

A STOCHASTIC ALGORITHM FOR PARAMETRIC SENSITIVITY IN SMOLUCHOWSKI'S COAGULATION EQUATION

ISMAËL F. BAILLEUL^{†‡}, PETER L. W. MAN[§] AND MARKUS KRAFT[§]

Abstract. In this article a stochastic particle system approximation to the parametric sensitivity in the Smoluchowski coagulation equation is introduced. The parametric sensitivity is the derivative of the solution to the equation with respect to some parameter, where the coagulation kernel depends on this parameter. It is proved that the particle system converges weakly to the sensitivity as the number of particles N increases. A Monte Carlo algorithm is developed and variance reduction techniques are applied. Numerical experiments are conducted for two kernels: the additive kernel and one which has been used for studying soot formation in a free molecular regime. It is shown empirically that the techniques for variance reduction are indeed very effective and that the order of convergence is $O(1/N)$. The algorithm is then compared to an algorithm based on a finite difference approximation to the sensitivity and it is found that the variance of the sensitivity estimators are considerably lower than that for the finite difference approach. Furthermore, two methods of establishing 'efficiency' are considered and the new algorithm is found to be significantly more efficient.

Key words. Smoluchowski coagulation equation, sensitivity, particle system, coupling, simulations.

AMS subject classifications. 65C05, 65C35, 68U20, 82C22, 60F05

1. Introduction. Smoluchowski's description of a coagulation process is made in terms of densities $\mu_t(x)$ of particles of mass $x = 1, 2, 3, \dots$ and takes the form of an infinite dimensional differential equation

$$\frac{d}{dt}\mu_t(x) = \frac{1}{2} \sum_{y=1}^{x-1} K(y, x-y) \mu_t(y) \mu_t(x-y) - \mu_t(x) \sum_{y=1}^{\infty} K(x, y) \mu_t(y). \quad (1.1)$$

The symmetric kernel $K(x, y)$ appearing in this equation should be understood as giving the rate at which two particles of mass x and y coagulate. One gets an equivalent and more symmetric equation considering $\mu_t(\cdot)$ as a measure on the set of non-negative integers and looking at the time evolution of observables of the form $(f, \mu_t) := \sum_x f(x) \mu_t(x)$; moments are examples of such observables. In these terms, equation (1.1) takes the form

$$(f, \mu_t) = (f, \mu_0) + \frac{1}{2} \int_0^t \left(\sum_{x, y \geq 1} \{f(x+y) - f(x) - f(y)\} K(x, y) \mu_s(x) \mu_s(y) \right) ds. \quad (1.2)$$

The basic problem we address is to derive a numerical scheme to understand how the solution to this equation depends on possible parameters in the kernel. We shall write K_λ to indicate that K depends on some d -dimensional parameter λ , and shall write μ_t^λ for the solution of equation (1.2). Formally differentiating this equation with respect to λ and setting $\sigma_t^\lambda = \partial_\lambda \mu_t^\lambda$ we get

$$\begin{aligned} (f, \sigma_t^\lambda) &= (f, \sigma_0^\lambda) + \frac{1}{2} \int_0^t \left(\sum_{x, y \geq 1} \{f(x+y) - f(x) - f(y)\} K_\lambda(x, y) \mu_s^\lambda(x) \sigma_s^\lambda(y) \right) ds \\ &\quad + \int_0^t \left(\sum_{x, y \geq 1} \{f(x+y) - f(x) - f(y)\} K'_\lambda(x, y) \mu_s^\lambda(x) \mu_s^\lambda(y) \right) ds. \end{aligned} \quad (1.3)$$

K'_λ is here the derivative of K_λ with respect to λ . Section 2 presents an algorithm which simulates the sensitivity σ_t^λ very accurately and in an efficient way.

[†]Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK (i.bailleul@statslab.cam.ac.uk)

[‡]This research was supported by EPSRC grant EP/E01772X/1.

[§]Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge, CB2 3RA, UK (mk306@cam.ac.uk)

There are two main motivations for performing sensitivity analysis. The first is for solving inverse problems. If some particle system is governed by a partial differential equation which in turn is dependent upon some unknown parameter, it is desirable to find this parameter. This can be achieved by choosing the parameter value which minimises some residual which is a function of experimentally realised quantities and its computational analogue, which varies with the parameter. The minimisation procedure often uses a gradient search, thus the value of computing parametric derivatives is apparent. Secondly, in considering a scientific model, we often wish to consider the smallest model which reasonably fits the data, in which case sensitivity analysis can be performed to discard parameters with small sensitivity.

Whilst the usual tools of solving differential equations (and their associated numerical schemes) are badly adapted to the above infinite dimensional framework, the stochastic approach of interaction particle systems (basically Markov chains) can be used efficiently, in this setting, as Marcus in [1], and later Lushnikov in [2], first realised. We follow their approach and give a stochastic particle approximation of the sensitivity σ_t^λ .

Before running any simulation, one should investigate the well-posedness of equation (1.3): if it had more than one solution it would be unclear what solution a numerical scheme approximates. The most general answers to this theoretical question for Smoluchowski equation were given by Jeon in [3] and Norris in [4], under a growth assumption on the interaction kernel and a moment condition on the initial condition μ_0 . Surprisingly enough, the existence and uniqueness problem for the sensitivity was only solved recently, by Bailleul [5], using methods developed by Kolokoltsov [6]. The algorithm developed in this article is the numerical counterpart of this theoretical work¹.

Three approaches to the simulation of the sensitivity by systems of particles have mainly been used up to now. The first uses weighted particles, as illustrated by Vikhansky and Kraft [7]. They approximate the family of solutions $\{\mu_t^\lambda\}_\lambda$ by Marcus-Lushnikov processes $\sum_{n \geq 0} w_n(t; \lambda) \delta_{x_n(t)}$ where the dependence on λ is entirely put on the weights $w_n(t; \lambda)$. A heuristic argument imposes to their derivative to satisfy a kind of Markov evolution rule. Despite its (numerically verified) convergence this approach essentially has the same speed of convergence and variance as the Marcus-Lushnikov process. Further, the paper does not any information regarding computation run times.

The second approach considers adjoint sensitivity [8]. A backward partial differential equation is used rather than a forward one, as used in most other methods. The advantage of this method is that sensitivity for any parameter value is immediate once the computation have been done whereas using the forward equation requires explicit calculation for each parameter value. The disadvantage is that one can only calculate the sensitivities for a particular functional of the particle ensemble.

In the third approach, devised by the authors with J. R. Norris in the forthcoming article [9], the sensitivity σ_t^λ is approximated by the ratio $(\mu_t^{\lambda+\delta\lambda; N} - \mu_t^{\lambda; N})/\delta\lambda$, where $\mu_t^{\lambda+\delta\lambda; N}$ and $\mu_t^{\lambda; N}$ are two Marcus-Lushnikov processes corresponding to close parameters, coupled so as to minimise the difference of their random fluctuations around $\mu_t^{\lambda+\delta\lambda}$ and μ_t^λ . This approach leads to a massive decrease of variance but does not improve the speed of convergence of the algorithm.

¹Consult this article for conditions under which existence and uniqueness of a solution to the sensitivity equation (1.3) holds.

The algorithm we propose improves the variance of the sensitivity estimator and requires a much smaller number of particles to converge. This is described in section 2. The reader who is not interested in mathematical details can skip sections 2.1 and 2.2 where it is proven that the particle system introduced in section 2 converges to the sensitivity. Section 3 presents the algorithm we have used to obtain the numerical results of section 4.

Notation. We shall prove convergence of the particle system in a general setting where masses of particles can take any positive real value. The densities of particles will then be represented by non-negative measures μ_t and all sums will be replaced by integrals. In this framework we shall write (f, μ) for $\int f(x)\mu(dx)$ and Smoluchowski's equation (1.2) will be written

$$(f, \mu_t) = (f, \mu_0) + \frac{1}{2} \int_0^t \int \{f(x+y) - f(x) - f(y)\} K(x, y) \mu_s(dx) \mu_s(dy) ds.$$

We shall formally write it as

$$\dot{\mu}_t^\lambda = \frac{1}{2} K_\lambda(\mu_t^\lambda, \mu_t^\lambda). \quad (1.4)$$

In the same way, we shall write formally equation (1.3) for the sensitivity as

$$\dot{\sigma}_t^\lambda = K_\lambda(\mu_t^\lambda, \sigma_t^\lambda) + \frac{1}{2} K'_\lambda(\mu_t^\lambda, \mu_t^\lambda). \quad (1.5)$$

The integral notation is adopted from now on.

2. Markov chain approximation. It is probably fair to say that although the Smoluchowski equation (1.2) is a deterministic evolution equation it should primarily be thought of as a deterministic large scale picture of a stochastic mesoscopic dynamics. Indeed, Smoluchowski obtained his equation from a representation of the coagulation process using 'particles' moving according to Brownian trajectories whose diffusivity depends on their mass and coagulate when they are close to each other. As explained in the article [10] of Chandrasekhar, section 6 of chapter III, in a region of space where the coagulating particles are well mixed, one can forget about their spatial location and obtain a mean-field evolution for their mass distribution. This mean-field picture is provided by Smoluchowski equation. Given in its simple form (1.1), it is not clear at first sight how one should simulate a solution to this infinite dimensional differential system.

The approach developed by Marcus in his seminal paper [1] in a sense comes back to the primary stochastic description of the coagulation phenomenon and relies on the intuitive content of Smoluchowski equation. Two particles of masses x and y coagulate at rate $K(x, y)$ to create a new particle of mass $x + y$: The particles x and y are removed from the system and the particle $x + y$ added. This motivated Marcus, and later Lushnikov, to represent a particle of mass x by a Dirac mass δ_x at x and to introduce a strong Markov jump process on the space of discrete non-negative measures with the following simple dynamics. Denote by $\mu_0^N = \frac{1}{N} \sum_i \delta_{x_i}$ its initial state and by μ_t^N its state at time t . Associate to each pair $1 \leq i < j \leq N$ independent exponential random times T_{ij} with parameter $\frac{K(x_i, x_j)}{N}$ and set

$$T = \min\{T_{ij}; 1 \leq i < j \leq N\}.$$

The process remains constant on the time interval $[0, T)$, and if $T = T_{pq}$ it has a jump $\frac{1}{N}(\delta_{x_p+x_q} - \delta_{x_p} - \delta_{x_q})$ at time T . The dynamics then starts afresh. Note that the new measure at time T is still non-negative, and that the above description leads to a mean jump of the process during a time interval $[t, t + \delta t]$ equal to²

$$\delta t \sum_{x, x'} (\delta_{x+x'} - \delta_x - \delta_{x'}) K(x, y) \mu_t^N(x) \mu_t^N(x')$$

up to terms of order $\frac{\delta t}{N}$ and $o(\delta t)$. This property makes it clear that the process converges to a solution of the Smoluchowski equation as N goes to infinity (under proper conditions), a fact which was used for simulation purposes long before it was proved under general conditions in [4].

