ON OPTIMAL DECAY ESTIMATES FOR ODES AND PDES WITH MODAL DECOMPOSITION

FRANZ ACHLEITNER, ANTON ARNOLD, AND BEATRICE SIGNORELLO

Abstract. We consider the Goldstein-Taylor model, which is a 2-velocity BGK model, and construct the "optimal" Lyapunov functional to quantify the convergence to the unique normalized steady state. The Lyapunov functional is optimal in the sense that it yields decay estimates in L^2-norm with the sharp exponential decay rate and minimal multiplicative constant. The modal decomposition of the Goldstein-Taylor model leads to the study of a family of 2-dimensional ODE systems. Therefore we discuss the characterization of "optimal" Lyapunov functionals for linear ODE systems with positive stable diagonalizable matrices. We give a complete answer for 2-dimensional ODE systems, and a partial answer for higher dimensional ODE systems.

1. Introduction

This note is concerned with optimal decay estimates of hypocoercive evolution equations that allow for a modal decomposition. The notion *hypocoercivity* was introduced by Villani in [14] for equations of the form $\frac{\mathrm{d}}{\mathrm{d}t}f = -Lf$ on some Hilbert space H, where the generator L is not coercive, but where solutions still exhibit exponential decay in time. More precisely, there should exist constants $\lambda > 0$ and $c \geq 1$, such that

$$\|e^{-Lt} f^I\|_{\widetilde{H}} \leq c\, e^{-\lambda t} \|f^I\|_{\widetilde{H}} \qquad \forall\, f^I \in \widetilde{H}\,, \tag{1.1}$$

where $\widetilde{H}$ is a second Hilbert space, densely embedded in $(\ker L)^\perp \subset H$.

The large-time behavior of many hypocoercive equations have been studied in recent years, including Fokker-Planck equations [3,5,14], kinetic equations [11] and BGK equations [2]. Determining the sharp (i.e. maximal) exponential decay rate λ was an issue in some of these works, in particular [2,5]. But finding at the same time the smallest multiplicative constant $c \geq 1$, is so far an open problem. And this is the topic of this note. For simple cases we shall describe a procedure to construct the "optimal" Lyapunov functional that will imply (1.1) with the sharp constants λ and c.

For illustration purposes we shall focus here only on the following 2-velocity BGK-model (referring to the physicists Bhatnagar, Gross and Krook [7]) for the two functions $f_\pm(x,t) \geq 0$ on the one-dimensional torus $x \in \mathbb{T}$

Key words and phrases. Lyapunov functionals, sharp decay estimates, Goldstein-Taylor model.

All authors were supported by the FWF-funded SFB #F65. The second author was partially supported by the FWF-doctoral school "Dissipation and dispersion in nonlinear partial differential equations".

and for $t \geq 0$. It reads

$$\begin{cases} \partial_t f_+ &= -\partial_x f_+ + \frac{1}{2}(f_- - f_+), \\ \partial_t f_- &= \partial_x f_- - \frac{1}{2}(f_- - f_+). \end{cases} \tag{1.2}$$

This system of two transport-reaction equations is also called *Goldstein-Taylor model.*

For initial conditions normalized as $\int_0^{2\pi} \left[f_+^I(x) + f_-^I(x) \right] dx = 2\pi$, the solution $f(t) = (f_+(t), f_-(t))^\top$ converges to its unique (normalized) steady state with $f_+^\infty = f_-^\infty = \frac{1}{2}$. An explicit exponential decay rate of this two velocity model by means of Lyapunov functionals was shown in [11, §1.4]. The sharp exponential decay estimate was found in [1, §4.1] via a refined functional, yielding the following result:

Theorem 1.1 ([1, Th. 6])**.** *Let* $f^I \in L^2(0, 2\pi; \mathbb{R}^2)$*. Then the solution to* (1.2) *satisfies*

$$\|f(t) - f^\infty\|_{L^2(0,2\pi;\mathbb{R}^2)} \leq c\, e^{-\lambda t} \|f^I - f^\infty\|_{L^2(0,2\pi;\mathbb{R}^2)} , \qquad t \geq 0 ,$$

with the optimal constants $\lambda = \frac{1}{2}$ *and* $c = \sqrt{3}$.

Remark 1.1.
a) The optimal c was not specified in [1], but will actually be the result of Lemma 3.2 below.
b) As we shall illustrate in §4, it does *not* make sense to optimize these two constants at the same time. The optimality in Theorem 1.1 refers to first maximizing the exponential rate λ, and then to minimize the multiplicative constant c.

The proof of Theorem 1.1 is based on the spatial Fourier transform of (1.2), cf. [1,11]. We denote the Fourier modes in the discrete velocity basis $\{\binom{1}{1}, \binom{1}{-1}\}$ by $u_k(t) \in \mathbb{C}^2$, $k \in \mathbb{Z}$. They evolve according to the ODE systems

$$\frac{\mathrm{d}}{\mathrm{d}t} u_k = -\mathbf{C}_k\, u_k\,, \quad \mathbf{C}_k = \begin{pmatrix} 0 & ik \\ ik & 1 \end{pmatrix}, \quad k \in \mathbb{Z}\,, \tag{1.3}$$

and their (normalized) steady states are

$$u_0^\infty = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad u_k^\infty = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad k \neq 0\,.$$

In the main body of this note we shall construct appropriate Lyapunov functionals for such ODEs, in order to obtain sharp decay rates of the form (1.1). In the context of the BGK-model (1.2), combining such decay estimates for all modes u_k then yields Theorem 1.1. We remark that the construction of Lyapunov functionals to reveal optimal decay rates in ODEs was already included in the classical textbook [4, §22.4], but optimality of the multiplicative constant c was not an issue there.

In this article we shall first review, from [1, 2], the construction of Lyapunov functionals for linear first order ODE systems that reveal the sharp decay rate. As these functionals are not uniquely determined, we shall then find the unique, "best Lyapunov" functional in §3 – by minimizing the constant c. We shall give a complete answer for 2-dimensional ODE systems, and a partial answer for higher dimensions. In the final section §4 we shall illustrate how to obtain a whole family of decay estimates – with suboptimal decay rates, but improved constant c. For small time this significantly improves the estimate obtained in §3.

