SAND2011-1844C

Optimization-based modeling with applications to transport. Part 2. The optimization algorithm

Joseph Young[1], Denis Ridzal[2], and Pavel Bochev[1]

[1] Numerical Analysis and Applications,
[2] Optimization and Uncertainty Quantification,
Sandia National Laboratories[3], Albuquerque, NM 87185-1320, USA,
{josyoun,dridzal,pbboche}@sandia.gov

Abstract. This paper is the second of three related articles that develop and demonstrate a new optimization-based framework for computational modeling. The framework uses optimization and control ideas to assemble and decompose multiphysics operators and to preserve their fundamental physical properties in the discretization process. One application of the framework is in the formulation of robust algorithms for optimization-based transport (OBT). Based on the theoretical foundations established in Part 1, this paper focuses on the development of an efficient optimization algorithm for the solution of the *remap subproblem* that is at the heart of OBT.

1 Introduction

In this and two companion papers [1, 2] we formulate and study a new optimization-based framework for computational modeling. One application of the framework, introduced in Part 1 [1], is in the formulation of a new class of optimization-based transport (OBT) schemes. OBT schemes combine *incremental remap* [3] with the reformulation of the *remap subproblem* as an inequality-constrained quadratic program (QP) [4]. In this paper we develop and analyze an efficient optimization algorithm for the solution of the remap subproblem.

Our algorithm is based on the *dual* formulation of the remap subproblem. Our previous work [4] uses the reflective Newton method by Coleman and Li [5] for the solution of the dual remap subproblem. The Coleman-Li approach handles general bound-constrained QPs and ensures convergence from remote starting points using a trust-region globalization. In this paper we focus solely on the derivation and solution of a first-order optimality system that is specific to the remap subproblem; in other words, we disregard globalization. In practice, see Part 3 [2], the resulting Newton method proves sufficiently accurate and robust in the context of incremental remapping where a nearly feasible and optimal initial guess for the remap subproblem is typically available.

[3] Sandia National Laboratories is a multi-program laboratory managed and 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.

2 The Remap Subproblem: Optimization Theory

Optimization-based transport, see Part 1 [1], requires the solution of the *remap subproblem*,

$$\begin{aligned}
&\min_{\widetilde{F}^h_{ij}} \sum_{i=1}^{K} \sum_{\widetilde{\kappa}_j \in \widetilde{N}_i} (\widetilde{F}^h_{ij} - \widetilde{F}^T_{ij})^2 \quad \text{subject to} \\
&\begin{cases} \widetilde{F}^h_{ij} = -\widetilde{F}^h_{ji} \\ m_i^{\min} \le \widetilde{m}_i + \sum_{\widetilde{\kappa}_j \in \widetilde{N}_i} \widetilde{F}^h_{ij} \le m_i^{\max} \ , \end{cases}
\end{aligned} \tag{1}$$

where $\widetilde{\kappa}_j \in \widetilde{N}_i$ denotes $j \in \{j : \widetilde{\kappa}_j \in \widetilde{N}_i\}$, $\widetilde{F}^h_{ij}$ are the unknown mass fluxes, $\widetilde{F}^T_{ij}$ are the given target fluxes, $\widetilde{\kappa}_i$ are the cells of the deformed mesh, $\widetilde{N}_i$ are the corresponding cell neighborhoods, $m_i^{\min}$ and $m_i^{\max}$ are the given local mass extrema, and $\widetilde{m}_i$ are the given masses on the deformed mesh. We enforce the antisymmetry constraint $\widetilde{F}^h_{ij} = -\widetilde{F}^h_{ji}$ explicitly by using only the fluxes $\widetilde{F}^h_{pq}$ for which $p < q$. This results in the simplified remap subproblem

$$\begin{aligned}
&\min_{\widetilde{F}^h_{ij}} \sum_{i=1}^{K} \sum_{\substack{\widetilde{\kappa}_j \in \widetilde{N}_i \\ i<j}} (\widetilde{F}^h_{ij} - \widetilde{F}^T_{ij})^2 \quad \text{subject to} \\
&m_i^{\min} - \widetilde{m}_i \le \sum_{\substack{\widetilde{\kappa}_j \in \widetilde{N}_i \\ i<j}} \widetilde{F}^h_{ij} - \sum_{\substack{\widetilde{\kappa}_j \in \widetilde{N}_i \\ i>j}} \widetilde{F}^h_{ji} \le m_i^{\max} - \widetilde{m}_i \ .
\end{aligned} \tag{2}$$