Following the heuristic approach of Marcus and Lushnikov, we are going to give in the next section a particle description of the sensitivity equation

$$\dot{\sigma}_t^\lambda = K_\lambda(\mu_t^\lambda, \sigma_t^\lambda) + \frac{1}{2} K'_\lambda(\mu_t^\lambda, \mu_t^\lambda). \quad (2.1)$$

To that end, introduce the notation $K'_+ := K' \vee 0$ and $K'_- := K' \wedge 0$ (dropping the index λ for it will be fixed), and write, for a signed measure σ ,

$$\sigma = \sigma \mathbf{1}_{\frac{d\sigma}{d|\sigma|} > 0} - |\sigma| \mathbf{1}_{\frac{d\sigma}{d|\sigma|} < 0} =: \sigma^+ - \sigma^-,$$

Using this notation, re-write equation (2.1) as

$$\dot{\sigma}_t^+ - \dot{\sigma}_t^- = \left(K_\lambda(\mu_t^\lambda, \sigma_t^+) + \frac{1}{2} K'_+(\mu_t^\lambda, \mu_t^\lambda) \right) - \left(K_\lambda(\mu_t^\lambda, \sigma_t^-) + \frac{1}{2} K'_-(\mu_t^\lambda, \mu_t^\lambda) \right) \quad (2.2)$$

This equation will motivate the introduction of the Markov chain described in the next section.

Notation. Given three non-negative measures μ, σ^+, σ^- on $(0, \infty)$ we shall adopt the notation $\mu \oplus \sigma^+ \oplus \sigma^-$ to denote the \mathbb{R}_+^3 -valued measure on $(0, \infty)^3$. It will clarify the notation to denote by $x \oplus y \oplus z$ the point of \mathbb{R}^3 with co-ordinates x, y and z . Given non-negative functions f, g, h on $(0, \infty)$ set

$$(f \oplus g \oplus h, \mu \oplus \sigma^+ \oplus \sigma^-) := (f, \mu) \oplus (g, \sigma^+) \oplus (h, \sigma^-).$$

As we shall simulate both μ_t and (σ_t^+, σ_t^-) at the same time, our approximating Markov chain will take values in the set

$$\mathcal{N} := \{ \mu \oplus \sigma^+ \oplus \sigma^- ; \mu, \sigma^+, \sigma^- \text{ non-negative discrete measures on } (0, \infty) \}.$$

2.1. Chain, generator. In the same way as the right hand side of Smoluchowski equation (1.4) can be interpreted as the coagulation of particles of μ_t of mass x and y at rate $K(x, y)$, we are going to follow what equation (2.2) suggest and interpret the term $K(\mu_t, \sigma_t^+)$ appearing there as the coagulation of a particle in μ_t of mass x with a particle in σ_t^+ of mass y at rate $K(x, y)$. Note that this leads to a jump $\delta_{x+y} - \delta_x - \delta_y$ of σ^+ which could transform the non-negative measure σ_t^+ into a signed measure, as the term δ_x does not necessarily appear inside σ_t^+ (while δ_y does). We shall take care of this by adding δ_x to the negative part σ_t^- of σ_t instead of subtracting it from σ_t^+ ;

² μ_t^N denotes the state of the process at time t .

as we are only interested in the difference $\sigma_t^+ - \sigma_t^- (= \sigma_t)$ this has no consequence. Note also that the particle δ_x from μ_t used in that coagulation event will not be removed from μ_t . Similar interpretations of the terms $K(\mu_t, \sigma_t^-)$ and $\frac{1}{2}K'_\pm(\mu_t, \mu_t)$ lead us to define the following Markov chain $\Theta_t = X_t \oplus Y_t \oplus Z_t$ on \mathcal{N} . Denote by $\Theta_0 = \left(\sum_{i=1..m} \delta_{x_i} \right) \oplus \left(\sum_{k=1..p} \delta_{y_k} \right) \oplus \left(\sum_{\ell=1..q} \delta_{z_\ell} \right)$ its starting point.

2.1.1. Dynamics. Associate to each pair

- $1 \leq i < j \leq m$, exponential random variables R_{ij}, S_{ij} and T_{ij} with respective parameters $K(x_i, x_j)$ and $K'_+(x_i, x_j)$ and $K'_-(x_i, x_j)$,
- $(i, k) \in \llbracket 1, m \rrbracket \times \llbracket 1, p \rrbracket$ an exponential random variable U_{ik} with parameter $K(x_i, y_k)$,
- $(i, \ell) \in \llbracket 1, m \rrbracket \times \llbracket 1, q \rrbracket$ an exponential random variable $V_{i\ell}$ with parameter $K(x_i, z_\ell)$.

All these random variables are supposed to be independent. Denoting by W the first event happening in the system

$$W = \min\{R_{ij}, S_{ij}, T_{ij}, U_{ik}, V_{i\ell} ; 1 \leq i < j \leq m, k \in \llbracket 1, p \rrbracket, \ell \in \llbracket 1, q \rrbracket\},$$

the jump $\Delta\Theta$ of the Markov chain depends on which of these exponential clocks rings first. For future reference, the different types of events that can happen are numbered. If

$$\begin{aligned} W = R_{ij}, & \text{ then } \Delta\Theta = (\delta_{x_i+x_j} - \delta_{x_i} - \delta_{x_j}) \oplus 0 \oplus 0 & (\text{event type: } 0) \\ W = S_{ij}, & \text{ then } \Delta\Theta = 0 \oplus \delta_{x_i+x_j} \oplus (\delta_{x_i} + \delta_{x_j}) & (\text{event type: } 1^+) \\ W = T_{ij}, & \text{ then } \Delta\Theta = 0 \oplus (\delta_{x_i} + \delta_{x_j}) \oplus \delta_{x_i+x_j} & (\text{event type: } 1^-) \\ W = U_{ik}, & \text{ then } \Delta\Theta = 0 \oplus (\delta_{x_i+y_k} - \delta_{y_k}) \oplus \delta_{x_i} & (\text{event type: } 2^+) \\ W = V_{i\ell}, & \text{ then } \Delta\Theta = 0 \oplus \delta_{x_i} \oplus (\delta_{x_i+z_\ell} - \delta_{z_\ell}) & (\text{event type: } 2^-) \end{aligned}$$

The process Θ_t will be constant on the time interval $[0, W)$ and have jump $\Delta\Theta$ at time W . The dynamics then starts afresh.

Remark. *It is clear from this description that for any function φ satisfying the relation $\varphi(a+b) \geq \varphi(a) - \varphi(b)$ for any $a, b > 0$, the function $(\varphi, Y_t + Z_t)$ increases with time. This fact is useful for the convergence result stated in theorem 2.2.*

Given any positive integer N , define $\frac{1}{N}\Theta_t$ as the element $\frac{1}{N}X_t \oplus \frac{1}{N}Y_t \oplus \frac{1}{N}Z_t$ of \mathcal{N} , and set

$$\Theta_t^N := \frac{1}{N}\Theta_{\frac{t}{N}} =: \mu_t^N \oplus \sigma_t^{+,N} \oplus \sigma_t^{-,N}.$$

Note that the first component of Θ_t^N is the usual Marcus-Lushnikov process. Set $\sigma_t^N = \sigma_t^{+,N} - \sigma_t^{-,N}$. We are going to prove in theorem 2.2 that σ_t^N converges in law to the sensitivity σ_t . *Those who do not care about the mathematical details of such a statement can skip the remaining of this section and go to section 3.*

2.1.2. Generator. The analytic description of the Markov chain $\{\Theta_t^N\}_{t \geq 0}$ in terms of its generator will be useful in proving theorem 2.2. Given a non-negative measure μ of the form $\frac{1}{N} \sum \delta_{x_i}$ define the rescaled counting measure on ordered pairs of masses of distinct particles by

$$\tilde{\mu}(A \times A') := \mu(A) \mu(A') - \frac{1}{N} \mu(A \cap A'),$$

and define the measure $\mathbf{G}^{(N)}(\mu)$ and the operator $\mathbf{P}^{(N)}(\mu)$ setting for any measurable bounded function f

$$\begin{aligned} (f, \mathbf{G}^{(N)}(\mu)) &= \frac{1}{2} \int \{f(x+x') - f(x) - f(x')\} K(x, x') \tilde{\mu}(dx, dx') \\ (f, \mathbf{P}^{(N)}(\mu)) &= \frac{1}{2} \int \{f(x+x') - f(x) - f(x')\}^2 K(x, x') \tilde{\mu}(dx, dx'). \end{aligned}$$

Given $x > 0$ and a non-negative measure γ on \mathbb{R}_+^* we shall write $K(x, \gamma)$ for the integral $\int K(x, y)\gamma(dy)$.

Denote by $\mathbf{H}^{(N)}$ the generator of the process $\{\Theta_t^N\}_{0 \leq t \leq T}$; for any bounded measurable functions f, g, h on $(0, \infty)$ the \mathbb{R}^3 -valued process

$$M_t^{f,g,h;N} := (f \oplus g \oplus h, \Theta_t^N) - (f \oplus g \oplus h, \Theta_0^N) - \int_0^t (f \oplus g \oplus h, \mathbf{H}^{(N)}(\Theta_s^N)) ds$$

is a martingale (with respect to its natural filtration). For a measure μ of the form $\frac{1}{N} \sum \delta_{x_i}$ and $\Theta = \mu \oplus \sigma^+ \oplus \sigma^- \in \mathcal{N}$ we have

$$\begin{aligned} (f \oplus g \oplus h, \mathbf{H}^{(N)}(\Theta)) &= \\ & (f, \mathbf{G}^{(N)}(\mu)) \oplus \\ & \left\{ \frac{1}{2} \int \{g(x+x')K'_+(x, x') + (g(x) + g(x'))K'_-(x, x')\} \tilde{\mu}(dx, dx') \right. \\ & + \int \left\{ (g(x+y) - g(y))K(x, y)\sigma^+(dy) + g(x)K(x, \sigma^-) \right\} \mu(dx) \left. \right\} \oplus \\ & \left\{ \frac{1}{2} \int \{h(x+x')K'_-(x, x') + (h(x) + h(x'))K'_+(x, x')\} \tilde{\mu}(dx, dx') \right. \\ & + \int \left\{ (h(x+z) - h(z))K(x, z)\sigma^-(dz) + h(x)K(x, \sigma^+) \right\} \mu(dx) \left. \right\} \end{aligned} \tag{2.3}$$

Compare this formula with the description of the dynamics given in the section 2.1.1.

