MATHEMATICS OF OPERATIONS RESEARCH
Vol. 26, No. 4, November 2001, pp. 741–768
Printed in U.S.A.

ESTIMATION OF DERIVATIVES OF NONSMOOTH PERFORMANCE MEASURES IN REGENERATIVE SYSTEMS

TITO HOMEM-DE-MELLO

We investigate the problem of estimating derivatives of expected steady-state performance measures in parametric systems. Unlike most of the existing work in the area, we allow those functions to be *nonsmooth* and study the estimation of *directional derivatives*. For the class of regenerative Markovian systems we provide conditions under which we can obtain *consistent* estimators of those directional derivatives. An example illustrates that the conditions imposed must be different from those in the differentiable case. The result also allows us to derive necessary and sufficient conditions for differentiability of the expected steady-state function. We then analyze the process formed by the *subdifferentials* of the original process, and show that the subdifferential set of the expected steady-state function can be expressed as an average of integrals of multifunctions, which is the approach commonly found in the literature for integrals of sets. The latter result can also be viewed as a limit theorem for more general compact-convex multivalued processes.

1. Introduction. In recent years a great deal of attention has been devoted to the computation of *derivatives* of performance measures in stochastic systems. The information provided by those quantities is essential to answer the important question: How much will the performance change if some parameters of the system are slightly changed? More formally, suppose we have a stochastic process, say $\{X_n(\theta)\}$, depending on a (vector-valued) parameter θ, and assume that the process converges to a steady-state $X_\infty(\theta)$. We would like to compute the gradient of the expected value of the process in equilibrium, i.e., $\nabla\mathbb{E}[X_\infty(\theta)]$. A typical example is a G/G/1 queue where the distribution of the service times depends on a parameter θ (e.g., its mean); we may be interested in computing the sensitivity of the expected waiting time $\mathbb{E}[W_\infty(\theta)]$ with respect to the parameter θ. Notice also that the computation (or estimation) of derivatives allows one to take an additional step and develop *optimization* procedures for the underlying performance measure. Such effort brings obviously numerous benefits, and in fact there have been several papers in the literature dealing with that issue. See, for instance, Chong and Ramadge (1994), L'Ecuyer and Glynn (1994), Suri and Leung (1989) and references therein for description of methods and further applications.

In general, however, closed-form expressions for the steady-state derivatives cannot be obtained, so one must resort to *simulation* methods like finite differences, perturbation analysis or likelihood ratios in order to estimate gradients (see, e.g., Glasserman 1991, Glynn 1989, L'Ecuyer 1990, Rubinstein and Shapiro 1993, Suri 1989 for discussions on that topic). In addition, it is necessary to show *consistency* of such estimators, since the steady-state performance measure of the system under scrutiny is a limiting quantity and hence so is its gradient. Extra assumptions that guarantee some type of uniform convergence, such as convexity (see Hu 1992, Shapiro and Wardi 1994, Robinson 1995), are often imposed

Received September 1, 1997; revised August 28, 1998; May 24, 1999, and March 15, 2001.
MSC 2000 subject classification. Primary: 60Gxx; secondary: 90C30.
OR/MS subject classification. Primary: Probability/stochastic model; secondary: programming/nondifferentiable.
Key words. Derivative estimation, nonsmooth optimization, regenerative processes, steady-state systems, multifunctions, convex analysis.

0364-765X/01/2604/0741/$05.00
1526-5471 electronic ISSN,

for that purpose. Further discussion on steady-state derivatives, especially on the topic of differentiability of *measures*, can be found in Pflug (1996) and Glynn and L'Ecuyer (1995).

A particularly neat situation occurs when the underlying process possesses a *regenerative* structure; i.e., it "restarts" itself at some points in time (see, for instance, Asmussen 1987). In such cases one can estimate steady-state quantities based on the behavior of the process over regenerative *cycles*, thus avoiding "warm-up" periods that are typically necessary in simulation (see Bratley, Fox and Schrage 1987). This is expressed in the formula

$$\mathbb{E}[X_\infty(\theta)] = \frac{\mathbb{E}\left[\sum_{n=0}^{\tau(\theta)-1} X_n(\theta)\right]}{\mathbb{E}\tau(\theta)},$$

where $\tau(\theta)$ is the length of a cycle. Notice however that the cycle length often depends on the parameter θ, thus making differentiation of those quantities a difficult task. In some cases this problem can be overcome by using the likelihood ratio technique (see, e.g., Rubinstein and Shapiro 1993, Glasserman and Glynn 1992). The situation is also remedied if the derivative process $\{\nabla X_n(\theta)\}$ regenerates at the same epochs as the original process; we then have that, under some additional assumptions,

$$\nabla\mathbb{E}[X_\infty(\theta)] = \frac{\mathbb{E}\left[\sum_{n=0}^{\tau(\theta)-1} \nabla X_n(\theta)\right]}{\mathbb{E}\tau(\theta)}. \tag{1.1}$$

Glasserman (1993) studies conditions under which (1.1) holds for Markov processes (in the one-dimensional case), thus providing a convenient way to estimate the desired derivatives.

Most of the work found in the literature treats the case where the expected steady-state function is *differentiable*. Indeed, in many situations this is actually the case, for instance in queueing networks in which the distributions of arrival and service times have densities (see Suri 1989, Wardi and Hu 1991). This differentiability property, however, does not hold in general; in fact, Shapiro and Wardi (1994) show by a simple example that not only can the expected steady-state function be nondifferentiable, but also the set of nondifferentiable points can be a *dense* subset of the domain. In such cases one is concerned with estimating *subgradients*, or more generally, *directional derivatives* (see §2 for definitions). An important contribution in that respect is the work by Robinson (1995): Under the assumption that the functions $X_n(\cdot, \omega)$ are convex w.p.1, Robinson shows that subgradients of the expected steady-state function can be consistently estimated; however, no claim can be made regarding the whole subdifferential set, since in that setting one can only show that limits of subgradients are *contained* in the subdifferential set of the expected steady-state function. An interesting application of nonsmooth optimization to the maximization of steady-state throughput in a tandem production line, using the techniques of Robinson (1995), can be found in Plambeck et al. (1996).

In this paper we investigate the problem of estimating directional derivatives of nonsmooth expected steady-state performance measures of Markov processes. We provide conditions under which we can obtain *consistent* estimators of those directional derivatives. The basic idea is to show that, under those conditions, the directional derivative function process $\{X_n'(\theta_0; \cdot)\}$ regenerates at the same points as the original process $\{X_n(\theta_0)\}$ and hence the directional derivatives of the expected value function can be expressed both as a long-run average and as a ratio formula. Here, regeneration plays an essential role: Without this condition, consistency is unlikely to hold, since typically pointwise convergence of functions does *not* imply convergence of directional derivatives. Notice that the situation is more restricted than in the differentiable case, where under proper assumptions—convexity is an example—convergence of derivatives follows from pointwise convergence. An example given in §3 shows that the analysis of regeneration of the directional derivative process is more complicated than in the differentiable case: We exhibit two random variables depending on θ, say $Y_1(\theta)$ and $Y_2(\theta)$, such that $Y_1(\theta)$ and $Y_2(\theta)$ are independent and identically distributed for each θ, but their subdifferential sets at $\theta = 0$ do not have the same distribution (the situation can be easily extended to regenerative processes in general). As a consequence, we give a necessary and sufficient condition for the differentiability of the expected steady-state function, thus extending a result by Shapiro and Wardi (1994).

We then rewrite the results obtained for directional derivatives in terms of subdifferential sets. Although this reformulation is an immediate consequence of the correspondence between sublinear functions and compact convex sets (cf. Hiriart-Urruty and Lemaréchal 1993a), the importance of subdifferentials in nonsmooth optimization theory, in terms of algorithms and optimality conditions (see Rockafellar 1970, Ioffe and Tihomirov 1979, Hiriart-Urruty and Lemaréchal 1993a,b) justify per se the re-statement of those results. This allows us to draw an important conclusion: The subdifferential of the expected steady-state function $\partial\mathbb{E}[X_\infty(\theta_0)]$ can be expressed both as a limiting average of sets ($N^{-1}\sum_{n=0}^{N-1}\partial X_n(\theta_0)$)—where the sum is understood as the Minkowski addition of sets $A+B=\{a+b : a\in A, b\in B\}$—and as an average of sets over a regenerative cycle ($\mathbb{E}[\sum_{n=0}^{\tau-1}\partial X_n(\theta)]/\mathbb{E}\tau$). The latter expression—which involves the integral of a set—can be computed by using the theory of integration of *multifunctions* found in the literature (see for instance Castaing and Valadier 1977, Hiai and Umegaki 1977, Rockafellar 1976). We show that this limiting result is valid in general for regenerative compact-convex multivalued processes.

The paper is organized as follows: In §2 we define the set-up for the problem and review concepts on regenerative processes, convex analysis and vector measures. In §3 we illustrate with an example some difficulties of the problem and give conditions to ensure regeneration of the directional derivative functions. In §4 we show that under some additional assumptions a ratio formula like (1.1) holds for the directional derivatives and then we exhibit consistent estimators of the derivatives of the expected steady-state function. As a consequence, we obtain a necessary and sufficient condition for the differentiability of that function. In §5 we restrict ourselves to the case of subdifferentiable functions and apply the theory of integrals of multifunctions to the results obtained in the previous sections. Section 6 presents two examples of application of the results, and in §7 we make some concluding remarks.

2. Background and basic assumptions. In this section we review some concepts that will be used in the sequel. The results presented are known, but some of them are proved here for the sake of completeness.

We start with concepts on regenerative processes. Following Asmussen (1987), we say that a vector-valued process $\{Y_n\}$ is *regenerative* if there exists a sequence of iid spacings $\{\tau_j\}$ such that, for each $j\geq 1$, the process

$$\{\tau_{j+1},\tau_{j+2},\ldots,\{Y_{\xi_j+m},\ m=0,1,\ldots\}\}$$

(where $\xi_j=\tau_1+\cdots+\tau_j$) is independent of $\tau_1,\ldots,\tau_j$ and its distribution does not depend on j. Observe that this definition does not impose that $\{Y_{\xi_j+m},\ m=0,1,\ldots\}$ be independent of $\{Y_m,\ m=0,\ldots,\xi_j-1\}$. Other definitions of regeneration exist; see, for instance, Thorisson (2000) for a detailed discussion.

The importance of regenerative processes has both theoretical and practical aspects. On the theoretical side, it can be shown that if $\mathbb{E}\tau_1<\infty$, then the process $\{Y_n\}$ has a limiting distribution defined as

$$F_\infty(y)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}P(Y_n\leq y). \tag{2.1}$$

Notice that we are following the concept of a limiting distribution in a *time-average* sense—as defined by Wolff (1989)—rather than in a pointwise sense. Typically, for the latter one to exist it is necessary to further assume other conditions, such as "τ_1 does not have a lattice distribution." Since the goal of this paper is to establish conditions for regeneration in some sense, we shall follow the time-average definition in (2.1) for simplicity. Nevertheless, all the results shown in this paper carry out to the pointwise case by imposing the extra assumptions mentioned above. We refer to Wolff (1989) for a comprehensive discussion on the different concepts of limiting distributions.

Let Y_∞ denote a random variable with the limiting distribution defined in (2.1), and let f be any measurable function. If $\mathbb{E}[\sum_{n=0}^{\tau_1-1}|f(Y_n)|]<\infty$, then we have that

$$\mathbb{E}[f(Y_\infty)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} f(Y_n)\big]}{\mathbb{E}\tau_1}. \tag{2.2}$$

Furthermore, the same quantity is given by time-averages; that is,

$$\mathbb{E}\,[f(Y_\infty)] = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(Y_n) \quad \text{w.p.1.} \tag{2.3}$$

On the practical side, formula (2.2) provides a way to estimate functions of the process in steady-state, yielding a procedure that is usually called *regenerative simulation*. Suppose we generate m cycles of sizes $\tau_1,\dots,\tau_m$. Then $w=\mathbb{E}[f(Y_\infty)]$ can be estimated by

$$\hat{w} = \frac{\sum_{i=1}^{m}\sum_{n=\xi_{i-1}}^{\xi_i-1} f(Y_n)}{\sum_{i=1}^{m}\tau_i}.$$

The variance of $\hat{w}$ can be estimated from the same sample; see for instance Shedler (1987) for details.

Let us now define formally the underlying processes and make some basic assumptions. Consider a vector-valued stochastic process $X(\theta) := \{X_n(\theta)\} := \{(X_n^1(\theta),\dots,X_n^K(\theta))\}$, $n=0,1,\dots,$ defined on a common probability space $(\Omega,\mathcal{F},P)$, and depending on an m-dimensional parameter θ belonging to some open set $\Theta\subset\mathbb{R}^m$. The choice for vector-valued rather than scalar-valued processes is driven mainly by the fact that many applications fall into that category, like queueing networks for instance. As we shall see later, the analysis is basically the same for both cases; there are, however, some exceptions, as in the discussion in the end of §3.

Since we are mostly interested in studying the derivatives at some fixed point, we shall fix from now on some $\theta_0\in\Theta$. Assume the following:

ASSUMPTION A1. $\{X_n(\theta)\}$ is a Markov process for each $\theta\in\Theta$ (possibly with continuous state space), and the process $\{X_n(\theta_0)\}$ is regenerative with regeneration epochs $\{\xi_m\}$, $m\ge 0$, and $\xi_0=0$.

Assumption A1 is the basic starting point, since our main goal is to find conditions under which the derivative process regenerates together with the original process at the points $\{\xi_m\}$. This assumption will be complemented later, as we shall see in §3. For now, only $\{X_n(\theta_0)\}$ is assumed to be regenerative; the processes $\{X_n(\theta)\}$ for $\theta\ne\theta_0$ are assumed to be just Markovian.

Before proceeding further, let us review some concepts of convex analysis which will be used in the sequel (basic references are Rockafellar 1970 and Hiriart-Urruty and Lemaréchal 1993a). Let $f:\Theta\to\mathbb{R}$ be an arbitrary function. For any $\theta\in\Theta$, the *directional derivative* of f at θ in the direction d, when it exists, is given by

$$f'(\theta;d) = \lim_{t\downarrow 0}\frac{f(\theta+td)-f(\theta)}{t}. \tag{2.4}$$

Notice that $f'(\theta;\cdot)$ is *positively homogeneous*, i.e., $f'(\theta;0)=0$ and $f'(\theta;\alpha d)=\alpha f'(\theta;d)$ for all $d\in\mathbb{R}^m$ and all $\alpha>0$.

DEFINITION. A function $f : \Theta \subset \mathbb{R}^m \to \mathbb{R}$ is said to be *directionally differentiable* at $\theta_0 \in \Theta$ if the directional derivative $f'(\theta_0; d)$ defined in (2.4) exists and is finite for all $d \in \mathbb{R}^m$, and the function $f'(\theta_0; \cdot)$ is continuous.

DEFINITION. A function $f : \Theta \subset \mathbb{R}^m \to \mathbb{R}$ is said to be *subdifferentiable* at $\theta_0 \in \Theta$ if it is directionally differentiable and, in addition, $f'(\theta_0; \cdot)$ is *convex*.

