Variational Analysis of Functionals of a Poisson Process

Ilya Molchanov, Sergei Zuyev

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Variational Analysis of Functionals of a Poisson Process

Ilya Molchanov — Sergei Zuyev

N° 3302

Novembre 1997

THÈME 1

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Variational Analysis of Functionals of a Poisson Process

Ilya Molchanov* , Sergei Zuyev†

Thème 1 — Réseaux et systèmes
Projet MISTRAL

Rapport de recherche n° 3302 — Novembre 1997 — 57 pages

Abstract: Let F be a functional of a Poisson process whose distribution is determined by the intensity measure μ. Considering the expectation $\mathbf{E}_\mu F$ as a function on the cone $\mathbb{M}$ of positive finite measures μ, we derive closed form expressions for the Fréchet derivatives of all orders that generalise the perturbation analysis formulae for Poisson processes. Variational methods developed for the space $\mathbb{M}$ allow us to obtain first and second-order sufficient conditions for different types of constrained optimisation problems for $\mathbf{E}_\mu F$. We study in detail optimisation in the class of measures with a fixed total mass a and develop technique that often allows us to obtain the asymptotic behaviour of the optimal intensity measure in the high intensity settings when a grows to infinity. We give applications of our methods to design of experiments, spline approximation of convex functions, optimal placement of stations in telecommunication studies and others. We scetch possible numerical algorithms of the steepest descend type based on the obtained explicit form of the gradient.

Key-words: Poisson process, variational methods, optimisation on measures, perturbation analysis, design of experiments, approximation of functions, telecommunication system modelling, stopping sets, gradient methods
AMS 1991 Subject Classification Primary : 60D05, 60G55, 49K45 Secondary : 49K27, 62K05 65D15, 65K10, 90B12,93E23

* Department of Statistics, University of Glasgow, Glasgow G12 8QW, Scotland, U.K. E-mail: ilya@stats.gla.ac.uk
† INRIA, 2004, r-te des Lucioles, 06902 Sophia-Antipolis, France. E-mail: sergei@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 93 65 77 77 - International : +33 4 93 65 77 77 — Fax : 04 93 65 77 65 - International : +33 4 93 65 77 65
à partir du 01/01/1998
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Analyse Variationnelle de Fonctionnelles de Processus de Poisson

Résumé : Soit F une fonctionnelle d'un processus de Poisson dont la distribution est définie par la mesure d'intensité μ. En considérant l'espérance $\mathbf{E}_\mu F$ comme une fonction sur le cône $\mathbb{M}$ des mesures positives finies, nous obtenons des expressions explicites pour les dérivés de Fréchet de tous ordres, ce qui généralise les formules d'analyse de perturbations pour les processus de Poisson. Ces méthodes variationnelles permettent d'obtenir des conditions nécessaires d'ordre un et deux pour plusieurs types de problèmes d'optimisation sous contraintes. Nous étudions en détails l'optimisation dans la classe des mesures de masse totale a fixée et développons une technique permettant l'obtention du comportement asymptotique de la mesure d'intensité optimale quand a tend vers l'infini. Nous donnons des applications de nos méthodes à la planification des expériences, à l'approximation des fonctions convexes par des splines et au problème de placement optimal des stations d'un système de télécommunications. Nous proposons des algorithmes numériques de type "steepest descent" basés sur la forme explicite du gradient.

Mots-clés : processus de Poisson, méthodes variationnelles, optimisation sur des mesures, analyse de perturbations, planification des expériences, approximation des fonctions, modélisation de systèmes de télécommunications, ensemble d'arrêt, méthodes de gradient

1 Introduction

Many stochastic systems are modelled using Poisson point processes. Given a σ-finite measure μ on a Polish locally compact space X, the Poisson point process Π on X with the *intensity measure* μ is the random set of points such that 1) the number of points, $\Pi(K)$, inside a Borel set K has a Poisson distribution with the parameter $\mu(K)$; and 2) the number of points inside disjoint sets are independent. These fundamental properties often make a Poisson framework ideal for modelling complex stochastic systems. On one hand, Poisson driven models are quite often directly computable, on the other hand, there is a large freedom of choice of the measure μ, which determines the distribution of the process. For example, the Poisson process serves as a reference (or benchmark) distribution for different interaction models that appear in statistical physics and spatial statistics, many objects in stochastic geometry are related to the Poisson processes in abstract spaces (sets of lines, planes, etc.).

Last decades new methods of analysis of functionals of Poisson processes have been developed including variational calculus on the Poisson space (cf. [20, 21, 1]), asymptotic expansions (cf. [6, 5]) and change of the phase space technique (cf. [5, 19]).

If Π is a Poisson point process with intensity measure μ, the corresponding expected value of a functional $F(\Pi)$, i. e. a function defined on realisations of the process, is denoted by $\mathbf{E}_\mu F(\Pi)$. The latter depends on μ and can be considered as a functional defined on a cone $\mathbb{M}$ of (non-negative) measures. In this paper we develop a variational calculus for such functionals of a Poisson point process directly on its parameter space, i. e. on $\mathbb{M}$. First, in Section 2 we obtain analyticity results that generalise the perturbation analysis formulae derived for such functionals in [3, 19]. In particular, we show that in the case of a finite intensity measure, the expectation $\mathbf{E}_\mu F(\Pi)$ is analytic on $\mathbb{M}$ for most 'reasonable' functionals F.

The obtained formulae for derivatives in the space $\mathbb{M}$ are useful to solve variational problems of minimising the above expectation $\mathbf{E}_\mu F(\Pi)$ over a class of admissible intensity measures. To outline the difficulties, it is worth noting that the cone of all positive measures has a very rich boundary. In fact, each measure lies on its boundary, because a subtraction of another (even 'small') measure may lead to a signed measure that does not belong to the cone in question. Section 3 relies substantially on new results obtained for optimisation problems in abstract Banach spaces (cf. [22, 29, 7]) in order to find useful first and second order necessary optimality conditions for the constrained optimisation problem on the cone $\mathbb{M}$.

Section 4 describes the form these conditions take in the most common cases like optimisation with finite number of linear constraints, including, in particular, optimisation in the class of measures with a fixed total mass, fixed barycentre, in the class of absolutely

continuous measures, optimisation of a mark distribution. In subsection 4.5 we study asymptotic behaviour of solutions with a fixed total mass, when this mass increases to infinity. We develop a method that often makes possible an explicit calculation of the asymptotically optimal measure in this high intensity framework.

In Section 5 we give many examples illustrating the developed technique. Applications include, in particular, optimal design of experiments, modelling of large telecommunications networks, approximation of convex functions and some stochastic geometry models of random sets.

Finally, in the last section we outline numerical optimisation methods based of the explicit form of the gradient obtained in the paper.

Note that our method differs fundamentally from optimisation technique for systems driven by point processes whose distribution depends on a *finite* number of parameters. For the latter, we address the reader to the book [23] and references therein.

In the paper we use of the following notation:

X : a Polish locally compact space with its Borel σ-algebra $\mathfrak{B}$;

$B^{\mathbf{c}}$: complement to B in X, i.e., $B^{\mathbf{c}} = X \setminus B$;

Y : a Banach space, often it will be Euclidean space $\mathbb{R}^m$;

Y^* : the space dual to Y;

$u \cdot y$: the canonical bilinear form on $u \in Y^*$ and $y \in Y$;

$\mathbb{M}$: the pointed cone of all non-negative finite measures on $\mathfrak{B}$ with the topology of the total variation convergence;

$\widetilde{\mathbb{M}}$: the Banach space of all signed measures on $\mathfrak{B}$ with bounded total variation and equipped with the total variation norm;

$\mathbb{A}$: a closed convex subset of $\mathbb{M}$;

μ^n $= \mu \times \cdots \times \mu$, the direct n-product of a measure μ;

$\mu^+ - \mu^-$: the Jordan decomposition of a signed measure $\mu \in \widetilde{\mathbb{M}}$;

$|\mu|$: the variation measure of a signed measure μ;

$\|\mu\|$ $= |\mu|(X)$, the total variation of $\mu \in \widetilde{\mathbb{M}}$;

$\mu|_B$: the restriction of a measure μ onto a Borel set B: $\mu|_B(\bullet) = \mu(\bullet \cap B)$;

$\mathbb{M}(B)$ $= \{\mu \in \mathbb{M} : \ \mu(B^{\mathbf{c}}) = 0\}$, respectively $\widetilde{\mathbb{M}}(B) = \{\mu \in \widetilde{\mathbb{M}} : \ \mu(B^{\mathbf{c}}) = 0\}$;

$\langle g, \mu \rangle$ $= \int g(x)\mu(dx)$ for a measurable $g : X \mapsto \mathbb{R}$ and $\mu \in \mathbb{M}$.

δ_x : the Dirac measure or the unit measure concentrated at $x \in X$: $\delta_x(B) = \mathbb{1}_{x \in B}$, $B \in \mathfrak{B}$;

$\mathcal{N}$: the set of counting measures on $\mathfrak{B}$ supplied with the standard σ-algebra $\mathfrak{N}$ which is generated by the system $\{\nu \in \mathcal{N} : \ \nu(B) = n\}$, $B \in \mathfrak{B}$, $n = 0, 1, \ldots$;

Π : a point process in the phase space X i.e. a measurable mapping from some abstract probability space to $[\mathcal{N}, \mathfrak{N}]$. If otherwise not stated explicitly, we identify this probability space with the space $[\mathcal{N}, \mathfrak{N}]$ itself and Π with the identical mapping. The same letter Π is used to denote the point set (with multiple points allowed) which is the support of the counting measure Π.

$\mathbf{P}_\mu$: the distribution of the Poisson process with intensity measure $\mu \in \mathbb{M}$;

$\mathbf{E}_\mu$: the expectation with respect to the measure $\mathbf{P}_\mu$ on $\mathfrak{N}$;

F : a real-valued measurable function on $\mathcal{N}$;

b_d $= 2\pi^{d/2} d^{-1} / \Gamma(d/2)$, the volume of a unit ball in $\mathbb{R}^d$.

2 Derivatives of functionals and perturbations

In this section we establish smoothness properties of the expectation $\mathbf{E}_\mu F(\Pi)$ with respect to μ. In particular, if μ is a finite measure and the functional F of a Poisson process' configurations admits an exponential type upper bound in the number of the process' points, then $\mathbf{E}_\mu F(\Pi)$ is an analytic function on the cone $\mathbb{M}$ of finite positive measures. Under additional conditions on F further generalisations to the case of an infinite intensity measure μ are obtained.

2.1 Analyticity on the cone of finite measures

The observations in this section is based on the following useful property of an a.s. finite Poisson process: if the expectation $\mathbf{E}_\mu F(\Pi)$ exists, then it can be written in the form

$$\mathbf{E}_\mu F(\Pi) = e^{-\mu(X)} \sum_{n=0}^{\infty} \frac{1}{n!} \int\limits_{X^n} F\Big(\sum_{i=1}^{n} \delta_{x_i}\Big)\, \mu(dx_1) \ldots \mu(dx_n) \tag{2.1}$$

(cf. [9, p. 129]). Here the term corresponding to $n = 0$ should be read as $e^{-\mu(X)} F(\emptyset)$, where $\emptyset$ is the empty point configuration. In what follows the function F is fixed, so that

(2.1) can be seen as the definition of $\mathbf{E}_\mu F(\Pi)$ as a function of the intensity measure $\mu \in \mathbb{M}$ of the underlying Poisson process.

To formulate the main result of this section we extend the definition of the functional $\mathbf{E}_\mu F(\Pi)$ to $\mu \in \widetilde{\mathbb{M}}$ by means of (2.1). Remember that $\widetilde{\mathbb{M}}$ consists of signed measures on $\mathfrak{B}$ with the finite total variation.

Theorem 2.1. *Let $\mu,\ \eta \in \widetilde{\mathbb{M}}$. Assume that there exist a function $g : X \mapsto [0,\infty)$ and a constant $b > 0$ such that $\langle g, |\mu| \rangle < \infty$, $\langle g, |\eta| \rangle < \infty$ and*

$$\left| F\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \right| \le b \prod_{i=1}^{n} g(x_i) \tag{2.2}$$

for all $n \ge 0$ and μ^n and η^n-almost all $(x_1, \dots, x_n) \in X^n$. Then the expectations $\mathbf{E}_\mu F(\Pi)$ and $\mathbf{E}_{\mu+\eta} F(\Pi)$ exist and

$$\mathbf{E}_{\mu+\eta} F(\Pi) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{X^n} \overline{\Delta^n_\mu}(x_1, \dots, x_n)\, \eta(dx_1) \dots \eta(dx_n)\,, \tag{2.3}$$

where

$$\overline{\Delta^n_\mu}(x_1, \dots, x_n) = \mathbf{E}_\mu\, \Delta^n_\mu(x_1, \dots, x_n; \Pi) = \mathbf{E}_\mu \left[\sum_{m=0}^{n} (-1)^{n-m} \binom{n}{m} F\Big(\Pi + \sum_{j=1}^{m} \delta_{x_j}\Big) \right]. \tag{2.4}$$

Proof. First of all note, that in view of (2.2) the absolute value of the n-th summand in the right-hand side of (2.1) is dominated by $(\langle g, |\mu| \rangle)^n/(n!)$ and thus $\mathbf{E}_\nu F(\Pi)$ exists whenever $\langle g, |\nu| \rangle < \infty$. In particular this is true for the measure $\nu = \mu + \eta$. Moreover, by Fubini's theorem we may interchange the order of integration in the summands.

When the measure μ is replaced by $\mu + \eta$, the n-th summand in (2.1) can be rewritten as

$$\begin{aligned}
&\int_{X^n} F\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \prod_{i=1}^{n} (\mu(dx_i) + \eta(dx_i)) \\
&\qquad = \int_{X^n} \sum_{k=0}^{n} F\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \sum_{\substack{\{i_1,\dots,i_k\} \\ \subseteq \{1,\dots,n\}}} \prod_{j=1}^{k} \eta(dx_{i_j}) \prod_{i_l \notin \{i_1,\dots,i_k\}} \mu(dx_{i_l}) \\
&\qquad = \int_{X^n} \sum_{k=0}^{n} F\Big(\sum_{i=1}^{k} \delta_{x_i} + \sum_{j=k+1}^{n} \delta_{x_j}\Big) \binom{n}{k} \prod_{i=1}^{k} \eta(dx_i) \prod_{j=k+1}^{n} \mu(dx_j)\,,
\end{aligned}$$

which in view of (2.1) coincides with

$$e^{\mu(X)} \sum_{k=0}^{n} \mathbf{E}_\mu \int\limits_{X^k} F\Big(\Pi + \sum_{i=1}^{k} \delta_{x_i}\Big)\, \mathbb{1}_{\Pi(X)=n-k} \binom{n}{k} (n-k)! \prod_{i=1}^{k} \eta(dx_i)\,.$$

(By convention $\prod_{i=1}^{0} = 1$). Therefore

$$\begin{aligned}
\mathbf{E}_{\mu+\eta} F(\Pi) &= e^{-\eta(X)} \sum_{n=0}^{\infty} \frac{e^{-\mu(X)}}{n!} \int\limits_{X^n} F\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \prod_{i=1}^{n} (\mu(dx_i) + \eta(dx_i)) \\
&= e^{-\eta(X)} \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=k}^{\infty} \mathbf{E}_\mu \int\limits_{X^k} F\Big(\Pi + \sum_{i=1}^{k} \delta_{x_i}\Big)\, \mathbb{1}_{\Pi(X)=n-k} \prod_{i=1}^{k} \eta(dx_i) \\
&= e^{-\eta(X)} \sum_{k=0}^{\infty} \frac{I_k}{k!}\,,
\end{aligned}$$

where

$$I_k = \mathbf{E}_\mu \int\limits_{X^k} F\Big(\Pi + \sum_{i=1}^{k} \delta_{x_i}\Big)\, \eta(dx_1) \ldots \eta(dx_k)\,.$$

Now writing the Taylor expansion for the exponent $e^{-\eta(X)}$ we find that the coefficient of $1/(n!)$ in the product above equals

$$\sum_{k=0}^{n} \binom{n}{k} \big(-\eta(X)\big)^{n-k} I_k = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} \mathbf{E}_\mu \int\limits_{X^n} F\Big(\Pi + \sum_{i=1}^{k} \delta_{x_i}\Big)\, \eta(dx_1) \ldots \eta(dx_n)$$

and we finally obtain (2.3). □

Note that if (2.3) holds for all η with sufficiently small $\|\eta\|$ then $\mathbf{E}_\mu F$ as a functional on $\widetilde{\mathbb{M}}$ is *analytic*, see [11].

Corollary 2.2. *If the conditions of Theorem 2.1 hold for all $\eta \in \mathbb{M}$ (e.g., this is the case when F is bounded or the function g from (2.2) is bounded), then the functional $\mu \mapsto \mathbf{E}_\mu F(\Pi)$ is analytic on $\widetilde{\mathbb{M}}$. In particular, for all n, it is n-times Fréchet differentiable and its n-th Fréchet derivative is the following n-linear form:*

$$D^n \mathbf{E}_\mu F(\Pi)[\eta_1, \ldots, \eta_n] = \int\limits_{X^n} \overline{\Delta_\mu^n}(x_1, \ldots, x_n)\, \eta_1(dx_1) \ldots \eta_n(dx_n)\,.$$

In particular, the first order Fréchet derivative is defined by the condition

$$\mathbf{E}_{\mu+\eta} F(\Pi) = \mathbf{E}_\mu F(\Pi) + D\mathbf{E}_\mu F(\Pi)[\eta] + o(\|\eta\|)\,.$$

In the framework of the Perturbation analysis, a random variable Φ is called a *stochastic gradient* of a random variable F for a family of probability distributions $\mathbf{P}_\theta$ if $\frac{d}{d\theta}\,\mathbf{E}_\theta F = \mathbf{E}_\theta \Phi$. Therefore under the conditions of Corollary 2.2 the n-linear form

$$\int\limits_{X^n} \Delta^n_\mu(x_1,\dots,x_n;\Pi)\,\eta_1(dx_1)\dots\eta_n(dx_n)$$

is an *n-th order stochastic gradient* of the random variable $F(\Pi)$.

Note that for Poisson processes

$$\mathbf{E}_\mu F(\Pi + \sum_{i=1}^k \delta_{x_i}) = \mathbf{E}^{x_1,\dots,x_k}_\mu F(\Pi),$$

where $\mathbf{E}^{x_1,\dots,x_k}_\mu$ is the expectation with respect to the *k-fold Palm distribution* $\mathbf{P}^{x_1,\dots,x_k}_\mu$ (cf. [12, p.110]). Introduce the following multi-index notation: for $s \subseteq \{1,\dots,n\}$ let $|s|$ denote the cardinality of s and $\boldsymbol{x}_s = \{x_{i_j} : i_j \in s\}$. Then it is easy to see that

$$\overline{\Delta^n_\mu}(x_1,\dots,x_n) = \sum_{s\subseteq\{1,\dots,n\}} (-1)^{n-|s|}\,\mathbf{E}^{\boldsymbol{x}_s} F\,. \tag{2.5}$$

An alternative expression can be obtained using an operator ∇_x acting on an expectation $\mathbf{E}$ as $\nabla_x \mathbf{E} = \mathbf{E}^x - \mathbf{E}$. Put $\nabla_{x_1\dots x_{k-1}x_k} = \nabla_{x_k}(\nabla_{x_1\dots x_{k-1}})$. Then the family of *operators* $\mathbf{E}_\mu$ is analytic for the functionals satisfying (2.2) and its n-th Fréchet derivative is the following n-linear form:

$$D^n\,\mathbf{E}_\mu F\,[\eta_1,\dots,\eta_n] = \int\limits_{X^n} \left(\nabla_{x_1\dots x_n}\,\mathbf{E}_\mu\right) F\,\eta_1(dx_1)\dots\eta_n(dx_n)\,. \tag{2.6}$$

Remark 2.1. Note that Theorem 2.1 yields the result of [19]. Indeed, if $\eta = c\mu$, then

$$\frac{d^n}{dc^n}\,\mathbf{E}_{\mu+c\mu} F\Big|_{c=0} = \int\limits_{X^n} \overline{\Delta^n_\mu}(x_1,\dots,x_n)\,\mu(dx_1)\dots\mu(dx_n)\,,$$

which is n-th Gateaux derivative of $\mathbf{E}_\mu F$ in 'direction' μ (compare with Corollary 2 of [19]).

In our considerations the central role play the first two derivatives given by

$$D\,\mathbf{E}_\mu\,F(\Pi)[\eta] = \int_X \mathbf{E}_\mu[F(\Pi+\delta_x) - F(\Pi)]\,\eta(x) \tag{2.7}$$

and

$$D^2\,\mathbf{E}_\mu\,F(\Pi)[\eta_1,\eta_2] = \int_X\int_X \mathbf{E}_\mu\left[F(\Pi+\delta_{x_1}+\delta_{x_2}) - 2F(\Pi+\delta_{x_1}) + F(\Pi)\right]\eta(x_1)\eta(x_2)\,. \tag{2.8}$$

The function

$$\Delta(x;\Pi) = \Delta^1(x;\Pi) = F(\Pi+\delta_x) - F(\Pi)\,, \quad x\in X\,, \tag{2.9}$$

is called the *first order difference* of $F(\Pi)$ and its mean

$$\overline{\Delta}_\mu\,(x) = \mathbf{E}_\mu\,\Delta(x;\Pi)$$

is called the *expected first difference*. In our notation (2.7) becomes

$$D\,\mathbf{E}_\mu\,F(\Pi)[\eta] = \left\langle \overline{\Delta}_\mu\,,\eta\right\rangle\,. \tag{2.10}$$

Similarly, the second order difference is defined by

$$\Delta^2(x,y;\Pi) = F(\Pi+\delta_x+\delta_y) - F(\Pi+\delta_x) - F(\Pi+\delta_y) + F(\Pi) \tag{2.11}$$

with $\overline{\Delta^2_\mu}\,(x,y) = \mathbf{E}_\mu\,\Delta^2(x,y;\Pi)$.