2. Lyapunov functionals for hypocoercive ODEs

In this section we review decay estimates for linear ODEs with constant coefficients of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} f = -\mathbf{C} f, \quad t \geq 0\,, \\ f(0) = f^I \in \mathbb{C}^n\,, \end{cases} \tag{2.1}$$

for some (typically non-Hermitian) matrix $\mathbf{C} \in \mathbb{C}^{n\times n}$. We assume that the matrix $\mathbf{C}$ is *hypocoercive* (i.e. positive stable, meaning that all eigenvalues have positive real part). Since we shall *not* require that $\mathbf{C}$ is coercive (meaning that its Hermitian part would be positive definite), we *cannot* expect that all solutions to (2.1) satisfy for the Euclidean norm: $\|f(t)\|_2 \leq e^{-\widetilde{\lambda} t}\|f^I\|_2$ for some $\widetilde{\lambda} > 0$. However, such an exponential decay estimate does hold in an adapted norm that can be used as a Lyapunov functional.

The construction of this Lyapunov functional is based on the following lemma:

Lemma 2.1 ([1, Lemma 2], [5, Lemma 4.3])**.** *For any fixed matrix $\mathbf{C} \in \mathbb{C}^{n\times n}$, let $\mu := \min\{\Re(\lambda) | \lambda$ is an eigenvalue of $\mathbf{C}\}$. Let $\{\lambda_j | 1 \leq j \leq j_0\}$ be all the eigenvalues of $\mathbf{C}$ with $\Re(\lambda_j) = \mu$. If all λ_j $(j = 1, \ldots, j_0)$ are non-defective*[1]*, then there exists a positive definite Hermitian matrix $\mathbf{P} \in \mathbb{C}^{n\times n}$ with*

$$\mathbf{C}^*\mathbf{P} + \mathbf{P}\mathbf{C} \geq 2\mu\mathbf{P}\,, \tag{2.2}$$

but $\mathbf{P}$ is not uniquely determined. Moreover, if all eigenvalues of $\mathbf{C}$ are non-defective, examples of such matrices $\mathbf{P}$ satisfying (2.2) *are given by*

$$\mathbf{P} := \sum_{j=1}^{n} b_j\, w_j \otimes w_j^*\,, \tag{2.3}$$

where $w_j \in \mathbb{C}^n$ $(j = 1, \ldots, n)$ denote the (right) normalized eigenvectors of $\mathbf{C}^*$ (i.e. $\mathbf{C}^* w_j = \bar{\lambda}_j w_j$), and $b_j \in \mathbb{R}^+$ $(j = 1, \ldots, n)$ are arbitrary weights.

For $n = 2$ all positive definite Hermitian matrices $\mathbf{P}$ satisfying (2.2) have the form (2.3), but for $n \geq 3$ this is not true (see Lemma 3.1 and a subsequent remark, respectively).

In this article, for simplicity, we shall only consider the case when all eigenvalues of $\mathbf{C}$ are non-defective. For the extension of Lemma 2.1 and of the corresponding decay estimates to the defective case we refer to [3, Prop. 2.2] and [6].

Due to the positive stability of $\mathbf{C}$, the origin is the unique and asymptotically stable steady state $f^\infty = 0$ of (2.1): Due to Lemma 2.1, there exists a positive definite Hermitian matrix $\mathbf{P} \in \mathbb{C}^{n\times n}$ such that $\mathbf{C}^*\mathbf{P} + \mathbf{P}\mathbf{C} \geq 2\mu\mathbf{P}$

[1] An eigenvalue is defective if its geometric multiplicity is strictly less than its algebraic multiplicity.

where $\mu = \min \Re(\lambda_j) > 0$. Thus, the time derivative of the adapted norm $\|f\|_{\mathbf{P}}^2 := \langle f \ , \mathbf{P}f\rangle$ along solutions of (2.1) satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f(t)\|_{\mathbf{P}}^2 \leq -2\mu\|f(t)\|_{\mathbf{P}}^2 \,,$$

which implies

$$\|f(t)\|_{\mathbf{P}}^2 \leq e^{-2\mu t}\|f^I\|_{\mathbf{P}}^2 \,, \qquad t \geq 0 \,.$$

Next we translate this decay in $\mathbf{P}$-norm into a decay in the Euclidean norm:

$$\|f(t)\|_2^2 \leq (\lambda_{\min}^{\mathbf{P}})^{-1}\|f(t)\|_{\mathbf{P}}^2 \leq (\lambda_{\min}^{\mathbf{P}})^{-1}e^{-2\mu t}\|f^I\|_{\mathbf{P}}^2 \leq \kappa(\mathbf{P})\,e^{-2\mu t}\|f^I\|_2^2 \,, \quad t \geq 0 \,, \tag{2.4}$$

where $\lambda_{\min}^{\mathbf{P}}$, $\lambda_{\max}^{\mathbf{P}}$ are, respectively, the (positive) smallest and largest eigenvalues of $\mathbf{P}$, and $\kappa(\mathbf{P}) = \lambda_{\max}^{\mathbf{P}}/\lambda_{\min}^{\mathbf{P}}$ is the (numerical) condition number of $\mathbf{P}$ with respect to the Euclidean norm.

Given the spectrum of $\mathbf{P}$, the exponential decay rate in (2.4) is optimal. But since the matrix $\mathbf{P}$ is not unique, the next goal is to minimize the multiplicative constant in (2.4). To this end we need to find the matrix $\mathbf{P}$ with minimal condition number that satisfies (2.2). Clearly, the answer can only be unique up to a positive multiplicative constant, since $\widetilde{\mathbf{P}} := \tau\mathbf{P}$ with $\tau > 0$ would reproduce the estimate (2.4).

As we shall prove in §3, the answer to this minimization problem is very easy in 2 dimensions: The best $\mathbf{P}$ corresponds to equal weights in (2.3), e.g. choosing $b_1 = b_2$.

3. Minimization of the condition number

In this section, we describe a procedure to construct the "optimal" Lyapunov functional that will imply (1.1) with the sharp constants λ and c.

We shall describe the procedure in case of ODEs (2.1) with positive stable matrices $\mathbf{C}$. For simplicity we confine ourselves to diagonalizable matrices. For diagonalizable matrices $\mathbf{C}$ (i.e. all eigenvalues are non-defective), Lemma 2.1 states that all matrices $\mathbf{P}$ of form (2.3) satisfy the matrix inequality (2.2). Then, (2.4) shows that the constant c in (1.1) is given by $c = \sqrt{\kappa(\mathbf{P})}$. Our strategy is now to minimize $\kappa(\mathbf{P})$ on the set of matrices of form (2.3) – for any $n \geq 2$. We shall prove that this actually yields the minimal constant c for $n = 2$ that is achievable via the chain of inequalities (2.4). But for $n \geq 3$ this procedure does not always yield the minimal κ for all matrices satisfying (2.2).

