Variational analysis of Poisson processes

Ilya Molchanov and Sergei Zuyev

Abstract The expected value of a functional $F(\eta)$ of a Poisson process η can be considered a function of its intensity measure μ. The paper surveys several results concerning differentiability properties of this functional on the space of signed measures with finite total variation. Then necessary conditions for μ being a local minima of the considered functional are elaborated taking into account possible constraints on μ, most importantly the case of μ with given total mass a. These necessary conditions can be phrased by requiring that the gradient of the functional (being the expected first difference $F(\eta+\delta_x)-F(\eta)$) is constant on the support of μ. In many important cases, the gradient depends only on the local structure of μ in a neighbourhood of x and so it is possible to work out the asymptotics of the minimising measure with the total mass a growing to infinity. Examples include the optimal approximation of convex functions, clustering problem, optimal search. In non-asymptotic cases, generally it is possible to find the optimal measure using steepest descent algorithms which are based on the obtained explicit form of the gradient.

1 Preliminaries

Importance of Poisson point processes for modelling various phenomena is impossible to overestimate. Perhaps, being the simplest mathematically treatable model, it, however, possesses a huge degree of flexibility since its the only parameter is a measure which describes the density of the process' points. It is amazing how many in-

Ilya Molchanov
University of Bern, Institute of Mathematical Statistics and Actuarial Science, Sidlerstrasse 5, CH-3012 Bern, Switzerland e-mail: ilya.molchanov@stat.unibe.ch

Sergei Zuyev
Chalmers University of Technology, Department of Mathematical Sciences, 412 96 Gothenburg, Sweden e-mail: sergei.zuyev@chalmers.se

triguing and deep properties such a seemingly simple model enjoys and new ones are constantly being discovered, as this monograph readily shows. Because the distribution of a Poisson point process depends on a measure, altering the measure changes the distribution which, in many cases, is a result of performing a certain transformation of the phase space or of the point configurations. Such approach is taken, for instance, in perturbation analysis of point process driven systems (see, e.g., [Ho and Cao(1991)] and the references therein) or in differential geometry of configuration spaces, see, e.g., [Albeverio et al(1998)Albeverio, Kondratiev, and Röckner] or [Röckner(1998)].

Rather than considering a change of the parameter measure induced by transformations of the phase space or of configurations, we take a more general approach by changing the parameter measure directly. A control over this change is made possible by a linear structure of the set of measures itself as we describe in details below.

The main subject of our study is a Poisson point process on a general locally compact separable metric space X with its Borel σ-algebra $\mathscr{X}$. Formally, a point process is a measurable mapping from some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ to $[\mathbf{N}, \mathscr{N}]$, where $\mathbf{N}$ is the set of locally finite counting measures on $\mathscr{X}$ called *configurations* and $\mathscr{N}$ is the minimal σ-algebra that makes all the mappings $\varphi \mapsto \varphi(B)$ measurable for any $B \in \mathscr{X}$. Any $\varphi \in \mathbf{N}$ can be represented as a sum of Dirac measures: $\varphi = \sum_i \delta_{x_i}$, where $\delta_x(B) = \mathbb{1}_B(x)$ for every $B \in \mathscr{X}$ and not necessarily all x_i's are distinct.

A point process η is Poisson with *intensity measure* μ on $\mathscr{X}$, if for any sequence of disjoint sets $B_1, \dots, B_n \in \mathscr{X}$, $n \geq 1$, the counts $\eta(B_1), \dots, \eta(B_n)$ are independent Poisson $\mathrm{Po}(\mu(B_1)), \dots, \mathrm{Po}(\mu(B_n))$ distributed random variables. The distribution of the Poisson point process with intensity measure μ will be denoted by $\mathbb{P}_\mu$ with the corresponding expectation $\mathbb{E}_\mu$. The term intensity measure is explained by the fact that, due to the definition, one has $\mathbb{E}\eta(B) = \mu(B)$ for any $B \in \mathscr{X}$. Notice that the Poisson process is finite, i.e. all its configurations with probability 1 contain only a finite number of points if and only if its intensity measure is finite: $\mu(X) < \infty$.

In what follows, we study the changes in the distributional characteristics of functionals of a configuration under perturbations of the intensity measure which we first assume finite. To this end, we consider the set $\widetilde{\mathbf{M}}_{\mathrm{f}}$ of all *signed* measures on $\mathscr{X}$ with a finite total variation and define operations of addition and multiplication by setting $(\mu + \nu)(B) = \mu(B) + \nu(B)$ and $(t\mu)(B) = t\mu(B)$ for any $B \in \mathscr{X}$. Recall that a signed measure ν can be represented as the difference $\nu = \nu^+ - \nu^-$ of two non-negative measures with disjoint supports (the *Lebesgue decomposition*) and that the total variation of ν is defined as $\|\nu\| = \nu^+(X) + \nu^-(X)$. Endowed with the total variation norm, $\widetilde{\mathbf{M}}_{\mathrm{f}}$ becomes a Banach space and the set $\mathbf{M}_{\mathrm{f}}$ of *finite non-negative measures* is a *pointed cone*, i.e. a set closed under addition and multiplication by non-negative numbers, see, e.g., [Dunford and Schwartz(1988)].

Given a function $F : \mathbf{N} \mapsto \mathbb{R}$ of a configuration, its expectation $\mathbb{E}_\mu F(\eta)$ with respect to the distribution $\mathbb{P}_\mu$ of a finite Poisson process η can be regarded as a function of the intensity measure μ and hence as a function on $\mathbf{M}_{\mathrm{f}}$. Therefore there is a reason to consider functions on $\widetilde{\mathbf{M}}_{\mathrm{f}}$ and their analytical properties in general.

2 Variational analysis on measures

Recall that a function f on a Banach space $\mathbb{B}$ is called *strongly* or *Fréchet differentiable* at $x \in \mathbb{B}$ if

$$f(x+y) = f(x) + L(x)[y] + o(\|y\|), \tag{1}$$

where $L(x)[\cdot] : \mathbb{B} \mapsto \mathbb{B}$ is a bounded linear functional called a *differential*. A function f called *weakly* or *Gateaux differentiable* at $x \in \mathbb{B}$ if for every $y \in \mathbb{B}$ there exists a limit

$$\partial_y f(x) = \lim_{t \downarrow 0} t^{-1}[f(x+ty) - f(x)] \tag{2}$$

which can be called the *directional derivative* along vector y. Strong differentiability implies that all weak derivatives also exist and that $\partial_y f(x) = L(x)[y]$. The converse is not true even for $\mathbb{B} = \mathbb{R}$. The same definitions apply to functions of a signed measure with finite total variation since $\widetilde{\mathbf{M}}_{\mathrm{f}}$ is a Banach space. A very wide class of differentiable functions of a measure possess a differential which has a form of an integral so that

$$f(\mu+\nu) = f(\mu) + \int_X g(x;\mu)\nu(dx) + o(\|\nu\|),\ \nu \in \widetilde{\mathbf{M}}_{\mathrm{f}}. \tag{3}$$

for some function $g(\cdot;\mu)$ called a *gradient function*. This name comes from the fact that when $X = \{1,\dots,d\}$ is a finite set, $\widetilde{\mathbf{M}}_{\mathrm{f}}$ is isomorphic to $\mathbb{R}^d$ and $g(\cdot;\mu) = (g_1(\mu),\dots,g_d(\mu))$ is a usual gradient, since

$$f(\mu+\nu) = f(\mu) + \langle g(\cdot;\mu),\nu\rangle + o(\|\nu\|),\ \mu,\nu \in \mathbb{R}^d.$$

Inline with this, we shall use from now on the notation $\langle f,\nu\rangle$ for the integral $\int f\,d\nu$. Not all differentiable functions of measures possess a gradient function (unless X is finite), but all practically important functions usually do. Notably, the expectation $\mathbb{E}_\mu F(\eta)$ as a function of $\mu \in \mathbf{M}_{\mathrm{f}}$ does possess a gradient function, as we will see in the next section. So it is not a severe restriction to assume that a differentiable function of a measure possesses a gradient function, as we often do below.

