UNIVERSITÀ DEGLI STUDI DI ROMA "LA SAPIENZA"

Dipartimento di Matematica, Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate

ISTITUTO NAZIONALE DI ALTA MATEMATICA "FRANCESCO SEVERI"

RENDICONTI DI MATEMATICA

E DELLE SUE APPLICAZIONI

Serie VII - Volume 25 - Tomo I - 2005

ADDRESS:
RENDICONTI DI MATEMATICA
DIPARTIMENTO DI MATEMATICA - PIAZZALE A. MORO, 2
00185 ROMA, ITALIA

Fax.: +39 06 49913052
e-mail: rendmat@mat.uniroma1.it
web page: www.dmmm.uniroma1.it/~rendiconti

Composizione tipografica effettuata con il programma TEX © dalla
CompoMat – Loc. Braccone – 02040 Configni (RI) – Telefax 39 0746 672240

Stampato a cura del *Centro Stampa d'Ateneo* dell'Università di Roma *"La Sapienza"*

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 1-16

Transformation groups and submanifold geometry

GUDLAUGUR THORBERGSSON

Abstract: *In the talk I give a survey on polar actions and generalizations of isoparametric hypersurfaces in space forms to more general ambient spaces.*

1 – Introduction

In this talk we will give a brief survey on generalizations of isoparametric hypersurfaces to submanifolds with higher codimension in various types of ambient spaces. We will also discuss the question when such submanifolds are homogeneous and introduce the isometric actions which have them as orbits.

A hypersurface M^n of a Riemannian manifold V^{n+1} is called *isoparametric* if M^n is locally a regular level set of a function f with the property that both $\| \operatorname{grad} f \|^2$ and Δf are constant on the level sets of f. One can show that M^n is an isoparametric hypersurface of V^{n+1} if and only if M^n and its parallel hypersurfaces have constant mean curvature.

The term 'isoparametric hypersurface' is due to Levi-Civita ([37]) and refers to the fact that $\| \operatorname{grad} f \|^2$ and Δf were at the time called the *first* and the *second differential parameter of* f respectively.

If the ambient space V^{n+1} is a real space form, then M^n can be shown to be an isoparametric hypersurface if and only if it has constant principal curvatures; see [9]. This characterization does not hold in more general ambient spaces; see [60] where counterexamples are given in complex projective spaces.

Key Words and Phrases: *Polar actions – Isoparametric submanifolds – Equifocal submanifolds.*
A.M.S. Classification: 53C30, 53C35, 53C40

Beniamino Segre proved the following theorem in [48]: let M^n be an isoparametric hypersurface in $\mathbf{R}^{n+1}$. Then M^n is a piece of a plane, of a sphere, or of a round cylinder. In particular it follows that M^n is homogeneous if it is complete. Conversely, it is clear that homogeneous hypersurfaces of $\mathbf{R}^{n+1}$ are isoparametric.

The case $n=2$ of the theorem of Segre was first proved by SOMIGLIANA ([49]) and later reproved in [47] and [37].

Cartan classified isoparametric hypersurfaces in hyperbolic spaces in [9] which also turn out to be homogeneous. He then turned to isoparametric hypersurfaces in spheres, see [10], [11], and [12], and noticed that the problem is much more difficult there than in the other real space forms. In [11] he asked three basic questions on isoparametric hypersurfaces in spheres. One of this questions was whether isoparametric hypersurfaces in spheres are homogeneous. A negative answer to Cartan's question was only given much later by OZEKI and TAKEUCHI in [42] who found inhomogeneous isoparametric hypersurfaces in spheres. These examples were later generalized by FERUS, KARCHER and MÜNZNER in [24].

I will not try to go further into the rich and beautiful theory of isoparametric hypersurfaces in spheres and refer to [58] for further information. Still I would like to mention the two highlights of the theory after the work of Cartan. The first are the papers [39] and [40] of Münzner where it is shown that the number g of principal curvatures of such a hypersurface can only be 1, 2, 3, 4 or 6. All of these numbers are known to occur. The second is the paper [50] of Stolz in which the possible multiplicities of the principal curvatures are determined. The contributions of Münzner and Stolz are important steps on the way to a full classification of isoparametric hypersurfaces in spheres, which is still an open problem.

2 – Polar actions

In this section we will discuss polar actions. The geometry of their principal orbits will serve us as a motivation in the generalizations of isoparametric hypersurfaces that we will present in the later sections.

Let V be a complete Riemannian manifold and let G be a Lie group acting on V by isometries. One says that the action is *polar* if there is a complete immersed submanifold Σ in V which meets all orbits of G in such a way that all intersections between Σ and orbits are perpendicular. The submanifold Σ is called a *section* of the action. It is rather easy to see that a section is totally geodesic; see [44] and [45], p. 95. The action is called *hyperpolar* if the section is flat.

One should think of a section as a set of canonical forms for the polar action as will be clear in the examples.

Example 2.1.

(i) Any isometric action with a hypersurface as an orbit is polar since a geodesic which meets one orbit orthogonally meets all orbits orthogonally.

(ii) Let V be the linear space $\mathcal{S}_0(n)$ consisting of real symmetric $n \times n$-matrices with zero trace endowed with the scalar product

$$\langle X, Y \rangle = \text{trace}(XY).$$

Let G be the group $\text{SO}(n)$ acting on V by conjugation. We let Σ denote the diagonal matrices in V. Then we know from linear algebra that every matrix X in V can be conjugated into Σ by an element of G. It is now an easy calculation to show that the intersections of conjugacy classes of matrices in V with Σ are all perpendicular. The action is therefore hyperpolar.

(iii) Let V be a compact connected Lie group G with a bi-invariant Riemannian metric acting on itself by conjugation. Let Σ be a maximal torus in G. The theorem on maximal tori says that all conjugacy classes in G meet Σ. An easy calculation shows that the intersections between conjugacy classes in G and Σ are all perpendicular. It follows that the action is hyperpolar since Σ is flat.

(iv) We now show how the examples (ii) and (iii) fit into the theory of symmetric spaces.

A *symmetric space* is a Riemannian manifold V such that for every point p in V there is an isometry σ_p of V fixing p and reversing the directions of the geodesics through p. We refer to the book [33] for what we will need from the theory of symmetric spaces. It is easy to show that symmetric spaces are homogeneous with respect to the isometry group. We can therefore write $V = G/K$ where G is the identity component of the isometry group of V and K is its isotropy group at some fixed point p_0 in V. Such a pair of groups (G, K) is called a *symmetric pair*.

Let Σ be a maximal flat and totally geodesic submanifold passing through p_0 in the symmetric space V. Then the action of K on V is hyperpolar with Σ as a section; see [32]. This example generalizes the one in (ii) since a compact connected Lie group K with a bi-invariant Riemannian metric is a symmetric space with a maximal torus as a maximal flat and totally geodesic submanifold. We can identify K with $K \times K/\Delta(K)$ where $\Delta(K)$ is the diagonal in $K \times K$ and it turns out that conjugation in K corresponds to the action of $\Delta(K)$ on $K \times K/\Delta(K)$.

One can generalize the action of K on the symmetric space $V = G/K$ as follows. Assume that (G, K_1) and (G, K_2) are symmetric pairs. Then one can show that the action of K_1 on $V = G/K_2$ is hyperpolar. This example was introduced by Hermann in [34] and we will refer to it as a *Hermann action*[(1)]. One gets concrete examples of this kind by considering

[(1)] Hermann proved in [34] that his examples are variationally complete and not that

Grassmann manifolds $G_k(\mathbf{C}^n) = \mathrm{SU}(n+1)/K_k$ where K_k is the stabilizer of $\mathbf{C}^k$ in $\mathbf{C}^n$. Then the actions of the groups $K_1, \ldots, K_{n-1}$ on $G_k(\mathbf{C}^n)$ are all hyperpolar.

Now the action of K on V induces an action of K on the tangent space $T_{p_0}V$ which is called the *isotropy representation of the symmetric space* V. This isotropy representation is hyperpolar with $T_{p_0}\Sigma$ as a section. The example in (ii) is a special case and corresponds to the symmetric space $V = G/K$, where $G = \mathrm{SL}(n, \mathbf{R})$ and $K = \mathrm{SO}(n)$. One clearly has the following direct sum decomposition

$$\mathfrak{sl}(n, \mathbf{R}) = \mathfrak{so}(n) \oplus \mathcal{S}_0(n)$$

into skew and symmetric matrices, and this decomposition is invariant under $\mathrm{Ad}_G(K)$. Hence one can identify $T_{p_0}V$ with $\mathcal{S}_0(n)$ and the scalar product on $\mathcal{S}_0(n)$ in (ii) extends to G-invariant Riemannian metric on V. The action in (ii) now corresponds to the isotropy representation of $\mathrm{SL}(n, \mathbf{R})/\mathrm{SO}(n)$.

(v) We finally give an example of a polar action which is not hyperpolar. We let V be the complex projective space $P^n(\mathbf{C})$ endowed with the Fubini-Study metric which is invariant under the action of $\mathrm{SU}(n+1)$. Now let T^n be the maximal torus in $\mathrm{SU}(n+1)$ consisting of diagonal matrices. Then it is not difficult to see that the action of T^n on $P^n(\mathbf{C})$ is polar with $P^n(\mathbf{R})$ as a section. This action is of course not hyperpolar since any two sections of a polar actions are isometric and there can therefore not be a flat section.

The complex projective space $P^n(\mathbf{C})$ with the Fubini-Study metric is an example of a rank one symmetric space. Polar actions on compact rank one symmetric spaces were classified in [46]. It turns out that the sections are always real projective spaces if their dimension is at least two.

The following two theorems show that some of the examples above describe in fact the most general situation. We will need the concept of orbit equivalent actions in the statement of the theorems. Let K_1 act isometrically on V_1 and let K_2 act isometrically on V_2. Then the actions of K_1 and K_2 are said to be *orbit equivalent* if there is an isometry $f : V_1 \to V_2$ such that $f(K_1 p) = K_2 f(p)$ for all p in V_1, i.e., the orbits of K_1 and K_2 correspond under f.

THEOREM 2.2 (Dadok [17]). *Let K be a compact group acting in a polar fashion on a Euclidean space V. Then the action of K is orbit equivalent to the isotropy representation of some symmetric space.*

they are hyperpolar. The relationship between the two concepts will be explained at the end of this section.

The *cohomogeneity* of an action is the minimal codimension of its orbits. Eschenburg and Heintze gave in [21] a proof of Dadok's theorem under the assumption that the cohomogeneity is at least three. Their proof does not use the classification of compact Lie groups. Lists of polar representations that are not isotropy representations of symmetric spaces can be found in [4], [20], and [25].

THEOREM 2.3 (Kollross [36]). *Let $V = G/K$ be a compact irreducible symmetric space and let H be a subgroup of G which acts in a hyperpolar fashion on V with cohomogeneity at least two. Then the action of H on V is orbit equivalent to a Hermann action.*

Kollross also classifies in [36] all cohomogeneity one actions on compact irreducible symmetric spaces V. The classification of such actions on spheres was already carried out in [35].

We now discuss the principal orbits of polar and hyperpolar actions from the point of view of submanifold geometry. This will serve as a motivation for the generalizations of isoparametric hypersurfaces in the later sections.

Let G be a Lie group acting by isometries on a Riemannian manifold V. A *principal orbit* of the action of G on a manifold V is by definition an orbit Gp with the property that there is a neighborhood U of p such that there is a G-equivariant map from Gp to Gq for all q in U. If Gp is principal, then p is said to be *regular*. The set of regular points is open and dense in V. Now let ξ_0 be an element of $\nu_p(Gp)$ where $\nu(Gp)$ denotes the normal bundle of Gp. Then $\xi(gp) = dg_p(\xi_0)$ is a well defined normal vector field if Gp is principal. We call such a normal vector field *equivariant*.

For a proof of the following proposition, see [45], p. 95-96, or [5], p. 44.

PROPOSITION 2.4. *Assume that the action of G on V is polar. Then the equivariant normal vector fields along a principal orbit Gp are parallel. In particular, the normal bundle is flat and has trivial normal holonomy.*

The next property of the principal orbits of polar actions that we would like to present has to do with focal points. Let M be a submanifold of the Riemannian manifold V and assume that γ is a geodesic that starts in M, i.e. $\gamma(0)$ lies in M, and that $\gamma'(0)$ is perpendicular to M. Suppose $\gamma_s(t)$ is a smooth variation of $\gamma = \gamma_0$ such that $\gamma_s(0) \in M$ and $\gamma_s'(0)$ is perpendicular to M for all s. Now let J be the variational vector field

$$J(t) = \left.\frac{\partial}{\partial s}\right|_{s=0} \gamma_s(t)$$

of γ_s. We call such a variational vector field an *M-Jacobi field along γ*. One can show that the M-Jacobi fields along γ form a vector space. A point $\gamma(t_0)$

is called a *focal point of M along γ* if there is a nonvanishing M-Jacobi field J with $J(t_0) = 0$. The dimension of the space of M-Jacobi fields vanishing in t_0 is called the *multiplicity* of the focal point $\gamma(t_0)$.

PROPOSITION 2.5. *Assume that the action of G on V is polar and let M be a principal orbit of G. Let ξ be a parallel normal field along M. Then the distances to the focal points and their multiplicities along the geodesic starting in direction $\xi(p)$ does not depend on p.*

If V is a Euclidean space then the focal points of M are determined by the principal curvatures. Let ξ be a normal vector field along M and X a tangent vector of M at p. We let $D_X\xi$ denote the directional derivative of ξ in direction X and denote the tangent component of $-D_X\xi$ by $A_\xi(X)$. It turns out that the map $A_\xi : T_pM \to T_pM$ that sends X to $A_\xi(X)$ is a selfadjoint linear endomorphism that depends only on the value of ξ at p. One calls A_ξ the *shape* (or *Weingarten*) *operator* of M in direction ξ_p. The eigenvalues of A_ξ are called the *principal curvatures of M in direction ξ_p*.

Now if ξ_p is a normal vector of M at p and λ is a nonvanishing principal curvature in direction ξ_p, then $p + (1/\lambda)\xi_p$ is a focal point of M along the line $\gamma(t) = p+t\xi_p$. Conversely if $p+(1/\lambda)\xi_p$ is a focal point of M along $\gamma(t) = p+t\xi_p$, then λ is a principal curvature in direction ξ_p.

We can therefore reformulate Proposition 2.5 as follows if the ambient space is Euclidean. Notice that a polar action on a Euclidean space is hyperpolar since the sections are affine subspaces.

PROPOSITION 2.6. *Let V be a Euclidean space on which a Lie group acts in a polar fashion. Let M be a principal orbit of G and let ξ be a parallel normal field along M. Then the principal curvatures in direction ξ_p do not depend on p.*

Before we end this section we would like to mention two classes of actions that are closely related to hyperpolar actions.

Variationally complete actions were introduced by BOTT in [6]; see also [7]. By definition an isometric action of a Lie group G on a Riemannian manifold V is called *variationally complete* if the following holds for all orbits M of G: let J be an M-Jacobi field along $\gamma(t)$ which vanishes at some point t_0. Then J is the variational vector field of a variation of the type $\phi_s(\gamma(t))$ where ϕ_s is a one-parameter subgroup of G. In other words, J is the restriction of a Killing field induced by the action of G to γ.

Conlon proved in [16] that a hyperpolar action on a complete Riemannian manifold is variationally complete. A partial converse was proved in [27]: a variationally complete action on a compact symmetric space is hyperpolar. It was previously proved by DI SCALA and OLMOS in [18], see also [25], that variationally complete representations are polar. Lytchak has conjectured that vari-

ationally complete actions on compact Riemannian manifolds with nonnegative sectional curvature are hyperpolar.

Variationally complete actions were introduced in [6] and [7] to study the Morse theory of geodesics on complete Riemannian manifolds and in particular on compact symmetric spaces. We next briefly review one of the main results of these papers.

Let M be a properly embedded submanifold of a Riemannian manifold V and p some point in V. We let $\mathcal{P} = \mathcal{P}(V, p \times M)$ denote the space of absolutely continuous paths $\gamma : [0,1] \to V$ that start in p and end in M and for which the so-called *energy*

$$E(\gamma) = \int_0^1 \| \gamma'(t) \|^2 \, dt$$

is finite. Then $\mathcal{P}$ is in a natural way a Hilbert manifold and E is a smooth functional on $\mathcal{P}$, see [43], whose critical points are the geodesics starting in p and meeting M perpendicularly. If p is not a focal point of M, then the energy functional E is a Morse function in the sense that it has only nondegenerate critical points. We say that the submanifold M is *taut* if the energy functional is perfect, meaning that the number of critical points of index k of E in $\mathcal{P}$ is equal to the k-th Betti number of $\mathcal{P}$ with respect to $\mathbf{Z}_2$-coefficients, or equivalently, that the $\mathbf{Z}_2$-Morse inequalities of E on $\mathcal{P}$ are equalities; see [56]. An isometric action is called *taut* if all of its orbits are taut.

One of the main theorems of Bott and Samelson in [7] can now be phrased in our terminology by saying that variationally complete actions are taut.

A taut action does not have to be variationally complete. It is proved in [25], [26], and [28] that there are precisely three irreducible taut representations of compact groups which are not variationally complete. These three representations happen to be precisely the cohomogeneity three representations which are not variationally complete.

3 – Isoparametric submanifolds of Euclidean spaces

Isoparametric submanifolds in Euclidean spaces with higher codimension were first introduced by Harle in [29]. Carter and West independently introduced and studied such submanifolds with codimension three in [13] and [14]. Terng then dealt with the case of general codimension in [52].

According to [52] a complete and connected submanifold M^n in $\mathbf{R}^{n+k}$ is called *isoparametric* if its normal bundle is flat and if the principal curvatures in the direction of any (locally defined) parallel normal vector field are constant. It is proved in [52] that the normal holonomy of M^n is trivial. A locally defined parallel normal curvature vector can therefore be extended to a globally defined one.

It is proved in [52] that a noncompact isoparametric submanifold is the product embedding of a compact isoparametric one with a Euclidean space. We will therefore always assume compactness in the following. A compact isoparametric submanifold is contained in a round hypersphere; see [52]. We can always assume that M^n is not contained in any proper affine subspace. Such submanifolds are called *full*. An isoparametric submanifold is said to be *irreducible* if it cannot be nontrivially written as the product embedding of two isoparametric submanifolds.

Propositions 2.4 and 2.6 imply that principal orbits of polar representations are isoparametric. Conversely, Palais and Terng proved in [44] that a homogeneous isoparametric submanifold is such an orbit. One can show that an isoparametric hypersurface M^n in S^{n+1} is isoparametric in $\mathbf{R}^{n+2}$. The inhomogeneous examples of Ferus, Karcher and Münzner in [24] that we already mentioned in the introduction therefore give us a further class of examples. All known examples of irreducible isoparametric submanifolds in Euclidean spaces belong to one of these two classes of examples.

Terng developed a beautiful structure theory of isoparametric hypersurfaces in [52]. We would like to mention some of these results since they have been a paradigm in the generalizations.

Let M^n be an isoparametric submanifold in $\mathbf{R}^{n+k}$ and let ξ be a parallel normal field along M^n. The *end-point map in direction* ξ is the map $\eta_\xi : M^n \to \mathbf{R}^{n+k}$ one gets by setting $\eta_\xi(p) = p + \xi_p$. It turns out that the image of M^n under η_ξ that we denote by M_ξ is either a submanifold of dimension n or one of a lower dimension. We call M_ξ the *parallel submanifold of* M^n *in direction* ξ. If the dimension of M_ξ is equal to that of M^n, then M_ξ is also isoparametric and η_ξ is a diffeomorphism between M^n and M_ξ. If the dimension of M_ξ is smaller than that of M^n, then M_ξ consists of focal points of M^n and η_ξ is a fibration from M^n onto M_ξ. In this case we will call M_ξ a *focal submanifold*. One can show that the set F of focal points of M^n is precisely the union over the focal submanifolds of M^n.

It is easy to see with help of Proposition 2.4 that if M^n is homogeneous and hence a principal orbit of a polar representation, then the parallel submanifolds are nothing but the other orbits of the representation.

If M^n is isoparametric, then $\mathcal{F} = \{M_\xi \,|\, \xi \text{ parallel along } M^n\}$ is a family of disjoint submanifolds that covers the whole ambient space $\mathbf{R}^{n+1}$. It is not difficult to show that the isoparametric submanifolds in $\mathcal{F}$ foliate $\mathbf{R}^{n+1} \setminus F$, the complement of the focal set F of M^n. One can in fact show much more than this: $\mathcal{F}$ is a singular Riemannian foliation in the sense of Molino. This is a consequence of a much more general result of Töben in [59] that we will explain in the last section; see also [45], Corollary 8.5.6.

Terng associated in [52] a Coxeter group to an isoparametric submanifold M as follows. Let $\nu_p M$ be the normal space of M at p considered as an affine subspace of $\mathbf{R}^{n+k}$ and consider the set $F_p = F \cap \nu_p M$ of focal points contained

in $\nu_p M$. Then F_p is a finite union over hyperplanes in $\nu_p M$ and the reflections in this hyperplanes generate a finite Coxeter group W that leaves the set F_p invariant. It then follows that the orbit of p under W is the intersection $M \cap \nu_p M$. The Coxeter group is implicit in Cartan's work for the codimension two case $M^n \subset S^{n+1} \subset \mathbf{R}^{n+1}$ since he proved that the focal points on the normal great circles to M^n in S^{n+1} are equidistant. In the codimension three case the Coxeter group was already found by CARTER and WEST in [13].

The following theorem proved in [57] shows that isoparametric submanifolds come close to characterize principal orbits of polar representations.

THEOREM 3.1. *Let M^n be an irreducible, full and compact isoparametric submanifold in M^{n+k} with $k \geq 3$. Then M^n is a principal orbit of a polar representation.*

Theorem 3.1 combined with Dadok's Theorem 2.2 gives a classification of irreducible isoparametric submanifolds with codimension at least three. The examples of Ferus, Karcher and Münzner are inhomogeneous with codimension $k = 2$. If the codimension is $k = 1$, then the round spheres are the only examples.

A new proof of Theorem 3.1 was given by OLMOS in [41] using his theory of the normal holonomy of submanifolds; see also [5], Section 7.3. A further proof was given by HEINTZE and LIU in [30] as a byproduct of a proof of an analogous theorem in Hilbert spaces that will play a role in the next section. Eschenburg gave a proof of the theorem in [19] that uses Lie triple products.

4 – Equifocal submanifolds

In the paper [55], equifocal submanifolds of compact symmetric spaces were introduced and their basic theory developed as a generalization of isoparametric hypersurfaces in spheres and an analogue of the isoparametric submanifolds in Euclidean spaces. For symmetric spaces see reference [33] and the remarks in Example 2.1 (iv) above.

The definition of an equifocal submanifold is based on the properties of principal orbits of polar actions in Propositions 2.4 and 2.5.

Let M^n be a compact submanifold of a compact symmetric space V^{n+k}. We say that M^n is *equifocal* if the following three conditions are satisfied:

(i) The normal bundle of M^n is flat and has trivial holonomy.
(ii) If ξ is a parallel normal vector field and $\eta_\xi(p_0) = \exp(\xi(p_0))$ is a multiplicity k focal point of M^n for some p_0 in M^n, then $\eta_\xi(p) = \exp(\xi(p))$ is a multiplicity k focal point of M^n for all p in M^n. (In other words, the focal data of M^n are invariant under normal parallel translation.)
(iii) The image $\exp(\nu_p(M^n))$ of the normal space $\nu_p(M^n)$ of M^n at p is contained in some flat of V^{n+k} for all p in M^n.

Principal orbits of polar actions satisfy conditions (i) and (ii) in the definition above, and all three conditions are satisfied for principal orbits of hyperpolar actions.

The third condition is of course always satisfies if M^n is a hypersurface. It follows from [31] that a hypersurface M^n in a compact symmetric space V^{n+1} is equifocal if and only if it is isoparametric in the sense of the definition given at the beginning of this paper. One can of course define equifocal hypersurfaces in more general ambient spaces than symmetric spaces; see the next section. If the ambient space has nonpositive sectional curvature one should take into account that there might be focal points 'beyond infinity'; see [22]. It is not to be expected that such generalizations are equivalent to the concept of an isoparametric hypersurface if the ambient space is not symmetric.

In [55] we show that if the compact symmetric space V^{n+1} is irreducible, then an equifocal hypersurface M^n in V^{n+1} has the property that any geodesic meeting M^n is closed. If V^{n+1} is simply connected, then one can show that the number of focal points on such a normal closed geodesic is an even number that we will denote by $2g$. If V^{n+1} is a sphere, then g is the number of principal curvatures of M^n which can only be one of the numbers 1, 2, 3, 4, and 6 as was proved by Münzner; see the introduction. One can now ask which values g can assume in general irreducible symmetric spaces, and what the possible values of the multiplicities of the focal points are; see [51] and [23] where this question is studied for rank one and two symmetric spaces.

One can prove more generally that the image of a normal space $\nu_p(M)$ of an equifocal submanifold M^n in an irreducible compact symmetric space V^{n+k} is a torus T^k; see [55]. One can associate to M^n an affine Coxeter group as follows. Let F denote the set of focal points of M^n in T^k and let $\tilde{F}$ be the preimage of F in the universal cover $\mathbf{R}^k$ of T^k. Then $\tilde{F}$ is a union of hyperplanes which are precisely the mirrors of an affine Coxeter group W acting on $\mathbf{R}^k$ and leaving $\tilde{F}$ invariant.

The next theorem which is analogous to Theorem 3.1 gives a characterization of the principal orbits of hyperpolar actions as equifocal submanifolds.

THEOREM 4.1 (Christ [15]). *Let M^n be an equifocal submanifold in an irreducible compact symmetric space V^{n+k}. Then M^n is the principal orbit of a hyperpolar action if $k \geq 2$.*

The theorem does not hold for $k = 1$ since the inhomogeneous isoparametric hypersurfaces in spheres are equifocal.

Theorem 4.1 and the results from [55] that we have been explaining are proved with the help of a generalization due to Terng of the theory of isoparametric submanifolds in Euclidean spaces to Hilbert spaces; see [53]. We end this section with a short explanation of this method.

Let V^{n+k} be a compact symmetric space that we write as a coset space $V^{n+k} = G/K$ where (G, K) is a symmetric pair. Let $\mathfrak{g}$ denote the Lie algebra of G and set $\mathcal{H} = L^2([0,1], \mathfrak{g})$, the Hilbert space of L^2-paths in $\mathfrak{g}$. Then there is a Riemannian submersion $\phi : \mathcal{H} \to V^{n+k}$ such that a submanifold M^n is equifocal in V^{n+k} if and only if the preimage $\mathcal{M} = \phi^{-1}(M^n)$ is isoparametric in $\mathcal{H}$. The main point is that it is easier to work in $\mathcal{H}$ than in V^{n+1} since $\mathcal{H}$ is linear, although infinite dimensional.

To define the Riemannian submersion ϕ, we need to introduce certain path spaces in G. Let B be a subset of $G \times G$ and let $\mathcal{P}(G, B)$ denote the space of absolutely continuous paths $\gamma : [0,1] \to G$ such that $(\gamma(0), \gamma(1)) \in B$ and such that the integral $\| \gamma' \|^2$ is finite. Here we assume G to be endowed with a bi-invariant Riemannian metric such that the projection $\pi : G \to V^{n+k}$ is a Riemannian submersion. Then $\mathcal{P}_e = \mathcal{P}(G, e \times G)$ is the space of paths in G starting at the identity e without a restriction on the end point.

Now it turns out that the map that sends a path γ in $\mathcal{P}_e$ to $\gamma^{-1}\gamma'$ in $\mathcal{H}$ is a diffeomorphism. Let $E : \mathcal{H} \to \mathcal{P}_e$ denote the inverse of this diffeomorphism. Now we can define a map $\psi : \mathcal{H} \to G$ by setting $\psi(u)$ equal to the endpoint of the curve $E(u)$, i.e. $\psi(u) = E(u)(1)$. It is proved in [55] that ψ is a Riemannian submersion. Now we define $\phi : \mathcal{H} \to V^{n+k}$ as $\phi = \pi \circ \psi$.

If H is a subgroup of G then $\mathcal{P}(G, H \times K)$ is an infinite dimensional Hilbert Lie group which acts on $\mathcal{H}$ by setting

$$\gamma * u = \gamma u \gamma^{-1} - \gamma' \gamma^{-1}$$

for γ in $\mathcal{P}(G, H \times K)$ and u in $\mathcal{H}$; see [54] where it is proved that the action of H on V^{n+k} is hyperpolar if and only if the action of $\mathcal{P}(G, H \times K)$ is polar on $\mathcal{H}$. It is also proved in [54] that the principal orbits of $\mathcal{P}(G, H \times K)$ are isoparametric if its action on $\mathcal{H}$ is polar.

A very important result of Heintze and Liu in [30] is that an irreducible isoparametric submanifold in an infinite dimensional Hilbert space is the principal orbit of a polar action if its codimension is at least two. This result of Heintze and Liu is one of the main steps in the proof of Theorem 4.1. The method of proof also works in finite dimensions if the codimension is at least three and can be used to prove Theorem 3.1.

One can also use the Hilbert space $\mathcal{H}$ to prove that an action on a compact symmetric space is hyperpolar if it is variationally complete; see [27] and Section 2. One shows that the action of a subgroup H of G is variationally complete (resp. hyperpolar) on V^{n+k} if and only if the action of $\mathcal{P}(G, H \times K)$ on $\mathcal{H}$ is variationally complete (resp. hyperpolar). One has now reduced the problem to an affine action on the linear space $\mathcal{H}$ and can argue in a similar way as as in [18].

5 – Submanifolds in Riemannian manifolds

In this last section we would like to mention some recent generalizations to Riemannian manifolds.

The orbits of a connected Lie group acting by isometries on a Riemannian manifold give an example of a singular Riemannian foliation in the sense of Molino; see [38], p. 189. By definition, a partition $\mathcal{F}$ of a Riemannian manifold V into connected immersed submanifolds, called *leaves*, is said to be a *singular Riemannian foliation* if the following two conditions are satisfied:

(i) The tangent space T_pM for every M in $\mathcal{F}$ and every p in M is generated by $\{X_p |\ X \in \Xi_{\mathcal{F}}\}$ where $\Xi_{\mathcal{F}}$ denotes the module of smooth vector fields on V that are tangent to the submanifolds in $\mathcal{F}$.
(ii) A geodesic that meets one leaf M in $\mathcal{F}$ perpendicularly, meets the leaves perpendicularly for all parameter values.

The leaves in $\mathcal{F}$ of maximal dimension are called *regular* and those of lower dimension *singular*.

If only the first condition is satisfied then one calls $\mathcal{F}$ a singular foliation. A singular foliation is a foliation in the usual sense if the leaves are all regular. The second condition means that the leaves are equidistant.

If $\mathcal{F}$ consists of the orbits of an action, then condition (ii) is satisfied since the vector fields it induces are contained in $\Xi_{\mathcal{F}}$ and condition (ii) is satisfied if the action is isometric.

Alexandrino studies singular Riemannian foliations that admit a section in [2], where a section is defined as for polar actions. Previously such foliations were studied by Boualem in [8]. Let $\mathcal{F}$ be such a singular foliation in a Riemannian manifold V, let L be a singular leaf in $\mathcal{F}$, and let T be a tubular neighborhood of L that is a union over leaves in $\mathcal{F}$. It is then proved in [2] that the foliation consisting of the intersections of the leaves of such a foliation $\mathcal{F}$ with the connected component of $\exp(\nu_p(L)) \cap T$ containing p is diffeomorphic to an isoparametric foliation in an open neighborhood of 0 in $\mathbf{R}^k$ where k is the codimension of L in V. This generalizes the slice theorems for polar actions and isoparametric submanifolds; see [45]. A further result of [2] is that the regular leaves of singular Riemannian foliations with a section have parallel focal structure, see also [59] for a different proof. Submanifolds with parallel focal structure were studied by Ewert in [22]. They generalize equifocal submanifolds in a similar way as polar actions generalize hyperpolar actions, see [59] for a precise definition. Töben gives in [59] a necessary and sufficient condition for a submanifold with parallel focal structure and finite normal holonomy to give rise to a singular Riemannian foliation with the leaves being parallel submanifolds. In [59] an action on the sections of a singular Riemannian foliations by a group called transversal holonomy group is introduced. This action generalizes the Weyl group action of polar actions.

In [1] Alexandrino studies transnormal maps. These are by definition maps from a Riemannian manifold into a Euclidean space with the property that its restrictions to sufficiently small neighborhoods of regular level sets are Riemannian submersions such that the normal spaces of the fibers form an integrable distribution on the neighborhood. The main result of [2] is that the level sets of an analytic transnormal map on a real analytic Riemannian manifold give rise to a singular Riemannian foliation with sections.

Further results along these lines can be found in [3].

REFERENCES

[1] M. Alexandrino: *Integrable Riemannian submersion with singularities*, to appear in Geometriae Dedicata.

[2] M. Alexandrino: *Singular Riemannian foliations with sections*, E-print math. DG/0311454, to appear in Illinois J. Math.

[3] M. Alexandrino – D. Töben: *Singular Riemannian foliations on simply connected spaces*, preprint 2004.

[4] I. Bergmann: *Reducible polar representations*, Manuscripta Math., **104** (2001), 309-324.

[5] J. Berndt – S. Console – C. Olmos: *Submanifolds and holonomy*, Research Notes in Mathematics, Vol. 434, Chapman & Hall/CRC, Boca Raton, 2003.

[6] R. Bott: *An application of the Morse theory to the topology of Lie-groups*, Bull. Soc. Math. France, **84** (1956), 251-281.

[7] R. Bott – H. Samelson: *Applications of the theory of Morse to symmetric spaces*, Amer. J. Math., **80** (1958), 964-1029, correction in Amer. J. Math., **83** (1961), 207-208.

[8] H. Boualem: *Feuilletages riemanniens singuliers transversalement intégrables*, Compositio Math., **95** (1995), 101-125.

[9] É. Cartan: *Familles de surfaces isoparamétriques dans les espaces à courbure constante*, Ann. di Mat., **17** (1938), 177-191.

[10] É. Cartan: *Sur des familles remarquables d'hypersurfaces isoparamétriques dans les espaces sphériques*, Math. Z., **45** (1939), 335-367.

[11] É. Cartan: *Sur quelques familles remarquables d'hypersurfaces*, C. R. Congrès Math. Liège, (1939), 30-41.

[12] É. Cartan: *Sur des familles d'hypersurfaces isoparamétriques des espaces sphériques à 5 et à 9 dimensions*, Revista Univ. Tucumán, **1** (1940), 5-22.

[13] S. Carter – A. West: *Generalised Cartan polynomials*, J. London Math. Soc., **32** (1985), 305-316.

[14] S. Carter – A. West: *Isoparametric systems and transnormality*, Proc. London Math. Soc., **51** (1985), 520-542.

[15] U. Christ: *Homogeneity of equifocal submanifolds*, J. Differential Geom., **62** (2002), 1-15.

[16] L. Conlon: *Variational completeness and K-transversal domains*, J. Differential Geom., **5** (1971), 135-147.

[17] J. Dadok: *Polar coordinates induced by actions of compact Lie groups*, Trans. Amer. Math. Soc., **288** (1985), 125-137.

[18] A. J. Di Scala – C. Olmos: *Variationally complete representations are polar*, Proc. Amer. Math. Soc., **129** (2001), 3445-3446.

[19] J.-H. Eschenburg: *Isoparametric submanifolds and symmetric spaces*, Geometry and topology: Aarhus (1998), Contemp. Math., Vol. 258, Amer. Math. Soc., Providence, RI, 2000, pp. 145-155.

[20] J.-H. Eschenburg – E. Heintze: *On the classification of polar representations*, Math. Z., **232** (1999), 391-398.

[21] J.-H. Eschenburg – E. Heintze: *Polar representations and symmetric spaces*, J. Reine. Angew. Math., **507** (1999), 93-106.

[22] H. Ewert: *Equifocal submanifolds in Riemannian symmetric spaces*, Ph.D. thesis, University of Cologne, 1998.

[23] F. Fang: *Equifocal hypersurfaces in symmetric spaces*, Chinese Ann. Math. Ser. B, **21** (2000), 473-478.

[24] D. Ferus – H. Karcher – H. F. Münzner: *Cliffordalgebren und neue isoparametrische Hyperflächen*, Math. Z., **177** (1981), 479-502.

[25] C. Gorodski – G. Thorbergsson: *Representations of compact Lie groups and the osculating spaces of their orbits*, preprint, University of Cologne, 2000 (also E-print math. DG/0203196).

[26] C. Gorodski – G. Thorbergsson: *Cycles of Bott-Samelson type for taut representations*, Ann. Global Anal. Geom., **21** (2002), 287-302.

[27] C. Gorodski – G. Thorbergsson: *Variationally complete actions on compact symmetric spaces*, J. Differential Geom., **62** (2002), 39-48.

[28] C. Gorodski – G. Thorbergsson: *The classification of taut irreducible representations*, J. Reine Angew. Math., **555** (2003), 187-235.

[29] C. E. Harle: *Isoparametric families of submanifolds*, Bol. Soc. Brasil. Mat., **13** (1982), 35-48.

[30] E. Heintze – X. Liu: *Homogeneity of infinite-dimensional isoparametric submanifolds*, Ann. of Math., **149** (1999), 149-181.

[31] E. Heintze – X. Liu – C. Olmos: *Isoparametric submanifolds and a Chevalley-type restriction theorem*, E-print math. DG/0004028.

[32] E. Heintze – R. S. Palais – C.-L. Terng – G. Thorbergsson: *Hyperpolar actions on symmetric spaces*, Geometry, Topology, and Physics for Raoul Bott, S. T. Yau (eds.), Conf. Proc. Lecture Notes Geom. Topology, IV, International Press, Cambridge, MA, 1995, pp. 214-245.

[33] S. Helgason: *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics, Vol. 80, Academic Press, New York-London, 1978.

[34] R. Hermann: *Variational completeness for compact symmetric spaces*, Proc. Amer. Math. Soc., **11** (1960), 544-546.

[35] W.-Y. Hsiang – H. B. Lawson Jr.: *Minimal submanifolds of low cohomogeneity*, J. Differential Geometry, **5** (1971), 1-38.

[36] A. Kollross: *A classification of hyperpolar and cohomogeneity one actions*, Trans. Amer. Math. Soc., **354** (2002), 571-612.

[37] T. Levi-Civita: *Famiglie di superficie isoparametriche nell'ordinario spazio euclideo*, Rend. Acc. Naz. Lincei, **XXVI** (1937), 657-664.

[38] P. Molino: *Riemannian foliations*, Progress in Mathematics, Vol. 73, Birkhäuser, Boston-Basel, 1988.

[39] H. F. Münzner: *Isoparametrische Hyperflächen in Sphären,* I, Math. Ann., **251** (1980), 57-71.

[40] H. F. Münzner: *Isoparametrische Hyperflächen in Sphären,* II, Math. Ann., **256** (1981), 215-232.

[41] C. Olmos: *Isoparametric submanifolds and their homogeneous structures*, J. Differential Geom., **38** (1993), 225-234.

[42] H. Ozeki – M. Takeuchi: *On some types of isoparametric hypersurfaces in spheres, i*, Tôhoko Math. J., **27** (1975), 515-559.

[43] R. Palais: *Morse theory on Hilbert manifolds*, Topology, **2** (1963), 299-340.

[44] R. Palais – C.-L. Terng: *A general theory of canonical forms*, Trans. Amer. Math. Soc., **300** (1987), 771-789.

[45] R. Palais – C.-L. Terng: *Critical point theory and submanifold geometry*, Lecture Notes in Mathematics, Vol. 1353, Springer-Verlag, Berlin, Heidelberg, New York, 1988.

[46] P. Podestà – G. Thorbergsson: *Polar actions on rank-one symmetric spaces*, J. Differential Geom., **53** (1999), 131-175.

[47] B. Segre: *Una proprietà caratteristica di tre sistemi* ∞^1 *di superficie*, Atti Acc. Sc. Torino, **LIX** (1924), 666-671.

[48] B. Segre: *Famiglie di ipersuperficie isoparametriche negli spazi euclidei ad un qualunque numero di dimensioni*, Rend. Acc. Naz. Lincei, **XXVII** (1938), 203-207.

[49] C. Somigliana: *Sulle relazione fra il principio di Huygens e l'ottica geometrica*, Atti Acc. Sc. Torino, **LIV** (1918-1919), 974-979.

[50] S. Stolz: *Multiplicities of Dupin hypersurfaces*, Invent. Math., **138** (1999), 253-279.

[51] Z. Tang: *Multiplicities of equifocal hypersurfaces in symmetric spaces*, Asian J. Math., **2** (1998), 181-213.

[52] C.-L. Terng: *Isoparametric submanifolds and their Coxeter groups*, J. Differential Geom., **21** (1985), 79-107.

[53] C.-L. Terng: *Proper Fredholm submanifolds of Hilbert space*, J. Differential Geom., **29** (1989), 9-47.

[54] C.-L. Terng: *Polar actions on Hilbert space*, J. Geom. Anal., **5** (1995), 129-150.

[55] C.-L. Terng – G. Thorbergsson: *Submanifold geometry in symmetric spaces*, J. Differential Geom., **42** (1995), 665-718.

[56] C.-L. Terng – G. Thorbergsson: *Taut immersions into complete Riemannian manifolds*, Tight and taut submanifolds (Berkeley, CA, 1994), T. E. Cecil, S. S. Chern (eds.), Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, Cambridge, 1997, pp. 181-228.

[57] G. Thorbergsson: *Isoparametric foliations and their buildings*, Ann. of Math., **133** (1991), 429-446.

[58] G. Thorbergsson: *A survey on isoparametric hypersurfaces and their generalizations*, Handbook of differential geometry, Vol. I, F. Dillen, L. Verstraelen (eds.), North-Holland, Amsterdam, 2000, pp. 963-995.

[59] D. Töben: *Parallel focal structure and singular Riemannian foliations*, E-print math. DG/0403050, to appear in Trans. Amer. Math. Soc.

[60] Q. M. Wang: *Isoparametric hypersurfaces in complex projective spaces*, Proceedings of the 1980 Beijing Symposium on Differential Geometry and Differential Equations, Vol. 1, 2, 3 (Beijing, 1980), Science Press, Beijing, 1982, pp. 1509-1523.

