



Reconstruction of low-rank aggregation kernels in univariate population balance equations

Robin Ahrens¹ · Sabine Le Borne¹ 

Received: 1 November 2019 / Accepted: 8 April 2021 / Published online: 2 May 2021
© The Author(s) 2021

Abstract

The dynamics of particle processes can be described by population balance equations which are governed by phenomena including growth, nucleation, breakage and aggregation. Estimating the kinetics of the aggregation phenomena from measured density data constitutes an ill-conditioned inverse problem. In this work, we focus on the aggregation problem and present an approach to estimate the aggregation kernel in discrete, low rank form from given (measured or simulated) data. The low-rank assumption for the kernel allows the application of fast techniques for the evaluation of the aggregation integral ($\mathcal{O}(n \log n)$ instead of $\mathcal{O}(n^2)$) where n denotes the number of unknowns in the discretization) and reduces the dimension of the optimization problem, allowing for efficient and accurate kernel reconstructions. We provide and compare two approaches which we will illustrate in numerical tests.

Keywords Population balance equation · Aggregation kernel · Inverse method · Low rank approximation

Mathematics Subject Classification (2010) 45J05 · 65D15 · 65F55 · 65M32

1 Introduction

The phenomenon of aggregation is seen in a wide variety of chemical processes like droplet coalescence, granulation and the aggregation of crystals. In this process, two particles with masses u and v , respectively, unite and form a new particle of mass $u + v$. This process is modelled by a convolution-type integral involving an aggregation kernel $\kappa(u, v)$ that is a non-negative symmetric function of the particle masses u

Communicated by: Youssef Marzouk

✉ Sabine Le Borne
leborne@tuhh.de

¹ Hamburg University of Technology, Institute of Mathematics, 21073 Hamburg, Germany

and v which takes into consideration the conditions of the underlying process (like temperature and saturation).

The number density distribution $f(v, t)$ throughout this pure aggregation process is governed by the integro-differential equation

$$\frac{df(v, t)}{dt} = Q_{\text{source}}(v, t) - Q_{\text{sink}}(v, t) \tag{1}$$

with a source term

$$Q_{\text{source}}(v, t) = \frac{1}{2} \int_0^v \kappa(u, v - u) f(u, t) f(v - u, t) du \tag{2}$$

and a sink term

$$Q_{\text{sink}}(v, t) = f(v, t) \int_0^\infty \kappa(u, v) f(u, t) du. \tag{3}$$

Starting from an initial distribution $f(v, 0) = f_{\text{init}}(v)$, the density distribution f will vary dynamically over time as smaller particles combine and form larger particles. The total mass of all particles (also called the first moment),

$$\mu^1(f, t) := \int_0^\infty v f(v, t) dv,$$

stays constant while the total number of all particles (the zero-th moment) $\mu^0(f, t) := \int_0^\infty f(v, t) dv$ decreases over time.

Here, we are concerned with the reconstruction of the kernel $\kappa(u, v)$ from given data $F(v, t_i)$ at $m + 1$ different time instances $t_i, i \in \{0, \dots, m\}$. We denote the reference distributions by a capital F while distributions computed with a reconstructed kernel are denoted by a lowercase f .

Previous work in this field is often concerned with fitting a single ([1], [15]) or a small number of parameters of a kernel that is assumed to be of a certain form, e.g. fitting exponents in rational functions [12]. Laurent polynomials were used in [7] to approximate the aggregation kernel. All these works determine coefficients for an approximation of a kernel on $(0, \infty) \times (0, \infty)$. A reconstruction on a bounded computational domain is presented in [5]. Here, the property space is divided into cells and a (piecewise) bilinear basis is used to approximate the kernel.

There exist some methods for parameter estimation in differential equations that are based on discontinuous data in the presence of measurement noise ([2, 13] or [14]) without a connection to population balances. However, the differential equations considered in these works have only a few degrees of freedom and more measurements available making them non-applicable in this setting.

When formulating the optimization problem, there exist (at least) the following two approaches:

- Compute approximations to the derivatives on the left hand side of (1) using the measured data and compute the right hand side of (1) using a reconstructed

kernel where the reconstruction should minimize the error between left and right hand side of (1) in some appropriate norm.

- Simulate the aggregation process with a reconstructed kernel where the reconstruction should minimize the error between the measured and simulated data.

The first approach minimizes the error of the derivative and is typically computationally less challenging but requires a high temporal resolution of the density distribution in order to provide useful results. This is often infeasible for applications of experimentally measured data where the time between measurements might be large. The second approach often challenges the selected optimization algorithm and is computationally expensive but—if successful—yields better results than the first approach. In this paper, we will pursue the second approach which will be feasible in view of a low-rank assumption imposed on the kernel. This assumption allows us to use the fast evaluation techniques for the discrete aggregation process developed in [8, 10]. In particular, this paper includes the following novel contributions:

- The reconstruction of a discrete kernel in low-rank representation, both for a fixed and variable kernel basis.
- Optimization with the objective to minimize the error between measured and simulated data with a new error-function based on the χ^2 -measure.
- Numerical tests of this new framework.

The remainder of this paper is organized as follows: In Section 2, we introduce the discretization of the property space and the low-rank assumption for the kernel. We introduce the optimization problem and elaborate its structure in 3. Section 4 is devoted to numerical results of the proposed optimization problem, showing the applicability of the reconstructed kernel in different simulations.

