Adjoint-based optimal control using meshfree discretizations

Jan Marburger [a], Nicole Marheineke [b,c,*], René Pinnau [b]

[a] *Fraunhofer ITWM Kaiserslautern, Germany*
[b] *TU Kaiserslautern, FB Mathematik, Germany*
[c] *FAU Erlangen-Nuernberg, Department Mathematik, Germany*

ARTICLE INFO

ABSTRACT

The paper at hand presents a combination of optimal control approaches for PDEs with meshless discretizations. Applying a classical Lagrangian type particle method to optimization problems with hyperbolic constraints, several adjoint-based strategies differing in the sequential order of optimization and discretization of the Lagrangian or Eulerian problem formulation are proposed and compared. The numerical results confirm the theoretically predicted independence principle of the optimization approaches and show the expected convergence behavior. Moreover, they exemplify the superiority of meshless methods over the conventional mesh-based approaches for the numerical handling and optimization of problems with time-dependent geometries and freely moving boundaries.

1. Introduction

An easy numerical handling of time-dependent problems with complicated geometries, freely moving boundaries and interfaces, or oscillating solutions is of great importance for many applications, e.g., in fluid dynamics (free surface and multiphase flows, fluid–structure interactions [1–3]), failure mechanics (crack growth and propagation [4]), magnetohydrodynamics (accretion disks, jets and cloud simulation [5]), biophysics and biochemistry. Appropriate discretizations, so-called meshless methods, have been developed during the last few decades to meet these challenging demands and to relieve the burden of remeshing and successive mesh generation being faced by the conventional mesh-based methods [6–8]. The prearranged mesh is an artificial constraint to ensure compatibility of the mesh-based interpolant schemes, that often conflicts with the real physical conditions of the continuum model. Then, remeshing becomes inevitable, which not only requires extreme amounts of time and storage but also is a source for numerical errors and hence the gradual loss of computational accuracy. Meshless methods overcome this problem as they do not need any mesh structure and hence it is possible to add and remove points freely, i.e. without computational effort, for adaptation reasons. Apart from this advantage, meshless methods also lead to fundamentally better approximations as regards aspects such as smoothness, nonlocal interpolation character, flexible connectivity, refinement and enrichment procedures [6]. The common idea of meshless methods is the discretization of the domain of interest by a finite set of independent, randomly distributed particles moving with a characteristic velocity of the problem. The location and distribution of the particles then account for the time-dependent description of the geometry, data and solution. Thereby, the global solution is linearly superposed from

---

* Corresponding author at: TU Kaiserslautern, FB Mathematik, 67653 Kaiserslautern, Germany.
 E-mail addresses: jan.marburger@itwm.fhg.de (J. Marburger), marheineke@mathematik.uni-kl.de, nicole.marheineke@am.uni-erlangen.de (N. Marheineke), pinnau@mathematik.uni-kl.de (R. Pinnau).

the local information carried by the particles. In classical particle methods [9,10], the respective weight functions are Dirac distributions which yield solutions in a distributional sense. In meshless Galerkin or partition of unity methods [8,11], in contrast, compact patches or volumes are attached to every particle, whose union forms an open covering of the domain. On the patches, local weight functions are constructed via data fitting. Representing the influence of the particles in their respective neighborhoods, these functions provide a basis for a Galerkin or collocation discretization process. Depending on the fitting ansatz, the methods generalize finite element, volume and difference approaches.

Since there is a tremendous and even increasing need of optimization and optimal control in the aforementioned applications [12,13], the goal of this paper is a first combination of conventional optimization strategies for PDEs with meshless discretizations. We focus on the application of classical Lagrangian type particle methods to balance laws, in particular hyperbolic equations, and study the optimization results formally obtained from different strategies using an adjoint approach [14]. The strategies investigated tackle hereby the question of the numerical effect and computational effort when varying the sequential order of optimization and discretization in Lagrangian as well as Eulerian problem formulations.

This paper is organized in the following manner. After a brief introduction to particle methods in Section 2, we present several adjoint-based optimization strategies differing in the sequential order of optimization and discretization (first optimization, then discretization versus first optimization, then discretization) for tracking type problems with hyperbolic partial differential constraints in Section 3. In particular, we formally derive the corresponding adjoint systems for a Lagrangian problem formulation, discuss their differences and compare them with the optimization results for the associated Eulerian formulation in Section 4. In Section 5 we apply the strategies as examples to one-dimensional problems, i.e., linear convection and nonlinear Burger equations. The numerical results confirm the theoretically predicted independence principle of the optimization approaches and show the expected convergence behavior. Moreover, they clearly emphasize the superiority of meshless methods over mesh-based ones in the case of time-dependent geometry changes and huge deformations.

2. The Lagrangian type particle method

Particle methods are special types of meshless methods. They originate from the numerical treatment of physics applications like Boltzmann equations [9,10] (molecular dynamics, direct simulation Monte Carlo; see [6] and the references within) where the quantities of the conservation problem are approximated on a set of randomly distributed particles. Classically, they have been truly Lagrangian methods, i.e. the particles move with the convective velocity of the problem. This ansatz has been extended to mixed Lagrangian–Eulerian considerations such that the dynamics of the particles might also be prescribed by another characteristic velocity, e.g., the deformation velocity given by a free surface motion. Thus, the approach provides an easy numerical treatment of freely moving boundaries and interfaces, complicated geometries and time dependencies, but is also restricted to (nonstationary) conservation problems. Stationary or elliptic problems might be handled by other meshless variants based on data fitting, e.g., Galerkin and partition of unity methods, such as smoothed particle hydrodynamics (SPH) [15,16], moving least square (MLS) [17,18] and moving least square particle hydrodynamics (MLSPH) [19], the reproducing kernel particle method (RKPM) [20], the partition of unity finite element method (PUFEM) [21,22] and the particle partition of unity method [23] or finite point-set method (FPM) [3] (for more details about meshless (particle) methods see [6–8]).

The approximation idea common to meshless methods can be summarized as follows. Let the quantity of interest $y : \Omega \times [0, T] \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$, be approximated by N particles carrying the information of their position and the specific local quantity $(x_i, y_i) : [0, T] \to \Omega \times \mathbb{R}^m$, $i = 1, \ldots, N$; then

$$y(x, t) \simeq y_h(x, t) := \sum_{i=1}^{N} y_i(t) S_i(x, t; \vec{x}(t)), \quad \vec{x}(t) := \{x_i(t)\}_{i=1,\ldots,N}.$$

The spatial interpolation functions S_i represent the influence of the respective particles in their local neighborhood. In particle methods they are given by Dirac distributions $S_i(x, t; \vec{x}(t)) = \delta(x - x_i(t))$, whereas in meshless Galerkin or partition of unity methods they are constructed via data fitting, satisfying $\sum_{i=1}^{N} S_i = 1$. Investigating particle methods on conservation laws, Raviart derived convergence results in appropriate function spaces (for details see [24]).