Lavoro pervenuto alla redazione il 14 ottobre 2004
ed accettato per la pubblicazione il 26 novembre 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo dell'Autore:

Gudlaugur Thorbergsson – Mathematisches Institut der Universität zu Köln – Weyertal 86-90 – 50931 Köln, Germany
E-mail: gthorber@mi.uni-koeln.de

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 17-32

Laurent type expansions of $\bar\partial$-closed (0, n − 1)-forms in $\mathbb{C}^n$

TELEMACHOS HATZIAFRATIS

ABSTRACT: *We characterize the multiple sequences $\varpi_{k_1,\dots,k_n}$ of complex numbers for which there exist $\bar\partial$-closed $(0,n-1)$-forms $\theta(\zeta)$, defined for $\zeta\in\mathbb{C}^n-\{|\zeta|\le R\}$, so that $\int_{|\zeta|=\rho}\zeta_1^{k_1}\cdots\zeta_n^{k_n}\theta(\zeta)\wedge d\zeta_1\wedge\ldots\wedge d\zeta_n=\varpi_{k_1,\dots,k_n}$ $(\rho>R)$. We also derive Laurent type expansions of such $\bar\partial$-closed $(0,n-1)$-forms in terms of the derivatives of the Bochner-Martinelli kernel and we discuss Mittag-Leffler type constructions in this setting.*

1 – Introduction

Let us recall that given a sequence ϖ_k, $k=0,1,2,\dots$, of complex numbers, there exists a holomorphic function $g(\zeta)$ defined for $\zeta\in\mathbb{C}-\{|\zeta|\le R\}$ (where $R\ge 0$) so that

$$\int\limits_{|\zeta|=\rho}\zeta^k g(\zeta)d\zeta=\varpi_k,\quad k=0,1,2,\dots\ (\rho>R),$$

if and only if

$$\limsup_{k\to\infty}\sqrt[k]{|\varpi_k|}\le R,$$

and that, moreover, such a function is of the form

$$g(\zeta)=\frac{1}{2\pi i}\sum_{k=0}^{\infty}\varpi_k\frac{1}{\zeta^{k+1}}+\ a\ holomorphic\ function\ in\ \mathbb{C}.$$

KEY WORDS AND PHRASES: *Laurent type expansions – $\bar\partial$-closed $(0,n-1)$-forms – Fourier-Laplace transform – Derivatives of the Bochner-Martinelli kernel – Mittag-Leffler type constructions.*
A.M.S. CLASSIFICATION: 32A26

In $\mathbb{C}^n$, we may consider systems $(g_1, \dots, g_n)$ of C^∞ functions, which satisfy the differential equation

$$\sum_{j=1}^{n}(-1)^{j-1}\frac{\partial g_j}{\partial \bar{\zeta}_j} = 0$$

(equivalently: the $(0, n-1)$-form

$$\theta = \sum_{j=1}^{n} g_j d\bar{\zeta}_1 \wedge \dots (j) \dots \wedge d\bar{\zeta}_n$$

is $\bar{\partial}$-closed), and pose an analogous question as follows: For which multiple sequences $\varpi_{k_1,\dots,k_n}$ of complex numbers, do there exist $\bar{\partial}$-closed $(0, n-1)$-forms $\theta(\zeta)$, defined for $\zeta \in \mathbb{C}^n - \{|\zeta| \le R\}$, so that

$$\int\limits_{|\zeta|=\rho} \zeta_1^{k_1} \cdots \zeta_n^{k_n}\theta(\zeta) \wedge d\zeta_1 \wedge \dots \wedge d\zeta_n = \varpi_{k_1,\dots,k_n} \quad (\rho > R)?$$

It turns out that we can characterize such sequences (see Theorem 2) and furthermore we can give an analogous expansion for these $\bar{\partial}$-closed $(0, n-1)$-forms θ, in terms of appropriate derivatives of the Bochner-Martinelli kernel (see Theorem 3). For background material, we refer to [2], [3], [4], and [7].

NOTATION. If D is an open subset of $\mathbb{C}^n$, we will denote by $Z_{\bar{\partial}}^{(0,n-1)}(D)$ the set of $\bar{\partial}$-closed $(0, n-1)$-forms with C^∞ coefficients in D and $H_{\bar{\partial}}^{(0,n-1)}(D)$ will denote the set of the corresponding $\bar{\partial}$-cohomology classes in D:

$$H_{\bar{\partial}}^{(0,n-1)}(D) = \{[\theta] : \theta \in Z_{\bar{\partial}}^{(0,n-1)}(D)\},$$

where $[\theta] = \{\theta + \bar{\partial} - exact\ (0, n-1) - forms\ in\ D\}$.

Also $\mathcal{O}(D)$ will denote the set of holomorphic functions in D.

2 – Fourier-Laplace transforms of $\bar{\partial}$-closed $(0, n-1)$-forms

Let E be a compact convex set in $\mathbb{C}^n$ and let $\xi \in Z_{\bar{\partial}}^{(0,n-1)}(\mathbb{C}^n - E)$. The Fourier-Laplace transform of ξ is the entire holomorphic function F_ξ defined by the integral

$$F_\xi(w) = \int\limits_{\zeta\in\partial U} e^{\langle\zeta,w\rangle}\xi(\zeta) \wedge \omega(\zeta), \quad w \in \mathbb{C}^n,$$

where $\langle\zeta, w\rangle = \sum \zeta_j w_j$, $\omega(\zeta) = d\zeta_1 \wedge \dots \wedge d\zeta_n$ and U is an open and bounded convex set with smooth boundary which contains E. Since the differential form

$e^{\langle\zeta,w\rangle}\xi(\zeta)\wedge\omega(\zeta)$ is $d-$closed, it follows from Stokes' theorem that the above integral is independent of the choice of U. Indeed, if V is a sufficiently large ball, then

$$\int\limits_{\zeta\in\partial V} e^{\langle\zeta,w\rangle}\xi(\zeta)\wedge\omega(\zeta) - \int\limits_{\zeta\in\partial U} e^{\langle\zeta,w\rangle}\xi(\zeta)\wedge\omega(\zeta) = \int\limits_{\zeta\in V-U} d[e^{\langle\zeta,w\rangle}\xi(\zeta)\wedge\omega(\zeta)] = 0.$$

Notice also that this integral depends only on the cohomology class $[\xi]\in H^{(0,n-1)}_{\bar\partial}(\mathbb{C}^n-E)$. For, if $\xi-\theta=\bar\partial u$ (where $\theta\in Z^{(0,n-1)}_{\bar\partial}(\mathbb{C}^n-E)$ and u is a $(0,n-2)-$form in $\mathbb{C}^n-E$), then

$$\int\limits_{\zeta\in\partial U} e^{\langle\zeta,w\rangle}\xi(\zeta)\wedge\omega(\zeta) - \int\limits_{\zeta\in\partial U} e^{\langle\zeta,w\rangle}\theta(\zeta)\wedge\omega(\zeta) = \int\limits_{\zeta\in\partial U} d[e^{\langle\zeta,w\rangle}u(\zeta)\wedge\omega(\zeta)] = 0.$$

Now it is easy to see that the function F_ξ is an entire function of exponential type. In fact,

$$|F_\xi(w)| \le \int\limits_{\zeta\in\partial U} e^{|\langle\zeta,w\rangle|}|\xi(\zeta)\wedge\omega(\zeta)| \le Ae^{R|w|} \quad for \quad w\in\mathbb{C}^n,$$

where A and R are positive constants.

Conversely, using the derivatives of the Bochner-Martinelli kernel, we will show that every entire function of exponential type is the Fourier-Laplace transform of a $\bar\partial$-closed $(0,n-1)$-form.

The derivatives of the Bochner-Martinelli kernel. For $\zeta\neq z$, set

$$M(\zeta,z) = \frac{\beta_n}{|\zeta-z|^{2n}}\sum_{j=1}^{n}(-1)^{j-1}(\bar\zeta_j-\bar z_j)d\bar\zeta_1\wedge\ldots(j)\ldots\wedge d\bar\zeta_n,$$

where $\beta_n=(-1)^{n(n-1)/2}(n-1)!/(2\pi i)^n$, and for each $k=(k_1,\ldots,k_n)$, where k_j are non-negative integers, let us define the $(0,n-1)$-forms

$$\eta_k(\zeta) = \left.\frac{\partial^{k_1+\cdots+k_n}M(\zeta,z)}{\partial z_1^{k_1}\cdots\partial z_n^{k_n}}\right|_{z=0}.$$

A simple computation shows that

$$\eta_k(\zeta) = \beta_n n(n+1)\cdots(n+k_1+\cdots+k_n-1)\frac{\bar\zeta_1^{k_1}\cdots\bar\zeta_n^{k_n}}{|\zeta|^{2(n+k_1+\cdots+k_n)}}\times$$
$$\times\sum_{j=1}^{n}(-1)^{j-1}\bar\zeta_j d\bar\zeta_1\wedge\ldots(j)\ldots\wedge d\bar\zeta_n.$$

Since $\bar\partial_\zeta[M(\zeta,z)]=0$, it follows that $\bar\partial\eta_k=0$. Thus $\eta_k\in Z^{(0,n-1)}_{\bar\partial}(\mathbb{C}^n-\{0\})$.

Now recall the Bochner-Martinelli formula: For $f \in \mathcal{O}(\mathbb{C}^n)$,

$$f(z) = \int\limits_{\zeta \in \mathbb{S}_\rho} f(\zeta) M(\zeta, z) \wedge \omega(\zeta), \quad when\ |z| < \rho, \tag{1}$$

where $\mathbb{S}_\rho = \{\zeta \in \mathbb{C}^n : \ |\zeta| = \rho\}$ and $\rho > 0$.

Applying to both sides of (1) the differentiation

$$\mathfrak{D}^k = \frac{\partial^{k_1+\cdots+k_n}}{\partial z_1^{k_1} \cdots \partial z_n^{k_n}}\bigg|_{z=0},$$

we obtain the formula

$$\mathfrak{D}^k f = \frac{\partial^{k_1+\cdots+k_n} f}{\partial z_1^{k_1} \cdots \partial z_n^{k_n}}\bigg|_{z=0} = \int\limits_{\zeta \in \mathbb{S}_\rho} f(\zeta) \eta_k(\zeta) \wedge \omega(\zeta). \tag{2}$$

Construction of $\bar\partial$-closed $(0, n-1)$-forms with prescribed Fourier-Laplace transform. Let F be an entire holomorphic function of the following exponential type:

$$|F(w)| \le A e^{R|w|}, \text{ for every } w \in \mathbb{C}^n, \tag{$\mathfrak{F}_R$}$$

where A and R are positive constants.

Now we will estimate the derivatives of F at zero, using Cauchy's formula in the polydisk:

$$\mathfrak{D}^k f = \frac{k_1! \ldots k_n!}{(2\pi i)^n} \int\limits_{\zeta \in \mathbb{T}_r} \frac{f(\zeta_1, ..., \zeta_n)}{\zeta_1^{k_1+1} \cdots \zeta_n^{k_n+1}}\, d\zeta_1 \wedge \ldots \wedge d\zeta_n, \tag{3}$$

where $\mathbb{T}_r$ is the torus of multi-radius $r = (r_1, ..., r_n)$:

$$\mathbb{T}_r = \{\zeta \in \mathbb{C}^n : \ |\zeta_1| = r_1, ..., |\zeta_n| = r_n\}.$$

Since for $\zeta \in \mathbb{T}_r$, $|F(\zeta)| \le A e^{R\sqrt{r_1^2+\cdots+r_n^2}}$, (3) implies that the coefficient σ_k, in the expansion $F(w) = \sum_k \sigma_k w^k$, satisfies the inequality

$$|\sigma_k| = \frac{1}{k_1! \ldots k_n!} |\mathfrak{D}^k F| \le A \frac{e^{R\sqrt{r_1^2+\cdots+r_n^2}}}{r_1^{k_1} \cdots r_n^{k_n}} \quad for\ every\ \ r_1, ..., r_n > 0.$$

Applying this inequality with

$$r_1 = \sqrt{k_1(k_1 + \cdots + k_n)}/R, ..., r_n = \sqrt{k_n(k_1 + \cdots + k_n)}/R,$$

we obtain

$$|\sigma_k| \le A\frac{(eR)^{k_1+\cdots+k_n}}{k_1^{k_1/2}\cdots k_n^{k_n/2}(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}} \quad \textit{for every } k_1,...,k_n. \tag{4}$$

(Convention: ${k_j}^{k_j/2}=1$, when $k_j=0$.)

Next let us recall that if $F(w)$ is to be the Fourier-Laplace transform of an analytic functional $\mathcal{T}$, then the action of $\mathcal{T}$ on a function $f\in\mathcal{O}(\mathbb{C}^n)$ will be given by the formula:

$$\mathcal{T}(f)=\sum_k \sigma_k\mathfrak{D}^k f.$$

Now we take an arbitrary $f\in\mathcal{O}(\mathbb{C}^n)$ and substitute the values of its derivatives $\mathfrak{D}^k f$ in the sum $\sum_k\sigma_k\mathfrak{D}^k f$, using formula (2). Interchanging the order of summation and integration, we obtain – at least formally – that

$$\sum_k \sigma_k\mathfrak{D}^k f=\int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\left(\sum_k\sigma_k\eta_k(\zeta)\right)\wedge\omega(\zeta). \tag{5}$$

We will show that the series $\sum_k\sigma_k\eta_k(\zeta)$ converges for $\zeta\in\mathbb{C}^n-\mathbb{B}_R$ (where $\mathbb{B}_R=\{|\zeta|\le R\}$), and defines a $\bar\partial$-closed $(0,n-1)$-form whose Fourier-Laplace transform is the given function F. In fact we will see that the convergence is uniform and absolute on compact subsets of $\mathbb{C}^n-\mathbb{B}_R$, and therefore (5) holds when $\rho>R$. In proving this, (4) will play the important role in conjunction with the following lemma.

LEMMA 1. *If $t_1,...,t_n>0$ and $t_1^2+\cdots+t_n^2<1$ then*

$$\sum_{k_1,\dots,k_n}\frac{e^{k_1+\cdots+k_n}(k_1+\cdots+k_n)!}{k_1^{k_1/2}\cdots k_n^{k_n/2}(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}}t_1^{k_1}...t_n^{k_n}<\infty.$$

PROOF. First let us keep in mind that the validity of the assertion is not affected if the general term of the sum is multiplied (or divided) by a quantity of the form $k_1^{s_1}...k_n^{s_n}$ (for some nonnegative constants $s_1,...,s_n$).

Now to prove the lemma, it suffices to show that the sum of the terms, whose indices $k_1,...,k_n$ are all even, is finite, i.e.,

$$\sum_{k_1,\dots,k_n}\frac{e^{2k_1+\cdots+2k_n}(2k_1+\cdots+2k_n)!}{(2k_1)^{k_1}...(2k_n)^{k_n}(2k_1+\cdots+2k_n)^{k_1+\cdots+k_n}}t_1^{2k_1}...t_n^{2k_n}<\infty \tag{6}$$

$$\text{for } t_1^2+\cdots+t_n^2<1.$$

To justify this reduction we split the sum according to the parity of the $k_1, ..., k_n$. More precisely, if we call $C(k_1, ..., k_n)$ the general term of the sum in the statement of the lemma then on the one hand it is clear that

$$\sum_{k_1,...,k_n} C(k_1, ..., k_n) = \sum_{(u_1,...,u_n)\in\{0,1\}^n} \sum_{k_1,...,k_n} C(2k_1 + u_1, ..., 2k_n + u_n),$$

and on the other hand it is easy to see (using the remark at the beginning of this proof) that (6) implies that each of the 2^n sums

$$\sum_{k_1,...,k_n} C(2k_1 + u_1, ..., 2k_n + u_n) \quad is\ finite$$

(i.e., when $u_1, ..., u_n \in \{0, 1\}$), and the reduction of the proof of the lemma to (6), follows.

Using the notation $|k| = k_1 + \cdots + k_n$, we have

$$\frac{e^{2k_1+\cdots+2k_n}(2k_1 + \cdots + 2k_n)!}{(2k_1)^{k_1}...(2k_n)^{k_n}(2k_1 + \cdots + 2k_n)^{k_1+\cdots+k_n}} =$$

$$= \left(\prod_{j=1}^{n} \frac{e^{k_j}k_j!}{k_j^{k_j}}\right) \frac{e^{|k|}|k|!}{|k|^{|k|}} \frac{(2|k|)!}{4^{|k|}(|k|!)^2} \frac{|k|!}{k_1!...k_n!}.$$

But from Stirling's formula, for $min\{s_j : 1 \le j \le n\}$ large enough,

$$\frac{e^{k_j}k_j!}{k_j^{k_j}} \approx \sqrt{2\pi k_j},\ \ \frac{e^{|k|}|k|!}{|k|^{|k|}} \approx \sqrt{2\pi|k|}\ \ and\ \ \frac{(2|k|)!}{4^{|k|}(|k|!)^2} \approx \frac{1}{2\sqrt{\pi|k|}}.$$

Therefore, using also the expansion

$$\sum_{k_1,...,k_n} \frac{(k_1 + \cdots + k_n)!}{k_1!...k_n!} t_1^{2k_1}...t_n^{2k_n} = \frac{1}{1 - (t_1^2 + \cdots + t_n^2)} \ \text{ valid for } t_1^2+\cdots+t_n^2 < 1,$$

we conclude that there is a positive integer N so that

$$\sum_{min\{k_1,...,k_n\}>N} \frac{e^{2k_1+\cdots+2k_n}(2k_1 + \cdots + 2k_n)!}{(2k_1)^{k_1}...(2k_n)^{k_n}(2k_1 + \cdots + 2k_n)^{k_1+\cdots+k_n}} t_1^{2k_1}...t_n^{2k_n} < \infty.$$

Now (6) can be proved by induction on n.

Some computations. For each $\psi = (\psi_1, ..., \psi_n) \in \mathbb{C}^n - \{0\}$, we define the region

$$\mathbb{G}_\psi = \left\{\zeta \in \mathbb{C}^n - \{0\} : \ \frac{|\zeta_j|}{|\zeta|^2} < \frac{|\psi_j|}{|\psi|^2} \ \ for\ j = 1, ..., n\right\}.$$

We will show that for $|\psi|>R$, the series $\sum_k |\sigma_k \mathfrak{C}_k(\zeta)|$ converges uniformly in $\zeta\in\mathbb{G}_\psi$, where $\mathfrak{C}_k(\zeta)$ is the main coefficient of $\eta_k(\zeta)$, i.e., the quantity

$$\mathfrak{C}_k(\zeta)=n(n+1)\cdots(n+k_1+\cdots+k_n-1)\frac{\bar\zeta_1^{k_1}\cdots\bar\zeta_n^{k_n}}{|\zeta|^{2(n+k_1+\cdots+k_n)}}.$$

First notice that $\mathbb{G}_\psi\subset\{\zeta\in\mathbb{C}^n:\,|\zeta|>|\psi|\}$, because

$$\frac{|\zeta_j|}{|\zeta|^2}<\frac{|\psi_j|}{|\psi|^2}\quad\Rightarrow\quad\sum_{j=1}^n\frac{|\zeta_j|^2}{|\zeta|^4}<\sum_{j=1}^n\frac{|\psi_j|^2}{|\psi|^4}\quad\Rightarrow\quad\frac{1}{|\zeta|^2}<\frac{1}{|\psi|^2}.$$

Also $u\psi\in\mathbb{G}_\psi$ for every $u>1$ (as it is easy to check) and for a fixed $\rho>R$,

$$\mathbb{C}^n-\mathbb{B}_\rho=\bigcup_{\psi\in\mathbb{S}_\rho}\mathbb{G}_\psi. \tag{7}$$

Indeed, if $\zeta\in\mathbb{C}^n-\mathbb{B}_\rho$ then it is easy to see that $\zeta\in\mathbb{G}_\psi$, where $\psi=\rho\zeta/|\zeta|$, and of course $\rho\zeta/|\zeta|\in\mathbb{S}_\rho$.

To prove the uniform convergence of the series $\sum_k|\sigma_k\mathfrak{C}_k(\zeta)|$ for $\zeta\in\mathbb{G}_\psi$ (with $|\psi|>R$), it suffices to notice that, since,

$$|\mathfrak{C}_k(\zeta)|=\frac{\prod_{l=1}^{n-1}(k_1+\cdots+k_n+l)}{(n-1)!|\zeta|^{2n}}(k_1+\cdots+k_n)!\left(\frac{|\zeta_1|}{|\zeta|^2}\right)^{k_1}\cdots\left(\frac{|\zeta_n|}{|\zeta|^2}\right)^{k_n},$$

inequality (4) implies that the series $\sum_k\sup\{|\sigma_k\mathfrak{C}_k(\zeta)|:\,\zeta\in\mathbb{G}_\psi\}$ is dominated by the convergent series

$$\sum_{k_1,\dots,k_n}\left[\frac{\prod_{l=1}^{n-1}(k_1+\cdots+k_n+l)}{(n-1)!|\psi|^{2n}}\frac{e^{|k|}(|k|)!}{k_1^{k_1/2}\cdots k_n^{k_n/2}|k|^{|k|/2}}\left(\frac{R|\psi_1|}{|\psi|^2}\right)^{k_1}\cdots\left(\frac{R|\psi_n|}{|\psi|^2}\right)^{k_n}\right].$$

The convergence of the above series follows from Lemma 1, since

$$\sum_{j=1}^n\left(\frac{R|\psi_j|}{|\psi|^2}\right)^2=\left(\frac{R}{|\psi|}\right)^2<1.$$

Now we can prove the following theorem which is a Paley-Wiener type theorem. As it is well-known such theorems deal with the question of representing entire functions of exponential type as Fourier-Laplace transforms of measures and the related literature is quite extensive. This particular theorem expresses such measures in terms of the Bochner-Martinelli kernel.

THEOREM 1. *If $F(w)=\sum_k \sigma_k w^k$ is an entire function, which satisfies $(\mathfrak{F}_R)$ for some $R>0$, then the series $\sum_k \sigma_k\eta_k(\zeta)$ defines a $\bar\partial$-closed $(0,n-1)$-form $\eta(\zeta)$, with C^∞ coefficients in $\zeta\in\mathbb{C}^n-\mathbb{B}_R$, and*

$$F(w)=\int\limits_{\zeta\in\mathbb{S}_\rho} e^{\langle\zeta,w\rangle}\eta(\zeta)\wedge\omega(\zeta),\quad \textit{for } w\in\mathbb{C}^n \textit{ and } \rho>R.$$

Thus an analytic functional $\mathcal{T}$, which is carried by the ball $\mathbb{B}_R$, is represented by the measure

$$d\lambda(\zeta)=\sum_k \sigma_k\eta_k(\zeta)\wedge\omega(\zeta)\big|_{\zeta\in\mathbb{S}_\rho},$$

supported by the sphere $\mathbb{S}_\rho$ $(\rho>R)$, where $\sigma_k=\mathcal{T}(z^k)/k!$.

In particular, any measure $d\mu$ (in $\mathbb{C}^n$ and with compact support) is analytically equivalent to $d\lambda$ (given by the above formula), where $\sigma_k=\int z^k d\mu(z)/k!$ and $\rho>\sup\{|z|:\ z\in supp(\mu)\}$

PROOF. Notice that

$$\eta_k(\zeta)=\sum_{j=1}^n(-1)^{j-1}\bar\zeta_j\mathfrak{C}_k(\zeta)d\bar\zeta_1\wedge...(j)...\wedge d\bar\zeta_n.$$

But if $\mathcal{P}$ is any derivative (of any order), with respect to $\zeta_1,...,\zeta_n,\bar\zeta_1,...,\bar\zeta_n$, then

$$\sum_k \sup\left\{|\sigma_k\mathcal{P}[\bar\zeta_j\mathfrak{C}_k(\zeta)]|:\ \zeta\in\mathbb{G}_\psi\right\}<\infty, \tag{8}$$

provided that $|\psi|>R$. This follows from Lemma 1, which implies that

$$\sum_{k_1,...,k_n} k_1^{s_1}...k_n^{s_n}\frac{e^{|k|}(|k|)!}{k_1^{k_1/2}\cdots k_n^{k_n/2}|k|^{|k|/2}}t_1^{k_1}...t_n^{k_n}<\infty\quad(t_1,...,t_n>0,\ t_1^2+\cdots+t_n^2<1),$$

for every nonnegative constants $s_1,...,s_n$. (At this point we use the fact that, since the function F satisfies the condition $(\mathfrak{F}_R)$, the coefficients σ_k satisfy (4), and, therefore, we can carry out computations, similar to the ones that follow the proof of Lemma 1, which lead to (8).)

But (8) implies that $\eta=\sum_k\sigma_k\eta_k$ has C^∞ coefficients $\mathbb{C}^n-\mathbb{B}_R$ and that

$$\bar\partial\eta=\sum_k\sigma_k\bar\partial\eta_k=0.$$

Furthermore, for $f \in \mathcal{O}(\mathbb{C}^n)$,

$$\int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\eta(\zeta)\wedge\omega(\zeta) = \int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\left(\sum_k \sigma_k\eta_k(\zeta)\right)\wedge\omega(\zeta) = \\ = \sum_k \sigma_k \int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\eta_k(\zeta)\wedge\omega(\zeta) = \sum_k \sigma_k \mathfrak{D}^k f,$$

where we also used (2). Applying the above formula with $f(\zeta) = e^{\langle\zeta,w\rangle}$ (for fixed w), we obtain

$$\int\limits_{\zeta\in\mathbb{S}_\rho} e^{\langle\zeta,w\rangle}\eta(\zeta)\wedge\omega(\zeta) = \sum_k \sigma_k w^k = F(w).$$

This completes the proof. ∎

3 – Laurent type expansions of $\bar\partial$-closed $(0,n-1)$-forms

The computations of the previous section lead also to the following theorem.

THEOREM 2. *Let $R \geq 0$. Suppose that for each $k = (k_1,...,k_n)$, where k_j are nonnegative integers, we are given a complex number $\varpi_k = \varpi_{k_1,...,k_n}$. Then a necessary and sufficient condition that there exist $\theta \in Z^{(0,n-1)}_{\bar\partial}(\mathbb{C}^n - \mathbb{B}_R)$ so that*

$$(\mathfrak{P}) \qquad \int\limits_{\zeta\in\mathbb{S}_\rho} \zeta_1^{k_1}\cdots\zeta_n^{k_n}\theta(\zeta)\wedge\omega(\zeta) = \varpi_{k_1,...,k_n}, \quad \text{for every } k \text{ (where } \rho > R),$$

is that the sequence $\varpi_k = \varpi_{k_1,...,k_n}$ satisfy the condition

$(\mathfrak{G}_R)$ *For every $\epsilon > 0$ there is a positive constant $A(\epsilon)$ so that*

$$|\varpi_k| \leq A(\epsilon)\frac{[e(R+\epsilon)]^{k_1+\cdots+k_n}k_1!...k_n!}{k_1^{k_1/2}\cdots k_n^{k_n/2}(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}} \quad \text{for every } k_1,...,k_n.$$

PROOF. Set

$$c_{k_1,\dots,k_n} = \frac{\varpi_{k_1,\dots,k_n}}{k_1!\dots k_n!}.$$

To prove the one direction, let us assume that $\theta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{B}_R)$ and satisfies $(\mathfrak{P})$. Then

$$c_{k_1,\dots,k_n} w_1^{k_1}\dots w_n^{k_n} = \int\limits_{\zeta\in\mathbb{S}_\rho} \frac{\zeta_1^{k_1}\cdots\zeta_n^{k_n} w_1^{k_1}\dots w_n^{k_n}}{k_1!\dots k_n!}\theta(\zeta)\wedge\omega(\zeta).$$

Since

$$\sum_{k_1,\dots,k_n} \frac{\zeta_1^{k_1}\cdots\zeta_n^{k_n} w_1^{k_1}\dots w_n^{k_n}}{k_1!\dots k_n!} = e^{\langle\zeta,w\rangle},$$

it follows that the series $F(w) = \sum_k c_k w^k$ converges, it defines an entire holomorphic function $F(w)$, and that this function is given by the integral:

$$F(w) = \int\limits_{\zeta\in\mathbb{S}_\rho} e^{\langle\zeta,w\rangle}\theta(\zeta)\wedge\omega(\zeta) \quad for\ \rho > R.$$

Applying this with $\rho = R+\epsilon$ (where $\epsilon > 0$), we see that

$$|F(w)| \le A(\epsilon)e^{(R+\epsilon)|w|},$$

where

$$A(\epsilon) = \int\limits_{|\zeta|=R+\epsilon} |\theta(\zeta)\wedge\omega(\zeta)|.$$

Now we can prove (in the same way we proved that $(\mathfrak{F}_R)$ implies (4)) that

$$|c_k| \le A(\epsilon)\frac{[e(R+\epsilon)]^{k_1+\dots+k_n}}{k_1^{k_1/2}\cdots k_n^{k_n/2}(k_1+\dots+k_n)^{(k_1+\dots+k_n)/2}},$$

and this proves $(\mathfrak{G}_R)$.

To prove the other direction, let us assume that the sequence ϖ_k satisfies $(\mathfrak{G}_R)$. Then, it follows from the proof of Theorem 1, that the series $\theta(\zeta) = \sum_k c_k\eta_k(\zeta)$ defines a $\bar\partial$-closed $(0,n-1)$-form with C^∞ coefficients in $\zeta\in\mathbb{C}^n - \mathbb{B}_{R+\epsilon}$, and this is true for every $\epsilon > 0$. Thus $\theta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{B}_R)$. Moreover

$$\int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\theta(\zeta)\wedge\omega(\zeta) = \sum_k c_k \int\limits_{\zeta\in\mathbb{S}_\rho} f(\zeta)\eta_k(\zeta)\wedge\omega(\zeta) = \sum_k c_k\mathfrak{D}^k f,$$

for $f \in \mathcal{O}(\mathbb{C}^n)$ and $\rho > R$. Applying this formula with $f(\zeta) = \zeta_1^{l_1} \cdots \zeta_n^{l_n}$ (with nonnegative integers $l_1, ..., l_n$), we see that, indeed, θ satisfies the required period condition $(\mathfrak{P})$. This completes the proof of the theorem. ∎

The following theorem is a variation of Theorem 2. It gives Laurent type expansions for $\bar\partial$-closed $(0,n-1)$-forms in $\mathbb{C}^n - \mathbb{B}_R$. (The case $R = 0$ of it, is in [2].)

THEOREM 3. *Every $\theta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{B}_R)$ has an expansion of the form*

$$\theta = \sum_k \frac{\varpi_k}{k!}\eta_k + \bar\partial\upsilon,$$

where the numbers ϖ_k are given by $(\mathfrak{P})$ and υ is a $(0,n-2)$-form with C^∞ coefficients in $\mathbb{C}^n - \mathbb{B}_R$.

PROOF. Given $\theta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{B}_R)$, we define ϖ_k by $(\mathfrak{P})$ and we set

$$\eta = \sum_k \frac{\varpi_k}{k!}\eta_k.$$

It follows from the proof of Theorem 2 that $\eta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{B}_R)$ and that, for $\rho > R$,

$$\int\limits_{\zeta\in\mathbb{S}_\rho} \zeta_1^{k_1} \cdots \zeta_n^{k_n}\eta(\zeta) \wedge \omega(\zeta) = \varpi_{k_1,...,k_n}.$$

Therefore

$$\int\limits_{\zeta\in\mathbb{S}_\rho} \zeta_1^{k_1} \cdots \zeta_n^{k_n}[\theta(\zeta) - \eta(\zeta)] \wedge \omega(\zeta) = 0, \quad \textit{for every } k_1, ..., k_n.$$

Now [1, Lemma 5] (see also Lemma 2, below) implies that there exists a $(0,n-2)$-form υ, with C^∞ coefficients in $\mathbb{C}^n - \mathbb{B}_R$, so that $\theta - \eta = \bar\partial\upsilon$. This gives the required expansion and completes the proof of the theorem. ∎

REMARKS. **1.** Writing the quantity

$$\frac{e^{k_1+\cdots+k_n}k_1!...k_n!}{k_1^{k_1/2}\cdots k_n^{k_n/2}(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}}$$

in the form

$$\left(\prod_{j=1}^n \frac{e^{k_j}k_j!}{k_j^{k_j}}\right)\frac{k_1^{k_1/2}\cdots k_n^{k_n/2}}{(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}},$$

and using Stirling's formula

$$\frac{e^{k_j}k_j!}{k_j^{k_j}} \approx \sqrt{2\pi k_j},$$

it is easy to see that a sequence ϖ_k satisfies the condition $(\mathfrak{G}_R)$ if and only if for every $\epsilon > 0$ there is a positive constant $\tilde{A}(\epsilon)$ so that

$$|\varpi_k| \leq \tilde{A}(\epsilon)\frac{(R+\epsilon)^{k_1+\cdots+k_n}k_1^{k_1/2}\cdots k_n^{k_n/2}}{(k_1+\cdots+k_n)^{(k_1+\cdots+k_n)/2}} \quad \textit{for every } k_1,...,k_n.$$

2. Let $\theta \in Z_{\bar{\partial}}^{(0,n-1)}(U - \mathbb{B}(a,R))$, where U is an open neighborhood of the closed ball $\mathbb{B}(a,R) = \{\zeta \in \mathbb{C}^n : |\zeta - a| \leq R\}$. Taking a $\rho > R$ so that $\mathbb{B}(a,\rho) \subset U$, we define the coefficients c_k by the formula:

$$c_k = \frac{1}{k_1!...k_n!}\int\limits_{\zeta\in\mathbb{S}_\rho} (\zeta_1 - a_1)^{k_1}\cdots(\zeta_n - a_n)^{k_n}\theta(\zeta)\wedge\omega(\zeta).$$

Let us also consider the differential forms $\eta_k(\cdot,a)$ defined by the formula

$$\begin{aligned}\eta_k(\zeta,a) &= \frac{\partial^{k_1+\cdots+k_n}M(\zeta,z)}{\partial z_1^{k_1}\cdots\partial z_n^{k_n}}\bigg|_{z=a} = \\ &= \beta_n n(n+1)\cdots(n+k_1+\cdots+k_n-1)\frac{(\bar{\zeta}_1-\bar{a}_1)^{k_1}\cdots(\bar{\zeta}_n-\bar{a}_n)^{k_n}}{|\zeta-a|^{2(n+k_1+\cdots+k_n)}}\times \\ &\quad\times\sum_{j=1}^{n}(-1)^{j-1}(\bar{\zeta}_j-\bar{a}_j)d\bar{\zeta}_1\wedge\ldots(j)\ldots\wedge d\bar{\zeta}_n.\end{aligned}$$

Then $\eta_k(\cdot,a) \in Z_{\bar{\partial}}^{(0,n-1)}(\mathbb{C}^n - \{a\})$ and they have properties analogous to those of η_k. We notice that although the differential form θ is defined only in $U - \mathbb{B}(a,R)$, the series

$$\sum_k c_k\eta_k(\zeta,a)$$

converges for $\zeta \in \mathbb{C}^n - \mathbb{B}(a,R)$ and defines there a $\bar{\partial}$-closed $(0,n-1)$-form.

Expansions in more general domains. Suppose that D is a pseudo-convex domain in $\mathbb{C}^n$, $a^1,...,a^N \in D$ and $R_1,...,R_N \geq 0$ so that

$$\mathbb{B}(a^j,R_j) \subset D \ \ (j=1,...N) \quad \textit{and} \quad \mathbb{B}(a^j,R_j)\cap\mathbb{B}(a^m,R_m) = \emptyset \ \ (j \neq m).$$

Let $\theta \in Z_{\bar\partial}^{(0,n-1)}\left(D-[\mathbb{B}(a^1,R_1)\cup\cdots\cup\mathbb{B}(a^N,R_N)]\right)$. Taking $\rho_j > R_j$ so that the balls $\mathbb{B}(a^j,\rho_j)$ are pairwise disjoint, we define

$$c_k^j = \frac{1}{k_1!\ldots k_n!}\int\limits_{\zeta\in\mathbb{S}_{\rho_j}} (\zeta_1-a_1^j)^{k_1}\cdots(\zeta_n-a_n^j)^{k_n}\theta(\zeta)\wedge\omega(\zeta).$$

Then, in view of the previous remark, $\sum_k c_k^j\eta_k(\zeta,a^j)\in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n-\mathbb{B}(a^j,R_j))$, and therefore

$$\xi \overset{\text{def}}{=} \theta-\sum_{j=1}^N\sum_k c_k^j\eta_k(\zeta,a^j)\in Z_{\bar\partial}^{(0,n-1)}\left(D-[\mathbb{B}(a^1,R_1)\cup\cdots\cup\mathbb{B}(a^N,R_N)]\right).$$

Moreover

$$\int\limits_{\zeta\in\mathbb{S}_{\rho_j}} (\zeta_1-a_1^j)^{k_1}\cdots(\zeta_n-a_n^j)^{k_n}\xi(\zeta)\wedge\omega(\zeta)=0 \quad \textit{for all } k \textit{ and } j.$$

It follows from Lemma 2 below that ξ is $\bar\partial$-exact in $D-[\mathbb{B}(a^1,R_1)\cup\cdots\cup\mathbb{B}(a^N,R_N)]$. The conclusion is that θ has the following expansion

$$\theta=\sum_{j=1}^N\sum_k c_k^j\eta_k(\zeta,a^j)+\bar\partial\upsilon,$$

for some $(0,n-2)$-form υ with C^∞ coefficients in $D-[\mathbb{B}(a^1,R_1)\cup\cdots\cup\mathbb{B}(a^N,R_N)]$.

LEMMA 2. *Let us consider an open set $\Omega\subset\mathbb{C}^n$ of the form $\Omega=D-(G_1\cup\ldots\cup G_N)$ where D is a pseudoconvex set and $G_1,\ldots,G_N$ are compact convex sets in $\mathbb{C}^n$ so that $G_j\subset D$ and $G_j\cap G_m=\emptyset$ for $j\neq m$. Let us also consider simple closed surfaces S_j, each one around the set G_j and close to it.*

Then a differential form $\chi\in Z_{\bar\partial}^{(0,n-1)}(\Omega)$ is $\bar\partial$-exact (in Ω) if and only if

$$\int\limits_{\zeta\in S_j} e^{\langle w,\zeta\rangle}\chi(\zeta)\wedge\omega(\zeta)=0, \quad \textit{for every } \ j=1,\ldots,N \ \textit{ and } \ w\in\mathbb{C}^n. \tag{9}$$

Notice that (9) is equivalent to

$$\int\limits_{\zeta\in S_j} f(\zeta)\chi(\zeta)\wedge\omega(\zeta)=0, \quad \textit{for } f\in\mathcal{O}(\mathbb{C}^n) \textit{ and } j=1,...,N,$$

because the set of the functions $e^{\langle w,\zeta\rangle}$, $w\in\mathbb{C}^n$, is dense in the space of entire functions (with the topology of uniform convergence on compact sets. Also this is equivalent to

$$\int\limits_{\zeta\in S_j} (\zeta_1-a_1^j)^{k_1}\cdots(\zeta_n-a_n^j)^{k_n}\chi(\zeta)\wedge\omega(\zeta)=0 \quad \textit{for all } k \textit{ and } j,$$

where a_j are any preassigned points.

Proof of Lemma 2. The one direction follows from Stokes's formula. The other direction is a generalization of [1, Lemma 5] and its proof is similar in this case too, so we will outline it.

First we exhaust the set Ω with a sequence of compact sets of the form

$$K=\{\lambda\le 0\}-(\{\rho_1<0\}\cup\ldots\cup\{\rho_N<0\}),$$

so that the set $\{\lambda<0\}$ is a bounded strictly pseudoconvex set with smooth boundary and the sets $\{\rho_1<0\},\ldots,\{\rho_N<0\}$ are strictly convex neighborhoods of the convex sets $G_1,\ldots,G_N$. In other words, the sets $\{\lambda<0\}$ should exhaust the pseudoconvex set D, while the set $\{\rho_j<0\}$ should shrink down to the set G_j, for $j=1,\ldots,N$.

Fixing such a set K, we consider the map $\gamma:(\partial K)\times\mathrm{int}(K)\to\mathbb{C}^n$ as follows: For $(\zeta,z)\in(\partial K)\times\mathrm{int}(K)$, $\{\gamma_l(\zeta,z)\}_{l=1}^n$ is defined to be a Henkin-Ramirez map of the strictly pseudoconvex set $\{\lambda<0\}$, if $\zeta\in\{\lambda=0\}$, and

$$\gamma_l(\zeta,z)=\frac{\partial\rho_j}{\partial\zeta_l}(z) \quad \textit{if } \zeta\in\{\rho_j=0\}.$$

(For exhaustions of pseudoconvex sets by strictly pseudoconvex domains and constructions of Henkin-Ramirez maps, see [5] and [6]).

Then

$$\sum_{l=1}^n(\zeta_l-z_l)\gamma_l(\zeta,z)\neq 0, \quad \textit{for } \ (\zeta,z)\in(\partial K)\times\mathrm{int}(K),$$

and therefore we may write down the Cauchy-Leray formula:

(10) $u=\bar\partial_z(T_{q-1}u)+T_q(\bar\partial u)+L_q^\gamma(u)$, *for $(0,q)$-forms u in a neighborhood of K*

(notation is as in [1, p. 912]).

Now if $\chi \in Z_{\bar\partial}^{(0,n-1)}(\Omega)$ satisfies (9), it follows, as in the proof of [1, Lemma 5], that $L_{n-1}^{\gamma}(\chi) = 0$, and therefore (10) gives

$$\chi = \bar\partial_z(T_{n-2}\chi), \quad in \ \ \mathrm{int}(K).$$

Now the conclusion that χ is $\bar\partial$-exact in Ω, follows from [1, Lemma 4], and this completes the proof of the lemma.

4 – Mittag-Leffler type constructions of $\bar\partial$-closed $(0,n-1)$-forms

In Theorem 2, we saw when and how we can construct a $\bar\partial$-closed $(0,n-1)$-form, in the complement of a closed ball, with prescribed certain weighted periods. The following theorem deals with the analogous question, when the closed ball is replaced by the union of an infinite sequence of pair-wise disjoint closed balls. Given the previous constructions, its proof is similar to the proof of [3,Theorem 2].

