Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

doi:10.3934/xx.xx.xx.xx

pp. **X–XX**

FOURIER-SPLITTING METHOD FOR SOLVING HYPERBOLIC LQR PROBLEMS

Petra Csomós*

Institute of Mathematics, Eötvös Loránd University Budapest
MTA-ELTE Numerical Analysis and Large Networks Research Group
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

Hermann Mena

School of Mathematical Sciences and Information Technology, Yachay Tech
Hacienda San José y Proyecto Yachay, EC100650 Urcuquí, Ecuador
Department of Mathematics, University of Innsbruck
Technikerstrasse 13, A-6020 Innsbruck, Austria

(Communicated by the associate editor name)

Abstract. We consider the numerical approximation to linear quadratic regulator problems for hyperbolic partial differential equations where the dynamics is driven by a strongly continuous semigroup. The optimal control is given in feedback form in terms of Riccati operator equations. The computational cost relies on solving the associated Riccati equation and computing the optimal state. In this paper we propose a novel approach based on operator splitting idea combined with Fourier's method to efficiently compute the optimal state. The Fourier's method allows to accurately approximate the exact flow making our approach computational efficient. Numerical experiments in one and two dimensions show the performance of the proposed method.

1. **Introduction.** Optimal control problems are used to determine the suitable way how to drive a system into its predefined state, meanwhile certain optimality criterion is fulfilled. Latter is usually achived by minimizing a cost functional being a function of the state and the control function. The optimal state is then obtained from a differential equation incorporating the operator which corresponds to the dynamics, and the optimal control function resulting from the minimization. A special case is the linear quadratic regulator (LQR) problem with linear operators for the dynamics and the control, and a quadratic continuous-time cost functional.

The infinite dimensional LQR problem has been extensively studied in the literature, a good survey can be found in, e.g. [5, 9, 31]. The purpose of the theoretical framework is to address optimal control of systems of partial differential equations (PDEs). For most systems, the controlling mechanism can only be applied from the

2010 *Mathematics Subject Classification.* Primary: 35Q93, 49J20; Secondary: 65M22.

Key words and phrases. Operator splitting method, Fourier method, Optimal feedback control, Hyperbolic LQR problem, Numerical solution.

The first author is supported by the National Research, Development and Innovation Fund (Hungary) under the grant PD121117.

* Corresponding author: P. Csomós.

interface of the system or at finitely many points or curves (see e.g. [6]) which needs the development of a framework for studying boundary/point control. Such control actions can mathematically be captured by using maps which are not bounded with respect to the state space, but take values in a larger dual space. The most natural class of problems, where such description has been used, are dynamics driven by analytic semigroups. For some classes of control systems which combine hyperbolic and parabolic dynamics, it has been observed that the control-to-state kernel satisfies a singular estimate which generalizes the case of analytic semigroup dynamics [30]. In both cases the feedback characterization of the optimal control is given in terms of Riccati equations.

The present paper focuses on the numerical solution of LQR problems associated to hyperbolic PDEs leading to the approximation of strongly continuous semigroups. Thus, one has to solve the large-scale Riccati equation arising from the space discretization of the problem, and then to compute the optimal state, i.e., to insert the control law in the original equation and to solve the resulting PDE. The Riccati equation can be algebraic or differential for infinite horizon and finite horizon problems, respectively. In recent years there has been a great algorithmic progress in solving large-scale algebraic Riccati equations see e.g. [13, 10] and the reference therein. The same occurs for differential Riccati equations see e.g [3, 11, 12, 29]. In this work we consider LQR problems in an infinite horizon, i.e., the control is given in terms of algebraic Riccati equations. The main contribution of this work is an efficient computation of the optimal state. Unfortunately, only relatively simple numerical schemes are usually applied for computing the optimal state in applications. Up to our knowledge there have been no attempts to compute the optimal state efficiently. Our approach is based on the operator splitting method (studied e.g. in [25], [20], [7], [8]) which is an efficient way to treat problems describing the combined effect of more phenomena by splitting the problem into more subproblems corresponding to the separate phenomena. One solves the sub-problems in a cycle by using the solution of the previous sub-problem as the initial condition to the next one. This approach makes it possible to approximate the optimal state, being the combined effect of the dynamics and the control, by the composition of the separately computed flows corresponding to the dynamics as well as to the control. More precisely, the optimal state is approximated via the product of the strongly continuous semigroups corresponding to these two separate flows. In many applications, the flow (or semigroup) corresponding to the dynamics can be almost accurately obtained by using Fourier transform. Moreover, in this approach the semigroup corresponding to the control should be computed and stored only once at the very beginning of the computation. The accurate modelling of the dynamics and the fast realization of the control make this method more efficient compared to those which approximate the optimal state in the usual way. We remark that the application of the operator splitting method introduces a new source of error being also investigated in the paper.

Since all our examples describe some kind of wave phenomena physically, we can consider their solution as a wave perturbation propagating on the surface of a water reservoir (lake, river, etc.). Since our goal is to avoid flood on the shore, the amplitude of these water waves should decrease, that is, the system should be driven to the zero state (i.e., without any water waves). Then the solution of the LQR problem gives the time-evolution of the water waves. Its convergence to zero

is determined by the optimal control function which depends on the actual control in the physical space. We will use two different types of control in our numerical experiments: distributed control acting at all points of the domain, and another one imitating a sink.

The paper is organized as follows. In Section 2 we briefly describe the abstract setting of the LQR problem in infinite dimensions, and introduce the hyperbolic systems we studied. In Section 3 our novel approach for computing the optimal state in the LQR framework is presented. Section 4 numerical experiments validating the efficiency of our approach are presented. Finally, conclusions and future work close the paper.

2. **Abstract LQR Problem.** In the following, we briefly describe the LQR problem in an abstract setting, for a detailed explanation we refer for example to [9, 31]. We consider systems governed by PDEs of hyperbolic type which can be written as an abstract Cauchy problem of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}x(t) = Ax(t) + Bu(t), \quad t > 0, \\ x(0) = x_0 \in \mathcal{H}, \end{cases} \tag{1}$$

where $x \in \mathrm{W}^{1,2}(0,\infty;\mathcal{H})$ is the state function, $u \in \mathrm{L}^2(0,\infty;\mathcal{U})$ is the control function, and $\mathcal{H}$, $\mathcal{U}$ are separable Hilbert spaces. In some applications, an output equation $y(t) = Dx(t)$ is also considered; $y\colon [0,+\infty) \to \mathcal{Y}$ and $D\colon \mathcal{H} \to \mathcal{Y}$, for a Hilbert space $\mathcal{Y}$. Without lost of generality the output equations is usually omitted. In addition, we consider the following quadratic cost functional:

$$J(x_0,u) = \frac{1}{2}\int_0^\infty \big(\langle x(t), Qx(t)\rangle_{\mathcal{H}} + \langle u(t), Ru(t)\rangle_{\mathcal{U}}\big)\mathrm{d}t, \tag{2}$$

where $\langle\cdot,\cdot\rangle$ represents an inner product in the corresponding Hilbert space, and the operators $Q\colon \mathcal{H} \to \mathcal{H}$ and $R\colon \mathcal{U} \to \mathcal{U}$ are weight operators particular of every application. The LQR problem is to minimize $J(x_0,u)$ with respect to u subject to the state equation (1).

Throughout the paper we suppose the following standard assumptions.

Assumptions 2.1.
1. Operator $A\colon \mathrm{dom}(A) \subset \mathcal{H} \to \mathcal{H}$ is the infinitesimal generator of the strongly continuous semigroup $(\mathrm{e}^{tA})_{t\geq 0}$ on $\mathcal{H}$.
2. Operator $B\colon \mathcal{U} \to \mathcal{H}$ is linear and bounded.
3. For every initial value $x_0 \in \mathcal{H}$, there exists a control function $u \in \mathrm{L}^2(0,\infty;\mathcal{U})$ for which $J(x_0,u) < \infty$ holds (then u is called an admissible control function).
4. Operator $Q : \mathcal{H} \to \mathcal{H}$ is positive semi-definite, and operator $R\colon \mathcal{U} \to \mathcal{U}$ is positive definite.

Note that the restrictive boundedness assumption on B can be weakened, see e.g. in [31]. These systems have the so-called singular estimate property and naturally arise from certain boundary control problems.

It was shown e.g. in [9, 31] that under Assumptions 2.1, the solution to the abstract LQR problem has a feedback form,

$$u(t) = -R^{-1}B^*Xx(t), \quad t \geq 0, \tag{3}$$

where $X\colon \mathcal{H} \to \mathcal{H}$ represents the solution to the algebraic operator Riccati equation

$$Q + A^*X + XA - XBR^{-1}B^*X = 0. \tag{4}$$

Under Assumptions 2.1, the solution of (4) is unique see e.g. in [9, Thm. V.3.1]. By inserting the optimal control u, derived as (3), into the state equation (1), the *optimal state* x is obtained from the system

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}x(t) = (A - BR^{-1}B^*X)x(t), \quad t \geq 0, \\ x(0) = x_0 \in \mathcal{H}. \end{cases} \tag{5}$$

In this paper we propose a novel approach to compute an approximate solution of (5) based on operator splitting and Fourier's method.

2.1. **Hyperbolic Systems.** In this section we introduce some hyperbolic problems which serve as examples for our numerical results. We present them in their dimensionless form, because this makes it possible to treat the systems as being independent of the actual parameters. Since we consider the problems on infinite intervals, we pose periodic boundary condition for the numerical experiments imitating the effect of the infinite interval. We note that Fourier's method is efficiently applicable in this case.

1D Advection. We consider the problem (1) on the Hilbert spaces $\mathcal{H} = \mathrm{L}^2(\mathbb{R})$ and $\mathcal{U} = \mathrm{L}^2(\mathbb{R})$ for the operators

$$\begin{aligned} Ax := -cx' \quad &\text{with} \quad \mathrm{dom}(A) := \mathrm{W}^{1,2}(\mathbb{R}), \quad \text{and} \\ B := \mathrm{I}_{\mathcal{U}} \quad &\text{with} \quad \mathrm{dom}(B) := \mathcal{U} \end{aligned} \tag{6}$$

with the parameter $c \in \mathbb{R}$. For $B = 0$, it corresponds to the one-dimensional advection equation for the unknown function $w\colon [0, +\infty) \times \mathbb{R} \to \mathbb{R}$:

$$\begin{cases} \partial_t w = -c\partial_\xi w, \quad t \geq 0,\ \xi \in \mathbb{R}, \\ w(0, \xi) = w_0(\xi), \quad \xi \in \mathbb{R}. \end{cases} \tag{7}$$

Its exact solution is given by $w(t, \xi) = w_0(\xi - ct)$ for all $t \geq 0$ and $\xi \in \mathbb{R}$, that is, the initial function is shifted to the right or to the left for $c > 0$ and $c < 0$, respectively. Hence, the solution corresponds to the right or left shift strongly continuous semigroup on $\mathrm{L}^2(\mathbb{R})$ (see [17, Chapter II.2.10]).

2D Advection. We consider the problem (1) on the Hilbert spaces $\mathcal{H} = \mathrm{L}^2(\mathbb{R}^2)$ and $\mathcal{U} = \mathrm{L}^2(\mathbb{R}^2)$ for the operators

$$\begin{aligned} Ax := -\langle \mathbf{c}, x' \rangle \quad &\text{with} \quad \mathrm{dom}(A) := \mathrm{W}^{1,2}(\mathbb{R}^2), \quad \text{and} \\ B := \mathrm{I}_{\mathcal{U}} \quad &\text{with} \quad \mathrm{dom}(B) := \mathcal{U} \end{aligned}$$

with the parameter $\mathbf{c} = (c_1, c_2) \in \mathbb{R}^2$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^2$. For $B = 0$, it corresponds to the two-dimensional advection equation for the unknown function $w\colon [0, +\infty) \times \mathbb{R}^2 \to \mathbb{R}$:

$$\begin{cases} \partial_t w = -c_1\partial_\xi w - c_2\partial_\eta w, \quad t \geq 0,\ (\xi, \eta) \in \mathbb{R}^2, \\ w(0, \xi, \eta) = w_0(\xi, \eta), \quad (\xi, \eta) \in \mathbb{R}. \end{cases} \tag{8}$$

Its exact solution is given by $w(t, \xi, \eta) = w_0(\xi - c_1 t, \eta - c_2 t)$ for all $t \geq 0$ and $(\xi, \eta) \in \mathbb{R}^2$, that is, the initial function is shifted to the direction specified by the vector $\mathbf{c}$.

