Albanian Journal of Mathematics

Për një Shqipëri të shkencës dhe kulturës.

Volume 7

www.albanian-j-math.com

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 1, Pages 3–23
ISSN: 1930-1235; (**2013**)

MORSE THEORY WITH LOW DIFFERENTIABILITY AND DEGENERATE CRITICAL POINTS FOR FUNCTIONAL ENERGY OF A FINSLER METRIC

Fausto Marçal de Souza
*Instituto de Matemática e Estatística,
Universidade Federal de Goiás,
CP 131, CEP 74001–970, Goiânia, GO,
Brazil.*
E-mail: fausto.turbo@brturbo.com.br

Abstract. The aim of this paper is to extend the Morse theory of $(\Lambda M, E)$ with low differentiability and degenerate critical points, where ΛM is the space of H^1-closed curves on an n-dimensional compact manifold M endowed with a Finsler metric $F : TM \to R$ and $E : \Lambda M \to R$ is the associated energy integral, or simply the energy.

1. Introduction

In the Morse theory with low differentiability and degenerate critical points, on Hilbert manifolds, the closed geodesics problem for Finsler metric can be developed as in the Riemannian case; see [1], [7]. In this theory a closed geodesic is a distinguished closed curve in the Hilbert manifold of H^1-closed curves and being a critical point of the functional integral energy of Finler metric F. The other aspect is to consider a closed geodesic (or more exactly the tangent vector field along a closed geodesic) as a periodic orbit in the geodesic flow on the cotangent bundle T^*M of a Finsler manifold $(M\,,\,F)$. In this case the geodesic flow is a special case of a Hamiltonian flow (see, [8], [23], [24]). One of the differences between the Morse theory for the Finsler and Riemannian cases is that the Finsler energy is not C^2. In fact it is twice differentiable at the critical points, and with strongly differentiable derivative on these points, but not, in general, outside the regular

2000 *Mathematics Subject Classification.* 58E05; 58E10.
Key words and phrases. strong differentiability; critical point; critical submanifold; critical group; H^1-closed curves; Finsler metric; closed geodesic.

closed curves (see, Theorem 8.1 of this paper). This is a peculiarity of the Finsler metrics since the energy E is C^2 if only if F is the norm of a Riemannian metric. Therefore, in order to have a Morse theory for the Finsler case we need a Morse lemma for functions with conditions of low differentiability that, although stronger than those in [10], are verified by the Finsler energy. This Morse lemma was done in the papers (see [1], [7]) for the case of isolated critical points of functions with low differentiability and in a very general context that beyond allowing the adaptation of Gromoll–Meyer arguments to our case, has probable utility in the treatment of more general variational problems. A very interesting and important question is to extend the Gromoll–Meyer theorem for Finsler metrics. In fact this is the case, and the first demonstration is due to Matthias; see [13]. The Matthias demonstration used an approach of finite dimension of the Morse theory for closed geodesics of the manifold of the H^1- closed curves, theory inspired by the treatment of Milnor (see [17]) of the problem of geodesics connecting two points. This type of argument, elegant and extremely efficient in the cases above, uses in an essential way the geometry of the metric and does not seem to be appropriate to the study of other more general 1-dimensional variational problems.

The purpose of this paper is to extend some results of the theory of critical points that allow to use of the original arguments of Gromoll–Meyer to prove the existence of infinitely many closed geodesics (non trivial and geometrically distinct) on a compact simply connected differentiable Finsler manifold whose cohomology ring is not generated by only one element (i.e., the cohomology ring is not isomorphic to the one of a compact symmetric space of rank one: a sphere or a complex projective space, quaternionic or of Cayley), provided that the sequence of the rational Betti numbers of the space ΛM of parametrized H^1-closed curves is unbounded. The Morse lemma that we proved in the papers [1] and [7] for degenerate critical points with low differentiability, is for a C^1- function defined on an open neighborhood of 0 in a Hilbert space H, where 0 is the only critical point of f, f twice differentiable at 0 with derivative f' strongly differentiable at the origin. This result generalizes the Morse lemma for a nondegenerate critical point of C^2 function that is due to Cambini (see [2]) and also the Morse lemma for degenerate case for a C^2 function that was extended by Mawhin and Willem; see [15], [16].

We extend to compact connected critical submanifold of a Hilbert manifold the notion of critical point and we compute the critical groups at an isolated critical submanifold of the Finsler energy E. In the paper [9], F. Mercuri describes the Liusternik–Schnirelman theory for the energy E of a Finsler metric F on a smooth manifold M, and the following properties are proved:

a. The energy E is C^1 and the differential is locally Lipschitzian.
b. If M is compact, E satisfy the condition (C) of Palais–Smale.

The our main purpose in this paper is to prove the following:

(1) Under the hypothesis of Proposition 3.2 (A Morse lemma for degenerate critical points with low differentiability), we prove that the function $\hat{f}$ associated to the degenerate part of the function f, is of class C^1, with derivative $(\hat{f})'$ strongly differentiable at $y = 0$ with $(\hat{f})''(0) = 0$, and also we compute the local critical groups.
(2) The derivative of the Finsler energy $E : \Lambda M \to \mathbb{R}$ is strongly differentiable on the regular curves and particularly on the closed geodesics, and therefore

E satisfies the hypothesis of the our Morse lemma for isolated critical points, possibly degenerate (Proposition 3.2, page 7).

(3) To generalize the Gromoll–Meyer theorem for Riemannian manifolds to Finsler manifolds.

A more direct approach using the Morse theory on Hilbert manifold now is possible since the energy functional E in Finsler case satisfies the hypothesis of our Morse lemma for degenerate critical point with low differentiability. Moreover, the index formula for Finsler manifolds works the same in the Riemannian case. The index form of the energy functional E in the Finsler case can be found already explicitly in standard text books and papers; see [7], [12], [25]. This index theory for closed geodesics on Finsler manifolds with our previous results permit the conclusion of the existence of infinitely many closed geodesics on a compact Finsler manifold, whose cohomology is not isomorphic to that of a compact symmetric space of rank one.

2. Preliminaries

Let M a Hilbert manifold and $f \in C^{2-}(M,\mathbb{R})$, i. e. f is C^1-function with differential df locally Lipschitzian. We assume that the manifold M is regular, i. e. every neighborhood of a point contains a closed neighborhood. Let u be an isolated critical point of f. The critical groups (over a field F) are defined by

$$C_n(f,u) = H_n(\{f \le c\} \cap U\,,\, \{f \le c\} \cap U - \{u\}), \quad n = 0,1,2,\ldots,$$

where $c = f(u)$,

$$\{f \le c\} = \{v \in M \,:\, f(v) \le c\},$$

U is a closed neighborhood of u, and $H_n(A\,,\,B)$ is the n-th singular homology group of the pair $(A\,,\,B)$ over a field F. By excision, the critical groups are independent of U. We recall that f satisfies the condition (C) of Palais–Smale over a closed subset S of M if, for any sequence z_n on S such that $f(z_n)$ is bounded and $|f'(z_n)|$ tends to zero, then (z_n) has a convergent subsequence and any limit point is a critical point of f. It is easy to see that a C^1 function f, twice differentiable at critical points that we suppose isolated, and admitting a nondegenerate critical point, always satisfies the condition (C) in a closed neighborhood of that point. A piecewise C^1 path from u_1 to u_2 is a piecewice C^1 mapping

$$\sigma \,:\, [a\,,\,b] \to M$$

that $\sigma(a) = u_1$ and $\sigma(b) = u_2$. We define the length of σ by

$$L(\sigma) = \int_a^b |\sigma'(t)|\, dt,$$

and the geodesic distance d on M by the

$$d(u_1,u_2) = inf\{L(\sigma);\ \sigma : [a,b] \to M,\ \sigma(a) = u_1,\ \sigma(b) = u_2,\ \sigma \text{ a piecewise } C^1\}.$$

A subset of M is complete if it is complete for geodesic distance. The gradient of f is the vector field $\triangledown f$ defined on M by

$$df(u)v = \langle \triangledown f(u)\,,\, v\rangle\ \ \forall v \in T_uM.$$

Proposition 2.1. *Let $f \in C^{2-}(M,\mathbb{R})$ and $\sigma(t) = \sigma(t,u)$, $\alpha(u) \le t \le \beta(u)$, the unique maximal solution of the equation $\dot\sigma(t) = -\nabla f(\sigma(t))$, $\sigma(0) = u$. Then the following conditions are true:*

(1) *Either* $f(\sigma(t)) = f(u)$ *or* $f \circ \sigma$ *is non-increasing for all* $t \geq 0$. *Moreover, for* $0 \leq s \leq t < \beta(u)$, *we have*

$$d(\sigma(t), \sigma(s)) \leq L(\sigma) \leq (t-s)^{1/2}(f(\sigma(s)) - f(\sigma(t)))^{1/2}.$$

(2) *If* $\beta(u)$ *is finite and the set* $\{\sigma(t) : 0 \leq t < \beta(u)\}$ *is contained in a complete subset of* M, *then* $f(\sigma(t)) \to -\infty$ *when* $t \to \beta(u)$.

(3) *If the set* $\{\sigma(t) : 0 \leq t < \beta(u)\}$ *is contained in a complete subset of* M *with* $f(\sigma(t)) > a$, *then* $\beta(u) = +\infty$.

3. Strongly Differentiable Functions

Recall that a function between two Banach spaces, $f : E \to F$, is said to be strongly differentiable at a when there is a linear map $T : E \to F$, such that for $x, y \in E$:

$$f(x) - f(y) = T(x-y) + R(x,y),$$

where $\lim\limits_{x,\, y \to a} \frac{R(x,y)}{\|x-y\|} = 0$.

Taking $y = a$, we see that a strongly differentiable function f at a is differentiable at a, with $T = f'(a)$. Another way to be equivalent to define a strongly differentiable function is as follows.

The function $f : E \to F$, is said to be *strongly differentiable* at a, if f is differentiable at a and $\lim\limits_{x,\, y \to a} \frac{r(x)-r(y)}{\|x-y\|} = 0$, where $r(y)$ is the remainder of the Taylor's formula for f around a.

In other words, f is strongly differentiable at a, if and only if, it is differentiable at a and, given $\varepsilon > 0$, there is a neighborhood $V$of a where $r(x)$ is $\varepsilon-Lipschitzian$, and therefore f satisfies the condition of Lipschitz, with constant $\|f'(a)\| + \varepsilon$.

It is clear also that if f is differentiable in a neighborhood of a and its differential f' is continuous at a, then f is strongly differentiable at a. Moreover, if f is continuous in a neighborhood of a and strongly differentiable at a, with invertible differential, then f is invertible around a. The proof is the same as the classical inverse function theorem; see [6, chapter 5].

Let $f : U \subset H \to \mathbb{R}$ be a C^1 function defined on a open set of a Hilbert space H and $0 \in U$. Suppose that f is twice differentiable at 0, having 0 as critical point. Let N be the kernel of the symmetric operator $A : H \to H$ given by

$$\langle Av, u \rangle = \frac{1}{2} d^2 f(0)(v,u).$$

If the image $Im(A)$ is closed, since A is symmetric, $N^{\perp} = Im(A)$ and H is decomposable in $H = N^{\perp} \oplus N$. Thus we can look at $z \in H$ as $x + y \in N^{\perp} \oplus N$. Also the corresponding version of the implicit function theorem gives the following proposition.

Proposition 3.1. *Using the conditions and notation above, let* 0 *be a critical point of* f *and suppose that* f' *is strongly differentiable at the origin. Then there is a continuous function*

$$g : U \subset N \to N^{\perp}$$

on an open set U *containing* 0 *such that* $f_x(g(y), y) \equiv 0$ *and* $g(0) = 0$, *where* f_x *denote the partial derivative with respect to* x. *Moreover,* g *is strongly differentiable at the origin and* $\mathrm{d}g(0) = 0$.

In the paper [1] we proved a degenerate-critical point version of the Morse lemma as in [4] with conditions of low differentiability that, although stronger than those in [9], are verified by the Finsler energy $E : \Lambda M \to \mathbb{R}$. These results are contained as part of the author PhD disertation; see [7].

Proposition 3.2 (A Morse Lemma for degenerate critical points with low differentiability)**.** *If f' is strongly differentiable at the origin, then there is a neighborhood V of 0 in H and a homeomorphism $\psi : V \to \psi(V) \subset H$, $\psi(0) = 0$ such that*

$$f(\psi(x, y)) \;=\; \frac{1}{2}\langle Ax, x\rangle \;+\; f(g(y), y),$$

where g is a function $g : V \bigcap N \to N^{\perp}$ strongly differentiable at $0 \in N$ with $g(0) = 0$, and $\mathrm{d}g(0) = 0$, the function ψ is differentiable at 0, with $\psi(0) = 0$ and $d\psi(0) = I$; see [7] and [1].

The next proposition that we will prove is very important for the computation of the local critical groups at a isolated critical point, that is reduced to a finite dimension problem.

Proposition 3.3. *Let $f : U \subset H \to R$ be a C^1 function defined on a open set and let $0 \in U$ be only critical point of f. Suppose that f is twice differentiable at 0 and that f' is strongly differentiable at the origin. Let*

$$z = x + y \in H = N^{\perp} \oplus N$$

and f_x, f_y be the partial derivatives with respect to x and y, and

$$g : B_r(0) \subset N \to N^{\perp}$$

be a unique continuous map defined on an open ball $B_r(0) = \{y \in N : |y| < r\}$, such that

$$f_x(g(y), y) \equiv 0,$$

g is strongly differentiable at the origin with $g(0) = 0$, $dg(0) = 0$ and g is Lipschitzian on $B_r(0)$. Then, the following hold

(1) *The function $\hat{f} : B_r(0) \subset N \to R, \hat{f}(y) = f(g(y), y)$ is of class C^1 with $(\hat{f})'(y) = f_y(g(y), y)$ where $0 \in B_r(0)$ is only critical point of $\hat{f}$ and $(\hat{f})'$ is Lipschitzian on $B_r(0)$.*
(2) *If f satisfies the condition (C) of Palais–Smale, then $\hat{f}$ also satisfies this condition.*
(3) *The function $\hat{f} : B_r(0) \subset N \to R,\ \hat{f}(y) = f(g(y), y)$ is twice differentiable at the origin with $(\hat{f})''(0) = 0$, and the derivative $(\hat{f})'$ is strongly differentiable at $y = 0$.*

Proof. We can choose the open set $U \subset H$ a convex open set

$$U = B_\delta(0) \oplus B_r(0),$$

with g Lipschitzian on $B_r(0)$, f_x *and* f_y Lipschitzian on $B_\delta(0) \oplus B_r(0)$. Let $M_1 > 0$, $M_2 > 0$, $M_3 > 0$ be the constants of Lipschitz for $g : B_r(0) \to B_\delta(0)$, $f_x : B_\delta(0) \oplus B_r(0) \to R$ *and* $f_y : B_\delta(0) \oplus B_r(0) \to R$, respectively.

Furthermore, let y_0, $y_0 + h \in B_r(0)$ with h sufficiently small and $f_x(g(y_0, y_0)) = f_x(g(y_0 + h), y_0 + h) \equiv 0$ and $r(h) = \hat{f}(y_0 + h) - \hat{f}(y_0) - f_y(g(y_0), y_0)h$ be the remainder of Taylor's formula. Then, we have

$$
\begin{aligned}
|r(h)| &= |f(g(y_0+h),y_0+h) - f(g(y_0,y_0)) - f_y(g(y_0),y_0)h| \\
&\leq |f(g(y_0+h),y_0+h) - f(g(y_0),y_0+h)| \\
&\quad +|f(g(y_0),y_0+h) - f(g(y_0,y_0) - f_y(g(y_0),y_0)h| \\
&\leq |g(y_0+h)-g(y_0)| \sup_{0\leq t\leq 1} |f_x(g(y_0)+t(g(y_0+h)-g(y_0)),y_0+h| \\
&\quad +|h| \sup_{0\leq t\leq 1} |f_y(g(y_0),y_0+th) - f_y(g(y_0),y_0)| \\
&\leq M_1|h| \sup_{0\leq t\leq 1} |f_x(g(y_0)+t(g(y_0+h)-g(y_0)),y_0+h) - f_x(g(y_0),y_0)| \\
&\quad +|h| \sup_{0\leq t\leq 1} |f_y(g(y_0),y_0+th) - f_y(g(y_0),y_0)| \\
&\leq M_1M_2|h| \sup_{0\leq t\leq 1} |(t(g(y_0+h)-g(y_0)),h)| + M_3|h| \sup_{0\leq t\leq 1} |(0,th)| \\
&\leq M_1M_2|h|(|g(y_0+h)-g(y_0)|+|h|) + M_3|h|^2 \\
&= (M_1M_2(M_1+1)+M_3)|h|^2 \ .
\end{aligned}
$$

Thus, $\lim_{h\to 0} \dfrac{|r(h)|}{|h|} = 0$ and therefore $\hat{f}$ is differentiable. If $y_0 \in B_r(0)$ is a critical point of $\hat{f}$, then

$$f_x(g(y_0),y_0) = 0, \quad f_y(g(y_0),y_0) = 0$$

and being $(x,y) = (0,0)$ only critical point of f on

$$U = B_\delta(0) \oplus B_r(0),$$

then $y_0 = 0$.

The proof that $(\hat{f})'$ is Lipschitzian on $B_r(0)$, and of items 2., 3. , will be omitted since it is completely elementary. This completes the proof.

□

4. Computation of the local critical groups

Under the hypotheses of the Morse Lemma, Proposition 3.2, if 0 is the only critical point of f and $f(0) = c$. Let V be a closed neighborhood of 0 in H, and a homeomorphism $\psi : V \to \psi(V) \subset H$, $\psi(0) = 0$, $d\psi(0) = I$, such that

$$(f\circ\psi)(x,y) = \frac{1}{2}\langle Ax,x\rangle + \hat{f}(y), \text{ with } g : V\cap N \to N^\perp \text{ and } \hat{f}(y) = f(g(y),y),$$

where $dg(0) = 0$, $g(0) = 0$. Let $C \subset V$ be a closed neighborhood of 0 with $f(0) = \hat{f}(0) = c$. Then the critical groups (over a field F) of V and C satisfy

$$
\begin{aligned}
C_n(f,0) &= H_n(\{f\leq c\}\cap\psi(C)\,,\{f\leq c\}\cap\psi(C)-\{0\}) \\
&\approx H_n(\{f\circ\psi\leq c\}\cap C\,,\{f\circ\psi\leq c\}\cap C-\{0\}) = C_n(f\circ\psi\,,0)
\end{aligned}
$$

The Proposition 3.3, allow us to prove the Shifting Theorem, that is due to Gromoll–Meyer [4]; see also K. C. Chang [18] and [19]. The proof that we give here is inspired by treatment of Mawhin–Willem [15, pg. 190].

Now we consider the following conditions:

(i) Let M a Hilbert manifold and $f \in C^{2-}(M,\mathbb{R})$ such the critical points are isolated;

(ii) $X \subset M$ is positively invariant for the flow of gradient field ∇f (i.e. $\sigma(t,u) \in X$ for all $u \in X$ and $0 \leq t < \beta(u)$);

(iii) $a < b$ are real numbers such that $f^{-1}([a,b]) \cap X$ is complete;
(iv) the condition (C) of Palais-Smale over $f^{-1}([a,b]) \cap X$ is satisfied.

Proposition 4.1 (Deformation Lemma)**.** *Under the above conditions, if $f^{-1}(a,b) \cap X$ is free of critical points. Then $\{f \leq a\} \cap X$ is a strong deformation retract of $\{f \leq b\} \cap X - K_b$, where $K_b = \{u \in X : f(u) = b,\ f'(u) = 0\}$; see* [15, pg. 181].

We shall use the following result of Relative Homology:

Lemma 4.1. *Let A be a subset of $\mathbb{R}^p$ containing 0 and let B^k be the $k-$ball. Then, for $k \geq 1$, $H_n(A \times B^k, (A \times B^k) - \{0\}) \approx H_{n-k}(A, A - \{0\})$.*
In particular, if $A = \{0\}$, we obtain

$$\begin{aligned} H_n(B^k, B^k - \{0\}) &\approx H_n(\{0\} \times B^k, (\{0\} \times B^k) - \{0\}) \\ &\approx H_{n-k}(\{0\}, \phi) = \delta_{n-k,\,0} F = \delta_{n,\,k} F. \end{aligned}$$

Now we prove that the critical groups at a critical point depend on the Morse index and the degenerate part.

Proposition 4.2 (Shifting Theorem)**.** *Let $f : U \subset H \to \mathbb{R}$ be a C^1 function defined on a open set and let $0 \in U$ be only critical point of f, $f(0) = c$. Suppose that f is twice differentiable at 0 and that $A = f''(0)$ is a Fredholm operator, with Morse index k finite, so that f' is strongly differentiable at the origin. Then $C_n(f,0) = C_{n-k}(\hat{f},0)$, $n = 0,1,2,\cdots$.*

Proof. By Proposition 3.2, we consider the function

$$(f \circ \psi)(x,y) = \frac{1}{2}\langle Ax,\, x\rangle + \hat{f}(y) \ \text{ with } \ f(0) = \hat{f}(0) = c.$$

By Proposition 3.3, $0 \in N = Ker(A)$ is the only critical point of the function $\hat{f} : W \subset N \to \mathbb{R}$, $\hat{f}(y) = f(g(y),y)$, W open set of N, $\hat{f}$ is of class C^1 with $(\hat{f})'$ Lipschitzian on W. Since $dimN$ is finite, $\hat{f}$ satisfy the Palais–Smalle condition over any closed ball. Let $\overline{B}_r(0) \subset W$ be a closed ball.

Now we consider the flow defined by the Cauchy problem

$$\dot{\sigma}(t,y) = -\nabla \hat{f}(\sigma(t,y)),\ \sigma(0,y) = y,\ y \in W.$$

Let $\varepsilon > 0$ be sufficiently small and $V = \overline{B}_{\frac{r}{2}}(0) \cap \{\hat{f} \leq c + \varepsilon\} \subset W$, such that, if $y \in V$, $\sigma(t,y)$ stays in $\overline{B}_r(0)$ for $0 \leq t < \beta(y)$ or $\sigma(t,y)$ stays in $\overline{B}_r(0)$ until $\hat{f}(\sigma(t,y)) \leq c - \varepsilon$ and, therefore the trajectory $\sigma(t,y)$ must cross the level $\hat{f}(y) \equiv c$ at an unique point. Let $X = \overline{Y}$ be the closure in W of the set $Y = \{\sigma(t,y) : y \in V,\ 0 \leq t < \beta(y)\}$. Then, X satisfy the following properties:

(i) X is a neighborhood of 0, closed in W, and X is positively invariant for the flow σ defined by $\dot{\sigma}(t,y) = -\nabla \hat{f}(\sigma(t,y))$, $\sigma(t,y) = y$.

(ii) $0 \in \hat{f}^{-1}([c-\varepsilon, c+\varepsilon]) \cap X$ is the only critical point of $\hat{f}$ and belonging to the interior of $\hat{f}^{-1}([c-\varepsilon, c+\varepsilon]) \cap X$.

(iii) $\hat{f}^{-1}([c-\varepsilon, c+\varepsilon]) \cap X$ is complete, because $\overline{B}_r(0)$ is closed in W, the set $\hat{f}^{-1}([c-\varepsilon, c+\varepsilon]) \cap X$ is contained in $\overline{B}_r(0)$ and closed in $\overline{B}_r(0)$ which is complete. The condition (C) of Palais–Smalle is satisfied over $\hat{f}^{-1}([c-\varepsilon, c+\varepsilon]) \cap X$, because $\hat{f}$ satisfy the condition (C) over $\overline{B}_r(0)$.

Now, we define $X^c = \{\hat{f} \le c\} \cap X$ and $X^{c+\varepsilon} = \{\hat{f} \le c+\varepsilon\} \cap X$. We observe that $\hat{f}$ is decreasing during the corresponding deformation σ, obtained of field

$$-\nabla \hat{f}: \;\; \dot{\sigma}(t,y) = -\nabla \hat{f}(t,y), \;\; \sigma(0,y) = y.$$

The set $\hat{f}^{-1}((c, c+\varepsilon]) \cap X = X^{c+\varepsilon} - X^c$ is free of critical points and union of two disjoint subsets S_1 and S_2 such that, for each $y \in S_1$, there is a unique $t(y)$ such that $\hat{f}(\sigma(t(y), y)) = c$ and, for each $y \in S_2$, $\hat{f}(\sigma(t,y)) \to 0$ and $, \sigma(t,y) \to 0$, as $t \to \infty$.

We observe that $\psi(t,y) = \hat{f}(\sigma(t,y))$ satisfies

$$\frac{\partial \psi}{\partial t}(t(y), y)) = -|\nabla \hat{f}(\sigma(t(y), y))|^2 \neq 0$$

and $t(y)$ is continuous by the Implicit Function Theorem. Proposition 4.1 implies X^c is a strong deformation retract of $X^{c+\varepsilon}$.

Let $H = H^- \oplus H^+ \oplus N$ be the orthogonal decomposition into subspaces spanned by the eigenvectors of Fredholm operator $A = f''(0)$ having eigenvalue negative, positive and zero, respectively. A is negative definite on H^-, positive definite on H^+, and $N = KerA$. Let $v = x + y = x^- + x^+ + y$ the corresponding decomposition of any $v \in H$. Define the deformation η of $C = H^- \oplus H^+ \oplus X^{c+\varepsilon}$ by

$$\begin{aligned} \eta : [0,1] \times H^- \times H^+ \times X^{\varepsilon} &\to H^- \times H^+ \times X^{c+\varepsilon} \\ \eta(t, x^-, x^+, y) &= x^- + (1-t)x^+ + \varphi(t,v). \end{aligned}$$

Thus, $H^- \times X^c$ (resp. $(H^- \times X^c) - \{0\}$) is a strong deformation retract of $H^- \times X^{c+\varepsilon}$ (resp. $(H^- \times X^{c+\varepsilon}) - \{0\}$) and we obtain:

$$\begin{aligned} C_n(f,0) &= C_n(f \circ \psi, 0) = H_n(\{f \circ \psi \le c\} \cap C, \{f \circ \psi \le c\} \cap C - \{0\}) \\ &\approx H_n(H^- \times X^c, (H^- \times X^c) - \{0\}). \end{aligned}$$

If $k = dimH^- \ge 0$, by Lemma 4.1, page 9, we obtain:

$$\begin{aligned} C_n(f,0) &\approx H_n(X^c \times \mathbb{R}^k, (X^c \times \mathbb{R}^k) - \{0\}) \\ &\approx H_n(X^c \times B^k, (X^c \times B^k) - \{0\}) \\ &= H_{n-k}(\{\hat{f} \le c\} \cap X, \{\hat{f} \le c\} \cap X - \{0\}) \\ &= C_{n-k}(\hat{f}, 0). \end{aligned}$$

This completes the proof.

□

5. Information about dimension of critical groups

It is a known fact in Morse theory: for a function f defined on open set U of a p–dimensional Euclidean space, $f \in C^2(U, \mathbb{R})$, where v is the only critical point, the function f, can be approximated with respect to the C^2 topology by a function $\tilde{f} \in C^2(U, \mathbb{R})$, with critical points in finite number and non degenerate, and that $dim\ C_n(f,v)$ is finite for every n and equal to zero for $n > p$. This fact is proved in the Lemma 8.6 and Theorem 8.5 of book [15] of J. Mawhin and M. Willem where the C^2 case is treated.

Now we consider a function $f : U \to \mathbb{R}$, with low differentiability, defined on open set U of a p–dimensional Euclidean space, $f \in C^1(U, \mathbb{R})$, where v is the only critical point of f, possible degenerate, and f with second derivative at v, and f' strongly differentiable at v. Under these conditions we can affirm that $dim\ C_n(f,v)$ is finite for every n and is zero for $n > p$.

The answer is still true and less assumptions are needed for the function f: If f is a function defined in an open set U of a Euclidean space of dimension p, f is continuously differentiable on U and v is an isolated critical point of f, then $dim\ C_n(f, v)$ is finite for every n and is zero for $n > p$.

This last result is contained in Theorem 3.2 of the paper [3] of C. Li, S. Li, and J. Liu, or in Theorem 1.1 of paper [22] of S. Cingolani and M. Degiovanni, or in Theorem 1.1 of paper [20] of M. Degiovanni (in finite dimension, all assumptions are trivially satisfied). This last result gives also the required information for $n > p$. Since one can consider Alexander-Spanier cohomology, critical groups can be obtained as a limit from open subsets of the Euclidean space of dimension p, hence must vanish for $n > p$.

By Shifting theorem the computation of the critical groups is reduced to a problem in finite dimension, and as consequence we have the following result: Under the assumptions of Proposition 3.2, if 0 is an isolated critical of f and $A = f''(0)$ is a Fredholm operator with finite Morse index k and nullity ν, then the following is true: $dim\ C_n(f, 0)$ is finite for every n and equal to zero for $n \notin \{k,\ k+1, \cdots,\ k+\nu\}$.

6. Critical submanifolds of a Hilbert manifold

Let M be a Hilbert manifold. We recall that a connected submanifold K is critical for a function $f \in C^1(M, \mathbb{R})$ if $df(x) = 0$, $\forall x \in K$. The tangent space to M at $x \in K$ admits the orthogonal decomposition

$$T_xM = T_xK \oplus H_1(x);\ \ H_1(x) = T_x^{\perp}K.$$

We will suppose f twice differentiable along K and for $x \in K$, $A(x) = \frac{1}{2}d^2f(x)$. If K is critical, $T_xK \subset Ker A(x)$ and since $A(x)$ is symmetric, $A(H_1(x)) \subset H_1(x)$.

Definition 6.1. *We will say that K is non-degenerate critical submanifold if $A(x)|H_1(x)$ is an isomorphism for all $x \in K$.*

By the preceding definition, if K is non-degenerate then $T_xK = Ker A(x)$; see [21]. Now, we will suppose that $f \in C^1(M, \mathbb{R})$, twice differentiable along a compact, connected critical submanifold K that can be degenerate, and $A(x)$ depends continuously on $x \in k$.

Definition 6.2. *Under the above hypotheses, if U is a sufficiently small closed neighborhood of K, $f \equiv 0$ in K, the critical groups of K are defined by*

$$C_n(f, K) = H_n(\{f \leq 0\} \cap U,\ \{f \leq 0\} \cap U - K),\ n = 0,\ 1,\ 2, ...$$

where F is a field of coefficients.(By excision, the critical groups are independent of U.)

7. The manifold of closed curves

The material covered in this item can be encountered in [23]. We denote by M an $n-$dimensional compact manifold endowed with a Riemannian metric $\langle\,,\rangle$. Let ∇ be the covariant derivative on TM, derived from the Levi-Cività connection. Let S be the parametrized circle $[0, 1]/\mathbb{Z}$. We will denote by ΛM and sometimes even simply Λ the set $H^1(S, M)$. Here a map $c : S \to M$ is called H^1, if it is absolutely continuous, and the derivative $\dot{c}(t)$ (which is defined almost everywhere) is square integrable with respect to the Riemannian metric on M : $\|\dot{c}(t)\| \in L^2(S)$, *i.e.*, $\int_S \langle \dot{c}(t), \dot{c}(t)\rangle_{c(t)} \mathrm{d}t < \infty$. We observe that ΛM has a smooth Riemannian manifold structure (modeled on a Hilbert space), which is associated in a natural way to the Riemannian metric on M. Let $c \in C^{\infty}(S, M)$ and consider the pull-back diagram:

$$\begin{array}{ccc} c^*TM & \xrightarrow{c^*_\pi} & TM \\ {\scriptstyle \pi^*_c}\big\downarrow & & \big\downarrow{\scriptstyle \pi} \\ \mathrm{S} & \xrightarrow[c]{} & M \end{array}$$

The Riemannian metric and connection on M can be pulled back to a Riemannian metric and a connection on π^*_c. We denote them by $\langle\, , \rangle_c$ and ∇_c, respectively. Let $\sum(\pi^*_c)$ be the set of all sections of π^*_c and

$$H^0(c^*TM) = L^2(c^*TM) = \{X \in \sum(\pi^*_c) : \|X(t)\|_c \in L^2(S)\}$$
$$H^1(c^*TM) = \{X \in C^0(\pi^*_c) : \nabla_c X \text{ exists a.e. and } \nabla_c X \in H^0(c^*TM)\}.$$

Naturally, $C^k(c^*TM)$ will have the usual meaning for $k = 0, 1, \ldots, \infty$. Then, $H^i(c^*TM)$, $i = 0, 1$, are Hilbert spaces (modulo the relation of being equal a. e.) with respect to the scalar products

$$\langle X\,, Y\rangle_0 \,=\int_S \langle X(t)\,, Y(t)\rangle_c\,\mathrm{d}t\,, \quad \langle X\,, Y\rangle_1 = \langle X\,, Y\rangle_0 + \langle \nabla_c X\,, \nabla_c Y\rangle_0\,,$$

and we denote by $\|\cdot\|_i$ the relative norms. We also consider

$$C^0(T^*M) = \text{ the set of continuous sections}$$

and endow this vector space with the norm $\|X\|_\infty = \sup\limits_{0\le t\le 1} \|X(t)\|$.

Proposition 7.1. *The inclusions*

$$H^1(c^*TM) \hookrightarrow C^0(c^*TM) \hookrightarrow H^0(c^*TM)$$

are continuous, the first one being compact. More precisely,

$$\|\cdot\|_0 \;\le\; \|\cdot\|_\infty \;\le \sqrt{2}\|\cdot\|_1.$$

For $i = 0, 1$, define $H^i(\Lambda M^*TM) = \bigcup_{c\,\in\Lambda M} H^i(c^*TM)$.

Proposition 7.2. $p_i : H^i(\Lambda M^*TM) \longrightarrow \Lambda M$, *where* $p_i(X)(t) = \pi(X(t))$, *has the structure of a (Hilbert) vector bundle over* ΛM *and*

$$p_1 : H^1(\Lambda M^*TM) \longrightarrow \Lambda M$$

is isomorphic to $T\Lambda M$.

Without going into details, we will produce an explicit local trivial structure for $H^i(\Lambda M^*TM)$. Let $\pi : TM \to M$ be the smooth vector bundle and

$$K : T(TM) \to TM$$

a connection. Then $T(TM)$ splits into its horizontal and vertical sub-bundle

$$(T^hTM)_x = \ker(K|T_xTM)\;,\; (T^vTM)_y = \ker(\mathrm{d}\pi)_y\;.$$

For $x \in TM, j = 1, 2$, define

$$(\nabla_j \exp)(x) : T_{\pi(x)}M \longrightarrow T_{\exp x}M,$$

by

$$(\nabla_1 \exp)(x) \cdot y = (d\exp)(x) \circ (d\pi|T^hTM)^{-1} \cdot y,$$
$$(\nabla_2 \exp)(x) \cdot y = (d\exp)(x) \circ K(x)^{-1} \cdot y,$$

where

$$K(x) : T_x^vTM \longrightarrow T_{\pi(x)}M$$

is the canonical identification.