In compact matrix / vector notation problem (2) has the form

$$\begin{aligned}
&\min_{\vec{F} \in \mathbb{R}^M} \frac{1}{2} (\vec{F} - \vec{F}^H)^\mathsf{T} (\vec{F} - \vec{F}^H) \quad \text{subject to} \\
&\vec{b}_{\min} \le \mathbf{A}\vec{F} \le \vec{b}_{\max} \, ,
\end{aligned} \tag{3}$$

where M denotes the number of unique flux variables, $\widetilde{F}^h_{ij}$. We also define $\vec{F} \in \mathbb{R}^M$, $\vec{F}^H \in \mathbb{R}^M$, $\vec{b}_{\min} \in \mathbb{R}^K$ and $\vec{b}_{\max} \in \mathbb{R}^K$ such that $\vec{F}_{\iota(i,j)} = \widetilde{F}^h_{ij}$, $\vec{F}^H_{\iota(i,j)} = \widetilde{F}^T_{ij}$, $(\vec{b}_{\min})_i = m_i^{\min} - \widetilde{m}_i$ and $(\vec{b}_{\max})_i = m_i^{\max} - \widetilde{m}_i$, respectively, where ι is an indexing function. Finally we let $\mathbf{A} \in \mathbb{R}^{K \times M}$ be a matrix with entries -1, 0 and 1 defining the inequality constraints in (2) or a related proxy (see swept-region approximation, [4, Sec. 4.1,4.2]). The matrix $\mathbf{A}$ is typically very sparse, with $M > K$ in 2D and 3D.

In what follows we use two conventions. First, we define the *Euclidean inner product*, $\langle \cdot, \cdot \rangle : \mathbb{R}^{2m} \to \mathbb{R}$, as $\langle \vec{x}, \vec{y} \rangle = \vec{x}^\mathsf{T} \vec{y}$, and the *Euclidian norm* $\|x\|_2^2 = \langle \vec{x}, \vec{x} \rangle = \vec{x}^\mathsf{T} \vec{x}$. Second, we abbreviate the *nonnegative orthant* as $\mathbb{R}^m_+ = \{\vec{x} \in \mathbb{R}^m : \vec{x} \ge 0\}$.

Rather than solving (3) directly, we focus on its dual formulation. This allows us to reformulate the problem into a simpler, *bound-constrained* optimization problem.

Theorem 1. *Given the definitions of* $\vec{F}^H \in \mathbb{R}^M$, $\vec{b}_{\min} \in \mathbb{R}^K$, $\vec{b}_{\max} \in \mathbb{R}^K$, *and* $\mathbf{A} \in \mathbb{R}^{K\times M}$ *from above, let us define* $J_p : \mathbb{R}^M \to \mathbb{R}$ *and* $J_d : \mathbb{R}^{2K} \to \mathbb{R}$ *as*

$$J_p(\vec{F}) = \frac{1}{2}\|\vec{F} - \vec{F}^H\|_2^2$$

and

$$J_d(\vec{\lambda},\vec{\mu}) = \frac{1}{2}\|\mathbf{A}^\mathsf{T}\vec{\lambda} - \mathbf{A}^\mathsf{T}\vec{\mu}\|_2^2 - \langle\vec{\lambda}, \vec{b}_{\min} - \mathbf{A}\vec{F}^H\rangle - \langle\vec{\mu}, -\vec{b}_{\max} + \mathbf{A}\vec{F}^H\rangle.$$

Then, we have that

$$\min_{F\in\mathbb{R}^M}\left\{J_p(\vec{F}) : \vec{b}_{\min} \le \mathbf{A}\vec{F} \le \vec{b}_{\max}\right\} = \min_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}^{2K}_+}\left\{J_d(\vec{\lambda},\vec{\mu})\right\}$$

where we call the first problem the primal and the second problem the dual. Furthermore,

$$\{\vec{F}^H + \mathbf{A}^\mathsf{T}(\vec{\lambda}^* - \vec{\mu}^*)\} = \arg\min_{F\in\mathbb{R}^M}\left\{J_p(\vec{F}) : \vec{b}_{\min} \le \mathbf{A}\vec{F} \le \vec{b}_{\max}\right\}$$

whenever

$$(\vec{\lambda}^*, \vec{\mu}^*) \in \arg\min_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}^{2K}_+}\left\{J_d(\vec{\lambda},\vec{\mu})\right\}.$$

