Journées

ÉQUATIONS AUX DÉRIVÉES PARTIELLES

Aussois, 19–23 juin 2023

Marie Doumic

Moments approaches for asymptotic inverse problems of depolymerisation and fragmentation systems

RÉSEAU THÉMATIQUE AEDP DU CNRS

Moments approaches for asymptotic inverse problems of depolymerisation and fragmentation systems

Marie Doumic

Abstract

Shrinkage of large particles, either through depolymerisation (i.e. progressive shortening) or through fragmentation (breakage into smaller pieces) may be modelled by discrete equations, of Becker–Döring type, or by continuous ones. In this note, we review two kinds of inverse problems: the first is the estimation of the initial size-distribution from moments measurements in a depolymerising system, in collaboration with Philippe Moireau and inspired by experiments carried out by Human Rezaei's team; the second is the inference of fragmentation characteristics from size distribution samples, in collaboration with Miguel Escobedo and Magali Tournus, based on biological questions and experiments of Wei-Feng Xue's team.

1. Introduction

Polymers are large macromolecules formed out of small molecular units, called monomers. They are ubiquitous in nature and industry – e.g. plastics, biopolymers such as DNA, actin filaments or protein fibrils. Their shortening, either through depolymerisation (i.e., loss of monomers) or through fragmentation into smaller pieces are dynamical phenomena which appear in many applications. More specifically, the departure point of our research has been protein fibrils depolymerisation and breakage, thought to be key mechanisms for many diseases (Parkinson's, Alzheimer's, Creuzfeldt–Jakob's etc.) as well as for many functional biomolecular systems (actin filaments). The dynamic nature of the experiments, as well as their nanoscale, makes it very challenging to estimate their features, i.e. their reaction rates or size distributions, leading to an urgent need for mathematical models and estimation methods to be developed. In this note, we consider two case studies of such inverse problems, directly inspired by modern experimental setups.

The first problem is based on experiments carried out in H. Rezaei's lab (INRAE, Jouy-en-Josas, France) where the observation consists in the time evolution of a moment of the size distribution of protein polymers called PrP, responsible for Prion diseases. The moment observed may be the total polymerised mass (first moment) or the average molecular weight (second moment). The original model is a discrete depolymerisation system, based on the constant coefficient case of the Becker–Döring equations, namely the reactions

$$\mathcal{C}_i \longrightarrow \mathcal{C}_{i-1} + \mathcal{C}_1. \tag{1.1}$$

Our aim is to estimate the initial polymer size distribution from such moment observations. We first evaluate the impact of using continuous approximations of the initial discrete model to solve this inverse problem. At first order, the model is approximated by a backward transport equation, for which the inverse problem turns to be mildly ill-posed (of order $k + 1$ when used to invert the time evolution of the kth-moment of the solution). This remains true when polymerisation is also

considered [10], as in the full Becker–Döring system, though the inversion reveals more intricate due to the fact that the problem becomes nonlinear.

At second order, the asymptotic model becomes an advection-diffusion equation, where the diffusion is a corrective term, complemented with an original transparent boundary condition at $x = 0$. This approximation is more accurate, but we face an accuracy versus stability trade-off: the inverse reconstruction reveals to be severely ill-posed. Thanks to Carleman inequalities and to log-convexity estimates, we prove observability results and error estimates for a Tikhonov regularisation. We then develop a Kalman-based observer approach, which reveals very efficient for the numerical solution. This is a joint work with Philippe Moireau (Inria) [14], inspired by depolymerisation experiments carried out by Human Rezaei and collaborators (INRAE). We sketch the main results of this study in Section 2.

The second problem is based on fragmentation experiments carried out on several protein polymers by W.-F. Xue's team (Univ. of Kent, Canterbury, United Kingdom), namely reactions of the kind

$$\mathcal{C}_i \longrightarrow \mathcal{C}_{i-j} + \mathcal{C}_j. \tag{1.2}$$

The biophysical question which interested our collaborators was to estimate the (size-dependent) fragmentation *rate*, as well as the so-called fragmentation *kernel*, which characterise the stability of the polymers and the places where they are more likely to break. We have proposed several approaches based on the continuous fragmentation equation, studying and making use either of the long-term, the transient or the short-term dynamics. Error estimates in Bounded Lipshitz norm are obtained for this last approach. This is a joint work with Miguel Escobedo and Magali Tournus [12], and the project is a long-standing collaboration with Wei-Feng Xue and collaborators [5, 11, 25], that we develop in Section 3.

2. Asymptotic inverse problems for depolymerisation systems

2.1. Original inverse problem: a discrete system

At the basis of this research program lie experimental protocols which follow the time dynamics of average quantities over a size distribution of polymers. Typical measurements consist either in the total polymerised mass (e.g. through a Thioflavine T or Th.T protocol [23]) or the average molecular weight (e.g. with Static Light Scattering [2]). In a discrete setting, denoting $C_i(\tau)$ the concentration of polymers containing i monomeric units, such measurement may be modelled by

$$M_k(\tau) := \sum_{i=i_0}^{\infty} i^k C_i(\tau), \tag{2.1}$$

with $k = 1$ for the polymerised mass, $k = 2$ linked to the average weight, $i_0 \geq 1$ either a detection level or the smallest stable polymer, and the measurement of M_k is made up to a noise. The first questions asked by our biologist collaborators were: what can we identify from such measurements, and what cannot we? We have addressed this question in different experimental settings [2, 3], and then decided to focus on the question of initial-state observability and estimation, both because it is interesting in itself and because it is a prerequisite to parameter estimation: even in cases where a priori knowledge of the initial state is known, it is always partial and noisy [2], requiring stability analysis.

We assume a constant depolymerisation rate $b > 0$, so that the mass balance equation of (1.1) leads to the elementary system

$$\frac{\mathrm{d}}{\mathrm{d}\tau} C_i = b(C_{i+1} - C_i), \quad C_i(0) = C_i^0, \quad i \geq i_0 \tag{2.2}$$

2.2. Asymptotic approximations by continuous equations

A fundamental characteristics of the systems we are interested in is that the average polymer size is very large - from some hundreds to some thousands of hundreds of monomers. For this reason, we rescale the system by accelerating the time, which is an alternative of size rescaling as done in [7] for instance. We define, for a given $\varepsilon \ll 1$, $t := \varepsilon\tau \in [0, T]$, $c_i^\varepsilon(t) := C_i(\frac{t}{\varepsilon})$, and obtain the

rescaled system

$$\frac{\mathrm{d}}{\mathrm{d}t}c_i^\varepsilon = \frac{b}{\varepsilon}(c_{i+1}^\varepsilon - c_i^\varepsilon), \qquad c_i^\varepsilon(0) = c_i^0, \qquad i \geq i_0. \tag{2.3}$$