The differentiability provides a useful tool for optimisation of functions. Necessary conditions for a local optimum are based on the notion of a tangent cone.

Definition 1. A *tangent cone* to a set $\mathbb{A} \subset \widetilde{\mathbf{M}}_{\mathrm{f}}$ at point $\nu \in \mathbb{A}$ is the set

$$T_{\mathbb{A}}(\nu) = \liminf_{t \downarrow 0} \frac{\mathbb{A} - \nu}{t}.$$

Recall that the lim inf in the right-hand side is the set of possible limits for all the sequences $\{\eta_n\}$ such that eventually $\nu + t_n\eta_n \in \mathbb{A}$ or all n when $t_n \downarrow 0$. It represents the closure of set of all *admissible* directions in which one can move from $\nu \in \mathbb{A}$ still staying in $\mathbb{A}$.

A first order necessary condition for an optimum in a constrained optimisation now takes the following form.

Theorem 1. *Assume $\mathbb{A} \subseteq \widetilde{\mathbf{M}}_{\mathrm{f}}$ is closed and convex in the total variation norm and that f is continuous on $\mathbb{A}$ and strongly differentiable at $\nu^* \in \mathbb{A}$. If ν^* provides a local minimum in the constrained optimisation problem*

$$f(\nu) \to \inf \quad \textit{subject to } \nu \in \mathbb{A}, \tag{4}$$

then

$$L(\nu^*)[\theta] \geq 0 \quad \textit{for all } \theta \in T_{\mathbb{A}}(\nu^*). \tag{5}$$

The proof of this general fact can be found, e.g., in [Ben-Tal and Zowe(1982)] for the case of a constraint set with non-empty interior. For the purpose of optimisation with respect to the intensity measure, the main constraint set is the cone $\mathbf{M}_{\mathrm{f}}$ of non-negative measures. However, $\mathbf{M}_{\mathrm{f}}$ does *not* have interior unless X is finite. The non-emptiness assumption on the interior was first dropped in [Cominetti(1990)][Theorem 4.1.(i)]. The next result proved in [Molchanov and Zuyev(2000a)] characterises the tangent cone to $\mathbf{M}_{\mathrm{f}}$.

Theorem 2. *The tangent cone to the set $\mathbf{M}_{\mathrm{f}}$ at $\mu \in \mathbf{M}_{\mathrm{f}}$ is the set of signed measures for which the negative part of their Lebesgue decomposition is absolutely continuous with respect to μ:*

$$T_{\mathbf{M}_{\mathrm{f}}}(\mu) = \{\theta \in \widetilde{\mathbf{M}}_{\mathrm{f}} : \ \theta^- \ll \mu\}.$$

Assume now that f possesses a gradient function and μ^* provides a local minimum on the constrain set $\mathbb{A} = \mathbf{M}_{\mathrm{f}}$. Applying necessary condition (5) with $\theta = \delta_x$ we immediately get that

$$L(\mu^*)[\delta_x] = g(x;\mu^*) \geq 0 \quad \text{for all } x \in X.$$

Now setting $\theta = -\mathbb{1}_B \mu^*$ for an arbitrary measurable $B \subset X$ leads to

$$L(\mu^*)[\delta_x] = \langle g(\cdot;\mu^*)\mathbb{1}_B, \mu^* \rangle \leq 0\,.$$

Combining both inequalities, proves the following result.

Theorem 3. *Assume that $\mu^* \in \mathbf{M}_{\mathrm{f}}$ provides a local minimum to f on $\mathbf{M}_{\mathrm{f}}$ and that f possesses a gradient function $g(\cdot;\mu^*)$ at μ^*. Then $g(\cdot;\mu^*) = 0$ μ^*-almost everywhere on X and $g(x;\mu^*) \geq 0$ for all $x \in X$.*

By considering an appropriate Lagrange function, one can generalise this statement to the case of optimisation over $\mathbf{M}_{\mathrm{f}}$ with additional constraints. Before we formulate the result, we need a notion of regularity.

Definition 2. Let Y be a Banach space and $\mathbb{A} \subseteq \mathbf{M}_{\mathrm{f}}$, $C \subseteq Y$ be closed convex sets. Let $f : \ \widetilde{\mathbf{M}}_{\mathrm{f}} \mapsto \mathbb{R}$ and $H : \ \widetilde{\mathbf{M}}_{\mathrm{f}} \mapsto Y$ be strongly differentiable. A measure $\nu \in \widetilde{\mathbf{M}}_{\mathrm{f}}$ is called *regular* for the optimisation problem

$$f(\nu) \to \inf \quad \text{subject to } \nu \in \mathbb{A},\ H(\nu) \in C, \tag{6}$$

if $0 \in \operatorname{core}\big(H(\nu)+L_H(\nu)[\mathbb{A}-\nu]-C\big)$, where L_H is the differential of H and $\operatorname{core}(B)$ for $B \subseteq Y$ is the set $\{b \in B:\ \forall y \in Y\ \exists t_1$ such that $b+ty \in B\ \forall t \in [0,t_1]\}$. For $Y=\mathbb{R}^d$, $\operatorname{core}(B)$ is just the interior of the set $B \subseteq \mathbb{R}^d$.

Consider the most common case of a finite number of equality and inequality constraints. In this case $Y=\mathbb{R}^k$ and $C=\{0\}^m \times \mathbb{R}_-^{k-m}$, $m \le k$, so that we have the following optimisation problem:

$$f(\mu) \to \inf \quad \text{subject to} \tag{7}$$

$$\begin{cases} \mu \in \mathbf{M}_{\mathrm{f}} \\ H_i(\mu)=0,\ i=1,\dots,m \\ H_j(\mu) \le 0,\ j=m+1,\dots,k \end{cases} \tag{8}$$

for some function $H:\ \mathbf{M}_{\mathrm{f}} \mapsto \mathbb{R}^k$. The following result and its generalisations can be found in [Molchanov and Zuyev(2000a)].

Theorem 4. *Let μ^* be a regular (in the sense of Definition 2) local minimum for the problem* (7) *for a function f which is continuous on $\mathbf{M}_{\mathrm{f}}$ and strongly differentiable at μ^* with a gradient function $g(x;\mu^*)$. Let $H=(H_1,\dots,H_k)$ also be strongly differentiable at μ^* with a gradient function $h(x;\mu)=(h_1(x;\mu),\dots,h_k(x;\mu))$. Then there exist* Langrange multipliers $u=(u_1,\dots,u_k)$ *with $u_j \le 0$ for those $j \in \{m+1,\dots,k\}$ for which $H_j(\mu^*)=0$ and $u_j=0$ if $H_j(\mu^*)<0$, such that*

$$\begin{cases} g(x;\mu^*)=\sum_{i=1}^k u_i h_i(x;\mu^*) & \mu^*-a.e., \\ g(x;\mu^*) \ge \sum_{i=1}^k u_i h_i(x;\mu^*) & \text{for all } x \in X. \end{cases} \tag{9}$$

When the functions f and H possess gradient functions, as in Theorem 4 above, the regularity condition becomes the so-called *Mangasarian–Fromowitz constraint qualification* that is a linear independence of the gradients $h_1(\,\cdot\,;\mu^*),\dots,h_k(\,\cdot\,;\mu^*)$ and the existence of a signed measure $\zeta \in \widetilde{\mathbf{M}}_{\mathrm{f}}$ such that

$$\begin{cases} \langle h_i,\zeta\rangle = 0 & \text{for all } i=1,\dots,m; \\ \langle h_j,\zeta\rangle < 0 & \text{for all } j \in \{m+1,\dots,k\} \text{ for which } H_j(\mu^*)=0. \end{cases} \tag{10}$$

Without inequality constraints, (10) holds trivially for ζ being the zero-measure and we come to the following important corollary giving the first-order necessary condition for optimisation with a fixed total mass.