1D Linearized Shallow Water Equations. The one-dimensional linearized shallow water equations describe the motion of a fluid which vertical scale is much less then its horizontal scale. Let $h(t,\xi)$ and $U(t,\xi)$ denote the deviations of the fluid's height and vertically averaged velocity, respectively, from the basic state described by the height $\overline{h} > 0$ and the vertically averaged velocity $\overline{U} \in \mathbb{R}$ at each point $\xi \in \mathbb{R}$ and for all time $t \geq 0$. Consider then the problem (see e.g. in [36, Section 3.6])

$$\begin{cases} \partial_t h = -\overline{U}\partial_\xi h - \overline{h}\partial_\xi U, \quad t > 0,\ \xi \in \mathbb{R}, \\ \partial_t U = -\overline{U}\partial_\xi U - c_{\mathrm{F}}\partial_\xi h \end{cases} \tag{9}$$

with parameter $c_F := \frac{1}{\mathrm{Fr}^2} > 0$ where $\mathrm{Fr} > 0$ is the Froude number describing the ratio between the velocity of the zonal and the gravity waves. By defining the unknown function x and the control function u, respectively, as

$$\big(x(t)\big)(\xi) := \begin{pmatrix} h(t,\xi) \\ U(t,\xi) \end{pmatrix} \quad \text{and} \quad \big(u(t)\big)(\xi) := \begin{pmatrix} u_h(t,\xi) \\ u_U(t,\xi) \end{pmatrix} \quad \text{for} \quad t \geq 0,\ \xi \in \mathbb{R},$$

we end up with problem (1) on the space $\mathcal{H} = (\mathrm{L}^2(\mathbb{R}))^2$ with the operator

$$A = -\begin{pmatrix} \overline{U}\partial_\xi & \overline{h}\partial_\xi \\ c_{\mathrm{F}}\partial_\xi & \overline{U}\partial_\xi \end{pmatrix} \tag{10}$$

on $\mathrm{dom}(A) = \{(h,U) \in (\mathrm{L}^2(\mathbb{R}))^2 \colon h \in \mathrm{H}^1(\mathbb{R}), U \in \mathrm{W}^{1,2}(\mathbb{R})\}$ and some suitable operator B on $\mathcal{U} = (\mathrm{L}^2(\mathbb{R}))^2$. We note that each entry of A corresponds to the advection operator from (6).

2D Linearized Shallow Water Equations. The two-dimensional linearized shallow water equations describe the motion of a fluid over $\mathbb{R}^2$. As before, $h(t,\xi,\eta)$ denotes the deviation of the fluid's height from $\overline{h} > 0$, while $U(t,\xi,\eta)$ and $V(t,\xi,\eta)$ denote the deviations of the vertically averaged velocities from $\overline{U} \in \mathbb{R}$ and $\overline{V} \in \mathbb{R}$ in ξ and η directions, respectively, for all $(\xi,\eta) \in \mathbb{R}^2$ and $t \geq 0$. We consider then the following problem (see e.g. in [36, Section 3.6]):

$$\begin{cases} \partial_t h = -\overline{U}\partial_\xi h - \overline{V}\partial_\eta h - \overline{h}\,(\partial_\xi U + \partial_\eta V), \quad t > 0,\ (\xi,\eta) \in \mathbb{R}^2, \\ \partial_t U = -\overline{U}\partial_\xi U - \overline{V}\partial_\eta U - c_{\mathrm{F}}\partial_\xi h + c_{\mathrm{R}} V, \\ \partial_t V = -\overline{U}\partial_\xi V - \overline{V}\partial_\eta V - c_{\mathrm{F}}\partial_\eta h - c_{\mathrm{R}} U, \end{cases} \tag{11}$$

with parameter $c_{\mathrm{G}} = \frac{1}{\mathrm{Ro}} > 0$ where $\mathrm{Ro} > 0$ stands for the Rossby number characterizing the ratio between the time scales of the Coriolis and the inertial forces. Similarly to the one-dimensional case, we set up problem (1) on the space $\mathcal{H} = (\mathrm{L}^2(\mathbb{R}))^3$ by defining the functions

$$\big(x(t)\big)(\xi,\eta) := \begin{pmatrix} h(t,\xi,\eta) \\ U(t,\xi,\eta) \\ V(t\xi,\eta) \end{pmatrix} \quad \text{and} \quad \big(u(t)\big)(\xi,\eta) := \begin{pmatrix} u_h(t,\xi,\eta) \\ u_U(t,\xi,\eta) \\ u_V(t\xi,\eta) \end{pmatrix}$$

for $t \geq 0$, $(\xi,\eta) \in \mathbb{R}$, and the operator

$$A = -\begin{pmatrix} \overline{U}\partial_\xi + \overline{V}\partial_\eta & \overline{h}\partial_\xi & \overline{h}\partial_\eta \\ c_{\mathrm{F}}\partial_\xi & \overline{U}\partial_\xi + \overline{V}\partial_\eta & -c_{\mathrm{R}}\mathrm{I}_{\mathrm{L}^2} \\ c_{\mathrm{F}}\partial_\eta & c_{\mathrm{R}}\mathrm{I}_{\mathrm{L}^2} & \overline{U}\partial_\xi + \overline{V}\partial_\eta \end{pmatrix} \tag{12}$$

on $\mathrm{dom}(A) = \{(h,U,V) \in (\mathrm{L}^2(\mathbb{R}))^3 \colon h \in \mathrm{H}^1(\mathbb{R}), \mathrm{div}\binom{U}{V} \in \mathrm{L}^2(\mathbb{R})\}$ and some suitable operator B on $\mathcal{U} = (\mathrm{L}^2(\mathbb{R}))^3$, where $\mathrm{I}_{\mathrm{L}^2}$ denotes the identity operator on $\mathrm{L}^2(\mathbb{R})$. We note that the entries of the operator matrix A contain the advection operator (6)

in directions ξ as well as η. In [14] it was shown that under some assumptions on operator A, it is the generator of a strongly continuous semigroup.

3. **Fourier-Splitting Method.** The aim of this paper is to present an efficient numerical method for approximating the optimal state, that is, to compute the numerical solution to problem (5).

Since $C := -BR^{-1}B^*X \colon \mathcal{H} \to \mathcal{H}$ is a bounded operator, it is the bounded perturbation of the generator A, therefore, by [17, Thm. III.1.3], the operator $A+C$, called the closed loop operator, generates the strongly continuous semigroup $(\mathrm{e}^{t(A+C)})_{t\geq 0}$. Hence, by [17, Prop. II.6.2], problem (5) is well-posed and its solution has the form $x(t) = \mathrm{e}^{t(A+C)}x_0$ for all $t \geq 0$. Moreover, by [17, Prop. I.3.5] operator C generates the strongly continuous semigroup $(\mathrm{e}^{tC})_{t\geq 0}$ on $\mathcal{H}$. Therefore, the following operator splitting methods are well-defined for $n \in \mathbb{N}$:

$$\begin{aligned} &\text{Sequential splitting:} && x_n^{\mathrm{sq}}(t) := \left(\mathrm{e}^{\frac{t}{n}A}\mathrm{e}^{\frac{t}{n}C}\right)^n x_0, \\ &\text{Strang splitting:} && x_n^{\mathrm{St}}(t) := \left(\mathrm{e}^{\frac{t}{2n}A}\mathrm{e}^{\frac{t}{n}C}\mathrm{e}^{\frac{t}{2n}A}\right)^n x_0. \end{aligned}$$

Under Assumptions 2.1, both are convergent time discretization methods, more precisely, it holds that

$$x(t) = \lim_{n\to\infty} x_n^{\mathrm{spl}}(t)$$

for all $x_0 \in \mathcal{H}$ uniformly for t in compact intervals, where "spl" stands for "sq" or "St". Therefore, in practice one defines a time step $\tau > 0$ and approximates the exact solution $x(n\tau)$ at time $t = n\tau$ by $x_n^{\mathrm{spl}}(n\tau)$ for all $n \in \mathbb{N}$, more precisely:

$$x(n\tau) \approx x_n^{\mathrm{seq}}(n\tau) = \left(\mathrm{e}^{\tau A}\mathrm{e}^{\tau C}\right)^n x_0 \qquad \text{for sequential splitting,} \tag{13}$$

$$x(n\tau) \approx x_n^{\mathrm{St}}(n\tau) = \left(\mathrm{e}^{\frac{\tau}{2}A}\mathrm{e}^{\tau C}\mathrm{e}^{\frac{\tau}{2}A}\right)^n x_0 \qquad \text{for Strang splitting.} \tag{14}$$

In [25] it was also shown that there exist constant $K_1, K_2 \geq 0$ being independent of n such that under certain restrictive assumptions on the operators and their commutators, the estimates

$$\begin{aligned} \|x(n\tau) - x_n^{\mathrm{seq}}(n\tau)\|_{\mathcal{H}} &\leq K_1\tau\log n, \\ \|x(n\tau) - x_n^{\mathrm{St}}(n\tau)\|_{\mathcal{H}} &\leq K_2\tau\log n^2 \end{aligned}$$

hold for all $n \in \mathbb{N}$, $\tau > 0$. Hence, the Strang splitting should give more accurate result if the semigroups are computed exactly.

For the examples in Section 2.1, the semigroup operators e^{tA}, $t \geq 0$, can be efficiently computed by using the Fourier transform F and its inverse F^{-1}, both being bounded linear operator from/onto $\mathcal{H}$. We define the operator $\widehat{A} \colon \mathbb{C}^m \to \mathbb{C}^m$ in the various cases as:

1D advection ($m = 1$): $\widehat{A} = -(c\mathrm{i}k)_{k\in\mathbb{Z}}$

2D advection ($m = 1$): $\widehat{A} = -(c_1\mathrm{i}k + c_2\mathrm{i}\ell)_{k,\ell\in\mathbb{Z}}$

1D shallow water ($m = 2$): $\widehat{A} = -\begin{pmatrix} \overline{U}\mathrm{i}k & \overline{h}\mathrm{i}k \\ c_{\mathrm{F}}\mathrm{i}k & \overline{U}\mathrm{i}k \end{pmatrix}_{k\in\mathbb{Z}}$

2D shallow water ($m = 3$): $\widehat{A} = -\begin{pmatrix} \overline{U}\mathrm{i}k + \overline{V}\mathrm{i}\ell & \overline{h}\mathrm{i}k & \overline{h}\mathrm{i}\ell \\ c_{\mathrm{F}}\mathrm{i}k & \overline{U}\mathrm{i}k + \overline{V}\mathrm{i}\ell & -c_{\mathrm{R}} \\ c_{\mathrm{F}}\mathrm{i}k & c_{\mathrm{R}} & \overline{U}\mathrm{i}k + \overline{V}\mathrm{i}\ell \end{pmatrix}_{k,\ell\in\mathbb{Z}}$

where i denotes the imaginary unit. Then we have $Ax = F^{-1}\widehat{A}Fx$ for all $x \in \mathrm{dom}(A)$. One can see that the range of applications can be further widened in a natural way in addition to the examples presented.

Since numerical experiments require a bounded and closed domain $\Omega \subset \mathbb{R}^2$, we consider real constants $a_\xi < b_\xi$, $a_\eta < b_\eta$ such that

$$\Omega = [a_\xi, b_\xi] \subset \mathbb{R} \quad \text{in one dimension, and}$$
$$\Omega = [a_\xi, b_\xi] \times [a_\eta, b_\eta] \subset \mathbb{R}^2 \quad \text{in two dimensions.}$$

In this case the solution is further approximated by using the discrete Fourier transform $F_N \colon \mathbb{C}^N \to \mathbb{C}^N$ defined for all fixed $t \geq 0$ and $\xi_p = p\frac{b_\xi - a_\xi}{N_\xi - 1}$, $\eta_q = q\frac{b_\eta - a_\eta}{N_\eta - 1}$ for some $N_\xi, N_\eta \in \mathbb{N}$ by the following formulae in one and two dimensions, respectively:

$$F_N w(t, \xi_p) = \sum_{k=-\frac{N_\xi}{2}}^{\frac{N_\xi}{2}-1} \widehat{w}_k(t)\mathrm{e}^{-\mathrm{i}k2\pi\xi_p},$$
$$p = 0, ..., N_\xi - 1,\ N = N_\xi,$$
$$F_N w(t, \xi_p, \eta_q) = \sum_{k=-\frac{N_\xi}{2}}^{\frac{N_\xi}{2}-1} \sum_{\ell=-\frac{N_\eta}{2}}^{\frac{N_\eta}{2}-1} \widehat{w}_{k,\ell}(t)\mathrm{e}^{-\mathrm{i}k2\pi\xi_p}\mathrm{e}^{-\mathrm{i}k2\pi\eta_q},$$
$$p = 0, ..., N_\xi - 1,\ q = 0, ..., N_\eta - 1,\ N = N_\xi N_\eta.$$

The truncated operator $\widehat{A}_N$ is defined then as $\widehat{A}_N = \widehat{A}$, but only for

$$k = -\tfrac{N_\xi}{2}, \ldots, \tfrac{N_\xi}{2} - 1 \quad \text{and} \quad \ell = -\tfrac{N_\eta}{2}, \ldots, \tfrac{N_\eta}{2} - 1.$$

To prove the convergence of the spatial discretized operators, we need the following operators.