THEOREM 4. *Let D be an open subset of $\mathbb{C}^n$ and $\mathbb{B}(a^j,R_j)$, $j=1,2,3,...$, a sequence of pair-wise disjoint closed balls, contained in D, with $R_j \geq 0$. Let us also assume that the set $\{a^1,a^2,a^3,...\}$ of the centers of these balls is discrete in D and set $\mathbb{M} = \bigcup_{j=1}^{\infty}\mathbb{B}(a^j,R_j)$. Suppose that for each j we are given a sequence $\varpi_k^j = \varpi_{k_1,..,k_n}^j$ of complex numbers which satisfies the condition $(\mathfrak{G}_{R_j})$. Then there exists $\theta \in Z_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n - \mathbb{M})$ so that*

$$(\mathfrak{M}) \qquad \int\limits_{\zeta\in\mathbb{S}_{\rho_j}} (\zeta_1-a_1^j)^{k_1}\cdots(\zeta_n-a_n^j)^{k_n}\theta(\zeta)\wedge\omega(\zeta) = \varpi_{k_1,..,k_n}^j, \quad \textit{for all } k \textit{ and } j,$$

where $\rho_j > R_j$, with the balls $\mathbb{B}(a^j,\rho_j)$ being pair-wise disjoint.

If we assume, in addition, that the open set D and the balls $\mathbb{B}(a^j,R_j)$ satisfy the condition

$(*)$ *D can be exshausted by a sequence of pseudoconvex sets G_ν $(\nu=1,2,3,\dots)$ so that $(\partial G_\nu)\cap\mathbb{M} = \emptyset$ $(\forall\,\nu)$,*

then the differential form θ, which satisfies $(\mathfrak{M})$, is unique up to a $\bar\partial$-exact $(0,n-1)$-form in $\mathbb{C}^n - \mathbb{M}$.

COROLLARY. *With the notation and under the assumptions of the above theorem (including condition $(*)$), we have an isomorphism:*

$$H_{\bar\partial}^{(0,n-1)}(\mathbb{C}^n-\mathbb{M}) \cong \prod_{j=1}^{\infty}\mathcal{O}(B_j),$$

where $B_j = \{\zeta\in\mathbb{C}^n : |\zeta| < 1/R_j\}$.

PROOF. To define this isomorphism, let us associate, to each cohomology class $[\theta] \in H^{(0,n-1)}_{\bar\partial}(\mathbb{C}^n - \mathbb{M})$, a sequence of holomorphic functions $(h_j)_{j=1}^{\infty}$ defined by the power series:

$$h_j(\tau) = \sum_k c_k^j \tau^k, \quad for\ \tau \in B_j,$$

where

$$c_k^j = \frac{1}{k_1!...k_n!} \int\limits_{\zeta \in \mathbb{S}_{\rho_j}} (\zeta_1 - a_1^j)^{k_1} \cdots (\zeta_n - a_n^j)^{k_n} \theta(\zeta) \wedge \omega(\zeta),$$

with the $\rho_j > R_j$ chosen so that the balls $\mathbb{B}(a^j, \rho_j)$ are pairwise disjoint.

Then it is easy to check (in view of the previous computations) that $h_j \in \mathcal{O}(B_j)$ and that the map

$$[\theta] \to (h_j)_{j=1}^{\infty},$$

gives the required isomorphism.

REFERENCES

[1] T. HATZIAFRATIS: *Note on the Fourier-Laplace transform of $\bar\partial$-cohomology classes*, Z. Anal. Anwendungen, **17** (1998), 907-915.

[2] T. HATZIAFRATIS: *Expansions of certain $\bar\partial$-closed forms via Fourier-Laplace transform*, Z. Anal. Anwendungen, **22** (2003), 289-298.

[3] T. HATZIAFRATIS: *Mittag-Leffler type expansions of $\bar\partial$-closed forms in certain domains in $\mathbb{C}^n$*, Comment.Math. Univ. Carolinae, **44** (2003), 347-358.

[4] L. HÖRMANDER: *An introduction to complex analysis in several variables*, North-Holland, Amsterdam, 1990.

[5] S. KRANTZ: *Function theory of several complex variables*, Wadsworth & Brooks/ Cole, California, 1992.

[6] R.M. RANGE: *Holomorphic functions and integral representations in several complex variables*, Springer-Verlag, New York, 1986.

[7] W. RUDIN: *Function theory in the unit ball of $\mathbb{C}^n$*, Springer-Verlag, New York, 1980.

Lavoro pervenuto alla redazione il 22 dicembre 2003
ed accettato per la pubblicazione il 5 aprile 2004.
Bozze licenziate il 14 febbraio 2005

INDIRIZZO DELL'AUTORE:

Telemachos Hatziafratis – University of Athens – Department of Mathematics – Panepistemiopolis 15784 – Athens, Greece
E-mail: thatziaf@math.uoa.gr

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 33-52

Matrix-free numerical torus bifurcation of periodic orbits

EUGENE ALLGOWER – ULF GARBOTZ – KURT GEORG†

Abstract: *We consider systems*

$$\dot{\varphi} = f(\varphi, \lambda)$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$. *Such systems often arise from space discretizations of parabolic PDEs. We are interested in branches (with respect to* λ*) of periodic solutions of such systems.*

In the present paper we describe a numerical continuation method for tracing such branches. Our methods are matrix-free, i.e., Jacobians are only implemented as actions, this enables us to allow for large n. *Of particular interest is the detection and precise numerical approximation of bifurcation points along such branches: especially period-doubling and torus bifurcation points. This will also be done in a matrix-free context combining Arnoldi iterations (to obtain coarse information) with the calculation of suitable test functions (for precise approximations). We illustrate the method with the one- and two-dimensional Brusselator.*

1 – Introduction

Recently, Georg [5] discussed a general setting for performing numerical continuation in a matrix-free setting. Transpose-free iterative linear solvers (see, e.g., [15]) can be effectively incorporated into such large-scale problems. A frequent application of numerical continuation concerns the detection of singularities and bifurcation points on a solution branch. By means of suitable test

Key Words and Phrases: *Matrix-free continuation – Numerical bifurcation – Torus bifurcation.*
A.M.S. Classification: 65P30 – 37M20 – 37G15

functions various bifurcation points can also be detected and approximated in a matrix free setting.

In this paper we describe how to numerically trace periodic orbits in a matrix-free way. We also present a test function for detecting and approximating torus bifurcations in a matrix-free setting. Numerical results for torus bifucation arising in the Brusselator equations in one and two dimensions are given.

The results reproduce the branches computed in [12], [13], where the Newton-Picard Gauss Seidel method was used. The test functions derived here to characterize bifurcations were used to obtain highly accurate approximations of the bifurcation points.

In order to describe the numerical tracing for the problems considered here, we review several of the ideas in [5].

2 – The Problem Setting

We consider systems

$$\dot{\varphi} = f(\varphi, \lambda) \tag{1}$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$. Such systems often arise from space discretizations of parabolic PDEs. We are interested in branches (with respect to λ) of periodic solutions of such systems.

We denote a solution of (1) which has initial value u for $t = 0$ with $\varphi(u, t, \lambda)$. We have that $t \mapsto \varphi(u, t, \lambda)$ is periodic with period T iff

$$\varphi(u, T, \lambda) - u = 0. \tag{2}$$

However, even fixing λ, all points u of the same periodic orbit would satisfy this equation, hence we need an additional *phase condition*, say

$$h(u, T, \lambda) = 0 \tag{3}$$

to single out, at least locally, one point per orbit (see, e.g., [16]). In our numerical example we used the Poincaré phase condition

$$(u - \hat{u})^T f(\hat{u}, \lambda) = 0 \tag{4}$$

where $\hat{u}$ is some current point close to some orbit for given λ. This point will, of course, need to be adapted regularly. Let us remark that the Poincaré phase condition has proven to be an effective choice, see [12], [13].

Let us form the equation

$$H(u, T, \lambda) := \begin{bmatrix} \varphi(u, T, \lambda) - u \\ h(u, T, \lambda) \end{bmatrix} = 0. \tag{5}$$

For almost all choices of λ there is a neighbourhood of the orbit such that 0 is a regular value of H, if $\hat{u}$ is in that neigbourhood.

Hence, the periodic orbits of (1) can be traced by using numerical continuation methods (arising from varying λ) on equation (5). In particular, a matrix-free approach is very suitable if we are interested in allowing large dimensions n, since an action of the (full) Jacobian H' can be readily obtained, as we will see in Section 8. On the other hand, an explicit evaluation of the full Jacobian for large n is prohibitively expensive (see, e.g., [12], [13]). Hence, in this case direct linear solving methods are generally out of the question.

A numerical continuation method traces the solution branches of $H^{-1}(0)$. The method is called *matrix-free* if the Jacobian of H is not calculated explicitly, but only its action on a vector is given via some efficient process. In connection with modern (transpose-free) iterative linear solvers, see, e.g., [15], this is often a suitable approach for large systems, in particular for those investigated here.

Our main interest will center on the precise numerical detection of bifurcation points along solution branches of $H^{-1}(0)$. These special points on a solution branch are characterized by an additional equation

$$\tau(u, T, \lambda) = 0$$

where $\tau : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ can be viewed as a so-called *test function*. Note that also the detection and approximation of bifurcation points is carried out here in a matrix-free context.

Let us briefly describe the different bifurcation scenarios we are interested in for the periodic solutions of the dynamical system (1). Here is a list, see also [16]:

1. We briefly mention the case of a simple bifurcation point. This case, however, is documented rather well in the literature.
2. A singularity which displays a characteristic feature of periodic solutions is a period-doubling bifurcation. It turns out that the approach is rather similar to the simple bifurcation.
3. The most intriguing bifurcation is a torus bifurcation, which is a bit similar to Hopf bifurcation (and is sometimes also called a Hopf bifurcation of limit cycles). This will be our main topic.

3 – Numerical Continuation

We consider the numerical tracing of a solution branch

$$s \mapsto (u(s), T(s), \lambda(s))$$

of Equation (5). For simplicity, we view s as an arc-length parameter. Numerically, we actually perform pseudo-arclength steps, see, e.g., [1], [8].

A numerical continuation (predictor-corrector) method repeats two steps:

1. A predictor step generates an approximate point further along the solution curve, typically by linear extrapolation.
2. A corrector step finds a point approximately on the solution curve and close to the predicted point, typically by Newton-like steps.

The following algorithm sketches a possible implementation of this idea. For a more compact notation, we use $x := (u, T, \lambda) \in \mathbb{R}^{n+2}$, $\tilde{x} := (\tilde{u}, \tilde{T}, \tilde{\lambda}) \in \mathbb{R}^{n+2}$.

– **Algorithm 6 (Matrix-Free Predictor-Corrector)**

1. **Initialization**
   choose x such that $H(x) \approx 0$
   choose approximate tangent S such that $H'(x)S \approx 0$, $||S|| = 1$
   choose step size $h > 0$
   choose small reduction factor $1 \gg \eta > 0$
2. repeat
   (a) **Predictor**
   $\tilde{x} \leftarrow x + hS$
   (b) **Corrector**
   find Δx such that

$$\left\| \begin{bmatrix} H(\tilde{x}) \\ 0 \end{bmatrix} + \begin{bmatrix} H'(\tilde{x}) \\ S^T \end{bmatrix} \Delta x \right\| \leq \eta \left\| \begin{bmatrix} H(\tilde{x}) \\ 0 \end{bmatrix} \right\|$$

   via a transpose-free iterative linear solver, see, e.g., [15]
   $\tilde{x} \leftarrow \tilde{x} + \Delta x$
   (c) determine new h
   $S \leftarrow (\tilde{x} - x)/||\tilde{x} - x||$
   $x \leftarrow \tilde{x}$

Remark 7. The corrector step approximately solves

$$\begin{bmatrix} H(x) \\ S^T(x - \tilde{x}) \end{bmatrix} = 0$$

for x using an inexact Newton step. In our numerical examples, we use several such Newton steps in fact while reducing η.

4 – Calculating Special Points

When tracing a solution branch

$$s \mapsto (u(s), T(s), \lambda(s)) =: x(s)$$

of (5), one is often interested in special points on this branch. They can be of various types. Our cases of interest are covered by requiring that a certain test function $\tau : \mathbb{R}^{n+2} \to \mathbb{R}$ changes sign. Hence we seek a point $x^* \in \mathbb{R}^{n+2}$ such that

$$\begin{bmatrix} H(x^*) \\ \tau(x^*) \end{bmatrix} = 0 \ . \tag{8}$$

The following Lemma is easy to see:

LEMMA 9. *Let $x^* = x(s^*)$ be a regular zero point of H, i.e., the Jacobian $H'(x^*)$ has maximal rank. Then the following statements are equivalent:*

1. *$\tau(x(s))$ has a simple zero at $s = s^*$.*
2. *x^* is a regular zero point of* (8).

Once an approximation of x^* is found, we could, of course, use an inexact Newton's method directly on (8) to obtain a better approximation, i.e., without continuing to follow a path. Note, however, that this places a calculation of τ into the innermost loop of the method, i.e., while evaluating the functions in (8).

In the context of bifurcation analysis an evaluation of τ may be rather costly, e.g., in the case of torus bifurcation, see Theorem 13 below. Therefore, in our path following context, we use a somewhat different approach, which places an evaluation of τ into the outermost loop: During the numerical continuation, we monitor the sign of τ. Assume that a situation $\tau(x_-)\tau(x_+) < 0$ is encountered for two subsequent points x_-, x_+ on the solution curve. Then we introduce the approximate tangent

$$S := (x_+ - x_-)/||x_+ - x_-||$$

and the linear approximations

$$p(s) = x_- + s\, S \approx x(s)$$

Now let $q(s)$ be the solution of

$$\begin{bmatrix} H(q) \\ S^T(q - p(s)) \end{bmatrix} = 0 \ .$$

Hence, $q(s)$ can be viewed as the corrector-point to the predictor point $p(s)$. Clearly, $q(s)$ can be approximated via an iterative nonlinear solver using a

matrix-free double loop, the outer loop consisting of a Newton iteration, and the inner loop being a transpose-free iterative linear solver. We now find a zero of the function $s \mapsto \tau(q(s))$ via a secant-like method (e.g., Brent's Method, see [2]).

The resulting method for calculating x^* is implemented as a matrix-free triple loop, the outer loop being the secant-like method. Note, however, that the iterative methods representing the two inner loops can be started with increasingly improved values. Alternatively, a modification of this approach can be implemented into the numerical continuation method as a steplength strategy, see [1, Section 8.1] for details. This modification permits a matrix-free implementation consisting of a double loop.

We will apply these ideas to test functions τ that signal certain types of bifurcation points.

5 – Simple Bifurcation Points

Bordered matrices are an important tool for a numerical unfolding of singularities. For example, this is one of the principal themes of the book [6]. A consequence of [6, Proposition 3.2.1] is Keller's Lemma, see [8]:

Let $A \in \mathbb{R}^{n\times n}$ have rank $n-k$, and let $B, C \in \mathbb{R}^{n\times k}$. Then

$$\begin{bmatrix} A & B \\ C^T & 0 \end{bmatrix}$$

is nonsingular if and only if

$$[\, A \quad B \,] \text{ and } \begin{bmatrix} A \\ C^T \end{bmatrix}$$

have full rank. Since the set of invertible matrices is open in the space of square matrices, the choice of matrices B, C such that the above matrices have full rank is usually easy to fulfill. However, for numerical purposes, one needs to take issues of condition into account, see Remark 20.

The following is a well-known fact, see, e.g., [5], [6]. A simple bifurcation point $x^* = x(s^*)$ is characterized by the fact that the determinant

$$\det \begin{bmatrix} H'(x(s)) \\ S^T \end{bmatrix}$$

changes sign at $s = s^*$. Here S has to be some approximate tangent, i.e., $S \approx \dot{x}(s^*)$. However, this is not a numerically suitable choice of a test function. A better choice is to consider the following bordered system:

$$\begin{bmatrix} \begin{bmatrix} H'(x(s)) \\ S^T \end{bmatrix} & a \\ b^T & 0 \end{bmatrix} \begin{bmatrix} \xi(s) \\ \tau(s) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

If $a, b \in \mathbb{R}^{n+2}$ are chosen such that

$$\begin{bmatrix} \begin{bmatrix} H'(x(s^*)) \\ S^T \end{bmatrix} & a \end{bmatrix} \text{ and } \begin{bmatrix} \begin{bmatrix} H'(x(s^*)) \\ S^T \end{bmatrix} \\ b^T \end{bmatrix}$$

have full rank, then the matrix of the bordered system is non-singular for $s \approx s^*$ where $x^* = x(s^*)$ is a simple bifurcation point. It is easy to see (e.g., via Cramer's Rule) that τ is a test-function for the simple bifurcation point x^*, i.e., τ has a simple zero at $s = s^*$.

Hence, in principle, we could use the method described in Section 4 to detect and approximate simple bifurcation points. However, the approximation cannot be executed very accurately since the Jacobian

$$\begin{bmatrix} H'(x(s)) \\ S^T \end{bmatrix}$$

becomes singular at $s = s^*$, and hence the numerical tracing of $x(s)$ becomes unstable for $s \approx s^*$. It is, however, possible to obtain a matrix-free stable method, see [5].

6 – Period-Doubling Bifurcation

If u is a T-periodic orbit of (1), i.e., $H(u, T, \lambda) = 0$, then $\partial_1 \varphi(u, T, \lambda)$ is called the monodromy matrix. A simple period-doubling bifurcation point $x^* = (u^*, T^*, \lambda^*)$ is characterized in the following way, see, e.g., [16]: For an algebraically simple real eigenvalue $\nu(s)$ of $s \mapsto \partial_1 \varphi(u(s), T(s), \lambda(s))$ there holds $\nu(s^*) = -1$ and $\nu'(s^*) \neq 0$. Note that this implies that the determinant of

$$\partial_1 \varphi\big(u(s), T(s), \lambda(s)\big) + \mathrm{I}$$

changes sign at $s = s^*$.

In analogy to Section 5 we obtain the following numerical test function for detecting such a point while numerically following the curve $s \mapsto x(s)$.

THEOREM 10. *Let $x^* = x(s^*)$ be a simple period-doubling bifurcation point as defined above. Suppose $b, c \in \mathbb{R}^n$ are chosen such that the bordered matrix in the following system is invertible*

$$\begin{bmatrix} \partial_1 \varphi\big(u(s), T(s), \lambda(s)\big) + \mathrm{I} & b \\ c^T & 0 \end{bmatrix} \begin{bmatrix} \xi(s) \\ \tau(s) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{11}$$

Then the system is non-singular for $s \approx s^*$, and $\tau(s)$ has a simple zero at $s = s^*$.

Contrary to a simple bifurcation point, however, a simple period-doubling bifurcation point is not a singular point on the curve $H^{-1}(0)$, and hence the method described in Section 4 applies. The only additional complexity of the problem comes from the calculation of τ which requires one matrix-free loop.

7 – Torus bifurcation

We consider again a local (i.e., for $s \approx s^*$) parametrization $s \mapsto (u(s),T(s),\lambda(s))$ of $H^{-1}(0)$, and define the monodromy matrix $A(s) := \partial_1 \varphi\big(u(s), T(s), \lambda(s)\big)$. A simple torus bifurcation point

$$(u^*, T^*, \lambda^*) = (u(s^*), T(s^*), \lambda(s^*))$$

is characterized in the following way, see, e.g., [6], [10], [16]: Let $\nu(s) + i\omega(s)$ be an algebraically simple complex eigenvalue of $A(s)$ for $s \approx s^*$. Hence

$$A(s)(v_1(s) + iv_2(s)) = (\nu(s) + i\omega(s))(v_1(s) + iv_2(s))$$

for linearly independent $v_1(s), v_2(s) \in \mathbb{R}^n$. Let the two eigenvalues furthermore cross the unit circle in the sense that

$$\epsilon(s) := \nu(s)^2 + \omega(s)^2 - 1$$

has a simple zero at $s = s^*$ with $\omega(s^*) \neq 0$. It follows that

$$(A - \nu)v_1 = -\omega v_2 \ ,$$

$$(A - \nu)v_2 = \omega v_1 \ .$$

Hence, if we consider the real vector space

$$E(s) = \operatorname{span}\{\, v_1(s), v_2(s) \,\},$$

then the kernel of

$$(A(s) - \nu(s)\ \mathrm{I})^2 + \omega(s)^2\ \mathrm{I} = A(s)^2 - 2\nu(s)A(s) + \ \mathrm{I} + \epsilon(s)\ \mathrm{I}$$

is $E(s)$. Also note that the two-dimensional space $E(s)$ is invariant under $A(s)$, and that $A(s)$ is bijective on $E(s)$ for $s \approx s^*$ since its two eigenvalues are close to the unit circle.

The following theorem describes a test function for a simple torus bifurcation point which we have implemented numerically. The introduction of the system (14) below was motivated by similar systems for Hopf bifurcation, such as [3], [17].

For the proof of the theorem we introduce the following notation:

DEFINITION 12. For $f, g : \mathbb{R}^{k_1} \to \mathbb{R}^{k_2}$ we define $f(z) \succ g(z)$ if there is an $\epsilon > 0$ such that $||f(z)|| \geq \epsilon ||g(z)||$ for sufficiently small $||z||$.

THEOREM 13. *Let $(u(s^*), T(s^*), \lambda(s^*))$ be a simple torus bifurcation point as described above. Assume that $c, d \in \mathbb{R}^n$ are chosen so that*

$$\begin{bmatrix} A(s^*)^2 - 2\nu(s^*)A(s^*) + \mathrm{I} \\ c^T \\ d^T \end{bmatrix} \text{ and } \begin{bmatrix} A(s^*) - (\nu(s^*) \pm i\omega(s^*))\ \mathrm{I} \\ d^T \end{bmatrix}$$

have full rank. Note that this implies that there exists a unique $e(s) \in E(s)$ with $c^T e(s) = 1$, $d^T e(s) = 0$ for $s \approx s^*$. Furthermore assume that $a, b \in \mathbb{R}^n$ are chosen so that

$$[\, A(s^*)^2 - 2\nu(s^*)A(s^*) + \mathrm{I} \quad a \quad b\,] \text{ and } [\, A(s^*)^2 - 2\nu(s^*)A(s^*) + \mathrm{I} \quad a \quad A(s^*)e(s^*)\,]$$

have full rank. Then the bordered matrix in the linear system

$$\begin{bmatrix} A(s)^2 - 2\mu A(s) + \mathrm{I} & a & b \\ c^T & 0 & 0 \\ d^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi(\mu, s) \\ \alpha(\mu, s) \\ \beta(\mu, s) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{14}$$

is non-singular for $s \approx s^*$ and $\mu \approx \nu(s^*)$. Hence

$$\begin{bmatrix} \xi(\mu, s) \\ \alpha(\mu, s) \\ \beta(\mu, s) \end{bmatrix}$$

is well-defined. Furthermore, the following holds:

1. $$\begin{bmatrix} \xi(\nu(s^*), s^*) \\ \alpha(\nu(s^*), s^*) \\ \beta(\nu(s^*), s^*) \end{bmatrix} = \begin{bmatrix} e(s^*) \\ 0 \\ 0 \end{bmatrix}.$$
2. *$\partial_1 \beta(\nu(s^*), s^*) \neq 0$, hence by the implicit function theorem the equation $\beta(\mu, s) = 0$ defines a parametrization $\mu(s)$ for $s \approx s^*$ such that $\beta(\mu(s), s) = 0$ and $\mu(s^*) = \nu(s^*)$.*
3. *$\tau(s) := \alpha(\mu(s), s)$ has a simple zero at $s = s^*$ and can hence be used as a test function for torus bifurcation.*

PROOF. The non-singularity of the linear system is an immediate consequence of the assumptions on a, b, c, d and that the complex eigenvalue $\nu(s^*) + i\omega(s^*)$ is simple.

1. The first claim follows from the uniqueness of the solution and the definition of $e(s)$.
2. Differentiating the linear system with respect to μ gives

$$\begin{bmatrix} -2A(s) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi(\mu,s) \\ \alpha(\mu,s) \\ \beta(\mu,s) \end{bmatrix} + \begin{bmatrix} A(s)^2 - 2\mu A(s) + \mathrm{I} & a & b \\ c^T & 0 & 0 \\ d^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \partial_1\xi(\mu,s) \\ \partial_1\alpha(\mu,s) \\ \partial_1\beta(\mu,s) \end{bmatrix} =$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Now we use Cramer's Rule to obtain that

$$\partial_1\beta(\nu(s^*),s^*) = \frac{\det \begin{bmatrix} A(s^*)^2 - 2\nu(s^*)A(s^*) + \mathrm{I} & a & 2A(s^*)\xi(\nu(s^*),s^*) \\ c^T & 0 & 0 \\ d^T & 0 & 0 \end{bmatrix}}{\det \begin{bmatrix} A(s^*)^2 - 2\nu(s^*)A(s^*) + \mathrm{I} & a & b \\ c^T & 0 & 0 \\ d^T & 0 & 0 \end{bmatrix}} \neq 0$$

because of the assumptions on a, b, c, d and since $A(s^*)\xi(\nu(s^*), s^*) = A(s^*)e(s^*) \neq 0$.

3. First we show that

$$\det \begin{bmatrix} \partial_1\alpha(\nu(s^*),s^*) & \partial_2\alpha(\nu(s^*),s^*) \\ \partial_1\beta(\nu(s^*),s^*) & \partial_2\beta(\nu(s^*),s^*) \end{bmatrix} \neq 0 \,. \tag{15}$$

For this purpose we write

$$A^2(s) - 2\mu A(s) + \mathrm{I} = \underbrace{(A(s) - (\mu - i\sqrt{1-\mu^2})\ \mathrm{I})}_{=:\,A_1(\mu,s)} \underbrace{(A(s) - (\mu + i\sqrt{1-\mu^2})\ \mathrm{I})}_{=:\,A_2(\mu,s)}$$

We will also make use of three lemmas which we list after this proof. According to [3, Proposition 1.1] we have to show two things to obtain (15):

(a) $A_1(\mu,s)v \succ v$ and $A_2(\mu,s)v \succ v$ uniformly in $(\mu\,,\,s) \approx (\nu(s^*)\,,\,s^*)$ for $v^T d = 0$. This is a fairly standard consequence of the fact that $\mu(s^*) + i\omega(s^*)$ is a simple eigenvalue of $A(s^*)$ and of the assumptions on d, see Lemma 19 for more details.

(b)

$$\sigma_{n-1}(A^2(s) - 2\mu A(s) + \ \mathrm{I}) \succ |s| + |\mu - \nu(s^*)| \tag{16}$$

Here $\sigma_1(B) \geq \sigma_2(B) \geq \ldots \sigma_n(B) \geq 0$ denote the singular values of a matrix $B \in \mathbb{R}^{n\times n}$. To prove (16), we first use Lemma 17 to obtain

$$\sigma_n(A(s) - (\mu + i\sqrt{1-\mu^2})\ \mathrm{I}) \succ (\mu - \nu(s)) + i(\sqrt{1-\mu^2} - \omega(s))\ .$$

We want to use Lemma 18 with

$$k(\mu,s) := \begin{bmatrix} \mu - \nu(s) \\ \sqrt{1-\mu^2} - \omega(s) \end{bmatrix}$$

and calculate the Jacobian:

$$k'(\nu(s^*),s^*) = \begin{bmatrix} 1 & -\nu'(s^*) \\ -\nu(s^*)/\omega(s^*) & -\omega'(s^*) \end{bmatrix}.$$

We obtain

$$\det k'(\nu(s^*),s^*) = \omega(s^*)^{-1}(-\omega'(s^*)\omega(s^*) - \nu'(s^*)\nu(s^*)) \neq 0\ .$$

Here we used the fact that $\nu^2(s) + \omega^2(s) - 1$ has a simple zero at $s = s^*$ and thus the derivative $2\nu(s^*)\nu'(s^*) + 2\omega(s^*)\omega'(s^*) \neq 0$. Hence, by using Lemma 18, we can continue our above estimate to conclude

$$\sigma_n(A_2(\mu,s)) = \sigma_n(A(s) - (\mu + i\sqrt{1-\mu^2})\ \mathrm{I}) \succ |s| + |\mu - \nu(s^*)|\ .$$

Combining this with the fact that

$$\sigma_{n-1}(A_1(\mu,s)) = \sigma_{n-1}(A(s) - (\mu - i\sqrt{1-\mu^2})\ \mathrm{I}) \succ 1$$

for $[s,\mu] \approx [0,\nu(s^*)]$, we now obtain from $\sigma_{n-1}(A_1A_2) \geq \sigma_{n-1}(A_1)\,\sigma_n(A_2)$ the nondegeneracy condition (16).
We have established (15) and use this in the following way. Let us define $\tilde{\alpha}(s) := \alpha(\mu(s),s)$. We know that $\tilde{\alpha}(s^*) = 0$ and have to show that $\tilde{\alpha}'(s^*) \neq 0$. We have

$$\tilde{\alpha}'(s^*) = \partial_1\alpha(\nu(s^*),s^*)\,\mu'(s^*) + \partial_2\alpha(\nu(s^*),s^*)\ .$$

From $\beta(\mu(s), s) = 0$ we obtain

$$0 = \partial_1\beta(\nu(s^*), s^*)\,\mu'(s^*) + \partial_2\beta(\nu(s^*), s^*)\ .$$

Note that $\partial_1\beta(\nu(s^*), s^*) \neq 0$ was already established. Eliminating $\mu'(s^*)$ in the last two equations leads to

$$\tilde{\alpha}'(s^*) = \partial_2\alpha(\nu(s^*), s^*) - \partial_1\alpha(\nu(s^*), s^*)\,\frac{\partial_2\beta(\nu(s^*), s^*)}{\partial_1\beta(\nu(s^*), s^*)}$$

which is different from zero because of (15). □

The previous proof used the following lemmas:

LEMMA 17. *Let $s \in \mathbb{R} \mapsto B(s) \in \mathbb{R}^{n\times n}$ and $s \in \mathbb{R} \mapsto \rho(s) \in \mathbb{C}$ be smooth and such that $\rho(s)$ is an eigenvalue of $B(s)$ and $\rho(s^*)$ is an algebraically simple eigenvalue of $B(s^*)$. Then $\sigma_n(B(s) - \rho(s)\ \mathrm{I}) \succ |\rho(s^*) - \rho(s)|$.*

A proof can be obtained by following arguments in [3, p. 533].

The next lemma is well-known:

LEMMA 18. *If $k : \mathbb{R}^2 \to \mathbb{R}^2$ is smooth, $k(y_0) = 0$, and $k'(y_0)$ non-singular, then $k(y) \succ y - y_0$.*

The following lemma seems to be fairly standard, see, e.g., the techniques used in [6].

LEMMA 19. *Let $B : \mathbb{R}^k \to \mathbb{R}^{m\times m}$ be continuous and such that $B(0)$ has 0 as an algebraically simple eigenvalue. Let $d \in \mathbb{R}^m$ be such that*

$$\begin{bmatrix} B(0) \\ d^T \end{bmatrix}$$

has full rank. Then

$$B(\lambda)v \succ v \text{ for } v \in \mathbb{R}^m \text{ with } v \perp d$$

uniformly for $\lambda \approx 0$.

REMARK 20. Note that the theorem gives a local result. We therefore propose to use it in conjunction with an Arnoldi iteration: While following a branch of periodic solutions, we occasionally apply an Arnoldi-type iteration (we used ARPACK [11]) to obtain snapshots of the dominant eigenvalues (Floquet multipliers) of the monodromy matrix $\partial_1\varphi(u,T,\lambda)$. The snapshot will show when a pair of conjugate complex eigenvalues passes through the unit circle. We then use the above test-function as described in Section 4 to approximate the torus bifurcation point more accurately. Note that the Arnoldi-like method not only gives us guesses for μ via the real part of the approximate eigenvalue, but also choices for a and b via the real and imaginary part of the approximate eigenvector. This proved to be an effective and efficient strategy.

8 – Implementing the action of $H'(u,T,\lambda)$

As has been seen in the preceding sections, in order to perform the predictor-corrector steps of numerical continuation and the evaluation of the test functions for bifurcations, it is necessary to implement the action of the Jacobian $H'(u,T,\lambda) = \begin{bmatrix} \partial_1\varphi - I & \partial_2\varphi & \partial_3\varphi \\ \partial_1 h & \partial_2 h & \partial_3 h \end{bmatrix}(u,T,\lambda)$ efficiently, see also [4], [12], [13].

1. The action $v \mapsto \partial_1\varphi(u,T,\lambda)v$ is obtained in the following way: Defining

$$z(t) := \partial_1\varphi(u,t,\lambda)v$$

and differentiating the equations

$$\partial_2\varphi(u,t,\lambda) = f(\varphi(u,t,\lambda),\lambda), \quad \varphi(u,0,\lambda) = u \tag{21}$$

with respect to u leads to the following description: We solve

$$\dot{z} = \partial_1 f(\varphi(u,t,\lambda),\lambda)\, z, \quad z(0) = v \tag{22}$$

and obtain

$$\partial_1\varphi(u,T,\lambda)v = z(T).$$

2. The vector $\partial_3\varphi(u,T,\lambda)$ is obtained in the following way: Defining

$$\xi(t) = \partial_3\varphi(u,t,\lambda)$$

and differentiating the equations (21) with respect to λ leads to the following description:

We solve

$$\dot{\xi} = \partial_1 f(\varphi(u,t,\lambda),\lambda)\,\xi + \partial_2 f(\varphi(u,t,\lambda),\lambda), \quad \xi(0) = 0 \tag{23}$$

and obtain

$$\partial_3\varphi(u,T,\lambda) = \xi(T).$$

3. The vector

$$\partial_2\varphi(u,T,\lambda) = f(\varphi(u,T,\lambda),\lambda) \tag{24}$$

is immediately obtained from (21)

The action of the other derivatives contained in $H'(u,T,\lambda)$ are even more obvious.

9 – The numerical calculation of φ and its derivatives

The previously indicated implementations all rely on an approximation of the orbit

$$t \mapsto \varphi(u,t,\lambda).$$

Our present aim is only to demonstrate the usefulness and applicability of the *matrix-free approach*. We need to approximate the orbit on a grid $t_0 = 0, t_1, \ldots, t_{m-1}, t_m = T$. For simplicity and convenience, we use an equidistant grid $t_i = ih$ in our numerical example, and note that for greater efficiency an adaptive grid must be taken into consideration.

Since we are mainly interested in cases where $\dot{\varphi} = f(\varphi,\lambda)$ is obtained from parabolic PDEs via space discretizations, we concentrate on the case where $\dot{\varphi} = f(\varphi,\lambda)$ is stiff, hence we have to consider the use of an implicit solver for the time steps. The main point here is that we need to solve a nonlinear system for each time step.

For simplicity, in our numerical example we consider the implicit midpoint rule, i.e., we approximate $\varphi(u,t_i,\lambda) \approx \varphi_i$ via

$$\frac{\varphi_i - \varphi_{i-1}}{h} = f\left(\frac{\varphi_i + \varphi_{i-1}}{2}, \lambda\right).$$

We solve this for φ_i using an inexact Newton's method, see, e.g., [9], where the linear equations (per Newton step) are solved with an iterative linear solver (see, e.g., [15]). For a preconditioner we use an (occasionally updated) sparse (possibly incomplete) LU factorization of the approximate Jacobian

$$\mathrm{I} - h\partial_1 f\,(v,\lambda) \;\text{ where } v \approx \varphi_i\ .$$

A few preconditioners are stored along an orbit, and are used again for small variations in u and λ.

The linear differential equation (22) is solved by the same implicit midpoint rule. This leads to the linear systems

$$\left(\mathrm{I}-\frac{h}{2}\partial_1 f\left(\varphi\left(u,\frac{t_{i+1}+t_i}{2},\lambda\right),\lambda\right)\right)z_{i+1}=\left(\mathrm{I}+\frac{h}{2}\partial_1 f\left(\varphi\left(u,\frac{t_{i+1}+t_i}{2},\lambda\right),\lambda\right)\right)z_i$$

which we have to solve for z_{i+1}. Typically we would replace the unknown $\varphi(u,\frac{t_{i+1}+t_i}{2},\lambda)$ with $\frac{\varphi_{i+1}+\varphi_i}{2}$.

The linear differential equation (23) is also solved by the implicit midpoint rule:

$$\left(\mathrm{I}-\frac{h}{2}\partial_1 f\left(\varphi\left(u,\frac{t_{i+1}+t_i}{2},\lambda\right),\lambda\right)\right)\xi_{i+1}=\left(\mathrm{I}+\frac{h}{2}\partial_1 f\left(\varphi\left(u,\frac{t_{i+1}+t_i}{2},\lambda\right),\lambda\right)\right)\xi_i+$$
$$+h\partial_2 f\left(\varphi\left(u,\frac{t_{i+1}+t_i}{2},\lambda\right),\lambda\right)$$

It is convenient to use the same iterative linear solver for all three cases, with the same preconditioner. We finally note that similar remarks would hold if we replaced the implicit midpoint rule with a higher order implicit solver for (21)-(23).

10 – Numerical Example

The Brusselator in one space dimension (z-variable) is modelled by the equations

$$\begin{aligned}\frac{\partial X}{\partial t}&=\frac{D_X}{L^2}\frac{\partial^2 X}{\partial z^2}+X^2Y-(B+1)X+A,\\ \frac{\partial Y}{\partial t}&=\frac{D_Y}{L^2}\frac{\partial^2 Y}{\partial z^2}-X^2Y+BX\end{aligned}\tag{25}$$

with Dirichlet boundary conditions

$$\begin{aligned}X(t,\,z=0)&=X(t,\,z=1)=A,\\ Y(t,\,z=0)&=Y(t,\,z=1)=B/A,\end{aligned}\tag{26}$$

see, e.g., [7]. As in [12], we use the characteristic length L as the bifurcation parameter while the other parameters are fixed at $A=2$, $B=5.45$, $D_X=0.008$ and $D_Y=0.004$.

It is known, see [12], that the first bifurcation from the trivial solution $X \equiv A$, $Y \equiv B/A$ is a Hopf bifurcation at $L \approx 0.513$ and the bifurcating branch of periodic orbits has two torus bifurcation points between 1.7 and 1.9.

Using a numerical continuation method as described in Section 3, we reconfirmed the bifurcation diagram published in [12].

As an example, we used our torus test function described in Theorem 13 together with the approach described in Section 4 to calculate the two torus bifurcation points more accurately. For this purpose, we chose $m = 100$ time steps to discretize the periodic orbit as described in Section 8. In the space coordinate z we used a central difference discretization with n (equidistant) interior points.

We note that this example is simple in that the solutions are known to be smooth. In fact, for more precise approximations, it would be adequate to use higher order discretizations in space and time. Also, in general, our approach should be modified to allow for adaptive meshes, in particular with respect to time.

However, here we just want to make the point that our matrix-free numerical approach is capable of handling large structures involving a variety of singularities. More complex approaches involving adaptive meshes and/or higher orders are currently under investigation.

For the first torus bifurcation we obtained

n	L
50	1.778310
100	1.783406
200	1.784757

For the second torus bifurcation we obtained

n	L
50	1.864434
100	1.872761
200	1.874973

This data supports the claim that the approximation is quadratic in the space discretization (neglecting the time discretization which was held fixed).
Now we consider the Brusselator in two space dimensions (x and y-coordinates)

$$\frac{\partial X}{\partial t} = \frac{D_X}{L^2}\left(\frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2}\right) + X^2 Y - (B+1)X + A,$$

$$\frac{\partial Y}{\partial t} = \frac{D_Y}{L^2}\left(\frac{\partial^2 Y}{\partial x^2} + \frac{\partial^2 Y}{\partial y^2}\right) - X^2 Y + BX$$

on the unit square $[0,1] \times [0,1]$ with the Dirichlet boundary conditions corresponding to (26) on all boundaries.

Fixing the same parameter values as above, the first branch of periodic orbits bifurcates from the trivial solution at $L \approx 0.72$.

This branch has been computed in [14] for similar parameter values. As in [14], we continued that branch and detected a torus bifurcation point at $L \approx 1.48$, see Figure 1 and Figure 2. We again used our torus test function described in Theorem 13 together with the approach described in Section 9 to precisely calculate this torus bifurcation point.

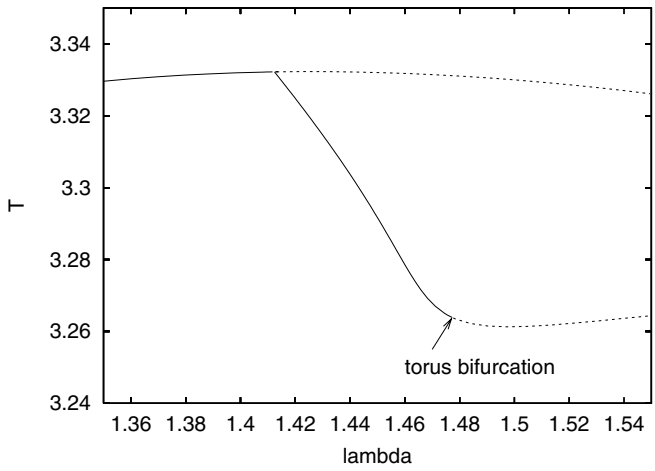

Fig. 1: The numerical continuation for a 20×20 space discretization.

For this purpose, we again chose $m = 100$ time steps to discretize the periodic orbit as described in Section 8. In the space co-ordinates x, y we used a central difference discretization with $n \times n$ (equidistant) interior points.

Note that the resulting computations are already relatively large for direct methods (i.e., generating the monodromy matrix). With our matrix-free approach, however, we were able to perform the required calculations for $n = 40$ on a 600MHz laptop.