Notice that our concept of directional differentiability is stronger than the usual one found in the literature, since we assume the directional derivative function to be continuous. When f is subdifferentiable, we can define the *subdifferential* of f at θ as the set supported by the function $f'(\theta; \cdot)$, that is,

$$\partial f(\theta) = \{\xi \in \mathbb{R}^m : \langle \xi, d\rangle \le f'(\theta; d) \quad \text{for all } d \in \mathbb{R}^m\}. \tag{2.5}$$

It follows that the subdifferential $\partial f(\theta)$ is a convex and compact set. Furthermore, it follows that in this case the directional derivative $f'(\theta; d)$ is a *sublinear* function of d (we say that a function g is sublinear if $g(\alpha_1 d_1 + \alpha_2 d_2) \le \alpha_1 g(d_1) + \alpha_2 g(d_2)$ for all $d_1, d_2 \in \mathbb{R}^m$ and all $\alpha_1, \alpha_2 > 0$. It can be shown—see Proposition V.1.1.4 in Hiriart-Urruty and Lemaréchal (1993a)—that g is sublinear if and only if g is positively homogeneous and convex). Notice that when f is locally Lipschitz (as defined below) the concept of subdifferentiability includes that of *regularity* introduced by Clarke (1990), that is, regular functions are subdifferentiable. Moreover, if f is regular then the subdifferential coincides with the so-called generalized gradient (see Clarke 1990). If f is convex, those two concepts also coincide with the standard subdifferential of convex analysis (see, e.g., Rockafellar 1970).

Relation (2.5) can actually be generalized to arbitrary finite-valued sublinear functions, resulting in an interesting connection between that class of functions and compact convex sets. Given a finite-valued sublinear function $\sigma(\cdot)$ on $\mathbb{R}^m$ we can define a compact convex set $C_\sigma \subset \mathbb{R}^m$ as

$$C_\sigma = \{\xi \in \mathbb{R}^m : \langle \xi, d\rangle \le \sigma(d) \quad \text{for all } d \in \mathbb{R}^m\}.$$

Conversely, given a compact convex set $C \subset \mathbb{R}^m$ we can construct a finite sublinear function $\sigma_C(\cdot)$ on $\mathbb{R}^m$ as

$$\sigma_C(d) = \max\{\langle \xi, d\rangle : \xi \in C\}$$

for all $d \in \mathbb{R}^m$. This correspondence can be shown to be an *isometric isomorphism*, and many results can be derived from this equivalence. In §V.3 of Hiriart-Urruty and Lemaréchal (1993a), for instance, one can find a comprehensive discussion of that topic.

An important situation occurs when the directional derivative function $f'(\theta_0; \cdot)$ is *linear*, i.e., there exists a vector $D_{\theta_0} \in \mathbb{R}^m$ such that $f'(\theta_0; d) = \langle D_{\theta_0}, d\rangle$ for all $d \in \mathbb{R}^m$. In that case f is said to be *Gâteaux-differentiable* at θ_0. A stronger concept is that of Fréchet differentiability: f is said to be *Fréchet-differentiable* at θ_0 if there exists a vector $D_{\theta_0} \in \mathbb{R}^m$ such that

$$\lim_{d\to 0} |f(\theta_0 + d) - f(\theta_0) - \langle D_{\theta_0}, d\rangle| / \|d\| = 0.$$

It is easy to see that Fréchet-differentiability implies Gâteaux-differentiability. The converse is true if f is *locally Lipschitz*, i.e., if there exists a positive constant M such that

$$|f(\theta_1) - f(\theta_2)| \le M\|\theta_1 - \theta_2\|$$

for all θ_1, θ_2 in a neighborhood of θ_0. If $f = (f_1, \ldots, f_K)$ is a mapping from Θ to $\mathbb{R}^K$ we say that f is directionally (resp. Gâteaux, Fréchet) differentiable at θ_0 if each f_i is directionally (resp. Gâteaux, Fréchet) differentiable at θ_0, $i = 1, \ldots, K$. For a comprehensive discussion on these concepts, including the generalization to infinite-dimensional spaces, see Shapiro (1990) and references therein.

It is clear from the above discussion that the set of directionally differentiable functions includes the Gâteaux-differentiable functions. Another important class included in that category are the convex continuous functions. More generally, functions that can be written as a *difference* of convex continuous functions (sometimes called DC-functions in the literature) are directionally differentiable. Notice also that composite functions of the form $f(\theta) = g(A(\theta))$, where $g(\cdot)$ is convex continuous and $A : \mathbb{R}^m \to \mathbb{R}^k$ is a Gâteaux-differentiable mapping, are directionally differentiable.

The ideas discussed so far suggest that a fairly general setting is obtained if we assume that the underlying functions are *directionally differentiable*. The goal is then to estimate the *directional derivatives* of the expected value function $x_\infty(\cdot) = \mathbb{E}[X_\infty(\cdot)]$ at θ_0, by using the directional derivatives of the process $\{X_n(\theta)\}$. In §5 we specialize our results for the subdifferentiable case.

In studying estimation procedures for the directional derivatives $x'_\infty(\theta_0; d)$, it will be useful to consider the *function* $x'_\infty(\theta_0; \cdot)$ rather than specific values of d. We proceed now in defining the *directional derivative process* $\{X'_n(\theta_0; \cdot)\}$ on an appropriate space. We can now state the second assumption:

ASSUMPTION B1. For each $n = 0, 1, 2, \ldots$ and P-almost all $\omega \in \Omega$, the sample-path functions $X_n(\cdot, \omega)$ are *directionally differentiable* at θ_0.

Assumption B1 guarantees, of course, that we can actually take the directional derivatives pathwise. As pointed out before, a fairly large class of functions satisfies that property. Observe that the state space of each $(X_n^k)'(\theta_0; \cdot)$, $k = 1, \ldots, K$, is the space $H^+(\mathbb{R}^m)$ of positively homogeneous continuous functions defined on $\mathbb{R}^m$. This space is clearly isomorphic to the space $C(S^{m-1})$ of continuous functions whose domain is the unit sphere S^{m-1} of $\mathbb{R}^m$ (i.e., the set $\{d \in \mathbb{R}^m : \|d\| = 1\}$) since we can always write $f(d) = f(\|d\| d/\|d\|) = \|d\| f(d/\|d\|)$ for $d \neq 0$. Endow $C(S^{m-1})$ and $H^+(\mathbb{R}^m)$ with the sup-norm

$$\|\sigma\| = \max_{d \in S^{m-1}} |\sigma(d)|, \tag{2.6}$$

and let $\mathcal{B}_H$ denote the corresponding collection of Borel sets in $H^+(\mathbb{R}^m)$. Now, $(H^+(\mathbb{R}^m), \mathcal{B}_H)$ is a measurable space and thus, in order to show that each $(X_n^k)'(\theta_0; \cdot)$ is well defined as a random variable with range in $H^+(\mathbb{R}^m)$ we must show that $(X_n^k)'(\theta_0; \cdot)$ is $\mathcal{F} - \mathcal{B}_H$ measurable, i.e., the inverse image of a set in $\mathcal{B}_H$ belongs to the σ-field $\mathcal{F}$. The lemma below shows that this is actually the case.

LEMMA 2.1. *Let $n \in \{0, 1, \ldots\}$ and $k \in \{1, \ldots, K\}$ be arbitrary. Then, under assumption* B1, *the function $\varphi : \Omega \to H^+(\mathbb{R}^m)$, defined by $\varphi(\omega)(\cdot) = (X_n^k)'(\theta_0; \cdot, \omega)$, is $\mathcal{F} - \mathcal{B}_H$ measurable.*

PROOF. Let $\varphi_\omega(\cdot)$ denote the function $\varphi(\omega)(\cdot)$. In order to show that φ is $\mathcal{F} - \mathcal{B}_H$ measurable, it is sufficient to show that, given any function $\sigma_0 \in H^+(\mathbb{R}^m)$ and any $\varepsilon > 0$, the set

$$A_{\sigma_0} = \{\omega \in \Omega : \|\varphi_\omega - \sigma_0\| \le \varepsilon\}$$

is $\mathcal{F}$-measurable. Notice that

$$\begin{aligned} A_{\sigma_0} &= \{\omega \in \Omega : |\varphi_\omega(d) - \sigma_0(d)| \le \varepsilon, \text{ for all } d \in S^{m-1}\} \\ &= \bigcap_{d \in S^{m-1}} \{\omega \in \Omega : |\varphi_\omega(d) - \sigma_0(d)| \le \varepsilon\}. \end{aligned} \tag{2.7}$$

Observe that, for a fixed $d \in S^{m-1}$, the set $\{\omega \in \Omega : |\varphi_\omega(d) - \sigma_0(d)| \le \varepsilon\}$ is $\mathcal{F}$-measurable. To see this, notice that $\varphi_\omega(d)$ is defined by the limit of the difference of two measurable functions (see the definition (2.4) of the directional derivative) and thus $\varphi_\omega(d)$ is measurable in ω. Next, let D be a countable dense subset of S^{m-1}. Since φ_ω and σ_0 are continuous, it follows that the intersection in (2.7) can be taken over the countable set D instead of S^{m-1} and hence we conclude that A_{σ_0} is $\mathcal{F}$-measurable. $\square$

Let $H_K^+(\mathbb{R}^m)$ be the cartesian product $H_K^+(\mathbb{R}^m) = H^+(\mathbb{R}^m) \times \cdots \times H^+(\mathbb{R}^m)$ (K times). Let $\mathscr{B}_{H_K}$ denote the corresponding collection of Borel sets of $H_K^+(\mathbb{R}^m)$. Notice that, because of the isomorphism between the metric spaces $(C(S^{m-1}), \|\|)$ and $(H^+(\mathbb{R}^m), \|\|)$, it follows that $H^+(\mathbb{R}^m)$ is a separable Banach space since so is $C(S^{m-1})$ (see, e.g., Royden 1988) and hence $H_K^+(\mathbb{R}^m)$ is also a separable Banach space. Thus, it makes sense to study the regeneration (or more generally, Harris recurrence) of $\{X_n'(\theta_0; \cdot)\} := \{((X_n^1)'(\theta_0; \cdot), \ldots, (X_n^K)'(\theta_0; \cdot))\}$ as a process on $(H_K^+(\mathbb{R}^m), \mathscr{B}_{H_K})$ (see Revuz 1984 or Nummelin 1984 for details on the latter topic). Note also that $\{X_n'(\theta_0; \cdot)\}$ is a Markov process with respect to the filtrations generated by $\{X_n(\cdot)\}$.

Consider now an arbitrary random variable φ on $(H^+(\mathbb{R}^m), \mathscr{B}_H)$. As seen above, φ takes on values in the separable Banach space $H^+(\mathbb{R}^m)$, so it is clear that the standard Lebesgue integral cannot be used to compute the expected value $\mathbb{E}[\varphi]$. Nevertheless we can resort to the so-called *Bochner integral*, which is in a sense an extension of the Lebesgue integral from the real line to a Banach space. Concepts and main results about the Bochner integral can be found for instance in Diestel and Uhl (1977) or Kelley and Srinivasan (1988). For now we need the following definitions (cf. Diestel and Uhl 1977, pp. 41, 45):

DEFINITION. A function $\psi : \Omega \to H^+(\mathbb{R}^m)$ is called *simple* if there exist $f_1, \ldots, f_N \in H^+(\mathbb{R}^m)$ and $E_1, \ldots, E_N \in \mathscr{F}$ such that $\psi(\omega) = \sum_{i=1}^N f_i\, 1_{E_i}(\omega)$.

DEFINITION. A function $\varphi : \Omega \to H^+(\mathbb{R}^m)$ is called *strongly measurable* if there exists a sequence of simple functions $\{\psi_n\}$ such that $\lim_{n\to\infty} \|\varphi(\omega) - \psi_n(\omega)\| = 0$ for P-almost all ω.

DEFINITION. A strongly measurable function $\varphi : \Omega \to H^+(\mathbb{R}^m)$ is called *Bochner integrable* if $\mathbb{E}[\|\varphi\|] < \infty$.

The importance of strongly measurable integrable functions lies in the fact that for this class of functions the Bochner integral is defined in a natural way. If $\psi = \sum_{i=1}^N f_i 1_{E_i}$ is a simple function, its Bochner integral is defined as $\int_\Omega \psi(\omega)P(d\omega) = \sum_{i=1}^N f_i P(E_i)$. Now let φ be a strongly measurable integrable function and let $\{\psi_n\}$ be a sequence of simple functions converging to φ. Then $\lim_n \int \psi_n$ exists (see, e.g., p. 45 in Diestel and Uhl 1977) and hence we can define

$$\int_\Omega \varphi(\omega)P(d\omega) := \lim_{n\to\infty} \int_\Omega \psi_n(\omega)P(d\omega). \tag{2.8}$$

It is important to note that, because of the separability of $H^+(\mathbb{R}^m)$, $\mathscr{F} - \mathscr{B}_H$ *measurable functions are strongly measurable*; for a proof of this fact, see Lemma 10.18 in Kelley and Srinivasan (1988) (observe that those authors use a different definition for strong measurability; the result stated above is translated in terms of the nomenclature adopted here). Hence, by Lemma 2.1 we have that, given $\theta_0 \in \Theta$, each $(X_n^k)'(\theta_0; \cdot)$ is strongly measurable. The following assumption will then ensure that $\mathbb{E}[(X_n^k)'(\theta_0; \cdot)]$ exists:

ASSUMPTION B2. For each $n = 0, 1, 2, \ldots$, each $k = 1, \ldots, K$, the $H^+(\mathbb{R}^m)$-valued random variables $(X_n^k)'(\theta_0; \cdot)$ are Bochner integrable.

Of course, definition (2.8) does not tell us how to evaluate the resulting function $\int \varphi$ pointwise unless we compute the approximating simple functions, which in general cannot be easily done. The following lemma shows that we can actually evaluate the integral function at a point by computing the integrals pointwise.

LEMMA 2.2. *Let φ be a Bochner integrable random variable on $(H^+(\mathbb{R}^m), \mathscr{B}_H)$, and let $\Phi = \int_\Omega \varphi(\omega)\, P(d\omega)$ be the Bochner integral of φ. Then, for each $d \in \mathbb{R}^m$ we have that*

$$\Phi(d) = \int_\Omega \varphi(\omega)(d)\, P(d\omega), \tag{2.9}$$

where the integral on the right-hand side is the standard Lebesgue integral, and $\varphi(\omega)(d)$ denotes the value of $\varphi(\omega)$ at d.

