THU-91/09

MORE (THOUGHTS ON) GRIBOV COPIES

Pierre van Baal *

Institute for Theoretical Physics
Princetonplein 5, P.O.Box 80.006
3508 TA Utrecht, The Netherlands

ABSTRACT

Inspired by an unambiguous observation of Gribov copies on the lattice an old continuum example due to Henyey is reanalysed in terms of bifurcations at the Gribov horizon. This gives yet another proof that there are in general Gribov copies within the horizon. Using results by Semenov-Tyan-Shanskii, Franke, Dell'Antonio and Zwanziger on a possible fundamental modular domain we argue that its boundary has both Gribov copies and points that coincide with the Gribov horizon.

May 1991; Revised August 1991

---

* KNAW fellow

1. Introduction

One way strong interactions in non-Abelian gauge theories manifest themselves is by a spreading of the wavefunctionals over such a large portion of configuration space, that the issue of Gribov copies [1] can no longer be ignored. To have a chance to understand the non-perturbative dynamics of the low-energy physics a better understanding of the physical configuration space is essential, particularly for describing how physics depends on the quantum numbers associated to the homotopically non-trivial gauge transformations. These issues are intimately connected to the dynamical questions involving tunnelling between different vacua. To make this a little more explicit we review an old argument due to Jackiw, Muzinich and Rebbi [2] in somewhat more generality. (In much of this paper we will be primarily interested in the Hamiltonian formulation, in which case the gauge fields are defined over a three dimensional space, but some of the discussion is relevant for the four dimensional situation too.) Let g be a homotopically non-trivial gauge transformation, such that $A_i = [g]\Theta_i = -ig\partial_i g^{-1}$ satisfies the Coulomb gauge $\partial_i A_i = 0$ (where throughout this paper Θ will denote the zero connection). Then the Faddeev-Popov operator (our gauge fields will be hermitian Lie-algebra elements and $\mathrm{ad}X(Y) \equiv [X, Y]$)

$$FP(A) = -\partial_i D_i(A) = -\partial_i(\partial_i + i\mathrm{ad}A_i) \qquad (1)$$

has a vanishing eigenvalue. The argument is simple: Constant gauge transformations leave the gauge condition invariant. The obviously related zero-modes, that correspond to constant Lie-algebra elements, are removed by defining the theory modulo constant gauge transformation and demanding the wavefunctionals to be trivial representations of the gauge group G. As g is a homotopically non-trivial gauge transformation, gXg^{-1} cannot be constant for all constant Lie-algebra elements X, but it is nevertheless a zero-mode of $FP(A)$. Thus the Gribov horizon will contain gauge copies of the classical vacuum, making it more urgent to understand the problem of gauge fixing.

It seemed therefore that Gribov copies are always associated to homotopically non-trivial gauge transformations. It was first demonstrated in an example by Henyey [3] that this is not always true. But probably the most simple example is provided by gauge fields on the torus in the abelian zero-momentum sector. For definiteness let us take $G = SU(2)$ (with the Pauli matrices τ_i as generators of the algebra and L the size of the torus) and

$A_i = \frac{C_i}{2L}\tau_3$, then the gauge transformation $g_{(k)} = \exp(-\pi i x_k \tau_3/L)$ maps C_k to $C_k + 2\pi$. As $g_{(k)}$ is anti-periodic it is homotopically non-trivial (they are 't Hooft's twisted gauge transformations [4]). However, $g^2_{(k)}$ will map C_k to $C_k + 4\pi$. As this gauge transformation is periodic and abelian, it is contractable to the identity and provides an example of gauge copies by homotopically trivial gauge transformations. This example also proves that in general there are gauge copies **within** the Gribov horizon, which occurs [5,6] at $|C_k| = 2\pi$. For example $\vec{C} = \vec{D} + (-\pi, 0, 0)$ is a gauge copy of $\vec{C} = \vec{D} + (\pi, 0, 0)$ and both occur inside the horizon if $|D_i| < \pi$. (It would seem the simplest case arises by taking $\vec{D} = 0$, however, in that case both configurations are related by a constant gauge transformation, which are to be divided out.)

Still, one might conjecture (as in the above example) that such copies within the horizon are always associated to homotopically non-trivial gauge transformations. Henyey's [3] example will be shown to provide an explicit example of gauge copies within the horizon, that are related by homotopically trivial gauge transformations (and that are related to a copy outside the horizon). In section 1 we will first give an argument based on Morse theory that in certain cases shows how, when moving from inside the horizon to outside (where the lowest eigenvalue of $FP(A)$ flips sign) two additional copies will be created for which $FP(A)$ is positive and which are therefore inside the horizon. As these copies coalesce at the horizon, the gauge transformation relating the two can be deformed to the identity. Our interest in this example was stirred by an unambiguous observation of Gribov copies on a lattice [7]. The structure of the relevant gauge transformations was analysed and some of them would have non-zero winding number, whereas others were typically of the form expected from ref. [3].

That there are Gribov copies within the Gribov horizon was first demonstrated by Semenov-Tyan-Shanskii and Franke [8] and analysed in more detail by Dell'Antonio and Zwanziger [9,10]. These authors also provide a recipe that will (almost) uniquely fix an element of the gauge orbit. In section 4 we make more precise in what sense this forms a fundamental modular domain. The interior is a convex subset of the transverse connections, that contains no Gribov copies. The boundary will contain Gribov copies, but can at most coincide with the horizon on a subset of the boundary of codimension one. The gauge transformations that relate the copies provides the identifications at the boundary that makes it into a fundamental modular domain for configuration space. These considerations are both relevant for the Hamiltonian formulation and for the recent path integral formulation on this fundamental modular domain by Zwanziger [11], Parinello and Jona-Lasinio [12]. In the case of SU(2) gauge theory on the torus the intersection of this fundamental modular domain with the constant abelian gauge fields is given by $|C_k| \leq \pi$ and $g_{(k)}$ provides the identification of the (opposite) points on the boundary. In this case the boundary is regular, but we will argue that in general so called singular boundary points (i.e. those that coincide with the horizon) do occur.