We recognise a first-order finite difference scheme for the transport equation, which leads us to define a stepwise interpolant and grid:

$$\forall\, i \geq i_0, \quad x_i^\varepsilon := \varepsilon(i - i_0), \qquad u^\varepsilon(t, x) := c_i^\varepsilon, \qquad t \geq 0,\ x \in [x_i^\varepsilon, x_{i+1}^\varepsilon). \tag{2.4}$$

A Taylor expansion leads us to the following backward transport equation as a first order approximation:

$$\begin{cases} \partial_t u_\infty - b\partial_x u_\infty = 0, & (t, x) \in [0, T] \times \mathbb{R}^+, \\ u_\infty(0, x) = u^0(x), & x \in \mathbb{R}^+, \end{cases} \tag{2.5}$$

whereas the second order approximation is given by

$$\begin{cases} \partial_t u_\infty^\varepsilon - b\partial_x u_\infty^\varepsilon - \dfrac{b\varepsilon}{2}\partial_x^2 u_\infty^\varepsilon = 0, & (t, x) \in [0, T] \times \mathbb{R}^+, \\ \partial_t u_\infty^\varepsilon(0, t) - b\partial_x u_\infty^\varepsilon(0, t) = 0, & t \in [0, T], \\ u_\infty^\varepsilon(0, x) = u^\varepsilon(0, x), & x \in \mathbb{R}^+. \end{cases} \tag{2.6}$$

The transport-diffusion equation comes from the Taylor expansion, whereas the transport boundary condition at $x = 0$ may be obtained by two considerations:

- as a boundary layer, we keep only the first order term, see e.g. [18],

- we require that the balance for the total number of polymers is preserved. This is given in the discrete setting by

$$\frac{\mathrm{d}}{\mathrm{d}t}M_0 = \frac{\mathrm{d}}{\mathrm{d}\tau}\sum_{i=i_0}^\infty C_i = -bC_{i_0}(t),$$

  and doing the asymptotic expansion on this equality, where we identify $c_{i_0}^\varepsilon$ with $u_\infty^\varepsilon(t, 0)$, leads to the transport boundary condition.

Well-posedness for (2.6) follows for instance from the Lumer–Phillips theorem applied to a well-chosen maximal accretive operator. To state error estimates, we introduce the following discrete norm

$$\|u\|_{2,\varepsilon}^2 := \sum_{i \geq i_0} \varepsilon|u(x_i^\varepsilon)|^2.$$

This norm is well defined for functions in $\mathrm{H}^1(\mathbb{R}_+^*)$ and is consistent with the $\mathrm{L}^2(\mathbb{R}_+)$-norm as ε tends to 0. Moreover, it is well defined and equal to the $\mathrm{L}^2(\mathbb{R}^+)$-norm for piecewise constant functions defined on the grid (x_i^ε), as $u^\varepsilon(t, \cdot)$ defined by (2.4) is.

We prove the following error estimates for the two approximate systems.

Proposition 1 ([14, Prop. 2.1. and 2.2.]). *Let* $u^0 \in \mathrm{H}^2(\mathbb{R}^+)$ *and* $u_\infty \in \mathrm{C}^0((0, T); \mathrm{H}^2(\mathbb{R}_+^*)) \cap \mathrm{C}^1((0, T); \mathrm{L}^2(\mathbb{R}_+))$ *solution to* (2.5), $u_\infty^\varepsilon \in \mathrm{C}^0((0, T); \mathrm{H}^2(\mathbb{R}^+)) \cap \mathrm{C}^1((0, T); \mathrm{L}^2(\mathbb{R}^+))$ *solution to* (2.6) *with* u^0 *as an initial condition. Let* $(c_i^{0,\varepsilon}) \in \ell_2(\mathbb{N})$, (c_i^ε) *solution to* (2.3) *and* u^ε *defined by* (2.4) *such that*

$$\|u^\varepsilon(0, \cdot) - u^0\|_{2,\varepsilon} \leq \varepsilon^\alpha,$$

for $\alpha > 0$. *There exists a constant* $C^{st} > 0$ *depending only on* b *such that for all* $t > 0$, *we have*

$$\|u^\varepsilon(t, \cdot) - u_\infty(t, \cdot)\|_{2,\varepsilon} \leq \varepsilon^\alpha + C^{st}\|u^0\|_{\mathrm{H}^2(0,L)}\,\varepsilon\, t,$$

and

$$\|u^\varepsilon(t, \cdot) - u_\infty^\varepsilon(t, \cdot)\|_{2,\varepsilon} \leq \varepsilon^\alpha + C^{st}\|u^0\|_{\mathrm{H}^2(0,L)}\,\varepsilon^{\frac{3}{2}}\,t^{\frac{1}{2}}.$$

The proof is obtained through the Taylor expansion, using the Cauchy–Schwarz inequality and a Gronwall lemma. For the second order approximation, we also use an energy estimate for $\partial_{xx}^2 u_\infty^\varepsilon$, which satisfies the same transport-diffusion equation as u_∞^ε complemented with an homogeneous Dirichlet boundary condition at $x = 0$.

2.3. Original and asymptotic inverse problems

The two approximations (2.5) and (2.6) may then be used to solve the original discrete inverse problem, which states: From a noisy measurement of (2.1) for $\tau \in [0, \mathcal{T}]$, how to estimate the initial distribution $C_i(0)$? We first apply the rescaling by $t = \varepsilon\tau$ and define

$$M_k^\varepsilon(t) := \varepsilon \sum_{i=i_0}^\infty (\varepsilon i)^k c_i^\varepsilon(t), \qquad (2.7)$$

so that we have the following moments dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t} M_k^\varepsilon = \begin{cases} -bc_{i_0}^\varepsilon, & \text{if } k = 0, \\ -bM_0^\varepsilon - b\varepsilon i_0 c_{i_0}^\varepsilon, & \text{if } k = 1, \\ -bk M_{k-1}^\varepsilon + O(\varepsilon^2) + \frac{k(k-1)}{2} b\varepsilon M_{k-2}^\varepsilon, & \text{if } k \geq 2. \end{cases} \qquad (2.8)$$

The term $O(\varepsilon^2)$ is formally obtained, but for a compactly supported initial condition, Proposition 1 implies that it is at most $O(\varepsilon^{3/2})$. We now restrict ourselves to such cases, namely $\varepsilon i \leq L$, or equivalently $x \in [0, L]$.