- (i) Event $\{W = R_{ij}\}$ corresponds to the term $(f, \mathbf{G}^{(N)}(\mu)) \oplus 0 \oplus 0$;
 - (ii) Event $\{W = S_{ij}\}$ corresponds to the term $\frac{1}{2} \int 0 \oplus g(x+y) \oplus (h(x) + h(y)) K'_+(x, y) \tilde{\mu}(dx, dy)$; a similar term corresponds to the event $\{W = T_{ij}\}$;
 - (iii) Event $\{W = U_{ik}\}$ corresponds to the term $\int \{0 \oplus (g(x+z) - g(z)) \oplus h(x)K(x, z)\sigma^+(dz)\} \mu(dx)$; a similar term corresponds to the event $\{W = V_{i\ell}\}$.
- The sum of all these terms gives $(f \oplus g \oplus h, \mathbf{H}^{(N)}(\Theta))$.

Following a classical approach, the study of martingales of the form $M^{f,g,h;N}$ will be our main tool in the proof of the convergence theorem. The explicit expression of the bracket of $M^{f,g,h;N}$ will be useful in that task. We have

$$\langle M^{f,g,h;N} \rangle_t = \frac{1}{N} \int_0^t (f \oplus g \oplus h, \mathbf{Q}^{(N)}(\Theta_s^N)) ds,$$

where $\mathbf{Q}^{(N)}(\Theta)$ is characterised on measures Θ of the form $(\frac{1}{N} \sum \delta_{x_i}) \oplus \sigma^+ \oplus \sigma^-$ by

the formula

$$\begin{aligned}
& \left(f \oplus g \oplus h, \mathbf{Q}^{(N)}(\Theta) \right) = \\
& \left(f, \mathbf{P}^{(N)}(\mu) \right) \oplus \\
& \left\{ \frac{1}{2} \int \left\{ g(x+x')^2 K'_+(x, x') + (g(x) + g(x'))^2 K'_-(x, x') \right\} \tilde{\mu}(dx, dx') \right. \\
& \left. + \int \left\{ (g(x+y) - g(y))^2 K(x, y) \sigma^+(dy) + g(x)^2 K(x, \sigma^-) \right\} \mu(dx) \right\} \oplus \\
& \left\{ \frac{1}{2} \int \left\{ h(x+x')^2 K'_-(x, x') + (h(x) + h(x'))^2 K'_+(x, x') \right\} \tilde{\mu}(dx, dx') \right. \\
& \left. + \int \left\{ (h(x+z) - h(z))^2 K(x, z) \sigma^-(dz) + h(x)^2 K(x, \sigma^+) \right\} \mu(dx) \right\}
\end{aligned}$$

2.2. Convergence theorem. Denote by \mathcal{U} a bounded open set of some \mathbb{R}^d indexing the family K_λ of kernels. Let $\varphi : (0, \infty) \rightarrow \mathbb{R}_+$ be a sublinear function: $\varphi(sx) \leq s\varphi(x)$ for any $s > 0$ and $x \in (0, \infty)$; such a function is also subadditive: $\varphi(x+y) \leq \varphi(x) + \varphi(y)$, for any $x, y \in (0, \infty)$. We shall suppose that the interaction kernels K_λ satisfy the growth condition

$$K_\lambda(x, y) \leq \varphi(x)\varphi(y)$$

for any $x, y \in (0, \infty)$, $\lambda \in \mathcal{U}$, and that the initial condition of Smoluchowski equation (1.2) (or better its ‘continuous mass version’) satisfies the moment condition

$$\int \varphi(x)^{4+\epsilon} \mu_0(dx) < \infty \quad (2.4)$$

for some (small) $\epsilon > 0$. We shall suppose in theorem 2.2 that φ^2 is sub-additive; together with the above moment condition (2.4) on μ_0 this implies that Smoluchowski equation has a unique strong solution³, defined for all non-negative times.

The following norm was used on the space \mathcal{M}_1 of signed Borel measures μ such that $\|\mu\|_1 := (\varphi, |\mu|) < \infty$, in the article [5] where the following key result about sensitivity is proved.

THEOREM 2.1. *Assume the moment condition (2.4) and that $K_\lambda(x, y)$ and $|K'_\lambda(x, y)|$ are both bounded above by $\varphi(x)\varphi(y)$ for any x, y . Then the map $(t, \lambda) \in [0, \infty) \times \mathcal{U} \mapsto \mu_t^\lambda \in (\mathcal{M}_1, \|\cdot\|_1)$, is a \mathcal{C}^1 function and its derivative σ_t^λ satisfies the following equation for any bounded measurable function f ⁴.*

$$\begin{aligned}
(f, \sigma_t^\lambda) &= (f, \sigma_0^\lambda) + \int_0^t \int \{f\}(x, y) K_\lambda(x, y) \mu_s^\lambda(dx) \sigma_s^\lambda(dy) ds \\
&+ \frac{1}{2} \int_0^t \int \{f\}(x, y) K'_\lambda(x, y) \mu_s^\lambda(dx) \mu_s^\lambda(dy) ds
\end{aligned}$$

The function σ^λ is the only $(\mathcal{M}_1, \|\cdot\|_1)$ -valued solution of this equation.

We shall consider here a weaker topology than the $\|\cdot\|_1$ -topology. We shall equip the space $\mathbb{R}_+^{\oplus 3}$ with the ℓ^1 -distance: $\|x \oplus y \oplus z - x' \oplus y' \oplus z'\| := |x - x'| + |y - y'| + |z - z'|$.

³In the sense defined in [4].

⁴We write here $\{f\}(x, y)$ for $f(x+y) - f(x) - f(y)$.

Write $\mathcal{M}^{\oplus 3}$ for the set of non-negative \mathbb{R}_+^* -valued measures on \mathbb{R}_+^* , and let d be any distance on $\mathcal{M}^{\oplus 3}$ metrising weak convergence: $\{\Theta_n\}_{n \geq 0}$ converges to Θ_∞ iff for any bounded continuous functions f, g, h on \mathbb{R}_+^* , we have $(f \oplus g \oplus h, \Theta_n) \rightarrow (f \oplus g \oplus h, \Theta_\infty)$. The space $(\mathcal{M}^{\oplus 3}, d)$ is a Polish space with \mathcal{N} as a dense subset.

Fix a positive time T . We shall state our convergence theorem in the functional setting $\mathcal{D}([0, T], (\mathcal{M}^{\oplus 3}, d))$ of càdlàg paths from $[0, T]$ to $(\mathcal{M}^{\oplus 3}, d)$. This space will be equipped with its Skorokhod topology, for which we refer the reader to the books [11] or [12] of Billingsley and Pollard. Last, we shall denote by d_0 any distance on the set of all non-negative Borel measures on $(0, \infty)$ metrising the following notion of convergence⁵: $\{\mu_n\}_{n \geq 0}$ converges to μ_∞ iff we have $(f, \mu_n) \rightarrow (f, \mu_\infty)$ for any bounded continuous measurable function f with bounded support.

The starting point Θ_0^N of Θ_t^N will be of the form $\frac{1}{N}X_0^N \oplus \frac{1}{N}Y_0^N \oplus \frac{1}{N}Z_0^N$ for some non-negative integer-valued finite measures X_0^N, Y_0^N, Z_0^N on $(0, \infty)$. To shorten the notation we shall denote by

$$\Theta_t^N =: \mu_t^N \oplus \sigma_t^{+,N} \oplus \sigma_t^{-,N}$$

the process starting from Θ_0^N constructed in section 2.1 and corresponding to a given parameter λ .

We shall suppose that the function φ controlling the kernels K_λ satisfies identity (2.5) below. As noted in the remark on page 5, this hypothesis implies that the function $(\varphi, \sigma_t^{+,N} + \sigma_t^{-,N})$ increases with time; this fact will enable us to control Θ_t^N . Note that this hypothesis is weaker than requiring that φ be increasing.

THEOREM 2.2 (Convergence of the particle system). *Let $K_\lambda(\cdot, \cdot) : \mathbb{R}_+^* \times \mathbb{R}_+^* \rightarrow [0, +\infty)$ be a family of symmetric kernels indexed by $\lambda \in \mathcal{U}$. We suppose the map $(\lambda; x, x') \mapsto K_\lambda(x, x')$ continuous and differentiable with respect to λ , with a derivative $K'_\lambda(x, x')$ continuous with respect to (x, x') . Let $\varphi \geq 1$ be a subadditive function whose square is also subadditive. Assume that*

$$\varphi(a+b) \geq \varphi(a) - \varphi(b), \quad \text{for any positive } a, b, \quad (2.5)$$

$$\begin{aligned} \forall \lambda \in \mathcal{U}, \forall x, x', y \in \mathbb{R}_+^*, \quad K_\lambda(x, x') &\leq \varphi(x) \varphi(x'), \\ |K'_\lambda(x, y)| &\leq \varphi(x) \varphi(y), \end{aligned} \quad (2.6)$$

$$\frac{K_\lambda(x, x')}{\varphi(x) \varphi(x')} \quad \text{and} \quad \frac{K'_\lambda(x, x')}{\varphi(x) \varphi(x')} \xrightarrow{x+x' \rightarrow \infty} 0 \quad (2.7)$$

Fix $\lambda \in \mathcal{U}$ and write Θ_t^N for the corresponding process in \mathcal{N} , started from $\mu_0^N \oplus \sigma_0^{+,N} \oplus \sigma_0^{-,N}$. Suppose that μ_0 satisfies the moment condition (2.4) for some (small) ϵ , that

$$d_0(\varphi \mu_0^N, \varphi \mu_0) \rightarrow 0, \quad (2.8)$$

and that there exists a positive constant C bigger than (φ^2, μ_0^N) and $(\varphi, \sigma_0^{+,N} + \sigma_0^{-,N})$ for any $N \geq 1$.