Theorem 5. *Let f be continuous on $\mathbf{M}_{\mathrm{f}}$ and strongly differentiable at $\mu^* \in \mathbf{M}_{\mathrm{f}}$ with a gradient function $g(x;\mu^*)$. If μ^* is a local minimum in the constrained optimisation problem*

$$f(\mu) \to \inf \quad \textit{subject to} \tag{11}$$

$$\begin{cases} \mu \in \mathbf{M}_{\mathrm{f}} \\ \mu(X)=a>0, \end{cases} \tag{12}$$

then there exists a real u such that

$$\begin{cases} g(x;\mu^*) = u & \mu^* - a.e., \\ g(x;\mu^*) \geq u & \text{for all } x \in X. \end{cases} \tag{13}$$

3 Analyticity of the expectation

The linear structure on the set of measures described in the previous section makes it possible to put analysis of variations of the intensity measure in the general framework of differential calculus on a Banach space. In this section we fix a functional $F: \mathbf{N} \mapsto \mathbb{R}$ on the configuration space and reagard its expectation $\mathbb{E}_\mu F(\eta)$ as a function of a measure μ. To explain the idea, we first consider a bounded functional F and the Banach space $\mathbf{M}_\mathrm{f}$ of finite measures and then discuss extensions to a wider class of functionals and to infinite measures.

It is a well-known fact that for a Poisson process η with a finite intensity measure μ, the conditional distribution of its points given their total number $\eta(X) = n$ corresponds to n points independently drawn from the distribution $(\mu(X))^{-1}\mu$. This observation after applying the total probability formula gives rise to the following expression for the expectation:

$$\mathbb{E}_\mu F(\eta) = F(\emptyset) + e^{-\mu(X)} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{X^n} F(\delta_{x_1} + \cdots + \delta_{x_n})\, \mu(dx_1) \ldots \mu(dx_n), \tag{14}$$

where $\emptyset$ stands for the empty configuration.

Substituting $\mu \leftarrow (\mu + \nu)$ for a signed measure $\nu \in \widetilde{\mathbf{M}}_\mathrm{f}$ such that $\mu + \nu \in \mathbf{M}_\mathrm{f}$ into (14),

$$\begin{aligned} \mathbb{E}_{\mu+\nu} F(\eta) &= e^{-\mu(X)}(1 - \nu(X) + o(\|\nu\|)) \times \\ &\qquad \Big[F(\emptyset) + \sum_{n=1} \frac{1}{n!} \int_{X^n} F(\textstyle\sum_{i=1}^{n} \delta_{x_i})\, (\mu+\nu)(dx_1) \ldots (\mu+\nu)(dx_n)\Big] \\ &= \mathbb{E}_\mu F + e^{-\mu(X)} \sum_{n=1} \frac{n}{n!} \int_{X^n} F(\textstyle\sum_{i=1}^{n} \delta_{x_i})\, \mu(dx_1) \ldots \mu(dx_{n-1})\nu(dx_n) \\ &\qquad - \nu(X) e^{-\mu(X)} \sum_{n=0} \frac{1}{n!} \int_{X^n} F(\textstyle\sum_{i=1}^{n} \delta_{x_i})\, \mu(dx_1) \ldots \mu(dx_n) + o(\|\nu\|). \end{aligned}$$

Thus

$$\begin{aligned}
&\mathbb{E}_{\mu+\nu}F(\eta)-\mathbb{E}_{\mu}F(\eta)\\
&\qquad=e^{-\mu(X)}\sum_{n=0}\frac{1}{n!}\int_{X^{n+1}}F\big(\textstyle\sum_{i=1}^{n}\delta_{x_i}+\delta_x\big)\,\mu(dx_1)\dots\mu(dx_n)\nu(dx)\\
&\quad-e^{-\mu(X)}\sum_{n=0}\frac{1}{n!}\int_{X^{n+1}}F\big(\textstyle\sum_{i=1}^{n}\delta_{x_i}\big)\,\mu(dx_1)\dots\mu(dx_n)\nu(dx)\;+o(\|\nu\|)\\
&\qquad\qquad=\mathbb{E}_{\mu}\int_X[F(\eta+\delta_x)-F(\eta)]\,\nu(dx)+o(\|\nu\|).
\end{aligned}$$

Denoting by D_x the *difference operator* $D_xF(\eta)=F(\eta+\delta_x)-F(\eta)$, we see that

$$\mathbb{E}_{\mu+\nu}F-\mathbb{E}_{\mu}F=\langle\mathbb{E}_{\mu}D_{\cdot}F,\nu\rangle+o(\|\nu\|). \tag{15}$$

Since F is bounded, so is $\mathbb{E}_{\mu}D_{\cdot}F$, hence $\mathbb{E}_{\mu}F$ is strongly differentiable on $\mathbf{M}_{\mathrm{f}}$ with the gradient function $\mathbb{E}_{\mu}D_xF$.

Using the infinite series Taylor expansion in $\nu(X)$, one can extend the above argument to show not only differentiability, but also *analyticity* of $\mathbb{E}_{\mu}F$ as a function of μ. Introduce iterations of the operator D_x by setting $D^0F=F$, $D^1_{x_1}F=D_{x_1}F$, $D^n_{x_1,\dots,x_n}F=D_{x_n}(D^{n-1}_{x_1,\dots,x_{n-1}}F)$ so that

$$D^n_{x_1,\dots,x_n}F(\eta)=\sum_{J\subseteq\{1,2,\dots,n\}}(-1)^{n-|J|}F\big(\eta+\textstyle\sum_{j\in J}\delta_{x_j}\big), \tag{16}$$

as it can be easily checked.

Theorem 6. *Assume that there exists a constant $b>0$ such that $|F\big(\Sigma_{i=1}^n\delta_{x_i}\big)|\le b^n$ for all $n\ge 0$ and $(x_1,\dots,x_n)\in X^n$. Then $\mathbb{E}_{\mu}F(\Pi)$ is analytic on $\mathbf{M}_{\mathrm{f}}$ and*

$$\mathbb{E}_{\mu+\nu}F=\sum_{n=0}^{\infty}\frac{1}{n!}\int_{X^n}\mathbb{E}_{\mu}D^n_{x_1,\dots,x_n}F\big(\eta+\textstyle\sum_{i=1}^{n}\delta_{x_i}\big)\,\nu(dx_1)\dots\nu(dx_n), \tag{17}$$

where the term corresponding to $n=0$ is, by convention, $\mathbb{E}_{\mu}F(\eta)$.

The proof can be found in [Molchanov and Zuyev(2000b)]. Notice that the integral above is an n-linear form of the n-th product measure (the n-th differential) and that

$$\mathbb{E}_{\mu}D^n_{x_1,\dots,x_n}F(\eta)=\sum_{m=0}^{n}(-1)^{n-m}\binom{n}{m}\mathbb{E}_{\mu}F\big(\eta+\textstyle\sum_{j=1}^{m}\delta_{x_j}\big) \tag{18}$$

because of the symmetry with respect to permutations of $x_1,\dots,x_n$.

