Formulation, analysis and numerical study of an optimization-based conservative interpolation (remap) of scalar fields for arbitrary Lagrangian-Eulerian methods

Pavel Bochev[1,*]

Applied Mathematics and Applications, MS-1320, Sandia National Laboratories, Albuquerque, NM 87185-1320

Denis Ridzal[1]

Optimization and Uncertainty Quantification, MS-1320, Sandia National Laboratories, Albuquerque, NM 87185-1320

Guglielmo Scovazzi[1]

Computational Shock and Multiphysics, MS-1319, Sandia National Laboratories, Albuquerque, NM 87185-1319

Mikhail Shashkov

XCP-4, Methods and Algorithms, MS-B284, Los Alamos National Laboratory, Los Alamos, NM, 87545

Abstract

In this report we use optimization ideas to develop and study conservative,

[*]Corresponding author

Email addresses: pbboche@sandia.gov (Pavel Bochev), dridzal@sandia.gov (Denis Ridzal), gscovaz@sandia.gov (Guglielmo Scovazzi), shashkov@lanl.gov (Mikhail Shashkov)

URL: http://www.sandia.gov/~pbboche/ (Pavel Bochev), http://www.sandia.gov/~dridzal/ (Denis Ridzal), http://www.sandia.gov/~gscovaz/ (Guglielmo Scovazzi), http://cnls.lanl.gov/~shashkov/ (Mikhail Shashkov)

[1]Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

bound-preserving algorithms for remap of a scalar conserved quantity (mass) between close meshes having the same connectivity. We formulate remap as an inequality-constrained optimization problem in which the objective is to minimize the distance between given high-order *target mass fluxes* and the mass exchanges between neighboring cells, subject to constraints derived from physically motivated bounds on the associated primitive variable (density). In doing so, we separate accuracy considerations, handled by the objective functional, from the enforcement of physical bounds, handled by the constraints.

A typical high-order remap algorithm enforces bounds by a direct manipulation of the reconstruction process using slope limiters, which is standard practice in numerical algorithms for the solution of advection problems. In contrast, the new optimization-based remap (OBR) finds the most accurate, with respect to the selected distance measure, remapped quantity from a feasible set defined by physical bounds. As a result, the OBR formulation can be easily applied to unstructured grids and grids comprising of arbitrary cell shapes. Moreover, under some additional assumptions on the grid motion, but not on the cell types, we prove that the OBR algorithm is linearity preserving in one, two and three dimensions.

The report also examines connections between OBR and the recently proposed flux-corrected remap (FCR) [1]. We show that FCR can be interpreted as a solution procedure for a modified version of the OBR problem (M-OBR) in which the same objective is minimized over a *subset* of the OBR feasible set. The modified feasible set is derived by considering a "worst-case" scenario that replaces the original constraints by a simplified set of box constraints. The simplified constraint set decouples M-OBR into a series of one-dimensional minimization problems that can be solved independently from each other. The resulting M-OBR solution coincides with the FCR solution. It thus follows that OBR is always at least as accurate as FCR.

The report concludes with a numerical study of the OBR and FCR formulations in one dimension. We compare qualitative properties, such as shape preservation, and estimate the rates of convergence using remap on a series of smooth and "hourglass" grid cycles. The study confirms that the larger feasible set of OBR delivers increased accuracy compared to FCR.

Keywords: constrained interpolation, remap, bound-preserving, FCT, conservative, optimization, inequality constraints, quadratic programming

---

1. Introduction

Transfer of data between different grids, subject to constraints, arises in many numerical algorithms. Mesh tying [2, 3], multiphysics problems [4] and prolongation and restriction operations in multilevel methods are just a few of the examples that require this capability.

Another important example, and the primary motivation for this work, are arbitrary Lagrangian-Eulerian (ALE) methods [5]. Typically, ALE methods are split into three separate phases comprising of (i) the Lagrangian update of the solution and the computational grid; (ii) rezoning of the computational grid in order to reduce grid distortion accrued during the Lagrangian motion; and (iii) conservative interpolation (remap) of the Lagrangian solution onto the rezoned grid. Formally, it is possible to run ALE algorithms primarily in the Lagrangian mode with the occasional rezone/remap taking place only when the grid becomes too distorted. However, an alternative computational strategy that combines the best properties of Eulerian and Lagrangian methods is to perform rezoning and remapping at every time step.

An important property of this so-called *continuous rezone* strategy is that individual grid movements can be limited to small perturbations of the Lagrangian (old) mesh, which means that conserved quantities are exchanged only between neighboring cells. This localizes the remap operation to neighborhoods of old mesh cells and eliminates expensive global search operations required to locate new cells in the old mesh. However, because remap is performed at every time step, the accuracy of the continuous-rezone ALE[2] strongly depends on the quality of the remap phase.

In this paper we focus on the remap of a single scalar conserved quantity defined at cell centers, which can be thought of as the mean value of a given scalar function over the cell. For clarity, the main ideas are introduced in the context of mass remap, in which case the scalar function is the density and the remapped quantity is the product of the mean cell density and the cell volume. In this setting the conserved variable (mass) is remapped to every new (rezoned) cell and the primitive variable (density) is approximated by the remapped mass divided by the volume of the rezoned cell.

In light of the importance of the remap phase for the continuous rezone

[2]In this paper we consider only standard ALE methods in which mesh connectivity does not change.

strategy, remap algorithms should posses the following three properties:

P1. Conservation of total mass;

P2. Preservation of linearity;

P3. Preservation of bounds for the primitive variable (density).

Because the conserved values are defined at the cell centers and we assume a continuous rezone strategy, the exchange of mass between neighboring cells can be expressed in flux form; see [6]. This is sufficient to satisfy the first property because the flux form of remap guarantees both global and local conservation. The second property is a statement of accuracy of the remapping. It requires the remap algorithm to recover exact masses in the rezoned cells whenever the old masses correspond to a linear density function. The third property accounts for the fact that physically motivated bounds are imposed on the primitive variable rather than on the conserved quantity. In the case of continuous rezone, local bounds for the density can be derived by noting that every rezoned cell is contained in the union of its Lagrangian prototype and its neighbors. Therefore, it is natural to require that the mean density in a rezoned cell be bounded by the minimum and maximum mean density values on these Lagrangian cells.

Explicit advection algorithms use information only from neighboring cells and can be modified to obtain local remap operators that satisfy the above three properties. The advection approach is not without a fault though, because it enforces (P3) using a slope-limited upwind reconstruction, which ties together accuracy and preservation of bounds. This tends to obscure the sources of discretization errors and complicates the analysis of the accuracy of the remap. Another drawback is that the extension of limiters to unstructured grids, typically used in ALE methods, and to arbitrary cell shapes can be quite tricky in practice.

An alternative approach, pursued in this paper, is to rephrase conservative remap as a global inequality-constrained optimization problem. The appropriateness of this idea becomes evident upon a closer examination of the last two[3] properties in (P1–P3), which can be mapped to an optimization objective and a set of constraints defining the feasible set, respectively.

[3]The first property, i.e., conservation, is "topological" feature that can be achieved by proper selection of the variables, e.g., using the flux form of remap as mentioned earlier.

More concretely, preservation of linearity can be enforced by minimizing the distance, measured in some suitable norm, between fluxes approximating the mass exchanges across adjacent cells and *target mass fluxes* computed using a piecewise linear reconstruction of the density. If the exact density is a linear function, then the global unconstrained minimizer of the distance functional corresponds to the fluxes of the exact density, thereby recovering the exact masses in each rezoned cell.

If, however, the density is not linear, then the global unconstrained minimizer will likely produce remapped masses and densities that violate local bounds, especially when the density is not smooth. To counter the tendency of the linear reconstruction to create artificial extrema, the local bounds for the primitive variable can be enforced during the minimization of the flux distance, i.e. the search for approximate mass fluxes can be confined to a feasible set defined by the local bounds. Effectively, this transforms remap into an inequality-constrained optimization problem for the fluxes.

We expect this strategy to be more accurate than the standard practice of limiting the slope during the reconstruction process because the latter is usually based on "worst-case" assumptions that may unduly restrict the accuracy of the remap. In contrast, optimization-based remap (OBR) finds an optimal solution from a feasible set defined by the local bounds, i.e. OBR always computes the best possible remapped quantity that also satisfies these bounds.

The choice of an optimization-based strategy for remap enables important theoretical and practical gains. For instance, as accuracy and physical bounds are enforced separately in OBR, the approach can be applied without difficulty to arbitrary cell shapes, as long as the density reconstruction remains exact for linear functions. In contrast, the traditional approach of enforcing bounds through slope limiters becomes increasingly complex on unstructured grids, and is difficult to formulate, analyze and implement on arbitrary cell shapes.

The idea to cast remap into an inequality-constrained optimization problem was first suggested in [1]. However, in [1] optimization was not used to develop an actual remap algorithm but only to motivate the formulation of the flux-corrected remap (FCR), which is based on ideas from flux-corrected transport (FCT) [7]. Specifically, in [1] FCR was interpreted as "a process of replacing a global constrained optimization problem by series of local constrained optimization problems by considering the worst case scenario". However, the precise connection between the global OBR problem and FCR

was not fully explored, nor was the question of linearity preservation settled for either one of the algorithms.

This paper continues the investigation of optimization strategies for remap, started in [1], by focusing attention on the global OBR problem, its connection with the FCR algorithm, and examination of the key distinctions between the two approaches. A rigorous proof of linearity preservation for OBR is one of the key results in this paper. Our analysis shows that under some additional assumptions on the mesh movement, OBR satisfies (P2) on arbitrary unstructured grids in one, two and three dimensions, including grids with polygonal or polyhedral cells.

Another key result is the precise quantification of the intuitive interpretation of FCR. Using a suitable change of variables we show that the FCR solution coincides with the solution of a modified version of the global OBR problem, M-OBR for short, in which the original set of constraints is replaced by a set of simpler *box* constraints. The latter define a feasible set that is a *subset* of the original OBR feasible set. It follows that FCR can be interpreted as an approximate solution procedure for OBR, which searches for minimizers in a reduced feasible set. One important conclusion from this observation is that the global OBR solution is *at least* as accurate as the FCR solution, and that the latter is not necessarily linearity preserving.

Our results also clarify the origins of the explicit constraint imposed on FCR fluxes, which requires them to be *convex* combinations of low and high-order fluxes. Because the global OBR problem does not include such a constraint, its appearance in FCR cannot be inferred directly from the former. Furthermore, one can easily construct examples for which the solution of the global OBR problem *is not* a convex combination of low and high-order fluxes. Consequently, the explicit inclusion of a convexity requirement in FCR is justified in [1] by intuitive arguments based on the parallels between FCR and FCT. Our analysis reveals that the convexity requirement is introduced implicitly when the original OBR inequality constraints are replaced by simpler box constraints. This restricts the optimal solution of the global M-OBR problem to convex combinations of low and high-order fluxes. Because FCR is a solution algorithm for the M-OBR problem, the convexity requirement becomes part of the "formula" for the optimal solution. Therefore, its inclusion in FCR becomes natural without any reference to flux-corrected transport.

The paper is organized as follows. Notation is introduced in Section 2.1, a formal statement of the remap problem is presented in Section 2, and

the new optimization-based formulation of remap is developed in Section 2.3. There we also establishes sufficient conditions for the preservation of linearity in OBR. Connections between OBR and FCR are examined in Section 3, while Section 4 compares and contrasts computational properties of the two formulations.

2. Optimization-based formulation of the remap problem

2.1. Notation

In what follows $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, denotes an open bounded domain with a Lipschitz continuous boundary $\partial\Omega$. The symbol $K_h(\Omega)$ stands for a conforming partition of Ω into K cells κ_i, $i = 1, \ldots, K$, with volumes and barycenters given by

$$V(\kappa_i) = \int_{\kappa_i} dV \quad \text{and} \quad \mathbf{b}_i = \frac{\int_{\kappa_i} \mathbf{x}\, dV}{V(\kappa_i)}\,, \tag{2.1}$$

respectively. We recall that conforming partitions of Ω consist of cells that cover the domain without gaps or overlaps. The partition $K_h(\Omega)$ can be uniform or nonuniform, and the cells are not required to have the same shape or to be convex. For instance, in two dimensions $K_h(\Omega)$ can contain triangles, quadrilaterals and convex and non-convex polygons. This makes our approach applicable to a wide range of grids and methodologies. For example, we can think of a two-dimensional AMR grid [8] as consisting from quadrilaterals and pentagons, while in three dimensions [9] such grids will contain cubes and polyhedrons.

We assume that Ω is endowed with two different grid partitions $K_h(\Omega)$ and $\widetilde{K}_h(\Omega)$ having the same connectivity. In what follows, quantities defined on the new grid will have the tilde accent, e.g. $\widetilde{f}$, whereas the quantities on $K_h(\Omega)$ will have no accent. In the context of ALE methods we refer to $K_h(\Omega)$ as the old or Lagrangian grid and $\widetilde{K}_h(\Omega)$ as the new or rezoned grid. Typically, the rezoned grid is close to the Lagrangian but has better geometric quality. The cells on the new grid are denoted by $\widetilde{\kappa}_i$, with barycenters $\widetilde{\mathbf{b}}_i$, $i = 1, \ldots, K$. Because $K_h(\Omega)$ and $\widetilde{K}_h(\Omega)$ have the same connectivity, it is convenient to assume that the new cells are numbered in the same order as the old cells. Therefore, the Lagrangian parent of the rezoned cell $\widetilde{\kappa}_i$ is the cell κ_i.