2 Kernel estimation

The first prerequisite for the numerical solution of this inverse problem is a suitable discretization of the property space. For this, we define a maximum particle property $v^{\max} \in \mathbb{R}_+$ and exclude any particle that is larger from further consideration, i. e. we assume that $f(v, t) = 0$ if $v \geq v^{\max}$.

For some $p \in \mathbb{N}$, we define pivots in $[0, v^{\max}]$,

$$v_j = \frac{v^{\max}}{n} \cdot \frac{2j - 1}{2}, \quad j = 1, \dots, n \quad \text{with } n := 2^p \tag{4}$$

with equal spacing $h := v^{\max}/n$, resulting in a grid $\mathcal{G}_p^{v^{\max}} := \{v_1, \dots, v_n\}$.

We will assume particles are concentrated at these equidistant pivots and discretize $f(v, t)$ by macroscopic variables

$$f_j(t) := \int_{(j-1)h}^{jh} f(v, t) \, dv$$

at any given time t which allows us to represent $f(v, t)$ as a non-negative, time-dependent vector-valued function

$$f(t) := (f_1(t) \dots f_n(t))^T \in \mathbb{R}_{\geq 0}^n.$$

We also represent the measured data $F(v, t_i)$ with respect to the grid \mathcal{G}_p^{\max} as a time-dependent vector $F(t_i) = (F_1(t_i), \dots, F_n(t_i))^T \in \mathbb{R}_{\geq 0}^n$.

In view of this discretization, the goal is to reconstruct a symmetric kernel matrix $K \in \mathbb{R}_{\geq 0}^{n \times n}$ with $K_{i,j} = \kappa(v_i, v_j)$ at the pivots of $\mathcal{G}_p^{\max} \times \mathcal{G}_p^{\max}$ for an (unknown) kernel function $\kappa(\cdot, \cdot)$.

Since several of the physically motivated kernel functions have separable representations or approximations of low rank, see for example [10] for an analysis of the Brownian, shear and kinetic kernels, and since a separable kernel function of rank k implies a kernel matrix of rank (at most) k , we make the following assumption.

Assumption 1 *To reduce the number of unknown coefficients in the kernel matrix $K \in \mathbb{R}^{n \times n}$, we assume it is of rank $k \ll n$, i.e. it can be represented (or approximated) in the form*

$$K = U \cdot S \cdot U^T \tag{5}$$

with matrices $U \in \mathbb{R}^{n \times k}$ and $S \in \mathbb{R}^{k \times k}$. To enforce the symmetry of K , we require that $S = S^T$.

This reduces the degrees of freedom in the kernel matrix K from $\frac{n(n+1)}{2}$ to $nk + \frac{k(k+1)}{2}$. It also allows us to use the algorithms introduced in [8, 10] to accelerate the calculations: They provide the efficient evaluation of the source and sink terms (2) and (3) in the case of a separable kernel. We will refer to the (thin/rectangular) matrix U as the *kernel basis*.

The above discretization leads to a discrete source term

$$Q_{\text{source}}(j, t; U, S) := \frac{1}{2} \sum_{v=1}^k \sum_{p=1}^{j-1} f_{j-p-1}(t) f_p(t) \cdot (US)_{p,v} U_{j-p-1,v} \tag{6}$$

of convolution type which can be evaluated for all $j = 1, \dots, n$ in complexity $\mathcal{O}(kn \log n)$ for given kernel factors U, S . The discrete sink term is given by

$$Q_{\text{sink}}(j, t; U, S) := \sum_{v=1}^k \sum_{p=1}^n f_j(t) f_p(t) \cdot (US)_{p,v} U_{p,v} \tag{7}$$

and evaluated in complexity $\mathcal{O}(kn)$.

Our framework for the kernel estimation is related to [5] and [7] where coefficients for kernel functions within a linear space spanned by a number of given basis functions, e.g. Laurent polynomials, are to be found.

The corresponding continuous kernel function in [7] can be expressed in the form $\kappa(u, v) = (b_1(u) \dots b_k(u)) S (b_1(v) \dots b_k(v))^T$ where some of the coefficients of the symmetric matrix $S \in \mathbb{R}^{k \times k}$ are subject to the optimization (while the others are fixed to zero). In our framework, this corresponds to a given (fixed) matrix U with its

entries given by the basis functions evaluated at the pivots, i. e. $U_{ij} = b_j(v_i)$. Here, we further generalize this framework by including U in the optimization process.

3 Optimization problem

With these prerequisites, we define the following minimization problem

$$\begin{aligned} \text{minimize}_{U,S} \quad & E_2(U, S) := \left(\sum_{i=0}^m \sum_{j=1}^n (F_j(t_i) - f_j(t_i))^2 \right)^{1/2} \\ \text{where} \quad & U \in \mathbb{R}^{n \times k}, S = S^T \in \mathbb{R}^{k \times k}. \end{aligned} \tag{8}$$

Here, $F_j(t_i)$ is the given (measured) data and $f_j(t_i)$ are computed by numerical simulation of

$$\frac{df_j(t)}{dt} = Q_{\text{source}}(j, t; U, S) - Q_{\text{sink}}(j, t; U, S)$$

on the grid $\mathcal{G}_p^{v_{\max}}$.