Clearly, for $x \in U \subset TM$, $(\nabla_2 \exp)(x)$ is an isomorphism; the maps

$$\widetilde{\phi}_c^i : H^1(U_c) \times H^i(c^*TM) \longrightarrow H^i(\Lambda M^*TM)$$

$$\widetilde{\phi}_c^i(X,Y)(t) = (\nabla_2 \exp)(c_\pi^* X(t)) \cdot (c_\pi^* Y(t))$$

give the required structure, and $(\widetilde{\phi}_c^1)^{-1} \circ (\widetilde{\phi}_d^1)$ is of form

$$(\widetilde{\phi}_c^{-1} \circ \widetilde{\phi}_d\,,\, d(\widetilde{\phi}_c^{-1} \circ \widetilde{\phi}_d)),$$

so the last assertion follows.

For any $c \in \Lambda M$, $\dot{c}(t) \in H^o(\Lambda M^*TM)$. This gives a section

$$\partial : \Lambda M \to H^o(\Lambda M^*TM).$$

Proposition 7.3. *∂ is a smooth map.*

Theorem 7.1. *The bundle $p_i : H^i(\Lambda M^*TM) \longrightarrow \Lambda M$ has a (unique) Riemannian metric characterized by the following property: For $c \in C^\infty(S,M)$ the metric on $p_i^{-1}(c) = H^i(c^*TM)$ is given by the scalar product $\langle\, ,\rangle_i$.*

Naturally we will keep denoting this Riemannian metric by $\langle\, ,\rangle_i$. In particular,

$$T\Lambda M \cong H^1(\Lambda M^*TM)$$

has a natural Riemannian structure that we will denote also by $\langle\, ,\rangle_i$.

7.1. **The integral energy of a Finsler manifold.** We consider the manifold $\Lambda M = H^1(S,M)$ of H^1-maps of the circle S into M. Let M be a compact manifold and $F : TM \to R$ a Finsler metric on M (see [14], chapter 1, for details). The function $L = F^2 : TM \to \mathbb{R}$ induces a map $E : \Lambda M \to \mathbb{R}$ by $E(c) = \int_S L(\dot{c}(t))\mathrm{d}t$ called energy integral or simply the energy. Let $c \in C^\infty(S,M)$ and $(\phi_c, H^1(U_c))$ a coordinate system near c and $E_c = E \circ \phi_c$, where $U_c = ({c_\pi^*}^{-1})U$, and U is an open set containing the zero section in TM. Then E_c is the composite of the following maps:

$$H^1(U_c) \xrightarrow{I\times\partial_c} H^1(U_c) \times H^0(c^*TM) \xrightarrow{\widetilde{\lambda}_c} L^1(S) \longrightarrow \mathbb{R}$$

where the last map is just integration and $\widetilde{\lambda}_c$ is induced by the fiber map

$$\lambda_c :\, U_c \oplus c^*TM \to S \times \mathbb{R},$$

$$\lambda_c(x,y) = (\pi_c^* x, L((\nabla_2 \exp)(c_\pi^* x) \cdot c_\pi^* y)), \text{where } L = F^2.$$

We note that $\widetilde{\lambda}_c$ is well-defined on all of $H^1(U_c) \times H^0(c^*TM)$. In fact, for $(X,Y) \in H^1(U_c) \times H^0(c^*M)$,

$$\int_S L((\nabla_2 \exp)(c_\pi^* X(t)) \cdot c_\pi^* Y(t))\mathrm{d}t \;\leq\; K \int_S \|(\nabla_2 \exp)(c_\pi^* X(t))\|^2 \|Y(t)\|^2 \mathrm{d}t$$

which is bounded since $\|X(t)\|_\infty$ is small and $Y(t) \in H^0(c^*TM) = L^2(c^*TM)$.

For any $t \in S$, consider the restriction of λ_c to the fiber

$$\lambda_t : (U_c)_t \oplus (c^*TM)_t \to \mathbb{R}.$$

If we denote by x, y the variables in the first and second factor, respectively, we have:

(1) λ_t and $\dfrac{\partial \lambda_t}{\partial x}$ are positively homogeneous of degree 2 in y,

(2) $\dfrac{\partial \lambda_t}{\partial y}$, $\dfrac{\partial^2 \lambda_t}{\partial x \partial y}$, $\dfrac{\partial^2 \lambda_t}{\partial y \partial x}$ are positively homogeneous of degree 1 in y,

(3) $\dfrac{\partial^2 \lambda_t}{\partial y^2}$ is positively homogeneous of degree zero in y.

The closed geodesics problem for the Finsler metric can be posed in an analogous manner to the one for the Riemannian metrics, and the critical points for the function

$$E : \Lambda M \to \mathbb{R}\,, \quad E(c) = \int_S F^2(\dot{c}(t))\mathrm{d}t$$

are exactly the closed geodesics (see [9]).

The gradient field of E is defined by

$$\langle \nabla E(c)\,,\, X\rangle_1 = dE(c).X \text{ for all } X \in T_c\Lambda M.$$

The energy function $E : \Lambda M \to \mathbb{R}$ possesses many properties that are necessary for the development of the theory of Morse of $(\Lambda M, E)$. Among these properties for the function $E : \Lambda M \to \mathbb{R}$ we have the following two:

i) E is C^{2-} (i.e. it is C^1 and the differential is locally Lipschitzian) and therefore E is strongly differentiable.

ii) If M is compact, E satisfy the condition (C) of Palais and Smale:
"Let (c_n) be a sequence on ΛM such that the sequence $(E(c_n))$ is bounded and $(\|\nabla E(c_n)\|_1)$ tends to zero. Then (c_n) has limit point and any limit point is a critical point of E".

Remark 7.1. *Condition (C) should be viewed as a substitute for the fact that ΛM is not locally compact.*

The proof that E is differentiable with differential locally Lipschitzian and that E satisfies the condition (C) of Palais–Smale is due to F. Mercuri and can be found in the paper [9]. To prove $i)$ it is sufficient to show that $\widetilde{\lambda}_c$ is C^{2-} with

$$(d\widetilde{\lambda}_c)(X,Y)(t) \;=\; (d_f\lambda_c)(X(t),Y(t)), \;\; (X,Y) \in H^1(U_c) \times H^0(c^*TM)$$

where d_f denotes derivative on the fiber and in this case

$$(d_f\lambda_c)(X(t),Y(t))(X_1(t),Y_1(t)) = \frac{\partial \lambda_t}{\partial x}(X(t),Y(t)) \cdot X_1(t) + \frac{\partial \lambda_t}{\partial y}(X(t),Y(t)) \cdot Y_1(t)$$

with $(X_1, Y_1) \in H^1(U_c) \times H^0(c^*TM)$. The other property that is necessary for the Morse theory of Finsler energy function is the following:

iii) The derivative of the Finsler energy function $E : \Lambda M \to \mathbb{R}$ is strongly differentiable on the regular curves and particularly on the closed geodesics.

Now our main purpose of this paper is to prove the property $iii)$ enunciated above.

Lemma 7.1. *Homogeneity lemma: Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be continuous, C^∞ on $\mathbb{R}^n - \{0\}$, and positively homogeneous of degree α. Then for all $x, y \in \mathbb{R}^n$:*

(a) *If $\alpha = 1$, there is a constant K with $\|f(x) - f(y)\| \leq K\|x - y\|$.*
(b) *If $\alpha = 2$, there are constants K_1, K_2 with*

$$\|f(x) - f(y)\| \leq K_1\|x - y\|^2 + K_2\|x - y\|\|y\|.$$

Also, the other statement in (b) is

$$\|f(x) - f(y)\| \leq K\,\|x - y\|.max(\|x\|, \|y\|),\ where\ K = max(\|4K_1\|, \|4K_2\|).$$

(c) *If $\alpha = 0$, then f is bounded.*

The proof of Lemma 7.1 will be omitted since it is completely elementary.

8. Strong differentiability of derivative of the Finsler energy in a critical sub-manifold

If we compute the second derivative of $E : \Lambda M \to \mathbb{R}$ we see that we need to use the second derivative of F^2 so that we can carry out the computation only at regular curves, in particular geodesics. The critical points of function $E : \Lambda M \to \mathbb{R}$ are exactly the closed geodesics. Now if c is a closed geodesic the orbit of c under the natural action of $SO(2)$ will give a critical sub-manifold.

Theorem 8.1. *The derivative of the Finsler energy function $E : \Lambda M \to \mathbb{R}$ is strongly differentiable on the regular curves and particularly on the closed geodesics.*

Proof. Let's now prove that the expression

$$\begin{aligned}[d^2\widetilde{\lambda}_c(A,B)](X_1,Y_1)(X_2,Y_2) &= \frac{\partial^2\lambda_t}{\partial x^2}(A,B)X_1X_2 \ + \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B)Y_1X_2 + \\ &+ \frac{\partial^2\lambda_t}{\partial y\partial x}(A,B)X_1Y_2 \ + \frac{\partial^2\lambda_t}{\partial y^2}(A,B)Y_1Y_2 \\ &= d_f^2\lambda_c(A,B)(X_1,Y_1)(X_2,Y_2),\end{aligned}$$

defines the second derivative of $\widetilde{\lambda}_c$. Then, for some $s \in [0,1]$ and

$$(X_1, Y_1) \in H^1(U_c) \times H^0(c^*TM),$$

small enough, we consider the remainder $R(X - W, Y - Z)$ of definition of strongly differentiable function:

$$\|R\,(X-W,Y-Z)\|_{L_1} =$$
$$= \left\| \left[d_f\lambda_c(X,Y)\right](X_1,Y_1) - \left[d_f\lambda_c(W,Z)\right](X_1,Y_1) - \left[d_f^2\lambda_c(A,B)\right](X-W,Y-Z)\,(X_1,Y_1) \right\|_{L_1}$$

$$= \|d_f^2\lambda_c(X+s(X-W),Y+s(Y-Z))(X-W,Y-Z)(X_1,Y_1) -$$
$$- d_f^2\lambda_c(A,B)(X-W,Y-Z)(X_1,Y_1)\|_{L_1} \leq$$

$$\leq \int_S \|\frac{\partial^2\lambda_t}{\partial x^2}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x^2}(A,B)\| \cdot \|X-W\| \cdot \|X_1\| \mathrm{d}t +$$
$$+ \int_S \|\frac{\partial^2\lambda_t}{\partial x\partial y}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B)\| \cdot \|Y-Z\| \cdot \|X_1\| \mathrm{d}t +$$
$$+ \int_S \|\frac{\partial^2\lambda_t}{\partial y\partial x}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial\lambda_t}{\partial y\partial x}(A,B)\| \cdot \|X-W\| \cdot \|Y_1\| \mathrm{d}t +$$
$$+ \int_S \|\frac{\partial^2\lambda_t}{\partial y^2}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial y^2}(A,B)\| \cdot \|Y-Z\| \cdot \|Y_1\| \mathrm{d}t$$

$$\leq \sqrt{2}\|X-W\|_1 \cdot \|X_1\|_\infty \{\int_S \|\frac{\partial^2\lambda_t}{\partial x^2}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x^2}(A,B)\|^2 \mathrm{d}t\}^{\frac{1}{2}} +$$
$$+ \|Y-Z\|_0 \cdot \|X_1\|_\infty \{\int_S \|\frac{\partial^2\lambda_t}{\partial x\partial y}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B)\|^2 \mathrm{d}t\}^{\frac{1}{2}} +$$
$$+ \sqrt{2}\|X-W\|_1 \cdot \|Y_1\|_\infty \{\int_S \|\frac{\partial^2\lambda_t}{\partial y\partial x}(X+s(X-W),Y+s(Y-Z)) -$$
$$- \frac{\partial^2\lambda_t}{\partial y\partial x}(A,B)\|^2 \mathrm{d}t\}^{\frac{1}{2}} + M\|Y_1\|_\infty \cdot \|Y-Z\|_0$$

$$\leq \sqrt{2} \cdot \|Y_1\|_\infty \cdot \|(X-W,Y-Z)\|_{H^1\times H^0}$$
$$\left[\int_S \|\frac{\partial^2\lambda_t}{\partial x^2}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x^2}(A,B)\|^2 \mathrm{d}t\right]^{\frac{1}{2}}$$
$$+ \|X_1\|_\infty \cdot \|(X-W,Y-Z)\|_{H^1\times H^0}$$
$$\left[\int_S \|\frac{\partial^2\lambda_t}{\partial x\partial y}(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B)\|^2 \mathrm{d}t\right]^{\frac{1}{2}}$$
$$+ \sqrt{2}\|Y_1\|_\infty \cdot \|(X-W,Y-Z)\|_{H^1\times H^0} \{\int_S \|\frac{\partial^2\lambda_t}{\partial y\partial x}(X+s(X-W),Y+s(Y-Z)) -$$
$$- \frac{\partial^2\lambda_t}{\partial y\partial x}(A,B)\|^2 \mathrm{d}t\}^{\frac{1}{2}} + M\|Y_1\|_\infty \cdot \|(X-W,Y-Z)\|_{H^1\times H^0}.$$

The proof that the derivative of the Finsler energy function $E : \Lambda M \to \mathbb{R}$ is strongly differentiable on the regular curves and particularly on the closed geodesics, will be concluded if we show that the expressions

$$\int_S \left\| \frac{\partial^2\lambda_t}{\partial x^2}(X+s(X-W),Y+s(Y-Z)) \; - \; \frac{\partial^2\lambda_t}{\partial x^2}(A,B) \right\|^2 \mathrm{d}t$$

and

$$\int_S \left\| \frac{\partial^2\lambda_t}{\partial x\partial y}\,(X+s(X-W),Y+s(Y-Z)) - \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B) \right\|^2 \mathrm{d}t$$

have limits equal to zero when $(X,Y) \to (A,B)$ and $(W,Z) \to (A,B)$.

On one hand,

$$\int_S \left\| \frac{\partial^2\lambda_t}{\partial x^2}(X + s(X-W), Y + s(Y-Z)) \,-\, \frac{\partial^2\lambda_t}{\partial x^2}(A,B) \right\|^2 \mathrm{d}t \leq$$
$$\leq \int_S [\| \frac{\partial^2\lambda_t}{\partial x^2}(X + s(X-W), Y + s(Y-Z)) \,-\, \frac{\partial^2\lambda_t}{\partial x^2}(X + s(X-W), B)\| +$$
$$+ \,\| \frac{\partial^2\lambda_t}{\partial x^2}(X + s(X-W), B) \,-\, \frac{\partial^2\lambda_t}{\partial x^2}(A,B)\|]^2 \mathrm{d}t$$
$$\leq \int_S [K_1 \cdot (\|Y-B\| \,+\, \|Y-Z\|)^2 \,+\, K_2\|B\| \cdot (\|Y-B\| \,+\, \|Y-Z\|) +$$
$$+ \,K_3(\|X-A\| \,+\, \|X-W\|)]^2 \mathrm{d}t.$$

On the other hand,

$$\int_S \left\| \frac{\partial^2\lambda_t}{\partial x\partial y}(X + s(X-W), Y + s(Y-Z)) \,-\, \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B) \right\|^2 \mathrm{d}t \leq$$
$$\leq \int_S [\| \frac{\partial^2\lambda_t}{\partial x\partial y}(X + s(X-W), Y + s(Y-Z)) \,-\, \frac{\partial^2\lambda_t}{\partial x\partial y}(X + s(X-W), B)\| \,+$$
$$+ \,\| \frac{\partial^2\lambda_t}{\partial x\partial y}(X + s(X-W), B)\| \,-\, \frac{\partial^2\lambda_t}{\partial x\partial y}(A,B)\|]^2 \mathrm{d}t$$
$$\leq \int_S [K_4\|Y-B\| \,+\, \|Y-Z\|) \,+\, K_5(\|X-A\| + \|X-W\|)]^2 \mathrm{d}t$$

□

9. Critical submanifolds of the Hilbert manifold ΛM for Finsler energy E

Let M an n-dimensional compact manifold endowed with a Riemannian metric $\langle\,,\rangle$. Let $c \in \Lambda M$ be a closed geodesic of a Finsler metric $F : TM \to \mathbb{R}$ of multiplicity $m \geq 1$, i.e. $c(t) \,=\, c(t + \frac{1}{m})$, for all $t \in [0,1]$. From the index form $d^2E(c)$ associated to the energy $E : \Lambda M \to \mathbb{R}$, $E(c) \,=\, \int_S F^2(\dot{c}(t))\mathrm{d}t$, we obtain in the usual way a self-adjoint operator $A_c :\, T_c\Lambda M \to T_c\Lambda M$ defined by

$$d^2E(c)(X,Y) \,=\, \langle A_cX, Y\rangle_1 \,=\, \langle X, A_cY\rangle_1$$

and A_c is of the form $A_c \,=\, I \,+\, K_c$, where I is the identity and K_c is a compact operator in the space $(T_c\Lambda M, \|\cdot\|_1)$. Therefore A_c is a Fredholm operator.

Let

$$T_c\Lambda M \,=\, T_c^+\Lambda M \,\oplus\, T_c^-\Lambda M \,\oplus\, T_c^0\Lambda M$$

be the orthogonal decomposition of tangent space $T_c\Lambda M$ into sub-spaces spanned by the eigenvectors of A_c having eigenvalue > 0, < 0 and $= 0$ respectively. Theses sub-spaces are $A-invariant$ and A is negative definite on $T_c^-\Lambda M$ and positive definite on $T_c^+\Lambda M$. Then $dim\, T_c^-\Lambda M$ and $dim\, T_c^0\Lambda M$ are finite. We call $dim\, T_c^-\Lambda M$ and $dim\, T_c^0\Lambda M \,-\, 1$ the index and the nullity of c, respectively. Denote by $T_c'\Lambda M$ the sub-space of $T_c\Lambda M$ of codimension 1 which is orthogonal to $\dot{c} \in T_c\Lambda M$. From the above decomposition we get the orthogonal decomposition

$$T_c'\Lambda M \,=\, T_c^+\Lambda M \,\oplus\, T_c^-\Lambda M \,\oplus\, T_c'^{\,0}\Lambda M,$$

where $T_c'^0\Lambda M = T_c^0\Lambda M \cap T_c'\Lambda M$ consists of the periodic Jacobi fields along c which are orthogonal to $\dot{c}$ and $index(c) = dim\, T_c'\Lambda M$, $nullity(c) = dim\, T_c'^0\Lambda M$. We do not assume that $nullity\,(c) = 0$, i.e. that c is non-degenerate.

The closed orbit $S\cdot c$ of a closed geodesic c, is a closed submanifold of ΛM critical for energy $E : \Lambda M \to R$, under the S-action:

$$S \times \Lambda M \to \Lambda M, \quad (z, d) \mapsto z \cdot d$$

with

$$z \cdot d(t) = d(t + r), \text{where } z = e^{2\pi i r} \in S, \; 0 \leq r \leq 1.$$

The critical sub-manifold $S.c$ is compact and connected. The connectedness of $S.c$ implies that the index and nullity of E along $S.c$ are well-defined, i.e. $T^-_{z.c}\Lambda M$ and $T'^0_{z.c}\Lambda M$ has constant dimensions (see [21] and [23]).

The closed geodesic c is non-degenerate, if and only if, the orbit $S \cdot c$ is non-degenerate. This is equivalent to saying that the $nullity(c) = 0$ or the

$$nullity(d^2E(c)) = dim\, T_c^0\Lambda M = 1.$$

Let $\mu = \mu(S \cdot c) : N \to S$ be the normal bundle of closed geodesic c over S, induced for embedding $z \in S \to z^{\frac{1}{m}} \cdot c \in \Lambda M$, $z^{\frac{1}{m}} \cdot c(t) = e^{2\pi i \frac{r}{m}} \cdot c(t) == c(t + \frac{r}{m})$ where $m =$multiplicity of c and $0 \leq r \leq 1$. Note that here we do not assume that $S \cdot c$ is a non-degenerate critical sub-manifold.

Let $\mu = \mu^+ \oplus \mu^- \oplus \mu^0$ be the splitting of the normal bundle, determined by the splitting

$$T_c'\Lambda M = T_c^+\Lambda M \oplus T_c^-\Lambda M \oplus T_c'^0\Lambda M$$

of the fiber.

Using the exponential map, exp, of Levi-Cività connection, we can identify the total space $D = D(S \cdot c)$ of a sufficiently small $\varepsilon - disc$ bundle $D_\varepsilon\mu$ of μ with a open neighborhood of $S \subset N$. The tangent bundle $T\Lambda M$ has in the usual way a Riemannian metric defined by

$$\langle X, Y\rangle_1 = \langle X, Y\rangle_0 + \langle \nabla_{\dot{c}} X, \nabla_{\dot{c}} Y\rangle_0.$$

We use the exponential map to pull the Riemannian metric and Finsler energy integral E back onto D. The action of S on D respects these quantities. For $z \in S$, we denote by D_z the fiber over z in $D\mu$. The restriction of E to D_z will be denoted by E_z. Actually, for our purposes a different metric on D is useful.

Let m be the multiplicity of c. Define on D the following modification of the Riemannian metric

$$\langle X, Y\rangle_m = m^2\langle X, Y\rangle_0 + \langle \nabla_{\dot{\sigma}} X, \nabla_{\dot{\sigma}} Y\rangle_0, \; X, Y \in T_\sigma\Lambda M.$$

On the tangent space of each $X \in D$, the metric $\langle\, ,\rangle_m$ is clearly equivalent to the metric $\langle\, ,\rangle_1$. The index and the nullity of c are not affected by this change of the metric. The function E_z is C^1 and E_z is twice differentiable at O_z with $\mathrm{d}E_z$ strongly differentiable at the origin O_z and $\mathrm{d}E_z(O_z) = 0$.

Using the previously concepts, we now obtain the generalized Morse lemma for degenerate critical points with low differentiability for $E_z = E|D_z$.

Proposition 9.1. *(Generalized Morse Lemma for $E_z = E|D_z$).*
Let c be a closed geodesic of a Finsler metric $F : TM \to \mathbb{R}$, of nullity $l \geq 0$ and

multiplicity $m \geq 1$. *We put* $\mu = \mu^+ \oplus \mu^- \oplus \mu^0$ *the splitting of normal bundle, determined by the splitting of the fiber* $T_c^{'}\Lambda M$:

$$T_c'\Lambda M = T_c^+\Lambda M \oplus T_c^-\Lambda M \oplus T_c'^{\,0}\Lambda M$$

where fiber dimension of μ^0 *is* l. *We put* $\mu^* = \mu^+ \oplus \mu^-$, *i.e.,* $\mu = \mu^* \oplus \mu^0$. *Denote by* O_+,O_-, O_0 *and* O_* *the zero sections of bundles* μ^+, μ^-, μ^0 *and* μ^*, *respectively. Then,*

(1) There is a local homeomorphism ψ *of* D, *such that* $\psi : V \to \psi(V) \subset D$, *where* V *is an open set,* $0 \in V$, $\psi(0) = 0$, *with* ψ *differentiable in* 0 *and* $d\psi(0) = I$.

(2) There is a continuous function $g : D_0 \to D_* = D_+ \oplus D_-$ *that is strongly differentiable in* $O_z \in D_0$, *and* $g(0_z) = 0_z$, $\mathrm{d}g(O_z) = 0$.

(3) There is a section $z \in S \to P_z \in L(\mu^*, \mu^*)$ *with*

$$P_z : X \in D_+ \oplus D_- \to P_z(X) \in D_+$$

being an orthogonal bundle projection such that, for $(X, Y) \in D_* \oplus D_0$

$$\begin{aligned} E_z(X,Y) &\equiv (E_z \circ \psi)(X,Y) = d_X^2 E_z(O_z)(X,X) + E_z(g(Y),Y) \\ &= \langle A_z(0,0)X, X\rangle_m + E_z(g(Y),Y) \\ &= \|P_z(X)\|_m^2 - \|(I - P_z)(X)\|_m^2 + E_z(g(Y),Y) \end{aligned}$$

where $\|X\|_m^2 = \langle A_z(0,0)X^+, X^+\rangle_m - \langle A_z(0,0)X^-, X^-\rangle_m$.
The homeomorphism

$$\psi|D_0 : D_0 \to \psi(D_0) \subset D$$
$$(\psi|D_0)(Y) = g(Y) + Y,\ d(\psi|D_0)(0) = I_N,\ N = T_c'^{\,0}\Lambda M,$$

define a topological sub-manifold $\psi(D_0) \subset D$, *called characteristic sub-manifold at* c.

10. Homological invariants of the energy E at the isolated critical submanifold

Now we will define homological invariants of a closed F- geodesic c of multiplicity $m \geq 1$ and homological invariants of isolated critical orbits $S \cdot c$ of the energy E on ΛM. Let c a closed geodesic of multiplicity m, and $c_0 \in \Lambda M$ a prime closed geodesic defined by $c_0(t) = c(\frac{t}{m})$. Put $E(c) = m^2 E(c_0) = k_m$. Let $\mu = \mu(S \cdot c)$ be the normal bundle over S induced from

$$z \in S \mapsto z^{\frac{1}{m}} \cdot c \in \Lambda M$$

and let

$$\mu(S \cdot c) = \mu^+(S \cdot c) \oplus \mu^-(S \cdot c) \oplus \mu^0(S \cdot c)$$

be the splitting according to the sign of the eigenvalues, introduced earlier.

We denote by $E_z(X, Y)$ the local representation of $E_z = E|D_z$ given by the generalized Morse lemma:

$$E_z : (D_z(S \cdot c), O_z(S \cdot c)) \to (\mathbb{R}, k_m)$$

$$E_z(X,Y) \equiv (E_z \circ \psi)(X,Y) = \|P_z(X)\|^2 - \|(I - P_z)(X)\|^2 + E_z(g(Y),Y)$$

and by $E_{0,z}$ the function given by

$$\widehat{E}_z : (D_{0,\,z}(S \cdot c), O_{0,\,z}(S \cdot c)) \to (\mathbb{R}, k_m)\,,\ \ \widehat{E}_z(Y) = E_z(g(Y),Y)$$

where $k_m = E(c) = m^2 E(c_0)$, $O_z(S \cdot c)$ denote the origin of fiber $D_z(S \cdot c)$ and $O_{0,z}(S \cdot c)$ is the origin of $D_{0,z}(S \cdot c)$.

The homology groups (over the field of rational numbers) defined by

$$C_i(E,c) \;=\; H_i([B_{z,\,\varepsilon}(c) \cap \{E_z \leq k_m\}],\, B_{z,\,\varepsilon}(c) \cap \{E_z \leq k_m\} - \{0\})$$

$$C_i(\widehat{E}_z,c) \;=\; H_i([D^0_{z,\,\varepsilon}(c) \cap \{\widehat{E}_z \leq k_m\}],\, D^0_{z,\,\varepsilon}(c) \cap \{\widehat{E}_z \leq k_m\} - \{0\})$$

are the homological invariants associated to the closed F-geodesic c, of multiplicity m, and $C_i(E,c)$ is the characteristic invariant, $B_{z,\,\varepsilon}(c)$ is an open disc with center at the origin of fiber D_z and radius $\varepsilon > 0$, sufficiently small, $D^0_{z,\,\varepsilon}(c)$ is a small open disc in $(D_0)_z$ of same center that $(D_0)_z$. By excision, these critical groups are independent of $B_{z,\,\varepsilon}(c)$ and $D^0_{z,\,\varepsilon}(c)$.

The numbers $b_i(c) = dim\, C_i(E,c)$ are called the i-th type number, and $b^0_i(c) = dim\, C_i(\widehat{E}_z,c)$ are the i-th singular type number of closed F-geodesic c. Since all constructions are made equivariantly with respect to the S-action on $D(S \cdot c)$, the homology groups and the type numbers are independent from the choice of $z \in S$. These homological invariants are independent the choice of the metric. Let $\langle\,,\rangle$ and $\widetilde{\langle\,,\rangle}$ be two Riemannian metrics in the compact manifold M, then are verified the following conditions:

(1) The unitary tangent bundles of M, in the metrics $\langle\,,\rangle$ and $\widetilde{\langle\,,\rangle}$ are compact. Therefore the induced norms $\|\,,\|$ and $\widetilde{\|\,,\|}$ in each tangent space T_xM are equivalents.

(2) A map $c : S \to M$ is H^1 in the metric $\langle\,,\rangle$, if and only if, c is H^1 in the metric $\widetilde{\langle\,,\rangle}$. Therefore the manifold ΛM and the vector spaces $H^0(c^*TM)$ and $T_c\Lambda M \;=\; H^1(c^*TM)$ do not depend on the used metric for defining them.

(3) In the space $T_c\Lambda M \;=\; H^1(c^*TM)$ with the scalar products

$$\langle X,Y\rangle_1 \;=\; \langle X(t),Y(t)\rangle_0 + \langle \nabla_c X(t), \nabla_c Y(t)\rangle_0$$

$$\widetilde{\langle X,Y\rangle}_1 \;=\; \widetilde{\langle X(t),Y(t)\rangle}_0 + \widetilde{\langle \widetilde{\nabla}_c X(t), \widetilde{\nabla}_c Y(t)\rangle}_0$$

the induced norms $\|\,,\|_1$ and $\widetilde{\|\,,\|}_1$ are equivalents. Therefore, the subspace $T^{'}_c\Lambda M$ of $T_c\Lambda M$ of codimension 1 which is orthogonal to $\dot{c} \in T_c\Lambda M$, and the subspaces $T^+_c\Lambda M$ and $T^-_c\Lambda M$ where $d^2E(c)$ is positive definite and negative definite, do not depend on the Riemannian metric of M.

(4)The homological invariants associated to a closed $F-$geodesic are independent of the choice of the metric, but $D_z,\ E_z$ depend on the metric.
Remark: The proof of above properties will be omitted since it is elementary.

An immediate consequence of Proposition 4.2, page 9, is the following proposition.

Proposition 10.1 (Shifting Theorem for Finsler energy)**.** *The characteristic invariant $C_i(E,c)$ of a closed F-geodesic c, together with the index $\lambda = dim\, T^-_c\Lambda M$, and nullity $l \geq 0$ determines $C_i(E,c)$ completely by*

$$C_i(E,c) \;=\; C_{i-\lambda}(\widehat{E}_z,c).$$

By Proposition 4.2 (Shifting theorem) the computation of the critical groups is reduced to a problem in finite dimension, and as consequence we have the following result: $dim\ C_i(E,c)$ is finite for every i and equal to zero for $i \notin \{\lambda,\, \lambda+1, \cdots,\, \lambda+l\}$ (see, [3], [20], [22]), see also papers on the Morse theory [4], [5], [18], [19].

Now will define a local homological invariant $C_i(E, S\cdot c)$ of the energy E at the isolated critical orbit $S.c$ by

$$C_i(E,S\cdot c) \,=\, H_i(B_\varepsilon(S\cdot c) \cap \{E(S\cdot c) \leq k\}\,,\, B_\varepsilon(S\cdot c) \cap \{E(S\cdot c) \leq k\} - O(S\cdot c))$$

where $B_\varepsilon(S\cdot c) \,=\, \bigcup_{z\in S} B_{z,\varepsilon}(S\cdot c)$ is a small tubular neighborhood, which is a normal bundle of small open discs.

The i-th type number of an isolated critical orbit $S\cdot c$ of energy E is defined by $b_i(S\cdot c) \,=\, dim\, C_i(E, S\cdot c)$. The type number $b_i(S\cdot c)$ of a critical orbit $S\cdot c$ and the singular type number of closed F-geodesic c satisfy the inequality

$$b_i(S\cdot c) \,\leq\, 2[b^\circ_{i-\lambda}(c) \,+\, b^\circ_{i-\lambda-1}(c)]$$

that is obtained making use of the homology theory of action of finite groups and all homological invariants are taken with respect to coefficients in a field of characteristic zero, which is necessary when we use the transfer map as is done here (see [5] and [11]). The inspection of the inequalities above is as follows: We denote by W_c and W_c^- the sets

$$W_c \,=\, [B_\varepsilon(c) \cap \{E \leq k\}]\,, \quad W_c^- \,=\, B_\varepsilon(c) \cap \{E \leq k\} - \{c\}$$

where $E(c) \,=\, k$ and by W and $W^-$the sets $S\cdot W_c$ and $S\cdot W_c^-$ respectively. Now we can write the pair

$$(W, W^-) \,=\, (S \times W_c,\, S\times W_c^-)/\Gamma \,=\, ((S\times W_c)/\Gamma,\, (S\times W_c^-)/\Gamma),$$

where the isotropy group Γ acts on trivial bundle $S \times W_c$ by covering transformations.

Hence,

$$H_i(W, W^-) \,=\, H_i((S\times W_c)/\Gamma,\, (S\times W_c^-)/\Gamma)$$

is isomorphic to the sub-space

$$H_i(S\times W_c,\, S\times W_c^-)^\Gamma,$$

of all elements in $H_i(S\times W_c, S\times W_c^-)$ which are kept fixed under the induced operation of Γ on the homology. Observing that Γ acts trivially on $H_i(S)$ we obtain

$$C_i(E, S\cdot c) \,=\, H_i(S) \otimes H_i(W_c,\, W_c^-)^\Gamma \subset H_i(S) \otimes C_i(E,c).$$

The invariant $C_i(E, S{\cdot}c)$ is of finite type as $C_i(E,c)$, i.e., C_i is finite dimensional and $C_i \,=\, 0$ for almost all i. The fact that $C_{i+\lambda}(E,c) \,=\, C_i(\widehat{E}, c)$ where $\lambda \,=\, index\,(c)$ the last equality above yields

$$C_i(E, S\cdot c) \subset V_i \oplus V_i\,, \quad V_i \,=\, C_{i-\lambda}(\widehat{E}, c) \,\oplus\, C_{i-\lambda-1}(\widehat{E}, c)$$

that in terms of numerical invariants $b_i(S\cdot c) \,=\, dim\, C_i(E, S\cdot c)$ and $b^\circ_i(c) \,=\, dim\, C_i(\widehat{E}, c)$ reads as

$$b_i(S\cdot c) \,\leq\, 2[b^\circ_{i-\lambda}(c) \,+\, b^\circ_{i-\lambda-1}(c)].$$

Let c be a closed F-geodesic and $S.c$ the associated critical sub-manifold. If those critical sub-manifolds are non-degenerate (which is a generic condition on the space of Finsler metrics, see [24]), in this case we compute the index of E and prove that index of the non-degenerate critical sub-manifold obtained by covering m times a given closed F-geodesic and its rotation, is a multiple of the original one.