The numerical continuation method and the evaluation of the test function have as the main computational expense the time integrations described in Section 9. The stiff solver used there makes use of an iterative linear solver which needs to be preconditioned. We chose some time points on the orbit to generate a sparse LU-factorization of the linear problem, and we used this LU-factorization as a preconditioner for neighboring times, and also for similar parameter val-

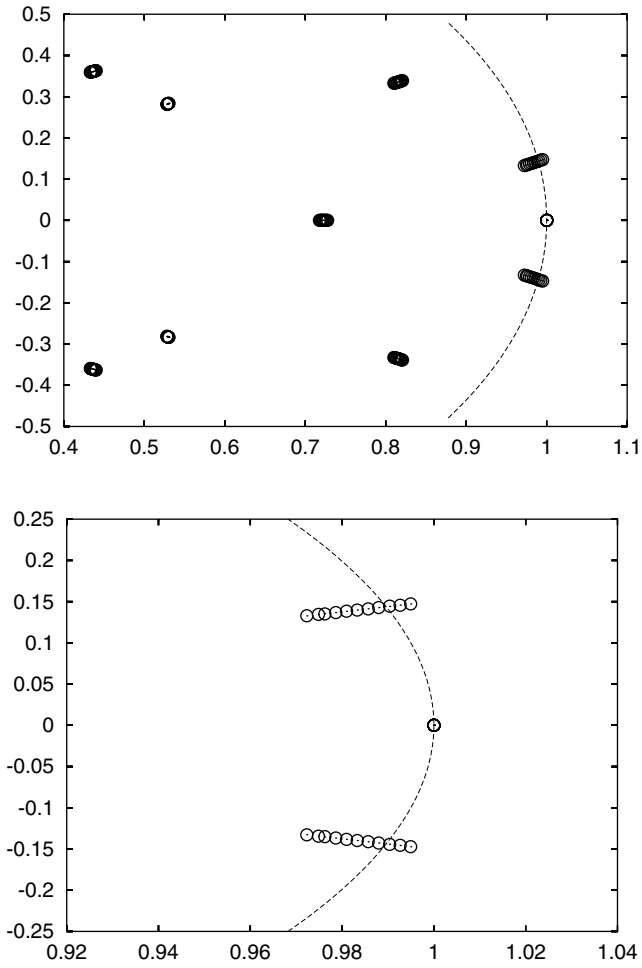

Fig. 2: The trajectories of the 10 eigenvalues with largest magnitude on the λ-interval $[1.47639, 1.48032]$ for a 40×40 space discretization.

ues during the continuation procedure. For finer space discretizations we would propose to use only incomplete LU-factorizations.

For the torus bifurcation we obtained

grid	L
10×10	1.47224
20×20	1.47756
40×40	1.47930

This data again supports the claim that the approximation is quadratic in the space discretization (neglecting the time discretization which was held fixed). Additional efficiencies could be effected by incorporating higher order spatial discretizations and variable time steps.

REFERENCES

[1] E. L. Allgower – K. Georg: *Numerical path following*, In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, volume 5, North-Holland, 1997, pp. 3-207.

[2] R. P. Brent: *Algorithms for minimization without derivatives*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1973. Prentice-Hall Series in Automatic Computation.

[3] K.-W. E. Chu – W. Govaerts – A. Spence: *Matrices with rank deficiency two in eigenvalue problems and dynamical systems*, SIAM J. Num.Anal., **3** (1994), 524-539.

[4] U. Garbotz: *Homokline Orbits in großen Systemen. Ein numerisches Verfahren für Mehrparameter-Probleme*, PhD thesis, University of Marburg, Department of Mathematics, 2001. http://archiv.ub.uni-marburg.de/diss/z2001/0274/.

[5] K. Georg: *Matrix-free numerical continuation and bifurcation*, Numerical Functional Analysis and Optimization, **22** (2001), 303-320.

[6] W. Govaerts: *Numerical Methods for Bifurcations of Dynamical Equilibria*, SIAM, 2000.

[7] M. Holodniok – P. Knedlík – M. Kubíček: *Continuation of periodic solutions in parabolic partial differential equations*, In T. Küpper, R. Seydel, and H. Troger, editors, *Bifurcations: Analysis, Algorithms, Applications*, volume 79 of ISNM, Birkhäuser Verlag, Basel, 1987, pp. 122-130.

[8] H. B. Keller: *Numerical solution of bifurcation and nonlinear eigenvalue problems*, In P. H. Rabinowitz, editor, *Applications of Bifurcation Theory*, Academic Press, New York, London, 1977, pp. 359-384.

[9] C. T. Kelley: *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[10] Yu. A. Kuznetsov.: *Elements of Applied Bifurcation Theory*, volume 112 of Appl. Math. Sci. Springer, second edition, 1998.

[11] R.B. Lehoucq – D.C. Sorensen – C. Yang: *ARPACK Users' Guide*, Software, Environments, Tools. Society for Industrial and Applied Mathematics, 1998.

[12] K. Lust – D. Roose: *Computation and bifurcation analysis of periodic solutions of large-scale systems*, In E. Doedel and L.S. Tuckerman, editors, *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems*. Springer-Verlag, 2000.

[13] K. Lust – D. Roose – A. Spence – A. R. Champneys: *An adaptive Newton-Picard algorithm with subspace iteration for computing periodic solutions*, SIAM Journal on Scientific Computing, **19** (1998), 1188-1209.

[14] D. Roose – S. Vandewalle: *Efficient parallel computation of periodic solutions of partial differential equations*, In R. Seydel and H. Troger, editors, *Bifurcations and Chaos: Analysis, Algorithms, Applications*, ISNM, Birkhäuser Verlag, Basel, 1991, pp. 278-288

[15] Y. Saad: *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, 1996.

[16] R. Seydel: *From Equilibrium to Chaos. Practical Bifurcation and Stability Analysis*, Elsevier, New York, 1988.

[17] B. Werner: *Computation of Hopf bifurcation with bordered matrices*, SIAM J. Numer. Anal., **33** (1996), 435-455.

Lavoro pervenuto alla redazione il 15 febbraio 2003
ed accettato per la pubblicazione il 21 aprile 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo degli Autori:

Eugene Allgower, Ulf Garbotz – Department of Mathematics – Colorado State University – Fort Collins, CO 80523
E-mail: allgower@math.colostate.edu ulf@garbotz.de

Work partially supported by NSF via grant # DMS-9870274.

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 53-67

Representation formulas and Fatou-Kato theorems for heat operators on stratified groups

ANDREA BONFIGLIOLI – FRANCESCO UGUZZONI

Abstract: *In this note, we provide a characterization of non-negative $\mathcal{L}$-caloric functions on strips, where $\mathcal{L}$ is a sub-Laplacian on a stratified group. We prove representation results, Fatou-type and uniqueness theorems analogous to the classical Poisson-Stieltjes formula and to Kato's theorem concerning with positive solutions to the heat equation.*

1 – Introduction and main results

A stratified group is a connected and simply connected Lie group $\mathbb{G}$ whose Lie algebra $\mathfrak{g}$ admits a stratification, i.e., a vector space direct sum decomposition $\mathfrak{g} = \mathfrak{G}_1 \oplus \ldots \oplus \mathfrak{G}_r$ with $[\mathfrak{G}_1, \mathfrak{G}_i] = \mathfrak{G}_{i+1}$, $[\mathfrak{G}_1, \mathfrak{G}_r] = \{0\}$. Stratified groups (also known as Carnot groups) have been introduced by Folland [7] and afterwards deeply studied by various authors, see e.g., Rothschild and Stein [15], Folland and Stein [8], Varopoulos, Saloff-Coste and Coulhon [18]. In particular, Rothschild and Stein pointed out that any Hörmander operator, sum of squares of vector fields, can be locally approximated by a sub-Laplacian on a stratified group. Recently, analysis on such groups has received new and significant impulses in many directions. Indeed, Carnot groups appear as tangent groups of subriemannian manifolds (see e.g., [12]) and they find many applications in mechanics and in control theory. We also refer to the rich bibliography in the recent monographs [1], [9].

Key Words and Phrases: *Carnot groups – Non-negative caloric functions – Fatou and Kato theorems – Uniqueness theorems.*
A.M.S. Classification: 31B25 – 35C15 – 35H20 – 43A80

In this paper, we give a contribution in the study of heat operators $\mathcal{H} = \mathcal{L} - \partial_t$, where $\mathcal{L}$ is a sub-Laplacian on $\mathbb{G}$. The study of parabolic-type operators on $\mathbb{G}$ has experienced an increasing interest, also in relation to some problems from image processing (see e.g., [5], [6], [13]) and from the geometric theory of several complex variables. In particular, the operator $\mathcal{H}$ (on a Carnot group $\mathbb{G}$) intervenes in the study of the linearizations of fully non-linear equations such as the Levi curvature equation [4], [11].

In this note, which is a natural sequel of a study started in [2] (and related to the above mentioned linearizations), we deal with a question left unanswered in [2], giving a characterization of non-negative $\mathcal{L}$-caloric functions and proving some representation formulas. We also prove some results analogous to the classical Fatou-type and uniqueness theorems of KATO [10] concerning with positive solutions to the heat equation. We point out that similar topics have been studied in [3], [14]: in [3], Fatou theory is generalized to the non-negative solutions of some sub-elliptic equations on non-tangentially accessible domain; in [14], Fatou-Kato results are obtained for a class of ultraparabolic Hörmander operators on different homogeneous Lie groups, making use of some Gaussian estimates of the fundamental solution analogous to the ones used here. Our main results are contained in Theorems 1.1, 1.3 and 1.4 below.

We point out that many results presented in this paper are valid in more general contexts. Indeed, the needed tools are mainly a local parabolic Harnack inequality and techniques related to Gaussian bounds (we refer to [18] for such results and related topics on general groups). However, our aim is only to answer to some questions arisen in the study of the above mentioned linearizations rather than to establish an axiomatic theory on the subject. Hence, we shall restrict to the setting of Carnot groups.

THEOREM 1.1. *Let u be a real valued function defined on a strip. The following statements are equivalent*:

(i) *u is a non-negative $\mathcal{L}$-caloric function in some strip $\mathbb{R}^N \times (0, \delta_1)$.*
(ii) *For some Radon measure σ on $\mathbb{R}^N$, u has the representation*

$$u(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\,\mathrm{d}\sigma(\xi), \tag{1}$$

for every (x,t) in some strip $\mathbb{R}^N \times (0,\delta_2)$.

Moreover, if (i)-(ii) *hold, then*

$$u(\cdot,t) \longrightarrow \sigma, \qquad \text{as} \quad t \to 0^+, \quad \text{in the weak sense of measures}, \tag{2}$$

$$u(x,t) \longrightarrow \varphi(x), \qquad \text{as} \quad t \to 0^+, \quad \text{for almost every } x \in \mathbb{R}^N, \tag{3}$$

where $\varphi \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^N)$ is the density of the absolutely continuous part of σ w.r.t. the Lebesgue measure.

Throughout the paper, we call $\mathcal{L}$-*caloric* any solution u of the equation $\mathcal{H}u = 0$, where $\mathcal{H}$ is the heat operator defined in (7) below. Moreover, a Radon measure is understood to be a positive (regular) Borel measure on $\mathbb{R}^N$, finite on compact sets. Finally, by (2) we mean $\int \psi(x)\, u(x,t)\, \mathrm{d}x \longrightarrow \int \psi(x)\, \mathrm{d}\sigma(x)$, as $t \to 0^+$, for every continuous function ψ with compact support. The other notations are explained below. The following remark shows how the strips in (i) and (ii) of the above theorem are related.

REMARK 1.2. If (i) holds then we have the representation (1) in the whole strip $\mathbb{R}^N \times (0, \delta_1)$ for a Radon measure σ on $\mathbb{R}^N$ satisfying the growth condition

$$\int_{\mathbb{R}^N} \exp(-\nu\, d^2(\xi))\, \mathrm{d}\sigma(\xi) < \infty, \tag{4}$$

where $\nu = \mathbf{c}/\delta_1$; vice-versa, if (ii) holds, then the measure σ satisfies (4) with $\nu = \mathbf{c}/\delta_2$ and (i) follows with $\delta_1 = \delta_2/\mathbf{c}^2$. Here $\mathbf{c} > 0$ is a structural constant only depending on $\mathcal{L}$.

The following result is a step in the proof of Theorem 1.1, beside being of its own interest.

THEOREM 1.3. *Let u be a non-negative $\mathcal{L}$-caloric function in $\mathbb{R}^N \times (0,T)$. Then, for every $\varepsilon > 0$, we have the following Poisson-Stieltjes type representation formula*

$$u(x, t+\varepsilon) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\, u(\xi,\varepsilon)\, \mathrm{d}\xi, \qquad (x,t) \in \mathbb{R}^N \times (0, T-\varepsilon). \tag{5}$$

From Theorem 1.1 and using as a main step Lemma 2.7 in the next section, we can also derive the following Kato-type uniqueness result.

THEOREM 1.4. *Let u be a non-negative $\mathcal{L}$-caloric function in $\mathbb{R}^N \times (0,T)$. If*

$$\lim_{t\to 0^+} u(x,t) = 0 \qquad \textit{for almost every } x \in \mathbb{R}^N,$$
$$\limsup_{t\to 0^+} u(x,t) < \infty \quad \textit{for every } x \in \mathbb{R}^N,$$

then u vanishes identically.

We explicitly remark that the lim sup-condition in Theorem 1.4 cannot be weakened, as one can easily realize taking $u = \Gamma$.

We now explain all the notation. First of all, we give an operative definition of Carnot group. Our definition is equivalent to the one of Folland, up to isomorphism. Let $\circ$ be an assigned Lie group law on $\mathbb{R}^N$. Suppose $\mathbb{R}^N$ is endowed with a homogeneous structure by a given family of Lie group automorphisms $\{\delta_\lambda\}_{\lambda>0}$ (called dilations) of the form

$$\delta_\lambda(x) = \delta_\lambda(x^{(1)}, \dots, x^{(r)}) = (\lambda x^{(1)}, \dots, \lambda^r x^{(r)}).$$

Here $x^{(i)} \in \mathbb{R}^{N_i}$ for $i = 1, \dots, r$ and $N_1 + \dots + N_r = N$. We denote by $\mathfrak{g}$ the Lie algebra of $(\mathbb{R}^N, \circ)$. For $i = 1, \dots, N_1$, let X_i be the vector field in $\mathfrak{g}$ that agrees at the origin with $\partial/\partial x_i^{(1)}$. We make the following assumption: the Lie algebra generated by $X_1, \dots, X_{N_1}$ is the whole $\mathfrak{g}$. With the above hypotheses, we call $\mathbb{G} = (\mathbb{R}^N, \circ, \delta_\lambda)$ a Carnot group. If $Y_1, \dots, Y_{N_1}$ is any basis for $\operatorname{span}\{X_1, \dots, X_{N_1}\}$, the second order differential operator

$$\mathcal{L} = \sum_{i=1}^{N_1} Y_i^2$$

is called a sub-Laplacian on $\mathbb{G}$. Since $X_1, \dots, X_{N_1}$ generate the whole $\mathfrak{g}$, which has rank N at every point, any sub-Laplacian $\mathcal{L}$ satisfies Hörmander's hypoellipticity condition. We denote by $Q = \sum_{j=1}^r jN_j$ the homogeneous dimension of $\mathbb{G}$. Then $|\delta_\lambda(E)| = \lambda^Q |E|$ for any measurable set E. Here and in the sequel, we denote by $|\cdot|$ the Lebesgue measure on $\mathbb{R}^N$. This measure is invariant w.r.t. the left and right translations on $\mathbb{G}$.

The simplest example of Carnot group is the additive Euclidean group $(\mathbb{R}^Q, +)$; in this case, the sub-Laplacians are exactly the constant coefficient elliptic operators. The most significant (and simple) non-abelian example of Carnot group is the Heisenberg group; in this case, a remarkable sub-Laplacian is the real part of the Kohn-Spencer Laplacian.

Throughout the paper, d will denote a fixed homogeneous norm on $\mathbb{G}$. For instance, we choose $d = \gamma^{1/(2-Q)}$, where γ denotes the fundamental solution of the sub-Laplacian $\sum_{i=1}^{N_1} X_i^2$. We recall that a homogeneous norm on $\mathbb{G}$ is a continuous function $d : \mathbb{R}^N \to [0, \infty)$, smooth away from the origin, such that $d(\delta_\lambda(x)) = \lambda\, d(x)$, $d(x^{-1}) = d(x)$, and $d(x) = 0$ iff $x = 0$. Hereafter, we also denote $d(y^{-1} \circ x)$ by $d(x, y)$ and use the notation $B_d(x, r)$ for the d-ball of center x and radius r. The following quasi-triangle inequality holds

$$d(x, y) \le \beta\,(d(x, z) + d(z, y)), \qquad x,\, y,\, z \in \mathbb{G}, \tag{6}$$

for a suitable constant β. Throughout the sequel, $\mathcal{L}$ will always denote a fixed sub-Laplacian on $\mathbb{G}$ and

$$\mathcal{H} = \mathcal{L} - \partial_t \tag{7}$$

the related heat operator on $\mathbb{G}\times\mathbb{R}\equiv\mathbb{R}^{N+1}$. Here $z=(x,t)$ is the point of $\mathbb{R}^{N+1}$ ($x\in\mathbb{G}$, $t\in\mathbb{R}$). The operator $\mathcal{H}$ is hypoelliptic by Hörmander theorem.

It is known that $\mathcal{H}$ possesses a fundamental solution with the properties recalled below (see [8], [18]; see also [2]). There exists a smooth function Γ on $\mathbb{R}^{N+1}\setminus\{0\}$ such that the fundamental solution for $\mathcal{H}$ is given by

$$\Gamma(x,t;\xi,\tau)=\Gamma(\xi^{-1}\circ x,t-\tau).$$

We have $\Gamma(x,t)\geq 0$ and $\Gamma(x,t)=0$ iff $t\leq 0$; moreover

$$\Gamma(x,t)=\Gamma(x^{-1},t),\qquad \Gamma(\delta_\lambda(x),\lambda^2 t)=\lambda^{-Q}\Gamma(x,t). \tag{8}$$

For every $\zeta\in\mathbb{R}^{N+1}$, $\Gamma(\cdot;\zeta)$ is locally integrable and $\mathcal{H}\Gamma(\cdot,\zeta)=-\delta_\zeta$ (the Dirac measure supported at $\{\zeta\}$). For every $x\in\mathbb{R}^N$, t, $\tau>0$, we have

$$\int_{\mathbb{R}^N}\Gamma(\xi,t)\,\mathrm{d}\xi=1,\qquad \Gamma(x,t+\tau)=\int_{\mathbb{R}^N}\Gamma(\xi^{-1}\circ x,t)\,\Gamma(\xi,\tau)\,\mathrm{d}\xi. \tag{9}$$

The main tool we shall employ in the proofs of our results is the following Gaussian estimate of Γ: there exists a positive constant $\mathbf{c}_0$ such that

$$\mathbf{c}_0^{-1}\,t^{-Q/2}\exp\Big(-\frac{\mathbf{c}_0\,d^2(x)}{t}\Big)\leq\Gamma(x,t)\leq\mathbf{c}_0\,t^{-Q/2}\exp\Big(-\frac{d^2(x)}{\mathbf{c}_0\,t}\Big), \tag{10}$$

for every $x\in\mathbb{R}^N$, $t>0$. We finally recall the following result, related to the Cauchy problem for $\mathcal{H}$ (for the proof we refer to the results in [18] and to the classical method of Aronson; see also [2]).

THEOREM 1.5. (i) *Let f be a continuous function on $\mathbb{R}^N$ satisfying the growth condition $|f(x)|\leq c\exp(\nu\,d^2(x))$, for some constants c, $\nu\geq 0$. Then the function*

$$u(x,t)=\int_{\mathbb{R}^N}\Gamma(x,t;\xi,0)\,f(\xi)\,\mathrm{d}\xi,\qquad x\in\mathbb{R}^N,\ t\in(0,(\mathbf{c}\nu)^{-1})$$

is well posed and is a classical solution to the Cauchy problem

$$\mathcal{H}u=0\ \text{ in }\ \mathbb{R}^N\times(0,(\mathbf{c}\nu)^{-1}),\qquad u(\cdot,0)=f.$$

Here $\mathbf{c}$ is a positive constant only depending on $\mathcal{L}$ and the structure of $\mathbb{G}$.

(ii) *Let u be a classical solution to the Cauchy problem*

$$\mathcal{H}u=0\ \text{ in }\ \mathbb{R}^N\times(0,r),\quad u(\cdot,0)=0.$$

Suppose that one of the following conditions holds: either u is non-negative or there exists $\nu>0$ such that

$$\int_0^r\int_{\mathbb{R}^N}\exp\big(-\nu\,d^2(x)\big)\,|u(x,t)|\,\mathrm{d}x\,\mathrm{d}t<\infty.$$

Then u vanishes identically.

2 – Fatou-Kato theorems

For the reader convenience, we first recall the following weak maximum principle on strips, whose proof is standard and will be omitted.

PROPOSITION 2.1. *Let $u \in C^2(\mathbb{R}^N \times (0,T))$. If $\mathcal{H}u \geq 0$, $\limsup u \leq 0$ both in $\mathbb{R}^N \times \{0\}$ and at infinity, then $u \leq 0$ in the whole strip.*

In the sequel, we shall need the following Harnack theorem for $\mathcal{H}$, whose proof easily follows from the Harnack inequality in [18].

THEOREM 2.2. *Let us fix $T > 0$ and set $S_T = \mathbb{R}^N \times (0,T)$.*

(i) *For every $z_0 = (x_0, t_0) \in S_T$ and for every compact set $K \subset S_{t_0}$, there exists a positive constant $\mathbf{c}$ such that*

$$\sup_K u \leq \mathbf{c}\, u(z_0),$$

for every non-negative function u, $\mathcal{L}$-caloric in S_T.

(ii) *Let $u_n \leq u_{n+1}$ be a monotone sequence of $\mathcal{L}$-caloric functions in S_T. If there exists $z_0 = (x_0, t_0) \in S_T$ such that $u_n(z_0)$ is bounded, then u_n converges uniformly on the compact subsets of S_{t_0} to a function u, $\mathcal{L}$-caloric in S_{t_0}.*

We are now able to prove the Poisson-Stieltjes type representation formula.

PROOF OF THEOREM 1.3. For every $n \in \mathbb{N}$, we set

$$v_n(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\, \psi\Big(\frac{d(\xi)}{n}\Big)\, u(\xi,\varepsilon)\, \mathrm{d}\xi,$$

where $\psi \in C^\infty(\mathbb{R})$ is a fixed non-increasing cut-off function such that $\psi(r) = 1$ if $r \leq 1$, $\psi(r) = 0$ if $r \geq 2$. By Theorem 1.5, we know that v_n is a solution to the Cauchy problem

$$\begin{cases} \mathcal{H} v_n = 0 \quad \text{in } \mathbb{R}^N \times (0,\infty), \\ v_n(\cdot,0) = \psi\Big(\dfrac{d(\cdot)}{n}\Big)\, u(\cdot,\varepsilon). \end{cases}$$

Moreover, for every $t \in (0,T)$, we have (by the estimates in (10))

$$\begin{aligned} 0 \leq v_n(x,t) &\leq \mathbf{c}_0\, t^{-Q/2} \int_{\mathbb{R}^N} \exp\Big(-\frac{d^2(x,\xi)}{\mathbf{c}_0\, t}\Big)\, \psi\Big(\frac{d(\xi)}{n}\Big)\, u(\xi,\varepsilon)\, \mathrm{d}\xi \leq \\ &\leq \mathbf{c}_0 \int_{d(x\circ\delta_{\sqrt{t}}\eta)\leq 2n} \exp\Big(-\frac{d^2(\eta)}{\mathbf{c}_0}\Big)\, u(x\circ\delta_{\sqrt{t}}\eta,\varepsilon)\, \mathrm{d}\eta \leq \\ &\leq \mathbf{c}_0 \max_{B_d(0,2n)} u(\cdot,\varepsilon) \int_{d(\eta)\geq(\beta^{-1}d(x)-2n)/\sqrt{T}} \exp(-d(\eta)^2/\mathbf{c}_0)\mathrm{d}\eta \to 0, \text{ as } d(x)\to\infty. \end{aligned}$$

We now apply the weak maximum principle for $\mathcal{H}$ to the $\mathcal{L}$-caloric function $w_n(x,t) = u(x,t+\varepsilon) - v_n(x,t)$ in the strip $\mathbb{R}^N \times (0, T-\varepsilon)$. Since $\psi \le 1$, we have $w_n(\cdot,0) \ge 0$. Moreover, we have proved that v_n vanishes at infinity in the strip. Hence, recalling that u is non-negative, we get $\liminf w_n \ge 0$ at infinity in the strip. The maximum principle of Proposition 2.1 then yields $w_n \ge 0$ in $\mathbb{R}^N \times (0, T-\varepsilon)$. Recalling the definition of w_n and letting n go to infinity, from the above inequality, we finally obtain

$$u(x,t+\varepsilon) \ge \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\, u(\xi,\varepsilon)\, \mathrm{d}\xi =: v_\infty(x,t), \quad (x,t) \in \mathbb{R}^N \times (0,T-\varepsilon),$$

since $v_n \nearrow v_\infty$ by monotone convergence. This proves in particular that v_∞ is finite in $\mathbb{R}^N \times (0, T-\varepsilon)$. Now, from the Harnack Theorem 2.2-(ii), it follows that v_∞ is $\mathcal{L}$-caloric in $\mathbb{R}^N \times (0, T-\varepsilon)$. Moreover, from the inequalities

$$v_n(x,t) \le v_\infty(x,t) \le u(x,t+\varepsilon), \quad (x,t) \in \mathbb{R}^N \times (0,T-\varepsilon),$$

and recalling that $v_n(x,0) = u(x,\varepsilon)$ if $d(x) \le n$, it follows that v_∞ is continuous in $\mathbb{R}^N \times [0, T-\varepsilon)$ and $v_\infty(\cdot,0) = u(\cdot,\varepsilon)$. As a consequence, setting $w_\infty(x,t) = u(x,t+\varepsilon) - v_\infty(x,t)$, w_∞ is a classical solution to $\mathcal{H} w_\infty = 0$ in $\mathbb{R}^N \times (0, T-\varepsilon)$, $w_\infty(\cdot,0) = 0$. Since moreover w_∞ is non-negative, it must vanish identically, by the uniqueness result in Theorem 1.5. This proves (5). ∎

We now turn to the proof of Theorem 1.1 which is split in various steps, starting with Lemma 2.3 below.

LEMMA 2.3. *Let u be a non-negative $\mathcal{L}$-caloric function in the strip $\mathbb{R}^N \times (0,T)$. Then, there exists a Radon measure σ on $\mathbb{R}^N$ such that*

$$u(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\, \mathrm{d}\sigma(\xi), \qquad (x,t) \in \mathbb{R}^N \times (0,T), \tag{11}$$

$$\int_{\mathbb{R}^N} \exp\Big(-\frac{2\mathbf{c}_0}{T}\, d^2(\xi)\Big)\, \mathrm{d}\sigma(\xi) < \infty, \tag{12}$$

where $\mathbf{c}_0 > 0$ is the constant in (10).

PROOF. Let us fix $t_0 \in (0,T)$ and choose $j_0 \in \mathbb{N}$ such that $t_0 < T - 1/j_0$. From Theorem 1.3, it follows that

$$u\Big(x, t+\frac{1}{j}\Big) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\, u\Big(\xi, \frac{1}{j}\Big)\, \mathrm{d}\xi, \qquad (x,t) \in \mathbb{R}^N \times \Big(0, T - \frac{1}{j_0}\Big), \quad j > j_0.$$

Since $u(0, t_0 + \frac{1}{j}) \longrightarrow u(0,t_0)$ as $j \to \infty$, the sequence of Radon measures on $\mathbb{R}^N$

$$\mathrm{d}\mu_j(\xi) = \Gamma(0,t_0;\xi,0)\, u\Big(\xi, \frac{1}{j}\Big)\, \mathrm{d}\xi, \qquad j > j_0,$$

is bounded and hence weakly converges (up to a subsequence) to a certain Radon measure μ (with $\mu(\mathbb{R}^N) < \infty$) in the sense that

$$\int_{\mathbb{R}^N} f(\xi)\,\mathrm{d}\mu_j(\xi) \longrightarrow \int_{\mathbb{R}^N} f(\xi)\,\mathrm{d}\mu(\xi), \quad \text{for every } f \in C_{(0)}(\mathbb{R}^N) \tag{13}$$

(we have denoted by $C_{(0)}(\mathbb{R}^N)$ the space of continuous functions in $\mathbb{R}^N$, vanishing at infinity). We now set $M = 2\,\beta^2\,\mathbf{c}_0^2$, where β is defined by (6) and $\mathbf{c}_0$ is the constant in (10). For every $(x,t) \in \mathbb{R}^N \times (0, t_0/M)$, we have

$$u(x,t) = \lim_{j\to\infty} u\Big(x, t+\frac{1}{j}\Big) = \lim_{j\to\infty} \int_{\mathbb{R}^N} \frac{\Gamma(x,t;\xi,0)}{\Gamma(0,t_0;\xi,0)}\,\mathrm{d}\mu_j(\xi) = \int_{\mathbb{R}^N} \frac{\Gamma(x,t;\xi,0)}{\Gamma(0,t_0;\xi,0)}\,\mathrm{d}\mu(\xi),$$

by (13), observing that $\Gamma(x,t;\cdot,0)/\Gamma(0,t_0;\cdot,0) \in C_{(0)}(\mathbb{R}^N)$, since the estimates in (10) give, for $d(\xi) \geq 4\beta\, d(x)$,

$$\begin{aligned} 0 < \frac{\Gamma(x,t;\xi,0)}{\Gamma(0,t_0;\xi,0)} &\leq \mathbf{c}(t,t_0)\,\exp\Big(\frac{\mathbf{c}_0\,d^2(\xi)}{t_0} - \frac{d^2(x,\xi)}{\mathbf{c}_0\,t}\Big) \leq \\ &\leq \mathbf{c}(t,t_0)\,\exp\Big(\frac{\mathbf{c}_0\,d^2(\xi)}{t_0} - \frac{1}{\mathbf{c}_0\,t}\Big(\frac{d^2(\xi)}{\beta^2} + d^2(x) - \frac{2\,d(x)\,d(\xi)}{\beta}\Big)\Big) \leq \\ &\leq \mathbf{c}(x,t,t_0)\,\exp\Big(-d^2(\xi)\Big(\frac{1}{2\,\beta^2\,\mathbf{c}_0\,t} - \frac{\mathbf{c}_0}{t_0}\Big)\Big) \longrightarrow 0, \end{aligned}$$

as $d(\xi) \to \infty$, if $t < t_0/M$. Choosing

$$\mathrm{d}\sigma(\xi) = \frac{\mathrm{d}\mu(\xi)}{\Gamma(0,t_0;\xi,0)}, \tag{14}$$

we get (11) in the strip $\mathbb{R}^N \times (0, t_0/M)$. In order to extend the representation formula to the whole strip $\mathbb{R}^N \times (0,T)$, we shall exploit (9). For fixed $T > t \geq t_0/M > \varepsilon > 0$, by Theorem 1.3 we have

$$\begin{aligned} u(x,t) &= \int_{\mathbb{R}^N} \Gamma(x,t-\varepsilon;\xi,0)\,u(\xi,\varepsilon)\,\mathrm{d}\xi = \\ &= \int_{\mathbb{R}^N} \Big(\int_{\mathbb{R}^N} \Gamma(x,t-\varepsilon;\xi,0)\,\Gamma(\xi,\varepsilon;y,0)\,\mathrm{d}\xi\Big)\mathrm{d}\sigma(y) = \int_{\mathbb{R}^N} \Gamma(x,t;y,0)\,\mathrm{d}\sigma(y). \end{aligned}$$

We explicitly remark that σ is finite on the compact sets by the estimates in (10) and recalling that $\mu(\mathbb{R}^N) < \infty$. Moreover, again using (10) and from (11), it follows that

$$u\Big(0,\frac{T}{2}\Big) \geq \mathbf{c}_0^{-1}\Big(\frac{T}{2}\Big)^{-\frac{Q}{2}} \int_{\mathbb{R}^N} \exp\Big(-\frac{2\,\mathbf{c}_0\,d^2(\xi)}{T}\Big)\,\mathrm{d}\sigma(\xi),$$

which gives (12). □

THEOREM 2.4. *Let u be a non-negative $\mathcal{L}$-caloric function in the strip $\mathbb{R}^N \times (0,T)$. Then, there exists a non-negative function $\varphi \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^N)$ such that*

$$u(x,t) \longrightarrow \varphi(x), \quad \text{as } t \to 0^+, \quad \text{for almost every } x \in \mathbb{R}^N.$$

PROOF. Let σ be the Radon measure found in Lemma 2.3. By the Lebesgue decomposition theorem, there exists a non-negative function $\varphi \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^N)$ and a singular Radon measure s on $\mathbb{R}^N$ such that

$$\mathrm{d}\sigma(\xi) = \varphi(\xi)\,\mathrm{d}\xi + \mathrm{d}s(\xi). \tag{15}$$

Moreover, for a.e. $x \in \mathbb{R}^N$ (w.r.t. the Lebesgue measure), we have

$$\frac{1}{|B_d(x,\rho)|}\int_{B_d(x,\rho)} |\varphi(\xi)-\varphi(x)|\,\mathrm{d}\xi \longrightarrow 0, \quad \frac{s(B_d(x,\rho))}{|B_d(x,\rho)|} \longrightarrow 0, \text{ as } \rho \to 0^+. \tag{16}$$

The proof of (16) will be omitted. It follows e.g. adapting the arguments in [16, Chapter 8], replacing the Euclidean metric by the quasi-distance d. The doubling property of the d-balls ensures, for instance, a suitable d-version of the Vitali covering lemma (see e.g. [17]).

Let us now fix an $x \in \mathbb{R}^N$ where (16) holds and set, for brevity, $\mathrm{d}\alpha(\xi) = |\varphi(\xi)-\varphi(x)|\,\mathrm{d}\xi + \mathrm{d}s(\xi)$. Also fix $\varepsilon > 0$. Then there exists $\rho_0 \in (0,\sqrt{T})$ such that

$$\frac{1}{|B_d(x,\rho)|}\int_{B_d(x,\rho)} \mathrm{d}\alpha(\xi) < \varepsilon, \qquad \text{for every } \rho \in (0, 2\,\rho_0]. \tag{17}$$

Let now $t \in (0,\rho_0^2)$ and let $N(t) \in \mathbb{N}$ be such that $2^{N(t)-1} \le \rho_o/\sqrt{t} < 2^{N(t)}$. From Lemma 2.3, (9) and (15), we obtain

$$\begin{aligned}
&|u(x,t)-\varphi(x)| \le \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\,\mathrm{d}\alpha(\xi) \le \\
&\le \Big(\int_{B_d(x,\sqrt{t})} + \sum_{j=1}^{N(t)} \int_{2^{j-1}\sqrt{t}\le d(x,\xi)<2^j\sqrt{t}} + \int_{d(x,\xi)>\rho_0} \Big)\Gamma(x,t;\xi,0)\,\mathrm{d}\alpha(\xi) = \\
&= \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3.
\end{aligned}$$

Using (17) and the estimates in (10), we get

$$\mathrm{I}_1 \le \frac{\mathbf{c}}{|B_d(x,\sqrt{t})|}\int_{B_d(x,\sqrt{t})} \exp\Big(-\frac{d^2(x,\xi)}{\mathbf{c}_0\,t}\Big)\,\mathrm{d}\alpha(\xi) \le \mathbf{c}\,\varepsilon,$$

recalling that $\sqrt{t} < \rho_0$. In the same way, we can prove the estimate

$$\begin{aligned} \mathrm{I}_2 &\le \mathbf{c} \sum_{j=1}^{N(t)} \frac{\exp(-4^{j-1}\mathbf{c}_0^{-1})\, 2^{j\,Q}}{|B_d(x, 2^j\sqrt{t})|} \int_{B_d(x,2^j\sqrt{t})} \mathrm{d}\alpha(\xi) \le \\ &\le \mathbf{c}\,\varepsilon \sum_{j=1}^{\infty} \exp(-4^{j-1}\mathbf{c}_0^{-1})\, 2^{j\,Q} = \mathbf{c}'\,\varepsilon, \end{aligned}$$

recalling that $2^j\sqrt{t} \le 2\,\rho_0$ for every $j \le N(t)$. Finally, using again (10) and recalling the definition (14) of σ, we have

$$\begin{aligned} \mathrm{I}_3 &\le \int_{d(x,\xi)>\rho_0} \Gamma(x,t;\xi,0)\,\mathrm{d}\sigma(\xi) + \varphi(x) \int_{d(x,\xi)>\rho_0} \Gamma(x,t;\xi,0)\,\mathrm{d}\xi \le \\ &\le \mathbf{c}(t_0) \int_{d(x,\xi)>\rho_0} t^{-\frac{Q}{2}}\, \exp\Big(-\frac{d^2(x,\xi)}{\mathbf{c}_0\, t} + \frac{\mathbf{c}_0\, d^2(\xi)}{t_0}\Big)\,\mathrm{d}\mu(\xi) + \\ &\quad + \mathbf{c}\,\varphi(x) \int_{d(\eta)>\frac{\rho_0}{\sqrt{t}}} \exp\big(-\mathbf{c}_0^{-1}\, d^2(\eta)\big)\,\mathrm{d}\eta, \end{aligned}$$

and then it is easy to see that I_3 vanishes as $t \to 0^+$. This concludes the proof. □

In order to complete the proof of Theorem 1.1, we are only left to prove Lemma 2.5 and Lemma 2.6 below.

LEMMA 2.5. *Let σ be a Radon measure on $\mathbb{R}^N$ satisfying the growth condition*

$$\int_{\mathbb{R}^N} e^{-\nu\, d^2(\xi)}\,\mathrm{d}\sigma(\xi) < \infty, \tag{18}$$

for some constant $\nu > 0$. Then the function

$$u(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\,\mathrm{d}\sigma(\xi) \tag{19}$$

is $\mathcal{L}$-caloric in the strip $\mathbb{R}^N \times (0, (\mathbf{c}^*\nu)^{-1})$, where $\mathbf{c}^*$ is a positive constant only depending on $\mathcal{L}$ and the structure of $\mathbb{G}$.

PROOF. From the estimate (10) and recalling that σ is finite on compact sets, it follows

$$\int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\,\mathrm{d}\sigma(\xi) \le \mathbf{c}_0 t^{-Q/2} \int_{\mathbb{R}^N} \exp\Big(-\frac{d^2(x,\xi)}{\mathbf{c}_0 t}\Big)\mathrm{d}\sigma(\xi) \le$$
$$\le \mathbf{c}(x,t) + \mathbf{c}'(t) \int_{d(\xi)>4\beta d(x)} \exp\Big(-\frac{d^2(\xi)}{4\mathbf{c}_0\beta^2 t}\Big)\,\mathrm{d}\sigma(\xi) < \infty,$$

if $0 < t < (4\mathbf{c}_0\beta^2\nu)^{-1} =: (\mathbf{c}^*\nu)^{-1}$. Moreover, by dominated convergence, it is easy to see that u is continuous on the strip $\mathbb{R}^N \times (0, (\mathbf{c}^*\nu)^{-1})$. In order to prove that u is $\mathcal{L}$-caloric, one can differentiate under the integral sign, making use of the estimates of the derivatives of Γ along the vector fields $X_1, \dots, X_{N_1}$ (see e.g., [18]; see also [2]). Alternatively, one can use the Harnack Theorem 2.2, following the lines of the proof of Theorem 1.3: the function

$$v_{n,\varepsilon}(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t-\varepsilon;\xi,0)\,\psi\Big(\frac{d(\xi)}{n}\Big)\,u(\xi,\varepsilon)\,\mathrm{d}\xi,$$

is a solution to $\mathcal{H}v_{n,\varepsilon} = 0$ in $\mathbb{R}^N \times (\varepsilon,\infty)$, $v_{n,\varepsilon}(x,\varepsilon) = \psi(d(x)/n)\,u(x,\varepsilon)$; moreover, recalling that $0 \le \psi \le 1$, (9) and the definition (19) of u, we have

$$v_{n,\varepsilon}(x,t) \le \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \Gamma(x,t-\varepsilon;\xi,0)\,\Gamma(\xi,\varepsilon;y,0)\,\mathrm{d}\xi\mathrm{d}\sigma(y) =$$
$$= \int_{\mathbb{R}^N} \Gamma(x,t;y,0)\,\mathrm{d}\sigma(y) = u(x,t) < \infty, \qquad \text{if } t < 1/(\mathbf{c}^*\nu);$$

hence, by Theorem 2.2, $v_{\infty,\varepsilon} = \lim_{n\to\infty} v_{n,\varepsilon}$ is $\mathcal{L}$-caloric in $\mathbb{R}^N \times (\varepsilon, 1/(\mathbf{c}^*\nu))$; finally, using again (9), we see that $v_{\infty,\varepsilon}(x,t) = \int_{\mathbb{R}^N} \Gamma(x,t-\varepsilon;\xi,0)\,u(\xi,\varepsilon)\,\mathrm{d}\xi = u(x,t)$ in $\mathbb{R}^N \times (\varepsilon, 1/(\mathbf{c}^*\nu))$; since ε is arbitrary, this ends the proof. □

LEMMA 2.6. *Under the hypotheses of Lemma* 2.5 *above, we have*

$$u(\cdot,t) \longrightarrow \sigma, \quad \text{as } t \to 0^+, \text{ in the weak sense of measures.}$$

PROOF. Let $f \in C_0(\mathbb{R}^N)$. We have to prove that

$$\int_{\mathbb{R}^N} f(x)\,u(x,t)\,\mathrm{d}x \longrightarrow \int_{\mathbb{R}^N} f(x)\,\mathrm{d}\sigma(x), \qquad \text{as } t \to 0^+.$$

For small $t > 0$, we have (see (8))

$$\int_{\mathbb{R}^N} f(x)\,u(x,t)\,\mathrm{d}x = \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \Gamma(x,t;\xi,0)\,f(x)\,\mathrm{d}x\,\mathrm{d}\sigma(\xi) =$$
$$= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \Gamma(\xi,t;x,0)\,f(x)\,\mathrm{d}x\,\mathrm{d}\sigma(\xi).$$

Moreover, $\int_{\mathbb{R}^N} \Gamma(\xi,t;x,0)\, f(x)\,\mathrm{d}x \longrightarrow f(\xi)$, as $t \to 0^+$, by Theorem 1.5. Hence it is sufficient to prove that

$$\Big| \int_{\mathbb{R}^N} \Gamma(\xi,t;x,0)\, f(x)\,\mathrm{d}x \Big| \le \mathbf{c}(f)\, e^{-\nu\, d^2(\xi)} \tag{20}$$

holds for every $\xi \in \mathbb{R}^N$ and for every small $t > 0$, and then to use the dominated convergence (we recall that (18) holds). Let us set $k_0 = 4\beta \max_{\mathrm{supp}(f)} d$. Since (9) holds, the integral in the left-hand side of (20) is clearly uniformly bounded for $d(\xi) \le k_0$. On the other hand, if $d(\xi) > k_0$, the estimate (10) gives

$$\begin{aligned}
&\Big| \int_{\mathbb{R}^N} \Gamma(\xi,t;x,0)\, f(x)\,\mathrm{d}x \Big| \le \\
&\le \mathbf{c}(f) \exp\Big(-\frac{d^2(\xi)}{\mathbf{c}_0\beta^2 t} + \frac{k_0\, d(\xi)}{2\mathbf{c}_0\beta^2 t} \Big) \int_{\mathrm{supp}(f)} t^{-Q/2} \exp\Big(-\frac{d^2(x)}{\mathbf{c}_0 t} \Big)\,\mathrm{d}x \le \\
&\le \mathbf{c}(f)\, \exp\Big(-\frac{d^2(\xi)}{2\mathbf{c}_0\beta^2 t} \Big) \int_{\mathbb{R}^N} e^{-d^2(\eta)/\mathbf{c}_0}\,\mathrm{d}\eta = \mathbf{c}'(f) \exp\Big(-\frac{d^2(\xi)}{2\mathbf{c}_0\beta^2 t} \Big),
\end{aligned}$$

which finally yields (20) for sufficiently small t. □

Proof of Theorem 1.1. It directly follows collecting Lemma 2.3, Theorem 2.4, Lemma 2.5 and Lemma 2.6. □

Finally, we have to prove Theorem 1.4; our main tool will be Lemma 2.7 below. First, we fix a notation. Given a Radon measure σ on $\mathbb{R}^N$, we define the *upper d-symmetric derivative* of σ at $x \in \mathbb{R}^N$,

$$d\text{-}\overline{D}_{\mathrm{sym}}\sigma(x) = \limsup_{\rho\to 0^+} \frac{\sigma(B_d(x,\rho))}{|B_d(x,\rho)|}. \tag{21}$$

The following result generalizes [10, Lemma 1].