The original inverse problem may be formulated as: Invert

$$\Psi_{T,\varepsilon}^{u^0 \to M_k^\varepsilon} : \begin{vmatrix} \mathrm{L}^2(0, L) \longrightarrow \mathrm{L}^2(0, T) \\ u^0 \longmapsto M_k^\varepsilon. \end{vmatrix}$$

The system (2.8) links M_k^ε to $c_{i_0}^\varepsilon(t) = u(t, 0)$, so that writing $\Psi_{T,\varepsilon}^{u^0 \to M_k^\varepsilon} = \Psi_{T,\varepsilon}^{Tr \to M_k^\varepsilon} \circ \Psi_{T,\varepsilon}^{u^0 \to Tr}$ we decompose our problem into two steps:

1. Invert

$$\Psi_{T,\varepsilon}^{u^0 \to Tr} : \begin{vmatrix} \mathrm{L}^2(0, L) \longrightarrow \mathrm{L}^2(0, T) \\ u^0 \longmapsto u(\,\cdot\,, 0) \in \mathbb{R}) \end{vmatrix}$$

2. Invert

$$\Psi_{T,\varepsilon}^{Tr \to M_k^\varepsilon} : \begin{vmatrix} \mathrm{L}^2(0, T) \longrightarrow \mathrm{L}^2(0, T) \\ u(\,\cdot\,, 0) \longmapsto M_k^\varepsilon. \end{vmatrix}$$

In the case of the first order inverse problem, the method of characteristics implies that $u(t, 0) = u^0(bt)$ so that inverting $\Psi_{T,\varepsilon}^{u^0 \to Tr}$ is a well-posed problem if and only if $bT \geq L$. Moreover, we may neglect the term $-b\varepsilon i_0 c_{i_0}^\varepsilon$ in (2.8) and obtain

$$\frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}} M_k^\varepsilon(t) = (-b)^{k+1} k! c_{i_0}^\varepsilon(t) + O(\varepsilon) = (-b)^{k+1} k! u^0(bt) + O(\varepsilon),$$

which shows that inverting $\Psi_{T,\varepsilon}^{Tr \to M_k^\varepsilon}$ is *moderately/mildly* ill-posed of degree $k + 1$, and from which we immediately infer the following observability/stability inequality

$$\forall T \geq T_0 := \frac{L}{b}, \quad \|u^0\|_{\mathrm{L}^2(0,L)}^2 \lesssim \left\| \frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}} M_k \right\|_{\mathrm{L}^2(0,T)}^2, \qquad (2.9)$$

see [1] for more details and a full solution of the inverse problem - theoretical, numerical with the implementation of a Kalman-type sequential approach, and also applied to experimental data. Let us only mention that for $u^0 \in \mathrm{H}^s(0, L)$, $s > 0$, an observation y^δ and a measurement error

$$\|M_k^\varepsilon - y^\delta\|_{\mathrm{L}^2(0,L)} \leq \delta,$$

we obtain an optimal error estimate in the order of $\varepsilon^{\frac{s}{k+s+1}}$ - as expected given the degree of ill-posedness $k + 1$.

In the case of the second order problem, Step 1 reveals *severely* ill-posed, as linked to the time-reversal of the (infinitely smoothing) heat equation. On the contrary, the second step, given by inverting (2.8), is slightly less ill-posed, thanks to the corrective term $-b\varepsilon i_0 c_{i_0}^\varepsilon$. In the following, we thus focus on Step 1: inverting $u^0 \mapsto u(\,\cdot\,, 0)$.

2.4. Observability of the second-order inverse problem

A necessary step before solving the inverse problem is to prove that it is observable - other said, that we have a unique solution in a certain space and continuity in a certain sense. To do so, we first restrict ourselves to a bounded domain $[0, L]$, and look for solutions to

$$
\begin{cases}
\partial_t u - b\partial_x u - \dfrac{b\varepsilon}{2}\partial_x^2 u = 0, & (t, x) \in [0, T] \times [0, L], \\
\partial_t u(t, 0) - b\partial_x u(t, 0) = 0, & t \in [0, T], \\
\qquad\quad u(t, L) = 0, & t \in [0, T], \\
\qquad\quad u(0, x) = u^0(x), & x \in (0, L).
\end{cases}
\tag{2.10}
$$

The homogeneous Dirichlet boundary condition at $x = L$ comes naturally from the discrete problem, for which, if initially, for $i \geq \varepsilon^{-1}L$, we have $c_i^\varepsilon(0) = 0$, then $c_i^\varepsilon(t) = 0$ for any $t \geq 0$. In this respect, the bounded problem is thus even more faithful than the unbounded one for compactly-supported initial data. The system is solved, as for the unbounded problem (2.6), by introducing an appropriate accretive operator, and the following observability inequality is obtained.

Theorem 2 ([14, Thm. 4.5]). *Let $u^0 \in \mathrm{H}_R^1(0, L) \cap \mathrm{H}^2(0, L)$, with $\|u^0\|_{\mathrm{H}^1(0,L)} \leq M$. Let $u \in \mathrm{C}^0((0, T); \mathrm{H}_R^1(0, L)) \cap \mathrm{C}^1((0, T); \mathrm{L}^2(0, L))$ the unique solution to (2.10). Defining the increasing function $\rho : x \mapsto xe^x$, we have the following observability inequality, for constants c, C depending only on b and L :*

$$
\|u(0, \cdot)\|_{L^2(0,L)}^2 + \varepsilon|u(0, 0)|^2 \lesssim \frac{Ce^{\frac{L}{\varepsilon}}M^2}{\rho^{-1}\left(CT\dfrac{e^{-\frac{c}{\varepsilon}(1+T^{-1})}}{1+T^{-4}}\dfrac{M^2}{\int_0^T |u|^2(t,0)\mathrm{d}t}\right)}\frac{T}{\varepsilon}.
\tag{2.11}
$$

Sketch of the proof. The inequality (2.11) relies on two main ingredients:

- a log-convexity inequality, inspired by [4, 22, 24], which gives a bound of the initial condition with respect to the final time solution. This allows us to bypass a more standard exponential stability estimate, which bounds the final time solution by the initial condition.

- A Carleman inequality, inspired by [8, 9], which proves a controllability inequality for a very similar system. This technical part is an adaptation of the proof of Proposition 10 of [8]. $\quad\square$

A first interest of the inequality (2.11) is to provide uniqueness for the inversion of $\Psi_{T,\varepsilon}^{u^0 \to \mathrm{Tr}}$ and stability of a reconstruction. Moreover, thanks to the diffusion term, we no longer have a lower-bound condition on T like the condition $T \geq T_0$ for (2.9). There is a price to pay however, which is the logarithmic rate of convergence given by $1/\rho^{-1}$.

2.5. Solution and error estimate for the second-order inverse problem

A second benefit of (2.11) is to indicate how a Tikhonov regularisation strategy may be built, and an estimate for the reconstruction. Since the bound M of the $\mathrm{H}^1(0, L)$ norm appears in (2.11), we define the following quadratic functional to minimise:

$$
\mathscr{J}_{|T}(u^0) := \frac{1}{2M^2}\|u^0\|_{H^1}^2 + \frac{1}{2\delta^2}\int_0^T \left|y^\delta(t) - u_{|u^0}(t, 0)\right|^2 \mathrm{d}t,
\tag{2.12}
$$

and we obtain the following error estimate.