Following Raviart's approach, we focus on the discretization via a classical Lagrangian type particle method in this paper. This discretization can be interpreted as a method of characteristics for hyperbolic equations which requires a specific coordinate transformation between Lagrangian (material) and Eulerian (spatial) coordinates. A coordinate transformation or motion Φ is given by an invertible, regular mapping

$$\Phi : \mathbb{R}^n \times (0, T) \to \mathbb{R}^n,$$

initialized with $\Phi(X, 0) = X$ at time $t = 0$. We denote its derivative with respect to X by $\mathcal{J} := \partial_X \Phi$ and assume $\mathcal{J} > 0$. For every function $\omega : \mathbb{R} \times (0, T) \to \mathbb{R}$, we define the associated transformed function by

$$\hat{\omega}(X, t) := \omega(\Phi(X, t), t).$$

Then, the following relations hold:

$$\partial_X \hat{\omega}(X, t) = \partial_x \omega(\Phi(X, t), t) \mathcal{J}(X, t),$$
$$\partial_t \hat{\omega}(X, t) = \partial_X \hat{\omega}(X, t)(\mathcal{J}(X, t))^{-1} \partial_t \Phi(X, t) + \partial_t \omega(\Phi(X, t), t).$$

3. Optimization strategies

In this section we study the two main adjoint-based approaches for the optimization of conservation problems using particle methods. In the first approach, we apply the adjoint-based optimization formalism directly to the continuous (Eulerian or Lagrangian) formulation of the optimization problem and derive the optimality system to be discretized via a particle method. In the second approach, we proceed from the discrete Lagrange functional to deduce the respective adjoint system. For convenience, we restrict the following investigations to scalar-valued one-dimensional hyperbolic PDEs in order to embed the particle method into the method of characteristics using the convective velocity as particle motion. Certainly, they can be directly extended to vector-valued, higher dimensional as well as parabolic equations.

We consider the following minimization problem:

$$\min J(y, u) \quad \text{subject to } e(y, u) = 0 \tag{1}$$

over $(y, u) \in Y \times U$, where the state space Y is a reflexive Banach space and the control space U is a Hilbert space such that $J : Y \times U \to \mathbb{R}$ is continuous and differentiable. The constraint is given by a mapping $e : Y \times U \to Z^*$, where the space of Lagrange multipliers Z is a reflexive Banach space. Further, we assume the Fréchet differentiability of e.

Throughout this paper we choose $U = L^2(0, T)$ and consider the minimization of the cost functional

$$J(y, u) := \frac{1}{2} \int_{\mathbb{R}} |y(x, T) - y_d(x)|^2 \mathrm{d}x + \frac{\alpha}{2} \int_0^T |u(t)|^2 \mathrm{d}t$$

with the hyperbolic constraint

$$\partial_t y + \frac{\mathrm{d}}{\mathrm{d}x} f(y, x, t) = b(x)u(t) \quad \text{in } \mathbb{R} \times (0, T) \tag{2}$$
$$y(x, 0) = y_0(x) \quad \text{in } \mathbb{R},$$

that is,

$$e(y, u)p := \int_0^T \int_{\mathbb{R}} \partial_t y p - f(y, x, t)\partial_x p - b(x)u(t)p \, \mathrm{d}x \, \mathrm{d}t + \int_{\mathbb{R}} (y(0) - y_0)p(0)\mathrm{d}x.$$

Here y_d denotes the desired state at final time T, b a spatial localization function, $\alpha > 0$ a regularization parameter in the tracking type problem and $f : \mathbb{R} \times \mathbb{R} \times (0, T) \to \mathbb{R}$ the flux function. For more details of the identification of the initial condition see e.g. [25].

Assumption 3.1. In the following we assume:

(a) The flux function f and localization function b is supposed to be sufficiently regular.
(b) The constraint $e(y, u) = 0$ has for any given $u \in U$ a unique solution $y \in Y$.
(c) The linearization $e_y(y, u)$ is surjective.

Remark 3.2. Examples for which the above assumptions are satisfied can be found for instance in [26].

We transform the constraint (2) into a Lagrangian representation by introducing a transformation $\Phi \in V$, where V is a reflexive Banach space, given by

$$\partial_t \Phi(X, t) = \partial_y f(y(\Phi(X, t), t), \Phi(X, t), t) \quad \text{in } \mathbb{R} \times (0, T)$$
$$\Phi(X, 0) = X \quad \text{in } \mathbb{R}.$$

Since Eq. (2) yields

$$\partial_t y + \partial_y f(y, x, t)\partial_x y + \partial_x f(y, x, t) = b(x)u(t)$$

we obtain

$$\partial_t \hat{y} + \partial_x f(\hat{y}, \Phi, t) = b(\Phi)u(t)$$

by setting $\hat{y}(X, t) := y(\Phi(X, t), t)$. Note that also $\hat{y} \in Y$. Hence, the transformed state equation reads

$$\partial_t \hat{y} = b(\Phi)u(t) - \partial_x f(\hat{y}, \Phi, t) \quad \text{in } \mathbb{R} \times (0, T)$$
$$\hat{y}(X, 0) = y_0(X) \quad \text{in } \mathbb{R}$$
$$\partial_t \Phi = \partial_y f(\hat{y}, \Phi, t) \quad \text{in } \mathbb{R} \times (0, T) \tag{3}$$
$$\Phi(X, 0) = X \quad \text{in } \mathbb{R}.$$

Remark 3.3. The above system is just an application of the method of characteristics to (2). Hence, if (2) has a unique solution, then also (3) is uniquely solvable. For details see [27].

This system is discretized by a particle method. For this, we randomly distribute N particles at the positions $\{X_i\}_{i=1,\dots,N}$. Then we obtain the corresponding discrete counterpart to (3) for the particle trajectories $\vec{\Phi} \in V_h$ and the local quantities carried, $\vec{y} \in Y_h$:

$$\partial_t y_i(t) = b(\Phi_i(t))u(t) - \partial_x f(y_i(t), \Phi_i(t), t)$$
$$y_i(0) = y_0(X_i)$$
$$\partial_t \Phi_i(t) = \partial_y f(y_i(t), \Phi_i(t), t)$$
$$\Phi_i(0) = X_i$$

(4)

and hence we obtain a system of $2N$ ODEs. The spaces Y_h and V_h denote the spatially discrete counterparts to Y and V, respectively. Again, Y_h and V_h are supposed to be reflexive Banach spaces.

Remark 3.4. For sufficiently high number of particles also the discrete forward system (4) has a unique solution if the continuous system (2) has a unique solution (cf. [27]).

3.1. The continuous Euler approach

Dealing with a constrained optimization problem [14], we formulate the Karush–Kuhn–Tucker system with the help of the Lagrangian functional associated with the minimization problem (1)

$$L(y, u, p) := J(y, u) + e(y, u)p$$

(5)

where $p \in Z$ denotes the adjoint variable. According to the Lagrange multiplier theorem [14], the variations of the Lagrange functional with respect to the state and control variable vanish at a critical point, which yields the first-order optimality system

$$\partial_y L(y, u, p)\varphi^y = 0 \qquad \partial_p L(y, u, p)\varphi^p = 0 \qquad \partial_u L(y, u, p)\varphi^u = 0$$

for all test functions $\varphi^y \in Y$, $\varphi^p \in Z$ and $\varphi^u \in U$.

Remark 3.5. The existence of a Lagrange multiplier p is guaranteed by Assumption 3.1(c).