Proof. We begin with the observation that J_p denotes a strictly convex, continuous function and that $\{\vec{F} \in \mathbb{R}^M : \vec{b}_{\min} \le \mathbf{A}\vec{F} \le \vec{b}_{\max}\}$ denotes a bounded, closed, convex set. Therefore, a unique minimum exists and is attained. Furthermore, since there exists an $\vec{F}$ such that $\vec{b}_{\min} < \mathbf{A}\vec{F} < \vec{b}_{\max}$ [1], we satisfy Slater's constraint qualification. This tells us that strong duality holds, which implies that the Lagrangian dual exists and possesses the same optimal value as the original problem.

Based on this knowledge, we notice that

$$\begin{aligned}
&\min_{F\in\mathbb{R}^M}\left\{J_p(\vec{F}) : \vec{b}_{\min} \le \mathbf{A}\vec{F} \le \vec{b}_{\max}\right\}\\
=&\min_{F\in\mathbb{R}^M}\max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}^{2K}_+}\left\{J_p(\vec{F}) - \langle\mathbf{A}\vec{F} - \vec{b}_{\min}, \vec{\lambda}\rangle - \langle\vec{b}_{\max} - \mathbf{A}F, \vec{\mu}\rangle\right\}\\
=&\max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}^{2K}_+}\min_{F\in\mathbb{R}^M}\left\{J_p(\vec{F}) - \langle\vec{F}, \mathbf{A}^\mathsf{T}(\vec{\lambda} - \vec{\mu})\rangle + \langle\vec{b}_{\min}, \vec{\lambda}\rangle - \langle\vec{b}_{\max}, \vec{\mu}\rangle\right\}.
\end{aligned}$$

Next, we consider the function $J : \mathbb{R}^M \to \mathbb{R}$ where

$$J(\vec{F}) = J_p(\vec{F}) - \langle\vec{F}, \mathbf{A}^\mathsf{T}(\vec{\lambda} - \vec{\mu})\rangle$$

and $(\vec{\lambda},\vec{\mu}) \in \mathbb{R}^{2K}$ are fixed. We see that J is strictly convex. Therefore, it attains its unique minimum when $\nabla J = 0$. Specifically, when

$$\vec{F} - \vec{F}^H - \mathbf{A}^\mathsf{T}(\vec{\lambda} - \vec{\mu}) = 0,$$

which occurs if and only if

$$\vec{F} = \vec{F}^H + \mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu}).$$

Therefore, we may find the optimal solution to our original problem with this equation when $(\vec{\lambda}, \vec{\mu})$ are optimal. In addition, we may use this knowledge to simplify our derivation of the dual. Let $\omega = \mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu})$ and notice that

$$\begin{aligned}
&\max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \min_{F\in\mathbb{R}^M} \left\{ J_p(\vec{F}) - \langle \vec{F}, \mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu})\rangle + \langle b_{\min}, \vec{\lambda}\rangle - \langle b_{\max}, \vec{\mu}\rangle \right\} \\
&= \max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ J_p(\vec{F}^H + \omega) - \langle \vec{F}^H + \omega, \omega)\rangle + \langle b_{\min}, \vec{\lambda}\rangle - \langle b_{\max}, \vec{\mu}\rangle \right\} \\
&= \max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ \frac{1}{2}\|\omega\|_2^2 - \langle \vec{F}^H, \omega\rangle - \|\omega\|_2^2 + \langle b_{\min}, \vec{\lambda}\rangle - \langle b_{\max}, \vec{\mu}\rangle \right\} \\
&= \max_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ -\frac{1}{2}\|\mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu})\|_2^2 - \langle \mathbf{A}\vec{F}^H, \vec{\lambda} - \vec{\mu}\rangle + \langle b_{\min}, \vec{\lambda}\rangle - \langle b_{\max}, \vec{\mu}\rangle \right\} \\
&= \min_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ \frac{1}{2}\|\mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu})\|_2^2 + \langle \mathbf{A}\vec{F}^H, \vec{\lambda} - \vec{\mu}\rangle - \langle b_{\min}, \vec{\lambda}\rangle + \langle b_{\max}, \vec{\mu}\rangle \right\} \\
&= \min_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ \frac{1}{2}\|\mathbf{A}^\intercal\vec{\lambda} - \mathbf{A}^\intercal\vec{\mu}\|_2^2 - \langle \vec{\lambda}, \vec{b}_{\min} - \mathbf{A}\vec{F}^H\rangle - \langle \vec{\mu}, -\vec{b}_{\max} + \mathbf{A}\vec{F}^H\rangle \right\} \\
&= \min_{(\vec{\lambda},\vec{\mu})\in\mathbb{R}_+^{2K}} \left\{ J_d(\vec{\lambda}, \vec{\mu}) \right\}.
\end{aligned}$$