At this point it is not difficult to obtain the following analogue of the Gromoll-Meyer theorem, without the non-degeneracy hypothesis:

Theorem 10.1. (*Gromoll-Meyer*)*: Let* (M, F) *be a n-dimensional compact simply-connected Finsler manifold satisfying:"If* b_k *denotes the k-th rational Betti number of* ΛM*, there is a sequence* $k_n \to \infty$ *such that* $b_{k_n} \to \infty$*". Then* M *has infinitely many closed F-geodesics (geometrically distinct).*

References

[1] Adriano A. Moura and Fausto M. Souza, *A Morse lemma for degenerate critical points with low differentiability*, Abstract and Applied Analysis 5:2(2002) 113–118, 2000 Mathematics subject classification: 53A07.

[2] A. Cambini, *Sul lemma di Morse*, Boll. Un. Mat. Ital.(4) 7, 87–93, 1973.

[3] C. Li, S. Li, and J. Liu, *Splitting theorem, Poincaré—Hopf theorem and jumping nonlinear problems*, Journal of Functional Analysis 221(2005), 439–455.

[4] D. Gromoll and Meyer, *On differentiable function with isolated critical points*, Topology 8(1969), 361–369. MR 39#7633. Zbl 212.28903.

[5] D. Gromoll and Meyer, *Periodic geodesics on compact, Riemannian manifolds*, J. Differential Geometry 3(1969), 493–510. MR 41#9143. Zbl 203.54401.

[6] E. Lages Lima, *Curso de Análise*, Volume 2, Terceira Edição. Projeto Euclides. Instituto de Matemática Pura e Aplicada (IMPA), CNPq, Rio de Janeiro, 1989.

[7] Fausto M. Souza, *Geodésicas fechadas em variedades de Finsler*, 1997, Doctoral Dissertation, Universidade Estadual de Campinas - UNICAMP, S.P., Brasil.

[8] Fausto M. Souza, *Local perturbation of a Finler metric with a closed geodesic without self-intersections.* (JIM)–Journal of Interdisciplinary Mathematics, Vol.11(April 2008), No.2, pp.247–264. Taru Publications, New Delhi, INDIA.

[9] F. Mercuri, *The critical points theory for the closed geodesics problem*, Math. Z. 156, 231–245(1977).

[10] F. Mercuri and G. Palmieri, *Morse theory with low differentiability*, Boll. Un. Mat. Ital. B(7)1(1987), no.3, 621–631. MR 89f: 58026. zbl 633.58014.

[11] G. Bredon, *Introduction to transformation groups*, Academic Press, 1972. Bull. Soc. Math. Belg. 37, 1985. C. R. Acad. Sci. Paris 230, (1950) p.427–429.

[12] H. B. Rademacher, *Morse Theorie und geschlossene Geodätische*, Bonner Math. Schriften Nr.229, Bonn, 1992.

[13] H. H. Matthias, *Zwei Verallgemeinerungen eines Satzes von Gromoll und Meyer*, Bonner Math. Schriften Nr. 126, 1980.

[14] H. Rund, *The differential geometry of Finsler space*. Grundlehren der Mathematischen Wissenchaften No. 101, Springer: Berlin-Heidelberg-New York 1959.

[15] J. Mawhin and M. Willem, *Critical Point Theory and Hamiltonian Systems*, Applied Mathematical Sciences, vol. 74, Springer-Verlag, New York Inc., 1989.

[16] J. Mawhin and M. Willem, *On the generalized Morse Lemma*, Bull. Soc. Math. Belgique (B)37, 23–29, 1985.

[17] J. Milnor, *Morse Theory*, Ann. Math. Studies Nr. 51, Princeton, N. J., Princeton Univ. Press, 1963.

[18] K. C. Chang, *Infinite dimensional Morse theory and its applications*. Séminaire de Mathématiques Supérieures. Séminaire Scientifique OTAN(Nato Advanced Study Institute). Les Presses de L'Université de Montréal,1985. MR 87m:58032

[19] K. C. Chang. *Infinite-dimensional Morse theory and multiple solution problems*, volume 6 of Progress in Nonlinear Differential Equations and their Applications. Birkhä user Boston Inc., Boston, MA, 1993. MR 94e:58023

[20] M. Degiovanni, *Critical groups of finite type for functionals defined on Banach spaces*, Commun. Appl. Anal. 13 (2009), no. 3, 395-410.

[21] R. Bott, *Non-degenerate critical manifolds*, Ann. of Math.(2)60(1954), 248–261.

[22] S. Cingolani and M. Degiovanni, *On the Poincaré–Hopf theorem for functionals defined on Banach spaces*, Adv. Nonlinear Stud. 9 (2009), no. 4, 679–699.

[23] W. Klingenberg, *Lectures on closed geodesics*, Springer-Verlag Berlin Heidelberg New York 1978.
[24] W. Ziller, *The geometry of the Katok example*, Ergodic Theory Dyn. Syst., 3(1983), 135–157.
[25] Zhongmin Shen, *Lectures on Finsler Geometry*, World Scientific Publishing Co. Pte. Ltd., PO Box 128, Farrer Road, Singapore 912805, 2001.

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 1, Pages 25–35
ISSN: 1930-1235; (**2013**)

GENERATING DIMENSION FORMULAS FOR MULTIVARIATE SPLINES

Simon Foucart[1]
Drexel University,
3141 Chestnut Street,
Philadelphia, PA 19104,
Email: foucart@math.drexel.edu

Tatyana Sorokina
Towson University,
7800 York Road,
Towson, MD 21252,
Email: tsorokina@towson.edu

Abstract. Dimensions of spaces of multivariate splines remain unknown in general. A computational method to obtain explicit formulas for the dimension of spline spaces on simplicial partitions is described. The method is based on Hilbert series and Hilbert polynomials. It is applied to conjecture the dimension formulas for splines on the Alfeld split of a simplex and on several other partitions.

1. Introduction

Let Δ_n denote a simplicial partition of a polyhedral domain $\Omega \subseteq \mathbb{R}^n$, so that if any two simplices in Δ_n intersect, then their intersection is a facet of Δ_n. The space of $\mathcal{C}^r$ splines of degree $\le d$ in n variables on Δ_n is

$$\mathcal{S}_d^r(\Delta_n) := \big\{ s \in \mathcal{C}^r(\Omega) : \; s|_T \in \mathcal{P}_{d,n} \text{ for each simplex } T \in \Delta_n \big\},$$

where $\mathcal{P}_{d,n}$ is the space of polynomials of degree $\le d$ in n variables.

2000 *Mathematics Subject Classification.* Primary 41A15, Secondary 14Q99.
Key words and phrases. multivariate splines, simplicial partition, dimension, Hilbert series, Hilbert polynomial, Alfeld split.
[1]This author is supported by the NSF grant DMS-1120622

We are interested in the dimension of the space $\mathcal{S}_d^r(\Delta_n)$. For fixed d and r, determining a closed formula for arbitrary partitions is still a major open problem, even in the bivariate case, see [9]. In this case, it is known [9, p. 240] that if Δ_2 is a shellable (regular with no holes) triangulation, then

$$(1)\quad \dim \mathcal{S}_d^r(\Delta_2) \geq \binom{d+2}{2} + E_I\binom{d+1-r}{2} - V_I\left[\binom{d+2}{2} - \binom{r+2}{2}\right] + \sum_{v\in\mathcal{V}_I}\sigma_v,$$

where E_I is the number of interior edges, V_I is the number of interior vertices, $\mathcal{V}_I$ is the set of interior vertices of Δ_2, and

$$\sigma_v := \sum_{j=1}^{d-r}\max\{r+j+1-jm_v, 0\},$$

where m_v = number of different edge slopes meeting at v.

The right-hand side of (1) is the correct expression for the dimension if $d \geq 3r+1$, see [9, p.247 and p.273]. Not much is known for $d \leq 3r$, and it is somewhat staggering that the dimensions of $\mathcal{S}_3^1(\Delta_2)$ and of $\mathcal{S}_2^1(\Delta_2)$ remain uncertain in general. Let us point out that the right-hand side of inequality (1) can be rewritten as a linear combination of binomial coefficients – a form that is favored in this paper:

$$\begin{aligned}\dim \mathcal{S}_d^r(\Delta_2) \geq{}& \binom{d+2}{2} + (E_I - 3V_I)\binom{d+1-r}{2}\\ &+ V_I\binom{d+1-\mu}{2} + V_I\binom{d+1-\nu}{2} + \Big(3V_I - \sum_{v\in\mathcal{V}_I} m_v\Big)\binom{\mu+1-r}{2},\end{aligned}$$

where

$$\mu := r + \left\lfloor\frac{r+1}{2}\right\rfloor, \quad \nu := r + \left\lceil\frac{r+1}{2}\right\rceil.$$

In the case of a cell C_2 – a triangulation with one interior vertex v – it is known that the lower bound is the correct dimension, namely

$$\begin{aligned}\dim \mathcal{S}_d^r(C_2) ={}& \binom{d+2}{2} + (E_I - 3)\binom{d+1-r}{2}\\ &+ \binom{d+1-\mu}{2} + \binom{d+1-\nu}{2} + (3-m)\binom{\mu+1-r}{2},\end{aligned}$$

where $m \leq E_I$ is the number of different slopes of the E_I interior edges meeting at v. In fact, the formula for the cell is the basis of the argument used to derive (1). This example demonstrates that the dimension depends not only on the combinatorics of Δ_n – number of vertices, edges, and other faces – but also on its exact geometry. The point of view adopted in this paper consists in fixing the partition and looking for dimension formulas valid for all d, r, and possibly n. The main experimental result, namely Conjecture 1, concerns the spline spaces on the Alfeld split of a single simplex. This split is a generalization of the Clough–Tocher split of a triangle to higher spacial dimensions. The Clough–Tocher split of a triangle has one interior vertex, three interior edges, and three subtriangles. The split of a tetrahedron with one interior vertex, four interior edges, six interior faces, and four subtetrahedra was introduced in [2]. We shall refer to the split of a simplex in $\mathbb{R}^n$ with $\binom{n+1}{k}$ interior k-dimensional faces, $0 \leq k \leq n$, as the Alfeld split A_n. The following is our conjecture on the dimension.

Conjecture 1. *The dimension of the space $\mathcal{S}_d^r(A_n)$ of splines of degree $\leq d$ in n variables over the Alfeld split A_n of a simplex is given by*

$$\dim \mathcal{S}_d^r(A_n) = \binom{d+n}{n} + \begin{cases} n\binom{d+n-\frac{r+1}{2}(n+1)}{n}, & \text{if } r \text{ is odd,} \\ \binom{d+n-1-\frac{r}{2}(n+1)}{n} + \cdots + \binom{d-\frac{r}{2}(n+1)}{n}, & \text{if } r \text{ is even.} \end{cases}$$

This formula was obtained using the computational method that we introduce in Section 2. In Section 3, we describe the steps leading to Conjecture 1, and report without details other formulas obtained via this method for several tetrahedral partitions. In Section 4, we discuss the potential of the method.

2. The computational method

In this section, we show how to derive an explicit formula for the dimension of $\mathcal{S}_d^r(\Delta_n)$, in the form of a linear combination of binomial coefficients, using computed values of this dimension for a finite number of parameters r and d. We first show why the sequence $\{\dim \mathcal{S}_d^r(\Delta_n)\}_{d\geq 0}$ depends only on a finite number of its values.

Let us for now fix the number n of variables, the simplicial partition Δ_n, and the smoothness parameter r. It is well-known that the dimension of $\mathcal{S}_d^r(\Delta_n)$ agrees with a polynomial of degree n in variable d when d is sufficiently large. This polynomial is called the Hilbert polynomial, and it is denoted by $H := H_{\Delta_n,r}$ throughout this paper.

We denote by $d^\star := d^\star_{\Delta_n,r}$ the smallest integer such that

$$\dim \mathcal{S}_d^r(\Delta_n) = H(d) \quad \text{for all } d \geq d^\star.$$

The sequence $\{\dim \mathcal{S}_d^r(\Delta_n)\}_{d\geq 0}$ is determined by its first $d^\star+n+1$ values. Indeed, the terms

$$\{\dim \mathcal{S}_d^r(\Delta_n),\ d^\star \leq d \leq d^\star + n\}$$

define $\{\dim \mathcal{S}_d^r(\Delta_n)\}_{d\geq d^\star}$ by interpolation of the Hilbert polynomial, while the values

$$\{\dim \mathcal{S}_d^r(\Delta_n),\ 0 \leq d \leq d^\star - 1\}$$

complete the first $d^\star$ terms of the sequence. The estimation of $d^\star$ remains a key question. Our method incorporates the widely accepted assumption that

$$d^\star_{\Delta_n,r} \leq r2^n + 1. \tag{2}$$

This is suggested by the technique of partitioning the minimal determining set into non-overlapping subsets associated with each face, see [4]. Moreover, for the subspace of $\mathcal{S}_d^r(\Delta_n)$ imposing additional (or super) smoothness $r2^{n-j-1}$ across every j-dimensional face of Δ_n, it was shown in [5] that the dimension is indeed a polynomial in d for $d \geq r2^n + 1$. The bound (2) is likely to be an overestimation, though. The examples of Section 3 and the improved bound $d^\star_{\Delta_2,r} \leq 3r+2$ obtained in [8] for shellable triangulations supports this belief. Reducing the bound would reduce the number of dimension values to be computed. Since splines with degrees not exceeding smoothness are simply polynomials, we have

$$\dim \mathcal{S}_d^r(\Delta_n) = \binom{d+n}{n} \quad \text{for } d \leq r.$$

Thus, assuming (2), only the $r(2^n-1)+n+1$ values $\{\dim \mathcal{S}_d^r(\Delta_n),\ r+1 \le d \le r2^n+n+1\}$ are left to be computed. An additional saving can be made by using the values for smaller degrees, since we have

$$\left[\dim \mathcal{S}_d^r(\Delta_n) = \dim \mathcal{P}_{d,n} = \binom{d+n}{n}\right] \Longrightarrow \left[\dim \mathcal{S}_k^r(\Delta_n) = \dim \mathcal{P}_{k,n} = \binom{k+n}{n} \text{ for } k \le d\right].$$

Assuming that computing $\dim \mathcal{S}_d^r(\Delta_n)$ is possible for any $d \ge 0$, the above described method gives us access to the whole sequence $\{\dim \mathcal{S}_d^r(\Delta_n)\}_{d\ge 0}$. To obtain an explicit formula, we rely on the concept of Hilbert series, i.e., the generating function of the sequence $\{\dim \mathcal{S}_d^r(\Delta_n)\}_{d\ge 0}$. According to [6, Theorem 2.8], it satisfies

$$\sum_{d\ge 0} \dim \mathcal{S}_d^r(\Delta_n)\, z^d = \frac{P(z)}{(1-z)^{n+1}}, \tag{3}$$

for some polynomial $P := P_{\Delta_n,r}$ with integer coefficients. Denoting these coefficients by $a_k = a_{k,\Delta_n,r}$, and denoting the degree of P by $k^\star = k^\star_{\Delta_n,r}$, that is,

$$P(z) = \sum_{k=0}^{k^\star} a_k z^k, \qquad a_{k^\star} \ne 0,$$

two further particulars are established in [6, Theorem 4.5]:

$$P(1) = \sum_{k=0}^{k^\star} a_k = N, \qquad P'(1) = \sum_{k=0}^{k^\star} k a_k = (r+1)\, F^{\text{int}}, \tag{4}$$

where N and F^{int} represent the number of simplices and interior facets of Δ_n, respectively. In the particular case when Δ_n is a single simplex, the space $\mathcal{S}_d^r(\Delta_n)$ is just the space $\mathcal{P}_{d,n}$ of polynomials of degree d in n variables. Then it can be seen that $P = 1$ from the identity

$$\sum_{d\ge 0} \binom{d+n}{n} z^d = \frac{1}{(1-z)^{n+1}}. \tag{5}$$

This identity is clear for $n = 0$ and is inductively obtained by successive differentiations with respect to z for $n \ge 1$. While the derivation of the polynomial P from the dimensions $\dim \mathcal{S}_d^r(\Delta_n)$ is straightforward, identity (5) conversely provides an explicit formula for the dimensions $\dim \mathcal{S}_d^r(\Delta_n)$ in terms of the coefficients of P. Indeed, the formula

$$\dim \mathcal{S}_d^r(\Delta_n) = \sum_{k=0}^{k^\star} a_k \binom{d+n-k}{n} \tag{6}$$

was isolated in [6] and it also follows from

$$\sum_{d\ge 0} \dim \mathcal{S}_d^r(\Delta_n)\, z^d = \sum_{k=0}^{k^\star} a_k \frac{z^k}{(1-z)^{n+1}} = \sum_{k=0}^{k^\star} \sum_{d\ge 0} a_k \binom{d+n}{n} z^{d+k} = \sum_{d\ge 0} \sum_{k=0}^{k^\star} a_k \binom{d+n-k}{n} z^d$$

by identifying the coefficients in front of each z^d.

Taking into account that

$$\binom{d+n-k}{n} = \begin{cases} \dfrac{(d-k+n)(d-k+n-1)\cdots(d-k+1)}{n!}, & \text{if } d \geq k, \\[2ex] 0 = \dfrac{(d-k+n)(d-k+n-1)\cdots(d-k+1)}{n!}, & \text{if } k-n \leq d \leq k-1, \\[2ex] 0 & \text{if } d \leq k-n-1, \end{cases}$$

we observe that, for $d \geq k^\star - n$, the dimension of $\mathcal{S}^r_d(\Delta_n)$ agrees with the Hilbert polynomial

$$H(d) := \sum_{k=0}^{k^\star} a_k \frac{(d-k+n)(d-k+n-1)\cdots(d-k+1)}{n!}.$$

Moreover, for $d = k^\star - n - 1$, we have

$$H(k^\star - n - 1) - \dim \mathcal{S}^r_{k^\star-n-1}(\Delta_n) = a_{k^\star}\left(\frac{(-1)(-2)\cdots(-n)}{n!} - 0\right) = (-1)^n a_{k^\star} \neq 0.$$

The definition of $d^\star$ therefore yields $d^\star = k^\star - n$, and consequently, we see that

$$k^\star = d^\star + n.$$

This was intuitively anticipated because the determination of the sequence

$$\{\dim \mathcal{S}^r_d(\Delta_n)\}_{d\geq 0}$$

requires $d^\star + n + 1$ pieces of information while the equivalent determination of the polynomial P requires the $k^\star + 1$ pieces of information corresponding to its coefficients.

Now we describe a practical way to determine these coefficients from the computed values $\{\dim \mathcal{S}^r_d(\Delta_n)\}_{d=0}^{d^\star+n}$. It is simply based on the observation that

$$\begin{aligned} a_k &= \frac{1}{k!}\frac{d^k P(z)}{dz^k}|_{z=0} = \frac{1}{k!}\frac{d^k}{dz^k}\Big((1-z)^{n+1}\sum_{d\geq 0}\dim \mathcal{S}^r_d(\Delta_n)\, z^d\Big)|_{z=0} \\ &= \frac{1}{k!}\sum_{\ell=0}^{k}\binom{k}{\ell}\frac{d^{k-\ell}}{dz^{k-\ell}}\Big((1-z)^{n+1}\Big)|_{z=0}\,\frac{d^\ell}{dz^\ell}\Big(\sum_{d\geq 0}\dim \mathcal{S}^r_d(\Delta_n)\, z^d\Big)|_{z=0} \\ &= \frac{1}{k!}\sum_{\ell=0}^{k}\binom{k}{\ell}(-1)^{k-\ell}\frac{(n+1)!}{(n+1-k+\ell)!}\,\ell!\dim \mathcal{S}^r_\ell(\Delta_n) \\ &= \sum_{\ell=0}^{k}(-1)^{k-\ell}\binom{n+1}{k-\ell}\dim \mathcal{S}^r_\ell(\Delta_n). \end{aligned} \tag{7}$$

In particular, the value $\dim \mathcal{S}^r_0(\Delta_n) = 1$ yields $a_0 = 1$, then the value of $\dim \mathcal{S}^r_1(\Delta_n)$ yields a_1, the values of $\dim \mathcal{S}^r_1(\Delta_n)$ and of $\dim \mathcal{S}^r_2(\Delta_n)$ yield a_2 and so on. This shows that the computation of the coefficients a_k can be performed sequentially, along with the computation of the dimensions $\dim \mathcal{S}^r_k(\Delta_n)$. As long as $\dim \mathcal{S}^r_k(\Delta_n)$ equals $\binom{k+n}{n}$, identity (5) ensures that the coefficients a_k agree with the coefficients of the constant polynomial $P = 1$:

$$a_0 = 1, \quad a_1 = 0, \quad a_2 = 0, \quad \cdots, \quad a_{d_\star} = 0,$$

where $d_\star$ denotes the largest integer such that $\dim \mathcal{S}^r_{d_\star}(\Delta_n) = \binom{d_\star+n}{n}$. As a matter of fact, applying (7) to a partition consisting of a single simplex, we obtain

$$0 = \sum_{\ell=0}^{k} (-1)^{k-\ell} \binom{n+1}{k-\ell} \binom{\ell+n}{n}, \quad k \geq 1.$$

We may therefore also express the coefficient a_k as

$$a_k = \sum_{\ell=0}^{k} (-1)^{k-\ell} \binom{n+1}{k-\ell} \delta^r_\ell(\Delta_n), \quad k \geq 1, \tag{8}$$

where $\delta^r_\ell(\Delta_n)$ is the codimension of the polynomial space $\mathcal{P}_{\ell,n}$ in the spline space $\mathcal{S}^r_\ell(\Delta_n)$, i.e.,

$$\delta^r_\ell(\Delta_n) := \dim \mathcal{S}^r_\ell(\Delta_n) - \binom{\ell+n}{n},$$

which is less costly to compute than the dimension of $\mathcal{S}^r_\ell(\Delta_n)$. We finally note that at most $\min\{n+2, k-d_\star\}$ nonzero terms enter the sum in (8), since the summand is nonzero only when $\ell \geq k-n-1$ and $\ell \geq d_\star + 1$.

The computational method described above exploits the specific form of the Hilbert series. As a conclusion to this section, we make the side observation that (3) can be derived by simple means. It suffices to set $u_d = \dim \mathcal{S}^r_d(\Delta_n)$ in the following lemma.

Lemma 1. *Let $\{u_d\}_{d\geq 0}$ be a sequence for which there is a polynomial Q of degree m such that $u_d = Q(d)$ whenever $d \geq \bar{d}$ for some $\bar{d}$. Then there exists a polynomial R such that*

$$\sum_{d\geq 0} u_d\, z^d = \frac{R(z)}{(1-z)^{m+1}}.$$

Proof. We write the polynomial Q as $Q(d) =: \sum_{k=0}^{m} q_k \binom{d+k}{k}$. Then, for the generating function of the sequence $\{u_d\}_{d\geq 0}$, we have

$$\begin{aligned}
\sum_{d\geq 0} u_d\, z^d &= \sum_{d\geq 0} Q(d)\, z^d + \sum_{d\geq 0} (u_d - Q(d))\, z^d \\
&= \sum_{k=0}^{m} q_k \sum_{d\geq 0} \binom{d+k}{k} z^d + \sum_{d=0}^{\bar{d}} (u_d - Q(d))\, z^d \\
&= \sum_{k=0}^{m} q_k \frac{1}{(1-z)^{k+1}} + \sum_{d=0}^{\bar{d}} (u_d - Q(d))\, z^d.
\end{aligned}$$

The latter indeed takes the form $\frac{R(z)}{(1-z)^{n+1}}$ for some polynomial R. □

The previous lemma also enables to reprove (4). Indeed, we notice that the lower and upper bounds derived in [3] yield

$$\dim \mathcal{S}^r_d(\Delta_n) = N \binom{d+n}{n} - (r+1) F^{\text{int}} \binom{d+n-1}{n-1} + \mathcal{O}(d^{n-2}).$$

Therefore, for d large enough, the quantity

$$u_d := \dim \mathcal{S}_d^r(\Delta_n) - N\binom{d+n}{n} + (r+1)F^{\mathrm{int}}\binom{d+n-1}{n-1} \tag{9}$$

reduces to a polynomial of degree $\leq n-2$. The claim implies that, for some polynomial R,

$$\frac{P_{\Delta_n,r}(z)}{(1-z)^{n+1}} - \frac{N}{(1-z)^{n+1}} + \frac{(r+1)F^{\mathrm{int}}}{(1-z)^n} = \sum_{d\geq 0} u_d\, z^d = \frac{R(z)}{(1-z)^{n-1}}.$$

Rearranging the latter, we obtain

$$P_{\Delta_n,r}(z) = N - (r+1)(1-z)F^{\mathrm{int}} + (1-z)^2 R(z),$$

which in turn shows that $P_{\Delta_n,r}(1) = N$ and $P'_{\Delta_n,r}(1) = (r+1)F^{\mathrm{int}}$.

3. Application of the method to specific partitions

In this section, we demonstrate the usefulness of our computational method on several specific partitions. We recall that our method relies on the computation of $\dim \mathcal{S}_d^r(\Delta_n)$ for fixed d, r, and Δ_n. This step was performed using the interactive applet [1] for $n = 3$, and other codes in Java and Fortran for $n > 3$, all written by Peter Alfeld.

3.1. **Alfeld split of a simplex.** We recall that the split of a simplex A_n in $\mathbb{R}^n$ with $\binom{n+1}{k}$ interior k-dimensional faces, $0 \leq k \leq n$, is the Alfeld split of A_n. For $n = 2$, Theorem 9.3 in [9] yields

$$\dim \mathcal{S}_d^r(A_2) = \binom{d+2}{2} + \binom{d+1-\mu}{2} + \binom{d+1-\nu}{2},\ \mu := r + \left\lfloor \frac{r+1}{2} \right\rfloor, \quad \nu := r + \left\lceil \frac{r+1}{2} \right\rceil.$$

For $n = 3$ and $r = 0, 1, 2, 3$, we were able to compute enough values of the dimensions to derive the sequence $\mathbf{a}^{(r)} := (a_0^{(r)}, a_1^{(r)}, a_2^{(r)}, \ldots)$ of coefficients of the polynomial $P_{A_3,r}$ with certainty. We obtained

$$\begin{aligned}
\mathbf{a}^{(0)} &= (1,1,1,1,0,\ldots)\\
\mathbf{a}^{(1)} &= (1,0,0,0,3,0,0,0,0,0,0,0,0,\ldots)\\
\mathbf{a}^{(2)} &= (1,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,\ldots)\\
\mathbf{a}^{(3)} &= (1,0,0,0,0,0,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,\ldots).
\end{aligned}$$

For $r \geq 4$, the dimensions we could compute yielded the start of the sequence $\mathbf{a}^{(r)}$ with certainty, but we cannot be totally sure that all nonzero coefficients have been found. We obtained

$$\mathbf{a}^{(4)} = (1,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,\ldots)$$
$$\mathbf{a}^{(5)} = (1,0,0,0,0,0,0,0,0,0,0,0,3,0,0,0,0,0,0,0,0,0,0,\ldots)$$
$$\mathbf{a}^{(6)} = (1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,\ldots)$$
$$\mathbf{a}^{(7)} = (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,0,0,0,0,\ldots)$$
$$\mathbf{a}^{(8)} = (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,\ldots).$$

Inspection of the sequences $\mathbf{a}^{(1)}$, …, $\mathbf{a}^{(8)}$ strongly suggests the pattern of nonzero coefficients

$$a^{(r)}_{1+2r} = a^{(r)}_{2+2r} = a^{(r)}_{3+2r} = 1 \;\text{ for } r \text{ even}, \qquad a^{(r)}_{2+2r} = 3 \;\text{ for } r \text{ odd}.$$

For $n = 4, 5, 6$, we were also able to compute some values of the dimensions for the space of $\mathcal{C}^r$-splines of degree $\le d$ over the Alfeld split A_n. These investigations lead us to the following conjecture:

(10)

$$\dim \mathcal{S}^r_d(A_n) = \binom{d+n}{n} + \begin{cases} n\dbinom{d+n-\frac{r+1}{2}(n+1)}{n}, & \text{if } r \text{ is odd},\\[2ex] \dbinom{d+n-1-\frac{r}{2}(n+1)}{n} + \cdots + \dbinom{d-\frac{r}{2}(n+1)}{n}, & \text{if } r \text{ is even}. \end{cases}$$

Let us note that for even values of r the formula can be expressed differently since

$$\sum_{j=1}^{n} \binom{d+n-\frac{r}{2}(n+1)-j}{n} = \binom{d+n-\frac{r}{2}(n+1)}{n+1} - \binom{d-\frac{r}{2}(n+1)}{n+1}.$$

The result can be equivalently formulated via the polynomial $P_{A_n,r}$ of (3) as

$$P_{A_n,r}(z) = \begin{cases} 1 + n\, z^{\frac{r+1}{2}(n+1)}, & r \text{ odd},\\[2ex] 1 + \sum_{j=1}^{n} z^{\frac{r}{2}(n+1)+j}, & r \text{ even}. \end{cases}$$

We report below further conjectures produced from our method. Descriptions and illustrations of the corresponding tetrahedral partitions are available in Alfeld's applet menu, see [1].

3.2. **Type-I split of a cube (B_I).** This partition of a cube consists of six tetrahedra, all sharing one main diagonal of the cube. This diagonal is the only interior edge of the partition. There are no interior split points. Type-I split has 6 interior triangular faces, and 18 boundary edges comprised of 12 edges of the cube and six diagonals of its faces. Based on computations for $r \le 8$, we conjecture that

$$\dim \mathcal{S}^r_d(B_I) = \binom{d+3}{3} + 3\binom{d+3-(r+1)}{3} + \begin{cases} 2\dbinom{d+3-\frac{3r+3}{2}}{3}, & r \text{ odd},\\[2ex] \dbinom{d+3-\frac{3r+2}{2}}{3} + \dbinom{d+3-\frac{3r+4}{2}}{3}, & r \text{ even}. \end{cases}$$

3.3. **Worsey–Farin split of a tetrahedron (WF).** This partition is a refinement of the Alfeld split A_3 of a tetrahedron obtained by applying the Clough–Tocher split A_2 to each face of the tetrahedron. The Worsey–Farin split consists of 12 subtetrahedra meeting at one interior point. This partition has 18 interior triangular faces and 8 interior edges. Based on computations for $r \le 8$, we conjecture that

$$\dim \mathcal{S}_d^r(WF) = \binom{d+3}{3} + \begin{cases} 8\binom{d+3-\frac{3r+3}{2}}{2} \\ 4\binom{d+3-\frac{3r+2}{2}}{3} + 4\binom{d+3-\frac{3r+4}{2}}{3} \end{cases}$$

$$+ \begin{cases} 3\binom{d+3-(2r+2)}{3}, & r \text{ odd}, \\ \binom{d+3-(2r+1)}{3} + \binom{d+3-(2r+2)}{3} + \binom{d+3-(2r+3)}{3}, & r \text{ even}. \end{cases}$$

Generic octahedron (OCT). This partition of an octahedron consists of eight tetrahedra meeting at one interior split point. This split point cannot be collinear with any two vertices of the octahedron. There are 12 interior triangular faces and 6 interior edges in this partition. Based on computations for $r \le 8$, we conjecture that

$$\dim \mathcal{S}_d^r(OCT) = \binom{d+3}{3}$$

$$+ \begin{cases} (r+1)\binom{d+3-(2r+1)}{3} + 7\binom{d+3-(2r+2)}{3} - (r+1)\binom{d+3-(2r+3)}{3}, & r = 2 \bmod 3, \\ (r+3)\binom{d+3-(2r+1)}{3} + 3\binom{d+3-(2r+2)}{3} - (r-1)\binom{d+3-(2r+3)}{3}, & \text{otherwise}. \end{cases}$$

Generic 8-cell (C_8). The easiest way to visualize this partition is to start with a refinement of the Alfeld split A_3 of a tetrahedron obtained by applying the Clough–Tocher split A_2 to two faces of the tetrahedron. Let us denote the new split points on the face u and v. This partition consists of 8 subtetrahedra meeting at one interior point. Note that the vertices u and v can be moved to the exterior of the original tetrahedron without changing the topology of the partition. This process results in a partition that has the same number of interior and boundary faces, edges, and vertices as the octahedral partition described above. However, connectivities of the faces are different. For example, each interior edge of the octahedral split is shared by exactly four tetrahedra. In the 8-cell, two interior edges are shared by five tetrahedra, another two are shared by four tetrahedra, and the remaining two edges are shared by three tetrahedra. Based on computations for $r \le 8$, we conjecture that, for $r \ge 2$,

$$\dim \mathcal{S}_d^r(C_8) = \binom{d+3}{3}$$

$$+\begin{cases} 2r\binom{d+3-(2r+1)}{3} - (2r-9)\binom{d+3-(2r+2)}{3} - 2\binom{d+3-(2r+3)}{3}, & r \text{ odd}, \\ (2r+1)\binom{d+3-(2r+1)}{3} - (2r-7)\binom{d+3-(2r+2)}{3} - \binom{d+3-(2r+3)}{3}, & r \text{ even}. \end{cases}$$

We note that the cases $r = 0$ and $r = 1$ do not follow the general pattern.

4. Discussion

4.1. **Towards theoretical improvements.** The main shortcomings of our method are its high complexity and limited reliability.

Complexity. At present, we need to compute an exponential in n number of values of spline dimensions. As n increases, the cost of computing each dimension goes up. This quickly becomes prohibitive. One way to resolve this issue is to lower the bound on $d^\star$. Ideally, it would be a drop from $r2^n + 1$ down to a quantity that is linear in n. Such estimate on the lower bound on $d^\star$ is supported by several observations. If $n = 2$, for shellable triangulations, we have $d^\star \leq 3r + 2$. When $n = 3$, reasonably low values of $d^\star$ can be inferred for the examples of Section 3. We also observed linear behavior in n of $d^\star$ for the Alfled splits. One can also envision that further theoretical information will help to reduce the number of computations. For instance, if a specific partition is known to yield nonnegative coefficients a_k, then we can stop computing $\dim \mathcal{S}_d^r(\Delta_n)$ as soon as the conditions $\sum_{k=0}^d a_k = F_n$ and $\sum_{k=0}^d k a_k = (r+1)F_{n-1}^{\text{int}}$ are satisfied.

Reliability. Even if all necessary values of $\dim \mathcal{S}_d^r(\Delta_n)$ are available for a fixed r, the formula we deduce is only valid for this fixed r. At present, the formula we infer for all values of r relies on a plausible guess. Some theoretical information on the type of dependence of $\dim \mathcal{S}_d^r(\Delta_n)$ on r would be decisive in this respect. The results of Section 3 suggest dependence on the parity of r, sometimes dependence on divisibility of r by 3, and occasionally the predicted dependence is not valid for smaller values of r.