It should be mentioned that for the torus in the presence of fields in the fundamental representation (quarks) only periodic gauge transformations are allowed. In that case it is easily seen that the intersection of the fundamental modular domain with the constant abelian gauge fields is given by the domain $|C_k| \leq 2\pi$, whose boundary coincides with the Gribov horizon. Nevertheless, this domain is a covering space of the domain $|C_k| \leq \pi$ and fields in the fundamental representation can be defined on the latter if allowed to be multivalued (an alternative formulation in terms of coordinate patches is described in ref. [6]). Later we will see that even for pure gauge theories, the wavefunctionals are still multivalued, but in that case the multivaluedness is restricted to a phasefactor determined by the the periodicity condition that arises as a consequence of the identifications at the boundary of the fundamental modular domain. The relevant "Bloch momenta" give the topological quantum numbers.

In section 5 we conclude with a discussion on the implications for the Hamiltonian formulation. We also re-emphasize [13] that for supersymmetric gauge theories on a torus the Gribov copies are essential for a proper evaluation of the Witten index, despite the fact that one can take the coupling constant (and hence the volume) as small as one likes. The problem remains open (despite the claim made in ref. [14]) and turns out to be related to the non-trivial issue of constructing Dirac vacuum bundles that incorporate the identifications associated to the Gribov copies. It seems to require one to introduce a multivalued vacuum wavefunctional that incorporates the non-conservation of chiral particle number, induced by the chiral U(1) anomaly [15]. Note that this multivaluedness is again of a different nature as discussed in the previous paragraph (the fermions in supersymmetric gauge theories are in the adjoint representation). These issues certainly deserve further study. However, this paper will concentrate on the pure gauge sector.

2. Morse theory and bifurcation of copies at the horizon

The Coulomb gauge condition $\partial_i A_i = 0$ can be formulated in terms of an action principle [8,9,16] (M is the manifold over which the gauge theory is defined, which we will mostly take three dimensional and compact. For $\mathbf{R}^3$ we assume the fields to have an asymptotics that allows compactification to the three sphere.)

$$I(g;A) \equiv \|\,[g]A\,\|^2 = \int_M \mathrm{Tr}(\{[g]A_i\}^2) = \int_M \mathrm{Tr}(\{A_i + ig^{-1}\partial_i g\}^2), \tag{2}$$

whose critical points satisfy the gauge condition and whose Hessian is precisely the Faddeev-Popov operator. This is most easily established by observing that

$$I(hg;A) = I(h;[g]A). \tag{3}$$

Writing $h = e^X$ and using

$$e^{-X}\partial_i e^X = \frac{1-\exp(-\mathrm{ad}X)}{\mathrm{ad}X}(\partial_i X) = \partial_i X + \frac{1}{2}[\partial_i X, X] + \frac{1}{6}[[\partial_i X, X], X] + \cdots, \tag{4}$$

one finds

$$I(e^X;A) = \|\,A\,\|^2 - 2i\int_M \mathrm{Tr}(X\partial_i A_i) + \int_M \mathrm{Tr}(X^\dagger FP(A)X) + \frac{i}{3}\int_M \mathrm{Tr}(X[[A_i,X],\partial_i X]) + \cdots. \tag{5}$$

We note that for $g \in G$ a constant group element $I(g;A) = I(1;A)$, thus $F_A(g)$ defined by

$$F_A(g) = I(g;A) \tag{6}$$

is for generic A a Morse function [18] on $\mathcal{G}/G$, where $\mathcal{G}$ is the group of local gauge transformations, i.e. the set of functions $g(x)$ that map M to G. (For Morse theory in relation to supersymmetric quantum mechanics, see [19].) The critical points of the Morse function are precisely the gauge functions g for which $[g]A$ is transverse (i.e. $\partial_i[g]A_i = 0$). The Hessian of the Morse function at the critical point is precisely the Faddeev-Popov operator $FP([g]A)$. The Morse index μ for the critical points, is defined as the number of negative eigenvalues of the Hessian. As F_A depends continuously on A (in the norm implicitly defined in eq. (2)), the alternating sum of the Morse indices over the critical points (which is the Euler characteristic of the manifold on which F_A is defined) is conserved.

This argument needs to be treated with caution, as $\mathcal{G}/G$ is not compact and there will in general be infinitely many critical points. Nevertheless, the property is topological in nature and is invariant under continuous deformations of F_A. If we will be interested in a local neighborhood of a particular critical point, we can if necessary deform F_A such that when changing A, there will only happen something to the critical points in this particular neighborhood. As the type of non of the other critical points will change, the change is finite and the alternating sum over the critical points in the chosen neighborhood will be conserved. However, in general singularities occur at reducible connections (these are connections for which there exists a non-trivial gauge function g, such that $[g]A = A$). Thus $\mathcal{G}$ does not act freely and although dividing out the constant gauge transformations will remove some of the singularities, it can be shown that $\mathcal{G}/G$ does not act freely either. Nevertheless, the presence of the remaining singularities does not affect our arguments. (For example, for SU(2) gauge theory on S^3 it can be shown that up to gauge transformations $A = 0$ is the only reducible connection.)

The Gribov region Ω^o is defined as the set of transverse gauge fields for which the Faddeev-Popov operator is (strictly) positive. Its boundary ($\partial\Omega$, $\Omega \equiv \Omega^o \cup \partial\Omega$) is the Gribov horizon, where the lowest eigenvalue vanishes. It is well known that Ω^o is convex [8,17]. This is essentially because $FP(A)$ is linear in A. Thus if $A_{(1)}$ and $A_{(2)}$ are in Ω^o,

$$FP(sA_{(1)} + (1-s)A_{(2)}) = sFP(A_{(1)}) + (1-s)FP(A_{(2)}) > 0 \tag{7}$$

for all $s \in [0,1]$. Thus, if A is on the horizon, the Morse index $\mu(sA)$ will be zero for $0 \leq s < 1$ and one for $s > 1$, at least in a small neighborhood of $s = 1$. Let us assume that as s approaches one from below, there is a sufficiently small neighborhood of the identity in $\mathcal{G}/G$ such that $F_A(g)$ has no further extrema, except the one at $g = 1$. The only way the alternating sum of Morse indices can be preserved is (when s increases beyond one) if two additional extrema are created, that are both local minima, see fig. 1. Thus for $s > 1$ F_A has extrema at $g_0(s) = 1$, $g_1(s)$ and $g_2(s)$, where $\lim_{s\to 1} g_1(s) = \lim_{s\to 1} g_2(s) = 1$, such that $g_{\{1,2\}}$ are homotopically trivial gauge transformations. Clearly $[g_{\{1,2\}}(s)]sA \in \Omega^o$ and are thus points inside the Gribov horizon (inside and outside is of course well defined since Ω is convex) that are related by a homotopically trivial gauge transformation.