A possible disadvantage of the above minimization problem results from the fact that absolute errors are considered. Cells with a small number of particles may have only a small influence on the kernel estimation since the error E_2 is dominated by index pairs (i, j) where $F_j(t_i)$ is large. To increase the sensitivity with respect to those cells with a small amount of particles, we define the error based on the χ^2 -measure leading to

$$\begin{aligned} \text{minimize}_{U,S} \quad & E(U, S) := \left(\sum_{i=0}^m \sum_{j=1}^n \frac{(F_j(t_i) - f_j(t_i))^2}{f_j(t_i) + \epsilon} \right)^{1/2} \\ \text{where} \quad & U \in \mathbb{R}^{n \times k}, S = S^T \in \mathbb{R}^{k \times k}. \end{aligned} \tag{9}$$

Here, we weigh the difference of simulated and observed particles higher when the simulation indicates a small number of particles in a class. Similar measures to curve-fitting are also used in machine learning [11]. We add $\epsilon = 10^{-10}$ to the denominator in (9) to ensure it is large enough to avoid numerical instabilities caused by the division.

We set $f_j(t_0) = F_j(t_0)$, $j = 1, \dots, n$, as the initial distribution which is always non-negative. The constraint $f_j(t_i) \geq 0$ will be satisfied throughout the simulation when step sizes in the time discretization of the differential equation are chosen sufficiently small.

Since the objective is to determine a kernel that minimizes the error between measured and simulated density distribution, every evaluation of $E(U, S)$ requires the solution of (1), making it mandatory that efficient computational techniques are available. In fact, there are kn degrees of freedom in U and $\frac{k(k+1)}{2}$ degrees of freedom in S , leading to this number of evaluations of $E(U, S)$. Most computational time during this optimization is spent in the evaluation of (6) which only becomes feasible with our Assumption 1 of a separable kernel and the fast algorithms designed in [8, 10].

We solve the minimization problem (9) using MATLAB’s optimization routine *lsqnonlin* and use the routine *ode45* to solve the underlying differential equation. The routine *lsqnonlin* uses the Levenberg-Marquardt algorithm to search for local optima via finite differences.

If U has full rank k , its columns can be chosen orthonormal which turned out favorable in our numerical tests. We obtain orthonormal columns in U by replacing the current estimate U, S by Q, RSR^T where Q and R are the QR -factors of U .

Remark 1 A kernel function $\kappa(u, v)$ is nonnegative, i.e. $\kappa(u, v) \geq 0$. This implies that the matrix K is elementwise non-negative as well, which cannot be guaranteed without imposing complicated constraints on U and S . It is possible to restrict U and S to non-negative matrices as well (which guarantees K to be non-negative). This, however, significantly reduces the search space for a fixed rank k and does not allow for orthonormal columns of U . Details about non-negative matrix factorization (NNMF) are available in [4] and [9] but will not be used in this work.

Low-rank optimization has seen some attention (e.g. [6, 16]). In [3], the influence of missing data was addressed.

4 Numerical results

This section is devoted to numerical results using the proposed method to reconstruct a kernel from given (measured or simulated) data. In this work, in order to be able to validate our results, we will reconstruct the following four different kernels from “measurements” $F(v_j, t_i)$ obtained through numerical simulation,

$$\text{Brownian} : \kappa_B(u, v) = \left(u^{1/3} + v^{1/3}\right) \cdot \left(u^{-1/3} + v^{-1/3}\right) \tag{10}$$

$$\text{Shear} : \kappa_S(u, v) = 2 \cdot \left(u^{1/3} + v^{1/3}\right)^{7/3} \tag{11}$$

$$\text{Sum} : \kappa_\Sigma(u, v) = 5 \cdot (u + v) \tag{12}$$

$$\text{Peglow} : \kappa_P(u, v) = 3 \cdot (u + v)^{0.7105} \cdot (uv)^{-0.062} \tag{13}$$

which are plotted in Fig. 1. We note that the kernels κ_B and κ_Σ are separable, i.e. the kernel matrices can be represented in factored form (5) with $k = 3$ and $k = 2$, respectively. The discretized Brownian and sum kernels κ_B, κ_Σ can be written in the form

$$K_B : U_{i,1} = v_i^{1/3}, U_{i,2} = 1, U_{i,3} = v_i^{-1/3}, S = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \tag{14}$$

$$K_\Sigma : U_{i,1} = v_i, U_{i,2} = 5, S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{15}$$

respectively, where $v_i, i = 1, \dots, n$, are the pivots defined in (4), giving rise to initialize our optimizations with an antidiagonal matrix S . The other two kernels