Theorem 3 ([14, Thm. 4.6]). *Let $u^0 \in \mathrm{H}_R^1(0, L)$ with $\|u^0\|_{H^1} \leq M$, $y = u_{u^0}(t, \cdot) = \Psi_{T,\varepsilon}^{u^0 \to \mathrm{Tr}}(u^0)$, $\delta > 0$. We assume that we observe y^δ such that*

$$
\|y - y^\delta\|_{L^2(0,T)} \leq \delta.
$$

Let $\widehat{u^0}$ an estimate for u^0 defined by

$$
\widehat{u^0} = \operatorname*{argmin}_{u^0 \in \mathrm{H}_R^1(0,L)} \mathscr{J}_{|T}(u^0).
\tag{2.13}
$$

Then there exist constants C_1 and C_2, depending only on the parameters L, b, T and ε, such that

$$\|u^0 - \bar{u}^0_{|T}\|^2_{L^2(0,L)} \le \frac{C_1 M^2}{\rho^{-1}(C_2 \frac{M^2}{\delta^2})}. \tag{2.14}$$

As expected, the speed of convergence is logarithmic with respect to the noise δ, to be compared with an algebraic rate of convergence for the first-order system. We give here the proof, which is very short and emblematic of error estimates for Tikhonov regularisation methods.

Proof. Let us denote $\widetilde{u}^0 = u^0 - \widehat{u^0}$ and $\widetilde{u} = u_{|\widetilde{u}^0}$ the solution of (2.10) with $\widetilde{u}^0$ as an initial condition. We apply (2.11) to $\widetilde{u}$ (and replace M by $\|\widetilde{u}^0\|_{H^1}$), and obtain

$$\|\widetilde{u}(\cdot,0)\|^2_{L^2(0,L)} + \varepsilon|\widetilde{u}(0,0)|^2 \lesssim \frac{Ce^{\frac{L}{\varepsilon}}\|\widetilde{u}^0\|^2_{H^1}}{\rho^{-1}\left(CT\frac{e^{-\frac{c}{\varepsilon}(1+T^{-1})}}{1+T^{-4}}\frac{\|\widetilde{u}^0\|^2_{H^1}}{\int_0^T |\widetilde{u}|^2(t,0)\mathrm{d}t}\right)}\frac{T}{\varepsilon}.$$

From $\mathscr{J}_{|T}(\widehat{u^0}) \le \mathscr{J}_{|T}(u^0) \le 1$ we have

$$\int_0^T |\widetilde{u}|^2(t,0)\mathrm{d}t \le 2\int_0^T |u_{|\widehat{u^0}} - y^\delta|^2(t,0)\mathrm{d}t + 2\int_0^T |y^\delta - y|^2(t,0)\mathrm{d}t \le 6\delta^2.$$

By the triangular inequality, we obtain

$$\|\widetilde{u}^0\|^2_{H^1} \le 2\|\widehat{u^0}_{|T}\|^2_{H^1} + 2\|u^0\|^2_{H^1} \le 2M^2 + 2\|\widehat{u^0}_{|T}\|^2_{H^1}.$$

To estimate $\|\widehat{u^0}_{|T}\|^2_{H^1}$, we use the fact that it minimises $\mathscr{J}_{|T}$: by definition, we have

$$\|\widehat{u^0}\|^2_{H^1} \le 2M^2 \mathscr{J}_{|T}(\bar{y}^0) \le 2M^2 \mathscr{J}_{|T}(u^0) \le 2M^2\left(\frac{1}{2M^2}\|u^0\|^2_{H^1} + \frac{1}{2}\right)\mathscr{J}_{|T}(\bar{y}^0) \le 2M^2,$$

hence

$$\|\widetilde{u}^0_{|T}\|^2_{H^1} \le 4M^2$$

and we conclude by the fact that $x \mapsto \frac{x}{\rho^{-1}(x)}$ is increasing on $(0, \infty)$. $\qquad\square$

To conclude, in [14] we propose a sequential approach based on Kalman filtering to solve the estimation problem. Despite the severely ill-posed character of the second-order problem, it reveals more accurate in simulation tests than the pure transport approximation. Moreover, the information contained in the corrective diffusion is visible in the fact that even for $T \le T_0 = \frac{L}{b}$ we are able to estimate the distribution for $x > bT$. How to refine the Carleman estimate cited above, in order to quantify precisely how this information content vanishes for small times and, on the contrary, stop improving for $T_0 \lesssim T$, remains a very interesting open problem.

3. Inverse problems for the fragmentation equation

Let us now turn to reactions of fragmentation type as given by (1.2). Writing a first-order asymptotics in the same spirit as (2.5) leads to the fragmentation equation:

$$\frac{\partial}{\partial t}u(t,x) = -\alpha x^\gamma u(t,x) + \alpha \int_x^\infty \kappa\left(\frac{x}{y}\right) y^{\gamma-1}u(t,y)\mathrm{d}y, \tag{3.1}$$

where αx^γ represents the breakage rate of a particle of size x, that we have assumed to be given by a power law, and $\kappa(z)$ is linked to the probability for a particle of size y to give rise to a particle of size $x = zy$. This model appeared as the best-fit one in experiments on β_2-microglobulin [27], where the experimental measurements consisted in samples of fibril sizes observed at several time points. This led us to formulating the following inverse problem: How to estimate the fragmentation features given by the parameters (α, γ, κ) from such data, with $\alpha, \gamma > 0$ and κ a probability measure on $[0,1]$?

For mass conservation considerations we assume here binary fragmentation and no atom at $z = 0$ and $z = 1$, hence

$$\kappa \in \mathcal{M}^+((0,1)), \qquad \int_0^1 \kappa(\mathrm{d}z) = 2, \qquad \int_0^1 z\kappa(\mathrm{d}z) = 1, \qquad \kappa(z) = \kappa(1-z). \tag{3.2}$$

For simplicity we have assumed binary fragmentation, which corresponds to $N = 2$ in [12], but the study is unchanged.

This section explains the various stages in our approach, and how we successively used asymptotic behaviour, then large-time dynamics, and finally short-time behaviour, to try to estimate κ, α and γ. The common thread running through our studies is the use of moments of the solution to interpret the data observed, either integer moments or through the Mellin transform.

3.1. Using the asymptotic profile

First inspired by our knowledge on long-time asymptotics for the fragmentation equation [16] and for the growth-fragmentation equation [20] (note that this field of research has been continuously active for several decades, with remarkable new results inspired by stochastic methods [6] or optimal transport [17]) and by previous studies of the inverse problem based on the asymptotic steady size distribution [13, 15, 21], we first studied the information content of the self-similar asymptotic profile g solution to

$$2g(z) + zg'(z) + \alpha\gamma z^\gamma g(z) = \alpha\gamma \int_z^\infty \kappa\left(\frac{z}{u}\right) u^{\gamma-1} g(u)\mathrm{d}u, \qquad \int g(z)\mathrm{d}z = 1. \qquad (3.3)$$

This idea was guided by the fact that, as proven in Theorem 3.2 of [16] and Theorem 3.2. of [20], under suitable assumptions on κ and for $\gamma > 0$, we have

$$\lim_{t \to \infty} \int_0^\infty \left| u(t,y) - t^{\frac{2}{\gamma}} g\left(t^{\frac{1}{\gamma}} y\right) \right| y\mathrm{d}y = 0, \qquad (3.4)$$

so that we can guess that, after rescaling, the profile g may be experimentally observed. The inverse problem can then be reduced to: From observations on g, is it possible to estimate (α, γ, κ)?