Hence, we see the equivalence between our two optimization problems and note that the equation $\vec{F} = \vec{F}^H + \mathbf{A}^\intercal(\vec{\lambda} - \vec{\mu})$ allows us to find an optimal primal solution given an optimal solution to the dual. □

Although the primal problem is strictly convex and possesses a unique optimal solution, the dual formulation does not. Rather, the dual problem is convex, but not strictly convex, so multiple minima may exist. Second, our formula for reconstructing the primal solution from the dual depends on an optimal dual solution. If the solution to the dual is not optimal, the reconstruction formula may generate infeasible solutions. With these points in mind, we require two additional definitions before we may proceed to our optimization algorithm.

Definition 1. *We define the diagonal operator,* $\mathrm{Diag} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$*, as*

$$[\mathrm{Diag}(\vec{x})]_{ij} = \begin{cases} \vec{x}_i & i = j \\ 0 & i \neq j \end{cases}.$$

Definition 2. *For some symmetric, positive semidefinite* $\mathbf{H} \in \mathbb{R}^{m\times m}$ *and some* $\vec{b} \in \mathbb{R}^m$*, we define the operator* $v_{\mathbf{H},\vec{b}} : \mathbb{R}^m \to \mathbb{R}^m$ *as*

$$v_{\mathbf{H},\vec{b}}(\vec{x}) = \left\{ \begin{matrix} \vec{x}_i & [\mathbf{H}\vec{x} + \vec{b}]_i \geq 0 \\ 1 & [\mathbf{H}\vec{x} + \vec{b}]_i < 0 \end{matrix} \right\}.$$

When both $\mathbf{H}$ *and* $\vec{b}$ *are clear from the context, we abbreviate this function as* v*.*

In order to solve the dual optimization problem, we use a simplified version of the locally convergent Coleman-Li algorithm [5]. The key to this algorithm follows from the following lemma.

Lemma 1. *Let* $\mathbf{H} \in \mathbb{R}^{m\times m}$ *be symmetric, positive semidefinite and let* $\vec{b} \in \mathbb{R}^m$*. Then, for some* $\vec{x}^* \geq 0$*, we have that*

$$\vec{x}^* \in \arg\min_{x\in\mathbb{R}^m_+} \left\{ \frac{1}{2}\langle \mathbf{H}\vec{x}, \vec{x}\rangle + \langle \vec{b}, \vec{x}\rangle \right\} \Longleftrightarrow \mathrm{Diag}(v(\vec{x}^*))(\mathbf{H}\vec{x}^* + \vec{b}) = 0.$$

Proof. We begin with the observation that since $\mathbf{H}$ is symmetric, positive semidefinite, the problem

$$\min_{x\in\mathbb{R}^m_+} \left\{ \frac{1}{2}\langle \mathbf{H}\vec{x}, \vec{x}\rangle + \langle \vec{b}, x\rangle \right\}$$

represents a convex optimization problem with a coercive objective and a closed, convex set of constraints. Therefore, a minimum exists and the first order optimality conditions become sufficient for optimality.

In the forward direction, we assume that we have an optimal pair $(\vec{x}^*, \vec{\lambda}^*)$ that satisfy the first order optimality conditions,

$$\begin{aligned} \mathbf{H}\vec{x}^* + \vec{b} - \vec{\lambda}^* &= 0 \\ \vec{x}^* \geq 0, \vec{\lambda}^* &\geq 0 \\ \mathrm{Diag}(\vec{x}^*)\vec{\lambda}^* &= 0. \end{aligned}$$

According to these equations, $\vec{\lambda}^* = \mathbf{H}\vec{x}^* + \vec{b}$ and $\vec{\lambda}^* \geq 0$. This implies that $\mathbf{H}\vec{x}^* + \vec{b} \geq 0$. Therefore, according to the definition of v, $[\mathrm{Diag}(v(\vec{x}^*))]_{ii} = \vec{x}^*_i$ for all i. This tells us that

$$[\mathrm{Diag}(v(\vec{x}^*))(\mathbf{H}\vec{x}^* + \vec{b})]_i = \vec{x}^*_i[\mathbf{H}\vec{x}^* + \vec{b}]_i = \vec{x}^*_i\vec{\lambda}^*_i = 0$$

where the final equality follows from our fourth optimality condition, complementary slackness.