⁵This notion of convergence, usually called vague convergence, is weaker than weak convergence.

Then the sequence of the laws of the processes Θ^N is tight and any (random) weak limit is almost surely of the form $\{\mu_t \oplus \sigma_t^{+, \infty} \oplus \sigma_t^{-, \infty}\}_{0 \leq t \leq T}$, with

$$\sigma_t^{+, \infty} - \sigma_t^{-, \infty} = \sigma_t.$$

Proof. The following estimate is essential in controlling the behaviour of the processes $\sigma^{+, N}$ and $\sigma^{-, N}$.

LEMMA 2.3. *There exists a positive constant C_1 such that*

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} (\varphi, \sigma_t^{+, N} + \sigma_t^{-, N}) \right] \leq C_1.$$

First decompose $(\varphi, \sigma_t^{+, N} + \sigma_t^{-, N})$ as the sum of a martingale $\{M_t\}_{0 \leq t \leq T}$ and a finite variation term:

$$\begin{aligned} (\varphi, \sigma_t^{+, N} + \sigma_t^{-, N}) &= (\varphi, \sigma_0^{+, N} + \sigma_0^{-, N}) + M_t \\ &+ \int_0^t \left(\int \{\varphi(x + x') + \varphi(x) + \varphi(x')\} K'(x, x') \tilde{\mu}_s^N(dx, dx') \right. \\ &\left. + \int \{\varphi(x + y) - \varphi(y) + \varphi(x)\} K(x, y) \mu_s^N(dx) (\sigma_s^{+, N} + \sigma_s^{-, N})(dy) \right) ds. \end{aligned}$$

From (2.6) we have for each $N \geq 1$ and $t \in [0, T]$

$$\begin{aligned} (\varphi, \sigma_t^{+, N} + \sigma_t^{-, N}) &\leq C + M_t + \int_0^t \int 2\{\varphi(x) + \varphi(x')\} \varphi(x) \varphi(x') \mu_s^N(dx) \mu_s^N(dx') ds \\ &+ 2 \int_0^t (\varphi^2, \mu_s^N) (\varphi, \sigma_s^{+, N} + \sigma_s^{-, N}) ds. \end{aligned}$$

This upper bound is simplified using the subadditivity of φ and φ^2 from which we have⁶

$$(\varphi, \mu_t^N) \leq (\varphi, \mu_0^N) \leq C \quad \text{and} \quad (\varphi^2, \mu_t^N) \leq (\varphi^2, \mu_0^N) \leq C.$$

This gives a Grönwall-type inequality

$$(\varphi, \sigma_t^{+, N} + \sigma_t^{-, N}) \leq C + M_t + 4C^2T + 2C \int_0^t (\varphi, \sigma_s^{+, N} + \sigma_s^{-, N}) ds$$

whose mean version gives a constant C_1 such that $\mathbb{E}[(\varphi, \sigma_t^{+, N} + \sigma_t^{-, N})] \leq C_1$ for any $0 \leq t \leq T$. We get the statement of the lemma recalling that hypothesis (2.5) implies that the function $t \mapsto (\varphi, \sigma_t^{+, N} + \sigma_t^{-, N})$ is increasing.

Given $\epsilon > 0$ define the compact subset

$$K_\epsilon = \left\{ \mu \oplus \sigma^+ \oplus \sigma^- \in \mathcal{M}^{\oplus 3}; \max\{(\varphi, \mu), (\varphi, \sigma^+), (\varphi, \sigma^-)\} \leq \frac{1}{\epsilon} \right\} \subset \mathcal{M}^{\oplus 3},$$

⁶Since $\varphi \geq 1$ we have $(\varphi, \mu_0^N) \leq (\varphi^2, \mu_0^N) \leq C$.

and denote by \mathbb{P}^N the law of Θ^N on $\mathcal{D}([0, T], (\mathcal{M}^{\oplus 3}, d))$.

COROLLARY 2.4 (Compactness). *Given $\eta > 0$, there exists $\epsilon > 0$ such that*

$$\mathbb{P}^N(D([0, T], K_\epsilon)) \geq 1 - \eta.$$

Now let f, g, h be bounded measurable functions on $(0, \infty)$ no greater than 1. By lemma 2.3 we have for all $s < t$

$$\begin{aligned} & \mathbb{E} \left[\int_s^t \|(f \oplus g \oplus h, \mathbf{H}^{(N)}(\Theta_s^N))\| ds \right] \\ & \leq 2C^2(t-s) + 2 \int_s^t \mathbb{E} \left[\frac{3C^2}{2} + 2C(\varphi, \sigma_r^{+,N} + \sigma_r^{-,N}) \right] dr \\ & \leq C_2(t-s) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E} [\langle M^{f,g,h;N} \rangle_t - \langle M^{f,g,h;N} \rangle_s] \\ & \leq \frac{1}{N} \mathbb{E} \left[\int_s^t \|(f \oplus g \oplus h, \mathbf{Q}^{(N)}(\Theta_r^N))\| ds \right] \\ & \leq \frac{4C^2}{N} + \frac{1}{N} \int_s^t 2\mathbb{E} \left[\frac{C^2 + 4C^2}{2} + 4C(\varphi, \sigma_r^{+,N}) + C(\varphi, \sigma_r^{-,N}) \right] dr \\ & \leq \frac{C_2}{N}(t-s), \end{aligned}$$

where C_2 is a positive constant depending only on C . So, by Doob's \mathbb{L}^2 -inequality, we have

$$\mathbb{E} \left[\sup_{s \leq r \leq t} \|(f \oplus g \oplus h, \Theta_r^N - \Theta_s^N)\|^2 \right] \leq C_3 \left((t-s)^2 + \frac{t-s}{N} \right) \quad (2.9)$$

for some positive constant C_3 depending only on C . It is then a standard fact that the equicontinuity inequality (2.9) together with corollary on compactness enable the use of Jakubowski's criterion⁷; so the sequence of laws of Θ^N in $\mathcal{D}([0, T], (\mathcal{M}^{\oplus 3}, d))$ has a convergent subsequence. Denote by $\Theta^\infty = \mu^\infty \oplus \sigma^{+, \infty} \oplus \sigma^{-, \infty}$ any limit point. Taking a subsequence and changing the probability space if necessary we can suppose without loss of generality that Θ^N converges almost surely to Θ^∞ in $\mathcal{D}([0, T], (\mathcal{M}^{\oplus 3}, d))$. As Θ^N makes jumps of size at most $\frac{3}{N}$, in the total variation distance, the limit process is a continuous process from $[0, T]$ to $(\mathcal{M}^{\oplus 3}, d)$.

It is proved in [4] that under conditions (2.8) and (2.4) the process μ^∞ is almost surely equal to the unique strong solution μ of Smoluchowski equation, and that we have almost surely $\sup_{s \leq t} d_0(\varphi \mu_s^N, \varphi \mu_s) \rightarrow 0$, as N goes to ∞ .

To prove that $\sigma^{+, \infty} - \sigma^{-, \infty}$ is equal to the unique solution of equation (1.5) it suffices to prove that it satisfies this equation for any bounded measurable function g

⁷See for instance Dawson's lecture notes [13].

with compact support, as a straightforward limit argument will give it for any bounded measurable function. We shall suppose without loss of generality that $\sigma_0^{+,N} - \sigma_0^{-,N} = 0$. We shall adopt the notation

$$\sigma_s^N := \sigma_s^{+,N} - \sigma_s^{-,N}, \quad |\sigma_s^N| := \sigma_s^{+,N} + \sigma_s^{-,N}$$

and

$$\sigma_s^\infty := \sigma_s^{+,\infty} - \sigma_s^{-,\infty}, \quad |\sigma_s^\infty| := \sigma_s^{+,\infty} + \sigma_s^{-,\infty}.$$

The conclusion of lemma 2.3 can now be re-written as $\mathbb{E} \left[\sup_{0 \leq t \leq T} (\varphi, |\sigma_t^N|) \right] \leq C_1$.

It can be seen from expression (2.3) for $\mathbf{H}^{(N)}$ that the real-valued process

$$\begin{aligned} B_t^{g;N} &= (g, \sigma_t^N) - \int_0^t \left(\int \frac{1}{2} \{g(x+x') - g(x') - g(x)\} K'(x, x') \tilde{\mu}_s^N(dx, dx') \right. \\ &\quad \left. + \int \{g(x+y) - g(y) - g(x)\} K(x, y) \mu_s^N(dx) \sigma_s^N(dy) \right) ds \end{aligned} \quad (2.10)$$

is a martingale with previsible increasing process

$$\begin{aligned} \langle B^{g;N} \rangle_t &= \frac{1}{N} \int_0^t \left(\int \frac{1}{2} \{g(x+x') - g(x') - g(x)\}^2 K'(x, x') \tilde{\mu}_s^N(dx, dx') \right. \\ &\quad \left. + \int \{g(x+y) - g(y) - g(x)\}^2 K(x, y) \mu_s^N(dx) \sigma_s^N(dy) \right) ds \end{aligned}$$

Using lemma 2.3 together with the almost sure inequality $(\varphi, \mu_s^N) \leq C$, it is seen that $\mathbb{E} \langle B^{g;N} \rangle_T$ converges to 0 as N goes to ∞ . So, to show that σ^∞ satisfies equation (1.5), it is sufficient to prove that the two integrals inside the right hand side of equation (2.10) converge almost surely to

$$\int \frac{1}{2} \{g(x+x') - g(x') - g(x)\} K'(x, x') \mu_s(dx) \mu_s(dx')$$

and

$$\int \{g(x+y) - g(y) - g(x)\} K(x, y) \mu_s(dx) \sigma_s^\infty(dy) \quad (2.11)$$

respectively, and that we have uniform bounds on them so that dominated convergence under the time integral can be used. The convergence of the first integral was proved in [4] using hypotheses (2.6) and (2.7), with K in place of K' ; the same argument applies here. This integral is bounded above by $\frac{3}{2} \|g\|_\infty C^2$, uniformly in $s \in [0, T]$ and $N \geq 1$.