Defining a matrix $\mathbf{W} := (w_1|\dots|w_n)$ whose columns are the (right) normalized eigenvectors of $\mathbf{C}^*$ allows to rewrite formula (2.3) as

$$\begin{aligned}\mathbf{P} &= \sum_{j=1}^n b_j\, w_j \otimes w_j^* = \mathbf{W}\,\mathrm{diag}(b_1, b_2, \dots, b_n)\mathbf{W}^* \\ &= \big(\mathbf{W}\,\mathrm{diag}(\sqrt{b_1}, \sqrt{b_2}, \dots, \sqrt{b_n})\big)\big(\mathbf{W}\,\mathrm{diag}(\sqrt{b_1}, \sqrt{b_2}, \dots, \sqrt{b_n})\big)^*\end{aligned} \tag{3.1}$$

with positive constants b_j $(j = 1, \dots, n)$. The identity

$$\widetilde{\mathbf{W}} := \mathbf{W}\,\mathrm{diag}(\sqrt{b_1}, \sqrt{b_2}, \dots, \sqrt{b_n}) = (\sqrt{b_1}w_1|\dots|\sqrt{b_n}w_n)$$

shows that the weights are just rescalings of the eigenvectors. Finally, the condition number of $\mathbf{P}$ is the squared condition number of $\widetilde{\mathbf{W}}$. Hence, to find matrices **P** of form (3.1) with minimal condition number, is equivalent to identifying (right) precondition matrices among the positive definite diagonal matrices which minimize the condition number of **W**. This minimization problem can be formulated as a convex optimization problem [9] based on the result [13]. Due to [10, Theorem 1], the minimum is attained (i.e. an optimal scaling matrix exists) since our matrix **W** is non-singular. (Note that its column vectors form a basis of $\mathbb{C}^n$.) The convex optimization problem can be solved by standard software providing also the exact scaling matrix which minimizes the condition number of **P**, see the discussion and references in [9]. For more information on convex optimization and numerical solvers, see e.g. [8].

In 2 dimensions this minimization problem can be solved very easily:

Lemma 3.1. *Let $\mathbf{C} \in \mathbb{C}^{2\times 2}$ be a diagonalizable positive stable matrix. Then all matrices* **P** *satisfying* (2.2) *are of the form* (2.3)*.*

Proof. We use again the matrix **W** whose columns are the normalized (right) eigenvectors of $\mathbf{C}^*$ such that

$$\mathbf{C}^*\mathbf{W} = \mathbf{W}\mathbf{D}^*\,, \tag{3.2}$$

with $\mathbf{D} = \operatorname{diag}(\lambda_1^{\mathbf{C}}, \lambda_2^{\mathbf{C}})$ where $\lambda_j^{\mathbf{C}}$ ($j \in \{1,2\}$) are the eigenvalues of **C**. Since **W** is regular, **P** can be written as

$$\mathbf{P} = \mathbf{W}\mathbf{B}\mathbf{W}^*\,,$$

with some positive definite Hermitian matrix **B**. Then the matrix inequality (2.2) can be written as

$$2\mu\mathbf{W}\mathbf{B}\mathbf{W}^* \le \mathbf{C}^*\mathbf{W}\mathbf{B}\mathbf{W}^* + \mathbf{W}\mathbf{B}\mathbf{W}^*\mathbf{C} = \mathbf{W}(\mathbf{D}^*\mathbf{B} + \mathbf{B}\mathbf{D})\mathbf{W}^*\,.$$

This matrix inequality is equivalent to

$$0 \le (\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I})\,. \tag{3.3}$$

Next we order the eigenvalues $\lambda_j^{\mathbf{C}}$ ($j \in \{1,2\}$) of **C** increasingly with respect to their real parts, such that $\Re(\lambda_1^{\mathbf{C}}) = \mu$. Moreover, we consider

$$\mathbf{B} = \begin{pmatrix} b_1 & \beta \\ \overline{\beta} & b_2 \end{pmatrix}$$

where $b_1, b_2 > 0$ and $\beta \in \mathbb{C}$ with $|\beta|^2 < b_1 b_2$. Then the right hand side of (3.3) is

$$(\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I}) = \begin{pmatrix} 0 & (\lambda_2^{\mathbf{C}} - \lambda_1^{\mathbf{C}})\beta \\ \overline{(\lambda_2^{\mathbf{C}} - \lambda_1^{\mathbf{C}})\beta} & 2b_2\Re(\lambda_2^{\mathbf{C}} - \lambda_1^{\mathbf{C}}) \end{pmatrix} \tag{3.4}$$

with $\operatorname{Tr}[(\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I})] = 2b_2\Re(\lambda_2^{\mathbf{C}} - \lambda_1^{\mathbf{C}})$ and

$$\det[(\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I})] = -\left|\lambda_2^{\mathbf{C}} - \lambda_1^{\mathbf{C}}\right|^2 |\beta|^2\,.$$

Condition (3.3) is satisfied if and only if $\operatorname{Tr}[(\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I})] \ge 0$ which holds due to our assumptions on $\lambda_2^{\mathbf{C}}$ and b_2, and $\det[(\mathbf{D}^* - \mu\mathbf{I})\mathbf{B} + \mathbf{B}(\mathbf{D} - \mu\mathbf{I})] \ge 0$. The last condition holds if and only if

$$\lambda_2^{\mathbf{C}} = \lambda_1^{\mathbf{C}} \quad \text{or } \beta = 0\,.$$

In the latter case **B** is diagonal and hence **P** is of the form (2.3). In the former case, (3.2) shows that $\mathbf{C} = \lambda_1^{\mathbf{C}}\mathbf{I}$, and the inequality (2.2) is trivial. Now any positive definite Hermitian matrix $\mathbf{P}$ has a diagonalization $\mathbf{P} = \mathbf{VEV}^*$, with a diagonal real matrix $\mathbf{E}$ and an orthogonal matrix $\mathbf{V}$, whose columns are –of course– eigenvectors of $\mathbf{C}$. Thus, $\mathbf{P}$ is again of the form (2.3). □

However, in dimensions $n \geq 3$, there exist matrices $\mathbf{P}$ satisfying (2.2) which are not of form (2.3). For example, consider the matrix $\mathbf{C} = \operatorname{diag}(1, 2, 3)$. Then, all matrices

$$\mathbf{P}(b_1, b_2, b_3, \beta) = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & \beta \\ 0 & \beta & b_3 \end{pmatrix} \tag{3.5}$$

with positive b_j ($j \in \{1, 2, 3\}$) and $\beta \in \mathbb{R}$ such that $8b_2b_3 - 9\beta^2 \geq 0$, are positive definite Hermitian matrices and satisfy (2.2) for $\mathbf{C} = \operatorname{diag}(1, 2, 3)$ and $\mu = 1$. But the eigenvectors of $\mathbf{C}^*$ are the canonical unit vectors. Hence, matrices of form (2.3) would all be diagonal.