Assumptions 3.1. Let $\mathcal{H}$ be a Hilbert space, $N \in \mathbb{N}$, and $P_N \colon \mathcal{H} \to \mathbb{R}^N$, $J_N \colon \mathbb{R}^N \to \mathcal{H}$ be linear and bounded operators with the following properties:

1. $P_N J_N = \mathrm{I}_N$, the identity operator in $\mathbb{R}^N$,
2. $\lim\limits_{N\to\infty} J_N P_N x = x$ for all $x \in \mathcal{H}$.

We choose operator P_N to be the projection onto $\mathbb{R}^N$, while J_N to be some spline interpolation (actually never needed for computations only for plotting the results). Then we have the following limit:

$$\lim_{N\to\infty} J_N F_N^{-1} \widehat{A}_N F_N P_N x = Ax \quad \text{for all} \quad x \in \mathrm{dom}(A), \tag{15}$$

and the Trotter–Kato theorem [24, Thm. 4.2, Prop. 4.3] implies that

$$\lim_{N\to\infty} J_N F_N^{-1} \mathrm{e}^{t\widehat{A}_N} F_N P_N x = \mathrm{e}^{tA} x \quad \text{for all} \quad x \in \mathcal{H}$$

uniformly for t in compact intervals.

One of the advantages of using the discrete Fourier transform is that there exist efficient methods for computing it. For the numerical experiments presented later on, we apply the `fft` and `fft2` functions of the programming language MATLAB implementing the Fast Fourier Transform for one and two dimensional array variables, respectively. Since this kind of approximation of the semigroup e^{tA} does not involve any further numerical technique, it is oppressed only by two sources of numerical errors: (i) approximation of the Fourier transform F by the discrete Fourier transform F_N, and (ii) implementation of the discrete Fourier transform F_N by the `fft`/`fft2` algorithm of the Fast Fourier Transform. The magnitude of both errors are related to the number of spatial grid points in the domain Ω and they are negligible compared to the case when any further numerical technique would be applied to approximate the semigroup e^{tA} (such as rational functions of the operator tA, etc.).

Since operator B describes the control in the physical space Ω, it is worth approximating the semigroup e^{tC} in the physical space instead of the Fourier space. To do so one needs to solve the algebraic operator Riccati equation (4) by using an approximation $A_N \colon \mathbb{R}^N \to \mathbb{R}^N$ to operator A fulfilling

$$\lim_{N\to\infty} J_N A_N P_N x = Ax \quad \text{for all} \quad x \in \mathrm{dom}(A). \tag{16}$$

For the above P_N, J_N, there could exist many possible operators A_N fulfilling the convergence property (16). For $c \in \mathbb{R}$, the differential operator $-c\partial_\zeta$ is approximated by the matrices $D_\zeta^{(c)} \in \mathbb{R}^{N_\zeta \times N_\zeta}$ for $\zeta = \xi, \eta$. The exact forms used in our numerical experiments will be given in Section 4. By using the notations

$$D_{1,\xi}^{(c)} := \tfrac{N_\xi}{2} D_\xi^{(c)} \otimes \mathrm{I}_{N_\eta} \in \mathbb{R}^{N_\xi N_\eta \times N_\xi N_\eta},$$
$$D_{\eta,1}^{(c)} := \tfrac{N_\eta}{2} \mathrm{I}_{N_\xi} \otimes D_\eta^{(c)} \in \mathbb{R}^{N_\xi N_\eta \times N_\xi N_\eta},$$

where $\otimes$ stands for the Kronecker product of the matrices, in the cases of the examples above we obtain $A_N \in \mathbb{R}^{N\times N}$ as:

1D advection: $\quad A_N = -D_\xi^{(c)}, \quad N = N_\xi,$

1D shallow water: $\quad A_N = -\begin{pmatrix} D_\xi^{(\overline{U})} & D_\xi^{(\overline{h})} \\ D_\xi^{(c_\mathrm{F})} & D_\xi^{(\overline{U})} \end{pmatrix}, \quad N = 2N_\xi,$

2D shallow water: $\quad A_N = -\begin{pmatrix} D_{1,\xi}^{(\overline{U})} + D_{\eta,1}^{(\overline{V})} & D_{1,\xi}^{(\overline{h})} & D_{\eta,1}^{(\overline{h})} \\ D_{1,\xi}^{(c_\mathrm{F})} & D_{1,\xi}^{(\overline{U})} + D_{\eta,1}^{(\overline{V})} & -c_\mathrm{R}\mathrm{I}_{N_\xi N_\eta} \\ D_{\eta,1}^{(c_\mathrm{F})} & c_\mathrm{R}\mathrm{I}_{N_\xi N_\eta} & D_{1,\xi}^{(\overline{U})} + D_{\eta,1}^{(\overline{V})} \end{pmatrix},$
$N = 3N_\xi N_\eta.$

By having operator $A_N \colon \mathbb{R}^N \to \mathbb{R}^N$ at hand, one solves the algebraic operator Riccati equation (4) with A_N instead of A, and obtains an approximation $C_N \in$

$\mathbb{R}^{N\times N}$ to the operator C such that

$$\lim_{N\to\infty} J_N C_N P_N x = Cx \quad \text{for all} \quad x\in\mathcal{H}, \tag{17}$$

and from the Trotter–Kato theorem [24, Thm. 4.2, Prop. 4.3] we have

$$\lim_{N\to\infty} J_N \mathrm{e}^{tC_N} P_N x = \mathrm{e}^{tC} x \quad \text{for all} \quad x\in\mathcal{H}$$

uniformly for t in compact intervals.

Remark 1. Since $\mathrm{e}^{tC_N}\in\mathbb{R}^{N\times N}$, we have the estimate $\|\mathrm{e}^{tC_N}\|\le \mathrm{e}^{t\|C_N\|}$ for all $N\in\mathbb{N}$. Numerical experiments suggest us to accept that there exists $\sup_{N\in\mathbb{N}}\|C_N\|<\infty$.

The combined Fourier-Splitting method to approximate the optimal state $x(n\tau)$ at time $t=n\tau$, can be written as

$$x(n\tau)\approx J_N(S_N^{\mathrm{spl}}(\tau))^n P_N x_0 \tag{18}$$

for all $x_0\in\mathcal{H}$, where "spl" stands for "seq" or "St". For time step $\tau>0$ and number $N\in\mathbb{N}$ of spatial grid points, the operators $S_N^{\mathrm{spl}}(\tau)\colon \mathbb{R}^{N\times N}\to\mathbb{R}^{N\times N}$ have the following forms:

$$S_N^{\mathrm{seq}}(\tau) = F_N^{-1}\mathrm{e}^{\tau\widehat{A}}F_N\mathrm{e}^{\tau C_N} \qquad \text{for sequential splitting,}$$
$$S_N^{\mathrm{St}}(\tau) = F_N^{-1}\mathrm{e}^{\frac{\tau}{2}\widehat{A}}F_N\mathrm{e}^{\tau C_N}F_N^{-1}\mathrm{e}^{\frac{\tau}{2}\widehat{A}}F_N \qquad \text{for Strang splitting.}$$

The convergence of the combined method (18) follows from the convergence results [8, Thm. 3.16 and 3.17].

Theorem 3.2. *Under Assumptions 2.1 and 3.1 the combined numerical methods defined in (18) are convergent, that is, for all $x_0\in\mathcal{H}$ initial value, we have the following limits:*

$$x(t) = \lim_{N,n\to\infty} J_N\big(S_N^{seq}(\tfrac{t}{n})\big)^n P_N x_0,$$
$$x(t) = \lim_{N,n\to\infty} J_N\big(S_N^{St}(\tfrac{t}{n})\big)^n P_N x_0,$$

where the convergence is uniform for t in compact intervals.

Proof. The consistency criterion in [8, Thm. 3.16 and 3.17] follows from the limits (15), (17) and from the boundedness of operators F_N, F_N^{-1}. The stability criterion holds since $\|\mathrm{e}^{t\widehat{A}_N}\|=1$ for all $N\in\mathbb{N}$, $t\ge 0$, and the strongly continous semigroups e^{tA}, e^{tC} are exponentially bounded. The exponential boundedness of the semigroup operators e^{tC_N} follows from Remark 1. □

3.1. **Algorithm.** In Section 2 we introduced the abstract LQR problem (1) for linear operators on Hilbert spaces. After semidiscretization, by finite difference or spectral method, finite dimensional LQR problems arise, where the operators B, R, Q have matrix representations B_N, R_N, Q_N. Then we may minimize the quadratic cost functional

$$J_N(P_N x_0, u_N) = \frac{1}{2}\int_0^\infty \big(\langle x_N(t), Q_N x_N(t)\rangle_{\mathcal{H}} + \langle u_N(t), R_N u_N(t)\rangle_{\mathcal{U}}\big)\mathrm{d}t, \tag{19}$$

subject to the finite dimensional Cauchy problem for $N\in\mathbb{N}$:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}x_N(t) = A_N x_N(t) + B_N u_N(t), \quad t>0, \\ x_N(0) = P_N x_0 \in \mathcal{H}. \end{cases} \tag{20}$$

Then, the optimal control is given in an analogous way to the infinite dimensional case as

$$u_N(t) = -R_N^{-1} B_N^T X_N x_N(t), \quad t \geq 0, \tag{21}$$

where X_N represents the solution to the matrix algebraic operator Riccati equation

$$Q_N + A_N^T X_N + X_N A_N - X_N B_N R_N^{-1} B_N^T X_N = 0. \tag{22}$$

A sketch of the main steps for computing the optimal state by the combined Fourier-Splitting method (18) is given below.

Algorithm 1 Computation of the optimal state

1: Define the number N of spatial grid points, the time step τ, and the number $\overline{n}$ of time integration steps.
2: Set up the initial condition x_0 for the state equation being a vector of size N.
3: Perform a finite difference approximation of the operators A, B, Q, R leading to matrices A_N, B_N, Q_N, R_N.
4: Solve the algebraic Riccati equation (22) by Newton's method.
5: Set $C_N := -B_N R_N^{-1} B_N^* X_N$.
6: Set up the initial vector $x_2^{(0)} := P_N x_0$.
7: **for** $n = 1 : \overline{n}$ **do**
8: Compute $x_1^{(n)} := \mathrm{e}^{\tau C_N} x_2^{(n-1)}$.
9: Set up $\widehat{x}_1^{(n)}$ as the discrete Fourier transform of $x_1^{(n)}$ by considering N points.
10: Compute $\widehat{x}_2^{(n)} := \mathrm{e}^{\tau \widehat{A}} \widehat{x}_1^{(n)}$.
11: Set up $x_2^{(n)}$ as the inverse discrete Fourier transform of $\widehat{x}_2^{(n)}$.
12: Set up $x_n^{\mathrm{seq}}(\overline{n}\tau) := x_2^{(\overline{n})}$ being the approximation to the optimal state at time $t = \overline{n}\tau$.

The application of Newton's method for solving Riccati equation is a standard approach, see e.g. [28]. It requires an initial value $X_N^{(0)}$ such that $A_N - B_N R_N^{-1} B_N^\top X_N^{(0)}$ is stable, i.e., all of its eigenvalues lie in the left half-plane. A possible way how to find such a starting matrix $X_N^{(0)}$ is described in [4]. Under suitable conditions, Newton's method converge quadratically, which means that its error in the actual step is proportional to the square of the error in the previous step. Our algorithm uses a stopping criterion, that is, (i) it iterates until the Frobenius norm of the left-hand size of Riccati equation (4) reaches a prediscribed a tolerance (0.01 in our numerical experiments) or (ii) the number of iteration steps reaches a maximal value (1000 in our numerical experiments). A detailed discussion on the error of this iterative method as well as its variants, like inexact Newton's method, is given in [13] and in the references therein.

4. **Numerical Experiments.** This section is devoted to present the numerical results obtained when applying our method (18) to solve the hyperbolic problems introduced in Section 2.1. We will demonstrate that it performs better than the usual grid-based methods used for solving hyperbolic problems coming from conservation laws.

