Ann Inst Stat Math (2018) 70:1–38
https://doi.org/10.1007/s10463-017-0628-7

INVITED REVIEW ARTICLE

Fold-up derivatives of set-valued functions and the change-set problem: A Survey

Estate Khmaladze[1] · **Wolfgang Weil**[2]

Received: 9 December 2016 / Revised: 14 July 2017 / Published online: 28 November 2017
© The Institute of Statistical Mathematics, Tokyo 2017

Abstract We give a survey on fold-up derivatives, a notion which was introduced by Khmaladze (J Math Anal Appl 334:1055–1072, 2007) and extended by Khmaladze and Weil (J Math Anal Appl 413:291–310, 2014) to describe infinitesimal changes in a set-valued function. We summarize the geometric background and discuss in detail applications in statistics, in particular to the change-set problem of spatial statistics, and show how the notion of fold-up derivatives leads to the theory of testing statistical hypotheses about the change-set. We formulate Poisson limit theorems for the log-likelihood ratio in two versions of this problem and present also the route to a central limit theorem.

Keywords Infinitesimal image analysis · Generalized functions · Fold-up derivatives · Local Steiner formula · Local point process · Set-valued mapping · Derivative set · Normal cylinder · Change-set problem

1 Introduction

The differentiation of set-valued functions encompasses a topic, which has seen several approaches by Artstein (1995, 2000), Aubin (1981), Aubin and Frankowska (1990),

✉ Estate Khmaladze
Estate.Khmaladze@vuw.ac.nz

Wolfgang Weil
wolfgang.weil@kit.edu

1 School of Mathematics and Statistics, Victoria University of Wellington, P.O.Box 600, Wellington, New Zealand

2 Institute of Stochastics, Department of Mathematics, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany

Bernardin (2003) and Borwein and Zhu (1999) and which has diverse and important applications, for example in the theory of optimal control and convex analysis, to name only two. As an illustration of probabilistic research connected with set-valued analysis, we refer to Kim and Kim (1999). Derivatives of set-valued functions also arise in statistics, in particular in connection with the change-set problem. For such applications, it turned out that a new approach was necessary, the fold-up derivatives, by which the infinitesimal changes in a set-valued function are described by a set in the normal cylinder of the limit set. This concept was introduced in Khmaladze (2007) for convex sets and extended to rather general closed sets in Khmaladze and Weil (2014). In the following, we give a survey on fold-up derivatives, describe the geometrical background and discuss its applications in statistics.

The text is organized as follows: In the remaining part of this Introduction, we discuss the nature of fold-up derivatives (Sects. 1.1 and 1.5) and describe the problem of spatial statistics, which stimulated the development of this notion. In Sect. 2, we present the local Steiner formula and its consequences, first for convex bodies and then for general solid sets (sets equal to the closure of their interior). In Sect. 3, we outline the theory of fold-up derivatives along with some examples. Then, applications to the change-set problem are collected in Sect. 4. Most of results in this section, although based on approaches developed earlier, are formulated here for the first time, unless it is explicitly stated otherwise. The final Sect. 5 compares the notion of fold-up derivatives with the notion of generalized functions and the so-called chimeric alternatives (Sects. 5.1 and 5.2), along with two further subsections on extensions of interest for future research.

1.1 Infinitesimal changes of sets and images

To motivate the notion of fold-up derivatives, we start with a possible application in the analysis of infinitesimal changes in images.

In image analysis, an image is often represented by a vector-valued function $\{f(x), x \in D\}$ on a rectangular array D—for example pixels x on a computer screen. The vector $f(x)$ describes certain properties of the image like the color and the intensity of this color in the pixel x. For simplicity, let us assume that f is one dimensional, given by the intensity of only one color, say the color "black." In order to apply analytic methods, it is advantageous to neglect the discrete structure of D and also the restriction to the two-dimensional setting and consider D as a subset of $\mathbb{R}^d$ and call a (real-valued) function $f = \{f(x), x \in D\}$ an image on D.

Consider now images f_t which change in time t in a continuous way. At the moment t_0, we have an image f_{t_0} and at time $t = t_0 + \varepsilon$ we have a small perturbation of f_{t_0}, if ε is small. To analyze this small change, as $\varepsilon \to 0$, we may end up with derivatives $\{\mathrm{d} f_t(x)/\mathrm{d}t, x \in D\}$ at $t = t_0$, as a function of x. This may be a natural approach to study continuous changes in images. As a *vector field* of velocities, this family of derivatives plays the key role, for example, in fluid mechanics, see, e.g., Landau and Lifshitz (1987). In the statistical "change-set problem," we are dealing with a different sort of changes. Namely, consider a set $F(t_0) \subset D$ and another set $F(t) \subset D$, which is a small deformation of $F(t_0)$, if $t - t_0$ is small. We will later assume that the sets

$F(t)$, $F(t_0)$ are compact and that $F(t) \to F(t_0)$ in the Hausdorff metric. Then let $f_{t_0}(x) = \mathbf{1}(x \in F(t_0))$ be the indicator function of $F(t_0)$ and let $f_t(x) = \mathbf{1}(x \in F(t))$ be the indicator function of $F(t)$. In this situation, there will be either no derivative $\mathrm{d}f_t(x)/\mathrm{d}t$ at x or the derivative will be trivial and equal to 0. How can one still consider the transition from $F(t)$ to $F(t_0)$ as smooth and differentiable? We will explain this in more detail in Sect. 1.4 below.

Let us slightly shift our attention. Instead of the indicator function $\mathbf{1}(x \in F)$, we now consider the set F itself as the object of interest; we could call it an image, but in the context of the statistical change-set problem we shall discuss, it is called a *change-set* (Khmaladze et al. 2006b).

In the change-set problem, we do not assume that we observe or know our change-set. All we have are random observations, the distribution of which depends in some particular way on the underlying set F. Then we would want to formulate hypotheses about the unknown F and try to test these hypotheses based on the observations we have. Moreover, we want to obtain statistical tests to discriminate between the null hypothetical set $F = F(t_0)$ and its small perturbation $F(t)$; the more observations we have, the smaller deviations we might be able to detect. Thus, the variable t in $F(t)$ is now a way to describe a family of possible alternative change-sets, which are approaching the set F, chosen as the primary candidate as a true change-set.

There is a vast literature on estimation of the change-set F based on random observations, to some of which we refer here: Carlstein and Krishnamoorthy (1992), Ripley and Rasson (1977), Khmaladze et al. (2006b), Müller and Song (1996), Ivanoff and Merzbach (2010) and Korostelev and Tsybakov (1993), which have more references. Among more recent ones, which study estimation of sets within nonparametric classes or functionals of sets, we refer to Baíllo and Cuevas (2001) and Cuevas et al. (2007).

However, results about the testing problems concerning change-sets are rather scarce. This unbalance can be explained by difficulties in the analysis of a neighborhood of a set and, in particular, by the lack of an appropriate notion of derivative of a set-valued function.

1.2 Testing statistical hypotheses: local tests, parametric families of distributions

Let us briefly recall how do we test a hypothesis within a parametric family of distributions, depending on some k-dimensional parameter θ. Suppose $\{P_\theta, \theta \in \Theta\}$ is such a family, where $\Theta \subset \mathbb{R}^k$ and each P_θ is a distribution in $\mathbb{R}^d$. Assume that θ_0 is an interior point of Θ and, given an i.i.d. sequence of d-dimensional random variables $\{X_i\}_{i=1}^n$, we take P_{θ_0} as a null hypothesis about the distribution of each X_i. As the alternative hypothesis to P_{θ_0}, we consider P_{θ_ε} and assume that $\theta_\varepsilon \to \theta_0$, as $\varepsilon \to 0$.

The log-likelihood ratio in this situation has the form

$$L_n(\theta_0, \theta_\varepsilon) = \sum_{i=1}^n \ln \frac{\mathrm{d}P_{\theta_\varepsilon}}{\mathrm{d}P_{\theta_0}}(X_i).$$

Assume that the parametric family is regular at θ_0, in the sense that the Taylor expansion is valid up to the second term,

$$L_n(\theta_0, \theta_\varepsilon) = (\theta_\varepsilon - \theta_0)^\top \sum_{i=1}^n l_{\theta_0,\theta_\varepsilon}(X_i) - \frac{1}{2}\sum_{i=1}^n [(\theta_\varepsilon - \theta_0)^\top l_{\theta_0,\theta_\varepsilon}(X_i)]^2 + o_P(\|\theta_\varepsilon - \theta_0\|^2), \quad (1)$$

where the k-dimensional vector l, the score function, is defined as

$$l_{\theta_0,\theta_\varepsilon}(X_i) = \frac{\mathrm{d}}{\mathrm{d}\theta_\varepsilon} \ln \frac{\mathrm{d}P_{\theta_\varepsilon}}{\mathrm{d}P_{\theta_0}}(X_i)\bigg|_{\theta_\varepsilon=\theta_0},$$

and is, we assume, square integrable with respect to P_{θ_0}. Expansions of this form (or of a more sophisticated form) can be found in the statistical literature through decades, say, from the textbook of Cramér (1999), first published in 1946, to the modern textbooks, such as van der Vaart (1998). Let us add to this setting another assumption, that θ_ε is differentiable in ε at θ_0,

$$\theta_\varepsilon - \theta_0 = \varepsilon\gamma + o(\varepsilon), \quad \varepsilon \to 0,$$

where the derivative γ is a fixed vector in $\mathbb{R}^k$. Then, the choice of $\varepsilon = 1/\sqrt{n}$ becomes clear and $L_n(\theta_0, \theta_\varepsilon)$ will attain a form convenient for asymptotic analysis,

$$L_n(\theta_0, \theta_\varepsilon) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma^\top l_{\theta_0,\theta_\varepsilon}(X_i) - \frac{1}{2n}\sum_{i=1}^n \left[\gamma^\top l_{\theta_0,\theta_\varepsilon}(X_i)\right]^2 + o_P(1),$$

so that $L_n(\theta_0, \theta_\varepsilon)$ converges in distribution to a well-defined and "visible" limiting object.

A full analysis of situations like this, when the sample size increases, but alternative and hypothetical distributions approach each other at the same time, is the subject of contiguity theory (see, e.g., Le Cam 1986, Le Cam and Lo Yang 2000, Chapter 3, Hajek and Shidak 1967, Chapter 7.1, and Oosterhoff and van Zwet 2012).

1.3 Local tests for the change-set problem (first version)

In the cases when θ is a functional parameter, as in the semi-parametric situation, expansions like (1) are still useful and differentiability of θ_ε in ε is well understood (see, e.g., Kosorok 2008). However, consider the class of statistical problems, where θ is another infinite dimensional parameter—a set. This is the case in the change-set problem below. In this problem, we have a family of distributions P_F, indexed by sets F, and we consider $P_{F(0)}$ as a null distribution and $P_{F(\varepsilon)}$ as the alternative distribution. So, to obtain the form of the local test statistics we will need to differentiate $P_{F(\varepsilon)}$ with respect to $F(\varepsilon)$ and $F(\varepsilon)$ with respect to ε, and the question is, how to do this?

Let us consider a first version of the change-set problem. For a measurable set $D \subset \mathbb{R}^d$, denote again $F(0) = F$ and let N_n be a counting Poisson measure (or Poisson process) on D with intensity measure $n\Lambda_F$, where

$$\Lambda_F(A) = \tilde{\Lambda}(A \cap F) + \Lambda(A \cap F^c), \quad A \subset D.$$

Here, Λ and $\tilde{\Lambda}$ are two intensity measures on D with densities (intensities) $\lambda(x)$ and $\tilde{\lambda}(x)$, $x \in D$, with respect to the Lebesgue measure μ_d in $\mathbb{R}^d$, and F^c is the complement of F (see Fig. 1). Then it is not difficult to deduce (see Daley and Vere-Jones 2005 and Karr 1991), that the log-likelihood ratio of the distribution of N_n, under $F(\varepsilon)$ and under $F(0)$, respectively, has the form

$$\begin{aligned} L_n(F, F(\varepsilon)) &= \int [\mathbf{1}\{z \in F(\varepsilon)\} - \mathbf{1}\{z \in F\}] \ln \frac{\tilde{\lambda}}{\lambda}(z) N_n(\mathrm{d}z) \\ &\quad - n \int [\mathbf{1}\{z \in F(\varepsilon)\} - \mathbf{1}\{z \in F\}](\tilde{\lambda} - \lambda)(z)\mu_d(\mathrm{d}z) \\ &= \int [\mathbf{1}\{z \in F(\varepsilon)\backslash F\} - \mathbf{1}\{z \in F\backslash F(\varepsilon)\}] \ln \frac{\tilde{\lambda}}{\lambda}(z) N_n(\mathrm{d}z) \\ &\quad - n \int [\mathbf{1}\{z \in F(\varepsilon)\backslash F\} - \mathbf{1}\{z \in F\backslash F(\varepsilon)\}](\tilde{\lambda} - \lambda)(z)\mu_d(\mathrm{d}z). \end{aligned} \tag{2}$$

Suppose now that F, $F(\varepsilon)$ are compact and $F(\varepsilon) \to F$ in the Hausdorff metric. Then, as $\varepsilon \to 0$, both sets $F(\varepsilon)\backslash F$ and $F\backslash F(\varepsilon)$ shrink toward the boundary ∂F of F. What can we say about the possible limit of the integral expressions on the right side of (2), when $n \to \infty$ and $\varepsilon \to 0$ with an appropriate rate?

An immediate attempt for the second integral in (2), which is to multiply and divide by ε, leads to

$$n\varepsilon \int \frac{\mathbf{1}\{z \in F(\varepsilon)\} - \mathbf{1}\{z \in F\}}{\varepsilon}(\tilde{\lambda} - \lambda)(z)\mu_d(\mathrm{d}z),$$

which, for $n\varepsilon \to const$ and one further condition (see Sect. 5.1), converges to a generalized function concentrated on the boundary ∂F of F. This is a very natural object in itself and will not require, as it may seem, a differentiation of $F(\varepsilon)$ in ε *per se* (in this connection see Weisshaupt (2001)). However, we will show in Sect. 5 that such a generalized function is unsuitable to describe the limiting object. This fact will be better visible when one divides the first integral in (2), taken with respect to N_n, by ε.