Example 2.1. (*Symmetric statistics.*) Let

$$F(\Pi) = \frac{1}{2}\sum_{x_i,x_j\in\Pi, i\neq j} f(x_i,x_j)$$

be the second-order symmetric statistic defined for the process Π, where f is a symmetric real-valued function on $X\times X$. Assume there exists a function $h(x)$ such that $|f(x,y)|\leq h(x)h(y)$ with $\langle h,\mu\rangle<\infty$. Then (2.2) holds with $g(x)=1+h(x)$ and

$$\Delta^1(x;\Pi) = \sum_{x_i\in\Pi} f(x,x_i)\,,$$

so that

$$\overline{\Delta}_\mu\,(x) = \int_X f(x,y)\mu(dy)\,.$$

Obviously, $\overline{\Delta^2_\mu}\,(x,y) = \Delta^2(x,y;\Pi) = f(x,y)$ in this example.

2.2 Infinite measure case

While the Taylor expansion (2.3) holds for Poisson processes with finite intensity measures, it is sometimes possible to extend it to measures with infinite total mass. Since X is locally compact, it can be represented as a countable union $\cup_k X_k$, where X_k are compact for all k. Everywhere in this section we assume that the reference measure μ is locally finite (i.e. $\mu(X_k)$ is finite for all k) and that a function $F : \mathcal{N} \mapsto \mathbb{R}$ is *continuous at infinity*, i.e. $F(N) = \lim_{k\to\infty} F(N|_{X_k})$ (remember that $N|_B$ is the restriction of a measure N onto a set B). Below we simply write μ_k instead of $\mu|_{X_k}$ and $F_k(N)$ instead of $F(N|_{X_k})$.

The key observation is the following. An important property of a Poisson process is that the distribution of a random measure $\Pi|_B$ under the probability law $\mathbf{P}_\mu$ coincides with the distribution of Π under $\mathbf{P}_{\mu|_B}$. Therefore $\mathbf{E}_\mu F_k(\Pi) = \mathbf{E}_{\mu_k} F(\Pi)$, measures μ_k are finite, so that Theorem 2.1 is applicable. Additional restrictions should, however, be imposed to assure the convergence of the corresponding terms to their infinite measure counterparts.

Theorem 2.3. *Assume that $\|\eta\| < \infty$ and there exists a constant M such that $|F(\sum_{i=1}^n \delta_{x_i})| \le M$ for all $n \ge 0$ and μ^n and η^n-almost all $(x_1, \dots, x_n) \in X^n$. Then the expansion (2.3) holds.*

Proof. Write the tail of the series (2.3) for the function F_k:

$$\left| \sum_{n=n_0}^{\infty} \frac{1}{n!} \int_{X_k^n} \overline{\Delta^n_{\mu_k}}(x_1, \dots, x_n)\, \eta(dx_1) \dots \eta(dx_n) \right|$$

$$\le \sum_{n=n_0}^{\infty} \frac{M}{n!} \sum_{m=0}^{n} \binom{n}{m} \|\eta\|^n = M \sum_{n=n_0}^{\infty} \frac{\big(2\|\eta\|\big)^n}{n!}. \qquad (2.12)$$

The expression above is the tail of $M \exp\{2\|\eta\|\}$ and therefore it is uniformly small in k. Now take the limit $k \to \infty$ in (2.3). By dominated convergence we can exchange the expectation and the limit in the left-hand side, swap the sum and the limit in the right-hand side and, finally, exchange the integral and the limit to obtain (2.3) for locally finite μ and finite η. □

The condition on the boundedness of the function F can be relaxed to conditions on the boundedness of its first order difference:

Theorem 2.4. *Assume that $|F(\sum_{i=1}^n \delta_{x_i} + \delta_x) - F(\sum_{i=1}^n \delta_{x_i})| \le 2M$ for all $n \ge 0$ and μ^n and η^n-almost all $(x_1, \dots, x_n) \in X^n$. If the family $F_k(\Pi) = F(\Pi|_{X_k})$ is uniformly integrable with respect to $\mathbf{P}_{\mu+\eta}$ for all η with $\|\eta\| < \infty$, then the expansion (2.3) holds.*

The proof is similar to Theorem 2.3, but the bound in (2.12) is obtained now by iteration of the difference functions.

3 Optimisation in the space of signed measures

In this section we derive first and second order necessary conditions for an extremum in constrained optimisation problems on the cone of positive measures.

3.1 Necessary conditions for extremum

Let Y be a Banach space, Y^* be its dual space and $u \cdot y$ be the canonical bilinear form defined for $y \in Y$ and $u \in Y^*$.

Consider the following optimisation problem:

$$f(\mu) \to \inf, \quad \mu \in \mathbb{A}, \quad H(\mu) \in C, \tag{3.1}$$

where $\mathbb{A}$ and C are closed convex subsets of $\mathbb{M}$ and Y, respectively, $f : \widetilde{\mathbb{M}} \mapsto \mathbb{R}$ and $H : \widetilde{\mathbb{M}} \mapsto Y$ are Fréchet differentiable functions.

The following is the definition of the Robinson's regularity condition [22, Theorem 1] which, as shown in [29], guarantees the existence and boundedness of the Lagrange multipliers (see also [17] for the discussion on different forms of the regularity condition).

Definition 3.1. A measure μ is called *regular* for Problem (3.1) if

$$0 \in \operatorname{core}\big\{H(\mu) - C + DH(\mu)[\mathbb{A} - \mu]\big\}, \tag{3.2}$$

where $\operatorname{core} A$ is the set of all $y \in A \subset Y$ such that $y + ty_1 \in A$ for all $y_1 \in Y$ and all sufficiently small positive t. Note that '+' (respectively '−') operation with sets designates all pairwise sums of (respectively differences between) the points from the corresponding sets.

Note that the lower limit of a system $\{S_t\}$ of subsets of a Banach space is defined to be the set of limits for all convergent sequences $\{x_t, t \geq 0\}$ such that $x_t \in S_t$ for all t.

The *first order tangent set* to a set B at a point x in a Banach space is defined as

$$T_B(x) = \liminf_{t \downarrow 0} \frac{B - x}{t}$$

and the *second order tangent set* to a set B at a point x in direction $z \in T_B(x)$ is the following set

$$T_B^2(x,z) = \liminf_{t \downarrow 0} \frac{B - x - tz}{t^2/2}\,.$$

Both tangent sets are closed and $T_B(x)$ is, in addition, a cone.

The following theorem, which is a synthesis of Theorems 4.1 and 4.2 of [7] applied to our framework, gives first and second order necessary conditions for a minimum in Problem (3.1).

Theorem 3.1. *Assume that both $f : \widetilde{\mathbb{M}} \mapsto \mathbb{R}$ and $H : \widetilde{\mathbb{M}} \mapsto Y$ are continuous on $\mathbb{A}$ and twice Fréchet differentiable at a regular $\overline{\mu} \in \mathbb{A}$ such that $H(\overline{\mu}) \in C$. Then, if $\overline{\mu}$ is a local minimum point in Problem* (3.1)*, the following (necessary) optimality conditions are satisfied for all $\eta \in T_{\mathbb{A} \cap H^{-1}(C)}(\overline{\mu})$:*

$$\textbf{(i)} \quad Df(\overline{\mu})[\eta] \geq 0\,; \tag{3.3}$$

$$\textbf{(ii)} \quad \text{if } Df(\overline{\mu})[\eta] = 0\,, \text{ then}$$
$$D^2 f(\overline{\mu})[\eta,\eta] \geq 0 \text{ for all } \nu \in T^2_{\mathbb{A} \cap H^{-1}(C)}(\overline{\mu},\eta)\,. \tag{3.4}$$

Moreover, for each $\overline{\eta}$ verifying

$$\textbf{(a)} \quad Df(\overline{\mu})[\overline{\eta}] = 0 \text{ and}$$

$$\textbf{(b)} \quad T^2_{\mathbb{A}}(\overline{\mu},\overline{\eta}) \neq \emptyset\,,\ T^2_C\big(H(\overline{\mu}), DH(\overline{\mu})[\overline{\eta}]\big) \neq \emptyset$$

the following dual conditions hold: there exists (a Lagrange multiplier or a Kuhn-Tucker vector) $u \in Y^*$ such that $u \cdot y \geq 0$ for any $y \in T_C\big(H(\overline{\mu})\big)$ and for the Lagrangian function $L(\mu) = Df(\mu) - u \cdot H(\mu)$ one has

$$\textbf{(iii)} \quad DL(\overline{\mu})[\eta] = Df(\overline{\mu})[\eta] - u \cdot DH(\overline{\mu})[\eta] \geq 0 \text{ for all } \eta \in T_{\mathbb{A}}(\overline{\mu})\,; \tag{3.5}$$

$$\textbf{(iv)} \quad DL(\overline{\mu})[\overline{\eta}] = Df(\overline{\mu})[\overline{\eta}] - u \cdot DH(\overline{\mu})[\overline{\eta}] = 0\,. \tag{3.6}$$

Remark 3.1. By replacing f by $-f$ we obtain necessary conditions for *maximum* that manifest in changing $\geq$ by $\leq$ in (3.3), (3.4) and (3.5).

3.2 Structure of tangent sets in the space of measures

By [7, Theorem 3.1] for a regular μ one has

$$T_{\mathbb{A}\cap H^{-1}(C)}(\mu) = T_{\mathbb{A}}(\mu) \cap \big(DH(\mu)\big)^{-1}\big[T_C(H(\mu))\big] \tag{3.7}$$

and

$$T^2_{\mathbb{A}\cap H^{-1}(C)}(\mu,\eta) = T^2_{\mathbb{A}}(\mu,\eta) \cap \big(DH(\mu)\big)^{-1}\big[T^2_C(H(\mu), DH(\mu)[\eta]) - (D^2H(\mu))[\eta,\eta]\big]\,. \tag{3.8}$$

Consider, for example, an important case of finitely many equality constraints $H_i(\mu) = 0,\ i = 1,\dots,k$, so that H is the vector function from $\mathbb{M}$ to $\mathbb{R}^k$. Often the differentials $DH(\mu)[\eta]$ admit the integral representation $\langle h_i, \eta\rangle$ for some functions $h_i : X \mapsto \mathbb{R}$. If $h = (h_1,\dots,h_k)$, then the tangent sets to the feasible set $\mathbb{A}\cap H^{-1}(0)$ are given by

$$T_{\mathbb{A}\cap H^{-1}(0)}\,(\mu) = T_{\mathbb{A}}(\mu) \cap \{\eta\ :\ \langle h,\eta\rangle = 0\} \tag{3.9}$$

and

$$T^2_{\mathbb{A}\cap H^{-1}(0)}\,(\mu,\eta) = T^2_{\mathbb{A}}(\mu,\eta) \cap \{\nu\ :\ \langle h,\nu\rangle = 0\}\,. \tag{3.10}$$

Different forms of constraints will be considered in Sections 3.3 and 4. Here we characterise the tangent sets $T_{\mathbb{A}}(\mu)$ and $T^2_{\mathbb{A}}(\mu,\eta)$ for the most common cases: for $\mathbb{A}$ being the whole cone $\mathbb{M}$, and for $\mathbb{A}$ which consists of measures absolutely continuous with respect to some reference measure λ.

Theorem 3.2. *Let $\mu \in \mathbb{M}$. Then*

$$T_{\mathbb{M}}(\mu) = \{\eta \in \widetilde{\mathbb{M}}\ :\ \eta^- \ll \mu\}\,; \tag{3.11}$$

$$T^2_{\mathbb{M}}(\mu,\eta) = \big\{\nu \in \widetilde{\mathbb{M}}\ :\ \nu^- \ll \mu \text{ or } \nu^- \ll \eta^+\big\} \quad \textit{if } \frac{d\eta^-}{d\mu} \textit{ is bounded}. \tag{3.12}$$

To prove this theorem we need the following result.

Lemma 3.3. *If $\mu \in \mathbb{M}$, then*

$$T_{\mathbb{M}}(\mu) \supseteq \Big\{\eta \in \widetilde{\mathbb{M}}\ :\ \eta^- \ll \mu \textit{ and } \frac{d\eta^-}{d\mu} \textit{ is bounded}\Big\}\,. \tag{3.13}$$

If, additionally, $\frac{d\eta^-}{d\mu}$ is bounded, then

$$T^2_{\mathbb{M}}(\mu,\eta) \supseteq \{\nu \in \widetilde{\mathbb{M}}\ :\ \nu^- \ll \mu \text{ or } \nu^- \ll \eta^+ \textit{ and at least one of the corresponding densities is bounded}\}\,. \tag{3.14}$$

Proof. Let η be such that $\eta^- \ll \mu$ and $\frac{d\eta^-}{d\mu} \le N$. Then for all $t \le N^{-1}$ the measure

$$\mu + \eta t = \int_{\bullet} \Big(1 - t\,\frac{d\eta^-}{d\mu}(x)\Big)\,\mu(dx) + \eta^+$$

is non-negative, so its distance to $\mathbb{M}$ is 0 and we get (3.13).

Similarly, $\nu \in T^2_{\mathbb{M}}(\mu,\eta)$ if the distance from $\mu + \eta t + \nu t^2/2$ to $\mathbb{M}$ has order $o(t^2)$ as $t \to 0$. Let now $\eta \in T_{\mathbb{M}}(\mu)$ be such that its density with respect to μ is bounded. Take ν such that either $\nu^- \ll \mu$ or $\nu^- \ll \eta^+$. Then writing

$$\mu + \eta t + \nu t^2/2 = \eta^+ t + \nu^+\,\frac{t^2}{2} + \int_{\bullet} \Big(1 - t\,\frac{d\eta^-}{d\mu} - \frac{t^2}{2}\,\frac{d\nu^-}{d\mu}\Big)\,d\mu$$

in the first case, or

$$\mu + \eta t + \nu t^2/2 = \nu^+\,\frac{t^2}{2} + \int_{\bullet} \Big(1 - t\,\frac{d\eta^-}{d\mu}\Big)\,d\mu + t\int_{\bullet} \Big(1 - \frac{t}{2}\,\frac{d\nu^-}{d\eta^+}\Big)\,d\eta^+$$

in the second case, we see that the measures above belong to $\mathbb{M}$ for all sufficiently small t if the density of ν^- with respect to μ or with respect to η^+ is bounded, that proves (3.14). □

Remark 3.2. The statement of Lemma 3.3 is, in fact, independent of the topology chosen in the set of measures. In contrast, the *exact* form of the tangent sets depends on it. Assume that the topology satisfies the following property: for any μ-integrable function h the sequence of measures $\mu_n(\bullet) = \int_{\bullet} \min\{h(x), n\}\,d\mu$ converges to $\int_{\bullet} h(x)\,d\mu$ (this is true, e.g., for the total variation convergence and hence for all weaker types of convergence: set-wise, weak etc.) Then the boundedness of the derivatives can be omitted from (3.13) and (3.14). Indeed, in this case the closure of the set in the right-hand side of (3.13) coincides with $\{\eta \in \widetilde{\mathbb{M}} : \eta^- \ll \mu\}$. But $T_{\mathbb{M}}(\mu)$ is closed, so we get the inclusion

$$T_{\mathbb{M}}(\mu) \supseteq \{\eta \in \widetilde{\mathbb{M}} : \eta^- \ll \mu\}\,.$$

The same argument applies to (3.14), since $T^2_{\mathbb{M}}(\mu,\eta)$ is also closed.

Proof of Theorem 3.2. By Lemma 3.3 and in view of Remark 3.2, it suffices to prove that the tangent cones are contained in the families in the right-hand sides of (3.11) and (3.12).

Assume that $\eta^- \not\ll \mu$, i.e. there exists $B \in \mathfrak{B}(X)$ such that $\mu(B) = 0$ and $\eta^-(B) = \varepsilon > 0$. Then $(\mu + t\eta)(B) = -t\varepsilon$. But $\upsilon(B) \ge 0$ for all $\upsilon \in \mathbb{M}$ so the total variation distance from $\mu + \eta t$ to $\mathbb{M}$ does not have order $o(t)$.

Similarly, if ν is such that neither $\nu \ll \eta^+$ nor $\nu^- \ll \mu$ holds, then there exists a Borel set B such that $\nu^-(B) = \varepsilon > 0$ but $\eta^+(B) = \mu(B) = 0$. Therefore $(\mu + \eta t + \nu t^2/2)(B) = -t^2/2\,\varepsilon$ and thus the distance between $\mu + \eta t + \nu t^2/2$ and $\mathbb{M}$ is of order at least $O(t^2)$. □

It is possible to show that Theorem 3.2 holds also for the space of measures with the topology induced by the set-wise convergence.

Consider now the family $\mathbb{M}_\lambda$ of measures which are absolutely continuous with respect to a certain reference measure λ and denote by $\mathbb{M}_\lambda^N$ its convex subset

$$\mathbb{M}_\lambda^N = \left\{ \mu \in \mathbb{M}_\lambda \ : \ \frac{d\mu}{d\lambda} \leq N \right\}. \tag{3.15}$$

If $\nu \in \mathbb{M}_\lambda$, then ϕ_ν stands for its Radon-Nikodym derivative with respect to λ. Note that the total variation distance between $\mu_1, \mu_2 \in \mathbb{M}_\lambda$ equals

$$d(\mu_1, \mu_2) = \|\mu_1 - \mu_2\| = \int_X |\phi_{\mu_1}(x) - \phi_{\mu_2}(x)|\, \lambda(dx)\,.$$

Theorem 3.4. *If $\mathbb{M}_\lambda$ is equipped with the topology of total variation convergence, then the first order tangent cones are given by:*

$$T_{\mathbb{M}_\lambda}(\mu) = \{\eta \ : \ \eta \ll \lambda \text{ and } \left\langle [t\phi_{\eta^-} - \phi_\mu]^+, \lambda \right\rangle = o(t) \text{ as } t \downarrow 0\}\,; \tag{3.16}$$

$$T_{\mathbb{M}_\lambda^N}(\mu) = \{\eta \ : \ \eta \ll \lambda \text{ and } \left\langle [t\phi_{\eta^-} - \phi_\mu]^+, \lambda \right\rangle + \left\langle [\phi_\mu + t\phi_{\eta^+} - N]^+, \lambda \right\rangle = o(t) \text{ as } t \downarrow 0\}\,, \tag{3.17}$$

where $[a]^+ = \max\{a, 0\}$. In particular,

$$\eta \in T_{\mathbb{M}_\lambda}(\mu) \text{ implies } \eta^- \ll \mu\,; \tag{3.18}$$

$$\eta \in T_{\mathbb{M}_\lambda^N}(\mu) \text{ implies } \eta^- \ll \mu \text{ and } \eta^+(\{x \ : \ \phi_\mu(x) = N\}) = 0\,. \tag{3.19}$$

Proof. If $\eta \not\ll \lambda$ then $d(\mu + t\eta, \mathbb{M}_\lambda)$ has order t, since $\mu \ll \lambda$. Therefore, such η does not belong to $T_{\mathbb{M}_\lambda}$. If $\eta \ll \lambda$, then

$$\begin{aligned} d(\mu + t\eta, \mathbb{M}_\lambda) &= \inf_{\psi \geq 0} \int_X |\phi_\mu(x) + t\phi_\eta(x) - \psi(x)|\, \lambda(dx) \\ &= \int_X |\phi_\mu(x) - t\phi_{\eta^-}(x)|\, \mathbb{1}_{\phi_\mu(x) - t\phi_{\eta^-}(x) < 0}\, \lambda(dx) \end{aligned}$$

that has order $o(t)$ by (3.16). Clearly, this is not the case if (3.18) is violated. Finally, (3.17) and (3.19) can be derived similarly. □

3.3 Equality and inequality constraints

Many interesting functionals of measures have derivatives which can be represented in the integral form. Then it is possible to simplify conditions of Theorem 3.1. Below we consider a practical case of finitely many constraints of the equality and inequality type:

$$\begin{cases} H_i(\mu) = 0\,, & i = 1, \dots, m\,; \\ H_j(\mu) \le 0\,, & j = m+1, \dots, k\,. \end{cases} \tag{3.20}$$

In this particular case one has $Y = Y^* = \mathbb{R}^k$, $H : \widetilde{\mathbb{M}} \mapsto \mathbb{R}^k$, $C = \{0_{\mathbb{R}^m}\} \times (\mathbb{R}_-)^{k-m}$, $\mathbb{R}_- = (-\infty, 0]$, and $u{\cdot}y$ is the scalar product of vectors u and y in $\mathbb{R}^k$.