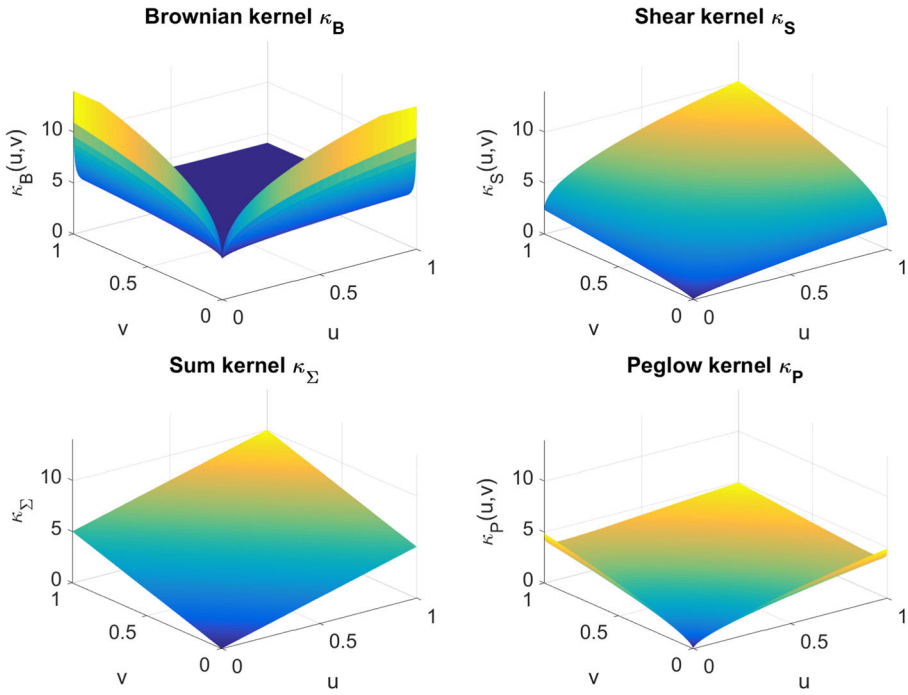


Fig. 1 Four different aggregation kernels $\kappa(u, v)$

are not separable but can be approximated using a low-rank k so that the resulting discretization error is dominated by the number of pivots n .

We use a bimodal initial distribution

$$f_{\text{init}}(v) = c \cdot \left(e^{-20v} + e^{-300 \cdot (v-0.2)^2} \right) \tag{16}$$

where the scaling coefficient c normalizes the function to first moment $\mu^1(f, t) = 10^{-2}$ with $v^{\text{max}} = 1$. We will use $v^{\text{max}} = 1$ in all our numerical tests and hence leave out the superscript in $\mathcal{G}_p := \mathcal{G}_p^1$.

We obtain reference solutions by solving (1) with respect to a very fine grid \mathcal{G}_{17} and take “measurements” for $m + 1 = 6$ equidistant time instances $t_i = i$ for $i \in \{0, \dots, 5\}$. We obtain $\tilde{F}(t_i) \in \mathbb{R}^{2^{17}}$ and consider it to be a distribution (perfectly) measured at time t_i . We coarsen it to the grid \mathcal{G}_{10} and gather the particles at the pivots by summing over each set of 2^7 entries to obtain $F(t_i) \in \mathbb{R}^{1024}$. Throughout this section, we will use the following notation for discrete density distributions:

Grid	Reconstructed kernel	Exact kernel
\mathcal{G}_{10}	\tilde{f}	F
\mathcal{G}_{17}	f	\tilde{F}

The distributions $F(v, t_i)$ for κ_S and κ_P are shown in Fig. 2. We see that the distributions have similar shapes and both have hardly any particles of mass greater than 0.5.

We present numerical tests for two variants of the optimization problem, one with a fixed matrix factor U , i.e. optimization only with respect to S , and one with both U and S (5) included in the optimization.

4.1 Optimization including the kernel basis U

For a variable U , we start the optimization process with rank $k = 5$ in (5) and

$$U_{j,\gamma} = v_j^{1/\gamma}, \quad S = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{17}$$

We use this initial kernel because it is the pointwise evaluation of a smooth function and gives only moderate aggregation rates in the considered domain. High rates of aggregation will result in very small time steps in the solution of the differential equation to ensure the positivity of f . Our choice of S mimics the matrix present in the kernels κ_B (14) and κ_Σ (15).

The kernel matrix that solves the optimization problem (9) is denoted by $K_{\text{est}} = U_{\text{est}} S_{\text{est}} U_{\text{est}}^T$ for each of the four kernels. We also compute the pointwise relative errors

$$E_{\text{rel}}(v_i, v_j) = \frac{|K_{\text{est}}(v_i, v_j) - K_{\text{true}}(v_i, v_j)|}{K_{\text{true}}(v_i, v_j)}, \quad 1 \leq i, j \leq n, \tag{18}$$

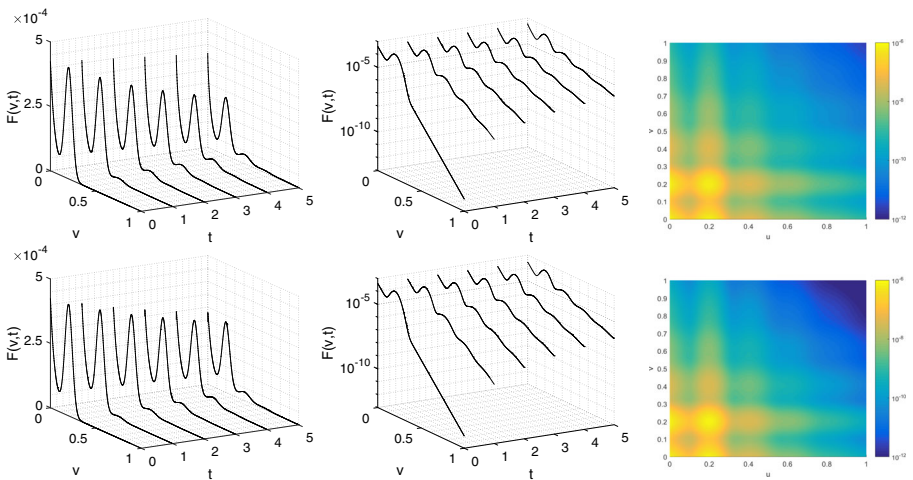


Fig. 2 Evolution of the particle distribution $F(v, t)$ for time instances $t_i = 0, \dots, 5$, with the shear kernel κ_S (top) and the Peglow kernel κ_P (bottom) for linear (left) and logarithmic (middle) scaling of the y -axis. The right column shows $F \cdot F^T$ on a logarithmic scale

and plot these in Fig. 3.