The adjoint (or backward) equation, obtained using $\partial_y L(y, u, p)$, reads

$$\partial_t p + \partial_y f(y, x, t)\partial_x p = 0 \quad \text{in } \mathbb{R}$$
$$p(x, T) = -(y(x, T) - y_d(x)) \quad \text{in } \mathbb{R}.$$

This system is also transformed into a Lagrangian representation; in particular,

$$\partial_t \hat{p} = 0 \quad \text{in } \mathbb{R} \times (0, T)$$
$$\hat{p}(T) = -(y(\bar{\Phi}(T), T) - y_d(\bar{\Phi}(T))) \quad \text{in } \mathbb{R}$$
$$\partial_t \bar{\Phi} = \partial_y f(y(\bar{\Phi}), \bar{\Phi}, t) \quad \text{in } \mathbb{R} \times (0, T)$$
$$\bar{\Phi}(X, 0) = X \quad \text{in } \mathbb{R}.$$

Obviously, $\bar{\Phi} = \Phi$ as the initial condition and the right hand side of the ordinary differential equation are equal. Hence, the solution to the adjoint equation is given by

$$\hat{p}(X, t) = -(\hat{y}(X, T) - y_d(\Phi(X, T))).$$

(6)

Using the same point set $\{X_i\}_{i=1,\dots,N}$ as for the discretization of the forward system (4) yields the discrete adjoint system

$$p_i(t) = -(y_i(T) - y_d(\Phi_i(T)))$$

(7)

where y_i and Φ_i are obtained by solving (4).

The gradient of the reduced cost functional $\tilde{J}(u) := J(y(u), u)$ is obtained by identifying

$$\langle D\tilde{J}(u), \psi^u \rangle_{U^*, U} = \langle \partial_u L(y, u, p), \psi^u \rangle_{U^*, U} = \int_0^T \left(\alpha u(t) - \int_{\mathbb{R}} b(x)p(x, t)\mathrm{d}x \right) \psi^u(t)\mathrm{d}t.$$

Hence, the gradient reads

$$\nabla \tilde{J}(u) = \alpha u(t) - \int_{\mathbb{R}} b(x)p(x, t)\mathrm{d}x = \alpha u(t) - \int_{\mathbb{R}} b(\Phi(X, t))\hat{p}(X, t)\mathcal{J}(X, t)\mathrm{d}X$$

(8)

or, using the discrete values of (7),

$$\nabla \tilde{J}^h(u) = \alpha u(t) - \sum_{i=1}^{N} b(\Phi_i(t))p_i(t)h_i(\vec{\Phi}(t)),$$

(9)

where the $h_i(\vec{\Phi}(t))$ denote quadrature weights. In particular we choose

$$h_i(\vec{\Phi}(t)) = \frac{1}{2}(\Phi_{i+1}(t) - \Phi_{i-1}(t)).$$

Note that we have

$$\mathcal{J}(X_i, t) = \partial_X \Phi(X_i, t) \simeq h_i(\vec{\Phi}) h_i(\vec{X})^{-1}.$$

3.2. The continuous Lagrangian approach

This approach is based on a reformulation of the optimization problem (1) itself. In particular, we solve

$$\min J_l(\hat{y}, \Phi, u) \quad \text{subject to } e_l(\hat{y}, \Phi, u, \hat{p}, z) = 0$$

which are obtained by applying the transformation Φ to the cost functional J and the constraint (2).

First, we transform the cost functional $J(y, u)$ by applying the coordinate transformation Φ, i.e.

$$J_l(\hat{y}, \Phi, u) = \frac{1}{2} \int_{\mathbb{R}} |\hat{y}(X, T) - y_d(\Phi(X, T))|^2 \mathcal{J}(X, T) \mathrm{d}X + \frac{\alpha}{2} \int_0^T |u(t)|^2 \mathrm{d}t.$$

Then we define the transformed constraint function $e_l : \hat{Y} \times U \to \hat{Z}^*$, where $\hat{Y} := Y \times V$ and $\hat{Z}$ is the space of Lagrange multipliers, by

$$
e_l(\hat{y}, \Phi, u)(\hat{p}, z) := \int_0^T \int_{\mathbb{R}} \partial_t \hat{y}\hat{p} + \partial_x f(\hat{y}, \Phi, t)\hat{p} - b(\Phi)u(t)\hat{p} + \Phi z - \partial_y f(\hat{y}, \Phi, t)z \, \mathrm{d}X \, \mathrm{d}t
$$
$$
+ \int_{\mathbb{R}} (\hat{y}(0) - y_0)\hat{p}(0) + (\Phi(0) - X)z(0)\mathrm{d}X.
$$

Assumption 3.6. In the following we assume:

(a) The constraint $e_l(\hat{y}, \Phi, u) = 0$ has for each $u \in U$ a unique solution $(\hat{y}, \Phi) \in \hat{Y}$ (cf. Remark 3.3).
(b) The linearization $\partial_y e_l(\hat{y}, \Phi, u)$ is surjective.

Finally, we state the transformed Lagrange functional L_l:

$$L_l(\hat{y}, \Phi, u, \hat{p}, z) = J_l(\hat{y}, \Phi, u) + e_l(\hat{y}, \Phi, u)(\hat{p}, z)$$

and, by analogy to the previous section, the first-order optimality system:

$$\partial_{\hat{y}} L_l(\hat{y}, \Phi, u, \hat{p}, z)\psi^{\hat{y}} = 0 \qquad \partial_{\Phi} L_l(\hat{y}, \Phi, u, \hat{p}, z)\psi^{\Phi} = 0 \qquad \partial_u L_l(\hat{y}, \Phi, u, \hat{p}, z)\psi^u = 0$$
$$\partial_{\hat{p}} L_l(\hat{y}, \Phi, u, \hat{p}, z)\psi^{\hat{p}} = 0 \qquad \partial_z L_l(\hat{y}, \Phi, u, \hat{p}, z)\psi^z = 0$$

for all test functions $\psi^{\hat{y}} \in Y$, $\psi^{\Phi} \in V$, $\psi^u \in U$ and $(\psi^{\hat{p}}, \psi^z) \in \hat{Z}$.

Remark 3.7. The existence of Lagrange multipliers $\hat{p}$ and z is guaranteed by Assumption 3.6(b).

The adjoint system, obtained using $\partial_{\hat{y}} L_l(\hat{y}, \Phi, u, \hat{p}, z) = 0$ and $\partial_{\Phi} L_l(\hat{y}, \Phi, u, \hat{p}, z) = 0$, reads

$$
\begin{aligned}
&\partial_t \hat{p} = \partial_{xy} f(\hat{y}, \Phi, t)\hat{p} - \partial_{yy} f(\hat{y}, \Phi, t)z \\
&\hat{p}(T) = -(\hat{y}(T) - y_d(\Phi(T)))\mathcal{J}(T) \\
&\partial_t z = \partial_{xx} f(\hat{y}, \Phi, t)\hat{p} - \partial_{xy} f(\hat{y}, \Phi, t)z - \partial_x b(\Phi)u\hat{p} \\
&z(T) = (\hat{y}(T) - y_d(\Phi(T)))\partial_X \hat{y}(T).
\end{aligned}
\tag{10}
$$

Of course, the adjoint system has the same trajectories as the forward system. Therefore, we discretize it by using the same particle set as for the forward system:

$$
\begin{aligned}
&\partial_t p_i(t) = \partial_{xy} f(y_i(t), \Phi_i(t), t)p_i(t) - \partial_{yy} f(y_i(t), \Phi_i(t), t)z_i(t) \\
&p_i(T) = -(y_i(T) - y_d(\Phi_i(T)))\mathcal{J}(X_i, T) \\
&\partial_t z_i(t) = \partial_{xx} f(y_i(t), \Phi_i(t), t)p_i(t) - \partial_{xy} f(y_i(t), \Phi_i(t), t)z_i(t) - \partial_x b(\Phi_i(t))u(t)p_i(t) \\
&z_i(T) = (y_i(T) - y_d(\Phi_i(T)))(\partial_X \hat{y}(T))_i.
\end{aligned}
\tag{11}
$$