PROOF. Let $(H^+(\mathbb{R}^m))^*$ denote the dual space of $H^+(\mathbb{R}^m)$, i.e., the space of all bounded linear functionals on $H^+(\mathbb{R}^m)$. Let $F \in (H^+(\mathbb{R}^m))^*$. From Lemma 10.21 in Kelley and Srinivasan (1988) we know that $F(\varphi)$ is Lebesgue integrable and

$$F(\Phi) = F\left(\int_\Omega \varphi(\omega)\, P(d\omega)\right) = \int_\Omega F(\varphi(\omega))\, P(d\omega). \tag{2.10}$$

Next, observe that because of the isomorphism between $H^+(\mathbb{R}^m)$ and $C(S^{m-1})$ we have that $(H^+(\mathbb{R}^m))^*$ is isomorphic to the space of Borel measures on S^{m-1}; see for instance p. 358 in Royden (1988). Hence, given a Borel measure ν on S^{m-1} there is a unique $F \in (H^+(\mathbb{R}^m))^*$ such that

$$F(\gamma) = \int_{S^{m-1}} \gamma\, d\nu$$

for all $\gamma \in H^+(\mathbb{R}^m)$. For each $d \in S^{m-1}$, let δ_d denote the atom measure with mass one at d, and let F_d denote the corresponding linear functional. Clearly, we have that, for any $\gamma \in H^+(\mathbb{R}^m)$,

$$F_d(\gamma) = \int_{S^{m-1}} \gamma\, d\delta_d = \gamma(d)$$

and by substituting F_d into equation (2.10) it follows that

$$\Phi(d) = F_d(\Phi) = \int_\Omega F_d(\varphi(\omega)) P(d\omega) = \int_\Omega \varphi(\omega)(d)\, P(d\omega),$$

as stated. □

Notice that the above lemma also implies that if φ is a Bochner integrable random variable on $(H^+(\mathbb{R}^m), \mathscr{B}_H)$, then each real-valued random variable $\varphi(\omega)(d)$, $d \in \mathbb{R}^m$, is (Lebesgue) integrable.

3. Local conditions for regeneration. As remarked in §1, our objective is to study the regeneration of subdifferentials (or, more generally, directional derivatives) based on the regeneration of the original process. A much simpler but fundamental question hidden under that is: Suppose $Y(\cdot)$ is a random function of a one-dimensional parameter θ. How does the distribution of $\partial Y(\theta_0)$ for some $\theta_0 \in \Theta$ relate to the distribution of the $Y(\theta)$'s? Let us consider initially the differentiable case. Suppose there is a set $A \subset \Omega$ such that $P(A) = 1$ and $Y(\cdot, \omega)$ is differentiable at θ_0 for all $\omega \in A$. Suppose that $Y(\theta_0)$ is constant, i.e., $Y(\theta_0, \omega) = y_0$ for some y_0 and all ω. Then we have that, for $\omega \in A$, $\partial Y(\theta_0) = \{Y'(\theta_0)\}$ and

$$Y'(\theta_0, \omega) = \lim_{h\to 0} \frac{Y(\theta_0 + h, \omega) - Y(\theta_0)}{h}.$$

Therefore, since convergence w.p.1 implies convergence in distribution, it follows that the distribution of $Y'(\theta_0)$ is determined by the *marginal* distribution of the process Y on a neighborhood of θ_0, that is, by the distribution of each $Y(\theta_0 + h)$ for $0 < h < \varepsilon$, where ε is any positive fixed number. We must stress here the importance of the assumption that $Y(\theta_0)$ is constant w.p.1 in the above argument; otherwise, different representations of the original process might lead to different derivatives, and in that case one must deal with the so-called *process derivatives* (see Pflug 1996 for details).

In the nondifferentiable case, however, more is needed, as can be verified from the following simple example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with $P(\omega_i) = \frac{1}{4}$, $i = 1, \ldots, 4$. Let $Y_1(\theta)$

and $Y_2(\theta)$ be random variables defined for each $\theta \in \Theta$ in the following way:

$$\begin{aligned} Y_1(\theta,\omega_1) &= |\theta|, & Y_2(\theta,\omega_1) &= \theta^+,\\ Y_1(\theta,\omega_2) &= |\theta|, & Y_2(\theta,\omega_2) &= \theta^-,\\ Y_1(\theta,\omega_3) &= 0, & Y_2(\theta,\omega_3) &= \theta^+,\\ Y_1(\theta,\omega_4) &= 0, & Y_2(\theta,\omega_4) &= \theta^-, \end{aligned}$$

where $\theta^+ = \max(\theta, 0)$, $\theta^- = \max(-\theta, 0)$. It is clear that $Y_1(\theta)$ and $Y_2(\theta)$ are independent for each θ. Notice that at $\theta_0 := 0$ we have $Y_1(\theta_0) = Y_2(\theta_0) = 0$ w.p.1 and, at $\theta \neq 0$ and $a \in \mathbb{R}$ we have that

$$P(Y_1(\theta) = a) = P(Y_2(\theta) = a) = \begin{cases} \frac{1}{2} & \text{if } a = |\theta| \text{ or } a = 0,\\ 0 & \text{otherwise,} \end{cases}$$

so $Y_1(\theta)$ and $Y_2(\theta)$ have the same distribution for any fixed θ. Furthermore, at any $\theta \neq 0$ $Y_1(\cdot)$ and $Y_2(\cdot)$ are differentiable w.p.1 and

$$P(Y_1'(\theta) = a) = P(Y_2'(\theta) = a) = \begin{cases} \frac{1}{2} & \text{if } a = \operatorname{sign}(\theta) \text{ or } a = 0,\\ 0 & \text{otherwise.} \end{cases}$$

However, at $\theta_0 = 0$ we have

$$P(\partial Y_1(0) = C) \;= \begin{cases} \frac{1}{2} & \text{if } C = [-1, 1] \text{ or } C = \{0\},\\ 0 & \text{otherwise,} \end{cases}$$

and

$$P(\partial Y_2(0) = C) \;= \begin{cases} \frac{1}{2} & \text{if } C = [-1, 0] \text{ or } C = [0, 1],\\ 0 & \text{otherwise,} \end{cases}$$

so $\partial Y_1(0)$ and $\partial Y_2(0)$ do not have the same distribution. Note that Y_1 is nondifferentiable at zero with probability $\frac{1}{2}$, whereas Y_2 is nondifferentiable at zero with probability 1.

The above example illustrates the nature of the problem. At a point θ where the functions are differentiable, the subdifferential is just a singleton and hence it is determined by any of the one-sided derivatives (i.e., the directional derivative at θ_0 in the direction $+1$ or -1). At the nondifferentiable point $\theta_0 = 0$, on the contrary, the subdifferential is given by the set whose support function is the directional derivative at 0, in this case the convex hull of left- *and* right-side derivatives, so its distribution depends on the *joint* distribution of $Y_i'(0; 1)$ and $Y_i'(0; -1)$. In summary, *knowledge of the distribution of each* $Y(\theta)$ *is not sufficient to determine the distribution of* $\partial Y(\theta_0)$. On the other hand, of course, the law of $\partial Y(\theta_0)$ can be computed if one knows the distribution of the *random function* $Y(\cdot)$, that is, all the finite-dimensional distributions $(Y(\theta_1), Y(\theta_2), \ldots, Y(\theta_l))$, where l and $\theta_1, \ldots, \theta_l$ are arbitrary.

The preceding discussion suggests that, in order to obtain conditions for regeneration of subdifferentials, one should impose conditions on the random *functions* $X_n(\cdot, \omega)$ rather than on each $X_n(\theta, \omega)$. This however appears to be quite restrictive, since regeneration of whole functions is unlikely to occur in real situations. A more flexible alternative is to study the behavior of the process $\{X_n(\cdot)\}$ on (random) neighborhoods of the fixed point θ_0. The key (although obvious) observation here is that the knowledge of a directionally differentiable function f on any neighborhood of θ_0 is sufficient to compute the directional derivative function $f'(\theta_0; \cdot)$, as can be verified directly from definition (2.4).

The lemma below establishes the connection between the distribution of the directional derivative and the distribution of the function on a *fixed* (i.e., independent of ω) neighborhood of θ_0.

LEMMA 3.1. *Let V be a neighborhood of θ_0, and let $Y(\cdot)$ be a real-valued directionally differentiable random function on Θ. Then, the distribution of $Y'(\theta_0;\cdot)$ is determined by the distribution of $Y|_V(\cdot)$ (the restriction of Y to V) on cylinder sets, that is, by the elements of the form*

$$P(Y(\theta_1)\in B_1,\ldots,Y(\theta_k)\in B_k),$$

for all $k\geq 1$, all $\theta_1,\ldots,\theta_k\in V$ and all $B_1,\ldots,B_k$ Borel sets of the real line.

PROOF. Notice first that the distribution of the random function $Y'(\theta_0;\cdot)$ is given by its finite-dimensional distributions

$$\big(Y'(\theta_0;d_1),Y'(\theta_0;d_2),\ldots,Y'(\theta_0;d_\ell)\big)$$

for all ℓ and all $d_1,\ldots,d_\ell$. Fix then $\ell, d_1,\ldots,d_\ell$ and, for each $t\geq 0$ and each $j=1,\ldots,\ell$, let $W_{d_j}(t)$ denote the random variable

$$W_{d_j}(t)=\frac{Y(\theta_0+td_j)-Y(\theta_0)}{t}.$$

By assumption, the joint distribution of $Y(\theta_0+td_1),\ldots,Y(\theta_0+td_\ell)$ and $Y(\theta_0)$ is known for t small enough. It follows that the joint distribution of

$$\big(W_{d_1}(t),\ldots,W_{d_\ell}(t)\big)$$

can be computed for t small enough. Moreover, it is clear that $W_{d_j}(t)\to Y'(\theta_0;d_j)$ w.p.1 as $t\to 0$ for each $j=1,\ldots,\ell$, and hence

$$\big(W_{d_1}(t),\ldots,W_{d_\ell}(t)\big)\xrightarrow{d}(Y'(\theta_0;d_1),\ldots,Y'(\theta_0;d_\ell))$$

(where $\xrightarrow{d}$ denotes convergence in distribution). The assertion of the lemma follows. □

Clearly, the above lemma implies that, if there exists a neighborhood V of θ_0 such that the *restricted functions* $X_n|_V(\cdot)$ regenerate, then $X_n'(\theta_0;\cdot)$ regenerates at the same epochs. We would like however to have a less strict condition, since the sample paths $X_n(\cdot,\omega)$, $\omega\in\Omega$, can be quite different from each other—so it is unlikely that they have the same properties around a fixed neighborhood. Similarly, the functions $X_n(\cdot,\omega)$ and $X_m(\cdot,\omega)$ may have different behaviors for $n\neq m$. In other words, we want to allow the neighborhood V to depend on n and ω, and still ensure some kind of regeneration. Assumption A2 below is a step in that direction.

Recall that, by Assumption A1, $\{\xi_m\}$ are regeneration points of $X_n(\theta_0)$. We now extend that assumption and state the main result.

ASSUMPTION A2. There exist K-dimensional vector-valued directionally differentiable random functions $\{F_n(\cdot)\}$, $n\geq 0$, which are identically distributed (i.e., have the same finite-dimensional distributions) and independent of the regeneration epochs $\{\xi_m\}$, and, for P-almost all $\omega\in\Omega$, neighborhoods $\{V_n(\omega)=(V_n^1(\omega),\ldots,V_n^K(\omega))\}$ of θ_0 such that, for all $m,n\geq 0$, on the event $\{\xi_m=n\}$ the restricted functions $X_n|_{V_n(\omega)}(\cdot,\omega)$ and $F_n|_{V_n(\omega)}(\cdot,\omega)$ coincide.

Verification of Assumption A2 must be done on a case-by-case basis. In §6 we shall see some typical examples of ways to check this condition.

THEOREM 3.1. *Suppose that Assumptions* A1, B1 *and* A2 *hold. Then, the process* $\{(X_n(\theta_0),\ X_n'(\theta_0;\cdot))\}$ *regenerates at the epochs* $\{\xi_m\}$, $m\geq 0$.

PROOF. Let $m \geq 0$. By Assumption A2, on the event $\{\xi_m = n\}$ we have $(X_n|_{V_n})'(\theta_0; \cdot) = (F_n|_{V_n})'(\theta_0; \cdot)$ and thus

$$X_n'(\theta_0; \cdot) = (X_n|_{V_n})'(\theta_0; \cdot) = (F_n|_{V_n})'(\theta_0; \cdot) = F_n'(\theta_0; \cdot),$$

where the above equalities are understood to hold for each ω such that $\xi_m = n$. But from Lemma 3.1 we have that $F_n'(\theta_0; \cdot) \stackrel{d}{=} F_1'(\theta_0; \cdot)$, since $F_n(\cdot) \stackrel{d}{=} F_1(\cdot)$. Now, since the functions $\{F_n(\cdot)\}$ are independent of the regeneration epochs, it follows that $(X_{\xi_m}(\theta_0), X'_{\xi_m}(\theta_0; \cdot))$ is independent of $\xi_1, \dots, \xi_{m-1}$ and independent of m. Moreover, by the Markovian Assumption A1, we have that $X'_{\xi_m+j}(\theta_0; \cdot)$ has the same distribution as $X_j'(\theta_0; \cdot)$, $j = 0, 1, \dots$. It follows that the process $\{(X_n(\theta_0), X_n'(\theta_0; \cdot))\}$ regenerates at ξ_m. $\square$

A particular situation occurs when the original process $X(\theta)$ satisfies a recursion of the type

$$X_{n+1}(\theta) = \varphi(X_n(\theta), U_n(\theta)), \tag{3.1}$$

where $\{U_n(\theta)\}$ for any fixed θ are iid random vectors independent of $X(\theta)$. In those cases we can often obtain an explicit regenerative structure. In Glasserman (1993), it is assumed that the original process $\{X_n(\theta_0)\}$ has the following structure: There are an integer $r \geq 1$, a recurrent set $B \subset \mathbb{R}^K$ ("recurrent" here means that B is visited infinitely often w.p.1), subsets $A_1, \dots, A_r$ of $\mathbb{R}^K$, and a function $h : \mathbb{R}^{rK} \to \mathbb{R}^K$ such that, if $X_n(\theta_0) \in B$ and $U_{n+i}(\theta_0) \in A_i$, $i = 1, \dots, r$, then $X_{n+r}(\theta_0) = h(U_n(\theta_0), \dots, U_{n+r}(\theta_0))$. This condition can be viewed as a more explicit version of the *splitting condition* for Harris recurrent chains. Recall that a Markov process $\{Z_n\}$ with state space $(S, \mathscr{B}_S)$ is a Harris chain if there exists a set $A \in S$ such that A is visited infinitely often (say, at times $T_1, T_2, \dots$) and there exist numbers $r > 0$, $p \in (0, 1]$ and a probability measure μ on $(S, \mathscr{B}_S)$ such that, for any set $B \in \mathscr{B}_S$,

$$P(Z_r \in B \mid Z_0 = x) \geq p\mu(B), \quad \text{for all } x \in A. \tag{3.2}$$

It is possible to show that the above condition implies that the original probability space can be extended to support 0-1 variables I_k with the following property: $P(I_k = 1 \mid Z_{T_k} = x) = p$ for $x \in A$ and, when the process reaches the set A for the kth time, it regenerates r steps later if $I_k = 1$. For details on this construction, see Thorisson (2000). It is easy to see that, in the present case, (3.2) holds for $\{X_n(\theta_0)\}$ with $p = \prod_{i=1}^r p_i$, where $p_i = P(U_1(\theta_0) \in A_i)$. Under this splitting condition, assuming that B and A_i are open and some other differentiability assumptions, Glasserman proves that the process $\{(X_n(\theta_0), X_n'(\theta_0))\}$ is regenerative. We refer again to Glasserman (1993) for details.