Thus under the assumption that $g = 1$ is the only extremum of F_A in a sufficiently small neighborhood in $\mathcal{G}/G$ there will be a bifurcation at the horizon of solutions to the Coulomb gauge condition into two stable and one unstable solution. The assumption just mentioned, however, is easily seen to imply that the third order term in eq. (5) has to vanish at the horizon for X the zero-mode of the Hessian (apart from a factor $\frac{1}{3}$ this term coincides with the definition of μ in ref. [9]). Generically this will not be the case, in which instance there will be a saddle point that coalesces with the local minimum at the horizon and the situation is as sketched in fig. 2, i.e. the saddle point with Morse index 1 will turn into a local minimum, such that the alternating sum of the Morse indices is again conserved. Note that, for $s > 1$ this local minimum is a homotopically trivial copy of the saddle point sA, such that the region just outside the Gribov horizon has copies just inside the Gribov horizon, which is what was already proved by Gribov [1]. In ref. [8,9] it is, however, proven that if the third order term of eq. (5) at the horizon is non-zero for X the zero-mode of the Hessian, then sA for $0 \leq 1 - s < \epsilon$ (with ϵ sufficiently small) cannot be an absolute minimum of F_{sA} and hence also in this case there will be a gauge copy within the horizon (in general by a large, possibly homotopically non-trivial gauge transformation).

3. The explicit example

As the Morse theory arguments of the previous section are formal and not easily made rigorous, it is useful to consider an explicit example, where all the features of the bifurcation can be checked in detail. We will consider Henyey's example [3] of an axial symmetric abelian gauge field for SU(2) gauge theory on $\mathbf{R}^3$, which obviously satisfies $\partial_i A_i = 0$

$$\vec{A} = a(r,\theta)\hat{\phi}\tau_3 \quad , \quad \hat{\phi} = (-\sin\phi, \cos\phi, 0), \tag{7}$$

where (r, θ, ϕ) are spherical coordinates. The following gauge transformation

$$g = \exp(\alpha(i\cos\phi\tau_1 + \sin\phi\tau_2)) \tag{8}$$

will leave the Coulomb gauge condition invariant if and only if [3]

$$2r^2\sin^2\theta\partial_i^2\alpha + \sin(2\alpha)(2a(r,\theta)r\sin\theta - 1) = 0. \tag{9}$$

Rather then solving this equation for α, Henyey's strategy was to choose α and solve for $a(r,\theta)$. He took

$$\alpha(r,\theta) = b(r) r \sin\theta, \tag{10}$$

which through eq. (9) yields

$$a(r,\theta) = \frac{1}{2r\sin\theta} - \frac{b(r) + r^2\sin^2\theta(\frac{d^2b(r)}{dr^2} + \frac{4}{r}\frac{db(r)}{dr})}{\sin(2rb(r)\sin\theta)}, \tag{11}$$

defining a proper gauge field, provided $b(r)$ satisfies the following three conditions [3]: $i)$ it is regular at the origin with vanishing first derivative $(db(0)/dr = 0)$, $ii)$ $2rb(r) < \pi$ for all r and $iii)$ $r^3 b(r)$ is bounded.

We note, as expected, that g comes always into pairs as a is invariant under a change of sign of b. Thus for this example $g_1 = g_2^{-1} = g$, cmp. fig. 1. To investigate the bifurcation structure of the Gribov copies, we replace b by βb, with β a constant. It satisfies the above mentioned conditions provided $|\beta| \leq 1$. Explicitly we therefore have

$$\begin{aligned}
\vec{A}(\beta) &= a(\beta)\hat{\phi}\tau_3, \\
a(\beta) &= \frac{1}{2r\sin\theta} - \frac{b + r^2\sin^2\theta(\frac{d^2b}{dr^2} + \frac{4}{r}\frac{db}{dr})}{\beta^{-1}\sin(2r\beta b\sin\theta)}, \\
g(\beta) &= \exp(i\beta rb\sin\theta(e^{i\phi}\tau_- + e^{i\phi}\tau_+)).
\end{aligned} \tag{12}$$

As $g(0)$ is the identity, the two gauge copies $[g(\pm\beta)]A(\beta)$ will coalesce at $\beta = 0$, and it should be such that $FP(A(0))$ has a zero-mode. We even know in advance what this zero-mode should be

$$X_o \propto \frac{\partial g}{\partial \beta}(0) = br\sin\theta(e^{-i\phi}\tau_+ + e^{i\phi}\tau_-). \tag{13}$$

This is easily checked explicitly. Introducing the function $f(r)$ by

$$f(r) = -\frac{a(\beta = 0)}{r\sin\theta} = \frac{1}{2b}(\frac{d^2b}{dr^2} + \frac{4}{r}\frac{db}{dr}), \tag{14}$$

one finds

$$FP(A(0))X_o = -(\partial_i^2 X_o - if(r)[\tau_3, \partial_\phi X_o]) = (-\partial_i^2 + 2f(r))X_o = 0.$$

Now we address the issue of the spectrum of $FP(A(\beta))$ for $\beta \neq 0$. We do this in perturbation theory to order β^2, restricting ourselves to the mode that coincides with X_o at

$\beta = 0$. As

$$FP(A(\beta)) = FP(A(0)) - i\beta^2 \frac{\Delta a}{r\sin\theta} \mathrm{ad}\tau_3 \partial_\phi + \cdots, \tag{15}$$

with

$$\Delta a = \frac{1}{2}\partial_\beta^2 a(\beta=0) = -\frac{1}{6}b^2 r\sin\theta(2+4fr^2\sin^2\theta), \tag{16}$$

the relevant eigenvalue is obtained from first order perturbation theory. With the innerproduct $< X|Y >= \int_M \mathrm{Tr}(X^\dagger Y)$ this becomes (using some partial integrations in the last step) to second order in β

$$\begin{aligned}\lambda &= \frac{\beta^2}{< X_o|X_o >} < X_o| - \frac{i\Delta a}{r\sin\theta}\mathrm{ad}\tau_3\partial_\phi|X_o > \\ &= \frac{\beta^2}{3}\frac{\int drd\theta\ r^4b^4\sin^3\theta(2+4f(r)r^2\sin^2\theta)}{\int drd\theta\ r^4b^2\sin^3\theta} \\ &= -\frac{2\beta^2}{5}\frac{\int drr^2(r^2b^4+\{\frac{dr^2b^2}{dr}\}^2)}{\int dr\ r^4b^2}.\end{aligned} \tag{17}$$

We thus confirm that $A(\beta)$ always corresponds to an unstable solution to the gauge condition.