Lemma 2.7. *Let σ be a Radon measure on $\mathbb{R}^N$ such that $d\text{-}\overline{D}_{\mathrm{sym}}\sigma(x) < \infty$ for every $x \in \mathbb{R}^N$. Then σ is absolutely continuous w.r.t. the Lebesgue measure.*

Proof. We assume by contradiction that there exists a Borel set $E \subseteq \mathbb{R}^N$ such that $|E| = 0$ and $\sigma(E) > 0$. From the hypotheses, we infer that $E = \cup_{n\in\mathbb{N}} E_n$, where $E_n = \{x \in E \,|\, d\text{-}\overline{D}_{\mathrm{sym}}\sigma(x) < n\}$. Hence there exists $n_0 \in \mathbb{N}$ such that $\sigma(E_{n_o}) > 0$. Moreover, by the definition of $d\text{-}\overline{D}_{\mathrm{sym}}$, we have $E_{n_0} = \cup_{j\in\mathbb{N}} A_j$, where we have set

$$A_j = \{x \in E_{n_0} \,|\, \sup_{0<\rho<1/j} \sigma(B_d(x,\rho))/|B(x,\rho)| < n_0\}.$$

Thus, there exists $j_0 \in \mathbb{N}$ such that $\sigma(A_{j_o}) > 0$. From the regularity of σ, it follows that there exists a compact set K such that $K \subseteq A_{j_0}$ ($\subseteq E_{n_0} \subseteq E$), $\sigma(K) > 0$. Clearly we have

$$\sigma(B_d(x,\rho)) < n_0\,|B_d(x,\rho)|, \qquad \text{for every } x \in K,\ 0 < \rho < 1/j_0. \tag{22}$$

We now fix $\varepsilon > 0$. Since $|E| = 0$ gives $|K| = 0$, there exists an open set V such that $K \subset V$, $|V| < \varepsilon$. We claim that there exists a disjoint family of d-balls $\{B_b(x_i,\delta)\}_{i=1}^p$ with the following properties: $x_i \in K$, $0 < \delta < (4\beta^2 j_0)^{-1}$, $K \subseteq \bigcup_{i=1}^p B_d(x_i, 4\beta^2\delta) \subseteq V$. As a consequence, by (22), we obtain

$$\begin{aligned} 0 < \sigma(K) &\le \sum_{i=1}^p \sigma(B_d(x_i,4\beta^2\delta)) \le n_0 \sum_{i=1}^p |B_d(x_i,4\beta^2\delta)| = \\ &= n_0\,(4\beta^2)^Q \sum_{i=1}^p |B_d(x_i,\delta)| = \\ &= n_0\,(4\beta^2)^Q \Big|\bigcup_{i=1}^p B_d(x_i,\delta)\Big| \le n_0\,(4\beta^2)^Q\,|V| < n_0\,(4\beta^2)^Q\,\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this gives a contradiction. Thus, in order to complete the proof, we only have to prove the claim. Let $\{\xi_n\}_n$ be a countable dense subset of K and let us choose a positive δ not exceeding $(4\beta^2 j_0)^{-1}$, such that $4\beta^2\delta < \min\{d(x,y) \mid x \in K,\, y \in \mathbb{R}^N \setminus V\}$ so that $B_d(x, 4\beta^2\delta) \subseteq V$ for every $x \in K$. We set $x_1 = \xi_1$. If $\{\xi_n\}_n \subset \overline{B_d(x_1, 2\beta\delta)}$, then $K = \overline{\{\xi_n\}_n} \subseteq \overline{B_d(x_1, 2\beta\delta)} \subset B_d(x_1, 4\beta^2\delta)$. Otherwise, let $n_2 \in \mathbb{N}$ be such that $\xi_1, \dots, \xi_{n_2-1} \in \overline{B_d(x_1, 2\beta\delta)}$, $\xi_{n_2} \notin \overline{B_d(x_1, 2\beta\delta)}$. Setting $x_2 = \xi_{n_2}$, we clearly have $B_d(x_1,\delta) \cap B_d(x_2,\delta) = \emptyset$ (we recall that β is defined by (6)). Iterating this procedure, we obtain a (possibly finite) subsequence $\{x_i = \xi_{n_i}\}_i$ of $\{\xi_n\}_n$ and a sequence of disjoint d-balls $\{B_d(x_i,\delta)\}_i$ such that $\{\xi_n\}_n \subset \cup_i \overline{B_d(x_i, 2\beta\delta)}$. This gives

$$K = \overline{\{\xi_n\}_n} \subseteq \overline{\bigcup_i B_d(x_i, 2\beta\delta)} \subset \bigcup_i B_d(x_i, 4\beta^2\delta)$$

(the radius of $B_d(x_i, 2\beta\delta)$ has been chosen not depending on i, in order to allow this last inclusion). The claim is proved by taking a finite sub-covering. ∎

With Lemma 2.7 at hands, we are able to prove our uniqueness result.

Proof of Theorem 1.4. By Theorem 1.5, it is sufficient to prove that $u = 0$ in $\mathbb{R}^N \times (0,\delta)$ for some small $\delta > 0$. Let σ be the Radon measure introduced in Lemma 2.3. Let us prove that $d\text{-}\overline{D}_{\rm sym}\sigma(x) < \infty$ for every $x \in \mathbb{R}^N$. Assuming by contradiction that for some $x \in \mathbb{R}^N$ one has $d\text{-}\overline{D}_{\rm sym}\sigma(x) = \infty$, there exists a sequence of radii $\rho_j \to 0^+$ such that $\sigma(B_d(x,\rho_j))/|B_d(x,\rho_j)| \longrightarrow \infty$, as $j \to \infty$.

From the representation formula (11) and the estimates of Γ in (10), it follows that

$$u(x,\rho_j^2) = \int_{\mathbb{R}^N} \Gamma(x,\rho_j^2;\xi,0)\,\mathrm{d}\sigma(\xi) \geq \mathbf{c}\,\rho_j^{-Q} \int_{B_d(x,\rho_j)} \exp\Big(-\mathbf{c}_0 \frac{d^2(x,\xi)}{\rho_j^2}\Big)\,\mathrm{d}\sigma(\xi) \geq$$
$$\geq \mathbf{c}'\,\sigma(B_d(x,\rho_j))/|B_d(x,\rho_j)| \longrightarrow \infty, \quad \text{as } j \to \infty.$$

This contradicts the hypothesis $\limsup_{t\to 0^+} u(x,t) < \infty$. Hence we can apply Lemma 2.7 and obtain that σ is absolutely continuous w.r.t. the Lebesgue measure. From the Lebesgue decomposition (15) $\mathrm{d}\sigma(\xi) = \varphi(\xi)\,\mathrm{d}\xi + \mathrm{d}s(\xi)$, it immediately follows that $s = 0$. Moreover, Theorem 2.4 gives $\varphi(x) = \lim_{t\to 0^+} u(x,t) = 0$ for almost every $x \in \mathbb{R}^N$, by hypothesis. Therefore, we obtain $\sigma = 0$. In order to complete the proof, it is now sufficient to recall the representation formula (11). □

Acknowledgement

The authors would like to thank Prof. E. Lanconelli for suggesting the study of these topics.

REFERENCES

[1] G. K. Alexopoulos: *Sub-Laplacians with drift on Lie groups of polynomial volume growth*, Mem. Amer. Math. Soc., **155** no. 739 (2002).

[2] A. Bonfiglioli – E. Lanconelli – F. Uguzzoni: *Uniform Gaussian estimates of the fundamental solutions for heat operators on Carnot groups*, Adv. Differential Equations, **7** (2002), 1153-1192.

[3] L. Capogna – N. Garofalo: *Boundary behavior of nonnegative solutions of subelliptic equations in NTA domains for Carnot-Carathéodory metrics*, J. Fourier Anal. Appl., **4** (1998), 403-432.

[4] G. Citti: *C^∞ regularity of solutions of a quasilinear equation related to the Levi operator*, Ann. Scuola Norm. Sup. Pisa Cl. Sci., **23** (4) (1996), 483-529.

[5] G. Citti – M. Manfredini – A. Sarti: *From neural oscillations to variational problems in the visual cortex*, Journal of Physiology, **97** (2003), 87-385.

[6] G. Citti – A. Sarti: *A cortical based model of perceptual completion in the Roto-Translation space*, preprint.

[7] G. B. Folland: *Subelliptic estimates and function spaces on nilpotent Lie groups*, Ark. Mat., **13** (1975), 161-207.

[8] G. B. Folland – E. M. Stein: *Hardy spaces on homogeneous groups*, Mathematical Notes, **28**, Princeton University Press, Princeton, 1982.

[9] P. Hajłasz – P. Koskela: *Sobolev met Poincaré*, Mem. Amer. Math. Soc., **145** no. 688, 2000.

[10] M. Kato: *On positive solutions of the heat equation*, Nagoya Math. J., **30** (1967), 203-207.

[11] A. Montanari: *Real hypersurfaces evolving by Levi curvature: smooth regularity of solutions to the parabolic Levi equation*, Comm. Partial Differential Equations, **26** (2001), 1633-1664.

[12] R. Montgomery: *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, **91**, American Mathematical Society, Providence, 2002.

[13] J. Petitot – Y. Tondut: *Vers une neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux*, Math. Inform. Sci. Humaines, **145** (1999), 5-101.

[14] S. Polidoro: *Uniqueness and representation theorems of Kolmogorov-Fokker-Planck equations*, Rendiconti di Matematica, Serie VII, **15** (1995), 535-560.

[15] L. P. Rothschild – E. M. Stein: *Hypoelliptic differential operators and nilpotent groups*, Acta Math., **137** (1976), 247-320.

[16] W. Rudin: *Real and Complex Analysis*, McGraw-Hill Book Co., New York-Toronto-London, 1966.

[17] E. M. Stein: *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series, **43**, Princeton University Press, Princeton, 1993.

[18] N. T. Varopoulos – L. Saloff-Coste – T. Coulhon: *Analysis and geometry on groups*, Cambridge Tracts in Mathematics, **100**, Cambridge University Press, Cambridge, 1992.

Lavoro pervenuto alla redazione il 23 febbraio 2004
ed accettato per la pubblicazione il 7 luglio 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo degli Autori:

Andrea Bonfiglioli, Francesco Uguzzoni – Dipartimento di Matematica – Università degli Studi di Bologna – Piazza di Porta S. Donato 5 – 40126 Bologna, Italy
E-mail: bonfigli@dm.unibo.it uguzzoni@dm.unibo.it

Investigation supported by University of Bologna. Funds for selected research topics.

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 69-84

An algorithm for estimating the optimal regularization parameter by the L-curve

G. RODRIGUEZ – D. THEIS

ABSTRACT: *In this paper we introduce a new algorithm to estimate the optimal regularization parameter in truncated singular value decomposition (TSVD) regularization methods for the numerical solution of severely ill-posed linear systems. The algorithm couples a geometrical approach to identify the corner of the L-curve associated to the problem with some heuristic rules. Numerical results are reported to highlight the performance of the algorithm with respect to other methods for the selection of the regularization parameter.*

1 – Introduction

A linear system of equations

$$A\mathbf{x} = \mathbf{b}$$

is considered *severely* ill-conditioned when the condition number

$$\kappa(A) := \|A\| \, \|A^{-1}\|,$$

in a given matrix norm, is of the same order of magnitude, or larger, than the reciprocal of the relative precision on the entries of the matrix A and of the right hand side vector $\mathbf{b}$.

KEY WORDS AND PHRASES: *ill-conditioned linear systems – Regularization – Truncated singular value decomposition (TSVD) – L-curve.*
A.M.S. CLASSIFICATION: 65F05 – 65F20 – 65F22

In this situation, no general purpose method is able to produce acceptable results, since ill-conditioning causes a huge amplification of errors in the solution of the linear system. Often, the only possibility for partially recovering the solution is the exploitation of *a priori* informations. Regularization methods consist of techniques that take advantage of informations on the regularity of the solution. Usually, these informations are formally expressed as the request for the solution $\mathbf{x}$ to belong to the kernel of a certain linear regularization operator H (in most cases, the discrete approximation of a differential operator). This approach is particularly effective when the solution of the linear system may be thought of as the sampling of a function which exhibits some degree of regularity. An example of great applicative interest is given by the linear systems arising in the discretization of first kind Fredholm integral equations with discrete data

$$\int_\Omega k(u_i, v)\, f(v)\, dv = g(u_i), \quad i = 1, \dots, m.$$

Each regularization method depends on at least one parameter, whose tuning is crucial for the quality of the numerical solution, since it balances the request of approximately satisfying the linear system with the regularity constraint.

The three most widely used regularization techniques are the *Truncated (Generalized) Singular Value Decomposition* (TSVD/TGSVD) [8], *Tikhonov regularization* [20], [21], [6] and regularizing iterative methods [4], [7]. A complete survey of the various regularization strategies and the available methods for the estimation of the optimal regularization parameter can be found in [12].

In this paper we will concentrate on the first mentioned regularization method, and on a particular strategy for choosing its parameter, the *L-curve* method. In Section 2 the TSVD and TGSVD are recalled, while in Section 3 the L-curve method is described. In Section 4 we introduce a new algorithm for the localization of the corner of the L-curve, which has already been applied, in a preliminary version, in some previous researches [2], [19]. Finally, in Section 5 the performance of the new algorithm is assessed on a set of test linear systems, and in Section 6 plans for future work are discussed.

2 – The truncated (G)SVD

Let us consider, as a model problem, the overdetermined linear system

$$A\mathbf{x} = \mathbf{b} \tag{2.1}$$

where $A \in \mathbb{R}^{m\times n}$, $m \geq n$, is a full-rank matrix.

The *singular value decomposition* (SVD) of A [1], [5] is given by

$$U^T A V = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix}, \quad \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n), \tag{2.2}$$

where $\sigma_1 \geq \cdots \geq \sigma_n > 0$ are the singular values and the orthogonal matrices

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \quad \text{and} \quad V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$$

contain the left and right singular vectors, respectively. Then, the least squares solution of (2.1) can be expressed in the form

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i. \tag{2.3}$$

Severe ill-conditioning can be restated by saying that A is numerically rank-deficient, i.e. there exists an integer $k \leq n$ such that, for a given tolerance ϵ,

$$\sigma_{k+1}, \ldots, \sigma_n < \epsilon.$$

This integer is, in fact, the numerical ϵ-rank of A, usually defined as

$$\operatorname{rank}_\epsilon(A) := \min_{\|E\|_2 \leq \epsilon} \operatorname{rank}(A + E).$$

When a singular value σ_i is approximately zero, the corresponding singular vector $\mathbf{v}_i$ belongs to the numerical kernel of A and we expect its coefficient in (2.3) to be negligible. If the system (2.1) is compatible this is certainly true, but the presence of noise on $\mathbf{b}$ may cause a huge growth in the norm of the solution $\mathbf{x}$.

To obtain a better estimate of the least squares solution the *truncated SVD* (TSVD) solution is often used. It is given by

$$\mathbf{x}_k = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i \tag{2.4}$$

and it coincides with the minimum 2-norm solution to the least squares problem

$$\min \|A_k \mathbf{x} - \mathbf{b}\|_2, \tag{2.5}$$

where A_k is the best rank k approximation to A in the 2-norm, obtainable by substituting $\sigma_j = 0$, $j = k+1, \ldots, n$, in (2.2).

It is then crucial to correctly tune the value of the regularization parameter k in order to avoid the numerical explosion of the norm of the solution while preserving, at the same time, all of its significant components.

When there is the *a priori* information that the solution (approximately) belongs to the kernel of a certain regularization matrix H, that is the number $\|H\mathbf{x}\|_2$ is small, it is more effective to compute the solution of (2.5) which minimizes the semi-norm $\|H\mathbf{x}\|_2$, instead than the norm $\|\mathbf{x}\|_2$.

The solution to this *non-standard* regularization problem can be obtained by the transformation $\mathbf{y} = H\mathbf{x}$, but while the case of H square nonsingular can be easily managed, if the regularization matrix is a non square $p \times n$ matrix ($p < n$) with rank p, the computation is a bit more cumbersome. A method for taking this problem to standard form has been described in [9] and starts by expressing the least-squares solution in the form

$$\mathbf{x} = H_A^{\dagger}\mathbf{y} + \mathbf{x}_0, \tag{2.6}$$

where $\mathbf{x}_0$ is in the null space of H and the matrix $H_A^{\dagger}$ is the *A-weighted pseudo inverse* of H, defined in the following.

Let the *generalized singular value decomposition* (GSVD) of the matrix pair (A, H) [1], [5] be the factorization

$$U^T A Z = \begin{bmatrix} D_A & 0 \\ 0 & I_{n-p} \\ 0 & 0 \end{bmatrix}$$

$$V^T H Z = [\, D_H \quad 0 \,]$$

with

$$D_A = \mathrm{diag}(d_1, \ldots, d_p), \quad 0 < d_1 \leq \cdots \leq d_p \leq 1,$$

$$D_H = \mathrm{diag}(h_1, \ldots, h_p), \quad 1 \geq h_1 \geq \cdots \geq h_p > 0,$$

and $d_i^2 + h_i^2 = 1$, $i = 1, \ldots, p$. The matrices U and V are orthogonal, $Z = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is nonsingular and the ratios $\gamma_i = d_i/h_i$, $i = 1, \ldots, p$, are called the *generalized singular values* of (A, H). Then, we define

$$H_A^{\dagger} := Z \begin{bmatrix} D_H^{-1} \\ 0 \end{bmatrix} V^T$$

and it is immediate to observe that it is a right inverse of H.

By substituting $\mathbf{y} = H\mathbf{x}$, with $\mathbf{x}$ given by (2.6), we come to the following standard form regularization problem: find the minimum norm solution to

$$\min \|\bar{A}_k \mathbf{y} - \bar{\mathbf{b}}\|_2,$$

where $\bar{A}_k$ is the best rank k approximation to $\bar{A} = AH_A^{\dagger}$ and $\bar{\mathbf{b}} = \mathbf{b} - A\mathbf{x}_0$. The solution to this problem, known as the *truncated GSVD (TGSVD) solution*, is given by

$$\mathbf{x}_k = \sum_{i=p-k+1}^{p} \frac{\mathbf{u}_i^T \mathbf{b}}{d_i} \mathbf{z}_i + \sum_{i=p+1}^{n} (\mathbf{u}_i^T \mathbf{b}) \mathbf{z}_i. \tag{2.7}$$

Again, to get a meaningful solution it is essential to correctly estimate the value of the regularization parameter k.

Several criteria are available for this task, some requiring the knowledge of the standard deviation of the noise affecting the data and some not requiring it. We will discuss the *L-curve* method in the next section, here we briefly recall some of the other most widely used techniques.

The *Generalized Cross Validation* (GCV) [3], [22] is a statistical method which estimates the optimal value of the regularization parameter, under the assumption that the data vector $\mathbf{b}$ is affected by normally distributed noise, by minimizing the functional

$$V(k) = \frac{\frac{1}{m}\|(I - A(k))\mathbf{b}\|^2}{\left[\frac{1}{m}\operatorname{trace}(I - A(k))\right]^2}. \tag{2.8}$$

The *influence matrix* $A(k)$ is defined by the identity

$$A\mathbf{x}_k = A(k)\mathbf{b}.$$

The GCV has some computationally relevant properties and, moreover, is a *predictive mean-square error criteria* [22], in the sense that it estimates the minimizer of the residual function

$$T(k) = \frac{1}{m}\|A(\mathbf{x}_k - \mathbf{x})\|^2.$$

If the standard deviation σ of the noise on the data is known, the following unbiased estimate [15] for the function $T(k)$ is also available

$$\hat{T}(k) = \frac{1}{m}\|(I - A(k))\mathbf{b}\|^2 - \frac{\sigma^2}{m}\operatorname{trace}(I - A(k))^2 + \frac{\sigma^2}{m}\operatorname{trace} A^2(k).$$

Mallows' criterion chooses the value of k which minimizes $\hat{T}(k)$. Another technique which makes use of the value of σ is *Morozov discrepancy principle* [17], which takes as optimal the value of k that satisfies the equation

$$\frac{1}{m}\|(I - A(k))\mathbf{b}\|^2 = \sigma^2.$$

However, numerical experiments reported in the literature (see e.g. [3], [18]) showed that these two procedures do not give better results than GCV, even when σ is exactly known.

3 – The L-curve method

The *L-curve* method [10], [14] consists of the analysis of the piecewise linear curve whose break-points are

$$(x_i, y_i) = (\log_{10} \|A\mathbf{x}_i - \mathbf{b}\|_2, \log_{10} \|H\mathbf{x}_i\|_2), \quad i = 1, \ldots, p$$

(p is the row dimension of the regularization matrix H).

This curve, in most cases, exhibits a typical "L" shape, and the optimal value of the regularization parameter k is considered to be the one corresponding to the *corner* of the "L" (see Figure 1).

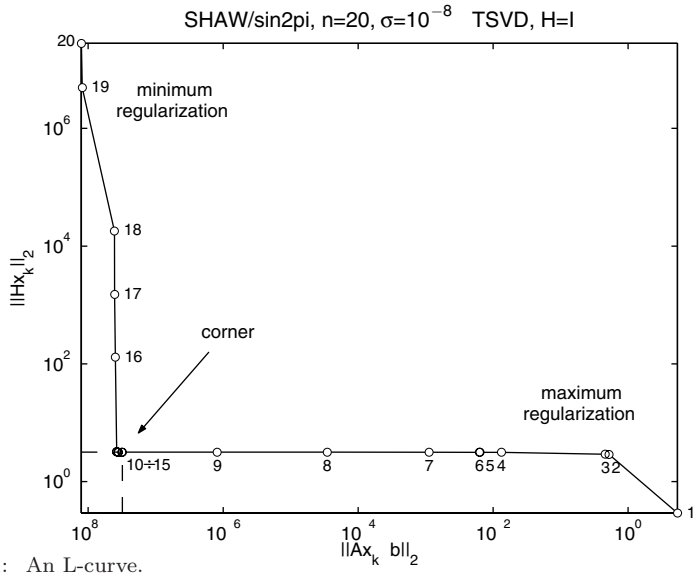

Fig. 1: An L-curve.

This choice is justified by the fact that while the regularized solution $\mathbf{x}_k$ of (2.1) coincides with the least-squares solution $\mathbf{x}$ when $k = p$, the ill-conditioning of A causes a strong growth in the weighted semi-norm $\|H\mathbf{x}_k\|_2$ when k exceeds a certain threshold (which is, in fact, the numerical ϵ-rank of A for a well-chosen ϵ). The corner of the L-curve marks this transition, since it represents a compromise between the minimization of the norm of the residual and the semi-norm of the solution. This is particular evident in Figure 1: the horizontal branch of the "L" is dominated by the regularization error, while the vertical branch shows the sharp increase in the semi-norm caused by propagation errors.

We spend some words to explain the symbols used in this and in the following figures. The heading of the graph displays informations on the test problem and on the method used for its solution. In this case we created a test linear system with the SHAW matrix, taken from [11], and with sample solution *sin2pi*, given

by $x_i = \sin \frac{2\pi i}{n}$, $i = 1, \ldots, n$ (other test matrices and solutions will be introduced in Section 5). The dimension of the system is n and the data vector $\mathbf{b}$ is affected by white noise with variance $n\sigma^2$. This system, as $H = I$, has been solved by TSVD. Each point on the graph stands for the particular regularized solution $\mathbf{x}_k$ whose index labels the point.

A numerical algorithm for the detection of the corner of the L-curve has been introduced in [14]. When the regularization method depends on a continuous parameter λ, like in Tikhonov regularization, then the L-curve is a smooth function, possibly twice differentiable, and this method selects the value which maximizes the curvature $\kappa(\lambda)$ of the L-curve. If, on the contrary, the regularization parameter is discrete, like in T(G)SVD or in iterative regularization methods, the algorithm selects the parameter closest to the point of maximum curvature of a cubic spline curve with knots resulting from a local smoothing of the L-curve points.

This method has some drawbacks, especially when applied to a discrete L-curve.

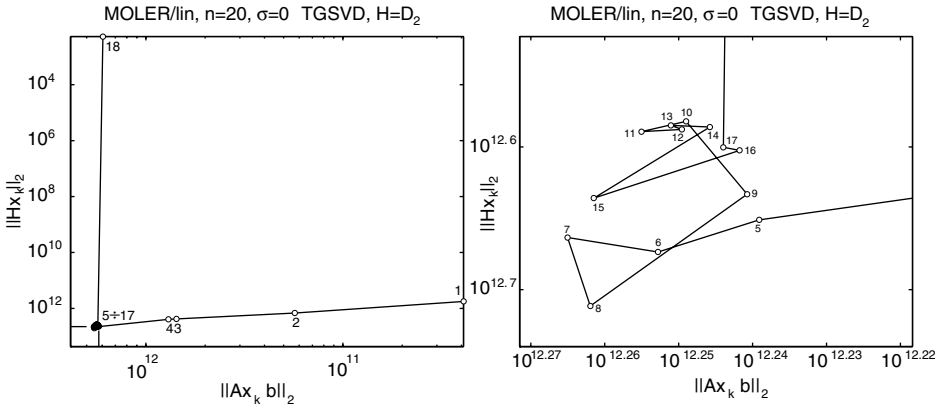

Fig. 2: A cluster of points.

In fact, in T(G)SVD regularization methods the points of the L-curve tend to cluster in a neighborhood of the corner. In this situation, errors due to floating point computations may produce false corners and loss of convexity, as illustrated in Figure 2 which shows an experimental L-curve together with a close-up of a neighborhood of its corner. The effect is that the spline which fits the L-curve often presents unexpected oscillations near the corner, which lead to an inaccurate estimate. This is particularly dangerous when the algorithm returns an over-estimation of the optimal value of k, which often causes a large increase in the norm of the computed solution.

Moreover, in many practical situations, some of which are depicted in Figure 3, the L-curve totally looses its "L" shape, making it difficult to choose a good value of the parameter without resorting to some heuristic rule.

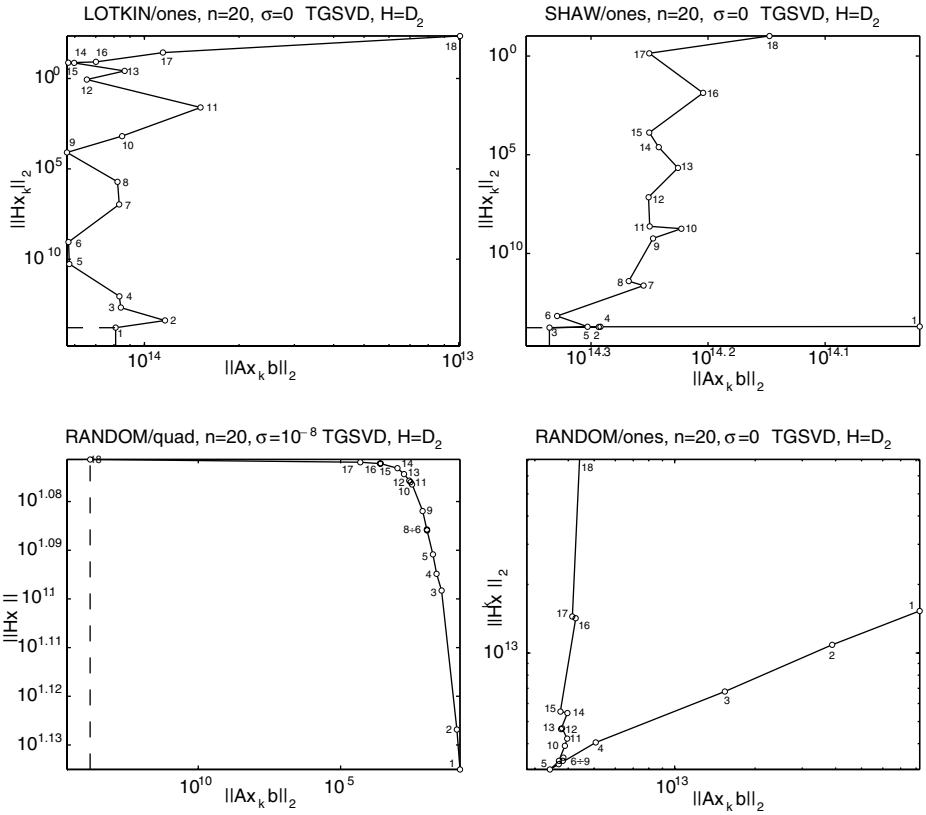

Fig. 3: Some bad L-curves.

We remark, anyway, that these L-curves are still rather informative about the problems we are trying to solve, and their interpretation will allow us, in the next section, to implement an effective strategy which exploits the informations they contain.

The two upper graphs in Figure 3, for example, exhibit a huge increase in the semi-norm of the solution and small changes in the residuals, mostly due to floating-point arithmetics. This situation is typical of linear systems whose solution exactly belongs to the kernel of the regularization matrix H and for which just a few generalized singular values are sufficient to compute a good approximation of the solution. In real applications, to get good results we are interested in using a regularization matrix H whose kernel contains the biggest possible component of the solution $\mathbf{x}$, so it is important that the parameter estimation routine could treat effectively this situation.

The lower left graph in Figure 3, instead, displays a monotonically decreasing residual associated to a negligible growth in the semi-norm, typical of a

well-conditioned (of mildly ill-conditioned) linear system, for which it is possible to use all the singular values in the computation of the solution. The last graph shows both the effects in the same test problem. Obviously, one would not apply regularization to a well conditioned matrix, but since in some applications the matrix A is severely ill-conditioned only for a certain range of dimensions, or in correspondence of particular values of some constants, it would be desirable if the regularization method would automatically detect a well-conditioned matrix, tuning correspondingly the regularization parameter.

4 – The *corner* algorithm

The algorithm we propose couples a simple geometrical approach to locate the corner of a discrete L-curve, preceded by a suitable numerical pre-processing of cluster of points, to some empirical rules aimed to recognize two classes of L-curves "without a unique corner" for which it is possible to predict a good value of the regularization parameter. These rules have been devised through analyzing, and interpreting, a large number of experimental L-curves.

The algorithm takes in input the residuals and the semi-norms associated to each regularized solution, namely

$$\|\mathbf{b} - A\mathbf{x}_i\|, \ \|H\mathbf{x}_i\|, \quad i = 1, \dots, p, \tag{4.1}$$

and can be decomposed into three phases.

1. Initially, we try to understand if the solution is approximately in the kernel of the regularization operator H. It is important to check this condition first, because if affirmative the L-curve would not be "L" shaped, and the search for a corner would be useless (see upper graphs in Figure 3). The task is performed through detecting extremely small semi-norms, with a test of the type
$$\frac{\min \|H\mathbf{x}_i\|}{\max \|H\mathbf{x}_i\|} < \tau_1.$$
The subroutine applies this test in conjunction with
$$\min \frac{\|H\mathbf{x}_i\|}{\|\mathbf{x}_i\|} < \tau_2$$
whenever the 2-norms of the regularized solutions are made available.
The tolerances τ_1 and τ_2 are two of the four constants the algorithm depends on. They have been fixed to 10^{-12} and 10^{-4} respectively, working in double precision, and the results do not seem much sensitive on changes in these parameters.

If the detection of small semi-norms is successful, the algorithm stops returning the index of the smaller one as an estimate of the optimal regularization parameter.

2. If the previous test is not verified, we compute the points of the L-curve

$$P_i = (\log_{10} \|\mathbf{b} - A\mathbf{x}_i\|, \log_{10} \|H\mathbf{x}_i\|), \quad i = 1, \dots, p,$$

and the vectors

$$\mathbf{v}_i = P_{i+1} - P_i, \quad i = 1, \dots, p-1.$$

To eliminate clusters (Figure 2) we delete all the "short" vectors, i.e. those verifying the condition

$$\|\mathbf{v}_i\| < \tau_3,$$

leaving q acceptable vectors ($q \leq p$). The constant τ_3, which we fix at $\|P_p - P_1\|/(2p)$, is rather important. Taking smaller values can give very good results, but may also lead to dangerous over-estimates. We are currently analyzing the possibility of choosing the value of τ_3 adaptively.
After normalizing the q remaining vectors (we are only interested in their orientation) the situation is similar to the one depicted in Figure 4.

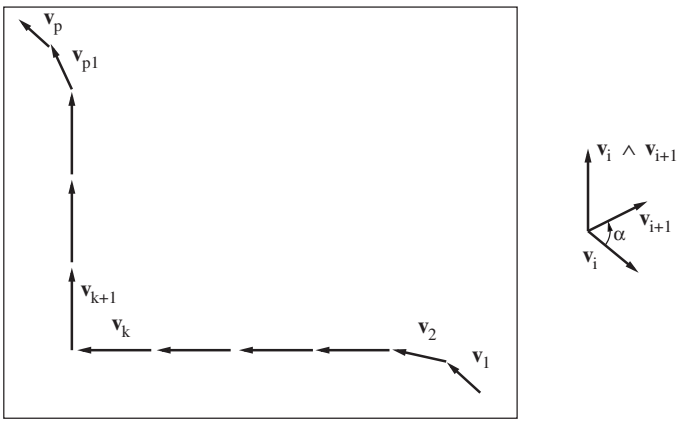

Fig. 4: L-curve and wedge products.

If we travel along the L-curve visiting the vectors $\mathbf{v}_i$ in ascending order, the corner is characterized by an angle $\alpha \simeq -\frac{\pi}{2}$ between $\mathbf{v}_k$ and $\mathbf{v}_{k+1}$. Then, the search for the corner can be carried out by finding the minimum z-coordinate of the wedge products between two succeeding vectors (See Figure 4)

$$w_i = (\mathbf{v}_i \wedge \mathbf{v}_{i+1})_z = \|\mathbf{v}_i\| \cdot \|\mathbf{v}_{i+1}\| \cdot \sin\alpha, \quad i = 1, \dots, q.$$

The following elementary property of wedge products

$$(\mathbf{v}_i \wedge \mathbf{v}_{i+1})_z = \det([\,\mathbf{v}_i \quad \mathbf{v}_{i+1}\,])$$

allows to compute the numbers w_i easily and with a small computational effort.

3. The minimum of the wedge products is accepted as a corner only if it verifies the condition

$$\min_{i=1,\dots,p} w_i < \tau_4, \quad (\tau_4 = -0.5). \tag{4.2}$$

The value of τ_4, like the first two constants of the algorithm, does not seem to be very critical for the performance of the method.
If condition (4.2) is not verified, the L-curve is considered to be *without a corner* and we check for the presence of a well conditioned (or mildly ill-conditioned) matrix by detecting a small change in the extremal semi-norms

$$|\log_{10}\|H\mathbf{x}_p\| - \log_{10}\|H\mathbf{x}_1\|\,| < 10.$$

In this case we return $k = p$ as the optimal parameter.

The failure of the last test is an error condition, which we still keep in the subroutine with the hope to trap unforeseen situations, i.e. L-curves not belonging to the three classes we have considered, and to further improve the algorithm.

The outline of the algorithm is reported in Figure 5. A Matlab [16] function is available upon request (send an email to `rodriguez@unica.it`).
We end up with a computational remark. It is known that the residuals and the semi-norms (4.1) can be expressed in terms of the singular system of the matrix A. In fact, from (2.4) and (2.7) it follows

$$\|\mathbf{x}_i\|^2 = \sum_{\ell=1}^{i} \left(\frac{\mathbf{u}_\ell^T\mathbf{b}}{\sigma_\ell}\right)^2 \quad \text{and} \quad \|H\mathbf{x}_i\|^2 = \sum_{\ell=p-i+1}^{p} \left(\frac{\mathbf{u}_\ell^T\mathbf{b}}{\gamma_\ell}\right)^2.$$

It is also possible, taking into account that

$$\mathbf{b} - A\mathbf{x}_k = A(\mathbf{x} - \mathbf{x}_k) = A(\mathbf{x}_p - \mathbf{x}_k),$$

to obtain a similar formula for the residuals. Anyway, even if this formula is computationally less expensive, we noticed that the residuals computed in this way are sometimes *too well computed* and give worse results, for what concerns the estimation of the regularization parameter, with respect to the residuals computed by implementing their definition. We feel that the reason for this is that the residuals should be affected by propagation errors in the same amount as the solution is, since we want to extract from them informations about the quality of results.

1. input $\|\mathbf{b} - A\mathbf{x}_i\|$, $\|H\mathbf{x}_i\|$, $\|\mathbf{x}_i\|$, $i = 1, \ldots, p$
2. $\tau_1 = 10^{-12}$, $\tau_2 = 10^{-4}$
3. if $\dfrac{\min\limits_{i=1,\ldots,p} \|H\mathbf{x}_i\|}{\max\limits_{i=1,\ldots,p} \|H\mathbf{x}_i\|} < \tau_1$ and $\min\limits_{i=1,\ldots,p} \dfrac{\|H\mathbf{x}_i\|}{\|\mathbf{x}_i\|} < \tau_2$
   1. $k = \arg \min\limits_{i=1,\ldots,p} \|H\mathbf{x}_i\|$
4. else
   1. for $i = 1, \ldots, p$
      1. $P_i = (\log_{10} \|\mathbf{b} - A\mathbf{x}_i\|, \log_{10} \|H\mathbf{x}_i\|)$
   2. $\tau_3 = \dfrac{\|P_p - P_1\|}{2p}$, $\tau_4 = -0.5$
   3. $q = 0$
   4. for $i = 1, \ldots, p$
      1. $\mathbf{v} = P_{i+1} - P_i$
      2. if $\|\mathbf{v}\| > \tau_3$
         1. $q = q + 1$
         2. $\mathbf{v}_q = \dfrac{\mathbf{v}}{\|\mathbf{v}\|}$
   5. for $i = 1, \ldots, q - 1$
      1. $w_i = \det\left(\begin{bmatrix}\mathbf{v}_i & \mathbf{v}_{i+1}\end{bmatrix}\right)$
   6. if $\min\limits_{i=1,\ldots,q} w_i < \tau_4$
      1. $k = \arg \min\limits_{i=1,\ldots,q} w_i$
   7. else
      1. if $|\log_{10} \|H\mathbf{x}_p\| - \log_{10} \|H\mathbf{x}_1\|| < 10$
         1. $k = p$
      2. else
         1. error 'corner not found'
5. output k

Fig. 5: The *corner* algorithm.

5 – Numerical experimentation

To investigate the performance of our algorithm we applied it to the estimatation of the optimal regularization parameter in a set of test problem, which we solved by TSVD or TGSVD when $H = I$ or $H \neq I$, respectively.

We considered eleven square test matrices, taken from the package *Regularization Tools* [11], [13] (HEAT(1), SHAW, SPIKES, BAART, ILAPLACE) and from Matlab [16] (HILBERT, PASCAL, LOTKIN, MOLER, PROLATE, RANDOM). For $n = 20$, most of these matrices are severely ill-conditioned (in the sense that their condition number exceeds the reciprocal of the *machine epsilon* $\epsilon_M \simeq 2.2 \cdot 10^{-16}$),

two of them are mildly ill-conditioned (MOLER and PROLATE) and one (the RANDOM matrix) is well conditioned. For $n = 80$ they are all severely ill-conditioned, except the RANDOM matrix.

For each test matrix, we constructed different linear systems by computing the right hand side $\mathbf{b}$ corresponding to the sample solutions listed in Table 1 (the *rtools* solution, which is the sample solution adopted in [11], is used only with the matrices coming from the *Regularization Tools*).

Table 1. Sample solutions of linear systems.

rtools	defined as in [11]
ones	$x_i = 1$
lin	$x_i = \dfrac{i}{n}$
quad	$x_i = \left(i - \left\lfloor \dfrac{n}{2} \right\rfloor\right)^2 / \left\lceil \dfrac{n}{2} \right\rceil^2$
sin2pi	$x_i = \sin \dfrac{2\pi(i-1)}{n}$
sinpi	$x_i = \sin \dfrac{\pi(i-1)}{n}$
lin+sinpi	$x_i = \dfrac{i}{n} + \sin \dfrac{\pi(i-1)}{n}$

The linear systems so obtained were solved both in the presence and in the absence of noise on the data. In practice, the data vector $\mathbf{b}$ was substituted by the vector $\tilde{\mathbf{b}}$, with components

$$\tilde{b}_i = b_i + \sigma\epsilon_i, \quad i = 1, \ldots, n,$$

with ϵ_i normally distributed with mean value 0 and variance 1. We considered $\sigma = 0, 10^{-8}, 10^{-4}$.