4.2. **Towards computational improvements.** To compute the dimension of $\mathcal{S}_d^r(\Delta_n)$, Alfeld's codes translate the set of smoothness conditions into a linear system for the Bernstein–Bézier coefficients, then the matrix of the system is reduced by Gaussian elimination, and its rank is determined. It may be possible to find faster alternatives. The discussion in [7] hints at a practical method using Gröbner bases. Additionally, when computing $\dim \mathcal{S}_d^r(\Delta_n)$, it should be possible to use the knowledge of the dimensions of the spaces with lower degree and smoothness, since the values $\{\dim \mathcal{S}_d^r(\Delta_n),\ 0 \leq d \leq d^\star + n\}$ are determined sequentially. Finally, to deduce the coefficients a_k, it may be sensible to compute only the quantities $\delta_d^r(\Delta_n)$ appearing in (8), or some suitable linear combinations of $\{\dim \mathcal{S}_d^r(\Delta_n),\ 0 \leq d \leq d^\star + n\}$. This latter approach could take advantage of the fact that the sequence $\{a_k\}$ appears to have only few nonzero terms.

4.3. **An optimistic final perspective.** Should the theoretical and computational improvements materialize, a stand-alone program for the explicit determination of the dimensions ought to be implemented. With modern (or future) computational power, the dimension formulas could be obtained for a wide variety of partitions. It is not unrealistic that some expressions for the coefficients a_k could then be inferred in terms of the smoothness r, the combinatorial parameters, and other topological parameters — especially in the generic case where the geometry does not play a role.

Acknowledgements. We thank Peter Alfeld for the essential computational tools he created and for the enlightening thoughts he shared with us.

References

[1] P. Alfeld, http://www.math.utah.edu/~pa/3DMDS/
[2] P. Alfeld, A trivariate Clough–Tocher scheme for tetrahedral data, Comput. Aided Geom. Design 1 (1984), 169–181.
[3] P. Alfeld, Upper and lower bounds on the dimension of multivariate spline spaces, SIAM J. Numer. Anal. 33 (1996), 571–588.
[4] P. Alfeld, L. L. Schumaker, and M. Sirvent, On dimension and existence of local bases for multivariate spline spaces, J. Approx. Theory 70 (1992), 243–264.
[5] P. Alfeld and M. Sirvent, A recursion formula for the dimension of superspline spaces of smoothness r and degree $d > r2^k$, Approximation Theory V, Proceedings of the Oberwolfach Meeting (1989), W. Schempp and K. Zeller (eds), Birkhäuser Verlag, 1-8.
[6] L. J. Billera and L. L. Rose, A dimension series for multivariate splines, Discrete and Computational Geometry 6 (1991), 107–128.
[7] L. J. Billera and L. L. Rose, Gröbner basis methods for multivariate splines, in: Mathematical methods in computer aided geometric design, T. Lyche and L. L. Schumaker (eds.), 1989, 93–104.
[8] D. Hong, Spaces of bivariate spline functions over triangulation, Approx. Theory Appl. 7 (1991), 56–75.
[9] M.-J. Lai and L. L. Schumaker, Spline functions on triangulations, Cambridge University Press, Cambridge, 2007.

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 1, Pages 37–48
ISSN: 1930-1235; (**2013**)

ABELIAN CROSSED MODULES AND STRICT PICARD CATEGORIES

Nguyen Tien Quang
Department of Mathematics,
Hanoi National University of Education,
Vietnam
E-mail: cn.nguyenquang@gmail.com

Che Thi Kim Phung
Department of Mathematics and Applications,
Saigon University, Vietnam
E-mail: kimphungk25@yahoo.com

Ngo Sy Tung
Department of Mathematics,
Vinh University, Vietnam
E-mail: ngositung@yahoo.com

Abstract. In this paper, we state the notion of morphisms in the category of abelian crossed modules and prove that this category is equivalent to the category of strict Picard categories and regular symmetric monoidal functors. The theory of obstructions for symmetric monoidal functors and symmetric cohomology groups are applied to show a treatment of the group extension problem of the type of an abelian crossed module.

1. Introduction

Crossed modules have been used widely, and in various contexts, since their definition by Whitehead [14] in his investigation of the algebraic structure of second relative homotopy groups. A brief summary of researches related to crossed modules was given in [4] in which Carrasco et al. obtained interesting results on the category

2000 *Mathematics Subject Classification.* Primary 20K40, 20K35; Secondary 18D10.
Key words and phrases. abelian crossed module, symmetric cohomology, group extension, Picard category, obstruction.

of *abelian* crossed modules. The notion of abelian crossed module was characterized by that of the *center* of a crossed module in the paper of Norrie [11].

Crossed modules are essentially the same as *strict categorical groups* (see [7, 3, 1, 10]). A *strict* categorical group is a categorical group in which the associativity, unit constraints are strict ($\mathbf{a} = id, \mathbf{l} = id = \mathbf{r}$) and, for each object x, there is an object y such that $x \otimes y = 1 = y \otimes x$. This concept is also called a *$\mathcal{G}$-groupoid* by Brown and Spencer [3], or a *2-group* by Noohi [10], or a *strict 2-group* by Baez and Lauda [1].

Brown and Spencer [3] (Theorem 1) published a proof that the category of *$\mathcal{G}$-groupoids* is equivalent to the category **CrossMd** of crossed modules (the morphisms in the first category are functors preserving the group structure, those in the second category are homomorphisms of crossed modules).

Another result on crossed modules, the group extension problem of the type of a crossed module, was presented by Brown and Mucuk in [2] (Theorem 5.2). This problem has attracted the attention of many mathematicians.

In our opinion, the above data can be considered for abelian crossed modules. At the beginning of Section 3, we show that each abelian crossed module is seen as a strict Picard category (as defined in Section 2). Therefore, we can apply Picard category theory to study abelian crossed modules and obtain results similar to the above results on crossed modules.

The content of this paper consists of two main results.

In Section 3, we prove that (Theorem 4) the category **Picstr** of strict Picard categories and regular symmetric monoidal functors is equivalent to the category **AbCross** of abelian crossed modules. Every morphism in the category **AbCross** consists of a homomorphism $(f_1, f_0) : \mathcal{M} \to \mathcal{M}'$ of abelian crossed modules and an element of the group of symmetric 2-cocycles $Z^2_s(\pi_0\mathcal{M}, \pi_1\mathcal{M}')$. This theorem is analogous to [3, Theorem 1].

In Section 4, we study the group extension problem of the type of an abelian crossed module. The theory of obstructions for symmetric monoidal functors is applied to show a treatment of this problem. Each abelian crossed module $B \xrightarrow{d} D$ defines a strict Picard category $\mathbb{P}$. The third invariant of $\mathbb{P}$ is an element $\overline{k} \in H^3_s(\operatorname{Coker} d, \operatorname{Ker} d)$. Then a group homomorphism $\psi : Q \to \operatorname{Coker} d$ induces $\overline{\psi^* k} \in Z^3_s(Q, \operatorname{Ker} d)$. Theorem 7 shows that the vanishing of $\psi^* k$ in $H^3_s(Q, \operatorname{Ker} d)$ is necessary and sufficient for there to exist a group extension of the type of an abelian crossed module $B \xrightarrow{d} D$. Each such extension induces a symmetric monoidal functor F: Dis $Q \to \mathbb{P}$. This correspondence determines a bijection (Theorem 6)

$$\Omega : \operatorname{Hom}^{Pic}_{(\psi,0)}[\operatorname{Dis}Q, \mathbb{P}_{B\to D}] \to \operatorname{Ext}^{ab}_{B\to D}(Q, B, \psi).$$

Theorem 7 is analogous to Theorem 5.2 [2].

2. Preliminaries

A *symmetric monoidal category* $\mathbb{P} := (\mathbb{P}, \otimes, I, \mathbf{a}, \mathbf{l}, \mathbf{r}, \mathbf{c})$ consists of a category $\mathbb{P}$, a functor $\otimes : \mathbb{P} \times \mathbb{P} \to \mathbb{P}$ and natural isomorphisms $\mathbf{a}_{X,Y,Z} : (X \otimes Y) \otimes Z \xrightarrow{\sim} X \otimes (Y \otimes Z), \mathbf{l}_X : I \otimes X \xrightarrow{\sim} X, \mathbf{r}_X : X \otimes I \xrightarrow{\sim} X$ and $\mathbf{c}_{X,Y} : X \otimes Y \xrightarrow{\sim} Y \otimes X$ such that, for any objects X, Y, Z, T of $\mathbb{P}$, the following coherence conditions hold:

i) $\mathbf{a}_{X,Y,Z\otimes T}\mathbf{a}_{X\otimes Y,Z,T} = (id_X \otimes \mathbf{a}_{Y,Z,T})\mathbf{a}_{X,Y\otimes Z,T}(\mathbf{a}_{X,Y,Z} \otimes id_T)$,

ii) $\mathbf{c}_{X,Y} \cdot \mathbf{c}_{Y,X} = id_{Y\otimes X}$,

iii) $(id_X \otimes \mathbf{l}_Y)\mathbf{a}_{X,I,Y} = \mathbf{r}_X \otimes id_Y$,
iv) $(id_Y \otimes \mathbf{c}_{X,Z})\mathbf{a}_{Y,X,Z}(\mathbf{c}_{X,Y} \otimes id_Z) = \mathbf{a}_{Y,Z,X}\mathbf{c}_{X,Y\otimes Z}\mathbf{a}_{X,Y,Z}$.

A *Picard category* is a symmetric monoidal category in which every morphism is invertible and, for each object X, there is an object Y with a morphism $X \otimes Y \to I$.

A Picard category is said to be *strict* when the constraints $\mathbf{a} = id$, $\mathbf{c} = id$, $\mathbf{l} = id = \mathbf{r}$ and, for each object X, there is an object Y such that $X \otimes Y = I$.

If $\mathbb{P}$, $\mathbb{P}'$ are symmetric monoidal categories, then a *symmetric monoidal functor* $F := (F, \widetilde{F}, F_*) : \mathbb{P} \to \mathbb{P}'$ consists of a functor $F : \mathbb{P} \to \mathbb{P}'$, natural isomorphisms $\widetilde{F}_{X,Y} : FX \otimes FY \to F(X \otimes Y)$ and an isomorphism $F_* : I' \to FI$, such that, for any objects X, Y, Z of $\mathbb{P}$, the following coherence conditions hold:

$$\widetilde{F}_{X,Y\otimes Z}(id_{FX} \otimes \widetilde{F}_{Y,Z})\mathbf{a}_{FX,FY,FZ} = F(\mathbf{a}_{X,Y,Z})\widetilde{F}_{X\otimes Y,Z}(\widetilde{F}_{X,Y} \otimes id_{FZ}),$$

$$F(\mathbf{r}_X)\widetilde{F}_{X,I}(id_{FX} \otimes F_*) = \mathbf{r}_{FX}\ ,\ F(\mathbf{l}_X)\widetilde{F}_{I,X}(F_* \otimes id_{FX}) = \mathbf{l}_{FX},$$

$$\widetilde{F}_{Y,X}\mathbf{c}_{FX,FY} = F(\mathbf{c}_{X,Y})\widetilde{F}_{X,Y}.$$

Note that if $F := (F, \widetilde{F}, F_*)$ is a symmetric monoidal functor between Picard categories, then the isomorphism $F_* : I' \to FI$ is implied from F and $\widetilde{F}$, so we can omit F_* when not necessary.

A *symmetric monoidal natural equivalence* between symmetric monoidal functors $(F, \widetilde{F}, F_*), (F', \widetilde{F}', F'_*) : \mathbb{P} \to \mathbb{P}'$ is a natural equivalence $\theta : F \xrightarrow{\sim} F'$ such that, for any objects X, Y of $\mathbb{P}$, the following coherence conditions hold:

$$\widetilde{F}'_{X,Y}(\theta_X \otimes \theta_Y) = \theta_{X\otimes Y}\widetilde{F}_{X,Y},\ \theta_I F_* = F'_*.$$

Let $\mathbb{P} := (\mathbb{P}, \otimes, I, \mathbf{a}, \mathbf{l}, \mathbf{r}, \mathbf{c})$ be a Picard category. According to Sinh [13], $\mathbb{P}$ is equivalent to its reduced Picard category $\mathbb{S} = S_{\mathbb{P}}$ thanks to *canonical* equivalences

$$G : \mathbb{P} \to \mathbb{S}, \quad H : \mathbb{S} \to \mathbb{P}.$$

For convenience, we briefly recall the construction of $\mathbb{S}$. Let $M = \pi_0\mathbb{P}$ be the abelian group of isomorphism classes of the objects in $\mathbb{P}$ where the operation is induced by the tensor product, $N = \pi_1\mathbb{P}$ be the abelian group of automorphisms of the unit object I of $\mathbb{P}$ where the operation is composition. Then, objects of $\mathbb{S}$ are elements $x \in M$, and its morphisms are automorphisms $(a, x) : x \to x,\ a \in N$. The composition of morphisms is given by

$$(a, x) \circ (b, x) = (a + b, x).$$

The tensor product is defined by

$$x \otimes y = x + y,$$
$$(a, x) \otimes (b, y) = (a + b, x + y).$$

The unit constraints in $\mathbb{S}$ are strict (in the sense that $\mathbf{l}_x = \mathbf{r}_x = id_x$), the associativity constraint ξ and the symmetry constraint η are, respectively, functions $M^3 \to N,\ M^2 \to N$ satisfying normalized condition:

$$\xi(0, y, z) = \xi(x, 0, z) = \xi(x, y, 0) = 0.$$

and satisfying the following relations:

i) $\xi(y, z, t) - \xi(x + y, z, t) + \xi(x, y + z, t) - \xi(x, y, z + t) + \xi(x, y, z) = 0$,
ii) $\eta(x, y) + \eta(y, x) = 0$,
iii) $\xi(x, y, z) - \xi(y, x, z) + \xi(y, z, x) + \eta(x, y + z) - \eta(x, y) - \eta(x, z) = 0$,

The pair (ξ, η) satisfying these relations is just an element in the group $Z_s^3(M, N)$ of *symmetric* 3-cocycles in the sense of [8]. We refer to $\mathbb{S}$ as *Picard category of type* (M, N).

Let $\mathbb{S} = (M, N, \xi, \eta), \mathbb{S}' = (M', N', \xi', \eta')$ be Picard categories. A functor $F : \mathbb{S} \to \mathbb{S}'$ is called a *functor of type* (φ, f) if there are group homomorphisms $\varphi : M \to M'$, $f : N \to N'$ satisfying

$$F(x) = \varphi(x), \quad F(a, x) = (f(a), \varphi(x)).$$

In this case, (φ, f) is called a *pair of homomorphisms*, and the function

$$k = \varphi^*(\xi', \eta') - f_*(\xi, \eta) \tag{1}$$

is called an *obstruction* of the functor $F : \mathbb{S} \to \mathbb{S}'$ of type (φ, f).

The following proposition is implied from the results on monoidal functors of type (φ, f) in [12] .

Proposition 1. *Let $\mathbb{P}, \mathbb{P}'$ be Picard categories and $\mathbb{S}, \mathbb{S}'$ be their reduced Picard categories, respectively.*

i) Any symmetric monoidal functor $(F, \widetilde{F}) : \mathbb{P} \to \mathbb{P}'$ induces a symmetric monoidal functor $\mathbb{S}_F : \mathbb{S} \to \mathbb{S}'$ of type (φ, f). Further, $\mathbb{S}_F = G'FH$, where H, G' are canonical equivalences.

ii) Any symmetric monoidal functor $(F, \widetilde{F}) : \mathbb{S} \to \mathbb{S}'$ is a functor of type (φ, f).

iii) The functor $F : \mathbb{S} \to \mathbb{S}'$ of type (φ, f) is realizable, i.e., there are isomorphisms $\widetilde{F}_{x,y}$ so that $(F, \widetilde{F})$ is a symmetric monoidal functor, if and only if its obstruction $\overline{k}$ vanishes in $H_s^3(M, N')$. Then, there is a bijection

$$\mathrm{Hom}^{Pic}_{(\varphi,f)}[\mathbb{S}, \mathbb{S}'] \leftrightarrow H_s^2(M, N'),$$

where $\mathrm{Hom}^{Pic}_{(\varphi,f)}[\mathbb{S}, \mathbb{S}']$ denotes the set of homotopy classes of symmetric monoidal functors of type (φ, f) from $\mathbb{S}$ to $\mathbb{S}'$.

3. Classification of abelian crossed modules by strict Picard categories

In this section, we will show a treatment of the problem on classification of abelian crossed modules due to 2-dimensional symmetric cohomology groups and regular symmetric monoidal functors.

We recall that a *crossed module* is a quadruple $\mathcal{M} = (B, D, d, \theta)$, where $d : B \to D$, $\theta : D \to \mathrm{Aut}B$ are group homomorphisms satisfying the following relations:

C_1. $\theta d = \mu$,

C_2. $d(\theta_x(b)) = \mu_x(d(b)), \quad x \in D, b \in B$,

where μ_x is an inner automorphism given by x.

In this paper, the crossed module (B, D, d, θ) is sometimes denoted by $B \xrightarrow{d} D$, or simply $B \to D$.

Standard consequences of the axioms are that $\mathrm{Ker}\, d$ is a left $\mathrm{Coker}\, d$-module under the action

$$sa = \theta_x(a), \quad a \in \mathrm{Ker}\, d, \ x \in s \in \mathrm{Coker}\, d.$$

The groups $\mathrm{Coker}\, d, \mathrm{Ker}\, d$ are also denoted by $\pi_0\mathcal{M}, \pi_1\mathcal{M}$, respectively.

We are interested in the case when B, D are abelian groups. Then, it follows from the condition C_1 that $\theta d = id$ (and hence $\mathrm{Im}d$ acts trivially on B). The condition C_2 leads to $\theta_x(b) - b \in \operatorname{Ker} d$. Therefore, θ determines a function $g : \operatorname{Coker} d \times \operatorname{Ker} d \to \operatorname{Ker} d$ by

$$g(s, b) = sb - b.$$

It is straightforward to see that g is a biadditive normalized function. Conversely, the data $(B \xrightarrow{d} D, g)$, where B, D are abelian, determines completely a crossed module. Particularly, if $g = 0$ we obtain the notion of *abelian crossed module*. In other words, abelian crossed modules are defined as follows.

Definition. A crossed module $\mathcal{M} = (B, D, d, \theta)$ is said to be *abelian* when B, D are abelian and $\theta = 0$.

For example, if $\mathcal{M}$ is a crossed module in which B, D are abelian and d is a monomorphism, then $\theta = 0$. Therefore, $\mathcal{M}$ is an abelian crossed module.

The notion of abelian crossed modules can be characterized by that of the center of crossed modules as in Norrie's work [11]. We say that the *center* $\xi\mathcal{M}$ of a crossed module $\mathcal{M} = (B, D, d, \theta)$ is a subcrossed module of $\mathcal{M}$ and defined by $(B^D, st_D(B) \cap Z(D), d, \theta)$, where B^D is the *fixed point subgroup* of B, $st_D(B)$ is the *stabilizer* in D of B, that is,

$$B^D = \{b \in B : \theta_x b = b \text{ for all } x \in D\},$$
$$st_D(B) = \{x \in D : \theta_x b = b \text{ for all } b \in B\},$$

and $Z(D)$ is the center of D (note that B^D is in the center of B). Then, the crossed module is termed *abelian* if $\xi(B, D, d, \theta) = (B, D, d, \theta)$.

It is well-known that crossed modules are the same as strict categorical groups (see [7], Remark 3.1). Now, we show that abelian crossed modules can be seen as strict Picard categories. We state this in detail.

• For any abelian crossed module $B \to D$, we can construct a strict Picard category $\mathbb{P}_{B\to D} = \mathbb{P}$, called the Picard category *associated to* the abelian crossed module $B \to D$, as follows.

$$\operatorname{Ob}(\mathbb{P}) = D, \quad \operatorname{Hom}(x, y) = \{b \in B \mid x = d(b) + y\},$$

for objects $x, y \in D$. The composition of two morphisms is given by

$$(x \xrightarrow{b} y \xrightarrow{c} z) = (x \xrightarrow{b+c} z). \tag{2}$$

The tensor operation on objects is given by the addition in the group D and, for two morphisms $(x \xrightarrow{b} y), (x' \xrightarrow{b'} y')$ in $\mathbb{P}$, one defines

$$(x \xrightarrow{b} y) \otimes (x' \xrightarrow{b'} y') = (x + x' \xrightarrow{b+b'} y + y'). \tag{3}$$

Associativity, commutativity and unit constraints are identities ($\mathbf{a} = id, \mathbf{c} = id, \mathbf{l} = id = \mathbf{r}$). By the definition of an abelian crossed module, it is easy to check that $\mathbb{P}$ is a strict Picard category.

• Conversely, for a strict Picard category $(\mathbb{P}, \otimes)$, we determine an *associated* abelian crossed module $\mathcal{M}_{\mathbb{P}} = (B, D, d)$ as follows. Set

$$D = \operatorname{Ob}(\mathbb{P}), \quad B = \{x \xrightarrow{b} 0 \mid x \in D\}.$$

The operations in D and B are, respectively, given by

$$x + y = x \otimes y, \quad b + c = b \otimes c.$$

Then D becomes an abelian group whose zero element is 0, and the inverse of x is $-x$ $(x \otimes (-x) = 0)$. B is a group whose zero element is id_0, and the inverse of $(x \xrightarrow{b} 0)$ is the morphism $(-x \xrightarrow{\bar{b}} 0)(b \otimes \bar{b} = id_0)$. Further, B is abelian due to the naturality of the commutativity constraint $\mathbf{c}= id$.

The homomorphism $d : B \to D$ is given by

$$d(x \xrightarrow{b} 0) = x.$$

Definition. A *homomorphism* $(f_1, f_0) : (B, D, d) \to (B', D', d')$ of abelian crossed modules consists of group homomorphisms $f_1 : B \to B'$, $f_0 : D \to D'$ such that

$$f_0 d = d' f_1.$$

Clearly, the category of abelian crossed modules is a full subcategory of the category of crossed modules.

In order to classify abelian crossed modules we establish the following lemmas.

Lemma 2. *Let $(f_1, f_0) : \mathcal{M} = (B, D, d) \to \mathcal{M}' = (B', D', d')$ be a homomorphism of abelian crossed modules. Let $\mathbb{P}$ and $\mathbb{P}'$ be strict Picard categories associated to $\mathcal{M}$ and $\mathcal{M}'$, respectively.*

i) There exists the functor $F : \mathbb{P} \to \mathbb{P}'$ defined by $F(x) = f_0(x)$, $F(b) = f_1(b)$, for $x \in D, b \in B$.

ii) Natural isomorphisms $\widetilde{F}_{x,y} : F(x) + F(y) \to F(x + y)$ together with F is a symmetric monoidal functor if and only if $\widetilde{F}_{x,y} = \varphi(\overline{x}, \overline{y})$, where φ is a symmetric 2-cocycle of the group $Z^2_s(\operatorname{Coker} d, \operatorname{Ker} d')$.

Proof. i) By the determination of a strict Picard category associated to an abelian crossed module and the fact that f_1 is a homomorphism, F is a functor.

ii) Since f_1, f_0 are group homomorphisms, for two morphisms $(x \xrightarrow{b} x')$, $(y \xrightarrow{c} y')$ in $\mathbb{P}$, we have

$$F(b \otimes c) = F(b) \otimes F(c).$$

On the other hand, since f_0 is a homomorphism and $F(x) = f_0(x)$, $\widetilde{F}_{x,y} : F(x) + F(y) \to F(x + y)$ is a morphism in $\mathbb{P}'$ if and only if $d'(\widetilde{F}_{x,y}) = 0'$, i.e.,

$$\widetilde{F}_{x,y} \in \operatorname{Ker} d'.$$

Then the naturality of $(F, \widetilde{F})$, that is the commutativity of the following diagram

$$\begin{tikzcd}
F(x) + F(y) \arrow[r, "\widetilde{F}_{x,y}"] \arrow[d, "F(b)\otimes F(c)"'] & F(x+y) \arrow[d, "F(b\otimes c)"] \\
F(x') + F(y') \arrow[r, "\widetilde{F}_{x',y'}"'] & F(x'+y'),
\end{tikzcd}$$

is equivalent to the relation $\widetilde{F}_{x,y} = \widetilde{F}_{x',y'}$, where $x = d(b) + x', y = d(c) + y'$. This determines a function $\varphi : \operatorname{Coker} d \times \operatorname{Coker} d \to \operatorname{Ker} d'$ by

$$\varphi(\overline{x}, \overline{y}) = \widetilde{F}_{x,y}.$$

By $F(0) = 0'$, the compatibility of $(F, \widetilde{F})$ with unit constraints is equivalent to the normalization of φ. From the relations (2) and (3), the compatibility of $(F, \widetilde{F})$ with associativity, commutativity constraints are, respectively, equivalent to relations

$$\widetilde{F}_{y,z} + \widetilde{F}_{x,y+z} = \widetilde{F}_{x,y} + \widetilde{F}_{x+y,z},$$

$$\widetilde{F}_{x,y} = \widetilde{F}_{y,x}.$$

This shows that $\varphi \in Z^2_s(\mathrm{Coker}\, d, \mathrm{Ker}\, d')$. □

Definition. A symmetric monoidal functor $(F, \widetilde{F}) : \mathbb{P} \to \mathbb{P}'$ between Picard categories $\mathbb{P}, \mathbb{P}'$ is termed *regular* if $F(x) \otimes F(y) = F(x \otimes y)$ for $x, y \in \mathrm{Ob}\mathbb{P}$.

Thanks to Lemma 2, we determine the category **AbCross** whose objects are abelian crossed modules and morphisms are triples (f_1, f_0, φ), where $(f_1, f_0) : (B \xrightarrow{d} D) \to (B' \xrightarrow{d'} D')$ is a homomorphism of abelian crossed modules and $\varphi \in Z^2_s(\mathrm{Coker}\, d, \mathrm{Ker}\, d')$.

Lemma 3. *Let $\mathbb{P}$, $\mathbb{P}'$ be corresponding strict Picard categories associated to abelian crossed modules (B, D, d), (B', D', d'), and let $(F, \widetilde{F}) : \mathbb{P} \to \mathbb{P}'$ be a regular symmetric monoidal functor. Then, the triple (f_1, f_0, φ), where*

$$f_1(b) = F(b),\ f_0(x) = F(x),\ \varphi(s_1, s_2) = \widetilde{F}_{x_1,x_2},$$

for $b \in B,\ x \in D,\ x_i \in s_i \in \mathrm{Coker}\, d, i = 1, 2,$ is a morphism in **AbCross***.*

Proof. Since F is regular, f_0 is a group homomorphism. Since F preserves the composition of morphisms, f_1 is a group homomorphism.

Any $b \in B$ can be considered as a morphism $(db \xrightarrow{b} 0)$ in $\mathbb{P}$, and hence $(F(db) \xrightarrow{F(b)} 0')$ is a morphism in $\mathbb{P}'$. This means that $f_0(d(b)) = d'(f_1(b))$, for all $b \in B$. Thus, (f_1, f_0) is a homomorphism of abelian crossed modules.

According to Lemma 2, $\widetilde{F}_{x_1,x_2}$ determines a function $\varphi \in Z^2_s(\mathrm{Coker}\, d, \mathrm{Ker}\, d')$ by

$$\varphi(s_1, s_2) = \widetilde{F}_{x_1,x_2},\quad x_i \in s_i \in \mathrm{Coker}\, d,\ i = 1, 2.$$

Therefore, (f_1, f_0, φ) is a morphism in **AbCross**. □

Let **Picstr** denote the category of strict Picard categories and regular symmetric monoidal functors, we obtain the following theorem.

Theorem 4 (Classification Theorem)**.** *There exists an equivalence*

$$\begin{array}{rcl} \Phi : \mathbf{AbCross} & \to & \mathbf{Picstr}, \\ (B \to D) & \mapsto & \mathbb{P}_{B \to D}, \\ (f_1, f_0, \varphi) & \mapsto & (F, \widetilde{F}), \end{array}$$

where $F(x) = f_0(x),\ F(b) = f_1(b),\ \widetilde{F}_{x_1,x_2} = \varphi(s_1, s_2)$ for $x \in D,\ b \in B,\ x_i \in s_i \in$ $\mathrm{Coker}\, d,\ i = 1, 2.$

Proof. Suppose that $\mathbb{P}$ and $\mathbb{P}'$ are Picard categories associated to abelian crossed modules $B \to D$ and $B' \to D'$, respectively. By Lemma 2, the correspondence $(f_1, f_0, \varphi) \mapsto (F, \widetilde{F})$ determines an injection on the homsets

$$\Phi : \mathrm{Hom}_{\mathbf{AbCross}}(B \to D, B' \to D') \to \mathrm{Hom}_{\mathbf{Picstr}}(\mathbb{P}_{B\to D}, \mathbb{P}_{B'\to D'}).$$

According to Lemma 3, Φ is surjective.

If $\mathbb{P}$ is a strict Picard category, and $\mathcal{M}_{\mathbb{P}}$ is an abelian crossed module associated to it, then $\Phi(\mathcal{M}_{\mathbb{P}}) = \mathbb{P}$ (not only isomorphic). Therefore, Φ is an equivalence. □

Remark. Denote by **AbCross*** the subcategory of **AbCross** whose morphisms are homomorphisms of abelian crossed modules ($\varphi = 0$), and denote by **Picstr*** the subcategory of **Picstr** whose morphisms are *strict* symmetric monoidal functors ($\widetilde{F} = id$). Then these two categories are equivalent via Φ. This result is analogous to Theorem 1 [3].

4. Classification of group extensions of the type of an abelian crossed module

The concept of group extension of the type of a crossed module was introduced by Dedecker [5] (see also [2]). This concept has a version for abelian crossed modules as follows.

Definition. Let $\mathcal{M} = (B \xrightarrow{d} D)$ be an abelian crossed module, and let Q be an abelian group. An *abelian extension of* B *by* Q *of type* $\mathcal{M}$ is the diagram of group homomorphisms

$$\begin{tikzcd} \mathcal{E}: & 0 \arrow[r] & B \arrow[r, "j"] \arrow[d, equal] & E \arrow[r, "p"] \arrow[d, "\varepsilon"] & Q \arrow[r] & 0, \\ & & B \arrow[r, "d"] & D & & \end{tikzcd} \tag{4}$$

where the top row is exact and (id_B, ε) is a homomorphism of abelian crossed modules.

So, any extension of the type of an abelian crossed module is an extension of the type of a crossed module.

Two extensions $\mathcal{E}, \mathcal{E}'$ of B by Q of type $\mathcal{M}$ are said to be *equivalent* if the following diagram commutes

$$\begin{tikzcd} \mathcal{E}: & 0 \arrow[r] & B \arrow[r, "j"] \arrow[d, equal] & E \arrow[r, "p"] \arrow[d, "\alpha"] & Q \arrow[r] \arrow[d, equal] & 0, & E \arrow[r, "\varepsilon"] & D \\ \mathcal{E}': & 0 \arrow[r] & B \arrow[r, "j'"] & E' \arrow[r, "p'"] & Q \arrow[r] & 0, & E' \arrow[r, "\varepsilon'"] & D \end{tikzcd} \tag{5}$$

and $\varepsilon'\alpha = \varepsilon$. Obviously, α is an isomorphism.

In the diagram

$$\begin{tikzcd} \mathcal{E}: & 0 \arrow[r] & B \arrow[r, "j"] \arrow[d, equal] & E \arrow[r, "p"] \arrow[d, "\varepsilon"] & Q \arrow[r] \arrow[d, dotted, "\psi"] & 0, \\ & & B \arrow[r, "d"] & D \arrow[r, "q"] & \mathrm{Coker} d & \end{tikzcd}$$

since the top row is exact and $q \circ \varepsilon \circ j = q \circ d = 0$, there is a homomorphism $\psi : Q \to \mathrm{Coker}\, d$ such that the right hand side square commutes. Moreover, ψ is dependent only on the equivalence class of the extension $\mathcal{E}$.

Our objective is to study the set

$$\mathrm{Ext}^{ab}_{B\to D}(Q, B, \psi)$$

of equivalence classes of extensions of B by Q of type $B \xrightarrow{d} D$ inducing $\psi : Q \to \mathrm{Coker}\, d$. It is well-known that the set $\mathrm{Ext}_{B\to D}(Q, B, \psi)$ for extensions of the type of a (not necessarily abelian) crossed module was classified by Brown and Mucuk. In the present paper, we use the obstruction theory of symmetric monoidal functors to prove Theorem 7 which is an abelian analogue of Theorem 5.2 in [2]. The second assertion of this theorem can be seen as a consequence of Schreier Theory (Theorem 6) due to symmetric monoidal functors between strict Picard categories $\mathbb{P}_{B\to D}$ and $\mathrm{Dis}\, Q$, where $\mathrm{Dis}\, Q$ is a Picard category of type $(Q, 0, 0)$.

Lemma 5. *Let $B \xrightarrow{d} D$ be an abelian crossed module, Q be an abelian group and $\psi : Q \to \operatorname{Coker} d$ be a group homomorphism. Then, for each symmetric monoidal functor $(F, \widetilde{F}) : \operatorname{Dis} Q \to \mathbb{P}$ which satisfies $F(0) = 0$ and induces the pair $(\psi, 0) : (Q, 0) \to (\operatorname{Coker} d, \operatorname{Ker} d)$, there exists an extension $\mathcal{E}_F$ of B by Q of type $B \to D$ inducing ψ.*

Such an extension $\mathcal{E}_F$ is called *associated* to a symmetric monoidal functor $(F, \widetilde{F})$.