Given $\delta \in (0, \infty]$, the function $\varphi^\delta(x) = \varphi(x) \mathbf{1}_{x \leq \delta}$ is subadditive. It comes from Fatou's lemma that the inequality

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} (\varphi^\delta, |\sigma_t^\infty|) \right] \leq C_1$$

holds for any $\delta \in (0, \infty]$. So, to any $\omega \in \Omega$ one can associate a positive constant $m(\delta; \omega)$ such that we have

$$(\varphi^\delta, |\sigma_t^\infty(\omega)|) \leq (\varphi^\delta, |\sigma_T^\infty(\omega)|) \leq m(\delta; \omega)$$

on the time interval $[0, T]$. One can choose this constant $m(\delta; \omega)$ so that it converges to 0 as δ decreases to 0. Taking ω in a subset Ω_1 of Ω of probability 1, for which $\Theta^N(\omega)$ converges to $\Theta^\infty(\omega)$ in $\mathcal{D}([0, T], (\mathcal{M}^{\oplus 3}, d))$, we get that

$$(\varphi^\delta, |\sigma_t^N(\omega)|) \leq (\varphi^\delta, |\sigma_T^N(\omega)|)$$

is arbitrarily small provided δ is small enough, and bounded above uniformly in $t \in [0, T]$, $N \geq 1$ and $\delta \in (0, \infty]$.

Proceed now as in [4] and write K as the sum of a kernel K_1 with compact support and a kernel K_2 with support in

$$F_1 \cup F_2 \cup F_3 := \{(x, y); x \leq \delta\} \cup \{(x, y); y \leq \delta\} \cup \{(x, y); \max\{x, y\} \geq \frac{1}{\delta}\}.$$

There is no problem in justifying the convergence of the integral in (2.11) corresponding to K_1 . For K_2 write, with $\{g\}(x, y) := g(x + y) - g(x) - g(y)$,

$$\begin{aligned} & \left| \int \{g\}(x, y) K_2(x, y) \left(\mu_s^N(dx) \sigma_s^N(dy) - \mu_s(dx) \sigma_s^\infty(dy) \right) \right| \\ & \leq \left| \int \{g\}(x, y) K_2(x, y) (\mu_s^N - \mu_s^\infty)(dx) \sigma_s^N(dy) \right| \\ & \quad + \left| \int \{g\}(x, y) K_2(x, y) \mu_s(dx) (\sigma_s^N - \sigma_s^\infty)(dy) \right| \end{aligned}$$

and deal with each term of the right hand side separately. The first term is bounded above by $d_0(\varphi \mu_s^N(\omega), \varphi \mu_s)(\varphi, |\sigma_s^N(\omega)|)$, up to a multiplicative constant. As the first factor converges to 0 (and is no greater than $2C$) while the second is uniformly bounded above, one can apply dominated convergence in the corresponding integral with respect to s . To deal with the second term, use the pointwise bounds⁸

$$\begin{aligned} \|K_2 \mathbf{1}_{F_1} \mu_s \oplus \sigma_s^N(\omega)\|_0 & \leq \gamma_\delta C(\varphi, |\sigma_s^N(\omega)|), \\ \|K_2 \mathbf{1}_{F_2} \mu_s \oplus \sigma_s^N(\omega)\|_0 & \leq C(\varphi^\delta, |\sigma_s^N(\omega)|), \\ \|K_2 \mathbf{1}_{F_3} \mu_s \oplus \sigma_s^N(\omega)\|_0 & \leq (\varphi^\delta, \mu_s)(\varphi, |\sigma_s^N(\omega)|), \end{aligned}$$

where $\gamma_\delta = \max\left\{\frac{K(x, y)}{\varphi(x)\varphi(y)}; (x, y) \in F_3\right\}$ converges to 0 as δ decreases to 0. As $(\varphi, |\sigma_s^N(\omega)|)$ is uniformly bounded above by a constant, and both $(\varphi^\delta, |\sigma_s^N(\omega)|)$ and (φ^δ, μ_s) can be made arbitrarily small for small enough δ , we have enough control to apply dominated convergence. \square

3. Algorithm. We describe in this section the algorithm used to simulate the particle system studied above; the numerical results are to be found in section 4. Two points of computational interest are first put forward in sections 3.1 and 3.2; the algorithm itself is described in section 3.3.

3.1. Coupling. The basic algorithm to simulate the sensitivity σ_t is given by the dynamics of the process Θ^N described in section 2.1. A fresh look at it reveals a potential computational drawback of this approach: It is seen from the explicit expression (2.3) of the generator of Θ^N that the mean number of particles inside σ^N satisfies a Grönwall-type inequality, which implies an exponential growth of this

⁸ $\|\cdot\|_0$ denotes total variation norm.

quantity. One should see in this exponential growth of the number of particles a good feature for the approximation qualities of our estimator σ_t^N of σ_t , especially regarding accuracy and variance. This should be opposed to what happens for the weighed and coupled particles systems described in the introduction, for which the number of particles in the system decreases with time⁹.

As an exponential growth of the quantity of information to consider is non-desirable for simulations, three kinds of tricks are used in order to reduce it.

(i) **Cancellation.** As we are only interested in the difference $\sigma_t^{+,N} - \sigma_t^{-,N}$ any particle which appears in both particle systems will be removed from both of them.

(ii) **Coupling.** A particle δ_x of μ^N coagulates with any particle of $\sigma_t^{+,N}$ at rate $\frac{1}{N}K(x, \sigma_t^{+,N}) = \frac{1}{N} \int K(x, y) \sigma_t^{+,N}(dy)$; it also coagulates with any particle of $\sigma_t^{-,N}$ at rate $\frac{1}{N}K(x, \sigma_t^{-,N})$. This particle is thus used in both systems at rate $\frac{1}{N}K(x, \sigma_t^{+,N}) \wedge K(x, \sigma_t^{-,N})$, in which case a cancellation removes the particles δ_x added to $\sigma_t^{-,N}$ and $\sigma_t^{+,N}$. This operation leaves the total number of particles in σ^N constant. The rest of the time δ_x is used in only one of the systems.

(iii) **Re-sampling.** A more drastic control of the number of particles in σ^N can be obtained using re-sampling. Let M and m be two integers depending on N , with $m \leq M$. Each time $\sigma_t^{+,N}$ or $\sigma_t^{-,N}$ has M particles, replace it by an iid sample of itself of size m ; this way the total number of particles in σ^N remains no greater than $2M$.

3.2. Majorant kernel. In order to treat information in a computationally efficient way, we have organized the data using tree structures. The use of a majorant kernel with a simple algebraic structure together with an acceptance/rejection step lead to an efficient updating of the data tree.

The choice of a majorant kernel $\widehat{K}(\cdot, \cdot)$ is made so that \widehat{K} is symmetric, no less than K and has the form

$$\widehat{K}(x_i, x_j) = \sum_{\beta} \widehat{K}_{\beta}(x_i, x_j) := \sum_{\beta} f_{\beta}(x_i) g_{\beta}(x_j) \quad (3.1)$$

for β in a finite set of indices [14]. This form of kernel leads to simple generation of probabilities of the form

$$\frac{\widehat{K}(x_i, x_j)}{\sum_{a \neq b} \widehat{K}(x_a, x_b)} = \sum_{\beta} \frac{\sum_{a \neq b} f_{\beta}(x_a) g_{\beta}(x_b)}{\sum_{a \neq b} \sum_{\beta'} f_{\beta'}(x_a) g_{\beta'}(x_b)} \frac{f_{\beta}(x_i)}{\sum_a f_{\beta}(x_a)} \frac{g_{\beta}(x_j)}{\sum_{b; b \neq a} g_{\beta}(x_b)}, \quad (3.2)$$

where a and b run in possibly different finite sets of indices. Identity (3.2) corresponds to choosing first an index β according to the probability specified by the first term of the right hand side and then choosing each particle x_i, x_j *separately*. The choice of a pair (x_i, x_j) according to the probability given the left hand side of formula (3.2) can thus be done in $O(N)$ operations rather than $O(N^2)$. All the required information can be held in binary tree structures (as described in [15]) whilst allowing an even

⁹This decrease is of the same order for the weighted particle system and for Marcus-Lushnikov's dynamics; it is worse for the coupled system. In this approach, σ_t is approximated by the ratio $(\mu_t^{\lambda + \frac{1}{2}\delta\lambda; N} - \mu_t^{\lambda - \frac{1}{2}\delta\lambda; N})/\delta\lambda$, where $\mu_t^{\lambda + \frac{1}{2}\delta\lambda; N}$ and $\mu_t^{\lambda - \frac{1}{2}\delta\lambda; N}$ are two coupled Markus-Lushnikov processes. So, the smaller $\delta\lambda$ is, the more $\mu_t^{\lambda + \frac{1}{2}\delta\lambda}$ and $\mu_t^{\lambda - \frac{1}{2}\delta\lambda}$ (and $\mu_t^{\lambda + \frac{1}{2}\delta\lambda; N}$ and $\mu_t^{\lambda - \frac{1}{2}\delta\lambda; N}$ with it) look the same. This means that the 'real' number of particles in the difference $\mu_t^{\lambda + \frac{1}{2}\delta\lambda; N} - \mu_t^{\lambda - \frac{1}{2}\delta\lambda; N}$ is a 'function' $f_{\delta\lambda}(N) \leq N$ of $\delta\lambda$ that decreases as $\delta\lambda$ goes to 0, a necessary condition for the ratio to be a good estimate of σ_t .

further reduction in the number of operations to choose each particle from $O(N)$ to $O(\log N)$. Updating this information also requires $O(\log N)$ operations. Further, the sums in the first fractions of the right hand side of (3.2) are automatically contained in the tree structure without further computation.

Note that in the theoretical framework used in section 2, the function $\varphi(x)\varphi(y)$ can be used as a unique majorant kernel. We have yet chosen to present the above general procedure as we shall consider situations in which the above theory does not apply directly.

3.3. Algorithm description. Recall Θ_t^N is of the form $\left(\frac{1}{N}X_t, \frac{1}{N}Y_t, \frac{1}{N}Z_t\right)$ for a Markov process $\Theta_t = (X_t, Y_t, Z_t)$ whose components are sums of Dirac masses and whose dynamics was described in section 2.1. What the algorithm really simulates is the discrete measure-valued process Θ_t ; a rescaling gives the time evolution of Θ_t^N . The algorithm is described in Algorithms 1 and 2 below.