For all four kernels, we see a minimum of the relative error around (0.2, 0.2) which we attribute to our choice of initial condition with a peak at 0.2. We are not concerned about larger relative errors in the upper right triangle ($v_i + v_j > 1$) since the aggregation of two particles to one with mass greater than 1 and hence out of our computational domain should not occur, hence there is no (or hardly any) data to estimate the kernel in this region. The relative error of the shear kernel κ_S is small over the entire domain, a similar result is observed for the Brownian kernel κ_B . The sum and Peglow kernel approximation errors are smallest where most data is available—this follows from comparison with the plots of $F \cdot F^T$ in Fig. 2 (right).

The sum kernel κ_Σ shows a large error for the aggregation rate involving very small particles even though there is enough data for an accurate estimation. We attribute this to the coarse discretization.

For additional validation of our kernel estimates, we use them for simulations with a different initial distribution and compare the obtained results with the simulations with the correct kernel. The approximation was obtained by fitting a given initial

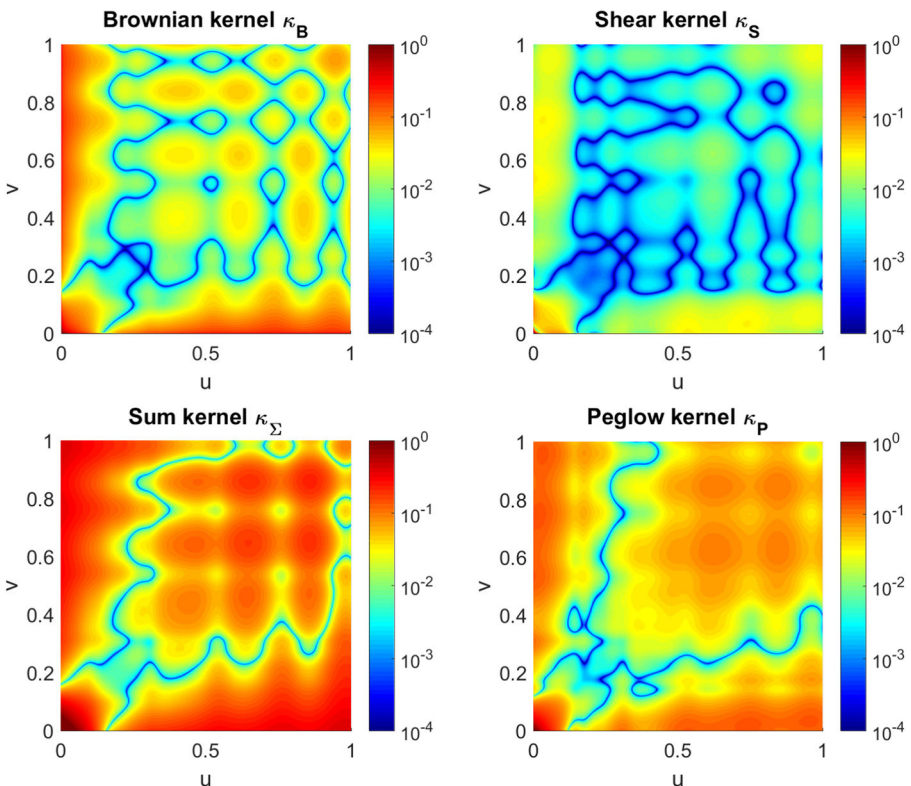


Fig. 3 (Logarithm of the) relative error (18) of the kernel with the kernel basis U included in the optimization

distribution f but should also hold a certain accuracy for other distributions to be of general use. We use the initial distribution

$$g(v, 0) = c \cdot v \cdot e^{-200(v-0.1)^2} \tag{19}$$

(the factor c is again used for normalization of $g(v, 0)$ to the first moment $\mu^1(g, t) = 10^{-2}$) and calculate $\tilde{G}(v, 10)$ with the reference kernel and $\tilde{g}(v, 10)$ with the approximated kernel factors U_{est} and S_{est} with respect to the very fine grid \mathcal{G}_{17} . We chose $g(v, 0)$ in view of its maximum at $v = 0.1$ which is near the local minimum of $f(v, 0)$. In Fig 4, we show $\tilde{G}(v, 10)$ and $\tilde{g}(v, 10)$ on the interval $[0, 0.4]$ (the simulation was computed on $[0, 1]$).

We also calculate the relative L_2 error

$$err = \left(\frac{\sum_{i=1}^{217} (\tilde{G}_i(10) - \tilde{g}_i(10))^2}{\sum_{i=1}^{217} \tilde{G}_i(10)^2} \right)^{1/2} \tag{20}$$