It remains to show that the corresponding eigenvalues for the Gribov copies, $[g(\beta)]A(\beta)$, stay positive. Explicitly we find

$$\begin{aligned}[g(\beta)]\vec{A}(\beta) =&\vec{A}_+(\beta)+\vec{A}_-(\beta), \\ \vec{A}_+(\beta) =&\{a(\beta)+\sin^2(\beta\alpha)(\frac{1}{r\sin\theta}-2a(\beta))\}\hat{\phi}\tau_3, \\ \vec{A}_-(\beta) =&-\beta\{\frac{\sin(2\beta\alpha)}{2\beta\alpha}\vec{D}(A(\beta))(X_o)+(1-\frac{\sin(2\beta\alpha)}{2\beta\alpha})X_o\vec{\partial}\ln(\alpha)\}.\end{aligned} \tag{18}$$

One easily verifies that $\vec{A}_-(\beta)$ does not contribute to second order in β, since

$$[\vec{A}_-(\beta), \vec{\partial}X_o] = 0 \tag{19}$$

Thus to second order in β the eigenvalue λ' of $FP([g(\beta)]A(\beta))$, that reduces to zero at $\beta = 0$, is given by

$$\lambda' = \frac{\beta^2}{< X_o|X_o >} < X_o|\frac{2i\Delta a}{r\sin\theta}\mathrm{ad}\tau_3\partial_\phi|X_o >= -2\lambda, \tag{21}$$

where we used the remarkable coincidence that ($\vec{A}_+ \equiv a_+\hat{\phi}\tau_3$)

$$a_+(\beta) = a(\beta) - 3\frac{\sin^2(\beta\alpha)}{\alpha^2}\Delta a = a(\beta=0) - 2\beta^2\Delta a + \mathcal{O}(\beta^4). \tag{21}$$

We thus confirm the bifurcation picture, but in general $A(\beta=0)$ need not be on the Gribov horizon as $FP(A(\beta=0))$ might have negative eigenvalues. To see this note that

$$FP(A(0)) = -\partial_i^2 + if(r)\mathrm{ad}\tau_3\partial_\phi \tag{22}$$

commutes with $\mathrm{ad}\tau_3$, L_z and L^2. Thus we can decompose the eigenfunctions as

$$X = X_+(r)Y_{\ell m}(\theta,\phi)\tau_+ + X_+^*(r)Y_{\ell m}^*(\theta,\phi)\tau_- + X_3(r)Y_{\ell m}(\theta,\phi)\tau_3. \tag{23}$$

Restricted to X_3 the Hessian is positive definite and only the "charged" sector will be relevant, for which the Hessian reduces to

$$H_{\ell,m} = -\frac{1}{r^2}\frac{d}{dr}r^2\frac{d}{dr} + (\frac{\ell(\ell+1)}{r^2} - 2mf(r)). \tag{24}$$

For given m and ℓ this is a one dimensional potential problem. It has a zero eigenvalue for $\ell = -m = 1$, with $X_+ = b$. Thus, if b has nodes there are lower (and hence negative) eigenvalues.

As an example without a node, we take [3]

$$b(r) = K(r^2 + r_o^2)^{-3/2} \quad , \quad f(r) = -\frac{15r_o^2}{(r_o^2 + r^2)^2}. \tag{25}$$

The Hamiltonian of eq. (24) can only have negative eigenvalues if its potential can become negative. This only leaves $(\ell, m) = (2,-2)$ and $(1,-1)$. As b has no nodes $H_{1,-1} \geq 0$, whereas $H_{2,-2} - H_{1,-1} \geq 0$ proves that also $H_{2,-2} \geq 0$. In conclusion, eq. (25) provides an example for which $A(\beta=0)$ is at the Gribov horizon, which concludes this section.

4. The fundamental modular domain

Let us consider the set

$$\Lambda = \{A | F_A(g) \geq F_A(1),\ \forall g \in \mathcal{G}\}. \tag{26}$$

Clearly one has Λ a subset of Ω. Furthermore, it was proven by Semenov-Tyan-Shanskii and Franke [8] and Dell'Antonio and Zwanziger [10] that Λ covers **all** gauge orbits. That is, given a connection A (with finite norm $||A||$), there exists a $g \in \mathcal{G}^c$ for which $F_A(g)$ is at its absolute minimum. The space $\mathcal{G}^c$ is the completion of $\mathcal{G}$ with respect to the norm

$\|\delta g\|_1^2 = \|\delta g\|^2 + \|d\delta g\|^2$, in which δg is viewed as a complex $N \times N$ matrix (in ref. [8] one works on $\mathbf{R}^3$ and a slightly different norm is used, so as to eliminate the constant gauge transformations). Let $\Lambda^o \subset \Lambda$ be the set where the minimum is unique, i.e. if $A \in \Lambda^o$ then $F_A(g) > F_A(1)$ for **all** non-constant g. Both Λ^o and Λ are convex [8,9], for let $A_{(1)}$, $A_{(2)} \in \Lambda^o(\Lambda)$ then

$$\begin{aligned} &\|[g](sA_{(1)} + (1-s)A_{(2)})\|^2 - \|sA_{(1)} + (1-s)A_{(2)}\|^2 \\ &\quad = s(\|[g]A_{(1)}\|^2 - \|A_{(1)}\|^2) + (1-s)(\|[g]A_{(2)}\|^2 - \|A_{(2)}\|^2) \end{aligned} \tag{27}$$

shows that $sA_{(1)} + (1-s)A_{(2)} \in \Lambda^o(\Lambda)$ for $s \in [0,1]$. Also clearly $A = \Theta = 0 \in \Lambda^o$. In [9] it was proved that the boundary of Λ^o ($\partial\Lambda^o$) is contained in Λ, and that Λ is closed (apply the "First step" lemma [9] to Λ^o and Λ), which is basically a continuity argument. Alternatively, take $A \in \Lambda - \Lambda^o$, and use eq. (27) for $A_{(1)} = 0$ and $A_{(2)} = A$, which implies that $sA \in \Lambda^o$ for all s, $0 \leq s < 1$ and hence $A \in \partial\Lambda^o$. Thus the boundary still exists of transverse gauge fields A such that $F_A(g)$ reaches its absolute minimum at $g = 1$, but this minimum need not be unique or might be on the Gribov horizon.