By this procedure we generated 213 test problems, which we solved for $n = 20$ and $n = 80$ either by TSVD ($H = I$) and by TGSVD for each of the regularization matrices $H = D_1, D_2, D_3$, being D_k the Toeplitz matrix of dimension $(n-k) \times n$ whose first row is the discrete approximation of the k-th derivative.

For each test problem, the optimal regularization parameter was compared with the estimates furnished by our algorithm (labelled as *corner* in Tables 2 and 3), by the routine *l_curve* from [13], which is an implementation of the *maximum curvature* algorithm described in [14], and by the routine *GCV*, coming from the same package, which returns the minimizer of the functional (2.8).

Table 2. Numerical tests, $n = 20$.

H	corner	l_curve	GCV
I	102(8/2)	38(76/37)	94(42/35)
D_1	89(21/4)	64(46/31)	76(74/54)
D_2	92(30/20)	50(50/39)	39(136/114)
D_3	101(44/35)	41(58/40)	28(119/100)

Table 2 lists the results obtained for $n = 20$ by applying the three mentioned methods with each of the four adopted regularization matrices. The first number in every entry of the table equals the number of tests in which the optimal parameter is exactly identified (the "full successes"), while the two numbers in parentheses indicate in how many tests an incorrect estimate of the regularization parameter produces an error in the solution which exceeds the optimal one by a factor 10^2 and 10^4, respectively (the "failures"). We remark that the algorithms should be considered successful at least when the error on the computed solution is smaller than 10^2 times the optimal error. Table 3 shows the same results for $n = 80$.

Table 3. Numerical tests, $n = 80$.

H	corner	l_curve	GCV
I	90(39/33)	50(66/40)	72(62/51)
D_1	71(27/24)	30(75/65)	56(84/39)
D_2	73(25/17)	20(125/98)	56(74/62)
D_3	84(34/17)	13(128/111)	44(108/77)

The numerical results show that the estimates furnished by our algorithm are significantly more trustworthy than the ones produced by the two other algorithms considered. Moreover, the number of the cases of total failure is rather small. We feel that the reason for this is also that our algorithm tends to under-estimate, rather than over-estimate, the optimal regularization parameter. These results confirm, in particular, the great efficacy of the L-curve as a tool for the analysis and the numerical resolution of ill-conditioned linear systems.

6 – Future work

In this section we expose the lines of research which we consider important in order to further improve the performance of our algorithm and to extend its range of applicability.

First of all, we are developing an adaptive procedure to choose the values of the constants on which the algorithm depends, in particular the one (τ_3) whose value seems to be the most sensible for the performance of the method.

We also plan to carry on a wider numerical experimentation, with the hope to identify some particular test problems leading to L-curves which our method actually does not recognize, i.e. cases which fall into the final error condition of the algorithm.

Finally we wish to extend the method in order to apply it to iterative regularization methods and to Tikhonov regularization. The difficulty, in the first case, is that the discrete regularization parameter, namely the iteration index, does not have an upper bound, unlike in SVD methods, and we think that a part of the algorithm should be repeated at each iteration to be able to track the overcoming of the corner of the L-curve.

In the second case, where the parameter is a positive real number, we plan to start with a coarse discretization of the L-curve and to add adaptively more points in a neighborhood of the corner until its position is identified up to a prescribed accuracy.

REFERENCES

[1] Å. Björck: *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[2] C. Brezinski – M. Redivo-Zaglia – G. Rodriguez – S. Seatzu: *Multi-parameter regularization techniques for ill-conditioned linear systems*, Numer. Math., **94** (2003), 203-228.

[3] P. Craven – G. Wahba: *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of Generalized Cross Validation*, Numer. Math., **31** (1979), 377-403.

[4] H.W. Engl – M. Hanke – A. Neubauer: *Regularization of Inverse Problems*, Kluwer, Dordrecht, the Netherlands, 1996.

[5] G.H. Golub – C.F. Van Loan: *Matrix Computations*, The John Hopkins University Press, Baltimore, third edition, 1996.

[6] Ch.W. Groetsch: *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Research Notes in Mathematics, Vol. 10, Pitman, London, 1984.

[7] M. Hanke: *Conjugate Gradient Type Methods for Ill-Posed Problems*, Pitman Research Notes in Mathematics, Vol. 327, Longman, Harlow, U.K., 1995.

[8] P.C. Hansen: *The truncated SVD as a method for regularization*, BIT, **27** (1987), 543-553.

[9] P.C. Hansen: *Regularization, GSVD and truncated GSVD*, BIT, **29** (1989), 491-504.

[10] P.C. Hansen: *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Review, **34** (1992), 561-580.

[11] P.C. Hansen: *Regularization Tools: A Matlab package for analysis and solution of discrete ill–posed problems*, Numerical Algorithms, **6** (1994), 1-35.

[12] P.C. Hansen: *Rank-deficient and discrete ill-posed problems*, SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[13] P.C. Hansen: *Regularization Tools version 3.0 for Matlab 5.2*, Numerical Algorithms, **20** (1999), 195-196.

[14] P.C. Hansen – D.P. O'Leary: *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., **14** (1993), 1487-1503.

[15] C. Mallows: *Some comments on C_p*, Technometrics, **15** (1973), 661-675.

[16] The MathWorks Inc.: South Natick, MA, MATLAB Version 6.5, 2002.

[17] V.A. Morozov: *The choice of parameter when solving functional equations by regularization*, Dokl. Akad. Nauk SSSR, **175** (1962), 1225-1228.

[18] G. Rodriguez – S. Seatzu: *Numerical solution of the finite moment problem in a reproducing kernel Hilbert space*, J. Comp. Appl. Math., **33** (3) (1990), 233-244.

[19] G. Rodriguez – S. Seatzu – D. Theis: *A new technique for ill-conditioned linear systems*, Numerical Algorithms, **33** (2003), 433-442.

[20] A.N. Tikhonov: *Regularization of incorrectly posed problems*, Soviet Math. Dokl., **4** (1963), 1624-1627.

[21] A.N. Tikhonov: *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl., **4** (1963), 1036-1038.

[22] G. Wahba: *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, PA, 1990.

Lavoro pervenuto alla redazione il 15 febbraio 2003
ed accettato per la pubblicazione il 6 aprile 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo degli Autori:

G. Rodriguez, D. Theis – Dipartimento di Matematica e Informatica – Università di Cagliari – viale Merello 92 – 09123 Cagliari, Italy
E-mail: rodriguez@unica.it theis@bugs.unica.it

This work was partially supported by MIUR, grant number 2002014121.

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 85-108

On the surface tension for non local energy functionals

CRISTIANA BISCEGLIA – EMANUELE ROSATELLI

ABSTRACT: *We consider the free energy functional $F_\varepsilon(m)$, $\varepsilon > 0$ a scaling parameter, $m \in L^\infty(\mathcal{T}; [-1, 1])$, $\mathcal{T}$ the unit torus, which has been derived in a continuum limit from Ising spin systems with Kac interactions, see* [8]. *In* [1] *it is proved that $F_\varepsilon(m)$ $\Gamma-$converges to a perimeter functional P. We study here the free energy functional with an additional term describing the interaction with an external magnetic field h. We suppose that h takes only the two values $\pm s$, $s > 0$. Calling E the region of the torus where the external field is negative and $F_{\varepsilon,s}(m; E)$ the new functional, we then define $G_{\varepsilon,s}(E) = \inf_m F_{\varepsilon,s}(m; E)$. We prove that $G_{\varepsilon,s}(\cdot)$ $\Gamma-$converges to a perimeter functional which as a function of s converges pointwise as $s \to 0$ to P.*

1 – Introduction

In this paper we consider the non local, excess, free energy functional defined for all m on $L^\infty(\mathbb{R}^d; [-1, 1])$, with values in $[0, +\infty]$, $+\infty$ included, by

$$\mathcal{F}_{\beta,h}(m) = \int_{\mathbb{R}^d} dr\, f_{\beta,h}\big(m(r)\big) + \frac{1}{4}\int_{\mathbb{R}^d} dr \int_{\mathbb{R}^d} dr' J(r, r')[m(r) - m(r')]^2 \tag{1.1}$$

where $h \in L^\infty(\mathbb{R}^d; \mathbb{R})$,

$$f_{\beta,h}(m) = \phi_{\beta,h}(m) - \min_{|m|\le 1} \phi_{\beta,h}(m) \tag{1.2}$$

$$\phi_{\beta,h}(m) = -\frac{m^2}{2} - hm - \frac{I(m)}{\beta}, \qquad m \in [-1, 1] \tag{1.3}$$

KEY WORDS AND PHRASES: *Coexistence of Phases – Surface Tension – Γ- convergence.*
A.M.S. CLASSIFICATION: 47H – 82C

$$I(m) = -\frac{1-m}{2}\log\frac{1-m}{2} - \frac{1+m}{2}\log\frac{1+m}{2}. \tag{1.4}$$

The interaction $J(r,r')$ is a translational invariant (i.e. $J(r,r') = J(0,r'-r)$), smooth, symmetric, probability kernel supported by $|r-r'| \leq 1$. As $\mathcal{F}_{\beta,h}$ depends symmetrically on $J(r,r')$ there is no loss of generality in assuming

$$J(r,r') = J(r',r), \quad \text{and, equivalently,} \quad J(0,r) = J(0,-r).$$

The expression (1.1) arises in the study of Gibbs measures in Ising spin systems with Kac interactions, see [4] and in their time evolution with Glauber dynamics , where it is derived in a continuum limit, [6]; m is then interpreted as a magnetization density and $\beta^{-1} = \kappa T$, T the absolute temperature and κ the Boltzmann constant, h is an external magnetic field.

Due to the positivity of J, the second term is minimized by any constant function, while the first one is minimal when the constant is set equal to a minimizer, call it $m_{\beta,h}$, of $f_{\beta,h}(s)$, $s \in [-1,1]$.

Thus $\mathcal{F}_{\beta,h}(m^*) = 0$ when $m^*(r) = m_{\beta,h}$ for all $r \in \mathbb{R}^d$: $m^*(r)$ is therefore called an equilibrium phase and $\mathcal{F}_{\beta,h}(m)$ measures the increase of free energy in magnetization profiles m which deviate from equilibrium.

Phase transitions are related to the lack of uniqueness of the minimizers of the free energy functional, which, for $\mathcal{F}_{\beta,h}$, occurs at $h = 0$ and $\beta > 1$. In such cases in fact the equilibrium magnetization $m_{\beta,0}$ can take two values, $\pm m_\beta$, solutions of the mean field equation

$$m_\beta = \tanh\{\beta m_\beta\}. \tag{1.5}$$

We now turn to the main object of this paper, surface tension and more generally, coexistence of phases. Roughly speaking, the surface tension is the excess free energy per unit area needed to create a state with two coexisting phases. The area in the definition refers to the interface which separates the two phases and the surface tension may depend on its orientation when the interaction is anisotropic. Thus, in a macroscopic description, characterized by the assumption of local thermodynamic equilibrium, at all points the magnetization is either equal to m_β or to $-m_\beta$. Let us restrict, for simplicity, to a unit torus $\mathcal{T}$ of $\mathbb{R}^d$ (in macroscopic units). Then a macroscopic state is a magnetization profile $u(r) \in \{\pm m_\beta\}$ for any $r \in \mathcal{T}$. Call E the region in $\mathcal{T}$ where $u = m_\beta$ and E^c its complement, where $u = -m_\beta$, then, if the boundary ∂E of E is regular, the macroscopic free energy of u is

$$P(u) = \int_{\partial E} dH^{d-1}(r)\theta_\beta(\nu(r)) \tag{1.6}$$

where $dH^{d-1}(r)$ is the Hausdorff area measure and $\theta_\beta(n) = \theta_\beta(-n)$ is the surface tension of a planar surface with normal n, $\nu(r)$ the unit normal to ∂E at r.

Regularity of E is not really necessary and in fact the above expression keeps its validity for all u in $BV(\mathcal{T}, \{\pm m_\beta\})$, as it will be discussed later on.

To relate the macroscopic theory to the functional (1.1) we interpret the latter as the result of a more accurate, microscopic, description of the system, where distances are magnified revealing deviations from the local equilibrium condition $|u(r)| = m_\beta$. Thus, calling ε^{-1} the magnifying factor of the blow up, a microscopic state is an element $m \in L^\infty(\varepsilon^{-1}\mathcal{T}, [-1,1])$ and its microscopic free energy is $\mathcal{F}^{\rm per}_{\varepsilon^{-1}\mathcal{T}}(m)$, where the latter is the functional (1.1) restricted to $m \in L^\infty(\varepsilon^{-1}\mathcal{T}); [-1,1])$ with J replaced by its periodization on $\varepsilon^{-1}\mathcal{T}$.

To compare with (1.6) we first need to have objects on a same space. Let $V_\varepsilon : L^\infty(\mathcal{T}; [-1,1]) \to L^\infty(\varepsilon^{-1}\mathcal{T}; [-1,1])$ be defined by

$$V_\varepsilon m(r) = m(\varepsilon r) =: m^{(\varepsilon)}(r), \qquad r \in \varepsilon^{-1}\mathcal{T}. \tag{1.7}$$

Then $\mathcal{F}^{\rm per}_{\varepsilon^{-1}\mathcal{T}} \circ V_\varepsilon$ becomes a functional on $L^\infty(\mathcal{T}; [-1,1])$ which associates to any given $m \in L^\infty(\mathcal{T}; [-1,1])$ a microscopic free energy, indexed by ε. Since we are interested in states with interface, their free energy must scale as an area, namely proportionally to ε^{-d+1}. We then define

$$\Phi_\varepsilon = \varepsilon^{d-1}\mathcal{F}^{\rm per}_{\varepsilon^{-1}\mathcal{T}} \circ V_\varepsilon \tag{1.8}$$

Φ_ε is the "normalized, microscopic free energy functional" which we want to compare with the macroscopic functional P of (1.6).

Φ_ε and P are defined on different functional spaces, and to establish a relation between them we follow De Giorgi and his definition of Γ convergence. We start by arguing that a microscopic profile which describes the macroscopic state $u \in BV(\mathcal{T}, \{\pm m_\beta\})$ should look more and more like u as $\varepsilon \to 0$. To make it quantitative, we use the $L^1(\mathcal{T})$ norm, which weights both the volume of the region where two profiles differ and the amount of their discrepancy: this is therefore a natural candidate to quantify distances. In this language, the physical apparatus used to prepare a macroscopic state u is then schematized as a constraint which imposes the microscopic states m to be in a $L^1(\mathcal{T})$-ball of u. Thus a state u which looks sharp at the macroscopic level, becomes fuzzy after the microscopic blow up and it is better represented by a set of states, a ball in L^1 with center u, rather than by a single profile. The radius of the ball is related to the accuracy of the physical apparatus used in the preparation of the state and we imagine that it can be taken arbitrarily small, as $\varepsilon \to 0$.

To conclude, we only need to determine the free energy to associate to the L^1 ball which represents a macroscopic interface $u \in BV(\mathcal{T}, \{\pm m_\beta\})$ at the microscopic level. By invoking thermodynamic principles, the equilibrium free energy under a given constraint is the minimal free energy of the states satisfying the constraint, hence calling $\delta > 0$ the accuracy parameter identified to the radius of the L^1-ball, we set

$$\Phi_{\delta,\varepsilon}(u) = \inf_{\|m-u\|_{L^1(\mathcal{T})} \le \delta} \Phi_\varepsilon(m) \tag{1.9}$$

and call

$$\Phi_0'(u) = \lim_{\delta\to 0} \liminf_{\varepsilon\to 0} \Phi_{\delta,\varepsilon}(u), \qquad \Phi_0''(u) = \lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \Phi_{\delta,\varepsilon}(u). \tag{1.10}$$

The equality $P(u) = \Phi_0'(u) = \Phi_0''(u)$ is the De Giorgi definition that Φ_ε Γ-converges to P, in such a case we will write $\Phi_\varepsilon \xrightarrow{\Gamma} P$. Alberti and Bellettini, [1], have proved for a class of functionals which includes Φ_ε that

THEOREM. *For any u in $BV(\mathcal{T}, \{\pm m_\beta\})$ $\Phi_\varepsilon \xrightarrow{\Gamma} P$ and, if u is regular (i.e. its discontinuity set ∂E is a regular surface), then $P(u)$ is given by the expression* (1.6) *with $\theta_\beta(\nu)$ a continuous function on the unit ball of $\mathbb{R}^d$. The general theory of BV functions, see* [1] *defines for any $u \in BV(\mathcal{T}, \{\pm m_\beta\})$ a set $\partial^* E \subset \partial E$, a measure $d\mu$ on $\partial^* E$ and a unit vector function $\nu(r)$ on $\partial^* E$. In terms of these quantities,*

$$P(u) = \int_{\partial^* E} d\mu(r)\theta_\beta(\nu(r)). \tag{1.11}$$

There are also results about the value of the surface tension $\theta_\beta(\nu)$, expressed in terms of the one dimensional free energy of standing fronts, see [2], see also [5] for a related model, a uniqueness theorem for such one dimensional fronts, [7], and a proof of strict convexity and regularity of the surface tension as a function of the direction ν, [9].

The motivation of this paper is about the actual implementation of the previous definition of surface tension in a physical experiment. For that we would need a physical apparatus which forces the minus phase in the set E and the plus one in E^c. The natural way is to use an external magnetic field and, with a great deal of idealization, we will suppose to be able to set the magnetic field equal to $-s$ in a region B and equal to $+s$ in the complement, with the additional assumption that E and B are close in the symmetric difference distance, namely that $|B \triangle E| \le \delta$ (the same accuracy parameter as before). Under such a space dependent magnetic field

$$h_B(r) := s\mathbf{1}_{B^c}(r) - s\mathbf{1}_B(r) \tag{1.12}$$

equilibrium will be reached by minimizing over all m the functional $F_{\varepsilon,s}(m; B)$, defined in (2.2) below.

For any Borel subset B of the torus, we call $G_{\varepsilon,s}(B)$ the infimum of $F_{\varepsilon,s}(m; B)$ over all $m \in L^\infty(\varepsilon^{-1}\mathcal{T}; [-1, 1])$.

Our main result in this paper is a proof that $G_{\varepsilon,s}(\cdot)$ $\Gamma-$converges to a perimeter functional P_s and $P_s \to P$ as $s \to 0$, thus justifying from an operational point of view, the original definition of surface tension via $\Gamma-$convergence.

The paper is organized as follows: in the next section we give precise definitions and in Theorem 2.2.1 we state the main results. We divide the proof of Theorem 2.2.1 in two sections, the lower and the upper bound. To prove the lower bound we need some results about "contours",the argument is treated in Section 5. Finally, in the last section, we will prove the convergence to P of the surface free energy functional $P_s(E)$, see (2.4).

2 – Definitions and results

Let $\mathcal{T}$ be the unit torus in $\mathbb{R}^d$, $s > 0$ and $\varepsilon > 0$. Furthermore let $\mathcal{B}$ be the set of all Borel measurable subsets of the torus equipped with L^1-distance, which is the same as the volume of the symmetric difference:

$$|A \triangle B| := vol\,((A\backslash B) \cup (B\backslash A)) = \int |\mathbf{1}_A - \mathbf{1}_B| dr. \tag{2.1}$$

For all $m \in L^\infty(\varepsilon^{-1}\mathcal{T}; [-1; 1])$ and $B \in \mathcal{B}$ we define

$$\begin{aligned} F_{\varepsilon,s}(m(r); B) : = & \int_{\varepsilon^{-1}\mathcal{T}} f_{\beta,h_B}(m(r))dr + \\ & + \frac{1}{4}\int_{\varepsilon^{-1}\mathcal{T}}\int_{\varepsilon^{-1}\mathcal{T}} J(r, r')(m(r) - m(r'))^2 dr dr' \end{aligned} \tag{2.2}$$

where, by an abuse of notation, J is the periodization on $\varepsilon^{-1}\mathcal{T}$ of the probability kernel in (1.1), $h_B(r)$ as in (1.12), f_{β,h_B}, ϕ_{β,h_B} and $I(m)$ as in (1.2)-(1.4).
We next define, for any $B \in \mathcal{B}$,

$$G_{\varepsilon,s}(B) = \varepsilon^{d-1} \inf_{m\in L^\infty(\varepsilon^{-1}\mathcal{T};[-1,1])} F_{\varepsilon,s}(m; B). \tag{2.3}$$

Our main result is

THEOREM 2.2.1. *For any s small enough, $G_{\varepsilon,s} \xrightarrow{\Gamma} P_s$ on $BV(\mathcal{T})$, where P_s is a perimeter functional in $BV(\mathcal{T})$. Namely for any $E \in BV(\mathcal{T})$, for any $m \in L^\infty(\varepsilon^{-1}\mathcal{T}; [-1, 1])$, and for any $\delta > 0$ there exists $s^\star > 0$ and a continuous function $\theta_{\beta,s}(\nu)$ on the unit ball of $\mathbb{R}^d$, such that for any $s \le s^\star$*

$$\liminf_{\delta\to 0} \liminf_{\varepsilon\to 0} \inf_{B\in BV(\mathcal{T}):|B\triangle E|\le\delta\varepsilon^{-d}} G_{\varepsilon,s}(B) = \int_{\partial^* E} \theta_{\beta,s}(\nu)d\mu := P_s(E). \tag{2.4}$$

Moreover

$$\lim_{s\to 0} \theta_{\beta,s}(\nu) = \theta_\beta(\nu) \tag{2.5}$$

and $\lim_{s\to 0} P_s(E) = P(\chi_E)$, with P as in (1.11) and $\chi_E = m_\beta \mathbf{1}_E - m_\beta \mathbf{1}_{E^c}$.

In the next two sections we prove (2.4), while (2.5) will be proved in the last section.

3 – Lower bound

In this section we will prove that

$$\lim_{\delta\to 0}\liminf_{\varepsilon\to 0}\inf_{B\in BV(\mathcal{T}):|B\triangle E|\le\delta\varepsilon^{-d}} G_{\varepsilon,s}(B)\ge\int_{\partial^* E} d\mu(r)\theta_{\beta,s}(\nu). \tag{3.1}$$

First of all, we need some basic notions and results for the theory of BV sets, which we state in the next subsection, for more details see, for example, [1].

3.1 – Geometric measure theory

We say that a function f on $\mathcal{T}$ has bounded variation, $f\in BV(\mathcal{T})$, if its gradient Df (in the sense of distributions) is a vector real valued Radon measure whose total variation measure has finite mass $\|\mu\|$:

$$\|\mu\|=\mu(\mathcal{T})=\sup_{\phi\in C^1(\mathcal{T},\mathbb{R}^d),\|\phi\|_\infty\le 1}\left|\int_{\mathcal{T}} dr\, f\,\mathrm{div}\,\phi\right|. \tag{3.2}$$

We say that E is a general BV set if $\mathbf{1}_E\in BV(\mathcal{T})$. If E is a C^1 set, the total variation $d\mu$ of $D\mathbf{1}_E$ is the usual Hausdorff measure $dH^{d-1}(r)$ on ∂E and for any $\phi\in C(\mathcal{T},\mathbb{R}^d)$

$$\int_{\mathcal{T}}\langle D\mathbf{1}_E,\phi\rangle=-\int_{\mathcal{T}} dr\,\mathbf{1}_E\,\mathrm{div}\,\phi=-\int_{\partial E} dH^{d-1}(r)\langle\nu(r),\phi\rangle \tag{3.3}$$

where $\nu(r)$ is the outward unit normal to ∂E at r.

If E is a general BV set, then there are a set $\partial^* E\subset\partial E$, called the reduced boundary of E, and a unit vector valued function $\nu(r)$ on $\partial^* E$ so that for any $\phi\in C(\mathcal{T},\mathbb{R}^d)$

$$\int_{\mathcal{T}}\langle D\mathbf{1}_E,\phi\rangle=-\int_{\partial^* E} dH^{d-1}(r)\langle\nu(r),\phi\rangle. \tag{3.4}$$

The following theorem states that BV sets can be regarded, measure theoretically, as C^1 sets:

THEOREM 3.3.1. *Let $E\in BV(\mathcal{T})$ and $D\mathbf{1}_E(r)=-d\mu\nu(r)$. Then for any $\varepsilon>0$ there are C^1 hyper-surfaces $S_1,\dots,S_m$ whose closures are disjoint from each other, and compact sets $K_1,\dots,K_m$ with $K_i\subset S_i\cap\partial^* E$, so that*

$$d\mu\big|_{K_i}=dH^{d-1}\big|_{K_i}\qquad \int_{\mathcal{T}} d\mu-\sum_{i=1}^m\int_{K_i} d\mu\le\varepsilon. \tag{3.5}$$

Moreover the normal to S_i ar $r \in K_i$, is the same as the unit vector $\nu(r)$ in (3.3) *and*

$$\max_{i=1\dots m} \max_{r,r' \in S_i} |\nu(r) - \nu(r')| \leq \varepsilon. \tag{3.6}$$

The next Theorem states that a BV set E is with good approximation made of essentially flat parts plus a small remainder. We set

$$u = C(\mathbf{1}_{E^c} - \mathbf{1}_E). \tag{3.7}$$

THEOREM 3.3.2. *For any $\varepsilon > 0$ there are $n \geq 1$ disjoint measurable sets Σ_i, each one contained in some $K_j^{(\varepsilon)}$, n cubes R_i, all of side h, and n unit vectors ν_i, ν_i orthogonal to a face of R_i, with the following proprieties so that*

$$\sup_{r \in \Sigma_i} |\nu(r) - \nu_i| < \varepsilon, \quad \left| h^{d-1} - \int_{\Sigma_i} d\mu \right| < \varepsilon h^{d-1}, \quad \left| nh^{d-1} - \int_{\mathcal{T}} d\mu \right| < \varepsilon. \tag{3.8}$$

Moreover calling $\chi(r) := C(\mathbf{1}_{R_i^+} - \mathbf{1}_{R_i^-})$, with $R_i^{\pm}$ the upper and lower halves of R_i with to the direction ν_i,

$$\int_{R_i} dr |\chi_{R_i} - u| < \varepsilon h^d, \quad i = 1, ..., n. \tag{3.9}$$

3.2 – Proof of (3.1)

Let $R_n(L; C)$ be the cylinder in $\mathbb{R}^d$ whose axis is directed along n and whose cross section is LC, C an unit cube of $\mathbb{R}^{d-1}$ and $L > 0$ a scaling parameter. We keep n and B fixed and to simplify notation we drop them, thus writing $R(L)$ and $R(L, k)$. We introduce coordinate axes with the origin the center of $R(L, k)$, x_d axis along n and the the others parallel to the side of C, so that C is a coordinate cube. Then

$$R(L,k) = \left\{ (x_1, .., x_d) \in \mathbb{R}^d : |x_d| \leq k, |x_i| \leq L, i = 1, \dots, d-1 \right\}$$

and denote with $R_{L,k}^{\pm}$ the upper and lower halves of $R(L, k)$ with respect to the direction n. Calling

$$\chi(r) = m_{\beta,s}^+ \mathbf{1}_{x_d \geq 0} + m_{\beta,s}^- \mathbf{1}_{x_d < 0} \tag{3.10}$$

we denote by χ_Δ, $\Delta \subset \mathbb{R}^d$, the restriction of χ to Δ and we define

(3.11)

$$\theta_{\beta,s}(L,k) := \frac{1}{L^{d-1}} \inf_{\substack{m \in L^\infty(R(L,k);[-1;1]) \\ B: |R_{L,k}^+ \cap (R_{L,k}^+ \triangle B^c)| \leq \delta, \ \ |R_{L,k}^- \cap (R_{L,k}^- \triangle B)| \leq \delta}} F_s\left(m_{R(L,k)} | \chi_{R(L,k)^c}; B\right).$$

We remember that the functional $F_{\varepsilon,s}(\cdot, B)$ is defined by making the interaction J periodic over each coordinate x_i, $i < d$, with period L, thus considering LB as a torus. We thus set

$$\theta_{\beta,s} := \liminf_{L\to\infty} \liminf_{k\to\infty} \theta_{\beta,s}(L,k). \tag{3.12}$$

In some cases, when the context is not clear, we indicate with $\theta_{\beta,s,\nu}$ the surface tension defined in the rectangle $R_\nu(L,k)$ directed along ν.

We consider the small parameter α and the cubes R_i as in Theorem 3.3.2 below, $i = 1, \dots, n$, all of side k so that if we call $\tilde{\chi}_i = s(\mathbf{1}_{\varepsilon^{-1}R_i^+} - \mathbf{1}_{\varepsilon^{-1}R_i^-})$, we have

$$\begin{aligned}\int_{\varepsilon^{-1}R_i} dr|h_B - \tilde{\chi}_i| &\le \int_{\varepsilon^{-1}R_i} dr|h_B - h_E| + \int_{\varepsilon^{-1}R_i} dr|h_E - \tilde{\chi}_i| \\ &\le \varepsilon^{-d}(\delta + \alpha k^d) \le 2\alpha k^d \varepsilon^{-d}.\end{aligned}$$

Hence $|(B \triangle \varepsilon^{-1}R_i) \cap \varepsilon^{-1}R_i| \le 2\alpha k^d \varepsilon^{-d}$.

We next write

$$\Delta = \bigcup_{i=1}^{n} \varepsilon^{-1} R_i \,.$$

Then

$$F_{\varepsilon,s}(m; B) = F_{\varepsilon,s}(m_{\Delta^c}; B) + \sum_{i=1}^{n} F_{\varepsilon,s}\left(m_{\varepsilon^{-1}R_i} | m_{\Delta^c}; B\right)$$

$$\varepsilon^{d-1} F_{\varepsilon,s}(m; B) \ge \sum_{i=1}^{n} F_{\varepsilon,s}\left(m_{\varepsilon^{-1}R_i} | m_{\Delta^c}; B \cap R_i\right) \tag{3.14}$$

where

$$\begin{aligned} F_{\varepsilon,s}(m_\Lambda | m_{\Lambda^c}; B) &:= F_{\varepsilon,s}(m_\Lambda; B) + \\ &+ \frac{1}{2} \int_{\varepsilon^{-1}\Lambda} \int_{\varepsilon^{-1}\Lambda^c} J(r, r')(m_\Lambda(r) - m_{\Lambda^c}(r'))^2 dr dr' \end{aligned} \tag{3.15}$$

$$\begin{aligned} F_{\varepsilon,s}(m_\Lambda; B) &:= \int_{\varepsilon^{-1}\Lambda} f_{\beta,h_B}(m_\Lambda(r)) dr + \\ &+ \frac{1}{4} \int_{\varepsilon^{-1}\Lambda} \int_{\varepsilon^{-1}\Lambda} J(r, r')(m_\Lambda(r) - m_\Lambda(r'))^2 dr dr'. \end{aligned}$$

At the end of section we are going to prove that

$$\inf_{\substack{m\in L^\infty(\mathbb{R}^d;[-1;1]) \\ B\in BV(\mathcal{T}):|B\triangle E|\le\delta\varepsilon^{-d}}} F_{\varepsilon,s}\,(m; B) = \inf_{\substack{m\in L^\infty(\mathbb{R}^d;[-1;1]):F_{\varepsilon,s}(m;B)\le\delta\varepsilon^{-d} \\ B\in BV(\mathcal{T}):|B\triangle E|\le\delta\varepsilon^{-d}}} F_{\varepsilon,s}\,(m; B) \tag{3.16}$$

therefore, using (3.14) and (3.16), we obtain that

$$\text{l.h.s. of (3.1)} \geq$$

$$\liminf_{\alpha\to 0}\sum_i \left\{ \liminf_{\varepsilon\to 0}\varepsilon^{d-1} \inf_{\substack{F_{\varepsilon,s}(m;B)\leq 2\alpha k^d\varepsilon^{-d}\\ B\in BV(\mathcal{T}):|(B\triangle\varepsilon^{-1}R_i)\cap\varepsilon^{-1}R_i|\leq 2\alpha k^d\varepsilon^{-d}}} F_{\varepsilon,s}\left(m_{\varepsilon^{-1}R_i}|m_{\varepsilon^{-1}R_i^c};B\right)\right\}.$$

Now we state two results that we will prove later. The first one gives us a constraint on the function m, the second one gives us a lower bound on each rectangle $C(L) = R(L, L/2)$, where

$$R(L, L/2) = \left\{x \in \mathbb{R}^d : |x_i| \leq L, i = 1...d\text{-}1, |x_d| \leq L/2\right\}.$$

Notational remark: when we consider function on $L^\infty(\mathcal{T};[-1,1])$ we write $F_s(m;B)$ instead of $F_{\varepsilon,s}(m;B)$, $B \in BV(\mathcal{T})$.

Proposition 3.3.3. *Let $C(L)$ be the cylinders of the form $R(L;L/2)$. Then for any $m \in L^\infty(C(L);[-1;1])$ such that $F_s\left(m_{C(L)};B\right) \leq \delta L^{-d}$ we have*

$$\int_{C(L)} |m(r) - \chi(r)|dr \leq \delta' L^d \tag{3.17}$$

$\chi(r) = m^+_{\beta,s}\mathbf{1}_{x_d\geq 0} + m^-_{\beta,s}\mathbf{1}_{x_d<0}$, $r \in C(L)$.

Theorem 3.3.4. *There is a $c > 0$ and a continuous function $\theta_{\beta,s}(\nu)$ on the unit ball, so that for any $\varepsilon > 0$ there is $L_\varepsilon > 0$ and for any $L \geq L_\varepsilon$*

$$F_s\left(m_{C(L)}|m_{C(L)^c};B\right) \geq L^{d-1}(\theta_{\beta,s}(\nu) - \varepsilon - c\sqrt{\delta}) \tag{3.18}$$

for any $\delta > 0$,for any m s.t. $\|m - \chi\|_{L^1(C(L))} \leq \delta' L^d$ and for any $B \in BV(\mathcal{T})$ such that

$$|(B^c \triangle C^-(L)) \cap C^-(L)| \leq \delta L^d \qquad |(B \triangle C^+(L)) \cap C^+(L)| \leq \delta L^d.$$

Then, using (3.18)

$$\text{l.h.s. of (3.1)} \geq \lim_{s\to 0} \liminf_{\alpha\to 0} \sum_i k^{d-1}(\theta_{\beta,s}(\nu_i) - cn\sqrt{\alpha}).$$

By (3.8) we have $nk^{d-1} \leq \mu(\mathcal{T}) + \alpha$ and for a suitable constant c',

$$|k^{d-1}\theta_{\beta,s}(\nu_i) - \int_{\Sigma_i} d\mu\theta_{\beta,s}(\nu)| \leq c' k^{d-1}\alpha \tag{3.19}$$

so, in conclusion $\exists \quad c'' > 0$ such that

$$\text{l.h.s. of (3.1)} \geq \lim_{s\to 0} \liminf_{\alpha\to 0} \sum_i \int_{\Sigma_i} d\mu\theta_{\beta,s}(\nu) - c''\sqrt{\alpha} \tag{3.20}$$

and, see the end of this section,

$$\lim_{\alpha\to 0} \sum_i \int_{\Sigma_i} d\mu\theta_{\beta,s}(\nu) \to \int_{\partial^* E} d\mu(r)\theta_{\beta,s}(\nu)$$

thus we obtain (3.1).

Proof of (3.16). It suffices to show that $\forall\delta > 0$ and for any $B \in BV(\mathcal{T})$ there exists $m \in L^\infty(\varepsilon^{-1}\mathcal{T}; [-1, 1])$ and $\delta' > 0$ such that

$$F_{\varepsilon,s}(m; B) \leq \delta'\varepsilon^{-d}.$$

It is enough to choose $\hat{m} = m^-_{\beta,s}\mathbf{1}_{B_n} + m^+_{\beta,s}\mathbf{1}_{B_n^c}$ where B_n are the polyedrical sets which approximate $B \in BV(\mathcal{T})$ in variation, namely $\mathbf{1}_{B_n}$ converges in variation to $\mathbf{1}_B$. Indeed, computing the functional $F_{\varepsilon,s}(\hat{m}; B)$

$$\begin{aligned} F_{\varepsilon,s}(\hat{m}; B) &= \int_{\varepsilon^{-1}\mathcal{T}} f_{\beta,h_B}(\hat{m})dr + \frac{1}{2}\int_{\varepsilon^{-1}B_n}\int_{\varepsilon^{-1}B_n^c} J(r, r')(m^+_{\beta,s} - m^-_{\beta,s})^2 dr dr' \leq \\ &\leq 2h\int_{\varepsilon^{-1}\mathcal{T}} dr|\mathbf{1}_{B_n} - \mathbf{1}_B| + \int_{\varepsilon^{-1}B_n}\int_{\varepsilon^{-1}B_n^c} J(r, r')dr dr' \leq \\ &\leq 2s\delta\varepsilon^{-d} + c_n\varepsilon^{-d+1} = \delta'\varepsilon^{-d} \end{aligned}$$

with $\delta' = 2s\delta + c_n\varepsilon$.

4 – Upper Bound

In this section we will prove that

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\inf_{B\in BV(\mathcal{T}):|B\triangle E|\le\delta\varepsilon^{-d}} G_{\varepsilon,s}(B)\le\int_{\partial^* E} d\mu(r)\theta_{\beta,s}(\nu(r)). \tag{4.1}$$

Given $E\in BV(\mathcal{T})$ we can approximate in the sense of variations the function h_E by functions h_{E_k} equal to $\pm s$ outside and inside polyhedral sets E_k with boundary ∂E_k. For each k we will construct functions $m^{(\varepsilon,k,L,t)}$ so that

$$\limsup_{L\to\infty}\limsup_{t\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^{d-1}F_{\varepsilon,s}(m^{(\varepsilon,k,L,t)};E_k)\le\int_{\partial^* E_k} d\mu_k(r)\theta_{\beta,s}(\nu(r)) \tag{4.2}$$

where $d\mu_k=d\mu|_{E_k}$ as in Theorem 3.3.1. Then by letting $k\to\infty$,

$$\begin{aligned}&\limsup_{k\to\infty}\limsup_{L\to\infty}\limsup_{t\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^{d-1}F_{\varepsilon,s}(m^{(\varepsilon,k,L,t)};E_k)\le\\ &\le\limsup_{k\to\infty}\int_{\partial^* E_k} d\mu_k(r)\theta_{\beta,s}(\nu)\end{aligned} \tag{4.3}$$

and

$$\lim_{k\to\infty}\int_{\partial^* E_k} d\mu_k(r)\theta_{\beta,s}(\nu(r))=\int_{\partial^* E} d\mu(r)\theta_{\beta,s}(\nu(r)). \tag{4.4}$$

Then, by (4.2) and (4.3) there are $L(\varepsilon)$, $t(\varepsilon)$, and $k(\varepsilon)$ so that the family $m^{(\varepsilon,k(\varepsilon),L(\varepsilon),t(\varepsilon))}$ satisfies (4.1). Thus the proof of (4.1) follows from the existence of a family $m^{(\varepsilon,k,L,t,)}$ satisfying (4.2), which is proved in the rest of the subsection.

We fix k and we will drop it from the notation in the sequel. Thus we denote with E a polyhedral set and with $h_E=s(\mathbf{1}_{E^c}-\mathbf{1}_E)$. The faces of E are called σ_i, $i=1,..,n$, and their normal ν_i, directed toward the plus magnetization. On each hyperplane which contains $\varepsilon^{-1}\Sigma_i$, we introduce a partition into $d-1$ dimensional cubes of side L, the orientation of the cubes of the partition being the same for all ε. We first define $m^{(\varepsilon,L,t)}$ around $\varepsilon^{-1}\Sigma_1$: on each rectangle $R_{\nu_1}(L,t)$ of height $2t$ and mid cross section a cube entirely contained in $\varepsilon^{-1}\Sigma_1$, we choose $m^{(\varepsilon,L,t)}$ so that

$$\frac{1}{L^{d-1}}F_s(m^{(\varepsilon,L,t)}_{R_{\nu_1}(L,t)}|\chi_{R^c_{\nu_1}(L,t)};E_k)\le\theta_{\beta,s,\nu_1}(L,t)+\varepsilon. \tag{4.5}$$

When the mid cross section of $R_{\nu_1}(L,t)$ is not entirely contained in $\varepsilon^{-1}\Sigma_1$, we set $m^{(\varepsilon)}=m^{\pm}_{\beta,s}$ in the part of $R_{\nu_1}(L,t)$ which is above and below $\varepsilon^{-1}\Sigma_1\sqcap R_{\nu_1}(L,t)$. we follow the same rule in the other faces, except for the points where $m^{(\varepsilon)}$ has

already been defined. On the remaining of the space we set $m^{(\varepsilon)}$ equal to $m^{\pm}_{\beta,s}$ outside and inside E respectively. If we fix t, if L is large enough, any rectangle $R_{\nu_i}(L,t)$ at distance $> L$ from the boundary of $\varepsilon^{-1}\Sigma_i$ has no intersection with any other rectangles, then, for a suitable constant c,

$$\varepsilon^{d-1}F_{\varepsilon,s}(m^{(\varepsilon)};E) \leq \sum_{i=1}^{n}\big([\theta_{\beta,s,\nu_i}(L,t)+\varepsilon]|\Sigma_i| + cLt\varepsilon\big). \tag{4.6}$$

Then (4.2) follows, and the proof of the upper bound is completed.