We consider the first-order Godunov and the second-order Lax–Wendroff schemes, both being developed for solving hyperbolic conservation laws, for deriving the approximations A_N shown in Section 3 for computing the matrix representation C_N

of the operator $C = -BR^{-1}B^*X$. Since all the resulting matrices have non-zero entries only in their main diagonal and in their upper and lower sub-diagonals (i.e., they are tridiagonal matrices), we will refer to them as tridiag(l, d, u), where the real numbers l, u, d correspond to the entries in the lower and upper sub-diagonals and in the main diagonal, respectively. We use again the notation $-c\partial_\zeta \approx D_\zeta^{(c)}$ for $c \in \mathbb{R}$ from Section 3. Then we have the following matrices for $\zeta = \xi, \eta$:

for Godunov's scheme:

$$D_\zeta^{(c)} = \min\{0, -c\}N_\zeta \cdot \text{tridiag}(-1, 1, 0) + \max\{0, -c\}N_\zeta \cdot \text{tridiag}(0, -1, 1),$$

for Lax–Wendroff scheme:

$$D_\zeta^{(c)} = -\frac{cN_\zeta}{2} \cdot \text{tridiag}(-1, 0, 1) + \frac{c^2 N_\zeta^2 \tau^2}{2} \cdot \text{tridiag}(1, -2, 1).$$

In agreement with the derivation of Godunov and Lax–Wendroff scheme, the time-stepping uses the explicit Euler method, that is, the exponential $\mathrm{e}^{\tau M}$, appearing in Steps 8 and 10 in Algorithm 1, is approximated by the term $\mathrm{I} + \tau M$ for any matrix M and for all $\tau \geq 0$, where I denotes the identity operator on the corresponding space. We remark that since only the action of the matrix exponential on a vector is needed in each time step, the idea and method presented in [2] could also be applied, however, it takes much more computational time than explicit Euler method and (due to the dominating spatial discretization error) does not give more accurate solution in these cases.

For $\Omega \subset \mathbb{R}^d$, we choose $\mathcal{H} = \mathcal{U} = (\mathrm{L}^2(\Omega))^d$ for $d = 1, 2, 3$, and fix the time step $\tau > 0$ and the number $N \in \mathbb{N}$ of grid points being specified for each example. We treat then the following three cases for the abstract LQR problem (1).

1. The choice $B = 0$ corresponds to the case when the dynamics is not controlled. It results in the matrix representation $B_N = 0 \in \mathbb{R}^{N\times N}$.
2. The choice $B = \mathrm{I}_\mathcal{U}$, the identity operator on $\mathcal{U}$, corresponds to a distributed control which acts at each spatial points in Ω. It results in the matrix representation $B_N = I_N \in \mathbb{R}^{N\times N}$, the identity matrix.
3. The choice $B = B_\Gamma$ corresponds to a control which acts only on a subset $\Gamma \subset \Omega$ of the spatial points representing a sink inside the spatial domain Ω. For our numerical experiments we choose

$$\Gamma_1 = [0.6(b_\xi - a_\xi), 0.8(b_\xi - a_\xi)] \quad \text{in one dimension,}$$

$$\Gamma_{2\ell} = \{(\xi, \eta) \in \mathbb{R}^2 \colon \xi \in [a_\xi, 0.33(b_\xi - a_\xi)],\ \eta \in [0.67(b_\xi - a_\xi), b_\xi]\}$$

in two dimensions (lake),

$$\Gamma_{2r} = \{(\xi, \eta) \in \mathbb{R}^2 \colon \xi \in [0.4(b_\xi - a_\xi), 0.5(b_\xi - a_\xi)],\ \eta \in [0.3(b_\xi - a_\xi), 0.7(b_\xi - a_\xi)]\}$$

in two dimensions (river).

The matrix representation in these cases results in a matrix having values 1 at places of grid points belonging to the corresponding subset Γ and zeros otherwise.

In the code we specify $Q_N = I_N$ and $\widetilde{R}_N^{-1} := B_N R_N^{-1} B_N^* = rI_N$ with $r \in \mathbb{R}$ for cases (ii)-(iv) above. Remember that due to the choice $B = 0$, there is no control process in case (i).

It is worth mentioning that, although their orders differ, we obtained the same results for both the sequential and the Strang splittings. The explanation of this phenomena is that the numerical error is dominated by the spatial discretization error originating from the finite difference approximation of operator A, and not from the error caused by the splitting procedures. We remark that the results presented in this section are the same also for smaller time steps as presented, which means again that the error oppressed the numerical solution is dominated by the error of the corresponding spatial discretization scheme.

4.1. **1D Advection.** We consider problem (7) with parameters $a_\xi = 0$, $b_\xi = 1$, $c = 1$, $r = 10$, and the initial function

$$h_0(\xi) = 10^{-4}\mathrm{e}^{-200(\xi-0.5)^2}. \tag{23}$$

This corresponds e.g. to a system with measures $a_\xi = 0km$, $b_\xi = 10km$ and velocity $c = 10\ km/h$ where the height of the initial function is $5\ m$.

Figures 1, 2 show the solution to the one-dimensional advection equation (7) without control ($B = 0$) at time $t = 3.2$ by using Godunov and Lax–Wendroff schemes, respectively. Hence, it corresponds to the right shift of the initial function (23). Due to the periodic boundary condition, the initial function should travel through the spatial interval $\Omega = [0, 1]$ more than three times without changing its shape. Figure 3 shows the time-evolution of the volume ratio

$$\mathcal{V}_1(t) := \frac{\int_\Omega w(t,\xi)\mathrm{d}\xi}{\int_\Omega w_0(\xi)\mathrm{d}\xi}, \quad t \geq 0.$$

Since the exact solution does not change its shape in time, the volume should remain constant, that is, $\mathcal{V}(t) = 1$ should hold. The Lax–Wendroff scheme introduces spuriuos oscillatios, even when the stability condition $\tau N_\xi |c| < 1$ is satisfied. Since Godunov's scheme is of first order, the shape of the solution is inaccurate. However, both methods preserve the volume exactly. The application of Fourier's method exactly preserves the shape and the volume of the initial function already for a small number of grid points ($N_\xi = 64$), that is, it coincides with the curves labelled "exact" on the figures.

Figures 4, 5, and 6 show the solution and the volume ratio for the distributed control ($B = \mathrm{I}_\mathcal{U}$). Since this case corresponds to a control acting at each spatial point, we should obtain the same picture as in the previous case $B = 0$, just with smaller values of water's height, and an exponentially decaying volume ratio. One can see that the Fourier-Splitting method overperforms the grid-based method in both cases. For the Fourier-Splitting method with first-order Godunov's scheme, one needs more grid points N to obtain an acceptable result. However, the application of the Fourier-Splitting method with second-order Lax–Wendroff scheme provides a sufficiently accurate result already for the case $N = 64$ which has oscillations without using splitting method. The volume ratio behaves in the desired way in all cases.

Figures 7, 8, and 9, 10 show the results for the solution and the volume ratio in the case of the sink-like control ($B = \mathrm{I}_{\Gamma_1}$). The solutions' shapes are similar to those in the case of distributed control. Again, the Fourier-Splitting method performs better than the grid-based methods without using splitting. The volume ratio behaves in time as expected, that is, it decreases as long as the material is above the sink, and stays constant when it is outside the sink. In Figures 11, 12 we present an enlargement of Figures 9, 10, respectively. One can see that Godunov's scheme

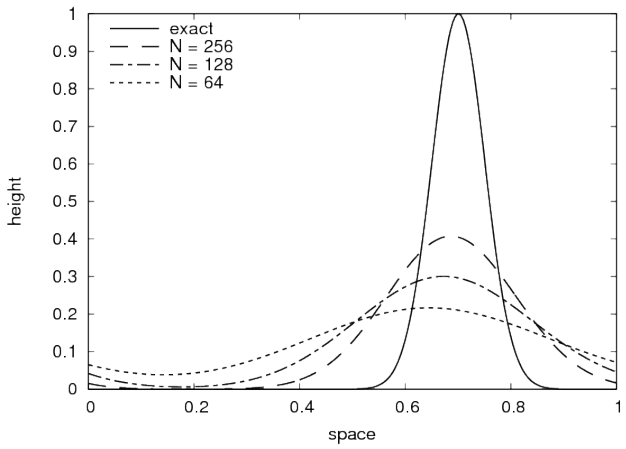

FIGURE 1. Solution $w(t,\xi)$ to the one-dimensional advection equation (7) without control ($B = 0$) at time $t = 3.2$ by using Godunov's scheme with time step $\tau = 10^{-3}$.

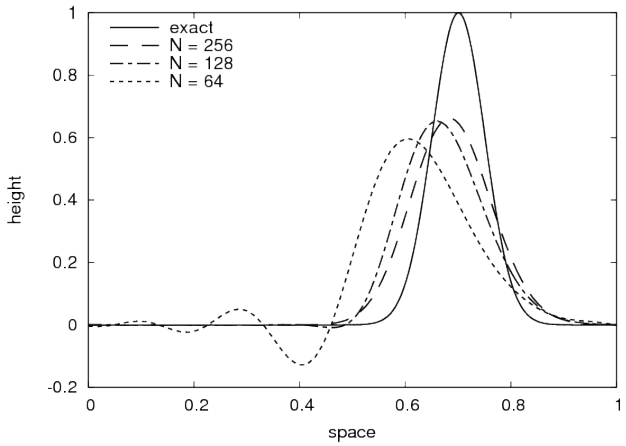

FIGURE 2. Solution $w(t,\xi)$ to the one-dimensional advection equation (7) without control ($B = 0$) at time $t = 3.2$ by using Lax–Wendroff scheme (right panel) with time step $\tau = 10^{-3}$.

introduces a first-order error, and becomes sufficiently accurate only for $N = 256$ grid points. Since the Lax–Wendroff scheme is of second order, its error remains smaller, but the Fourier-Splitting method overperforms the grid-based variant also in this case. The Fourier-Splitting methods shows the expected behaviour.

4.2. **2D Advection Equation.** We consider problem (8) with parameters $a_\xi = 0$, $b_\xi = 1$, $\mathbf{c} = (-0.1, 0.1)$, $r = 10^2$, and the initial function

$$h_0(\xi,\eta) = 10^{-4}\mathrm{e}^{-200((\xi-0.5)^2+(\eta-0.5)^2}. \tag{24}$$

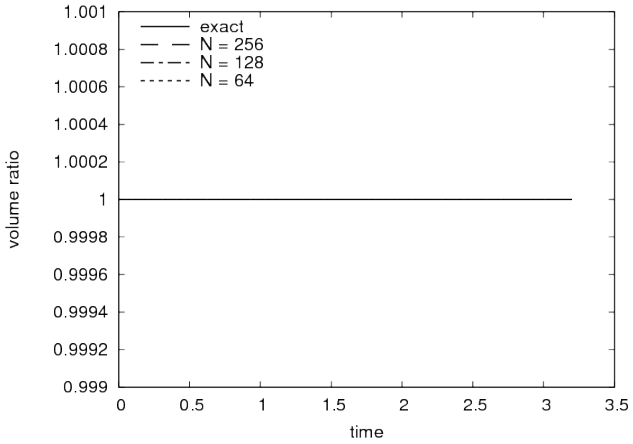

FIGURE 3. Volume ratio $\mathcal{V}_1(t)$ of the one-dimensional advection equation (7) without control ($B = 0$) by using Godunov's scheme or Lax–Wendroff scheme with time step $\tau = 10^{-3}$.

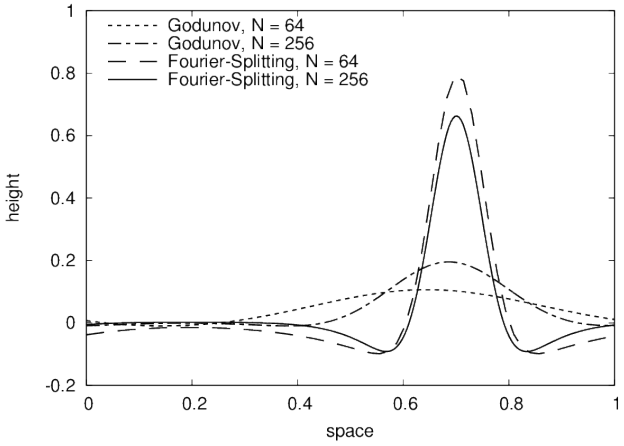

FIGURE 4. Solution $w(t, \xi)$ to the one-dimensional advection equation (7) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 3.2$ by using Godunov's scheme with time step $\tau = 10^{-3}$.

Figure 13 shows the solution to the two-dimensional advection equation (8) without control ($B = 0$) at time $t = 3.2$ by using Fourier's method (left column), considered to be the exact one, and Lax–Wendroff scheme (right column) for $N_\xi = N_\eta = 32, 64, 128$, respectively. Hence, it corresponds to the left-up shift of the initial function (24). One can see that the Lax–Wendroff scheme needs at least 128×128 grid points for reaching a nearly accurate solution. When there is no control ($B = 0$), both methods preserve the volume exactly (by construction,

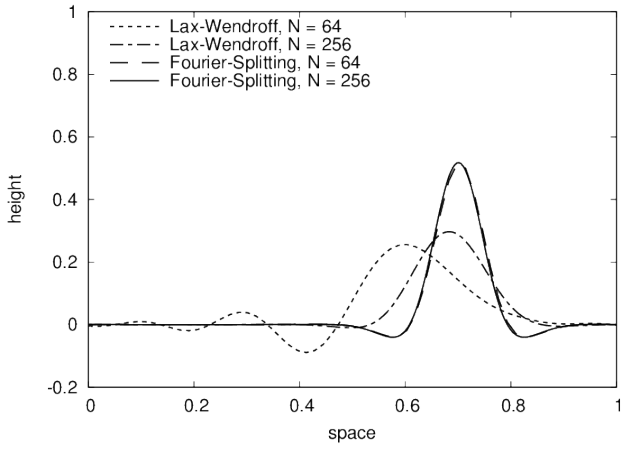

FIGURE 5. Solution $w(t,\xi)$ to the one-dimensional advection equation (7) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 3.2$ by using Lax–Wendroff scheme with time step $\tau = 10^{-3}$.