Notice that the assumption that B and A_i are open is actually equivalent to assuming pointwise regeneration of $X(\theta)$ on neighborhoods of θ_0, since in that case the splitting condition can be applied to $\{X_n(\theta)\}$ for θ near θ_0. By strengthening the assumption on the input functions $U_n(\cdot)$, Theorem 3.6 in Glasserman (1993) can be modified as follows:

THEOREM 3.2. *Assume* A1 *and* B1, *and suppose there is an integer* $r \geq 1$, *a recurrent set* $B \subset \mathbb{R}^K$, *subsets* $A_1, \dots, A_r$ *of* $\mathbb{R}^K$, *a Fréchet-differentiable function* $h : \mathbb{R}^{rK} \to \mathbb{R}^K$ *and neighborhoods* $\{V_n(\omega) = (V_n^1(\omega), \dots, V_n^k(\omega))\}$ *of* θ_0 *such that*

$$\begin{aligned} &\text{if } X_n(\theta_0, \omega) \in B \text{ and } U_{n+i}(\theta_0, \omega) \in A_i,\ i = 1, \dots, r, \\ &\text{then } X_{n+r}(\cdot, \omega) = h(U_n(\cdot, \omega), \dots, U_{n+r}(\cdot, \omega)) \text{ on } V_{n+r}(\omega). \end{aligned} \tag{3.3}$$

Suppose further that the input functions $\{U_n(\cdot)\}$, $n \geq 0$, *are independent and identically distributed, and pathwise directionally differentiable. Then,* $\{(X_n(\theta_0), X_n'(\theta_0; \cdot))\}$ *is regenerative.*

PROOF. Let $\{\sigma_j,\ j \geq 1\}$ be the regeneration epochs

$$\sigma_j = \inf\{q > \sigma_{j-1} : X_{q-r}(\theta_0) \in B \quad \text{and} \quad U_{q-r+i}(\theta_0) \in A_i,\ i = 1, \ldots, r\}.$$

On $\{\sigma_j = q\}$ we have that $X_q(\cdot, \omega) = h(U_{q-r}(\cdot, \omega), \ldots, U_q(\cdot, \omega))$ on each neighborhood $V_q(\omega)$. Now, let F_q be a random function from Θ to R^K defined by $F_q(\cdot) = h(U_{q-r}(\cdot), \ldots, U_q(\cdot))$. It can be shown that F is pathwise directionally differentiable, since each $U_n(\cdot)$ is directionally differentiable and h is Fréchet-differentiable (see Proposition 3.6 in Shapiro 1990). Moreover, the $F_q(\cdot)$, $q \geq 0$, are identically distributed. It follows that Assumption A2 is satisfied and hence the result follows from Theorem 3.1. □

4. Obtaining consistent derivative estimators. One of the most useful applications of regenerative processes is the estimation of steady-state quantities by ratio-type formulas like (2.2). Indeed, from a simulation standpoint, by using the regenerative structure one avoids "warm-up" periods typically necessary when computing time-averages like (2.3) and, furthermore, variances and consequently confidence intervals can be constructed by standard application of the Central Limit Theorem. In many systems, however, the regenerative cycles can be extremely long, thus making the use of ratio formulas not feasible from a practical point of view. Nevertheless, even in such cases regeneration plays an important role, namely that of ensuring the existence of a steady-state under mild assumptions. In this section we discuss these issues with respect to the directional derivatives. As we shall see, the assumptions of the previous sections, together with some regularity conditions, guarantee that the directional derivatives of the expected value function can be expressed both as ratio-type and as time-average formulas.

Suppose that the processes $\{X_n(\theta)\}$ (for all θ in a neighborhood of θ_0) and $\{X_n'(\theta_0; \cdot)\}$ are regenerative with the same regeneration epochs, and suppose that the iid cycle times $\{\tau_n\}$ are such that $\mathbb{E}\tau_1 < \infty$, so a limiting distribution exists. In order to simplify the discussion, we assume throughout this section that $K = 1$, i.e., $X_n(\theta)$ is a scalar. To ensure the existence of ratio-type results, we shall also extend Assumption B2 as follows:

ASSUMPTION B3. For all θ in a neighborhood of θ_0, the random variable $\sum_{n=0}^{\tau_1-1} X_n(\theta)$ is *integrable*. Moreover, the $H^+(\mathbb{R}^m)$-valued random variable $\sum_{n=0}^{\tau_1-1} X_n'(\theta_0; \cdot)$ is *Bochner integrable*.

Let $x_n(\theta) = \mathbb{E}[X_n(\theta)]$, $0 \leq n \leq \infty$, where $X_\infty(\theta)$ denotes the weak limit of $X_n(\theta)$. Recall that our main goal is to estimate the steady-state directional derivative $x_\infty'(\theta_0; \cdot)$ provided it exists. A direct application of the ratio formula (2.2) gives us

$$x_\infty(\theta) = \mathbb{E}[X_\infty(\theta)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} X_n(\theta)\big]}{\mathbb{E}\tau_1} \tag{4.1}$$

and

$$\mathbb{E}[\phi_{\theta_0}(\cdot)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} X_n'(\theta_0; \cdot)\big]}{\mathbb{E}\tau_1}, \tag{4.2}$$

where $\phi_{\theta_0}(\cdot)$ is the weak limit of $X_n'(\theta_0; \cdot)$, and the expected value in (4.2) is understood as the Bochner integral (see §2). Notice however that neither of the above formulas seems to be appropriate to compute $x_\infty'(\theta_0; \cdot)$: Differentiation of (4.1) would require differentiating $\mathbb{E}\tau_1$ (which despite the notation is also a function of θ), whereas in (4.2) it does not necessarily hold that $\phi_{\theta_0}(\cdot) = X_\infty'(\theta_0; \cdot)$—and even when it does, we still have to be concerned about interchanging the expectation and the differentiation operators. Indeed, it is not even clear whether X_∞ is directionally differentiable. Those difficulties do not arise from the nonsmoothness of the functions; they exist in the differentiable case as well, as pointed out by Glasserman (1993).

In order to overcome these problems, we shall impose some stronger conditions. The goal of Assumption A3 below is to ensure that the regeneration epochs are constant in a neighborhood of θ_0 with some positive probability. It has the same spirit of the splitting condition in Harris chains discussed before Theorem 3.2. Assumption A4, in turn, imposes independence of the original process between cycles. This assumption is sometimes called "classical regeneration" in the literature. Assumption B4 is used to ensure that the interchange of limits and expectations is valid.

ASSUMPTION A3. Assumption A2 holds and there exists an $\epsilon \in (0, 1)$ such that, for any $\delta \in (0, \epsilon)$, there exists a neighborhood V_δ of θ_0 such that

$$P\big(X_{\xi_k}|_{V_\delta} = F_{\xi_k}|_{V_\delta}\big) \geq 1 - \delta \quad \text{for all } k = 1, 2, \ldots . \tag{4.3}$$

ASSUMPTION A4. For each $j \geq 1$, the post-ξ_j process $\{X_{\xi_j+k}(\theta),\ \ k = 0, 1, \ldots\}$ is independent of the pre-ξ_j process.

ASSUMPTION B4. *Either* each sample-path function $X_n(\cdot, \omega)$, $n = 0, 1, 2, \ldots$, is convex for P-almost all $\omega \in \Omega$, *or* each $X_n(\cdot, \omega)$ is almost surely *locally Lipschitz* at θ_0 with constant $M_n(\omega)$; that is,

$$|X_n(\theta_1, \omega) - X_n(\theta_2, \omega)| \leq M_n(\omega)|\theta_1 - \theta_2| \quad \text{for all } \theta_1, \theta_2 \in V_n(\omega),$$

for P-almost all $\omega \in \Omega$. In the latter case, $\{M_n\}$ satisfies $\mathbb{E}[\sum_{n=0}^{\tau-1} M_n] < \infty$.

As with Assumption A2, verification of Assumption A3 may involve a thorough study of the system. However, as we shall see in §6, typically that reduces to showing that the family of functions $X_n(\cdot, \omega)$ is *equicontinuous* for all n and all ω in some set of arbitrarily high probability, which can be accomplished by exploiting the structure of the system. Assumption A4, on the other hand, is satisfied in most regenerative systems. Assumption B4 covers the *strong stochastic convexity* assumption usually found in the literature (see, e.g., Hu 1992, Robinson 1995) as well as an alternative condition in case convexity is not present. Notice that the condition $\mathbb{E}[\sum_{n=0}^{\tau-1} M_n] < \infty$ holds in particular when the M_n's are all equal and deterministic, or when $\{M_n\}$ is independent of $\{X_n\}$ and $\mathbb{E}M_n < \infty$. As we shall see later, however, there are cases where the flexibility provided by Assumption B4 is convenient.

THEOREM 4.1. *Suppose that Assumptions* B1–B4, A2–A4 *hold. Also, assume that the iid cycle times* $\{\tau_n\}$ *have finite expectation. Then, the expected value function* $x_\infty(\cdot) = \mathbb{E}[X_\infty(\cdot)]$ *is directionally differentiable at* θ_0 *and*

$$x'_\infty(\theta_0; \cdot) = \mathbb{E}[\phi_{\theta_0}(\cdot)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} X'_n(\theta_0; \cdot)\big]}{\mathbb{E}\tau_1}$$

where, as above, $\phi_{\theta_0}(\cdot)$ *is the weak limit of* $X'_n(\theta_0; \cdot)$, *and the expectation corresponds to the Bochner integral.*

PROOF. Fix $\delta \in (0, 1)$, and let V_δ be the neighborhood of θ_0 given by Assumption A3. We will show initially that, for any $\theta \in V_\delta$, $\{X_n(\theta)\}$ regenerates at a subset of the regeneration epochs of $\{X_n(\theta_0)\}$.

Consider an extension of the original probability space to support 0-1 random variables I_k such that $P(I_k = 1) = 1 - \delta$ and $X_{\xi_k}|_{V_\delta} = F_{\xi_k}|_{V_\delta}$ on $\{I_k = 1\}$. This can be accomplished by the virtue of Assumption A3 and a construction similar to that in Corollary 3.5.1 in Thorisson (2000). Next, let ν_δ be the number of trials between two successive occurrences of $I_k = 1$. Clearly, ν_δ has a geometric distribution with success probability $q = 1 - \delta > 0$.

Consider now a subset $\{\xi_{k_j}\}$ of the regeneration points such that $I_j = 1$, i.e., $X_{\xi_{k_j}}|_{V_\delta} = F_{\xi_{k_j}}|_{V_\delta}$. Clearly, the length of these new cycles is given by

$$\rho := \xi_{\nu_\delta} = \sum_{i=1}^{\nu_\delta} \tau_i$$

(we drop the subscript δ from ρ to simplify the notation). By Assumption A4 we can easily see that, for any $j \geq 1$, the event $\{\nu_\delta \leq j-1\}$ is independent of the random variable τ_j. Therefore we can apply Wald's identity (see, e.g., Chung 1974) to obtain

$$\mathbb{E}\rho = \mathbb{E}[\nu_\delta]\mathbb{E}[\tau] = \frac{\mathbb{E}\tau}{q} = \frac{\mathbb{E}\tau}{1-\delta} < \infty.$$

We conclude that, for any $\theta \in V_\delta$, the process $\{X_n(\theta)\}$ regenerates at a subset of the regeneration epochs of $\{X_n(\theta_0)\}$. Hence, for any sufficiently small $t > 0$ we have that

$$\begin{aligned}\lim_{t\downarrow 0} \frac{x_\infty(\theta_0+td) - x_\infty(\theta_0)}{t} &= \\ \lim_{t\downarrow 0} \frac{1}{t}\left(\frac{\mathbb{E}\left[\sum_{n=0}^{\rho_1-1} X_n(\theta_0+td)\right]}{\mathbb{E}\rho_1} - \frac{\mathbb{E}\left[\sum_{n=0}^{\rho_1-1} X_n(\theta_0)\right]}{\mathbb{E}\rho_1}\right) \\ = \frac{1}{\mathbb{E}\rho_1}\lim_{t\downarrow 0}\mathbb{E}\left[\sum_{n=0}^{\rho_1-1}\frac{X_n(\theta_0+td)-X_n(\theta_0)}{t}\right].\end{aligned} \tag{4.4}$$

By Assumption B4 (in case the X_n's are Lipschitz), we have that $(1/t)|X_n(\theta_0+td) - X_n(\theta_0)| \leq M_n\|d\|$. Furthermore, since $\mathbb{E}[\sum_{n=0}^{\tau-1} M_n] < \infty$ by assumption, it follows that $\mathbb{E}[\sum_{n=0}^{\rho_1-1} M_n] < \infty$ and thus, by the Dominated Convergence Theorem we have that the limit and the expectation in (4.4) can be interchanged; that is,

$$\begin{aligned}x'_\infty(\theta_0; d) &= \lim_{t\downarrow 0}\frac{x_\infty(\theta_0+td) - x_\infty(\theta_0)}{t} \\ &= \frac{1}{\mathbb{E}\rho_1}\mathbb{E}\left[\sum_{n=0}^{\rho_1-1}\lim_{t\downarrow 0}\frac{X_n(\theta_0+td)-X_n(\theta_0)}{t}\right] \\ &= \frac{\mathbb{E}\left[\sum_{n=0}^{\rho_1-1} X'_n(\theta_0; d)\right]}{\mathbb{E}\rho_1} \qquad (4.5) \\ &= \mathbb{E}[\phi_{\theta_0}(d)], \qquad (4.6)\end{aligned}$$

the last equality following from the fact that, as seen before, $\{X'_n(\theta_0; d)\}$ also regenerates at the same points as $\{X_n(\theta_0)\}$. Observe that in case $X_n(\cdot, \omega)$ is convex w.p.1, clearly $\sum_{n=0}^{\rho_1-1} X_n$ is convex as well and hence the interchange of the derivative and expectation operators follows from the Monotone Convergence Theorem (cf., Proposition 2.1 in Shapiro and Wardi 1994). Next, notice that equations (4.5) and (4.5) hold for all $d \in \mathbb{R}^m$ and hence by Lemma 2.2 we conclude that $x'_\infty(\theta_0; \cdot)$ is the Bochner integral of the corresponding functions on the right-hand sides of those equations. Finally, since $P(\nu_\delta = 1) \geq 1-\delta$ goes to one as δ goes to zero, it follows that $P(\rho = \tau) \to 1$ as $\delta \to 0$. The assertion of the theorem now follows. □

REMARK. A simple situation where Assumption A3 is satisfied occurs when the "regeneration neighborhoods" of Assumption A2 do not depend on n or ω. In that case, (4.3) holds with $\delta = 0$. Quite often, however, we cannot ensure that such a strong condition holds. Suppose for instance that $\Omega = \mathbb{R}_+$, $\theta_0 \in \mathbb{R}$ and $\tau(\theta_0) = 1$, the regeneration occurring for all $\theta \in [\theta_0 - 1/\omega, \theta_0 + 1/\omega]$, where ω has exponential distribution. Then, for any positive t, there exists no neighborhood of θ_0 such that $\{X_n(\theta_0+t)\}$ regenerates together with $\{X_n(\theta_0)\}$. Assumption A3, in turn, covers this case, thus allowing one to apply Theorem 4.1. As we shall see in §6, there are systems, for which we cannot ensure regeneration in a global neighborhood of θ_0, that satisfy Assumption A3.