1.4 The change-set problem (second version)

Let us consider another formulation of the change-set problem. It is graphically illustrated in Fig. 2. Suppose we have an i.i.d sequence of pairs $(X_i, Y_i)_{i=1}^n$, where $X_i \in D$ is a random location and Y_i is a corresponding mark (see Mammen and Tsybakov 1995 and Khmaladze et al. 2006a). This mark can be one dimensional or it can be

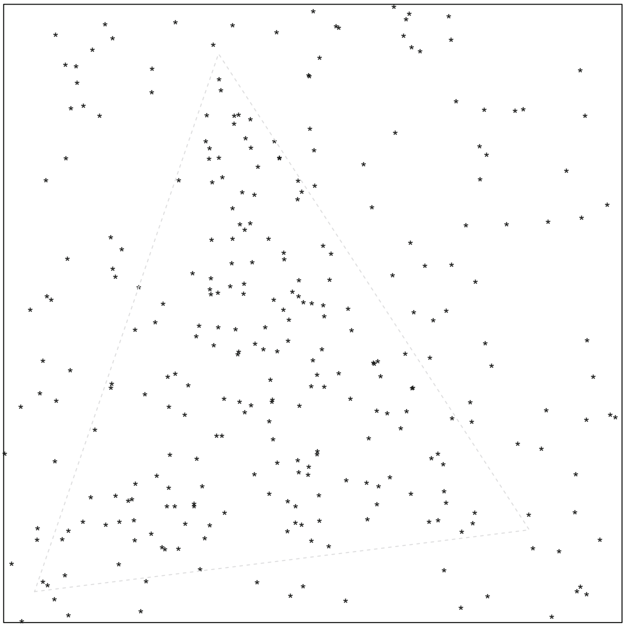

Fig. 1 The points shown here form a realization of a Poisson process with constant intensity inside F and three times higher than outside, where it is also constant. The set F is a faintly shown triangle, and the total number of points is 300. Without looking on F, an eye may have difficulties identifying this triangle as a change-set. The problem is to test whether the change-set is this triangle or another nearby shape

very high dimensional, listing, for example, the concentration of several minerals at different depths in a well at location X_i. It is enough for us to assume, however, that Y_i is a one-dimensional random variable. The defining property of the change-set F is that, for locations X_i in this set, the distribution of Y_i is some probability measure $\tilde{P}$, while for locations X_i outside F the mark Y_i has a different "gray-level" distribution P. The (marginal) distribution of X_i on D is some absolutely continuous Q, unrelated to the possible change-set F. As before, F is the parameter of interest in the problem. Then, the differential of the joint distribution of the pair (X_i, Y_i) is

$$\tilde{P}(\mathrm{d}y)^{\mathbf{1}\{x \in F\}} P(\mathrm{d}y)^{1-\mathbf{1}\{x \in F\}} Q(\mathrm{d}x),$$

and if we take a particular set $F = F(0)$ as a hypothetical change-set and another set $F(\varepsilon)$ as its alternative, then the log-likelihood ratio of the two corresponding distributions becomes

$$L_n(F, F(\varepsilon)) = \sum_{i=1}^{n} [\mathbf{1}\{X_i \in F(\varepsilon)\} - \mathbf{1}\{X_i \in F\}] \ln \frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(Y_i). \tag{3}$$

Here, we implicitly assumed that $\tilde{P}$ is absolutely continuous with respect to P, which looks like an additional regularity assumption but is of only little consequence for

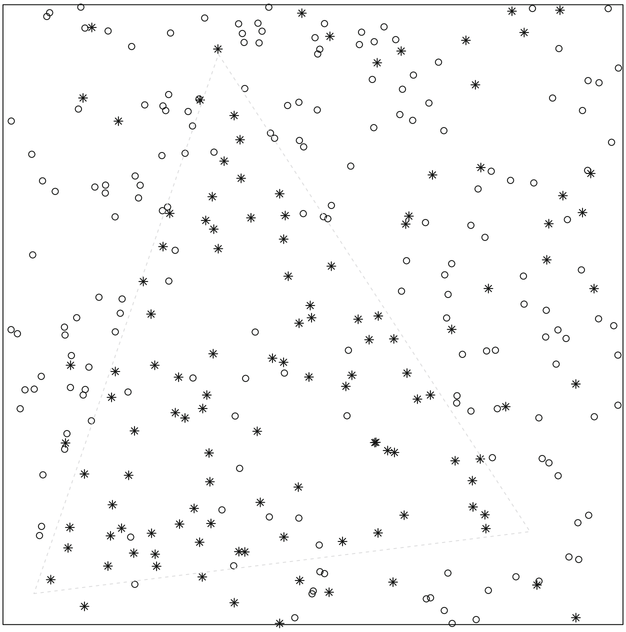

Fig. 2 About 300 points are scattered now uniformly, and points in the same triangle F as in Fig. 1, are shown as "stars" with probability 0.8; points outside the triangle are shown as "stars" with probability 0.3

us. Note that, if $\tilde{P}$ and P have mutually singular parts, the statistical problem of discrimination between $F(0)$ and $F(\varepsilon)$ will become only easier.

Let now N_n denote the binomial process generated by the pairs $(X_i, Y_i)_{i=1}^n$,

$$N_n(y, C) = \sum_{i=1}^{n} \mathbf{1}\{Y_i \leq y\}\mathbf{1}\{X_i \in C\}, \quad y \in \mathbb{R}, C \subset D. \tag{4}$$

Then,

$$\begin{aligned} L_n(F, F(\varepsilon)) = &\int_{\mathbb{R}\times(F(\varepsilon)\setminus F)} \ln \frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) N_n(\mathrm{d}y, \mathrm{d}x) \\ &- \int_{\mathbb{R}\times(F\setminus F(\varepsilon))} \ln \frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) N_n(\mathrm{d}y, \mathrm{d}x). \end{aligned} \tag{5}$$

There is a clear similarity between the form of the log-likelihood ratios in (2) and (5), and in both of them, in order to establish the limiting object for $L_n(F, F(\varepsilon))$, we need to define such a limiting object for the process N_n on shrinking sets $F(\varepsilon)\setminus F$ and $F\setminus F(\varepsilon)$. Again, as in (2), in (5) it will not be true that the limiting process should live on the boundary of F.

The set $F(\varepsilon)$, given for all small $\varepsilon \geq 0$, is a set-valued function. Looking on the change-set problem in the breadth it requires, we should speak not about one set-

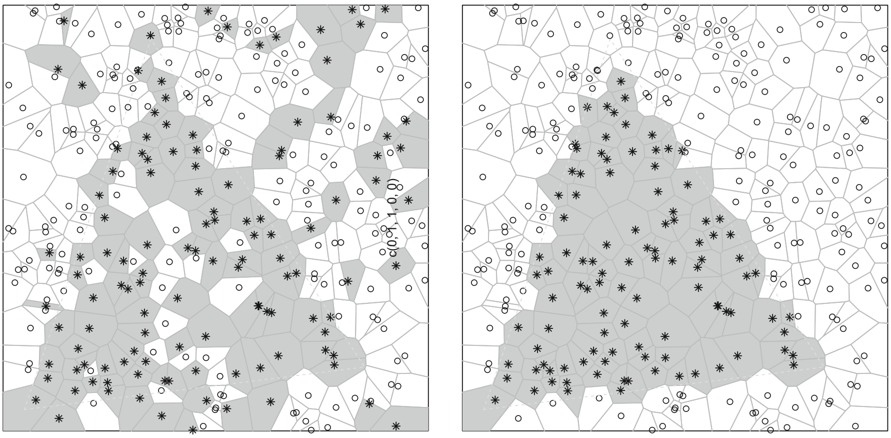

Fig. 3 In comparison, a "nonparametric estimation" of F is illustrated here: for the same points as in Fig. 2, the Voronoi tessellation is constructed and the union of the tiles with "stars" as nuclei is shaded. On the left, the marking of points as "stars" or "circles" is the same as in Fig. 2, while on the right the marking is changed—all points in F, and none outside F, are marked as stars

valued function passing through $F(0)$ at $\varepsilon = 0$, but about a class of such set-valued functions, giving rise to many likelihoods, asymptotically connected with the class of derivatives of $F(\varepsilon)$. We can anticipate that the limiting process will live on the class of these properly defined derivatives.

Before we move on to set-valued derivatives and compare the testing problem with the problem of estimation of sets, we show in Fig. 3 two nonparametric estimations of the set F. Both of them are the maximum likelihood estimators within their corresponding models. The one on the left is, we believe, not consistent. The one on the right is certainly consistent. This fact and the rate of its convergence were the matter of investigation in Khmaladze and Toronjadze (2001), Penrose (2007), Reitzner et al. (2012) and Thäle and Yukich (2016), see also Schneider and Weil (2008), p. 482.

1.5 Set-valued derivatives

Suppose that for each $\varepsilon \in \mathbb{R}$ we are given a Borel set $F(\varepsilon) \in \mathbb{R}^d$. Thus, we have a set-valued function. To consider a relatively general set-up, assume that each $F(\varepsilon)$ is a *solid* set, a compact set which is the closure of its interior points and such that $\mu_d(\partial F(\varepsilon)) = 0$. We also assume that $F(\varepsilon)$ is continuous in ε in the Hausdorff metric, although more general forms of continuity were considered in Khmaladze (2007) and Khmaladze and Weil (2014). Since we are interested in differentiability of $F(\varepsilon)$ in ε at some particular value ε_0, we may choose $\varepsilon_0 = 0$ and consider our set-valued function in some small interval $[0, \varepsilon T]$ with constant T.

Differentiation of set-valued functions is not a new topic and does not start with our attempt to introduce a new type of derivative. The topic has a long history and several approaches to the problem of differentiation form now an important and well-developed mathematical theory. In Khmaladze (2007) and Khmaladze and Weil

(2014), we referred to literature sources in differential inclusions, such as Aubin and Cellina (1984), in differentials of $F(\varepsilon)$ understood as different forms of affine mappings, such as Artstein (1995), Artstein (2000) and Lemaréchal and Zowe (1991), and in derivatives considered as tangential cones, an approach which is particularly interesting for problems in convex analysis, see Aubin (1981), Aubin and Frankowska (1990), Borwein and Zhu (1999) and Pflug (1996). We will not compare our method with the various existing notions (see (Khmaladze 2007, Section 4) for some results in this direction), but merely say that it was surprising to see that the change-set problem of statistics required still another approach.

The derivative of $F(\varepsilon)$ at $\varepsilon = 0$ which we introduce may be called a *fold-up derivative*, since it lifts points in $\mathbb{R}^d$ to a cylinder. It uses the natural decomposition of a point $z \notin F = F(0)$ in the form

$$z = x + tu, \tag{6}$$

where x is the point in the boundary ∂F nearest to z, t is the distance of z from ∂F and u the direction from x to z. Since we want to allow deviations $F(\varepsilon)$ from F not only to the outside but also to the inside of F, a corresponding decomposition (6) has to be performed on F as well. The existence and uniqueness of the decomposition (6), and the decomposition of the Lebesgue measure it induces, lead to interesting and deep questions in geometric measure theory. We will discuss these geometric aspects in Sect. 2, but mention here that at the basis there is the Steiner formula from convex geometry and its generalization to closed sets provided in Hug et al. (2004).

For the asymptotic analysis of $F(\varepsilon)$, as $\varepsilon \to 0$, we need the local magnification map, introduced in Khmaladze (2007) as

$$\tau_\varepsilon : z \mapsto (t/\varepsilon, x, u),$$

by which the outside of F is mapped, or folded up, to a part of what we call the *normal cylinder* $\Sigma = \mathbb{R} \times \mathrm{Nor}(F)$, where the normal bundle $\mathrm{Nor}(F)$ of F consists of the pairs (x, u) arising from (6). The derivative of $F(\varepsilon)$ at $\varepsilon = 0$ will then be a subset B in Σ.

A strong support for the use of the normal cylinder comes from the description of the change-set problem above. Namely, let F be a solid set of reach $\varepsilon > 0$. This means that for all points $z \notin F$ which have distance smaller than ε from F, the nearest point in F is unique, whereas for $\delta > \varepsilon$ there are points z with distance δ to F which have at least two nearest points in F. Let $A_1(\varepsilon) = A_1$, $A_2(\varepsilon) = A_2$ be two sets in the ε-neighborhood of F and let $B_1 = \tau_\varepsilon(A_1)$ and $B_2 = \tau_\varepsilon(A_2)$ be the corresponding sets of magnified points as above. Then, if A_1 and A_2 are disjoint, the magnified sets B_1 and B_2 will also be disjoint. If N_n is, for each n, a Poisson process on $\mathbb{R}^d$, the number of its jumps in A_1 and in A_2 are two independent Poisson random variables. If we use τ_ε to map these jump points Z onto random points (ζ, X, U) in the cylinder Σ, there will be no apparent controversy and the image process $\tau_\varepsilon(N_n)$ on Σ is still a Poisson process. However, if we map Z onto $\overline{Z} = X + \zeta U$, the point Z stretched in $\mathbb{R}^d$, the image points from N_n on A_1 and on A_2 may lie in overlapping sets, see Fig. 4 , which would be incompatible with the independence properties of spatial Poisson processes.

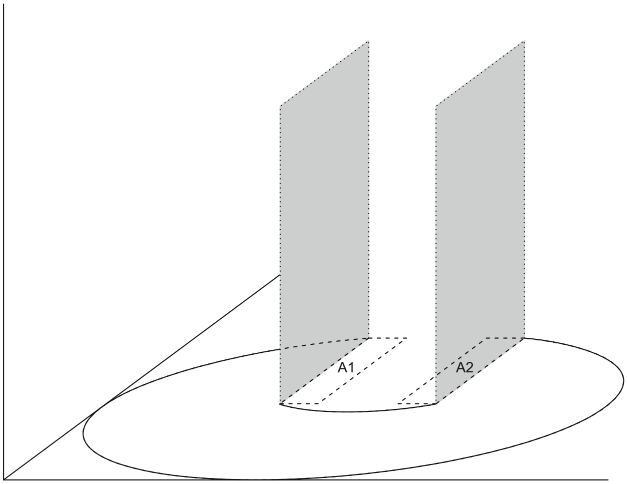

Fig. 4 The circle with the part in "horseshoe" shape cut out is F. The union of F and the narrow strips protruding in this cutoff region is $F(\varepsilon)$. The symmetric difference $F(\varepsilon)\Delta F$ cannot be magnified inside the plane without causing the images of the disjoint strips to overlap. This would be in conflict with the theory in many respects. Another dimension is necessary to describe the derivative, as shown in the figure: the strips A_1 and A_2 are folded up and magnified

Later we will see that we can consider the cylinder $\Gamma = \mathbb{R}\times\partial F$, which is convenient for visualization, and project derivative sets from Σ to Γ (this is already used in Fig. 4).

2 Geometric background

For a set-valued derivative of a family $F(\varepsilon)$ at a set $F \subset \mathbb{R}^d$, as we have it in mind, the points in the neighborhood of F have to be inspected. A quantitative description of the neighborhood of a set, in case the set is compact and convex (a *convex body* K), has been obtained as early as 1840 by Jacob Steiner, with his now famous Steiner formula, see Gruber (1993) and Schneider (2013), p. 223. We describe this first and then turn to local versions and generalizations, for convex bodies K and after that for quite general closed sets F.

2.1 The classical Steiner formula

For a convex body $K \subset \mathbb{R}^d$, we consider the outer parallel body

$$K_t = \{x \in \mathbb{R}^d : d_K(x) \le t\}, \quad t > 0,$$

which is built by all points x which have Euclidean distance $d_K(x)$ to K less or equal t. The *Steiner formula* expresses the volume $V_d(K_t)$ of K_t as a polynomial in t,

$$V_d(K_t) = \sum_{j=0}^{d} t^{d-j}\kappa_{d-j}V_j(K). \tag{7}$$

Here, κ_{d-j} is the $(d-j)$-dimensional volume of the unit ball B^{d-j} in $\mathbb{R}^{d-j}$. The most interesting aspect of this formula is the coefficients $V_j(K)$ which describe the geometric structure of K and its boundary ∂K.