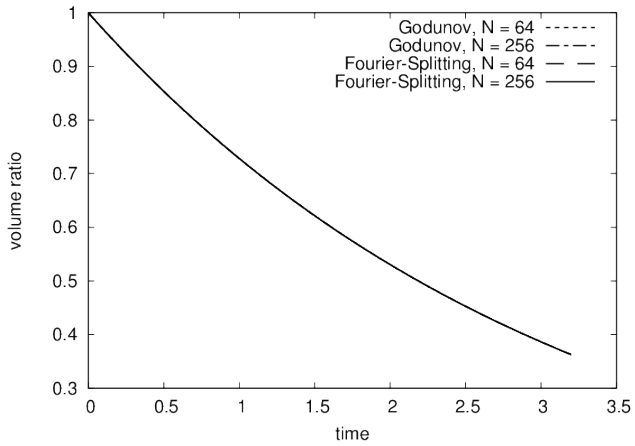

FIGURE 6. Volume ratio $\mathcal{V}_1(t)$ of the one-dimensional advection equation (7) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 3.2$ by using Godunov's scheme or Lax–Wendroff scheme with time step $\tau = 10^{-3}$.

and numerically as well), that is, the volume ratio

$$\mathcal{V}_2(t) := \frac{\iint_\Omega w(t,\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta}{\iint_\Omega w_0(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta}, \quad t \geq 0,$$

stays constant one for the whole integration time, therefore, we omit the figure showing its time behaviour (cf Figure 3).

Figure 14 shows the solution to the two-dimensional advection equation (8) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 3.2$ by using Lax–Wendroff scheme and

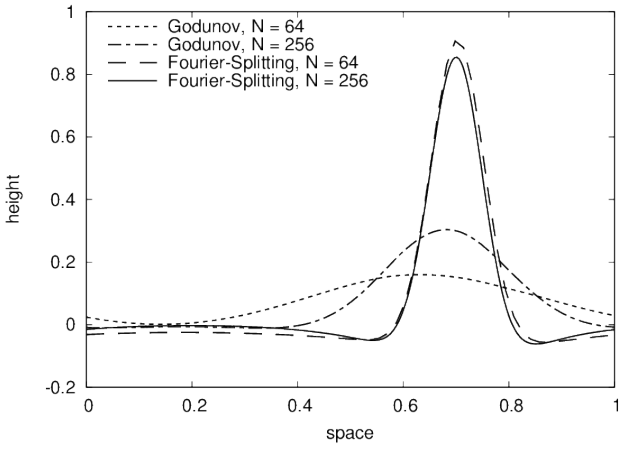

FIGURE 7. Solution $w(t,\xi)$ to the one-dimensional advection equation (7) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 3.2$ by using Godunov's scheme with time step $\tau = 10^{-3}$.

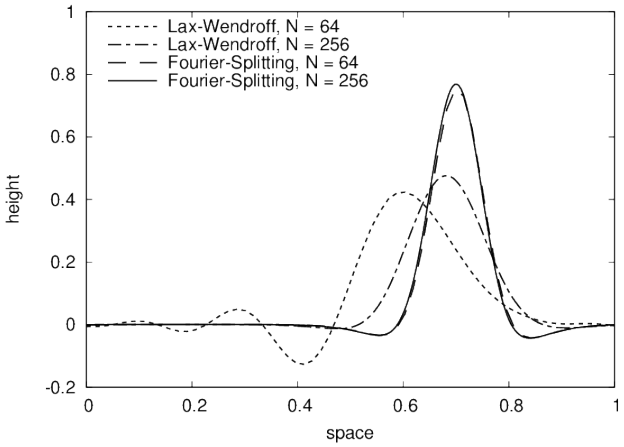

FIGURE 8. Solution $w(t,\xi)$ to the one-dimensional advection equation (7) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 3.2$ by using Lax–Wendroff scheme with time step $\tau = 10^{-3}$.

Fourier-Splitting. One can see that the Fourier-Splitting yields an accurate solution already for $N_\xi = N_\eta = 32$.

In Figure 15 the time-evolution of the volume ratio $\mathcal{V}_2$ is showed for distributed control ($B = \mathrm{I}_{\mathcal{U}}$). Although it is hard to see, the curves of the Fourier-Splitting follow the curves of the corresponding curves of the Lax–Wendroff methods: The upper curve corresponds to the case $N_\xi = 32$, and lower curve to $N_\xi = 64$.

4.3. **1D Linearized Shallow Water Equations.** We consider problem (9) with parameters $a_\xi = 0$, $b_\xi = 1$, $\overline{h} = 5 \cdot 10^{-4}$, $\overline{U} = 2 \cdot 10^{-4}$, $\mathrm{Fr} = 0.32$ ($c_\mathrm{F} \approx 9.77$),

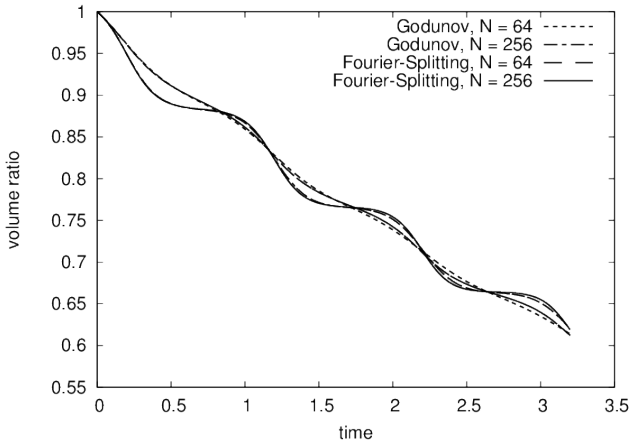

FIGURE 9. Volume ratio $\mathcal{V}_1(t)$ of the one-dimensional advection equation (7) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 3.2$ by using Godunov's scheme with time step $\tau = 10^{-3}$.

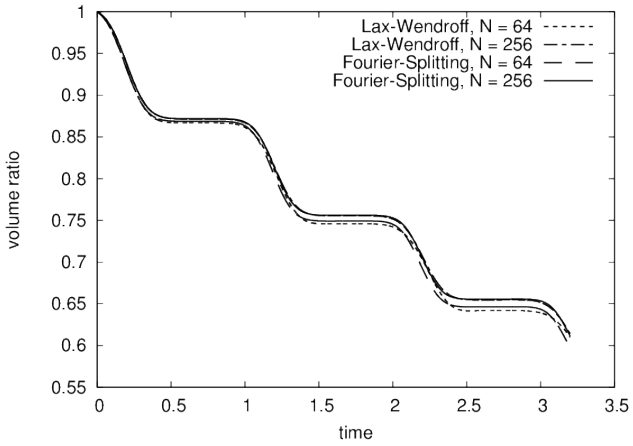

FIGURE 10. Volume ratio $\mathcal{V}_1(t)$ of the one-dimensional advection equation (7) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 3.2$ by using Lax–Wendroff scheme with time step $\tau = 10^{-3}$.

$r = 10^8$, and the initial functions

$$h_0(\xi) = 10^{-4}\mathrm{e}^{-200(\xi-0.5)^2} \quad \text{and} \quad u_0(\xi) = 0.$$

This corresponds e.g. to a system with measures $a_\xi = 0\,km$, $b_\xi = 10\,km$ where the height of the initial function is $5\,m$.

We note that similarly to the case of the advection equations, we obtained the same numerical results for the sequential and the Strang splittings, indicating that the error oppressing the solution is dominated by the spatial discretization error.

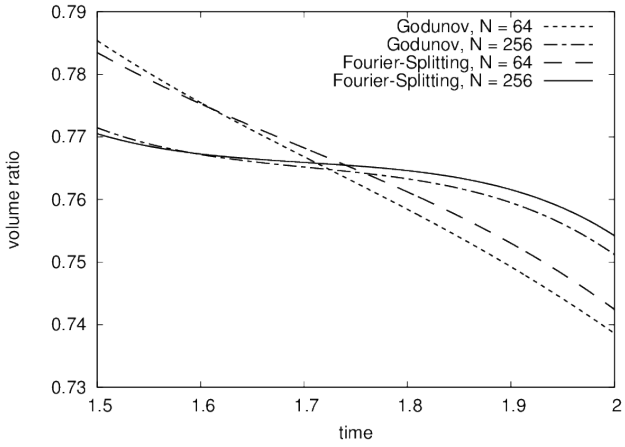

FIGURE 11. Enlargements of Figure 9.

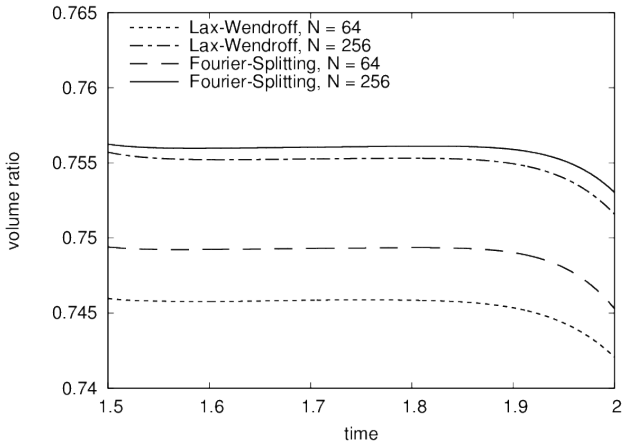

FIGURE 12. Enlargements of Figure 10.

Since $h(t,\xi)$ describes the deviation from $\overline{h}$, the volume ratio is defined in this case as

$$\mathcal{V}_3(t) := \frac{\overline{h} + \int_\Omega h(t,\xi)\mathrm{d}\xi}{\overline{h} + \int_\Omega h_0(\xi)\mathrm{d}\xi}, \quad t \geq 0.$$

The shallow water equations preserve the water's volume as well, therefore, $V_3(t) = 1$ should hold for all $t \geq 0$ when the system is not controlled ($B = 0$).

Figures 16–21 show the solutions and time-evolution of the volume ratio $\mathcal{V}_3$ in the cases of no control ($B = 0$), distributed control ($B = \mathrm{I}_{\mathcal{U}}$), and sink-like control ($B = \mathrm{I}_{\Gamma_1}$), respectively. It is clearly seen that, as for the advection equations, that the Fourier-Splitting yields an accurate solution already for $N_\xi = 32$ (and the same for $N_\xi = 128$), while the Lax–Wendroff method reaches this accuracy for $N_\xi = 128$ or even more grid points. Although it is hard to see from the figures, the volume ratio of Fourier-Splitting behaves in the same way as that of the Lax–Wendroff