Lemma 3.2. *Let $\mathbf{C} \in \mathbb{C}^{2\times 2}$ be a diagonalizable positive stable matrix. Then the condition number of the associated matrix $\mathbf{P}$ in (2.3) is minimal by choosing the weights $b_1 = b_2 = 1$.*

Proof. A diagonalizable matrix $\mathbf{C}$ has only non-defective eigenvalues. Up to a unitary transformation, we can assume w.l.o.g. that the eigenvectors of $\mathbf{C}^*$ are

$$w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \quad w_2 = \begin{pmatrix} \alpha \\ \sqrt{1-\alpha^2} \end{pmatrix} \quad \text{for some } \alpha \in [0, 1). \tag{3.6}$$

This unitary transformation describes the change of the coordinate system. To construct the new basis, we choose one of the normalized eigenvectors v_1 as first basis vector, and recall that the second normalized eigenvector v_2 is only determined up to a scalar factor $\gamma \in \mathbb{C}$ with $|\gamma| = 1$. The right choice for the scalar factor γ allows to fulfill the above restriction on α.

We use the representation of the positive definite matrix $\mathbf{P}$ in (3.1):

$$\mathbf{P} = \mathbf{W} \operatorname{diag}(b_1, b_2)\mathbf{W}^* \quad \text{with } \mathbf{W} = \begin{pmatrix} 1 & \alpha \\ 0 & \sqrt{1-\alpha^2} \end{pmatrix} .$$

Since $\mathbf{P}$ and $\tau\mathbf{P}$ have the same condition number, we consider w.l.o.g. $b_1 = 1/b$ and $b_2 = b$. Thus, we have to determine the positive parameter $b > 0$ which minimizes the condition number of

$$\mathbf{P}(b) = \mathbf{W} \operatorname{diag}(1/b, b)\mathbf{W}^* = \begin{pmatrix} \frac{1}{b} + b\alpha^2 & b\alpha\sqrt{1-\alpha^2} \\ b\alpha\sqrt{1-\alpha^2} & b(1-\alpha^2) \end{pmatrix} .$$

The condition number of matrix $\mathbf{P}(b)$ is given by

$$\kappa(\mathbf{P}(b)) = \lambda_+^{\mathbf{P}}(b)/\lambda_-^{\mathbf{P}}(b) \geq 1 ,$$

where

$$\lambda_\pm^{\mathbf{P}}(b) = \frac{\operatorname{Tr}\mathbf{P}(b) \pm \sqrt{(\operatorname{Tr}\mathbf{P}(b))^2 - 4\det\mathbf{P}(b)}}{2}$$

are the (positive) eigenvalues of $\mathbf{P}(b)$. We notice that $\operatorname{Tr}\mathbf{P}(b) = b + 1/b$ is independent of α and is a convex function of $b \in (0, \infty)$ which attains its minimum for $b = 1$. Moreover, $\det \mathbf{P}(b) = 1 - \alpha^2$ is independent of b. This implies that the condition number

$$\kappa(\mathbf{P}(b)) = \frac{\lambda_+^{\mathbf{P}}(b)}{\lambda_-^{\mathbf{P}}(b)} = \frac{1 + \sqrt{1 - \frac{4\det \mathbf{P}(b)}{(\operatorname{Tr} \mathbf{P}(b))^2}}}{1 - \sqrt{1 - \frac{4\det \mathbf{P}(b)}{(\operatorname{Tr} \mathbf{P}(b))^2}}}$$

attains its minimum at $b = 1$, taking the value $\kappa_{\min} = \frac{1+\alpha}{1-\alpha}$. □

This lemma now allows us to identify the minimal constant c in Theorem 1.1. For the matrices $\mathbf{C}_k$, $k \neq 0$ from (1.3), the transformation matrices $\mathbf{P}_k$ with $b_1 = b_2 = 1$ are given by $\mathbf{P}_0 = \mathbf{I}$ and

$$\mathbf{P}_k = \begin{pmatrix} 1 & -\frac{i}{2k} \\ \frac{i}{2k} & 1 \end{pmatrix}, \qquad \text{with} \qquad \kappa(\mathbf{P}_k) = \frac{2|k| + 1}{2|k| - 1}, \qquad k \neq 0\,.$$

Combining the decay estimates for all Fourier modes $u_k(t)$ shows that the minimal multiplicative constant in Theorem 1.1 is given by $c = \sqrt{\kappa(\mathbf{P}_{\pm 1})} = \sqrt{3}$. For a more detailed presentation how to recombine the modal estimates we refer to §4.1 in [1].

For diagonalizable positive stable matrices $\mathbf{C}$ in higher dimensions, the same construction as in 2D (using normalized eigenvectors of $\mathbf{C}^*$ and equal weights b_j) may *not* yield a matrix $\mathbf{P}$ with the lowest condition number. We give a counterexample in 3 dimensions: For some $\mathbf{C}^*$, consider its eigenvector matrix

$$\mathbf{W} := \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \operatorname{diag}\left(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}\right), \tag{3.7}$$

which has normalized column vectors. We define the matrices $\mathbf{P}(b_1, b_2, b_3) := \mathbf{W} \operatorname{diag}(b_1, b_2, b_3) \mathbf{W}^*$ for positive parameters b_1, b_2 and b_3, which are of form (2.3) and hence satisfy the inequality (2.2). In case of equal weights $b_1 = b_2 = b_3$ the condition number is $\kappa(\mathbf{P}(b_1, b_1, b_1)) \approx 15.12825876$. Using [12, Theorem 3.3], the minimal condition number $\min_{b_j} \kappa(\mathbf{P}(b_1, b_2, b_3)) \approx 13.92820324$ is attained for the scaling/weights $b_1 = 2$, $b_2 = 4$ and $b_3 = 3$.

