
[6] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D **10** (1974) 2445.
[7] E. Witten, *Quantum field theory and the Jones polynomial*, Commun. Math. Phys. **121** (1989) 351.
[8] I. Chepelev, *Non-Abelian Wilson surfaces*, JHEP **0202** (2002) 013 [hep-th/0111018].
[9] A. Alekseev, O. Chekeres, and P. Mnev, *Wilson surface observables from equivariant cohomology*, JHEP **1511** (2015) 093 [1507.06343 [hep-th]].
[10] J. C. Baez, *Higher Yang–Mills theory*, hep-th/0206130.
[11] J. C. Baez and J. Huerta, *An invitation to higher gauge theory*, Gen. Relativ. Gravit. **43** (2011) 2335 [1003.4485 [hep-th]].
[12] R. Zucchini, *On higher holonomy invariants in higher gauge theory I*, Int. J. Geom. Meth. Mod. Phys. **13** (2016) 1650090 [1505.02121 [hep-th]].
[13] R. Zucchini, *On higher holonomy invariants in higher gauge theory II*, Int. J. Geom. Meth. Mod. Phys. **13** (2016) 1650091 [1505.02122 [hep-th]].
[14] A. Caetano and R. F. Picken, *An axiomatic definition of holonomy*, Int. J. Math. **5** (1994) 835.
[15] J. C. Baez and U. Schreiber, *Higher gauge theory: 2-connections on 2-bundles*, hep-th/0412325.
[16] J. C. Baez and U. Schreiber, *Higher gauge theory*, Contemp. Math. **431** (2007) 7 [math.DG/0511710].
[17] U. Schreiber and K. Waldorf, *Parallel transport and functors*, J. Homot. Relat. Struct. **4** (2009) 187 [0705.0452 [math.DG]].
[18] U. Schreiber and K. Waldorf, *Smooth functors vs. differential forms*, Homol. Homot. App. **13** (2011) 143 [0802.0663 [math.DG]].
[19] U. Schreiber and K. Waldorf, *Connections on non-Abelian gerbes and their holonomy*, Theor. Appl. Categor. **28** (2013) 476 [0808.1923 [math.DG]].
[20] J. F. Martins and R. Picken, *On two-dimensional holonomy*, Trans. Am. Math. Soc. **362** (2010) 5657 [0710.4310 [math.DG]].
[21] J. F. Martins and R. Picken, *Surface holonomy for non-Abelian 2-bundles via double groupoids*, Adv. Math. **226** (2011) 3309 [0808.3964 [math.CT]].
[22] S. Chatterjee, A. Lahiri, and A. N. Sengupta, *Parallel transport over path spaces*, Rev. Math. Phys. **22** (2010) 1033 [0906.1864 [math-ph]].
[23] S. Chatterjee, A. Lahiri, and A. N. Sengupta, *Double category related to path space parallel transport and representations of Lie 2-groups*, Springer Proc. Math. Stat. **85** (2014) 379.
[24] S. Chatterjee, A. Lahiri, and A. N. Sengupta, *Path space forms and surface holonomy*, AIP Conf. Proc. **1191** (2009) 66 [1007.1159 [math-ph]].
[25] E. Soncini and R. Zucchini, *A new formulation of higher parallel transport in higher gauge theory*, J. Geom. Phys. **95** (2015) 28 [1410.0775 [hep-th]].
[26] H. Abbaspour and F. Wagemann, *On 2-holonomy*, 1202.2292 [math.AT].
[27] C. A. Abad and F. Schaetz, *Higher holonomies: comparing two constructions*, Diff. Geo. Appl. **40** (2015) 14 [1404.0729 [math.AT]].
[28] C. A. Abad and F. Schaetz, *Holonomies for connections with values in L_∞-algebras*, Homol. Homot. App. **16** (2014) 89 [1404.0727 [math.AT]].
[29] A. S. Cattaneo and C. A. Rossi, *Wilson surfaces and higher dimensional knot invariants*, Commun. Math. Phys. **256** (2005) 513 [math-ph/0210037].
[30] A. S. Crans, *Lie 2-algebras*, PhD thesis, University of California Riverside (2004) [math.QA/0409602].
[31] A. S. Crans and F. Wagemann, *Crossed modules of racks*, Homol. Homot. App. **16** (2014) 85 [1310.4705 [math.QA]].
[32] R. Zucchini, *AKSZ models of semistrict higher gauge theory*, JHEP **1303** (2013) 014 [1112.2819 [hep-th]].
[33] R. Zucchini, *A Lie based 4-dimensional higher Chern–Simons theory*, J. Math. Phys. **57** (2016) 052301 [1512.05977 [hep-th]].
[34] E. Soncini and R. Zucchini, *4d semistrict higher Chern–Simons theory I*, JHEP **1410** (2014) 79 [1406.2197 [hep-th]].
[35] F. Hosokawa and A. Kawauchi, *Proposals for unknotted surfaces in four-spaces*, Osaka J. Math. **16** (1979) 233.