COROLLARY 4.1. *Under the assumptions of Theorem* 4.1, *for any* $d \in \mathbb{R}^m$ *the random variables*

$$\frac{1}{N}\sum_{n=0}^{N-1} X_n'(\theta_0; d) \tag{4.7}$$

and

$$\frac{\sum_{m=1}^{M}\sum_{n=\xi_{m-1}}^{\xi_m-1} X_n'(\theta_0; d)}{\sum_{m=1}^{M}(\xi_m-\xi_{m-1})}\left(=\frac{\sum_{n=0}^{\xi_M-1} X_n'(\theta_0; d)}{\xi_M}\right) \tag{4.8}$$

are consistent estimators of $x_\infty'(\theta_0; d)$ (*respectively as N and M go to infinity*).

PROOF. Since $X_n'(\theta_0; \cdot)$ regenerates at the points $\{\xi_m\}$, it follows that both above expressions converge to $\mathbb{E}[\phi_{\theta_0}(d)]$. By Theorem 4.1, this quantity is equal to $x_\infty'(\theta_0; d)$. □

Corollary 4.1 provides a way to estimate the directional derivative of x_∞ at θ_0 in any given direction d. Suppose we simulate the system for M regenerative cycles, and recall that $\xi_0=0, \xi_1, \xi_2, \ldots$ are the regeneration epochs. Then (4.8) gives an estimator of $x_\infty'(\theta_0; d)$, and by using the regenerative property it is possible to estimate the *variance* of that estimator (for details, see for instance Shedler 1987). Alternatively, one can fix a run length N and use the estimator (4.7) in that case, variances can be estimated by the batch means method (see, e.g., Bratley et al. 1987).

The importance of the above result arises from the fact that, in general, (4.7) and (4.8) are *not* consistent estimators of $x_\infty'(\theta_0; d)$, because typically pointwise convergence does not imply convergence of directional derivatives. For instance, the functions $f_n(\theta)=\max(\theta-1/n, 0)$ are all differentiable at 0 and $f_n'(0)=0$; however, f_n converges (uniformly) to $f(\theta)=\max(\theta, 0)$ and f is not differentiable at 0—we have $f'(0; 1)=1$, $f'(0; -1)=0$. In this sense, assumptions of Theorem 4.1 can be viewed as sufficient conditions for convergence of directional derivatives *even in the deterministic case*.

As another consequence of Theorem 4.1, we can derive *necessary and sufficient* conditions for nondifferentiability of the steady-state function in case all X_n are *subdifferentiable*. This nondifferentiability phenomenon was observed by Shapiro and Wardi (1994). In their paper, they assume that each $X_n(\theta)$ is subdifferentiable and that $\{X_n(\theta)\}$ and $\{\partial X_n(\theta)\}$ are regenerative for each θ. They then give a sufficient condition for $\mathbb{E}[X_\infty(\theta_0)]$ not to be differentiable at some fixed point θ_0, namely, that there exists a nonsingleton convex compact set C such that

$$P(C \subset_a \partial X_k(\theta_0), \text{ for some } k \le \tau_1\ ) > 0,$$

where $\subset_a$ means inclusion up to an additive constant and, as before, τ_1 is the length of the first regenerative cycle; see Proposition 2.2 in Shapiro and Wardi (1994) for details. The present setting allows us to extend that result, as follows. Recall from §2 that $\{X_n'(\theta_0; \cdot)\}$ is a process on $(H^+(\mathbb{R}^m), \mathcal{B}_H)$. Let $A \subset H^+(\mathbb{R}^m)$ denote the set

$$A=\{\varphi \in H^+(\mathbb{R}^m) : \varphi(\cdot) \text{ is not linear}\}. \tag{4.9}$$

The next lemma shows that $P(X_n'(\theta_0; \cdot) \in A)$ is well defined.

LEMMA 4.1. *The set A defined above is a $\mathscr{B}_H$-Borel set.*

PROOF. For any $r \in \mathbb{R}^m$, let $L_r(\cdot) \in H^+(\mathbb{R}^m)$ denote the linear function $\langle r, \cdot\rangle$. Let $D \subset H^+(\mathbb{R}^m)$ denote the set

$$D = \bigcap_{n\geq 1} \bigcup_{r\in\mathbb{Q}^m} B(L_r; 1/n),$$

where $B(L_r; 1/n) \subset H^+(\mathbb{R}^m)$ is the closed ball $B(L_r; 1/n) = \{\varphi \in H^+(\mathbb{R}^m) : \|\varphi - L_r\| \leq 1/n\}$. We claim that $A^c = D$. Indeed, let $\varphi \in A^c$, i.e., $\varphi = L_x$ for some $x \in \mathbb{R}^m$. Let $n \geq 1$ be arbitrary. Since $\mathbb{Q}^m$ is dense in $\mathbb{R}^m$ there exists an $r_n \in \mathbb{Q}^m$ such that $\|r_n - x\|_2 \leq 1/n$ and therefore we have

$$\begin{aligned}
\|L_x - L_{r_n}\| &= \sup_{d\in S^{m-1}} |\langle x - r_n, d\rangle| \\
&\leq \sup_{d\in S^{m-1}} \|x - r_n\|_2 \|d\|_2 \\
&= \|x - r_n\|_2 \leq 1/n,
\end{aligned}$$

so $\varphi \in B(L_{r_n}; 1/n)$. Thus, $\varphi \in \bigcup_{r\in\mathbb{Q}^m} B(L_r; 1/n)$ for all $n \geq 1$ and hence $\varphi \in D$, so $A^c \subset D$.

Conversely, let $\gamma \in D$. Then we know that given any $n \geq 1$ there exists an $r_n \in \mathbb{Q}^m$ such that $\|\gamma - L_{r_n}\| \leq 1/n$. Fix now an arbitrary $N \geq 1$. Let d_1, d_2 be arbitrary points in S^{m-1}, and let α_1, α_2 be any real numbers. Let $n = n(\alpha_1, \alpha_2, N)$ be any integer greater than or equal to $2N(|\alpha_1| + |\alpha_2|)$. Then we have that

$$\begin{aligned}
&|\gamma(\alpha_1 d_1 + \alpha_2 d_2) - \alpha_1\gamma(d_1) - \alpha_2\gamma(d_2)| \\
&\quad = |\gamma(\alpha_1 d_1 + \alpha_2 d_2) - L_{r_n}(\alpha_1 d_1 + \alpha_2 d_2) \\
&\qquad + \alpha_1 L_{r_n}(d_1) + \alpha_2 L_{r_n}(d_2) - \alpha_1\gamma(d_1) - \alpha_2\gamma(d_2)| \\
&\quad \leq |\gamma(\alpha_1 d_1 + \alpha_2 d_2) - L_{r_n}(\alpha_1 d_1 + \alpha_2 d_2)| \\
&\qquad + |\alpha_1||L_{r_n}(d_1) - \gamma(d_1)| + |\alpha_2||L_{r_n}(d_2) - \gamma(d_2)| \\
&\quad \leq \|\alpha_1 d_1 + \alpha_2 d_2\|_2 \|\gamma - L_{r_n}\| + |\alpha_1|\|L_{r_n} - \gamma\| + |\alpha_2|\|L_{r_n} - \gamma\| \\
&\quad \leq \frac{1}{n}(\|\alpha_1 d_1 + \alpha_2 d_2\|_2 + |\alpha_1| + |\alpha_2|) \\
&\quad \leq \frac{1}{n}(2|\alpha_1| + 2|\alpha_2|) \leq \frac{1}{N}.
\end{aligned}$$

Thus, we have that $|\gamma(\alpha_1 d_1 + \alpha_2 d_2) - \alpha_1\gamma(d_1) - \alpha_2\gamma(d_2)| \leq 1/N$ for any $N \geq 1$, so we conclude that $\gamma(\alpha_1 d_1 + \alpha_2 d_2) = \alpha_1\gamma(d_1) + \alpha_2\gamma(d_2)$, i.e., γ is linear. Therefore, $D \subset A^c$ and hence $D = A^c$. Since D is clearly a $\mathscr{B}_H$-Borel set, we conclude that so are A^c and A. □

We can now state the result.

THEOREM 4.2. *Suppose the assumptions of Theorem 4.1 hold, and suppose furthermore that $X_n(\cdot)$ is subdifferentiable at $\theta_0 \in \Theta$ w.p.1 for all $n \geq 0$. Then, the steady-state function $x_\infty(\cdot) := \mathbb{E}[X_\infty(\cdot)]$ is (Gâteaux, Fréchet) differentiable at θ_0 if and only if*

$$P(\{\omega : \exists\, j,\ 0 \leq j \leq \tau_1(\omega),\ \text{s.t. } X_j(\cdot, \omega) \text{ is not differentiable at } \theta_0\}) = 0. \tag{4.10}$$

For the proof, we shall need the following lemma:

LEMMA 4.2. *Let $\varphi(\cdot)$ be a a random variable on $(H^+(\mathbb{R}^m), \mathscr{B}_H)$ such that $\varphi(\cdot)$ is sublinear, i.e., $\varphi(\alpha_1 d_1 + \alpha_2 d_2) \leq \alpha_1\varphi(d_1) + \alpha_2\varphi(d_2)$ w.p.1 for all $d_1, d_2 \in \mathbb{R}^m$ and all $\alpha_1, \alpha_2 > 0$. Then, $\mathbb{E}\varphi(\cdot)$ is linear if and only if $P(\{\omega : \varphi(\cdot)$ is linear$\}) = 0$.*

PROOF. Let A be defined as in (4.9). Suppose first that $\mathbb{E}[\varphi(\cdot)]$ is linear. Then for any $d_1, d_2 \in \mathbb{R}^m$ we have that

$$\mathbb{E}\varphi(d_1 + d_2) = \mathbb{E}\varphi(d_1) + \mathbb{E}\varphi(d_2)$$

(notice the implicit use of Lemma 2.2 here) and hence $\mathbb{E}[\varphi(d_1 + d_2) - \varphi(d_1) - \varphi(d_2)] = 0$. Since the integrand is almost always nonpositive, it follows that $\varphi(d_1 + d_2) = \varphi(d_1) + \varphi(d_2)$ w.p.1. Moreover, for any $d \in \mathbb{R}^m$ and any $\alpha > 0$ we have that $0 = \varphi(0) = \varphi(\alpha d + (-\alpha)d) = \alpha\varphi(d) + \varphi((-\alpha)d)$ and thus $\varphi((-\alpha)d) = (-\alpha)\varphi(d)$. We conclude that φ is linear w.p.1, i.e., $P(\varphi(\cdot) \in A) = 0$.

Conversely, suppose that $\varphi(\cdot)$ is linear w.p.1. Then, by linearity of the integral we have that $\mathbb{E}[\varphi(\cdot)]$ is also linear; that is, $\mathbb{E}[\varphi(\cdot)] \notin A$. □

PROOF OF THEOREM 4.2. Let A be defined as in (4.9). Note initially that $X_j(\cdot, \omega)$ is nondifferentiable at θ_0 if and only if $X_j'(\theta_0; \cdot, \omega) \in A$. By Lemma 4.1, the set in (4.10) is measurable.

Let φ be a random variable on $(H^+(\mathbb{R}^m), \mathscr{B}_H)$ defined as $\varphi(\cdot) = \sum_{n=0}^{\tau_1 - 1} X_n'(\theta_0; \cdot)$. From Theorem 4.1 we know that

$$x_\infty'(\theta_0; \cdot) = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1 - 1} X_n'(\theta_0; \cdot)\big]}{\mathbb{E}\tau_1} = \frac{\mathbb{E}\varphi(\cdot)}{\mathbb{E}\tau_1}. \tag{4.11}$$

Notice that the assumption of subdifferentiability of X_n implies that $X_n'(\theta_0; \cdot)$ is *convex* w.p.1 (see §2) and hence so is $\varphi(\cdot)$. Furthermore, since $\varphi(\cdot)$ is positively homogeneous, it follows that φ is sublinear. Let E denote the set $\{\omega : \exists\, j \in [0, \tau_1(\omega)] \text{ s.t. } X_j'(\theta_0; \cdot, \omega) \in A\}$. We shall prove now that

$$E = \{\omega : \varphi(\cdot, \omega) \in A\}. \tag{4.12}$$

Indeed, let $\omega \in E$. Then, there exists some $j \le \tau_1$ such that $X_j'(\theta_0; \cdot, \omega) \in A$ and hence $\varphi(\cdot, \omega)$ is nonlinear; otherwise, $\varphi(\cdot, \omega) - X_j'(\theta_0; \cdot, \omega)$ would be *concave*. Thus, $\omega \in \{\varphi(\cdot) \in A\}$. Conversely, suppose that $\omega \notin E$. This means that $X_j'(\theta_0; \cdot, \omega)$ is linear for all $j \le \tau_1$ and hence $\varphi(\cdot, \omega) = \sum_{n=0}^{\tau_1(\omega)-1} X_n'(\theta_0; \cdot, \omega)$ is linear, i.e., $\omega \notin \{\varphi(\cdot) \in A\}$. Therefore, (4.12) holds.

Finally, from (4.11) we see immediately that $x_\infty'(\theta_0; \cdot)$ is linear (i.e., $x_\infty(\cdot)$ is Gâteaux-differentiable at θ_0) if and only if $\mathbb{E}[\varphi(\cdot)]$ is linear. Furthermore, as seen in the proof of Theorem 4.1, the properties of Assumption B4 carry out to $x_\infty(\cdot)$; hence Gâteaux-differentiability of $x_\infty(\cdot)$ is equivalent to Fréchet differentiability. The assertion of the theorem follows now from (4.12) and Lemma 4.2. □

5. The subdifferential process. The results presented in the previous sections are based on the assumption of directional differentiability of the functions $X_n(\cdot)$. As we have seen, this includes a fairly general class of functions. Suppose now that we restrict ourselves to *subdifferentiable* processes, that is, directionally differentiable processes whose directional derivative functions are convex (and hence sublinear). In dealing with optimization algorithms, one often encounters the problem of computing the subdifferential set (or at least obtaining an approximation by using some computed subgradients) in order to check optimality conditions or generate a next iterate. In this sense, it is important to formulate the results discussed in the previous sections in terms of the subdifferentials $\partial X_n(\theta_0)$, $n \ge 0$. The basic property that makes this possible is the one-to-one correspondence between compact convex sets and finite sublinear functions mentioned in §2—every finite sublinear function is the support function of a unique compact convex set. In particular, the subdifferential $\partial X_n(\theta_0)$ (which is compact and convex) corresponds to the directional derivative function $X_n'(\theta_0; \cdot)$.