Finally, we will combine our two examples of matrices in $\mathbb{R}^{3\times 3}$: Consider the matrix defined as

$$\widetilde{\mathbf{C}} = (\mathbf{W}^*)^{-1} \operatorname{diag}(1, 2, 3) \mathbf{W}^*$$

with $\mathbf{W}$, the eigenvector matrix of $\widetilde{\mathbf{C}}^*$, given by (3.7). Then the matrices $\widetilde{\mathbf{C}}$ and

$$\widetilde{\mathbf{P}}(b_1, b_2, b_3, \beta) := \mathbf{W} \mathbf{P}(b_1, b_2, b_3, \beta) \mathbf{W}^*$$

with matrix $\mathbf{P}(b_1, b_2, b_3, \beta)$ in (3.5) satisfy the matrix inequality (2.2) with $\mu = 1$. However, the condition number $\kappa(\widetilde{\mathbf{P}}(b_1, b_2, b_3, \beta)) \approx 5.82842780720132$ for the scaling/weights $b_1 = 2$, $b_2 = 4$, $b_3 = 3$, and $\beta = -2.45$, is much lower than for the previous case with $\beta = 0$. This shows that the matrix $\mathbf{P}$ minimizing $\kappa(\mathbf{P})$ within the set of matrices that satisfy the inequality (2.2) with maximal μ, is not necessarily of the form (2.3).

4. A family of decay estimates for hypocoercive ODEs

In this section we shall illustrate the interdependence of maximizing the decay rate λ and minimizing the multiplicative constant c in estimates like (1.1). For the ODE-system (2.1), the procedure described in Remark 1.1(b) yields the optimal bound for large time, with the sharp decay rate $\mu := \min\{\Re(\lambda) | \lambda \text{ is an eigenvalue of } \mathbf{C}\}$. But for non-coercive $\mathbf{C}$ we must have $c > 1$. Hence, such a bound cannot be sharp for short time. As a counterexample we consider the simple energy estimate (obtained by premultiplying (2.1) with f^*)

$$\|f(t)\|_2 \le e^{-\mu_s t} \|f^I\|_2 , \quad t \ge 0 ,$$

with $\mathbf{C}_s := \frac{1}{2}(\mathbf{C} + \mathbf{C}^*)$ and $\mu_s := \min\{\lambda | \lambda \text{ is an eigenvalue of } \mathbf{C}_s\}$.

The goal of this section is to derive decay estimates for (2.1) with rates in between this weakest rate μ_s and the optimal rate μ from (2.4). At the same time we shall also present lower bounds on $\|f(t)\|_2$. The energy method again provides the simplest example of it, in the form

$$\|f(t)\|_2 \ge e^{-\nu_s t} \|f^I\|_2 , \quad t \ge 0 ,$$

with $\nu_s := \max\{\lambda | \lambda \text{ is an eigenvalue of } \mathbf{C}_s\}$. Clearly, estimates with decay rates outside of $[\mu_s, \nu_s]$ are irrelevant.

We present our main result only for the two-dimensional case, as the best multiplicative constant is not yet known explicitly in higher dimensions (cf. §3):

Theorem 4.1. *Let $\mathbf{C} \in \mathbb{C}^{2\times 2}$ be a diagonalizable positive stable matrix with spectral gap $\mu := \min\{\Re(\lambda_j^{\mathbf{C}}) | \, j = 1, 2\}$. Then, all solutions to (2.1) satisfy the following upper and lower bounds:*

a)

$$\|f(t)\|_2 \le c_1(\tilde\mu)\, e^{-\tilde\mu t} \|f^I\|_2 , \quad t \ge 0 , \quad \mu_s \le \tilde\mu \le \mu , \tag{4.1}$$

with

$$c_1^2(\tilde\mu) = \kappa_{\min}(\beta(\tilde\mu))$$

given explicitly in (4.8) below. There, $\alpha \in [0, 1)$ is the cos *of the (minimal) angle of the eigenvectors of $\mathbf{C}^*$ (cf. the proof of Lemma 3.2), and $\beta(\tilde\mu) = \max(-\alpha, -\beta_0)$, with β_0 defined in (4.6), (4.7) below.*

b)

$$\|f(t)\|_2 \ge c_2(\tilde\mu)\, e^{-\tilde\mu t} \|f^I\|_2 , \quad t \ge 0 , \quad \nu \le \tilde\mu \le \nu_s , \tag{4.2}$$

with $\nu := \max\{\Re(\lambda_j^{\mathbf{C}}) | \, j = 1, 2\}$. The maximal constant

$$c_2^2(\tilde\mu) = \kappa_{\min}(\beta(\tilde\mu))^{-1}$$

is given again by (4.8), with α, $\beta(\tilde\mu)$ defined as in Part (a).

Proof. Part (a): For a fixed $\tilde\mu \in [\mu_s, \mu]$ we have to determine the smallest constant c_1 for the estimate (4.1), following the strategy of proof from §3. To this end, we use a unitary transformation of the coordinate system and write $\mathbf{P}(\tilde\mu) = \mathbf{W}\mathbf{B}_u\mathbf{W}^*$ with

$$\mathbf{W} = \begin{pmatrix} 1 & \alpha \\ 0 & \sqrt{1-\alpha^2} \end{pmatrix} , \quad \mathbf{B}_u = \begin{pmatrix} 1/b & \beta(\tilde\mu) \\ \bar\beta(\tilde\mu) & b \end{pmatrix} , \tag{4.3}$$

where we set w.l.o.g. $b_1 = 1/b$, $b_2 = b$ with $b > 0$. Moreover, $|\beta|^2 < 1$ has to hold. Now, we have to find the positive definite Hermitian matrix $\mathbf{B}_u$, such that the analog of (3.3), (3.4) holds, i.e.:

$$\mathbf{A} := \begin{pmatrix} 2(\Re(\lambda_1^{\mathbf{C}}) - \tilde{\mu})/b & (\bar{\lambda}_1^{\mathbf{C}} + \lambda_2^{\mathbf{C}} - 2\tilde{\mu})\beta \\ (\lambda_1^{\mathbf{C}} + \bar{\lambda}_2^{\mathbf{C}} - 2\tilde{\mu})\bar{\beta} & 2(\Re(\lambda_2^{\mathbf{C}}) - \tilde{\mu})b \end{pmatrix} \geq 0\,, \tag{4.4}$$

As in the proof of Lemma 3.1, we assume that the eigenvalues of $\mathbf{C}$ are ordered as $\Re(\lambda_2^{\mathbf{C}}) \geq \Re(\lambda_1^{\mathbf{C}}) = \mu \geq \tilde{\mu}$. Hence, $\operatorname{Tr}\mathbf{A} \geq 0$. For the non-negativity of the determinant to hold, i.e.

$$\det \mathbf{A} = 4\big(\Re(\lambda_1^{\mathbf{C}}) - \tilde{\mu}\big)\big(\Re(\lambda_2^{\mathbf{C}}) - \tilde{\mu}\big) - |\lambda_1^{\mathbf{C}} + \bar{\lambda}_2^{\mathbf{C}} - 2\tilde{\mu}|^2|\beta|^2 \geq 0\,, \tag{4.5}$$

we have the following restriction on β:

$$|\beta|^2 \leq \beta_0^2 := \frac{4\big(\Re(\lambda_1^{\mathbf{C}}) - \tilde{\mu}\big)\big(\Re(\lambda_2^{\mathbf{C}}) - \tilde{\mu}\big)}{|\lambda_1^{\mathbf{C}} + \bar{\lambda}_2^{\mathbf{C}} - 2\tilde{\mu}|^2}\,. \tag{4.6}$$