In [11], we used the Mellin transform, defined for measure-valued functions by

$$M[g](s) := \int_0^{+\infty} x^{s-1} g(x)\mathrm{d}x, \qquad M[\kappa](s) := \int_0^1 x^{s-1} \kappa(\mathrm{d}x), \qquad (3.5)$$

and the multiplicative convolution (which plays a similar role for the Mellin transform as the standard convolution product for the Fourier transform)

$$(f * g)(x) := \int_{\mathbb{R}^+} f(y)\, g\left(\frac{x}{y}\right)\frac{dy}{y}, \qquad M[f * g](s) = M[f](s).M[g](s). \qquad (3.6)$$

We moreover notice that $M[z \mapsto z^\gamma g(z)](s) = M[g](s + \gamma)$, hence taking the Mellin transform of (3.3) leads to an explicit formula for κ, namely

$$M[\kappa](s) = 1 + \frac{(2-s)M[g](s)}{\alpha\gamma M[g](s+\gamma)}. \qquad (3.7)$$

To justify rigorously (3.7), we need to prove that $M[g](s+\gamma)$ never vanishes on a vertical strip; we then also need to define its inverse Mellin transform. To do so, we carried out a detailed analytical study in the complex plane. We also proved a uniqueness result which states that for a wide class of kernels κ, for a given g, there exists at most one triplet (α, γ, κ) such that g is solution to (3.3) ([11, Thm. 1 and 2]). These results however revealed of little practical use, for two main reasons.

- Though (α, γ) are uniquely determined, they are obtained by using the asymptotic behaviour of $g(z)$ for $z \to \infty$. Since we have only access to size samples, we have information on g only on a compact set of $(0, +\infty)$.

- The formula involves Mellin and inverse Mellin transform, so that the inverse problem, as for the second-order asymptotic inverse problem seen above (Theorem 3), is severely ill-posed. When numerical implemented, the results failed to be satisfactory.

3.2. Using the long-term dynamics

We thus turned to the time-evolution dynamics given by (3.4). From a sample $(x_1(t_i), \ldots, x_{n_i}(t_i))$ at time t_i, we estimate the average polymer size $M_1(t_i) = \frac{\int_0^\infty x u(t,x) \mathrm{d}x}{\int_0^\infty u(t,x) \mathrm{d}x}$ by the empirical first moment, i.e.

$$\widehat{M_1}(t_i) = \sum_{k=1}^{n_i} x_k(t_i) \tag{3.8}$$

and we can relate it to $u(t,x)$ to estimate (α, γ). The limit given by (3.4) implies

$$M_1(t) \approx_{t \gg 1} \frac{\int t^{\frac{2}{\gamma}} y g(t^{\frac{1}{\gamma}} y) \mathrm{d}y}{\int t^{\frac{2}{\gamma}} g(t^{\frac{1}{\gamma}}) \mathrm{d}y} = C t^{-\frac{1}{\gamma}},$$

Similarly, for the estimation of α, we notice that integrating (3.3) and using the mass conservation property $\int \kappa(\mathrm{d}z) = 1$ provides the equality

$$1 = \alpha \gamma \int z^\gamma g(z) \mathrm{d}z. \tag{3.9}$$

We thus use the γ−th moment $M_\gamma(t) = \frac{\int_0^\infty x^\gamma u(t,x) \mathrm{d}x}{\int_0^\infty u(t,x) \mathrm{d}x}$, approximated by an empirical $\widehat{M_{\hat\gamma}}(t_i)$ defined by

$$\widehat{M_{\hat\gamma}}(t_i) = \sum_{k=1}^{n_i} (x_k(t_i))^{\hat\gamma} \tag{3.10}$$

and the equality

$$M_\gamma(t) \approx_{t \gg 1} \frac{\int t^{\frac{2}{\gamma}} y^\gamma g(t^{\frac{1}{\gamma}} y) \mathrm{d}y}{\int t^{\frac{2}{\gamma}} g(t^{\frac{1}{\gamma}}) \mathrm{d}y} = \frac{\int z^\gamma g(z) \mathrm{d}z}{t} \implies \alpha \approx_{t \gg 1} \frac{1}{\gamma t} \frac{1}{M_\gamma(t)} \tag{3.11}$$

so that in [5] we designed the following estimation protocol for (α, γ):

1. On observed data $(\log(t_i), \log(\widehat{M_1}(t_i))_{1 \le i \le n}$ defined by (3.8), fit the three parameters $(C, t_{\mathrm{asymp}}, \hat\gamma)$ of a curve

$$\left(\log(t), C - \frac{1}{\hat\gamma} \log\left(\frac{t}{t_{\mathrm{asymp}}} \right) \mathbb{1}_{t \ge t_{\mathrm{asymp}}} \right),$$

   and choose $\hat\gamma$ as an estimator for γ. Notice here that we only use a small part of the available data, namely the average size and not the whole size distribution. For each of the four protein fibrils analysed (α−Synuclein, associated with Parkinson, bovine β−lactoglobulin, chicken egg lysozyme and β_2-microglobulin – this last amyloids being involved in systemic dialysis-related amyloidosis), we had from 7 to 14 time points t_i. The fits appeared very satisfactory (see [5, Fig. 5]).

2. From $\hat\gamma$ and t_{asymp} computed in Step 1, use $\widehat{M_{\hat\gamma}}(t_i)$ defined by (3.10) for $t_i \ge t_{\mathrm{asymp}}$ and define an estimator of α from (3.11) by

$$\hat\alpha := \frac{1}{\hat\gamma t_i} \frac{1}{\widehat{M_{\hat\gamma}}(t_i)}.$$

To validate a posteriori the estimation obtained, we then use the full dataset as follows.

1. At initial time $t_1 = 0$, use the size sample $(x_1(0), \ldots, x_{n_0}(0))$ to obtain an estimate $\widehat{f}(0, x)$ of the initial size distribution $f(0, x) = \frac{u(0,x)}{\int u(0,x) \mathrm{d}x}$ by a kernel density estimation method [26],

2. Simulate (3.1) with $u(0, x) = \widehat{f}(0, x)$, parameters $\hat\alpha$, $\hat\gamma$ and a given kernel κ,

3. at times t_i of experimental observations, compare the simulated $f(t_i, x) = \frac{u(t_i,x)}{\int u(t_i,x) \mathrm{d}x}$ with the density $\widehat{f}(t_i, x)$ estimated from the size sample $(x_1(t_i), \ldots, x_{n_i}(t_i))$ by a kernel density estimation method.

We obtained the following quantitative results.

- We observed a remarkable agreement between the data observed and simulated (see [5, Fig. 7]). This is all the more satisfactory that only two parameters α and γ are necessary to fit the whole time dynamics.