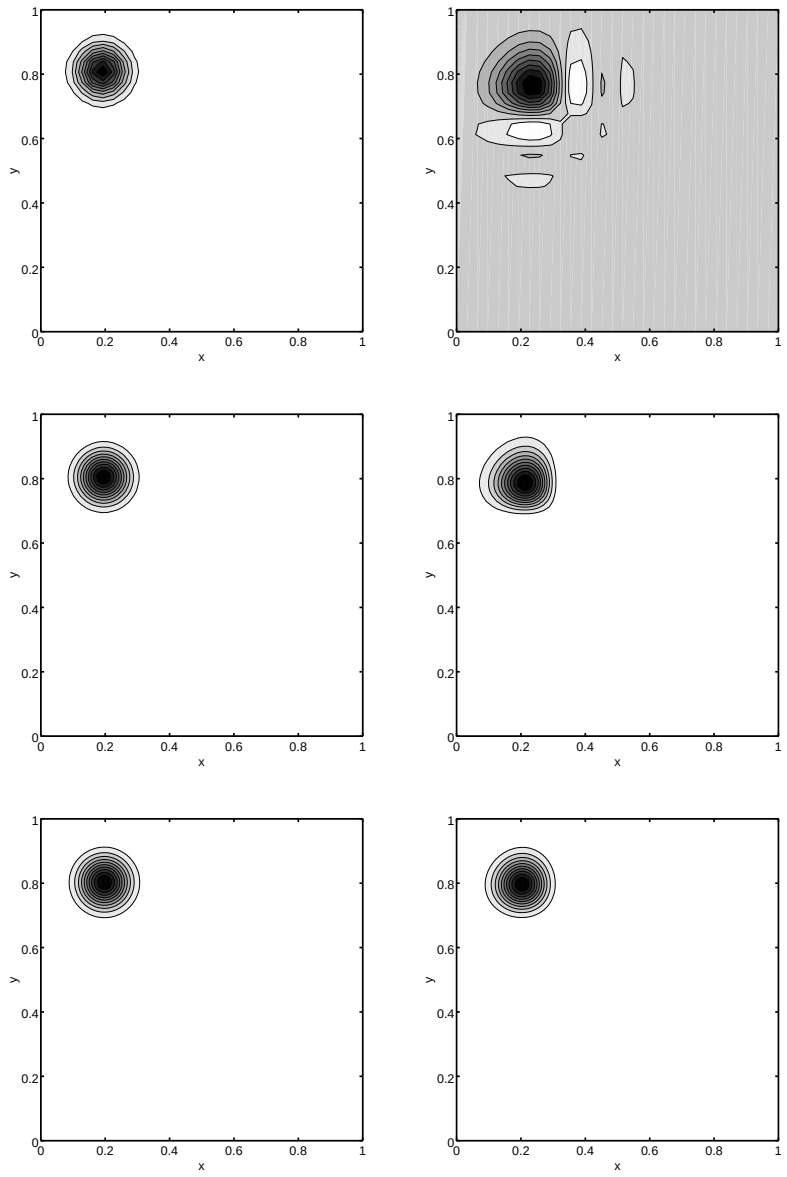

FIGURE 13. Solution to two-dimensional advection equation (8) at time $t = 3$ without control ($B = 0$) by using Fourier's method (left column) and Lax–Wendroff scheme (right column) with time step $\tau = 10^{-3}$ and number of grid points $N_\xi = N_\eta = 32, 64, 128$ from top to bottom, respectively.

method with the same number of grid points: The upper curve corresponds to the case $N_\xi = 32$, and lower curve to $N\xi = 128$. Hence, the Fourier-Splitting method performs better then the purely grid-based method. We remark that the local increase of volume ratio is due to the numerical error introduced by the effect of

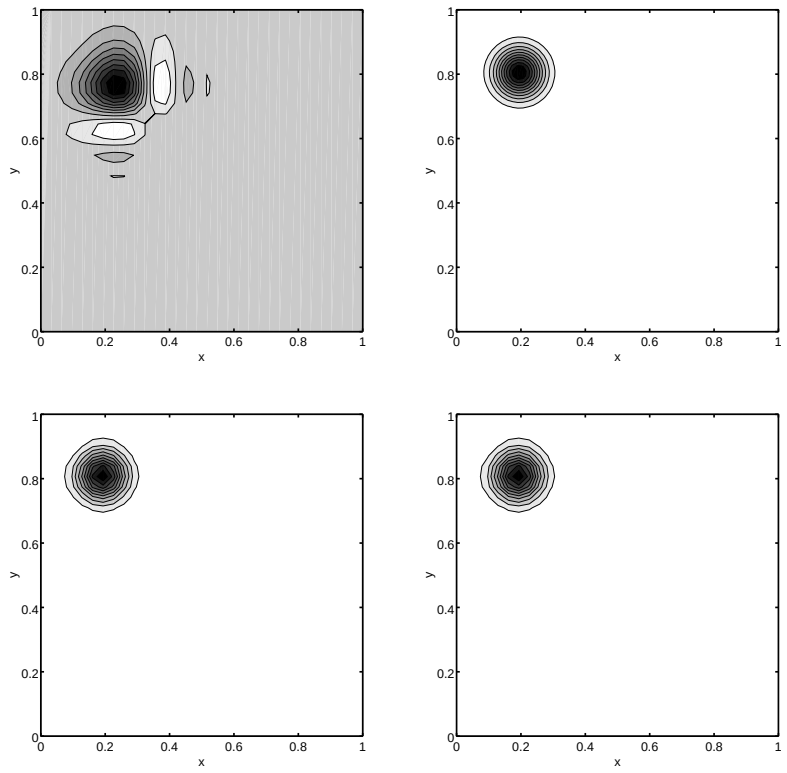

FIGURE 14. Solution to two-dimensional advection equation (8) at time $t = 3$ with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) by using Lax–Wendroff scheme (left column) and Fourier-Splitting (right column) with time step $\tau = 10^{-3}$ and number of grid points $N_\xi = N_\eta = 32, 64$ from top to bottom, respectively.

matrix C_N, and it is present in case of both methods. Since Godunov method is of first order, its solution to problem (9) is more diffusive as that of the second-order Lax–Wendroff method, depicted on Figure 18. Otherwise, Godunov method yields similar results, therefore, we omit them.

One can see that the results obtained for one-dimensional shallow water equations (9) are very similar to the ones for the advection equations (7) and (8). It is because the advection term (6) appears in each element of the operator matrix (10) corresponding to equations (9). Hence, we expect similar behaviour in the case of the two-dimensional shallow water equations, being our main example, as well.

4.4. **2D Linearized Shallow Water Equations.** In order to numerically test the Fourier-Splitting method given by Algorithm 1 for the two-dimensional shallow water equations, we consider two problems. The first problem illustrates a water drop in the middle of the spatial domain, which models e.g. a heavy rainfall over a lake. The second problem describes the time-evolution of an initial wave, which models a flood along a river segment.

Heavy rainfall over a lake.

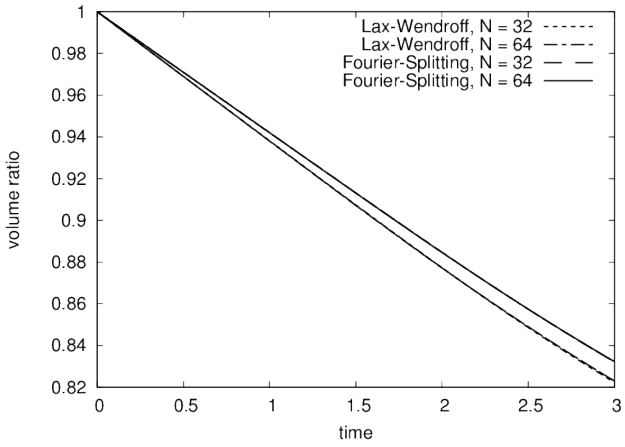

FIGURE 15. Volume ratio $\mathcal{V}_2(t)$ of two-dimensional advection equation (8) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) by using Lax–Wendroff scheme and Fourier-Splitting with time step $\tau = 10^{-3}$.

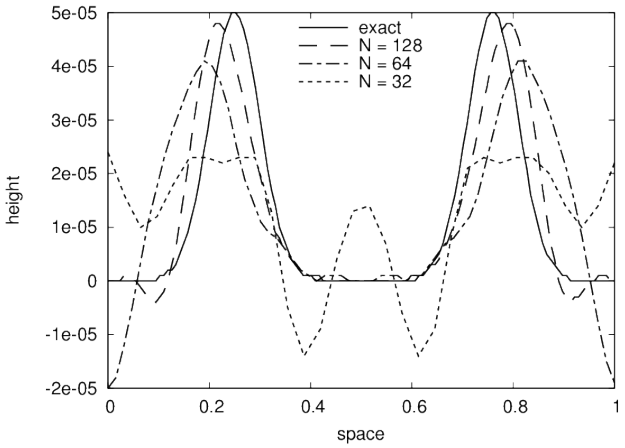

FIGURE 16. Solution to the one-dimensional linearized shallow water equations (9) without control ($B = 0$) at time $t = 25$ with time step $\tau = 10^{-4}$.

We consider now problem (11) with parameters $a_\xi = 0$, $b_\xi = 1$, $a_\eta = 0$, $b\eta = 1$, $\overline{h} = 5 \cdot 10^{-4}$, $\overline{u} = -10^{-2}$, $\overline{v} = 10^{-2}$, Fr $= 0.32$, Ro $= 0.1$, $r = 10^3$, and the initial function

$$\begin{aligned} h_0(\xi,\eta) &= 10^{-4} \cdot \mathrm{e}^{-200((\xi-0.5)^2+(\eta-0.5)^2)} \\ U_0(\xi,\eta) &= 0, \\ V_0(\xi,\eta) &= 0, \end{aligned}$$

which simulates a water drop in the middle of the spatial domain $\Omega = [0,1] \times [0,1]$. This case corresponds to a lake or sea with a heavy rainfall over its middle area.

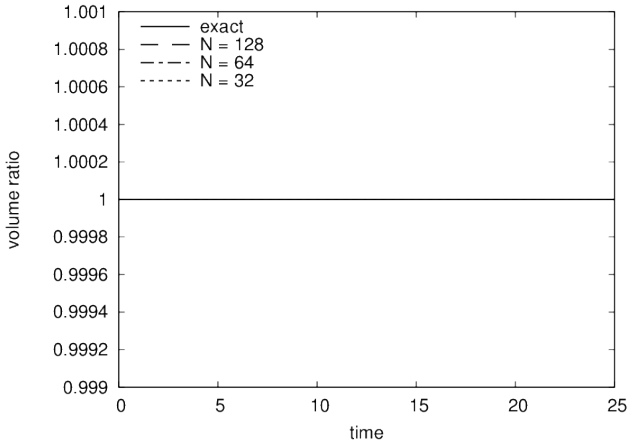

FIGURE 17. Time-evolution of the volume ratio $\mathcal{V}_3$ for the one-dimensional linearized shallow water equations (9) without control ($B = 0$) at time $t = 25$ with time step $\tau = 10^{-4}$.

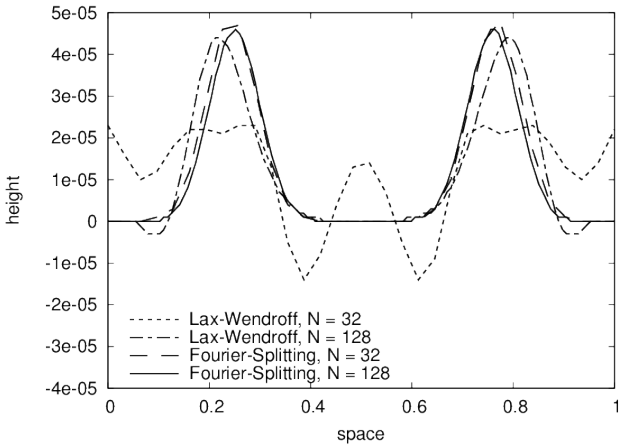

FIGURE 18. Solution to the one-dimensional linearized shallow water equations (9) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 25$ with time step $\tau = 10^{-4}$.

Figure 22 shows the solution to the two-dimensional shallow water equations (12) without control ($B = 0$) at time $t = 4.5$ by using Fourier's method (left column) and Lax–Wendroff scheme (right column) for $N_\xi = N_\eta = 16, 32, 64$, respectively. One can see that the Fourier's method provides an accurate solution already for $N_\xi = N_\eta = 64$, while the Lax–Wendroff scheme needs more than 128×64 grid points for reaching that accuracy. The Lax–Wendroff scheme's dissipativity is clearly seen in this case as well. When there is no control ($B = 0$), both methods preserve the

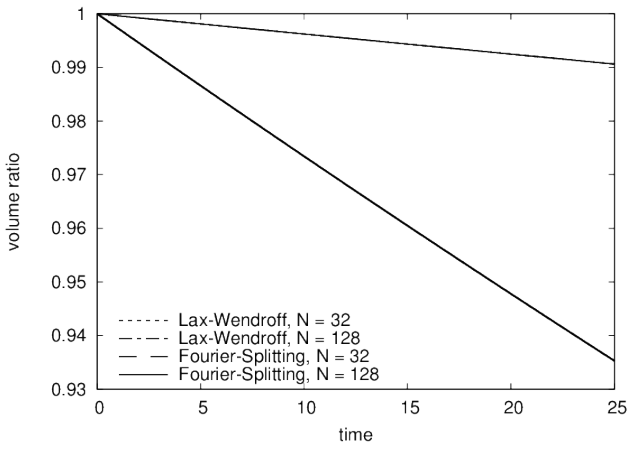

FIGURE 19. Time-evolution of the volume ratio $\mathcal{V}_3$ for the one-dimensional linearized shallow water equations (9) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 25$ with time step $\tau = 10^{-4}$.