If $\lambda_1^{\mathbf{C}} + \bar{\lambda}_2^{\mathbf{C}} - 2\tilde{\mu} = 0$, we conclude $\lambda_1^{\mathbf{C}} = \lambda_2^{\mathbf{C}}$ and that we have chosen the sharp decay rate $\tilde{\mu} = \mu$. As the associated, minimal condition number $\kappa(\mathbf{P})$ was already determined in Lemma 3.2, we shall not rediscuss this case here. But to include this case into the statement of the theorem, we set

$$\beta_0 := 1\ , \quad \text{if } \lambda_1^{\mathbf{C}} = \lambda_2^{\mathbf{C}} \text{ and } \tilde{\mu} = \mu\,. \tag{4.7}$$

From (4.6) we conclude that $\beta_0 \in [0, 1]$. Note that $\beta_0 = 1$ is only possible for $\tilde{\mu} = \mu$ and $\lambda_1^{\mathbf{C}} = \lambda_2^{\mathbf{C}}$, i.e. the case that we just sorted out. For the rest of the proof we hence assume that condition (4.6) holds with $\beta_0 \in [0, 1)$.

For admissible matrices $\mathbf{B}_u$ (i.e. with $b > 0$ and $|\beta| \leq \beta_0$) it remains to determine the matrix

$$\mathbf{P}(b, \beta) = \mathbf{W}\mathbf{B}_u\mathbf{W}^* = \begin{pmatrix} \frac{1}{b} + 2\alpha\Re\beta + b\alpha^2 & (\beta + b\alpha)\sqrt{1-\alpha^2} \\ (\bar{\beta} + b\alpha)\sqrt{1-\alpha^2} & b(1-\alpha^2) \end{pmatrix}\,,$$

(with $\mathbf{W}$ and $\mathbf{B}_u$ given in (4.3)), having the minimal condition number $\kappa\big(\mathbf{P}(b,\beta)\big) = \lambda_+^{\mathbf{P}}(b,\beta)/\lambda_-^{\mathbf{P}}(b,\beta)$. Here

$$\lambda_\pm^{\mathbf{P}}(b,\beta) = \frac{\operatorname{Tr}\mathbf{P}(b,\beta) \pm \sqrt{(\operatorname{Tr}\mathbf{P}(b,\beta))^2 - 4\det\mathbf{P}(b,\beta)}}{2}$$

are the (positive) eigenvalues of $\mathbf{P}(b, \beta)$.

As a first step we shall minimize $\kappa\big(\mathbf{P}(b,\beta)\big)$ w.r.t. b (and for β fixed), since $\operatorname{argmin}_{b>0} \kappa\big(\mathbf{P}(b,\beta)\big)$ will turn out to be independent of β. We notice that $\operatorname{Tr}\mathbf{P}(b,\beta) = b + 2\alpha\Re\beta + 1/b$ is a convex function of $b \in (0,\infty)$ which attains its minimum for $b = 1$. Moreover, $\det\mathbf{P}(b,\beta) = (1-\alpha^2)(1-|\beta|^2) > 0$ is independent of b. This yields the condition number

$$\kappa_{\min}(\beta) = \frac{\lambda_+^{\mathbf{P}}(1,\beta)}{\lambda_-^{\mathbf{P}}(1,\beta)} = \frac{1 + \sqrt{1 - \frac{(1-\alpha^2)(1-|\beta|^2)}{(1+\alpha\Re\beta)^2}}}{1 - \sqrt{1 - \frac{(1-\alpha^2)(1-|\beta|^2)}{(1+\alpha\Re\beta)^2}}}\,.$$

As a second step we minimize $\kappa_{\min}(\beta)$ on the disk $|\beta| \leq \beta_0$. To this end, the quotient $\frac{(1-\alpha^2)(1-|\beta|^2)}{(1+\alpha\Re\beta)^2}$ should be as large as possible. For any fixed $|\beta| \leq \beta_0$, this happens by choosing $\beta = -|\beta|$, since $\alpha \in [0, 1)$. Hence it remains to maximize the function $g(\beta) := \frac{1-\beta^2}{(1+\alpha\beta)^2}$ on the interval $[-\beta_0, 0]$.

It is elementary to verify that g is maximal at $\tilde{\beta} := \max(-\alpha, -\beta_0)$. Then, the minimal condition number is

$$\kappa_{\min}(\tilde{\beta}) = \kappa\big(\mathbf{P}(1,\tilde{\beta})\big) = \frac{1+\sqrt{1-\frac{(1-\alpha^2)(1-\tilde{\beta}^2)}{(1+\alpha\tilde{\beta})^2}}}{1-\sqrt{1-\frac{(1-\alpha^2)(1-\tilde{\beta}^2)}{(1+\alpha\tilde{\beta})^2}}} \,. \tag{4.8}$$

Part (b): Since the proof of the lower bound is very similar to Part (a), we shall just sketch it. For a fixed $\tilde{\mu} \in [\nu, \nu_s]$ we have to determine the largest constant c_2 for the estimate (4.2). To this end we need to satisfy the inequality

$$\mathbf{C}^*\mathbf{P} + \mathbf{P}\mathbf{C} \leq 2\tilde{\mu}\mathbf{P}$$

with a positive definite Hermitian matrix $\mathbf{P}$ with minimal condition number $\kappa(\mathbf{P})$. In analogy to §2 this would imply

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f(t)\|^2_{\mathbf{P}} \geq -2\tilde{\mu}\|f(t)\|^2_{\mathbf{P}} \,,$$

and hence the desired lower bound

$$\|f(t)\|_2^2 \geq (\lambda^{\mathbf{P}}_{\max})^{-1}\|f(t)\|^2_{\mathbf{P}} \geq (\lambda^{\mathbf{P}}_{\max})^{-1}e^{-2\tilde{\mu}t}\|f^I\|^2_{\mathbf{P}} \geq (\kappa(\mathbf{P}))^{-1}\, e^{-2\tilde{\mu}t}\|f^I\|_2^2 \,.$$

For minimizing $\kappa(\mathbf{P})$, we again use a unitary transformation of the coordinate system and write $\mathbf{P}$ as $\mathbf{P}(\tilde{\mu}) = \mathbf{W}\mathbf{B}_l\mathbf{W}^*$, with $\mathbf{W}$ from (4.3) and the positive definite Hermitian matrix

$$\mathbf{B}_l = \begin{pmatrix} 1/b & \beta(\tilde{\mu}) \\ \bar{\beta}(\tilde{\mu}) & b \end{pmatrix} \,,$$

with $b > 0$ and $|\beta|^2 < 1$. Then, the matrix $\mathbf{A}$ from (4.4) has to satisfy $\mathbf{A} \leq 0$. Since we chose the eigenvalues of $\mathbf{C}$ to be ordered as $\Re(\lambda_1^{\mathbf{C}}) \leq \Re(\lambda_2^{\mathbf{C}}) = \nu \leq \tilde{\mu}$, we have $\operatorname{Tr}\mathbf{A} \leq 0$. The necessary non-negativity of its determinant again reads as (4.5).