The neighborhood $N(\kappa_i)$ of κ_i comprises of the cell κ_i itself and all its neighbors, i.e. those cells in $K_h(\Omega)$ that share a vertex (in 1D), vertex or edge (in 2D) and vertex, edge or face (in 3D) with κ_i. The remap problem is stated under the assumption that the rezoned grid satisfies the *locality condition*

$$\widetilde{\kappa}_i \subset N(\kappa_i), \quad \text{for all } \; i = 1, \ldots, K \,, \tag{2.2}$$

that is, each rezoned cell $\widetilde{\kappa}_i$ is contained in $N(\kappa_i)$, the neighborhood of its Lagrangian parent. Here the relation $\widetilde{\kappa}_i \subset N(\kappa_i)$ is interpreted geometrically (in contrast to its set-relational definition).[4] In the context of ALE methods, assumption (2.2) corresponds to using the continuous rezone strategy. Finally, $\mathscr{I}$ denotes the operator that returns the index of a cell, i.e. $\mathscr{I}(\kappa_i) = \mathscr{I}(\widetilde{\kappa}_i) = i$. The extension of this operator to sets of cells is natural, e.g.

$$\mathscr{I}(N(\kappa_i)) = \{\mathscr{I}(\kappa_i) \,|\, \kappa_i \in N(\kappa_i)\}$$

is the set of all indices of the cells in $N(\kappa_i)$.

For completeness, we review the specialization of some notation to one-dimensional domains $\Omega = [a, b]$ where $a < b$ are real numbers. In this case $K_h(\Omega)$ is defined by a set of $K+1$ points $a = x_0 < x_1 < \ldots < x_{K-1} < x_K = b$, the Lagrangian cells are the intervals $\kappa_i = [x_{i-1}, x_i]$ and their volumes are $V(\kappa_i) = h_i = x_i - x_{i-1}$. The new grid $\widetilde{K}_h(\Omega)$ comprises of rezoned cells $\widetilde{\kappa}_i = [\widetilde{x}_{i-1}, \widetilde{x}_i]$ such that $a = \widetilde{x}_0 < \widetilde{x}_1 < \ldots < \widetilde{x}_{K-1} < \widetilde{x}_K = b$. In one dimension, (2.2) assumes a particularly simple form:

$$\begin{gathered}\widetilde{\kappa}_i \subset (\kappa_{i-1} \cup \kappa_i \cup \kappa_{i+1}) \quad \text{for } \; i = 2, \ldots, K-1, \\ \widetilde{\kappa}_1 \subset (\kappa_1 \cup \kappa_2) \quad \text{and} \quad \widetilde{\kappa}_K \subset (\kappa_{K-1} \cup \kappa_K) \,,\end{gathered}$$

or

$$\begin{gathered}\widetilde{\kappa}_i \subset [x_{i-2}, x_{i+1}] \quad \text{for } \; i = 2, \ldots, K-1, \\ \widetilde{\kappa}_1 \subset [a, x_2] \quad \text{and} \quad \widetilde{\kappa}_K \subset [x_{K-2}, b] \,.\end{gathered}$$

An equivalent form of the locality condition is given by

$$x_{i-1} \le \widetilde{x}_i \le x_{i+1} \,, \quad i = 1, \ldots, K-1 \,. \tag{2.3}$$

[4] In what follows, we use the set-relational definitions and the corresponding geometric interpretations of $\subset$, $\subseteq$, $\cup$, $\cap$, $\setminus$, $\in$ interchangeably. Their meaning will be clear from the context. In particular, relations between entities defined on $\widetilde{K}_h(\Omega)$ and those defined on $K_h(\Omega)$ only make sense when interpreted geometrically relative to the common domain Ω.

2.2. Statement of the remap problem

For convenience, we recall the formal statement of mass-density remap following [6, 1]. We assume that there is a positive function $\rho(\mathbf{x}) > 0$, referred to as *density*, that is defined on Ω and whose values on the boundary $\partial\Omega$ are known. The only information given about ρ in the interior of Ω is its mean value on the old cells:

$$\rho_i = \frac{\int_{\kappa_i} \rho(\mathbf{x}) dV}{V(\kappa_i)} \,.$$

Equivalently, we can write

$$\rho_i = \frac{m_i}{V(\kappa_i)} \quad \text{or} \quad m_i = \rho_i V(\kappa_i) \tag{2.4}$$

where

$$m_i = \int_{\kappa_i} \rho(\mathbf{x}) dV$$

is the (old) cell mass. The total mass is

$$M = \int_\Omega \rho(\mathbf{x}) dV = \sum_{i=1}^{K} \int_{\kappa_i} \rho(\mathbf{x}) dV = \sum_{i=1}^{K} m_i = \sum_{i=1}^{K} \rho_i V(\kappa_i) \,.$$

For further reference we note that the mean density on every Lagrangian cell κ_i trivially satisfies the bounds

$$\rho_i^{\min} \le \rho_i \le \rho_i^{\max}, \tag{2.5}$$

where

$$\rho_i^{\min} = \begin{cases} \min\limits_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\} & \text{if } \kappa_i \cap \partial\Omega = \emptyset \\ \min \left\{ \min\limits_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\}, \min\limits_{\mathbf{x} \in N(\kappa_i) \cap \partial\Omega} \rho(\mathbf{x}) \right\} & \text{if } \kappa_i \cap \partial\Omega \neq \emptyset \end{cases} \tag{2.6}$$

and

$$\rho_i^{\max} = \begin{cases} \max\limits_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\} & \text{if } \kappa_i \cap \partial\Omega = \emptyset \\ \max \left\{ \max\limits_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\}, \max\limits_{\mathbf{x} \in N(\kappa_i) \cap \partial\Omega} \rho(\mathbf{x}) \right\} & \text{if } \kappa_i \cap \partial\Omega \neq \emptyset \,. \end{cases} \tag{2.7}$$

In words, for cells that do not intersect the boundary $\partial\Omega$, the values of $\rho_i^{\min}$ and $\rho_i^{\max}$ give the smallest and the largest mean densities in the neighborhood of κ_i, respectively. For cells adjacent to the boundary, $\rho_i^{\min}$ is the smaller of the smallest mean cell density in the cell neighborhood and the smallest density on the boundary segment $N(\kappa_i) \cap \partial\Omega$; $\rho_i^{\max}$ is defined analogously. For every cell κ_i the cell masses trivially satisfy the bounds

$$\rho_i^{\min} V(\kappa_i) = m_i^{\min} \leq m_i \leq m_i^{\max} = \rho_i^{\max} V(\kappa_i)\,. \tag{2.8}$$

A formal statement of the mass-density remap problem is as follows.

Definition 2.1 (Remapping of mass-density)**.** Given mean density values ρ_i on the *old* grid cells κ_i, find accurate approximations $\widetilde{m}_i$ for the masses of the *new* cells $\widetilde{\kappa}_i$,

$$\widetilde{m}_i \approx \widetilde{m}_i^{\text{ex}} = \int_{\widetilde{\kappa}_i} \rho(\mathbf{x}) dV\,; \quad i = 1, \ldots, K\,, \tag{2.9}$$

such that the following conditions hold:

R1. The total mass is conserved:

$$\sum_{i=0}^{K} \widetilde{m}_i = \sum_{i=0}^{K} m_i = M\,.$$

R2. If the exact density $\rho(\mathbf{x})$ is a linear function on all of Ω, then the remapped masses are exact:

$$\widetilde{m}_i = \widetilde{m}_i^{\text{ex}} = \int_{\widetilde{\kappa}_i} \rho(\mathbf{x}) dV\,; \quad i = 1, \ldots, K\,. \tag{2.10}$$

R3. Given approximate masses $\widetilde{m}_i$ on the new cells, define $\widetilde{\rho}_i$ as in (2.4). Let $\rho_i^{\min}$ and $\rho_i^{\max}$ be the numbers defined in (2.6)–(2.7). Then the bounds

$$\rho_i^{\min} \leq \widetilde{\rho}_i \leq \rho_i^{\max}$$

and

$$\rho_i^{\min} V(\widetilde{\kappa}_i) = \widetilde{m}_i^{\min} \leq \widetilde{m}_i \leq \widetilde{m}_i^{\max} = \rho_i^{\max} V(\widetilde{\kappa}_i) \tag{2.11}$$

hold on every new cell $\widetilde{\kappa}_i$. □

Requirements (R1–R3) in Definition 2.1 are derived from the desired remap properties stated in (P1–P3). Obviously, (R1) and (R2) are just formal statements of (P1) and (P2), whereas (R3) follows from the bounds in (2.5) and (2.8), and the locality assumption (2.2). Therefore, the last requirement is specific to a continuous rezone strategy and may have to be modified for other settings. Such a modification is beyond the scope of this paper.

2.3. Optimization formulation of the remap problem

In this section we develop an inequality-constrained optimization formulation of remap that satisfies requirements (R1–R3). In preparation for this task we examine sufficient conditions for (R1–R3), starting with (R1), the conservation of mass. Owing to the locality assumption (2.2), the exact masses of the new cells can be expressed in *flux form* (see [6]),

$$\widetilde{m}_i^{\mathrm{ex}} = m_i + \sum_{j \in \mathscr{I}(N(\kappa_i))} F_{ij}^{\mathrm{ex}}\,, \tag{2.12}$$

where the (exact) fluxes are given by

$$F_{ij}^{\mathrm{ex}} = \int_{\widetilde{\kappa}_i \cap \kappa_j} \rho(\mathbf{x}) dV - \int_{\kappa_i \cap \widetilde{\kappa}_j} \rho(\mathbf{x}) dV\,. \tag{2.13}$$

The flux form (2.12) is a consequence of the identity

$$\widetilde{\kappa}_i = \left(\kappa_i \ \cup \bigcup_{j \in \mathscr{I}(N(\kappa_i))} \widetilde{\kappa}_i \cap \kappa_j \right) \setminus \bigcup_{j \in \mathscr{I}(N(\kappa_i))} \kappa_i \cap \widetilde{\kappa}_j\,,$$

which holds for any two grids that satisfy the locality assumption (2.2). From (2.13) it follows that the exact mass fluxes are antisymmetric: $F_{ij}^{\mathrm{ex}} = -F_{ji}^{\mathrm{ex}}$. Assume now that we are given approximations F_{ij}^h of the exact fluxes that have the same property, i.e.

$$F_{ij}^h = -F_{ji}^h\,. \tag{2.14}$$

Substituting F_{ij}^h in (2.12) yields the following approximation[5] for the new cell masses:

$$\widetilde{m}_i = m_i + \sum_{j \in \mathscr{I}(N(\kappa_i))} F_{ij}^h\,. \tag{2.15}$$

[5]A simplified version of (2.15) can be obtained by limiting flux exchanges to cells that share a side. We refer to [10, 6] for further details.

The use of this formula with approximate fluxes that satisfy (2.14) guarantees global mass conservation, i.e. (R1) in Definition 2.1.

A standard way to compute F_{ij}^h is to reconstruct an approximate density $\rho_i^h(\mathbf{x})$ from the mean values on the old mesh and then integrate the result over the regions in (2.13):

$$F_{ij}^h = \int_{\widetilde{\kappa}_i \cap \kappa_j} \rho_i^h(\mathbf{x}) dV - \int_{\kappa_i \cap \widetilde{\kappa}_j} \rho_i^h(\mathbf{x}) dV \,. \tag{2.16}$$

Therefore, sufficient conditions for (R2) are that (i) the integrals in (2.16) are computed exactly[6] for linear functions and that (ii) the reconstruction procedure defining ρ_i^h is exact for linear polynomials. Indeed, if these two conditions hold and ρ is a linear polynomial on the entire domain Ω, then $\rho_i^h = \rho$ for all $1 \leq i \leq K$ and

$$F_{ij}^h = \int_{\widetilde{\kappa}_i \cap \kappa_j} \rho_i^h(\mathbf{x}) dV - \int_{\kappa_i \cap \widetilde{\kappa}_j} \rho_i^h(\mathbf{x}) dV = F_{ij}^{\text{ex}} \,,$$

that is, exact and approximate fluxes coincide. Consequently, formula (2.15) gives the exact masses on the new cells, i.e. (2.10) holds.

Finally, sufficient (and necessary) conditions for (R3) can be readily obtained by substituting the approximate mass in (2.11) with the flux form formula (2.15). The result is a global system of linear inequalities for the mass fluxes:

$$\widetilde{m}_i^{\min} \leq m_i + \sum_{j \in \mathscr{I}(N(\kappa_i))} F_{ij}^h \leq \widetilde{m}_i^{\max} \,; \quad i = 1, \ldots, K \,. \tag{2.17}$$

Having identified sufficient conditions for (R1)–(R3) in Definition 2.1 we formulate a constrained optimization problem which fulfills these requirements. To this end, we assume that for every old cell $\kappa_i \in K_h(\Omega)$ there is a density reconstruction $\rho_i^h(\mathbf{x})$ that is exact for linear functions. We recall that this is one of the two sufficient conditions for linearity preservation, (R2). Given $\rho_i^h(\mathbf{x})$ we define the *target* fluxes according to

$$F_{ij}^T = \int_{\widetilde{\kappa}_i \cap \kappa_j} \rho_i^h(\mathbf{x}) dV - \int_{\kappa_i \cap \widetilde{\kappa}_j} \rho_i^h(\mathbf{x}) dV \,,$$

[6] In practice this means that the integrals in (2.16) should be approximated by quadratures that are exact for linear functions.

and regard the fluxes in formula (2.15) as unknowns. We define the objective (cost) functional to be the sum of squares[7] of the differences between target and unknown fluxes, so that its unconstrained minimizer is given by F_{ij}^T. Finally, to fulfill (R1) and (R3), we constrain the minimization process by the antisymmetry condition (2.14) and the local bounds (2.17). Succinctly, we have expressed remap as the constrained optimization problem

$$\begin{cases} \min\limits_{F_{ij}^h} \displaystyle\sum_{i=1}^{K} \sum_{j\in\mathscr{I}(N(\kappa_i))} (F_{ij}^h - F_{ij}^T)^2 \quad \text{subject to} \\ F_{ij}^h = -F_{ji}^h \qquad\qquad\qquad\qquad i = 1,\dots,K, \quad j \in \mathscr{I}(N(\kappa_i)) \\ \widetilde{m}_i^{\min} \le m_i + \displaystyle\sum_{j\in\mathscr{I}(N(\kappa_i))} F_{ij}^h \le \widetilde{m}_i^{\max} \quad i = 1,\dots,K\,. \end{cases} \tag{2.18}$$

In practice, in lieu of (2.18) one works with the equivalent version

$$\begin{cases} \min\limits_{F_{ij}^h} \displaystyle\sum_{i=1}^{K} \sum_{\substack{j\in\mathscr{I}(N(\kappa_i)) \\ i<j}} (F_{ij}^h - F_{ij}^T)^2 \quad \text{subject to} \\ \widetilde{m}_i^{\min} \le m_i + \displaystyle\sum_{\substack{j\in\mathscr{I}(N(\kappa_i)) \\ i<j}} F_{ij}^h - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i)) \\ i>j}} F_{ji}^h \le \widetilde{m}_i^{\max} \quad i = 1,\dots,K\,, \end{cases} \tag{2.19}$$

in which the antisymmetry constraint is explicitly enforced by using only the fluxes F_{pq}^h for which $p < q$.