Consider the space $\mathscr{C}$ of all compact convex subsets of $\mathbb{R}^m$, endowed with the Hausdorff metric Δ_H. It is known (see Debreu 1966) that $(\mathscr{C}, \Delta_H)$ is a complete and separable metric space, so in principle we can construct the corresponding collection of Borel sets $\mathscr{B}_{\mathscr{C}}$ and hence, assuming measurability, we can consider $\{\partial X_n(\theta_0)\}$ as a process on $(\mathscr{C}, \mathscr{B}_{\mathscr{C}})$. Notice that this measurability assumption follows immediately from Assumption B1 and Lemma 2.1. Indeed, the aforementioned correspondence between $\mathscr{C}$ and the space $\mathscr{S} \subset H^+(\mathbb{R}^m)$ of finite sublinear functions implies that $\partial X_n(\theta_0)$ is $\mathscr{F} - \mathscr{B}_{\mathscr{C}}$ measurable if and only if $X_n'(\theta_0; \cdot)$ is $\mathscr{F} - \mathscr{B}_{\mathscr{S}}$ measurable, where $\mathscr{B}_{\mathscr{S}}$ are the Borel sets of $\mathscr{S}$. But the sets of $\mathscr{B}_{\mathscr{S}}$ are of the form $B \cap \mathscr{S}$ with $B \in \mathscr{B}_H$. Since $X_n'(\theta_0; \cdot)$ is $\mathscr{F} - \mathscr{B}_H$ measurable by Lemma 2.1 and $X_n'(\theta_0; \cdot) \in \mathscr{S}$, the conclusion follows.

All results from §3 can be easily reformulated into this new setting. Notice however that, in principle, we cannot apply directly the theory of the Bochner integral to compute the expected value of a random variable on $(\mathscr{C}, \mathscr{B}_{\mathscr{C}})$ as done in §2, since that theory is constructed for *Banach* spaces—and $\mathscr{C}$ does not have a linear structure. Nevertheless, as pointed out by Hiai and Umegaki (1977) (see also Radström 1952), $\mathscr{C}$ can be embedded as a convex cone in a Banach space, and this suffices to allow the theory of Bochner integration to be applied to $\mathscr{C}$-valued functions.

As with the space $(H^+(\mathbb{R}^m), \mathscr{B}_H)$, we say that a function $\Upsilon : \Omega \to \mathscr{C}$ is *simple* if there exist sets $C_1, \ldots, C_N \in \mathscr{C}$ and $E_1, \ldots, E_N \in \mathscr{F}$ such that $\Upsilon(\omega) = \sum_{i=1}^N C_i 1_{E_i}(\omega)$. Also, we say that a function $\Gamma : \Omega \to \mathscr{C}$ is *strongly measurable* if there exists a sequence of simple functions $\{\Upsilon_n\}$ such that $\lim_{n\to\infty} \Delta_H(\Gamma(\omega), \Upsilon_n(\omega)) = 0$ for P-almost all ω. Analogously to the case of positive homogeneous functions studied in §2, it follows from the separability of $\mathscr{C}$ that a random variable Γ on $(\mathscr{C}, \mathscr{B}_{\mathscr{C}})$ is strongly measurable. Then let $\Upsilon_1, \Upsilon_2, \ldots$ (with $\Upsilon_n = \sum_{i=1}^{N_n} C_i^n 1_{E_i^n}$) be $\mathscr{C}$-valued simple functions such that, for P-almost all $\omega \in \Omega$,

$$\Gamma(\omega) = \lim_{n\to\infty} \Upsilon_n(\omega) = \lim_{n\to\infty} \sum_{i=1}^{N_n} C_i^n 1_{E_i^n}(\omega), \tag{5.1}$$

where the limit is understood to be with respect to the Hausdorff metric. We define the Bochner integral of Γ as

$$\int \Gamma \, dP := \lim_{n\to\infty} \int \Upsilon_n \, dP = \lim_{n\to\infty} \sum_{i=1}^{N_n} C_i^n P(E_i^n).$$

It is interesting to notice that the integral defined above for a "random set" Γ differs from the standard construction of integrals of *multifunctions* found in the literature (see e.g., Castaing and Valadier 1977, Hiai and Umegaki 1977, Debreu 1966, Rockafellar 1976). In that approach, Γ is viewed as a multifunction $\Omega \to \mathbb{R}^m$ (we shall consider here only closed-valued multifunctions); Γ is said to be *measurable* if for each closed subset B of $\mathbb{R}^m$, the set

$$\Gamma^{-1}(B) := \{\omega \in \Omega : \Gamma(\omega) \cap B \neq \varnothing\}$$

belongs to $\mathscr{F}$. A *measurable selection* of Γ is a measurable function $v : \Omega \to \mathbb{R}^m$ such that $v(\omega) \in \Gamma(\omega)$ for all $\omega \in \Omega$. It can be shown that a multifunction Γ is measurable if and only if there is a countable family $\{v_i : i \in I\}$ of measurable selections of Γ such that

$$\Gamma(\omega) = \mathrm{cl}\{v_i(\omega) : i \in I\}$$

for all $\omega \in \Omega$. The integral of Γ is defined as the set of integrals of (integrable) measurable selections of Γ, that is,

$$\int \Gamma \, dP := \left\{ \int v(\omega) \, P(d\omega) : v \in L_1(P), \ v \text{ is a measurable selection of } \Gamma \right\}, \tag{5.2}$$

where $L_1(P)$ is the set of integrable functions with respect to the measure P. Notice that definition (5.2) is general in that no assumptions on compactness or convexity of Γ are imposed.

A third way to view the integral of a random compact convex set Γ is by using the correspondence between that class of sets and sublinear functions as discussed in §2. Indeed, let $\gamma(\omega)(\cdot) \in H^+(\mathbb{R}^m)$ be the support function of $\Gamma(\omega)$ (so γ is a random variable on $(H^+(\mathbb{R}^m), \mathscr{B}_H))$, and suppose that γ is Bochner integrable. Let $\lambda = \int_\Omega \gamma(\omega)P(d\omega)$ be the Bochner integral of γ. Then, $\lambda(\cdot)$ is finite-valued and sublinear, so there exists a compact convex set Λ whose support function is λ. We can then take Λ to be the integral of Γ. Observe that, by Lemma 2.2, the computation of the Bochner integral $\int \gamma(\omega)(\cdot)P(d\omega)$ reduces to the computation of the Lebesgue integrals $\int \gamma(\omega)(d)P(d\omega)$. A particular advantage of this approach in the present case is that, since the support function of the subdifferential $\partial X_n(\theta_0)$ is the directional derivative function $X_n'(\theta_0; \cdot)$, we can easily reinterpret the results of the previous sections.

A natural question of course is: Do the integrals discussed above coincide? As it turns out, the answer is affirmative. The equivalence between Bochner and multifunction integrals is studied in Theorem 4.5 in Hiai and Umegaki (1977) and §6.5 in Debreu (1966), whereas the correspondence between multifunction integrals and integrals of support functions can be found in Theorem V-14 in Castaing and Valadier (1977) and Theorem 2.2 in Hiai and Umegaki (1977). It must be stressed that those equivalences hold for compact-convex-valued multifunctions (which is the present case), since otherwise the Bochner integral is not well defined and the correspondence with support functions does not hold.

We now re-state the results of §4 in terms of subdifferentials. First we derive a result that, although well known (see Rockafellar and Wets 1982, Ioffe and Tihomirov 1969), illustrates an application of the equivalence between integrals discussed above:

PROPOSITION 5.1. *Suppose that, for P-almost $\omega \in \Omega$, each $X_n(\cdot, \omega)$ is subdifferentiable at θ_0. Suppose also that Assumptions* B2 *and* B4 *hold. Then, for any finite $n \geq 0$,*

$$\partial\mathbb{E}[X_n(\theta_0)] = \mathbb{E}[\partial X_n(\theta_0)], \tag{5.3}$$

where the expected value is understood as the multifunction integral defined in (5.2).

PROOF. Let $x_n(\theta) = \mathbb{E}[X_n(\theta)]$. By the equivalence of multifunction integrals with integrals of supporting functions, (5.3) holds if and only if $x_n'(\theta_0; \cdot) = \mathbb{E}[X_n'(\theta_0; \cdot)]$, i.e.,

$$x_n'(\theta_0; d) = \mathbb{E}[X_n'(\theta_0; d)] \quad \text{for all } d \in \mathbb{R}^m.$$

The above equation holds if we can interchange integrals and limits, which as seen in the proof of Theorem 4.1 follows from Assumption B4. □

Another interesting consequence comes from Lemma 4.2:

PROPOSITION 5.2. *Suppose that the assumptions of Proposition* 5.1 *hold. Then, for any finite $n \geq 0$, $\mathbb{E}[X_n(\cdot)]$ is differentiable at θ_0 if and only if $X_n(\cdot)$ is differentiable at θ_0 with probability one.*

PROOF. First notice that, under the assumptions of the proposition, $\mathbb{E}[X_n(\cdot)]$ is differentiable at θ_0 if and only if $\partial\mathbb{E}[X_n(\theta_0)]$ is a singleton, i.e., if the directional derivative of $\mathbb{E}[X_n(\cdot)]$ at θ_0 is a linear function. By Lemma 4.2, this occurs if and only if each $X_n'(\theta_0; \cdot)$ is linear w.p.1, that is, if and only if $X_n(\cdot)$ is differentiable at θ_0 w.p.1. □

The following is an immediate consequence of Theorem 4.1 and Corollary 4.1.

PROPOSITION 5.3. *Suppose that the assumptions of Theorem* 4.1 *hold. Then, the following equations hold*:

$$\partial \mathbb{E}[X_\infty(\theta_0)] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \partial X_n(\theta_0) \quad w.p.1, \tag{5.4}$$

$$\partial \mathbb{E}[X_\infty(\theta_0)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} \partial X_n(\theta_0)\big]}{\mathbb{E}\tau_1}, \tag{5.5}$$

where the sums on the right-hand sides are understood as Minkowski addition of sets, the expected-values are again understood as the multifunction integrals defined in (5.2), *and the limit in* (5.4) *is understood in the Hausdorff sense.*

As with the directional derivatives, equation (5.4) provides a way to estimate the subdifferential of $\mathbb{E}[X_\infty]$ at θ_0. Again, suppose we simulate the system for M regenerative cycles, and let $\xi_0 = 0, \xi_1, \xi_2, \ldots$ be the regeneration epochs. An estimator for $\partial \mathbb{E}[X_\infty(\theta_0)]$ is given by

$$\widehat{\Gamma}_{\theta_0} = \frac{\sum_{m=1}^{M} \sum_{n=\xi_{m-1}}^{\xi_m - 1} \partial X_n(\theta_0)}{\sum_{m=1}^{M} (\xi_m - \xi_{m-1})}.$$

Clearly, from the above formula we see that the choice of any particular subgradient of each set $\partial X_n(\theta_0)$ yields an estimator for a subgradient of $\mathbb{E}[X_\infty]$ at θ_0. Furthermore, as remarked before, we can also estimate the variance of the resulting estimators.

Observe that Proposition 5.3 can actually be stated in terms of a general $\mathscr{C}$-valued process $\{C_n\}$. Let $\sigma_n(\cdot)$ denote the support function of C_n. As seen in §2, $\sigma_n(\cdot)$ is a finite-valued sublinear (i.e., convex and positively homogeneous) function. Moreover, it is easy to see that $\sigma_n(d) = \sigma_n'(0; d)$ for all $d \in \mathbb{R}^m$, whence we have that $C_n = \partial \sigma_n(0)$. It follows that if the process $\{\sigma_n(\cdot)\}$ satisfies assumptions A3–A4 and B1–B4 discussed in the previous sections (for θ_0=0), then $\{C_n\}$ is regenerative and hence we can apply Proposition 5.3. Notice that Assumptions B2 and B3 are equivalent to assuming Bochner integrability of $\sigma_n(\cdot)$ (and therefore of C_n), and that Assumptions B1 and B4 are automatically satisfied in this case since $\sigma_n(\cdot)$ is convex. Below we summarize this result.

COROLLARY 5.1. *Let $\{C_n\}$ be a process on $(\mathscr{C}, \mathscr{B}_C)$, and let $\sigma_n(\cdot)$ denote the support function of C_n. Suppose that $\{\sigma_n(\cdot)\}$ satisfies Assumptions* A3 *and* A4 *for θ_0=0. Suppose also that the cycle times $\{\tau_n\}$ have finite expectation and that $\sum_{n=0}^{\tau_1-1} C_n$ is* (*Bochner*) *integrable. Then, $\{C_n\}$ is regenerative, has a weak limit C_∞, and the following equations hold*:

$$\mathbb{E}[C_\infty] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} C_n \quad w.p.1, \tag{5.6}$$

$$\mathbb{E}[C_\infty] = \frac{\mathbb{E}\big[\sum_{n=0}^{\tau_1-1} C_n\big]}{\mathbb{E}\tau_1}. \tag{5.7}$$

The above corollary can also be viewed as a result for compact-convex-valued multifunction processes. Other types of limiting results have been studied in the literature for multivalued processes: in more general settings (i.e., without imposing convexity and compactness) Hiai (1984) gives various forms of the Strong Law of Large Numbers when the C_n are independent, whereas Hiai and Umegaki (1977) study martingales formed by multivalued processes and provide some convergence theorems. We refer to those papers and references therein for details. To the best of our knowledge, however, there have been no results on *regenerative* multivalued processes as given by Corollary 5.1.

6. Examples.

(1) *A G/D/1 queue.* The following system was presented by Shapiro and Wardi (1994) to illustrate the nondifferentiability of the mean steady-state function. Consider a G/D/1 queue where the distribution of the interarrival times A_n has atoms at two points b and c with $b < c$. For simplicity, we take $P(A_n = b) = P(A_n = c) = \frac{1}{2}$. Suppose that the deterministic service time is a parameter $\theta \in \Theta = (b, (b+c)/2)$. Notice that the assumption $\theta < (b+c)/2$ guarantees that the queue is stable and hence regenerative.

Denote by $T_n(\theta)$ the system time of customer n (i.e., waiting time plus service time). Then, for $n \geq 1$, $T_n(\theta)$ satisfies the recursion

$$T_n(\theta) = \theta + [T_{n-1}(\theta) - A_n]^+,$$

where $[x]^+ = \max\{x, 0\}$ (assume $T_0(\cdot) \equiv 0$). Notice that $T_n(\cdot)$ is defined by the maximum of linear functions and therefore it is *convex*. Given $\theta_0 \in \Theta$, define random variables $\{\sigma_m\}$, $m = 0, 1, 2, \ldots$, with $\sigma_0 = 0$ and

$$\sigma_m = \inf\{n > \sigma_{m-1} : T_{n-1}(\theta_0) \leq A_n\}.$$

Observe that $\{\sigma_m\}$, $m = 0, 1, 2, \ldots$ are the epochs at which an arriving customer finds the queue empty; hence $\{T_n(\theta_0)\}$ regenerates at those points. Notice also that on the event $\{T_{n-1}(\theta_0) \leq A_n\}$ we have that $T_n(\theta_0) = \theta_0$. Shapiro and Wardi (1994) show that on $\{T_{n-1}(\theta_0) = A_n\}$ the function $T_n(\cdot)$ is not differentiable at θ_0 and, furthermore, the event $\{T_{n-1}(\theta_0) = A_n,$ for some $n \leq \sigma_1\}$ has positive probability, which is their basic condition for nondifferentiability of the mean steady-state function $\mathbb{E}[T_\infty(\cdot)]$ at θ_0.