In the special case $\lambda_1^{\mathbf{C}} + \bar{\lambda}_2^{\mathbf{C}} - 2\tilde{\mu} = 0$, we conclude again $\lambda_1^{\mathbf{C}} = \lambda_2^{\mathbf{C}}$ and $\tilde{\mu} = \nu$. Hence $\mathbf{A} = 0$. Since β is then only restricted by $|\beta| < 1$, we can again set $\beta_0 = 1$ and obtain the minimal $\kappa(\mathbf{P})$ for $\tilde{\beta}(\nu) = -\alpha$, as in Part (a).

In the generic case, the minimal $\kappa(\mathbf{P})$ is obtained for $\tilde{\beta} = \max(-\alpha, -\beta_0)$ with β_0 given in (4.6). Hence, the maximal constant in the lower bound (4.2) is $c_2^2(\tilde{\mu}) = \kappa_{\min}(\tilde{\beta})^{-1}$ where $\kappa_{\min}$ is given by (4.8). This finishes the proof. □

We illustrate the results of Theorem 4.1 with two examples.

Example 4.1. We consider ODE (2.1) with the matrix

$$\mathbf{C} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

which has eigenvalues $\lambda_\pm = (1 \pm i\sqrt{3})/2$, and some normalized eigenvectors of $\mathbf{C}^*$ are, e.g.

$$v_+ = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ \lambda_- \end{pmatrix}, \qquad v_- = \frac{1}{\sqrt{2}}\begin{pmatrix} -\lambda_- \\ 1 \end{pmatrix}. \tag{4.9}$$

The optimal decay rate is $\mu = 1/2$, whereas the minimal and maximal eigenvalues of $\mathbf{C}_s$ are $\mu_s = 0$ and $\nu_s = 1$, respectively. To bring the eigenvectors of $\mathbf{C}^*$ in the canonical form used in the proof of Theorem 4.1, we fix the eigenvector v_+, and choose the unitary multiplicative factor for the second eigenvector v_- as in (4.9) such that $\langle v_+, v_- \rangle = 1/2$ is a real number. Finally, we use the Gram-Schmidt process to obtain a new orthonormal basis such that the eigenvectors of $\mathbf{C}^*$ in the new orthonormal basis are of the form (3.6) with $\alpha = 1/2$. Then, the upper and lower bounds for the Euclidean norm of a solution of (2.1) are plotted in Fig. 1 and Fig. 2. For both

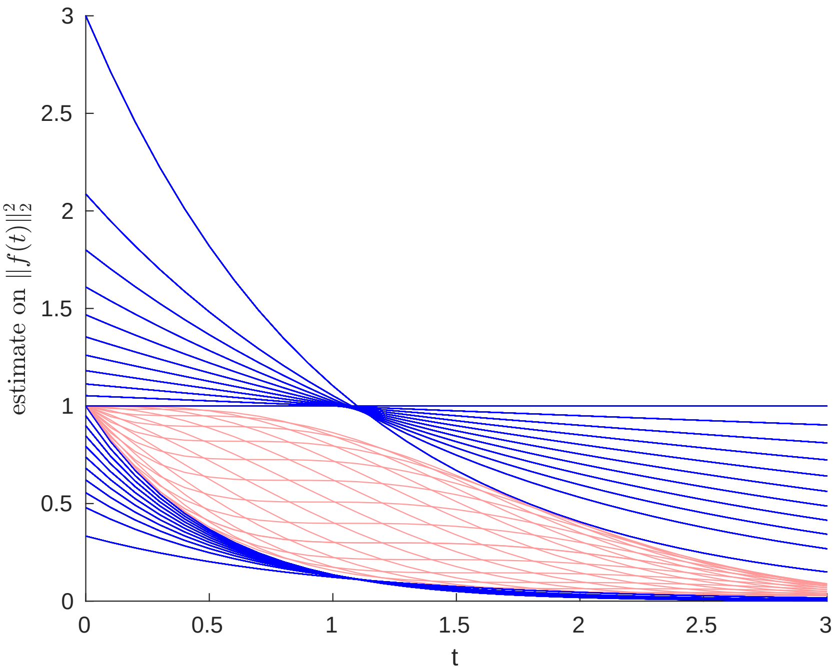

FIGURE 1. The red (grey) curves are the squared norm of solutions $f(t)$ for ODE (2.1) with matrix $\mathbf{C} = [1, -1; 1, 0]$ and various initial data f^I with norm 1. The blue (black) curves are the lower and upper bounds for the squared norm of solutions. Note: The curves are colored only in the electronic version of this article.

the upper and lower bounds, the respective family of decay curves does *not* intersect in a single point (see Fig. 2). Hence, the whole family of estimates provides a (slightly) better estimate on $\|f(t)\|_2$ than if just considering the two extremal decay rates, i.e. μ and μ_s for the upper bound and ν and ν_s for the lower bound.

Note that upper bound with the sharp decay rate $\mu = \frac{1}{2}$, i.e. $3e^{-t}$ carries the (unconditionally) optimal multiplicative constant $c = 3$, as it touches the set of solutions (see Fig. 1). But this is not true for the estimates with smaller decay rates (except of $\tilde{\mu} = 0$). Neither does it hold for the estimates in Fig. 3.