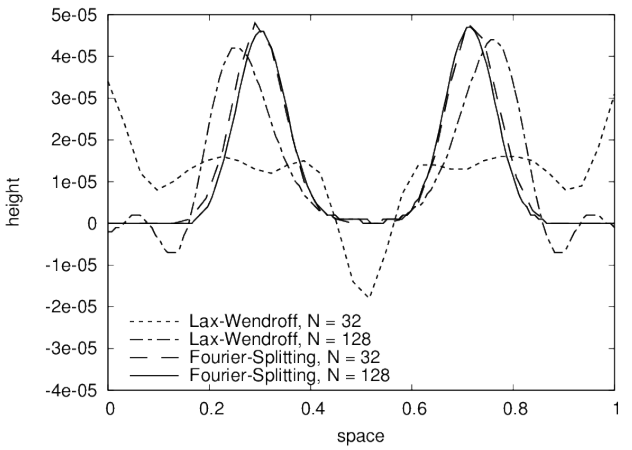

FIGURE 20. Solution to the one-dimensional linearized shallow water equations (9) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 40$ with time step $\tau = 10^{-4}$.

volume exactly (by construction, and numerically as well), that is, the volume ratio

$$\mathcal{V}_4(t) := \frac{\overline{h} + \iint_\Omega h(t,\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta}{\overline{h} + \iint_\Omega h_0(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta}, \quad t \geq 0,$$

stays constant one for the whole integration time.

Figure 23 shows the solution to the two-dimensional shallow water equations (12) with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) at time $t = 4.5$ by using Lax–Wendroff scheme (left column) and Fourier-Splitting (right column) for $N_\xi = N_\eta = 16, 32$, respectively. As before, the Fourier-Splitting yields an accurate solution for fewer grid points than the Lax–Wendroff scheme. Figure 24 shows the time-evolution of

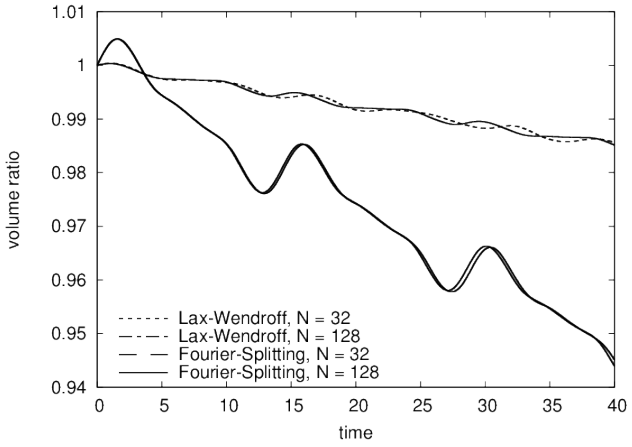

FIGURE 21. Time-evolution of the volume ratio $\mathcal{V}_3$ for the one-dimensional linearized shallow water equations (9) with sink-like control ($B = \mathrm{I}_{\Gamma_1}$) at time $t = 40$ with time step $\tau = 10^{-4}$.

the volume ratio V_4 for distributed control ($B = \mathrm{I}_{\mathcal{U}}$), which behaves in the same way for both methods: The upper curve corresponds to $N_\xi = N_\eta = 16$, the lower to $N_\xi = N_\eta = 32$.

Figure 25 shows the solution to the two-dimensional shallow water equations (12) with sink-like control ($B = \mathrm{I}_{\Gamma_{2\ell}}$) at time $t = 4.5$ by using Lax–Wendroff scheme (left column) and Fourier-Splitting (right column) for $N_\xi = N_\eta = 16, 32$, respectively. One can see that both methods yield unstable solutions for $N_\xi = N_\eta = 16$, meanwhile, they both capture the correct shape of the solution for $N_\xi = N_\eta = 32$. Also this example shows that the Fourier-Splitting method results in a more accurate solution for $N_\xi = N_\eta = 32$ as the Lax–Wendroff scheme. We remark, however, that the star shaped phenomenon in the upper left corner of the figure is due to the numerical error originating at the place of the "sink", being visible for both methods. Since the case $N_\xi = N_\eta = 16$ leads to unstable solutions for both methods, Figure 26 shows the time-evolution of the volume ratio V_4 for the sink-like control ($B = \mathrm{I}_{\Gamma_1}$) only for $N_\xi = N_\eta = 32$. The two curves differ more significantly in this case as before, however, the Fourier-Splitting shows the expected behaviour. Flood along a river segment.

As another example we consider problem (11) with parameters $a_\xi = 0$, $b_\xi = 10$, $a_\eta = 0$, $b\eta = 1$, $\overline{h} = 1$, $\mathrm{Fr} = \frac{1}{\sqrt{9.80665}}$, $\mathrm{Ro} = \frac{1}{60}10^4$ and the initial function which simulates a wave at the very left of the spatial domain $\Omega = [0, 10] \times [0, 1]$ spreading to the right-hand side. This case represents a more practical example: It models a flood along a river segment.

Figure 27 presents our results when the sink-like control $B = \mathrm{I}_{\mathcal{U}}$ is applied in the domain Γ_{2r}, with the purpose to represent a sink that sucks in the water once the wave reaches that point. The left column shows the time-evolution of the wave without control ($B = 0$), while the other two columns shows the results obtained by the Fourier-Splitting method with sequential and Strang splittings, respectively. It is easy to visualize that the wave at the right-hand side of the river segment is

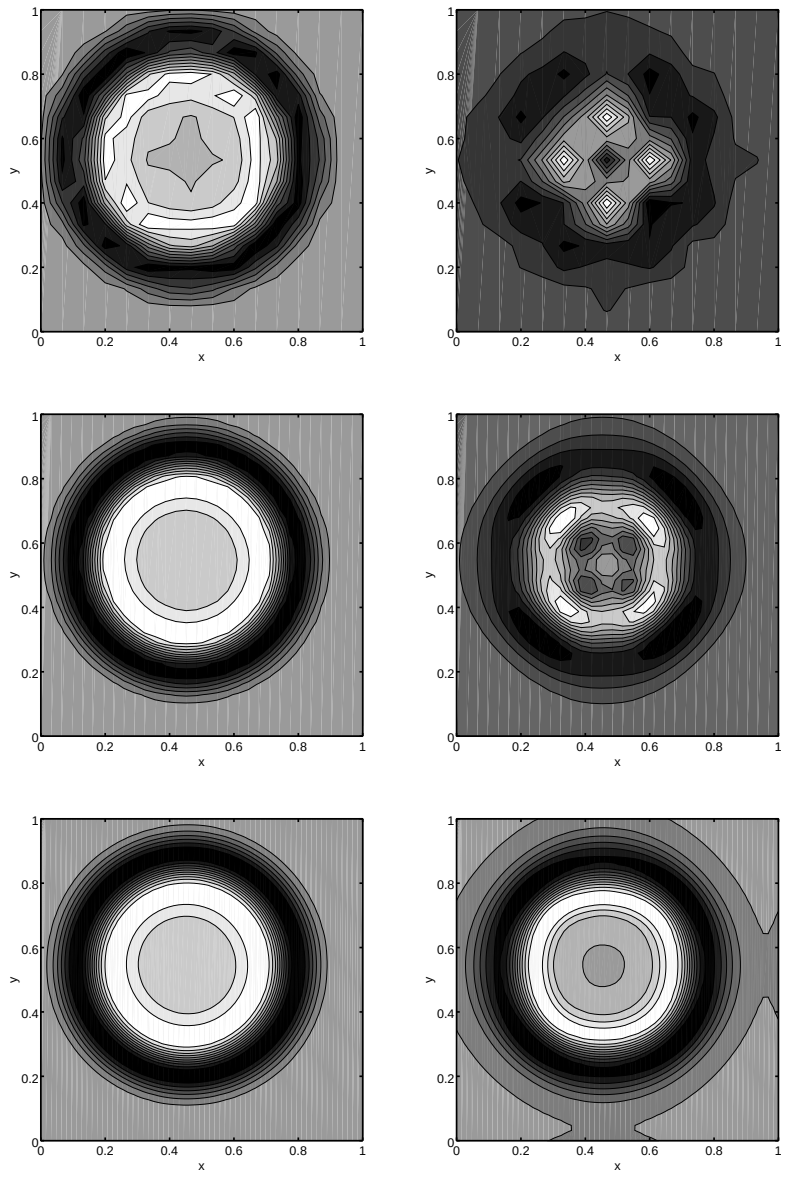

FIGURE 22. Solution to two-dimensional shallow water equations (12) at time $t = 4.5$ without control ($B = 0$) by using Fourier's method (left column) and Lax–Wendroff scheme (right column) with time step $\tau = 10^{-4}$ and number of grid points $N_\xi = N_\eta = 16, 32, 64$ from top to bottom, respectively.

significantly smaller with control than without control, and it bevahaves in time as expected.

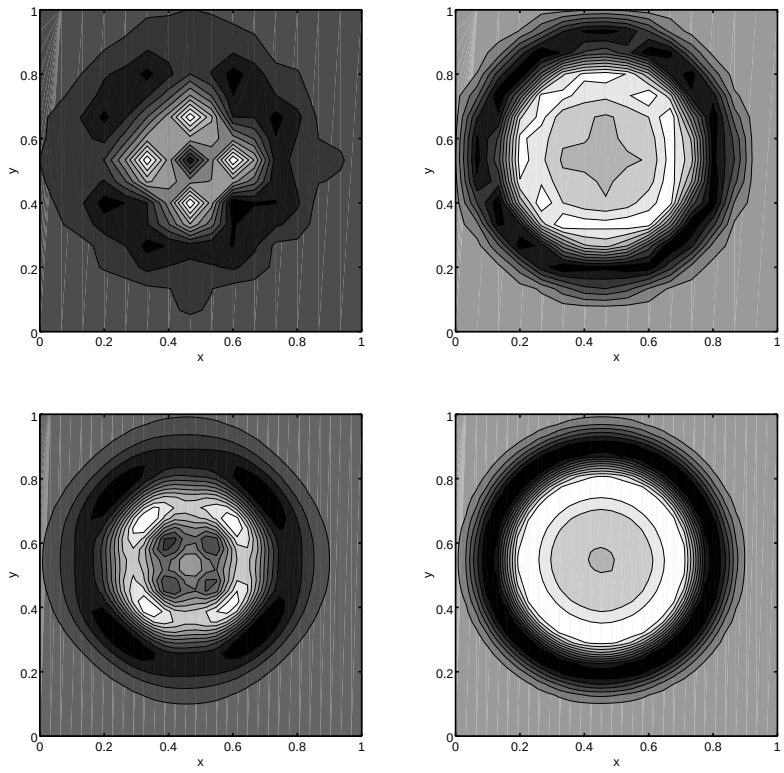

FIGURE 23. Solution to two-dimensional shallow water equations (12) at time $t = 4.5$ with distributed control ($B = \mathrm{I}_{\mathcal{U}}$) by using Lax–Wendroff scheme (left column) and Fourier-Splitting (right column) with time step $\tau = 10^{-4}$ and number of grid points $N_\xi = N_\eta = 16, 32$ from top to bottom, respectively.

As expected, we obtained similar results for the two-dimensional linearized shallow water equations as for the one-dimensional advection and shallow water equations. This happens due to the advection-like terms (6) in the matrix (12).

5. **Conclusions and future work.** Numerical experiments show that the Fourier-Splitting method proposed in this paper performs better than standard grid-based methods, developed for hyperbolic conservation laws, for approximating the optimal state of the linear quadratic optimal control problem governed by partial differential equations. Meanwhile the Fourier-Splitting yields the same time-evolution of the solution's spatial integral (water's volume in the examples), it leads to more accurate results for a fixed number of spatial grid points, thus, it can achieve a prediscribed accuracy for fewer spatial grid points than the purely grid-based methods.

Our approach will play, therefore, a key role when solving *nonlinear* optimal control problems by model predictive control/receding horizon techniques. Since many linear optimal control problems have to be solved in given subintervals, a fast solver for computing the optimal state allows one to deal with the nonlinear

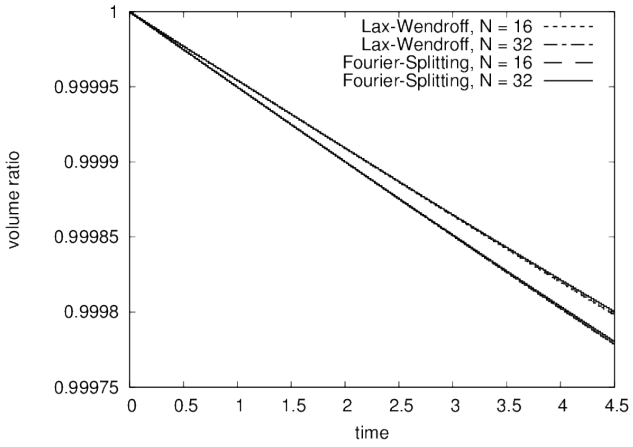

FIGURE 24. Volume ratio $\mathcal{V}_4$ for two-dimensional shallow water equations (12) for distributed control ($B = \mathrm{I}_{\mathcal{U}}$) by using time step $\tau = 10^{-4}$.

problem efficiently in this context. Particularly, we are interested in the optimal control problem where the dynamics corresponds to the shallow water equations. We intend to study its application for flood protecion in our future work.