5 – Contours and dynamics

In this section we give a generalized definition of contours and we study some proprieties of the evolution. For this purpose we define three basic objects. The first one is the family of partitions of $\mathbb{R}^d$

$$\big\{\mathcal{D}^\ell,\ \ell = 2^n,\ n\in\mathbb{Z}\big\}$$

$\mathcal{D}^\ell$ is a decreasing sequence of partitions into cubes C^ℓ of side ℓ. $C^{(\ell)}_r$ denotes the cube of $\mathcal{D}^\ell$ which contains r. Another basic object is the coarse-grained image of $m\in L^\infty(\mathbb{R}^d;[-1,1])$ with grain ℓ, $Av^{(\ell)}(m;r)$

$$Av^{(\ell)}(m;r) = \frac{1}{C^{(\ell)}}\int_{C^{(\ell)}_r} dr' m(r')\ , \quad |C^{(\ell)}| = \ell^d,\ m\in L^\infty(\mathbb{R}^d;[-1;1]). \tag{5.1}$$

The last basic object is the "block spin" function

$$\eta^{(\zeta,\ell)}(m;r) = \begin{cases} \pm 1 & \text{if } |Av^{(\ell)}(m;r) - m^{\pm}_{\beta,s}| \leq \zeta, \\ 0 & \text{otherwise.}\end{cases} \tag{5.2}$$

where $\zeta > 0$ and $\ell < 1$. Using these quantities we define:

- *Outer and inner boundaries.*
  The $\mathcal{D}^\ell-$outer boundary of a $\mathcal{D}^\ell-$measurable region Λ, denoted by $\delta^\ell_{out}[\Lambda]$, is the union of all the cubes C of $\mathcal{D}^\ell$ not in Λ which are connected to Λ. The $\mathcal{D}^\ell-$inner boundary $\delta^\ell_{out}[\Lambda]$ is the $\mathcal{D}^\ell-$outer boundary of Λ^c.
- *Phase Indicator.*
  Denoted by $\Theta^{(\zeta,\ell_-,\ell_+,s)}(m,B;r)$, $\ell_- < 1$, $\ell_+ > 10$, $\zeta > 0$, it is defined as $\Theta^{(\zeta,\ell_-,\ell_+,s)}(m,B;r) = \pm 1$ if $\eta^{(\zeta,\ell_-)}(m;r') = \pm 1$ for all $r'\in C^{(\ell_+)}_r \cup \delta^\ell_{out}[C^{(\ell_+)}_r]$ and $|C^{(\ell_-)}_r \cap (C^{(\ell_-)}_r \triangle B^c[B])| \leq \zeta$.
  Elsewhere $\Theta^{(\zeta,\ell_-,\ell_+)}_B((m,h);r) = 0$

- *Correct points.*
  The $\pm$ correct points of m for a set B are the points r where, respectively $\{\Theta^{(\zeta,\ell_-,\ell_+,s)}(m,B;r)=\pm 1\}$.
  The set $\{\Theta^{(\zeta,\ell_-,\ell_+,s)}(m,B;r)=0\}$, is instead the union of the spatial support of all the contours of m.
- *Approximate, local equilibrium phase spaces.*
  These are the spaces with elements m for which all points of Λ are $\pm$ correct. Such spaces are denoted by $M_{\zeta,\ell,\ell_+,\pm,\Lambda}$ and we drop Λ when $\Lambda=\mathbb{R}^d$.

5.1 – Invariance under evolution

In this subsection we will prove that the local equilibrium ensembles $M_{\zeta,\ell,\ell_+,\pm,\Lambda}$ are invariant under the partial dynamics and that the minimizers of free energy in $M_{\zeta,\ell,\ell_+,\pm,\Lambda}$ is pointwise close to $m^+_{\beta,s}$ [or to $m^-_{\beta,s}$], the closeness being exponentially with the distance from the boundaries. By simmetry, it is sufficient to prove the statement for the + ensemble, to which in the sequel we restrict.

We consider the Cauchy problem obtained, after a suitable scaling limit, by the Glauber dynamics, applied to Ising systems with Kac potentials,

$$\text{(5.3)}\qquad \begin{cases} \dfrac{dm(r,t)}{dt} = -m(r,t) + \tanh\left\{\beta[J\star m(r,t)+h_B]\right\}, & r\in\mathbb{R}^d,\ t>0;\\ m(r,0)=m(r) & r\in\mathbb{R}^d. \end{cases}$$

We also consider dynamics where, outside region Λ, the function is frozen and it acts as a boundary condition for the evolution inside Λ. Namely, we define a partial dynamics in Λ by setting

$$\text{(5.4)}\qquad \begin{cases} \dfrac{dm^{(\Lambda)}(r,t)}{dt} = -m^{(\Lambda)}(r,t)+\tanh\left\{\beta[J\star m^{(\Lambda)}+h_B]\right\}, & (r,t)\in\Lambda\times\{t>0\};\\ m^{(\Lambda)}(r,t)=m(r), & (r,t)\in(\Lambda^c\times\{t>0\})\cup(\mathbb{R}^d\times\{t=0\}). \end{cases}$$

Definition 5.5.1. Let T^Λ_t be the semigroup on $L^\infty(\mathbb{R}^d,[-1,1])$ defined by setting

$$\text{(5.5)}\qquad T^\Lambda_t(m) = \text{solution of (5.4).}$$

With similar arguments as in [5] it is possible to prove that the orbits $T^\Lambda_t(m)$ converge by subsequences as $t\to\infty$ and that the limits points satisfy the mean field equation

$$\text{(5.6)}\qquad m^{(\Lambda)}(r)=\tanh\left\{\beta[J\star m^{(\Lambda)}+h_B]\right\}\quad r\in\Lambda.$$

LEMMA 5.5.2. *There are ζ_0', κ_0 and $s^\star$ positive, so that if $\zeta < \zeta_0'$ and $\ell < \ell_0(\zeta) = \kappa_0\zeta$, then, for any $m \in M_{\zeta,\ell,\ell_+,+,\Lambda}$, with $s < s^\star$ and $r \in \Lambda$*

$$|J \star m(r) - m^+_{\beta,s}| \le 2\zeta \tag{5.7}$$

$$|\tanh\{\beta[J \star m(r) + h_B]\} - m^+_{\beta,s}| \le \zeta - \varepsilon_0(\zeta), \qquad \varepsilon_0(\zeta) = \kappa_0\zeta. \tag{5.8}$$

PROOF. Calling

$$J^{(\ell)}(r, r') = Av^{(\ell)}(J(r, \cdot); r')$$

the average of $J(r, \cdot)$ over its second variable, for ℓ small enough

$$|J(r, r') - J^{(\ell)}(r, r')| \le c\ell \mathbf{1}_{|r-r'|\le 2}, \qquad c := d\|\nabla J\|_\infty < \infty. \tag{5.9}$$

Then

$$|J \star m - J^{(\ell)} \star m| \le 2^d c\ell$$

and since

$$J^{(\ell)} \star m = J^{(\ell)} \star u, \qquad u(r) \equiv Av^{(\ell)}(m; r)$$

$$|J \star m - J^{(\ell)} \star u| \le 2^d c\ell.$$

On other hand, by assumption, $|u(r) - m^+_{\beta,s}| \le \zeta$ for all r at distance ≤ 2 from Λ, hence

$$|J^{(\ell)} \star u(r) - m^+_{\beta,s}| \le \zeta \quad r \in \Lambda$$

thus concluding

$$|J \star m(r) - m^+_{\beta,s}| \le \zeta + 2^d c\ell. \tag{5.10}$$

By choosing κ_0 so small that $\kappa_0 2^d c < 1$ we derive (5.7) from (5.8). Since

$$\frac{d}{dm}\tanh\{\beta m\}\big|_{m=m^+_{\beta,s}} \le a < 1$$

$$\begin{aligned} |\tanh\{\beta[m(r) + h_B]\} - m^+_{\beta,s}| &\le a|J \star m(r) - m^+_{\beta,s} + (h_B - s)| \\ &\le a|\zeta + 2^d c\ell + (h_B - s)|. \end{aligned}$$

Choosing $s^\star = 2^{d-1}c\kappa_0\zeta$ for any $s < s^\star$ and $\kappa_0 \le (1-a)/(1+2^d c)$

$$|\tanh\{\beta[m(r) + h_B]\} - m^+_{\beta,s}| \le \zeta\big(1 - [(1-a) - 2^d c\kappa_0]\big) \le \zeta - \kappa_0\zeta.$$

The lemma is proved. ∎

The next Lemma proves the invariance of $M_{\zeta,\ell,\ell_+,+,\Lambda}$ under the partial dynamics T_t^Λ. We omit the proof.

LEMMA 5.5.3. *If ζ, ℓ, $s^\star$ and Λ are as in Lemma 5.5.2, T_t^Λ, $t>0$, maps $M_{\zeta,\ell,\ell_+,+,\Lambda}$ into itself.*

We call

$$\mathcal{X}_{\Lambda,m}=\left\{u\in M_{\zeta,\ell,\ell_+,+,\Lambda}\ :\ u_{\Lambda^c}=m_{\Lambda^c}\right\} \tag{5.11}$$

ψ_Δ standing for the restriction of a function ψ to a set Δ.

THEOREM 5.5.4. *There are $\zeta_0<\zeta_0'$ (ζ_0', $\ell_0(\zeta)$ and $s^\star$ as in Lemma 5.5.2), ω and c_ω all positive, such that for any $B\in BV(\mathcal{T})$, $m\in M_{\zeta,\ell,\ell_+,+,\Lambda}$, and for any $s\le s^\star$, the following holds:*

$$\inf_{u\in M_{\zeta,\ell,\ell_+,+,\Lambda}} F_s\left(u_\Lambda|m_{\Lambda^c};B\right)=F_s(\psi|m_{\Lambda^c}) \tag{5.12}$$

where $\psi(r)$ is the unique solution of the mean field:

$$\psi(r)=\tanh\{\beta[J\star\psi(r)+s]\} \tag{5.13}$$

and

$$\begin{aligned}&\psi_\Lambda\in C^\infty(\Lambda,[m^+_{\beta,s}-\zeta,m^+_{\beta,s}+\zeta])\\ &\qquad|\psi_\Lambda(r)-m^+_{\beta,s}|\le c_\omega e^{-\omega\,\mathrm{dist}(r,\Lambda^c_{\neq})}\end{aligned} \tag{5.14}$$

where $\Lambda^c_{\neq}=\{r\in\Lambda^c\ \ \mathrm{dist}(r,\Lambda)\le 1;\ m_{\Lambda^c}(r)\neq m^+_{\beta,s}\}$.

In (5.12) $F_s(\cdot)$ means that the magnetic field is constantly equals to s on the whole space. Moreover when $s=0$, we simply write $F(\cdot)$.

PROOF. By Lemma 5.5.3 T_t^Λ leaves $\mathcal{X}_{\Lambda,m}$ invariant and since $\mathcal{X}_{\Lambda,m}$ is closed under uniform convergence on the compacts, for any $u\in\mathcal{X}_{\Lambda,m}$, $T_t^\Lambda u$ converges by subsequences to an element ψ of $X^0_{\Lambda,m}$:

$$X^0_{\Lambda,m}=\{\psi\in\mathcal{X}_{\Lambda,m}:\ \psi\text{ solves (5.6) }\}$$

and $F_s(u_\Lambda|m_{\Lambda^c};B)\ge F_s(\psi_\Lambda|\psi_{\Lambda^c};B)$, the inequality being strict unless $u\in X^0_{\Lambda,m}$. Therefore

$$F_s(u_\Lambda|m_{\Lambda^c};B)>\inf_{\psi\in X^0_{\Lambda,m}} F_s(\psi_\Lambda|m_{\Lambda^c};B),\qquad\text{for any } u\in\mathcal{X}_{\Lambda,m}\backslash X^0_{\Lambda,m}.$$

By (5.8), any $\psi \in X^0_{\Lambda,m}$ satisfies the first condition in (5.14). We show that if ζ is small enough then $X^0_{\Lambda,m}$ consists of only one element, ψ, which is therefore the strict minimizer of $F_s(u_\Lambda|m_{\Lambda^c};B)$. Suppose ψ and ϕ are both in $\mathcal{X}_{\Lambda,m}$, then by (5.7), $J\star\psi(r)$ and $J\star\phi(r)$, $r\in\Lambda$, are in $[m^+_{\beta,s}-2\zeta, m^+_{\beta,s}+2\zeta]$ so that, recalling that $s\le s^\star$

$$|\tanh\{\beta J\star\psi(r)+\beta h_B(r)\}-\tanh\{\beta J\star\phi(r)+\beta h_B(r)\}| \le \\ \le \frac{\beta}{\cosh^2\{\beta(m^+_{\beta,s}-2\zeta')\}}\Big(\int_\Lambda dr' J(r,r')|\psi(r')-\phi(r')|\Big)$$

since $\beta\cosh^{-2}\{\beta(m^+_{\beta,s})\}<1$ we have for $r\in\Lambda$ and a suitable constant $c<1$,

$$|\tanh\{\beta J\star\psi(r)+\beta h_B(r)\}-\tanh\{\beta J\star\phi(r)+\beta h_B(r)\}| \le c\sup_{r'\in\Lambda}|\psi(r')-\phi(r')|$$

which implies that $\phi=\psi$ in Λ, hence everywhere. By (5.8) applied to ψ_Λ

$$F_s(\psi_\Lambda|m_{\Lambda^c};B)\ge F_s(\psi_\Lambda|m_{\Lambda^c}).$$

Now we can repeat the same arguments and with $h_B=s$ every where and we obtain

$$F_s(\psi_\Lambda|m_{\Lambda^c})\ge F_s(\bar\psi_\Lambda|m_{\Lambda^c}) \tag{5.15}$$

where $\bar\psi$ satisfies (5.13). To prove the last inequality in (5.14), let $\psi\in X^0_{\Lambda,m}$ and $\phi\in X^0_{\Lambda,n}$, then, for $r\in\Lambda$

$$|\phi(r)-\psi(r)|\le e^{-2\omega}\left(\int_\Lambda dr' J(r,r')|\psi(r')-\phi(r')|+\int_\Lambda dr' J(r,r')|m(r')-n(r')|\right) \tag{5.16}$$

where we have chosen ζ_0 so small that

$$e^{-2\omega}:=\frac{\beta}{\cosh^2\{\beta(m^+_{\beta,s}-2\zeta_0)\}}<1.$$

Calling n_0 the smallest integer larger or equal to $\mathrm{dist}(r,\Lambda^c_{\not=})$, by iterating (5.16) we get

$$|\phi(r)-\psi(r)|\le\sum_{n\ge n_0}e^{-2\omega n}2\le\Big(2\sum_{n\ge0}e^{-\omega n}\Big)e^{-\omega n_0}$$

which yields (5.14) with $c_\omega:=2/(1-e^{-\omega})$ and $n(r)=m^+_{\beta,s}$. ∎

5.2 – Free energy of Contours

We call the triple (ζ, ℓ_-, ℓ_+) good if the following holds:

- The pair (ζ, ℓ_-) is good if $\zeta < \zeta_0/2$ and $\ell_- < \ell^\star(\zeta)$, $\ell^\star(\zeta) = \kappa_0\zeta$ with ζ_0 and κ_0 as in Theorem 5.5.4.
- The triple (ζ, ℓ_-, ℓ_+) is good if he pair (ζ, ℓ_-) is good, $\ell_+ > 100$ and
$$c\ell_-^d\zeta^2 \geq 2^{d+3}\ell_+^d[c_\omega e^{-\ell_+\omega/6}]^2$$
with c a suitable positive constant and ω and c_ω as in Theorem 5.5.4.

The contours of a profile m relative to the parameters (ζ, ℓ_-, ℓ_+), are the pairs $\Gamma = (sp(\Gamma), \eta_\Gamma)$, where $sp(\Gamma)$, the spatial support of Γ, is a maximal connected component of $\{r \in \mathbb{R}^d : \Theta^{(\zeta,\ell_-,\ell_+,s)}(m, B; \cdot) = 0\}$ and η_Γ is the restriction of $\eta^{(\zeta,\ell)}(m; r)$ to $sp(\Gamma)$. Γ is a bounded contour if $sp(\Gamma)$ is bounded. If Γ is bounded we set

$$K = \delta_{in}^{\ell_+}[sp(\Gamma)], \qquad A = \delta_{out}^{\ell_+}[sp(\Gamma)] \tag{5.17}$$

K is the "safety zone" of Γ.

A_0 is the maximal connected component of A contained in the unbounded component of $sp(\Gamma)^c$. K^0 the maximal connected component of K which is connected to A_0; $\eta_\Gamma \equiv 1$ or $\eta_\Gamma \equiv -1$ on K_0; in the former case Γ is a $+$ contour, in the latter a $-$ contour. The othe maximal connected components of K, if they exist, are denoted by $K_i^\pm$ $i = 1, .., n_\pm$, labelled so that $\eta_\Gamma = 1$ on K_i^+ and $\eta_\Gamma = -1$ on K_i^-. The maximal connected component of A connected to $K_i^\pm$ is called $A_i^\pm$. The maximal connected component of $sp(\Gamma)$ which contains $A_i^\pm$ is called $int_i^\pm(\Gamma)$ and we write

$$\begin{aligned} int^\pm(\Gamma) &= \bigcup_{i=1}^{n_\pm} int_i^\pm(\Gamma), \\ int(\Gamma) &= int^+(\Gamma) \cup int^-(\Gamma), \\ C(\Gamma) &= int(\Gamma) \cup sp(\Gamma) \end{aligned} \tag{5.18}$$

in the sequel we will choose ℓ_- "very small" and ℓ_+ very large, so that a correct point r is always inside a "large" region, where $\eta^{(\zeta,\ell_-)}(m; \cdot)$ is constantly equal to 1 or -1. At the same time, the region of correct points and the red zone where the deviations from equilibrium are localized, are separated by the safety zone, where $\eta^{(\zeta,\ell)}(m; r)$ has a constant non zero value.

THEOREM 5.5.5. *Let (ζ, ℓ_-, ℓ_+) be good, $m \in L^\infty(\mathbb{R}^d, [-1, 1])$, $s^\star$ as in Lemma 5.5.2, $B \in BV(\mathcal{T})$ and Γ a (ζ, ℓ_-, ℓ_+), $+$ bounded contour for m, then for any $s < s^\star$ there is $\psi \in L^\infty(\mathbb{R}^d, [-1, 1]$ equal to m on $C(\Gamma)^c$, to $m_{\beta,s}^+$ on $C(\Gamma)\backslash K_0$ and with ψ with values in $[m_{\beta,s}^+ - \zeta + \varepsilon, m_{\beta,s}^+ + \zeta - \varepsilon]$ on K_0 such that*

$$F_s\left(m_{C(\Gamma)}|m_{C(\Gamma)^c}; B\right) \geq F_s\left(\psi_{C(\Gamma)}|\psi_{C(\Gamma)^c}\right) \tag{5.19}$$

PROOF. We need to prove that

$$F_s\left(m;B\right) \geq F_s\left(\psi\right). \tag{5.20}$$

Let Σ_0 be a $\mathcal{D}^{(\ell'_-)}$-measurable circuit contained in K_0 whose complement is made of two unconnected components at mutual distance ≥ 1, calling $ext(\Sigma_0)$ the one which contains A_0. We also suppose that Σ_0 has distance $\leq \ell'/3$ from $S_0 := \delta^1_{in}[K_0]$. By Theorem 5.5.4 applied to $K_0\backslash S_0$ with boundary conditions the restriction of m to S_0 there is ϕ equal to m outside $K_0\backslash S_0$, which, on $K_0\backslash S_0$ has values in $[m^+_{\beta,s} - \zeta + \varepsilon, m^+_{\beta,s} + \zeta - \varepsilon]$, $\varepsilon = \varepsilon_0(\zeta')$ and such that

$$F_s(\phi_{K_0\backslash S_0}|m_{S_0}) \leq F_s(m_{K_0\backslash S_0}|m_{S_0};B)$$

$$|\phi(r) - m^+_{\beta,s}| \leq c_\omega e^{-\omega\ell'/3} \quad \text{on } \Sigma_0 \tag{5.21}$$

setting $\Delta = \Sigma_0 \cup ext(\Sigma_0)$, we have

$$F_s(\phi) = F_s(\phi_{\Delta^c}|\phi_\Delta) + F_s(\phi_\Delta) \geq F_s(\phi_\Delta).$$

Set $\psi = \phi$ on Δ and equal to $m^+_{\beta,s}$ on Δ^c, we are going to prove that

$$F_s(\phi_{\Delta^c}|\phi_\Delta) \geq F_s(\psi_{\Delta^c}|\psi_\Delta). \tag{5.22}$$

Indeed, since $F_s(\psi_{\Delta^c}) = 0$, we have

$$\begin{aligned} F_s(\psi_{\Delta^c}|\psi_\Delta) &= \frac{1}{2}\int_{\Sigma_0} dr \int_{\Delta^C} dr' J(r,r')(\phi(r) - m^+_{\beta,s})^2 \leq \\ &\leq F_s(\phi_{\Delta^c}) + \frac{1}{2}\int_{\Delta} dr \int_{\Delta^C} dr' J(r,r')(\phi(r) - \phi(r'))^2. \end{aligned} \tag{5.23}$$

The last inequality follows from the fact that the interaction between ψ_Δ and ψ_{Δ^c} is very small, i.e.

$$\int_{\Sigma_0} dr \int_{\Delta^C} dr' J(r,r')(\phi(r) - m^+_{\beta,s})^2 \leq \frac{|\Sigma_0|}{2}[c_\omega e^{-\ell'_+/3}]^2.$$

Then, since $F_s(\phi_\Delta) = F_s(\psi_\Delta)$ and from (5.23)

$$F_s(\phi) = F_s(\phi_\Delta) + F_s(\phi_{\Delta^c}|\phi_\Delta) \geq F_s(\psi_\Delta) + F_s(\psi_{\Delta^c}|\psi_\Delta) = F_s(\psi)$$

and then the theorem is proved. □

In the proof of Theorem 3.3.4 we use the following Corollary, whose for brevity we omit the proof.

COROLLARY 5.5.6. *Let (ζ, ℓ_-, ℓ_+) be good, $s^\star$ as in Lemma 5.5.2, Λ and $\Delta \subset \Lambda$ two bounded, $\mathcal{D}^{(\ell_+)}$-measurable regions; $m \in L^\infty(\mathbb{R}^d, [-1,1])$ with $\eta^{(\zeta,\ell_-)}(m;r) = 1$, $r \in \delta^{\ell_+}_{out}[\Lambda] \cup \delta^{\ell_+}_{in}[\Lambda]$, $B \in BV(\mathcal{T})$ with $|C_r^{(\ell_-)} \cap (C_r^{(\ell_-)} \triangle B^c)| \leq \zeta \ell_-^d$, $r \in \delta^{\ell_+}_{out}[\Lambda] \cup \delta^{\ell_+}_{in}[\Lambda]$. Then there is a $\phi \in L^\infty(\mathbb{R}^d, [-1,1])$ so that $\phi = m$ on Λ^c, $\phi = m^+_{\beta,s}$ on Δ, $\eta^{(\zeta,\ell_-)}(\phi;r) = 1$ on Λ and for any $s < s^\star$, calling*

$$\delta\Delta = \{r \in \Delta : dist(r, \Delta^c) \leq 1\}, \quad \Lambda^c_{\neq} = \{r \in \Lambda^c, m(r) \neq m^+_{\beta,s},\ \mathrm{dist}(r,\Lambda) \leq 1\}$$

$$F_s\left(m_\Lambda | m_{\Lambda^c}; B\right) \geq F_s\left(\phi_\Lambda | \phi_{\Lambda^c}\right) - (2c\omega \exp^\omega |\delta\Delta|) \exp^{-\omega\, \mathrm{dist}(\Delta, \Lambda^c_{\neq})} \tag{5.24}$$

6 – The surface tension

In this section we prove (2.5), Proposition 3.3.3 and Theorem 3.3.4.
Now we prove that

$$\lim_{s\to 0} \theta_{\beta,s} = \liminf_{k\to\infty} \lim_{s\to 0} \liminf_{L\to\infty} \theta_{\beta,s}(L,k) = \theta_\beta(\nu) \tag{6.1}$$

which clearly implies (2.5). We observe that (6.1) shows also that it is possible to obtain the same value by taking limits in the reverse order. To simplify the notation we omit the dependence on β writing θ_s instead of $\theta_{\beta,s}$. First of all we want to prove that

$$\lim_{s\to 0} \theta_s(L,K) = \theta(L,K) := \inf_{m \in L^\infty(R(L,K);[-1;1])} F\left(m_{R(L,k)} | \chi_{R(L,k)^c}\right). \tag{6.2}$$

Let m and B be so that

$$L^{d-1}\theta_s(L,k) = F_s\left(m_{R(L,k)} | \chi_{R(L,k)^c}; B\right).$$

Then for s small enough there exists ε

$$F_s\left(m_{R(L,k)} | \chi_{R(L,k)^c}; B\right) \geq F\left(m_{R(L,k)} | \chi^0_{R(L,k)^c}\right) - \varepsilon \geq L^{d-1}\theta(L,K) - \varepsilon$$

where $\chi^0(r) = m_\beta \mathbf{1}_{x_d \geq 0} - m_\beta \mathbf{1}_{x_d \leq 0}$.

On the other hand let $\tilde{m}$ such that

$$L^{d-1}\theta(L,k) = F\left(\tilde{m}_{R(L,k)} | \chi^0_{R(L,k)^c}\right).$$

By straigthforward computation it is possible to show that

$$|F\left(\tilde{m}_{R(L,k)}|\chi^0_{R(L,k)^c}\right) - F_s\left(\tilde{m}_{R(L,k)}|\chi^0_{R(L,k)^c}; B\right)| \leq sKL^{d-1} \tag{6.3}$$

and

$$F_s\left(\tilde{m}_{R(L,k)}|\chi^0_{R(L,k)^c}; B\right) \geq F_s\left(\tilde{m}_{R(L,k)}|\chi_{R(L,k)^c}; B\right) + s^2L^d. \tag{6.4}$$

Then, using (6.3) and (6.4)

$$\theta(L,K) \geq \lim_{s\to 0} \theta_s(L,K).$$

Hence (6.2) is proved.

The next step is

$$\lim_{s\to 0}\theta_s \leq \liminf_{k\to\infty}\lim_{s\to 0}\liminf_{L\to\infty}\theta_s(L,k) \leq \liminf_{k\to\infty}\liminf_{L\to\infty}\theta(L,k). \tag{6.5}$$

It is easy to check that $\theta_s(L,k)$ is a non increasing function of K, i.e.

$$\liminf_{k\to\infty}\theta_s(L,k) = \inf_K \theta_s(L,k) := \theta_s(L).$$

This implies that

$$\theta_s(L) \leq \theta_s(L,k) \qquad \liminf_{L\to\infty}\theta_s(L) \leq \liminf_{L\to\infty}\theta_s(L,k).$$

By letting first $s \to 0$ and then $k \to \infty$ we obtain the first inequality in (6.5). The last inequality follows from (6.2) and by letting the limits in the following order: first $L \to \infty$, $s \to 0$ and then $k \to \infty$. Using again (6.2) we can obtain

$$\liminf_{k\to\infty}\liminf_{L\to\infty}\theta(L,k) \leq \lim_{s\to 0}\theta_s \tag{6.6}$$

that together with (6.5) completes the proof of (2.5).

6.1 – Proof of Proposition 3.3.3

By definition of χ,

$$\int_{C(L)} |m(r) - \chi(r)|dr = \int_{C^-(L)} |m(r) - m^-_{\beta,s}|dr + \int_{C^+(L)} |m(r) - m^+_{\beta,s}|dr.$$

We define

$$A_\zeta = \left\{ r \in C^-(L) \text{ s.t. } |m(r) - m^-_{\beta,s}| \leq \zeta \right\}$$

and A_η the analogous on $C^+(L)$. Then

$$\int_{C(L)} |m(r)-\chi)|dr \le (\zeta+\eta)L^d + \int_{A^c_\zeta} |m-m^-_{\beta,s}|dr + \int_{A^c_\eta} |m-m^+_{\beta,s}|dr.$$

In A^c_ζ we have that

$$\int_{A^c_\zeta} |m-m^-_{\beta,s}|dr \le \frac{c}{(\zeta^2\wedge h)} \int_{A^c_\zeta} f^-_h(m) \le \frac{c}{(\zeta^2\wedge s)}\delta L^d.$$

With same arguments in A^c_η we obtain

$$\int_{C(L)} |m(r)-\chi(r)|dr \le (\zeta+\eta)L^d + \frac{c}{(\zeta^2\wedge s)}\delta L^d + \frac{c}{(\eta^2\wedge s)}\delta L^d \le \delta' L^d.$$

6.2 – Proof of the Theorem 3.3.4

We define the i-th layer, $i\in\mathbb{Z}$,

$$S_i = \{x\in C(L) : (x_d - \ell_+ i)\in[-\ell_+/2, \ell_+/2)\}. \tag{6.7}$$

Let

$$N = \min\{n\in\mathbb{N} : 2n\ell_+ \ge \sqrt{\delta}L\}. \tag{6.8}$$

Supposing $\sqrt{\delta}$ small enough, we define, for any $1\le n\le N$

$$\Sigma_n := S_{2n-1}\cup S_{2n}\cup S_{2n-1+2N}\cup S_{2n+2N}$$

and in the same way Σ_{-n}, observing that $|\Sigma_n\cup\Sigma_{-n}| = 8|S_0|$.

We will use the estimates of Section 5 in the boxes delimited by Σ_n and Σ_{-n} respectively, to conclude that in the center layer of the boxes we can replace m by $m^+_{\beta,s}$ and $m^-_{\beta,s}$, and h by $\pm s$. Let

$$a_n = \frac{1}{8|S_0|}\left\{\int_{\Sigma_n\cup\Sigma_{-n}} dr|m-\chi| + \int_{\Sigma_n\cup\Sigma_{-n}} dr|h_B-\tilde\chi|\right\}. \tag{6.9}$$

Where $\tilde\chi$ is defined as (3.13) with $C^\pm(L)$ instead of $R^\pm_i$. Then

$$a = \min_{n\le N} a_n \le C\sqrt{\delta}. \tag{6.10}$$

In fact, by assumption

$$3\delta L^d \geq \int_{C(L)} dr|m-\chi| + \int_{C(L)} dr|h_B - \tilde{\chi}| \geq$$
$$\geq \sum_{i=1}^{N} 8|S_0|a_n \geq 8|S_0|aN = 8L^{d-1}N\ell^+ a \geq 4\sqrt{\delta}L^d a$$

which proves (6.10).

Call n the integer where the minimum in (6.10) is achieved. Now we are going to use the analysis of Section 5. We shorthand $\eta(\cdot;\cdot)$ for $\eta^{(\zeta,\ell_-)}(\cdot;\cdot)$ and we define:

- $C^0(L)$ is the union of all cubes $C \in \mathcal{D}^{(\ell_+)}$ such that both C and $\delta^{\ell_+}_{out}[C]$are in $C(L)$.
- $\mathcal{M}_n$ is the union of all cubes $C \in \mathcal{D}^{(\ell_-)}$ contained in Σ_n where $\eta(m;\cdot) < 1$ and $C \cap B \neq 0$, of those in Σ_{-n} where $\eta(m;\cdot) > -1$ and $C \cap B^c \neq 0$ and of the set
$$\delta^{\ell_+}_{out}[C^0(L)] \sqcap \{\sqcup_{|j|\leq 4N} S_j\}.$$

We want to estimate the free energy cost changing m and h_B into new functions ϕ and $\tilde{h}_B$ set respectively equal to χ and $\tilde{\chi}$ on $\mathcal{M}_n$ and unchanged everywhere else. We need an estimate on the volume $|\mathcal{M}_n|$. It's easy to prove that for a suitable constant c the following estimate holds:

$$|\mathcal{M}_n| \leq c\sqrt{\delta}L^{d-1}. \tag{6.11}$$

Then there is a constant $c_0 > 0$ so that

$$F_s(m_{C(L)}|m_{C(L)^c}; B) \geq F_{\tilde{s}}(\phi_{C(L)}|\phi_{C(L)^c}; B) - c_0|\mathcal{M}_n|. \tag{6.12}$$

Indeed the first term in the functional does not increase when replacing m by ϕ and the other changes are proportional to the volume where have been made.

Recalling the definition of $C^0(L)$ and since

$$F_{\tilde{s}}(\phi_{C(L)}|\phi_{C(L)^c}; B) \geq F_{\tilde{s}}(\phi_{C^0(L)}|\phi_{C^0(L)^c}; B).$$

We have

$$F_s(m_{C(L)}|m_{C(L)^c}; B) \geq F_{\tilde{s}}(\phi_{C^0(L)}|\phi_{C^0(L)^c}; B) - C_0 c\sqrt{\delta}L^{d-1}. \tag{6.13}$$

Let then Λ_+ be the box in $C^0(L)$ union of all $S_j \sqcap C^0(L)$ with $2n < j \leq 2n+2N-1$ and let Λ_- be its reflection around $x_d = 0$. We are going to apply the Corollary 5.5.6 with $\Lambda = \Lambda_+$ and $\Delta = S_{2n+N} \sqcap C^0(L)$ and then with their images under

reflecion around $x_d = 0$. By symmetry we only consider the former and drop the sub fix $+$. The hypotheses of Corollary are here met because:

$$\eta(\phi; r) = 1,\ r \in \delta^{\ell_+}_{out}[\Lambda] \qquad \Lambda^c_{\neq} \subset \{S_{2n-1} \cup S_{2n+2N}\} \tag{6.14}$$

thus

$$\mathrm{dist}(\Delta, \Lambda^c_{\neq}) \geq \ell_+ N/2. \tag{6.15}$$

There is ψ equal to ϕ outside $\Lambda_\pm$ and equal to χ on $S_{2n+N} \sqcap C^0(L)$ and $S_{-2n-N} \sqcap C^0(L)$ such that

$$F_{\tilde{s}}(\phi_{C^0(L)}|\phi_{C^0(L)^c}; B) \geq F_{\tilde{s}}(\psi_{C^0(L)}|\psi_{C^0(L)^c}; B) - (2c_\omega e^\omega |S_0|)e^{-\omega \ell_+ N/2}. \tag{6.16}$$

Setting

$$U := \bigcup_{|j|<2n+N} \{S_j \sqcap C^0(L)\}. \tag{6.17}$$

We get

$$F_{\tilde{s}}(\psi_{C^0(L)}|\psi_{C^0(L)^c}; B) \geq F_{\tilde{s}}(\psi_U|\psi_{U^c}; B) = F_{\tilde{s}}(\psi_U|\chi_{U^c}; B) \tag{6.18}$$

U is a rectangle whose basis is a cube of side b, $L \geq b \geq L - 2\ell_+$; denoting by k the height of U we then have, recalling (3.12),

$$F_{\tilde{s}}(\psi_U|\chi_{U^c}; B) \geq b^{d-1}\theta_{\beta,s}(b, k). \tag{6.19}$$

Given $\varepsilon > 0$, we may choose $L_\varepsilon > 0$ so large that $\theta_{\beta,s}(b, k) > \theta_{\beta,s}(k) - \varepsilon/2$, and by letting $k \to \infty$ and using (6.13) we obtain (3.18).

Acknowledgements

We are indebted to A. De Masi and E. Presutti for helpful discussions and comments.

REFERENCES

[1] G. Alberti – G. Bellettini: *A non-local anisotropic model for phase transition: asymptotic behaviour of rescaled energies*, Euro. Journal of Applied Mathematics, **9** (1998), 261-284.

[2] G. Alberti – G. Bellettini: *A non-local anisotropic model for phase transition: the optimal profile problem*, Math. Annalen, **310** (1998), 527-569.

[3] G. Bellettini – P. Buttà – E. Presutti: *Sharp interface limits for non local anisotropic interactions*, Arch. Rational Mech. Anal., **159** (2001), 109-135.

[4] T. Bodineau: *Phase coexistence for the Kac Ising models*, Progr. Probab., Birkhuser Boston, Boston, MA, **51** (2002), 75-111.

[5] A. De Masi – T. Gobron: *Liquid-vapor interfaces and surface tension in a mesoscopic model of fluid with nonlocal interactions*, preprint, 2003.

[6] A. De Masi – E. Orlandi – E. Presutti – L. Triolo: *Glauber evolution with Kac potentials. I. Mesoscopic and macroscopic limits, interface dynamics*, Nonlinearity, **7** (1994), 1-67.

[7] A. De Masi – E. Orlandi – E. Presutti – L. Triolo: *Uniqueness and global stability of the instanton in non local evolution equations*, Rend. di Matematica, **14** (1993).

[8] J. L. Lebowitz – O. Penrose: *Rigorous treatment of the van der Waals Maxwell theory of the liquid vapour transition*, J. Math. Phys., **7** (1966), 98-113.

Lavoro pervenuto alla redazione il 29 agosto 2003
ed accettato per la pubblicazione il 8 aprile 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo degli Autori:

Cristiana Bisceglia, Emanuele Rosatelli – Dipartimento di Matematica Pura ed Applicata – Università de L'Aquila – Via Vetoio (Coppito) 67100 L'Aquila, Italy
E-mail: biscegli, rosatell@univaq.it.

Rendiconti di Matematica, Serie VII
Volume 25, Roma (2005), 109-125

Analysis of the Laplacian of an incomplete manifold with almost polar boundary

JUN MASAMUNE

Abstract: *Motivated by recent interest in global analysis of singular manifolds, we establish the essential self-adjointness of a Laplacian, a Liouville property of sub-harmonic functions, conservativeness and parabolicity of an incomplete manifold. These results are applicable for manifolds with fractal Cauchy boundary.*

Let M be a connected $C^{1,1}$-Riemann manifold without boundary. In our previous paper [13], we had studied that if the Cauchy boundary $\partial M = \overline{M} \setminus M$, where $\overline{M}$ is the completion of M, is almost polar (see Definition 3), then a Laplace-Beltrami operator (hereafter, Laplacian, in short) is essentially self-adjoint. We call such a manifold a *manifold with almost polar boundary*. The present paper is a continuation of this previous work. Here we will investigate the spectral theory of an incomplete manifold such as: The essential self-adjointness of the Laplacian, conservativeness and parabolicity of the manifold, a Liouville property of sub-harmonic functions.

The typical example $M = N \setminus \Sigma$ is a complete manifold N deleted a closed manifold Σ of co-dimension ≥ 2. More crucial example is: M itself is a manifold but the completion $\overline{M}$ is no more a manifold. For example, $\overline{M}$ may be an algebraic variety with singular set, a football, an orbifold, a Met_1-surface, so called *singular manifolds*. We allow ∂M to be a fractal (see Section 5 for examples).

Key Words and Phrases: *Laplacian – Essential self-adjointness – Lioville property – Conservativeness – Parabolicity – Incomplete manifold – Singular manifold.*
A.M.S. Classification: 58J35 – 60J65 – 53C43

We consider a Laplacian $\Delta = \operatorname{div} \cdot \nabla$ with the following domain: the domain $D(\nabla)$ of ∇ is the set of $C^{1,1}$-functions f such that both f and ∇f are square integrable. Similarly, the domain $D(\operatorname{div})$ of div is the set of $C^{1,1}$-vector fields X such that X and $\operatorname{div} X$ are square integrable. Then let the domain $D(\Delta)$ of Δ be the set of functions $f \in D(\nabla)$ such that $\nabla f \in D(\operatorname{div})$. M is said to have *negligible boundary* if $-\operatorname{div}$ is a formal adjoint of ∇. The following is our main result.

THEOREM 1. *Let M be a connected $C^{1,1}$-Riemann manifold without boundary. If ∂M is almost polar, then M has negligible boundary. Moreover,*

(i) *If M has negligible boundary, then the Laplacian Δ is essentially self-adjoint.*
(ii) *Δ is essentially self-adjoint if and only if two Sobolev spaces W_0 and W coincide, and moreover, then the L^2-closure $\overline{\Delta}$ coincides with both Dirichlet and Neumann Laplacians.*
In the sequel, assume Δ to be essentially self-adjoint.
(iii) *If the volume $v(r)$ of the ball $B(x,r)$ of radius r centered at an arbitrary but fixed point $x \in M$ satisfies*

$$\int^{\infty} \frac{r}{\log v(r)} = \infty, \tag{1}$$

then M is conservative. If $v(r)$ satisfies

$$\int^{\infty} \frac{r}{v(r)} = \infty, \tag{2}$$

then M is parabolic.
(iv) *Every sub-harmonic function f belonging to $D(\overline{\Delta})$ or to $L^{\infty} \cap L^p$ for an arbitrary $p > 1$, is a constant.*

It is known that condition (1) (resp. (2)) of Theorem 1 implies the conservativeness (resp. parabolicity) of a complete manifold [8].

An immediate application of Theorem 1 is

COROLLARY 1. *Let us assume M satisfies the condition of Theorem 1 and Δ is essentially self-adjoint. If M is not bounded and the Ricci curvature is non-negative, then the following holds.*

(i) *Every harmonic 1-form α such that $|\alpha| \in D(\overline{\Delta})$ is 0.*
(ii) *Every harmonic map $f : M \to N$ such that the energy is in $D(\overline{\Delta})$ is constant, where N is a complete smooth manifold whose sectional curvature is non-positive. The condition such that M is not bounded and the Ricci curvature is non-negative may be replaced to that the Ricci curvature is positive at some point.*

We organize the paper in the following manner.

In Section 1, we establish notation.

In Section 2, we discuss Sobolev spaces $W_0^{1,2}$ and $W^{1,2}$. The nature of the research through the paper is the fact two Sobolev spaces $W_0^{1,2}$ and $W^{1,2}$ coincide if ∂M is almost polar (Theorem 2). This result has been well known for an incomplete manifold $M = N \setminus \Sigma$, where N is a complete manifold and Σ is a closed almost polar set. Let us explain why our setting covers this case. Since Σ is almost polar, it has volume 0, hence any ball centered at arbitrary point $x \in \Sigma$ has intersection with N. Therefore, there exists a sequence in N that converges to x. This shows $N = \overline{M}$, because N is complete. The contribution of our study is to generalize the previous result to an incomplete manifold such that whose completion is no more a manifold.

In Section 3, we study the essential self-adjointness of the Laplacian. In general, if a symmetric operator on a Hilbert space has a unique self-adjoint extensions, it is called *essentially self-adjoint*. This problem has been introduced to Riemannian geometry by M. P. Gaffney [6]. He established a sufficient condition for a manifold called M *has negligible boundary* so that the Laplacian Δ on forms is essentially self-adjoint. We present two alternative proofs of Gaffney theorem. Subsequently, together with the main result in [5], Gaffney proved the essential self-adjointness of the Laplacian on forms of complete manifolds. We prove the Laplacian is essentially self-adjoint if and only if $W = W_0$.

In Section 4, we study the conservativeness, parabolicity, and a Liouville property. The idea of our study of conservativeness and parabolicity bases on the following fact. Consider again $M = N \setminus \Sigma$, where N is a complete manifold and Σ is a closed almost polar set. Then *the Brownian motion of N does not hit Σ, accordingly, if N is conservative or parabolic, then so is M.* On our setting, we discuss without asking if Brownian motion hits ∂M or not, because we do not know if the Brownian motion could be extended to $\overline{M}$.