In the reverse direction, we assume that $\mathrm{Diag}(v(\vec{x}^*))(\mathbf{H}\vec{x}^* + \vec{b}) = 0$ for some $\vec{x}^* \in \mathbb{R}^m_+$. Since the problem

$$\min_{x\in\mathbb{R}^m_+} \left\{ \frac{1}{2}\langle \mathbf{H}\vec{x}, \vec{x}\rangle + \langle \vec{b}, \vec{x}\rangle \right\}$$

represents a convex optimization problem, it is sufficient to show that the first order optimality conditions hold for $\vec{x}^*$ and some $\vec{\lambda}^*$. Of course, we immediately see that we satisfy primal feasibility since $\vec{x}^* \geq 0$ by assumption.

Due to the definition of v, our initial assumption implies that $\mathbf{H}\vec{x}^* + \vec{b} \geq 0$. If this were not the case, then there would exist an i such that $[\mathbf{H}\vec{x}^* + \vec{b}]_i < 0$. In this case, we see that $[v(\vec{x}^*)]_i = 1$ and that $[\mathrm{Diag}(v(\vec{x}^*))(\mathbf{H}\vec{x}^* + \vec{b})]_i = [\mathbf{H}\vec{x}^* + \vec{b}]_i < 0$, which contradicts our initial assumption. Therefore, $\mathbf{H}\vec{x}^* + \vec{b} \geq 0$. As a result, let us set $\vec{\lambda}^* = \mathbf{H}\vec{x}^* + \vec{b}$. This allows us to satisfy our first optimality condition, $\mathbf{H}\vec{x}^* + \vec{b} - \vec{\lambda}^* = 0$ as well as our third, $\vec{\lambda}^* \geq 0$.

In order to show that we satisfy complementary slackness, we combine our initial assumption as well as our knowledge that $\mathbf{H}\vec{x}^* + \vec{b} \geq 0$ to see that

$$\begin{aligned} 0 &= \mathrm{Diag}(v(\vec{x}^*))(\mathbf{H}\vec{x}^* + \vec{b}) \\ &= \mathrm{Diag}(\vec{x}^*)(\mathbf{H}\vec{x}^* + \vec{b}) \\ &= \mathrm{Diag}(\vec{x}^*)\vec{\lambda}^*. \end{aligned}$$

Therefore, we satisfy our final optimality condition and, hence, $\vec{x}^*$ denotes an optimal solution to the optimization problem. □

The above lemma allows us to recast a bound-constrained, convex quadratic optimization problem into a piecewise differentiable system of equations. In order to solve this system of equations, we apply Newton's method. Before we do so, we require one additional definition and a lemma.

Definition 3. *For some symmetric, positive semidefinite* $\mathbf{H} \in \mathbb{R}^{m\times m}$ *and some* $\vec{b} \in \mathbb{R}^m$*, we define the operator* $K_{\mathbf{H},\vec{b}} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ *as*

$$[K_{\mathbf{H},\vec{b}}(\vec{x})]_{ij} = \begin{cases} 1 & [\mathbf{H}\vec{x} + \vec{b}]_i \geq 0 \\ 0 & [\mathbf{H}\vec{x} + \vec{b}]_i < 0 \end{cases}.$$

When both $\mathbf{H}$ *and* $\vec{b}$ *are clear from the context, we abbreviate this operator as* K*.*

Lemma 2. *Let* $\mathbf{H} \in \mathbb{R}^{m\times m}$ *be symmetric, positive definite,* $\vec{b} \in \mathbb{R}^m$*, and define the function* $J : \mathbb{R}^m \to \mathbb{R}$ *as*

$$J(\vec{x}) = \mathrm{Diag}(v(\vec{x}))(\mathbf{H}\vec{x} + \vec{b}).$$

Then, we have that

$$J'(\vec{x}) = K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + \mathrm{Diag}(v(x))H.$$