Note that there may be up to three different majorant kernels — for K , K'_+ and K'_- . Therefore, we slice up the total majorant rates according to the event type $\alpha \in \{0, 1^+, 1^-, 2^+, 2^-\}$ to occur. We then have $\widehat{K}_{\alpha\beta}$ such that $\sum_{\beta} \widehat{K}_{\alpha\beta} = \widehat{K}_{\alpha}$ (from eq. 3.1), where $\widehat{K}_{\alpha} \in \{\widehat{K}, \widehat{K}'_+, \widehat{K}'_-\}$. This gives the corresponding rates $\widehat{\rho}_{\alpha\beta}$ and $\widehat{\rho}_{\alpha}$.

To order to describe numerical results it provides, we shall denote by L the number of simulations with the same initial conditions and by t_{run} the computational time taken to run the algorithm (CPU time in seconds).

4. Numerical Results. We have chosen to illustrate our approach in situations where the theoretical results of section 2 do not apply, so as to show its robustness. The main motivation of this article is to produce a stochastic estimate of the sensitivity σ_t whose variance is smaller than that given by existing methods. One step in this direction was done in [9], where σ_t was approximated by the ratio $(\mu_t^{\lambda+\frac{1}{2}\delta\lambda; N} - \mu_t^{\lambda-\frac{1}{2}\delta\lambda; N})/\delta\lambda$, for two Marcus-Lushnikov processes with slightly different parameters. The method there called for coupling them so as to reduce the variance of this estimator as much as can be done; this was done in the same spirit as the coupling used above. We shall refer to this algorithm as the **CD** algorithm (for central difference). The variance reduction obtained by this method is significant; we shall thus compare our results with those given by the CD algorithm. As our algorithm simulates σ_t directly, it will be called **Exact**; and depending on whether or not we use the coupling step we shall talk of the **ExactCoupling** or **ExactIndep** algorithm.

The data presented deal with the additive kernel $K(x, y) = \lambda(x + y)$ and a kernel that is used in modelling soot formation in a free molecular regime [16, 17, 18], thus we shall call it the ‘Soot Kernel’:

$$K(x, y) = \left(\frac{1}{x} + \frac{1}{y}\right)^{\frac{1}{2}} \left(x^{\frac{1}{\lambda}} + y^{\frac{1}{\lambda}}\right)^2;$$

both are considered in the *discrete setting where masses are integers*. The reference value of λ for the additive kernel will be 1 and for the soot kernel 2.1. We shall always take as initial condition for the Marcus-Lushnikov process N particles with mass equal to 1, and $\sigma_0^{+,N} = \sigma_0^{-,N} = 0$.

Smoluchowski equation has an explicit analytic solution for an additive interaction (see the review by Aldous [19] for instance) we can compare our results with it; it will be convenient to write σ_t^∞ for σ_t in this case. No analytic solution of Smoluchowski equation or its sensitivity equation is available for the soot kernel; we shall thus compare our estimators σ_t^N with what the ExactCoupling algorithm gives us for very

Algorithm 1: The ExactCoupling algorithm - Part 1

- 1 Set $t = 0$. **while** $t < t_{end}$ **do**
- 2 Generate a realisation of the holding time Δt with exponential law of parameter $\frac{1}{N} \sum_\alpha \widehat{\rho}_\alpha$, and set $t \leftarrow t + \Delta t$.
- 3 Choose event type $\alpha \in \{0, 1^+, 1^-, 2^+, 2^-\}$ to occur with distribution $\frac{\widehat{\rho}_\alpha}{\sum_\alpha \widehat{\rho}_\alpha}$.
- 4 Choose process β with distribution $\frac{\widehat{\rho}_{\alpha\beta}}{\widehat{\rho}_\alpha}$.
- 5 Given α and β , choose a pair of particles using the **index distribution**

$$\frac{\widehat{K}_{\alpha\beta}(x_i, x_j)}{\widehat{\rho}_{\alpha\beta}} = \frac{f_{\alpha\beta}(x_i)}{\sum_a f_{\alpha\beta}(x_a)} \frac{g_{\alpha\beta}(x_j)}{\sum_b g_{\alpha\beta}(x_b)} \quad (3.3)$$

where (x_i, x_j) are the masses of particles sampled from the appropriate ensembles $(\mu^N, \sigma^{+,N}$ or $\sigma^{-,N})$ depending on α .

- 6 Perform the coagulation step which depends on α :
switch the value of α chosen do
- 7 **case** $\alpha = 0$; *this part is the original Marcus-Lushnikov process.*
 The chosen pair of particles is of the form (x_i, x_j) .
 With probability $\frac{K_\alpha}{\widehat{K}_\alpha}$ make the jump
 $\Delta\Theta^N = (\delta_{x_i+x_j} - \delta_{x_i} - \delta_{x_j}) \oplus 0 \oplus 0$.
- 9 **case** $\alpha = 1^+$ or 1^-
 The chosen pair of particles is of the form (x_i, x_j) . Set
 $p = \frac{\max\{K_{2^+}, K_{2^-}\}}{K_{2^+} + K_{2^-}}$, and generate a realisation of a uniform
 random variable \mathbf{U} in $(0, 1)$.
 if $\mathbf{U} \leq p$ **then**
 if $K_{2^+} > K_{2^-}$ **then**
 make the jump $\Delta\Theta^N = 0 \oplus \delta_{x_i+x_j} \oplus \delta_{x_i} + \delta_{x_j}$.
 else
 make the jump $\Delta\Theta^N = 0 \oplus \delta_{x_i} + \delta_{x_j} \oplus \delta_{x_i+x_j}$.
 else Go to Step 15.
- 14 **case** $\alpha = 2^+$ or 2^- ; Go to Algorithm 2.
- 15 For **each** particle of σ^N that has just been involved in a coagulation or newly formed, do a cancellation operation if it can be done.

16 STOP.

Algorithm 2: The ExactCoupling algorithm - Part 2 (Cases $\alpha = 2^+, 2^-$ only)

1 **case** $\alpha = 2^+$ or $\alpha = 2^-$
 The chosen ordered pair of particles contains one particle of μ^N and one particle of σ^N , in either order.
if the pair is of the form (x_i, \cdot) where x_i is the mass of a particle from μ^N **then**
 2 **if** the second particle belongs to $\sigma^{+,N}$ **then**
 Choose a particle of $\sigma^{-,N}$ according to the distribution

$$\frac{g_{\alpha\beta}(\cdot)}{\sum_{\ell \in [1, \dots, q]} g_{\alpha\beta}(z_\ell)}. \quad (3.4)$$

else
 Choose a particle of $\sigma^{+,N}$ according to the distribution

$$\frac{g_{\alpha\beta}(\cdot)}{\sum_{k \in [1, \dots, p]} g_{\alpha\beta}(y_k)}. \quad (3.5)$$

 3 Set

$$r_+ := \sum_{k \in [1, \dots, p]} g_{\alpha\beta}(y_k) \quad , \quad r_- := \sum_{\ell \in [1, \dots, q]} g_{\alpha\beta}(z_\ell) \quad (3.6)$$

else
 4 \lfloor Do the symmetrical operation, swapping $g_{\alpha\beta}$ with $f_{\alpha\beta}$.
 5 The preceding steps produce a triple (x_i, y_k, z_ℓ) of particles from $\mu^N \oplus \sigma^{+,N} \oplus \sigma^{-,N}$. Set

$$p_{\min} = \frac{\min\{r_+, r_-\} K}{r_+ + r_- \widehat{K}} \quad , \quad p_{\max} = \frac{\max\{r_+, r_-\} K}{r_+ + r_- \widehat{K}}. \quad (3.7)$$

 6 Generate realisation of a uniform random variable U in $(0, 1)$.
if $0 < U \leq p_{\min}$ **then**
 7 make the jump

$$\Delta\Theta^N = 0 \oplus (\delta_{x_i+y_k} - \delta_{y_k}) \oplus (\delta_{x_i+z_\ell} - \delta_{z_\ell}).$$

else if $p_{\min} < U \leq p_{\max}$ **then**
if $r_+ > r_-$ **then**
 8 \lfloor make the jump $\Delta\Theta^N = 0 \oplus (\delta_{x_i+y_k} - \delta_{y_k}) \oplus \delta_{x_i}$.
else
 9 \lfloor make the jump $\Delta\Theta^N = 0 \oplus \delta_{x_i} \oplus (\delta_{x_i+z_\ell} - \delta_{z_\ell})$.
else
 10 \lfloor Go to Step 15 of Algorithm 1.
 11 Go to Step 15 of Algorithm 1.

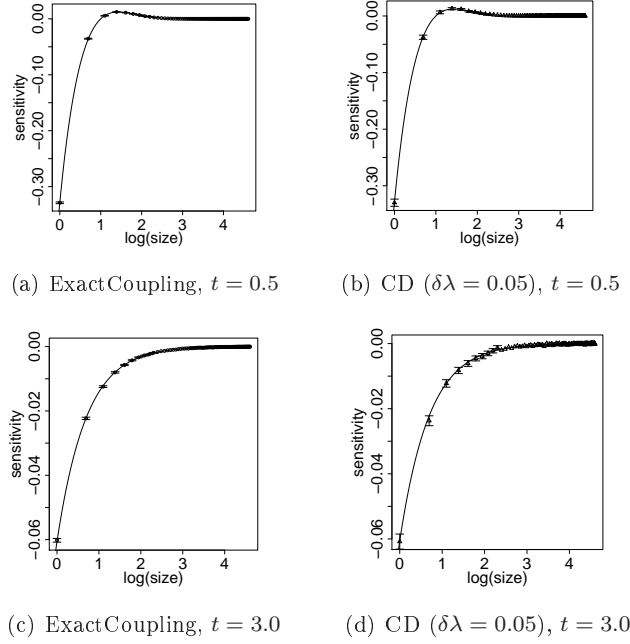


FIGURE 4.1. Sensitivity for additive kernel, $\lambda = 1.0$, $N = 10^3$, $L = 1000$. The confidence intervals for the larger particle sizes are omitted for pictorial clarity.

high settings, say $N = 3 \times 10^6$ and $L = 10^3$ simulations. Given any N , the l^{th} run of the algorithm produces an estimator of σ_t which we shall denote by $\sigma_t^{l,N}$. We shall set $\sigma_t^\infty := 10^{-3} \sum_{l=1, \dots, 10^3} \sigma_t^{l,10^6}$. Figures 4.1 and 4.2 show the empirical estimate of σ_t

given after L runs, at different times. The line represents σ_t^∞ . For comparison, the results given by the CD algorithm for the same setting, with $\delta\lambda = 0.05$, are plotted using stars. Also, Figure 4.3 shows what the solution to the original Smoluchowski equation looks like.