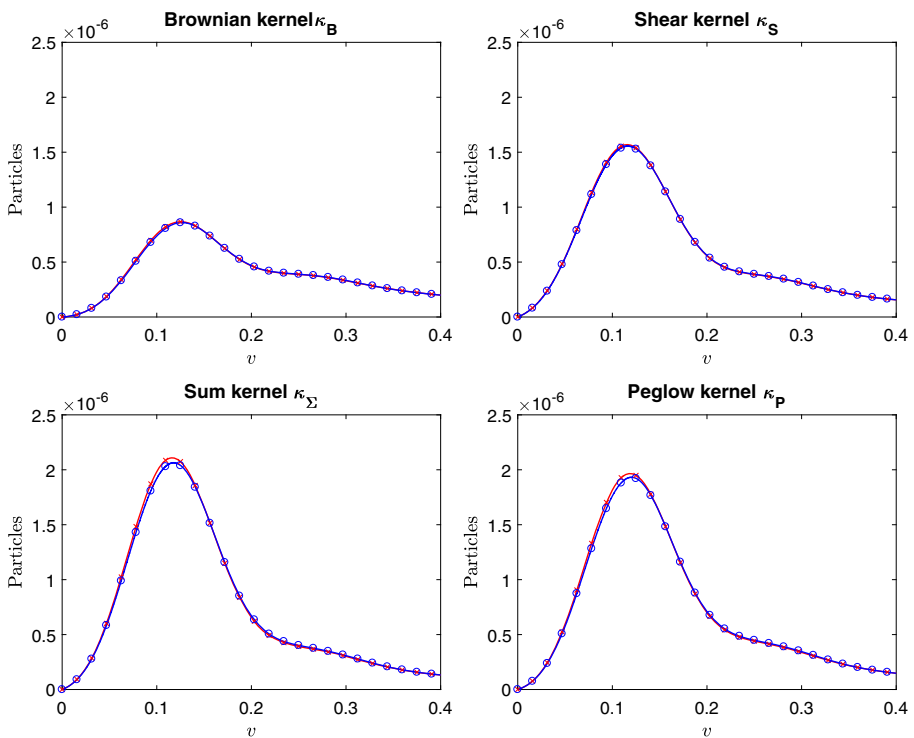


Fig. 4 Particle distributions $\tilde{G}(v, 10)$ (red) and $\tilde{g}(v, 10)$ (blue) for initial distribution g (19) for kernels and their approximations based on initial distribution f (16) with kernel basis U subject to optimization for $v \in [0, 0.4]$

Table 1 Relative L_2 error (20) between \tilde{G} and \tilde{g} with the kernel basis U included in the optimization for kernel approximations based on different grids \mathcal{G}_p

kernel \ p	κ_B	κ_S	κ_Σ	κ_P
5	$4.53 \cdot 10^{-2}$	$3.23 \cdot 10^{-2}$	$3.89 \cdot 10^{-2}$	$3.82 \cdot 10^{-2}$
6	$1.35 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$3.01 \cdot 10^{-2}$	$2.47 \cdot 10^{-2}$
7	$1.22 \cdot 10^{-2}$	$9.80 \cdot 10^{-3}$	$2.79 \cdot 10^{-2}$	$2.50 \cdot 10^{-2}$
8	$1.13 \cdot 10^{-2}$	$6.59 \cdot 10^{-3}$	$2.64 \cdot 10^{-2}$	$1.43 \cdot 10^{-2}$
9	$1.40 \cdot 10^{-2}$	$5.06 \cdot 10^{-3}$	$2.60 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$
10	$1.40 \cdot 10^{-2}$	$4.59 \cdot 10^{-3}$	$2.59 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$

between $\tilde{G}(v, 10)$ and $\tilde{g}(v, 10)$ for each kernel based on approximations for different grids \mathcal{G}_5 to \mathcal{G}_{10} and present the results in Table 1. On the positive side, we have relative approximation accuracies of order $\mathcal{O}(10^{-2})$ on a relatively coarse grid with 2^5 pivots, i.e. a gridwidth of $h = 2^{-5} = 0.03125$. However, we observe hardly any improvement with respect to further refinement of the grid even though, as we

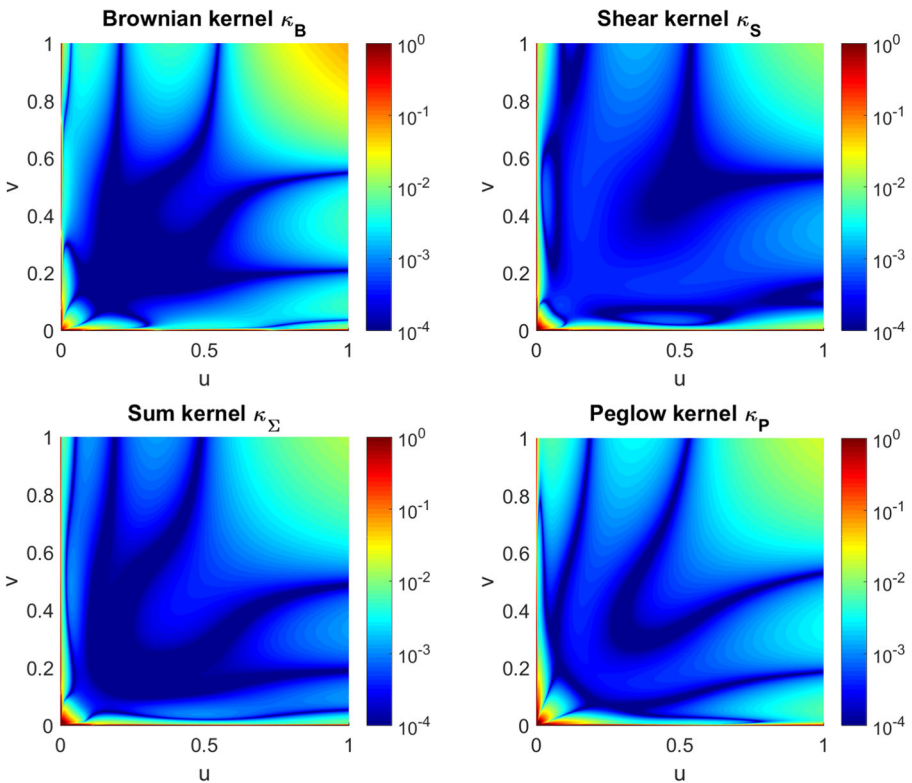


Fig. 5 (Logarithm of the) relative error (18) of the kernel approximation for a fixed kernel basis U

will see in the following subsection, the framework would indeed allow for higher accuracies.