Proof. Let us begin by assessing the derivative of v. We notice that

$$[v(\vec{x} + t\vec{\eta})]_i = \begin{cases} \vec{x}_i + t\vec{\eta}_i & [\mathbf{H}\vec{x} + b]_i \geq 0 \\ 1 & [\mathbf{H}\vec{x} + b]_i < 0 \end{cases}.$$

Therefore, from a piecewise application of Taylor's theorem, we see that

$$[v'(\vec{x})\vec{\eta}]_i = \begin{cases} \vec{\eta}_i & [\mathbf{H}\vec{x} + b]_i \geq 0 \\ 0 & [\mathbf{H}\vec{x} + b]_i < 0 \end{cases}.$$

Next, we apply a similar technique to J. Let us define $g : \mathbb{R}^m \to \mathbb{R}$ so that $g(\vec{x}) = \mathbf{H}\vec{x} + \vec{b}$. Then, we see that

$$\begin{aligned} J(\vec{x} + t\vec{\eta}) =& \mathrm{Diag}(v(\vec{x} + t\vec{\eta}))(\mathbf{H}(\vec{x} + t\vec{\eta}) + \vec{b}) \\ =& \mathrm{Diag}(v(\vec{x}) + tv'(\vec{x})\vec{\eta} + o(|t|))(\mathbf{H}\vec{x} + \vec{b} + t\vec{\eta}) \\ =& \mathrm{Diag}(v(\vec{x}))g(\bar{x}) + t\left(\mathrm{Diag}(v(\vec{x}))\mathbf{H}\vec{\eta} + \mathrm{Diag}(v'(\vec{x})\vec{\eta})g(\bar{x})\right) + o(|t|). \end{aligned}$$

Hence, from a piecewise application of Taylor's theorem, we have that

$$\begin{aligned} J'(\vec{x})\vec{\eta} &= \mathrm{Diag}(v(\vec{x}))\mathbf{H}\vec{\eta} + \mathrm{Diag}(v'(\vec{x})\vec{\eta})(\mathbf{H}\vec{x} + \vec{b}) \\ &= \mathrm{Diag}(v(\vec{x}))\mathbf{H}\vec{\eta} + K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b})\vec{\eta}. \end{aligned}$$

Therefore, $J'(\vec{x}) = K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + \mathrm{Diag}(v(\vec{x}))\mathbf{H}$. $\square$

The preceding lemma allows us to formulate Newton's method where we seek a step $\vec{p} \in \mathbb{R}^m$ such that $J'(\vec{x})\vec{p} = -J(\vec{x})$. Although the operator $J'(\vec{x})$ is well structured, it is nonsymmetric. We symmetrize the system as follows.

Definition 4. *For some symmetric, positive semidefinite $\mathbf{H} \in \mathbb{R}^{m\times m}$ and some $\vec{b} \in \mathbb{R}^m$, we define the operator $D_{\mathbf{H},\vec{b}} : \mathbb{R}^m_+ \to \mathbb{R}^{m\times m}$ as*

$$D_{\mathbf{H},\vec{b}}(\vec{x}) = \mathrm{Diag}(v_{\mathbf{H},\vec{b}}(\vec{x}))^{1/2}.$$

When both $\mathbf{H}$ and $\vec{b}$ are clear from the context, we abbreviate this operator as D.

Lemma 3. *Let $\mathbf{H} \in \mathbb{R}^{m\times m}$ be symmetric, positive semidefinite and let $\vec{b} \in \mathbb{R}^m$. Then, we have that*

$$\begin{aligned} &(K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + \mathrm{Diag}(v(\vec{x}))\mathbf{H})\vec{p} = -\mathrm{Diag}(v(\vec{x}))(\mathbf{H}\vec{x} + \vec{b}) \\ \Longleftrightarrow &(K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H}D(\vec{x}))\vec{q} = -D(\vec{x})(\mathbf{H}\vec{x} + \vec{b}) \end{aligned}$$

where $\vec{p} = D(x)\vec{q}$.