In the formulation of (7), we used modern terminology. Steiner proved the result for polytopes and smooth bodies K in dimensions $d = 2$ and $d = 3$, where the coefficients had a simple geometric interpretation. For $d = 3$, the volume, the surface area, the integral mean curvature and the Euler characteristic arise. The general situation was prepared by Minkowski, who used a similar expansion of the volume of a sum set

$$V_d(t_1 K_1 + t_2 K_2 + \cdots + t_k K_k) = \sum_{i_1=1}^{k} \cdots \sum_{i_d=1}^{k} t_{i_1} \cdots t_{i_d} V(K_{i_1}, \ldots, K_{i_d})$$

for $t_i > 0$ and convex bodies K_i, to introduce *mixed volumes* $V(K_1, \ldots, K_d)$, a notion which is at the heart of the Brunn–Minkowski theory in convex geometry. Notice that the Steiner formula is a special case, since $K_t = K + tB^d$, where $B^d \subset \mathbb{R}^d$ is the unit ball. We refer to the book of Schneider (2013), for an up-to-date survey on the Brunn–Minkowski theory, including variants of the Steiner formula and historical remarks on the development of the theory, and for further details of most notions and results which we present in this section.

From Minkowski's approach, it turns out that the coefficients in the polynomial expansion (7) are special mixed volumes of K and B^d. Such quantities also showed up later in integral geometric formulas, a fact which motivated to call them *quermass-integrals*. Nowadays it is more popular to use a rescaled version of these functionals, the *intrinsic volumes* $V_j(K)$, since they are independent of the dimension of the ambient space. Hence, for a j-dimensional body K in $\mathbb{R}^d$, $0 \le j \le d$, the value $V_j(K)$ is just the j-dimensional volume of K. Moreover, the subscript j corresponds to the degree of homogeneity of V_j, $V_j(\alpha K) = \alpha^j V_j(K)$ for $\alpha \ge 0$. We emphasize that $V_d(K)$ is the (d-dimensional) volume of K, $V_{d-1}(K)$ is half the surface area, $V_1(K)$ is proportional to the mean width and $V_0(K)$ is the Euler characteristic, thus $V_0(K) = 0$, if $K = \emptyset$, and $V_0(K) = 1$, if $K \neq \emptyset$. The remaining functionals $V_j(K)$ can be expressed as certain curvature integrals over ∂K, if K has a smooth boundary. For example, $V_{d-2}(K)$ is then up to a constant the integral mean curvature of K.

Polynomial expansions of volumes of parallel sets have been later studied for other set classes as well, an example is Weyl's tube formula for smooth manifolds (Weyl (1939)) or Federer's formula for sets of positive reach (Federer (1959)). We will come to such a general result in a moment, but describe the convex situation a bit further.

Convex bodies have nice geometric properties, and their boundary structure is well understood. They include sets with a rather discrete boundary structure like convex polytopes, which are convex hulls of finitely many points, but also convex sets which are bounded by a smooth manifold. The boundary ∂K of a convex body K has finite $(d-1)$-dimensional Hausdorff measure $\mathcal{H}^{d-1}$ and ∂K determines K uniquely. In each boundary point x there is at least one supporting hyperplane which leads to an outer normal $u(x)$, but a point $x \in \partial K$ can have more than one, and thus infinitely many, normals. This behavior makes convex bodies especially useful for a local description of their neighborhood. In particular, we get a local version of (7) in a very natural way. Such local Steiner formulas have been proved in 1938 by Fenchel and Jessen, introducing the *area measures* of K, and in 1959 by Federer, establishing the *curvature measures*, actually for a larger class of sets K, the sets of positive reach.

2.2 The local Steiner formula in the convex case

We describe the local result using a common generalization of area and curvature measures, the *support measures* due to Schneider (1979). For a convex body $K \subset \mathbb{R}^d$, we choose a Borel set $A \subset \mathbb{R}^d \times S^{d-1}$, where S^{d-1} denotes the unit sphere in $\mathbb{R}^d$. As we have already indicated above, each point $z \in \mathbb{R}^d \setminus K$ has a (unique) decomposition

$$z = x + tu, \tag{8}$$

with $x = p_K(z) \in \partial K$, $u = u_K(z) = \frac{z-x}{\|z-x\|} \in S^{d-1}$ and $t = d_K(z) > 0$. Here, $p_K(z)$ is the point in K nearest to z (the *metric projection* of z onto K) and $u_K(z)$ is an outer normal of K in the point $p_K(z)$. The property that each z outside K has a unique nearest point in K is of course due to the convexity of K, in fact it characterizes convex sets by Motzkin's theorem (Motzkin (1935)). We now define the local outer parallel set of A,

$$A_t = \{z \in K_t \setminus K : (p_K(z), u_K(z)) \in A\}, \quad t > 0.$$

Then, A_t is a Borel set in $\mathbb{R}^d$ and the following *local Steiner formula* holds for the Lebesgue measure of A_t,

$$\mu_d(A_t) = \frac{1}{d} \sum_{j=1}^{d} \binom{d}{j} t^j \Theta_{d-j}(K, A) \tag{9}$$

with finite (nonnegative) measures $\Theta_0(K, \cdot), \ldots, \Theta_{d-1}(K, \cdot)$ on $\mathbb{R}^d \times S^{d-1}$, the support measures of K. Actually, the measures $\Theta_i(K, \cdot)$ are concentrated on the *normal bundle*

$$\mathrm{Nor}(K) = \{(x, u) : x \in \partial K, u \text{ is an outer normal of } K \text{ at } x\}.$$

The image measure of $\Theta_i(K, \cdot)$ under the projection $(x, u) \mapsto x$ yields the *curvature measure* $C_i(K, \cdot)$ and the image under $(x, u) \mapsto u$ is the *area measure* $S_i(K, \cdot)$ of K. A comparison of (9) and (7) shows that

$$\Theta_i(K, \mathrm{Nor}(K)) = C_i(K, \mathbb{R}^d) = S_i(K, S^{d-1}) = \frac{d\kappa_{d-i}}{\binom{d}{i}} V_i(K),$$

for $i = 0, \ldots, d-1$.

If we consider a deviation $K(\varepsilon)$ as a (not necessarily convex) set in the neighborhood of a convex body K, it would be a too narrow model to allow local changes of K only to the outside. Hence, we should also consider a Steiner-type decomposition of K itself. Here, we can make use of the fact that, for μ_d-almost all $z \in K$, the metric projection $p_{\partial K}(z)$ onto the boundary ∂K is unique. The set S_K of points $z \in K$ which have more than one nearest point in ∂K is called the *(inner) skeleton* of K. The *inner parallel body* K_{-t} of K, $t \geq 0$, is defined as

$$K_{-t} = \{z \in K : z + tB^d \subset K\}.$$

Notice that $K_{-t} + tB^d \subset K$, but in general we do not have equality here. The largest value $r = r(K) \geq 0$ such that $(K_{-r})_r = K$ is called the *interior reach* of K. As a local counterpart, we define the *local (interior) reach* $r(x)$ of a boundary point $x \in \partial K$ as the largest $r \geq 0$ such that x is in the boundary of a ball $B(y, r)$ with center y and radius r and with $B(y, r) \subset K$ (here $r(x) = 0$ means that there is no such ball). Then $r(K) = \min_{x \in \partial K} r(x)$. If K has no interior points, we have $r(x) = 0$ for all $x \in \partial K = K$, hence $r(K) = 0$, but we can have $r(K) = 0$ in many other cases, for example if K is a convex polytope. Then $r(x) = 0$ for all $x \in \partial K$, which are not in the relative interior of a facet of K. The following result is the most general version of a (local) Steiner formula for convex bodies. For any μ_d-integrable real function f on $\mathbb{R}^d$, we have

$$\int_{\mathbb{R}^d} f(z)\mu_d(\mathrm{d}z) = \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{\mathrm{Nor}(K)} \int_{-r(x)}^{\infty} f(x+tu)t^{j-1}\mathrm{d}t\,\Theta_{d-j}(K, \mathrm{d}(x,u)) \tag{10}$$

(see, e.g., Theorem 1 in Khmaladze and Weil (2008)).

2.3 Extension to solid sets

Although the assumption of a convex body underlying the statistical situation is very convenient from a geometrical point of view, for applications it will be useful to consider more general set classes. For polyconvex sets (finite unions of convex bodies) or sets of positive reach, extensions of the concepts and results described above in the convex case are possible with appropriate modifications. We now consider a rather general framework allowing closed sets $F \subset \mathbb{R}^d$ with only a few regularity properties. The approach uses a general Steiner formula for closed sets (Theorem 2.1 in Hug et al. (2004)).

In general, closed sets F can have quite a complicated structure. They need not have a defined inner and outer part. Even for compact F, the boundary ∂F can have infinite $(d-1)$-dimensional Hausdorff measure $\mathcal{H}^{d-1}(\partial F) = \infty$ or positive Lebesgue measure $\mu_d(\partial F) > 0$. Boundary points $x \in \partial F$ need not have any normal, but also can have one, two or infinitely many normals. Consequently, the normal bundle $\mathrm{Nor}(F)$ of F (or $\mathrm{Nor}(\partial F)$ of ∂F), as it was defined in Hug et al. (2004) as an extension of the same notion for convex bodies, can also have a rather complicated structure. Moreover, the support measures of F, which were introduced in Hug et al. (2004) as ingredients of the general Steiner formula, are no longer finite nonnegative measures but signed

Radon-type measures. They are finite only on sets in the normal bundle with local reach bounded from below.

In the following, we concentrate on *solid* sets, that is, compact sets F which are the closure of their interior and satisfy $\mu_d(\partial F) = 0$. The assumption of compactness is convenient but not essential here. Since we only work with concepts which are locally defined, an extension to unbounded closed sets (satisfying the appropriate conditions) is easily possible.

For a solid set F, in contrast to the convex case, the nearest point map $z \mapsto p_F(z) \in \partial F$ need not be defined for all $z \in \mathbb{R}^d \setminus F$ anymore, since the smallest distance $d_F(z)$ can be attained in several points of F. Fortunately, the (outer) skeleton of F,

$$S_F = \{z \in \mathbb{R}^d \setminus F : \text{ a point in } F \text{ nearest to } z \text{ is not unique}\},$$

is a set of Lebesgue measure 0 (see Hug et al. 2004). For $z \notin F \cup S_F$, the metric projection $p_F(z)$ exists uniquely and we can define the corresponding outer normal

$$u_F(z) = \frac{z - p_F(z)}{\|z - p_F(z)\|}.$$

As in the convex case, we get

$$z = x + tu \tag{11}$$

with $x = p_F(z)$, $u = u_F(z)$ and $t = d_F(z)$. We define the (outer) normal bundle $\mathrm{Nor}_+(F)$ by

$$\mathrm{Nor}_+(F) = \{(x, u) : x \in \partial F, u \text{ is an outer normal of } F \text{ at } x\}$$

and remark that a point $x \in \partial F$ can have more than one outer normal (for example, it can have two opposite outer normals). In contrast to the convex case, there can be also boundary points $x \in \partial F$ without an outer normal. Those boundary points then do not contribute to the outer normal bundle. Another important difference to the convex situation is that we need not have $p_F(x + tu) = x$ for $(x, u) \in \mathrm{Nor}_+(F)$ and all $t > 0$. This fact gives rise to the *outer reach function* $r_+ = r_{+,F}$ of F, which is defined on $\mathrm{Nor}_+(F)$,

$$r_+(x, u) = \sup\{s > 0 : p_F(x + su) = x\}.$$

Of course, convex bodies K have reach function $r_{+,K} = \infty$.

Since F is the closure of its interior, we can extend the decomposition (11) to the interior of F, as we did in the full-dimensional convex case. This then involves an *inner reach function* r_- of F. The situation is most easily solved if we define the inner reach function r_- of F as the outer reach function of F^*, the closed complement of F, and obey the reflection $R : (x, u) \mapsto (x, -u)$. The fact that, for compact F, the set F^* is not compact, is not a problem here, since we work with locally defined notions. A problem which does occur comes from the fact that the outer normal bundles of F and F^* need not fit together. Namely, a boundary point x of F which has an outer normal u with respect to F appears in a pair $(x, u) \in \mathrm{Nor}_+(F)$. Of course, x is also a boundary point of F^*, but it need not have an outer normal with respect to F^*, and hence $(x, -u)$ might not be a point in $\mathrm{Nor}_+(F^*)$. Therefore, we define the *(extended) normal bundle* $\mathrm{Nor}(F)$ of F as

$$\mathrm{Nor}(F) = \mathrm{Nor}_+(F) \cup R(\mathrm{Nor}_+(F^*))$$

and extend the outer and inner reach functions appropriately (by 0). Notice that in Hug et al. (2004) and Khmaladze and Weil (2008) a slightly different notation was used. The following local Steiner formula for solid sets F is then a consequence of Theorem 5.2 in Hug et al. (2004). It reads

$$\int_{\mathbb{R}^d} f(z)\mu_d(\mathrm{d}z) \tag{12}$$
$$= \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{\mathrm{Nor}(F)} \int_{-r_-(x,u)}^{r_+(x,u)} f(x+tu)t^{j-1}dt\Theta_{d-j}(F, \mathrm{d}(x,u))$$

and holds for any measurable bounded function f with bounded support on $\mathbb{R}^d$ and for certain set functions $\Theta_i(F, \cdot)$, $i = 0, \ldots, d-1$, on the right side which we call the *support measures* of F.

Full-dimensional convex bodies K are solid and for them (12) just reduces to (10). Moreover, we then have $\mathrm{Nor}(K) = \mathrm{Nor}_+(K)$, $r_+(x, u) = \infty$ for all $(x, u) \in \mathrm{Nor}(K)$, and $r_-(x, u) = r(x)$. Moreover, the support measures in (10) are those defined by (9).

For non-convex sets F, the situation is more complicated since the set functions $\Theta_i(F, \cdot)$ need not be finite Borel measures anymore. First, they may have positive and negative values; hence, they are signed functions, for example, if F has convex and concave pieces in the boundary. Moreover, $\Theta_i(F, A)$ is not defined for all Borel sets $A \subset \mathrm{Nor}(F)$, but only for those for which the (outer and inner) reach functions are bounded from below by a positive constant (r*-bounded* sets). Such set functions $\Theta_i(F, \cdot)$ are called r*-measures*. The situation can be compared to (signed) Radon measures in functional analysis which are also not defined on all Borel sets of a space, but only on bounded sets. For the details on r-bounded sets and r-measures, which we leave out here, we refer to Hug et al. (2004).

2.4 First-order terms, regular points and the normal cylinder

Because of the polynomial-like nature of (12), the set-valued derivatives which we describe in the next section will be driven by the support measure $\Theta_{d-1}(F, \cdot)$ and they will live on the normal cylinder of F. Therefore, we will have a closer look at the structure of this $(d-1)$st support measure and we also introduce the normal cylinder Σ of F.