To analyse these two options, take $A \in \Omega - \Lambda$, then the ray sA will cross the boundary of Λ where necessarily an absolute minimum turns into a relative minimum. At the boundary $F_A(g)$ therefore has degenerate absolute minima, that are related by in general large gauge transformations. We will call these points on $\partial\Lambda$ **regular boundary points**. The remaining points of the boundary will necessarily be on the Gribov horizon and are called **singular boundary points**. That the set of regular boundary points is non-empty is easily established by taking the example of $A = [g]0$ for g homotopically non-trivial, such that A is on the Gribov horizon. The path sA will pass $\partial\Lambda$ at (necessarily) a regular point (as the absolute minimum of F_A does not occur at constant gauge functions, but at g^{-1}, where it vanishes). The gauge transformation that relates the two copies at the boundary of Λ is essentially g^{-1} and is thus homotopically non-trivial (as mentioned before, in the torus example the fundamental modular domain restricted to the abelian constant modes is given by $|C_k| \leq \pi$, modulo the action of the Weyl group $\vec{C} \to -\vec{C}$ and all boundary points are easily seen to be regular [5,6]). Strictly speaking therefore, also Λ is **not** a fundamental modular domain. Yet it will be, once we have appropriately identified the boundary points. It is these boundary identifications that will give the fundamental modular domain the topology of the full configuration space which is the basis of Singer's [20] argument why for gauge theories on a compact three or four dimensional manifold M, there necessarily have to be gauge copies.

Observe that homotopical non-trivial gauge transformations are in one to one correspondence with non-contractable loops in configuration space, which give rise to conserved quantum numbers. The quantum numbers are like the Bloch momenta in a periodic potential and have to be representations of the homotopy group of gauge transformations. On the fundamental modular domain the non-contractable loops arise through identifications of boundary points (as is quite explicit for the torus in the zero-momentum sector). Although slightly more hidden, the fundamental modular domain will therefore contain all the information relevant for the topological quantum numbers (i.e. it doesn't have to be "put in by hand"). Sufficient accurate knowledge of the boundary identifications will allow, however, for an efficient and natural projection on the various superselection sectors (i.e. by choosing the appropriate "Bloch wavefunctionals"). All these features were at the heart of the finite volume analysis on the torus [5] and we see that they can in principle naturally be extended to the full theory, thereby including the desired θ dependence. In ref. [6] we proposed formulating the Hamiltonian theory on coordinate patches with homotopically non-trivial gauge transformations as transition functions. We can shrink these patches almost to Λ (and their associated gauge copies, with the homotopically non-trivial gauge transformations that relate the inequivalent classical vacua). If there would be no singular boundary points we would avoid any points on the Gribov horizon, thereby defining open sets that do cover the whole configuration space. These two formulations are therefore equivalent.

However, it is essential to note that the topology of the configuration space is **not** described entirely by non-contractable loops. One also needs to consider non-contractable spheres of any dimension. It is the non-contractable spheres that in general will be responsible for singular boundary points in the fundamental modular domain. As the interior is convex, non-contractable d-spheres can only arise if the boundary contains a (d-1)-sphere on which all points are identified. These correspond to gauge orbits for which F_A is degenerate along this (d-1)-sphere embedded in $\mathcal{G}$. Thus, these A necessarily coincide with the horizon. One can introduce a regular coordinate patches in the neighborhood of these singular points to eliminate the singularities, as observed by Singer and Nahm [20,21]. In this case, though, transition functions can not be described in terms of gauge transformations [6]. For example, when $G = SU(2)$ these non-contractable spheres do actually occur [20]. Thus it seems that we cannot avoid singular boundary points (as was suggested by fig. 1 of ref. [11]). We do not claim that our arguments present a proof for the existence of singular boundary points, as we implicitly assumed that the topology of configuration space is unaffected by the completion with respect to the norm $\|A\|$. These issues can be quite intricate, as for example the winding number of a C^1 gauge transformation, when defined through

$$\nu(g) = \frac{1}{24\pi^2} \int_M \mathrm{Tr}((g^{-1}dg)^3), \tag{28}$$

is not continuous in the norm $\|\delta g\|_1$. However, continuity is sufficient to define homotopy types. Relevant for this issues is the Sobolev embedding theorem $W_p^m \subset C^k$ for $k < m - n/p$ (see for example [22]), where n is the dimension of the manifold, W_p^m is the Sobolev space of functions for which the first m distributional derivatives are in L^p and C^k is the set of k times continuously differentiable functions. In the one dimensional case, functions in W_p^m for $p > 1, \quad m \geq 1$ are guaranteed to be continuous (as can be easily deduced explicitly by using Hölder's inequality). In the present case $n = 3$, $m = 1$ and $p = 2$, which unfortunately does **not** imply continuity. Indeed, for example, one can construct a series of maps $g_n : S^3 \cong SU(2) \to SU(2)$ that for all n have winding number zero but converges in the norm $\|\delta g\|_1$ to the identity map with winding number one. Yet, one must remember that the gauge fields in the fundamental modular domain satisfy the Coulomb condition (in the weak sense) from which one might deduce stronger smoothness properties. Thus, it is possible that using more sophisticated results from functional analysis will allow one to make stronger claims than we are willing to commit ourselves to here. One should address these issues primarily in the light of the physically more relevant dynamical questions and in that context we certainly intend to come back to this in the future.

In conclusion, it might be that the "hole" in configuration space, due to a non-contractable loop or sphere, is of zero size in the norm $\|A\|$. We consider this unlikely, but cannot exclude it. Actually, it might be the mechanism through which lattice gauge theories in the continuum limit reproduce the various topological sectors associated with the winding numbers. By this we mean that by demanding the fields to be "smooth", we will "cut" the necessary holes in configuration space. This requires an appropriate understanding of what it means to take the continuum limit, for which the "dislocations" [23] play an important role. Concerning gauge fixing in lattice gauge theory we only wish to remark that one can similarly impose a Coulomb (or Landau) type gauge through an action principle [24]. Our statement that Λ requires identifications at the boundary is equally valid in this case. Also one has to realize that, although there are no homotopically non-trivial gauge transformations associated with the winding number, twisted gauge transformations [4] are still well defined and cannot be deformed to the identity.