3.1 Margulis–Russo type formula for Poisson process

An important case of perturbations of the intensity measure is when the increment is proportional to the measure itself. So fix a $\mu\in\mathbf{M}_{\mathrm{f}}$ and consider $\nu=t\mu$ for a small $t\in(-1,1)$. Substituting this into (17) gives a power series in t:

$$\mathbb{E}_{\mu+t\mu}F = \sum_{n=0}^{\infty} \frac{t^n}{n!} \int_{X^n} \mathbb{E}_{\mu} D^n_{x_1,\dots,x_n} F\big(\eta + \sum_{i=1}^{n} \delta_{x_i}\big)\, \mu(dx_1)\dots\mu(dx_n). \qquad (19)$$

In particular,

$$\frac{d}{ds}\mathbb{E}_{s\mu}F(\eta) = \int_X \mathbb{E}_{s\mu} D_x F(\eta)\, \mu(dx) = \int_X \mathbb{E}_{s\mu}[F(\eta+\delta_x) - F(\eta)]\, \mu(dx). \qquad (20)$$

Let $F(\Pi) = \mathbb{1}_{\Xi}(\eta)$ be an indicator of some event Ξ. The integration in the last expression can be restricted to the (random) set $\Upsilon(\eta) = \{x \in X :\ \mathbb{1}_{\Xi}(\eta+\delta_x) \neq \mathbb{1}_{\Xi}(\eta)\}$ leading to

$$\frac{d}{ds}\mathbb{P}_{s\mu}(\Xi) = \mathbb{E}_{s\mu} \int_X \mathbb{1}_{\Xi}(\eta+\delta_x)\mathbb{1}_{\Upsilon(\eta)}(x)\, \mu(dx) - \mathbb{E}_{s\mu} \int_X \mathbb{1}_{\Xi}(\eta)\mathbb{1}_{\Upsilon(\eta)}(x)\, \mu(dx).$$

The last term is obviously $\mathbb{E}_{s\mu}\mathbb{1}_{\Xi}(\eta)\mu(\Upsilon(\eta))$. For the first one, we apply the Refined Campbell theorem together with the Slivnyak–Mecke formula

$$\mathbb{E}_{\mu} \int_X f(x,\eta)\eta(dx) = \mathbb{E}_{\mu} \int_X f(x,\eta+\delta_x)\mu(dx) \qquad (21)$$

valid for any measurable $f : X \times \mathscr{X} \mapsto \mathbb{R}_+$ which characterises the Poisson process, see, e.g., Prop. 13.1.IV and Prop. 13.1.VII in [Daley and Vere-Jones(2008)]. Using (21),

$$\mathbb{E}_{s\mu} \int_X \mathbb{1}_{\Xi}(\eta+\delta_x)\mathbb{1}_{\Upsilon(\eta)}(x)\, \mu(dx) = \frac{1}{s} \int_X \mathbb{1}_{\Xi}(\eta)\mathbb{1}_{\{x:\ \mathbb{1}_{\Xi}(\eta)\neq\mathbb{1}_{\Xi}(\eta-\delta_x)\}}(x)\, \eta(dx).$$

Combining all together,

$$\frac{d}{ds}\mathbb{P}_{s\mu}(\Xi) = \frac{1}{s}\mathbb{E}_{s\mu}\mathbb{1}_{\Xi}(\eta)N_{\Xi}(\eta) - \mathbb{E}_{s\mu}\mathbb{1}_{\Xi}(\eta)V_{\Xi}(\eta). \qquad (22)$$

Here $V_{\Xi}(\eta) = \mu\{x \in X :\ \mathbb{1}_{\Xi}(\eta+\delta_x) \neq \mathbb{1}_{\Xi}(\eta)\}$ is the μ-content of the set where adding a new point to configuration η would change the occurrence of Ξ, so the elements of this set are called *pivotal locations* for event Ξ in configuration η. While $N_{\Xi}(\eta) = \int_X \mathbb{1}\{x \in \eta :\ \mathbb{1}_{\Xi}(\eta) \neq \mathbb{1}_{\Xi}(\eta)\}\eta(dx)$, in the case of non-atomic μ, is equal to the number of points in configuration η which removal would change the occurrence of Ξ. Such configuration points are called *pivotal points* for event Ξ in configuration η. This geometric interpretation is a key to usefulness of this formula which is a counterpart of the Margulis–Russo formula for Bernoulli fields first proved in [Margulis(1974)] and independently in [Russo(1981)]. Identity (22) was shown in [Zuyev(1992)] in more restrictive settings.

Let us mention two useful implications of (22):

$$\frac{d}{ds}\log\mathbb{P}_{s\mu}(\Xi) = \frac{1}{s}\mathbb{E}_{s\mu}[N_{\Xi}(\eta) \;\big|\; \Xi] - \mathbb{E}_{s\mu}[V_{\Xi}(\eta) \;\big|\; \Xi]$$

obtained by dividing both parts by $\mathbb{P}_{s\mu}(\Xi)$, and consequently,

$$\mathbb{P}_{s_2\mu}(\Xi) = \mathbb{P}_{s_1\mu}(\Xi)\exp\Big\{\int_{s_1}^{s_2} \mathbb{E}_{s\mu}[s^{-1}N_\Xi(\eta) - V_\Xi(\eta) \;\big|\; \Xi]\,ds\Big\}$$

providing a way to control the change in the probability of an event in terms of the control over the number of pivotal points versus the μ-content of the pivotal locations.

3.2 Infinite measures

To extend the formula (17), or at least its first k-th term expansion, to infinite mass measures one must put additional assumptions on the functional F as there are examples of a bounded functional which expectation is, however, not differentiable. A notable example is the indicator that the origin belongs to an infinite cluster in a Boolean model of spheres in $\mathbb{R}^d$, $d \geq 2$. Its expectation is the density of the infinite cluster which is *not* differentiable at the percolation threshold.

One possible approach is to consider a growing sequence of compact sets $\{X_n\}$ such that $\cup_n X_n = X$ and the corresponding restrictions η_n of the Poisson process η onto X_n. Since these are finite processes, then if $F(\eta_n)$ converges to $F(\eta)$ (such functionals are called *continuous at infinity*), then by controlling this convergence it is possible to assure that the corresponding derivatives also converge. This approach was taken in [Molchanov and Zuyev(2000b)] where, in particular, it was shown that if F is bounded and continuous at infinity then (17) holds for a locally-finite μ and a finite ν such that $\mu+\nu$ is a positive measure, see [Molchanov and Zuyev(2000b), Th.2.2]. Note that the indicator function that the origin is in an infinite cluster is not continuous at infinity.

A more subtle method is based on the Fock space representation and it allows to extend the expansion formula to square-integrable functionals. Consider two locally finite non-negative measures μ and $\lambda = \mu + \nu$ and another measure ρ dominating their sum. Denote by h_λ and h_ν the corresponding Radon–Nikodym densities. The following result is proved in [Last(2013)].

Theorem 7. *Assume that*

$$\langle (1-h_\nu)^2, \rho\rangle + \langle (1-h_\lambda)^2, \rho\rangle < \infty.$$

Let F be such that $\mathbb{E}_\rho F(\eta)^2 < \infty$. Then (17) *holds, all the integrals there exist and the series converges absolutely.*

Perhaps, the most important case is when the increment measure ν is absolutely continuous with respect to μ with the corresponding density h_ν. Then the above theorem implies that for F such that $\mathbb{E}_{\mu+\nu}F^2(\eta) < \infty$, condition $\langle h_\nu(1+h_\nu)^{-1}, \mu\rangle < \infty$ is sufficient for (17) to hold.