If K is a convex body, the $(d-1)$st curvature measure $C_{d-1}(K, \cdot)$ is the Hausdorff measure $\mathcal{H}^{d-1}$ on the boundary and the area measure $S_{d-1}(K, \cdot)$ is the image of this Hausdorff measure under the Gauss map $\gamma : \partial K \to S^{d-1}, x \mapsto u(x)$. Remember here that both of these measures are the image measures of $\Theta_{d-1}(K, \cdot)$ under the projections $(x, u) \mapsto x$, resp. $(x, u) \mapsto u$. Now, it is an important fact, that $\mathcal{H}^{d-1}$-almost all boundary points x of a convex body K are *regular points*, that is, they have one and only one outer normal $u(x)$ and so the Gauss map is defined almost everywhere (see, e.g., p. 92 in Schneider (2013)).

With some adjustments, similar results hold also for closed sets F. Here, we call a point $x \in \partial F$ *regular*, if F has one outer normal u or two opposite (outer) normals $u, -u$ in x (the latter usually occurs in flat parts of F). For a solid set F, the set $\partial^2 F$ of regular points $x \in \partial F$ with two opposite outer normals seems to be negligible, since there should be no flat parts. However Example 1 in Ambrosio et al. (2008) shows that $\mathcal{H}^{d-1}(\partial^2(F)) > 0$ can occur. For an integral representation of $\Theta_{d-1}(F, \cdot)$ over the boundary ∂F (see (14)), we consider $\operatorname{reg}(F)$, the union of all regular points of F and of F^*. We also need a further restriction on solid sets, namely that "flat points" of F or F^* are negligible. Thus, we add

$$\mathcal{H}^{d-1}(\partial^2(F) \cup \partial^2(F^*)) = 0 \tag{13}$$

to our conditions on F. Then we have

$$\Theta_{d-1}(F, \cdot) = \int_{\operatorname{reg}(F)} \mathbf{1}\{(x, u(x)) \in \cdot\} \mathcal{H}^{d-1}(\mathrm{d}x). \tag{14}$$

Note that, due to condition (13), the normal vector $u(x)$ is uniquely determined for $\mathcal{H}^{d-1}$-almost all $x \in \operatorname{reg}(F)$ ($u(x)$ is either the outer normal of F in x or $-u(X)$ is the inner normal or both properties hold). Notice however that, in contrast to the convex case, we may still have $\mathcal{H}^{d-1}(\partial F \setminus \operatorname{reg}(F)) > 0$ (see Hug et al. 2004).

Given a solid set $F \subset \mathbb{R}^d$, we can use the decomposition (11) and the local magnification map (see Sect. 1.5) and the next section, to map a point z to the triple (t, x, u) representing it. This works for $z \notin S_{\partial F} \cup \partial F$, and the corresponding triple lies in the *normal cylinder*

$$\Sigma = \Sigma(F) = \mathbb{R} \times \operatorname{Nor}(F).$$

Here $t = d_F(z)/\varepsilon > 0$, if $z \notin F$, and $t = -d_{\partial F}(z)/\varepsilon < 0$, if $z \in F$. In the convex case, the whole upper part $\Sigma^+ = \{(t, x, u) \in \Sigma : t > 0\}$ of Σ appears as the image, whereas in the lower part $\Sigma^- = \{(t, x, u) \in \Sigma : t < 0\}$ the images build a bounded subset. For general solid F, the image set in the upper half cylinder Σ^+ can have bounded and unbounded parts, whereas the images in Σ^- are again bounded.

The normal cylinder Σ may be difficult to visualize. Namely, since F may have more than one normal in $x \in \partial F$, it is in general not enough to think of Σ as the cylinder over the base ∂F. However, (14) allows us to simplify the situation. Indeed, consider the measure

$$M = \mu_1 \otimes \Theta_{d-1}(F, \cdot) \tag{15}$$

(where μ_1 is Lebesgue measure in $\mathbb{R}$), which will play a prominent role in the definition of derivatives below. If F is a solid set (and (13) is satisfied), (14) shows that for M-almost all points $(t, x, u) \in \Sigma$, the mapping $(t, x, u) \mapsto (t, x)$ is injective and the image measure of M under this mapping is the measure

$$m = \mu_1 \otimes C_{d-1}(F, \cdot)$$

on the cylinder $\Gamma = \mathbb{R} \times \partial F$. Here, $C_{d-1}(F, \cdot)$ is the image of $\Theta_{d-1}(F, \cdot)$, the $(d-1)$st curvature measure of F. Hence, with respect to M, the cylinders Σ and Γ can be identified and M can be replaced by m. For a further simplification, we notice that the curvature measure $C_{d-1}(F, \cdot)$ is the Hausdorff measure $\mathcal{H}^{d-1}$ restricted to the set $\mathrm{reg}(F)$. Thus, if we add

$$\mathcal{H}^{d-1}(\partial F \setminus \mathrm{reg}(F)) = 0 \tag{16}$$

as a final condition for our solid sets F, we get

$$m = \mu_1 \otimes \mathcal{H}^{d-1},$$

as in the convex case.

To summarize, the step from convex bodies to general solid sets faces some difficulties due to the more complicated structure of the boundary, which led to the additional assumptions (13) and (16).

3 Fold-up derivatives

We return to the situation which we described in Introduction. Namely, we consider a set-valued function $(F(\varepsilon), 0 \leq \varepsilon \leq 1)$, with $F(\varepsilon)$ in $\mathbb{R}^d$ and want to define the derivative at $F = F(0)$. The approach to define a set-valued derivative was developed in Khmaladze (2007) for convex bodies and sets of positive reach as F, and later was extended in Khmaladze and Weil (2014) to a class of general closed sets—to the solid sets of the previous section. To explain the essential ideas, we may concentrate on solid sets $F(\varepsilon)$ which satisfy (13) and (16) and which converge to F in the Hausdorff metric, as $\varepsilon \to 0$. Moreover, we require that the symmetric difference $F(\varepsilon)\Delta F$ lies in the neighborhood $(\partial F)_{\varepsilon T}$ of the boundary ∂F, with some constant $T > 0$, for small enough ε. For simplicity, we assume $F(\varepsilon)\Delta F \subset (\partial F)_{\varepsilon T}$ for $0 \leq \varepsilon \leq 1$.

3.1 The definition

In order to define the derivative of $F(\varepsilon)$ at F, we use the representation (11) and define the *local magnification map* τ_ε,

$$\tau_\varepsilon(z) = \left(\frac{d_F(z)}{\varepsilon}, p_F(z), u_F(z) \right),$$

for $z \notin F \cup S_F$, and

$$\tau_\varepsilon(z) = \left(-\frac{d_{F^*}(z)}{\varepsilon}, p_{F^*}(z), -u_{F^*}(z) \right),$$

for $z \in F \backslash (\partial F \cup S_F)$. Remark that $\tau_\varepsilon(z)$ lies in the normal cylinder Σ of F. In fact, τ_ε is bicontinuous and one-to-one as a mapping from $\mathbb{R}^d \backslash (S_{\partial F} \cup \partial F)$ onto the set $\tau_\varepsilon(\mathbb{R}^d \backslash (S_{\partial F} \cup \partial F)) \subset \Sigma$. Consider the image

$$B(\varepsilon) = \tau_\varepsilon(A(\varepsilon))$$

of $A(\varepsilon) = F(\varepsilon) \Delta F$ under the local magnification map. If the sets $B(\varepsilon)$ converge, as $\varepsilon \to 0$, in a reasonable way to a set $B \subset \Sigma$, then B will be our derivative set.

In order to motivate the appropriate notion of convergence on Σ, consider the image $\tau_\varepsilon \circ \mu_d$ of the Lebesgue measure μ_d on $(\partial F)_{\varepsilon T}$ under τ_ε. Suppose that the reach functions of F satisfy $r_+, r_- \geq \varepsilon$. Then, for a Borel set $A \subset (\partial F)_{\varepsilon T}$, such that $C = \tau_\varepsilon(A) = [-T, T] \times \tilde{A}$, $\tilde{A} \subset \mathrm{Nor}(F)$, the local Steiner formula (12) yields

$$\begin{aligned}(\tau_\varepsilon \circ \mu_d)(C) &= \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{\mathrm{Nor}(F)} \int_{-\varepsilon T}^{\varepsilon T} \mathbf{1}\{(t, x, u) \in C\} t^{j-1} \mathrm{d}t\, \Theta_{d-j}(F, \mathrm{d}(x, u)) \\ &= \sum_{j=1}^{d} \frac{2}{j} \binom{d-1}{j-1} (\varepsilon T)^j \Theta_{d-j}(F, \tilde{A}).\end{aligned}$$

Here, the leading term in ε is $2T\Theta_{d-1}(F, \tilde{A})$. Therefore, it seems natural to use the measure M from (15) on Σ and define $B(\varepsilon) \to B$ by $M(B(\varepsilon) \Delta B) \to 0$.

Definition (Khmaladze 2007, see also Khmaladze and Weil 2014). For $0 \leq \varepsilon \leq 1$, let $A(\varepsilon)$ be a Borel set with $A(\varepsilon) \subset (\partial F)_{\varepsilon T}$. The set-valued mapping $A(\varepsilon)$ is *differentiable* at ∂F, for $\varepsilon = 0$, if there exists a Borel set $B \subset \Sigma$ such that

$$M(\tau_\varepsilon(A(\varepsilon)) \Delta B) \to 0, \quad \text{as } \varepsilon \to 0.$$

The set-valued function $F(\varepsilon)$ is *differentiable* at F, for $\varepsilon = 0$, if $A(\varepsilon) = F(\varepsilon) \Delta F$ is differentiable at ∂F. The set B is then called the *fold-up derivative* of $A(\varepsilon)$ at ∂F (respectively of $F(\varepsilon)$ at F) and we write

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} F(\varepsilon)|_{\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} A(\varepsilon)|_{\varepsilon=0} = B.$$

Note that in Khmaladze (2007) and Khmaladze and Weil (2014) a condition of essential boundedness of the sets $A(\varepsilon)$ was used which is automatically fulfilled here, since we assumed $A(\varepsilon) \subset (\partial F)_{\varepsilon T}$.

If we consider the image $\tilde{B}$ of the derivative set B under the M-almost everywhere defined map $(t, x, u) \mapsto (t, x)$, then $\tilde{B}$ sits in the cylinder $\Gamma = \mathbb{R} \times \partial F$ and is in fact in one-to-one correspondence with B. Let us map now the set $\tilde{B}$ onto $\mathbb{R}^d$, by $(t, x) \mapsto x + tu(x)$. Here, $u(x)$ is the unique normal in x, if $x \in \mathrm{reg}(F)$ and $-u(x)$ is the normal (with respect to F^*), if $x \in \mathrm{reg}(F^*) \backslash \mathrm{reg}(F)$. The corresponding image $\hat{B}$ will only represent B, if we can distinguish overlapping points $x_1 + t_1 u(x_1) = x_2 + t_2 u(x_2)$ coming from different boundary parts $x_1 \neq x_2$ of F. Figure 4 illustrates this situation.

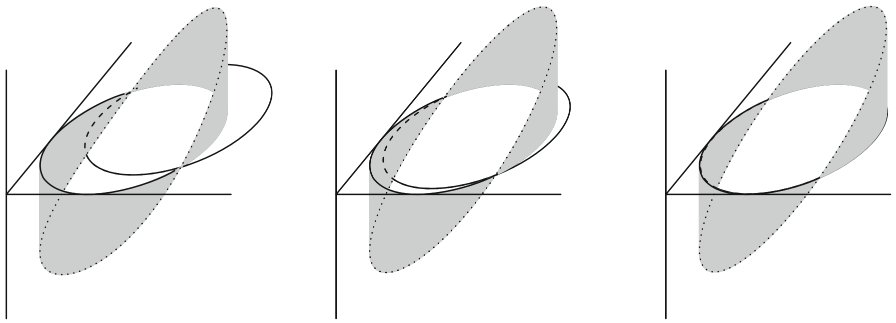

Fig. 5 Shifted circles converge to the initial one. The first shifted circle is quite far, the next is nearer, but the last is almost indistinguishable from the initial one. However, the fold-up sets change little and the convergence to the derivative is visible

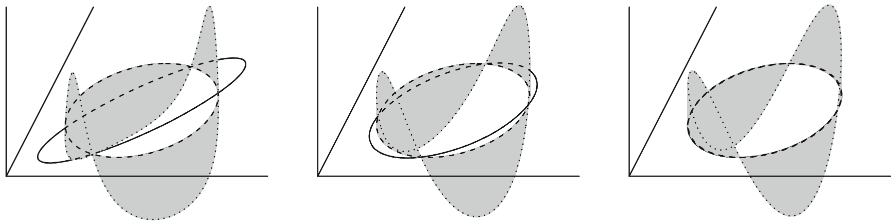

Fig. 6 Ellipses approaching a circle. The first ellipse is quite far, the next is nearer, but the last ellipse is almost indistinguishable from the original circle. However, again, the fold-up sets change little, visualizing the convergence to the derivative

Figures 5 and 6 show the fold-up derivative in two simple situations, shifted circles converging to the original circle and ellipses converging to a circle.

3.2 Derivative in measure

The fold-up derivative $(d/d\varepsilon)F(\varepsilon)$ at $\varepsilon = 0$ can also be called the *derivative in measure*, not only since the symmetric difference metric with respect to M is used in the definition, but also for a reason which we explain now.

Let $\mathbb{P}$ be an absolutely continuous measure on $\mathbb{R}^d$ with density $f \geq 0$ and let $F \subset \mathbb{R}^d$ be a solid set. We assume that $f(z), z \in \mathbb{R}^d$, can be approximated in the neighborhood of ∂F by functions $\overline{f}_+ \geq 0$ from outside and $\overline{f}_- \geq 0$ from inside, defined on ∂F and depending only on $p_{\partial F}(z)$. More precisely, we assume that

$$\begin{aligned} &\frac{1}{\varepsilon}\int_{\mathbb{R}^d} \mathbf{1}\{0 < \mathrm{d}(F, z) \leq \varepsilon\}|f(z) - \bar{f}_+(p_F(z))|\mu_d(\mathrm{d}z) \to 0, \\ &\frac{1}{\varepsilon}\int_{\mathbb{R}^d} \mathbf{1}\{0 < \mathrm{d}(F^*, z) \leq \varepsilon\}|f(z) - \bar{f}_-(p_{F^*}(z))|\mu_d(\mathrm{d}z) \to 0, \end{aligned} \tag{17}$$

as $\varepsilon \to 0$. Now define a measure $\mathbb{Q}$ on Σ by

$$\mathbb{Q}(\mathrm{d}(s,x,u)) = ds \times \bar{f}_+(x)\Theta_{d-1}(F, \mathrm{d}(x,u)) \quad \text{on} \quad \Sigma^+,$$

and

$$\mathbb{Q}(\mathrm{d}(s,x,u)) = \mathrm{d}s \times \bar{f}_-(x)\Theta_{d-1}(F, \mathrm{d}(x,u)) \quad \text{on} \quad \Sigma^-.$$

Theorem 1 (*Khmaladze 2007, see also Khmaladze and Weil 2014*) *Suppose that the measure $\mathbb{P}$ satisfies condition (17) and suppose that the functions $\bar{f}_-$, $\bar{f}_+$ are integrable with respect to $|\Theta_i|(F,\cdot)$, for $i = 0,\dots,d-1$. Let $A(\varepsilon) \subset (\partial F)_{\varepsilon T}$ be differentiable at ∂F (with derivative $B \subset \Sigma$). Then*

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{P}(A(\varepsilon))|_{\varepsilon=0} = \mathbb{Q}\left(\frac{\mathrm{d}}{\mathrm{d}\varepsilon}A(\varepsilon)|_{\varepsilon=0}\right) = \mathbb{Q}(B). \tag{18}$$

Equation (18) highlights the fact that the fold-up derivative of a set-valued function is a set-valued function and shows how to interchange the differentiation in ε with taking measure. For the d-dimensional volume V_d, Equation (18) brings us back to the statement

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V(A(\varepsilon))|_{\varepsilon=0} = M\left(\frac{\mathrm{d}}{\mathrm{d}\varepsilon}A(\varepsilon)|_{\varepsilon=0}\right) = M(B),$$

which underlies our definition of differentiability.