4.2 Optimization with a fixed kernel basis U

We now fix the kernel basis to $U = (u_{j\gamma}) \in \mathbb{R}^{n,k}$ with $k = 7$ and entries $u_{j,\gamma} = v_j^{(\gamma-4)/3}$ for $\gamma = 1, \dots, 7$. The pivots $v_j, j = 1, \dots, n$, have been defined in (4). The optimization occurs now only with respect to the symmetric matrix $S \in \mathbb{R}^{k \times k}$. We initialize S by $S = 0$ since we experienced better results compared to an initialization with an antitriangular matrix (which was not the case in the previous setting where U was included in the optimization). We are able to allow for a larger rank k in this setting since the number of degrees of freedom is no longer linear in the number of pivots n but only quadratic in the maximum rank k of the kernel matrix.

In Fig. 5, we show the relative approximation errors (18).

We observe a clear improvement compared to the kernels estimated with variable U whose errors were shown in Fig. 3 and offer two interpretations of this result: (i) The fixed basis U has been chosen to span a space that allows for accurate approximations of the exact kernels; hence, the search space has been reduced significantly,

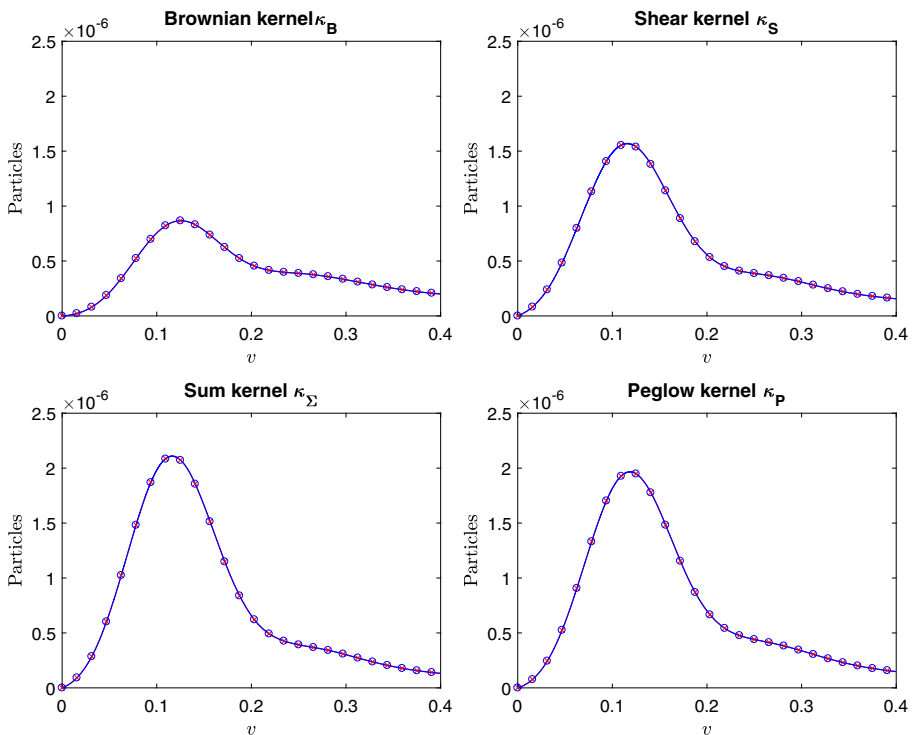


Fig. 6 Particle distributions $\tilde{G}(v, 10)$ (red) and $\tilde{g}(v, 10)$ (blue) for initial distribution g (19) for kernels and their approximations based on initial distribution f (16) with fixed basis U for $v \in [0, 0.4]$

Table 2 Relative L_2 error (20) between \tilde{G} and \tilde{g} for a fixed kernel basis U with approximations based on different grids \mathcal{G}_p

kernel \ p	κ_B	κ_S	κ_Σ	κ_P
5	$3.94 \cdot 10^{-1}$	$4.75 \cdot 10^{-1}$	$5.66 \cdot 10^{-1}$	$5.52 \cdot 10^{-1}$
6	$1.04 \cdot 10^{-1}$	$2.93 \cdot 10^{-1}$	$2.30 \cdot 10^{-1}$	$2.49 \cdot 10^{-1}$
7	$2.97 \cdot 10^{-2}$	$2.58 \cdot 10^{-2}$	$1.51 \cdot 10^{-2}$	$3.57 \cdot 10^{-2}$
8	$4.64 \cdot 10^{-3}$	$5.46 \cdot 10^{-3}$	$3.39 \cdot 10^{-3}$	$3.00 \cdot 10^{-2}$
9	$1.32 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	$9.66 \cdot 10^{-3}$
10	$2.32 \cdot 10^{-4}$	$8.59 \cdot 10^{-4}$	$7.03 \cdot 10^{-4}$	$1.12 \cdot 10^{-3}$

allowing for an easier optimization. (ii) Kernel estimation is an inverse problem which we propose to solve using a (non-convex) optimization. There could exist several (local) minima. In the end, we are estimating a kernel that simulates results close to the measured results, not a kernel that is close to another kernel (in our experiments given, but in practice unknown).

Using these estimated kernels together with an initial distribution $g(v, 0) = c \cdot v \cdot e^{-200(v-0.1)^2}$, the resulting distributions $\tilde{g}(v, 10)$ are shown in Fig. 6. They are indeed more accurate than those obtained for a variable U which were shown in Fig. 4. The relative L_2 errors for approximations based on grids \mathcal{G}_5 to \mathcal{G}_{10} are shown in Table 2. Comparing to the respective results for an optimized U in Table 1, we see less accurate kernel approximations on the coarser grids ($p = 5, 6$) but then a decrease of order up to $\mathcal{O}(h^2)$ with respect to the grid width $h = 2^{-p}$ (errors decrease by factors close to 4 when the grid is refined, i. e. when h is divided by 2).