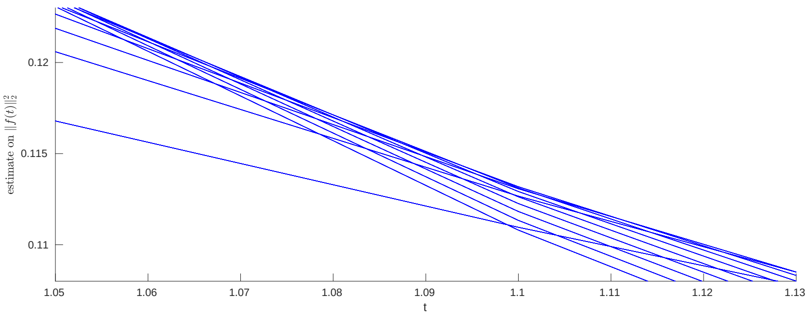

FIGURE 2. Zoom of Fig. 1: The curves are the lower bounds for the squared norm of solutions for ODE (2.1) with matrix $\mathbf{C} = [1, -1; 1, 0]$ and various initial data f^I with norm 1. This plot shows that these lower bounds do not intersect in a single point.

Example 4.2. We consider ODE (2.1) with the matrix

$$\mathbf{C} = \begin{pmatrix} 19/20 & -3/10 \\ 3/10 & -1/20 \end{pmatrix}$$

which has the eigenvalues $\lambda_1 = 1/20$ and $\lambda_2 = 17/20$, and some normalized eigenvectors of $\mathbf{C}^*$ are, e.g.

$$v_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ -3 \end{pmatrix}, \qquad v_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

The optimal decay rate is $\mu = 1/20$, whereas the minimal and maximal eigenvalues of $\mathbf{C}_s$ are $\mu_s = -1/20$ and $\nu_s = 19/20$, respectively. Since the matrix $\mathbf{C}$ and its eigenvalues are real valued, the eigenvectors of $\mathbf{C}^*$ are already in the canonical form used in the Gram-Schmidt process to obtain a new orthogonal basis such that the eigenvectors of $\mathbf{C}^*$ in the new basis are of the form (3.6) with $\alpha = 3/5$. Then, the upper and lower bounds for the Euclidean norm of a solution of (2.1) are plotted in Fig. 3. Since $\mu_s < 0$, solutions $f(t)$ to this example may initially increase in norm.

REFERENCES

[1] Achleitner, F., Arnold, A., Carlen, E.A.: On linear hypocoercive BGK models. In: From particle systems to partial differential equations. III, *Springer Proc. Math. Stat.*, vol. 162, pp. 1–37. Springer (2016).

[2] Achleitner, F., Arnold, A., Carlen, E.A.: On multi-dimensional hypocoercive BGK models. arXiv preprint, arXiv:1711.07360 (2017).

[3] Achleitner, F., Arnold, A., Stürzer, D.: Large-Time Behavior in Non-Symmetric Fokker-Planck Equations. Rivista di Matematica della Università di Parma **6**, 1–68 (2015).

[4] Arnold, V.I.: Ordinary differential equations. MIT Press, Cambridge, Mass.-London (1978).

[5] Arnold, A., Erb, J.: Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift, arXiv preprint, arXiv:1409.5425 (2014).

[6] Arnold, A., Jin, S., Wöhrer, T.: Sharp Decay Estimates in Defective Evolution Equations: from ODEs to Kinetic BGK Equations, preprint, (2018).

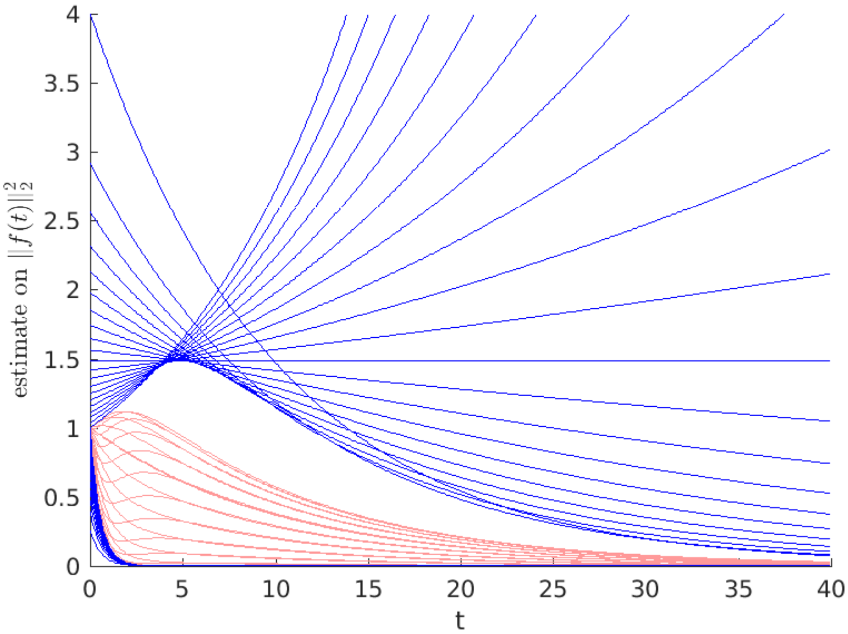

FIGURE 3. The red (grey) curves are the squared norm of solutions $f(t)$ for ODE (2.1) with matrix $\mathbf{C} = [19/20, -3/10; 3/10, -1/20]$ and various initial data f^I with norm 1. The blue (black) curves are the lower and upper bounds for the squared norm of solutions. Note: The curves are colored only in the electronic version of this article.

[7] Bhatnagar, P. L., Gross, E. P., Krook, M.: A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems. Phys. Rev. **94**, no. 3, 511–525 (1954).

[8] Boyd, S. P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004).

[9] Braatz, R. D., Morari, M.: Minimizing the Euclidean condition number. SIAM J. Control Optim. **32**, no. 6, 1763–1768 (1994).

[10] Businger, P. A.: Matrices which can be optimally scaled. Numer. Math. **12**, 346–348 (1968).

[11] Dolbeault, J., Mouhot, C., Schmeiser, C.: Hypocoercivity for linear kinetic equations conserving mass. Transactions of the American Mathematical Society **367**(6), 3807–3828 (2015).

[12] Kolotilina, L.Yu.: Solution of the problem of optimal diagonal scaling for quasireal Hermitian positive definite 3×3 matrices. Journal of Mathematical Sciences **132**, no. 2, 190–213 (2006).

[13] Sezginer, R. S., Overton, M. L.: The largest singular value of $e^X A_0 e^{-X}$ is convex on convex sets of commuting matrices, IEEE Transactions on Automatic Control **35**, no. 2, 229–230 (1990).

[14] Villani, C.: Hypocoercivity. Mem. Amer. Math. Soc., **202** (2009).

Faculty of Mathematics, University of Vienna,, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
E-mail address: franz.achleitner@univie.ac.at

TU Wien, Institute for Analysis and Scientific Computing, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria
E-mail address: anton.arnold@tuwien.ac.at

TU Wien, Institute for Analysis and Scientific Computing, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria
E-mail address: beatrice.signorello@tuwien.ac.at