For the *proof* of Theorem 18, the asymptotic behavior of $\varepsilon^{-1}\mathbb{P}(A(\varepsilon))$ has to be established, since $\mathbb{P}(A(0)) = 0$ due to our assumption $\mu_d(\partial F) = 0$. Condition (17) allows to replace here $\mathbb{P}$ by the absolutely continuous measure $\overline{\mathbb{P}}$ on $(\partial F)_{\varepsilon T}$ with density $\overline{f}_+$ outside F and density $\overline{f}_-$ in F. Now the outside and inside parts $A^+(\varepsilon) = A(\varepsilon)\backslash F$ and $A^-(\varepsilon) = A(\varepsilon) \cap F$ can be treated separately, in a totally analogous way.

The Steiner formula (12) shows that

$$\begin{aligned}\bar{\mathbb{P}}(A^+(\varepsilon)) = &\int_{\mathrm{Nor}(F)} \int_0^{r_+(x,u)\wedge\varepsilon} \bar{f}_+(x)\mathbf{1}_{A^+(\varepsilon)}(x+tu)\mathrm{d}t\, \Theta_{d-1}(F, \mathrm{d}(x,u)) \\ &+ \sum_{j=2}^{d} \binom{d-1}{j-1} \int_{\mathrm{Nor}(F)} \int_0^{r_+(x,u)\wedge\varepsilon} \bar{f}_+(x)\mathbf{1}_{A^+(\varepsilon)}(x+tu) \\ &\times t^{j-1}\mathrm{d}t\, \Theta_{d-j}(F, \mathrm{d}(x,u)).\end{aligned} \tag{19}$$

The sum of the higher-order terms is $o(\varepsilon)$. For the first summand in (19), we have

$$\begin{aligned}&\frac{1}{\varepsilon}\int_{\mathrm{Nor}(F)} \int_0^{r_+(x,u)\wedge\varepsilon} \bar{f}_+(x)\mathbf{1}_{A^+(\varepsilon)}(x+tu)\mathrm{d}t\, \Theta_{d-1}(F, \mathrm{d}(x,u)) \\ &\quad = \int_\Sigma \mathbf{1}\left\{0 \le t \le \frac{r_+(x,u)}{\varepsilon} \wedge 1\right\} \bar{f}_+(x)\mathbf{1}_{B^+(\varepsilon)}(t,x,u)M(\mathrm{d}(t,x,u))\end{aligned}$$

with $B^+(\varepsilon) = \tau_\varepsilon(A^+(\varepsilon))$. The differentiability of $A(\varepsilon)$ implies that of $A^+(\varepsilon)$ (with limit B^+). Therefore, the function $|\mathbf{1}_{B^+(\varepsilon)}(t, x, u) - \mathbf{1}_{B^+}(t, x, u)|$ tends to 0 M−a.e. on Σ. The dominated convergence theorem then implies that

$$\frac{1}{\varepsilon}\bar{\mathbb{P}}(A^+(\varepsilon)) \to \int_\Sigma \mathbf{1}\{0 \le t \le 1\}\ \bar{f}_+(x)\mathbf{1}_{B^+}(t, x, u)M(\mathrm{d}(t, x, u)) = \mathbb{Q}(B^+).$$

We used in this proof the fact that the differentiability of $A(\varepsilon)$ implies the differentiability of $A^+(\varepsilon)$ and $A^-(\varepsilon)$ and vice versa. For this and further algebraic properties of fold-up derivatives, we refer to (Khmaladze 2007, Lemma 2) and (Khmaladze and Weil 2014, Lemma 3).

3.3 Subgraphs and other examples

What are natural examples of fold-up derivatives? We describe one class here, the subgraphs, and then use this to give further examples, the outer parallel sets.

We start with a solid set $F = F(0)$ and consider a family $(h_\varepsilon, 0 \le \varepsilon \le 1)$ of nonnegative measurable functions on $\mathrm{Nor}(F)$ (with $h_0 = 0$). As $A(\varepsilon)$ we take the *subgraph*

$$(h_\varepsilon)_{\mathrm{sub}} = \{z = x + tu : (x, u) \in \mathrm{Nor}(F), 0 \le t \le h_\varepsilon(x, u) \wedge r_+(x, u)\},$$

and assume the following two conditions.

(a) For each $(x, u) \in \mathrm{Nor}(F)$, $h_\varepsilon(x, u)$ is differentiable at $\varepsilon = 0$ with derivative $g(x, u)$. Thus

$$\frac{h_\varepsilon(x, u)}{\varepsilon} \to g(x, u), \quad \varepsilon \to 0.$$

(b) There is a $\delta > 0$, such that the function $\max_{0<\varepsilon\le\delta}(h_\varepsilon/\varepsilon)$ is bounded and integrable with respect to $\Theta_{d-1}(F, \cdot)$. Hence,

$$\max_{0<\varepsilon\le\delta} \frac{h_\varepsilon(x, u)}{\varepsilon} \le T, \tag{20}$$

for some $T > 0$ and

$$\int_{\mathrm{Nor}(F)} \max_{0<\varepsilon\le\delta} \frac{h_\varepsilon(x, u)}{\varepsilon}\ \Theta_{d-1}(F, \mathrm{d}(x, u)) < \infty.$$

Proposition 2 (*Khmaladze 2007, see also Khmaladze and Weil 2014*) *Let F be solid and let h_ε, $0 \le \varepsilon \le 1$, be a family of nonnegative measurable functions on* $\mathrm{Nor}(F)$ *satisfying conditions (a) and (b). Then, $A(\varepsilon) = (h_\varepsilon)_{\mathrm{sub}}$ is differentiable at ∂F and*

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(h_\varepsilon)_{sub}|_{\varepsilon=0} = \left(\frac{\mathrm{d}}{\mathrm{d}\varepsilon}h_\varepsilon|_{\varepsilon=0}\right)_{sub} = (g)_{sub}.$$

Here

$$(g)_{sub} = \{(t, x, u) : 0 < t \le g(x, u), (x, u) \in \mathrm{Nor}(F)\}.$$

A particular simple case is given by $h_\varepsilon = \varepsilon g$,

$$g(x, u) = h_K(u), \quad (x, u) \in \mathrm{Nor}(F),$$

where h_K is the support function of a convex body $K \subset \mathbb{R}^d$ with $0 \in K$. Condition (a) is here obvious and (b) reduces to the integrability of h_K with respect to $\Theta_{d-1}(F, \cdot)$. The derivative set B is then

$$B = \{(t, x, u) : 0 < t \le h_K(u), (x, u) \in \mathrm{Nor}(F)\} = g_{\mathrm{sub}}. \tag{21}$$

Notice that the subgraph $(h_\varepsilon)_{\mathrm{sub}}$, obtained in this case, is different in general from the outer parallel strip $F + \varepsilon K \backslash F$. However, the latter family has the same derivative B given by (21). This follows from (Khmaladze and Weil 2008, Theorem 12), which needed in addition that $\Theta_{d-1}(F, \mathrm{Nor}(F)) < \infty$ and that $\mathcal{H}^{d-1}$-almost all points $x \in \partial F$ are *normal* (which means that there is some ball $C \subset F$ with $x \in C$).

There is also a local parallel set arising from the local reach function,

$$F_{\varepsilon,\mathrm{loc}} = F \cup \{z = x + tu : (x, u) \in \mathrm{Nor}(F), 0 < t \le \varepsilon r(x, u) \wedge \varepsilon\}.$$

This set is the subgraph of εh, $h(x, u) = r(x, u) \wedge 1$, $(x, u) \in \mathrm{Nor}(F)$. Here, $F_{\varepsilon,\mathrm{loc}}$ is differentiable with derivative

$$B = \{(t, x, u) : (x, u) \in \mathrm{Nor}(F), 0 \le t \le r(x, u) \wedge 1\},$$

see (Khmaladze and Weil 2014, Corollary 11).

A further natural situation would be to consider the *sublevel set* $F(c) = \{x : g(x) \le c\}$ of a function g on $\mathbb{R}^d$. We believe that the fold-up derivative of $F(c + \varepsilon)$ at $\varepsilon = 0$ is the subgraph of the gradient of g at the level c. However, this still has to be proved.

4 Convergence of likelihood ratios

Coming back to the change-set problem, as it was described in Introduction, we explain now the role of the fold-up derivatives in this setting. Recall that we consider a family $(F(\varepsilon), 0 \le \varepsilon \le 1)$ of solid sets with $F = F(0)$ and $A(\varepsilon) = F(\varepsilon) \Delta F \subset (\partial F)_{\varepsilon T}$. It may seem that as soon as the functional convergence in distribution of the local processes $N_n(A(\varepsilon))$ is established, it will not be difficult to state the convergence in distribution results for the local likelihood ratio processes in the change-set problems—at least, as they were formulated in Introduction. However, this requires a more detailed argument and we will clarify this point below.

As we have seen in the previous section, in construction of fold-up derivatives we can restrict the local magnification map to the points $z \in \mathbb{R}^d$, which project to regular points of the boundary, i.e., to the points with unique outer normal u and with $-u$ being the inner normal. This, in its turn, allows to map such z directly onto cylinder Γ, which is much easier to visualize:

$$\text{if } z = x \pm tu(x), \text{ then } \tau_\varepsilon(z) = \left(\pm\frac{t}{\varepsilon}, x\right).$$

We use this adjustment throughout this section.

To avoid possible ambiguity, let us agree that, although we will speak about random points, or random jumps, in $(\partial F)_{\varepsilon T}$ and Γ, the term *point process* and the notation N with various indices will be used for random counting measures which, for a given set, count the jump points in it.

4.1 Local processes in the Poissonian case

For the first formulation of the change-set problem and for the Poissonian case, when $n \to \infty$, $\varepsilon \to 0$ and $n\varepsilon \to S > 0$, it may seem that the limit theorem was basically established in (Khmaladze and Weil 2008, Theorem 2). However, let us consider the situation more closely.

We may assume that the intensity $\tilde{\lambda}$ on the change-set F and the gray-level intensity λ on the closed complement F^* are continuous functions in the neighborhood $(\partial F)_{\varepsilon T}$ of the boundary ∂F. That is, we assume that there are limits from inside and outside of F,

$$\begin{aligned} &\frac{1}{\varepsilon}\int_{F(\varepsilon)\setminus F} |\lambda(z) - \lambda(p_F(z))|\mu_d(\mathrm{d}z) \to 0, \\ &\frac{1}{\varepsilon}\int_{F\setminus F(\varepsilon)} |\tilde{\lambda}(z) - \tilde{\lambda}(p_F(z))|\mu_d(\mathrm{d}z) \to 0, \end{aligned} \tag{22}$$

as $\varepsilon \to 0$, and λ and $\tilde{\lambda}$ are different functions on ∂F.

Now let $N_n(A)$, $A \subset (\partial F)_{\varepsilon T}$, be the Poisson process in the narrow strip $(\partial F)_{\varepsilon T}$ with intensity measure

$$\Lambda_n(A) = n\int_{A\cap F} \tilde{\lambda}(z)\mu_d(\mathrm{d}z) + n\int_{A\cap F^*} \lambda(z)\mu_d(\mathrm{d}z).$$

Let us split the jump points $Z \in (\partial F)_{\varepsilon T}$ of N_n into those which project onto regular points on the boundary ∂F, $p_F(Z) \in \mathrm{reg}(F)$, and those which have non-regular projections. Let $N_{n,s}$ denote the part of the process N_n with jump points such that their projections are not regular. Map the jump points with regular projections onto the cylinder $\Gamma = \mathbb{R} \times \partial F$, $\tau_\varepsilon(Z) = (\zeta, \Xi)$, where $\Xi = p_F(Z) \in \partial F$, and let $N_{n,\varepsilon}$ be the process defined by these images. This process is the image of $N_n - N_{n,s}$ under the local magnification map. Then the following result holds true.

Theorem 3 *If $n \to \infty$, $\varepsilon \to 0$, so that $n\varepsilon \to S > 0$, and if (22) is satisfied, then the point process N_{ns} converges to the zero measure in probability, while the point* *process $N_{n,\varepsilon}$ converges in the total variation norm to the Poisson point process N_∞ on Γ with intensity measure*

$$\Lambda_F(B) = S\left[\int_{B^+} \lambda(x)dt\mathcal{H}^{d-1}(dx) + \int_{B^-} \tilde{\lambda}(x)\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x)\right].$$

This result is, basically, equivalent to Theorem 2 in Khmaladze and Weil (2008). The *proof* of it uses the following arguments. The local magnification map τ_ε maps $(\partial F)_{\varepsilon T}$ into $\Gamma_T = [-T, T] \times \partial F$ so that the Borel σ-algebra in $(\partial F)_{\varepsilon T}$ is mapped into the Borel σ-algebra in Γ_T. The thinned Poisson process $N_n - N_{n,s}$ is mapped to the Poisson process $N_{n,\varepsilon}$ on Γ_T, with intensity measure, which is the image of the intensity measure Λ_n. Omitting the higher-order terms, cf. (12), this measure becomes

$$\begin{aligned}&\Lambda_{F,n,\varepsilon}(B)\\ &\quad = S\left[\int_{B^+} \lambda(x+t\varepsilon u(x))\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x) + \int_{B^-} \tilde{\lambda}(x-t\varepsilon u(x))\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x)\right].\end{aligned}$$

Then it follows that $\Lambda_{F,n,\varepsilon}$ converges in total variation to Λ_F. This implies the convergence of the Poisson distribution of $N_{n,\varepsilon}$ to a Poisson distribution with intensity measure Λ_F (see, e.g., Daley and Vere-Jones 2005, Chapter 11 in vol. 2, or Karr 1991).

Note now that in Theorem 3 there is no mention of differentiation and the sets B there are just Borel sets in Γ_T. Thus the statement seems general and sufficient for all purposes. However, given a particular set-valued function $F(\varepsilon)$, or a finite number of such functions, and sets $A(\varepsilon) = F(\varepsilon)\Delta F$, this theorem does not tell us what will be the limit distribution of random variables $N_n(A(\varepsilon))$. In the time when Khmaladze and Weil (2008) was submitted the notion of fold-up derivative of Khmaladze (2007) did not exist yet and thus an unusual situation occurred: there was a functional limit theorem, but no corresponding finite-dimensional limit result. Using the notion of differentiability, the one-dimensional limit theorem below has a very simple proof. Simple as it is, it requires fold-up derivatives.