Proof. Notice that

$$\begin{aligned} 0 &= (K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + \mathrm{Diag}(v(\vec{x}))\mathbf{H})\vec{p} + \mathrm{Diag}(v(\vec{x}))(\mathbf{H}\vec{x} + \vec{b}) \\ &= (K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})^2\mathbf{H})\vec{p} + D(\vec{x})^2(\mathbf{H}\vec{x} + \vec{b}) \\ &= D(\vec{x})((D(\vec{x})^{-1}K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H})\vec{p} + D(\vec{x})(\mathbf{H}\vec{x} + \vec{b})) \\ &= D(\vec{x})((D(\vec{x})^{-1}K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H})D(\vec{x})\vec{q} + D(\vec{x})(\mathbf{H}\vec{x} + \vec{b})) \\ &= D(\vec{x})((K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H}D(\vec{x}))\vec{q} + D(\vec{x})(\mathbf{H}\vec{x} + \vec{b})), \end{aligned}$$

which occurs if and only if

$$0 = (K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H}D(\vec{x}))\vec{q} + D(\vec{x})(\mathbf{H}\vec{x} + \vec{b})$$

since $D(\vec{x})$ is nonsingular. $\square$

Properly, we require a line search to ensure feasible iterates. However, we can be far more aggressive in practice. In order to initialize the algorithm, we use the starting iterate of $(\vec{\lambda}, \vec{\mu}) = (\vec{0}, \vec{0})$. This corresponds to a primal solution where $\vec{F} = \vec{F}^H$. Since the optimal solution to the primal problem is close to the target $\vec{F}^H$, we expect the optimal solution to the dual problem to reside in a neighborhood close to zero. As a result, Newton's method should converge quadratically to the solution with a step size equal to one. Therefore, we ignore the feasibility constraint and always use a unit step size. Sometimes, this allows the dual solution to become slightly infeasible, but the amount of infeasibility tends to be small. In practice, the corresponding primal solution is always feasible and produces good results. In order to allow infeasible solutions, we must use the original formulation of Newton's method rather than the symmetric reformulation. Namely, the operator D becomes ill-defined for infeasible points.

When we combine the above pieces, we arrive at the final algorithm.

Algorithm 1: Dual algorithm for the solution of the remap subproblem

1. Define $H \in \mathbb{R}^{2K\times 2K}$ and $b \in \mathbb{R}^{2K}$ as

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}\mathbf{A}^\mathsf{T} & -\mathbf{A}\mathbf{A}^\mathsf{T} \\ -\mathbf{A}\mathbf{A}^\mathsf{T} & \mathbf{A}\mathbf{A}^\mathsf{T} \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} \mathbf{A}\vec{F}^H - \vec{b}_{\min} \\ -\mathbf{A}\vec{F}^H + \vec{b}_{\max} \end{bmatrix}.$$

2. Initialize $\vec{x} = \vec{0}$.
3. Until $\|\mathrm{Diag}(v(\vec{x}))(\mathbf{H}\vec{x} + \vec{b})\|$ becomes small or we exceed a fixed number of iterations.
   (a) When feasible, solve

$$(K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + D(\vec{x})\mathbf{H}D(\vec{x}))\vec{q} = -D(\vec{x})(\mathbf{H}\vec{x} + \vec{b})$$

   and set $\vec{p} = D(x)\vec{q}$. Otherwise, solve

$$(K(\vec{x})\mathrm{Diag}(\mathbf{H}\vec{x} + \vec{b}) + \mathrm{Diag}(v(\vec{x}))\mathbf{H})\vec{p} = -\mathrm{Diag}(v(\vec{x}))(\mathbf{H}\vec{x} + \vec{b}).$$

   (b) Set $\vec{x} = \vec{x} + \vec{p}$.

References

1. Bochev, P., Ridzal, D., Young, D.: Optimization–based modeling with applications to transport. Part 1. Abstract formulation. In Lirkov, I., Margenov, S., Wasniewski, J., eds.: Proceedings of LSSC 2011. Springer Lecture Notes in Computer Science (Submitted 2011)
2. Ridzal, D., Bochev, P., Young, J., Peterson, K.: Optimization–based modeling with applications to transport. part 3. Implementation and computational studies. In Lirkov, I., Margenov, S., Wasniewski, J., eds.: Proceedings of LSSC 2011. Springer Lecture Notes in Computer Science (Submitted 2011)
3. Dukowicz, J.K., Baumgardner, J.R.: Incremental remapping as a transport/advection algorithm. Journal of Computational Physics **160**(1) (2000) 318 – 335
4. Bochev, P., Ridzal, D., Scovazzi, G., Shashkov, M.: Formulation, analysis and numerical study of an optimization-based conservative interpolation (remap) of scalar fields for arbitrary lagrangian-eulerian methods. Journal of Computational Physics **Submitted** (2011)
5. Coleman, T.F., Li, Y.: A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM Journal on Optimization **6**(4) (1996) 1040–1058