By construction, any solution of (2.19) satisfies (R1) and (R3). However, it is not immediately clear that (2.19) is a linearity-preserving formulation. To establish (R2) one has to show that the target fluxes F_{ij}^T are in the feasible set of (2.19) whenever the exact density ρ is a linear function, i.e. that the inequalities

$$\widetilde{m}_i^{\min} \le m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i)) \\ i<j}} F_{ij}^T - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i)) \\ i>j}} F_{ji}^T \le \widetilde{m}_i^{\max} \quad i = 1,\dots,K\,, \tag{2.20}$$

hold for linear ρ. The proof of this fact requires a simple technical result.

[7]This definition is used for simplicity. The cost functional can be defined using any valid norm function.

Lemma 2.1. *Let $n > 0$ be an integer and let $\mathbf{c} \in \mathbb{R}^n$ be an arbitrary fixed vector. For any closed and bounded set of points $P \subset \mathbb{R}^n$*

$$\min_{\mathbf{x}\in P}(\mathbf{c}^\mathsf{T}\mathbf{x}) = \min_{\mathbf{x}\in \mathscr{H}(P)}(\mathbf{c}^\mathsf{T}\mathbf{x}) \quad \textit{and} \quad \max_{\mathbf{x}\in P}(\mathbf{c}^\mathsf{T}\mathbf{x}) = \max_{\mathbf{x}\in \mathscr{H}(P)}(\mathbf{c}^\mathsf{T}\mathbf{x})\,, \tag{2.21}$$

where $\mathscr{H}(P)$ is the convex hull of P.

Proof. The real-valued function $\mathbf{c}^\mathsf{T}\mathbf{x}$ is continuous on $\mathbb{R}^n$. The set P is closed and bounded, which implies that $\mathbf{c}^\mathsf{T}\mathbf{x}$ attains its minimum and maximum over P. Since the convex hull of a closed and bounded set is closed and bounded, see [11, Theorem 17.2], the same is true for $\mathscr{H}(P)$. Therefore, the statement (2.21) is well defined.

The function $\mathbf{c}^\mathsf{T}\mathbf{x}$ is linear, hence both convex and concave. The claim of the lemma follows from a standard result on the supremum of convex (infimum of concave) functions, see e.g. [11, Theorem 32.2]. □

The following theorem provides sufficient conditions for (2.20) to hold.

Theorem 2.2. *Assume the locality condition (2.2) and suppose that the exact density ρ is linear in all of Ω. Let B_i denote the set of barycenters of the Lagrangian cells in $N(\kappa_i)$,*

$$B_i = \{\mathbf{b}_j \,|\, j \in \mathscr{I}(N(\kappa_i))\},$$

and let $\widetilde{\mathbf{b}}_i$ be the barycenter of the rezoned cell $\widetilde{\kappa}_i$. Sufficient conditions for the target fluxes to be in the feasible set of (2.19), that is for (2.20) to hold, are

$$\widetilde{\mathbf{b}}_i \in \mathscr{H}(B_i) \qquad \textit{if } \kappa_i \cap \partial\Omega = \emptyset, \tag{2.22}$$

$$\widetilde{\mathbf{b}}_i \in \mathscr{H}\left(B_i \cup (N(\kappa_i) \cap \partial\Omega)\right) \qquad \textit{if } \kappa_i \cap \partial\Omega \neq \emptyset, \tag{2.23}$$

where $\mathscr{H}(\cdot)$ denotes the convex hull.

Proof. Because ρ is linear and the density reconstruction is exact for linear functions it follows that the remapped mass equals the exact mass on every rezoned cell $\widetilde{\kappa}_i$:

$$\widetilde{m}_i = m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} F_{ij}^T - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} F_{ji}^T = m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} F_{ij}^{\text{ex}} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} F_{ji}^{\text{ex}} = \widetilde{m}_i^{\text{ex}}\,.$$

Therefore, proving that (2.20) holds reduces to showing that

$$\widetilde{m}_i^{\min} \le \widetilde{m}_i^{\mathrm{ex}} \le \widetilde{m}_i^{\max} \qquad \text{for all } \; i = 1, \ldots, K\,. \tag{2.24}$$

Recalling $\rho(\mathbf{x}) = c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}$ and using the barycenter formula (2.1) yields

$$\begin{aligned}\widetilde{m}_i^{\mathrm{ex}} &= \int_{\widetilde{\kappa}_i} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) dV = c_0 V(\widetilde{\kappa}_i) + \mathbf{c}^{\mathsf{T}} \left[\int_{\widetilde{\kappa}_i} \mathbf{x}\, dV \right] \\ &= c_0 V(\widetilde{\kappa}_i) + \mathbf{c}^{\mathsf{T}} \left[\frac{\int_{\widetilde{\kappa}_i} \mathbf{x}\, dV}{V(\widetilde{\kappa}_i)} \right] V(\widetilde{\kappa}_i) = (c_0 + \mathbf{c}^{\mathsf{T}}\, \widetilde{\mathbf{b}}_i) V(\widetilde{\kappa}_i)\,.\end{aligned}$$

We consider two cases, $\kappa_i \cap \partial\Omega = \emptyset$ and $\kappa_i \cap \partial\Omega \neq \emptyset$.

Case 1: Suppose $\kappa_i \cap \partial\Omega = \emptyset$. Using

$$\rho_i^{\min} = \min_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\} \quad \text{and} \quad \rho_i^{\max} = \max_{j \in \mathscr{I}(N(\kappa_i))} \{\rho_j\}\,,$$

the barycenter formula yields

$$\widetilde{m}_i^{\min} = \min_{j \in \mathscr{I}(N(\kappa_i))} \left[\frac{\int_{\kappa_j} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) dV}{V(\kappa_j)} \right] V(\widetilde{\kappa}_i) = \min_{\mathbf{b}_j \in B_i} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{b}_j) V(\widetilde{\kappa}_i)$$

for the lower bound and

$$\widetilde{m}_i^{\max} = \max_{j \in \mathscr{I}(N(\kappa_i))} \left[\frac{\int_{\kappa_j} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) dV}{V(\kappa_j)} \right] V(\widetilde{\kappa}_i) = \max_{\mathbf{b}_j \in B_i} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{b}_j) V(\widetilde{\kappa}_i)$$

for the upper bound in (2.24). From Lemma 2.1 it follows that

$$\min_{\mathbf{b}_j \in B_i} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{b}_j) = \min_{\mathbf{x} \in \mathscr{H}(B_i)} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \quad \text{and} \quad \max_{\mathbf{b}_j \in B_i} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{b}_j) = \max_{\mathbf{x} \in \mathscr{H}(B_i)} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x})\,. \tag{2.25}$$

Consequently, whenever $\kappa_i \cap \partial\Omega = \emptyset$, (2.24) is equivalent to

$$\min_{\mathbf{x} \in \mathscr{H}(B_i)} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \le (c_0 + \mathbf{c}^{\mathsf{T}}\widetilde{\mathbf{b}}_i) \le \max_{\mathbf{x} \in \mathscr{H}(B_i)} (c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x})\,. \tag{2.26}$$

A sufficient condition for (2.26) is given by (2.22).

Case 2: Suppose $\kappa_i \cap \partial\Omega \neq \emptyset$. We have

$$\rho_i^{\min} = \min\left\{ \min_{j\in\mathscr{I}(N(\kappa_i))}\{\rho_j\}, \min_{\mathbf{x}\in N(\kappa_i)\cap\partial\Omega}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \right\}$$

and

$$\rho_i^{\max} = \max\left\{ \max_{j\in\mathscr{I}(N(\kappa_i))}\{\rho_j\}, \max_{\mathbf{x}\in N(\kappa_i)\cap\partial\Omega}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \right\}.$$

Using again the barycenter formula, we obtain

$$\widetilde{m}_i^{\min} = \min\left\{ \min_{\mathbf{x}\in B_i}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}), \min_{\mathbf{x}\in N(\kappa_i)\cap\partial\Omega}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \right\}$$

and

$$\widetilde{m}_i^{\max} = \max\left\{ \max_{\mathbf{x}\in B_i}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}), \max_{\mathbf{x}\in N(\kappa_i)\cap\partial\Omega}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}) \right\}.$$

In other words,

$$\widetilde{m}_i^{\min} = \min_{\mathbf{x}\in B_i\cup(N(\kappa_i)\cap\partial\Omega)}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x})$$

and

$$\widetilde{m}_i^{\max} = \max_{\mathbf{x}\in B_i\cup(N(\kappa_i)\cap\partial\Omega)}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}).$$

Treating $B_i \cup (N(\kappa_i) \cap \partial\Omega)$ as a set of points in $\mathbb{R}^n$, another application of Lemma 2.1 gives

$$\widetilde{m}_i^{\min} = \min_{\mathbf{x}\in\mathscr{H}(B_i\cup(N(\kappa_i)\cap\partial\Omega))}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x})$$

and

$$\widetilde{m}_i^{\max} = \max_{\mathbf{x}\in\mathscr{H}(B_i\cup(N(\kappa_i)\cap\partial\Omega))}(c_0 + \mathbf{c}^{\mathsf{T}}\mathbf{x}).$$

Therefore, whenever $\kappa_i \cap \partial\Omega \neq \emptyset$, a sufficient condition for (2.24) is given by (2.23). This concludes the proof. □

Remark 2.1. We note that the sufficient condition (2.23) can be replaced by generally more restrictive conditions of the type

$$\widetilde{\mathbf{b}}_i \in \mathscr{H}(B_i \cup S_i) \qquad \text{if } \kappa_i \cap \partial\Omega \neq \emptyset,$$

where $S_i \subseteq (N(\kappa_i) \cap \partial\Omega)$, i.e. S_i is any (for example, finite) set of points taken from the boundary segment $N(\kappa_i) \cap \partial\Omega$.

In one dimension conclusions of Theorem 2.2 can be strengthened to show that (2.22) and (2.23) are *necessary and sufficient* for linearity preservation.

Corollary 2.3. *Assume that $\Omega = [a, b]$ is endowed with an old grid $K_h(\Omega)$ and $\widetilde{K}_h(\Omega)$ is a rezoned grid that satisfies locality condition (2.3). If the exact density is linear, but not constant, i.e. $\rho(x) = c_0 + c_1 x$ with $c_1 \neq 0$, then the target fluxes are in the feasible set of the optimization problem (2.19) if and only if (2.22) and (2.23) hold for the rezoned mesh.*

Proof. Let $B = \{b_1, \ldots, b_K\}$ be the set of barycenters (midpoints) of all cells in $K_h(\Omega)$. Let $b_0 = x_0$ and $b_{K+1} = x_K$, and let $\widetilde{b}_i$ be the barycenter (midpoint) of the rezoned cell $\widetilde{\kappa}_i$. With this notation, conditions (2.22) and (2.23) are equivalent to the single condition

$$\widetilde{b}_i \in [b_{i-1}, b_{i+1}] \quad \text{for all } \; i = 1, \ldots, K. \tag{2.27}$$

The extrema of $\rho(x) = c_0 + c_1 x$ are attained at the endpoints of these intervals. Therefore, following the argument that leads to statement (2.26) in the proof of Theorem 2.2, the target fluxes are in the feasible set of the optimization problem (2.19) if and only if

$$\min\{c_1 b_{i-1}, c_1 b_{i+1}\} \leq c_1 \widetilde{b}_i \leq \max\{c_1 b_{i-1}, c_1 b_{i+1}\} \quad \text{for all } \; i = 1, \ldots, K.$$

For $c_1 > 0$, dividing all terms by c_1 yields the inequalities

$$\min\{b_{i-1}, b_{i+1}\} \leq \widetilde{b}_i \leq \max\{b_{i-1}, b_{i+1}\}.$$

For $c_1 < 0$, dividing all terms by c_1 yields the (equivalent) inequalities

$$\max\{b_{i-1}, b_{i+1}\} \geq \widetilde{b}_i \geq \min\{b_{i-1}, b_{i+1}\}.$$

These inequalities are equivalent to (2.27). □

Additionally, in one dimension the locality condition (2.3) is sufficient to guarantee linearity preservation.