The stochastic variant of the infinite dimensional LQR problem was studied in the same setting in [21, 15, 18, 34]. Recently, a theoretical framework for this problem has been laid for singular estimates control systems in the presence of noise in the control and in the case of finite time penalization in the performance index, see [19]. Moreover, considering this setting in [33], an approximation scheme for solving the control problem and the associated Riccati equation has been proposed in [35]. Therefore, based on a novel idea that generalizes splitting methods to stochastic PDEs in [27], the results of this paper can be extended for solving the optimal state of the stochastic LQR problems, too.

Acknowledgments. The authors thank Johannes Schwaighofer (Innsbruck) for his collaboration in the early phase of the work. The financial support of the Stiftung Aktion Österreich-Ungarn (project nr. 89öu11) is kindly acknowledged.

REFERENCES

[1] H. Abou-Kandil, G. Freiling, V. Ionescu and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser, Basel, Switzerland, 2003.
[2] A.H. Al-Mohy and N.J. Higham, Computing the action of the matrix exponential, with an application to exponential integrators, *SIAM J. Sci. Comput.*, **33** (2011), 488–511.
[3] E. Arias, V. Hernández, J. Ibanes and J. Peinado, A family of BDF algorithms for solving differential matrix Riccati equations using adaptive techniques, *Procedia Computer Science*, **1** (2010), 2569–2577.
[4] E. Armstrong, An extension of Bass' algorithm for stabilizing linear continuous constant systems, *IEEE Trans. Automatic Control*, **AC-20** (1975), 153–154.
[5] A. Balakrishnan, *Applied Functional Analysis*, Springer-Verlag, New York, 1981.
[6] H. Banks, R. Smith and Y. Wang, The modeling of piezoceramic patch interactions with shells, plates and beams, *Quart. Appl. Math.*, **53** (1995), 353–381.
[7] A. Bátkai, P. Csomós, B. Farkas and G. Nickel, Operator splitting for non-autonomous evolution equations, *J. Funct. Anal.* **260** (2011), 2163–2192.

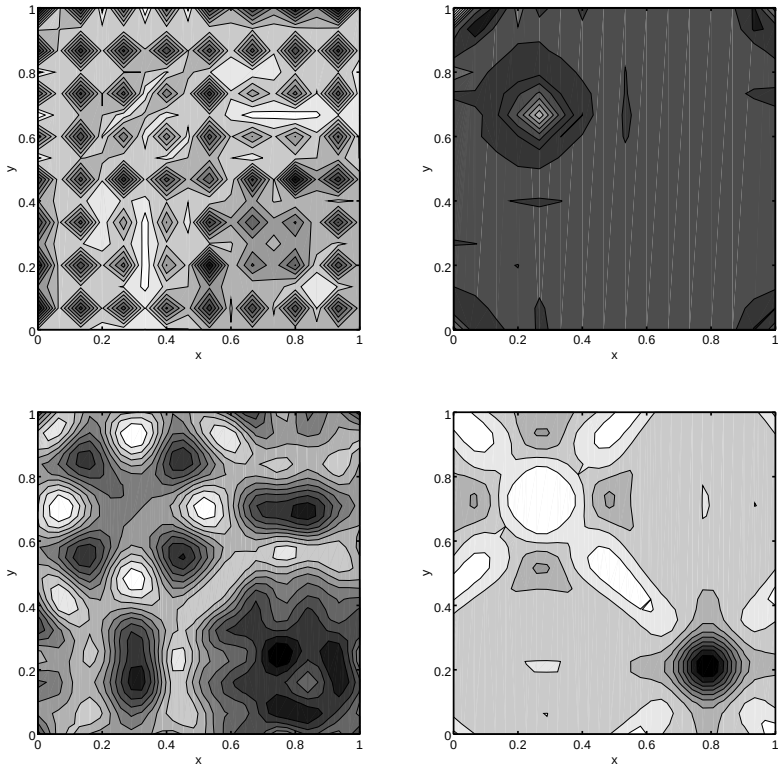

FIGURE 25. Solution to two-dimensional shallow water equations (12) at time $t = 4.5$ with sink-like control ($B = \mathrm{I}_{\Gamma_{2\ell}}$) by using Lax–Wendroff scheme (left column) and Fourier-Splitting (right column) with time step $\tau = 10^{-4}$ and number of grid points $N_\xi = N_\eta = 16, 32$ from top to bottom, respectively.

[8] A. Bátkai, P. Csomós and G. Nickel, perator splittings and spatial approximations for evolution equations, *J. Evol. Eqs.*, **9** (2009), 613–636.
[9] A. Bensoussan, G. Da Prato, M. Delfour and S. Mitter, *Representation and Control of Infinite Dimensional Systems*, Birkhauser, 1993
[10] P. Benner, P. Ezzatti, H. Mena, E. S. Quintana-Ortí and A. Remón, Solving matrix equations on multi-core and many-core architectures, *Algorithms*, **6** (2013), 857–870.
[11] P. Benner and H. Mena, Numerical solution of the infinite-dimensional LQR-problem and the associated differential Riccati equations, *Journal of Numerical Mathematics*, (2016), in press.
[12] P. Benner and H. Mena, Rosenbrock methods for solving differential Riccati equations, *IEEE Transactions on Automatic Control*, **58** (2013), 2950–2957.
[13] P. Benner and J. Saak, Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey, *GAMM Mitteilungen*, **36**, (2013), 32–52.
[14] P. Csomós and J. Winckler. A semigroup proof for the well-posedness of the linearised shallow water equations, *J. Anal. Math.*, **43** (2017), 445–459.
[15] G. Da Prato, Direct solution of a Riccati equation arising in stochastic control theory, *Appl. Math. Optim.*, **11** (1984), 191–208.
[16] G. Da Prato, P. Kunstmann, I. Lasiecka, A. Lunardi, R. Schnaubelt and L. Weis, *Functional Analytic Methods for Evolution Equations*, Springer-Verlag, Berlin, 2004.

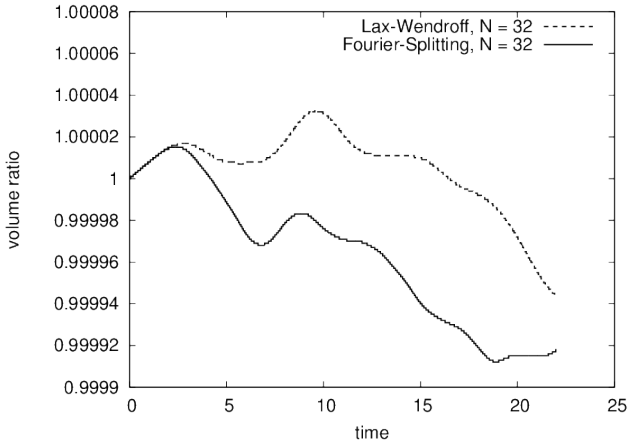

FIGURE 26. Volume ratio $\mathcal{V}_4$ for two-dimensional shallow water equations (12) for the sink-like control ($B = \mathrm{I}_{\Gamma_1}$).

[17] K. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics, Springer-Verlag, New York, 2000.

[18] F. Flandoli, Direct solution of a Riccati equation arising in a stochastic control problem with control and observation on the boundary, *Appl. Math. Optim.*, **14** (1986), 107–129.

[19] C. Hafizoglu, I. Lasiecka, T. Levajković, H. Mena and A. Tuffaha, The stochastic linear quadratic problem with singular estimates, *SIAM J. Control Optim.*, **55** (2017), 595-626

[20] E. Hansen and A. Ostermann, Exponential splitting for unbounded operators, *Math. Comput.*, **78** (2009), 1485–1496.

[21] A. Ichikawa, Dynamic programming approach to stochastic evolution equation, *SIAM J. Control. Optim.* **17**, (1979) 152–174.

[22] A. Ichikawa and H. Katayama, Remarks on the time-varying H_∞ Riccati equations, *Sys. Cont. Lett.*, **37** (1999), 335–345.

[23] O. Iftime and M. Opmeer, A representation of all bounded selfadjoint solutions of the algebraic Riccati equation for systems with an unbounded observation operator, *Proceedings of the 43rd IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, December 14–07 (2004), 2865–2870.

[24] K. Ito and F. Kappel, *Evolution Equations and Approximations*, World Scientific, Singapore, 2002.

[25] T. Jahnke and Ch. Lubich, Error bounds for exponential operator splittings, *BIT*, **40** (2000), 735–744.

[26] D. Kleinman, On an iterative technique for Riccati equation computations, *IEEE Trans. Automatic Control*, **AC-13** (1968), 114–115.

[27] A. Kofler, H. Mena and A. Ostermann, Splitting Methods for Stochastic Partial Differential Equations, preprint

[28] V. Mehrmann, *The Autonomous Linear Quadratic Control Problem*, Springer-Verlag, Berlin, 1991.

[29] N. Lang, H. Mena and J. Saak, On the benefits of the LDL factorization for large-scale differential matrix equation solvers, *Linear Algebra and its Applications*, **480** (2015), 44–71.

[30] I. Lasiecka, Optimal control problems and Riccati equations for systems with unbounded controls and partially analytic generators: applications to boundary and point control problems, in *Functional Analytic Methods for Evolution Equations* (eds. M. Iannelli, R. Nagel, S. Piazzera), Lecture Notes in Mathematics, **1855**, Springer, Berlin, Heidelberg (2004), 313–369.

[31] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories II. Abstract Hyperbolic-like Systems over a Finite Time Horizon*, Cambridge University Press, Cambridge, UK, 2000.

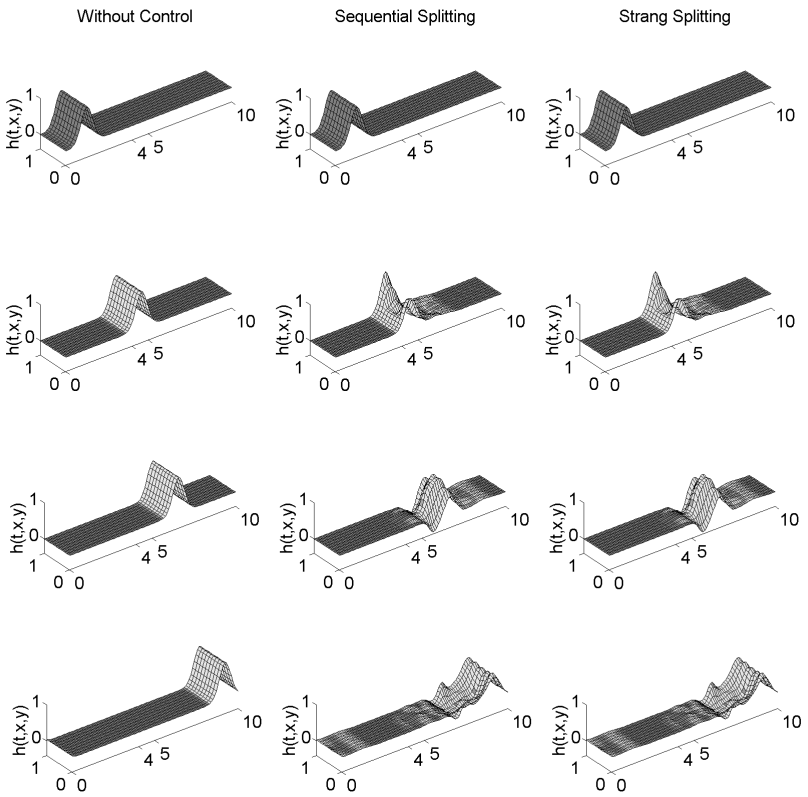

FIGURE 27. Two-dimensional linearized shallow water equations with control matrix $B = \Gamma_{2r}$ representing a sink. Comparison between no control (left column), sequential splitting (column in the middle) and Strang splitting (right column).

[32] I. Lasiecka and A. Tuffaha, Riccati equations for the Bolza problem arising in boundary/point control problems governed by C_0-semigroups satisfying a singular estimate, *J. Optim. Theory Appl.*, **136** (2008), 229–246.

[33] T. Levajković and H. Mena, On deterministic and stochastic linear quadratic control problem, in *Current Trends in Analysis and Its Applications. Trends in Mathematics.* (eds. V. Mityushev, M. Ruzhansky), Birkhuser, Cham (2015), 315–322.

[34] T. Levajković, H. Mena and A. Tuffaha, The Stochastic Linear Quadratic Control Problem: A Chaos Expansion Approach, *Evolution Equations and Control Theory*, **5** (2016), 105–134.

[35] T. Levajković, H. Mena and A. Tuffaha, A numerical approximation framework for the stochastic linear quadratic regulator problem on Hilbert spaces, *Applied Mathematics and Optimization*, **75** (2017), 499–523.

[36] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer-Verlag, New York, 1987.

[37] I. Petersen, V. Ugrinovskii and A. Savkin, *Robust control design using H^∞ methods*, Springer-Verlag, London, 2000.

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: csomos@cs.elte.hu
E-mail address: mena@yachaytech.edu.ec, hermann.mena@uibk.ac.at