Note an interesting fact on the validity of the expansion formula. Each general increment measure ν can be represented as $\nu = \nu_1 + \nu_2$, where ν_1 is absolutely continuous with respect to μ and ν_2 is orthogonal to it. In order for (17) to hold

for *all bounded* F, it is necessary that $\nu_2(X) < \infty$! This and other results on infinite measure case can be found in [Last(2013)].

4 Asymptotics in the high-intensity setting

Consider the minimisation problem

$$f(\mu) = \mathbb{E}_\mu F(\eta) \to \inf \quad \text{subject to } \mu \in \mathbf{M}_{\mathrm{f}} \text{ and } \mu(X) = a\,, \tag{23}$$

where F is a functional satisfying the conditions of Theorem 6. For simplicity, we consider only the case of the fixed total mass and refer to [Molchanov and Zuyev(2000b)] for more general cases.

It is rearly possible to find analytic solution to (23), but Theorem 5 opens a possibility to use gradient descent type methods in order to numerically solve it as described later in Section 5. However, when the total mass a is large, in many cases it is possible to come up with asymptotic properties of the optimal measure that solves the optimisation problem (23) for a that grows to the infinity.

The key idea is to rescale the optimal measure around some point x, so it looks like proportional to the Lebesgue measure. In the case of a stationary point processes, it is then easier to calculate the first difference in order to equate it to a constant to satisfy the necessary condition (13) for the minimum.

Assume that X is a compact subset of $\mathbb{R}^d$ that coincides with the closure of its interior and let $\gamma_a^x(y) = x + a^{1/d}(y-x)$ denote the rescaling around the point $x \in \mathbb{R}^d$, so that $\gamma_a^x\eta$ consists of points $\gamma_{a^{-1}}^x x_i$ for $\eta = \{x_i\}$. Consider a solution to (23) which we represent in the form $a\mu_a$ for some probability measure μ_a. In particular,

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

where $\hat{\mu}_a^x(\cdot) = a\mu_a(\gamma_{a^{-1}}^x \cdot)$. Assume that μ_a is absolutely continuous with density p_a with respect to the Lebesgue measure ℓ_d. Then $\hat{\mu}_a^x(\cdot)$ has density $p_a(\gamma_{a^{-1}}^x y)$ on $\gamma_a^x X$. The key idea is that in some situations the expected first difference

$$\mathbb{E}_{a\mu_a} D_x(\eta) = \mathbb{E}_{\hat{\mu}_a^x} D_x(\gamma_a^x\eta) \propto g(a)\mathbb{E}_{p(x)\ell_d} \Gamma(x;\eta)$$

for a function $\Gamma(x;\eta)$ that depends on η locally in a possibly random neighbourhood of x. Then the gradient function used in Theorem 5 can be calculated for a stationary Poisson process with intensity $p(x)$ which is generally easier.

To make precise the local structure of $\Gamma(x;\eta)$, we need the concept of a stopping set that is a multidimensional analogue of a stopping time, see [Zuyev(1999)]. Let $\mathscr{A}_B$ be the σ-algebra generated by events $\{\eta(B) = n\}$ for $n \geq 0$ and a Borel set B. A random compact set S is called a *stopping set* if $\{S \subseteq K\} \in \mathscr{A}_K$ for any compact set K in $\mathbb{R}^d$. The stopping σ-algebra is the collection of events $A \in \mathscr{A}$ such that $A \cap \{S \subseteq K\} \in \mathscr{A}_K$ for all compact K.

The following result is proved in [Molchanov and Zuyev(2000b)].

Theorem 8. *Let $a\mu_a$ be a measure solving* (11) *for the fixed total mass a. Assume that for an inner point x of X the following condition holds.*

(M) For all sufficiently large a, μ_a is absolutely continuous with respect to ℓ_d with densities p_a, and there exists a finite double limit

$$\lim_{y\to x,\, a\to\infty} p_a(y) = p(x) > 0\,. \tag{24}$$

Furthermore, assume that for the same x, the first difference D_xF satisfies the following conditions.

(D) For some positive function $g(a)$, the random variable

$$\Gamma_a = \Gamma_a(x;\eta) = D_x(\gamma_a^x\eta)/g(a)$$

converges to $\Gamma = \Gamma(x;\eta)$ as $a\to\infty$ for almost all realisations of the stationary Poisson process η with unit intensity, and

$$0 < \mathbb{E}_{p(x)\ell_d}\Gamma(x;\eta) < \infty\,.$$

(L) There exist a family of stopping sets $S_a = S_a(x;\eta)$ and a stopping set $S(x;\eta)$ such that $\Gamma_a(x;\eta)$ is $\mathscr{A}_{S_a}$-measurable for all sufficiently large a; $\Gamma(x;\eta)$ is $\mathscr{A}_S$-measurable; and for every compact set W containing x in its interior

$$\mathbb{1}_{S_a(x;\eta)\subseteq W} \to \mathbb{1}_{S(x;\eta)\subseteq W} \quad \text{as } a\to\infty$$

for almost all realisations of a stationary unit intensity Poisson process η.

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

$$\lim_{a\to\infty,\, n\to\infty} \mathbb{E}_{\hat{\mu}_\xi^x}|\Gamma_a(x;\eta)|\mathbb{1}_{S_a\subseteq W} = 0$$

and there exists a constant $M = M(W,b)$ such that $|\Gamma_a(x;\eta)| \le M$ for all sufficiently large a and η such that $S_a(x;\eta)\subset\gamma_b^xW$.

Then

$$\lim_{a\to\infty} |\mathbb{E}_{\hat{\mu}_a^x}\Gamma_a(x;\eta) - \mathbb{E}_{p(x)\ell_d}\Gamma(x;\eta)| = 0$$

and

$$\lim_{a\to\infty} \frac{\mathbb{E}_{a\mu_a}D_xF}{\mathbb{E}_{ap(x)\ell_d}D_xF} = 1\,.$$

The uniform integrability condition **(UI)** can be efficiently verified for stopping sets S_a and S that satisfy the condition $\ell_d(B)\ge\alpha\ell_d(S_a)$ for some fixed α and almost all x, see [Molchanov and Zuyev(2000b), Th. 5.4]. We now show how this theorem applies to various problems of a practical interest.

4.1 Approximation of functions

Consider a strictly convex function $f(x)$, $x \in [a,b] \subset \mathbb{R}$, and its linear spline approximation $s(x;\eta)$ build on the grid of points $a \le x_1 \le x_2 \le \cdots \le x_N \le b$, where $\{x_1,\dots,x_N\}$ form a Poisson point process η on $[a,b]$. Since the end-points are included as the spline knots, the spline approximation is well-defined even if η is empty. The quality of approximation is measured in the L^1-distance as

$$F(\eta) = \int_a^b (s(x;\eta) - f(x))dx.$$

If instead of a Poisson process η one takes a set of deterministic points, the problem of determining the best locations of those points have been considered in [McClure and Vitale(1975)] (in relation to approximation of convex sets), see also [Schneider(1988)]. It is well known that the empirical probability measure generated by the best deterministic points converges weakly to the measure with density proportional to $f''(x)^{1/3}$.