The gradient of the reduced cost functional $\tilde{J}_l(u) := J_l(\hat{y}(u), \Phi(u), u)$ is obtained by identifying

$$\langle D\tilde{J}_l(u), \psi^u \rangle_{U*,U} = \langle \partial_u L_l(\hat{y}, \Phi, u, \hat{p}, z), \psi^u \rangle_{U*,U}$$
$$= \int_0^T \left(\alpha u(t) - \int_{\mathbb{R}} b(\Phi(X, t))\hat{p}(X, t)dX \right) \psi^u(t)dt.$$

Hence, the gradient reads

$$\nabla\tilde{J}_l(u) = \alpha u(t) - \int_{\mathbb{R}} b(\Phi(X, t))\hat{p}(X, t)dX \tag{12}$$

or, using the discrete values of (11),

$$\nabla\tilde{J}_l^h(u) = \alpha u(t) - \sum_{i=1}^N b(\Phi_i(t))p_i(t)h_i(\vec{X}) \tag{13}$$

where the $h_i(\vec{X})$ denote quadrature weights as defined in the previous section.

3.3. The discrete Lagrangian approach

Here we use the approach of "first discretize then optimize", that is, we perform the optimization on the discretized system (4). Further, we also discretize the cost functional

$$J_h(\vec{y}, \vec{\Phi}, u) = \frac{1}{2} \sum_{i=1}^N |y_i(T) - y_d(\Phi_i(T))|^2 h_i(\vec{\Phi}) + \frac{\alpha}{2} \int_0^T |u(t)|^2 dt.$$

The associated discrete constraint function $e_h : \hat{Y}_h \times U \to \hat{Z}_h^*$, where $\hat{Y}_h := Y_h \times V_h$ and $\hat{Z}_h$ is the space of Lagrange multipliers, reads

$$e_h(\vec{y}, \vec{\Phi}, u)(\vec{p}, \vec{z}) = \sum_{j=1}^N \left[(y_j(0) - y_0(X_j))p_j(0) + (\Phi_j(0) - X_j)z_j(0) \right] h_j(\vec{X})$$
$$+ \int_0^T \sum_{j=1}^N \left[(\partial_t y_j - b(\Phi_j)u(t)p_j + \partial_x f(y_j, \Phi_j, t))p_j + (\partial_t \Phi_j - \partial_y f(y_j, \Phi_j, t))z_j \right] h_j(\vec{X})dt.$$

Assumption 3.8. In the following we assume:

(a) The constraint $e_h(\vec{y}, \vec{\Phi}, u) = 0$ has for each $u \in U$ a unique solution $(\vec{y}, \vec{\Phi}) \in \hat{Y}$ (cf. Remark 3.4).
(b) The Fréchet derivative of $e_h(\vec{y}, \vec{\Phi}, u)$ is surjective.

The discrete Lagrange functional is given by

$$L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z}) = J_h(\vec{y}, \vec{\Phi}, u) + e_h(\vec{y}, \vec{\Phi}, u)(\vec{p}, \vec{z})$$

and, by analogy to the previous section, the first-order optimality system is

$$\partial_{\vec{y}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z})\psi^y = 0 \qquad \partial_{\vec{\Phi}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z})\psi^\Phi = 0 \qquad \partial_u L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z})\psi^u = 0$$
$$\partial_{\vec{p}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z})\psi^p = 0 \qquad \partial_{\vec{z}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z})\psi^z = 0$$

for all test functions $\psi^y \in Y_h$, $\psi^\Phi \in V_h$, $\psi^u \in U$ and $(\psi^p, \psi^z) \in \hat{Z}_h$.

Remark 3.9. The existence of Lagrange multipliers $\vec{p}$ and $\vec{z}$ is guaranteed by Assumption 3.8(b).

The adjoint system, obtained using $\partial_{\vec{y}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z}) = 0$ and $\partial_{\vec{\Phi}} L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z}) = 0$, reads

$$\partial_t p_i = \partial_{xy} f(y_i, \Phi_i, t)p_i - \partial_{yy} f(y_i, \Phi_i, t)z_i$$
$$p_i(T) = -(y_i(T) - y_d(\Phi_i(T)))h_i(\vec{\Phi}(T))h_i(\vec{X})^{-1}$$
$$\partial_t z_i = \partial_{xx} f(y_i, \Phi_i, t)p_i - \partial_{xy} f(y_i, \Phi_i, t)z_i - \partial_x b(\Phi_i)u(t)p_i$$
$$z_i(T) = (y_i(T) - y_d(\Phi_i(T)))\partial_x y_d(\Phi_i(T))h_i(\vec{\Phi}(T))h_i(\vec{X})^{-1} \tag{14}$$
$$+ \frac{1}{4}\left((y_{i+1}(T) - y_d(\Phi_{i+1}(T)))^2 - (y_{i-1}(T) - y_d(\Phi_{i-1}(T)))^2 \right)h_i(\vec{X})^{-1}.$$

The gradient of the reduced cost functional $\tilde{J}_h(u) := J_h(\vec{y}(u), \vec{\Phi}(u), u)$ is obtained by identifying

$$\langle D\tilde{J}_h(u), \psi^u \rangle_{U^*,U} = \langle \partial_u L_h(\vec{y}, \vec{\Phi}, u, \vec{p}, \vec{z}), \psi^u \rangle_{U^*,U}$$

$$= \int_0^T \left(\alpha u(t) - \sum_{i=1}^N b(\Phi_i(t)) p_i(t) h_i(\vec{X}) \right) \psi^u(t) \mathrm{d}t.$$

Hence, the gradient reads

$$\nabla \tilde{J}_h(u) = \alpha u(t) - \sum_{i=1}^N b(\Phi_i(t)) p_i(t) h_i(\vec{X}).$$

4. Comparison of the three approaches

In the following we discuss the relation of the three optimization approaches and show that they are either equivalent or differ by at most an error of the order of the discretization error.

4.1. Comparison of the Lagrangian optimization approaches

The varying sequential order of optimization and discretization does not affect the forward system. It is the same in the two approaches, in particular for (4). Hence, starting with the same particle distribution $\{X_i\}_{i=1,\dots,N}$, the particle trajectories Φ_i and quantities carried y_i are identical. Also the forms of the gradients and the right hand sides of the evolution equations for the adjoints p_i and z_i in (11) and (14) coincide. The only apparent differences are in the initial conditions $p_i(T)$ and $z_i(T)$. To estimate their effect, we reintroduce the dependence on X of the particle quantities, i.e., we reinterpret $\Phi_i(t) = \Phi(X_i, t)$ and $y_i(t) = \hat{y}(X_i, t)$. Moreover,

$$\mathcal{J}(X_i, t) = (\partial_X \Phi)_i = \frac{\Phi_{i+1} - \Phi_{i-1}}{X_{i+1} - X_{i-1}} + \mathcal{O}(h^2) = h_i(\vec{\Phi}) h_i(\vec{X})^{-1} + \mathcal{O}(h^2)$$

holds for $h = \max_i h_i(\vec{X})$. Then, the initial conditions $p_i(T)$ and $z_i(T)$ of (14) can be rewritten as

$$p_i(T) = -(y_i(T) - y_d(\Phi_i(T))) J(X_i, t) + \mathcal{O}(h^2)$$

$$z_i(T) = (y_i(T) - y_d(\Phi_i(T))) \partial_x y_d(\Phi_i(T)) \mathcal{J}_i(T) + \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}X} \left((\hat{y}(X_i, T) - y_d(\Phi(X_i, T)))^2 \right) + \mathcal{O}(h^2)$$

$$= (y_i(T) - y_d(\Phi_i(T))) (\partial_X \hat{y}(T))_i + \mathcal{O}(h^2).$$

Hence, (14) just differs from (11) in a discretization error of order $\mathcal{O}(h^2)$, that enters the system through the approximation of the derivatives. Note that h refers to the initial particle distribution and is thus independent of time. Consequently, the two strategies yield consistent discretizations of the continuous problem, as expected.