5. Discussion

We have reconsidered the proposed [8,10,11,12] fundamental modular domain Λ consisting of the absolute minima of the norm $||\,[g]A\,||$ on a gauge orbit and shown it is a fundamental modular domain provided **necessary** gauge identifications at the boundary are taken into account. As these identifications determine uniquely the topology of the configuration space, it can be argued, due to the presence of non-contractable spheres in the configuration space (this is true both for three and four dimensional compact manifolds M over which the gauge theory is defined) that in general there will be so-called singular boundary points, on which the Faddeev-Popov determinant will vanish. One can still attempt to formulate the standard Hamiltonian [25] on this fundamental modular domain. Usually one rescales the wavefunctional with $\sqrt{\det'(FP(A))}$ which will be strictly positive everywhere, expect at a subset of the boundary to the fundamental modular domain of codimension 1, where its vanishing is associated to a coordinate singularity due to a non-contractable sphere in configuration space. As topological quantum numbers are only associated to non-contractable loops, it might be that it is sufficient to simply demand the (rescaled) wavefunctional to vanish at the singular boundary points. This requires further study as subtle effects can complicate the issue, especially in the presence of fermions. We only need to remind the reader of the global SU(2) anomaly [26], which on a three dimensional manifold M will be associated with a two-dimensional non-contractable sphere in configuration space. In this context we can recommend ref. [27] for a clear description of the issue of anomalies in the Hamiltonian formulation.

The issue of gauge copies has played an important role in the analysis of the spectrum of the low energy states for SU(2) gauge theory on the torus in a finite volume. As was mentioned before, in the sector of the abelian constant modes $A_k = \frac{C_k}{2L}\tau_3$, which forms the "vacuum or toron valley" along which the classical energy vanishes, Λ is described by $|C_k| \leq \pi$. The gauge transformation that maps $C_k = -\pi$ to $C_k = \pi$ is given by $g_{(k)} = \exp(-\pi i\tau_3 x_k/L)$, which due to its anti-periodicity is homotopically non-trivial and provides the required boundary identifications. In this subsector all boundary points are easily seen to be regular. The "Bloch momenta" label 't Hooft's electric flux quantum numbers [4] $\Psi(C_k = -\pi) = \exp(\pi i e_k)\Psi(C_k = \pi)$. Note that the phasefactor is not arbitrary, but ± 1. This is because $g^2_{(k)}$ is homotopically trivial. It thus looks like as if we have to put the topological structure in by hand after all, however, one should realize that considering a slice of Λ will obscure some of the topological features. A loop that winds around the slice twice is contractable in Λ as soon as it is allowed to leave the slice. Indeed including the lowest modes transverse to this slice will make the Z_2 nature of the relevant homotopy group evident [5,6]. It shows, that not dynamically motivated truncations, can obscure things. For example, a recent reduction to spherically symmetric gauge fields [28] is only of limited value if the non spherical fluctuations (which can in principle lead to Gribov horizons) are not taken into account.

In weak coupling Lüscher [29] showed unambiguously that the wavefunctionals are localized around $A = 0$, that they are normalizable and that the spectrum is discrete. In this limit the spectrum is insensitive to the boundary identifications (giving rise to a degeneracy in the topological quantum numbers). At stronger coupling the wavefunctional spreads out over the vacuum valley and the boundary conditions drastically change the spectrum [5]. In supersymmetric gauge theory the situation is even more dramatic. In the bosonic case, what localizes the wavefunctional in weak coupling is an induced potential barrier due to the zero-point fluctuations of the modes transverse to the vacuum valley. Due to the supersymmetry this induced barrier is expected to be canceled exactly by the fermionic contribution and the wavefunctional is expected to spread out over the whole vacuum valley. The problem is, however, that transverse fluctuations become singular near $A = 0$, preventing a reduction to the vacuum valley. As in the bosonic sector one can hope that a truncation to the zero-momentum sector will be possible, but as the wavefunctional is expected to spread out over the vacuum valley and as the gauge copies that thus arise will bring one outside of the zero-momentum sector, this would not be a consistent truncation either. Indeed, it was rigorously proven in ref. [13] that the spectrum of the zero-momentum Yang-Mills Hamiltonian is continuous down to zero-energy. Apart from the fact that such a continuous spectrum would make the Witten index ill-defined, it is not compatible with the fact that the theory was originally defined in a finite volume. We consider this as a strong indication for the spreading of the wavefunctional beyond the Gribov copies. The appropriate identifications due to these copies should lead to the desired discrete spectrum. In the bosonic sector there is a **dynamical** reduction to a finite number of degrees of freedom [5,6]. But in the fermionic sector it requires one to construct vacuum Dirac bundles that incorporate the identifications at the boundary of the fundamental modular domain. Our problem is, that this does not seem to allow for a dynamical reduction to a finite number of degrees of freedom. As the results of ref. [14] rely on the truncation to the zero-momentum sector, without addressing the Gribov copy problem, we do not understand how the results in that paper can solve the problem of constructing the zero-energy states (unfortunately the construction of the wavefunction in ref. [14] is rather implicit and incomplete, which makes it hard to pin down exactly what might make it unsuitable as a zero-energy groundstate wavefunction). Thus it will remain an interesting and unfortunately open problem, whose solution will shed light on the discrepancy between the naive Witten index calculation [30] on the torus for O(N) ($N > 6$) (giving an index equal to the rank of O(N) plus one) and the value deduced form the gluino condensate calculations in an infinite volume [31] (yielding the value $N-2$). When it persists, this would imply that the Witten index in this case will have discontinuities, which will have interesting consequences for the non-perturbative vacuum in these theories.

Acknowledgements

I thank Daniel Zwanziger for sending me ref. [10] and Maarten Golterman, Hari-Dass, Andreas Kronfeld, Morton Laursen, Peter Lepage, Jeff Mandula, Paul Mackenzie, Michael Ogilvie, Stepen Sharpe, Jan Smit, Arjan van der Sijs and Jac Verbaarschot for discussions on Gribov copies and horizons at various occasions. I also thank Bernard de Wit, Hiroshi Itoyama, Bob Razzaghe-Ashrafi, Arkadi Vainshtein and in particular M. Shifman for discussions on the Witten index calculation on the torus. I am grateful to Karl Isler for his insights about the Dirac vacuum bundles.