In order to establish the conservativeness, we decompose M into M_1 and M_2, where $\partial M \subset \partial M_1$, M_1 has finite volume. We impose Neumann boundary condition to both manifolds. Then both manifolds are conservative (A. Grigor'yan [8] proved that (1) in Theorem 1 is a sufficient condition for the conservativeness of a complete manifold or a manifold with boundary of Neumann condition). As we had seen M has no boundary condition (Theorem 2), we will obtain the conservativeness of M. One may prove the parabolicity in the same way, however, we present a different proof.

In Section 5, we provide examples. P. Li and G. Tian [11] proved the essential self-adjointness of the Laplacian, the conservativeness of an incomplete manifold $M \setminus \Sigma$, where $M \subset \mathbb{CP}^n$ is an algebraic variety deleted the singular set Σ of co-dimension not less than 3. Since M has finite volume and the singular set is almost polar, $M \setminus \Sigma$ is not only conservative but more strongly, parabolic. See more detailed proof for their result in [24]. The main reason why we study the manifold of class $C^{1,1}$ is because the most simple but non-trivial manifold

with fractal boundary is merely $C^{1,1}$ (Example 7). Finally, let us introduce some related topics to our result. The Laplacian Δ_C on the set of forms with compact support of a complete manifold is essentially self-adjoint [20]. This result contains the Gaffney theorem on a complete manifold because $\Delta_C \subset \Delta$. However, Δ_C is not necessarily essentially self-adjoint on a manifold with polar boundary, while the corresponding Markov form $(\mathcal{E}, C_0^\infty)$ has unique Dirichlet extension. Indeed, Δ_C on $\mathbb{R}^3 \setminus \{0\}$, where $\{0\}$ is almost polar, has infinitely many self-adjoint extensions [9], [2]. The Laplacian Δ_C on $M \setminus N$, where M is a complete and N is a closed sub-manifold, is essentially self-adjoint if and only if the co-dimension of N is greater than 3 [13].

S. Ozawa [16], studied the behavior of the first eigenvalue λ_ϵ of $M \setminus B_\epsilon$ where M is a compact manifold as $\epsilon \to 0$. P. Li and G. Tian established an eigenvalue estimate of an algebraic variety deleted the singular set [11]. The author and W. Rossman proved the Weyl's asymptotic formula for an incomplete manifold [14]. G. C. Papanicolaou and S. R. S. Varadhan [17] analyzed the asymptotic behavior of the solution of the heat equation on a domain of an Euclidean space punched out small balls.

1 – Notation

We list up notation for convenience of reading. Most of them will be explained also in the main body of the paper.

M - a connected Riemann manifold of class $C^{1,1}$ without boundary. M admits an atlas such that every coordinate transformation is $C^{1,1}$ and Riemann metric is Lipschitz on every compact set.

- μ - the Riemann measure.
- d - the intrinsic distance of M - see Section 2.
- $r := d(\cdot, x)$ - the radius function from the point $x \in M$.
- $\overline{M}$ - the completion of M with respect to d.
- $\partial M := \overline{M} \setminus M$.
- $B(\Sigma, r) := \{x \in \overline{M} \,|\, d(\Sigma, x) < r\}$ - the r-neighbourhood of the set $\Sigma \subset \overline{M}$.
- $\Delta := \operatorname{div} \nabla$ - the Laplace-Beltrami operator on M.
- $\overline{\Delta}$ - the L^2-closure of Δ - see Section 2.
- Δ_D (resp. Δ_N) - Dirichlet (resp. Neumann) Laplacian - see Section 3.
- $p(t, x, y)$ - the heat kernel associated with $\frac{1}{2}\Delta$ - see Section 4.
- Ω - a bounded domain of M.
- C^l - the set of real-valued functions of class l on M.
- $C_0^l(\Omega)$ - the set of functions $f \in C^l$ with compact support in Ω.
- V^l - the set of real-valued vector fields of class l on M.

- $V_0^l(\Omega)$ - the set of vector fields in V^l with compact support in Ω.
- $f|_\Omega$ - the function f restricted to Ω.
- $L^p := L^p(M,\mu)$ - the completion of $C_0^{1,1}$ with respect to the norm $\|f\|_p := (\int f^p)^{1/p} := (\int_M f^p(x)\,d\mu(x))^{1/p}$. Especially $\langle f,g\rangle := \int fg$ for $f,g \in L^2$.
- e^{tT} - the semi-group generated by a non-positive self-adjoint operator T on L^2.
- $W := W^{1,2}(M,\mu)$, $W_0 := W_0^{1,2}(M,\mu)$, and $H := H_2^1(M,\mu)$ - Sobolev spaces of order $(1,2)$ - see Section 2.
- $\mathcal{E}(f,g) := \langle \nabla f, \nabla g\rangle$ - the Dirichlet integral of $f,g \in W$.
- $\mathrm{Cap}(\Sigma)$ - the capacity of a Borel set $\Sigma \subset \overline{M}$ - see Section 2.

2 – Sobolev spaces

The main purpose of this section is to prove; if ∂M is almost polar, then $W = W_0$. On a complete manifold, where the Cauchy boundary is empty, this goes back to Gaffney [5]. He cuts off the function $f \in W$ out side of a ball $B(r)$, and prove that the modified function f_r belongs to W_0 and converges to f as $r \to \infty$. If a manifold is incomplete, one should cut off f also near the Cauchy boundary ∂M. We will prove that if ∂M is almost polar, then this modified function f_n belongs to W_0 and converges to f (Theorem 2).

Definition 1. Denote by W the completion of the set of real-valued $C^{1,1}$-functions f on M such that $\|f\|_{1,2} = \|f\|_2 + \|\nabla f\|_2 < \infty$, where $\|\cdot\|_2$ stands for the L^2-norm, with respect to the norm $\|\cdot\|_{1,2}$. The set W_0 is the completion of the set of functions in $C^{1,1}$ with compact support $C_0^{1,1}$ in W. Another Sobolev type space H consists of measurable functions f such that both f and ∇f are square integrable.

The Riemann distance does not work on a $C^{1,1}$-manifold, so we work with the intrinsic distance [1].

Definition 2. The *intrinsic distance* d is defined by

$$d(x,y) = \sup\{\psi(x) - \psi(y) \,|\, \psi \in C^{1,1},\ \|\nabla\psi\|_\infty \le 1\} \text{ for } x,y \in M.$$

We impose

Assumption 1. d is non-degenerate and generates the original topology of M.

Remark 1. It is known that d coincides with the Riemann distance, if the manifold is class $C^{2,1}$ [7].

EXAMPLE 1. K. TH. STURM [21] developed the conservativeness, parabolicity, and L^p-Liouville property of a local Dirichlet space utilizing the canonical intrinsic distance associated to the Dirichlet form. A local Dirichlet space is a generalization of a Riemann manifold, so his result covers complete manifolds, however, his assumption excludes incomplete manifolds.

We define the capacity for $\overline{M}$.

DEFINITION 3. Let $\Sigma \subset \overline{M}$ be a Borel set. Denote by $\mathcal{O}$ the family of open sets O of $\overline{M}$ such that $\Sigma \subset O$. Let $L(O)$ be the set of functions $f \in W_0$ such that

$$0 \leq f \leq 1 \text{ and } f|_O = 1.$$

The capacity $\mathrm{Cap}(\Sigma)$ of Σ is

$$\mathrm{Cap}(\Sigma) = \inf_{O \in \mathcal{O}} \mathrm{Cap}(O),$$

where

$$\mathrm{Cap}(O) = \inf_{f \in L(O)} \|f\|_{1,2}.$$

We say Σ is *almost polar* if $\mathrm{Cap}(\Sigma) = 0$.

REMARK 2. The Brownian motion on M hits $\Sigma \subset M$ if and only if $\mathrm{Cap}(\Sigma) > 0$, so the Brownian motion on M and that of $M \setminus \Sigma$ (in order to make $M \setminus \Sigma$ a manifold, Σ should be closed) are the same almost surely, if Σ is almost polar. If Σ is a manifold and co-dimension is not less than 2, or a fractal with Hausdorff co-dimension greater than 2, then it is almost polar.

EXAMPLE 2. See Section 5 for examples of manifolds M with almost polar boundary such that $\overline{M}$ is not a manifold.

Recall the definition of the *closure* of an operator.

DEFINITION 4. An operator $S : H_1 \to H_2$, where H_1 and H_2 are Hilbert spaces, is called *closed* if the graph $G(S)$ is closed in $H_1 \times H_2$. S is called *closable* if it has a closed extension. The operator T whose graph $G(T)$ coincides with the the completion of $G(S)$ in $H_1 \times H_2$ is called the L^2*-closure* (hereafter, closure, for short) of S and denoted by $\overline{S}$.

It is well known in functional analysis that

PROPOSITION 1. *Every closable operator S has its closure $\overline{S}$.*

We may state the main result of this section.

THEOREM 2. *Let M be a $C^{1,1}$-manifold without boundary. Then the following holds.*

(i) $W = H$.
(ii) *If* $\mathrm{Cap}(\partial M) = 0$, *then* $W_0 = W$.

PROOF. We start to prove (i). For $f \in W$, let $f_n \in C^1$ be a sequence such that $f_n \to f$ in W as $n \to \infty$. By Stokes theorem,

$$\langle f, \operatorname{div} X\rangle = \lim_{n\to\infty} \langle f_n, \operatorname{div} X\rangle = \lim_{n\to\infty} \langle \nabla f_n, X\rangle = \langle \overline{\nabla} f, X\rangle$$

for every $X \in V_0^1$. This shows $f \in H$.

Conversely, let f be in H. Let $\{U_\alpha, \psi_\alpha\}_{\alpha>0}$ be a local chart, where each U_α is relative compact, and $\{\rho_\alpha\}_\alpha$ be an associated partition of unity such that $\rho_\alpha \in C_0^{1,1}(U_\alpha)$. We claim $f_\alpha := \rho_\alpha f \in H$. Let X be the weak derivative of f. Then it holds

$$\langle \rho_\alpha X + f\nabla\rho_\alpha, Y\rangle = \langle f, -\operatorname{div}(\rho_\alpha Y)\rangle + \langle f\nabla\rho_\alpha, Y\rangle = \langle f_\alpha, \operatorname{div} Y\rangle$$

for every $Y \in V_0^1$. Hence $-(\rho_\alpha X + f\nabla\rho_\alpha) \in L^2$ is the weak derivative of f_α. So $f_\alpha \in H$. Denote by J_ϵ the Friedrich mollifier. Define $J_\epsilon f_\alpha \in C^{1,1}$ by

$$J_\epsilon f_\alpha := \int J_\epsilon(\cdot, y) f_\alpha(y)\, \mu(dy).$$

Since $J_\epsilon(x, \cdot)$ has support in $B(x, \epsilon)$, for every $\alpha > 0$, there exists $\epsilon_\alpha > 0$ such that $J_\epsilon f_\alpha \in C_0^{1,1}(U_\alpha)$ for every $0 < \epsilon < \epsilon_\alpha$. Due to compactness argument,

$$\|J_\epsilon f_\alpha - f_\alpha\|_{1,2} \to 0$$

as $\epsilon \to 0$. For $\alpha > 0$ and $\delta > 0$, let $\epsilon_\alpha > 0$ be such that

$$\|J_{\epsilon_\alpha} f_\alpha - f_\alpha\|_{1,2} < 2^{-\alpha}\delta.$$

Then $f_\delta = \sum_\alpha J_{\epsilon_\alpha} f_\alpha \in C^{1,1}$ satisfies

$$\|f_\delta - f\|_{1,2} < \delta.$$

This shows $f \in W$. Now we have completed the proof of (i).

Next, we prove (ii). We would like to show that for every $f \in W$ there exists $f_n \in W_0$ such that $f_n \to f$ in W as $n \to \infty$. First, we claim that we may assume f to be bounded. Define

$$f \vee g := \max\{f, g\},$$
$$f \wedge g := \min\{f, g\}.$$

Then $f_n := (f \vee (-l)) \wedge l \to f$ in W as $l \to \infty$ [11]. Hence, hereafter we assume f is bounded.

Next we claim that we may assume that f is 0 on some neighbourhood of ∂M. Since ∂M is almost polar, there exists a sequence $e_n \in W$ such that

(a) $0 \leq e_n \leq 1$,
(b) there exists an open set $\partial M \subset O_n \subset \overline{M}$ such that $e_n|_{O_n}=1$,
(c) $\|e_n\|_{1,2} \to$ as $n \to \infty$.

From condition (c), we may assume that e_n tends to 0 almost everywhere as $n \to \infty$. Set $f_n = (1 - e_n)f$. Because of condition (a), $f_n \in W$. Then

$$\|f - f_n\|_{1,2} \leq \|e_n f\|_2 + \|e_n \nabla f\|_2 + \|f \nabla e_n\|_2. \tag{3}$$

The first and second terms of R.H.S. of (3) tends to 0 as $n \to \infty$ by Lebesgue theorem. The third term of R.H.S. of (3) tends to 0 as $n \to \infty$ because f is bounded. Due to (b), hereafter, we assume f is 0 on some neighbourhood of ∂M.

Finally, we are going to cut off f outside of a big ball. Define a function η_n by

$$\eta_n(r) = ((2 - n^{-1}r) \vee 0) \wedge 1, \tag{4}$$

where r is the radius function from an arbitrary but fixed point $x \in M$. Put $B(r) := B(x, r)$. We note that $B(r)$ has finite volume for every $r > 0$. Indeed, since ∂M is almost polar, there exists an open set $\partial M \subset O$ with finite volume, and as $\overline{B(r)} \setminus O \subset M$ is compact because of Assumption 1, it has finite volume. Due to the definition of the intrinsic distance, $\|\nabla r\|_\infty \leq 1$, and thus $\eta_n(r) \in W$ for every $n > 0$. As η_n is bounded,

$$f_n := f\eta_n \in W \text{ for every } n > 0.$$

Since $\|\nabla \eta_n\|_\infty \leq 1/n$, we have

$$\|f - f_n\|_{1,2} \leq \|(1 - \eta_n)f\|_2 + n^{-1}\|f\|_2 + \|(1 - \eta_n)\nabla f\|_2. \tag{5}$$

By Lebesgue theorem, R.H.S. of (5) tends to 0 as $n \to \infty$. Now we may assume f has compact support. By the mollifier techniques as in the proof of (i) above, we obtain a sequence $f_n \in C_0^{1,1}$ such that $f_n \to f$ in W. This shows $W_0 = W$, and we have completed the proof. □

3 – Essential self-adjointness

A symmetric operator is called essentially self-adjoint if it has a unique self-adjoint extension. In a series of papers, M. P. Gaffney studied this problem for the Laplacian of a manifold. In [6] he established a criterion called *M has negligible boundary* (see Definition 5) so that the Laplacian is essentially self-adjoint. Subsequently, in [5] he showed that a complete manifold has negligible boundary. In our previous paper [13] we had showed that if $W = W_0$, then M has negligible boundary. In this section, we will prove that the converse is also true, namely

THEOREM 3. *The Laplacian is essentially self-adjoint if and only if* $W = W_0$.

First, we will present alternative two different proofs of Gaffney theorem (Theorem 4). Then Theorem 3 will follow immediately.

In order to make Δ symmetric, we need

DEFINITION 5. We say *M has negligible boundary* if

$$\int \operatorname{div}(fX) = 0$$

for every $f \in D(\nabla)$ and $X \in D(\operatorname{div})$.

The following Gaffney theorem says the assumption such that M has negligible boundary makes Δ not only symmetric but also essentially self-adjoint.

THEOREM 4. *Δ is essentially self-adjoint if and only if M has negligible boundary.*

Before starting the proof, let us present a corollary which is the (i) of Theorem 1.

COROLLARY 2. *If M has almost polar boundary, then Δ is essentially self-adjoint.*

PROOF. Let $f \in D(\nabla)$ and $X \in D(\operatorname{div})$. Due to Theorem 2, there exists $f_n \in C_0^{1,1}$ such that $f_n \to f$ in W. Then

$$\langle \nabla f, X \rangle = \lim_{n\to\infty} \langle \nabla f_n, X \rangle = -\langle f, \operatorname{div} X \rangle.$$

Hence M has negligible boundary, and Δ is essentially self-adjoint by Theorem 4. ∎

The first proof is to combine Lemma 1 and 2. We prove that the closure $\overline{\Delta}$ is self-adjoint, because if T is a self-adjoint extension of Δ then

$$\Delta \subset \overline{\Delta} \subset T = T^* \subset \overline{\Delta}^* = \Delta^*,$$

which shows $\overline{\Delta} = \overline{\Delta}^*$ implies $\overline{\Delta} = T$, the essential self-adjointness of Δ.

LEMMA 1. *If M has negligible boundary, then $\overline{\operatorname{div}} \cdot \overline{\nabla}$ is self-adjoint.*

PROOF. We would like to show $-\overline{\operatorname{div}} = \nabla^*$, where ∇^* stands for the adjoint-operator of ∇, because then by Von Neumann theorem, $\overline{\operatorname{div}} \cdot \overline{\nabla} = -\nabla^* \cdot \overline{\nabla}$ is self-adjoint. One direction is obvious, because by the definition of negligible boundary, $-\operatorname{div} \subset \nabla^*$ and since ∇^* is closed, $-\overline{\operatorname{div}} \subset \nabla^*$.

Conversely, let $X \in D(\nabla^*)$. Because $D(\nabla_C^*) = H = W$ where ∇_C is the restriction of ∇ to the set of functions with compact support, there exists a sequence $X_n \in D(\operatorname{div})$ such that $X_n \to X$ in W as $n \to \infty$. This means $X \in D(\overline{\operatorname{div}})$. ∎

LEMMA 2. *If M has negligible boundary, then $\overline{\Delta} = \overline{\operatorname{div}} \cdot \overline{\nabla}$.*

PROOF. As Lemma 1 says $\overline{\Delta} \subset \overline{\operatorname{div}} \cdot \overline{\nabla}$, we only prove the converse. Let $\{U_\alpha, \psi_\alpha\}_{\alpha>0}$ be a local chart, where each U_α is relative compact, and $\{\rho_\alpha\}_\alpha$ be an associated partition of unity such that $\rho_\alpha \in C_0^{1,1}(U_\alpha)$. Suppose $f \in D(\overline{\operatorname{div}} \cdot \overline{\nabla})$. Then $f_\alpha := \rho_\alpha f$ belongs to $D(\overline{\operatorname{div}} \cdot \overline{\nabla})$. By the definition of a closed operator, for every $\alpha > 0$ and $\epsilon > 0$, there exists a vector field $X_{\alpha,\epsilon} \in V_0^{1,1}(U_\alpha)$ such that

$$\|\overline{\nabla} f_\alpha - X_{\alpha,\epsilon}\|_{1,2} < \epsilon. \tag{6}$$

Due to Kodaira-Morrey-Eells decomposition, there exist $f_{\alpha,\epsilon} \in D(\nabla)$ and $Y_{\alpha,\epsilon} \in \operatorname{div}^{-1}(0)$ such that

$$X_{\alpha,\epsilon} = \nabla f_{\alpha,\epsilon} + Y_{\alpha,\epsilon}.$$

Since the $\operatorname{div}^{-1}(0)$ component of $\nabla f_{\alpha,\epsilon}$ is 0, we have

$$\|\overline{\nabla} f_\alpha - \nabla f_{\alpha,\epsilon}\|_2 \le \|\overline{\nabla} f_\alpha - X_{\alpha,\epsilon}\|_2 \tag{7}$$

and

$$\|\overline{\operatorname{div}} \cdot \overline{\nabla} f_\alpha - \operatorname{div} \cdot \nabla f_{\alpha,\epsilon}\|_2 = \|\overline{\operatorname{div}} \cdot \overline{\nabla} f_\alpha - \operatorname{div} X_{\alpha,\epsilon}\|_2. \tag{8}$$

By (6), (7), (8) and Poincaré inequality, for every $\alpha > 0$ and $\epsilon > 0$, there exists $h_{\alpha,\epsilon} \in D(\Delta)$ such that

$$\|f_\alpha - h_{\alpha,\epsilon}\|_2 + \|\overline{\operatorname{div}} \cdot \overline{\nabla}(f_\alpha - h_{\alpha,\epsilon})\|_2 < 2^{-\alpha}\epsilon.$$

Define a function $f_\epsilon \in D(\Delta)$ by

$$f_\epsilon = \sum_\alpha f_{\alpha,\epsilon}.$$

Then we have

$$\|f - f_\epsilon\|_2 + \|\overline{\operatorname{div}} \cdot \overline{\nabla} f - \Delta f_\epsilon\|_2 < \epsilon.$$

Since $\epsilon > 0$ is arbitrary, now we have completed the proof. ∎

We proceed to the second proof. Let us recall two self-adjoint Laplacians. Both of them are an extension of the Laplacian on $C_0^{1,1}$. The Dirichlet Laplacian Δ_D is the self-adjoint operator defined on the set of functions $f \in W_0$ such that $\Delta f \in L^2$. The Neumann Laplacian Δ_N is the self-adjoint operator defined on the set of functions $f \in W$ such that $\langle \Delta f, \psi \rangle = -\langle \nabla f, \nabla \psi \rangle$ for every $\psi \in W$ [18]. Denote by $\Delta_{D,0}$ (resp. $\Delta_{N,0}$) the Laplacian defined on $C^{1,1} \cap D(\Delta_D)$ (resp. $C^{1,1} \cap D(\Delta_N)$).

LEMMA 3. *$\Delta_{D,0}$ (resp. $\Delta_{D,0}$) is essentially self-adjoint and its closure is Δ_D (resp. Δ_N).*

The proof is similar to that of main result of [15]. For the sake of completeness, we present the proof.

PROOF. We would like to show $\overline{\Delta_{D,0}} = \Delta_D$. Let $f \in D(\Delta_D)$. Then by hypo-ellipticity of Δ,

$$f_t = e^{t\Delta_D} f \in D(\Delta_{D,0}).$$

By definition of $e^{t\Delta_D}$,

$$f_t \to f \text{ in } L^2 \text{ as } t \to 0,$$

and

$$\Delta f_t = e^{t\Delta_D} \Delta_D f \to \Delta_D f \text{ in } L^2 \text{ as } t \to 0.$$

Hence $f_t \in D(\Delta_{D,0})$ is a Cauchy sequence with respect to the graph norm of $\Delta_{D,0}$, and thus, $f = \lim_{n\to\infty} f_t \in D(\overline{\Delta_{D,0}})$. Obviously, the same proof applies for $\Delta_{N,0}$. □

As $\Delta_{D,0} \subset \Delta$, if Δ is symmetric, then it is essentially self-adjoint. This completes the second proof of Theorem 4.

Now we prove Theorem 3.

PROOF. In the proof above, we have observed that if M has negligible boundary, then Δ is essentially self-adjoint and the closure $\overline{\Delta}$ is the Dirichlet Laplacian. Now we would like to see that the closure coincides also to the Neumann Laplacian, because if it is true, then the quadratic forms $\sqrt{-\Delta_D}$ and $\sqrt{-\Delta_N}$ coincide, where

$$D(\sqrt{-\Delta_D}) = W_0 \text{ and } D(\sqrt{-\Delta_N}) = W,$$

and accordingly, $W = W_0$ [18].

Suppose f is in $D(\Delta_{N,0})$. Then

$$\langle \Delta f, \psi \rangle = -\langle \nabla f, \nabla \psi \rangle \text{ for every } \psi \in W$$

shows $\nabla f \in D(\nabla^*)$. Hence $\nabla f \in D(\overline{\text{div}})$ by Lemma 1, and $f \in D(\overline{\Delta})$ by Lemma 2. Thus, $\Delta_D = \overline{\Delta} = \Delta_N$ and we have the proof. □

REMARK 3. Consider the Laplacian $\Delta = \partial\delta + \delta\partial$ on forms with following domain [6]. The domain $D(\partial)$ of ∂ is the set of C^1-forms α such that both α and $\partial\alpha$ are square integrable. Similarly, the domain $D(\delta)$ of δ is the set of C^1-forms α such that both α and $\delta\alpha$ are square integrable. Then let the domain $D(\Delta)$ of the Laplacian Δ be the set of C^1-forms $\alpha \in D(\partial) \cap D(\delta)$ such that $\partial\alpha \in D(\delta)$ and $\delta\alpha \in D(\partial)$. One may prove the essential self-adjointness of Δ by a similar method of the second proof presented above. In fact, assume Δ is symmetric. Then, for the Friedrich self-adjoint extension Δ_F of Δ,

$$\alpha_t = e^{-t\Delta_F}\alpha \in D(\Delta),$$

and both α_t and $\Delta\alpha_t$ converges to α and $\Delta\alpha$, respectively, as $t \to 0$.

4 – Conservative, parabolic, and Liouville property

In this section we prove conservativeness, parabolicity and a Liouville type property. Let us start from definitions. The *heat kernel* p associated to $\frac{1}{2}\Delta$ is the smallest positive fundamental solution to the heat equation

$$\frac{1}{2}\Delta u_t = \frac{\partial}{\partial t}u_t.$$

J. DODZIUK [3] showed that every Riemann manifold (whether it is complete or incomplete) admits the heat kernel. Let us recall

DEFINITION 6. A manifold M is called *conservative* if the heat kernel p satisfies

$$\int p(t,x,y)\,d\mu(y) = 1$$

for every $t > 0$ and $x \in M$.

Let $(\mathcal{E}, D(\mathcal{E}))$ be a Dirichlet form on L^2 and T be the generator. $(\mathcal{E}, D(\mathcal{E}))$ is called *conservative* if for every $f_n \in L^2$ such that $0 \le f_n \le 1$, $f_n \to 1$, it holds $e^{tT}f_n \to 1$ as $n \to \infty$ for every $t > 0$. For further study of a Dirichlet form, we refer [4]. Denote by Δ_D the Dirichlet Laplacian. Then, since

$$e^{t\Delta_D}f = \int p(t,\cdot,y)f(y)\,d\mu(y) \text{ for every } f \in L^2,$$

the conservativeness of M is equivalent to that of the Dirichlet form $(\mathcal{E}, W_0)$. As $(\mathcal{E}, W_0)$ generates the Brownian motion, M is conservative if and only if the Brownian motion X_t starting from an arbitrary point of M may be found on M almost surely at every time $t > 0$. The terminology *conservative* originates on this fact.

A manifold with boundary is never conservative in the sense above, because the Brownian motion will be absorbed at the boundary. So instead of the heat kernel, we consider the Neumann heat kernel. In such case, the corresponding Dirichlet form is $(\mathcal{E}, W)$ and the Brownian motion is reflected at the boundary. A manifold with boundary of Neumann condition is conservative if it has volume growth condition (1) of Theorem 1 [8].

We start the proof of conservativeness.

PROOF. Our argument bases on the following [4]. □

THEOREM 5. *The Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ is conservative if and only if there exists a sequence $f_n \in D(\mathcal{E})$ such that*

$$0 \le f_n \le 1, \lim_{n\to\infty} f_n = 1, \text{ and } \lim_{n\to\infty} \mathcal{E}(f_n, \psi) = 0 \text{ for every} \psi \in D(\mathcal{E}) \cap L^1.$$

Let $\partial M \subset O_n$ be a decreasing family of open sets of $\overline{M}$ such that $\partial O_n \cap M$ is $C^{1,1}$ for every $n > 1$, and $\mathrm{Cap}(O_n) \to 0$ as $n \to \infty$. The manifold with boundary $M \setminus O_n$ is conservative with Neumann condition [8] for every $n > 1$. Hence, by Theorem 5, for every $n > 0$ there exists a sequence $f_{n,l} \in W(M \setminus O_n)$ such that $0 \le f_{n,l} \le 1$ on M,

$$\lim_{l\to\infty} f_{n,l} = 1, \text{ and } \lim_{l\to\infty} \mathcal{E}(f_{n,l}, \psi) = 0 \text{ for every } \psi \in W(M \setminus O_n) \cap L^1.$$

Set $h_{n,l} = e_n \vee f_{n,l}$, where e_n is the equilibrium potential of $O_n \cap M$, that is the function $e_n \in W_0$ such that $0 \le e_n \le 1$ on M,

$$e_n|_{O_n \cap M} = 1, \text{ and } \mathrm{Cap}(O_n \cap M) = \|e_n\|_{1,2}.$$

Then $h_{n,l} \in W$, and

$$|\mathcal{E}(h_{n,l}, \psi)| \le |\langle \nabla e_n, \nabla \psi \rangle| + \Big| \int_{M \setminus O_n} \langle \nabla f_{n,l}, \nabla \psi \rangle \Big|. \tag{9}$$

The R.H.S. of (9) tends to 0 as $n, l \to \infty$. Since $h_{n,l} \in W_0$ for every $n, l > 1$, and $h_{n,l} \to 1$ as $n, l \to \infty$, M is conservative by Theorem 5. □

A manifold is said to be *parabolic* if it does not admit a non-negative Green function G. By definition,

$$G(x, y) = \int_0^\infty p(t, x, y)\, dt, \text{ for every } x, y \in M.$$

This shows M is parabolic if and only if the Brownian motion is recurrent. The concept of recurrence may be extended to general Dirichlet form, and it holds [4].

THEOREM 6. *The Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ is recurrent if and only if there exists a sequence $f_n \in D(\mathcal{E})$ such that*

$$0 \le f_n \le 1, \lim_{n\to\infty} f_n = 1, \text{ and } \lim_{n\to\infty} \mathcal{E}(f_n, f_n) = 0.$$

We proceed to prove the parabolicity.

PROOF. The volume growth condition (2) of Theorem 1 implies

$$\sum_{n>0} \frac{2^{n+1}}{v(2^{n+1}) - v(2^n)} = \infty,$$

where $v(r)$ stands for the volume of the ball $B(x,r)$ with arbitrary but fixed $x \in M$. Hence,

$$\lim_{n\to\infty} \frac{v(2^{n+1}) - v(2^n)}{2^{2n}} = 0. \tag{10}$$

Let $\eta_n \in W$ be the function defined by (4). Then

$$\mathcal{E}(\eta_n, \eta_n) = \frac{v(2^{n+1}) - v(2^n)}{2^{2n}}.$$

By (10) and Theorem 6, the Brownian motion is recurrent. □

REMARK 4. The parabolicity may be proved by the same idea of the proof of conservativeness presented above. More precisely, decompose $M = M_1 \cup M_2$, where $\partial M \subset M_1$ and M_1 has finite volume. Imposing Neumann condition to both M_1 and M_2, by condition (2) of Theorem 1, both manifolds are parabolic. As M does not have boundary condition, it is parabolic.

Finally, we prove a Liouville property.

PROOF. Let $\eta_n \in W$ be the function defined by (4). Let f be a non-negative sub-harmonic function. Assume $f \in D(\overline{\Delta})$. Since $\overline{\Delta} = \Delta_D$ ((ii) of Theorem 1),

$$0 \le \langle \Delta f, \eta_n^2 f \rangle = -2\langle \nabla f, \eta_n f \nabla \eta_n \rangle - \langle \nabla f, \eta_n^2 \nabla f \rangle,$$

and hence

$$\|\eta_n \nabla f\|_2 \le 2\|f \nabla \eta_n\|_2. \tag{11}$$

The R.H.S. of (11) tents to 0 as $n \to \infty$ by Lebesgue theorem. Therefore $\nabla f = 0$.

If one puts $h = f^{p/2}$, then h is sub-harmonic and $f \in L^2$. Essentially the same proof described above applies to show $\nabla h = 0$, so we omit it. □

REMARK 5. It is known that there exists non-constant non-negative sub-harmonic function such that $f \notin L^p$ for any $p > 1$ on a complete manifold. The next two examples show that we may not remove the additional assumption such that f is bounded and ∂M is almost polar.

Consider $M = \mathbb{R}^3 \setminus \{0\}$ and $f = (r^{-2} - 1) \vee 0$, where r is the radius from the origin. Obviously, M has almost polar, f is sub-harmonic, and $f \in L^{5/4}$. Then,

a standard smoothing technique of sub-harmonic functions on Euclidean space yields a non-negative, non-bounded, non-constant C^2-sub-harmonic function in $L^{5/4}$.

Consider $M = \mathbb{R} \setminus \{0\}$ and $f = ((r+1)^{-1} - 4) \vee 0$. Then $\{0\}$ is not almost polar, f is bounded sub-harmonic and belongs to L^p with every $p > 1$.

REMARK 6. A similar estimate in the proof of parabolicity may be found in [10]. The conservativeness and parabolicity can be proved by the main result in [10]. The original proof of Liouville property of complete manifolds can be found in [22].

5 – Examples

In this section, we consider some examples. Let us recall a sufficient condition of ∂M to be almost polar [13].

DEFINITION 7. The lower Minkowski co-dimension of a Cauchy boundary ∂M is

$$\liminf_{\epsilon \to 0} \frac{\log \mathrm{vol}(N_\epsilon)}{\log \epsilon}$$

where N_ϵ is the ϵ-tubular neighbourhood of ∂M.

THEOREM 7. *If the lower Minkowski co-dimension of ∂M is greater than 2, then it is almost polar. In particular, if ∂M is a manifold, and its lower Minkowski co-dimension is not less than 2, then the same conclusion holds.*

EXAMPLE 3. Consider the incomplete manifold $M \setminus \Sigma$, where $M \subset \mathbb{CP}^n$ is an algebraic variety in complex projective space with singular set of co-dimension 2. P. LI and G. TIAN [11] showed the essential self-adjointness of the Laplacian, conservativeness, and established an estimate of eigenvalues. As they proved that Σ is almost polar (Theorem 4.1), and M has finite volume, their manifold is not only conservative but more strongly, also parabolic. For a detailed proof, we refer [24].

EXAMPLE 4. Let M be a compact orbifold with singular locus Σ of codimension ≥ 2. In the same way as in [11], one may show that the Riemann manifold $M \setminus \Sigma$ has almost polar Cauchy boundary. The spectrum of the L^2-closure of the Laplacian is studied in [19].

EXAMPLE 5. Consider an incomplete 2-dimensional manifold of finite volume having constant curvature = 1 with isolated conical singularities. Such a manifold is called Met_1-surface and is important for the study of minimal surfaces [12].

EXAMPLE 6. Consider the so called "football". Set $M := \mathbb{C} \setminus \{0\}$

$$g = \frac{4(dx^2 + dy^2)}{(1+r^2)^2},\ f = \frac{\mu r^{\mu-1}(1+r^2)}{1+r^{2\mu}},\ \mu \in \mathbb{R}_+,\ h = f^2 g,$$

where r is the distance from the origin. Then $\{0\}$ is almost polar. In [14], we show that the Laplacian has pure point spectrum and it satisfies Weyl's asymptotic formula.

The next example has fractal Cauchy boundary.

EXAMPLE 7. Let us recall the Cantor set Σ in a real-line. Consider the union of 2^n-segments

$$\Sigma_n = [0, 3^{-n}] \cup [2 \cdot 3^{-n}, 3^{-n+1}] \cup \ldots \cup [1 - 3^{-n}, 1] \subset \mathbb{R}.$$

Then Σ is defined by $\Sigma = \lim_{n\to\infty} \cap_{1<l<n} \Sigma_l$. Let $(M, g) = (N \setminus \Sigma, f g_o)$ be an incomplete manifold defined as follows; N is a 3-dimensional complete C^2-manifold with metric $g_o \in C^2$, Σ is a Cantor set, by this we mean there exists a local chart (U, ψ) of N such that $C \subset U$ and $\psi(\Sigma)$ is a Cantor set in $\mathbb{R}$. The metric $g = f g_o$ is defined as follows. Denote by r the distance from Σ with respect to g_o, and set

$$f(x) = \begin{cases} r^{2\epsilon}, & \text{if } x \in B; \\ 1, & \text{otherwise,} \end{cases}$$

where $\epsilon > \frac{\log 2 - \log 3}{2\log 3 - \log 2}$ and $B = \{x \in N |\, r(x) < 1\}$.

PROPOSITION 2. *M is a $C^{1,1}$-manifold with almost polar boundary.*

PROOF. Since M is of class C^2 and g is Lipschitz on every compact set, M is of class $C^{1,1}$. Let us assume $\psi(\Sigma)$ lies in x-axis. We claim $\partial M = \Sigma$. Indeed, for every $a = (x, 0, 0) \in \Sigma$ (hereafter we identify Σ with $\psi(\Sigma)$) and $a_n = (x, 0, 1/n) \in M$, there exists $C > 0$ such that

$$d(a - a_n) < C \int_0^{\frac{1}{n}} z^\epsilon \, dt \to 0 \text{ as } n \to \infty,$$

where d is the Riemann distance with respect to g. As N is complete, $\partial M = \Sigma$. Denote by V_n the volume of $B(\Sigma_n, 3^{-1})$ with respect to g. By an explicit computation of V_n and letting $n \to \infty$, Minkowski dimension of Σ is

$$\dim(\Sigma) = 3 + \frac{\log V_n}{n \log 3} \leq 1 + \frac{\log 2}{\log 3} - \frac{2\epsilon + 1}{\epsilon + 1}.$$

Hence, co-dimension of $\Sigma \subset M$ is greater than 2 if $\epsilon > \frac{\log 2 - \log 3}{2\log 3 - \log 2}$. By Theorem 7, ∂M is almost polar. ∎

Acknowledgements

The author wishes to show his gratitude towards Professors W. Rossman and H. Urakawa for many important suggestions and for providing him with interesting examples of incomplete manifolds that are studied from a different point of view in their joint work [14]. He wishes to thank to the referee for valuable comments, which improved the original manuscript in the present form, and inspired the author to study Theorem 3. Finally, he appreciates it to Professors M. Bordoni, S. Marchiafava, U. Mosco, and A. Savo for their warm hospitality and useful discussion during the author stayed at Rome University.

REFERENCES

[1] M. Biroli – U. Mosco: *A Saint-Venant type principle for Dirichlet forms on discontinuous media*, Annali di Matematica pura ed applicatata, **IV** (1995), 125-181.

[2] Y. Colin de Verdiére: *Pseudo-Laplaciens.* 1, Ann. Inst. Fourier., **32** (1982), 275-286.

[3] J. Dodziuk: *Maximum principle for parabolic inequalities and the heat flow on open manifolds*, Indiana Univ. Math. J., **32** (1983), 703-716.

[4] M. Fukushima – Y. Oshima – M. Takeda: *Dirichlet forms and Markov processes*, North Hooland, Kodansha, 1980.

[5] M. P. Gaffney: *Special Stokes' theorem for complete Riemannian manifolds*, Ann. of Math., **60** (1954), 140-145.

[6] M. P. Gaffney: *Hilbert space methods in the theory of harmonic integral*, Trans. Amer. Math. Soc., **78** (1955), 426-444.

[7] G. D. Maso – R. Gulliver – U. Mosco: *Asymptotic spectrum of manifolds of increasing topological type*, preprint.

[8] A. Grigor'yan: *Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds*, Bull. Amer. Math. Soc., **36** No. 2 (1999), 135-249.

[9] S. T. Kuroda: *An introduction to scatting theorey*, Lecture notes, series No. 51, Aarhus Univ, 1978.

[10] K. Kuwae: *Recurrence and conservativeness of symmetric diffusion processes by Girsanov transformation*, J. Math. Kyoto Univ., **37** No. 4 (1997), 735-756.

[11] P. Li – G. Tian: *On the heat kernel of Bergman metric on algebraic variety*, J. Amer. Math. Soc., **8** (1995), 857-877.

[12] L. L. D. Lima – W. Rossman: *On the index of constant mean curvature 1 surfaces in hyperbolic space*, Indiana Univ. Math. J., **47** (1998), 685-723.

[13] J. Masamune: *Essential self-adjointness of Laplacians on Riemannian manifolds with fractal boundary*, Comm. Partial Differential Equations, **24** No. 3-4 (1999), 749-757.

[14] J. Masamune – W. Rossman: *Discrete spectrum and Weyl's asymptotic formula for incomplete manifolds. Minimal surfaces, geometric analysis and symplectic geometry 2000*, Adv. Stud. Pure Math., **29** (2000), 1-11.

[15] J. Masamune: *Essential self-adjointness of a sub-Laplacian via heat equation*, preprint.

[16] S. Ozawa: *The first eigenvalue of the Laplacian on two dimensional Riemannian manifolds*, Tohoku Math. Journ., **34** (1982), 7-14.

[17] G. C. Papanicolaou – S. R. S. Varadhan: *Diffusion in region with many small holes*, Lecture notes in control and information science, Springer, **25** (1980), 190-206.

[18] M. Reed – B. Simon: *Methods of modern mathematical physics*, Academic Press, New york and London, 1972.

[19] T. Shioya: *Eigenvalues and suspension structure of compact Riemannian orbifolds with positive Ricci curvature*, Manuscripta Math., **99** (1999), 509-516.

[20] R. S. Strichartz: *Analysis of Laplacian on the complete Riemannian manifold*, Journal of functional analysis, **52** (1983), 48-79.

[21] K-Th. Sturm: *Analysis on local Dirichlet spaces I. Recurrence, conservativeness and -Liouville properties*, J. Reine. Angew. Math., **45** (1994), 173-196.

[22] S. T. Yau: *Some function-theoretic properties of complete Riemannian manifolds and their applications to geometry*, Indiana Math J., **25** (1976), 659-670.

[23] S. T. Yau: *On the heat kernel of a complete Riemannian manifold*, J. Math. Pure Appl., **57** ser. 9 (1978), 191-201.

[24] K. Yoshikawa: *Degeneration of algebraic manifolds and spectrum of Laplacians*, Nagoya Math. J., **146** (1997), 83-129.

Lavoro pervenuto alla redazione il 28 giugno 2002
ed accettato per la pubblicazione il 15 novembre 2004.
Bozze licenziate il 14 febbraio 2005

Indirizzo dell'Autore:

Jun Nasmaune – Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate – Via A. Scarpa, 16 – 00161 Roma, Italy
E-mail: masamune@dmmm.uniroma1.it

This work was partially supported by the Italian education, university and research ministry (project Cofin 2000 "Proprietà geometriche delle varietà reali e complesse"), I.N.D.A.M., and by J.S.P.S. (Japan society for the promotion of science).