4.2. Comparison of the Lagrangian ansatz and the Eulerian ansatz

The Eulerian and Lagrangian formulations are transformable into each other using the coordinate transformation Φ. Hence, applying the same Lagrangian particle method, the respective discrete forward systems are identical. The backward systems (7) and (11) as well as the corresponding gradients (9) and (13), in contrast, seem to differ at first glance, but there also exists a transformation between the adjoints as we will show in the following.

We consider the continuous formulation of the adjoints and gradients as they only differ by $\mathcal{O}(h^2)$ from the discrete formulations. In particular, we compare the continuous adjoints (6) and (10) and the corresponding gradients (8) and (12), respectively.

We introduce the quantity $\tilde{p} = \hat{p} \mathcal{J}^{-1}$ and apply it to the Lagrangian approach described in Section 3.2. The gradient (12) becomes

$$\nabla J_l(u) = \alpha u(t) - \int_{\mathbb{R}} b(\Phi) \hat{p} \, \mathrm{d}X = \alpha u(t) - \int_{\mathbb{R}} b(\Phi) \tilde{p} \mathcal{J} \, \mathrm{d}X$$

which obviously coincides with (8). Rewriting the evolution equation for $\hat{p}$ of (10), we obtain

$$\partial_t \tilde{p} = \partial_{xy} f(\hat{y}, \Phi, t) \tilde{p} - \tilde{p} \mathcal{J}^{-1} \partial_t \mathcal{J} - \partial_{yy} f(\hat{y}, \Phi, t) z \mathcal{J}^{-1}$$

$$\tilde{p}(T) = -(\hat{y}(T) - y_d(\Phi(T)))$$

as we assume $\mathcal{J} \neq 0$. With the regularity of the coordinate transformation we obtain

$$\partial_t \mathcal{J} = \partial_t (\partial_X \Phi) = \partial_X (\partial_t \Phi) = \frac{\mathrm{d}}{\mathrm{d}X} \left(\partial_y f(\hat{y}, \Phi, t) \right) = \partial_{yy} f(\hat{y}, \Phi, t) \partial_X \hat{y} + \partial_{xy} f(\hat{y}, \Phi, t) \mathcal{J}.$$
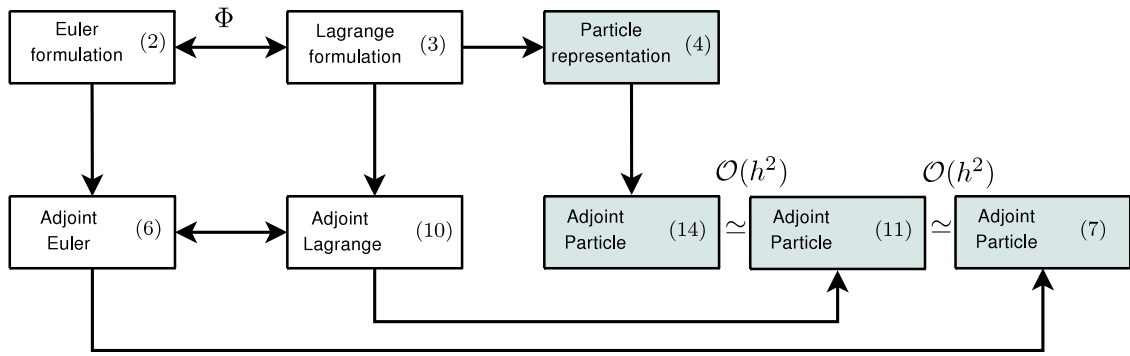
Fig. 1. Independence principle: the conformity of optimization strategies.

Inserting this relation yields

$$\partial_t \tilde{p} = -\partial_{yy}f(\hat{y}, \Phi, t)\mathcal{J}^{-1}(\tilde{p}\partial_X\hat{y} + z)$$
$$\tilde{p}(T) = -(\hat{y}(T) - y_d(\Phi(T)))$$

(15)

for an arbitrary flux function f. Obviously, it holds that $(\tilde{p}\partial_X\hat{y} + z)(T) = 0$ using (10) and (15). To check whether this is true for all $t \in [0, T]$ we derive

$$\partial_t(\tilde{p}\partial_X y + z) = \partial_t\tilde{p}\partial_X\hat{y} + \tilde{p}\partial_X(\partial_t\hat{y}) + \partial_t z$$

$$\overset{(3)}{=} \left(\partial_x b(\Phi)\mathcal{J}u(t) - \partial_{xy}f(\hat{y}, \Phi, t)\partial_X\hat{y} - \partial_{xx}f(\hat{y}, \Phi, t)\mathcal{J}\right)\tilde{p} + \partial_X\hat{y}\partial_t\tilde{p}v + \partial_t z$$

$$\overset{(10)}{=} \partial_X y\partial_t\tilde{p} - \partial_{xy}f(\hat{y}, \Phi, t)(\partial_t\tilde{p}\partial_X\hat{y} + z)$$

$$\overset{(15)}{=} -\frac{d}{dX}\left(\partial_y f(\hat{y}, \Phi, t)\right)\mathcal{J}^{-1}(\tilde{p}\partial_X y + z).$$

This results in an ordinary differential equation on the compact interval $[0, T]$, whose right hand side is continuous in t and Lipschitz continuous in $(\tilde{p}\partial_X y + z)$. Hence, it has the unique solution $(\tilde{p}\partial_X y + z) = 0$, due to the initial value. Consequently, $\partial_t\tilde{p} = 0$ which yields

$$\tilde{p} = -(\hat{y}(T) - y_d(\Phi(T))).$$

Hence, $\tilde{p} = \hat{p}\mathcal{J}^{-1}$ is the solution to the Eulerian approach and therefore the Lagrangian and Eulerian ansatz coincide, up to $\mathcal{O}(h^2)$, as expected.

The results of this section are summarized in Fig. 1 which illustrates the independence principle investigated for the optimization approaches. The Eulerian formulation (2) can be transformed into the Lagrangian formulation (3) by applying the coordinate transformation Φ. Discretizing the Lagrangian system yields the desired particle representation (4). For the optimization process, in particular taking the adjoints and gradients, it makes no difference whether we derive the adjoint from the Eulerian or from the Lagrangian system, as stated in Section 4.2. Discretizing these adjoints by a particle method yields, up to $\mathcal{O}(h^2)$, the same equations as performing the optimization from the discrete forward system (2) directly, as stated in Section 4.1.