Theorem 3.5. *Assume that the conditions of Theorem 3.1 are satisfied and there exist measurable functions $\nabla f : X \mapsto \mathbb{R}$ and $h_i : X \mapsto \mathbb{R}$, $i = 1, \dots, k$, such that for all $\eta \in \widetilde{\mathbb{M}}$*

$$Df(\overline{\mu})[\eta] = \langle \nabla f, \eta \rangle \quad \text{and} \quad DH(\overline{\mu})[\eta] = \langle h, \eta \rangle\ , \tag{3.21}$$

where $H = (H_1, \dots, H_k)$ and $h = (h_1, \dots, h_k)$. Let $\overline{\mu}$ be a regular local minimum of f subject to (3.20). *Assume that there exists $\varepsilon > 0$ such that*

$$\textbf{(i)} \quad (1-\varepsilon)\,\overline{\mu} \in \mathbb{A} \text{ and } (1+\varepsilon)\,\overline{\mu} \in \mathbb{A}; \tag{3.22}$$

$$\textbf{(ii)} \quad \overline{\mu} + t\delta_x \in \mathbb{A} \text{ for all } x \in X \text{ and } t \in (0, \varepsilon]\,. \tag{3.23}$$

Then there exists $u \in \mathbb{R}^m \times (\mathbb{R}_-)^{k-m}$ such that

$$\begin{cases} \nabla f(x) = u{\cdot}h(x) & \overline{\mu} - \text{a.e.}\,, \\ \nabla f(x) \ge u{\cdot}h(x) & \text{for all } x \in X\,. \end{cases} \tag{3.24}$$

Proof. First of all note that a measure η belongs to $T_{\mathbb{A}}(\overline{\mu})$ if and only if the distance from $\overline{\mu} + t\eta$ to the set $\mathbb{A}$ is of order $o(t)$ as $t \downarrow 0$. Condition (3.23) implies that $\overline{\mu} + t\varepsilon\delta_x \in \mathbb{A}$ for all $t \le 1$, so that $\varepsilon\delta_x \in T_{\mathbb{A}}(\overline{\mu})$ for all $x \in X$. Furthermore, (3.22) and convexity of $\mathbb{A}$ imply that $s\overline{\mu} \in \mathbb{A}$ for all $s \in [1-\varepsilon, 1+\varepsilon]$, whence both $\varepsilon\overline{\mu}$ and $-\varepsilon\overline{\mu}$ belong to $T_{\mathbb{A}}(\overline{\mu})$. Therefore, inequality (3.5) writes

$$\langle \nabla f - u{\cdot}h, \eta \rangle \ge 0 \quad \text{for all } \ \eta \in T_{\mathbb{A}}(\overline{\mu})\,,$$

so substituting $\eta = \varepsilon\delta_x$ yields the inequality in (3.24). Finally, choosing $\eta = \varepsilon\overline{\mu}$ and $\eta = -\varepsilon\overline{\mu}$ consequently yields $\langle \nabla f - u{\cdot}h, \overline{\mu} \rangle = 0$, whence the non-negative function $\nabla f(x) - u{\cdot}h(x)$ vanishes for $\overline{\mu}$-almost all x. □

Remark 3.3. In the case of constraints (3.20) satisfying (3.21), the regularity condition (3.2) becomes the *Mangasarian-Fromowitz constraint qualification* (cf. [7, p.274]) that is the linear independence of the gradients $h_1, \dots, h_m$ and the existence of a measure $\eta \in \widetilde{\mathbb{M}}$ such that

$$\begin{cases} \langle h_i, \eta \rangle = 0 & \text{for all } i = 1, \dots, m\,, \\ \langle h_j, \eta \rangle < 0 & \text{for all } j \in \{m+1, \dots, k\} \text{ verifying } H_j(\overline{\mu}) = 0\,. \end{cases} \tag{3.25}$$

Without the inequality constraints (3.25) trivially holds for η being the zero measure.

Remark 3.4. Conditions (3.22) and (3.23) can be replaced by the condition that $\pm t\overline{\mu} \in T_{\mathbb{A}}(\overline{\mu})$ and $t\delta_x \in T_{\mathbb{A}}(\overline{\mu})$ for all $x \in X$ and sufficiently small t. Theorem 3.2 shows that this is indeed the case if $\mathbb{A} = \mathbb{M}$.

Remark 3.5. Relation (3.24) has been obtained by applying (3.5) to some *specific* measures η from the tangent cone $T_{\mathbb{A}}(\overline{\mu})$. It is easy to see that if, in addition, $\overline{\mu}|_B \in T_{\mathbb{A}}(\overline{\mu})$ for all Borel sets B (for example, if $\mathbb{A} = \mathbb{M}$ or $\mathbb{A} = \mathbb{M}_\lambda$), then the first order conditions from Theorem 3.1 are, in fact, *equivalent* to (3.24).

4 Special optimisation problems

In this section, the general results of Section 3 are applied to derive specific necessary conditions for extremum of the functional $f(\mu) = \mathbf{E}_\mu F(\Pi)$ defined on $\mu \in \mathbb{M}$ by means of (2.1). We address problems with various types of constraints: fixed total mass, fixed barycentre, optimisation of the mark distribution and discuss the case of a high intensity. The class $\mathbb{A}$ in which we search the optimum will be the set $\mathbb{M}$ of all positive measures, the set $\mathbb{M}_\lambda$ of the measures that are absolutely continuous with respect to some measure λ and finally, the class of product measures with one fixed component (representing marked point processes with a fixed distribution in the position space).

4.1 Measures with a fixed total mass

In most examples considered in Section 5 below we deal with the following optimisation problem:

$$\mathbf{E}_\mu F(\Pi) \to \inf\,, \quad \mu \in \mathbb{M}, \quad \mu(X) = a\,, \tag{4.1}$$

called the *fixed total mass problem* with the total mass a. We assume that the condition of Corollary 2.2 holds, so that $\mathbf{E}_\mu F$ has two Fréchet derivatives. If $H(\mu) = \mu(X) - a$, then

the condition $H(\mu) = 0$ singles out the family of measures with the total mass a. In this case $DH(\mu)[\eta] = \langle 1, \eta \rangle = \eta(X)$ and Theorem 3.5 applies with $\mathbb{A} = \mathbb{M}$, $m = k = 1$ and $h_1 = 1$. Namely, if $\overline{\mu} \in \mathbb{M}$ solves (4.1), then there exists a constant $u \in \mathbb{R}$ such that

$$\begin{cases} \overline{\Delta}_{\overline{\mu}}(x) = u & \overline{\mu} - \text{a.e.}, \\ \overline{\Delta}_{\overline{\mu}}(x) \geq u & \text{for all } x \in X. \end{cases} \tag{4.2}$$

Take now a measure $\eta \in T_{\mathbb{M} \cap H^{-1}(0)}(\overline{\mu})$ such that $\langle \overline{\Delta}_{\overline{\mu}}, \eta \rangle = 0$. The second order optimality condition (3.4) now reads

$$\langle \overline{\Delta}_{\overline{\mu}}, \nu \rangle + \iint \overline{\Delta^2_{\overline{\mu}}}(x_1, x_2)\, \eta(dx_1)\eta(dx_2) \geq 0 \tag{4.3}$$

for all $\nu \in T^2_{\mathbb{M} \cap H^{-1}(0)}(\overline{\mu}, \eta)$.

By (3.9) $\eta(X) = 0$, whence

$$\begin{aligned} \int \overline{\Delta^2_{\overline{\mu}}}(x_1, x_2)\, \eta(dx_1) &= \int \left(\mathbf{E}_{\overline{\mu}}[F(\Pi + \delta_{x_1} + \delta_{x_2}) - F(\Pi)] + 2\overline{\Delta}_{\overline{\mu}}(x_1) \right) \eta(dx_1) \\ &= \int \mathbf{E}_{\overline{\mu}} F(\Pi + \delta_{x_1} + \delta_{x_2})\, \eta(dx_1). \end{aligned}$$

and (4.3) transforms into

$$\langle \overline{\Delta}_{\overline{\mu}}, \nu \rangle + \iint \mathbf{E}_{\overline{\mu}} F(\Pi + \delta_{x_1} + \delta_{x_2})\, \eta(dx_1)\eta(dx_2) \geq 0. \tag{4.4}$$

Choose

$$\eta(\bullet) = \delta_{x_1} - \overline{\mu}(\bullet \cap B_r(x_2))/\overline{\mu}(B_r(x_2)),$$

where $\overline{\Delta}_{\overline{\mu}}(x_1) = \overline{\Delta}_{\overline{\mu}}(x_2) = u$ and $B_r(x)$ is a ball of radius r centred in x. Then by Lemma 3.3 and (3.9) we have

$$\eta \in \{\eta^- \ll \overline{\mu},\ \frac{d\eta^-}{d\overline{\mu}} \text{ is bounded and } \eta(X) = 0\} \subseteq T_{\mathbb{M} \cap H^{-1}(0)}(\overline{\mu}),$$

and, in addition, $\langle \overline{\Delta}_{\overline{\mu}}, \eta \rangle = 0$. Moreover, η^- converges weakly to δ_{x_2} as r vanishes, so if the function $\mathbf{E}_{\overline{\mu}} F(\Pi + \delta_{x_1} + \delta_{x_2})$ is continuous for $\overline{\mu}^2$-almost all (x_1, x_2), then (4.4) gives the following second order necessary condition for minimum:

$$\mathbf{E}_{\overline{\mu}}[F(\Pi + 2\delta_{x_1}) + F(\Pi + 2\delta_{x_2})] \geq 2\, \mathbf{E}_{\overline{\mu}} F(\Pi + \delta_{x_1} + \delta_{x_2}) \tag{4.5}$$

for $\overline{\mu}$-almost all x_1 and x_2. If F does not depend on multiple points, then (4.5) becomes

$$\mathbf{E}_{\overline{\mu}}[F(\Pi+\delta_{x_1})-F(\Pi)]+\mathbf{E}_{\overline{\mu}}[F(\Pi+\delta_{x_2})-F(\Pi)]\geq 2\,\mathbf{E}_{\overline{\mu}}[F(\Pi+\delta_{x_1}+\delta_{x_2})-F(\Pi)]\,.$$

Thus, under (4.2), the condition (4.5) is equivalent to

$$\mathbf{E}_{\overline{\mu}}\,F(\Pi+\delta_{x_1})\geq \mathbf{E}_{\overline{\mu}}\,F(\Pi+\delta_{x_1}+\delta_{x_2})\quad \text{for } \overline{\mu}^2\text{-almost all } (x_1,x_2). \tag{4.6}$$

Remark 4.1. One can give a probabilistic meaning of the constant u intervening in (4.2). Assume that $\overline{\mu}$ satisfies (4.2). Then for $\overline{\mu}$-almost all x Slivnyak's theorem (see, e.g. [26, p.121] and [25]) yields

$$u=\overline{\Delta}_{\overline{\mu}}(x)=\mathbf{E}_{\overline{\mu}}[F(\Pi+\delta_x)-F(\Pi)]=\mathbf{E}^x_{\overline{\mu}}F(\Pi)-\mathbf{E}_{\overline{\mu}}F(\Pi)\,,$$

where $\mathbf{E}^x_{\overline{\mu}}$ is the local Palm expectation (the expectation 'given a point of the process at location x'). Therefore by Campbell's formula (see, e.g. [26, p.119])

$$\begin{aligned}0=\int(\mathbf{E}^{x_1}_{\overline{\mu}}\,F(\Pi)-\mathbf{E}^{x_2}_{\overline{\mu}}\,F(\Pi))\,\overline{\mu}(dx_1)&=\int\mathbf{E}^{x_1}_{\overline{\mu}}\,F(\Pi)\,\overline{\mu}(dx_1)-\overline{\mu}(X)\;\mathbf{E}^{x_2}_{\overline{\mu}}\,F(\Pi)\\&=\mathbf{E}_{\overline{\mu}}[F(\Pi)\Pi(X)]-a\,\mathbf{E}^{x_2}_{\overline{\mu}}\,F(\Pi)\,.\end{aligned}$$

Thus the constant u in (4.2) equals

$$\mathbf{E}_{\overline{\mu}}\Big[F(\Pi)(a^{-1}\Pi(X)-1)\Big]=\mathbf{cov}_{\overline{\mu}}\{F(\Pi),a^{-1}\Pi(X)-1\}\,.$$

4.2 Measures with a fixed barycentre

Let $X\subset\mathbb{R}^d$, $Y=\mathbb{R}^d$, and

$$H(\mu)=\int_X x\mu(dx)\,.$$

The condition $H(\mu)=z$ for some $z\in\mathbb{R}^d$ means that the barycentre of μ is located at z. Clearly, $DH(\mu)[\eta]=\langle h,\eta\rangle$ for $h(x)=x$, $x\in\mathbb{R}^d$. By Theorem 3.5, if $\overline{\mu}$ minimises $\mathbf{E}_\mu F$ under the condition $H(\mu)=z$, then

$$\begin{cases}\overline{\Delta}_{\overline{\mu}}(x)=v\cdot x & \overline{\mu}-\text{a.e.},\\ \overline{\Delta}_{\overline{\mu}}(x)\geq v\cdot x & \text{for all } x\in X\end{cases} \tag{4.7}$$

for some $v\in\mathbb{R}^d$. Thus, (4.7) means that $\overline{\Delta}_{\overline{\mu}}(x)$ is *linear* for $\overline{\mu}$-almost all x.

One can show similarly, that if $\overline{\mu}$ provides a minimum over all measures with a given total mass and a fixed barycentre, then the function $\overline{\Delta}_{\overline{\mu}}(x)$ must be *affine* for $\overline{\mu}$-almost all x.

4.3 Optimisation in the class of absolutely continuous measures

In this section we consider optimisation problems in the spaces $\mathbb{M}_\lambda$ and $\mathbb{M}_\lambda^N$ of absolutely continuous with respect to λ measures introduced in Section 3.2 (see (3.15)). As above, ϕ_ν denotes the Radon-Nikodym derivative of a measure ν with respect to λ. Recall the definitions of essential extrema:

$$\lambda\text{–ess-inf}_B f(x) = \sup_{E:\,\lambda(E)=0} \inf_{B\setminus E} f(x)\,;$$
$$\lambda\text{–ess-sup}_B f(x) = \inf_{E:\,\lambda(E)=0} \sup_{B\setminus E} f(x)\,.$$

Note that neither $\mathbb{M}_\lambda$ nor $\mathbb{M}_\lambda^N$ satisfies (3.23) so that Theorem 3.5 is not applicable. However, a modification of its proof allows us to obtain the following result.

Theorem 4.1. *Assume that the conditions of Theorem 3.1 and condition* (3.21) *are satisfied. Let $\overline{\mu}$ be a regular local minimum of $\mathbf{E}_\mu F$ over $\mu \in \mathbb{A}$ subject to* (3.20). *Then there exists $u \in \mathbb{R}^m \times (\mathbb{R}_-)^{k-m}$ such that*

(i) *in the case $\mathbb{A} = \mathbb{M}_\lambda$ one has*

$$\begin{cases} \overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x) = 0 \qquad \overline{\mu}-\text{a.e.}\,, \\ \lambda\text{–ess-inf}\left(\overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x)\right) \ge 0\,; \end{cases} \tag{4.8}$$

(ii) *and in the case $\mathbb{A} = \mathbb{M}_\lambda^N$ one has*

$$\begin{cases} \overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x) = 0 \qquad \overline{\mu}|_{\{\phi_{\overline{\mu}}<N\}}-\text{a.e.}\,, \\ \lambda\text{–ess-sup}_{\{\phi_{\overline{\mu}}=N\}}\left(\overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x)\right) \le 0\,; \\ \lambda\text{–ess-inf}_{\{\phi_{\overline{\mu}}<N\}}\left(\overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x)\right) \ge 0\,. \end{cases} \tag{4.9}$$

Proof. Let $\rho(x) = \overline{\Delta}_{\overline{\mu}}(x) - u\cdot h(x)$. By Theorem 3.4 we have $\lambda|_B \in T_{\mathbb{M}_\lambda}$ for all Borel sets B. So taking $\eta = \lambda|_B$ in (3.5) we obtain $\langle \rho\,\mathbb{1}_B, \lambda\rangle \ge 0$ for any B that is equivalent to the second statement of (4.8). Put now consequently $\eta = \pm\varepsilon\overline{\mu}|_{B\cap\{\phi_\mu>\varepsilon\}} \in T_{\mathbb{M}_\lambda}$. Then (3.5) gives $\rho(x) = 0$ for $\overline{\mu}$-almost all x from $\{\phi_{\overline{\mu}} > \varepsilon\}$. Since ε is arbitrary, we get the first statement of (4.8). If ρ satisfies (4.8) and $\eta \in T_{\mathbb{M}_\lambda}$, then (3.16) and (3.18) yield

$$\langle \rho, \eta\rangle = \left\langle \rho\phi_{\eta^+}, \lambda\right\rangle - \left\langle \rho\frac{d\eta^-}{d\overline{\mu}}, \overline{\mu}\right\rangle \ge 0\,,$$

because the first summand is non-negative and the second is 0. Thus, (3.5) turns into (4.8).

The case $\mathbb{A} = \mathbb{M}^N_\lambda$ is treated similarly. Taking $\eta = \lambda|_{B\cap\{\phi_{\overline{\mu}}\le N-\varepsilon\}} \in T_{\mathbb{M}^N_\lambda}$ gives

$$\lambda\text{–ess-inf}_{\{\phi_{\overline{\mu}}\le N-\varepsilon\}} \rho(x) \ge 0$$

and by arbitrariness of ε, it implies the third statement of (4.9). Then $\eta = -\varepsilon\overline{\mu}|_{B\cap\{\phi_{\overline{\mu}}=N\}}$ gives the second one. Finally, $\eta = \pm\varepsilon\overline{\mu}|_{B\cap\{\varepsilon\le\phi_\mu\le N-\varepsilon\}}$ yields $\rho(x) = 0$ for any $\varepsilon > 0$ and $\overline{\mu}$-almost all x such that $\varepsilon \le \phi_{\overline{\mu}}(x) \le N - \varepsilon$, that implies the first statement in (4.9).

Again, if ρ satisfies (4.9) and $\eta \in T_{\mathbb{M}^N_\lambda}$, then by (3.17) and (3.19) one has

$$\langle\rho,\lambda\rangle = \left\langle \rho\phi_{\eta^+}\, \mathbb{1}_{\phi_{\overline{\mu}}<N}, \lambda\right\rangle - \left\langle \rho\frac{d\eta^-}{d\overline{\mu}}\, \mathbb{1}_{\phi_{\overline{\mu}}<N}, \overline{\mu}\right\rangle - \left\langle \rho\frac{d\eta^-}{d\overline{\mu}}\, \mathbb{1}_{\phi_{\overline{\mu}}=N}, \overline{\mu}\right\rangle \ge 0\,,$$

so that (3.5) implies (4.9). □

4.4 Optimisation of the mark distribution

Assume that the phase space X is represented as a direct product of a Polish space Z called *position space* and a complete separable metric space Q – *mark space*. We are looking for a minimum value of $\mathbf{E}_\mu F$ on the set of such measures μ that are represented as a direct product of a fixed σ-finite measure λ on Borel σ-algebra $\mathfrak{B}(Z)$ of subsets of Z and a probability measure on $\mathfrak{B}(Q)$. Note that λ may not be a finite measure on Z. Let $\mathbb{A}$ be the family of measures $\mu = \lambda\times\kappa$ obtained for all finite measures κ. Consider the minimisation problem

$$\mathbf{E}_\mu F \to \inf \quad \text{for } \mu \in \mathbb{A}, \tag{4.10}$$

subject to

$$\begin{cases} H_0(\mu) = \mu(E\times Q) - \lambda(E) = 0\,; & \\ H_i(\mu) = 0, & i = 1,\dots,m\,; \\ H_j(\mu) \le 0 & j = m+1,\dots,k\,. \end{cases} \tag{4.11}$$

for some $E \subseteq Z$ with $0 < \lambda(E) < \infty$. Note that the first constraint yields $\kappa(Q) = 1$. Assume that for all Π one has $F(\Pi) = F(\Pi|_W)$, where $W \subseteq X$ is a fixed Borel set with $\mu(W) < \infty$. Assume that (2.2) holds with a bounded g, so that Corollary 2.2 is applicable and $\mathbf{E}_\mu F(\Pi)$ is differentiable with respect to μ.

Similarly to Lemma 3.3, it is possible to deduce that

$$T_{\mathbb{A}}(\lambda\times\overline{\kappa}) \supseteq \{\mu=\lambda\times\kappa\,:\,\kappa^-\ll\overline{\kappa} \text{ and } \frac{d\kappa^-}{d\overline{\kappa}} \text{ is bounded}\}\,.$$

Following the lines of the proof of Theorem 3.5 we find that if $\overline{\mu}=\lambda\times\overline{\kappa}$ is a regular minimum point for Problem (4.10) and the derivatives of the constraints can be represented in the form $H'(\overline{\mu})=\langle h,\bullet\rangle$ for some $h=(\mathbb{1}_{E\times Q},h_1,,\dots,h_k)\,:\,X\to\mathbb{R}^{k+1}$, then there exists $u=(u_0,u_1,\dots,u_k)\in\mathbb{R}^{m+1}\times\mathbb{R}^{k-m}_-$ such that

$$\begin{cases}\int_Z[\overline{\Delta}_{\overline{\mu}}(z,q)-u\cdot h(z,q)]\,\lambda(dz)=0 & \text{for } \overline{\kappa}\text{-almost all } q\in Q\,,\\ \int_Z[\overline{\Delta}_{\overline{\mu}}(z,q)-u\cdot h(z,q)]\,\lambda(dz)\ge 0 & \text{for all } q\in Q\,.\end{cases} \tag{4.12}$$

Here the inequality is obtained for $\eta=\lambda\times\delta_q$, while the identity can be deduced taking $\eta=\pm\,\lambda\times\varepsilon\overline{\kappa}$. If $k=0$, then (4.12) implies

$$\begin{cases}\int_Z\overline{\Delta}_{\overline{\mu}}(z,q)\lambda(dz)=u & \text{for } \overline{\kappa}-\text{almost all } q\in Q\,,\\ \int_Z\overline{\Delta}_{\overline{\mu}}(z,q)\lambda(dz)\ge u & \text{for all } q\in Q\end{cases} \tag{4.13}$$

for some $u\in\mathbb{R}$, which may depend on E.