Theorem 4 (*Khmaladze (2007), Theorem 11*) *Suppose the conditions of Theorem 3 are satisfied. Suppose also that $F(\varepsilon)$ is differentiable at F and B is its fold-up derivative. Then, the random variables $N_n(A(\varepsilon))$ converge in distribution to a Poisson random variable $N_\infty(B)$, where*

$$EN_\infty(B) = \Lambda_F(B).$$

The *proof* proceeds as follows. The image of the thinned random variable $N_n(A(\varepsilon)) - N_{ns}(A(\varepsilon))$ under the local magnification map is the random variable $N_{n\varepsilon}(B(\varepsilon))$ where $B(\varepsilon) = \tau_\varepsilon(A(\varepsilon))$, just as in Theorem 3. The expected value of $N_{n,\varepsilon}(B(\varepsilon))$ is $\Lambda_{F,n,\varepsilon}(B(\varepsilon))$ and the measure $\Lambda_{F,n,\varepsilon}$ converges in total variation to Λ_F. However, our differentiability assumption guarantees that $B(\varepsilon)$ has a limit in measure M. The intensity measure Λ_F is absolutely continuous with respect to M, and therefore $\Lambda_F(B(\varepsilon)) \to \Lambda_F(B)$, which completes the proof.

The reader will notice that the argument used here is very close to the one which is behind (18).

As to how local empirical processes have been constructed and studied without the notion of differentiability we refer to Deheuvels and Mason (1995),Einmahl (1997) and Einmahl and Mason (1997) as, perhaps, the closest to the present paper.

4.2 Convergence of the log-likelihood ratio (first version)

We note that the random part of the likelihood function in (2) is an integral with respect to the point process N_n. We use now Theorem 3 and a differentiability assumption for $F(\varepsilon)$ and deduce the following statement on the limit distribution of the likelihood in (2).

Corollary 5 *Suppose $n \to \infty, \varepsilon \to 0$ so that $n\varepsilon \to S > 0$, and suppose (22) is satisfied. Then if $F(\varepsilon)$ is differentiable and B is its fold-up derivative, the log-likelihood statistic $L_n(F, F(\varepsilon))$ converges in distribution to the random variable*

$$\int_{B^+} \ln \frac{\tilde{\lambda}}{\lambda}(x) N_\infty(\mathrm{d}t, \mathrm{d}x) - S \int_{B^+} (\tilde{\lambda} - \lambda)(x) \mathrm{d}t \mathcal{H}^{d-1}(\mathrm{d}x)$$
$$- \int_{B^-} \ln \frac{\tilde{\lambda}}{\lambda}(x) N_\infty(\mathrm{d}t, \mathrm{d}x) + S \int_{B^-} (\tilde{\lambda} - \lambda)(x) dt \mathcal{H}^{\mathrm{d}-1}(\mathrm{d}x), \tag{23}$$

and the Poisson process N_∞ has intensity measure Λ_F.

In the simple but practically interesting situation, when $\tilde{\lambda}(x) = c\lambda(x)$ on the boundary ∂F, we obtain

$$\ln(c)\, N_\infty(B^+) - (c-1)S \int_{B^+} \lambda(x) \mathrm{d}t \mathcal{H}^{d-1}(\mathrm{d}x)$$
$$- \ln(c) N_\infty(B^-) + (c-1)S \int_{B^-} \lambda(x) \mathrm{d}t \mathcal{H}^{d-1}(\mathrm{d}x).$$

Suppose we want to obtain the limit of $L_n(F, F(\varepsilon))$ under the sequence of alternatives, when the true change-sets are now $F(\varepsilon)$ and we still have $n \to \infty, \varepsilon \to 0, n\varepsilon \to S$. Such a limit theorem is necessary, if we want to speak about the power of the statistical test based on the likelihood ratio. Then we notice that nothing will change in the geometric structure of the problem; $F(\varepsilon)$ still remains differentiable at F with the same derivative. What will change is the intensity measure which drives the Poisson process N_n on the strip $(\partial F)_{\varepsilon T}$. Then, we obtain the following statement.

Corollary 6 *Under the conditions of the previous corollary, but with (22) replaced by*

$$\frac{1}{\varepsilon} \int_{F(\varepsilon)\setminus F} |\tilde{\lambda}(z) - \tilde{\lambda}(p_F(z))| \mu_d(\mathrm{d}z) \to 0,$$

$$\frac{1}{\varepsilon}\int_{F\setminus F(\varepsilon)}|\lambda(z)-\lambda(p_F(z))|\mu_d(\mathrm{d}z)\to 0,$$

the statistic $L_n(F, F(\varepsilon))$ converges in distribution, under alternatives $F(\varepsilon)$, to the random variable (23), while the Poisson process N_∞ has intensity measure

$$\Lambda_{F,\mathrm{alt}}(B) = S\left[\int_{B^+}\tilde{\lambda}(x)\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x)+\int_{B^-}\lambda(x)\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x)\right].$$

For the *proof*, note that if the true change-set is $F(\varepsilon)$, then the measure Λ_n will change to

$$\Lambda_{n,\mathrm{alt}}(A) = n\int_{A\cap F(\varepsilon)}\tilde{\lambda}(z)\mu_d(\mathrm{d}z)+n\int_{A\cap F(\varepsilon)^*}\lambda(z)\mu_d(\mathrm{d}z).$$

Mapped by the local magnification map onto the normal cylinder Σ and then projected onto Γ, it will converge to $\Lambda_{F,\mathrm{alt}}$, which, therefore, is the intensity measure of N_∞ under the alternatives.

The corollary implies, by the way, that the expected value of (23) will change by the quantity

$$S\int_B \ln\frac{\tilde{\lambda}}{\lambda}(x)[\tilde{\lambda}(x)-\lambda(x)]\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x).$$

Therefore, the local likelihood ratio test will have some power for alternatives, converging to the null hypothesis with the rate $\varepsilon = 1/n$, a property it shares with the change-point problems on the line (see , e.g., Brodsky and Darkhovsky 1993).

4.3 Convergence of the log-likelihood ratio (second version)

The situation with the asymptotic behavior of the likelihood (5) is, in many respects, similar. First, let us replace the assumption that the number n of observations $(X_i, Y_i)_{i=1}^n$ is fixed and assume that it is a Poisson random variable ν with expected value n. This is not an important change in the present context, but it makes the process in (4), with n replaced by ν, again a Poisson process N_ν. Its intensity measure under the null hypothesis is

$$\mathbb{E}N_\nu(y, A) = n\tilde{P}(y)Q(A\cap F)+nP(y)Q(A\cap F^*). \tag{24}$$

If we consider only the part $N_\nu(\infty, A)$ of $N_\nu(y, A)$, which leaves the marks Y_i out, then all what was said about local processes in Theorems 3 and 4 will still be valid. Restricted to the strip $(\partial F)_{\varepsilon T}$, the process $N_\nu(\infty,\cdot)$ will be Poisson process with intensity measure now equal to measure nQ on $(\partial F)_{\varepsilon T}$; splitting this process again into asymptotically negligible part $N_{\nu,s}(\infty,\cdot)$ and the leading part $N_\nu(\infty,\cdot)-N_{\nu,s}(\infty,\cdot)$, we again can map this part to Poisson process on Γ_T. Denote it by $N_{\nu,\varepsilon}$. The intensity measure of this Poisson process will, certainly, be the image on nQ, so that

$$\Lambda_{n,\varepsilon}(B(\varepsilon)) = nQ(A(\varepsilon)),$$

and, as in Theorems 3 and 4, we conclude that the limit of $\Lambda_{n,\varepsilon}$ is the measure

$$\Lambda(B) = S\int_B q(x)dt\mathcal{H}^{d-1}(\mathrm{d}x),$$

where $q(x)$ is the density of Q at $x \in \partial F$. Hence $N_{\nu,\varepsilon}$ converges in distribution to the Poisson process on Γ, with the intensity measure Λ.

The dependence of Λ on F is now "weak"—there is no need to split B into B^+ and B^-.

Going back to $N_\nu(y, A)$, note that each location X_i, which projects to a regular point on ∂F, will be mapped to the pair (ζ_i, Ξ_i) in Γ, with $p_F(X_i) = \Xi_i$, but the mark Y_i will not be altered. The conditional distribution of the marks, given ζ_i, is

$$\tilde{P}(dy)^{\mathbf{1}\{\zeta_i \geq 0\}} P(dy)^{\mathbf{1}\{\zeta_i < 0\}}.$$

Therefore, as soon as the process $N_{\nu,\varepsilon}$ on Γ, based on the pairs (ζ_i, Ξ_i), converges to a Poisson process, the point process based on triples (Y_i, ζ_i, Ξ_i),

$$N_{\nu,\varepsilon}(y, B) = \sum_{i=1}^{\nu(\Gamma)} \mathbf{1}\{Y_i \leq y\}\mathbf{1}\{(\zeta_i, \Xi_i) \in B\}, \quad \text{with} \quad \nu(\Gamma) = N_{\nu,\varepsilon}(\infty, \Gamma),$$

will converge to a Poisson process N_∞ on $\mathbb{R} \times \Gamma$, and its intensity measure will be

$$\Pi((-\infty, y] \times B) = S\int_\Gamma \mathbf{1}\{(t, x) \in B\} P(y)^{\mathbf{1}\{t \geq 0\}} \tilde{P}(y)^{\mathbf{1}\{t \leq 0\}} q(x)\mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x).$$

The statement on the convergence in distribution of the log-likelihood statistics follows, here we abbreviate (t, x) by z.

Corollary 7 (*Asymptotic null distribution of statistics* (5)) *Under the conditions of Theorem 4, the log-likelihood statistic $L_n(F, F(\varepsilon))$ of* (5) *converges in distribution to the random variable*

$$\int_{\mathbb{R}\times B} \ln \frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) N_\infty(\mathrm{d}y, \mathrm{d}z), \tag{25}$$

where the intensity measure of N_∞ is Π.

The case when both $\tilde{P}$ and P are Bernoulli distributions is illustrated in Fig. 2. The data points on the right-hand side of Fig. 3 illustrate the extreme case of degenerate Bernoulli distributions, with $\tilde{p}(1) = 1$ and $p(1) = 0$.

4.4 Central limit theorems

In the situation, where ε is asymptotically not as small as $1/n$, but larger, so that $n\varepsilon \to \infty$, if the intensities $\tilde{\lambda}$ and λ in the first formulation of the change-set problem, and the distributions $\tilde{P}$ and P in the second formulation, remain fixed, the power of the test based on the statistic $L_n(F, F(\varepsilon))$ will converge to 1, making discrimination between F and $F(\varepsilon)$ asymptotically obvious. In order to stay within the more difficult situation when some power, although not power 1, can be retained, the intensities $\tilde{\lambda}$ and λ or the distributions $\tilde{P}$ and P should be allowed to change with n and approach each other. If this mutual convergence is not too quick, the tests based on (2) and (5) will have some power against such alternatives.

If $n\varepsilon \to \infty$, the number of jump points of N_n will increase unboundedly and the limit theorems for this process should be Gaussian and not Poisson. For the first change-set problem, the Gaussian limit theorems for N_n have been studied in the neighborhood of convex bodies in Einmahl and Khmaladze (2011). Here, let us consider the second change-set problem.

Suppose that the distribution $\tilde{P}$ depends on some one-dimensional parameter δ in a smooth way, such that the expansion analogous to (1) is true,

$$\ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) = \delta l(y) - \frac{\delta^2}{2} l^2(y) + o_P(\delta^2), \tag{26}$$

and the function l is such that that

$$\int l(y) P(\mathrm{d}y) = 0, \quad \int l^2(y) P(\mathrm{d}y) < \infty.$$

Thus, as $\delta = \delta_n \to 0$, the distribution $\tilde{P}_\delta$ approaches the distribution P from the "direction" l. We want now to find the rates of $\delta = \delta_n$ and $\varepsilon = \varepsilon_n$ such that the statistic (5) converges to a proper random variable.

Let us subtract from N_ν in the right-hand side of (5) its expected value (24) and add it, so that the non-central term of our statistic $L_\nu(F, F(\varepsilon))$ will be

$$n \int_{\mathbb{R}} \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) P(\mathrm{d}y)\ Q(F(\varepsilon)\setminus F) - n \int_{\mathbb{R}} \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) \tilde{P}(\mathrm{d}y)\ Q(F\setminus F(\varepsilon)). \tag{27}$$

Using expansion (26), we can evaluate the expected values of $\ln(d\tilde{P}_\delta/dP)(Y)$,

$$\begin{aligned}
\int \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) P(\mathrm{d}y) &= -\frac{\delta^2}{2} \int l^2(y) P(\mathrm{d}y) + o(\delta^2), \\
\int \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) \tilde{P}(\mathrm{d}y) &= \int \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) P(\mathrm{d}y) \\
&= \frac{\delta^2}{2} \int l^2(y) P(\mathrm{d}y) + o(\delta^2).
\end{aligned}$$

These relationships are easy to establish heuristically, while their formal justification in broader context can be found, for example, in Janssen (1995), Janssen (2000) and van der Vaart (1998).

Thus, the shift part (27) is of order $n\delta^2\varepsilon$, and it is necessary that this quantity stays bounded. Hence, we assume that

$$n \to \infty, \varepsilon \to 0, \delta \to 0, \quad \text{such that} \quad n\delta^2\varepsilon = S < \infty.$$

Now consider the central part. If we put

$$z_\nu(y, A) = \frac{1}{\sqrt{n\varepsilon}} \left(N_\nu(y, A) - n\left[\tilde{P}(y)Q(A \cap F) + P(y)Q(A \cap F^c \right] \right),$$

where $A \subset (\partial F)_{\varepsilon T}$, then this centered part can be re-written as

$$\sqrt{n\varepsilon}\left(\int_{\mathbb{R}\times(F(\varepsilon)\setminus F)} \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) z_\nu(\mathrm{d}y, \mathrm{d}x) - \int_{\mathbb{R}\times(F\setminus F(\varepsilon))} \ln \frac{\mathrm{d}\tilde{P}_\delta}{\mathrm{d}P}(y) z_\nu(\mathrm{d}y, \mathrm{d}x) \right).$$

An asymptotically equivalent form of this expression is

$$\sqrt{n\varepsilon}\delta \left(\int_{\mathbb{R}\times(F(\varepsilon)\setminus F)} l(y) z_\nu(\mathrm{d}y, \mathrm{d}x) - \int_{\mathbb{R}\times(F\setminus F(\varepsilon))} l(y) z_\nu(\mathrm{d}y, \mathrm{d}x) \right).$$

Since z_ν is the centered and normalized form of the Poisson process N_ν in $\mathbb{R}\times(\partial F)_{\varepsilon T}$, it retains the property of having independent increments on disjoint sets. This implies that the two integrals above are independent random variables. The variance of the first integral (including the factor $\sqrt{n\varepsilon}\delta$) is

$$S \int l^2(y) P(\mathrm{d}y) \frac{Q(F(\varepsilon)\setminus F)}{\varepsilon},$$

while for the variance of the second integral we obtain

$$\begin{aligned} &S \int l^2(y) \tilde{P}(\mathrm{d}y) \frac{Q(F\setminus F(\varepsilon))}{\varepsilon} \\ &\quad = S \int l^2(y) P(\mathrm{d}y) \frac{Q(F\setminus F(\varepsilon))}{\varepsilon} + o_P(1), \quad n \to \infty. \end{aligned}$$

We see that both variances are finite as soon as S is fixed.