5. Numerical results

After giving a short outline of the gradient algorithm used for the numerical implementation, we study the computational performance and correlations of the described optimization approaches in this section. Therefore, we apply the strategies, denoted by giving the section number where they were introduced, to different test cases, i.e. linear advection and the nonlinear Burgers equation.
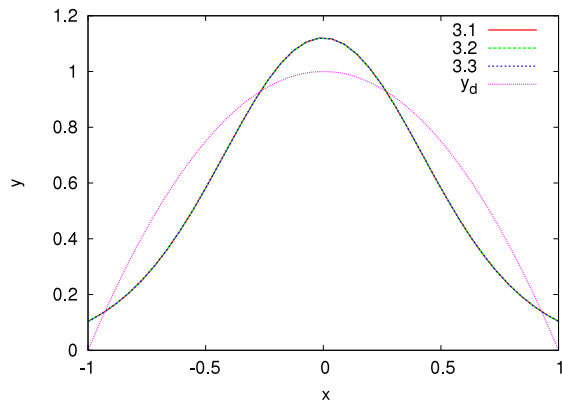
Gradient algorithm

(a) Initialize control function $u^{(0)}$, iteration counter $k = 0$, tolerance ϵ and set desired state y_d
(b) Generate particle set $\{X_i\}_{i=1,...,N}$
(c) Do
    (a) Given $u^{(k)}$ solve forward system (4) for $y_i^{(k)}$, $\Phi_i^{(k)}$
    (b) Solve backward system (7), (11) or (14) for $p_i^{(k)}$, $z_i^{(k)}$
    (c) Determine a step size parameter $\tau > 0$, e.g., by Armijo's rule
    (d) Set $u^{(k+1)} = u^{(k)} - \tau \nabla\tilde{J}_h(u^{(k)})$
    while $\|\nabla J(u^{(k)})\|/\|\nabla J(u^{(0)})\| \leq \epsilon$.
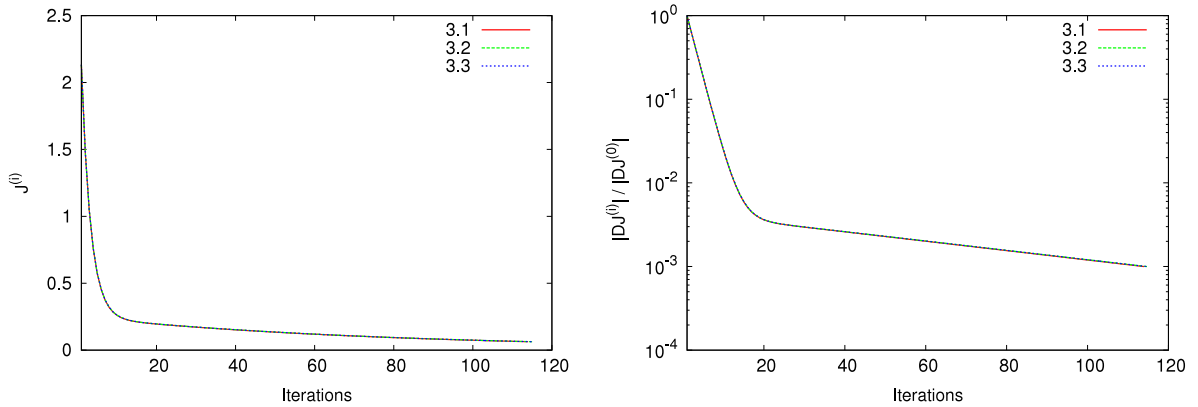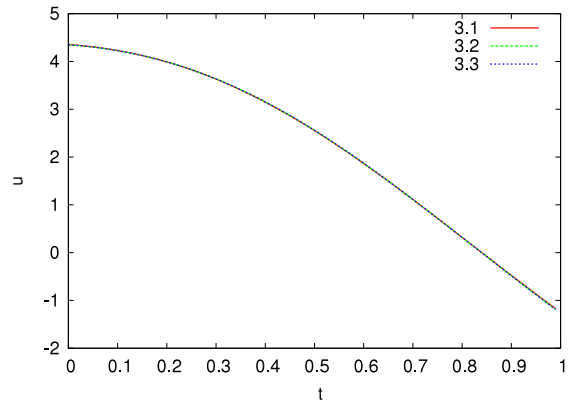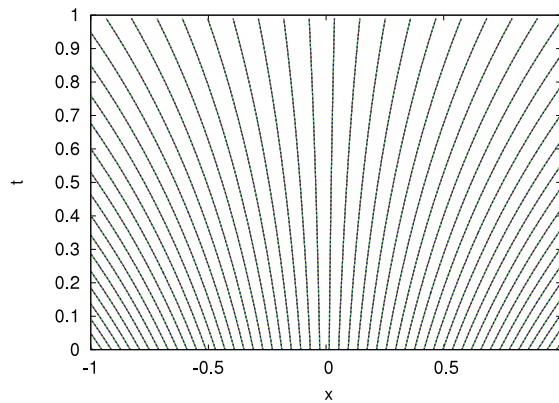
(a) Cost functional.

(b) Norm of gradient (relative).
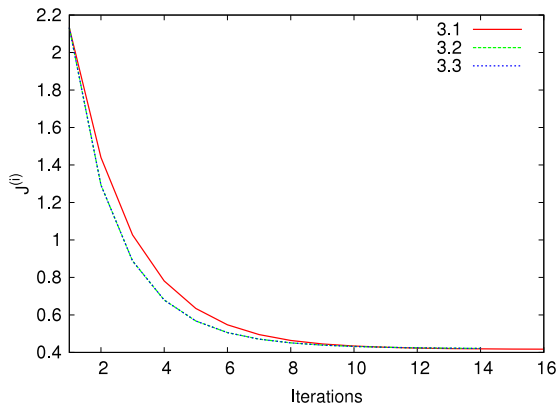
(c) State at final time T.

(d) Control function.

(e) Trajectories.

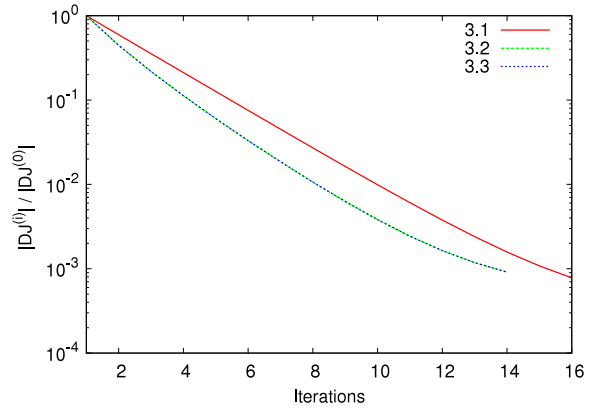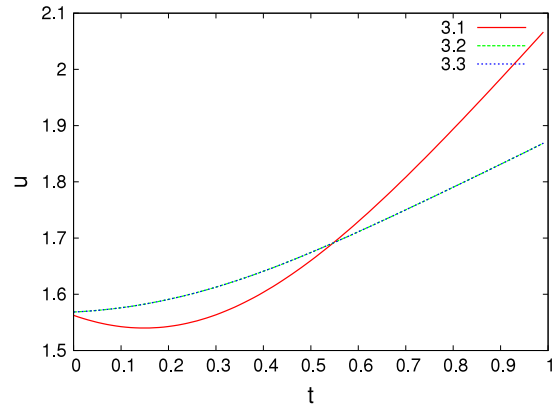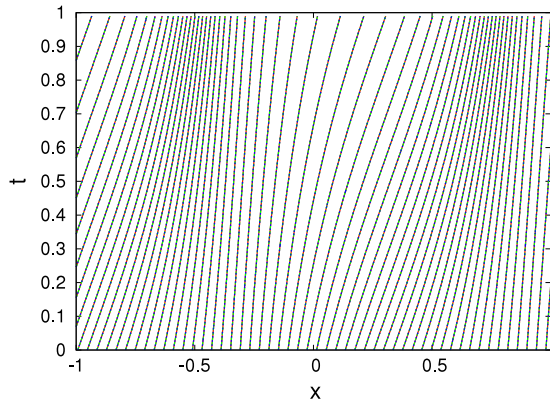**Fig. 2.** Linear advection. Flux function $f(y, x, t) = 2xy$.