- The choice of κ appears to have only little influence on the goodness of fit, even if we were able to draw some conclusion on whether the breakage occurs more at the edges or at the centre of the fibrils ([5, Fig. 8]).

3.3. Using the short-time dynamics

Intrigued by the influence of κ on the time dynamics, we carried out a thorough numerical investigation in [25]. Specifically, we developed a statistical test to quantify how different fragmentation kernels influence the size distribution of particles over time ([25, Fig. 4]), when all dynamics depart from the same initial distribution. We noticed first that the sharper the initial condition, the greater the influence of κ; and second, that this influence occurs mainly during an early time-window – even if not too early either, since the initial condition is taken identical in all simulations. Following this study, we turned to a short-time asymptotic development, carried out from a Dirac delta function as an initial distribution. A formal computation shows:

$$u(t + \Delta t, x) \approx u(t, x) - \alpha \Delta t x^\gamma u(t, x) + \alpha \Delta t \int_x^\infty \kappa \left(\frac{x}{y} \right) y^{\gamma - 1} u(t, y) \mathrm{d}y + o(\Delta t). \tag{3.12}$$

Departing at time $t = 0$ from a Dirac delta function $\delta(x - 1)$ at $x = 1$, we have

$$u(\Delta t, x) \approx \delta(x - 1)(1 - \alpha \Delta t) + \alpha \kappa(x) \Delta t + o(\Delta t),$$

and thus we obtain an estimation formula for the kernel κ: if we have $\widehat{u}(\Delta t, \cdot)$ an estimate, obtained from observations on the size distribution $u(\Delta t, \cdot)$, we could define an estimate for κ as

$$\widehat{\kappa}(x) := \frac{1}{\alpha \Delta t} \left(u(\Delta t, x) - (1 - \alpha \Delta t) \delta(x - 1) \right) + o(1), \qquad \Delta t \ll 1. \tag{3.13}$$

How to evaluate the error made between κ and $\widehat{\kappa}$? We first recall the definitions of the Total Variation norm and the Bounded Lipshitz norm, the latter being more appropriate to evaluate the discrepancy between two measures.

Definition 4. Let $\mathcal{M}(\mathbb{R}^+)$ be the set of Radon measures whose support belongs to $\mathbb{R}^+$. The Total Variation (TV) norm of the signed measure $\mu \in \mathcal{M}(\mathbb{R}^+)$ is defined as

$$\|\mu\|_{TV} := \sup \left\{ \int_{R^+} \varphi(x) \mathrm{d}\mu(x), \quad \varphi \in \mathcal{C}(\mathbb{R}^+) \cap L^1(\mathrm{d}|\mu|), \ \|\varphi\|_\infty \leq 1 \right\}, \tag{3.14}$$

whereas its Bounded Lipshitz (BL) norm is defined by

$$\|\mu\|_{BL} := \sup \left\{ \int_{R^+} \varphi(x) \mathrm{d}\mu(x), \quad \varphi \in \mathcal{C}(\mathbb{R}^+) \cap L^1(\mathrm{d}|\mu|), \ \|\varphi\|_\infty \leq 1, \ \|\varphi'\|_\infty \leq 1 \right\}. \tag{3.15}$$

A second preliminary step is to define measure-valued solutions to (3.12), since we want to depart from a measure-valued initial condition. This is given by the following definition.

Definition 5 (Weak solution for (3.1) – [12, Def. 2.1]). A family $(\mu_t)_{t \geq 0} \subset \mathcal{M}(\mathbb{R}^+)$, where $\mathcal{M}(\mathbb{R}^+)$ denotes the space of Radon measures on $\mathbb{R}_+$, is called a measure-valued solution to (3.1), with initial data $\mu_0 \in \mathcal{M}(\mathbb{R}^+)$ satisfying $\mathrm{supp}(\mu_0) \subset [0, L]$, if the mapping $t \to \mu_t$ is narrowly continuous and for all $\varphi \in \mathcal{C}(\mathbb{R}^+)$ such that $x \mapsto \varphi(x)/(1 + x)$ is bounded on $[0, \infty)$, and all $t \geq 0$,

$$\int_{\mathbb{R}^+} \varphi(x) \mathrm{d}\mu_t(x) = \int_{\mathbb{R}^+} \varphi(x) \mathrm{d}\mu_0(x) + \int_0^t \mathrm{d}s \int_{\mathbb{R}^+} \mathrm{d}\mu_s(x) \alpha x^\gamma \left(-\varphi(x) + \int_0^1 \mathrm{d}\kappa(z) \varphi(xz) \right). \tag{3.16}$$

The class of functions used to define the weak solutions ensures the finiteness of $\int (1 + x) \mathrm{d}\mu_t(x)$. Despite the numerous studies carried out on fragmentation equations, the question of existence and uniqueness of measure-valued solutions to (3.1) for a large class of kernels κ appeared yet unsolved. We thus formulated the following result, which moreover provides an explicit decomposition of such solutions.

Theorem 6 (From Theorems 2.2, 2.4 and 2.6 of [12])**.** *Let κ satisfy* (3.2), *$\alpha > 0$, $\gamma \geq 0$ and $\mu_0 \in \mathcal{M}^+((0,L))$ for some $L > 0$. There exists a unique solution $\mu \in \mathcal{C}(\mathbb{R}^+, \mathcal{M}^+(\mathbb{R}_+))$ to* (1.2) *in the sense of Definition 4. Moreover, this unique solution*

- *preserves the mass: $\int x\mu_t(\mathrm{d}x) = \int x\mu_0(\mathrm{d}x)$ for any $t \geq 0$,*

- *is nonnegative,*

- *satisfies $\mathrm{supp}(\mu_t) \subset (0,L)$ for any $t \geq 0$,*

- *is explicitly defined as the following series, which is absolutely convergent in the TV norm for any $t \geq 0$:*

$$\mu_t = e^{-\alpha x^\gamma t}\mu_0 + \sum_{n=0}^{\infty}(\alpha t)^n \int_0^\infty \ell^{n\gamma}a_n\left(\frac{x}{\ell}\right)\mu_0(\ell)\frac{\mathrm{d}\ell}{\ell}, \tag{3.17}$$

where the sequence a_n is defined by induction, for $x \in [0,1]$, by

$$a_0(x) = 0, \quad a_{n+1}(x) = \frac{1}{n+1}\left(-x^\gamma a_n(x) + 2\int_x^\infty y^{\gamma-1}\kappa\left(\frac{x}{y}\right)a_n(y)\mathrm{d}y + \kappa(x)\frac{(-1)^n}{n!}\right). \tag{3.18}$$

Proof. Let us sketch the main steps of the proof.

Step 1. Uniqueness and, if the solution is proved to be nonnegative, $\mathrm{supp}(\mu_t) \subset (0,L)$ ([12, Thm. 2.2]). By choosing appropriate weight functions, we prove that a nonnegative solution satisfies $\mathrm{supp}(\mu_t) \subset (0,L)$ for any $t \geq 0$, and then, with a Gronwall lemma, we prove the estimate

$$\|\mu_t\|_{TV} \leq \|\mu_0\|_{TV}e^{\alpha(2L)^\gamma 3t},$$

which implies uniqueness.