Lemma 2.4. *Assume that $\Omega = [a, b]$ is endowed with an old grid $K_h(\Omega)$ and $\widetilde{K}_h(\Omega)$ is a rezoned grid that satisfies locality condition (2.3). Then (2.27) holds for $\widetilde{K}_h(\Omega)$.*

Proof. As in Corollary 2.3, let $B = \{b_1, \ldots, b_K\}$ be the set of midpoints of all cells in $K_h(\Omega)$. Let $b_0 = x_0$ and $b_{K+1} = x_K$, and let $\widetilde{b}_i$ be the midpoint of the rezoned cell $\widetilde{\kappa}_i$. In addition, define $x_{-1} = x_0$ and $x_{K+1} = x_K$. With this notation, the locality condition (2.3) directly implies

$$\frac{x_{i-2} + x_{i-1}}{2} \leq \frac{\widetilde{x}_{i-1} + \widetilde{x}_i}{2} \leq \frac{x_i + x_{i+1}}{2} \quad \text{for all } \; i = 1, \ldots, K\,,$$

which is equivalent to the statement $b_{i-1} \leq \widetilde{b}_i \leq b_{i+1}$, i.e. condition (2.27). □

Remark 2.2. The converse is not true, i.e. the locality condition (2.3) is not necessary for the barycenter condition (2.27) to hold. A simple example is as follows. Let $h > 0$ and define the old mesh using the nodes $x_i = ih$, $i = 0, 1, 2, 3$. The midpoints of the old cells are

$$b_1 = \frac{h}{2}; \quad b_2 = \frac{3h}{2}; \quad b_3 = \frac{5h}{2}\,.$$

Define the rezoned mesh using the nodes

$$\widetilde{x}_0 = x_0 = 0; \quad \widetilde{x}_1 = \frac{9h}{4}; \quad \widetilde{x}_2 = b_2 = \frac{5h}{2}; \quad \widetilde{x}_3 = x_3 = 3h\,.$$

The locality condition is violated because $\widetilde{x}_1 = 9h/4 > 2h = x_2$. However, the barycenters of the rezoned cells are

$$\widetilde{b}_1 = \frac{9h}{8}; \quad \widetilde{b}_2 = \frac{19h}{8}; \quad \widetilde{b}_3 = \frac{22h}{8}\,,$$

and (2.27) still holds.

Remark 2.3. We note that the proof of Theorem 2.2 does not depend in any way on the type of the cells in the new and old grids. In particular, the sufficient conditions (2.22) and (2.23) for linearity preservation remain in force for grids comprising of cells such as nonconvex polygons, in two dimensions, and nonconvex polyhedra, in three dimensions. In one dimension, from Corollary 2.3 we know that these conditions become necessary and sufficient, and that they are implied by the locality assumption (2.2), i.e. when the mesh nodes satisfy (2.3).

Remark 2.4. The observations of Remark 2.3 related to one-dimensional domains cannot be extended to higher dimensions. Specifically, in two and three dimensions, (2.22) and (2.23) are sufficient but not necessary for the target

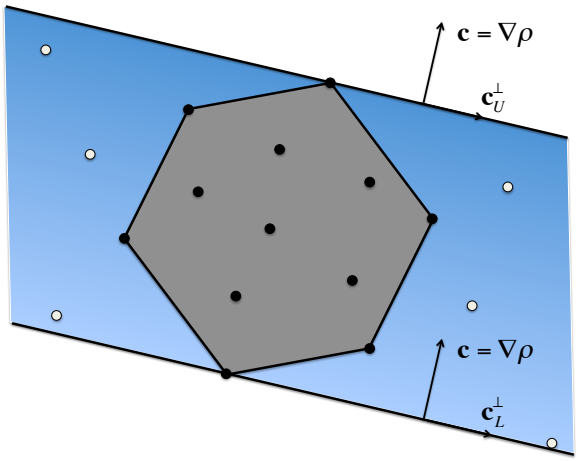

Figure 1: Corollary 2.3 gives a sufficient *and* necessary condition for linearity preservation in one dimension; in two dimensions the level sets of $\rho(\mathbf{x}) = c_0 + \mathbf{c}^\mathsf{T}\mathbf{x}$ are perpendicular to $\nabla\rho(\mathbf{x}) = \mathbf{c}$ and the extrema of $\rho(\mathbf{x})$ are achieved along the parallel lines $\mathbf{c}_L^\perp$ and $\mathbf{c}_U^\perp$ shown in the plot. Therefore, inequality (2.26) holds for all points between the two lines, while (2.22) requires $\widetilde{b}_i$ to remain in the convex hull $\mathscr{H}(B_i)$ (the gray hexagon).

fluxes to be in the feasible set of (2.19). It is easy to see, as shown in Figure 1, that in more than one dimension (2.26) can hold even if the barycenter of $\widetilde{\kappa}_i$ is not in the convex hull $\mathscr{H}(B_i)$. However, to take advantage of this fact, so as to allow a wider range of mesh motions, would require knowledge of the exact linear density, which of course is not a realistic assumption.

Simple examples showing mesh motions that comply with or violate condition (2.22) are shown in Figure 2. It is worth pointing out that a similar but more restrictive condition $\widetilde{\kappa}_i \subset \mathscr{H}(B_i)$ is necessary and sufficient for linear functions to be preserved under Van Leer slope limiting; see [12]. The center panel in Figure 2 shows an example for which $\widetilde{\kappa}_i \nsubseteq \mathscr{H}(B_i)$ but $\widetilde{\mathbf{b}}_i \in \mathscr{H}(B_i)$, i.e. condition (2.22) in Theorem 2.2 holds.

3. Connection with flux-corrected remap

In this section we examine the connections between the global OBR problem (2.19) and the FCR algorithm [1]. The first step in this process is to rewrite (2.19) in terms of the low-order and high-order fluxes employed by FCR. The reformulation of OBR amounts to a change of variables that leaves

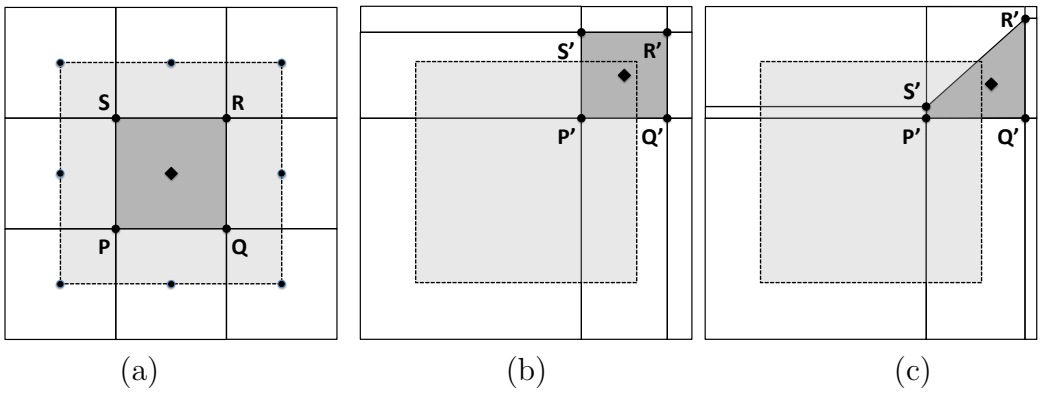

Figure 2: Examples of admissible and inadmissible mesh motions for linearity preservation: (a) the neighborhood $N(\kappa_i)$ consisting of 9 square cells, the Lagrangian parent of $\widetilde{\kappa}_i$ with vertices (**P**,**Q**,**R**,**S**), its barycenter (the diamond), the set B_i (the solid dots), and its convex hull $\mathscr{H}(B_i)$ (the dotted square); (b) an *admissible* rezoned grid for which $\widetilde{\mathbf{b}}_i \in \mathscr{H}(B_i)$; (c) an *inadmissible* rezoned grid for which $\widetilde{\mathbf{b}}_i \notin \mathscr{H}(B_i)$. In (b) and (c) $\widetilde{\kappa}_i$ is the cell with vertices (**P**',**Q**',**R**',**S**'). All cells in (a)–(c) satisfy the locality condition (2.2). Note that the rezoned cell in (b) violates $\widetilde{\kappa}_i \subset \mathscr{H}(B_i)$ which is necessary and sufficient for Van Leer slope limiting to recover linear functions [12], but which is not required for the OBR formulation.

the solution of (2.19) intact but places the OBR problem in a form that can be compared with FCR. The second step replaces the constraints in OBR by a set of inequalities which are sufficient for the original constraints to hold but have a simpler structure. This step gives rise to a modified version of OBR, termed M-OBR, in which the original objective is minimized over a subset of the original OBR feasible set. The final step entails showing that the optimal solution of M-OBR coincides with the FCR solution.

3.1. Reformulation of the optimization-based remap

The low-order fluxes in FCR are defined by the formula

$$F_{ij}^L = \int_{\widetilde{\kappa}_i \cap \kappa_j} \rho_i^h(\mathbf{x})\, dV - \int_{\kappa_i \cap \widetilde{\kappa}_j} \rho_i^h(\mathbf{x})\, dV\,,$$

using a piecewise constant reconstruction $\rho_i^h(\mathbf{x})$ of the old mesh values ρ_i, i.e.

$$\rho_i^h(\mathbf{x}) = \rho_i \quad \forall \mathbf{x}\,, \quad i = 1, \ldots, K\,.$$

We seek the solution of (2.19) as a linear combination of the low-order fluxes and the high-order target fluxes

$$F_{ij}^h = (1 - a_{ij})F_{ij}^L + a_{ij}F_{ij}^T = F_{ij}^L + a_{ij}dF_{ij}\,, \tag{3.1}$$

where $dF_{ij} = F_{ij}^T - F_{ij}^L$. The coefficients a_{ij} are the new variables for the optimization problem. Except for the symmetry condition $a_{ij} = a_{ji}$, the new variables are not subject to any additional constraints. As before, we enforce the symmetry constraint by using only coefficients a_{pq} for which $p < q$.

Under the change of variables (3.1) the objective and the constraints in (2.19) transform as follows. Using (3.1), each term in the objective functional in (2.19) can be written as

$$F_{ij}^h - F_{ij}^T = F_{ij}^L + a_{ij}dF_{ij} - F_{ij}^T = (a_{ij} - 1)dF_{ij}\,.$$

To transform the constraints, note that in terms of the new variables the remapped mass is given by the formula

$$\begin{aligned}
\widetilde{m}_i &= m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} F_{ij}^h - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} F_{ji}^h \\
&= m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} \left(F_{ij}^L + a_{ij}dF_{ij}\right) - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} \left(F_{ji}^L + a_{ji}dF_{ji}\right) \\
&= \Big(m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} F_{ij}^L - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} F_{ji}^L\Big) + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} a_{ji}dF_{ji} \\
&= \widetilde{m}_i^L + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} a_{ji}dF_{ji}\,,
\end{aligned}$$

where

$$\widetilde{m}_i^L = m_i + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} F_{ij}^L - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} F_{ji}^L$$

is a low-order approximation of the mass in the rezoned cell $\widetilde{\kappa}_i$. If the new mesh satisfies the locality condition (2.2), it is not hard to prove that

$$\widetilde{m}_i^{\min} \le \widetilde{m}_i^L \le \widetilde{m}_i^{\max}\,.$$

Therefore,

$$\widetilde{Q}_i^{\min} := \widetilde{m}_i^{\min} - \widetilde{m}_i^L \le 0 \quad \text{and} \quad \widetilde{Q}_i^{\max} := \widetilde{m}_i^{\max} - \widetilde{m}_i^L \ge 0\,,$$

and the transformed constraints can be written in the form

$$\widetilde{Q}_i^{\min} \leq \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} a_{ji}dF_{ji} \leq \widetilde{Q}_i^{\max} \qquad i=1,\ldots,K\,. \quad (3.2)$$

In summary, after changing variables according to (3.1), the OBR problem (2.19) assumes the form

$$\begin{cases} \min\limits_{a_{ij}} \sum\limits_{i=1}^{K} \sum\limits_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} (1-a_{ij})^2(dF_{ij})^2 \quad \text{subject to} \\ \widetilde{Q}_i^{\min} \leq \sum\limits_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} a_{ij}dF_{ij} - \sum\limits_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} a_{ji}dF_{ji} \leq \widetilde{Q}_i^{\max} \quad i=1,\ldots,K\,. \end{cases} \quad (3.3)$$

Problems (2.19) and (3.3) are completely equivalent. For example, the global minimizer $a_{ij}=1$ of (3.3), sans constraints, corresponds to $F_{ij}^h = F_{ij}^T$, which is the global minimizer of (2.19), sans constraints. The sufficient conditions in Theorem 2.2 guarantee that $a_{ij}=1$ are in the feasible set of (3.3) when the exact density $\rho(\mathbf{x})$ is linear function in all of Ω.

3.2. The modified optimization-based remap formulation

In this section we modify (3.3) to another inequality-constrained optimization problem, termed M-OBR, in which the same objective is minimized subject to a set of simple box constraints. The box constraints are sufficient for the original inequality constraints in (3.3) to hold and are derived by following the same reasoning as in [1]. To this end, we define the quantities

$$P_i^- = \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\leq 0} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\geq 0} dF_{ji} \leq 0\,; \qquad P_i^+ = \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\geq 0} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\leq 0} dF_{ji} \geq 0\,; \quad (3.4)$$

$$D_i^- = \begin{cases} \dfrac{\widetilde{Q}_i^{\min}}{P_i^-} & \text{if } P_i^- < 0 \\ 0 & \text{if } P_i^- = 0 \end{cases} \quad \text{and} \quad D_i^+ = \begin{cases} \dfrac{\widetilde{Q}_i^{\max}}{P_i^+} & \text{if } P_i^+ > 0 \\ 0 & \text{if } P_i^+ = 0 \end{cases}.$$

Using these quantities we reduce the constraints in (3.3) to a set of box constraints in three steps.