Proof. Suppose that $(F, \widetilde{F}) :\ \operatorname{Dis} Q \to \mathbb{P}$ is a symmetric monoidal functor. Then, we set a function $f : Q \times Q \to B$ as follows

$$f(u, v) = \widetilde{F}_{u,v}.$$

Because $\widetilde{F}_{u,v}$ is a morphism in $\mathbb{P}$, one has

$$F(u) + F(v) = df(u, v) + F(u + v).$$

Since $F(0) = 0$ and $(F, \widetilde{F})$ is compatible with the strict constraints of Dis$Q$ and $\mathbb{P}$, f is a normalized function satisfying

$$f(v, t) + f(u, v + t) = f(u, v) + f(u + v, t), \tag{6}$$

$$f(u, v) = f(v, u). \tag{7}$$

Now we construct the semidirect product $E_0 = [B, f, Q]$, that is, $E_0 = B \times Q$ with the operation

$$(b, u) + (c, v) = (b + c + f(u, v), u + v).$$

The set E_0 is an abelian group due to the normalization of f and the relations (6), (7), the zero element is $(0, 0)$ and $-(b, u) = (-b - f(u, -u), -u)$. Then, we have an exact sequence of abelian groups

$$\mathcal{E}_F :\ \ 0 \to B \xrightarrow{j_0} E_0 \xrightarrow{p_0} Q \to 0,$$

where

$$j_0(b) = (b, 0),\ \ p_0(b, u) = u,\ \ b \in B, u \in Q.$$

The map $\varepsilon : E_0 \to D$ given by

$$\varepsilon(b, u) = db + F(u),\ \ (b, u) \in E_0,$$

is a homomorphism, and hence the pair (id_B, ε) is a homomorphism of abelian crossed modules. Therefore, one obtains an extension of the type of an abelian crossed module $\mathcal{E}_F$ satisfying the diagram (4). For all $u \in Q$, one has

$$q\varepsilon(b, u) = q(db + F(u)) = qF(u) = \psi(u),$$

i.e., this extension induces $\psi : Q \to \operatorname{Coker} d$. □

Under the assumptions of Lemma 5, we have

Theorem 6 (Schreier Theory for group extensions of the type of an abelian crossed module)**.** *There exists a bijection*

$$\Omega : \operatorname{Hom}^{Pic}_{(\psi,0)}[\text{Dis}Q, \mathbb{P}_{B\to D}] \to \operatorname{Ext}^{ab}_{B\to D}(Q, B, \psi)$$

if one of the above sets is nonempty.

Proof. Step 1: Symmetric monoidal functors $(F,\widetilde{F})$, $(F',\widetilde{F}')$ are homotopic if and only if the corresponding associated extensions $\mathcal{E}_F, \mathcal{E}_{F'}$ are equivalent.

First, since every symmetric monoidal functor $(F,\widetilde{F})$ is homotopic to one $(G,\widetilde{G})$ in which $G(0)=0$, the following symmetric monoidal functors are regarded as the functors which have this property .

Suppose that $F, F' : \mathrm{Dis}Q \to \mathbb{P}_{B\to D}$ are homotopic by a homotopy $\alpha : F \to F'$. By Lemma 5, there exist the extensions $\mathcal{E}_F$ and $\mathcal{E}_{F'}$ associated to F and F', respectively. Then, it follows from the definition of a homotopy that $\alpha_0 = 0$ and the following diagram commutes

$$\begin{tikzcd} Fu+Fv \arrow[r, "\widetilde{F}_{u,v}"] \arrow[d, "\alpha_u\otimes\alpha_v"'] & F(u+v) \arrow[d, "\alpha_{u+v}"] \\ F'u+F'v \arrow[r, "\widetilde{F}'_{u,v}"'] & F'(u+v), \end{tikzcd}$$

that is,

$$\widetilde{F}_{u,v} + \alpha_{u+v} = \alpha_u \otimes \alpha_v + \widetilde{F}'_{u,v}.$$

By the relation (3), one has

$$f(u,v) + \alpha_{u+v} = \alpha_u + \alpha_v + f'(u,v), \tag{8}$$

where $f(u,v) = \widetilde{F}_{u,v}, f'(u,v) = \widetilde{F}'_{u,v}$. Now we set

$$\begin{aligned} \alpha^* : E_F &\to E_{F'} \\ (b,u) &\mapsto (b+\alpha_u, u). \end{aligned}$$

Then α^* is a homomorphism thanks to the relation (8). Further, the diagram (5) commutes. It remains to show that $\varepsilon'\alpha^* = \varepsilon$. Since $\alpha : F \to F'$ is a homotopy, $F(u) = d(\alpha_u) + F'(u)$. Then,

$$\begin{aligned} \varepsilon'\alpha^*(b,u) &= \varepsilon'(b+\alpha_u,u) = d(b+\alpha_u) + F'(u) \\ &= d(b) + d(\alpha_u) + F'(u) = d(b) + F(u) = \varepsilon(b,u). \end{aligned}$$

Therefore, $\mathcal{E}_F$ and $\mathcal{E}_{F'}$ are equivalent.

Conversely, if an isomorphism $\alpha^* : E_F \to E_{F'}$ satisfying the triple (id_B, α^*, id_Q) is an equivalence of extensions, then it is easy to see that

$$\alpha^*(b,u) = (b+\alpha_u, u),$$

where $\alpha : Q \to B$ is a function satisfying $\alpha_0 = 0$. By retracing our steps, α is a homotopy between F and F'.

Step 2: Ω is surjective.

Assume that $\mathcal{E}$ is an extension E of B by Q of type $B \to D$ inducing $\psi : Q \to \mathrm{Coker}\, d$. We prove that $\mathcal{E}$ is equivalent to the semidirect product extension $\mathcal{E}_F$ which is associated to a symmetric monoidal functor $(F,\widetilde{F}) :\ \mathrm{Dis}Q \to \mathbb{P}_{B\to D}$.

For any $u \in Q$, choose a representative $e_u \in E$ such that $p(e_u) = u,\ e_0 = 0$. Each element of E can be represented uniquely as $b + e_u$ for $b \in B, u \in Q$. The representatives $\{e_u\}$ induces a normalized function $f : Q \times Q \to B$ by

$$e_u + e_v = f(u,v) + e_{u+v}. \tag{9}$$

Then, the group structure of E can be described by

$$(b+e_u) + (c+e_v) = b + c + f(u,v) + e_{u+v}.$$

Now, we construct a symmetric monoidal functor $(F, \widetilde{F}):\ \mathrm{Dis}\,Q \to \mathbb{P}$ as follows. Since $\psi(u) = \psi p(e_u) = q\varepsilon(e_u)$, $\varepsilon(e_u)$ is a representative of $\psi(u)$ in D. Thus, we set

$$F(u) = \varepsilon(e_u),\ \widetilde{F}_{u,v} = f(u,v).$$

The relation (9) shows that $\widetilde{F}_{u,v}$ are actually morphisms in $\mathbb{P}$. Obviously, $F(0) = 0$. This together with the normalization condition of the function f implies the compatibility of $(F, \widetilde{F})$ with the unit constraints. The associativity and commutativity laws of the operation in E lead to the relations (6), (7), respectively. These relations prove that $(F, \widetilde{F})$ is compatible with the associativity and commutativity constraints of Dis$Q$ and $\mathbb{P}$, respectively. The naturality of $\widetilde{F}_{u,v}$ and the condition of F preserving the composition of morphisms are obvious.

Finally, it is easy to check that the semidirect product extension $\mathcal{E}_F$ associated to $(F, \widetilde{F})$ is equivalent to the extension $\mathcal{E}$ by the isomorphism $\beta : (b, u) \mapsto b + e_u$. □

Let $\mathbb{P} = \mathbb{P}_{B\to D}$ be a strict Picard category associated to an abelian crossed module $B \to D$. Since $\pi_0(\mathbb{P}) = \mathrm{Coker}\, d$ and $\pi_1(\mathbb{P}) = \mathrm{Ker}\, d$, the reduced Picard category $S_{\mathbb{P}}$ is of form

$$S_{\mathbb{P}} = (\mathrm{Coker}d, \mathrm{Ker}d, k),\ \overline{k} \in H^3_s(\mathrm{Coker}d, \mathrm{Ker}d).$$

Then, by the relation (1), the pair of homomorphisms $(\psi, 0) : (Q, 0) \to (\mathrm{Coker}d, \mathrm{Ker}\, d)$ induces an *obstruction*

$$\psi^* k \in Z^3_s(Q, \mathrm{Ker}\, d).$$

Under this notion of obstruction, we state and prove the following theorem.

Theorem 7. *Let (B, D, d) be an abelian crossed module, and let $\psi : Q \to \mathrm{Coker}\, d$ be a homomorphism of abelian groups. Then, the vanishing of $\overline{\psi^* k}$ in $H^3_s(Q, \mathrm{Ker}\, d)$ is necessary and sufficient for there to exist an extension of B by Q of type $B \to D$ inducing ψ. Further, if $\overline{\psi^* k}$ vanishes, then the set of equivalence classes of such extensions is bijective with $H^2_s(Q, \mathrm{Ker}\, d)$.*

Proof. By the assumption $\overline{\psi^* k} = 0$, it follows by Proposition 1 that there is a symmetric monoidal functor $(\Psi, \widetilde{\Psi}) : \mathrm{Dis}\, Q \to S_{\mathbb{P}}$. Then the composition of $(\Psi, \widetilde{\Psi})$ and the canonical symmetric monoidal functor $(H, \widetilde{H}) : S_{\mathbb{P}} \to \mathbb{P}$ is a symmetric monoidal functor $(F, \widetilde{F}) : \mathrm{Dis}\, Q \to \mathbb{P}$, and hence by Lemma 5, we obtain an associated extension $\mathcal{E}_F$.

Conversely, suppose that there is an extension as in the diagram (4). Let $\mathbb{P}'$ be a strict Picard category associated to the abelian crossed module $B \to E$. Then, according to Lemma 2, there is a symmetric monoidal functor $F : \mathbb{P}' \to \mathbb{P}$. Since the reduced Picard category of $\mathbb{P}'$ is $\mathrm{Dis}\, Q$, by Proposition 1 *i*), F induces a symmetric monoidal functor of type $(\psi, 0)$ from $\mathrm{Dis}\, Q$ to $S_{\mathbb{P}} = (\mathrm{Coker}\, d, \mathrm{Ker}\, d, k)$. Now, thanks to Proposition 1 *iii*), the obstruction of the pair $(\psi, 0)$ vanishes in $H^3_s(Q, \mathrm{Ker}\, d)$, i.e., $\overline{\psi^* k} = 0$.

The final assertion of the theorem is obtained from Theorem 6. First, there is a natural bijection

$$\mathrm{Hom}^{Pic}_{(\psi,0)}[\mathrm{Dis}\, Q, \mathbb{P}] \leftrightarrow \mathrm{Hom}^{Pic}_{(\psi,0)}[\mathrm{Dis}\, Q, S_{\mathbb{P}}].$$

Since $\pi_0(\mathrm{Dis}\, Q) = Q, \pi_1(S_{\mathbb{P}}) = \mathrm{Ker}\, d$, the bijection

$$\mathrm{Ext}^{ab}_{B\to D}(Q, B, \psi) \leftrightarrow H^2_s(Q, \mathrm{Ker}\, d)$$

follows from Theorem 6 and Proposition 1. □

In the case when the homomorphism d of the abelian crossed module $\mathcal{M}$ is a monomorphism, then the diagram (4) shows that the extension $(\mathcal{E} : B \to E \to Q)$ is obtained from the extension $(\mathcal{D} : B \to D \to \operatorname{Coker} d)$ and ψ, i.e., $\mathcal{E} = \mathcal{D}\psi$ (see [9, 6]). Since $\operatorname{Ker} d = 0$, by Theorem 7, we obtain a well-known result as follows.

Corollary 8. *Let $(\mathcal{D} : B \to D \to C)$ be an extension of abelian groups and $\psi : Q \to C$ be a homomorphism of abelian groups. Then, there is an extension $\mathcal{D}\psi$ determined uniquely up to equivalence.*

Acknowledgement The authors are much indebted to the referee, whose useful observations greatly improved our exposition.

References

[1] J. C. Baez and A. D. Lauda, *Higher Dimensional Algebra V: 2-groups*, Theory Appl. Categ. 12 (2004) 423–491.

[2] R. Brown and O. Mucuk, *Covering groups of non-connected topological groups revisited*, Math. Proc. Camb. Phil. Soc. 115 (1994) 97–110.

[3] R. Brown and C. B. Spencer, *$\mathcal{G}$-groupoids, crossed modules and the fundamental groupoid of a topological group*, Proc. Konn. Ned. Akad. v. Wet. 79 (1976) 296–302.

[4] P. Carrasco, A. M. Cegarra and A. R. -Grandjeán, *(Co)homology of crossed modules*, Category theory 1999 (Coimbra). J. Pure Appl. Algebra 168 no. 2-3 (2002) 147–176.

[5] P. Dedecker, *Les foncteurs Ext_Π, H^2_Π et H^2_Π non abéliens*, C. R. Acad. Sci. Paris 258 (1964) 4891–4894.

[6] L. Fuchs, *Infinite Abelian Groups*, Vol. 1, Academic Press New York and London 1970.

[7] A. Joyal and R. Street, *Braided tensor categories,* Advances in Mathematics 102 (1993) 20–78.

[8] S. Mac Lane, *Cohomology theory of Abelian groups,* from: "Proceedings of the International Congress of Mathematicians, Cambridge, Mass., Vol. 2", Amer. Math. Soc. (1952) 8–14.

[9] S. Mac Lane, *Homology,* Springer-Verlag 1963.

[10] B. Noohi, *Notes on 2-groupoids, 2-groups and crossed modules,* Homology, Homotopy Appl. 9 (2007) 75–106.

[11] K. Norrie, *Actions and automrphisms of crossed modules*, Bull. Coc. Math. France Tome 118 No 2 (1990) 129–146.

[12] N. T. Quang, N. T. Thuy and P. T. Cuc, *Monoidal functors between (braided) Gr-categories and their applications*, East-West J. of Mathematics, Vol. 13 No 2 (2011) 170–193.

[13] H. X. Sinh, *Gr-catégories*, Thèse de doctorat, Université Paris VII 1975.

[14] J. H. C. Whitehead, *Combinatorial homotopy II*, Bull. Amer. Math. Soc 55 (1949) 453–496.

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 2, Pages 51–60
ISSN: 1930-1235; (**2013**)

QUASICONFORMAL HARMONIC MAPPINGS ONTO A CONVEX DOMAIN REVISITED

DAVID KALAJ
Faculty of Natural Sciences and Mathematics,
University of Montenegro,
Cetinjski put b.b. 81000
Podgorica, Montenegro
E-mail: *davidk@t-com.me*

KEN-ICHI SAKAN
Osaka City University,
Graduate School of Science,
Department of Mathematics Sugimoto,
Sumiyoshi-ku, Osaka, 558, Japan
E-mail: *ksakan@sci.osaka-cu.ac.jp*

ABSTRACT. We give an explicit dependence of quasiconformal constant on its boundary function, provided that the mapping is quasiconformal harmonic and maps the unit disk onto a strictly convex domain. This result refines some earlier results obtain by the first author and Pavlović ([11, 27]).

1. INTRODUCTION AND STATEMENT OF THE MAIN RESULTS

1.0.1. *Harmonic mappings.* The function

$$P(r,t) = \frac{1-r^2}{2\pi(1-2r\cos t + r^2)}, \quad 0 \le r < 1, \ \ t \in [0, 2\pi]$$

is called the Poisson kernel. Let $\mathbf{U} = \{z : |z| < 1\}$ be the unit disk and $\mathbf{T} = \partial \mathbf{U}$ is the unit circle. The Poisson integral of a complex function $F \in L^1(\mathbf{T})$ is a complex harmonic mapping given by

$$w(z) = u(z) + iv(z) = P[F](z) = \int_0^{2\pi} P(r, t-\tau) F(e^{it}) dt, \tag{1.1}$$

where $z = re^{i\tau} \in \mathbf{U}$. If w is a bounded harmonic mapping, then there exists a function $F \in L^\infty(\mathbf{T})$, such that $w(z) = P[F](z)$ (see e.g. [4, Theorem 3.13 b), $p = \infty$]). From now on we will identify $F(t)$ with $F(e^{it})$ and $F'(t)$ with $\frac{dF(e^{it})}{dt}$.

We refer to Axler, Bourdon and Ramey [4] for good setting of harmonic mappings.

2000 *Mathematics Subject Classification.* Primary 58E20.
Key words and phrases. Harmonic mappings, Harmonic quasiconformal mappings, Strictly convex domains, curvature.

1.0.2. *Quasiconformal mappings.* A sense-preserving injective harmonic mapping $w = u + iv$ is called K-*quasiconformal* (K-q.c), $K \geq 1$, if

$$|w_{\bar z}| \leq k|w_z| \tag{1.2}$$

on $\mathbf{U}$ where $k = (K-1)/(K+1)$. Notice that, since

$$|\nabla w(z)| := \max\{|\nabla w(z)h| : |h| = 1\} = |w_z(z)| + |w_{\bar z}(z)|,$$

and

$$l(\nabla w(z)) := \min\{|\nabla w(z)h| : |h| = 1\} = \| w_z(z)| - |w_{\bar z}(z)\|.$$

The condition (1.2) is equivalent with

$$|\nabla w(z)| \leq K l(\nabla w(z)). \tag{1.3}$$

For a general definition of quasiregular mappings and quasiconformal mappings we refer to the book of Ahlfors [1].

For a background on the topic of quasiconformal harmonic mappings we refer [5], [8]-[22], [23], [26], [27]. In this paper we obtain some new results concerning a characterization of this class. We will restrict ourselves to the class of q.c. harmonic mappings w between the unit disk $\mathbf{U}$ and a convex Jordan domain D. The unit disk is taken because of simplicity. Namely, if $w : \Omega \to D$ is q.c. harmonic, and $a : \mathbf{U} \to \Omega$ is conformal, then $w \circ a$, is also q.c. harmonic. However the image domain D cannot be replaced by the unit disk.

To state the main result of the paper, we make use of Hilbert transforms formalism. It provides a necessary and a sufficient condition for the harmonic extension of a homeomorphism from the unit circle to a smooth convex Jordan curve γ to be a q.c mapping. It is an extension of the corresponding result [11, Theorem 3.1] related to convex Jordan domains. The Hilbert transformation of a function $\chi \in \mathrm{L}^1(\mathbf{T})$ is defined by the formula

$$\tilde\chi(\tau) = H[\chi](\tau) = -\frac{1}{\pi}\int_{0+}^{\pi} \frac{\chi(\tau+t) - \chi(\tau-t)}{2\tan(t/2)}\mathrm{d}t. \tag{1.4}$$

Here $\int_{0+}^{\pi}\Phi(t)dt := \lim_{\epsilon\to 0^+}\int_{\epsilon}^{\pi}\Phi(t)dt$. This integral is improper and converges for a.e. $\tau \in [0, 2\pi]$; this and other facts concerning the operator H used in this paper can be found in the book of Zygmund [31, Chapter VII]. If $f = u + iv$ is a harmonic function defined in the unit disk then a harmonic function $\tilde f = \tilde u + i\tilde v$ is called the harmonic conjugate of f if $u + i\tilde u$ and $v + i\tilde v$ are analytic functions and $\tilde u(0) = \tilde v(0) = 0$. Let $\chi, \tilde\chi \in L^1(\mathbf{T})$. Then

$$P[\tilde\chi] = \widetilde{P[\chi]}, \tag{1.5}$$

where $\tilde k(z)$ is the harmonic conjugate of $k(z)$ (see e.g. [28, Theorem 6.1.3]).

Let D be a strictly convex domain with C^2 Jordan boundary γ. By κ_z we denote the curvature of γ at $z \in \gamma$. We now state a theorem that concerns with quasiconformal harmonic mappings between the unit disk and strictly convex domains.

Theorem 1.1. (I) *Let γ be a $C^{1,\alpha}$ convex Jordan curve and let F be an arbitrary absolutely continuous parametrization.*

Then $w = P[F]$ is a quasiconformal mapping if and only if

$$0 < m = \operatorname{ess}\inf_{\tau} |F'(\tau)|, \tag{1.6}$$

$$M = \|F'\|_\infty := \operatorname{ess}\sup_{\tau} |F'(\tau)| < \infty \tag{1.7}$$

and

$$H = \|H(F')\|_\infty := \operatorname{ess\,sup}_\tau |H(F')(\tau)| < \infty. \tag{1.8}$$

(II) *Let γ be a C^2 convex Jordan curve and κ_z be the curvature of γ at $z \in \gamma$. Further let $\kappa_0 = \min_{z\in\gamma} \kappa_z$ and $\kappa_1 = \max_{z\in\gamma} \kappa_z$. If F satisfies the conditions (1.6), (1.7) and (1.8), and γ is strictly convex, then $w = P[F]$ is K quasiconformal, where*

$$K \le \frac{\kappa_1(M^2 + H^2) + \sqrt{(\kappa_1(M^2 + H^2))^2 - (2\kappa_0^2 m^3)^2}}{2\kappa_0^2 m^3}. \tag{1.9}$$

The constant K is the best possible in the following sense, if w is the identity or it is a mapping close to the identity, then $K = 1$ or K close to 1 (respectively).

2. Preliminaries

Suppose γ is a rectifiable, directed, differentiable curve given by its arc-length parametrization $g(s)$, $0 \le s \le l$, where $l = |\gamma|$ is the length of γ. Then $|g'(s)| = 1$ and $s = \int_0^s |g'(t)|dt$, for all $s \in [0, l]$. We say that $\gamma \in C^{1,\alpha}$ if $g \in C^{1,\alpha}$.

If γ is a twice-differentiable curve, then the curvature of γ at a point $p = g(s)$ is given by $\kappa_\gamma(p) = |g''(s)|$. Let

$$K(s,t) = \mathrm{Re}\,[\overline{(g(t) - g(s))} \cdot ig'(s)] \tag{2.1}$$

be a function defined on $[0, l] \times [0, l]$. By $K(s \pm l, t \pm l) = K(s,t)$ we extend it on $\mathbb{R} \times \mathbb{R}$. Note that $ig'(s)$ is the unit normal vector of γ at $g(s)$ and therefore, if γ is convex then

$$K(s,t) \ge 0 \text{ for every } s \text{ and } t. \tag{2.2}$$

Suppose now that $F : \mathbb{R} \mapsto \gamma$ is an arbitrary 2π periodic Lipschitz function such that $F|_{[0,2\pi)} : [0, 2\pi) \mapsto \gamma$ is an orientation preserving bijective function.

Then there exists an increasing continuous function $f : [0, 2\pi] \mapsto [0, l]$ such that

$$F(\tau) = g(f(\tau)). \tag{2.3}$$

In the remainder of this paper we will identify $[0, 2\pi)$ with the unit circle S^1, and $F(s)$ with $F(e^{is})$. In view of the previous convention we have

$$F'(\tau) = g'(f(\tau)) \cdot f'(\tau),$$

and therefore

$$|F'(\tau)| = |g'(f(\tau))| \cdot |f'(\tau)| = f'(\tau).$$

Along with the function K we will also consider the function K_F defined by

$$K_F(t,\tau) = \mathrm{Re}\,[\overline{(F(t) - F(\tau))} \cdot iF'(\tau)].$$

It is easy to see that

$$K_F(t,\tau) = f'(\tau) K(f(t), f(\tau)). \tag{2.4}$$

Lemma 2.1. [12] *If $w = P[F]$ is a harmonic mapping, such that F is a Lipschitz homeomorphism from the unit circle onto a Jordan curve of the class $C^{1,\alpha}$ ($0 < \alpha < 1$), then for almost every $\tau \in [0, 2\pi]$ there exists*

$$J_w(e^{i\tau}) := \lim_{r\to 1^-} J_w(re^{i\tau})$$

and there hold the formula

$$J_w(e^{i\tau}) = f'(\tau)\int_0^{2\pi} \frac{\mathrm{Re}\,[\overline{(g(f(t)) - g(f(\tau)))}\cdot ig'(f(\tau))]}{2\sin^2\frac{t-\tau}{2}}dt. \tag{2.5}$$

Lemma 2.2. *If $\varphi : \mathbf{R} \to \mathbf{R}$ is a $(\ell,\mathcal{L})$ bi-Lipschitz mapping, such that $\varphi(x+a) = \varphi(x) + b$ for some a and b and every x, then there exists a sequence of $(\ell,\mathcal{L})$ bi-Lipschitz diffeomorphisms (respectively a sequence of diffeomorphisms) $\varphi_n : \mathbf{R} \to \mathbf{R}$ such that φ_n converges uniformly to φ, and $\varphi_n(x+a) = \varphi_n(x) + b$.*

Proof. We introduce appropriate mollifiers: Fix a smooth function $\rho : \mathbb{R} \to [0,1]$ which is compactly supported in the interval $(-1,1)$ and satisfies $\int_{\mathbb{R}} \rho = 1$. For $\varepsilon = 1/n$ consider the mollifier

$$\rho_\varepsilon(t) := \frac{1}{\varepsilon}\,\rho\left(\frac{t}{\varepsilon}\right). \tag{2.6}$$

It is compactly supported in the interval $(-\varepsilon,\varepsilon)$ and satisfies $\int_{\mathbb{R}} \rho_\varepsilon = 1$. Define

$$\varphi_\varepsilon(x) = \varphi * \rho_\varepsilon = \int_{\mathbf{R}} \varphi(y)\frac{1}{\varepsilon}\rho(\frac{x-y}{\varepsilon})dy = \int_{\mathbf{R}} \varphi(x-\varepsilon z)\rho(z)dz,$$

then

$$\varphi'_\varepsilon(x) = \int_{\mathbf{R}} \varphi'(x-\varepsilon z)\rho(z)dz.$$

It follows that

$$\ell\int_{\mathbf{R}} \rho(z)dz = \ell \le |\varphi'_\varepsilon(x)| \le \mathcal{L}\int_{\mathbf{R}} \rho(z)dz = \mathcal{L}.$$

The fact that $\varphi_\varepsilon(x)$ converges uniformly to φ follows by Arzela-Ascoli theorem. $\square$

Lemma 2.3. *For every bi-Lipschitz mapping $\phi : [0,\pi] \to [0,\pi]$, $\phi'(0) = \phi'(\pi)$ we have*

$$\operatorname{ess\,inf}(\phi'(x))^2 \le \frac{\sin^2\phi(x)}{\sin^2 x} \le \operatorname{ess\,sup}(\phi'(x))^2.$$

Proof. Assume first that, ϕ is a diffeomorphism such that $\phi'(0) = \phi'(\pi)$. Let

$$h(x) = \frac{\sin\phi(x)}{\sin x}.$$

Then h is differentiable in $[0,\pi]$. The stationary points of h satisfy the equation

$$\phi'\frac{\cos\phi(x)}{\sin x} - \frac{\cos x}{\sin x}h = 0.$$

Therefore

$$h^2(x) = (\phi'(x))^2\cos^2\phi(x) + \sin^2\phi(x).$$

Since

$$\phi(2\pi) - \phi(0) = \int_0^{2\pi} \phi'(x)dx,$$

we have that $\min_x(\phi'(x)) \le 1 \le \max_x(\phi'(x))$. It follows that

$$\min_x(\phi'(x))^2 \le h^2(x) \le \max_x(\phi'(x))^2.$$

The general case follows from Lemma 2.2. $\square$

3. The proof of Theorem 1.1

We begin by the following lemma

Lemma 3.1. *Let γ be a C^2 strictly convex Jordan curve and let F be an arbitrary parametrization. Let $m = \min_{\tau\in[0,2\pi]} |F'(\tau)|$ and $M = \max_{\tau\in[0,2\pi]} |F'(\tau)|$. Then we have the following double inequalities:*

$$\frac{\kappa_0^2}{\kappa_1} \le \frac{K(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \frac{\kappa_1^2}{\kappa_0}, \tag{3.1}$$

and

$$\frac{\kappa_0^2}{\kappa_1} m^3 \le \frac{K_F(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \frac{\kappa_1^2}{\kappa_0} M^3, \tag{3.2}$$

where K and K_F are defined in (2.1) and (2.4). If γ is in addition a symmetric Jordan curve then we have the better estimates

$$\kappa_0 \le \frac{K(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \kappa_1, \tag{3.3}$$

and

$$\kappa_0 m^3 \le \frac{K_F(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \kappa_1 M^3. \tag{3.4}$$

Proof. Let $\tilde{g}$ be a arch length parametrization function of the curve $\tilde{\gamma} = \frac{1}{|\gamma|}\gamma$, where $|\gamma|$ is the length of γ. Let $\tilde{\kappa}_0 = \min_{z\in\tilde{\gamma}} \tilde{\kappa}_z$ and $\tilde{\kappa}_1 = \max_{z\in\tilde{\gamma}} \tilde{\kappa}_z$, where $\tilde{\kappa}_z$ is the curvature of $\tilde{\gamma}$ at z. It is clear that

$$|\gamma|\kappa_{|\gamma|z} = \tilde{\kappa}_z. \tag{3.5}$$

Let

$$G(\sigma,\varsigma) := \frac{\langle \tilde{g}(\sigma) - \tilde{g}(\varsigma), i\tilde{g}'(\varsigma)\rangle}{2\sin^2\frac{\sigma-\varsigma}{2}}.$$

Since $\tilde{g}'(\varsigma)$ is a unit vector and γ is a C^2 strictly convex curve, there exists a diffeomorphism $\beta : \mathbb{R}\to\mathbb{R}$, $\beta(0) = 0$, $\beta(2\pi+\sigma) = 2\pi + \beta(\sigma)$ such that

$$\tilde{g}'(\sigma) = e^{i\beta(\sigma)}. \tag{3.6}$$

Therefore

$$G(\sigma,\varsigma) = \frac{\int_\varsigma^\sigma \sin(\beta(\tau)-\beta(\varsigma))d\tau}{2\sin^2\frac{\sigma-\varsigma}{2}}. \tag{3.7}$$

On the other hand from

$$\tilde{g}''(\tau) = i\beta'(\tau)e^{i\beta(\tau)}$$

it follows that

$$\kappa_{\tilde{g}(\tau)} = \beta'(\tau). \tag{3.8}$$

According to (3.6), we obtain first that

$$\int_0^{2\pi} e^{i\beta(\sigma)}d\sigma = \tilde{g}(0) - \tilde{g}(2\pi) = 0. \tag{3.9}$$

Thus

$$\int_0^{2\pi} \sin(\beta(\sigma))d\sigma = \int_0^{2\pi} \cos(\beta(\sigma))d\sigma = 0. \tag{3.10}$$

Therefore

$$\int_\varsigma^\sigma \sin(\beta(\tau) - \beta(\varsigma))d\tau = \int_{[0,2\pi]\setminus[\varsigma,\sigma]} \sin(\beta(\varsigma) - \beta(\tau))d\tau.$$

As β is a diffeomorphism it follows that at least one of the following relations hold

$$\sin(\beta(\tau) - \beta(\varsigma)) \geq 0 \text{ for } \tau \in [\varsigma, \sigma] \tag{3.11}$$

or

$$\sin(\beta(\varsigma) - \beta(\tau)) \geq 0 \text{ for } \tau \in [0, 2\pi] \setminus [\varsigma, \sigma]. \tag{3.12}$$

Introducing the change $a = \beta(\tau)$ we obtain in the case (3.11) that

$$\begin{aligned} \int_\varsigma^\sigma \sin(\beta(\tau) - \beta(\varsigma))d\tau &= \int_{\beta(\varsigma)}^{\beta(\sigma)} \sin(a - \beta(\varsigma))\frac{da}{\beta'(\tau)} \\ &\geq (\leq)\frac{1}{\max_\tau(\min_\tau)\beta'(\tau)} \int_{\beta(\varsigma)}^{\beta(\sigma)} \sin(a - \beta(\varsigma))da \\ &= \frac{2}{\max_\tau(\min_\tau)\beta'(\tau)} \sin^2(\frac{\beta(\sigma) - \beta(\varsigma)}{2}). \end{aligned} \tag{3.13}$$

Therefore

$$\frac{1}{\max_\tau \beta'(\tau)} \frac{\sin^2(\frac{\beta(\sigma)-\beta(\varsigma)}{2})}{\sin^2 \frac{\sigma-\varsigma}{2}} \leq G(\sigma, \varsigma) \leq \frac{1}{\min_\tau \beta'(\tau)} \frac{\sin^2(\frac{\beta(\sigma)-\beta(\varsigma)}{2})}{\sin^2 \frac{\sigma-\varsigma}{2}}. \tag{3.14}$$

The case (3.12) can be consider similarly. In this case we apply the fact that $\beta(2\pi + \sigma) = 2\pi + \beta(\sigma)$ and in the same way obtain (3.14).

By taking $u = \frac{\sigma-\varsigma}{2}$ and $\phi(u) = \frac{\beta(2u+\varsigma)-\beta(\varsigma)}{2}$, and using Lemma 2.3 we obtain that

$$\frac{(\min_\tau \beta'(\tau))^2}{\max_\tau \beta'(\tau)} \leq G(\sigma, \varsigma) \leq \frac{(\max_\tau \beta'(\tau))^2}{\min_\tau \beta'(\tau)}. \tag{3.15}$$

From (3.15) we obtain

$$\frac{\tilde{\kappa}_0^2}{\tilde{\kappa}_1} \leq G(\sigma, \varsigma) \leq \frac{\tilde{\kappa}_1^2}{\tilde{\kappa}_0}. \tag{3.16}$$

On the other hand there exists a diffeomorphism $\sigma : [0, 2\pi] \to [0, 2\pi]$ such that

$$F(\tau) = |\gamma|\tilde{g}(\sigma(\tau)).$$

Thus

$$F'(\tau) = |\gamma|\sigma'(\tau)g'(\sigma(\tau)) \tag{3.17}$$

and

$$|F'(\tau)| = |\gamma|\sigma'(\tau). \tag{3.18}$$

Thus

$$K_F(t,\tau) = \left\langle \overline{F(t)-F(\tau)}, iF'(\tau)\right\rangle \tag{3.19}$$
$$= |\gamma|^2\sigma'(\tau)\left\langle \overline{\tilde g(\sigma(\tau)) - \tilde g(\sigma(t))}, i\tilde g'(\sigma(\tau))\right\rangle$$
$$= |\gamma|^2\sigma'(\tau)G(\sigma(t),\sigma(\tau))\cdot 2\sin^2\frac{\sigma(\tau)-\sigma(t)}{2}.$$

By applying again Lemma 2.3 we obtain

$$\min_t(\sigma'(t))^2 \le \frac{2\sin^2\frac{\sigma(\tau)-\sigma(t)}{2}}{2\sin^2\frac{\tau-t}{2}} \le \max_t(\sigma'(t))^2. \tag{3.20}$$

Combining (3.16), (3.19) and (3.20) we obtain

$$\min_t(\sigma'(t))^2\frac{|\gamma|^2\sigma'(t)\tilde\kappa_0^2}{\tilde\kappa_1} \le \frac{K_F(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \max_t(\sigma'(t))^2\frac{|\gamma|^2\sigma'(t)\tilde\kappa_1^2}{\tilde\kappa_0}. \tag{3.21}$$

Combining (3.21), (3.5) and (3.18) we obtain

$$\frac{\kappa_0^2 m^3}{\kappa_1} \le \frac{K_F(t,\tau)}{2\sin^2\frac{\tau-t}{2}} \le \frac{\kappa_1^2 M^3}{\kappa_0}.$$

This yields (3.2). In particular, if $F=g$, where g is natural parametrization of γ we obtain (3.1). In order to prove the statement for symmetric domain, we differentiate (3.7). Then we have

$$G_\sigma(\sigma,\varsigma) = \frac{\sin(\beta(\sigma)-\beta(\varsigma))}{2\sin^2\frac{\sigma-\varsigma}{2}} - \frac{\int_\varsigma^\sigma \sin(\beta(\tau)-\beta(\varsigma))d\tau}{2\sin^2\frac{\sigma-\varsigma}{2}}\cdot\cot\frac{\sigma-\varsigma}{2}. \tag{3.22}$$

So $G_\sigma(\tilde\sigma,\tilde\varsigma)=0$ if and only if

$$G(\tilde\sigma,\tilde\varsigma) = \frac{\sin(\beta(\tilde\sigma)-\beta(\tilde\varsigma))}{\sin(\tilde\sigma-\tilde\varsigma)}.$$

Define the function

$$H(\sigma,\varsigma) = \frac{\sin(\beta(\sigma)-\beta(\varsigma))}{\sin(\sigma-\varsigma)}, 0<|\sigma-\varsigma|\neq\pi.$$

Then it can be extended in $[0,2\pi]\times[0,2\pi]$ because of symmetry of γ. Namely if $\sigma-\varsigma=\pi$, we have $\beta(\sigma)-\beta(\varsigma)=\pi$. Thus by L'Hopital's rule we have $H(\sigma,\sigma+\pi)=\beta'(\sigma)=H(\sigma,\sigma)$. By putting $x=\sigma-\varsigma\in[0,\pi]$ and $\phi(x)=\beta(x+\varsigma)-\beta(\varsigma)$ and applying Lemma (2.3), instead of (3.16) we obtain

$$\tilde\kappa_0 \le H(\sigma,\varsigma) \le \tilde\kappa_1, \tag{3.23}$$

and consequently

$$\tilde\kappa_0 \le G(\sigma,\varsigma) \le \tilde\kappa_1. \tag{3.24}$$

By repeating the previous proof we obtain (3.3) and (3.4). □

From Lemma 3.1 it follows at once the following theorem.