The situation simplifies if λ is a finite Haar measure for a group of shifts in Z. Define $\theta_z B=z+B$ for any $B\in\mathfrak{B}(Z)$ and $z\in Z$. This group of shifts in Z generates a family of shifts acting on a measure γ in Z by the rule: $(\theta_z\gamma)(B)=\gamma(\theta_z B)$ for all $B\in\mathfrak{B}(Z)$; and a family of shifts acting on $\mu=\gamma\times\kappa\in\mathbb{M}$ by $(\theta_z\mu)(B\times K)=\mu(\theta_z B\times K)$ for all $B\times K\in\mathfrak{B}(Z)\otimes\mathfrak{B}(Q)$.

Theorem 4.2. *Assume that for all $z\in Z$ one has $\theta_z\lambda=\lambda$, $F(\theta_z\Pi)=F(\Pi)$ and the derivatives of the constraints $h_i(z,q)$, $i=1,\dots,k$, do not depend on z. Then if $\overline{\mu}=\lambda\times\overline{\kappa}$ is a regular local minimum in Problem (4.10), $H=(H_0,H_1,,\dots,H_k)$ is twice Fréchet differentiable at $\overline{\mu}$ with the derivatives allowing representation $H'(\overline{\mu})=\langle h,\bullet\rangle$ for some $h=(1,h_1,,\dots,h_k)\,:\,Q\to\mathbb{R}^{k+1}$, then $\overline{\Delta}_{\overline{\mu}}(z,q)$ does not depend on z and there exists $u\in\mathbb{R}^{m+1}\times\mathbb{R}^{k-m}_-$ such that*

$$\begin{cases}\overline{\Delta}_{\overline{\mu}}(z,q)=u\cdot h(q) & \overline{\kappa}-\text{a.e.}\,,\\ \overline{\Delta}_{\overline{\mu}}(z,q)\ge u\cdot h(q) & \text{for all } q\in Q\,.\end{cases} \tag{4.14}$$

Proof. It suffices to note that

$$\mathbf{E}_{\lambda\times\kappa}\,F(\theta_z\Pi)=\mathbf{E}_{\theta_z\lambda\times\kappa}\,F(\Pi)=\mathbf{E}_{\lambda\times\kappa}\,F(\Pi)\,.$$

Thus the integral in (4.12) equals $\lambda(Z)(\overline{\Delta}_{\overline{\mu}}(z,q)-u\cdot h(q))$ and the statement of the lemma follows easily. □

4.5 The high intensity case

Consider optimisation problem (4.1) and assume that a measure $a\mu_a$ minimises $\mathbf{E}_\nu F(\Pi)$ over all measures ν with the total mass a. Unfortunately, in most cases the problem of finding explicitly the optimal measure is hardly achievable, as examples in Section 5 readily illustrate. However, even if the measures μ_a are not available for each fixed a, in many cases it is possible to find their limit μ as $a \to \infty$ (assuming that μ_a converge to μ in some sense). It is natural to call this limiting measure μ a *high intensity solution* of Problem (4.1), so that $a\mu$ becomes 'asymptotically optimal solution' of the problem with the fixed large total mass a.

We concentrate here on the case when the phase space X is a compact subset of Euclidean space $\mathbb{R}^d$ such that X coincides with the closure of its interior $\overset{\circ}{X}$. In this section λ denotes the d-dimensional Lebesgue measure.

Let γ_a^x denote the homothetic transformation of $\mathbb{R}^d$ with the centre $x \in \mathbb{R}^d$ and the coefficient $a^{1/d}$. These transformations of the phase space generate corresponding transformations of measures defined by $(\gamma_a^x \nu)(\bullet) = \nu(\gamma_a^x \bullet)$. In particular, if $\Pi = \sum_i \delta_{x_i}$ then $\gamma_a^x \Pi = \sum_i \delta_{\gamma_{a^{-1}}^x x_i}$. It is easy to see that for any measure ν the distribution of $\gamma_a^x \Pi$ under $\mathbf{P}_\nu$ coincides with the distribution of Π under $\mathbf{P}_{\gamma_a^x \nu}$ (cf. [1, Prop.2.1]). Therefore for any measurable functional $F(\Pi)$ one has

$$\mathbf{E}_{a\mu_a} F(\Pi) = \mathbf{E}_{\hat{\mu}_a^x} F(\gamma_a^x \Pi)\,, \tag{4.15}$$

where $\hat{\mu}_a^x(\bullet) = a\mu_a(\gamma_{a^{-1}}^x \bullet)$. This change of the phase space formula is the key to our method. The idea is the following. Assume that μ_a are absolutely continuous with respect to λ for all sufficiently large a with the corresponding densities p_a. For each a, the measure $\hat{\mu}_a^x$ is concentrated on the set $\gamma_a^x X$ and has the density $p_a(\gamma_{a^{-1}}^x y)$ there. If, for simplicity, all $p_a(x)$ are continuous at x and converge as $a \to \infty$ to some limit $p(x) > 0$, then the restriction $\hat{\mu}_a^x|_W$ of the measure $\hat{\mu}_a^x$ onto any compact set $W \subseteq X$ containing x inside, converges in total variation to the measure $p(x)\lambda|_W$ being the uniform measure with density $p(x)$ on W. One may expect that if $\Delta(x; \gamma_a^x \Pi) \propto g(a)\Gamma(x; \Pi)$ for some normalising function $g(a)$ and a random variable $\Gamma(x; \Pi)$ that 'depends mainly on $\Pi|_W$', then

$$\overline{\Delta}_{a\mu_a}(x) = \mathbf{E}_{a\mu_a} \Delta(x; \Pi) = \mathbf{E}_{\hat{\mu}_a^x} \Delta(x; \gamma_a^x \Pi) \propto g(a)\, \mathbf{E}_{p(x)\lambda} \Gamma(x; \Pi)\,.$$

Since by (4.2) the left-hand side is a constant, then it might be possible to find $p(x)$ by equating the right-hand side $\mathbf{E}_{p(x)\lambda} \Gamma(x; \Pi)$ to a constant.

To make explicit what is meant by 'depends mainly on $\Pi|_W$', we need the concept of a stopping set that is a multi-dimensional analogue of a stopping (or optional) time, well known for temporal stochastic processes.

Let $\mathcal{F}_B$ be the σ-algebra generated by the events of the form $\{\nu \in \mathcal{N} : \nu(A \cap B) = n\}$, where A is Borel and $n \geq 0$. A *random* compact set $S = S(\Pi)$ is called a *stopping set* (with respect to filtration $\{\mathcal{F}_K\}$) if $\{S \subseteq K\} \in \mathcal{F}_K$ for any compact $K \subseteq \mathbb{R}^d$. Each deterministic compact set is a stopping set. Probably, the simplest non-trivial example of a stopping set is the random closed ball centred at a fixed point x_0 with the radius given by the distance from x_0 to its k-th closest neighbour from Π (if Π contains less than k points, $S(\Pi)$ is equal to a fixed compact set, e.g., X). The stopping σ-algebra is the following collection:

$$\mathcal{F}_S = \{\Sigma \in \mathcal{F}_X \, : \, \Sigma \cap \{S \subseteq K\} \in \mathcal{F}_K \text{ for all compact } K\} \, .$$

(For details and properties of the stopping sets in the Poisson framework see [30]).

Now we are ready to formulate the main result of this section.

Theorem 4.3. *Let $a\mu_a$ be a family of measures solving the Problem (4.1) for the fixed total mass a. Assume that for an inner point $x \in \overset{\circ}{X}$ the following condition on μ_a holds:*

(M) *for all sufficiently large a, μ_a are absolutely continuous with respect to λ with densities p_a; and there exists a finite double limit*

$$\lim_{\substack{y \to x \\ a \to \infty}} p_a(y) = p(x) > 0 \, . \tag{4.16}$$

Furthermore, assume that for the same x the random variable $\Delta(x; \Pi)$ satisfies the following conditions:

(D) *(Path differentiability) For some positive function $g(a) = g(a, x)$ and $\mathbf{P}_\lambda$-almost all Π the variable $\Gamma_a = \Gamma_a(x; \Pi) = g^{-1}(a)\Delta(x; \gamma_a^x \Pi)$ has a limit $\Gamma = \Gamma(x; \Pi)$ as $a \to \infty$ such that*

$$0 < \mathbf{E}_{p(x)\lambda} \, \Gamma(x; \Pi) < \infty \, . \tag{4.17}$$

(L) *(Localisation) There exist a family of compact stopping sets $S_a = S_a(x; \Pi)$ and a compact stopping set $S = S(x; \Pi)$ such that $\Gamma_a(x; \Pi)$ is $\mathcal{F}_{S_a}$-measurable for all $a \geq A$; $\Gamma(x; \Pi)$ is $\mathcal{F}_S$-measurable; and for any compact set W containing x in its interior*

$$\mathbb{1}_{S_a(x;\Pi) \subseteq W} \to \mathbb{1}_{S(x;\Pi) \subseteq W} \text{ as } a \to \infty \text{ for } \mathbf{P}_\lambda\text{-almost all } \Pi \, . \tag{4.18}$$

(UI) *(Uniform integrability) There exists a compact set W containing x in its interior such that*

$$\lim_{\substack{a \to \infty \\ b \to \infty}} \mathbf{E}_{\hat{\mu}_a^x} |\Gamma_a| \, \mathbb{1}_{S_a \not\subseteq \gamma_b^x W} = 0 \, . \tag{4.19}$$

In addition, there exists a constant $M = M(W, b)$ *such that*

$$|\Gamma_a(x;\Pi)| \le M \text{ for all } a \ge A \text{ and } \Pi \text{ such that } S_a(x;\Pi) \subseteq \gamma_b^x W. \tag{4.20}$$

Then,

$$\lim_{a\to\infty} \left|\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a(x;\Pi) - \mathbf{E}_{p(x)\lambda}\,\Gamma(x;\Pi)\right| = 0 \tag{4.21}$$

and

$$\lim_{a\to\infty} \frac{\overline{\Delta}_{a\mu_a}(x)}{\overline{\Delta}_{ap(x)\lambda}(x)} = 1\,. \tag{4.22}$$

Proof. Let the conditions of the theorem hold for $x \in \overset{\circ}{X}$. Consider W from **(UI)** and denote $W_b = \gamma_b^x W$, for short. Then

$$\begin{aligned}|\,\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a - \mathbf{E}_{p(x)\lambda}\,\Gamma| \le |\,\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \mathbf{E}_{p(x)\lambda}\,\Gamma\,\mathbb{1}_{S\subseteq W_b}\,| \\ + \mathbf{E}_{\hat\mu_a^x}\,|\Gamma_a|\,\mathbb{1}_{S_a\not\subseteq W_b} + \mathbf{E}_{p(x)\lambda}\,|\Gamma|\,\mathbb{1}_{S\not\subseteq W_b}\,.\end{aligned}$$

By (4.17) and (4.19) the last two terms can be made arbitrarily small for all sufficiently large a by an appropriate choice of b. For that b the first summand is bounded from above by

$$\begin{aligned}&|\,\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \mathbf{E}_{p(x)\lambda}\,\Gamma\,\mathbb{1}_{S\subseteq W_b}\,| \\ &\quad\le |\,\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \mathbf{E}_{p(x)\lambda}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b}\,| + \mathbf{E}_{p(x)\lambda}\,|\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \Gamma\,\mathbb{1}_{S\subseteq W_b}\,|\end{aligned} \tag{4.23}$$

The second term in the right-hand side of (4.23) vanishes as $a \to \infty$ by (4.18) and the bounded convergence theorem.

To estimate the first summand in (4.23) we can use the stopping property, due to which $\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b}$ is $\mathcal{F}_{W_b}$-measurable. Since for Poisson processes the restriction of $\mathbf{P}_{\hat\mu_a^x}$ onto $\mathcal{F}_{W_b}$ coincides with $\mathbf{P}_{\hat\mu_a^x|_{W_b}}$, then (4.20) yields

$$\begin{aligned}|\,\mathbf{E}_{\hat\mu_a^x}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \mathbf{E}_{p(x)\lambda}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b}\,| &= |\,\mathbf{E}_{\hat\mu_a^x|_{W_b}}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b} - \mathbf{E}_{p(x)\lambda|_{W_b}}\,\Gamma_a\,\mathbb{1}_{S_a\subseteq W_b}\,| \\ &\le M\|\mathbf{P}_{\hat\mu_a^x|_{W_b}} - \mathbf{P}_{p(x)\lambda|_{W_b}}\| \le 2M\|\hat\mu_a^x|_{W_b} - p(x)\lambda|_{W_b}\|\end{aligned} \tag{4.24}$$

(cf. [13, Prop. 1.6.26] for the last inequality). But

$$\|\hat\mu_a^x|_{W_b} - p(x)\lambda|_{W_b}\| = \int\limits_{W_b} |p_a(\gamma_{a^{-1}}^x y) - p(x)|\,dy\,.$$

By Condition **(M)** the function under the integral is uniformly bounded for all sufficiently large a. Thus by Fatou's lemma

$$\limsup_{a\to\infty} \|\hat{\mu}_a^x|_{W_b} - p(x)\lambda|_{W_b}\| \le \int_{W_b} \limsup_{a\to\infty} |p_a(\gamma_{a^{-1}}^x y) - p(x)|\, dy\,.$$

Again, Condition **(M)** guarantees that the expression under the integral is zero and thus (4.21) follows.

Finally, (4.17) yields

$$\frac{\overline{\Delta}_{a\mu_a}(x)}{\overline{\Delta}_{ap(x)\lambda}(x)} = \frac{\mathbf{E}_{\hat{\mu}_a^x}\Gamma_a(x;\Pi)}{\mathbf{E}_{p(x)\lambda}\Gamma(x;\Pi)}\,,$$

and (4.22) follows from (4.21). □

Corollary 4.4. *For λ-almost all x satisfying the conditions of Theorem 4.3 with the function $g(a)$ not depending on x, one has*

$$\mathbf{E}_{p(x)\lambda}\Gamma(x;\Pi) = \text{const}\,. \tag{4.25}$$

Proof. Let X_0 denote the set of all $x \in X$ satisfying the conditions of Theorem 4.3 with the function $g(a)$ not depending on x. Assume that $\lambda(X_0) > 0$ (otherwise, the statement of the corollary is trivial). Let

$$X_\varepsilon = \bigcup_{A>0} \bigcap_{a\ge A} \{x:\ p_a(x) > \varepsilon\} \cap X_0\,.$$

Since Condition **(M)** is satisfied for λ-almost all $x \in X_0$ then up to a null-set we have $X_\varepsilon \uparrow X_0$ as $\varepsilon \downarrow 0$. Thus there exists $\varepsilon > 0$ such that $\lambda(X_\varepsilon) > 0$. By (4.2)

$$\overline{\Delta}_{a\mu_a}(x) = g(a)\,\mathbf{E}_{\hat{\mu}_a^x}\Gamma_a(x;\Pi) = \text{const}$$

for μ_a-almost all x. Therefore, $\mathbf{E}_{\hat{\mu}_a^x}\Gamma_a(x;\Pi) = C_a$ μ_a-almost everywhere for some constants C_a. Take a sequence $\sigma = \{a_n\}$ such that $a_n \to \infty$. Since $\mu_a \ll \lambda$ then 'for μ_a-almost all x' effectively means 'for λ-almost all x such that $p_a(x) > 0$'. Therefore there exists a set $\Theta(\sigma)$ with $\lambda(\Theta(\sigma)) = 0$ such that $\mathbf{E}_{\hat{\mu}_{a_n}^x}\Gamma_{a_n}(x;\Pi) = C_{a_n}$ and (4.21) hold for all $x \in \Theta^{\mathbf{c}}(\sigma) \cap X_\varepsilon$. Thus there exists $\lim_{n\to\infty} C_{a_n} = C(\sigma)$ that is finite by (4.17), so that the functions $\mathbf{E}_{\hat{\mu}_{a_n}^x}\Gamma_{a_n}(x;\Pi)$ are uniformly bounded for λ-almost all $x \in X_\varepsilon$. Therefore,

understanding the limits below as λ-almost everywhere limits, for any measurable $B \subseteq X_\varepsilon$ we have

$$\lambda(B)C(\sigma) = \lambda(B) \lim_{n\to\infty} C_{a_n} = \lim_{n\to\infty} \int_B \mathbf{E}_{\hat{\mu}^x_{a_n}} \Gamma_{a_n}(x;\Pi)\, \lambda(dx)$$
$$= \int_B \lim_{n\to\infty} \mathbf{E}_{\hat{\mu}^x_{a_n}} \Gamma_{a_n}(x;\Pi)\, \lambda(dx) = \int_B \mathbf{E}_{p(x)\lambda}\, \Gamma(x;\Pi)\, \lambda(dx)\,, \quad (4.26)$$

implying $\mathbf{E}_{p(x)\lambda}\, \Gamma(x;\Pi) = C(\sigma)$ for λ-almost all $x \in X_\varepsilon$. But the last quantity in (4.26) does not depend on σ, hence $C(\sigma)$ does not neither. Thus (4.25) holds for λ-almost all $x \in X_\varepsilon$ and, thereupon, for λ-almost all $x \in X_0$ if $\varepsilon \downarrow 0$. □

The formula (4.25) is useful because the expectation $\mathbf{E}_{p(x)\lambda}\, \Gamma(x;\Pi)$ is usually simpler to obtain explicitly, since this expectation is calculated for a *homogeneous* Poisson process. As we show in the examples below, the density, $p(x)$, of the high-intensity solution μ, can be easily derived from (4.25).

Note, also, that Condition **(M)** can be relaxed to

(M′) there exists a number $p(x) > 0$ such that

$$\kappa_{d-1}\{s \in \mathbb{S}^{d-1} :\ \limsup_{a\to\infty} |p_a(x + a^{-1/d}s) - p(x)| > 0\} = 0\,,$$

where κ_{d-1} is a standard Haar measure on the sphere $\mathbb{S}^{d-1}$.

It should be noted that the absolute continuity of the measures μ_a imposed by Condition **(M)** was used only in showing that the term (4.24) vanishes. This can also be the case under rather weaker conditions like the weak convergence of $\hat{\mu}^x_a$ to $\mu = p(x)\lambda$ and additional assumptions of the smoothness of the functionals $\Gamma_a(x;\Pi)$. If it is true then Corollary (4.4) could be used to find $p(x)$, even if the measures μ_a are atomic but the atoms are accumulating with the density $p(x)$ as $a \to \infty$.

Usually a difficult part in specific applications of Theorem 4.3 is establishing the uniform integrability Condition **(UI)**. Theorem 4.6 below gives useful sufficient conditions that simplify verification of **(UI)** for the measures μ_a satisfying Condition **(MG)** below and a class of stopping sets that we call *thick*. Note that **(MG)** is a 'global' variant of **(M)**.

(MG) There exists a positive constant N such that

(i) $p_a(x) \le N$ for all $x \in X$ and all sufficiently large a;
(ii) Condition **(M)** is satisfied for λ-almost all $x \in X$.

Definition 4.1. Compact set K is called (x,α)-*thick* if there exists a closed ball $B \subseteq K$ containing x such that $\lambda(B) \geq \alpha\lambda(K)$.

Lemma 4.5. *Assume that* **(MG)** *is satisfied, x is such that $p(x) > 0$ and a compact K containing x is (x,α)-thick for some $\alpha > 0$. Then there exists $\delta > 0$ such that*

$$\hat{\mu}_a^x(K) \geq \delta\lambda(K)\,. \tag{4.27}$$

Proof of Lemma. Denote $U_m = \{z \in X : \liminf_{a\to\infty} p_a(z) \geq 2m\}$. By Condition **(MG)** for sufficiently small $\varepsilon > 0$ there exists $0 < m' < p(x)/2$ such that

(i) $B_{2\varepsilon}(x) \subset U_m$ for all $m < m'$.

Since $U_m^{\mathbf{c}} \downarrow \Theta = \{z \in X : \liminf_{a\to\infty} p_a(z) = 0\}$ as $m \downarrow 0$, then **(MG)** implies $\lim_{a\to\infty} \lambda(U_m^{\mathbf{c}}) = \lambda(\Theta) = 0$. Therefore

(ii) $X \setminus U_m$ does not contain a ball of radius $\varepsilon > 0$ for all sufficiently small m.