If η is a Poisson process of total intensity a, then the optimisation problem aims to determine the asymptotic behaviour of the intensity measure $\mu = a\mu_a$ that minimises $\mathbb{E}_\mu F(\eta)$. The key observation is that the first difference $D_x F(\eta)$ equals the area of the triangle with vertices at $(x, f(x))$, $(x^-, f(x^-))$ and $(x^+, f(x^+))$, where x^- and x^+ are left and right neighbours to x from η. Denoting $r_x^- = x - x^-$ and $r_x^+ = x^+ - x$, we arrive at the expected first difference (the gradient function) given by

$$g(x;\mu) = \mathbb{E}_\mu D_x F = -f(x)[\mathbb{E}_\mu r_x^- + \mathbb{E} r_x^+] + \mathbb{E}_\mu r_x^- \mathbb{E}_\mu f(x+r_x^+) + \mathbb{E}_\mu r_x^+ \mathbb{E}_\mu f(x+r_x^-).$$

If μ is an optimal measure, then the strict convexity and continuity properties imply that (13) holds for all $x \in [a,b]$. It is easy to write down the distributions of r_x^- and r_x^+ in terms of μ. Then the requirement $g(x;\mu) = \text{const}$ turns into a sytem of four differential equations. However, one is interested in the asymptotic solution when a is large, so the high intensity framework is very much relevant in this setting. Notice that here

$$\Gamma(x;\eta) = -\frac{1}{4} f''(x) r_x^+ r_x^- (r_x^+ + r_x^-)$$

depends only on the stopping set $[x^-, x^+]$ that shrinks to $\{x\}$ as the total mass a of the measure $\mu = a\mu_a$ grows. If μ is proportional to the Lebesgue measure ℓ_1, then it is easy to calculate the first difference explicitly as

$$\mathbb{E}_{p(x)\ell_1} \Gamma(x;\eta) \propto -f''(x) p(x)^{-3}.$$

By Theorem 8, if (24) holds, then it is possible to equate the right-hand side to a constant, so that the density of the optimal measure μ_a is asymptotically proportional to $f''(x)^{1/3}$, exactly as it is in the deterministic case. The same argument applies to a strictly convex function $f(x)$ for x from a convex compact subset of $\mathbb{R}^d$ and leads to the asymptotically optimal measure with density propor-

tional to $K(x)^{1/(2+d)}$, where $K(x)$ is the Gaussian curvature of f at point x, see [Molchanov and Zuyev(1997)]. The multidimensional optimal approximation results for deterministic sets of points (including also the Bezier approximation) are also studied in [Molchanov and Tontchev(2007a)].

4.2 Clustering

Consider the data set $\{y_1,\dots,y_m\}$ in $\mathbb{R}^d$. One of objectives in the cluster analysis consists in determining cluster centres $\eta = \{x_1,\dots,x_k\} \subset \mathbb{R}^d$ for some given k. Each cluster centre x_i is associated with the data points (also referred to as daughter points) which are nearest to it, i.e. lie within the corresponding *Voronoi cell* C_{x_i} (see, e.g., [Okabe et al(2000)Okabe, Boots, Sugihara, and Chiu] for the definition and properties of the Voronoi tessellations). The cluster centres can be determined using the Ward-type criterion by minimising

$$F(\eta) = \sum_{x_i\in\eta} \sum_{y_j\in C_{x_i}} \|x_i - y_j\|^2 ,$$

which is also the trace of the pooled within groups sum of squares matrix. In view of this criterion function, the optimal set of k cluster centres is also called the k-means of the data, see [Molchanov et al(2001)Molchanov, van Lieshaut, and Zuyev] for further references on this topic. In most applications, the number k is predetermined and then a steepest descent algorithm is employed to find the cluster centres. It should be noted that the functional $F(\eta)$ is not convex and so the descent algorithms might well end up in a local rather than a global minimum.

Alternatively, if the cluster centres are regarded as points of a Poisson point process with intensity measure μ and the mean of $F(\eta)$ is taken as an objective function, then

$$\mathbb{E}_\mu F(\eta) = \mathbb{E}_\mu \left[\sum_{x_i\in\eta} \sum_{y_j\in C_{x_i}} \|x_i - y_j\|^2 \right] = \sum_{j=1}^m \mathbb{E}_\mu \rho(y_j,\eta)^2 ,$$

where $\rho(y,\eta)$ is the Euclidean distance from y to the nearest point of η. Since η can be empty, we have to assign a certain (typically large) value u to $\rho(y,\emptyset)$. Since η is a Poisson process, it is easy to compute the latter expectation in order to arrive at

$$\mathbb{E}_\mu F(\eta) = \sum_{j=1}^m \int_0^{u^2} \exp\{-\mu(B_{\sqrt{t}}(y_j)\}dt , \tag{25}$$

which is a convex functional of μ. Since taking the expectation in the Poissonised variant of the clustering problem yields a convex objective function, the steepest descent algorithm applied in this situation would always converge to the global

minimum. The optimal measure μ can be termed as the solution of the P-means problem.

In the asymptotic setting, it is assumed that the total mass a of the optimal measure $a\mu_a$ is growing to infinity and the data points are sampled from a probability distribution with density p_ν, so that the empty configurations η are no longer relevant and the objective function becomes

$$\mathbb{E}_\mu F(\eta) = \int_{\mathbb{R}^d} \mathbb{E}_\mu[\rho(y,\eta)^2] p_\nu(y) dy\,.$$

Adding an extra cluster point x affects only the data points within the so-called Voronoi flower of x, see [Okabe et al(2000)Okabe, Boots, Sugihara, and Chiu]. The Voronoi flower is a stopping set that satisfies the conditions of Theorem 8. Since $\mathbb{E}_{a\ell_d} D_x F$ is proportional to $p_\nu(x) a^{-1-2/d}$, the high intensity solution has the density proportional to $p_\nu(y)^{d/(d+2)}$.

A similar problems appears in the telecommunication setting, where the data points y_j represent the customers and $x_1,\dots,x_k$ are the locations of server stations. If the connection cost of a customer to the server is proportional to the β-power of the Euclidean distance between them (so that $\beta = 2$ in the clustering application), then the density of the high intensity solution is proportional to $p_\nu(y)^{d/(d+\beta)}$, see [Molchanov and Zuyev(1997), Molchanov and Zuyev(2000b)]. This problem is also known in computational geometry under the name of the mail-box problem. Another similar application is the optimal stratification in Monte Carlo integration, see, e.g., XXX (ref needed!)

4.3 Optimal quantisation

The optimal server placement problem from the previous section can be thought of as a representation of a measure ν on $\mathbb{R}^d$ (that describes the probability distribution of customers) by another (discrete) measure with k atoms. This is a well-known optimal quantisation problem, see [Graf and Luschgy(2000), Gruber(2004)]. Apart from finding the optimal quantiser, it is important to know the asymptotic behaviour of the quantisation error, which is the infimum of the objective function. The classical quantisation theory concerns the case when the quantiser is deterministic. We follow a variant of this problem for quantising points that form a Poisson point process of total intensity a studied in [Molchanov and Tontchev(2007b)].

Let $p(y)$, $y \in \mathbb{R}^d$ be a Riemann integrable function with bounded support K that is proportional to the density of the probability measure to be approximated by a discrete one. The objective functional for the optimal Poisson quantisation problem is then

$$E(p;\mu) = \int_{\mathbb{R}^d} \mathbb{E}_\mu \rho(y,\eta)^\beta p(y) dy\,.$$

Denote

$$E_a(p) = n^{\beta/d} \inf_{\mu \in \mathbf{M}_{\mathrm{f}},\, \mu(\mathbb{R}^d)=a} E(p;\mu)\,.$$

Theorem 9. *The limit of $E_n(p)$ as $n \to \infty$ exists and*

$$\lim_{n\to\infty} E_n(p) = \mathscr{I}\|p\|_{d/(d+\beta)} = \mathscr{I}\left(\int_{\mathbb{R}^d} p(y)^{d/(d+\beta)}dy\right)^{1+\beta/d}$$

for a certain constant $\mathscr{I}$ that depends only on β and dimension d. If $a\mu_a$ is supported by K and minimises $E(p;\mu)$ over all measures with the total mass a, then μ_a weakly converges as $a \to \infty$ to the probability measure with density proportional to $p(y)^{d/(d+\beta)}$.