In the first step we replace the upper and lower bounds in the constraints of (3.3) by $D_i^- P_i^-$ and $D_i^+ P_i^+$, respectively:

$$D_i^- P_i^- \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}} a_{ji}dF_{ji} \le D_i^+ P_i^+ \qquad i=1,\dots,K\,. \quad (3.5)$$

In the second step we split (3.5) into two parts, according to the signs of the flux differentials:

$$\begin{aligned} &(a)\quad D_i^- P_i^- \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} a_{ji}dF_{ji} \le 0 \\ &(b)\quad 0 \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\ge 0} a_{ij}dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\le 0} a_{ji}dF_{ji} \le D_i^+ P_i^+ \end{aligned} \qquad i=1,\dots,K\,. \quad (3.6)$$

Finally, using definition (3.4), we reduce (3.6) to a set of box constraints by applying the upper and the lower bounds componentwise:

$$\begin{aligned} &(a)\begin{cases} D_i^- dF_{ij} \le a_{ij}dF_{ij} \le 0 & \text{for } i<j,\ dF_{ij}\le 0 \\ D_i^- dF_{ji} \ge a_{ji}dF_{ji} \ge 0 & \text{for } i>j,\ dF_{ji}\ge 0 \end{cases} \\ &(b)\begin{cases} 0 \le a_{ij}dF_{ij} \le D_i^+ dF_{ij} & \text{for } i<j,\ dF_{ij}\ge 0 \\ 0 \ge a_{ji}dF_{ji} \ge D_i^+ dF_{ji} & \text{for } i>j,\ dF_{ji}\le 0 \end{cases} \end{aligned} \qquad \begin{matrix} i=1,\dots,K \\ j\in\mathscr{I}(N(\kappa_i)) \end{matrix}\,. \quad (3.7)$$

Using the box constraints (3.7) in lieu of the original set of inequalities in (3.3) yields the modified OBR problem (M-OBR)

$$\begin{cases} \min\limits_{a_{ij}} \displaystyle\sum_{i=1}^{K} \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}} (1-a_{ij})^2 (dF_{ij})^2 \quad \text{subject to} \\ \quad (a)\begin{cases} D_i^- dF_{ij} \le a_{ij}dF_{ij} \le 0 & \text{for } i<j, dF_{ij}\le 0 \\ D_i^- dF_{ji} \ge a_{ji}dF_{ji} \ge 0 & \text{for } i>j, dF_{ji}\ge 0 \end{cases} \\ \quad (b)\begin{cases} 0 \le a_{ij}dF_{ij} \le D_i^+ dF_{ij} & \text{for } i<j, dF_{ij}\ge 0 \\ 0 \ge a_{ji}dF_{ji} \ge D_i^+ dF_{ji} & \text{for } i>j, dF_{ji}\le 0 \end{cases} \end{cases} \qquad \begin{matrix} i=1,\dots,K \\ j\in\mathscr{I}(N(\kappa_i)) \end{matrix}\,. \quad (3.8)$$

3.2.1. Properties of the M-OBR formulation

In this section we study the global M-OBR formulation (3.8) and its connections to the OBR problem (3.3). The first result shows that (3.8) always has a solution.

Proposition 3.1. *The feasible set of the modified OBR problem (3.8) is non-empty.*

Proof. The inequalities in (3.8) are always satisfied for $a_{ij} = 0$ because $D_i^- \geq 0$ and $D_i^+ \geq 0$ for all $i = 1, \ldots, K$. Therefore, the feasible set of (3.8) always contains at least one point. □

We note that $a_{ij} = 0$ results in $F_{ij}^h = F_{ij}^L$, which corresponds to a low-order mass remap or, using an advection parlance, to a "donor-cell" solution of the remap problem. Thus, at the least, the M-OBR problem admits the same solution as a conventional low-order local remapper.

The following theorem examines the relationship between M-OBR and OBR.

Theorem 3.2. *The feasible set of the M-OBR formulation (3.8) is a subset of the feasible set of the OBR formulation (3.3).*

Proof. The feasible sets of the OBR and M-OBR problems are given by

$$\mathscr{U}_O = \{a_{ij} \in \mathbb{R} \,|\, (3.2) \text{ hold for } i = 1, \ldots, K \text{ and } j \in \mathscr{I}(N(\kappa_i))\}\,,$$

and

$$\mathscr{U}_M = \{a_{ij} \in \mathbb{R} \,|\, (3.7) \text{ hold for } i = 1, \ldots, K \text{ and } j \in \mathscr{I}(N(\kappa_i))\}\,,$$

respectively. To show that $\mathscr{U}_M \subseteq \mathscr{U}_O$ define the intermediate sets

$$\mathscr{U}_A = \{a_{ij} \in \mathbb{R} \,|\, (3.5) \text{ hold for } i = 1, \ldots, K \text{ and } j \in \mathscr{I}(N(\kappa_i))\}\,,$$

and

$$\mathscr{U}_B = \{a_{ij} \in \mathbb{R} \,|\, (3.6) \text{ hold for } i = 1, \ldots, K \text{ and } j \in \mathscr{I}(N(\kappa_i))\}\,,$$

corresponding to the first and the second stages in the transformation of the OBR constraints to the box constraints of M-OBR.

To prove the theorem we will show that

$$\mathscr{U}_M \subseteq \mathscr{U}_B \subseteq \mathscr{U}_A \subseteq \mathscr{U}_O\,.$$

Step 1: $\mathscr{U}_M \subseteq \mathscr{U}_B$.. Let $\{a_{ij}\} \in \mathscr{U}_M$. Summing up the inequalities in (3.7) yields

$$\sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} D_i^- dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} D_i^- dF_{ji} \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} a_{ij} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} a_{ji} dF_{ji} \le 0\,,$$

$$0 \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} a_{ij} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} a_{ji} dF_{ji} \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} D_i^+ dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} D_i^+ dF_{ji}\,.$$

From (3.4) we see that the left hand side in the first inequality equals $D_i^- P_i^-$ and the right hand side in the second inequality is $D_i^+ P_i^+$. Therefore, inequalities (3.6) hold for $\{a_{ij}\}$, i.e. $\{a_{ij}\} \in \mathscr{U}_B$. This proves the inclusion $\mathscr{U}_M \subseteq \mathscr{U}_B$.

Step 2: $\mathscr{U}_B \subseteq \mathscr{U}_A$.. Assume that $\{a_{ij}\} \in \mathscr{U}_B$. Summing up inequalities (a) and (b) in (3.6) gives

$$D_i^- P_i^- \le \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\le 0} a_{ij} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\ge 0} a_{ji} dF_{ji} + \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i<j}}^{dF_{ij}\ge 0} a_{ij} dF_{ij} - \sum_{\substack{j\in\mathscr{I}(N(\kappa_i))\\ i>j}}^{dF_{ji}\le 0} a_{ji} dF_{ji} \le D_i^+ P_i^+$$

from where it follows that (3.5) hold for $\{a_{ij}\}$, i.e. $\{a_{ij}\} \in \mathscr{U}_A$. This proves the inclusion $\mathscr{U}_B \subseteq \mathscr{U}_A$.

Step 3: $\mathscr{U}_B \subseteq \mathscr{U}_O$.. Finally, let $\{a_{ij}\} \in \mathscr{U}_A$. Note that

$$\widetilde{Q}_i^{\min} \le D_i^- P_i^- \quad \text{and} \quad D_i^+ P_i^+ \le \widetilde{Q}_i^{\max}\,.$$

Therefore, inequalities (3.2) hold for $\{a_{ij}\}$, i.e. $\{a_{ij}\} \in \mathscr{U}_O$. This proves the inclusion $\mathscr{U}_A \subseteq \mathscr{U}_O$. □

Remark 3.1. Since the M-OBR feasible set is contained in the OBR feasible set due to Theorem 3.2, it follows that the OBR solution is always at least as accurate as the M-OBR solution.

3.2.2. Connection with the FCR algorithm

In this section we show that the M-OBR formulation is equivalent to the recently proposed FCR algorithm. For convenience, below we summarize the FCR formulation for the mass-density remap. Full details can be found in [1, Section 3].

The original motivation for FCR is to replace a global optimization problem such as OBR by a series of local problems. To this end, FCR restricts the mass fluxes in (2.15) to *convex* combinations of the low-order and the high-order target fluxes, i.e.

$$F_{ij}^h = (1 - a_{ij})F_{ij}^L + a_{ij}F_{ij}^T = F_{ij}^L + a_{ij}dF_{ij}\,, \tag{3.9}$$

where $a_{ij} = a_{ji}$ and $0 \le a_{ij} \le 1$. The convexity assumption is motivated by analogies with the FCT approach [7] for advection. Except for this requirement, formula (3.9) is identical to the change of variables in (3.1). In the FCR algorithm the approximate mass flux exchanges in (3.9) are computed using the following values for the unknown coefficients:

$$a_{ij} = \begin{cases} \min\{D_i^+, D_j^-, 1\} & \text{if } dF_{ij} > 0 \\ \min\{D_i^-, D_j^+, 1\} & \text{if } dF_{ij} < 0 \end{cases} \qquad \begin{array}{c} 1 \le i, j \le K \\ \text{and } i < j \end{array} \,. \tag{3.10}$$

For completeness, one can set $a_{ij} = 1$ whenever $dF_{ij} = 0$. In [1] it is shown that (3.10) is sufficient for the local mass-density bounds in (3.2) to hold.

We proceed to show that the solution of the global M-OBR problem is also given by (3.10). This fact establishes the equivalence of FCR and M-OBR and is a direct consequence of the following theorem.

Theorem 3.3. *The M-OBR formulation (3.8) is equivalent to the following set of independent, single-variable, constrained optimization problems: for $1 \le i, j \le K$ and $i < j$ solve*

$$\begin{cases} \min\limits_{a_{ij}} \ (1 - a_{ij})^2 (dF_{ij})^2 \quad \textit{subject to} \\ \\ 0 \le a_{ij} \le \begin{cases} \min\{D_i^+, D_j^-\} & \textit{if } dF_{ij} > 0 \\ \min\{D_i^-, D_j^+\} & \textit{if } dF_{ij} < 0\,. \end{cases} \end{cases} \tag{3.11}$$

Proof. A flux differential dF_{ij}, $i < j$, can be negative, zero or positive. If $dF_{ij} = 0$, we denote the variable a_{ij} as *free*, because the box constraint

in (3.7) holds for any value of a_{ij}. Note that the terms associated with free variables do not contribute to the objective, because $(1 - a_{ij})^2(dF_{ij})^2 = 0$. It follows that all free variables can be eliminated[8] from the optimization problem. Thus, without loss of generality we may assume that $dF_{ij} \neq 0$.

It is easy to see that whenever $dF_{ij} \neq 0$, the associated variable a_{ij} enters in exactly one constraint of type (a) and one constraint of type (b). Solving the inequalities for a_{ij} gives

$$0 \leq a_{ij} \leq D_i^+ \quad \text{and} \quad 0 \leq a_{ij} \leq D_j^-$$

for $i < j$ and $dF_{ij} > 0$, and

$$0 \leq a_{ij} \leq D_i^- \quad \text{and} \quad 0 \leq a_{ij} \leq D_j^+$$

for $i < j$ and $dF_{ij} < 0$. Succinctly,

$$0 \leq a_{ij} \leq \begin{cases} \min\{D_i^+, D_j^-\} & \text{if } dF_{ij} > 0 \\ \min\{D_i^-, D_j^+\} & \text{if } dF_{ij} < 0 \end{cases} \qquad \begin{array}{c} 1 \leq i, j \leq K \\ \text{and } i < j \end{array}$$

is a new set of box constraints that is completely equivalent to (3.8). Because each of the terms in the objective functional depends on only one variable, it follows that (3.8) decouples into the set of independent, single-variable, constrained optimization problems given in (3.11). □

The equivalence of FCR and M-OBR easily follows.

Corollary 3.4. *The solution $\{a_{ij}\}$ of the M-OBR problem (3.8) is given by the FCR formula (3.10).*

Proof. To find the solution of the M-OBR problem we set all free variables to 1. The rest of the variables are computed by solving the decoupled optimization problems in (3.11). For a given pair of indices $i < j$ let $D_{ij} \geq 0$ denote the upper bound in the constraint of the optimization problem for the variable a_{ij}. The cost functional $(1 - a_{ij})^2(dF_{ij})^2$ in this problem represents a parabola with the vertex at (1,0). Therefore, the constrained minimum is achieved at the smaller of the two values $a_{ij} = 1$ or $a_{ij} = D_{ij}$. It follows that whenever $dF_{ij} \neq 0$, the solution of the optimization problem in (3.11) is given by formula (3.10). □

[8]For a complete match with FCR we can set all free variables to 1.

4. Computational studies

In this section we present computational examples in one space dimension which examine the qualitative and quantitative properties of the OBR and M-OBR formulations. Because, as shown in Corollary 3.4, the solution of the M-OBR problem (3.8) is equivalent to the one given by the FCR algorithm, the studies in this section effectively compare and contrast the computational properties of OBR and FCR. Nonetheless, because FCR is not an optimization formulation and does not posses a notion of a feasible set, in the discussion we continue to differentiate between FCR and the M-OBR whenever the role of the feasible set needs to be highlighted. For simplicity, in all tests the computational domain Ω is the unit interval $[0, 1]$. We refer to Section 2.3 for relevant notation.

4.1. Optimization techniques for the solution of the OBR problem

In one space dimension, the OBR problem (2.19) reduces to

$$\begin{cases} \min\limits_{F^h_{i,i+1}} \ \sum\limits_{i=1}^{K-1} (F^h_{i,i+1} - F^T_{i,i+1})^2 \quad \text{subject to} \\ \\ F_1^{\min} \leq \ F^h_{1,2} \ \leq F_1^{\max}, \\ F_i^{\min} \leq \ F^h_{i,i+1} - F^h_{i-1,i} \ \leq F_i^{\max} \qquad i = 2, \ldots, K-1\,, \\ F_K^{\min} \leq \ -F^h_{K-1,K} \ \leq F_K^{\max}, \end{cases} \tag{4.1}$$

where $F_i^{\min} := \widetilde{m}_i^{\min} - m_i$, $F_i^{\max} := \widetilde{m}_i^{\max} - m_i$. This quadratic programming problem (QP) can be solved numerically using a variety of well-established techniques for inequality-constrained optimization, such as interior-point and active-set methods, see [13, Ch.16] and references therein. A detailed investigation of the performance of interior-point and active-set algorithms in the context of OBR problems is beyond the scope of this paper. Instead, we offer two alternatives that rely on either (i) an equivalent reformulation of the problem (4.1) into a *singly linearly constrained QP*, or (ii) an asymptotically equivalent reformulation into a sequence of *box-constrained QPs*. For both classes of problems, specialized solution approaches are available that take full advantage of the inherent structure, see [14, 15, 16] and [17, 18], and therefore can be extremely efficient. In particular, the box-constrained reformulation allows us to develop an algorithm that in practice exhibits the same $\mathscr{O}(K)$ complexity as the state-of-the-art remap techniques, such as FCR.