5 Conclusions and future work

We have presented a novel framework for the approximation of aggregation kernels in population balance equations from measured or previously simulated data. We do not require a high temporal resolution of the measured data since we do *not* use it to approximate the (time) derivative on the left hand side on (1) as is done in several previous works. This allows estimation of an aggregation kernel from population data that is spaced in time without information at intermediate time instances.

The main idea is the assumption of a discrete low-rank kernel of the form $K = USU^T$ which allows the fast evaluation of aggregation integrals introduced in [8, 10] in nonlinear optimization procedures (here: MATLAB’s *lsqnonlin*). We presented two approaches, one with a fixed kernel basis U and optimization with respect to $\frac{1}{2}k(k+1)$ entries in the symmetric matrix S , and one with U included in the optimization. Our numerical tests indicate that the approach with a variable U is preferable only for a rather small number of pivots (here less than 100) whereas using a fixed basis U yields approximation results that depend (up to) quadratically on the grid with h .

In addition to producing results of higher accuracy, the approach with a fixed kernel basis U is also preferable with respect to computational complexity which is $\mathcal{O}(kn \log n)$ compared to $\mathcal{O}(kn^2 \log n)$ for a variable U . Presenting some actual simulation timings, using a variable U took 2 min for $n = 2^7 = 128$ pivots and 22 minutes for $n = 2^{10} = 1024$. Using a fixed kernel, using $n = 2^{10} = 1024$ pivots can still be done in under a minute.

Possible extensions in the future include the analysis for noisy data as it might occur in actual measurements for physical particles.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors gratefully acknowledge the financial support of this work by the Deutsche Forschungsgemeinschaft (DFG) under the Grant BO 4141/1-3 in the framework of the research priority program SPP 1679 “Dynamische Simulation vernetzter Feststoffprozesse”.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bramley, A.S., Hounslow, M.J., Ryall, R.L.: Aggregation during precipitation from solution: A method for extracting rates from experimental data. *J. Colloid Interface Sci.* **183**(1), 155–165 (1996)
2. Brunel, N.J.-B.: Parameter estimation of ODEs via nonparametric estimators. *Electronic Journal of Statistics* **2**(0), 1242–1267 (2008)
3. Buchanan, A., Fitzgibbon, A.: Damped Newton algorithms for matrix factorization with missing data. 2012 IEEE Conference on Computer Vision and Pattern Recognition **2**, 316–322 (2005)
4. Catral, M., Han, L., Neumann, M., Plemmons, R.J.: On reduced rank nonnegative matrix factorization for symmetric nonnegative matrices. *Linear Algebra and its Applications* **393**, 107–126 (2004)
5. Chakraborty, J., Kumar, J., Singh, M., Mahoney, A., Ramkrishna, D.: Inverse problems in population balances. Determination of aggregation kernel by weighted residuals. *Industrial & Engineering Chemistry Research* **54**(42), 10530–10538 (2015)
6. Chi, Y., Lu, Y.M., Chen, Y.: Nonconvex optimization meets low-rank matrix factorization: an overview. *IEEE Trans. Signal Process* **67**, 5239–5269 (2018)
7. Eisenschmidt, H., Soumaya, M., Bajcinca, N., Le Borne, S., Sundmacher, K.: Estimation of aggregation kernels based on Laurent polynomial approximation. *Comput. Chem. Eng.* **103**, 210–217 (2017)
8. Hackbusch, W.: On the efficient evaluation of coalescence integrals in population balance models. *Computing* **78**, 145–159 (2006)
9. Huang, K., Sidiropoulos, N.D., Swami, A.: Non-negative matrix factorization revisited: uniqueness and algorithm for symmetric decomposition. *IEEE Trans. Signal Process.* **62**(1), 211–224 (2014)
10. Le Borne, S., Shahmuradyan, L., Sundmacher, K.: Fast evaluation of univariate aggregation integrals on equidistant grids. *Comput. Chem Eng.* **74**, 115–127 (2015)
11. Vidyullatha, P., Rajeswara Rao, D.: Machine learning techniques on multidimensional curve fitting data based on r-square and chi-square methods. *International Journal of Electrical and Computer Engineering (IJECE)* **6**(3), 974 (2016)
12. Peglow, M., Kumar, J., Warnecke, G., Heinrich, S., Mörl, L.: A new technique to determine rate constants for growth and agglomeration with size- and time-dependent nuclei formation. *Chem. Eng. Sci.* **61**(1), 282–292 (2006)

13. Peifer, M., Timmer, J.: Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *IET Syst. Biol.* **1**(2), 78–88 (2007)
14. Poyton, A.A., Varziri, M.S., McAuley, K.B., McLellan, P.J., Ramsay, J.O.: Parameter estimation in continuous-time dynamic models using principal differential analysis. *Comput. Chem. Eng.* **30**(4), 698–708 (2006)
15. Ramachandran, R., Barton, P.I.: Effective parameter estimation within a multi-dimensional population balance model framework. *Chem. Eng. Sci.* **65**(16), 4884–4893 (2010)
16. Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y.: Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. In: Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I., Culotta, A. (eds.) *Advances in neural information processing systems*, vol. 22, pp. 2080–2088. Curran Associates, Inc. (2009)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.