Theorem 3.2. *If $w=P[F]$ is a harmonic diffeomorphism of the unit disk onto a (symmetric) convex Jordan domain $D=\mathrm{int}\gamma\in C^2$, such that F is (m,M) bi-Lipschitz, then*

$$\left(\kappa_0 m^3 \le J_w(e^{i\tau}) \le \kappa_1 M^3\right), \frac{\kappa_0^2 m^3}{\kappa_1} \le J_w(e^{i\tau}) \le \frac{\kappa_1^2 M^3}{\kappa_0}. \tag{3.25}$$

Proof. From (2.5) we obtain

$$J_w(e^{i\tau}) = \int_0^{2\pi} \frac{K_F(t,\tau)}{2\sin^2\frac{\tau-t}{2}}\frac{dt}{2\pi}. \tag{3.26}$$

From (3.2) and (3.4) we obtain (3.25). □

Proof of Theorem 1.1. The part (I) of this theorem coincides with [11, Theorem 3.1]. Prove the part (II). We have to prove that under the conditions (1.6), (1.7) and (1.8) w is $K-$ quasiconformal, where K is given by (1.9). This means that, according to (1.3), we need to prove that the function

$$K(z) = \frac{|w_z|+|w_{\bar z}|}{|w_z|-|w_{\bar z}|} = \frac{1+|\mu|}{1-|\mu|} \tag{3.27}$$

is bounded by K.

It follows from (1.1) that w_φ is equals to the Poisson-Stieltjes integral of F':

$$w_\varphi(re^{i\tau}) = \frac{1}{2\pi}\int_0^{2\pi} P(r,\tau-t)dF(t).$$

Hence, by Fatou's theorem, the radial limits of F_τ exist almost everywhere and $\lim_{r\to 1-} F_\tau(re^{i\tau}) = F_0'(\tau)$ a.e., where F_0 is the absolutely continuous part of F.

As rw_r is harmonic conjugate of w_τ, it turns out that if F is absolutely continuous, then

$$\lim_{r\to 1-} F_r(re^{i\tau}) = H(F')(\tau)\ (a.e.),$$

and

$$\lim_{r\to 1-} F_\varphi(re^{i\tau}) = F'(\tau).$$

As

$$|w_z|^2+|w_{\bar z}|^2 = \frac{1}{2}\left(|w_r|^2 + \frac{|w_\varphi|^2}{r^2}\right)$$

it follows that

$$\lim_{r\to 1^-}\left(|w_z|^2+|w_{\bar z}|^2\right) \le \frac{1}{2}(\|F'\|^2_\infty + \|H(F')\|^2_\infty). \tag{3.28}$$

On the other hand, by (3.25)

$$\lim_{r\to 1^-}\left(|w_z|^2-|w_{\bar z}|^2\right) \ge \frac{\kappa_0^2 m^3}{\kappa_1}. \tag{3.29}$$

From (3.28) and (3.29) we obtain

$$\lim_{r\to 1-}\frac{|w_z|^2+|w_{\bar z}|^2}{|w_z|^2-|w_{\bar z}|^2} \le C := \frac{\kappa_1(\|F'\|^2_\infty + \|H(F')\|^2_\infty)}{2\kappa_0^2 m^3}, \tag{3.30}$$

i.e.

$$\lim_{r\to 1-}\frac{|w_{\bar z}|}{|w_z|} \le \sqrt{\frac{C-1}{C+1}}. \tag{3.31}$$

By Lewy' theorem, $\frac{|w_{\bar z}|}{|w_z|}$ is a subharmonic function bounded by 1. From (3.31) it follows that

$$\frac{|w_{\bar z}|}{|w_z|} \le \sqrt{\frac{C-1}{C+1}}.$$

Further

$$K = \frac{\sqrt{C+1}+\sqrt{C-1}}{\sqrt{C+1}-\sqrt{C-1}} = C+\sqrt{C^2-1}$$
$$= \frac{\kappa_1(\|F'\|_\infty^2+\|H(F')\|_\infty^2)+\sqrt{(\kappa_1(\|F'\|_\infty^2+\|H(F')\|_\infty^2))^2-(2\kappa_0^2m^3)^2}}{2\kappa_0^2m^3}.$$

The last quantity is equal to 1 for F being identity because all the constants appearing at the quantity are 1 in this special case. Moreover, if F is close to identity in C^2 norm, then the quantity is close to 1. □

Remark 3.3. For symmetric domains, in view of Theorem 3.2, instead of (1.9) we can obtain the following estimate

$$K \leq \frac{\|F'\|_\infty^2+\|H(F')\|_\infty^2+\sqrt{(\|F'\|_\infty^2+\|H(F')\|_\infty^2)^2-(2\kappa_0m^3)^2}}{2\kappa_0m^3}.$$

Example 3.4. If F is the arc-parametrization of a C^2 convex Jordan curve γ, then $m = \|F'\|_\infty = 1$. We assume w.l.g. that the length of γ is 2π. Furthermore since $F'(s) = e^{i\beta(s)}$, by applying Lemma 2.3 again we obtain

$$\begin{aligned}|H[F'](\tau)| &= \left|-\frac{1}{\pi}\int_{0+}^{\pi}\frac{F'(\tau+t)-F'(\tau-t)}{2\tan(t/2)}\mathrm{d}t\right|\\ &\leq \frac{1}{\pi}\int_{0+}^{\pi}\frac{|e^{i\beta(\tau+t)}-e^{i\beta(\tau-t)}|}{2\tan(t/2)}\mathrm{d}t\\ &= \frac{1}{\pi}\int_{0+}^{\pi}\frac{2\left|\sin(\frac{\beta(\tau+t)-\beta(\tau-t)}{2})\right|}{2\tan(t/2)}\mathrm{d}t\\ &\leq \sup|F''(s)|\frac{1}{\pi}\int_0^{\pi}\frac{\sin t}{\tan(t/2)}dt = \kappa_1.\end{aligned}$$

So

$$K \leq \frac{\kappa_1(1+\kappa_1^2)+\sqrt{(\kappa_1(1+\kappa_1^2))^2-4\kappa_0^4}}{2\kappa_0^2}$$

and for symmetric domains

$$K \leq \frac{1+\kappa_1^2+\sqrt{(1+\kappa_1^2)^2-4\kappa_0^2}}{2\kappa_0}.$$

If γ is the unit circle, then $\kappa_0 = 1 = \kappa_1$. Both estimates are asymptotically sharp; if the curve γ approaches in C^2 topology to the unit circle centered at origin, then the quasiconformal constant tends to 1.

In particular if γ is the ellipse $\gamma = \{(x,y) : x^2/a^2+y^2/b^2 = 1\}$, $a \leq b$, $|\gamma| = 2\pi$, then $\kappa_0 = 1/b$ and $\kappa_1 = 1/a$.

References

[1] L. Ahlfors: *Lectures on Quasiconformal mappings,* Van Nostrand Mathematical Studies, D. Van Nostrand 1966.

[2] G. Alessandrini, V. Nesi: *Invertible harmonic mappings, beyond Kneser.* Ann. Scuola Norm. Sup. Pisa, Cl. Sci. (5) VIII (2009), 451-468.

[3] G. Alessandrini, V. Nesi: *Univalent σ-harmonic mappings*. (English summary) Arch. Ration. Mech. Anal. 158 (2001), no. 2, 155–171.

[4] Sheldon Axler, Paul Bourdon, Wade Ramey, *Harmonic function theory*, Springer Verlag New York 1992.

[5] W. Hengartner and G. Schober: *Harmonic mappings with given dilatation.* J. London Math. Soc. (2) 33 (1986), no. 3, 473–483.
[6] J. Garnett: *Bounded analytic functions.* Pure and Applied Mathematics, 96. Academic Press, Inc., New York-London, 1981. xvi+467 pp.
[7] J. Jost: *Univalency of harmonic mappings between surfaces.* J. Reine Angew. Math. 324 (1981), 141–153.
[8] D. Kalaj, M. Pavlovic: *On quasiconformal self-mappings of the unit disk satisfying the Poisson equation,* Trans. Amer. Math. Soc. 363 (2011) 4043–4061.
[9] D. Kalaj, M. Pavlovic: *Boundary correspondence under harmonic quasiconformal homeomorfisms of a half-plane.* Ann. Acad. Sci. Fenn., Math. 30:1 (2005), 159-165.
[10] D. Kalaj, M. Mateljević: *Inner estimate and quasiconformal harmonic maps between smooth domains.* Journal d'Analise Math. 100 (2006), 117-132.
[11] D. Kalaj: *Quasiconformal harmonic mapping between Jordan domains* Math. Z. 260:2 (2008), 237-252.
[12] ______: *Invertible harmonic mappings beyond the Kneser theorem and quasiconformal harmonic mappings.* Stud. Math. **207**, 117–136 (2011).
[13] ______: *Harmonic quasiconformal mappings and Lipschitz spaces.* Ann. Acad. Sci. Fenn., Math. 34:2 (2009), 475-485.
[14] ______: *On quasiregular mappings between smooth Jordan domains*, Journal of Mathematical Analysis and Applications, 362:1 (2010), 58-63.
[15] ______: *Quasiconformal harmonic functions between convex domains*, Publ. Inst. Math., Nouv. Ser. 76(90), 3-20 (2004).
[16] ______: *Harmonic mappings and distance function,* Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5) 10, No. 3, 669-681 (2011).
[17] M. Kneževic, M. Mateljević: *On the quasi-isometries of harmonic quasiconformal mappings.* Journal of Mathematical Analysis and Applications, 334: 1 (2007), 404-413.
[18] H. Kneser: *Lösung der Aufgabe 41*, Jahresber. Deutsch. Math.-Verein. 35 (1926) 123-124.
[19] O. Kellogg: *Harmonic functions and Green's integral.* Trans. Amer. Math. Soc. 13 (1912), 109-132.
[20] H. Lewy: *On the non-vanishing of the Jacobian in certain in one-to-one mappings.* Bull. Amer. Math. Soc. 42 (1936), 689-692.
[21] F. D. Lesley; S. E. Warschawski: *Boundary behavior of the Riemann mapping function of asymptotically conformal curves.* Math. Z. 179 (1982), 299-323.
[22] V. Manojlović: *Bi-lipshicity of quasiconformal harmonic mappings in the plane*, Filomat 23:1 (2009), 85–89.
[23] O. Martio: *On harmonic quasiconformal mappings*, Ann. Acad. Sci. Fenn., Ser. A I 425 (1968), 3-10.
[24] M. Mateljević and M. Vuorinen: *On harmonic quasiconformal quasi-isometries.* Journal of Inequalities and Applications,Volume 2010, Article ID 178732, 19 pages doi:10.1155/2010/1787;
[25] C. Pommerenke: *Univalent functions.* Vanderhoeck & Riprecht, 1975.
[26] D. Partyka and K. Sakan: *On bi-Lipschitz type inequalities for quasiconformal harmonic mappings.* Ann. Acad. Sci. Fenn. Math. Vol 32 (2007), 579-594.
[27] M. Pavlović: *Boundary correspondence under harmonic quasiconformal homeomorfisms of the unit disc.* Ann. Acad. Sci. Fenn., Vol 27 (2002) 365-372.
[28] M. Pavlović: *Introduction to function spaces on the disk.* 20. Matematički Institut SANU, Belgrade, 2004. vi+184 pp.
[29] S. E. Warschawski: *On Differentiability at the Boundary in Conformal Mapping.* Proceedings of the American Mathematical Society, 12:4 (1961), 614-620.
[30] A. Weil: *L' Intégration dans les groupes topologiques et ses applications,* Actualités Sci et. Ind. 869, 1145, Hermann et Cie, 1941 and 1951.
[31] A. Zygmund: *Trigonometric Series I.* Cambrige University Press, 1958.
[32] Peter L. Duren. *Theorey of H^p Spaces*. Academic Press New York 1970.

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 2, Pages 61–75
ISSN: 1930-1235; (**2013**)

ALGORITHMS FOR REAL NUMERICAL VARIETIES WITH APPLICATION TO PARAMETERIZING QUADRATIC SURFACE INTERSECTION CURVES

Barry H. Dayton
Department of Mathematics,
Northeastern Illinois University,
Chicago, IL 60625
Email: bhdayton@neiu.edu

Abstract. A numerical method for analyzing and parameterizing quadratic surface intersection curves (QSIC) is proposed. This method is based on numerical fractional linear transformations, numerical irreducible decomposition and a numerical version of a classical method for replacing an irreducible QSIC with a numerically birationally equivalent cubic plane curve. Ultimately each component of the QSIC is parameterized by a plane curve of the form $y = u(x)$ or $y^2 = u(x)$ where $u(x)$ is a numerical polynomial of degree 3 or less.

1. Introduction

Although possibly the simplest space curves, the study of QSIC has been the subject of recent research [8, 14, 15]. From the point of view of parameterizing QSIC L. Dupont, D. Lazard, S. Lazard and S. Petitjean [8] essentially solve the problem. Although their method requires starting with exact systems, takes 65 pages and involves looking at many cases the accompanying software [9] is extremely fast and accurate. The coefficients, which must be integers, can be quite large so can adequately approximate most numerical systems.

We describe a numerical based method. Although it will not achieve the black box and speed of [9] the method is straight forward and can be described using standard methods of numerical curves with the one step specific to QSIC described in the few pages of Section 5. Moreover this section simply reformulates a classical argument into numerics. Although, for ease of replication, examples are given exactly, the method immediately switches to an equivalent numerical system so examples could be given numerically. Unlike [8] which uses the projective line as a parameter space we use simple affine real plane curves of the form $y = u(x)$ or $y^2 = u(x)$ for a real numerical polynomial $u(x)$ of degree at most 3.

The tools used consist of (1) fractional linear transformations [1] given by a matrix based presentation, (2) Macaulay and Sylvester matrix based computations for decomposing numerical curves into irreducible components and finding equations

1991 *Mathematics Subject Classification.* Primary 14Q05; Secondary 65D18, 68U05.
Key words and phrases. Real Numerical Varieties; Quadratic Surface; Intersection Curves.

for images of curves under polynomial maps, and (3) methods involving numerical polynomial system solving to find real points on numerical algebraic varieties. For (3) these method are quite recent and this is the first exposition of these methods.

The method for identifying and parameterizing QSICs has the following steps:

(1) Do a (complex) numerical irreducible decomposition to identify algebraic components. If all components are of degree 1 and/or 2 go to Step 5.
(2) Find a random real nonsingular real point, if any, of the QSIC.
(3) Using the original system, find a numerical cubic plane curve birationally equivalent to a union of components of the QSIC (Main Theorem on QSIC) and the birational transformations.
(4) Separate the cubic into irreducible components, use 1) to check whether all components of original QSIC with real points are accounted for. Otherwise the missing line will come from Step 1.
(5) Transform each component via fractional linear transformations to parameter curves of form $y = u(x)$ or $y^2 = u(x)$.
(6) Analyze the rational parameterizations on the parameter curves to obtain practical parameterizations of the QSIC.

Note this last step is not covered in [9]. The only step specific to QSICs is step 3.

In §8 below we give a complete example using this method. The next three sections deal with general techniques. The phrase *algebraic set* will refer to the point set $\mathcal{X}$ in $\mathbb{R}^s$ of solutions of a system of real polynomials in s-variables. On the other hand *algebraic variety* will refer to an ideal $\mathcal{I}$ of $\mathbb{R}[x_1, \ldots, x_s]$ such that $\mathcal{X} = V(\mathcal{I})$.

2. Preliminaries

In this section we outline the general methods we will use.

2.1. H-bases and Duality Method.

For numerical work the equivalalent of a Gröbner Basis is an *H-basis* [12, 13], also known as a *Macaulay basis* [11, §4.2]. An H-basis of an affine ring, eg. ring of the form $A = \mathbb{C}[x_1, \ldots, x_s]/\mathcal{I}$, is a set $\{f_1, \ldots, f_n\} \subseteq \mathbb{C}[\mathbf{x}] = \mathbb{C}[x_1, \ldots, x_s]$ such that if $f \in \mathcal{I}$ then there exist $g_1, \ldots, g_n \in \mathbb{C}[\mathbf{x}]$ such that $f = g_1 f_1 + \ldots g_n f_n$ where for each i $\deg(g_i f_i) \leq \deg(f)$. Note that if $B = \{f_1, \ldots, f_n\}$ is a homogeneous basis of $\mathcal{I}$ or is a Gröbner basis with respect to a positive degree ordering of $\mathcal{I}$ then B is an H-basis.

A theory of local-global duality is outlined in [6], a more recent summary has been given in [7]. In particular there are two numerical algorithms that we will use extensively below

Algorithm 1: Given a basis B, not necessarily H-basis, for ideal $\mathcal{I} \subseteq \mathbb{R}[\mathbf{x}]$ and points $\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_k \in V(\mathbb{R}[\mathbf{x}]/\mathcal{I})$ an H-basis is returned for the variety $Y = V(\mathcal{J})$ which is the union of irreducible components of $V(\mathcal{I})$ which contain one or more of the points $\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_k$.

Algorithm 2: Given a real H-basis for the ideal of variety $\mathcal{X} = V(\mathcal{I})$ and an algebraic map $\phi = \{\phi_1, \ldots, \phi_s\} : \mathbb{R}^s \mapsto \mathbb{R}^p$, the $\phi_i \in \mathbb{R}[x_1, \ldots, x_p]$, a real H-Basis will be returned for the variety $V(\mathcal{J})$ which is the Zariski closure of $\phi(\mathcal{X})$.

It should be noted that these algorithms require the user to supply both a numerical tolerance and appropriate degree to assure an H-basis. If the user has his or her own favorite algorithms to do these calculations they may be substituted provided that they do work for ideals defined by floating point polynomials.

2.2. **Fractional Linear Transformations.** We will make use of fractional linear transformations [1], that is transformations of the form

$$\mathbf{x} = (x_1, \dots, x_s) \mapsto \left(\frac{\alpha_1(\mathbf{x})}{\delta(\mathbf{x})}, \dots, \frac{\alpha_s(\mathbf{x})}{\delta(\mathbf{x})} \right)$$

where the α_i and δ are linear functions in s variables, i.e. $\alpha_i = \alpha_{i,1}x_1 + \cdots + \alpha_{i,s}x_s + \alpha_{i,s+1}$. If A is the $(s+1) \times (s+1)$ matrix

$$A = \begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,s} & \alpha_{1,s+1} \\ \dots & \dots & \dots & \dots \\ \alpha_{s,1} & \dots & \alpha_{1,s} & \alpha_{s,s+1} \\ \delta_1 & \dots & \delta_s & \delta_{s+1} \end{bmatrix}$$

then this is the transformation

$$\mathbf{x} \mapsto A \begin{bmatrix} x_1 \\ \vdots \\ x_s \\ 1 \end{bmatrix} = (y_1, \dots, y_s, y_{s+1}) \mapsto (y_1, \dots, y_s)/y_{s+1} \tag{1}$$

In effect we are homogenizing, applying a linear projective transformation and then specializing again. Note that if we compose the fractional linear transformation given by A with the one given by B we get the fractional linear transformation given by BA. In particular if A is invertible then the fractional linear transformation given by A is birational.

To find the action of the fractional linear transformation associated with matrix A on a curve just follow the instructions in (1), homogenizing, transforming, dehomogenizing, using Algorithm 2.

3. Finding Real Points on Curves

The main algorithm below starts with finding a random real point on the curve. The first thing to try is to intersect the curve with random real hyperplane and check for real solutions. One may repeat several times if a real point has not been found. However, in general the real locus of an algebraic set can be quite small or even a finite set so this simple method may not produce a real point.

In the case of a plane $f(x, y) = 0$ curve a very efficient way to find real points is to look for real solutions of the system $\{f, \mathcal{J}(f, \ell)\}$ where $\ell = ax + by$ for random or chosen real numbers a, b not both zero and $\mathcal{J}(f, \ell)$ is the determinant of the Jacobian of $\{f, \ell\}$. This idea was motivated by [4]. The picture in Figure 1 shows how $\mathcal{J}(f, \ell)$ grabs the real locus of the curve. The points found include all those where the tangent line to the curve is parallel to the line ℓ as well as any singular points, including any isolated points of the curve. If the curve has an *oval*, which here means a non-empty bounded topological component, then at least one real point will be found. Further, letting $\ell = x$ or $\ell = y$ will find the x, y bounds of the oval.

In higher dimensions we can try the following based on [3]. Let $F = \{f_1, \dots, f_{s-1}\}$ define a curve in $\mathbb{R}^s$. Pick a non-zero linear form $k = k_1x_1 + \cdots + k_sx_s$ randomly or purposefully. Let $J(f_1, \dots, f_s, k)$ be the Jacobian matrix, B a $s \times s$ orthogonal

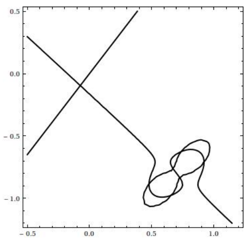


FIGURE 1. Real plane Curve (peanut), Jacobian Determinant curve and ℓ

matrix, and $C = [1, c_1, \dots c_{s-1}]$ a $1 \times s$ matrix where the c_i are new variables. Then $A = C \cdot B \cdot J$ is a $1 \times s$ matrix $A = [a_1, \dots, a_s]$, let $G = \{a_1, \dots, a_s, f_1, \dots, f_{s-1}\}$. Then G is a system of $2s-1$ equations in the $2s-1$ variables $x_1, \dots, x_s, c_1, \dots, c_{s-1}$. Find the real solutions for this square system G, and discard the last $s-1$ coordinates corresponding to the $c's$. If s is small and the system is exact one could use Gröbner bases with an elimination order to get an $s \times s$ system. The solution points in $\mathbb{R}^s$ should be the points where the curve intersects a hyperplane parallel to $V(k)$ singularly, in particular isolated points should be found. If the curve is an oval any k should be successful otherwise different k should be tried. Failure of this method may be caused by incomplete identification of the real solutions of G. Since this can be a common problem of solvers which find all complex solutions, what makes this method appealing is that we obtain a square system which is more likely to properly suggest real solutions than when trying to solve over or under-determined systems.

Example 1 (Example 4 of §9 below)**.** *Consider the QSIC $V(\{f, g\})$ given by*

$$f = x^2 + z^2 - 2y, \qquad g = -3x^2 + y^2 - z^2.$$

This space curve and the birationally equivalent plane curve h of Example 4 both have isolated real points. To find an isolated zero of h we choose, randomly,

$$l = 0.816353x - 0.273704y$$

and solve the system $\{\mathcal{J}(h, l), h\}$ getting 6 real solutions of which 2 are the multiple zero

$$(0.103102, -0.0989506)$$

indicating a singular solution. Further inspection reveals that this is an isolated zero.

For the space curve $V(\{f, g\})$ we pick a random real linear form

$$k = -0.668293x - 0.286001y - 0.214335z$$

and form $A = C\cdot B\cdot J$ where B is a random orthogonal matrix and J is the Jacobian matrix of the system $\{f, g, k\}$

$$A =$$
$$[1, c, d]\begin{bmatrix} -0.658953 & -0.557442 & -0.505014 \\ 0.70223 & -0.696511 & -0.147464 \\ -0.269545 & -0.451808 & 0.850421 \end{bmatrix}\begin{bmatrix} 2x & -2 & 2z \\ -6x & 2y & -2z \\ 0.816353 & -0.273704 & 0 \end{bmatrix}$$
$$= \big[-0.41227 - 0.120382c + 0.694244d + 2.02675x + 5.58353cx + 2.17175dx,$$
$$1.45613 - 1.3641c + 0.306327d - 1.11488y - 1.39302cy - 0.903615dy,$$
$$-0.20302z + 2.79748cz + 0.364524dz\big]$$

Solving the system consisting of the three polynomials above in x, y, z, c, d and f, g by Mathematica *and* Bertini *4 real solutions are identified by the two solvers which give essentially the same results. Two of these zeros essentially agree and give a multiple solution with $(x, y, z) = (0, 0, 0)$ indicating a singular point which is, in fact, isolated. The two other solutions for x, y, z are $(\pm 3.4641, 6.0000, 0)$, giving random real points on the 1 dimensional real component of the solution.*

Remark 1. *We should mention that discussion of this example with Daniel Lichtblau motivated the material in this section.*

4. Main Reduction

Here we give a constructive proof of a numerical simplification of a classically known fact: a QSIC is generically birationally equivalent to a degree 3 plane curve plus, perhaps, a line.

Before stating this formally we define a *real affine QSIC* to be a QSIC $V(\{f, g\})$ such that every complex projective component has a real positive dimensional affine solution set.

Theorem 1. *Let $\mathcal{C}$ be a real affine QSIC. There is a plane cubic h and rational maps $\Phi : V(h) \to \mathcal{C}$, $\Psi : \mathcal{C} \to V(h)$ such that $\Psi \circ \Phi = id_{V(h)}$. In particular $V(h)$ is birationally equivalent to a Zariski closed subset of $\mathcal{C}$.*

Proof. We prove this by explictly giving Φ and Ψ. The method used follows the main case of [5, §8, case (iv)].

So let f, g be quadratic functions in the three variables x, y, z, $\mathcal{C} = V(\{f, g\})$. Pick a random real solution $\hat{\mathbf{x}}$, as in §4, of the system $\{f = 0, g = 0\}$, such a solution exists by our assumption. Now homogenize by $f_h = t^2 * f(\frac{x}{t}, \frac{y}{t}, \frac{z}{t})$ and similarly for g_h. Although we are thinking "projective curve" we really will be working in affine x, y, z, t space.

Now construct a random orthogonal 4×4 matrix A satisfying $A\hat{\mathbf{x}} = [0, 0, 0, 1]^T$. Using Algorithm 2 of §2 get homogeneous equations $\{f_1, g_1\}$ for the QSIC which is the image of $\mathcal{C}$ under the linear transformation $[x, y, z, t]^T \mapsto A[x, y, z, t]^T$. Since $[0, 0, 0, 1]$ is a solution of this system the coefficient of t^2 is 0 for both f_1, g_1. Collect the terms involving t and write

$$\begin{aligned} f_1 &= tL + R \\ g_1 &= tM + S \end{aligned} \tag{2}$$

where L, M are linear in x, y, z and R, S are homogeneous quadrics in x, y, z. The linear polynomials L, M will be independent, and hence both non-zero, with probability 1; if this does not happen try a different random point and/or orthogonal matrix. Then $h_h = LS - RM$ is a hogeneneous cubic. Finally $h(x, y) = h_h(x, y, 1)$ is the desired plane cubic.

The forward map is

$$\Psi : \qquad [x, y, z] \mapsto A[x, y, z, 1]^T = [\check{x}, \check{y}, \check{z}, \check{t}\ ]^T \mapsto [\check{x}, \check{y}]/\check{z} \tag{3}$$

Whereas the backwards map makes use of (2) to recover t satisfying f_h, g_h from $[x, y, z]$:

$$\Phi : \quad [x, y] \mapsto A^{-1}\Big[x, y, 1, -\frac{R(x, y, 1)}{L(x, y, 1)}\Big]^T = [\tilde{x}, \tilde{y}, \tilde{z}, \tilde{t}\ ]^T \mapsto [\tilde{x}, \tilde{y}, \tilde{z}]/\tilde{t} \tag{4}$$

It is a straightforward check that, given our assumption, these maps do have the indicated domain and codomain with perhaps the exception of finitely many points.

To finish the proof we need to calculate $\Psi \circ \Phi$. If we plug the last expression in the definition of Φ, $[\tilde{x}, \tilde{y}, \tilde{z}\ ]/\tilde{t}$, in for the first expression $[x, y, z]$ in the definition of Ψ then the first term is

$$\frac{1}{\tilde{t}}[\tilde{x}, \tilde{y}, \tilde{z}, \tilde{t}\ ]^T = \frac{1}{\tilde{t}}A^{-1}\left[x, y, 1, \tilde{t}\right]$$

Now this gets multiplied by A giving

$$\frac{1}{\tilde{t}}[x, y, 1, t] = \left[\frac{x}{\tilde{t}}, \frac{y}{\tilde{t}}, \frac{1}{\tilde{t}}, \frac{t}{\tilde{t}}\ \right]$$

by definition of Φ. But dropping the last two coordinates and dividing by the third leaves us with just $[x, y]$.

□

We note that if our hypothesis of real affine QSIC does not hold the result is still true interpreted correctly in the complex projective situation. Even without this hypothesis the maps Φ, Ψ are still useful, however Φ may not be onto.

5. Parameterizing plane quadratics and cubics

Since a line is immediately parameterizable, our problem is reduced to finding parameter patches for a plane cubic. We first check whether the plane cubic given by $h = 0$ is reducible. Using Algorithm 2 above with just one point we can check for proper components and, since plane curves are solutions of single equations, components are associated with factors of the polynomial h. Any factors can be divided out making it easier to find any other factors. Thus we could still have to deal at this level with degree 1, 2 or 3 irreducible curves. Degree 1 is immediate. We will assume the irreducible curve is still called h.

Suppose $h = 0$ is the real equation of a plane curve of any degree and $\hat{\mathbf{p}}$ is a real non-singular point in $V(h)$. We will first use a fractional linear transformation to move $\hat{\mathbf{p}}$ to $[0, 1, 0]$ in $\mathbb{P}^2$ with the infinite line as tangent to the curve at this points. To this end we homogenize $h(x, y)$ to $h_h(x, y, z)$ and $\hat{\mathbf{p}}$ to $\hat{\mathbf{p}}_h$, in the latter case simply by adding a third coordinate 1. We define the normal vector at $\hat{\mathbf{p}}_h$ by

$$\mathbf{n} = \left[\frac{\partial h_h}{\partial x}, \frac{\partial h_h}{\partial y}, \frac{\partial h_h}{\partial z}\right]\Bigg|_{\hat{\mathbf{p}}}$$

It is well known, eg. [10, Chapter 4], that for a projective curve h_h at projective point $\hat{\mathbf{p}}_h$ then then $\mathbf{n} \cdot \hat{\mathbf{p}}_h = 0$.

Let $\mathbf{c} = \mathbf{n} \times \hat{\mathbf{p}}_h$ be the cross product and, simply for better numerical stability, let $\mathbf{n}, \hat{\mathbf{p}}_h, \mathbf{c}$ be normalized as $\bar{\mathbf{n}}, \overline{\hat{\mathbf{p}}_h}$ and $\bar{\mathbf{c}}$. Set B to be the 3×3 matrix

$$B = \begin{bmatrix} \bar{\mathbf{c}} \\ \overline{\hat{\mathbf{p}}_h} \\ \bar{\mathbf{n}} \end{bmatrix}$$

Then the fractional linear transformation given by B

$$\Theta : (x, y) \mapsto B[x, y, 1]^T = [\check{x}, \check{y}, \check{z}\ ]^T \mapsto [\check{x}, \check{y}\ ]/\check{z} \tag{5}$$

is our desired transformation. One should note that B is an orthogonal matrix.

Now if h is quadratic then by picking $\hat{\mathbf{p}}$ to be any point on $V(h)$ we arrive at a parabola $y = u(x)$ where $u(x) = ax^2 + bx + c$ is a quadratic in x since we have the unique infinite point $[0, 1, 0]$. In fact this is a real parabola even if h or $\hat{\mathbf{p}}$ are complex.

If h is an irreducible cubic curve, possibly singular, then we pick $\hat{\mathbf{p}}$ more carefully. In principle if h is singular we could pick $\hat{\mathbf{p}}$ to be the unique singular point. Now $\mathbf{n} = 0$ so we cannot use B above but any matrix B with $B.\hat{\mathbf{p}}_h = [0, 1, 0]$ would transform h into the form

$$y(dx + e) = ax^2 + bx + c \quad \text{ or } \quad y = \frac{ax^2 + bx + c}{dx + e}$$

giving a rational parameterization. In theory this should give satisfactory results but the hyperelliptic form below will behave much better numerically.

To get hyperelliptic form we take the Hessian curve H of h [10, §4.4] and find the intersections with h. As recently as 2000 when [15] was written this was computationally difficult, with modern numerical algebraic geometry, and even Mathematica's numerical Gröbner basis, this is now routine. If h is irreducible and non-singular we are guarenteed at least one real point on $V(h_h) \cap V(H)$. If h is irreducible and singular but generated by the random procedure of Theorem 1 then it is still quite likely that there be such a non-singular point. This point will have an inflectional tangent. Using the fractional linear transformation from B above we move this point to $[0, 1, 0]$ with inflectional tangent the line at infinity, call the resulting homogeneous curve (use Algorithm 4) j_h. In theory the coefficients of y^3 and xy^2 of j_h are easily seen to be 0, it is known [5] that the coefficient of x^2y will also vanish because of the inflectional tangent. In numerical practice these coefficients will be very tiny so we can discard these terms. Specializing by setting $z = 1$ and dividing by the coefficient of y^2 gives

$$j = ax^3 + bx^2 + cx + dxy + ey + y^2 = y^2 + (dx + e)y + v(x)$$

For fixed x setting $j = 0$ the quadratic equation in y on the right has two solutions which add to $-dx - e$. In other words the line $y = -(\frac{d}{2}x + \frac{e}{2})$ lies on the midpoints of the two solutions of $j = 0$ for given x, possibly complex. The fractional linear transformation given by matrix

$$C = \begin{bmatrix} 1 & 0 & 0 \\ \frac{d}{2} & 1 & \frac{e}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

sends this line to the x-axis giving an equation of the form $y^2 = u(x)$ for some cubic polynomial $u(x)$. As in (5) the map from the h to the parameter curve is

$$\Theta : (x, y) \mapsto CB[x, y, 1]^T = [\check{x}, \check{y}, \check{z}\ ]^T \mapsto [\check{x}, \check{y}\ ]/\check{z}$$

A few more steps, using classical techniques actualized as fractional linear transformations, will put $y^2 = u(x)$ in the form $y^2 = x^3 + ax + b$ for suitable real a, b. But this is not necessary for our next step so we will not go into details here.