Many useful comments on the manuscript and during seminars have helped me to clarify a number of issues, for which I thank Sidney Coleman, Jan de Boer, Philippe de Forcrand, Jim Hetrick, Key-Fei Liu, Herman Verlinde and Daniel Zwanziger. I am grateful to the Cern Theory Division for their hospitality during my visit in July. This research has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

References

[1] V. Gribov, Nucl. Phys. B139(1978)1

[2] R. Jackiw, I. Muzinich and C. Rebbi, Phys. Rev. D17(1978)1576

[3] F.S. Henyey, Phys. Rev. D20(1979)1460

[4] G. 't Hooft, Nucl. Phys. B153(1979)141

[5] J. Koller and P. van Baal, Nucl. Phys. B302(1988)1

[6] P. van Baal, in: Probabilistic Methods in Quantum Field Theory and Quantum Gravity, ed. P.H. Damgaard et al, (Plenum Press, New York, 1990) p.131; P. van Baal, in: Frontiers in Non-perturbative Field Theory, ed. Z. Horvath et al, (World Scientific, Singapore, 1989) p.204

[7] Ph. de Forcrand et al, Nucl. Phys. B[Proc.Suppl]20(1991)194

[8] M.A. Semenov-Tyan-Shanskii and V.A. Franke, Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V.A. Steklov AN SSSR, 120(1982)159. Translation: (Plenum Press, New York, 1986) p.1999

[9] G. Dell'Antonio and D. Zwanziger, in: Probabilistic Methods in Quantum Field Theory and Quantum Gravity, ed. P.H. Damgaard et al, (Plenum Press, New York, 1990) p.107

[10] G. Dell'Antonio and D. Zwanziger, Comm. Math. Phys. 138(1991)291

[11] D. Zwanziger, Nucl. Phys. B345(1990)461

[12] C. Parinello and G. Jona-Lasinio, Phys. Lett. B251(1990)175

[13] B. de Wit, M. Lüscher and H. Nicolai, Nucl. Phys. B320(1989)135

[14] H. Itoyama and B. Razzaghe-Ashrafi, Nucl. Phys. B354(1991)85

[15] G. 't Hooft, Phys. Rev. Lett. 37(1976)8; Phys. Rev. D14(1976)3432

[16] K. Wilson, in: Recent developments in Gauge Theories, ed. G. 't Hooft, et al (Plenum

Press, New York, 1980) p. 363; G. 't Hooft, Nucl. Phys. B190[FS3](1981)455

[17] D. Zwanziger, Phys. Lett. 114B(1982)337; Nucl. Phys. B209(1982)336

[18] J. Milnor, Morse Theory, Ann. of Math. Studies 51, Princeton U.P., 1973

[19] E. Witten, J. Diff. Geom. 17(1982)661; P. van Baal, Acta Phys. Pol. B21(1990)73

[20] I. Singer, Comm. Math. Phys. 60(1978)7

[21] W. Nahm, in: IV Warsaw Symp. on Elem. Part. Phys., ed. Z. Ajduk, p.275 (Warsaw, 1981)

[22] Y. Choquet-Bruhat et al, Analysis, Manifolds and Physics (North-Holland, Amsterdam, 1977) p.412

[23] D.J.R. Pugh and M. Teper, Phys. Lett. 224B(1989)159

[24] J. Mandula and M. Ogilvie, Phys. Lett. B185(1987)127

[25] N.M. Christ and T.D. Lee, Phys. Rev. D22(1980)939

[26] E. Witten, Phys. Lett. 117B(1982)324

[27] P. Nelson and L. Alvarez-Gaumé, Comm. Math. Phys. 99(1985)103

[28] D. Schütte, The problem of gauge fixing for spherically symmetric Yang-Mills fields, Bonn preprint, April 1991

[29] M. Lüscher, Nucl. Phys. B219(1983)233 ; M. Lüscher and G. Münster, Nucl. Phys. B232(1984)445