The proof from [Molchanov and Tontchev(2007b)] does not rely on Theorem 8. Theorem 9 is proved first for the uniform distribution $p(y) \equiv const$ and then extended to a non-uniform case. The main idea is the firewall construction from [Graf and Luschgy(2000)] that ensures the additivity of the objective functional for indicators of disjoint sets. The main new feature in the Poisson case is that the firewalls constructed by adding extra cluster points in the stochastic case correspond to the changes in the intensity and so may be empty. Bounds on the coverage probabilities from [Hall(1985)] are used in oder to ensure that the firewalls are established with a high probability. The constant $\mathscr{I}$ is the limit of the quantisation error for the uniform distribution on the unit cube.

Note that laws of large numbers for functionals of point processes have been considered in [Penrose and Yukich(2003)]. They make it possible to obtain the limit of a functional of a Poisson process with intensity measure $a\mu$ for any given μ as $a \to \infty$. However, [Penrose and Yukich(2003)] does not contain any results about convergence of minimal values and minimisers. By examining the proof of [Penrose and Yukich(2003), Lemma 3.1] it is possible to justify the uniform convergence of the rescaled functional of $a\mu_a$ for a measure μ_a with density p_a (and so arrive at the convergence results for minimal values) if

$$a\int_{\|y-x\|\le a^{-1/d}} |p_a(y) - p_a(x)|dy \to 0 \quad \text{as } a \to \infty \tag{26}$$

for all $x \in \mathbb{R}^d$. If $p_a(x) \to p(x)$ as $a \to \infty$, (26) implies the validity of the double limit condition (24).

4.4 Optimal search

Let Y be a random closed subset of $\mathbb{R}^d$ that is independent of the Poisson process η. The aim is to determine the intensity measure μ that maximises the coverage probability $\mathbb{P}\{\eta(Y) > 0\}$ meaning that at least one point of η hits Y. Equivalently, it is possible to minimise the avoidance probability

$$\mathbb{E}_\mu \mathbb{1}_{\eta(Y)=0} = \mathbb{E}_\mu e^{-\mu(Y)}.$$

The expected first difference is given by

$$g(x;\mu) = \mathbb{E}_\mu D_x F = -\mathbb{E}_\mu \left[e^{-\mu(Y)} \mathbb{1}_{x\in Y} \right].$$

If Y is a subset of a countable space, it is possible to determine μ explicitly, see [Molchanov and Zuyev(1997), Sec. 5.5]. Otherwise, the high intensity approach applies. For instance, if $Y = B_\zeta(\xi)$ is a random ball of radius ζ centred at an independent ξ with probability densities p_ζ and p_ξ, then

$$\mathbb{E}_{p(x)\ell_d} \Gamma(x;\eta) \propto -\kappa_d p_\xi(x) \left[\frac{p_\eta(0)(d+1)\Gamma(1+1/d)}{(ap(x)\kappa_d)^{1+1/d}} + \frac{p'_\eta(0)(d+2)\Gamma(1+2/d)}{(ap(x)\kappa_d)^{1+2/d}} + \cdots \right],$$

where κ_d is the volume of a unit ball in $\mathbb{R}^d$. Thus, the density of the asymptotically optimal measure is proportional to $(p_\xi)^{d/(d+1)}$ if $p_\eta(0) \neq 0$, and to $(p_\xi)^{d/(d+2)}$ if $p_\eta(0) = 0$ and $p'_\eta(0) \neq 0$, etc.

5 Steepest descent algorithms

Algorithms of the steepest descent type are widely used in the optimisation literature see, e.g., [Polak(1997)]. The basic steepest descent algorithm consists in moving from a measure μ_n (approximate solution at step n) to $\mu_{n+1} = \mu_n + \nu_n$, where ν_n minimises the directional derivative, which in our context becomes $L(\mu)[\nu] = \langle g(\cdot;\mu), \nu \rangle$ with $g(x;\mu) = \mathbb{E}_\mu D_x F(\eta)$.

The general description of the steepest descent direction from [Molchanov and Zuyev(2002), Th. 4.1] in the case of optimisation over intensity measures with a fixed total mass yields the following result.

Theorem 10. *The minimum of $L(\mu)[\nu]$ over all $\nu \in \widetilde{\mathbf{M}}_{\mathrm{f}}$ with $\|\nu\| \le \varepsilon$ is achieved on a signed measure ν such that ν^+ is the positive measure with total mass $\varepsilon/2$ concentrated on the points of the global minima of $g(x;\mu)$ and $\nu^- = \mu|_{M(t_\varepsilon)} + \delta\mu|_{M(s_\varepsilon)\setminus M(t_\varepsilon)}$, where*

$$M(p) = \{x \in X : \ g(x;\mu) \ge p\},$$

and

$$t_\varepsilon = \inf\{p : \ \mu(M(p)) < \varepsilon/2\},$$
$$s_\varepsilon = \sup\{p : \ \mu(M(p)) \ge \varepsilon/2\}.$$

The factor δ is chosen in such a way that $\mu(M(t_\varepsilon)) + \delta\mu(s_\varepsilon)) = \varepsilon/2$.

This result means that the mass of μ is eliminated at high gradient locations, while μ acquires extra atoms at locations where the gradient is the smallest.

In a numeric implementation, the space X is discretised and the discrete variant of μ is considered. The corresponding steepest descent algorithms are used in R-libraries mefista (for optimisation with a fixed mass) and *medea* (for optimisation with many linear equality constraints) available from the authors' web-pages. The increment step size in these algorithms is chosen by either the Armijo method described in [Polak(1997), Sec. 1.3.2] or by taking into account the difference between the supremum and the infimum of $g(x;\mu_n)$ over the support of μ_n.

Numeric computations of an optimal measure relies on effective evaluation of the gradient function which is possible to obtain in many cases as the next sections demonstrate.

5.1 Design of experiments

The basic problem in the theory of linear optimal design of experiments [Atkinson and Donev(1992)] aims to find positions of design (observation) points x_i in order to minimise the determinant of the covariance matrix of estimators of coefficients β_j in the linear regression model

$$y_i = \sum_{j=1}^{k} \beta_j r_j(x_i) + \varepsilon_i \,,$$

where $r = (r_1, \ldots, r_k)^\top$ is a column vector of linearly independent functions and ε_i are i.i.d. centred errors. If the design points are produced from a probability distribution $\mu(dx)$ reflecting the frequency of taking x as an observation point, the objective function can be expressed as

$$f(\mu) = -\log\det \int M(\mu) \,,$$

where the covariance matrix M is given by

$$M(\mu) = r(x)^\top r(x) \mu(dx) \,.$$

For the optimisation purpose, it is possible to discard the logarithm, so that the gradient function in this model becomes

$$g(x;\mu) = -r(x) M^{-1}(x) r^\top(x) \,,$$

see [Molchanov and Zuyev(2000c), Molchanov and Zuyev(2004)]. It is also possible to consider the Poissonised variant of the optimal design problem. It should be noted however that adding an extra design point has a non-local effect and so the high-intensity approach from Section 4 does not apply in these problems.