Remark 5.1. The hyperbolic constraint can be very difficult to solve, e.g. due to shocks. At least we have to forbid crossing characteristics. More numerical issues are discussed in e.g. [26].

The following tests are evaluated with an initial point distance $h = 0.02$ and regularization $\alpha = 10^{-8}$. The time integrations are performed with an explicit Euler scheme using 100 time steps with $T = 1$. Moreover, the state and control function are initialized by $y_0 \equiv 0$, $u^{(0)} \equiv 0$. The localization function b is defined by $b(x) = \exp(-10x^2)$ and the tolerance parameter ϵ is set to $\epsilon = 10^{-3}$.

As exemplified for linear advection and the nonlinear Burgers equation in Figs. 2–4, respectively, all optimization strategies yield similar particle trajectories Φ_i and solutions y_i while having comparable computational effort in the

(a) Cost functional.

(b) Norm of gradient (relative).
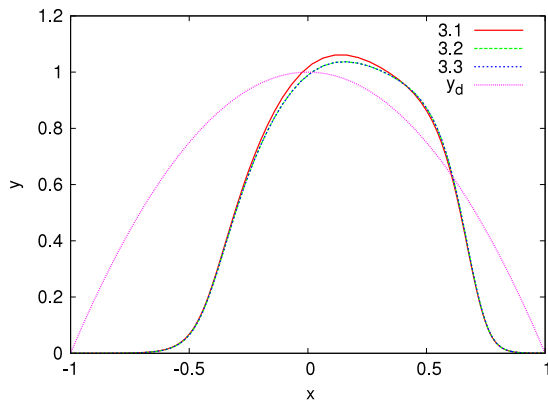
(c) State at final time T.
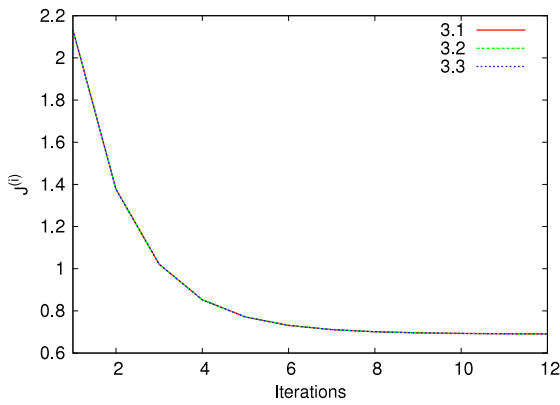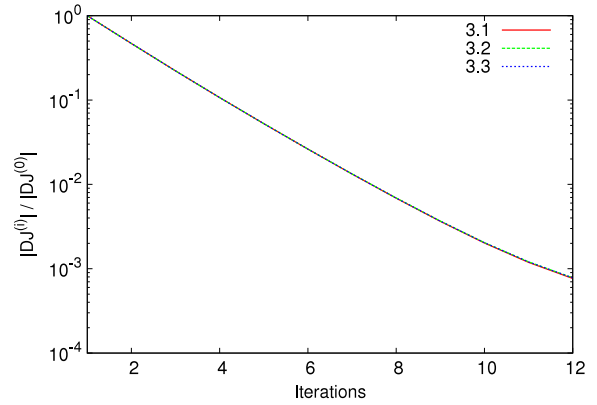
(d) Control function.

(e) Trajectories.

Fig. 3. Linear advection. Flux function $f(y, x, t) = 0.2(\sin(5x) + 1.5)y$.

performance of gradient iterations and reductions in step size. Only the evaluation of the adjoints in approach 3.1 is obviously faster, since no ODEs for p_i and z_i need to be solved as in the approaches 3.2 and 3.3. Furthermore, the optimizations are in general independent of the number of particles. Just very big particle sets implying small distances to neighbor particles could cause an exceptional behavior, since additional reductions in step size might be performed in the gradient algorithm to exclude the artificial crossing of characteristics; cf. Remark 5.1.

The numerical deviation of the Lagrangian approaches, which vary only in the sequential order of optimization and discretization, lies in the order of the discretization $\mathcal{O}(h^2)$ as theoretically predicted. This is illustrated for $f(y, x, t) = y^4/4$ in Fig. 5, where the resulting controls are compared in the L^2 sense. Also for the Eulerian approach, compared with the
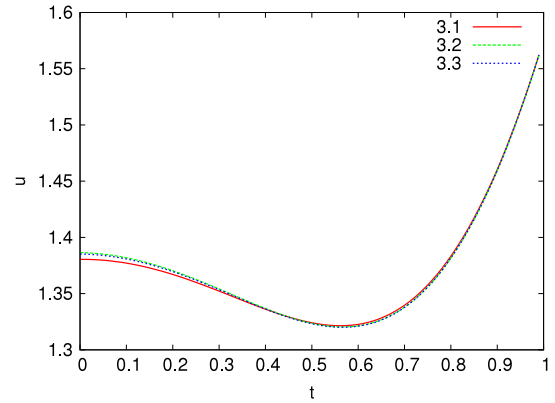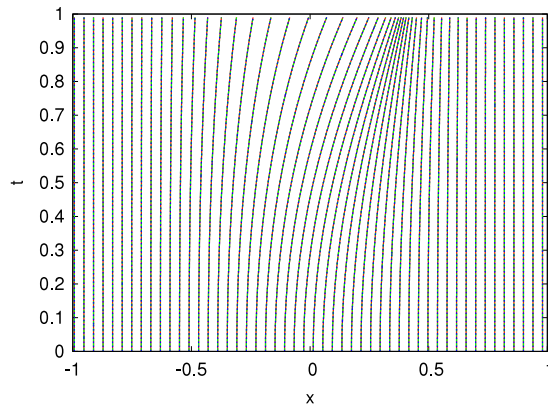
(a) Cost functional.

(b) Norm of gradient (relative).
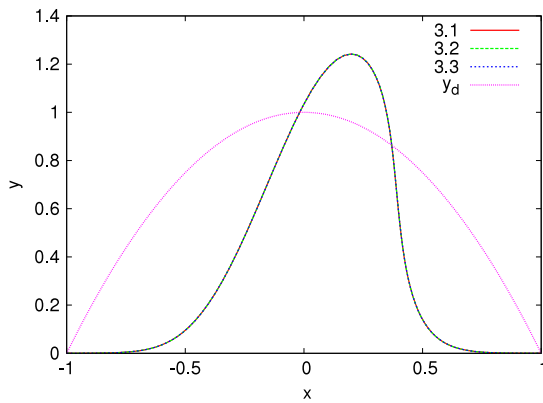
(c) State at final time T.