It is straightforward to verify that the change of variables

$$f_1^h = F_{1,2}^h, \qquad f_i^h = F_{i,i+1}^h - F_{i-1,i}^h, \;\; i = 2, \ldots, K-1, \qquad f_K^h = -F_{K-1,K}^h$$

yields the reformulation

$$\begin{cases} \min\limits_{f_i^h} \; \sum\limits_{i=1}^{K-1} \left(\sum\limits_{j=1}^{i} f_j^h - F_{i,i+1}^T \right)^2 & \text{subject to} \\ F_i^{\min} \leq \; f_i^h \; \leq F_i^{\max} \qquad i = 1, \ldots, K, \\ \sum_{i=1}^{K} f_i^h = 0, \end{cases} \tag{4.2}$$

which is equivalent to (4.1). We call this the *singly linearly constrained 1D-OBR QP*. Algorithms developed in [14, 15, 16] can be used for an efficient solution of (4.2). We plan to investigate their effectiveness in a future publication.

The single equality constraint in (4.2) can be enforced via a quadratic penalty term, see [13, Ch.17], which gives the *box-constrained 1D-OBR QP*

$$\begin{cases} \min\limits_{f_i^h} \; \sum\limits_{i=1}^{K-1} \left(\sum\limits_{j=1}^{i} f_j^h - F_{i,i+1}^T \right)^2 + \dfrac{1}{\gamma} \left(\sum\limits_{i=1}^{K} f_i^h \right)^2 & \text{subject to} \\ F_i^{\min} \leq \; f_i^h \; \leq F_i^{\max} \qquad i = 1, \ldots, K, \end{cases} \tag{4.3}$$

where $\gamma > 0$. In a penalty approach, one typically drives γ to zero by solving a sequence of box-constrained QPs, however, in remap applications, setting γ to a sufficiently small value and solving a single QP proves as effective as solving a sequence of QPs with a decreasing sequence $\{\gamma_k\}$. For the solution of (4.3) we adapt the algorithm for box-constrained QPs developed by Coleman and Hulbert in [17]. In contrast to Coleman and Hulbert, who suggest the use of Cholesky factorizations of a matrix that has the structure of the system matrix given by the objective functional in 4.3, we take full advantage of the structure of the system matrix and develop a procedure for the application of its inverse in $\mathscr{O}(K)$ operations. The details of the linear-algebraic considerations are beyond the scope of this paper. The application of the inverse of the system matrix dominates the cost of each optimization iteration. The outer optimization loop converges to machine precision in 6 to 15 Newton iterations for all examples, which agrees with the observations made

by Coleman and Hulbert for a variety of box-constrained QPs. Therefore, as demonstrated in Section 4.4, the total computational cost is $\mathscr{O}(K)$.

In the following we use $\gamma = 10^{-3}$. For all numerical results presented here, it can be verified that the solutions of (4.2) and (4.3) agree to at least six significant digits in the reported norms.

4.2. Shape preservation

In this section we examine the preservation of the shape of the density function under the OBR and M-OBR formulations. The goal is to show that the smaller feasible set of the M-OBR formulation (3.8) can limit its ability to accurately preserve the shape of a given density distribution. To this end we design a "torture" test example which shows how the shape of a given "peak" density distribution can be changed by M-OBR into a step-function profile. Of course, because M-OBR and FCR are equivalent, the same will hold true for the FCR solution.

A schematic of the torture test is shown in Figure 3. The old mesh $K_h(\Omega)$ is defined by a uniform partition of the unit interval into 3 cells using the vertices $x_1 = 0$, $x_2 = 1/3$, $x_3 = 2/3$ and $x_4 = 1$. The nodes of the new mesh $\widetilde{K}_h(\Omega)$ are set to $\widetilde{x}_1 = x_1$, $\widetilde{x}_2 = x_2 + \Delta_1$, $\widetilde{x}_3 = x_3 - \Delta_2$ and $\widetilde{x}_4 = x_4$, where $\Delta_1 > 0$ and $\Delta_2 > 0$ are such that $\Delta_1 + \Delta_2 < 1/3$; see Figure 3. In other words, the new mesh is defined by compressing the middle cell of the old mesh. Note that $\widetilde{K}_h(\Omega)$ satisfies the locality assumption (2.3) and that

$$x_2 < \widetilde{x}_2 \quad \text{and} \quad \widetilde{x}_3 < x_3 \,. \tag{4.4}$$

To complete the specification of the torture test we prescribe the mean density values ρ_1, ρ_2, ρ_3 on the old cells and boundary values $\rho_1^b = 0$, $\rho_3^b = 0$ at the endpoints. The mean density values are subject to the conditions

$$\rho_1 > \rho_3 \quad \text{and} \quad \rho_2 = \max\{\rho_1, \rho_2, \rho_3\} \,. \tag{4.5}$$

To explain these choices it is necessary to examine the structure of the feasible set of (3.3) and its modification (3.8), specialized to the torture test. As before, we follow the rule that the antisymmetry of fluxes and the symmetry of coefficients are enforced by using index pairs $\{i, j\}$ for which $i < j$. In the case of the torture test, which has three cells, there are two such pairs, given by $\{1, 2\}$ and $\{2, 3\}$. Therefore, the independent fluxes are F_{12}^h and F_{23}^h, the unknown coefficients are a_{12} and a_{23}, and the OBR problem (3.3) specializes

to

$$\begin{cases} \min\limits_{a_{12},a_{23}} \ (1-a_{12})^2(dF_{12})^2 + (1-a_{23})^2(dF_{23})^2 & \text{subject to} \\ \widetilde{Q}_1^{\min} \ \leq \ a_{12}dF_{12} \qquad\qquad\qquad \leq \ \widetilde{Q}_1^{\max} & (1) \\ \widetilde{Q}_2^{\min} \ \leq \ a_{23}dF_{23} \ - \ a_{12}dF_{12} \ \leq \ \widetilde{Q}_2^{\max} & (2) \\ \widetilde{Q}_3^{\min} \ \leq \qquad\qquad - \ a_{23}dF_{23} \ \leq \ \widetilde{Q}_3^{\max} & (3) \end{cases} \tag{4.6}$$

Regarding the M-OBR formulation (3.3), a simple but tedious calculation shows that $dF_{12} > 0$ and $dF_{23} > 0$ whenever (i) the middle cell is compressed, i.e. (4.4) holds, and (ii) the first condition in (4.5) holds, i.e. $\rho_1 > \rho_3$. As a result, the M-OBR problem assumes the form

$$\begin{cases} \min\limits_{a_{12},a_{23}} \ (1-a_{12})^2(dF_{12})^2 + (1-a_{23})^2(dF_{23})^2 & \text{subject to} \\ 0 \leq a_{12} \leq \min\{D_1^+, D_2^-\} & (1) \\ 0 \leq a_{23} \leq \min\{D_2^+, D_3^-\} & (2) \end{cases} \tag{4.7}$$

The left and the right panels in Figure 4 show cartoons of the feasible sets of (4.6) and (4.7), respectively. The horizontal and the vertical axes in these plots correspond to the unknowns a_{12} and a_{23}, respectively. The strips between the pairs of lines marked by $OBR(1)$, $OBR(2)$ and $OBR(3)$ correspond to the three inequality constraints in (4.6). Note that the slope of the lines marked by $OBR(2)$ is given by dF_{12}/dF_{23} and is therefore positive. The lines marked by *M-OBR(U1)* and *M-OBR(U2)* represent the two upper bounds in the two inequality constraints in (4.7), respectively. The lower bounds coincide with the coordinate axes and are marked by *M-OBR(L1)* and *M-OBR(L2)*, respectively.

The relation between the two feasible sets can be understood by examining the points **A**, **B**, **C**, **D**, **E** and **F**. The first pair of points corresponds to the lower and upper bounds on a_{12} imposed by the first constraint in (4.6). The second pair, i.e., **C**, and **D**, corresponds to the lower and upper bounds on a_{23} imposed by the third constraint in (4.6). The last two points correspond to the intercepts of the lines associated with the upper and lower bounds in the second constraint in (4.6) with the vertical and horizontal coordinate axes, respectively. The definitions of these points and their values corresponding to the actual test data used in the study are summarized in Table 1.

Point	**A**	**B**	**C**	**D**	**E**	**F**
Definition	$\dfrac{\widetilde{Q}_1^{\min}}{dF_{12}}$	$\dfrac{\widetilde{Q}_1^{\max}}{dF_{12}}$	$\dfrac{\widetilde{Q}_3^{\max}}{-dF_{23}}$	$\dfrac{\widetilde{Q}_3^{\min}}{-dF_{23}}$	$\dfrac{\widetilde{Q}_2^{\max}}{dF_{23}}$	$\dfrac{\widetilde{Q}_2^{\min}}{-dF_{12}}$
Value	-25.04	4.10	-20.53	8.62	0.00	3.28

Table 1: Control points for the feasible sets of the OBR (4.6) and the M-OBR (4.7) problems and their values for $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$ and $\rho_3 = 0$.

To explain the construction of the torture test, note that the shape of the M-OBR feasible set is completely determined by the positions of **E** and **F** along the vertical and the horizontal coordinate axes. This is a consequence of the worst-case analysis used to derive the constraints of (4.7). Consequently, by moving **E** to the origin the M-OBR feasible set can be reduced to a line extending from the origin to point **F**. This removes the point $(1, 1)$ from the feasible set and forces the M-OBR formulation to pick a solution that corresponds to remap by low-order fluxes. By moving **E** to the origin we also shrink the OBR feasible set. However, because the lines corresponding to the second constraint have positive slopes, they can be chosen in such a way that $(1, 1)$ remains in this feasible set.

In order to move **E** to the origin we need to set $\widetilde{Q}_2^{\max}/dF_{23} = 0$. It is not hard to see that this is true whenever (i) the middle cell is compressed, i.e. (4.4) holds, and (ii) the second condition in (4.5), i.e. $\rho_2^{\max} = \rho_2$ holds.

Figure 5 compares the OBR and M-OBR solutions on the new mesh for $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$, $\rho_3 = 0$, and boundary values $\rho_1^b = \rho_3^b = 0$. Table 2 shows the corresponding values of the lower and the upper inequality bounds as well as the values of the flux differentials in (4.6)–(4.7).

The initial density distribution has the shape of a "peak" and is shown in the top panel of Figure 5. The bottom panel in Figure 5 shows clearly that the OBR solution preserves this shape on the new mesh. However, as one can see from the middle panel in Figure 5, the M-OBR solution changes the shape of the peak to a step-function profile on the new mesh. Of course, owing to the equivalence of M-OBR and FCR the same will happen with the FCR solution.

The constraint sets of (4.6) and (4.7) for this example are compared in Figure 6. We see that $(1, 1)$ is included in the former but not in the

	$i=1$	$i=2$	$i=3$
$\widetilde{Q}_i^{\min}$	-40.66	-5.33	-14.00
$\widetilde{Q}_i^{\max}$	6.66	0.00	33.33
$dF_{i,i+1}$	1.62	1.62	—

Table 2: Numerical values for the lower and the upper bounds and the flux differentials in (4.6)–(4.7) corresponding to $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$, $\rho_3 = 0$, and $\rho_1^b = \rho_3^b = 0$.

latter. This is a consequence of the worst-case analysis used to obtain the constraint set in (4.7).

4.3. Convergence study of the OBR and the M-OBR formulations

The convergence studies in this section are designed to asses the asymptotic accuracy of the OBR and M-OBR (FCR) algorithms in the context of a continuous rezone strategy. In this case, the appropriate notion of remap error and convergence rates can be defined with the help of a *cyclic remap* test. The precise methodology used in the paper is described below.

4.3.1. Methodology for estimation of convergence rates of remap algorithms

A cyclic remap test simulates continuous rezone by performing remap over a parameterized sequence of grids $K_h^r(\Omega)$, $r = 0, \ldots, R$, such that the following three condition are satisfied:

- Every $K_h^r(\Omega)$, $r = 1, \ldots, R$, is topologically equivalent to the initial grid $K_h^0(\Omega)$, i.e. all grids in the sequence have the same number of cells and the same connectivity as $K_h^0(\Omega)$.
- Any two consecutive grids $K_h^{r-1}(\Omega)$, $K_h^r(\Omega)$ satisfy the locality assumption (2.2).
- The first and the last grids coincide, i.e., $K_h^0(\Omega) = K_h^R(\Omega)$.

The integer R is the number of remap steps. Its reciprocal $1/R$ can be thought of as a "pseudo-time" step which defines the temporal resolution of the cyclic remap test. The total resolution of the test is specified by the pair (K, R), where K is the number of cells in $K_h^0(\Omega)$.

Given a cyclic mesh sequence $\{K_h^r(\Omega)\}_{r=0}^R$, called a *cyclic grid*, with total resolution (K, R), let $\vec{\rho}^r \in \mathbb{R}^K$ denote the approximate density solution on

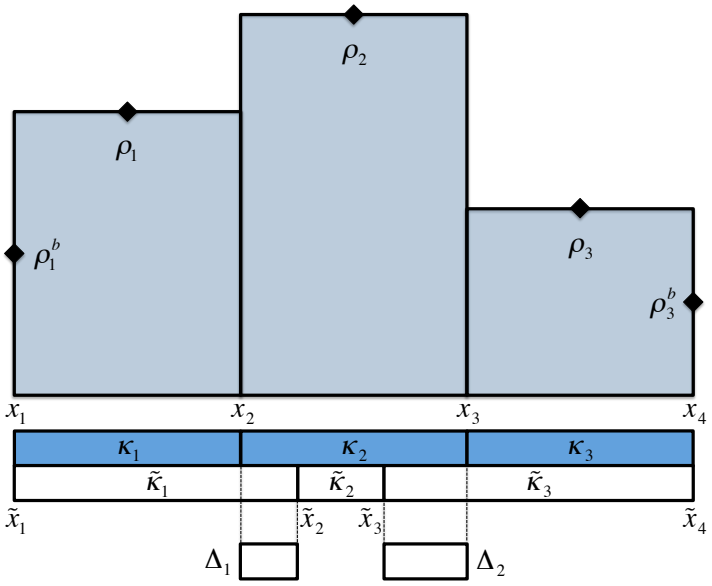

Figure 3: Specification of the "torture" test for shape preservation. The new mesh is defined by compressing the middle cell of the old mesh. The mean density values are subject to the conditions that $\rho_1 > \rho_3$ and that ρ_2 is the largest value. The results reported in this section correspond to $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$, $\rho_3 = 0$, and $\rho_1^b = \rho_3^b = 0$.