To come now to the object which will be asymptotically Gaussian, recall that the mapping of N_ν onto $\mathbb{R} \times \varGamma$ produces the Poisson process $N_{\nu,\varepsilon}$ introduced in the previous subsection, and its intensity measure is

$$\varPi_{n\varepsilon,\delta}(y, B) = n\varepsilon \int_B \tilde{P}_\delta(y)^{\mathbf{1}\{t\geq 0\}} P(y)^{\mathbf{1}\{t<0\}}\, \mathrm{d}t\, \mathcal{H}^{d-1}(\mathrm{d}x), \tag{28}$$

but now $n\varepsilon \to \infty$. The image of the normalized process z_ν, which we denote by $z_{\nu,\varepsilon}$, is the process $N_{\nu,\varepsilon}$ centered by the measure $\Pi_{n\varepsilon,\delta}$ and normalized by $\sqrt{n\varepsilon}$,

$$z_{\nu,\varepsilon}(y, B) = \frac{1}{\sqrt{n\varepsilon}}\left[N_{\nu,\varepsilon}(y, B) - \Pi_{n\varepsilon,\delta}(y, B)\right],$$

where B is a Borel set in Γ. As in the similar case treated in Einmahl and Khmaladze (2011), it is natural to expect now that this normalized Poisson process $z_{\nu,\varepsilon}$ converges in distribution to a Brownian motion, as the intensity increases. However, we need to take one delicate point into account.

Consider the class $\mathcal{A}$ of set-valued functions

$$\{A(\varepsilon) : A(\varepsilon) \subset (\partial F)_{\varepsilon T}, 0 \le \varepsilon \le 1\},$$

which are differentiable at ∂F, for $\varepsilon = 0$. This, or the corresponding class of $\{F(\varepsilon), 0 \le \varepsilon \le 1\}$, is the class of alternatives to be chosen for the statistical analysis. For each $\varepsilon > 0$, let $\mathcal{A}_\varepsilon$ be the class of values of our set-valued functions $\mathcal{A}_\varepsilon = \{A(\varepsilon) : A(\varepsilon) \subset (\partial F)_{\varepsilon T}\}$ at a given ε and let $\mathcal{B}_\varepsilon$ be the class of images of the sets $A(\varepsilon)$ under the local magnification map τ_ε onto Γ,

$$\mathcal{B}_\varepsilon = \{B(\varepsilon) = \tau_\varepsilon(A(\varepsilon)), A(\varepsilon) \in \mathcal{A}_\varepsilon\}.$$

Then, a functional limit theorem for z_ν on the class of shrinking sets $\mathcal{A}_\varepsilon$ is not directly possible. Instead, we may consider a functional limit theorem for $z_{\nu,\varepsilon}$ on the class $\mathcal{B}_\varepsilon$. However, this class changes with ε.

In this situation, the following approach was suggested in Einmahl and Khmaladze (2011). Between the sets C and C' in $\mathbb{R} \times \Gamma$, introduce a distance as $d(C, C') = (\mu_1 \otimes \mu_1 \otimes \mathcal{H}^{d-1})(C \Delta C')$. Using this distance one can introduce the Hausdorff distance between classes of such sets. Now assume that

a) there is a fixed class $\mathcal{B}$, such that the Hausdorff distance between $\mathcal{B}_\varepsilon$ and $\mathcal{B}$ tends to zero as $\varepsilon \to 0$, and
b) the class $\mathcal{B}$ is a Donsker class (see, e.g., van der Vaart and Wellner 1996, Sec. 2.11.3).

Then it can be proved that for $\eta_n > 0$

$$\sup_{B(\varepsilon)\in\mathcal{B}_\varepsilon, B\in\mathcal{B}, d(B(\varepsilon),B)\le\eta_n} |z_{\nu,\varepsilon}(B(\varepsilon)) - z_{\nu,\varepsilon}(B)| \xrightarrow{P} 0, \quad \text{as } \eta_n \to 0,$$

see Einmahl and Khmaladze (2011), for details. The class of functionals of z_ν, or rather functionals of $z_{\nu,\varepsilon}$, which can then be used as various test statistics, also needed some discussion in Einmahl and Khmaladze (2011), because these functionals can change with n as well. The final step, the convergence

$$z_{\nu,\varepsilon} \xrightarrow{d} z_\infty \quad \text{on} \quad \mathcal{B},$$

where z_∞ is a set-parametric Brownian motion on $\mathcal{B}$, is then relatively well understood. Since the conditions needed here, like metric entropy condition, exponential inequalities, and others, are not specific for the subject of this review, but will require some space, we abstain from a formal statement of the general central limit theorem here, but state one special case in the proposition below. We only mention that the variance measure of z_∞ is given by Π,

$$Ez_\infty^2(y, B) = \Pi(y, B).$$

Let us come back to the situation we are mainly interested in this section, with one given $F(\varepsilon)$, i.e., one alternative, and one test statistic $L_\nu(F, F(\varepsilon))$. For this statistic we can formulate the following result.

Proposition 8 *If $F(\varepsilon)$ is differentiable at F and expansion (26) is satisfied, then for $n\varepsilon\delta^2 = S$,*

$$\begin{aligned} L_\nu(F, F(\varepsilon)) \xrightarrow{d} S &\left[\int_{\mathbb{R}\times B^+} l(y) z_\infty(\mathrm{d}y, \mathrm{d}x) - \int_{\mathbb{R}\times B^-} l(y) z_\infty(dy, dx)\right] \\ - S &\left[\int_{\mathbb{R}} l^2(y) P(\mathrm{d}y) \int_{B^+} q(x)\mathcal{H}^{d-1}(\mathrm{d}x) + \int_{\mathbb{R}} l^2(y) P(\mathrm{d}y) \int_{B^-} q(x)\mathcal{H}^{d-1}(\mathrm{d}x)\right]. \end{aligned}$$

We remark that, for the asymptotic normality of the statistic $L_\nu(F, F(\varepsilon))$, we do not need the functional convergence. The one-dimensional convergence for one set-valued function $A(\varepsilon) = F(\varepsilon)\Delta F$ is sufficient. Moreover, one could prove this asymptotic normality of $L_\nu(F, F(\varepsilon))$, even without the notion of derivative of $F(\varepsilon)\Delta F$. However, if we want to understand properly why the distribution of $L_\nu(F, F(\varepsilon))$ is asymptotically Gaussian, and more importantly, what would be the asymptotic “structure” of many statistics $L_\nu(F, F(\varepsilon))$ for many deviations $F(\varepsilon)$ from F, the notion of fold-up derivatives of these sets and the notion of Brownian motion on these derivatives is indeed necessary.

5 Further remarks and outlook

In this final section, we collect some remarks on possible extensions and variants of the differentiability approach which we have presented.

5.1 Fold-up derivatives versus generalized functions

In Introduction, we already considered the tempting possibility to use the difference

$$\mathbf{1}\{z \in F(\varepsilon)\} - \mathbf{1}\{z \in F(0)\} = \mathbf{1}\{z \in F(\varepsilon)\setminus F(0)\} - \mathbf{1}\{z \in F(0)\setminus F(\varepsilon)\},$$

divide it by ε and consider the limit to describe the shrinkage of $F(\varepsilon)\Delta F(0)$ as $\varepsilon \to 0$. Indeed, if $F(\varepsilon)$ is differentiable at $F(0) = F$ with the fold-up derivative B, and if φ is from the class $C_c^\infty(\mathbb{R}^d)$ of smooth functions with compact support on $\mathbb{R}^d$ (*test* *functions*), then both integrals in

$$\frac{1}{\varepsilon}\int_{\mathbb{R}^d}\varphi(z)\mathbf{1}\{z\in F(\varepsilon)\backslash F\}\mu_d(\mathrm{d}z)-\frac{1}{\varepsilon}\int_{\mathbb{R}^d}\varphi(z)\mathbf{1}\{z\in F\backslash F(\varepsilon)\}\mu_d(\mathrm{d}z)$$

will converge to limits $g_{B^+}(\varphi)$, $g_{B^-}(\varphi)$, which yields generalized functions g_{B^+}, g_{B^-} on $\mathbb{R}^d$, concentrated on ∂F. The proof of this convergence uses the Steiner formula and the assumption that $F(\varepsilon)$ is differentiable. In fact an asymptotic analysis, similar to the arguments which led to Theorem 1, shows that g_{B^+} and g_{B^-}, which are formally linear functionals on $C_c^\infty(\mathbb{R}^d)$, can actually be realized as measurable, nonnegative functions on ∂F, such that

$$g_{B^+}(\varphi)=\int_{\partial F}\varphi(x)\gamma_{B^+}(x)\mathcal{H}^{d-1}(\mathrm{d}x),\;\; g_{B^-}(\varphi)=\int_{\partial F}\varphi(x)\gamma_{B^-}(x)\mathcal{H}^{d-1}(\mathrm{d}x).$$

Moreover, the functions γ_{B^+} and γ_{B^-} are related to the positive and negative parts $B^+\subset\Gamma^+$ and $B^-\subset\Gamma^-$ of the derivative B and, for $\mathcal{H}^{d-1}$-almost all $x\in\partial F$, the values $\gamma_{B^+}(x)$, $\gamma_{B^-}(x)$ are given by the lengths of the intersections $B\cap(\mathbb{R}_+\times\{x\})$ and $B\cap(\mathbb{R}_-\times\{x\})$.

It may look very natural to use the pair g_{B^+}, g_{B^-} or even their difference $g_B = g_{B^+}-g_{B^-}$ to describe the limiting processes, whether in Poisson or Gaussian asymptotics. However, this is not appropriate. As the arguments below show, there are infinitely many fold-up derivatives B, rather distinct from the point of view of the local processes, but which correspond to the same pair g_{B^+}, g_{B^-}. In other words, the language of generalized functions is too coarse for our needs in the change-set problem.

This is already visible in a one-dimensional situation. Indeed, let $F=[-a,0]$ and $F(\varepsilon)=F_1(\varepsilon)=[-a,\varepsilon]$, for some $a>0$. Thus, $F_1(\varepsilon)\backslash F=(0,\varepsilon]$, while $F\backslash F_1(\varepsilon)=\emptyset$. The local magnification map will produce the set $(0,1]\times\{0\}\subset\Gamma$, see Fig. 7, which does not depend on ε and is the fold-up derivative of F_1. If we now take $F_2(\varepsilon)=[-a,0]\cup(\varepsilon,2\varepsilon]$, then the fold-up derivative will be the set $(1,2]\times\{0\}$.

If now $X_1,\dots,X_n$ are independent uniform random variables on, say, $[0,1]$, then the classical local binomial process

$$N_n([0,t\varepsilon])=\sum\mathbf{1}\{X_i\le t\varepsilon\}$$

is mapped into $N_{n,\varepsilon}([0,t])$, which for $n\varepsilon\to1$ converges to a Poisson process with intensity 1, or expected value t, and the number of points in $F_1(\varepsilon)\backslash F$ and $F_2(\varepsilon)\backslash F$, which are $N_n((0,\varepsilon])$ and $N_n((\varepsilon,2\varepsilon])$, will be mapped to $N_{n,\varepsilon}((0,1])$ and $N_{n,\varepsilon}((1,2])$ and both converge to Poisson random variables with the same parameter 1. The differences $F_1(\varepsilon)\backslash F$ and $F_2(\varepsilon)\backslash F$ are disjoint and so are their fold-up derivatives. Hence the two limiting Poisson random variables are independent. This aspect is, however, lost as soon as we turn to generalized functions. Both integrals

$$\frac{1}{\varepsilon}\int_0^{t\varepsilon}\varphi(z)\mathbf{1}\{z\in(0,t\varepsilon]\}\mathrm{d}z\quad\text{and}\quad\frac{1}{\varepsilon}\int_{t\varepsilon}^{2t\varepsilon}\varphi(z)\mathbf{1}\{z\in[t\varepsilon,2t\varepsilon]\}\mathrm{d}z$$

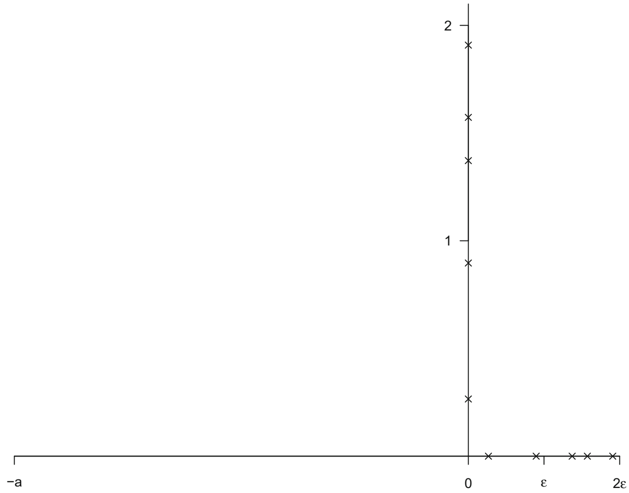

Fig. 7 The random points in $(0, \varepsilon]$ and $(\varepsilon, 2\varepsilon]$ on the X-axis are mapped by the local magnification map into random points in $(0, 1]$ and $(1, 2]$ on the Y-axis

converge to $t\varphi(0)$, and thus define the same generalized function at the boundary point $z = 0$. This example can be easily extended to the d-dimensional situation.

5.2 The change-set problem and chimeric alternatives

The concept of *chimeric alternatives* was introduced in Khmaladze (1998). These are the alternatives, which remain on a certain non-diminishing Hellinger distance from the hypothetical distribution, but which, as far as the empirical process is concerned, are asymptotically undetectable.

More exactly, consider a sequence of distributions P_n, alternatives to the distribution P, with density with respect to the distribution P of the form

$$\sqrt{\frac{\mathrm{d}P_n}{\mathrm{d}P}(z)} = 1 + \frac{1}{2\sqrt{n}} h_n(z), \tag{29}$$

where

$$\lim_{n\to\infty} \int h_n^2(z) P(\mathrm{d}z) = const > 0,$$

while

$$\int h_n(z)\phi(z) P(\mathrm{d}z) \to 0 \quad \text{for any fixed } \phi \in L_2(P).$$

The last property of h_n says that this sequence runs away from the space, it does not have a limiting point in $L_2(P)$. There are many ways of visualizing such sequences.

One is when functions h_n oscillate more and more with increasing n. One other, the *spike alternatives* in Khmaladze (1998), is when the functions h_n are concentrated on subsets of P-probabilities which converge to zero as $n \to \infty$.