To quantify the convergence of the empirical sensitivity

$$\bar{\sigma}_t^{L;N} := \frac{1}{L} \sum_{l=1..L} \sigma_t^{l,N}$$

to σ_t^∞ as N increases we have plotted in Figure 4.4 the quantity

$$d_{\text{var}}(N) = \sum_j \sum_{i \geq 1} \left| (\bar{\sigma}_{t_j}^{L;N} - \sigma_{t_j}^\infty)(i) \right|,$$

where $\bar{\sigma}_t^{L;N}(i)$ and $\sigma_t^\infty(i)$ represent the empirical and real sensitivities at particle mass $i \in \mathbb{N}$ respectively, and $d_{\text{var}}(N)$ represents the total variation distance between the empirical sensitivity and the sensitivity itself summed over some chosen time points¹⁰ $\{t_j\}$. These results empirically confirm Theorem 2.2 (in this case where it does not apply), and quantify the speed of convergence as being of order $\frac{1}{N}$. The analogue result for the CD algorithm is given in [9].

¹⁰For Figure 4.4, the times points $\{t_j\}$ were chosen to be $0.125j$ for $j = 1, \dots, 56$

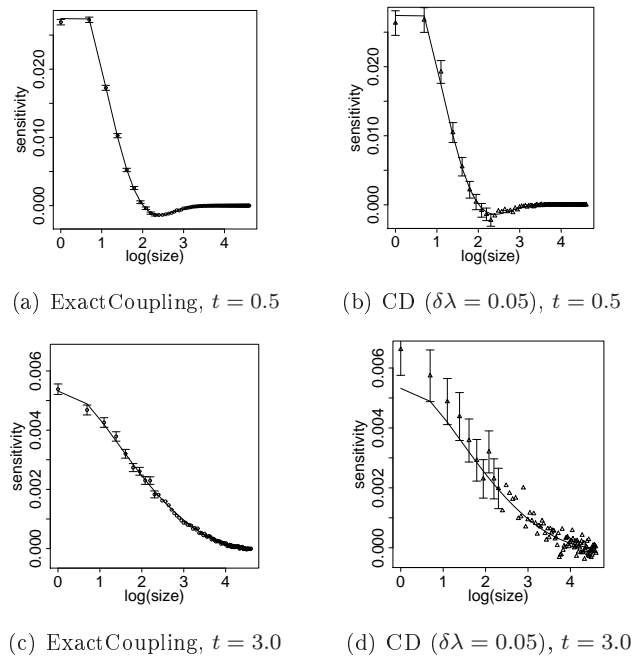


FIGURE 4.2. Sensitivity for soot kernel, $\lambda = 2.1$, $N = 10^3$, $L = 1000$. The confidence intervals for the larger particle sizes are omitted for pictorial clarity.

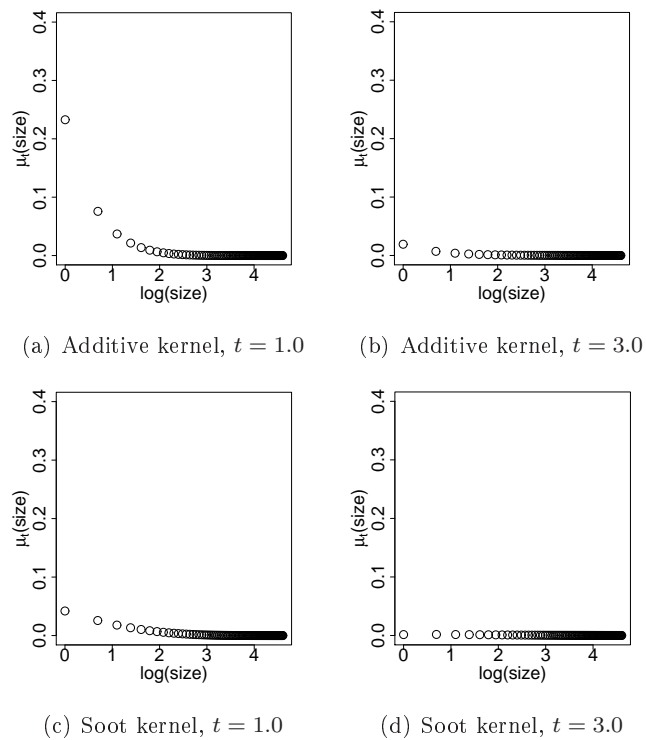


FIGURE 4.3. μ_t as a function of $\log(\text{particle size})$

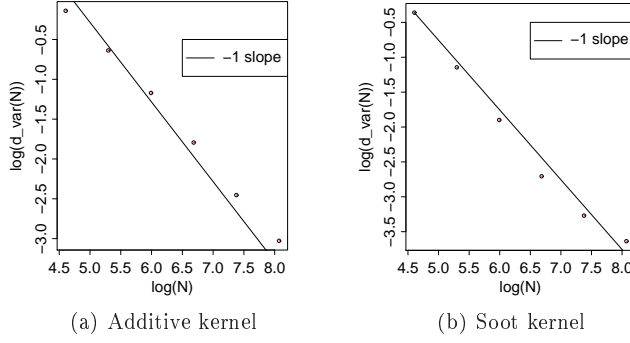


FIGURE 4.4. Convergence in N of the *ExactCoupling* algorithm, $N = 100 \times 2^i$ for $i = 0, \dots, 5$, $NL = 2 \times 10^8$.

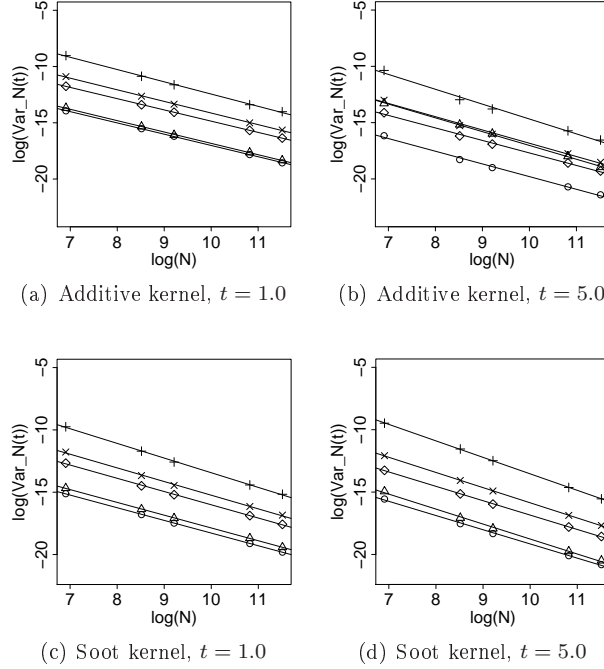


FIGURE 4.5. $\log \text{Var}_N(t)$ as a function of N . The meaning of the symbols are as follows: Circles = *ExactCoupling*, Diamonds = *ExactIndep*, Triangles = *CD*($\delta\lambda = 0.10$), Crosses = *CD*($\delta\lambda = 0.05$), Pluses = *CD*($\delta\lambda = 0.01$).

4.1. Variance. To analyse the variance of the random output of the algorithm we shall define the empirical variance at particle mass $i \in \mathbb{N}$ and time t as

$$\text{Var}_N(i, t) := \frac{1}{L-1} \sum_{l=1}^L ((\sigma_t^{l,N} - \bar{\sigma}_t^{L;N})(i))^2$$

and shall take as a measure of the variance the quantity

$$\text{Var}_N(t) := \sum_{i \geq 1} \text{Var}_N(i, t). \quad (4.1)$$

Figure 4.5 represents its graph as a function of N using different algorithms. It shows that the ExactCoupling algorithm achieves a variance reduction by a factor 10^3 compared to the CD algorithm. The plots also show that $\text{Var}_N(t)$ is proportional to $\frac{1}{N}$, a fact that should be related to a central limit theorem.

4.2. Computational efficiency. Although section 4.1 indicates that the ExactCoupling algorithm produces very accurate estimators of the sensitivity, it comes at the price of a computational time greater than the one needed by the CD algorithm. This comes from the fact that the latter algorithm being essentially a Marcus-Lushnikov algorithm, it uses a generally decreasing amount of information, as the number of sensitivity particles decreases with time. On the other hand, the ExactCoupling algorithm has to deal with more and more sensitivity particles, whose number tends to grow exponentially. To see whether the gain of accuracy given by the ExactCoupling algorithm is worth the effort we propose two criteria.

4.2.1. CPU time to reach a certain level of accuracy. Fix the observation time t (we choose large enough t so that the particle system has experienced many jumps, and therefore the variances are expected to be larger - see Figure 4.3). Given a certain level of accuracy, v , find for each algorithm the smallest N for which $\text{Var}_N(t)$ is smaller than v . See what computational time is needed to run the algorithm for this N (during an evolution time t for the particle system). Tables 4.1 and 4.2 show that the ExactCoupling algorithm remains mostly better than the CD algorithm. It also shows that it converges much quicker to the true sensitivity than the CD algorithm does. Note that for the soot kernel the CD algorithm with $\delta\lambda = 0.1$, 10^5 initial particles are not sufficient to reach the given level of accuracy; this setup already requires a CPU time equal to 1058.91 seconds. The comparison with the corresponding time for the ExactCoupling algorithm is greatly in favour of the latter.

TABLE 4.1
Additive kernel, $v = 1.43 \times 10^{-4}$

t	1.0	1.0	3.0	3.0
algorithm	ExactCoupling	CD ($\delta\lambda = 0.10$)	ExactCoupling	CD ($\delta\lambda = 0.10$)
N	6500	55000	2100	16250
t_{run} (secs)	281.15	593.99	99.22	213.34

TABLE 4.2
Soot kernel, $v = 2.57 \times 10^{-5}$

t	1.0	1.0	3.0	3.0
algorithm	ExactCoupling	CD ($\delta\lambda = 0.10$)	ExactCoupling	CD ($\delta\lambda = 0.10$)
N	10000	100000	6350	55000
t_{run} (secs)	379.01	1058.91 (v not reached)	382.15	1104.24

4.2.2. Gain factor. Eibeck and Wagner introduced in [14] another quantity to compare the relative efficiency of two algorithms. Fix the observation time t . Given a setup $(K(\cdot, \cdot), N, L)$, denote by $T^{\text{EC}}(t)$ and $T^{\text{CD}}(t)$ the empirical mean CPU time needed by the ExactCoupling and CD algorithms to be run up to time t . Denote also by $\text{Var}_N^{\text{EC}}(t)$ and $\text{Var}_N^{\text{alg}}(t)$ the empirical variances given by formula (4.1) when computed using ExactCoupling and the given algorithm ‘alg’ respectively. The *gain factor* of an algorithm *over ExactCoupling*, similar to that as introduced by Eibeck and Wagner, is defined here by the ratio

$$\frac{T^{\text{EC}}(t) \text{Var}_N^{\text{EC}}(t)}{T^{\text{alg}}(t) \text{Var}_N^{\text{alg}}(t)}$$

It is related in some way to the analysis made in section 4.2.1. See section 5 of [14]. Figures 4.6 and 4.7 plot the reciprocal gain (its logarithm) as a function of time. Triangles, pluses and crosses represent data of the CD algorithm, for $\delta\lambda = 0.01, 0.05$ and 0.10 respectively, circles represent data of the ExactIndep algorithm, and the horizontal line at zero represents the threshold for ExactCoupling.