$K_h^r(\Omega)$, and $\|\cdot\|$ be a given norm on $\mathbb{R}^K$. The remap error on $\{K_h^r(\Omega)\}_{r=0}^R$ is defined by the norm of the density difference on the first and the last grids in the sequence, i.e.

$$\mathscr{E}(\rho; \|\cdot\|, K, R) = \|\vec{\rho}^0 - \vec{\rho}^r\| \,. \tag{4.8}$$

This definition is justified by the fact that $K_h^0(\Omega) = K_h^R(\Omega)$, and so the difference between the first and last solutions provides a measure of the total error accrued by the remap algorithm.

To compute the remap error $\mathscr{E}(\rho; \|\cdot\|, K, R)$ in (4.8) we use three norms suggested in [6]. Given an arbitrary vector $\vec{\phi} \in \mathbb{R}^K$ these norms are defined as follows:

$$\|\vec{\phi}\|_2 = \left(\sum_{i=1}^K \phi_i^2 h_i\right)^{1/2}, \quad \|\vec{\phi}\|_1 = \sum_{i=1}^K |\phi_i| h_i \,, \quad \|\vec{\phi}\|_\infty = \max_{0\le i\le K} |\phi_i| \,. \tag{4.9}$$

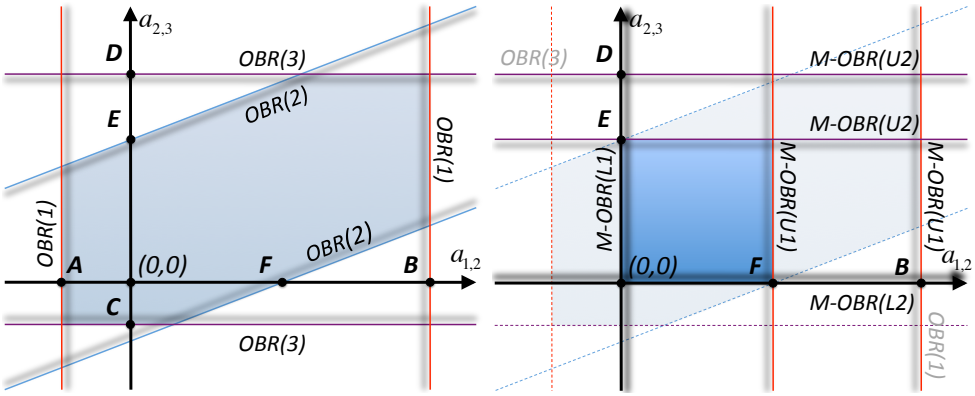

Figure 4: Structure of the OBR (left panel) and M-OBR (right panel) feasible sets when dF_{12} and dF_{23} are positive. The strips between the pairs of lines marked by $OBR(1)$, $OBR(2)$ and $OBR(3)$ correspond to the three inequality constraints in (4.6). The lines marked by $OBR(2)$ have positive slopes given by dF_{12}/dF_{23}. The lines marked by *M-OBR(U1)* and *M-OBR(U2)* represent the two upper upper bounds in the two inequality constraints in (4.7), respectively. The lower bounds correspond to the coordinate axes and are identified by *M-OBR(L1)* and *M-OBR(L2)*, respectively. The shadows point towards the interiors of the domains defined by the constraints.

If $\vec{\phi}$ is a piecewise constant approximation of a given scalar function $\phi(x)$, then these norms are discrete approximations of the L_2, L_1 and L_∞ norms on Ω, respectively.

Once the appropriate notion of remap error is defined, the estimate of convergence rates proceeds in the usual fashion: we compute remap errors using a sequence of cyclic grids with increasing resolution and then estimate the slope of the curve representing the log-log plot of the remap error versus the spatial resolution of the cyclic grid. To this end we use least-squares regression fit. Specifically, for a sequence of cyclic grids with resolutions (K_i, R_i), $i = 0, \ldots, Q$ and corresponding remap errors $\mathscr{E}_i = \mathscr{E}(\rho; \|\cdot\|, K_i, R_i)$, the rate of convergence ν_Q is estimated by least-squares regression, i.e. by solving the minimization problem

$$\{\nu_Q, \omega_Q\} = \arg\min \sum_{i=1}^{Q} \left(\log \mathscr{E}_i + \nu \log R_i - \omega\right)^2 . \tag{4.10}$$

4.3.2. Convergence study on smooth cyclic grids

The cyclic grids and the density functions for this study are adopted from [6]. Specifically, for $r = 0, \ldots, R$ the mesh node positions in $K_h^r(\Omega)$ are

given by

$$x_k^r = g(x_k^0, t_r)\,; \qquad k = 0, \ldots, K\,, \tag{4.11}$$

where

$$x_k^0 = \frac{k}{K}; \quad k = 0, \ldots, K \qquad \text{and} \qquad t_r = \frac{r}{R}; \quad r = 0, \ldots, R$$

are the uniform initial grid and sequence of pseudo-time steps, respectively, and

$$g(x,t) = (1 - \alpha(t))x + \alpha(t)x^3; \quad \alpha(t) = \frac{\sin(4\pi t)}{2} \tag{4.12}$$

is the grid mapping. One can show that for any $0 \le t \le 1$ the grids generated by this mapping are valid [6]. As a result, for any $0 \le r \le R$ grids $K_h^r(\Omega)$ satisfy

$$0 = x_0^r < x_1^r < \ldots < x_K^r = 1\,.$$

If R is sufficiently large the locality condition (2.3) also holds for every pair of consecutive grids.

Convergence rates are estimated as follows. First, we use (4.11) to define a sequence of $Q = 4$ cyclic grids with total resolutions (K_i, R_i) given by $(64, 320)$, $(256, 1280)$, $(1024, 5120)$, and $(4096, 20480)$, respectively. Thus, the resolution is increased by a factor of four in every subsequent set. Then, for every norm in (4.9) we compute the errors $\mathscr{E}_i = \mathscr{E}(\rho; \|\cdot\|, K_i, R_i)$, $i = 1, 2, 3, 4$, and solve (4.10) with $\{\mathscr{E}_1, \mathscr{E}_2\}$, $\{\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3\}$, and $\{\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3, \mathscr{E}_4\}$. This approach yields three increasingly accurate estimates of the convergence rates in each norm.

This estimation procedure is applied to three different density distributions suggested in [6]: the "sine"

$$\rho(x) = 2 + \sin(2\pi x)\,, \tag{4.13}$$

the "peak"

$$\rho(x) = \begin{cases} 0 & x < 0.25 \text{ or } x > 0.75 \\ \max\{0.001, 4(x - 0.25)\} & 0.25 \le x \le 0.50 \\ \max\{0.001, 4(0.75 - x)\} & 0.50 \le x \le 0.75 \end{cases} \tag{4.14}$$

and the "shock"

$$\rho(x) = \begin{cases} 1 & 0.0 \le x \le 0.5 \\ 0 & 0.5 \le x \le 1.0 \end{cases}. \tag{4.15}$$

OBR							
#cells	#remaps	L_2 err	L_1 err	L_∞ err	L_2 rate	L_1 rate	L_∞ rate
64	320	1.68e-3	9.17e-4	6.65e-3	—	—	—
256	1280	8.32e-5	3.03e-5	5.82e-4	2.17	2.47	1.76
1024	5120	4.47e-6	9.30e-7	5.50e-5	2.14	2.49	1.72
4096	20480	3.12e-7	3.46e-8	8.14e-6	2.07	2.46	1.62
M-OBR (FCR)							
64	320	1.99e-3	1.09e-3	7.47e-3	—	—	—
256	1280	1.30e-4	4.58e-5	8.34e-4	1.97	2.29	1.58
1024	5120	1.05e-5	2.17e-6	1.06e-4	1.89	2.24	1.53
4096	20480	9.96e-7	1.24e-7	1.56e-5	1.83	2.19	1.48

Table 3: OBR and M-OBR (FCR) errors and convergence rate estimates for the "sine" density (4.13) using 4 cyclic grids defined by (4.12).

OBR							
#cells	#remaps	L_2 err	L_1 err	L_∞ err	L_2 rate	L_1 rate	L_∞ rate
64	320	1.48e-2	7.94e-3	6.35e-2	—	—	—
256	1280	3.08e-3	1.01e-3	2.46e-2	1.13	1.49	0.68
1024	5120	6.49e-4	1.27e-4	9.25e-3	1.13	1.49	0.69
4096	20480	1.35e-4	1.61e-5	3.40e-3	1.13	1.49	0.70
M-OBR (FCR)							
64	320	1.48e-2	7.94e-3	6.40e-2	—	—	—
256	1280	3.13e-3	1.02e-3	2.52e-2	1.12	1.49	0.67
1024	5120	6.73e-4	1.31e-4	9.65e-3	1.11	1.48	0.68
4096	20480	1.46e-4	1.71e-5	3.66e-3	1.11	1.48	0.69

Table 4: OBR and M-OBR (FCR) errors and convergence rate estimates for the "peak" density (4.14) using 4 cyclic grids defined by (4.12).

Errors of the OBR and M-OBR (FCR) algorithms and the corresponding convergence rates are presented in Tables 3–5. From the data in these tables we see that in most cases the convergence rates of OBR and M-OBR are very close to each other. The only exception is the sine density for which the L_2 and L_∞ rates of OBR are better by 0.2. Overall, however, these results seem to suggest that OBR and M-OBR have roughly the same accuracy. In

OBR							
#cells	#remaps	L_2 err	L_1 err	L_∞ err	L_2 rate	L_1 rate	L_∞ rate
64	320	8.67e-2	2.47e-2	4.14e-1	—	—	—
256	1280	5.23e-2	8.97e-3	4.42e-1	0.36	0.73	-0.05
1024	5120	3.13e-2	3.20e-3	4.63e-1	0.37	0.74	-0.04
4096	20480	1.88e-2	1.15e-3	4.79e-1	0.37	0.74	-0.03
M-OBR (FCR)							
64	320	8.67e-2	2.47e-2	4.14e-1	—	—	—
256	1280	5.22e-2	8.87e-3	4.41e-1	0.37	0.74	-0.05
1024	5120	3.13e-2	3.18e-3	4.62e-1	0.37	0.74	-0.04
4096	20480	1.88e-2	1.15e-3	4.78e-1	0.37	0.74	-0.03

Table 5: OBR and M-OBR (FCR) errors and convergence rate estimates for the "shock" density (4.15) using 4 cyclic grids defined by (4.12).

the next section we show that this is not the case and that the accuracy of M-OBR can degrade for certain types of mesh motions.

4.3.3. Convergence study on hourglass cyclic grids

Theorem 3.2 asserts that the feasible set of M-OBR is always a subset of the feasible set of the OBR formulation. This suggests that (3.3) may be more accurate than (3.8), and by virtue of the equivalence of M-OBR and FCR, this conclusion extends to the latter as well. The examples in this section show that this is indeed the case and that the smaller feasible set of (3.8) can impact adversely the accuracy of M-OBR (FCR).

To this end, we compare convergence rates of the OBR and M-OBR (FCR) algorithms for the sine density (4.13) on a sequence of cyclic grids defined by the discrete "hourglass" grid mapping

$$x_k^r = g(x_k^0, t_r) = \begin{cases} x_k^0 & \text{if } r \text{ is even, for all } k, \textit{ otherwise}: \\ x_k^0 & \text{if } k \equiv 0 \pmod 3, \quad \text{or if } k = K, \\ x_k^0 + \Delta_{(K,R)} & \text{if } k \equiv 1 \pmod 3, \quad \text{for } k < K, \\ x_k^0 - \Delta_{(K,R)} & \text{if } k \equiv 2 \pmod 3, \quad \text{for } k < K\,. \end{cases} \tag{4.16}$$

As before, the initial grid K_h^0 is a uniform grid on the unit interval. For every pair (K, R) we set

$$\Delta_{(K,R)} = \frac{19}{40}\,(x_1^0 - x_0^0),$$

resulting in a constant compression ratio of 1:20 for every third grid cell (starting with cell κ_2) whenever r is odd. For even r the grid is relaxed to its original position.

Estimates of the convergence rates of OBR and M-OBR (FCR) are presented in Table 6. The first observation is that the performance of the OBR algorithm on the hourglass cyclic grid is comparable to its performance on the smooth cyclic mesh reported in Table 3. In particular, the convergence rates of OBR in all three norms equal the best possible theoretical rates for a linearity-preserving scheme.

In contrast, it is clear that the convergence rates of M-OBR (FCR) suffer on the hourglass cyclic grid. The estimates in all three norms show a consistent drop from second to first order. Moreover, an examination of the error values in Table 6 reveals that on the finest mesh the M-OBR (FCR) errors are two to three orders of magnitude greater than the OBR errors.