Consider now the event $E = \{\sigma_m = n\}$, and observe that we can partition E as $E_< \cup E_=$, where $E_< = \{T_{n-1}(\theta_0) < A_n\}$ and $E_= = \{T_{n-1}(\theta_0) = A_n\}$. Notice that the continuity of $T_n(\cdot)$ implies that for any $\omega \in E_<$ there is a neighborhood $V_n(\omega)$ of θ_0 such that $T_{n-1}(\theta) < A_n$ for all $\theta \in V_n(\omega)$ and hence $T_n(\theta) = \theta$ on that neighborhood. On the other hand, on $E_=$ it might happen that $T_{n-1}(\theta) > A_n$ for θ arbitrarily close to θ_0 and hence $T_n(\theta) = \theta + T_{n-1}(\theta) - A_n$. Since the event $\{T_{n-1}(\theta_0) = A_n,$ for some $n \geq 1\}$ happens at some σ_k with probability one, it follows that we cannot ensure the existence of neighborhoods $\{V_n(\omega)\}$ satisfying Assumption A2 and thus we cannot apply Theorem 3.1 directly.

We can nevertheless overcome that problem by considering a *subset* of the regeneration points as follows. Let $\varepsilon > 0$ be such that $\rho(A_n - \varepsilon \leq T_{n-1}(\theta_0) < A_n) = 0$, which exists since A_n takes on a finite number of values. Let $\{\rho_m\}$, $m = 0, 1, 2, \ldots$ (with $\rho_0 = 0$) denote the epochs

$$\rho_m = \inf\{n > \rho_{m-1} : T_{n-1}(\theta_0) < A_n - \varepsilon\}.$$

As argued before, by continuity of the functions T_n there exist neighborhoods $V_n(\omega)$ of θ_0, $\omega \in \Omega$, such that on the event $F = \{\rho_m = n\}$ we have that $T_n(\theta) = \theta$ on $V_n(\omega)$. Therefore, Assumption A2 is satisfied and hence from Theorem 3.1 it follows that $(T_n(\theta_0), \partial T_n(\theta_0))$ regenerates at each ρ_m.

Let us show now that Assumption A3 also holds in this case. Note initially that, since $\mathbb{E}\rho_1 < \infty$, it follows that for any $\delta \in (0, 1)$ there exists an N such that $P(\rho_1 > N) < \delta$. Now, on $\{\rho_m = n\}$ we can write

$$T_{n-1}(\theta_0) = (n - 1 - \rho_{m-1})\theta_0 - \sum_{k=\rho_{m-1}}^{n-1} A_k < A_n - \varepsilon.$$

Define the set

$$V := \left\{\theta \in \Theta : |\theta - \theta_0| < \frac{\varepsilon}{N}\right\}.$$

Consider now a point $\theta \in V$. Since $\theta \in \Theta$, it follows that the system with service times θ is stable and therefore it regenerates at epochs, say, $\tilde{\rho}_1, \tilde{\rho}_2, \ldots$. For each $m = 0, 1, \ldots$, let B_m denote the event

$$B_m := \{\text{there exists a } \tilde{\rho}_k \text{ such that } \rho_{m-1} \le \tilde{\rho}_k \le \rho_m\}.$$

It follows that, for any $\theta \in V$, on $\{\rho_m = n\} \cap B_m$ we can write

$$\begin{aligned} T_{n-1}(\theta) - T_{n-1}(\theta_0) &= (n-1)(\theta - \theta_0) + \rho_{m-1}\theta_0 - \tilde{\rho}_k\theta + \sum_{j=\rho_{m-1}}^{\tilde{\rho}_k - 1} A_j \\ &= (n-1-\tilde{\rho}_k)(\theta - \theta_0) - \left[(\tilde{\rho}_k - \rho_{m-1})\theta_0 - \sum_{j=\rho_{m-1}}^{\tilde{\rho}_k - 1} A_j\right] \\ &\le (n-1-\tilde{\rho}_k)(\theta - \theta_0) \\ &\le (\rho_m - \rho_{m-1})(\theta - \theta_0), \end{aligned} \tag{6.1}$$

where the next-to-last inequality follows from the fact that $\sum_{j=\rho_{m-1}}^{\ell-1} A_j \le (\ell - \rho_{m-1})\theta_0$ for all $\ell \in (\rho_{m-1}, \rho_m]$. Therefore we have that

$$|T_{n-1}(\theta) - T_{n-1}(\theta_0)| \le (\rho_m - \rho_{m-1})\frac{\varepsilon}{N}$$

and thus, on $\{\rho_m = n\} \cap B_m \cap \{\rho_m - \rho_{m-1} \le N\}$ we have

$$|T_{n-1}(\theta) - T_{n-1}(\theta_0)| \le \varepsilon.$$

It follows that there exists some $\tilde{\delta} \in (0, 1)$ such that $\tilde{\delta} \to 0$ as $\delta \to 0$ and

$$\begin{aligned} &P\big(\big|T_{\rho_m - 1}(\theta) - T_{\rho_m - 1}(\theta_0)\big| \le \varepsilon\big) \\ &\quad\ge P\big(\big|T_{\rho_m - 1}(\theta) - T_{\rho_m - 1}(\theta_0)\big| \le \varepsilon \mid B_m, \rho_m - \rho_{m-1} \le N\big)\ P\big(B_m, \rho_m - \rho_{m-1} \le N\big) \\ &\quad\ge 1 - \tilde{\delta} \end{aligned}$$

for any $\theta \in V$. Notice that the condition $|T_{\rho_m - 1}(\theta) - T_{\rho_m - 1}(\theta_0)| < \varepsilon$ implies that $T_{\rho_m - 1}(\theta) < A_j$, which in turn implies that $T_{\rho_m}(\theta) = \theta$. Therefore, Assumption A3 holds. Incidentally, notice that (6.1) illustrates a case where Assumption B4 holds with variable Lipschitz constants M_n—in this case, $M_n = n$—provided that $\mathbb{E}\rho_1^2 < \infty$.

Notice that $T_{\rho_m}(\cdot)$ is differentiable at θ_0 for all $m \ge 0$—indeed, $T'_{\rho_m}(\theta_0) = 1$—but for the indices n such that $\{T_{n-1}(\theta_0) = A_n\}$ we have that

$$\partial T_n(\theta_0) = \operatorname{conv}\big(\{1\} \cup \{1\} + \partial T_{n-1}(\theta_0)\big) \tag{6.2}$$

(where $\operatorname{conv}(A)$ denotes the convex hull of A). For all other indices j such that $0 < j < \rho_1$, we have

$$\partial T_j(\theta_0) = \begin{cases} \{1\} + \partial T_{j-1}(\theta_0) & \text{if } T_{j-1}(\theta_0) > A_j, \\ \{1\} & \text{if } T_{j-1}(\theta_0) < A_j. \end{cases} \tag{6.3}$$

Observe that with positive probability (i.e., the probability that the event $\{T_{n-1}(\theta_0) = A_n\}$ occurs before $\{T_{n-1}(\theta_0) < A_n\}$) there exists $n < \rho_1$ such that $T_n(\cdot)$ is nondifferentiable at θ_0. Thus, we see that the sum of the sets $\partial T_i(\theta_0)$ for i within a cycle will be a nonsingleton with positive probability, and hence by Theorem 4.2 it follows $\mathbb{E}[T_\infty(\cdot)]$ is not differentiable at θ_0.

We now check Assumptions B2 and B3. It is not difficult to see from (6.2) and (6.3) that $\|\partial T_j(\theta_0)\| \le j$, for all $0 < j < \rho_1$. It follows immediately that $\partial T_j(\theta_0)$ is Bochner integrable if $\mathbb{E}\rho_1 < \infty$, and $\sum_{n=0}^{\rho_1-1} \partial T_n(\theta_0)$ is Bochner integrable if $\mathbb{E}\rho_1^2 < \infty$. Assuming these conditions hold, by applying Proposition 5.3 we obtain

$$\partial\mathbb{E}\big[T_\infty(\theta_0)\big] = \frac{\mathbb{E}\big[\sum_{n=0}^{\rho_1-1} \partial T_n(\theta_0)\big]}{\mathbb{E}\rho_1}.$$

A consistent estimator of $\partial\mathbb{E}[T_\infty(\theta_0)]$ can now be computed as follows. Fix some $M > 0$, and simulate the system for M cycles. Let $\rho_0, \dots, \rho_M$ be the regeneration points (without loss of generality, assume we start with an empty system, i.e., $\rho_0 = 0$). Then,

$$\frac{\sum_{i=1}^{M} \sum_{j=\rho_{i-1}}^{\rho_i - 1} \partial T_j(\theta_0)}{\sum_{i=1}^{M} (\rho_i - \rho_{i-1})}$$

is an estimator of $\partial\mathbb{E}[T_\infty(\theta_0)]$, where the $\partial T_j(\theta_0)$ are given by (6.2) and (6.3), and $\partial T_0(\theta_0) = \partial T_{\rho_i}(\theta_0) = \{1\}$. Another consistent estimator is obtained by simulating the system for N periods and computing

$$\frac{1}{N} \sum_{i=0}^{N-1} \partial T_i(\theta_0).$$

(2) *A tandem queue model.* Consider a series of K single-server queues, with general distributions for the inter-arrival times A_n and the service times $\{S_n^k(\theta)\}$, $n = 0, 1, \dots,$ $k = 1, \dots, K$. Suppose that all service time vectors $S_n(\theta) := (S_n^1(\theta), \dots, S_n^K(\theta))$ are iid, and that the network is stable. Such system has been studied in the literature in the case the $S_n(\cdot)$ are differentiable: Some papers deal with optimization issues and consequently the computation of derivatives of steady-state quantities, often using perturbation analysis techniques (see, for instance, Wardi and Hu 1991, Hu 1992, Chong and Ramadge 1994), whereas Glasserman (1993) shows that the waiting times $W_n(\theta)$ and the derivative process $W_n'(\theta)$ regenerate at the same epochs. Here, we do not assume differentiability; instead we assume that $S_n^k(\theta)$ is only a *subdifferentiable* function of θ for each n and each k, and that the functions $S_n(\cdot) := (S_n^1(\cdot), \dots, S_n^K(\cdot))$ are iid. We also assume that the service times satisfy the following property: For any $\delta \in (0, 1)$ there exists a constant $M_\delta > 0$ such that, for any θ in a neighborhood of θ_0,

$$P\big(|S_n^k(\theta) - S_n^k(\theta_0)| \le M_\delta \,\|\theta - \theta_0\|\big) > 1 - \delta \tag{6.4}$$

for all n and k. This condition is satisfied for instance if $S_n^k(\theta)$ has the form $\max(Y_n^k\theta, Z_n^k)$, where Y_n^k and Z_n^k are random variables with finite expectation.

Let $T_n^k(\theta)$ denote the *system time* of job n upon its completion on server k. It is clear that one set of regeneration points for the process $T_n(\theta) := (T_n^1(\theta), \dots, T_n^K(\theta))$ consists of the epochs at which an arriving customer finds the *whole network* empty. However, as pointed out by Nummelin (1981), those points may never occur, even though the system is stable. Alternatively, with mild assumptions, Nummelin provides regeneration epochs for the waiting time process $W_n^k(\theta)$ and shows that the corresponding regeneration cycles have expected finite length. By using an argument similar to Nummelin (1981), it can be shown that under proper assumptions the system time process $\{T_n(\theta)\}$ is a regenerative Markov chain with regeneration points given by

$$\sigma_j(\theta) = \inf\{n > \sigma_{j-1}(\theta) : T_{n-1}^k - A_n < b^k(\theta) - \varepsilon,\ T_n^{k-1} > b^k(\theta) + \varepsilon,\ k = 1, \dots, K\},$$

where $b(\theta) = (b^1(\theta), \dots, b^K(\theta))$ is a deterministic continuous function of θ, and ε is an arbitrary positive number. Notice that on $\{\sigma_j(\theta) = n\}$ the total waiting time of job n is zero and hence we have that

$$T_n^k(\theta) = S_n^1(\theta) + \cdots + S_n^k(\theta), \qquad k = 1, \ldots, K.$$

For $\theta_0 \in \Theta$, let $B_0 \subset \mathbb{R}^K \times \mathbb{R}$ denote the set

$$B_0 = \{(x, a) : x^k - a < b^k(\theta_0) - \varepsilon, k = 1, \ldots, K\}.$$

and let $D_0 \subset \mathbb{R}^K$ denote the set

$$D_0 = \{y : y^1 + \cdots + y^{k-1} > b^k(\theta_0) + \varepsilon, k = 1, \ldots, K\}.$$

It is easy to see that on $\{(T_n(\theta_0), A_{n+1}) \in B_0\} \cap \{S_{n+1}(\theta_0) \in D_0\}$ we have that $T_n^k(\theta) - A_{n+1} < b^k(\theta) - \varepsilon$, $T_{n+1}^{k-1} > b^k(\theta) + \varepsilon$ for all $k = 1, \ldots, K$ and all θ in some neighborhood $V_n(\omega)$ of θ_0, so it follows that

$$T_{n+1}^k(\cdot) = S_{n+1}^1(\theta) + \cdots + S_{n+1}^k(\theta), \qquad k = 1, \ldots, K$$

on $V_n(\omega)$. Furthermore, by writing a recursive expression for $T_n^k(\theta)$ it is readily seen that $T_n^k(\cdot)$ is subdifferentiable, since so are the service times. Therefore, the conditions of Theorem 3.2 are satisfied and hence we conclude that $\{T_n(\theta_0), \partial T_n(\theta_0)\}$ regenerates at the epochs $\{\sigma_j(\theta_0)\}$.