Step 2. The representation formula (3.17) *([12, Thm. 2.4]).* We first obtain the representation formula for the fundamental solution μ_t^F, defined as a solution departing from $\mu_0(x) = \delta(x-1)$, namely:

$$\mu_t^F(x) = e^{-\alpha t}\delta(x-1) + \sum_{n=1}^{\infty}(\alpha t)^n a_n(x). \tag{3.19}$$

By a scaling property, we deduce the solution for $\mu_0(x) = \delta(x-\ell)$, and we conclude by the superposition principle.

Step 3. Nonnegativity of the representation formula (3.17) *([12, Thm. 2.6]).* This last step is more delicate than it can seem at first sight, and we have not found nonnegativity results for cases where both the solution and the fragmentation kernel are measure-valued functions in the literature. We prove this result by an approximation strategy on the fundamental solution μ_t^F. Regularising both κ and the initial condition $\delta(x-1)$, we can apply a nonnegativity result [19]; we then conclude by weak convergence of the approximate sequence. $\qquad\square$

Equipped with this well-defined notion of measure-valued solutions, we then turn to our inverse problem, and prove successive error estimates:

1. error estimate, in TV norm, for perfectly observed data (no noise), with perfectly monodisperse initial condition ($\mu_0 = \delta(x-1)$),

2. error estimate, in BL norm, for noisy data and nearly monodisperse initial condition ($\|\mu_0 - \delta(x-1)\|_{BL}$ small),

3. error estimate, via the Mellin transform, for perfectly observed data (no noise) and disperse initial condition (arbitrary μ_0).

Error estimate (1): monodisperse initial condition, no noise

From (3.19), we are led to precise the formal approximation (3.13), and define

$$\kappa^{est}(t) := \frac{\mu_t^F - e^{-\alpha t}\delta(x-1)}{\alpha t}. \tag{3.20}$$

From (3.17) we immediately obtain ([12, Thm. 3.1]):

$$\left\|\kappa^{est} - \kappa\right\|_{TV} \leq t\left(\alpha \sum_{n=0}^{\infty}(\alpha T)^n \|a_{n+2}\|_{TV}\right), \ \forall \, t \in (0,T]. \tag{3.21}$$

Error estimate (2): nearly monodisperse initial condition, noisy data

From (3.21) it seems that the smaller the observation time t, the more precise the estimate κ^{est}. This fails in practice, where the noise comes from at least two sources:

- the initial condition cannot be as precise as a Dirac delta function. Even if we could imagine an experimental setup selecting polymers of the same size, there will always remain some heterogeneity.

- Typical experimental measurements being polymer size samples, μ_t as well as μ_0 can only be observed up to a noise.

For these two sources of noise, the distance in BL-norm is well-adapted. We thus formulate the following error estimate result.

Theorem 7 ([12, Thm. 3.5]). *Let κ satisfy (3.2) and $\mathrm{supp}(\mu_0^{\varepsilon_0}) \subset [0,L]$ such that*

$$\|\mu_0^{\varepsilon_0} - \delta(x-1)\|_{BL} \leq \varepsilon_0.$$

Let $\mu_t^{\varepsilon_0}$ the unique solution to the fragmentation equation (1.2) with initial condition $\mu_0^{\varepsilon_0}$. Let $\mu_0^{\varepsilon_0,\varepsilon_1}$ and $\mu_t^{\varepsilon_0,\varepsilon_2}$ be noisy observations of the respective measures $\mu_0^{\varepsilon_0}$ and $\mu_t^{\varepsilon_0}$ such that

$$\|\mu_0^{\varepsilon_0,\varepsilon_1} - \mu_0^{\varepsilon_0}\|_{BL} \leq \varepsilon_1, \qquad \|\mu_t^{\varepsilon_0,\varepsilon_2} - \mu_t^{\varepsilon_0}\|_{BL} \leq \varepsilon_2.$$

Assume moreover either $\gamma \geq 1$ or $\mathrm{supp}(\mu_0) \subset [m,L]$ with $m > 0$. Then, for all $0 \leq t \leq T$, there exist constants $C_1(T,\alpha)$ and $C_2(L,T,\alpha,\gamma,m^{\gamma-1})$ such that

$$\left\|\frac{\mu_t^{\varepsilon_0,\varepsilon_2} - e^{-\alpha t}\mu_0^{\varepsilon_0,\varepsilon_1}}{\alpha t} - \kappa\right\|_{BL} \leq C_1 t + \frac{C_2\varepsilon_0 + \varepsilon_1 + \varepsilon_2}{\alpha t}. \tag{3.22}$$

The proof is immediate once a stability result for the solution to the fragmentation equation in BL-norm is obtained ([12, Thm. 2.11]): this stability is obtained only for $\gamma \geq 1$ or $\mathrm{supp}(\mu_0) \subset [m,L]$ with $m > 0$, hence these extra assumptions. The form of the inequality (3.22) is very interesting, since it has the exact form of the usual bias-variance decomposition for regularisation of inverse problems as well as nonparametric density estimation: a balance between two terms, one vanishing with the regularisation parameter, the other exploding with it but multiplied by the noise level, so that the observation time t identifies with a regularisation parameter, and is optimal if $t = O(\sqrt{\varepsilon_0 + \varepsilon_1 + \varepsilon_2})$. Intuitively, this corresponds to a time large enough so that "enough" (compared to the noise level) polymers have divided once, but small enough so that not too many have divided twice (term $C_1 t$, which shows that the distance increases with time).

Error estimate (3): arbitrary initial condition, no noise

In many practical cases, it seems very difficult, if not impossible, to have monodisperse or almost monodisperse initial condition, so that μ_0 is far away from a Dirac delta function. The same strategy can however be followed.