5.2 Mixtures

Let $\{p_x(\cdot)\}$ be a family of probability densities indexed by $x \in X$. For a probability measure μ on X define the mixture

$$p_\mu(y) = \int_X p_x(y)\mu(dx)\,.$$

The estimation of the mixing distribution μ is a well-studied topic in statistics. The steepest descent algorithm in the space of measures yields a pure non-parametric approach to the estimation of μ based on maximising the log-likelihood

$$f(\mu) = \sum_{i=1}^{n} \log p_\mu(y_i)$$

based on a sample $y_1, \ldots, y_n$. The gradient function is

$$g(x;\mu) = \sum_{i=1}^{n} \frac{p_x(y_i)}{\int p_x(y)\mu(dx)}\,.$$

5.3 P-means

A direct computation shows that the gradient of the functional (25) is given by

$$g(x;\mu) = -\sum_{y_j} \int_{\|x-y_j\|^\beta}^{u^2} \exp\{-\mu(B_{\sqrt{t}}(y_j))\}dt\,.$$

5.4 Maximisation of the covered colume

Let η be a Poisson process in $X \subset \mathbb{R}^d$ with intensity measure μ. If $B_r(x)$ is a ball of radius r centred at x, then

$$\Xi = \bigcup_{x_i \in \eta} B_r(x_i)$$

is called a *Boolean model*, see [Molchanov(1997), Stoyan et al(1995)Stoyan, Kendall, and Mecke]. The ball of radius r is referred to as the *typical grain*, which can be also a rather general random compact set. Then

$$\mathbb{P}\{x \notin \Xi\} = \exp\{-\mu(B_r(x))\}\,.$$

Fubini's theorem yields that the expected uncovered volume is given by

$$f(\mu) = \int_X \mathbb{P}\{x \notin \Xi\} dx = \int_X \exp\{-\mu(B_r(x))\} dx.$$

A minimiser of $f(\mu)$ yields the intensity of a Poisson process with the largest coverage. The gradient is directly computed as

$$g(x;\mu) = -\int_{B_r(x)} \exp\{-\mu(B_r(z))\} dz.$$

Further related problems are discussed in [Molchanov et al(2000)Molchanov, Chiu, and Zuyev] in relation to design of materials with given properties. This problem does not admit the high-intensity solution, since adding an extra ball affects the configuration within distance r which does not go to zero as the intensity of the Poisson process grows.

Acknowledgements The authors thank the organisers of the Oberwolfach Mini-Workshop: Stochastic Analysis for Poisson point processes (Feb. 2013) and the hospitality of the Oberwolfach Mathematical Institute for an excellent forum to exchange ideas partly presented in this paper.

References

[Albeverio et al(1998)Albeverio, Kondratiev, and Röckner] Albeverio S, Kondratiev YG, Röckner M (1998) Analysis and geometry on configuration spaces. J Funct Anal 154(2):444–500

[Atkinson and Donev(1992)] Atkinson AC, Donev AN (1992) Optimum Experimental Designs. Clarendon Press, Oxford

[Ben-Tal and Zowe(1982)] Ben-Tal A, Zowe J (1982) A unified theory of first and second order conditions for extremum problems in topological vector spaces. Math Programming Stud 19:39–76, optimality and stability in mathematical programming

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

[Daley and Vere-Jones(2008)] Daley DJ, Vere-Jones D (2008) An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure, 2nd edn. Springer, New York

[Dunford and Schwartz(1988)] Dunford N, Schwartz JT (1988) Linear Operators. 1. General Theory. Wiley

[Graf and Luschgy(2000)] Graf S, Luschgy H (2000) Foundations of Quantization for Probability Distributions, Lect. Notes Math., vol 1730. Springer, Berlin

[Gruber(2004)] Gruber PM (2004) Optimum quantization and its applications. Adv Math 186:456–497

[Hall(1985)] Hall P (1985) On the coverage of k-dimensional space by k-dimensional spheres. Ann Probab 13:991–1002

[Ho and Cao(1991)] Ho Y, Cao X (1991) Perturbation analysis of discrete event dynamic systems. Kluwer Acad. Publ., Boston

[Last(2013)] Last G (2013) Perturbation analysis of poisson processes. Bernoulli

[Margulis(1974)] Margulis GA (1974) Probabilistic characteristics of graphs with large connectivity. Problemy Peredači Informacii 10(2):101–108, in Russian

[McClure and Vitale(1975)] McClure DE, Vitale RA (1975) Polygonal approximation of plane convex bodies. J Math Anal Appl 51:326–358

[Molchanov and Tontchev(2007a)] Molchanov I, Tontchev N (2007a) Optimal approximation and quantisation. J Math Anal Appl 325:1410–1429

[Molchanov and Tontchev(2007b)] Molchanov I, Tontchev N (2007b) Optimal Poisson quantisation. Statist Prob Lett 77:1123–1132

[Molchanov and Zuyev(2000a)] Molchanov I, Zuyev S (2000a) Tangent sets in the space of measures: with applications to variational analysis. J Math Anal Appl 249:539–552

[Molchanov and Zuyev(2000b)] Molchanov I, Zuyev S (2000b) Variational analysis of functionals of Poisson processes. Math Oper Research 25(3):485–508

[Molchanov and Zuyev(2000c)] Molchanov I, Zuyev S (2000c) Variational calculus in space of measures and optimal design. In: Atkinson A, Bogacka B, Zhigljavsky A (eds) Optimum Experimental Design: Prospects for the New Millennium., Kluwer, pp 79–90

[Molchanov and Zuyev(2002)] Molchanov I, Zuyev S (2002) Steepest descent algorithms in space of measures. Statistics and Computing 12:115–123

[Molchanov and Zuyev(2004)] Molchanov I, Zuyev S (2004) Optimisation in space of measures and optimal design. ESAIM: Probability and Statistics 8:12–24

[Molchanov et al(2000)Molchanov, Chiu, and Zuyev] Molchanov I, Chiu S, Zuyev S (2000) Design of inhomogeneous materials with given structural properties. Phys Rev E 62:4544–4552

[Molchanov et al(2001)Molchanov, van Lieshaut, and Zuyev] Molchanov I, van Lieshaut M, Zuyev S (2001) Clustering methods based on variational analysis in the space of measures. Biometrika 88(4):1021–1033

[Molchanov(1997)] Molchanov IS (1997) Statistics of the Boolean Model for Practitioners and Mathematicians. Wiley, Chichester

[Molchanov and Zuyev(1997)] Molchanov IS, Zuyev SA (1997) Variational analysis of functionals of a Poisson process. Tech. Rep. RR-3302, INRIA, Sophia-Antipolis, ftp://ftp.inria.fr/INRIA/publication/RR/RR-3302.ps.gz.

[Okabe et al(2000)Okabe, Boots, Sugihara, and Chiu] Okabe A, Boots B, Sugihara K, Chiu SN (2000) Spatial Tessellations — Concepts and Applications of Voronoi Diagrams, 2nd edn. Wiley, Chichester

[Penrose and Yukich(2003)] Penrose MD, Yukich JE (2003) Weak laws in geometric probability. Ann Appl Probab 13:277–303

[Polak(1997)] Polak E (1997) Optimization: Algorithms and Consistent Approximations. Springer, New York

[Röckner(1998)] Röckner M (1998) Stochastic analysis on configuration spaces: basic ideas and recent results. In: New directions in Dirichlet forms, AMS/IP Stud. Adv. Math., vol 8, Amer. Math. Soc., Providence, RI, pp 157–231

[Russo(1981)] Russo L (1981) On the critical percolation probabilities. Z Wahrsch verw Geb 56:229–237

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

[Stoyan et al(1995)Stoyan, Kendall, and Mecke] Stoyan D, Kendall WS, Mecke J (1995) Stochastic Geometry and its Applications, 2nd edn. Wiley, Chichester

[Zuyev(1999)] Zuyev S (1999) Stopping sets: Gamma-type results and hitting properties. Adv Appl Probab 31:355–366

[Zuyev(1992)] Zuyev SA (1992) Russo's formula for the Poisson point processes and its applications. Diskretnaya matematika 4(3):149–160, (In Russian). English translation: *Discrete Math. and Applications*, **3**, 63-73, 1993