Let us show now that Assumption A3 is satisfied. It is possible to show that, given $\omega \in \Omega$, $\theta \in \Theta$, $k \le K$ and $n \ge 0$, there exist sets $\mathcal{J}$ and $\mathcal{M}$ (depending on ω, θ, n and k) such that $\mathcal{J} \subset \{1, \ldots, K\}$, $\mathcal{M} \subset \{\sigma_{\ell-1}(\theta), \ldots, \sigma_\ell(\theta) - 1\}$ for some $\ell > 0$, $\sigma_{\ell-1}(\theta) \le n \le \sigma_\ell(\theta) - 1$ and we can write

$$T_n^k(\theta, \omega) = \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} S_m^j(\theta, \omega) - \sum_{m=r}^{n} A_m \tag{6.5}$$

for some $r \in [\sigma_{\ell-1}(\theta), \sigma_\ell(\theta) - 1]$. The sets $\mathcal{J}$ and $\mathcal{M}$ correspond to the solution of a longest path problem in a graph, as shown in Homem-de-Mello, Shapiro and Spearman (1999); we refer to that paper for details. Let $\mathcal{J}_0$, $\mathcal{M}_0$, and r_0 be the elements corresponding to $T_n^k(\theta_0, \omega)$, i.e.,

$$T_n^k(\theta_0, \omega) = \sum_{j \in \mathcal{J}_0} \sum_{m \in \mathcal{M}_0} S_m^j(\theta_0, \omega) - \sum_{m=r_0}^{n} A_m.$$

By the property of longest paths we have that

$$\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} S_m^j(\theta, \omega) \ge \sum_{j \in \mathcal{J}_0} \sum_{m \in \mathcal{M}_0} S_m^j(\theta, \omega),$$

$$\sum_{j \in \mathcal{J}_0} \sum_{m \in \mathcal{M}_0} S_m^j(\theta_0, \omega) \ge \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} S_m^j(\theta_0, \omega),$$

and thus

$$\begin{aligned}\sum_{j \in \mathcal{J}_0} \sum_{m \in \mathcal{M}_0} \left[S_m^j(\theta, \omega) - S_m^j(\theta_0, \omega)\right] &\le T_n^k(\theta, \omega) - T_n^k(\theta_0, \omega) \\ &\le \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} \left[S_m^j(\theta, \omega) - S_m^j(\theta_0, \omega)\right].\end{aligned}$$

Notice that the term $\sum_{m=r}^{n} A_m$ disappears in the above inequality—this follows from an argument similar to the one used in Example 1. It follows that

$$|T_n^k(\theta,\omega) - T_n^k(\theta_0,\omega)| \le \max\Bigg(\Bigg| \sum_{j\in\mathcal{J}_0} \sum_{m\in\mathcal{M}_0} [S_m^j(\theta,\omega) - S_m^j(\theta_0,\omega)] \Bigg|, \Bigg| \sum_{j\in\mathcal{J}} \sum_{m\in\mathcal{M}} [S_m^j(\theta,\omega) - S_m^j(\theta_0,\omega)] \Bigg| \Bigg). \tag{6.6}$$

Next, by the assumption (6.4) on service times it follows from (6.6) that, for any $\delta \in (0,1)$, there exists an event B_n^k with positive probability such that, on B_n^k, we have

$$|T_n^k(\theta) - T_n^k(\theta_0)| \le \sum_{j\in\mathcal{J}_1} \sum_{m\in\mathcal{M}_1} M_\delta \|\theta - \theta_0\|,$$

where $\mathcal{J}_1$ and $\mathcal{M}_1$ correspond to the maximizer of the right-hand side of (6.6). Let $\nu = \operatorname{card}(\mathcal{J}_1)\operatorname{card}(\mathcal{M}_1)$. Then, we have that $P(B_n^k \mid \nu = r) \ge (1-\delta)^r$. Moreover, since $\mathcal{J}_1 \subset \{1,\dots,K\}$ and $\mathcal{M}_1 \subset \{\sigma_{\ell-1},\dots,\sigma_\ell\}$ for some $\ell > 0$, it follows that $\operatorname{card}(\mathcal{J}_1) \le K$ and $\operatorname{card}(\mathcal{M}_1) \le \tau_\ell$, where τ_ℓ is the length of the ℓth cycle (notice that here σ and τ have arguments θ or θ_0, according to the maximizer of the right-hand side of (6.6)). Since $\mathbb{E}\tau_\ell < \infty$, there exists a constant $Q_\delta > 0$ such that

$$P\big(\operatorname{card}(\mathcal{M}_1) \le Q_\delta\big) > 1 - \delta,$$

which implies that

$$P\big(\nu \le KQ_\delta\big) > 1 - \delta.$$

It follows that on $B_n^k \cap \{\nu \le KQ_\delta\}$ we can write

$$|T_n^k(\theta) - T_n^k(\theta_0)| \le \nu M_\delta \|\theta - \theta_0\| \le KQ_\delta M_\delta \|\theta - \theta_0\|. \tag{6.7}$$

Let V denote the neighborhood $V = \{\theta : |\theta - \theta_0| < \varepsilon / KQ_\delta M_\delta\}$. Then, from (6.7) we see that, on $B_n^k \cap \{\nu \le KQ_\delta\}$, we have, for all $\theta \in V$, $|T_n^k(\theta) - T_n^k(\theta_0)| \le \varepsilon$. Finally, since

$$P\big(B_n^k \cap \{\nu \le KQ_\delta\}\big) = \sum_{r=1}^{KQ_\delta} P\big(B_n^k \mid \nu = r\big) P(\nu = r) \ge \sum_{r=1}^{KQ_\delta} (1-\delta)^r P(\nu = r),$$

it follows that there exists some $\tilde{\delta} \in (0,1)$ such that $\tilde{\delta} \to 0$ as $\delta \to 0$ and

$$P\big(|T_n^k(\theta) - T_n^k(\theta_0)| \le \varepsilon\big) > 1 - \tilde{\delta},$$

for all n and k, and all $\theta \in V$. We conclude that

$$P\big(T_{\sigma_m}^k(\theta)\big|_V = S_{\sigma_m}^1(\theta) + \cdots + S_{\sigma_m}^k(\theta)\big) > 1 - \tilde{\delta},$$

for all m and k. Therefore, Assumption A3 holds.

Notice also that from (6.5) we can see that if $\partial S_n^k(\theta)$ is integrable for all n and k—which happens, for example, if $S_n^k(\theta)$ has the form $Y_n^k \theta$, where Y_n^k is a random variable with finite expectation—then Assumptions B2 and B3 will follow, as in the previous example, from finiteness of first and second moments of the cycle lengths. Thus, by applying Proposition 5.3 we obtain

$$\partial \mathbb{E}[T_\infty^k(\theta_0)] = \frac{\mathbb{E}\big[\sum_{n=0}^{\sigma_1 - 1} \partial T_n^k(\theta_0)\big]}{\mathbb{E}\sigma_1}.$$

We can now estimate an element of $\partial\mathbb{E}[T_\infty^k(\theta_0)]$, as follows. Fix some $M > 0$, and simulate the system for M cycles. Let $\sigma_0, \ldots, \sigma_M$ be the regeneration points (without loss of generality, assume we start with an empty system, i.e., $\sigma_0 = 0$). Then,

$$\frac{\sum_{i=1}^{M}\sum_{j=\sigma_{i-1}}^{\sigma_i-1}\partial T_j^k(\theta_0)}{\sum_{i=1}^{M}(\sigma_i-\sigma_{i-1})}$$

is an estimator of $\partial\mathbb{E}[T_\infty^k(\theta_0)]$. Notice that for any $j \geq 0$ we have

$$T_{\sigma_j}^k(\theta_0) = S_{\sigma_j}^1(\theta_0) + \cdots + S_{\sigma_j}^k(\theta_0), \qquad k = 1, \ldots, K,$$

and hence

$$\nabla T_{\sigma_j}^k(\theta_0) = \nabla S_{\sigma_j}^1(\theta_0) + \cdots + \nabla S_{\sigma_j}^k(\theta_0), \qquad k = 1, \ldots, K,$$

where ∇f denotes an arbitrary subgradient of f. For the indices n such that $\sigma_{j-1} < n < \sigma_j$ a subgradient of $T_{\sigma_j}^k$ at θ_0 can be computed by using the corresponding solution of the longest-path problem mentioned above, together with subgradients of the service times; we refer again to Homem-de-Mello et al. (1999) for details. Another consistent estimator is obtained by simulating the system for N periods and computing

$$\frac{1}{N}\sum_{i=0}^{N-1}\nabla T_i^k(\theta_0).$$

7. Concluding remarks. In this paper we provided results that can potentially enlarge the scope of applications of sensitivity analysis and optimization in stochastic systems to include "nonsmooth" processes. In particular, we have shown that, under some assumptions, one can consistently estimate directional derivatives (and consequently subdifferential sets and subgradients) of expected steady-state functions both by ratio-type and long-run average formulas. Those formulas are convenient in that they allow the computation of the derivatives to be done simultaneously with the original process during the simulation. We have given some examples showing potential applications of these results.

From the more theoretical viewpoint, our contribution is twofold: First, we extended the result in Shapiro and Wardi (1994) by exhibiting a necessary and sufficient condition for the differentiability of the expected steady-state function. Second, we showed a limit theorem for compact-convex-valued multifunctions by proving that, under proper assumptions, the average of a sequence of *regenerative* random multifunctions converge with probability one.

We hope this work will allow many results from the well-studied fields of nonsmooth analysis and multifunctions to be incorporated into the area of stochastic processes.

Acknowledgments. The author is grateful to Alexander Shapiro for some enlightening discussions during the preparation of this work. The author also thanks three anonymous referees for some detailed comments and suggestions that greatly helped to improve the quality of the paper, and David McDonald and A. Ioffe for discussions on the Bochner integral. This work was supported, in part, by CNP_q, Brasília, Brazil, through a Doctoral Fellowship under Grant 200595/93-8. It was also supported in part by grant Grant DMI-9713878 from the National Science Foundation.

References

Asmussen, S. 1987. *Applied Probability and Queues.* Wiley, New York.

Bratley, P., B. L. Fox, L. E. Schrage. 1987. *A Guide to Simulation*, 2nd. Edition. Springer-Verlag, New York.

Castaing, C., M. Valadier. 1977. *Convex Analysis and Measurable Multifunctions.* Springer-Verlag, Berlin, Germany.

Chong, E. K. P., P. J. Ramadge. 1994. Stochastic optimization of regenerative systems using infinitesimal perturbation analysis. *IEEE Trans. Automat. Control* **39** (7) 1400–1410.

Chung, K. L. 1974. *A Course in Probability Theory*, 2nd Edition. Academic Press, San Diego, CA.

Clarke, F. H. 1990. *Optimization and Nonsmooth Analysis*. Reprint, SIAM (originally published by Wiley, New York, 1983).

Debreu, G. 1966. Integration of correspondences. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* Vol. II, Part I, 351–372.

Diestel, J., J. J. Uhl, Jr. 1977. *Vector Measures*. American Mathematical Society, Providence, RI.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Kluwer, Norwell, MA.

———. 1993. Regenerative derivatives of regenerative sequences. *Adv. Appl. Probab.* **25** 116–139.

———, P. W. Glynn. 1992. Gradient estimation for regenerative processes. *Proc. 1992 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 280–288.

Glynn, P. W. 1989. Optimization of stochastic systems via simulation. *Proc. 1989 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 90–105.

———, P. L'Ecuyer. 1995. Likelihood ratio gradient estimation for stochastic recursions. *Adv. Appl. Probab.* **27** 1019–1053.

Hiai, F. 1984. Strong laws of large numbers for multivalued random variables. A. Dold and B. Eckmann, Eds. *Multifunctions and Integrands*. Springer-Verlag, Berlin, Germany.

———, H. Umegaki. 1977. Integrals, conditional expectations and martingales of multivalued functions. *J. Multivariate Anal.* **7** 149–182.

Hiriart-Urruty, J.-B., C. Lemaréchal. 1993. *Convex Analysis and Minimization Algorithms* I. Springer-Verlag, Berlin, Germany.

———, ———. 1993a. *Convex Analysis and Minimization Algorithms* II. Springer-Verlag, Berlin, Germany.

Homem-de-Mello, T., A. Shapiro, M. L. Spearman. 1999. Finding optimal material release times using simulation based optimization. *Management Sci.* **45** 86–102.

Hu, J. Q. 1992. Convexity of sample path performance and strong consistency of infinitesimal perturbation analysis. *IEEE Trans. Automat. Control* **37** (2) 258–262.

Ioffe, A. D., V. M. Tihomirov. 1969. On the minimization of integral functionals. *Funct. Anal. Appl.* **3** 218–227.

———, ———. 1979. *Theory of Extremal Problems*. North-Holland, Amsterdam, The Netherlands.

Kelley, J. L., T. P. Srinivasan. 1988. *Measure and Integral*. Springer-Verlag, New York.

L'Ecuyer, P. 1990. A unified view of the IPA, SF and LR gradient estimation techniques. *Management Sci.* **36** 1364–1383.

———, P. W. Glynn. 1994. Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state. *Management Sci.* **40** 1562–1578.

Nummelin, E. 1981. Regeneration in tandem queues. *Adv. Appl. Probab.* **13**, 221–230.

———. 1984. *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge, MA.

Pflug, G. Ch. 1996. *Optimization of Stochastic Models*. Kluwer, Norwell, MA.

Plambeck, E. L., B. R. Fu, S. M. Robinson, R. Suri. 1996. Sample-path optimization of convex stochastic performance functions. *Math. Programming Ser.* B **75** 137–176.

Radström, H. 1952. An embedding theorem for spaces of convex sets. *Proc. Amer. Math. Soc.* **3** 165–169.

Revuz, D. 1984. *Markov Chains*. North-Holland, Amsterdam, The Netherlands.

Robinson, S. M. 1995. Convergence of subdifferentials under strong stochastic convexity. *Management Sci.* **41** 1397–1401.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

———. 1976. Integral functionals, normal integrands and measurable selections. J.P. Gossez, E.J. Lami Dozo, J. Mawhin and L. Waelbroeck, Eds. *Nonlinear Operators and the Calculus of Variations*. Springer-Verlag, Berlin, Germany.

———, R. J.-B. Wets. 1982. On the interchange of subdifferentiation and conditional expectation for convex functionals. *Stochastics* **7** 173–182.

Royden, H. 1988. *Real Analysis*. Macmillan, New York.

Rubinstein, R. Y., A. Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. John Wiley & Sons, New York.

Shapiro, A. 1990. On concepts of directional differentiability. *J. Optim. Theory Appl.* **66** (3) 477–487.

———, Y. Wardi. 1994. Nondifferentiability of the steady-state function in discrete event dynamical systems. *IEEE Trans. Automat. Control* **39** (8) 1707–1711.

Shedler, G. S. 1987. *Regeneration and Networks of Queues*. Springer-Verlag, New York.

Suri, R. 1989. Perturbation analysis: The state of the art and research issues explained via the GI/G/1 queue. *Proc. IEEE* **77** 114–137.

———, Y. T. Leung. 1989. Single run optimization of discrete event simulations—an empirical study using the M/M/1 queue. *IIE Trans.* **21** 35–49.

Thorisson, H. 2000. *Coupling, Stationarity and Regeneration*. Springer-Verlag, New York.

Wardi, Y., J. Q. Hu. 1991. Strong consistency of infinitesimal perturbation analysis for tandem queueing networks. *J. Discrete Event Dynamic Systems: Theory and Applications* **1** 37–59.
Wolff, R. W. 1989. *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, NJ.

T. Homem-de-Mello: Department of Industrial, Welding & Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, Ohio 43210; e-mail: homem-de-mello.1@osu.edu