6. Parameterization and analysis of real QSIC

From the previous two sections we can parameterize each irreducible algebraic component of the QSIC in one of the following ways where u is a polynomial function of t of specified degree and the result is a rational function $\mathbb{R}^1 \to \mathbb{R}^3$ where each coordinate has the same denominator. The maximum degrees of numerator and denominator are given.

degree u	parameterization	max degrees
1	$\Phi(\Theta^{-1}(t, u(t)))$	degree 3 in t
2	$\Phi(\Theta^{-1}(t, u(t)))$	degree 6 in t
3	$\Phi(\Theta^{-1}(t, \pm\sqrt{u(t)}))$	degree 3 in t and $\sqrt{u}$

For degrees 1 and 2 we need only find the poles, i.e. zeros of the common denominator. The QSIC is of degree 4 so unless a component lies in the infinite plane of $\mathbb{P}^3$ it can interesect this plane in at most 4 points so at most 4 of these poles will be essential counting possible poles at the ends of the parameter lines. Thus the real projective line is divided into at most 4 intervals which would give up to 4 affine topological components. As the parameter goes to $\pm\infty$ and poles, even the inessential ones, the parameterization may become numerically inaccurate, thus it may be useful to use different choices and create an additional parameterization or two that would overlap to cover finite points of the QSIC that could be misssed.

For degree 3 the situation is more complicated as for real QSIC the domain of the parameterizations is only part of the real line. Further because of the square root in the denominator it is harder to find the zeros. The first problem is easy as the domain is $\{t|u(t) \geq 0\}$ and it is only necessary to know the zeros of $u(t)$ to calculate this set. For the second collecting powers of $\sqrt{u(t)}$ one gets

$$\begin{aligned} v_0(t) + v_1(t)\sqrt{u(t)} + v_2(t)\sqrt{u(t)}^2 + v_3(t)\sqrt{u(t)}^3 \\ = v_0(t) + v_2(t)u(t) + \big(v_1(t) + v_3(t)u(t)\big)\sqrt{u(t)} \end{aligned}$$

Setting this last expression equal to 0 gives

$$v_0(t) + v_2(t)u(t) = -\big(v_1(t) + v_3(t)u(t)\big)\sqrt{u(t)}$$

Squaring both sides gives

$$\big(v_0(t) + v_2(t)u(t)\big)^2 - \big(v_1(t) + v_3(t)u(t)\big)^2 u(t) = 0 \tag{6}$$

a single variable polynomial equation which is easily solved numerically. The zeros of the denominator are contained in the zero set of this equation, and note that this same zero set will apply to both parameterizations $y = \sqrt{u(t)}$ and $y = -\sqrt{u(t)}$ but denominators will have different zeros of (6). Combining this result with the calculation of the domain allows us to identify the affine topological components.

One posibility not yet considered is that two algebraic components may intersect. This will be a singular point of the QSIC Alternatively the image of a singular point in the QSIC likely appeared as a singular point in $V(h)$. One can identify the component curves containing this point then use the appropriate Θ to the parameter curves. If this singular point did not show up in $V(h)$ it is because of a line of the QSIC not in the image of Φ, which would show up in Step 1.

7. A fully worked out Example

In this section we do a complete example using the method outlined in the introduction. This is one of the more complicated examples.

Consider the QSIC given by $\mathcal{C} = V(\{f, g\})$ where

$$f = x^2 + z^2 - 2z - y^2, \quad g = 2x^2 - 2xy - 2z - 3xz \tag{7}$$

Step 1: Do numerical irreducible decomposition to identify complex algebraic components.

I use Bertini [2] on the system above which finds two algebraic components with non-singular witness points (rounded to 6 significant digits for display but given by 16 significant digits).

$$\hat{\mathbf{p}}_1 = (-.07985493 + .259423\imath, -.07985493 + .259423\imath, 0)$$
$$\hat{\mathbf{p}}_2 = (-.0978825 + .161370\imath, -.0164075 + .2438801\imath, .0215337 - .0120534\imath)$$

Bertini calculates the degree of the component containing $\hat{\mathbf{p}}_1$ to be 1, thus a line, and the degree of the other component to be 3.

The remaining steps are done with Mathematica with default precison, generally 17 digits, and linear algebra tolerance 10^{-12}.

Step 2: For this unbounded curve we can intersect with a random real plane to get the random real point $\hat{\mathbf{p}} = (1.133057, -0.5986927, 0.7268414)$ again rounded to 7 significant digits but used with 17 digits in the calculations.

Step 3: We next apply Theorem 1 to the system (7) to obtain the plane cubic

$$\begin{aligned} h = & -0.277749 + 0.471423x + 0.0905586x^2 - 0.00582857x^3 + 0.129608y \\ & - 0.0407206xy + 0.0475988x^2y + 0.362871y^2 - 0.0760948xy^2 + 0.0966969y^3 \end{aligned}$$

Here

$$A = \begin{bmatrix} 0.655457 & 0.407462 & 0.160413 & -0.61532 \\ 0.0261099 & -0.845622 & -0.00374475 & -0.53313 \\ 0.405925 & -0.0764678 & -0.898681 & 0.147482 \\ 0.636333 & -0.33623 & 0.408199 & 0.561607 \end{bmatrix}$$

$$L = -0.511267x + 0.252642y + 0.229098z$$

$$R = -0.16528x^2 - 0.17456xy + 0.29100y^2 + 0.50610xz + 0.40022yz - 0.25201z^2$$

Thus from (3) the map $\Psi : \mathcal{C} \to V(h)$ is given by

$$\Psi(x, y, z) = \left(\frac{\alpha_1(x, y, z)}{\gamma(x, y, z)}, \frac{\alpha_2(x, y, z)}{\gamma(x, y, z)} \right)$$

where

$$\begin{aligned}\alpha_1(x,y,z) =& -0.61532 + 0.655457x + 0.407462y + 0.160413z\\ \alpha_2(x,y,z) =& -0.53313 + 0.0261099x - 0.845622y - 0.00374475z\\ \gamma(x,y,z) =& 0.147482 + 0.405925x - 0.0764678y - 0.898681z\end{aligned}$$

And from (4) the map $\Phi : V(h) \to \mathcal{C}$ is given by

$$\Phi(x,y) = \left(\frac{\beta_1(x,y)}{\delta(x,y)}, \frac{\beta_1(x,y)}{\delta(x,y)}, \frac{\beta_1(x,y)}{\delta(x,y)}\right)$$

where

$$\begin{aligned}\beta_1(x,y) =& 0.25336 - 0.37942x - 0.22994x^2 - 0.14614y + 0.26333xy - 0.17858y^2\\ \beta_2(x,y) =& -0.10225 + 0.30261x - 0.26390x^2 - 0.078482y + 0.47659xy - 0.11580y^2\\ \beta_3(x,y) =& -0.10301 + 0.28963x - 0.014546x^2 - 0.39127y + 0.11370xy - 0.11973y^2\\ \delta(x,y) =& 0.17532 - 0.50051x + 0.40742x^2 - 0.30965y + 0.2152xy - 0.29812y^2\end{aligned}$$

Step 4: We pick random linear form $k = 0.647639x + 0.134229y$ then the Jacobian Determinant is the ellipse

$$\mathcal{J} = -0.0206603 + 0.0506834x - 0.0331739x^2 - 0.475485y + 0.111342xy - 0.198088y^2$$

which intersects h in a singular point $(-4.17438, -3.61523)$ and points $q_1 = (0.537437, -0.00725009), q_2 = (-6.75208, -2.73808)$. Applying Algorithm 1 to q_1 decomposes $h = h_1h_2$ into the two curves

$$\begin{aligned}h_1 =& 0.407209 - 0.670328x - 0.167057x^2 - 0.326707y + 0.267999xy - 0.422342y^2\\ h_2 =& -0.682079 + 0.0348897x - 0.228954y\end{aligned}$$

Applying Ψ to the two points $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2$ we find that $\hat{\mathbf{p}}_1$ maps to a point on $V(h_2)$ while $\hat{\mathbf{p}}_2$ maps to a point on h_1. Thus Ψ respects the decompositions of $\mathcal{C}$ and $V(h)$ so we can conclude that Ψ is a numerical birational equivalence of curves $\mathcal{C}, V(h)$ with inverse Φ.

Step 5: Using the method of §6 we get a fractional linear transformation Θ with matrix

$$B = \begin{bmatrix} 0.545499 & 0.668587 & 0.505394\\ -0.832815 & 0.364724 & 0.416408\\ -0.0940751 & 0.64805 & -0.755766\end{bmatrix}$$

which takes $V(h_1)$ to the parabola

$$y = 0.145433 - 0.209493x - 0.322363x^2 = u(x)$$

Then our parameterization of the second component of $\mathcal{C}$ is

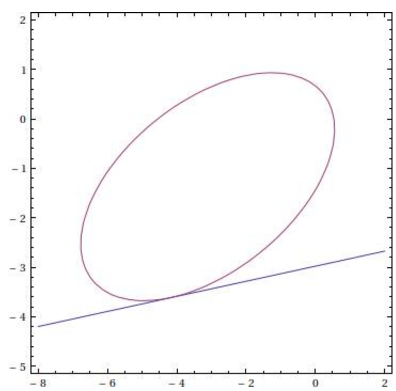
$$\Phi\Theta^{-1}((t, u(t)) = \left(\frac{\xi_1(t)}{\rho(t)}, \frac{\xi_2(t)}{\rho(t)}, \frac{\xi_3(t)}{\rho(t)}\right)$$

Where

$$\begin{aligned}\xi_1(t) =& -0.00123344 + 0.116394t + 0.0455282t^2 - 0.0600163t^3 - 0.0114211t^4\\ \xi_2(t) =& -0.106938 + 0.0555114t + 0.165079t^2 - 0.0241627t^3 - 0.0496511t^4\\ \xi_3(t) =& 0.107571 - 0.116773t - 0.0431373t^2 + 0.0859286t^3 - 0.0247614t^4\\ \rho(t) =& 0.000638119 - 0.06223t + 0.165178t^2 + 0.227341t^3 + 0.0347652t^4\end{aligned}$$

The first component of $\mathcal{C}$ is easily seen to be the line $y = x, z = 0$. The two components meet at the origin of $\mathbb{R}^3$ which is Φ of the singular point of $V(h)$ where the two components of $V(h)$ meet tangentially.

Step 6: Solving $\rho(t) = 0$ we get $\{-5.64066, -1.17225, 0.010554, 0.26302\}$, the root $t = -1.17225$ is not an essential pole but the others are. To graph the second component we need only plot the line and $\Phi\Theta^{-1}(t, u(t))$ on the real intervals $[-100, -5.67], [-5.6, .01], [.011, .26], [.27, 5]$. There is a tiny gap near the singular point $(0, 0, 0)$ but it is not very noticible on the graph (Figure 2).



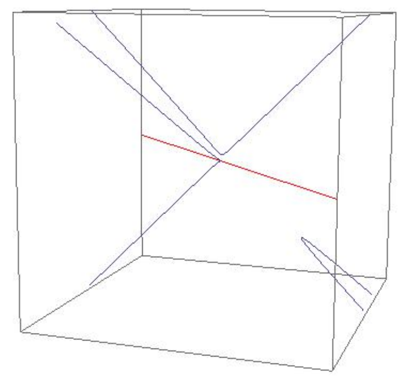


Figure 2. $V(h)$ (left) and the QSIC inside $-40 \leq x, y, z \leq 40$.

0ne further comment about this example is required. The algebraic component which is not a line is of degree 3 but not planar. Thus it cannot be a complete intersection curve in $\mathbb{C}^3$. In fact the curve given by

$$\gamma = 0.40756x - 0.649881x^2 - 0.40756y - 0.391031xy - 0.181769y^2 \\ - 0.143229z - 0.0770954xz + 0.181769z^2$$

is linearly independent of f, g but also contains this degree 3 component. Thus the decomposition of Step 1 is not helpful in finding a parameterization for this component. The decompositions of Step 1 and Step 4 serve different purposes, that of Step 1 identifies components to see if Ψ is onto while Step 4 decomposes the curve into useful complete intersection components. In this example both of these are necessary.

In [14] the classification of QSIC gives 23 of the 35 total types which are reducible with only planar components. In these 23 cases one could go directly from Step 1 to Step 5. In some of the other cases the QSIC is irreducible in which case the information from Step 1 allows us to skip Step 4. But for a complete description of all QSIC we need this 6 step method.

8. More Examples

Examples have been calculated with a default of approximately 17 digits in Mathematica 8 with a numerical linear algebra tolerance of 10^{-12}. Generally the calculations are accurate to about 11 digits but for display only about 6 digits are shown here. It should be noted that because a number of random choices are made in this approach that the details of these examples can not be replicated without knowing the choices. But the point set and properties of the QSIC obtained will be the same with different random choices.

Example 2. *Generically a QSIC will be a genus 1 space curve. A typical example is* $V(\langle f, g\rangle)$

$$f = x^2 + y^2 + z^2 - 16, \quad g = 57 - 12x + 4x^2 + y^2 - 64z + 16z^2 \tag{8}$$

An application of Theorem 4 gives

$$\begin{aligned} h = 0.0442427 + 0.140313x + 0.116615x^2 + 0.0294054x^3 \\ - 0.217169y - 0.402909xy - 0.172791x^2y + 0.344722y^2 \\ + 0.330695xy^2 - 0.208248y^3 \end{aligned}$$

Checking h *there are no singularities and three real inflectional tangents two of which, pictured below in the top lefthand plot in Figure 1, then give, using the method of §6, the two curves*

$$y^2 = 38.6067 + 16.0287x + 6.07442x^2 + 0.716794x^3$$
$$y^2 = -2.33478 + 1.87741x - 0.375185x^2 + 0.0245423x^3$$

both of which look like the upper right picture of Figure 3 but with very different scales. Note these curves are both birationally equivalent to the QSIC (8) *and so are birationally equivalent to each other. Each of these curves gives two patches (positive and negative* y*) the the four patches cover QSIC* (8)

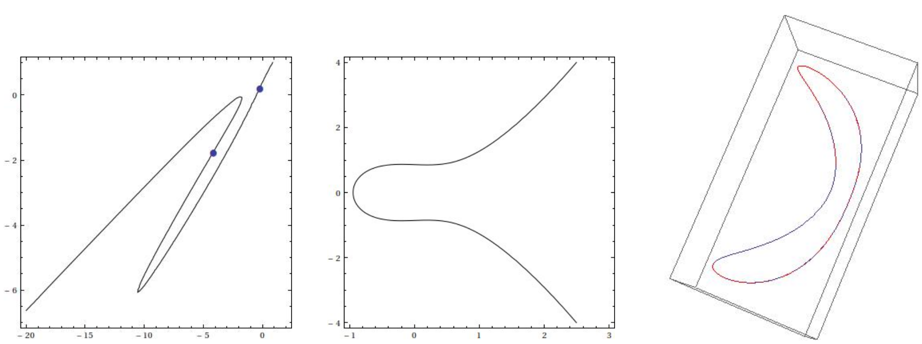


FIGURE 3. Example 2, bottom curve is QSIC

Example 3. *A more complicated version of the above is the curve* $V(\langle f, g\rangle)$

$$f = x^2 - y^2 + z^2 - 1 \quad g = x^2 - z^2 - 4 \tag{9}$$

The plane cubic obtained is

$$\begin{aligned} h = -0.0947177 - 0.263213x + 0.267739x^2 - 0.0855912x^3 \\ - 0.308632y + 0.0559627xy - 0.194258x^2y \\ - 0.122697y^2 + 0.387761xy^2 - 0.0316547y^3 \end{aligned}$$

which is transformed into

$$y^2 = -0.0029057 - 0.494465x + 0.252846x^2 + 0.45218x^3 = u(x)$$

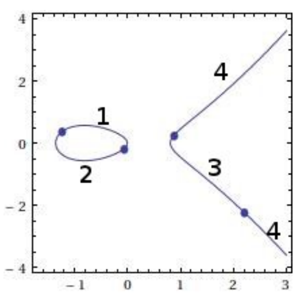
The parameter domain is then

$$\{-1.359 \leq t \leq 0.00586\} \cup \{0.8065 \leq t\}$$

and the possible zeros from (6) *of the parameter functions* $y = \pm\sqrt{u(t)}$ *are*

$$t = -1.2287, -.05820, 0.8936, 1.347, 20.925.$$

The values $t = -1.2287, .8936$ *are poles of* $y = \Phi\Theta^{-1}(t, \sqrt{u(t)})$ *while* $t = -.05197, 20.925$ *are poles of* $y = \Phi\Theta^{-1}(t, -\sqrt{u(t)})$ *but* $t = 1.3465$ *turns out not to be an essential pole of either. Thus our parameter curve breaks into 4 intervals as shown, dots are poles, on the left in Figure 3 below. The images under* $y = \Phi\Theta^{-1}(t, \pm\sqrt{u(t)})$ *of the intervals are the 4 affine topological components of the QSIC as shown on the right of Figure 4 after projection by* $(x, y, z) \mapsto (x + .5z, y)$*. Note we used Algorithm 4 to find the equation of the projection and plotted with a contour plot.*



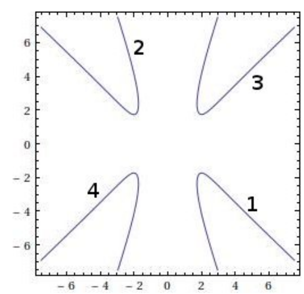


FIGURE 4. top parameter curve, bottom projection of QSIC in Example 3

Example 4. *Consider curve number 6 in* [14, Table 1] $V(\langle f, g\rangle)$ *where*

$$f = -x^2 - z^2 + 2y, \quad g = -3x^2 + y^2 - z^2 \tag{10}$$

This curve has a bounded dimension 1 real component and an isolated real point. Theorem 4 gives the plane cubic

$$\begin{aligned} h = 0.00491366 - 0.0192757x + 0.304379x^2 + 0.13571x^3 \\ + 0.0803394y + 0.449203xy + 0.11988x^2y \\ + 0.661842y^2 - 0.0941065xy^2 + 0.0385216y^3 \end{aligned}$$

This h *also has an isolated real point which is a singular point. Letting* h_h *be the homgenation of* h MATHEMATICA'S `NSolve`*, in a rare miss, does not see the common zero* $\hat{\mathbf{x}} = [0.103102, -0.0989506, 1]$ *of* $\{\frac{\partial h_h}{\partial x}, \frac{\partial h_h}{\partial y}, h_h\}$ *and also fails to identify* $\hat{\mathbf{x}}$ *as a multiple real zero of the intersection of* h_h *and its Hessian, seeing instead two close complex zeros. Thus it is not unexpected that one may mistake* h *for a non-singular cubic. Since there are real inflectional points the method of §4 gives a birational equvalence of* h *with* $y^2 = 0.498 + 2.55083x + 2.54258x^2 - 1.94738x^3$ *which is a singular curve with isolated real singular point at* $(-0.356012, 0)$*. This curve maps* *birationally onto the QSIC (10) including the isolated point and we get, a good parameter patch.*

With the method of §4 the isolated point of both the original QSIC and h are easily found, see the example in §4.

So h must be a singular curve but from the equation alone it is difficult in some cases to identify whether a numerical curve is singular. Thus this suggested method which does not treat singular and non-singular curves diffently is preferable to a method which treats them as separate cases. The singular isolated point is more readily identified in the final $y^2 = u(x)$ form.

The reader should notice the difference here between working numerically and exactly. In an exact calculation h could be a curve of the form $y - v(x)$ in which case it would make no sense transforming it to the cusp $y^2 - u(x)$ to get a parameterization. In our numerical case the chance of h being of the form $y - v(x)$ is virtually nil so it is not worth even considering that possibility.

Example 5. *The easiest example of a QSIC, the intersection of two spheres, is one not covered by Theorem 4 as stated, since it is not an affine QSIC. for instance*

$$f = x^2 + y^2 + z^2 - 4, \qquad g = -2 - 2x + x^2 + y^2 + z^2 \tag{11}$$

Note $f - g = 2x - 2$ so the affine solution is $\{x = 1, y^2 + z^2 = 3\}$. However applying the technique of Theorem 4 gave the cubic

$$\begin{aligned} h = 0.15022 + 0.10600x + 0.0724682x^2 - 0.00712392x^3 \\ + 0.44953y + 0.31137xy + 0.156986x^2y \\ + 0.403442y^2 + 0.165618xy^2 + 0.206231y^3 \end{aligned}$$

The curve $V(h)$ is reducible with components $V(h_1), V(h_2)$

$$h_1 = 0.448791 - 0.038654x + 0.8928y$$

$$\begin{aligned} h_2 = 0.334722 + 0.265018x + 0.1843x^2 + 0.335768y \\ +0.195505xy + 0.230994y^2 \end{aligned}$$

where Φ applied to the line $h_1 = 0$ parameterizes the real affine solution circle but the quadratic h_2 has no real solutions. However Φ does not take $V(h_2)$ to $V(\{f, g\})$. Instead if we work projectively then the projective closure of $V(h_2)$ goes to the complex ideal curve in $\mathbb{P}^3$ given by $\{t = 0, x^3 + y^3 + z^3 = 0\}$ which is in the projective closure of (11).

The system $\{x^2 - y^2, z^2 - 1\}$ [14, curve 28] consists of 4 real affine lines and does satisfy the hypotheses of Theorem 4 but it is impossible for Φ to be a birational equivalence because a cubic cannot have 4 components.

9. Conclusion

We have shown that working numerically instead of exactly greatly simplifies the problem and yet we are still able, in our experiments, to distinguish the different types of QSIC, even when there are singular points involved. It is possible, of course, that very sensitive examples may be found where our method may need higher precision or arithmetic or may fail.

Using the entire algorithm, especially Step 5, we can specialize Theorem 1 as follows:

Theorem 2. *Suppose $\mathcal{X}$ is an irreducibe real affine QSIC, possibly singular. Then $\mathcal{X}$ is numerically birationally equivalent to a plane cubic $\mathcal{Y}$ in hyperelliptic form $y^2 = x^3 + ax + b$ for real numbers a, b.*

Since Theorem 1 is the only place we have specifically addressed QSIC it is possible that this method can be extended to other space curves. It is not clear what parameterizable curves should be used as cononical models for plane curves of degree greater than 3. Fractional linear transformations may not be general enough in higher degrees. Nor is it clear how to generalize Theorem 1. These are all questions for further research.

References

[1] S.S. Abhyankar, *Algebraic Geometry for Scientists and Engineers*, Math. Surveys and Monographs Vol. 35, AMS, 1990.

[2] D.J. Bates, J.D. Hauenstein, A.J. Sommese, C. W. Wampler II, *Bertini: Software for Numerical Algebraic Geometry*, version 1.3.1 (32 bit Linux) January 2012, `www.nd.edu/ sommese/bertini`.

[3] D.J. Bates, J.D. Hauenstein, C. Peterson and A.J. Sommese, *Numerical Decomposition of the Rank-Deficiency Set of a Matrix of Multvariate Polynomials*, L. Robbiano and J.Abbott (eds.) *Approximate Commutative Algebra*, Vol 14 Texts and Manuscripts in Symbolic Computation, Springer-Verlag, 2009, pp. 55-77. (Or see this author's review in Math. Sci. Net, MR2641156).

[4] D.J. Bates and F. Sottile, *Khovanskii-Rolle Continuation for Real Solutions*, Found. Comput. Math (2011) pp. 11:563-587.

[5] J.W.S. Cassels, *Lectures on Elliptic Curves*, London Math. Soc. Student Texts 24, Cambridge University Press, 1991.

[6] B.H. Dayton, *Numerical Calculation of H-bases for Positive Dimensional Varieties*, in Proceedings of the 2011 International Workshop on Symbolic-Numeric Computation, ACM, 2012.

[7] B.H. Dayton, *Numerical Algebraic Geometry via Macaulay's Perspective*, talk at SIAM AN12 special section, June 2012, at `www.neiu.edu/∼bhdayton/siamAN12.pdf`

[8] L. Dupont, D. Lazard, S. Lazard and S. Petitjean,*Near-optimal parameterization of the intersection of quadrics, Parts I,II,III*, J. Symbolic Comput. 3(43), 2008, pp. 168–232.

[9] L. Dupont, D. Lazard, S. Lazard, S. Petitjean and G. Hanrot, *Quadric Intersection Online Calculation Server*, `http://vegas.loria.fr/qi/server`.

[10] G. Fischer, *Plane Algebraic Curves*, Student Math. Lib. 15, (English translation), AMS, 2001.

[11] M. Kreuzer and L. Robbiano,*Computational Commutative Algebra 2*, Springer, 2005.

[12] F.S. Macaulay *The Algebraic Theory of Modular Systems*, Cambridge University Press, 1916.

[13] H.M. Möller and T. Sauer, *H-bases for polynomial interpolation and system solving*, Adv. in Comp. Math. 12 (2000) pp. 335–362.

[14] C. Tu, W. Wang, B. Mourrain, J. Wang, *Using signature sequences to classify intersection curves of two quadrics*, Computer Aided Geometric Design 26 (2009), pp. 317-335.

[15] W. Wang, B. Joe, R. Goldman, *Computing quadric surface intersections based on an analysis of plane cubic curves*, Graph. Models 64 (2003) 335-367.

ALBANIAN JOURNAL OF MATHEMATICS
Volume 7, Number 2, Pages 77–92
ISSN: 1930-1235; (**2013**)

LINEAR BLOCK AND ARRAY CODES CORRECTING REPEATED CT BURST ERRORS

FATİH TEMİZ
Department of Mathematics,
Yildiz Technical University,
Turkey
Email: ftemiz@yildiz.edu.tr

VEDAT ŞİAP
Department of Mathematical Engineering,
Yildiz Technical University,
Turkey
Email: vedatsiap@gmail.com

Abstract. Burst errors are very common in practice. There have been many designs in order to correct or at least detect such errors. Recently, a new kind of burst error which is termed as repeated burst error has been introduced in order to detect or correct errors that occurs in very busy communication channels. In this paper, we extend the definition of repeated burst errors for block and array codes endowed with a homogeneous metric. We also obtain some upper bounds on the number of parity check digits for these codes correcting all repeated burst errors.

1. Introduction

The early studies in coding theory based on the detection and correction of errors have been introduced for detection and correction of random errors [7]. In the applications of codes to various communication channels, errors do not occur in independently but are in clustered, that is, the error patterns are mostly in the form of bursts. This led to the study of burst error correcting codes, depicted by Fire [6] and Reiger [8]. Because of the nature of applications to communication channels, several definitions regarding the concept of burst error have been introduced by many researchers. Chien and Tang [3] introduced the concept of Chien and Tang (shortly CT) burst errors for block codes. These burst errors have found applications in error analysis experiment on telephone lines [1]. Later, Jain [10] extended the notion of CT burst errors for array codes by endowing a non-Hamming metric [9]. In order to solve the same problem in [10] with a novel approach Siap [11] introduced a CT burst error weight enumerator. The result obtained over finite fields in [10] was extended to array codes over finite rings [12].

2000 *Mathematics Subject Classification.* 94B05, 94B20.
Key words and phrases. burst error, 2−repeated burst error, Chieng-Tang burst error, linear code, array code.

During the process of transmission over very noisy communication channels, errors repeat themselves. Dass [4] *et al.* introduced another type of error pattern called 2−repeated burst error for block codes and obtained an upper bound on the number of parity check digits for codes correcting such errors. It is pointed out that 2−repeated burst error correcting linear block codes provide a good source for detecting and correcting these burst errors in very busy communication channels. In these type of channels, burst errors may repeat more frequently. Therefore, it is useful to consider more than 2−repeated burst errors. Dass & Verma [5] termed such a burst error as $m-$repeated burst error for block codes. In this paper, firstly, we obtain some bounds on the parameters of linear block codes correcting all CT burst errors and all $m-$repeated burst errors with respect to homogeneous metric, respectively. Moreover, the study of these burst errors in terms of homogeneous weight is given. Later, we derive some bounds on the parameters of array codes with respect to this metric by using the definition of the extended CT burst error given in [10] for array coding systems. Furthermore, we combine these two interesting topics: CT burst error and repeated burst error for array codes and we introduce the concept of 2−repeated CT burst error for array codes. Finally, we obtain an upper bound on the parameters of linear array codes correcting all 2−repeated CT burst errors in terms of homogeneous weight.

The organization of this paper is as follows: In Section 2, we develop some basic terminology and cover some preliminary definitions. In Section 3, we study on $m-$repeated CT burst errors and CT burst errors with homogeneous weight constraint in linear block codes. In Sections 4 and 5, some new bounds on the parameters of array linear codes correcting all CT burst errors and all 2−repeated CT burst errors with respect to homogeneous metric are given, respectively.

2. Definitions and notations

Let $\mathbb{Z}_{q^l}$ be the ring of integer modulo q^l, where q is a prime. Let $\mathbb{Z}_{q^l}^n$ be the space of all $n-$tuples with entries from a ring $\mathbb{Z}_{q^l}$. Then $\mathbb{Z}_{q^l}^n$ is a module over $\mathbb{Z}_{q^l}$. C is said to be an $(n, M)-$linear block code if and only if C is a submodule of $\mathbb{Z}_{q^l}^n$ of size M. If C is a $k-$free (with a basis of k elements) submodule with length n, then C is called an $[n, k]-$linear block code. A linear array code C is a linear $\mathbb{Z}_{q^l}-$submodulo of the space $Mat_{m\times s}\left(\mathbb{Z}_{q^l}\right)$, the space of all $m \times s$ matrices with entries from a ring $\mathbb{Z}_{q^l}$.

The homogeneous weight w_{hom} on $\mathbb{Z}_{q^l}$ is defined as

$$w_{hom}(x) = \begin{cases} 0 & \text{if } x = 0 \\ q^{l-1} & \text{if } x \in \ (q^{l-1}) \setminus \{0\} \\ (q-1)q^{l-2} & otherwise \end{cases} \tag{1}$$

where (q^{l-1}) denotes the ideal of $\mathbb{Z}_{q^l}$ generated by q^{l-1}.

For $u = (u_1, u_2, ..., u_n) \in \mathbb{Z}_{q^l}^n$, we have

$$w_{hom}\left(u\right) = \sum_{i=1}^{n} w_{hom}\left(u_i\right). \tag{2}$$

For any $u, v \in \mathbb{Z}_{q^l}^n$, the homogeneous distance d_{hom} is given by

$$d_{hom}\left(u, v\right) = w_{hom}\left(u - v\right). \tag{3}$$

Note that there are $q-1$ elements of weight q^{l-1} and q^l-q elements of weight $(q-1)q^{l-2}$ in $\mathbb{Z}_{q^l}$.

There exist various types of burst errors in order to construct error detecting/correcting codes in literature. We first give the following definition of CT burst error.

Definition 2.1. [3] *A CT burst error of length b is a vector whose nonzero components are confined to some b consecutive components, with the first component being nonzero.*

A $2-$repeated burst error of length b is defined as follows:

Definition 2.2. *A $2-$repeated burst error of length b is a vector of length n whose only nonzero components are confined to two distinct sets of b consecutive components, the first and the last component of each set being nonzero* [4].

Dass & Verma [5] extended the above definition into the definition of $m-$repeated CT burst error as follows:

Definition 2.3. *An $m-$repeated CT burst error of length b is a vector of length n whose only non-zero components are confined to m distinct sets of b consecutive components, the first component of each set being non-zero.*

3. Linear block codes with repeated CT burst errors

In this section, we consider the definitions of CT burst errors given in Definition 2.1 and $m-$repeated CT burst errors given in Definition 2.3 for linear block codes with respect to homogeneous metric, respectively. We determine the number of these burst errors of a given weight w_{hom} and also derive some bounds on the parameters of linear block codes correcting all these burst errors.

We first introduce the following lemma which will be used in the proof of Theorem 3.2.

Lemma 3.1. *The number of all CT burst errors of length b in $\mathbb{Z}_{q^l}^n$ is given by*

$$\mathbf{B}_n^b\left(\mathbb{Z}_{q^l}\right)=(n-b+1)\left(q^l-1\right)\left(q^l\right)^{b-1}. \tag{4}$$

Proof. Choosing b consecutive positions among n positions can be done in $n-b+1$ ways. Then, by Definition 2.1, the first component of these b components can be any q^l-1 nonzero elements of the ring $\mathbb{Z}_{q^l}$ and the other $b-1$ elements can be any q^l elements of the ring $\mathbb{Z}_{q^l}$. □

We present some properties for an $[n,k]-$linear code over $\mathbb{Z}_{q^l}$ that is going to appear in the statement of Theorem 3.2.

Theorem 3.1. [2] *A nonzero $[n,k]-$linear code C over $\mathbb{Z}_{q^l}$ has a generator matrix which after a suitable permutation of the coordinates can be written in the form*

$$G=\begin{pmatrix} I & A_{0,1} & A_{0,2} & A_{0,3} & \ldots & A_{0,l-1} & A_{0,l} \\ 0 & qI & qA_{1,2} & qA_{1,3} & \ldots & qA_{1,l-1} & qA_{1,l} \\ 0 & 0 & q^2I & q^2A_{2,3} & \ldots & q^2A_{2,l-1} & q^2A_{2,l} \\ . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & \ldots & q^{l-1}I & q^{l-1}A_{l-1,l} \end{pmatrix}. \tag{5}$$

Here the columns are grouped into blocks of sizes $k_0, k_1, ..., k_{l-1}, k_l$, and k_i are non-negative integers adding to n. This means that C consists of all codewords

$$[v_0 \ v_1 \ v_2 \ ... \ v_{l-1}]\, G,$$

where each v_i is a vector of length k_i with components from $\mathbb{Z}_{q^{l-i}}$, so that C contains q^k codewords, $A_{i,j}$ $(0 \leq i < j \leq l)$ are matrices over $\mathbb{Z}_{q^{l-i}}$ and

$$k = \sum_{i=0}^{l-1} (l-i)\, k_i.$$

We say that C has type

$$1^{k_0} q^{k_1} \left(q^2\right)^{k_2} ... \left(q^{l-1}\right)^{k_{l-1}}.$$

The following theorem gives a bound in order to correct of all CT burst errors in linear codes with respect to homogeneous metric.