For simplicity, let $\kappa = \kappa(x)\mathrm{d}x$ and $\mu_0 = u_0(x)\mathrm{d}x$ admit densities with respect to the Lebesgue measure, denote $\mu_t(\mathrm{d}x) = u(t,x)\mathrm{d}x$ and define

$$F^{est}(u_0,\kappa;t,x) = \frac{u(t,x) - e^{-\alpha t x^{\gamma}}u_0(x)}{\alpha t}. \tag{3.23}$$

Thanks to the superposition principle, we have the following inequality ([12, Cor. 3]), for a given constant C,

$$\left\| F^{est}(u_0, \kappa; t) - w_0 * \kappa(x) \right\|_{TV} \le C L^{2\gamma} \|u_0\|_{TV}\, t, \ \forall\, t \in (0, T], \qquad (3.24)$$

where $*$ denotes the multiplicative convolution defined by (3.6) and $w_0(x) = x^\gamma u_0(x)$. To have an estimate for κ, we thus need to invert the multiplicative convolution. Recalling the formula for the Mellin transforms (3.5), a candidate to estimate $M[\kappa]$ is thus

$$M[\kappa]^{est}(s; t) := \frac{M[u(t, \cdot)](s) - M\left[x \mapsto e^{-\alpha t x^\gamma} u_0(x)\right](s)}{\alpha t M[u_0](s + \gamma)}. \qquad (3.25)$$

The common point between (3.7) and (3.25) is the presence in the denominator of the Mellin transform, taken in $s + \gamma$, of a size distribution observed. For large γ, this may reveal very noisy, since this gives a lot of weight to the largest particles. However, contrarily to the case of formula (3.7), we managed to prove an error estimate between $M[\kappa]^{est}(s; t)$ and $M[\kappa](s)$, thanks to the Mellin transform of the series representation (3.17) and under regularity assumptions on κ and u_0. We refer the reader to Theorems 4.4 and 4.6 in [12] for a detailed presentation of the results, which contain an estimate of the distance between κ and $\kappa^{est} := M^{-1}[M[\kappa]^{est}(\cdot; t)]$, in a certain weighted norm, in the order of $C^{st}\Delta t$. This order of magnitude is coherent with the estimate (3.21); however the constant C^{st} is not explicit, and the observation noise is not yet taken into account. This question of stability with respect to noise, together with a numerical investigation, are perspectives for future work.

References

[1] A. Armiento, M. Doumic, P. Moireau & H. Rezaei, "Estimation from Moments Measurements for Amyloid Depolymerisation", *J. Theor. Biol.* **397** (2016), p. 68-88.

[2] A. Armiento, P. Moireau, D. Martin, N. Lepejova, M. Doumic & H. Rezaei, "The mechanism of monomer transfer between two structurally distinct PrP oligomers", *PLoS ONE* **12** (2017), no. 7, article no. e0180538.

[3] H. T. Banks, M. Doumic, C. Kruse, S. Prigent & H. Rezaei, "Information content in data sets for a nucleated-polymerisation model", *J. Biol. Dyn.* **9** (2015), no. 1, p. 172-197.

[4] C. Bardos & L. Tartar, "Sur l'unicité rétrograde des équations paraboliques et quelques questions voisines", *Arch. Ration. Mech. Anal.* **50** (1973), no. 1, p. 10-25.

[5] D. M. Beal, M. Tournus, R. Marchante, T. Purton, D. P. Smith, M. F. Tuite, M. Doumic & W.-F. Xue, "The division of amyloid fibrils", *iScience* **23** (2020), no. 9, article no. 101512.

[6] J. A. Cañizo, P. Gabriel & H. Yoldas, "Spectral gap for the growth-fragmentation equation via Harris's theorem", *SIAM J. Math. Anal.* **53** (2021), no. 5, p. 5185-5214.

[7] J.-F. Collet, T. Goudon, F. Poupaud & A. Vasseur, "The Becker–Döring system and its Lifshitz–Slyozov limit", *SIAM J. Appl. Math.* **62** (2002), no. 5, p. 1488-1500.

[8] P. Cornilleau & S. Guerrero, "Controllability and observability of an artificial advection-diffusion problem", *Math. Control Signals Syst.* **24** (2012), no. 3, p. 265-294.

[9] J.-M. Coron & S. Guerrero, "Singular optimal control: A linear 1-D parabolic-hyperbolic example", *Asymptotic Anal.* **44** (2005), no. 3-4, p. 237-257.

[10] C. Della Valle, M. Doumic & P. Moireau, "State estimation of a backward transport equation by moment measurements", working paper.

[11] M. Doumic, M. Escobedo & M. Tournus, "Estimating the division rate and kernel in the fragmentation equation", *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* **35** (2018), no. 7, p. 1847-1884.

[12] ———, "An inverse problem: recovering the fragmentation kernel from the short-time behaviour of the fragmentation equation", `https://hal.science/hal-03494439v2`, to appear in *Ann. Henri Lebesgue*, 2024.

[13] M. Doumic, M. Hoffmann, P. Reynaud-Bouret & V. Rivoirard, "Nonparametric estimation of the division rate of a size-structured population", *SIAM J. Numer. Anal.* **50** (2012), no. 2, p. 925-950.

[14] M. Doumic & P. Moireau, "Asymptotic approaches in inverse problems for depolymerization estimation", working paper.

[15] M. Doumic, B. Perthame & J. P. Zubelli, "Numerical solution of an inverse problem in size-structured population dynamics", *Inverse Probl.* **25** (2009), no. 4, article no. 045008 (25 pages).

[16] M. Escobedo, S. Mischler & M. Rodriguez Ricard, "On self-similarity and stationary problem for fragmentation and coagulation models", *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* **22** (2005), no. 1, p. 99-125.

[17] N. Fournier & B. Perthame, "A nonexpanding transport distance for some structured equations", *SIAM J. Math. Anal.* **53** (2021), no. 6, p. 6847-6872.

[18] L. Halpern, "Artificial boundary conditions for the linear advection diffusion equation", *Math. Comput.* **46** (1986), p. 425-438.

[19] Z. A. Melzak, "A scalar transport equation", *Trans. Am. Math. Soc.* **85** (1957), p. 547-560.

[20] P. Michel, S. Mischler & B. Perthame, "General relative entropy inequality: an illustration on growth models", *J. Math. Pures Appl.* **84** (2005), no. 9, p. 1235-1260.

[21] B. Perthame & J. P. Zubelli, "On the inverse problem for a size-structured population model", *Inverse Probl.* **23** (2007), no. 3, p. 1037-1052.

[22] K.-D. Phung, "Note on the cost of the approximate controllability for the heat equation with potential", *J. Math. Anal. Appl.* **295** (2004), no. 2, p. 527-538.

[23] S. Prigent, A. Ballesta, F. Charles, N. Lenuzza, P. Gabriel, L. M. Tine, H. Rezaei & M. Doumic, "An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation", *PLoS ONE* **7** (2012), no. 11, article no. e43273.

[24] Thi Minh Nhat Vo, "Construction of a control and reconstruction of a source for linear and nonlinear heat equations", PhD Thesis, Université d'Orléans, 2018.

[25] M. Tournus, M. Escobedo, W.-F. Xue & M. Doumic, "Insights into the dynamic trajectories of protein filament division revealed by numerical investigation into the mathematical model of pure fragmentation", *PLoS Comput. Biol.* **17** (2021), no. 9, article no. e1008964.

[26] A. B. Tsybakov, "Nonparametric estimators", in *Introduction to Nonparametric Estimation*, Springer Series in Statistics, Springer, 2009, p. 1-76.

[27] W.-F. Xue & S. E. Radford, "An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior", *Biophys. J.* **105** (2013), no. 12, p. 2811-2819.

Marie Doumic
MERGE, CMAP, Inria,
IP Paris, Ecole polytechnique, CNRS,
91128 Palaiseau cedex
FRANCE
`marie.doumic@inria.fr`