Choose now m so that both **(i)** and **(ii)** above hold. Since X coincides with the closure of its interior, the function

$$f(r,y) = \frac{\lambda(B_r(y) \cap U_m)}{\lambda(B_r(y) \cap X)}$$

is continuous for $r > 0$ and $y \in X$. By the assumption of the lemma, K is (x,α)-thick, so that there exists a ball $B_r(y) \subseteq K$ containing x. Let us show that $\inf\{f(r,y) : x \in B_r(y)\} > 0$. For this, note that if $r < \varepsilon$ and $x \in B_r(y)$, then $B_r(y) \subset B_{2\varepsilon}(x)$ whence $f(r,y) = 1$. Also, $f(r,y) = \lambda(U_m)/\lambda(X) = \text{const} > 0$ if $r > \operatorname{diam} X$. Furthermore, $\{(r,y) : \varepsilon \leq r \leq \operatorname{diam} X,\ y \in X, x \in B_r(y)\}$ is compact in $[0,\infty) \times \mathbb{R}^d$. Since the function $f(r,y)$ is continuous and non-vanishing on this set, its infimum is strictly greater than zero. Thus,

$$f(r,y) \geq \delta' > 0$$

for all $r > 0$ and $y \in X$ such that $x \in B_r(y)$. By definition of U_m, the functions $p_a(x)$ are bounded from below by m for all sufficiently large a and $x \in U_m$ that yields the following sequence of estimates:

$$\begin{aligned}\hat{\mu}_a^x(K) \geq \hat{\mu}_a^x(B) &= a\mu_a(\gamma_{a^{-1}}^x B) \geq a\mu_a(\gamma_{a^{-1}}^x B \cap U_m) \\ &\geq am\lambda(\gamma_{a^{-1}}^x B \cap U_m) \geq am\delta'\lambda(\gamma_{a^{-1}}^x B) = m\delta'\lambda(B) \geq \alpha m\delta'\lambda(K)\,,\end{aligned}$$

where $B = B_r(y)$, for short. So the statement of the lemma follows with $\delta = \alpha m\delta' > 0$. □

Theorem 4.6. *Assume that the measures μ_a satisfy Condition* ***(MG)****. Assume also that in notation of Theorem 4.3 for λ-almost all $x \in X$*

(i) there exists a compact set W containing x in its interior such that for all $\theta > 0$ and $\varepsilon > 0$

$$\lim_{\substack{a\to\infty \\ b\to\infty}} \mathbf{E}_{\theta\lambda} \, |\Gamma_a|^{1+\varepsilon} \, \mathbb{1}_{S \not\subseteq \gamma_b^x W} = 0 \, ; \tag{4.28}$$

(ii) there exists $\alpha > 0$ such that for $\mathbf{P}_\lambda$-almost all Π and all sufficiently large a one has $S_a(x;\Pi) \subseteq S(x;\Pi)$ and $S(x;\Pi)$ is (x,α)-thick.

Finally, assume that at least one of the following two conditions is satisfied:

(iii_1) for δ given by Lemma 4.5 with $K = S(x;\Pi)$ and N as in ***(MG)*** *there exists $\varepsilon > 0$ such that*

$$\mathbf{E}_{\delta\lambda} \left(\frac{N}{\delta}\right)^{(1+\varepsilon)\Pi(S(x;\Pi))} < \infty \, ;$$

(iii_2) for N as in ***(MG)*** *there exists $\varepsilon > 0$ such that*

$$\mathbf{E}_{N\lambda} \, e^{(1-\varepsilon)N\lambda(S(x;\Pi))} < \infty \, .$$

Then Condition ***(UI)*** *is satisfied for λ-almost all $x \in X$.*

Proof. For the proof we need the following change of the intensity measure formula proved in [30, Prop.2]. Let S be a compact stopping set, $f(\Pi)$ is $\mathcal{F}_S$-measurable and ν_1, ν_2 be two σ-finite measures such that $\nu_1 \ll \nu_2$ with density $q(x)$. Then

$$\mathbf{E}_{\nu_1} f(\Pi) = \mathbf{E}_{\nu_2} f(\Pi) \exp\big\{\nu_2(S(\Pi)) - \nu_1(S(\Pi))\big\} \prod_{x_i \in \Pi \cap S(\Pi)} q(x_i) \, . \tag{4.29}$$

By *(ii)* we have $S_a = S_a(x;\Pi) \subseteq S(x;\Pi) = S$. Therefore $\mathcal{F}_{S_a} \subseteq \mathcal{F}_S$, whence Γ_a is $\mathcal{F}_S$-measurable for all a. Using (4.29) and Lemma 4.5 for the set S we can write

$$\begin{aligned} \mathbf{E}_{\hat\mu_a^x} |\Gamma_a| \, \mathbb{1}_{S_a \not\subseteq W_b} \le \mathbf{E}_{\hat\mu_a^x} |\Gamma_a| \, \mathbb{1}_{S \not\subseteq W_b} &= \mathbf{E}_\lambda |\Gamma_a| \, \mathbb{1}_{S \not\subseteq W_b} \exp\{\lambda(S) - \hat\mu_a^x(S)\} \prod_{x_i \in \Pi \cap S} p_a(\gamma_{a^{-1}}^x x_i) \\ &\le \mathbf{E}_\lambda |\Gamma_a| \, \mathbb{1}_{S \not\subseteq W_b} \exp\{(1-\delta)\lambda(S)\} N^{\Pi(S)} \, , \end{aligned} \tag{4.30}$$

where N is the same as in Condition **(MG)** and $W_b = \gamma_b^x W$, for short. Then, writing $N^{\Pi(S)}$ as $\delta^{\Pi(S)}\,(N/\delta)^{\Pi(S)}$ we can again apply (4.29) to see that (4.30) equals

$$\mathbf{E}_{\delta\lambda}\,|\Gamma_a|\,\mathbb{1}_{S\not\subseteq W_b}\left(\frac{N}{\delta}\right)^{\Pi(S)} \le \left[\mathbf{E}_{\delta\lambda}\,|\Gamma_a|^{\frac{1+\varepsilon}{\varepsilon}}\,\mathbb{1}_{S\not\subseteq W_b}\right]^{\frac{\varepsilon}{1+\varepsilon}}\,\left[\mathbf{E}_{\delta\lambda}(N/\delta)^{(1+\varepsilon)\Pi(S)}\right]^{\frac{1}{1+\varepsilon}},$$

for any $\varepsilon > 0$ by Hölder's inequality. The second factor is finite if Condition *(iii$_1$)* holds, while the first is uniformly in $a \ge A$ small, so Condition **(UI)** is satisfied.

If Condition *(iii$_2$)* is satisfied then apply (4.29) to write (4.30) as

$$\mathbf{E}_{N\lambda}\,|\Gamma_a|\,\mathbb{1}_{S\not\subseteq W_b}\exp\{(N-\delta)\lambda(S)\} \le \left[\mathbf{E}_{N\lambda}\,|\Gamma_a|^{\frac{1+\varepsilon}{\varepsilon}}\,\mathbb{1}_{S\not\subseteq W_b}\right]^{\frac{\varepsilon}{1+\varepsilon}}\,\left[\mathbf{E}_{N\lambda}\exp\{(1+\varepsilon)(N-\delta)\lambda(S)\}\right]^{\frac{1}{1+\varepsilon}}.$$

Here the second factor is finite by *(iii$_2$)* provided ε is chosen so small that $(1+\varepsilon)(N-\delta) < N$ and **(UI)** follows similarly. □

5 Applications and examples

In this section we give, first, a few illustrative examples of applications of our methods before passing to more serious models.

5.1 Symmetric statistics

Consider two examples where the functional F is related to symmetric statistics defined on a Poisson process.

Example 5.1. (*First-order symmetric statistics.*) Let X be a subset of $\mathbb{R}^d$. The first order symmetric statistic $F(\Pi) = \sum_{x_i\in\Pi} f(x_i)$ with a bounded function $f(x)$, $x \in X$, is a trivial example. In this case $\overline{\Delta}_\mu(x) = f(x)$ and $\overline{\Delta^2_\mu}(x,y) = 0$ for all μ. The fixed total mass problem is trivial, so that the necessary condition is validated by any measure concentrated on the set of global minimum points of $f(x)$, $x \in X$.

Example 5.2. (*Hard-core type functional.*) Let X be a convex closed subset of $\mathbb{R}^d$ and let

$$F(\Pi) = \frac{1}{2}\sum_{x_i,x_j\in\Pi,\,i\ne j} \mathbb{1}_{\|x_i-x_j\|\le r}$$

for some $r > 0$. The functional $F(\Pi)$ appears, for instance, in the definition of the Strauss process [27]. This functional can be represented as the second order symmetric statistic

on Π (see Example 2.1) with $f(x,y) = \mathbb{1}_{\|x-y\|\le r}$. Clearly, (2.2) holds and $\overline{\Delta}_\mu(x) = \mu(B_r(x)\cap X)$, where $B_r(x)$ is the ball of radius r centred at x. The necessary condition for the problem with the fixed total mass is

$$\begin{cases} \overline{\mu}(B_r(x)\cap X) = u & \overline{\mu}-\text{a.e.}, \\ \overline{\mu}(B_r(x)\cap X) \ge u & \text{for all } x\in X\,. \end{cases}$$

It is hard to express explicitly the optimal measure in this model. A numerical solution can, however, be obtained using the steepest descent type algorithm described in Section 6 below.

5.2 Optimal design of experiments

Consider the following linear regression model. Assume that at any $x\in X$ it is possible to observe a random variable $y(x)$ represented as

$$y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + dw(x) = f(x)\beta + dw(x)\,, \tag{5.1}$$

where $\beta = (\beta_1,\dots,\beta_k)^\top$ is a vector of unknown parameters, $f(x) = (f_1(x),\dots,f_k(x))$ is a row of linearly independent on X functions and $dw(x)$ is an orthogonal white noise, so that $\mathbf{E}\, w(x) = 0$ and $\mathbf{cov}(w(x_1), w(x_2)) = \sigma^2\, \mathbb{1}_{x_1=x_2}$. Given n observation points $x_1,\dots,x_n$ the least square estimator $\widehat{\beta} = \widehat{\beta}(x_1,\dots,x_n)$ minimises

$$\sum_{i=1}^{n} \Big(y(x_i) - \sum_{j=1}^{k} \beta_j f_j(x_i) \Big)^2 . \tag{5.2}$$

The theory of experimental design addresses the problem of choice of the observation points in order to achieve better properties of the estimator $\widehat{\beta}$. As a criterion for optimality one may take, for example, the so-called *generalised variance* that is the determinant of the covariance matrix $\|\,\mathbf{cov}(\widehat{\beta}_i,\widehat{\beta}_j)\|$. Specifically, let $\xi(dx)$ be a probability distribution on X describing the frequency of taking x as an observation point. Then it can be shown that the above covariance matrix equals $\sigma^2 M(\xi)^{-1}$, where

$$M(\xi) = \int f(x)^\top f(x)\, \xi(dx)$$

is the *information matrix* (cf. [2] for details). The measure ξ that maximises $\det M(\xi)$ is called D-*optimal design measure*. Note that the objective functional is a concave functional

on the space of measures, so that any maximum found will certainly be global. The general equivalence theorem by Kiefer and Wolfowitz [15] implies, in particular, that the measure ξ provides the D-optimal design if and only if the function

$$d(x,\xi) = f(x)M^{-1}(\xi)f(x)^\top$$

called the standardised variance of the predicted response at point x, achieves its maxima at the atoms of ξ.

In general, given a goal function $\Psi : \mathbb{R}^{k^2} \mapsto \mathbb{R}$, a *$\Psi$-optimal design measure* ξ minimises $\Psi(M(\xi))$ over all probability measures on X. The technique developed in Section 3 allows us to obtain general equivalence type theorems. Indeed, if $m_{ij} = \int f_i(x)f_j(x)\,\xi(dx)$ denotes the (i,j)-h entry of the information matrix, then the Fréchet derivative of m_{ij} considered as a functional of ξ equals $Dm_{ij}[\bullet] = \langle f_i f_j, \bullet\rangle$. Therefore

$$D\Psi(M(\xi))[\bullet] = \sum_{i,j} \frac{\partial\Psi(M)}{\partial m_{ij}}(\xi) Dm_{ij}(\xi)[\bullet] = \sum_{i,j} \frac{\partial\Psi(M)}{\partial m_{ij}}(\xi)\,\langle f_i f_j, \bullet\rangle = \left\langle f D\Psi(\xi) f^\top, \bullet \right\rangle ,$$

where

$$D\Psi(M(\xi)) = \left\| \frac{\partial\Psi(M)}{\partial m_{ij}}(\xi) \right\|_{ij} .$$

Since the total mass of ξ is fixed to 1 then applying (3.24) we obtain, similarly to (4.2), that the optimal ξ satisfies

$$\begin{cases} f(x)D\Psi(M(\xi))f^\top(x) = u & \xi - \text{a.e.}, \\ f(x)D\Psi(M(\xi))f^\top(x) \ge u & \text{for all } x \in X . \end{cases} \tag{5.3}$$

In particular, if $\Psi(M) = -\log\det M$ then $D\Psi = M^{-1}$ and we obtain the Kiefer-Wolfowitz theorem for D-optimal designs, cf [14]. An advantage of this approach is that one can easily integrate additional constraints on the design measure, like, for instance, the boundedness of the density discussed in Section 4.3, see also [8]. These issues will be discussed in details in a separate paper.

Consider now a problem where the observation points are distributed according to a point process N and one can control them only through the process' intensity measure μ. In this case the fixed total number of points n is replaced by the fixed total mass constraint $\mu(X) = a$. Then it is possible to choose the estimator $\widehat{\beta} = \widehat{\beta}(\mu)$ in order to minimise

$$\mathbf{E}_N \sum_{x_i \in N} \Big(y(x_i) - \sum_{j=1}^k \beta_j f_j(x_i) \Big)^2 = \int_X \Big(y(x) - \sum_{j=1}^k \beta_j f_j(x) \Big)^2 \mu(dx) \tag{5.4}$$

(where we have used the Campbell's theorem for the last equality) and to choose μ afterwards to minimise the generalised variance of $\widehat{\beta}$. From (5.4) it is clear that this approach is equivalent to the construction of a D-optimal design, so that the solution is given by $\mu = a\xi$.

It is also possible to use another approach which is more natural: define $\widehat{\beta}$ as in the standard regression model (5.2) thus letting it to be a functional of a configuration of the process N, and choose then $\overline{\mu}$ that maximises $\mathbf{E} \det M(N)$, where

$$M(N) = \int f(x)^\top f(x)\, N(dx)\,.$$

We will call such $\overline{\mu}$ a P-*optimal design measure*.

If the process $N = \Pi$ is Poisson then putting $F(\Pi) = \det M(\Pi)$ in (4.2) and assuming, for simplicity, that all $f_j(x)$ are bounded on X, we get the following necessary condition for the P-optimal design measure $\overline{\mu}$ (note that this is a *maximisation* problem)

$$\begin{cases} \overline{\Delta}_{\overline{\mu}}(x) = u & \overline{\mu} - \text{a.e.}\,, \\ \overline{\Delta}_{\overline{\mu}}(x) \le u & \text{for all } x \in X\,. \end{cases} \tag{5.5}$$

This condition provides an analogue of the general equivalence theorem for P-optimal designs. The stochastic gradient in this problem is given by

$$\Delta(x;\Pi) = \det M(\Pi + \delta_x) - \det M(\Pi) = \det(M(\Pi) + f(x)^\top f(x)) - \det M(\Pi)\,.$$

Recall that for any $k \times k$ positive definite matrix A and a $1 \times k$-row b one has

$$\frac{\det(A + b^\top b)}{\det A} - 1 = bA^{-1}b^\top\,,$$

see, e.g., [28]. Therefore

$$\begin{aligned} \overline{\Delta}_\mu(x) &= \mathbf{E}_\mu \det M(\Pi)\, f(x) M^{-1}(\Pi) f(x)^\top = \mathbf{E}_\mu \det M(\Pi)\, d(x,\Pi) \\ &= \sum_{i,j=1}^{k} (-1)^{i+j} f_i(x) f_j(x)\, \mathbf{E}_\mu \det \int f_{\hat{\imath}\hat{\jmath}}(y)^\top f_{\hat{\imath}\hat{\jmath}}(y)\, \Pi(dy)\,, \end{aligned}$$

where $f_{\hat{\imath}\hat{\jmath}}(y)$ is the row of $f_m(y)$, $m \in \{1,\dots,k\} \setminus \{i,j\}$. In particular, if $k = 2$, then

$$\overline{\Delta}_\mu(x) = f_1^2(x) \int f_2^2(t)\mu(dt) + f_2^2(x) \int f_1^2(t)\mu(dt) - 2f_1(x)f_2(x) \int f_1(t)f_2(t)\mu(dt)\,. \tag{5.6}$$

If X is a compact subset of $\mathbb{R}$ and all $f_1, \dots, f_k$ are polynomials of the highest order n, then $\overline{\Delta}_\mu(x)$ is a polynomial of order $2n$. Therefore, (5.5) implies that $\overline{\mu}$ has at most $(n+1)$ atoms.

Example 5.3. (Simple linear regression.) Let X be a subset of the real line. If $f_1(x) = 1$ and $f_2(x) = x$, then (5.1) is the simple linear regression model. Then (5.6) yields

$$\overline{\Delta}_\mu(x) = x^2 - 2a_1 x + a_2\,,$$

where

$$a_i = \int_X t^i \mu(dt)\,, \quad i \geq 1\,. \tag{5.7}$$

Therefore if $X = [s,t]$ is a segment, then the optimum design is concentrated at one or both its end-points. If $\mu(\{s\}) = w$, then $\overline{\Delta}_\mu(s) = (s+t)^2(1-w)$ and $\overline{\Delta}_\mu(t) = (s+t)^2 w$. If $w = 0$, then $\overline{\Delta}_\mu(t) < \overline{\Delta}_\mu(s)$, which violates (5.5) because t is a support point of μ. Similarly, the case $w = 1$ is excluded. Thus, $0 < w < 1$, whence $\overline{\Delta}_\mu(s) = \overline{\Delta}_\mu(t)$ and $w = 1/2$. Therefore, the P-optimal μ assigns equal weights to the end-points of the segment independently of its location.

Example 5.4. Let $X = [s,t] \subset \mathbb{R}$ and let $f_1(x) = 1$ and $f_2(x) = g(x)$ for some function g. If g is monotone, it is possible to show that the P-optimal design $\overline{\mu}$ assigns equal weights to the end-points of X. The problem is more interesting if g is not monotone on X. Consider a particular case when $g(x) = x^2$ and $s < 0 < t$. By analysing the expected first order difference

$$\overline{\Delta}_\mu(x) = x^4 - 2a_2 x^2 + a_4$$

it is possible to show that if $|s| < |t|$, then the P-optimal design $\overline{\mu}$ assigns equal weights to 0 and to t. Assume that $X = [-t,t]$ is a symmetric interval. If $t \leq 2^{-1/4}$, then $\overline{\mu}$ assigns equal weights to the end-points of X. Otherwise, $\overline{\mu}(\{-t\}) = \overline{\mu}(\{t\}) = 1/(4t^4)$ and $\overline{\mu}(\{0\}) = 1 - 1/(2t^4)$.

Example 5.5. (Quadratic regression through the origin.) If $X = [s,t] \subset \mathbb{R}$, $f_1(x) = x$ and $f_2(x) = x^2$, then (5.1) yields the quadratic regression through the origin. Now (5.6) yields

$$\overline{\Delta}_\mu(x) = a_2 x^4 - 2a_3 x^3 + a_4 x^2\,,$$

where a_i are given by (5.7). Therefore, the P-optimal design is atomic and has at most 3 atoms. If $X = [0,1]$, then the optimal μ has at most two atoms, one at $\alpha \in (0,1)$ and the

other at point 1. It is easy to see that a one-point design does not satisfy the necessary condition (5.5). Assume that $\mu(\{1\}) = w$. Then $a_i = \alpha^i(1-w) + w$, whence

$$\overline{\Delta}_\mu(x) = (\alpha^2(1-w)+w)x^4 - (\alpha^3(1-w)+w)x^3 + (\alpha^4(1-w)+w)x^2 .$$

Then the derivative of $\overline{\Delta}_\mu$ taken at $x = \alpha$ is given by $2(1-w)\alpha(2\alpha^2 - 3\alpha + 1)$. Since α is the point of a local maximum of $\overline{\Delta}_\mu(x)$, we get $\alpha = 1/2$. By comparison of $\overline{\Delta}_\mu(\alpha)$ and $\overline{\Delta}_\mu(1)$, we can easily find that $w = 1/2$, so that $\overline{\mu}$ gives equal weight to $\alpha = 1/2$ and 1. It is easy to check that

$$\overline{\Delta}_{\overline{\mu}}(x) = \frac{5}{8}x^4 - \frac{9}{8}x^3 + \frac{17}{32}x^2$$

satisfies (5.5).

5.3 Optimal placement of stations

Point processes find applications in modelling of telecommunication networks. In [4] and [10] the following model of the topology of a two level telecommunication system was considered.