OBR							
#cells	#remaps	L_2 err	L_1 err	L_∞ err	L_2 rate	L_1 rate	L_∞ rate
64	320	1.52e-3	1.23e-3	3.87e-3	—	—	—
256	1280	8.96e-5	7.50e-5	2.44e-4	2.04	2.02	1.99
1024	5120	5.54e-6	4.68e-6	1.54e-5	2.03	2.01	1.99
4096	20480	3.45e-7	2.93e-7	1.39e-6	2.02	2.01	1.92
M-OBR (FCR)							
64	320	7.71e-3	5.96e-3	1.57e-2	—	—	—
256	1280	1.78e-3	1.31e-3	3.81e-3	1.06	1.09	1.02
1024	5120	4.42e-4	3.25e-4	9.51e-4	1.03	1.05	1.01
4096	20480	1.10e-4	8.10e-5	2.38e-4	1.02	1.03	1.01

Table 6: OBR and M-OBR (FCR) errors and convergence rate estimates for the sine density (4.13) using 4 cyclic hourglass grids defined by (4.16).

4.4. Computational Cost

From Theorem 3.3 we know that (3.8) decouples into a set of independent single-variable inequality-constrained optimization problems whose solution is given by (3.10), i.e. M-OBR (FCR) is quite cheap computationally. On the other hand, the OBR formulation is a globally coupled inequality-constrained optimization problem. It is therefore of considerable practical interest to assess the performance cost incurred by the need to solve a global optimization

problem. Owing to the equivalence of M-OBR and FCR, our study also provides information about the performance cost of OBR relative to FCR.

Table 7 presents preliminary results using a Matlab^{TM} implementations of OBR and M-OBR. While additional studies with more efficient implementations of M-OBR (FCR) and, especially, OBR are needed, we can already see that the cost of OBR is only a constant factor times the cost of M-OBR (FCR). The worst ratio occurs for the sine density where OBR costs approximately 6.5 times more than M-OBR (FCR). However, for density distributions such as the shock, the OBR algorithm efficiently eliminates redundant (fixed) optimization variables that are due to flat regions in the density distribution, which leads to OBR actually outperforming M-OBR (FCR).

Sine				
# cells	# remaps	M-OBR(sec)	OBR(sec)	ratio
262,144	10	6.35	43.11	6.8
524,288	10	12.60	85.56	6.8
1,048,576	10	25.33	165.67	6.5
Peak				
262,144	10	6.09	24.88	4.1
524,288	10	12.08	49.79	4.1
1,048,576	10	23.73	106.35	4.5
Shock				
262,144	10	6.12	5.28	0.86
524,288	10	12.11	10.07	0.83
1,048,576	10	23.76	19.77	0.83

Table 7: Comparison of computational costs of the OBR and FCR algorithms, as measured by wall-clock times, for the density distributions defined in (4.13), (4.14) and (4.15).

5. Conclusions

We formulate and study a new, optimization-based, conservative, bound and linearity preserving remap algorithm (OBR). The use of an optimization setting allows us to separate accuracy considerations from the enforcement of physical bounds by making the former the objective of optimization, while the latter is used to define the constraints in the optimization problem. In so

doing we obtain a scheme that is provably linearity preserving and monotone on arbitrary unstructured grids, including grids with non-convex polygonal or polyhedral cells.

The new OBR approach is compared and contrasted with the FCR algorithm [1]. We show that the FCR solution coincides with the solution of another inequality-constrained optimization problem, termed M-OBR, which is derived from OBR by replacing its constraints by a set of simpler sufficient conditions for the local bounds. These conditions are represented by box constraints obtained using a worst-case local analysis to simplify the original inequality constraints. As a result, we prove that the feasible set of M-OBR is always contained in the feasible set of OBR. It follows that OBR is always at least as accurate as M-OBR and owing to the equivalence of M-OBR and FCR, at least as accurate as the latter.

Computational examples show that for relatively smooth cyclic grids there are no significant differences in the accuracy and the convergence rates of M-OBR (FCR) and OBR. However, our study shows that on less smooth cyclic grids such as the "hourglass" grid, the smaller feasible set of M-OBR can adversely impact its accuracy. In particular, we demonstrate that on such grids M-OBR (FCR) defaults to a first-order accurate scheme, while OBR achieves the theoretically best possible accuracy (second order) for a linearity-preserving scheme. Furthermore, a "torture" test reveals that under certain conditions the smaller feasible set of M-OBR can lead to the loss of qualitative information about the shape of the remapped density distribution. Owing to the equivalence of M-OBR and FCR, these conclusions extend to the latter.

Preliminary studies show that for a set of standard remap test problems the cost of OBR is a constant factor times the cost of M-OBR (FCR). This suggests that OBR can be competitive in practical applications where a (i) provably linearity-preserving (and otherwise *optimally* accurate) and (ii) monotone method is desired.

Extension of the OBR approach to systems, its efficient implementation in 2D and further theoretical and computational studies, including a comparison with *iterated* FCR, will be the subject of a forthcoming paper.

Acknowledgments

The authors acknowledge funding by the DOE Office of Science Advanced Scientific Computing Research Program and the NNSA ASC program and

the NNSA Climate Modeling and Carbon Measurement project.

Our colleagues Richard Liska, Dmitri Kuzmin, Kara Peterson, John Shadid and Pavel Váchal provided many comments and valuable insights that helped improve this work.

References

[1] R. Liska, M. Shashkov, P. Váchal, B. Wendroff, Optimization-based synchronized flux-corrected conservative interpolation (remapping) of mass and momentum for arbitrary lagrangian-eulerian methods, Journal of Computational Physics 229 (2010) 1467–1497.

[2] T. Laursen, M. Heinstein, A three dimensional surface-to-surface projection algorithm for non-coincident domains, Comm. Numer. Meth. Eng. 19 (2003) 421–432.

[3] P. Bochev, D. Day, Analysis and computation of a least-squares method for consistent mesh tying., J. Comp. Appl. Math 218 (2008) 21–33.

[4] G. Carey, G. Bicken, V. Carey, C. Berger, J. Sanchez, Locally constrained projections on grids, Int. J. Num. Meth. Engrg. 50 (2001) 549–577.

[5] C. Hirt, A. Amsden, J. Cook, An arbitrary Lagrangian-Eulerian computing method for all flow speeds, Journal of Computational Physics 14 (1974) 227–253.

[6] L. G. Margolin, M. Shashkov, Second-order sign-preserving conservative interpolation (remapping) on general grids, J. Comput. Phys. 184 (2003) 266–298.

[7] D. Kuzmin, R. Lhner, S. Turek (Eds.), Flux-Corrected Transport. Principles, Algorithms and Applications, Springer Verlag, Berlin, Heidelberg, 2005.

[8] M. J. Berger, P. Colella, Local adaptive mesh refinement for shock hydrodynamics, J. Comput. Phys. 82 (1989) 64–84.

[9] J. Bell, M. Berger, J. Saltzman, M. Welcome, Three-dimensional adaptive mesh refinement for hyperbolic conservation laws, SIAM J. Sci. Comput. 15 (1994) 127–138.

[10] M. Kucharik, M. Shashkov, B. Wendroff, An efficient linearity-and-bound-preserving remapping method, Journal of Computational Physics 188 (2003) 462 – 471.

[11] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, N.J., 1970.

[12] B. Swartz, Good neighborhoods for multidimensional van leer limiting, Journal of Computational Physics 154 (1999) 237 – 241.

[13] J. Nocedal, S. J. Wright, Numerical Optimization, Springer Verlag, Berlin, Heidelberg, New York, first edition, 1999.

[14] Y.-H. Dai, R. Fletcher, New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds, Math. Program. 106 (2006) 403–421.

[15] J. P. Dussault, J. A. Ferland, B. Lemaire, Convex quadratic programming with one constraint and bounded variables, Mathematical Programming 36 (1986) 90–104.

[16] R. Helgason, J. Kennington, H. Lall, A polynomially bounded algorithm for a singly constrained quadratic program, Mathematical Programming 18 (1980) 338–343.

[17] T. F. Coleman, L. A. Hulbert, A globally and superlinearly convergent algorithm for convex quadratic programs with simple bbounds, SIAM Journal on Optimization 3 (1993) 298–321.

[18] T. F. Coleman, Y. Li, A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables, SIAM Journal on Optimization 6 (1996) 1040–1058.

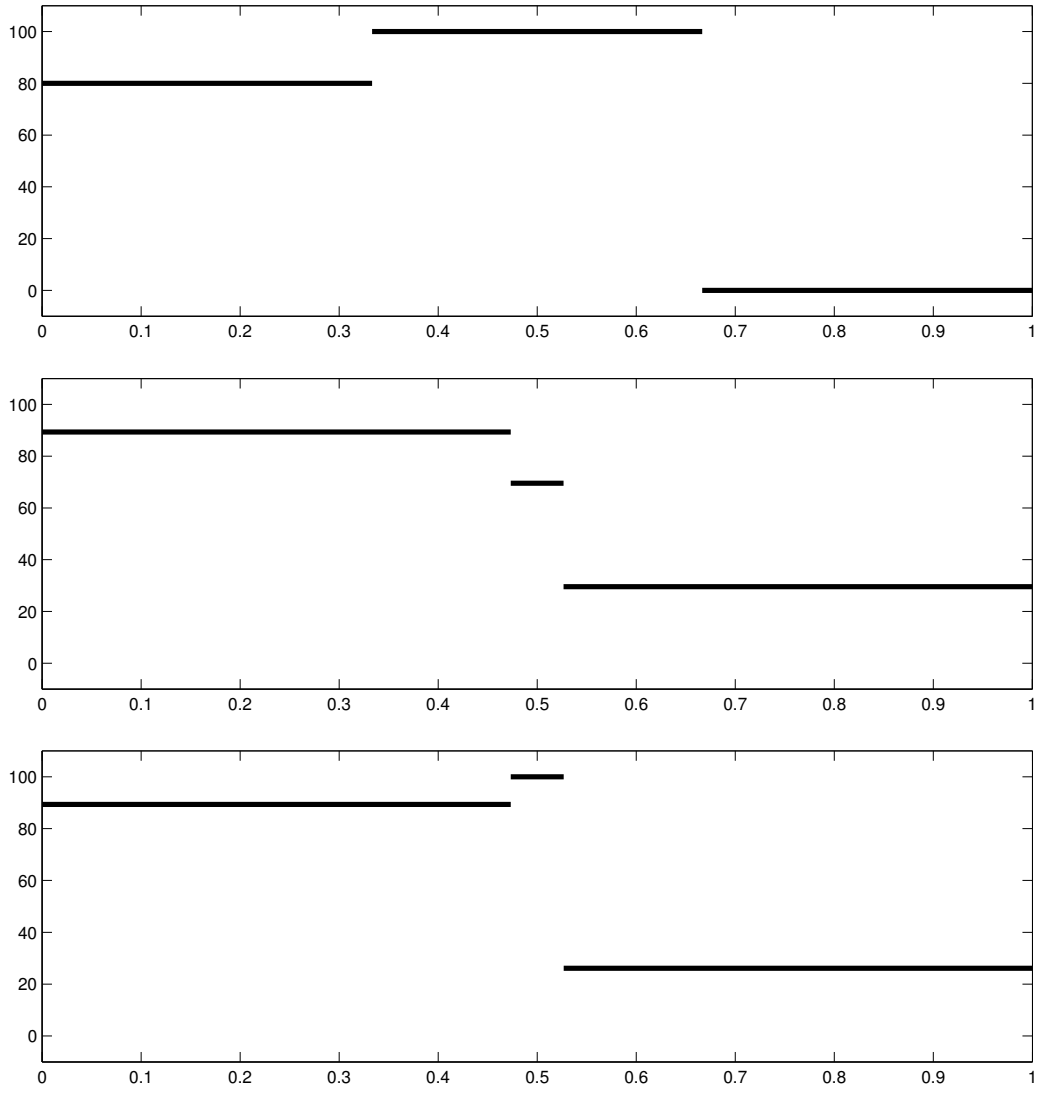

Figure 5: Initial density distribution (top panel), M-OBR solution (middle panel) and OBR solution (bottom panel) for $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$, $\rho_3 = 0$, and $\rho_1^b = \rho_3^b = 0$. The OBR solution preserves the shape of the original density distribution, while the M-OBR (FCR) solution does not.

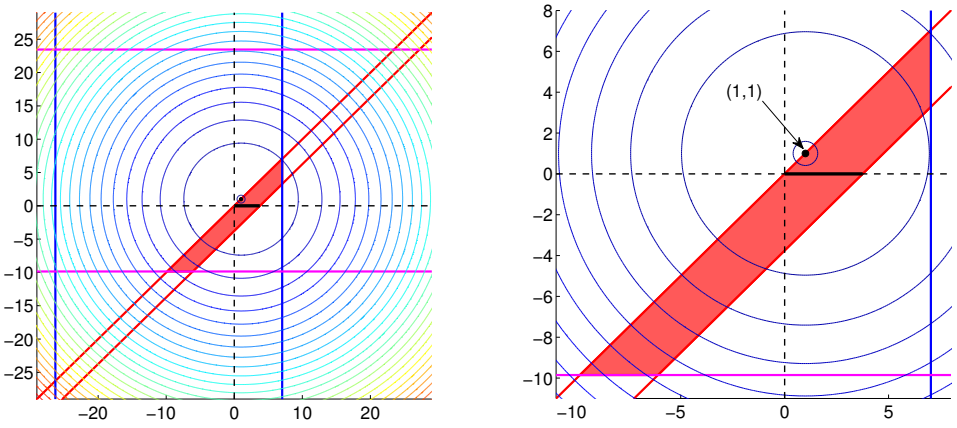

Figure 6: Level sets of the objective functional and the feasible sets of problems (4.6) and (4.7) for $\Delta_1 = \Delta_2 = 0.14$, $\rho_1 = 80$, $\rho_2 = 100$, $\rho_3 = 0$, and $\rho_1^b = \rho_3^b = 0$. The regions between horizontal (magenta), slanted (red) and vertical (blue) lines on the left panel correspond to the first, second and third constraints in the OBR problem (4.6). Their intersection (red region) gives the OBR feasible set which contains the point $(1, 1)$. The feasible set of M-OBR is given by the solid horizontal segment (black) and does not contain the point $(1, 1)$. The right panel shows a zoom of the OBR and M-OBR feasible sets.