Given a sequence of i.i.d. random variables $\{Z_i\}_{i=1}^n$, a function-parametric empirical process to test whether P is indeed the distribution of each Z_i is defined as

$$v_n(\phi) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\phi(Z_i) - \int \phi(z)P(\mathrm{d}z)\right],\ \phi \in \Phi \subset L_2(P).$$

The class Φ of square integrable functions, on which $v_n(\phi)$ is considered, is a part of the setting and depends on the user. Functionals from this process, like, for example, $\sup_{\phi\in\Phi}|v_n(\phi)|$ are used as test statistics. In order that v_n converge in distribution to a Brownian bridge, the class Φ has to satisfy certain metric entropy conditions, but we simply assume that these conditions are satisfied. Moreover, we are willing to assume that all ϕ are bounded functions. What we want to clarify here is what will be the distribution of our empirical process under a chimeric alternative. Under the null hypothesis, $Ev_n(\phi) = 0$ and $Ev_n^2(\phi) = \|\phi\|_P^2$ and $v_n(\phi)$ is asymptotically normal with these parameters. Under a chimeric alternative

$$\begin{aligned} Ev_n(\phi) &= \int \phi(z)\,(P_n(\mathrm{d}z) - P(\mathrm{d}z)) \\ &= \int \phi(z)\left(h_n(z) + \frac{1}{4\sqrt{n}}h_n^2(z)\right)P(\mathrm{d}z) \to 0, \end{aligned}$$

as it follows from the definition of chimeric alternatives and boundedness of ϕ; as a consequence

$$\begin{aligned} Ev_n^2(\phi) &= \int \phi^2(z)P(\mathrm{d}z) + \int \phi^2(z)\left(h_n(z) + \frac{1}{4\sqrt{n}}h_n^2(z)\right)P(\mathrm{d}z) \\ &= \int \phi^2(z)P(\mathrm{d}z) + o(1), \end{aligned} \tag{30}$$

and therefore under chimeric alternatives the random variable $v_n(\phi)$, for any ϕ, has asymptotically the same Gaussian distribution as under the hypothesis. So, tests based on v_n will asymptotically have no power.

Now let us see what is the corresponding situation in the change-set problem. Consider, for example, its second formulation. As it can be seen from Sect. 1.4, the square root of the likelihood ratio of distributions of each pair (X_i, Y_i) under $F(\varepsilon)$ and under F is

$$\begin{aligned} &\left(\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y)\right)^{(\mathbf{1}(x\in F(\varepsilon))-\mathbf{1}(x\in F))/2} \\ &\quad = 1 + \mathbf{1}(x \in A(\varepsilon))\left(\left(\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y)\right)^{(\mathbf{1}(x\in A^+(\varepsilon))-\mathbf{1}(x\in A^-(\varepsilon)))/2} - 1\right) \end{aligned} \tag{31}$$

and comparison with (29) shows that here

$$h_n(x, y) = \sqrt{n}\mathbf{1}(x \in A(\varepsilon)) \left(\left(\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) \right)^{(\mathbf{1}(x \in A^+(\varepsilon)) - \mathbf{1}(x \in A^-(\varepsilon)))/2} - 1 \right).$$

This function is nonzero only on the shrinking set $A(\varepsilon)$ and its L_2-norm under the null distribution is

$$\begin{aligned} & nQ(A^+(\varepsilon)) \int_{\mathbb{R}} \left(\left(\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) \right)^{1/2} - 1 \right)^2 \mathrm{d}P(y) \\ & + nQ(A^-(\varepsilon)) \int_{\mathbb{R}} \left(\left(\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}(y) \right)^{-1/2} - 1 \right)^2 \mathrm{d}\tilde{P}(y). \end{aligned}$$

We already know that if $F(\varepsilon)$ is differentiable at F and its fold-up derivative is B, and if $n\varepsilon \to S$, then

$$nQ(A^+(\varepsilon)) \to S \int_{B^+} \mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x) \quad \text{and} \quad nQ(A^-(\varepsilon)) \to S \int_{B^-} \mathrm{d}t\mathcal{H}^{d-1}(\mathrm{d}x),$$

and therefore the L_2-norm of h_n has a positive limit. This implies that the set $F(\varepsilon)$ indeed creates a chimeric alternative to F, and that the conventional approach based on empirical processes would not be useful.

5.3 Boundary sets

As a second class of sets, the *boundary sets* F are considered in Khmaladze and Weil (2014). These are non-empty compact sets $F \subset \mathbb{R}^d$ with $F = \partial F$ and $\mu_d(F) = 0$. We may also assume $\mu_{d-1}(F) > 0$. Hence $F^* = \mathbb{R}^d$ and $r_- = 0$. For a boundary set F, the Steiner formula (12) holds, but consists only of the outside part. Consequently, we only need the upper part Σ_+ of the normal cylinder Σ. The definition of the fold-up derivative follows the same lines as in the solid case, the distinction between F and ∂F is not necessary here. The support measure $\Theta_{d-1}(F, \cdot)$ satisfies

$$\Theta_{d-1}(F, \cdot) = \int_{\mathrm{reg}(F)} [\mathbf{1}\{(x, \nu(F, x)) \in \cdot\} + \mathbf{1}\{(x, -\nu(F, x)) \in \cdot\}] \mathcal{H}^{d-1}(\mathrm{d}x), \qquad (32)$$

see (Hug et al. 2004, Prop. 4.1). Notice that there are still topological phenomena, also for boundary sets, which are counter-intuitive. One expects that in most points x of a boundary set F there are two normals $\nu_F(x), -\nu_F(x)$, but there are examples where $\mathcal{H}^{d-1}(\partial^1 F) > 0$ and $\mathcal{H}^{d-1}(\partial^2 F) = 0$. Here, $\partial^i F$ is the set of boundary points with precisely i normals, $i = 1, 2$.

Since the values $r_+(x,u)$, $r_+(x,-u)$ are different, in general, the normal cylinder Σ cannot be identified with the cylinder $\Gamma = \mathbb{R} \times \partial F$ in a natural way, but we would need two copies Γ_1^+, Γ_2^+ of the upper part of Γ.

Otherwise, the properties of the derivative, Theorem 1 and most of the considerations made above for solid sets carry over to boundary sets with obvious modifications (see Khmaladze and Weil 2014, for details).

5.4 Variations of solid sets

With respect to the local Steiner formula, various set classes have been considered in the literature, which generalize convex sets one one hand and are not as general as solid sets on the other hand. The purpose is to see which additional structure of the support measures $\Theta_i(F,\cdot)$, $i = 0,\dots,d-1$, (often called *curvature measures* or *Lipschitz–Killing curvatures*) can be obtained if the sets F have further regularity properties. One example is the question, for which sets F the support measures satisfy (locally or globally) a kinematic formula like the classical principal kinematic formula in integral geometry or the Crofton formula, see Schneider and Weil (2008), Chapter 5. In this direction, the most general set class at the moment are the DC-sets of Fu et al. (2017). Another question is, whether or under which additional conditions on F the support measures are connected to further local and global quantities in geometry, like the lower-order Hausdorff measures, the Minkowski content, or the perimeter. Here, the paper Ambrosio et al. (2008) gives a good account of the various relations.

From the probabilistic point of view, it is a natural question, which of the geometric results, relevant to set-differentiation, hold for graphs or subgraphs of trajectories of stochastic processes, like the Wiener process. However, we are not aware of investigations in this direction.

Acknowledgements The authors thank two anonymous referees for their useful remarks on a previous version of this work.

References

Ambrosio, L., Colesanti, A., Villa, E. (2008). Outer Minkowski content for some classes of closed sets. *Mathematische Annalen*, *342*, 727–748.

Artstein, Z. (1995). A calculus of set-valued maps and set-valued evolution equations. *Set-Valued Analysis*, *3*, 213–261.

Artstein, Z. (2000). Invariant measures of set-valued maps. *Journal of Mathematical Analysis and Applications*, *252*, 696–709.

Aubin, J.-P. (1981). Contingent derivatives of set-valued maps and existence of solutions to non-linear inclusions and differential inclusions. In L.Nachbin (Ed.), *Mathematical Analysis and Applications, Part A* (pp. 160–232). *Advances in Mathematics: Supplimentary Studies, 7A*. New York: Academic Press

Aubin, J.-P., Cellina, A. (1984). *Differential Inclusions, Set-Valued Maps and Viability Theory*. Grundlehren der mathematischen Wissenschaften. Berlin: Springer

Aubin, J.-P., Frankowska, H. (1990). *Set-valued Analysis*. Basel: Birkhäuser.

Baíllo, A., Cuevas, A. (2001). On the estimation of a star-shaped set. *Advances in Applied Probability*, *33*, 1–10.

Bernardin, F. (2003). Multivalued stochastic differential equations: convergence of a numerical scheme. *Set-Valued Analysis*, *11*, 393–415.
Borwein, J. M., Zhu, Q. J. (1999). A survey of sub-differential calculus with applications. *Nonlinear Analysis*, *38*, 687–773.
Brodsky, B. E., Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-Point Problems*. Dordrecht: Kluwer Academic Publishers.
Carlstein, E., Krishnamoorthy, C. (1992). Boundary estimation. *Journal of the American Statististical Association*, *87*, 430–438.
Cramér, H. (1999). *Mathematical Methods of Statistics, 19th printing*. Princeton: Princeton University Press.
Cuevas, A., Fraiman, R., Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Annals of Statistics*, *35*, 1031–1051.
Daley, D., Vere-Jones, D. (2005). *An Introduction to the Theory of Point Processes,* 2nd ed. 2003, Corrected 2nd printing. New York: Springer
Deheuvels, P., Mason, D. M. (1995). Nonstandard local empirical processes indexed by sets. *Journal of Statistical Planning and Inference*, *45*, 91–112.
Einmahl, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Processes and Applications*, *70*, 31–58.
Einmahl, U., Mason, D. M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields*, *107*, 283–311.
Einmahl, J. H. J., Khmaladze, E. (2011). Central limit theorems for local empirical processes near boundaries of sets. *Bernoulli*, *17*, 545–561.
Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society*, *93*, 418–491.
Fu, J. H. G., Pokorny, D., & Rataj, J. (2017). Kinematic formulas for sets defined by differences of convex functions. *Advances in Mathematics*, *311*, 796–832.
Gruber, P. M. (1993). History of convexity. In P. M. Gruber J. M. Wills (Eds.), *Handbook of Convex Geometry* (Vol. A, pp. 3–15). Amsterdam: North Holland.
Hajek, Ja, Shidak, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.
Hug, D., Last, G., & Weil, W. (2004). A local Steiner-type formula for general closed sets and applications. *Mathematische Zeitschrift*, *246*, 237–272.
Ivanoff, B. G., Merzbach, E. (2010). Optimal detection of a change-set in a spatial Poisson process. *Annals of Applied Probability*, *20*, 640–659.
Janssen, A. (1995). Principal component decomposition of non-parametric tests. *Probability Theory and Related Fields*, *101*, 193–209.
Janssen, A. (2000). Global power functions of goodness of fit tests. *Annals of Statistics*, *28*, 239–253.
Karr, A. F. (1991). *Point Processes and their Statistical Inference* (2nd ed.). New York: Marcel Dekker.
Khmaladze, E. (1998). Goodness of fit tests for “chimeric” alternatives. *Statistica Neerlandica*, *52*, 90–111.
Khmaladze, E. (2007). Differentiation of sets in measure. *Journal of Mathematical Analysis and Applications*, *334*, 1055–1072.
Khmaladze, E., Toronjadze, N. (2001). On the almost sure coverage property of Voronoi tessellation: the R^1 case. *Advances in Applied Probability*, *33*, 756–764.
Khmaladze, E., Weil, W. (2008). Local empirical processes near boundaries of convex bodies. *Annals of the Institute of Statistical Mathematics*, *60*, 813–842.
Khmaladze, E., Weil, W. (2014). Differentiation of sets - The general case. *Journal of Mathematical Analysis and Applications*, *413*, 291–310.
Khmaladze, E., Mnatsakanov, R., Toronjadze, N. (2006a). The change-set problem for Vapnik-Červonenkis classes. *Mathematical Methods of Statistics*, *15*, 224–231.
Khmaladze, E., Mnatsakanov, R., Toronjadze, N. (2006b). The change-set problem and local covering numbers. *Mathematical Methods of Statistics*, *15*, 289–308.
Kim, B. K., Kim, J. H. (1999). Stochastic integrals of set-valued processes and fuzzy processes. *Journal of Mathematical Analysis and Applications*, *236*, 480–502.
Korostelev, A. P., Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstructions*. Lecture Notes in Statistics vol. 82. New York: Springer.
Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.
Landau, L. D., Lifshitz, E. M. (1987). *Fluid Mechanics* (2nd ed.). Oxford-Burlington: Butterworth-Heinemann.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer.
Le Cam, L., Lo Yang, G. (2000). *Asymptotics in Statistics*. New York: Springer.
Lemaréchal, C., Zowe, J. (1991). The eclipsing concept to approximate a multi-valued mapping. *Optimization*, *22*, 3–37.
Mammen, E., Tsybakov, A. B. (1995). Asymptotic minimax recovery of sets with smooth boundaries. *Annals of Statistics*, *23*, 502–524.
Motzkin, T. (1935). Sur quelques propriétés charactéristiques des ensemble convexes. Atti della Accademia Nazionale dei Lincei. *Classe di Scienze Fisiche, Matematiche e Naturali*, *21*, 562–567.
Müller, H. G., Song, K. S. (1996). A set-indexed process in a two-region image. *Stochastic Processes and Applications*, *62*, 87–101.
Oosterhoff, J., van Zwet, W.R. (2012) A note on contiguity and Hellinger distance. In *Selected Works of Willem van Zwet* (pp. 63–72). New York: Springer
Penrose, M. D. (2007). Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli*, *13*, 1124–1150.
Pflug, G Ch. (1996). *Optimization of Stochastic Models*. Dordrecht: Kluwer Academic Publishers.
Reitzner, M., Spodarev, E., Zaporozhets, D. (2012). Set reconstruction by Voronoi cells. *Advances in Applied Probability*, *44*, 938–953.
Ripley, B. D., Rasson, J.-P. (1977). Finding the edge of a Poisson forest. *Journal of Applied Probability*, *14*, 483–491.
Schneider, R. (1979). Bestimmung konvexer Körper durch Krümmungsmaße. *Commentarii Mathematici Helvetici*, *54*, 42–60.
Schneider, R. (2013). *Convex Bodies: the Brunn-Minkowski Theory*, 2nd expanded ed., Encyclopedia of Mathematics and its Applications vol. 44. Cambridge, UK: Cambridge University Press
Schneider, R., Weil, W. (2008). *Stochastic and Integral Geometry*. Berlin: Springer.
Thäle, C., Yukich, J. E. (2016). Asymptotic theory for statistics of the Poisson-Voronoi approximation. *Bernoulli*, *22*, 2372–2400.
van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
van der Vaart, A., Wellner, J. A. (1996). *Weak Convergence of Empirical processes*. New York: Springer.
Weisshaupt, H. (2001). A measure-valued approach to convex set-valued dynamics. *Set-Valued Analysis*, *9*, 337–373.
Weyl, H. (1939). On the volume of tubes. *American Journal of Mathematics*, *61*, 461–472.