Let Π_0, Π be two independent Poisson process on a compact Borel set $X \subset \mathbb{R}^d$ (usually $\mathbb{R}^2$) with the intensity measures ν and μ, respectively. The process Π_0 represents the *subscribers* of telecommunication services and Π represents the *stations* or concentration points. Each station $x_i \in \Pi$ serves its zone $Z_i = Z_i(\Pi)$ consisting of those points of X which are closer to x_i than to any other $x_j \in \Pi$. This system of zones constitutes a *Voronoi tessellation* of X with respect to the point set Π, see, e.g., [26]. Let $f(y_j, x_i)$ denote the cost of the connection of a subscriber $y_j \in \Pi_0$ situated in a zone Z_i to its serving station x_i. Our goal is to find the intensity measure μ of the process of stations Π which minimises the expected total connection cost:

$$\mathbf{E} \sum_{y_j \in \Pi_0} \sum_{x_i \in \Pi} f(y_j, x_i)\, \mathbb{1}_{Z_i}(y_j) \longrightarrow \inf \quad \text{subject to} \quad \mu(X) = a\,, \tag{5.8}$$

where $\mathbf{E}$ is the expectation with respect to the product of the distributions $\mathbf{P}_\nu$ and $\mathbf{P}_\mu$ of the processes Π_0 and Π, respectively.

We start with the simplest one-dimensional variant of this problem where the exact analytical solution is possible to obtain. Assume that $X = [0,1]$; the subscribers are uniformly distributed, i.e. $\nu(dy) = \lambda_0 dy$; and the cost function $f(y,x) = \gamma|y-x|$ is proportional to the distance for some $\gamma > 0$. To exclude the degenerate case when $\Pi([0,1]) = 0$, assume that there are two fixed stations in positions 0 and 1.

Conditioning on Π and then applying Campbell's theorem to the process Π_0 in (5.8), it is easy to reduce (5.8) to the following equivalent problem:

$$\mathbf{E}_\mu F(\Pi) = \lambda_0 \gamma \, \mathbf{E}_\mu \int_0^1 R(y,\Pi) dy \longrightarrow \inf \quad \text{subject to} \quad \mu(X) = a \,, \tag{5.9}$$

where $R(y,\Pi) = \min\{|y - x_i| : x_i \in \Pi \cup \{0,1\}\}$ is the distance from x to the set $\Pi \cup \{0,1\}$.

It is straightforward to verify that the first order difference equals

$$\Delta(x;\Pi) = -\frac{\gamma}{2} r_x^- \, r_x^+ \,, \tag{5.10}$$

where r_x^- is the distance from a point x to its nearest to the left point of the set $\Pi \cup \{0\}$, and r_x^+ is the distance from x to its nearest to the right point of the set $\Pi \cup \{1\}$. By the well known property of Poisson processes, the variables r_x^- and r_x^+ are independent so that $-2\gamma^{-1}\overline{\Delta}_\mu(x) = \mathbf{E}_\mu \, r_x^- \, \mathbf{E}_\mu \, r_x^+$. Assume that $\overline{\mu}$ solves (5.9). Then, by the first order necessary condition (4.2), $\overline{\Delta}_{\overline{\mu}}(x)$ should be a constant, say, u, for $\overline{\mu}$-almost all $x \in [0,1]$.

Note that the functions $\mathbf{E}_\mu \, r_x^-$ and $\mathbf{E}_\mu \, r_x^+$, $x \in (0,1)$, are strictly positive and bounded for all $\mu \in \mathbb{M}$. Since the product of these expectations approaches 0 when x tends either to 0 or to 1, the support of $\overline{\mu}$ must be separated from these points unless we have $u = 0$. Therefore, we see that $\overline{\mu}$ is concentrated on $[\varepsilon_1, \varepsilon_2]$ for some $0 < \varepsilon_1 < \varepsilon_2 < 1$.

Suppose $\overline{\Delta}_{\overline{\mu}}(x_1) = \overline{\Delta}_{\overline{\mu}}(x_2) = u$ for some $x_1 < x_2$. If $\overline{\mu}((x_1,x_2)) = 0$, then

$$r_x^- = r_{x_1}^- + x - x_1 \quad \text{and} \quad r_x^+ = r_{x_2}^+ + x_2 - x$$

for all $x \in (x_1, x_2)$, whence

$$\overline{\Delta}_{\overline{\mu}}(x) = -\frac{\gamma}{2}(\mathbf{E}_{\overline{\mu}} \, r_{x_1}^- + x - x_1)(\mathbf{E}_{\overline{\mu}} \, r_{x_2}^+ + x_2 - x) \,. \tag{5.11}$$

Equation (5.11) shows that the function $\overline{\Delta}_{\overline{\mu}}(x)$, $x_1 \le x \le x_2$, is a convex parabola taking the value u in the points x_1 and x_2. Thus $\overline{\Delta}_{\overline{\mu}}(x) < u$ for all $x \in (x_1, x_2)$ that contradicts the necessary condition (4.2). Thus, the optimal measure $\overline{\mu}$ has a connected support, and $\overline{\Delta}_{\overline{\mu}}(x) = u$ for all x from some segment.

We have

$$\begin{aligned} \mathbf{E}_{\overline{\mu}} \, r_x^- &= \int_0^x \exp\bigl\{-\overline{\mu}[x-y,x]\bigr\} \, dy = \exp\{-\overline{\mu}[0,x]\} \int_0^x \exp\bigl\{\overline{\mu}[0,x-y)\bigr\} \, dy \\ &= \exp\{-\overline{\mu}[0,x]\} \int_0^x \exp\bigl\{\overline{\mu}[0,y)\bigr\} \, dy \,. \end{aligned} \tag{5.12}$$

Similarly,

$$\mathbf{E}_{\overline{\mu}}\, r_x^+ = \int_0^{1-x} \exp\bigl\{-\overline{\mu}[x, x+y]\bigr\}\, dy = \exp\{\overline{\mu}[0,x)\} \int_0^{1-x} \exp\bigl\{-\overline{\mu}[0, x+y]\bigr\}\, dy$$

$$= \exp\{\overline{\mu}[0,x)\} \int_x^1 \exp\bigl\{-\overline{\mu}[0,y]\bigr\}\, dy\,. \tag{5.13}$$

Therefore, introducing functions

$$U(x) = \int_0^x \exp\bigl\{\overline{\mu}[0,y)\bigr\}\, dy \quad \text{and} \quad V(x) = \int_x^1 \exp\bigl\{-\overline{\mu}[0,y]\bigr\}\, dy$$

we obtain

$$\mathbf{E}_{\overline{\mu}}\, r_x^-\, \mathbf{E}_{\overline{\mu}}\, r_x^+ = \exp\bigl\{-\overline{\mu}\{x\}\bigr\} U(x)V(x) = u \quad \overline{\mu} - \text{a.e.}\,. \tag{5.14}$$

Therefore, $\overline{\mu}(\{x\})$ is continuous on a segment which is the support of $\overline{\mu}$, whence $\overline{\mu}(\{x\})$ vanishes identically and $\overline{\mu}$ has no atoms. To meet the identity (5.14) at the boundary points of the support of $\overline{\mu}$, it is necessary that they were symmetric with respect to $1/2$, so that the support of $\overline{\mu}$ is $[\varepsilon, 1-\varepsilon]$ for $0 \le \varepsilon \le 1/2$.

Now (5.14) yields $U(x)V(x) = u$ for all $x \in [\varepsilon, 1-\varepsilon]$ and thus

$$U'(x)V(x) + U(x)V'(X) = uU'(x)/U(x) - U(x)/U'(x) = 0\,,$$

since $V'(x) = -\exp\bigl\{-\overline{\mu}[0,x]\bigr\} = -1/U'(x)$. The solution is given by $U(x) = C/\sqrt{u}\exp\bigl\{x/\sqrt{u}\bigr\}$ implying that the measure $\overline{\mu}$ is *uniform* on $[\varepsilon, 1-\varepsilon]$ with density ε^{-1}. The condition on the total mass $\overline{\mu}([0,1]) = a$ allows us to determine $\varepsilon = (a+2)^{-1}$.

It is also easy to verify that $\overline{\Delta}_{\overline{\mu}}(x) > -u\gamma/2 = \overline{\Delta}_{\overline{\mu}}(\varepsilon)$ for all $x \in [0,\varepsilon) \cup (1-\varepsilon, 1]$, thus condition (4.2) is satisfied for that measure $\overline{\mu}$.

Remark 5.1. If we denote the solution for the problem with the total mass a by $a\mu_a$, then μ_a converges in total variation as $a \to \infty$ to the *uniform* distribution on [0,1] that is the high intensity solution for this problem.

Consider now a multidimensional variant of the problem with non-homogeneous distribution of the subscribers. In this setup even the one-dimensional problem becomes quite difficult to solve analytically and leads to a differential equation of the fourth order. Below we will find the high intensity solution in the multidimensional case when X is a compact subset of $\mathbb{R}^d$ such that X coincides with the closure of its interior.

Assume that the connection cost is equal to the Euclidean distance between the subscriber and the server raised to the power $\beta > 0$ and the intensity measure ν of subscribers has an almost everywhere continuous bounded density p_ν (with respect to the Lebesgue measure λ in $\mathbb{R}^d$) that is positive almost everywhere in X. Finally, assume that Condition **(MG)** of Section 4.5 holds.

It is easy to see that

$$F(\Pi) = \int \rho^\beta(z,\Pi)\nu(dz)\,,$$

where $\rho(z,\Pi)$ is the minimal distance from z to the points of Π. Then for $x \in X$

$$\begin{aligned}\Delta(x;\Pi) &= -\int [\rho^\beta(z,\Pi) - |z-x|^\beta]\, \mathbb{1}_{\rho(z,\Pi)-|z-x|>0}\, \nu(dz) \\ &= -\int\limits_{\mathcal{C}_x(\Pi)} [\rho^\beta(z,\Pi) - |z-x|^\beta]\, \nu(dz)\,,\end{aligned}$$

where $\mathcal{C}_x(\Pi)$ is the Voronoi cell with nucleus x constructed with respect to $\Pi + \delta_x$. Below we use the notation of Theorem 4.3. Putting $z = \gamma^x_{a^{-1}} y$ and using $\mathcal{C}_x(\gamma^x_a \Pi) = \gamma^x_{a^{-1}} \mathcal{C}_x(\Pi)$ we obtain

$$-\Delta(x;\gamma^x_a\Pi) = a^{-1-\beta/d} \int\limits_{\mathcal{C}_x(\Pi)} [\rho^\beta(y,\Pi) - |y-x|^\beta]\, \hat\nu^x_a(dy) = a^{-1-\beta/d}\Gamma_a(x;\Pi)\,, \qquad (5.15)$$

where $\hat\nu^x_a(\bullet) = a\nu(\gamma^x_{a^{-1}}\bullet)$. Since at the continuity points of p_ν

$$\frac{\hat\nu^x_a(dy)}{dy} = p_\nu(\gamma^x_{a^{-1}} y) = p_\nu(x) + o(1)\,,$$

we have $g(a) = a^{-1-\beta/d}$ and

$$\Gamma(x;\Pi) = p_\nu(x) \int\limits_{\mathcal{C}_x(\Pi)} [\rho^\beta(y,\Pi) - |y-x|^\beta]\, dy\,.$$

Applying the change of space formula (4.15) we get for some constant $c > 0$

$$\begin{aligned}\mathbf{E}_{p(x)\lambda}\, \Gamma(x;\Pi) &= p_\nu(x)\, \mathbf{E}_\lambda \int\limits_{\mathcal{C}_x(\gamma^x_{p(x)}\Pi)} [\rho^\beta(y,\gamma^x_{p(x)}\Pi) - |y-x|^\beta]\, dy \\ &= p_\nu(x) p(x)^{-1-\beta/d}\, \mathbf{E}_\lambda \int\limits_{\mathcal{C}_x(\Pi)} [\rho^\beta(z,\Pi) - |z-x|^\beta]\, dz \\ &= c\, p_\nu(x) p(x)^{-1-\beta/d} > 0 \qquad (5.16)\end{aligned}$$

after change $z = \gamma^x_{p(x)} y$, so that the path differentiability condition **(D)** of Theorem 4.3 holds.

The measure $\hat{\nu}^x_a$ is concentrated on the set $\gamma^x_a X$, whence the domain of integration in (5.15) can be reduced to $\mathcal{C}_x(\Pi) \cap \gamma^x_a X$. Let $V_a(x;\Pi)$ denote the set of vertices of the cell $\mathcal{C}_x(\Pi)$ and the intersection points of the cell's one-dimensional faces with the boundary of $\gamma^x_a X$. The set $\mathcal{C}_x(\Pi) \cap \gamma^x_a X$ can be associated with a random compact set $S_a = S_a(x;\Pi)$ being the intersection of $\gamma^x_a X$ with the union of the closed balls containing x in their boundaries and centred in the points of $V_a(x;\Pi)$ (see Figure 1). By definition of the Voronoi cell all these balls have no points of the Poisson process inside. It is readily seen that $S_a(x;\Pi)$ is a stopping set and that it completely determines the geometry of $\mathcal{C}_x(\Pi) \cap \gamma^x_a X$. Therefore, $\Gamma_a(x;\Pi)$ is $\mathcal{F}_{S_a}$-measurable. If all the balls forming S_a lie entirely inside $\gamma^x_a X$ then $S_a(x;\Pi)$ coincides with the so-called *fundamental region* $S(x;\Pi)$ or *the Voronoi flower* associated with the cell (cf. [18]). Since $\mathbf{P}_\lambda$-almost all Voronoi cells are compact then (4.18) is satisfied and thus the localisation property holds.

To show the uniform integrability of the family $\Gamma_a(x;\Pi)$ we use Theorem 4.6. By construction $S_a(x;\Pi) \subseteq S(x;\Pi)$ for all sufficiently large a. Given a compact set K, let $r_x(K) = \sup_{y\in K} |y-x|$. It is easy to see that $r_x(S(x;\Pi))$ equals the diameter of the largest ball, call it $D(x;\Pi)$, forming $S(x;\Pi)$, and is twice $r_x(\mathcal{C}_x(\Pi))$. Since

$$\lambda(S(x;\Pi)) \le b_d r_x(S(x;\Pi))^d = 2^d \lambda(D(x;\Pi))$$

then $\lambda(D(x;\Pi)) \ge 2^{-d}\lambda(S(x;\Pi))$ so that $S(x;\Pi)$ is $(x, 2^{-d})$-thick and thus condition *(ii)* of Theorem 4.6 satisfied.

Next,

$$\mathbf{P}_{\theta\lambda}\{\lambda(S(x;\Pi)) > t\} \le \mathbf{P}_{\theta\lambda}\{\lambda(D(x;\Pi)) > 2^{-d}t\} < e^{-c_d\theta t}$$

for some positive constant c_d depending on the dimension d of the space only, see [10, Lemma 1 and Remark 5]. This bound implies

$$\mathbf{E}_{\theta\lambda}\, e^{\gamma\lambda(S(x;\Pi))} < \infty$$

for all θ and all $\gamma < \theta$, so *(iii$_2$)* is also satisfied.

To show validity of condition *(i)*, note that p_ν is bounded, so that $\sup_{y\in X} |p_\nu(y)| \le L$ for some $L < \infty$. Therefore

$$|\Gamma_a(x;\Pi)| < L \int\limits_{\mathcal{C}_x(\Pi)} \rho^\beta(y,\Pi)\, dy \le L r_x^\beta(\mathcal{C}_x(\Pi))\lambda(\mathcal{C}_x(\Pi)) \le L b_d r_x^{\beta+d}(\mathcal{C}_x(\Pi))\,. \qquad (5.17)$$

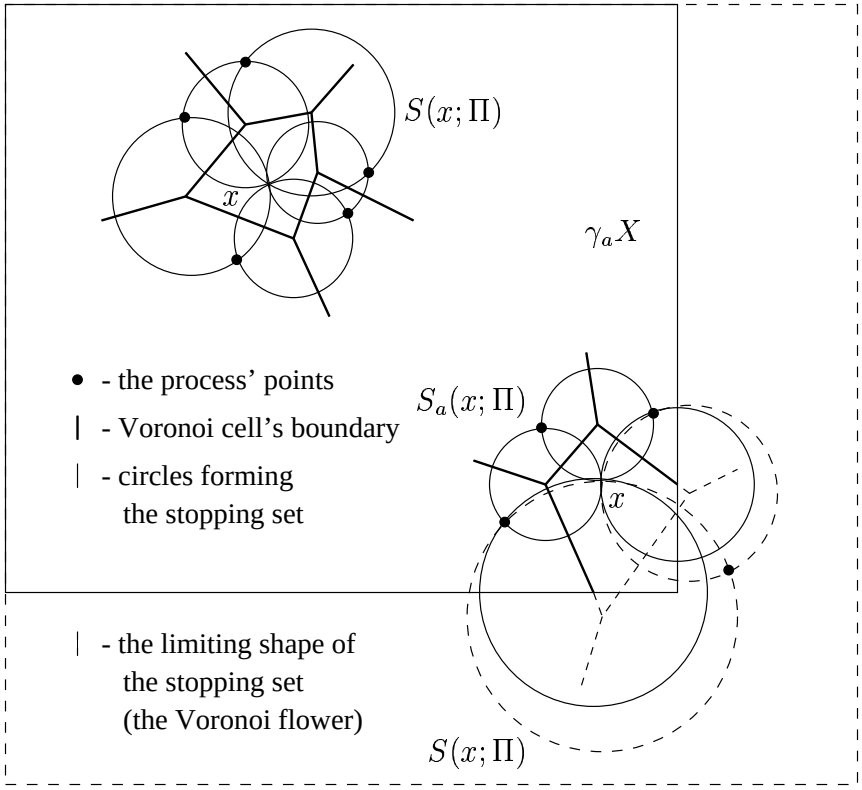

Figure 1: Stopping sets in the optimal placement of stations example.

But

$$\mathbf{P}_{\theta\lambda}\{r_x(\mathcal{C}_x(\Pi)) > t\} \le \mathbf{P}_{\theta\lambda}\{r_x(D(x;\Pi)) > 2t\} = \mathbf{P}_{\theta\lambda}\{\lambda(D(x;\Pi)) > 2^d b_d t^d\} < e^{-2^d b_d \theta t^d}$$

that implies $\mathbf{E}_{\theta\lambda}\,|\Gamma_a|^n < \infty$ for any $n > 0$. We have verified all conditions of Theorem 4.6 that implies **(UI)** condition for λ-almost all $x \in X$.

Finally, (5.17) implies (4.20), Theorem 4.3 is applicable, so that (4.25) and (5.16) yield the density of the high intensity solution

$$p(x) \propto (p_\nu(x))^{d/(d+\beta)}\,.$$

In particular, in the planar case with a linear cost function ($\beta = 1$) the asymptotically optimal density of stations should be proportional to the density of subscribers raised to the power 2/3.

5.4 Approximation of convex functions

The following example is related to $\mathscr{L}_1$-approximation of convex functions. Let $f(x)$, $0 \le x \le 1$, be a twice continuously differentiable convex function. For a point configuration Π on $X = [0,1]$ denote by $g(x;\Pi)$ the piecewise linear function such that $g(x;\Pi) = f(x)$ for all x from $\Pi \cup \{0,1\}$, and otherwise obtained by linear interpolation. By convexity, $g(x;\Pi) \ge f(x)$ for all $x \in [0,1]$.

The functional

$$F(\Pi) = \int_0^1 (g(x;\Pi) - f(x))dx \tag{5.18}$$

characterises the quality of approximation in the $\mathscr{L}_1$-metric. This functional is bounded and, therefore its expectation is analytic as a function on $\mathbb{M}$. Furthermore,

$$\Delta(x;\Pi) = F(\Pi + \delta_x) - F(\Pi) = -\frac{1}{2}\left[-(r_x^- + r_x^+)f(x) + r_x^- f(x + r_x^+) + r_x^+ f(x - r_x^-)\right], \tag{5.19}$$

where $r_x^+ = r_x^+(\Pi)$ and $r_x^- = r_x^-(\Pi)$ have the same meaning as in (5.10). Since r_x^+ and r_x^- are independent,

$$\begin{aligned} -2\overline{\Delta}_\mu(x) &= \Big[- f(x)(\mathbf{E}_\mu\, r_x^- + \mathbf{E}_\mu\, r_x^+) + \mathbf{E}_\mu\, r_x^-\, \mathbf{E}_\mu\, f(x + r_x^+) + \mathbf{E}_\mu\, r_x^+\, \mathbf{E}_\mu\, f(x - r_x^-)\Big] \\ &= \mathbf{E}_\mu\, r_x^- \left[\mathbf{E}_\mu\, f(x + r_x^+) - f(x)\right] + \mathbf{E}_\mu\, r_x^+ \left[\mathbf{E}_\mu\, f(x - r_x^-) - f(x)\right]. \end{aligned}$$

Similarly to (5.12) and (5.13) it is easy to show that

$$\begin{aligned} \mathbf{E}_\mu[f(x - r_x^-) - f(x)] &= -\int_0^x e^{-\mu([x-y,y])} f'(x-y)dy = -\int_0^x e^{-\mu([z,x])} f'(z)dz\,, \\ \mathbf{E}_\mu[f(x + r_x^+) - f(x)] &= \int_0^{1-x} e^{-\mu([x,x+y])} f'(x+y)dy = \int_x^1 e^{-\mu([x,z])} f'(z)dz\,. \end{aligned}$$

Then

$$-2\overline{\Delta}_\mu(x) = e^{-\mu(\{x\})}\Big(U(x)V_f(x) - V(x)U_f(x)\Big), \tag{5.20}$$

where

$$U(x) = \int_0^x e^{\mu([0,y))}\, dy\,, \quad U_f(x) = \int_0^x f'(y) e^{\mu([0,y))}\, dy\,,$$
$$V(x) = \int_x^1 e^{-\mu([0,y])}\, dy\,, \quad V_f(x) = \int_x^1 f'(y) e^{-\mu([0,y])}\, dy\,.$$

It is possible to show that the optimal measure in the minimisation problem (4.1) has no atoms and its support is a connected set. For each fixed a, it is possible to reduce (5.20) to a system of four differential equations by using relationships between the derivatives of U, V and U_f, V_f. In general, it requires numerical solution for any given f and a.