(d) Control function.

(e) Trajectories.

Fig. 4. Nonlinear Burgers equation. Flux function $f(y, x, t) = \frac{1}{4}y^2$.

Lagrangian approaches, we find an agreement in the range of $\mathcal{O}(h^2)$. The small discrepancy that is finally kept between 3.2/ 3.1 and 3.3 for decreasing particle distance h is carried into the simulations by the time discretization. The good agreement of the approaches is not surprising, but in fact physically desirable since the optimization results should be independent of the problem formulation and the chosen strategy (Fig. 1). This independence principle is the basis for all black-box optimization tools applied in mesh-based methods in practice so far.

Apart from the approximation quality, the test cases show clearly the superiority of particle (meshless) methods over the conventional mesh-based methods. The particles are spread according to the coordinate transformation/convective motion and yield thus an appropriate discretization that varies over time. In the case of equation-driven oscillations of the solution,
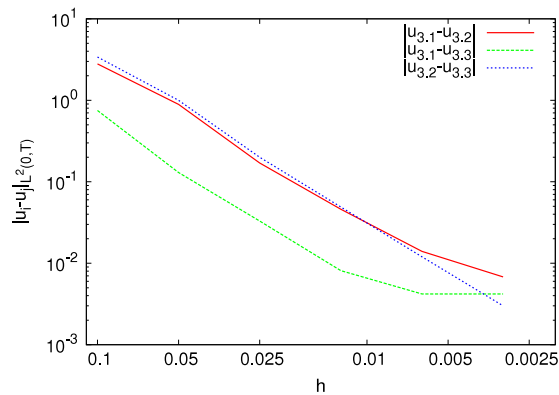
Fig. 5. Convergence of the approaches to each other.

the resolution of particle methods is automatically adapted to the problem. However, adaptive strategies on the market that generate new particles and delete redundant ones may certainly even improve performance and approximation quality. The fact that the used Lagrangian type particle method yields ordinary differential equations in time has moreover the advantage that no accuracy is lost by the approximation of spatial derivatives of the system variables. But this advantage is annulled when parabolic equations or hyperbolic systems are considered.

6. Conclusion

In this paper we combined optimal control approaches with meshless discretizations. Applying a classical Lagrangian type particle method to hyperbolic constrained optimization problems, we have presented several adjoint-based strategies differing in the sequential order of optimization and discretization of the Lagrangian or Eulerian problem formulation. Their theoretical investigation and numerical application to linear and nonlinear test cases show their suitability for the handling of practically relevant and evidently more demanding optimization problems arising in the context of fluid dynamics, failure mechanics, biophysics and biochemistry where time dependencies, complex geometries, freely moving boundaries and interfaces matter tremendously.

References

 [1] E.S. Oran, J.P. Boris, Numerical Simulation of Reactive Flows, Elsevier, 1987.
 [2] N. Marheineke, Mesh-less method for the homogeneous handling of fiber–fluid interactions, in: K.J. Bathe (Ed.), Computational Fluid and Solid Mechanics, vol. 2, Elsevier, 2003, pp. 2067–2070.
 [3] S. Tiwari, J. Kuhnert, Modeling of two-phase flows with surface tension by finite pointset method (FPM), J. Comput. Appl. Math. 203 (2) (2007) 376–386.
 [4] T. Belytschko, Y.Y. Lu, L. Gu, Element-free Galerkin methods, Internat. J. Numer. Methods Engrg. 37 (1994) 229–256.
 [5] W. Dai, P. Woodward, Extension of the piecewise parabolic method to multidimensional ideal magnetohydrodynamics, J. Comput. Phys. 115 (1994) 485–514.
 [6] S. Li, W.K. Liu, Meshfree Particle Methods, Springer, 2007.
 [7] M. Griebel, M.A. Schweitzer, Meshfree Methods for Partial Differential Equations III, in: Lecture Notes in Computational Science and Engineering, vol. 57, Springer, 2006.
 [8] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, P. Krysl, Meshless methods: an overview and recent developments, Comput. Methods Appl. Mech. Engrg. 139 (1996) 3–47.
 [9] K. Nanbu, Direct simulation scheme derived from the Boltzmann equation, J. Phys. Soc. Japan 49 (1980) 20–49.
[10] H. Neunzert, J. Struckmeier, Particle methods for Boltzmann equation, Acta Numer. (1995) 417–457.
[11] C.A.M. Duarte, A review of some meshless methods to solve partial differential equations, Tech. Rep. 95-06, TICAM, University of Texas, 1995.
[12] G. Carpentieri, B. Koren, M.J.L. van Tooren, Adjoint-based aerodynamic shape optimization on unstructured meshes, J. Comput. Phys. 224 (1) (2007) 267–287.
[13] M. Hinze, S. Ziegenbalg, Optimal control of the free boundary in a two-phase Stefan problem with flow driven by convection, Z. Angew. Math. Mech. 87 (6) (2007) 430–448.
[14] M. Hinze, R. Pinnau, M. Ulbrich, S. Ulbrich, Optimization with PDE Constraints, in: Mathematical Modelling: Theory and Applications, vol. 23, Springer, 2008.
[15] R.A. Gingold, J.J. Monaghan, Smoothed particle hydrodynamics: theory and application to non-spherical stars, Mon. Not. R. Astron. Soc. 181 (1977) 375–389.
[16] J.J. Monaghan, An introduction to SPH, Comput. Phys. Comm. 48 (1977) 89–96.
[17] P. Lancaster, Moving weighted least-squares methods, in: B.N. Sahney (Ed.), Polynomial and Spline Approximation, vol. 49, NATO Adv. Stud. Series C, Dordrecht, Boston, London, 1979, pp. 103–120.
[18] P. Lancaster, K. Salkauskas, Surfaces generated by moving least squares methods, Math. Comp. 37 (1981) 141–158.
[19] G.A. Dilts, Moving-least-square-particle hydrodynamics I: consistency and stability, Internat. J. Numer. Methods Engrg. 44 (1999) 1115–1155.
[20] W.K. Liu, S. Jun, Y.F. Zhang, Reproducing kernel particle methods, Internat. J. Numer. Methods Fluids 20 (1995) 1081–1106.
[21] I. Babuška, J.M. Melenk, The partition of unity finite element method: basic theory and applications, Comput. Methods Appl. Mech. Engrg. 139 (1996) 289–314.
[22] I. Babuška, J.M. Melenk, The partition of unity method, Internat. J. Numer. Methods Engrg. 40 (1997) 727–758.

[23] M. Griebel, M.A. Schweitzer, A particle-partition of unity method for the solution of elliptic, parabolic, and hyperbolic PDEs, SIAM J. Sci. Comput. 22 (3) (2000) 853–890.
[24] P.A. Raviart, An analysis of particle methods, in: Numerical Methods in Fluid Dynamics, Springer, Berlin, New York, 1985, pp. 243–324.
[25] F. Tröltzsch, Optimale Steuerung Partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen, Vieweg + Teubner, 2005.
[26] S. Ulbrich, Optimal control of nonlinear hyperbolic conservation laws with source terms, Habilitationsschrift, Zentrum Mathematik, Technische Universität München, 2001.
[27] E. Godlewski, P. Raviart, Numerical Approximation of Hyperbolic Systems of Conservation Laws, Springer, 1996.