[30] E. Witten, Nucl. Phys. B202(1982)253

[31] M. Shifman and A. Vainshtein, Nucl. Phys. B296(1988)455

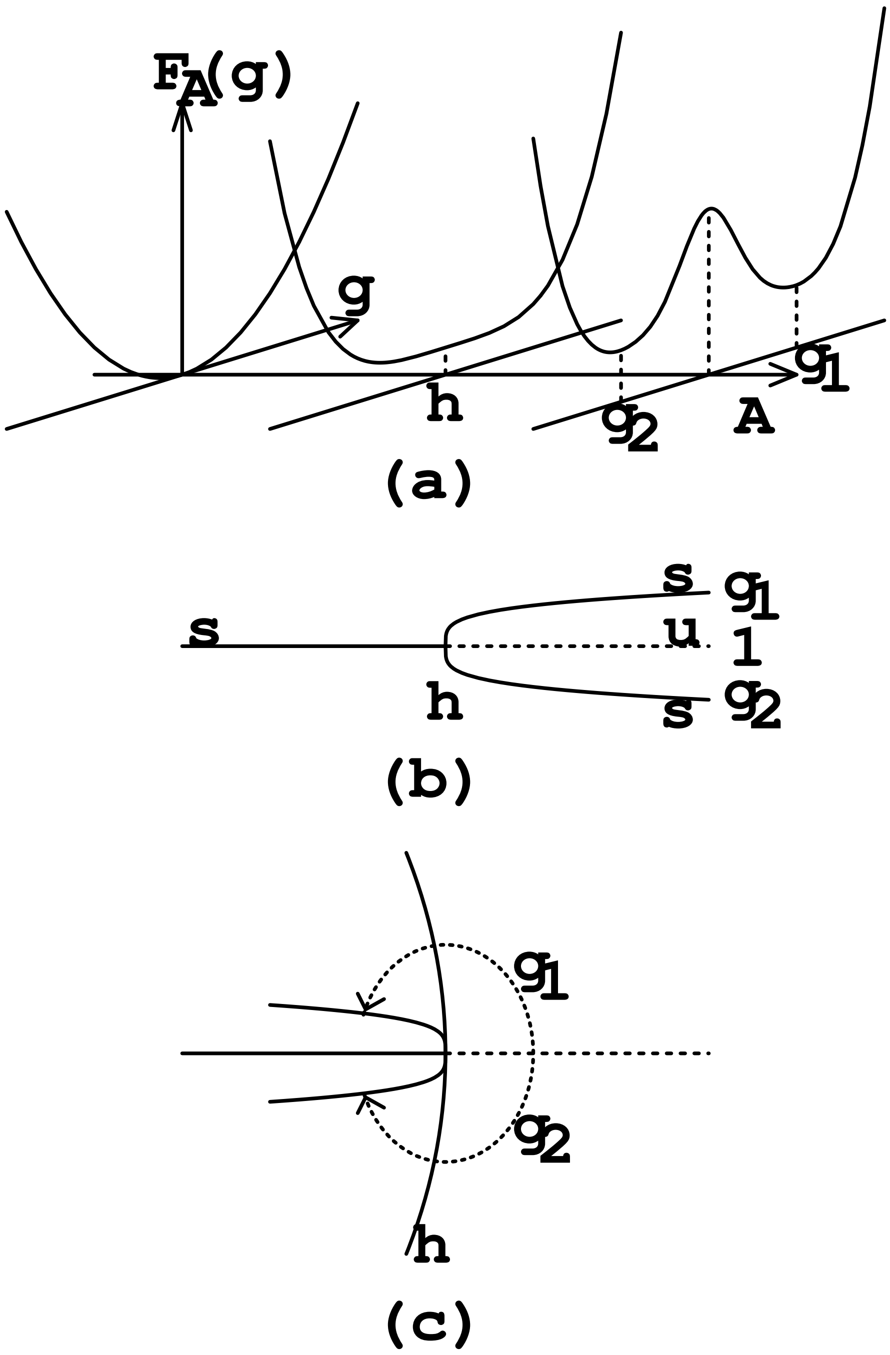

Figure 1: Here we sketch the bifurcation at the horizon by three different representations. In a. we show what happens to $F_A(g)$ when we pass the horizon (h), b. is the well-known bifurcation picture, where s stands for stable (a local minimum) and u stands for unstable (a saddle point). Finally, in c. we sketch the situation regarding the Gribov region.

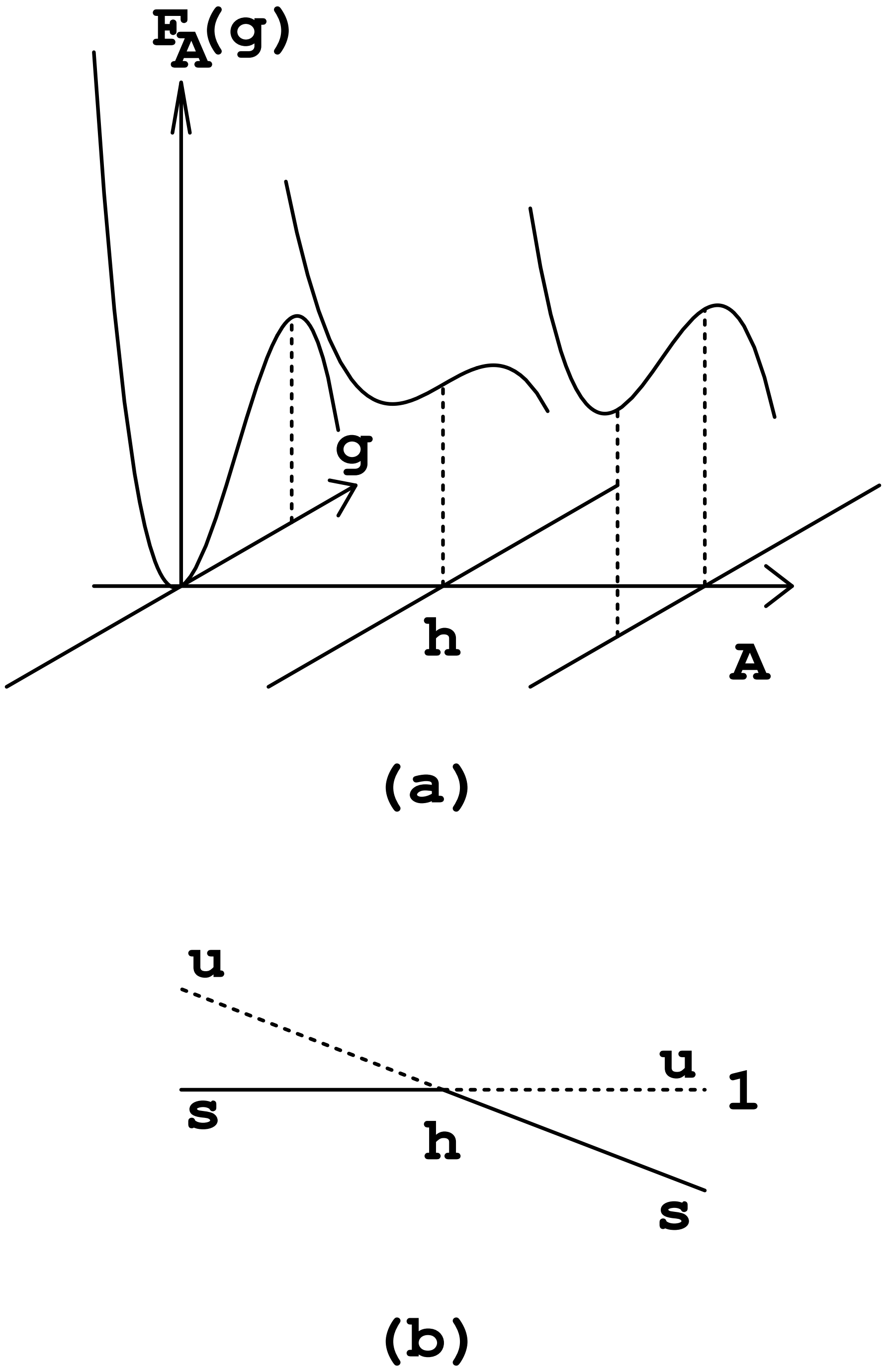

Figure 2: Here we sketch the generic situation at the horizon (h), when a (local) minimum coalesces with a saddle point.