Below we find the high intensity solution that reduces to a routine check of the conditions of Theorem 4.3 and Theorem 4.6. Assume that condition **(MG)** of Section 4.5 is satisfied. Denoting γ_a^x the homothetic transformation with the centre x and the coefficient a, and

$$\begin{aligned} \xi_x^+ &= \xi_x^+(\Pi) = \inf\{y - x :\ y \ge x,\ y \in \Pi\}\,, \\ \xi_x^- &= \xi_x^-(\Pi) = \inf\{x - y :\ y \le x,\ y \in \Pi\}\,, \end{aligned} \tag{5.21}$$

one may write

$$r_x^-(\gamma_a^x \Pi) = a^{-1}\xi_x^-(\Pi)\, \mathbb{1}_{\Pi([-ax+x,x))>0} + x\, \mathbb{1}_{\Pi([-ax+x,x))=0}\,.$$

Similarly,

$$r_x^+(\gamma_a^x \Pi) = a^{-1}\xi_x^+(\Pi)\, \mathbb{1}_{\Pi((x,a(1-x)+x])>0} + (1-x)\, \mathbb{1}_{\Pi((x,a(1-x)+x])=0}\,.$$

Since the second derivative of f is continuous,

$$\begin{aligned} f(x + r_x^+(\gamma_a^x\Pi)) - f(x) = \big[a^{-1}f'(x)\xi_x^+ + a^{-2}\frac{1}{2}f''(x+\theta_x^+)(\xi_x^+)^2\big]\, \mathbb{1}_{\Pi((x,a(1-x)+x])>0} \\ + [f(1) - f(x)]\, \mathbb{1}_{\Pi((x,a(1-x)+x])=0}\,, \end{aligned}$$

where $0 \le \theta_x^+ \le a^{-1}\xi_x^+$. Using a similar expansion for $f(x - r_x^-) - f(x)$, we get

$$\begin{aligned} a^3\Delta(x;\gamma_a^x\Pi) = -\frac{1}{4}\big[f''(x+\theta_x^+)\xi_x^-(\xi_x^+)^2 + f''(x-\theta_x^-)(\xi_x^-)^2\xi_x^+\big]\, \mathbb{1}_{\Pi([-ax+x,x))>0}\, \mathbb{1}_{\Pi((x,a(1-x)+x])>0} \\ + a^3\Delta(x;\gamma_a^x\Pi)\big[1 - \mathbb{1}_{\Pi([-ax+x,x))>0}\, \mathbb{1}_{\Pi((x,a(1-x)+x])>0}\big] = \Gamma_a(x;\Pi)\,, \end{aligned} \tag{5.22}$$

where $0 \le \theta_x^- \le a^{-1}\xi_x^-$. The right-hand side of the last identity is the function $\Gamma_a(x;\Pi)$ of Theorem 4.3. Note that $\Gamma_a(x;\Pi)$ is $\mathcal{F}_{S_a}$-measurable for the stopping set $S_a = [-ax + x, a(1-x)+x] \cap [\xi_x^-, \xi_x^+]$. For $\mathbf{P}_\lambda$-almost all Π the second summand in $\Gamma_a(x;\Pi)$ vanishes for all sufficiently large a. Therefore conditions **(D)** and **(L)** of Theorem 4.3 hold with $g(a) = a^{-3}$, $S(x;\Pi) = [\xi_x^-(\Pi), \xi_x^+(\Pi)]$ and

$$\Gamma(x;\Pi) = -\frac{1}{4} f''(x) \Big[\xi_x^+(\Pi)^2\, \xi_x^-(\Pi) + \xi_x^+(\Pi)\, \xi_x^-(\Pi)^2 \Big] .$$

To check Condition **(UI)** for λ-almost all $x \in [0,1]$ we use Theorem 4.6. Since $S(x;\Pi)$ is a segment, it is $(x,1)$-thick and Condition *(ii)* of Theorem 4.6 holds by definition of S_a. Next, $\Pi(S(x;\Pi)) \le 2$ for all Π, so *(iii$_1$)* is also satisfied. To show *(i)*, note that there exists a constant c such that $|f(x)| \le c$ and $|f''(x)| \le c$ for all $x \in [0,1]$. Therefore by (5.22) and (5.19)

$$|\Gamma_a(x;\Pi)| \le \frac{1}{4} c \Big[(\xi_x^-)^2 \xi_x^+ + \xi_x^- (\xi_x^+)^2 \Big] \mathbb{1}_{\Pi([-ax+x,x))>0}\, \mathbb{1}_{\Pi((x,a(1-x)+x])>0} + 2ca^4 \Big[\mathbb{1}_{\Pi([-ax+x,x))=0} + \mathbb{1}_{\Pi((x,a(1-x)+x])=0} \Big] .$$

Take $W = [x-1, x+1]$ so that $\gamma_b^x W = [x-b, x+b]$. For any positive power n one has

$$|\Gamma_a(x;\Pi)|^n\, \mathbb{1}_{S \not\subseteq \gamma_b^x W} \le |\Gamma_a(x;\Pi)|^n [\mathbb{1}_{\xi_x^+ > b} + \mathbb{1}_{\xi_x^- > b}] .$$

Furthermore, for any positive k and any $\theta > 0$

$$\mathbf{E}_{\theta\lambda} (\xi_x^+)^k\, \mathbb{1}_{\xi_x^+ > b} = \theta \int_b^\infty y^k e^{-\theta y} dy \to 0 \quad \text{as } b \to \infty ;$$

$$a^k\, \mathbf{E}_{\theta\lambda} [\mathbb{1}_{\Pi([-ax+x,x))=0} + \mathbb{1}_{\Pi((x,a(1-x)+x])=0}] = a^k (e^{-\theta a x} + e^{-\theta a (1-x)}) \to 0 \quad \text{as } a \to \infty .$$

The random variables ξ_x^-, ξ_x^+ are independent and under $\mathbf{P}_{\theta\lambda}$ have the exponential distribution with parameter θ, so that $\mathbf{E}_{\theta\lambda} (\xi_x^-)^k < \infty$ for all $k > 0$. Thus

$$\lim_{\substack{a\to\infty \\ b\to\infty}} \mathbf{E}_{\theta\lambda}\, |\Gamma_a|^n\, \mathbb{1}_{\xi_x^+ > b} = 0$$

for all n. Similarly,

$$\lim_{\substack{a\to\infty \\ b\to\infty}} \mathbf{E}_{\theta\lambda}\, |\Gamma_a|^n\, \mathbb{1}_{\xi_x^- > b} = 0$$

and condition *(i)* of Theorem 4.6 is satisfied implying Condition **(UI)** for λ-almost all $x \in [0,1]$.

Finally,

$$\mathbf{E}_{p(x)\lambda}\, \xi_x^{\pm} = p(x)^{-1} \quad \text{and} \quad \mathbf{E}_{p(x)\lambda}(\xi_x^{\pm})^2 = 2p(x)^{-2}$$

giving

$$\mathbf{E}_{p(x)\lambda}\, \Gamma(x) = -f''(x)p(x)^{-3}\,.$$

From (4.25) we immediately obtain that $p(x)$ is proportional to $(f''(x))^{1/3}$. Since, the integral of $p(x)$ is 1,

$$p(x) = \frac{(f''(x))^{1/3}}{\int_0^1 (f''(x))^{1/3} dx}\,, \quad x \in [0,1]\,,$$

is the density of the measure that provides a high intensity solution of the minimisation problem.

Clearly, the functional $F(\Pi)$ given by (5.18) characterises the quality of numerical integration of f by trapezoidal rule using Poisson points. If f is non-increasing and the integral is computed by step-function approximations using the left end-points only, then

$$F(\Pi) = \sum_{i=0}^{n} f(x_i)(x_{i+1} - x_i) - \int_0^1 f(x)dx\,.$$

In this case $\Delta(x;\Pi) = r_x^+(f(x) - f(x - r_x^-))$. It is easy to show that the high intensity solution here has the density $p(x)$ proportional to $(|f'(x)|)^{1/2}$.

Using the same technique it is possible to find the best Poisson points approximations of convex functions in the uniform metric. A similar method is also applicable for the study of approximations of convex sets by inscribed polyhedra if the error is measured as the Hausdorff distance or as the difference between the areas (or perimeters) of the set and the polygon. For example, if a planar compact convex set K with a twice differentiable boundary ∂K is approximated by a polygon obtained as the convex hull of a high intensity Poisson point process on its boundary, then the best area approximation is achieved if the density is proportional to $(R(x))^{2/3}$, where $R(x)$ is the radius of curvature at point $x \in \partial K$. This is the same as the best density if the number of points is fixed and tends to infinity, see [24]. Note that the quality of approximation in the high intensity Poisson case does not differ from the quality of approximation using n random points with $n \to \infty$.

An alternative method of approximation of a real-valued twice continuously differentiable convex function f defined on a convex set $X \subset \mathbb{R}^d$ is based on approximation from below. This tangent approximation is defined by

$$y(x, \Pi) = \max_{x_i \in \Pi} \Big\{ f(x_i) + \nabla f(x_i) \cdot (x - x_i) \Big\} .$$

The graph of $y(x, \Pi)$, $x \in X$, is composed of tangent planes drawn at the points of Π to the graph of the function f. As above, we take the $\mathscr{L}_1$-distance between $f(x)$ and $y(x, \Pi)$ as the object function F. In addition we assume that the function f does not contain parabolic points, so that the Gaussian curvature of its graph does not vanish. It is easy to see that

$$-\Delta(x_0; \Pi) = \int_X dx \int_{-\infty}^{f(x_0) + \nabla f(x_0) \cdot (x - x_0)} \prod_{x_i \in \Pi} \mathbb{1}_{y > f(x_i) + \nabla f(x_i) \cdot (x - x_i)} \, dy .$$

Since for a Poisson process with the intensity measure μ,

$$\mathbf{E}_\mu \prod_{x_i \in \Pi} h(x_i) = \exp\{ - \int (h(z) - 1) \mu(dz) \}$$

for a suitable function $h : X \to \mathbb{R}_+$. Therefore,

$$\begin{aligned} -\overline{\Delta}_\mu (x_0) &= \int_X dx \int_{-\infty}^{f(x_0) + \nabla f(x_0) \cdot (x - x_0)} \exp \Big\{ - \int_X \mathbb{1}_{y \le f(z) + \nabla f(z) \cdot (x - z)} \, \mu(dz) \Big\} dy \\ &= \int_X dx \int_0^\infty \exp \Big\{ - \int_X \mathbb{1}_{-t \le \tilde{f}(z) + \nabla \tilde{f}(z) \cdot (x - z)} \, \mu(dz) \Big\} dt , \end{aligned}$$

where $\tilde{f}(z) = f(z) - f(x_0) - \nabla f(x_0) \cdot (z - x_0)$.

The set $A(x, t) = \{ z : \ \tilde{f}(z) + \nabla \tilde{f}(z) \cdot (x - z) \ge -t \}$ is the projection onto X of the part of the cone lying below the graph of $\tilde{f}$ and formed by the tangent rays to the graph of $y = \tilde{f}(x)$ drawn from the point $(x, -t)$ (see Fig. 2). Introduce the polar coordinates $x = x_0 + ru$, where $r \ge 0$ and u belongs to the unit sphere, $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$. Let $x + \rho(s)s = x_0 + ru + \rho(s)s$ describe the boundary of the set $A(x, t)$ for $\rho(s) = \rho_{x,t}(s) \ge 0$ and $s \in \mathbb{S}^{d-1}$. At the boundary points $x + \rho(s)s$ the segment with the end-points $(x, -t)$ and $P = (x + \rho(s)s, \tilde{f}(x + \rho(s)s))$ is orthogonal to the normal vector to the surface $\tilde{f}$ at P, so that $\rho = \rho(s)$ can be found from the equation

$$\nabla \tilde{f}(x_0 + ru + \rho s) \cdot \rho s - \tilde{f}(x_0 + ru + \rho s) - t = 0 . \tag{5.23}$$

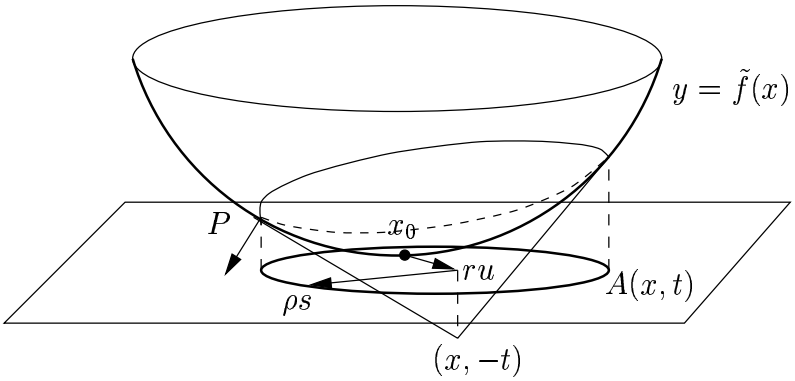

Figure 2: Tangent approximation.

Although we do not give details here, one can show that since $\Delta(x_0;\Pi)$ depends only on the nearest to x_0 points of Π, there exist compact stopping sets $S_a(x_0;\Pi)$ such that $\Delta(x_0;\gamma_a^{x_0}\Pi)$ is $\mathcal{F}_{S_a}$-measurable. Moreover, $\mathbb{1}_{S_a(x_0;\Pi)} \to \mathbb{1}_{S(x_0;\Pi)}$, where $S(x_0;\Pi)$ is the Voronoi flower centred in x_0 defined in Section 5.3 (see Fig. 1). If we assume that **(MG)** holds, then, by (4.22) and (4.25), $p(x_0)$ can be found from $g(a)\overline{\Delta}_{ap(x_0)}(x_0) = \text{const}$ for a suitable normalising function $g(a)$.

Since

$$-\overline{\Delta}_{ap(x_0)}(x_0) = \int_X dx \int_0^\infty \exp\{-ap(x_0)\lambda(A(x,t))\}dt$$
$$= \int_{\mathbb{S}^{d-1}} du \int_0^\infty r^{d-1}dr \int_0^\infty \exp\left\{-\frac{1}{d}ap(x_0)\int_{\mathbb{S}^{d-1}} \rho^d_{x,t}(s)ds\right\}dt\,,$$

then the asymptotic behaviour of $\overline{\Delta}_{ap(x_0)}(x_0)$ is determined by the behaviour of the exponent when $\lambda(A(x,t)) \downarrow 0$, i.e. when r and t are close to 0. In this settings, (5.23) writes

$$D^2 f(x_0)(ru+\rho s)\cdot\rho s - \frac{1}{2}D^2 f(x_0)(ru+\rho s)\cdot(ru+\rho s) - t + o(t^2+r^2) = 0\,,$$

or

$$\rho^2 k_s - r^2 k_u - 2t + o(t^2+r^2) = 0\,,$$

where $k_s = D^2 f(x_0)s\cdot s$ is the curvature of the normal section of the graph of $\tilde{f}$ at x_0 in the direction of $s \in \mathbb{S}^{d-1}$ (similarly, $k_u = D^2 f(x_0)u\cdot u$). Therefore, up to the smaller terms

$$\lambda(A(x,t)) = \frac{1}{d}\int\limits_{\mathbb{S}^{d-1}} \left(\frac{2t + k_u r^2}{k_s}\right)^{d/2} ds = \frac{1}{d}(2t + k_u r^2)^{d/2} \int\limits_{\mathbb{S}^{d-1}} \frac{ds}{k_s^{d/2}}\,.$$

(remember that the graph does not contain parabolic points, so that $k_s > 0$ for all s from the unit sphere). To evaluate the last integral, note that the function $R(s) = k_s^{d/2}$, $s \in \mathbb{S}^{d-1}$, represents in polar coordinates the boundary of an ellipsoid E oriented along the directions of the principal curvatures of the surface $y = f(x)$ at the point $(x_0, f(x_0))$. The semi-axes of E are given by $\sqrt{k_i}$, where k_i, $i = 1, \ldots, d$ are the corresponding principal curvatures (to see this, write the positive quadratic form $D^2 f(x_0)s\cdot s$ in the coordinate system corresponding to the principal directions where it takes the canonical form $\sum_{i=1}^d k_i s_i^2$). Therefore,

$$\int\limits_{\mathbb{S}^{d-1}} \frac{ds}{k_s^{d/2}} = d \int\limits_{\mathbb{S}^{d-1}} ds \int\limits_0^{1/R(s)} r^{d-1} dr\,,$$

that is d-times the volume of the body obtained by the inversion of E with respect to the unit sphere. Since this body is again the ellipsoid with the semi-axes $k_i^{-1/2}$,

$$\int\limits_{\mathbb{S}^{d-1}} \frac{ds}{k_s^{d/2}} = \frac{db_d}{\sqrt{k_1 \cdots k_d}} = db_d K(x_0)^{-1/2}\,,$$

where $K(x_0) = \det \|D^2 f(x_0)\|$ is the Gaussian curvature of the surface $y = f(x)$ at the point $(x_0, f(x_0))$.

Substitution

$$z = b_d\, p(x_0) K(x_0)^{-1/2} (2t + k_u r^2)^{d/2}$$

reduces the integral $\int_0^\infty \exp\{-ap(x_0)\lambda(A(x,t))\}dt$ to the incomplete Gamma function

$$\frac{1}{dB^{2/d}} \int\limits_{Bk_u^{d/2} r^d}^\infty z^{2/d-1} e^{-az} dz = I_1\,,$$

where $B = b_d\, p(x_0) K(x_0)^{-1/2}$. Using

$$\int\limits_b^\infty y^\beta e^{-ay} dy = \frac{b^\beta e^{-ab}}{a}\big(1 + O(1/a)\big)\,, \quad b > 0\,,$$

we get

$$I_1 \sim \frac{k_u^{1-d/2} r^{2-d}}{adB} e^{-aBk_u^{d/2} r^d} \quad \text{as } a \to \infty .$$

Therefore,

$$\int_0^\infty r^{d-1} I_1 dr \sim \frac{k_u^{1-d/2}}{adB} \int_0^\infty r e^{-aBk_u^{d/2} r^d} dr = \frac{\Gamma(2/d)}{a^{1+2/d} dB^{1+2/d} k_u^{d/2}} \quad \text{as } a \to \infty .$$

Finally,

$$-\overline{\Delta}_{ap(x_0)}(x_0) \sim \frac{\Gamma(2/d)\, K(x_0)^{1/d}}{a^{1+2/d}\, b_d^{2/d} p(x_0)^{1+2/d}} , \tag{5.24}$$

whence

$$p(x) \propto K(x)^{\frac{1}{2+d}} .$$

In particular, if $d = 1$, then $p(x) \propto (f''(x))^{1/3}$ that coincides with the optimal density for the polygonal approximation from above for one-dimensional convex functions.

5.5 How to 'catch' a random set by random points?

Let Y be a random closed subset of X (see [16, 26]) that is independent of the Poisson process Π. We say that Y is *trapped* by Π if $\Pi(Y) > 0$ (or $\Pi \cap Y \neq \emptyset$, since we identify Π and its support). The distribution of Y is fixed, so the subscript μ denotes the probability and expectation for Π having the intensity measure μ. By independence of Y and Π, we get

$$\mathbf{P}_\mu\{\Pi \cap Y \neq \emptyset\} = \mathbf{E}\, \mathbf{P}_\mu\{\Pi \cap Y \neq \emptyset \;\big|\; Y\} = \mathbf{E}\left[1 - e^{-\mu(Y)}\right] . \tag{5.25}$$

We seek μ that maximises the trapping probability $\mathbf{P}\{\Pi \cap Y \neq \emptyset\}$. Equivalently, we can minimise $\mathbf{E}_\mu F$, where

$$F(\Pi) = \mathbb{1}_{\Pi \cap Y = \emptyset} .$$

Then

$$\mathbf{E}_\mu F(\Pi) = \mathbf{E}\, e^{-\mu(Y)} = \mathbf{E}_\mu Q_Y(\Pi) ,$$

where $Q_Y(K) = \mathbf{P}\{Y \cap K = \emptyset\}$ is the *avoiding functional* of Y. Below we consider the problem with the fixed total mass $\mu(X) = a$.