Theorem 3.2. *An $[n,k]-$linear code C over $\mathbb{Z}_{q^l}$ that corrects all CT burst errors of length b must satisfy the bound*

$$q^{ln-k} \geq 1 + \mathbf{B}_n^b \left(\mathbb{Z}_{q^l}\right), \tag{6}$$

where $\mathbf{B}_n^b \left(\mathbb{Z}_{q^l}\right)$ is the number of all CT burst errors of length b in $\mathbb{Z}_{q^l}^n$.

Proof. The proof is based on the fact that the number of available cosets must be greater than or equal to the number of correctable CT burst errors having the vector of all zeros. By Lemma 3.1, the number of correctable CT burst errors having the vector of all zeros is

$$1 + \mathbf{B}_n^b \left(\mathbb{Z}_{q^l}\right).$$

By Theorem 3.1, an $[n,k]-$linear code C over $\mathbb{Z}_{q^l}$ has q^k codewords. So the number of available cosets is q^{ln-k}. □

We now derive a bound for the correction of all CT burst errors of length b or less with respect to homogeneous metric.

Theorem 3.3. *An $[n,k]-$linear code C over $\mathbb{Z}_{q^l}$ that corrects all CT burst errors of length b or less must satisfy the bound*

$$q^{ln-k} \geq 1 + \sum_{a=1}^{b} \mathbf{B}_n^a \left(\mathbb{Z}_{q^l}\right), \tag{7}$$

$\mathbf{B}_n^a \left(\mathbb{Z}_{q^l}\right)$ is the number of all CT burst errors of length a in $\mathbb{Z}_{q^l}^n$.

Proof. Using Theorem 3.2, its proof is straightforward. □

In Theorem 3.4, we obtain a bound on the number of parity check digits for a linear code over $\mathbb{Z}_{q^l}$ correcting all CT burst errors of length b or less having homogeneous weight w_{hom} or less. For this purpose, we first prove the following lemma that counts the number of all CT burst errors of length b having homogeneous weight w_{hom}.

Lemma 3.2. *The number of all CT burst errors of length b having homogeneous weight w_{hom} in $\mathbb{Z}_{q^l}^n$ is given by*

$$\begin{aligned}\mathbf{B}_n^b\left(\mathbb{Z}_{q^l}, w_{hom}\right) = \quad & (n-b+1)\Bigg[\sum_{u,v}\binom{b-1}{u}\binom{b-1-u}{v}(q-1)^{u+1}(q^l-q)^v \\ + \quad & \sum_{u',v'}\binom{b-1}{u'}\binom{b-1-u'}{v'}(q-1)^{u'}(q^l-q)^{v'+1}\Bigg],\end{aligned} \tag{8}$$

where u, v and u', v' are nonnegative integers such that

$$u+v \leq b-1,$$
$$u'+v' \leq b-1,$$

$$w_{hom} - q^{l-1} = u(q^{l-1}) + v(q-1)(q^{l-2}),$$
$$w_{hom} - (q-1)(q^{l-2}) = u'(q^{l-1}) + v'(q-1)(q^{l-2}).$$

Note that

$$\begin{aligned}&\textstyle\sum_{u,v}\binom{b-1}{u}\binom{b-1-u}{v}(q-1)^{u+1}(q^l-q)^v = 0 && \text{if } A,\\ &\textstyle\sum_{u',v'}\binom{b-1}{u'}\binom{b-1-u'}{v'}(q-1)^{u'}(q^l-q)^{v'+1} = 0 && \text{if } B,\end{aligned}$$

where

$$A := w_{hom} \neq (u+1)\,(q^{l-1}) + v(q-1)(q^{l-2})$$
$$B := w_{hom} \neq (u')\,(q^{l-1}) + (v'+1)\,(q-1)(q^{l-2}).$$

Proof. Consider a CT burst error of length b having homogeneous weight w_{hom}. Since the first component of the b consecutive positions in which nonzero components are clustered, it must be nonzero and since there exist two different nonzero homogeneous weights in $\mathbb{Z}_{q^l}$, we can investigate our proof in two cases:

Case 1: If the first nonzero component is of weight q^{l-1} then there can be u components of weight q^{l-1} in rest of b components and there can be v components of weight $(q-1)(q^{l-2})$ in rest of $b-u$ components such that $u+v \leq b-1$ and $w_{hom} = (u+1)(q^{l-1}) + v(q-1)(q^{l-2})$. For every values of u and v satisfying inequality and equality at the same time, we have some CT burst errors of length b having homogeneous weight w_{hom} in $\mathbb{Z}_{q^l}^n$. For convenient values of u and v we can choose u positions among $b-1$ positions and then we can choose v positions among $b-1-u$ positions. Afterwards, since there are $u+1$ components of weight q^{l-1} including the first component and since there are v components of weight $(q-1)q^{l-2}$, regarding there are $q-1$ elements of weight q^{l-1} and q^l-q elements of $(q-1)q^{l-2}$ in $\mathbb{Z}_{q^l}$, we have

$$(n-b+1)\binom{b-1}{u}\binom{b-1-u}{v}(q-1)^{u+1}(q^l-q)^v$$

CT burst errors of length b having homogeneous weight w_{hom}.

Case 2: When the first nonzero component is of weight $(q-1)q^{l-2}$ similarly, for convenient nonzero integers u' and v' satisfying $u'+v' \leq b-1$ and $w = u(q^{l-1}) + (v+1)(q-1)(q^{l-2})$, we have

$$(n-b+1)\binom{b-1}{u'}\binom{b-1-u}{v'}(q-1)^u(q^l-q)(v+1)$$

CT burst errors of length b having homogeneous weight w_{hom}. □

We give an example in order to illustrate Lemma 3.2.

Example 3.1. *Let $q = l = 2$, $b = 3$, $n = 5$ and $w_{hom} = 4$. We first determine nonnegative integers u, v and u', v' satisfying the conditions in Lemma 3.2 in order to obtain the number of all CT burst errors of length 3 having homogeneous weight 4 in $\mathbb{Z}^5_{2^2}$. For this we consider the following conditions for integers u, v and u', v':*

$$\begin{aligned} 4 &= (u+1)2^{2-1} + v(2-1)(2^{2-2}), \\ 4 &= u'2^{2-1} + (v'+1)(2-1)(2^{2-2}), \\ 2 &\geq u+v, \\ 2 &\geq u'+v'. \end{aligned} \tag{9}$$

Then, nonnegative integer solutions u, v and u', v' satisfying the conditions given in (9) are as follows, respectively:

$$u = 0, v = 2 \text{ and } u = 1, v = 0$$

and also

$$u' = 1, v' = 1.$$

Substituting these nonnegative integers into Eq.(8) in Lemma 3.2, we get the following result:

$$3\left[\binom{2}{0}\binom{2}{2}1^1 2^2 + \binom{2}{1}\binom{1}{0}1^2 2^0 + \binom{2}{1}\binom{1}{1}1^1 2^2\right] = 42.$$

In fact, all CT burst errors of length 3 having homogeneous weight 4 in $\mathbb{Z}^5_4$ are as follows:

11200	01120	00112	13200	01320	00132	12100
01210	00121	12300	01230	00123	31200	03120
00312	33200	03320	00332	32100	03210	00321
32300	03230	00323	22000	02200	00220	20200
02020	00202	21100	02110	00211	21300	02130
00213	23100	02310	00231	23300	02330	00233.

Theorem 3.4. *An $[n, k]-$linear code C over $\mathbb{Z}_{q^l}$ that corrects all CT burst errors of length b or less and having homogeneous weight w_{hom} or less must satisfy the bound*

$$q^{ln-k} \geq 1 + \sum_{a=1}^{b}\sum_{j=1}^{a(q^{l-1})} \mathbf{B}^a_n\left(\mathbb{Z}_{q^l}, j\right) \tag{10}$$

where $\mathbf{B}_n^a\left(\mathbb{Z}_{q^l}, j\right)$ *is the number of all CT burst errors of length* a *in* $\mathbb{Z}_{q^l}^n$ *having homogeneous weight* j*.*

Proof. The proof is straightforward from Lemma 3.2. □

Now, we can obtain some results on the correction of $m-$repeated CT burst errors in linear codes with respect to homogeneous metric.

Lemma 3.3. *The number of all* $m-$*repeated CT burst errors of length* b *in* $\mathbb{Z}_{q^l}^n$ *is given by*

$$\mathbf{B}_n^{m,b}\left(\mathbb{Z}_{q^l}\right) = \frac{(n-mb+1)\cdots(n-mb+m)}{m!}\left(q^l-1\right)^m\left(q^l\right)^{mb-m}.$$

Proof. Clearly, m distinct sets of b consecutive positions can be chosen in

$$\frac{(n-mb+1)\cdots(n-mb+m)}{m!}$$

ways among n positions. Later, each vector can be constructed such a way that the first components of m sets of b consecutive positions in which all the nonzero components are clustered must be nonzero, and the rest $b-1$ components may be any element of $\mathbb{Z}_{q^l}$ for each m distinct set. Hence the lemma. □

We enumerate all $m-$repeated CT burst errors with weight constraint in the following lemma.

Lemma 3.4. *The number of all* $m-$ *repeated CT burst errors of length* b *having homogeneous weight* w_{hom} *in* $\mathbb{Z}_{q^l}^n$ *is given by*

$$\mathbf{B}_n^{m,b}\left(\mathbb{Z}_{q^l}, w_{hom}\right) = \frac{(n-mb+1)\cdots(n-mb+m)}{m!}\times$$
$$\left[\sum_{i=0}^{m}\sum_{u_i,v_i}\binom{m}{i}\binom{mb-m}{u_i}\binom{mb-m-u_i}{v_i}(q-1)^{u_i+i}(q^l-q)^{v_i+m-i}\right],$$

where u_i *and* v_i *are nonnegative integers for each* $0 \le i \le m$ *such that*

$$u_i + v_i \le m(b-1),$$

and for each $0 \le i \le m$

$$w_{hom} = (u_i+i)(q^{l-1}) + (v_i+m-i)(q-1)(q^{l-2}).$$

Say

$$\mathbf{B}_n^{m,b}\left(\mathbb{Z}_{q^l}, w_{hom}\right) = \frac{(n-mb+1)\cdots(n-mb+m)}{m!}\times\left[\mathbf{V}_b^m\left(\mathbb{Z}_{q^l}, w_{hom}\right)\right].$$

Proof. Considering m distinct sets of b consecutive components, the sketch of the proof can be investigated in $m+1$ cases depending upon weights of the first components of each b consecutive positions.

Case 1: Assuming all the first components of each m distinct sets of b consecutive positions are of weight $(q-1)q^{l-2}$ we have

$$\binom{mb-m}{u_0}\binom{mb-m-u_0}{v_0}(q-1)^{u_0}(q^l)^{v_0+m}$$

$m-$repeated CT burst errors of length b for convenient pairwise u_0 and v_0.

Case 2: Assuming one of the first components of m distinct sets of b consecutive positions is of weight q^{l-1} and all the others are of weight $(q-1)q^{l-2}$ we have

$$\binom{m}{1}\binom{mb-m}{u_1}\binom{mb-m-u}{v_1}(q-1)^{u_1+1}(q^l)^{v_1+m-1}$$

$m-$repeated CT burst errors of length b for convenient pairwise u_1 and v_1.

Applying similar arguments, we have

$$\binom{m}{i}\binom{mb-m}{u_i}\binom{mb-m-u_i}{v_i}(q-1)^{u_i+i}(q^l)^{v_i+m-i}$$

$m-$repeated CT burst errors of length b for convenient u_i and v_i whenever i $(0 \leq i \leq m)$ of the first components of m distinct sets of b consecutive positions is of weight q^{l-1} and the rest are of weight $(q-1)q^{l-2}$. Hence the result. $\square$

Example 3.2. *Take $q = l = m = b = 2, w = 3,$ and $n = 5$ in Lemma 3.4. Then, nonnegative integers u_i and v_i for each $0 \leq i \leq m$ such that*

$$3 = (u_i + i)(2^{2-1}) + (v_i + 2 - i)(2-1)(2^{2-2}), \text{ and } 2(2-1) \geq u_i + v_i$$

are given by

$$u_0 = 0, v_0 = 1 \text{ and } u_1 = v_1 = 0$$

and there is no solution for u_2 and v_2 and u_3 and v_3. Therefore, our formula gives a direct computation

$$\frac{2\cdot 3}{2!}\left[\binom{2}{0}\binom{2}{0}\binom{2}{1}1^0 2^3 + \binom{2}{1}\binom{2}{0}\binom{2}{0}1^1 2^1\right] = 60.$$

In fact, all $2-$repeated CT burst errors of length 2 having homogeneous weight 3 in $\mathbb{Z}_4^5$ are as follows:

01020	10020	10200
02010	20010	20100
03020	30020	30200
02030	30020	30200
01110	11010	11100
03330	33030	33300
01011	10011	10110
03033	30033	30330
01130	11030	11300
03310	33010	33100
01013	10013	10130
03031	30031	30310
01310	13010	13100
03130	31030	31300
01013	10013	10130
03031	30031	30310
03110	31010	31100
01310	13010	13100
03031	30031	30310
01013	10013	10130.

In Theorem 3.5, we obtain a bound for the correction of all $m-$repeated CT burst errors of length b or less and having homogeneous weight w or less.

Theorem 3.5. *An $[n,k]-$linear code C over $\mathbb{Z}_{q^l}$ that corrects all $m-$repeated CT burst errors of length b or less and having homogeneous weight w_{hom} or less must satisfy the bound*

$$q^{ln-k} \geq 1 + \sum_{a=1}^{b} \sum_{j=m(q-1)q^{l-2}}^{ma(q^{l-1})} \mathbf{B}_n^{m,a}\left(\mathbb{Z}_{q^l}, j\right). \tag{11}$$

Proof. It follows directly from Lemma 3.4. □

4. Linear array codes with respect to homogeneous weight

In this section, we extend the notion of CT burst errors for linear array codes with respect to homogeneous metric, originally in given in [10].

Definition 4.1. *A CT burst error of order $p \times r$ ($1 \leq p \leq m$, $1 \leq r \leq s$) in $Mat_{m\times s}(\mathbb{Z}_{q^l})$ is an $m \times s$ matrix in which all the nonzero entries are confined to some $p \times r$ submatrix which has nonzero first row and first column.*

We first enumerate all CT burst errors having homogeneous weight w_{hom} for linear array codes.

Lemma 4.1. *The number of all CT burst errors of order $p\times r$ having homogeneous weight w_{hom} in $Mat_{m\times s}(\mathbb{Z}_{q^l})$ is given by*

$$\begin{aligned}
\mathbf{B}_{m\times s}^{p\times r}\left(\mathbb{Z}_{q^l}, w\right) &= (m-p+1)(s-r+1) \\
&\times \Bigg[\mathbf{V}_{pr}^{1}\left(\mathbb{Z}_{q^l}, w_{hom}\right) + (p-1)(r-1) \\
&\times \Bigg[\sum_{a,b}\binom{pr-3}{a}\binom{pr-3-a}{b}(q-1)^{a+1}(q^l-q)^{b+1} \\
&+ \sum_{c,d}\binom{pr-3}{c}\binom{pr-3-c}{d}(q-1)^{c+2}(q^l-q)^{d} \\
&+ \sum_{e,f}\binom{pr-3}{e}\binom{pr-3-e}{f}(q-1)^{e}(q^l-q)^{f+2}\Bigg]\Bigg]
\end{aligned}$$

where a,b,c,d,e,f are nonnegative integers satisfying

$$a+b \leq pr-3,$$
$$c+d \leq pr-3,$$
$$e+f \leq pr-3,$$

and

$$w_{hom} = (a+1)(q^{l-1}) + (b+1)(q-1)(q^{l-2}),$$
$$w_{hom} = (c+2)(q^{l-1}) + d(q-1)(q^{l-2}),$$
$$w_{hom} = c(q^{l-1}) + (d+2)(q-1)(q^{l-2}).$$

Proof. Take any CT burst error $A \in Mat_{m\times s}(\mathbb{Z}_{q^l})$ of order $p\times r$. Let us denote $p\times r$ nonzero submatrix whose first row and first column being nonzero by B. The row number of starting positions for B can vary between 1 and $m-p+1$ and the column number of starting positions for B can vary between 1 and $s-r+1$. Therefore, choosing the location of B regardless of its entries can be done in $(m-p+1)(s-r+1)$ ways.

The selection of entries of the submatrix B such that having homogeneous weight w_{hom} can be considered in two cases.

Case 1: If the entry b_{11} is different from zero, then the first row and the first column of B will be nonzero automatically. For this case, constructing the submatrix B can be achieved in $\mathbf{V}^1_{pr}$ ways since the work is the same that constructing a CT burst error of length pr in one array.

Case 2: If the entry b_{11} is zero, then we must have at least one nonzero component in the first row and at least one nonzero component in the first column of B. There are $(p-1)(r-1)$ options for choosing these nonzero components. Since the rest of $pr-3$ entries will be selected depending upon the weights of these two nonzero components, the constructing of the submatrix B can be considered in three cases as given in the lemma. Hence, the proof is completed. □

Theorem 4.1. *An $[n,k]-$linear array code C over $\mathbb{Z}_{q^l}$ that corrects all CT burst errors of order $p\times r$ or less and having homogeneous weight w_{hom} or less must satisfy the bound*

$$q^{lms-k} \geq 1+\sum_{a=1}^{p}\sum_{b=1}^{r}\sum_{j=(q-1)q^{l-2}}^{ab(q^{l-1})} \mathbf{B}^{a\times b}_{m\times s}\left(\mathbb{Z}_{q^l}, j\right). \tag{12}$$

Proof. It follows directly from Lemma 4.1. □

5. Linear array codes with 2−repeated CT burst errors

In this section, we present a new transmission model for linear array codes correcting all types of burst errors introduced in Definition 5.1. Suppose that a message is an $s-$tuple of $m-$tuples of symbols from $\mathbb{Z}_{q^l}$. We assume that this message is sent over m parallel channels. When this coded message is transmitted through m parallel channels, it may get corrupted and errors may occur. These errors are not scattered randomly but occur in clusters in the code matrix. In this code matrix, burst errors may repeat usually themselves in the submatrix parts of the code matrix. These errors appear due to very busy communication channels for array coding systems. So, we define the notion of 2−repeated CT burst errors for linear array codes.

Definition 5.1. *A 2−repeated CT burst error of order $p\times r$ in $Mat_{m\times s}(\mathbb{Z}_{q^l})$ is an $m\times s$ matrix in which all the nonzero entries are confined to two distinct $p\times r$ submatrices, with the first row and the first column of each submatrix being nonzero.*

Lemma 5.1. *The number of all 2-repeated CT burst errors of order $p \times r$ in $Mat_{m\times s}(\mathbb{Z}_{q^l})$ is given by*

$$\mathbf{B}^{2,p\times r}_{m\times s}\left(\mathbb{Z}_{q^l}\right) = \Bigg[(m-2p+1)\left(p\sum_{i=1}^{s-2r+1} i\right) + (m-2p+1)\left((p-1)\sum_{j=1}^{s-2r+1} j\right) + (s-r+1)^2\left(\sum_{k=1}^{m-2p+1} k\right) + \sum_{u=1}^{s-2r+1} u\sum_{v=1}^{p} v + \sum_{z=1}^{s-2r+1} z\sum_{t=0}^{p-1} t\Bigg]\times (q^l)^{2r(p-1)}\left[\left((q^l)^r-1\right)-\left((q^l)^{r-1}-1\right)q^{l(1-p)}\right]^2$$

and say

$$\mathbf{B}^{2,p\times r}_{m\times s}(\mathbb{Z}_{q^l}) = \mathbf{B}^{2,p\times r}_{m\times s} \times (q^l)^{2r(p-1)}\left[\left((q^l)^r-1\right)-\left((q^l)^{r-1}-1\right)q^{l(1-p)}\right]^2.$$

Proof. Consider a matrix A of order $m\times s$ having distinct submatrices B and C of order $p\times r$, with the first row and the first column of each submatrices being nonzero.

For the sake of avoiding complexity, we will take into account the arrangement of submatrices. For this purpose we will set a submatrix as the first one if the row number of its starting position is less than the row number of the starting position of the other one. If starting positions of submatrices are located at the same row then the left one is set as the first one.

In order to obtain 2-repeated burst errors of order $p\times r$, the starting position of the first submatrix can be a_{ij} where $1\le i\le(m-p+1)$ and $1\le j\le(s-r+1)$. Note that the column number of starting position of the first submatrix vary between 1 and $s-2r+1$ when its row number is $m-p+1$.

For $m-p+1$ possible row number of the possible starting positions of the first submatrix, we will consider $m-p+1$ steps to count the number of starting positions of the second submatrix. We will also consider several cases in these steps. In this manner, first consider the matrix having two submatrices irrespective of their entries, in which the starting position of the first submatrix is a_{11}, then the number of all possible starting positions for the second matrix will be

$$(s-2r+1)\cdot p + (s-r+1)(m-2p+1).$$

Afterwards consider the matrix including two submatrices, in which the starting position of the first submatrix is a_{12}, this time the number of all possible starting positions for the second submatrix will be

$$(s-2r)\cdot p + (s-r+1)(m-2p+1).$$

In this wise, until the starting position of the first submatrix will be a_{1r}, the number of possible starting positions for the second submatrix will decrease p according to the previous case as the column number increase. Because until this case, the second submatrix C can never appear on the left side of the first submatrix B entirely and hence the number of possible starting positions will decrease p for p rows of B.

After this step, the number of possible starting positions for C will decrease 1 according to the previous case each time. Because after this step, the second submatrix C can appear on the left side of B entirely on condition that the row number of its starting position is greater than the row number of the starting position of B. Here, we can think this number as decreasing p and increasing $p-1$ for each case. Hence the number of all possible starting positions for the second submatrix will be

$$(s-3r+1)\cdot p + 1\cdot(p-1) + (s-r+1)(m-2p+1)$$

whenever the first submatrix starts from $a_{1,r+1}$, and

$$1\cdot p + (s-3r+1)(p-1) + (s-r+1)(m-2p+1)$$

whenever B starts from $a_{1,s-2r+1}$

After the case the starting position of the first submatrix is $a_{1,s-2r+2}$, the number of the possible starting positions for the second submatrix will not change until the last case of our first step that the starting position of the first submatrix is located in the first row of the matrix $A_{m\times s}$ i.e. the starting position of B is being $a_{1,s-r+1}$ Thus, the number of all possible starting positions for the second submatrix will be

$$(s-2r+1)(p-1) + (s-r+1)(m-2p+1)$$

in this last case.

Up to now, we have considered all 2-repeated burst errors where the starting position of the first submatrix varies in the first row of the matrix A.

In the same way, let us consider all 2-repeated burst errors where this time the starting position of B varies in the second row of the matrix A. In these cases, clearly, the number of possible starting positions for C will decrease $(s-r+1)$ for each time. When B starts from a_{21}, for instance, the number of all possible starting positions for the second submatrix will be

$$(s-2r+1)\cdot p + (s-r+1)(m-2p)$$

and

$$(s-2r+1)(p-1) + (s-r+1)(m-2p)$$

when B starts from $a_{2,s-r+1}$.

As the row number of the starting position of B increases until the row number $(m-2p+1)$ i.e. in each step, the number of possible starting positions for C will be decrease $(s-r+1)$ each time according to the previous row number. Hence the number of all possible starting positions for C will be

$$(s-2r+1)\cdot p$$

when B starts from $a_{m-2p+2,1}$ and

$$(s-2r+1)(p-1)$$

when B starts from $a_{m-2p+2,s-r+1}$.

After this step, the $(s-r+1)$ decreasing of the number of possible starting positions for the second submatrix for each time will be invalid because the second submatrix will not be able to located under the first submatrix any more.

This combinatoric process proves that the number of all two distinct submatrices of order $p\times r$ in a matrix of order $m\times s$ is

$$\mathbf{B}_{m\times s}^{2,p\times r} = \Bigg[(m-2p+1)\left(p\sum_{i=1}^{s-2r+1} i\right) + (m-2p+1)\left((p-1)\sum_{j=1}^{s-2r+1} j\right) + (s-r+1)^2\left(\sum_{k=1}^{m-2p+1} k\right) + \sum_{u=1}^{s-2r+1} u\sum_{v=1}^{p} v + \sum_{z=1}^{s-2r+1} z\sum_{t=0}^{p-1} t\Bigg]$$

Now, let us compute the number of ways that elements of submatrices B and C can be selected. This selection can be done in

$$\left((q^l)^r - 1\right)(q^l)^{r(p-1)} \tag{13}$$

ways if the first row of the submatrix is not entirely zero. However, to consider a CT burst error, the first column of the submatrix must also be nonzero. In this manner, we should subtract the number of cases included in the equation (13) which the first column of the submatrix being zero. This number is given by

$$\left((q^l)^{r-1} - 1\right)(q^l)^{(r-1)(p-1)}. \tag{14}$$

By subtracting (13) from (14) we obtain

$$(q^l)^{r(p-1)}\left[\left((q^l)^r - 1\right) - \left((q^l)^{r-1} - 1\right)(q^l)^{1-p}\right] \tag{15}$$

and also note that A is a 2-repeated burst error. Hence the lemma. □

Example 5.1. *Let us show all $2-$repeated CT burst errors of order 2×2 in $Mat_{4\times 4}(\mathbb{Z}_{q^l})$ such that all of the entries of the submatrices B and C are 1.*

$$\begin{pmatrix}1&1&1&1\\1&1&1&1\\0&0&0&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}1&1&0&0\\1&1&1&1\\0&0&1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}1&1&0&0\\1&1&0&0\\1&1&0&0\\1&1&0&0\end{pmatrix}\begin{pmatrix}1&1&0&0\\1&1&0&0\\0&1&1&0\\0&1&1&0\end{pmatrix}$$

$$\begin{pmatrix}1&1&0&0\\1&1&0&0\\0&0&1&1\\0&0&1&1\end{pmatrix}\begin{pmatrix}0&1&1&0\\0&1&1&0\\1&1&0&0\\1&1&0&0\end{pmatrix}\begin{pmatrix}0&1&1&0\\0&1&1&0\\0&1&1&0\\0&1&1&0\end{pmatrix}\begin{pmatrix}0&1&1&0\\0&1&1&0\\0&0&1&1\\0&0&1&1\end{pmatrix}$$

$$\begin{pmatrix}0&0&1&1\\1&1&1&1\\1&1&0&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}0&0&1&1\\0&0&1&1\\1&1&0&0\\1&1&0&0\end{pmatrix}\begin{pmatrix}0&0&1&1\\0&0&1&1\\0&1&1&0\\0&1&1&0\end{pmatrix}\begin{pmatrix}0&0&1&1\\0&0&1&1\\0&0&1&1\\0&0&1&1\end{pmatrix}$$

$$\begin{pmatrix}0&0&0&0\\1&1&1&1\\1&1&1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}0&0&0&0\\1&1&0&0\\1&1&1&1\\0&0&1&1\end{pmatrix}\begin{pmatrix}0&0&0&0\\0&0&1&1\\1&1&1&1\\1&1&0&0\end{pmatrix}\begin{pmatrix}0&0&0&0\\0&0&0&0\\1&1&1&1\\1&1&1&1\end{pmatrix}.$$

Note that this number is precisely the same as $\mathbf{B}_{4\times 4}^{2,2\times 2}$ since the selection of entires was done uniquely.

Theorem 5.1. *The number of all 2-repeated CT burst errors of order $p\times r$ having homogeneous weight w_{hom} in $Mat_{m\times s}(\mathbb{Z}_{q^l})$ is given by*

$$\mathbf{B}^{2,p\times r}_{m\times s}\left(\mathbb{Z}_{q^l}, w_{hom}\right) = \mathbf{B}^{2,p\times r}_{m\times s}\left(\mathbb{Z}_{q^l}\right) \times \Bigg[\mathbf{V}^2_{pr}(\mathbb{Z}_{q^l}, w)+$$

$$2(p-1)(r-1)\times\sum_{i=0}^{3}\sum_{u_i,v_i}\binom{2pr-3}{u_i}\binom{2pr-3-u_i}{v_i}(q-1)^{u_i+i}(q^l-q)^{v_i+3-i}+$$

$$(p-1)^2(r-1)^2\times\sum_{j=0}^{4}\sum_{u'_j,v'_j}\binom{2pr-4}{u'_j}\binom{2pr-4-u'_j}{v'_j}(q-1)^{u'_j+j}(q^l-q)^{v'_j+4-j}\Bigg],$$

where u_i, v_i *and* u'_j, v'_j *are all nonnegative pairwise integers for* $0 \leq i \leq 3$ *and* $0 \leq j \leq 4$ *satisfying*

$$u_i + v_i \leq 2pr - 3$$

$$u'_j + v'_j \leq 2pr - 4$$

also for each $0 \leq i \leq 3$ *and* $0 \leq j \leq 4$ *satisfying* $w_{hom} = (u_i + i)(q^{l-1}) + (v_i + 3 - i)(q-1)(q^{l-2})$ *and* $w_{hom} = (u'_j + j)(q^{l-1}) + (v'_j + 4 - j)(q-1)(q^{l-2})$.

Proof. Let A be a matrix of order $m \times s$ having two distinct submatrices B and C with first row and first column of each being nonzero. The places of these two submatrices can be chosen in $\mathbf{B}^{2,p\times r}_{m\times s}$ ways by the lemma 5.1. After choosing arrays of submatrices we have three cases with respect to the weights of the starting positions of B and C.

Case 1: When the starting positions of B and C are both nonzero then we have $\mathbf{V}^2_{pr}(\mathbb{Z}_{q^l}, w)$ CT burst errors.

Case 2: When only one of the starting positions of B or C is zero then we choose one position from the first row and one position from the first column of submatrix whose starting position is zero. Afterwards, we use the same technique for counting burst errors by taking into account the weights of three chosen nonzero positions.

Case 3: When the starting positions of B and C are both zero, then after choosing one position from the first row and one position from the first column for each submatrix, again we count the burst errors taking into account the weights of four chosen nonzero positions. □

Example 5.2. *Let us show all possible submatrices* B *and* C *each of order* 2×2 *and having homogeneous weight* 3 *altogether in* $Mat_{m\times s}(\mathbb{Z}_{2^2})$.

$$\begin{pmatrix}1&0\\1&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix} \quad \begin{pmatrix}1&1\\0&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix} \quad \begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}$$

$$\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}1&1\\0&0\end{pmatrix} \quad \begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix} \quad \begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}1&0\\1&0\end{pmatrix}$$

$$\begin{pmatrix}3&0\\3&0\end{pmatrix}\begin{pmatrix}3&0\\0&0\end{pmatrix} \quad \begin{pmatrix}3&3\\0&0\end{pmatrix}\begin{pmatrix}3&0\\0&0\end{pmatrix} \quad \begin{pmatrix}3&0\\0&3\end{pmatrix}\begin{pmatrix}3&0\\0&0\end{pmatrix}$$

$$\begin{pmatrix}3&0\\0&0\end{pmatrix}\begin{pmatrix}3&3\\0&0\end{pmatrix} \quad \begin{pmatrix}3&0\\0&0\end{pmatrix}\begin{pmatrix}3&0\\0&3\end{pmatrix} \quad \begin{pmatrix}3&0\\0&0\end{pmatrix}\begin{pmatrix}3&0\\3&0\end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 3 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 3 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that we showed that there are 16 *possible places for locating these submatrices in a* 4×4 *matrix in example 5.1 and hence there are* $16 \times 68 = 1088$ $2-$*repeated CT burst errors of order* 2×2 *in* $Mat_{4\times 4}(\mathbb{Z}_4)$*.*

Theorem 5.2. *An* $[n, k]-$*linear array code over* $\mathbb{Z}_{q^l}$ *that corrects all* $2-$*repeated CT burst errors of order* $p \times r$ *having homogeneous weight* w_{hom} *must satisfy the bound*

$$q^{lms-k} \geq 1 + \sum_{a=1}^{p}\sum_{b=1}^{r}\sum_{j=2(q-1)(q^{l-2})}^{2ab(q^{l-1})} \mathbf{B}_{m\times s}^{p\times r,2}\left(\mathbb{Z}_{q^l}, j\right). \tag{16}$$

Proof. It follows directly from Lemma 5.1 and Theorem 5.1. □

References

[1] A.A. Alexander, R.M. Gryb, D.W. Nast, *Capabilities of the telephone network for data transmission*, Bell System Tech. J. 39 (1960), 431-476.
[2] A.R. Calderbank, N.J.A. Sloane, *Modular and $p-$adic cyclic codes*, Des. Codes Cryptogr. 6 (1995), 21-35.
[3] R.T. Chien, D.T. Tang, *On the definition of a burst*, IBM J. Res. Develop. 9 (1965), 292-293.
[4] B.K. Dass, R. Verma, L. Berardi, *On 2-repeated burst error detecting codes*, J. Stat. Theory Pract. 3 (2009), 381-391.
[5] B.K. Dass, R. Verma, *Repeated burst error correcting linear codes*, Asian-Eur. J. Math. 1 (2008), 303-335.
[6] P. Fire, *A class of multiple-error-correcting binary codes for non-independent errors*, Sylvania Report RSL-E-2, Sylvania Reconnaissance System Laboratory, Mountain View, Calif, (1959).
[7] R.W. Hamming, *Error-detecting and error-correcting codes*, Bell System Tech. J. 29 (1950), 147-160.
[8] S.H. Reiger, *Codes for the correction of clustered errors*, IRE Trans. Inform. Theory. IT-6 (1960), 16-21.
[9] M.Y. Rosenbloom, M.A. Tsfasman, *Codes for the $m-$metric*, Problems Inform. Transmission. 33 (1997), 45-52.
[10] S. Jain, *CT bursts from classical to array coding*, Discrete Math. 308 (2008), 1489-1499.
[11] I. Siap, *CT burst error weight enumerator of array codes*, Albanian J. Math. 2 (2008), 171-178.
[12] I. Siap, *Burst error enumeration of $m-$array codes over rings and its applications*, Eur. J. Pure Appl. Math. 3 (2010), 653-669.

Albanian Journal of Mathematics (ISSN: 1930-1235) was founded by T. Shaska in 2007 with the idea to support Albanian mathematicians in Albania and abroad.

The journal is not associated with any government institutions in Albania or any public or private universities in Albania or abroad. The journal does not charge any fees to the authors and has always been an open access journal. The journal supports itself with private donations and voluntary work from its staff. Its main office is in Vlora, Albania.

www.albanian-j-math.com