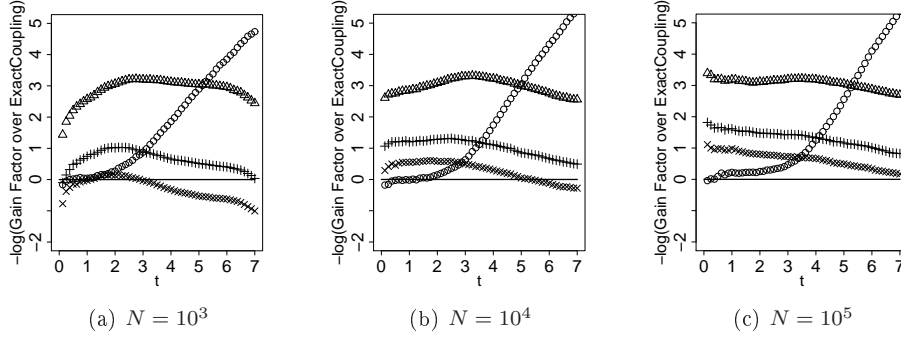


FIGURE 4.6. Additive kernel: $\log(\text{Gain factor}^{-1})$ as a function of t . The meanings of the symbols are as follows: Circles = ExactIndep, Crosses = $CD(\delta\lambda = 0.10)$, Pluses = $CD(\delta\lambda = 0.05)$, Triangles = $CD(\delta\lambda = 0.01)$. The horizontal line is the threshold value 1.0 for ExactCoupling.

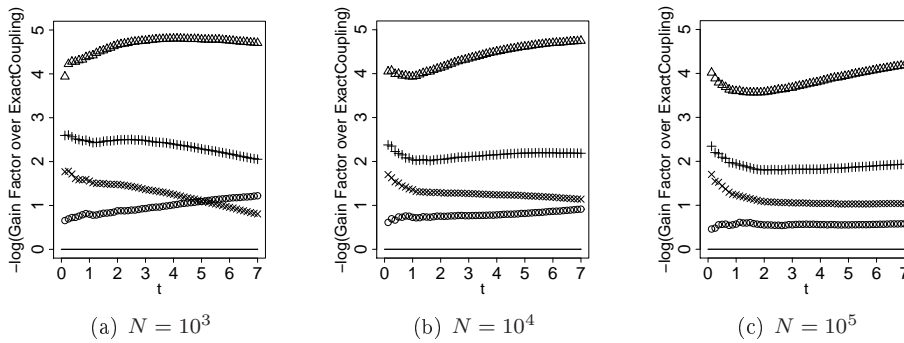


FIGURE 4.7. Soot kernel: $\log(\text{Gain factor}^{-1})$ as a function of t . The meanings of the symbols are as follows: Circles = ExactIndep, Crosses = $CD(\delta\lambda = 0.10)$, Pluses = $CD(\delta\lambda = 0.05)$, Triangles = $CD(\delta\lambda = 0.01)$. The horizontal line is the threshold value 1.0 for ExactCoupling.

Figures 4.6 and 4.7 show good results. By and large, the CD algorithms appear to be considerably inferior to the ExactCoupling algorithm for the Soot kernel, and the ExactIndep algorithm in either performs slightly better than the CD ($\delta\lambda = 0.10$). There appears to be little to moderate difference in behaviour over different values of N .

The picture is different for the Additive kernel. For $N = 1000$, we find that the CD ($\delta\lambda = 0.1$) is better than the ExactCoupling, at least for very small or large times. This disadvantage gradually disappears over larger N — this is due to the increased probability of cancellations for larger N which reduces the number of particles in the ensembles and therefore the CPU times. Other than this, the ExactCoupling algorithm maintains a substantial lead over the other algorithms.

5. Conclusions. A stochastic particle system approximation to the parametric sensitivity in Smoluchowski’s coagulation equation was introduced. Rather than taking a finite difference approach to calculating sensitivities, we considered the direct parametric derivative of (1.2), and developed a Monte-Carlo algorithm which would approximate its solution. The particle system approximation was proved to converge weakly to the solution of the sensitivity equation (1.2), as the number of particles increases indefinitely.

The first algorithm developed (ExactIndep) allows for an exponential increase in the number of sensitivity particles. We sought to reduce this increase using several tricks: Cancellation removes ‘unnecessary’ sensitivity particles which are needed to describe it, whilst coupling prevents their creation. These make a significant reduction to the number of particles in the ensemble. Furthermore, the resampling method puts a cap on the total number of sensitivity particles, thus stopping their exponential escalation. This gives us the ExactCoupling algorithm.

In the Numerical Results section, it was empirically confirmed that the order of convergence is $O(1/N)$ where N is the number of initial particles. We then compared the Exact algorithms with those found in [9], named here CD algorithms. It was shown that the variance of the sensitivity estimators were orders of magnitude smaller for the ExactCoupling algorithm than for the CD algorithms. However this came at the price of longer CPU run times. Two measures of efficiency, taking both the variance and the CPU time into account, were then considered. The ExactCoupling algorithm happens to require much smaller time to reach a fixed level of error than any CD algorithm, and the gain factor, as defined in [14], also happens to be in favour of the ExactCoupling algorithm, most of the time. This definitely gives a clear advantage of our approach over finite difference methods.

However, both methods have some inherent drawback: unlike the adjoint method [8], they are unidimensional in nature and compute sensitivity only for a fixed value of the parameter. It would be useful to construct a particle system approximation which do not have these weaknesses. Also, although the convergence theorem established in section 2 in a general framework is quite encouraging, it is not clear whether the algorithm will be as efficient as above if particles’s masses can take any positive value. We leave the investigation of these questions for future work.

REFERENCES

- [1] A. H. MARCUS, *Stochastic coalescence*, Technometrics, 10 (1968), pp. 133–143.
- [2] Y. R. DOMILOVSKIY, A. A. LUSHNIKOV AND V. N. PISKUROV, *Monte Carlo simulation of coagulation processes*, Izv. Atmos. Ocean. Phys., 15 (1979).

- [3] I. JEON, *Existence of gelling solutions for coagulation-fragmentation equations*, Comm. Math. Phys., 194 (1998), pp. 541–567.
- [4] J. R. NORRIS, *Smoluchowski's Coagulation Equation: Uniqueness, Nonuniqueness and a Hydrodynamic Limit for the Stochastic Coalescent*, Ann. Appl. Probab., 9 (1999), pp. 78–109.
- [5] I. BAILLEUL, *Sensitivity for Smoluchowski equation*, Submitted, (2009).
- [6] V. N. KOLOKOLTSOV, *The central limit theorem for the Smolukovski coagulation model*, to appear in Probab. Theory Related Fields, 2007.
- [7] A. VIKHANSKY AND M. KRAFT, *A Monte Carlo methods for identification and sensitivity analysis of coagulation processes*, J. Comput. Phys., 200 (2004), pp. 50–59.
- [8] A. VIKHANSKY AND M. KRAFT, *Two methods of sensitivity analysis of coagulation processes in population balances by a Monte Carlo method*, Chem. Eng. Sci., 61 (2006), pp. 4966–4972.
- [9] P. L. W. MAN, J. R. NORRIS, I. BAILLEUL AND M. KRAFT, *Coupling Algorithms for calculating sensitivities of Smoluchowski's Coagulation equation*, Technical Report 70, c4e-Preprint Series, Cambridge, Submitted to SIAM J. Sci. Comput. for publication, (2009).
- [10] S. CHANDRASEKHAR, *Stochastic problems in astrophysics and astronomy*, Rev. Modern Phys., 15 (1943), pp. 1–89.
- [11] P. BILLINGSLEY, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, second edition, 1999.
- [12] D. POLLARD, *Convergence of stochastic processes*, Springer Series in Statistics, Springer-Verlag, New York, 1984.
- [13] D. A. DAWSON, B. MAISONNEUVE AND J. SPENCER, *École d'Été de Probabilités de Saint-Flour XXI—1991*, Lecture Notes in Mathematics, Springer-Verlag, 1541 (1993).
- [14] A. EIBECK AND W. WAGNER, *An Efficient Stochastic Algorithm for Studying Coagulation Dynamics and Gelation Phenomena*, SIAM J. Sci. Comput., 22 (2000), pp. 802–821.
- [15] R. I. A. PATTERSON, J. SINGH, M. BALTHASAR, M. KRAFT AND J. R. NORRIS, *The Linear Process Deferment Algorithm: A new technique for solving population balance equations*, SIAM J. Sci. Comput., 28 (2006), pp. 303–320.
- [16] M. GOODSON AND M. KRAFT, *An Efficient Stochastic Algorithm for Simulating Nano-particle Dynamics*, J. Comput. Phys., 183 (2002), pp. 210–232.
- [17] A. EIBECK AND W. WAGNER, *Stochastic Particle Approximations for Smoluchoski's Coagulation Equation*, Ann. Appl. Probab., 11 (2001), pp. 1137–1165.
- [18] R. I. A. PATTERSON, J. SINGH, M. BALTHASAR, M. KRAFT AND W. WAGNER, *Extending stochastic soot simulation to higher pressures*, Combustion and Flame, 145 (2006), pp. 638–642.
- [19] D. J. ALDOUS, *Deterministic and Stochastic Models for Coalescence (Aggregation and Coagulation) : a Review of the Mean-Field Theory for Probabilists*, Bernoulli, 5 (1999), pp. 3–48.