The expected first difference can be written as

$$\overline{\Delta}_\mu(x) = -\mathbf{E}\left[e^{-\mu(Y)}\, \mathbb{1}_{x \in Y}\right] \tag{5.26}$$

or, equivalently, as

$$\overline{\Delta}_\mu(x) = \mathbf{E}_\mu[Q_Y(\Pi \cup \{x\}) - Q_Y(\Pi)]\,. \tag{5.27}$$

Note that $\overline{\Delta}_\mu(x) \le 0$ for all $x \in X$. If $\overline{\mu}$ is a solution of the fixed total mass problem, then (4.2) yields

$$\begin{cases} \overline{\Delta}_{\overline{\mu}}(x) = -u & \overline{\mu} - \text{a.e.}\,, \\ \overline{\Delta}_{\overline{\mu}}(x) \ge -u & \text{for all } x \in X\,, \end{cases} \tag{5.28}$$

for a constant $u \ge 0$. Note that $u = 0$ yields $\overline{\Delta}_{\overline{\mu}}(x) = 0$ everywhere, which is possible only if $\mathbf{P}\{x \in Y\} = 0$ for all x. In the sequel we exclude this trivial case.

Assume that there are at least two points $x \in X$ such that $\mathbf{P}\{x \in Y\} > 0$. Otherwise Y either cannot be trapped at all using a Poisson point process on X, or the trapping strategy is trivial. Furthermore, assume that Y has no fixed points, i.e. $\mathbf{P}\{x \in Y\} < 1$ for all $x \in X$.

Example 5.6. Let $X = \{x_1, x_2\}$, and let Y take values $\{x_1\}$, $\{x_2\}$ or $\{x_1, x_2\}$ with probabilities p_1, p_2, p_{12} such that $p_1, p_2 > 0$. Note that $\mathbf{P}\{Y = \emptyset\} = 1 - p_1 - p_2 - p_{12}$. Each measure μ on X is represented by two non-negative numbers (m_1, m_2). Then

$$\overline{\Delta}_\mu(x_i) = -p_i e^{-m_i} - p_{12} e^{-m_1 - m_2}\,, \quad i = 1, 2\,.$$

If $\overline{\mu} = (m_1, m_2)$ minimises the trapping probability, then (4.2) yields

$$p_1 e^{-m_1} = p_2 e^{-m_2}\,,$$

so that $m_2 = m_1 + \log(p_2/p_1)$. If $a = 1$, then $m_1 + m_2 = 1$, whence

$$m_1 = (1 - \log(p_2/p_1))/2\,, \quad m_2 = (1 + \log(p_2/p_1))/2\,.$$

Example 5.7. Let X be a countable space $\{x_1, x_2, \dots\}$, and let $Y = \{\xi\}$ be a random singleton. Renumbering, if necessary, the points of X, we may assume that $p_i = \mathbf{P}\{\xi = x_i\}$, $i \ge 1$, is a non-increasing sequence. By (5.28), there exists $u > 0$ such that

$$\overline{\Delta}_{\overline{\mu}}(x_i) = -\mathbf{E}\left[e^{-\overline{\mu}(\{\xi\})}\, \mathbb{1}_{\xi = x_i}\right] = -e^{-m_i} p_i \ge -u$$

for all i and $\overline{\Delta}_{\overline{\mu}}(x_i) = -u$ if $m_i = \overline{\mu}(\{x_i\}) > 0$. Thus,

$$\begin{cases} m_i = \log(p_i/u)\,, & x_i \in S_{\overline{\mu}}\,, \\ p_i \le u\,, & x_i \notin S_{\overline{\mu}}\,, \end{cases} \tag{5.29}$$

where $S_{\overline{\mu}} \subset X$ is the support of $\overline{\mu}$, so that $\overline{\mu}(S_{\overline{\mu}}) = a$. Note that $S_{\overline{\mu}} = X$ is impossible, since this leads to negative values of m_i as $p_i \to 0$ for $i \to \infty$. Thus $S_{\overline{\mu}} = \{x_1, \dots, x_k\}$ for some number $k \ge 1$. Then

$$\mathbf{P}_\mu\{\Pi(Y) > 0\} = \sum_{i=1}^{k}(1 - e^{-m_i})p_i = \sum_{i=1}^{k}(p_i - u) = \sum_{i=1}^{k} p_i - ku\,.$$

If the total mass of μ is a, then

$$\sum_{i=1}^{k} \log(p_i/u) = a\,,$$

so that

$$\log u = \frac{1}{k}\Big[\sum_{i=1}^{k} \log p_i - a\Big]\,.$$

Therefore,

$$\mathbf{P}_\mu\{\Pi(Y) > 0\} = \sum_{i=1}^{k} p_i - k \exp\left\{\frac{1}{k}\Big[\sum_{i=1}^{k} \log p_i - a\Big]\right\} = g(k)\,.$$

The optimal value for k can be found by maximising the function $g(k)$. Clearly, $k \le \min\{i : p_i = 0\}$. Moreover, the second condition of (5.29) implies $p_{k+1} \le u \le p_k$, so that

$$k \log p_{k+1} \le \sum_{i=1}^{k} \log p_i - a \le k \log p_k\,. \tag{5.30}$$

Consider the case when $\xi = x_\eta$ with a geometrically distributed η, so that $p_i = pq^{i-1}$ with $p + q = 1$. Then (5.30) yields

$$k(k-1) \le -2a/\log q \le k(k+1)\,.$$

Therefore, k is equal to the integer part of $(1 + \sqrt{1 - 8a/\log q})/2$. The corresponding hitting probability equals

$$g(k) = 1 - q^k - kpq^{(k-1)/2}e^{-a/k} .$$

For example, if $p = q = 0.5$ and $a = 1$, then $k = 2$ with $\overline{\mu} = (0.847, 0.153)$, and $g(2) = 0.3211$. Since the total mass is equal to 1, on average we have only one point of the process. It is interesting to note that the fixed point $\{x_1\}$ traps Y with probability $p = 0.5$, which is larger than $g(2)$. Indeed, the Poisson point process with the unit intensity measure concentrated in x_1 may have zero points, while multiple points at x_1 do not increase the trapping probability.

In the high intensity case it is possible to come up with further results.

Example 5.8. Let $Y = [\min(\xi, \eta), \max(\xi, \eta)]$ for two random variables ξ and η with the joint probability density function $p_{\xi\eta}(\cdot, \cdot)$ and a compact support $X \in \mathbb{R}$. Using representation (5.27) we get

$$\Delta(x; \Pi) = -\mathbf{P}\{x - \xi_x^-(\Pi) < \min(\xi, \eta) < x < \max(\xi, \eta) < x + \xi_x^+(\Pi)\} ,$$

where $\xi_x^-(\Pi)$, $\xi_x^+(\Pi)$ are defined in (5.21). Thus all $\Gamma_a(x; \Pi) = a^2\Delta(x; \gamma_a^x\Pi)$ are $\mathcal{F}_S$-measurable for the stopping set $S = [x - \xi_x^-, x + \xi_x^+]$. Similarly to as we have done it in Section 5.4, one can show that the conditions of Corollary 4.4 are satisfied for $\Gamma(x; \Pi) = 2p_{\xi\eta}(x, x)\xi_x^-(\Pi)\xi_x^+(\Pi)$ and all x from the set $\{x : \ p_{\xi\eta}(x, x) > 0\}$. If $p_{\xi\eta}(x, x) > 0$, then

$$-\mathbf{E}_{p(x)\lambda}\,\Gamma = -2p_{\xi\eta}(x, x)(p(x))^{-2} ,$$

where $p(x)$ is the density of the high intensity solution. Thus, $p(x) \propto \sqrt{p_{\xi\eta}(x, x)}$. For instance, if ξ and η are independent and identically distributed, then $p(x) = p_\xi(x)$.

Example 5.9. Let $X \subseteq \mathbb{R}^d$ be an open set, and let $Y = B_\rho(\xi)$ be a random ball, where the centre ξ and the radius ρ are independent and have probability density functions p_ξ and p_ρ respectively. We have

$$\Delta(x; \Pi) = \mathbf{P}\{\Pi(B_\rho(\xi)) = 0, \ x \in B_\rho(\xi)\} .$$

Under this event $B_\rho(\xi) \subseteq S(x; \Pi)$, where the stopping set $S = S(x; \Pi)$ is the Voronoi flower constructed with respect to $\Pi + \delta_x$ defined in Section 5.3 (see Fig. 1). Thus $\Delta(x; \gamma_a^x\Pi)$ is $\mathcal{F}_S$-measurable and equals

$$\Delta(x; \gamma_a^x\Pi) = \mathbf{P}\{x \in B_\rho(\xi) \subseteq \gamma_{a^{-1}}^x S(x; \Pi)\} .$$

If $p_\rho^{(k)}(0)r^k$ is the first non-zero term in the Taylor expansion of the density $p_\rho(r)$ at $r=0$ and p_ρ is at least k times continuously differentiable in a neighbourhood of $r=0$, then the above quantity behaves like $O\bigl(a^{-k}p_\xi(x)\,p_\rho^{(k)}(0)\bigr)$. As in Section 5.3, it is possible to show that the conditions of Corollary 4.4 are satisfied on the set $\{x:\ p_\xi(x)>0\}$. Therefore, for any $\varepsilon>0$,

$$\begin{aligned} a^k\overline{\Delta}_{a\mu_a}(x) \propto a^k\overline{\Delta}_{ap(x)\lambda} &= -a^k\int_0^\infty p_\rho(r)dr\int_{B_r(x)} e^{-ap(x)\lambda(B_r(y))}p_\xi(y)\,dy \\ &\propto -p_\xi(x)\int_0^\varepsilon e^{-ap(x)b_dr^d}b_dr^d\left[p_\rho^{(k)}(0)r^k+O(a^{-1}r)\right]dr \\ &\sim -p_\xi(x)b_dd^{-1}\,\frac{p_\rho^{(k)}(0)\,(d+k+1)\,\Gamma(1+(k+1)/d)}{(ap(x)b_d)^{1+(k+1)/d}}, \end{aligned}$$

that allows us to find $p(x)\propto(p_\xi(x))^{d/(d+k+1)}$.

6 Gradient method for minimisation of functional

Even if the minimisation problem does not admit an analytical solution, it may be possible to find an approximate solution using gradient based numerical methods. Because of the form of the directional derivative, the function $\overline{\Delta}_\mu(x)$ plays the same role as the gradient in the classical steepest descent method. We consider the fixed total mass problem in the space of all positive measures.

Let $\varepsilon>0$ be a constant that controls the maximal size of a step of the algorithm. If μ is a starting measure, then in order to minimise $\mathbf{E}_\mu F(\Pi)$ it is natural to move from μ to $\mu+\eta$ where $\eta(X)=0$, $\|\eta\|=\varepsilon$ and η minimises the directional derivative $\langle\overline{\Delta}_\mu,\eta\rangle$. Clearly, such η can be obtained by putting the atom of weight $-\varepsilon$ in the point of maximum of $\overline{\Delta}_\mu(x)$ and the atom ε in the point of minimum of $\overline{\Delta}_\mu(x)$. However, often this is not possible, because $\mu+\eta$ may not be a non-negative measure. If $\mu+\eta\in\mathbb{M}$, then η is called an *admissible direction*. It is easy to see that the 'steepest' admissible direction η can be constructed as follows.

Let N_μ be the set of all global minima points of $\overline{\Delta}_\mu(x)$. For each $t\in\mathbb{R}$ denote

$$E_t=\{x\in X:\ \overline{\Delta}_\mu(x)\ge t\}.$$

the level sets of the function $\overline{\Delta}_\mu(x)$ and introduce

$$h_\mu(\varepsilon) = \inf\{t \in \mathbb{R} : \ \mu(E_t) \le \varepsilon\}\,,$$
$$E(\varepsilon) = E_{h_\mu(\varepsilon)}\,.$$

Now construct the measure

$$\eta = \mu\big(E(\varepsilon)\big)\,\nu_{N_\mu} - \mu|_{E(\varepsilon)}\,,$$

where ν_{N_μ} is any measure concentrated on N_μ with the total mass 1, and $\mu|_A$ is the restriction of μ onto a set A. It is easy to see that any such η serves as the steepest admissible direction for the gradient method.

The necessary condition (4.2) can be used as a stopping rule for the described algorithm. The difference between the supremum and infimum of $\overline{\Delta}_\mu(x)$ over the support of μ can be used to adjust the value of ε of the next step. The series $\sum_k \varepsilon_k$ of values ε_k on k-th step should, of course, diverge.

In many practical situations the optimal measure is atomic. Then it is possible to stop the descent when the mass starts accumulating around atom locations of the optimal measure. An additional analysis is however, necessary to justify that the found local minimum is the global one.

Similar numerical methods are used to find an optimal experiment design, see [2, Section 11.2]. However, the latter methods are designed to minimise a special type of functionals on the space of measures and have no direct relation to the Poisson process setup.

In order to check the feasibility of the steepest descent method, we have considered the following minimisation problem.

Example 6.1. Consider again the hard-core type functional $F(\Pi)$ defined in Example 5.2 with the phase space $X = [0,1]$. We are looking for a measure μ on $[0,1]$ with $\mu([0,1]) = a$ that minimises $\mathbf{E}_\mu F$. In this problem the function $\overline{\Delta}_\mu(x)$ is computed directly and equals

$$\overline{\Delta}_\mu(x) = \frac{1}{2}\mu([0,1] \cap [x-r, x+r])\,.$$

Intuitively it is clear, that the optimal μ should repulse the points in order to keep the distance between them of order r. Figure 3 shows the resulting functions $\overline{\Delta}_\mu(x)$ (the upper graph) and the density of μ (the lower graph) as the algorithm runs 0, 20 and 40 steps. The initial distribution was taken uniform on [0,1], $r = 0.1$ and $a = 10$. One can see that the function $\overline{\Delta}_\mu(x)$ after 40 steps is almost a constant on the support of the measure μ and is greater beyond it, so the necessary condition (4.2) is fulfilled within the error 0.1 for that measure.

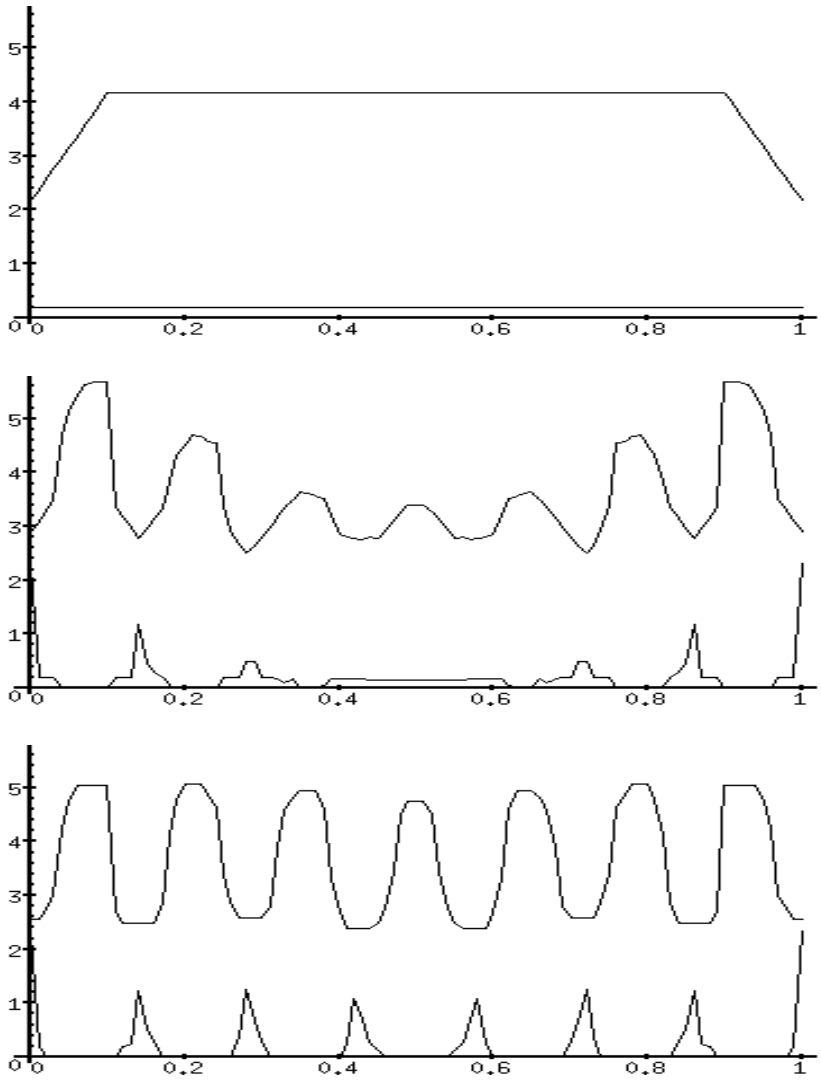

Figure 3: Implementation of the steepest descent algorithm.

We observed that depending on the initial measure the algorithm converges to different local minima. Indeed, one can easily check that, for instance, the uniform distribution among k equally spaced atoms for $k = 5, 6, \ldots, 9$ satisfies the necessary condition (4.2).

More detailed analysis of the gradient type algorithms will be reported elsewhere.

Acknowledgements

This work would be impossible without moral support of our families which enabled the authors to find time and efforts for this research. Many people helped us with their advice. In particular, we thank François Baccelli for helpful discussions, Frédéric Bonnans for consultation on optimisation in Banach spaces, Ben Torsney who has drawn our attention to relationships with optimal design theory and other people for their interest in this work. The authors acknowledge sincere hospitality of CWI (Amsterdam) where a part of these research has been carried out, INRIA and University of Glasgow who supported our exchange visits.

References

[1] Albeverio, S., Kondratiev, Y., and Röckner, M. (1996) Differential geometry on Poisson spaces. *C. R. Acad. Sci., Paris, Ser. I* **323**, 1129–1134.

[2] Atkinson, A. and Donev, A. (1992) *Optimum Experimental Designs.* Clarendon Press, Oxford.

[3] Baccelli, F. and Brémaud, P. (1993) Virtual customers in sensitivity and light traffic analysis via Campbell's formula for point processes. *Adv. in Appl. Probab.* **25**, 221–234.

[4] Baccelli, F., Klein, M., Lebourges, M., and Zuyev, S. (1996) Géométrie aléatoire et architecture de réseaux de communications. *Annales des Télécommunications* **51**, 158–179.

[5] Baccelli, F., Klein, M., and Zuyev, S. (1995) Perturbation analysis of functionals of random measures. *Adv. in Appl. Probab.* **27**, 306–325.

[6] Błaszczyszyn, B. (1995) Factorial moment expansion for stochastic systems. *Stochastic Process. Appl.* **45**, 321–335.

[7] Cominetti, R. (1990) Metric regularity, tangent sets, and second-order optimality conditions. *Appl. Math. Optim.* **21**, 265–287.

[8] Cook, D. and Fedorov, V. (1995) Constrained optimization of experimental design. *Statistics* **26**, 129–178.

[9] Daley, D. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes.* Springer, New York.

[10] Foss, S. and Zuyev, S. (1996) On a Voronoi aggregative process related to a bivariate Poisson process. *Adv. in Appl. Probab.* **28**, 965–981.

[11] Hille, E. and Phillips, R. (1957) *Functional Analysis and Semigroups*, volume XXXI of *AMS Colloquium Publications.* American Mathematical Society, Providence.

[12] Kallenberg, O. (1983) *Random Measures.* Akademie-Verlag/Academic Press, Berlin/New York, 3rd edition.

[13] Kerstan, J., Matthes, K., and Mecke, J. (1974) *Unbegrenzt teilbare Punktprozesse.* Akademie-Verlag, Berlin.

[14] Kiefer, J. (1974) General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–879.

[15] Kiefer, J. and Wolfowitz, J. (1960) The equivalence of two extremal problems. *Canad. J. Math.* **14**, 363–366.

[16] Matheron, G. (1975) *Random Sets and Integral Geometry.* Wiley, New York.

[17] Maurer, H. and Zowe, J. (1979) First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Programming* **16**, 98–110.

[18] Miles, R. and Maillardet, R. (1982) The basic structures of Voronoi and generalized Voronoi polygons. *J. Appl. Probab.* **19A**, 97–111.

[19] Møller, J. and Zuyev, S. (1996) Gamma-type results and other related properties of Poisson processes. *Adv. in Appl. Probab.* **28**, 662–673.

INRIA

[20] Nualart, D. and Vives, J. (1990) Anticipative calculus for the Poisson process based on the Fock space. *Lect. Notes Math.* **1426**, 154–165.

[21] Privault, N. (1994) Chaotic and variational calculus in discrete and continuous time for the Poisson process. *Stochastics Stoch. Rep.* **51**, 83–109.

[22] Robinson, S. (1976) First order conditions for general nonlinear optimization. *SIAM J. Appl. Math.* **30**, 597–607.

[23] Rubinstein, R. and Shapiro, A. (1993) *Discrete event systems. Sensitivity analysis and stochastic optimization by score function method.* Wiley, Chichester.

[24] Schneider, R. (1988) Random approximations of convex sets. *J. Microscopy* **151**, 211–227.

[25] Slivnyak, I. (1962) Some properties of stationary flows of homogeneous random events. *Theory Probab. Appl.* **7**, 336–341.

[26] Stoyan, D., Kendall, W., and Mecke, J. (1995) *Stochastic Geometry and Its Applications.* Wiley, Chichester, 2nd edition.

[27] Strauss, D. (1975) A model for clustering. *Biometrika* **63**, 467–475.

[28] Wynn, H.-P. (1972) Results in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **34**, 133–147.

[29] Zowe, J. and Kurcyusz, S. (1979) Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.* **5**, 49–62.

[30] Zuyev, S. (1997) Stopping sets: Gamma-type results and hitting properties. Technical Report 3146, INRIA, Sopia Antipolis.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399