Open access • Journal Article • DOI:10.1137/070681867

Iterative Solution of Piecewise Linear Systems — Source link

Luigi Brugnano, Vincenzo Casulli

Institutions: University of Florence, University of Trento

Published on: 01 Nov 2007 - SIAM Journal on Scientific Computing (Society for Industrial and Applied Mathematics)

Topics: Piecewise, Piecewise linear function, Coefficient matrix, Linear system and Numerical linear algebra

ITERATIVE SOLUTION OF PIECEWISE LINEAR SYSTEMS

LUIGI BRUGNANO* AND VINCENZO CASULLI†

Abstract. The correct formulation of numerical models for free-surface hydrodynamics often requires the solution of special linear systems whose coefficient matrix is a piecewise constant function of the solution itself. In so doing one may prevent the development of unrealistic negative water depths. The resulting *piecewise linear systems* are equivalent to particular linear complementarity problems whose solution could be obtained by using, for example, interior point methods. These methods may have a favorable convergence property but they are purely iterative and convergence to the exact solution is proven only in the limit of an infinite number of iterations. In the present paper a simple Newton-type procedure for certain piecewise linear systems is derived and discussed. This procedure is shown to have a finite termination property, *i.e.*, it converges to the exact solution in a finite number of steps and, actually, it converges very quickly, as confirmed by a few numerical tests.

Key words. piecewise linear systems, linear complementarity problems, Newton-type methods, free-surface hydrodynamics, wetting and drying.

AMS subject classifications. 90C33, 90C53, 90C06, 76M20.

1. Introduction. In numerical simulation of *free-surface hydrodynamics*, semi-implicit methods for the time integration of the governing partial differential equations are often used (see, e.g., [3, 4]). At every time step, these methods require the solution of a large, sparse, symmetric and positive definite linear system whose solution identifies the location of the water surface elevation at the grid points of a chosen mesh.

When *wetting and drying* is being simulated, the resulting water depth may become negative with the consequence that mass conservation is compromised. In Reference [19] this problem has been carefully investigated and a time step limitation has been derived to prevent the development of negative water depths.

Alternatively, as shall be seen later in our second test case, a correct formulation of numerical methods for free-surface hydrodynamics, that guarantees nonnegative water depths for any time step, requires the solution of large systems that can be written in the following form

$$\max\{\mathbf{0}, \mathbf{x}\} + T\mathbf{x} = \mathbf{b}, \tag{1.1}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad \max\{\mathbf{0}, \mathbf{x}\} = \begin{pmatrix} \max\{0, x_1\} \\ \vdots \\ \max\{0, x_n\} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

$\mathbf{b}$ is known and T is an irreducible, symmetric, and (at least) positive semidefinite matrix of size $n \times n$ satisfying either one of the following properties:

*Dipartimento di Matematica "U. Dini", Università di Firenze, Viale Morgagni 67/A, 50134 Firenze, Italy (brugnano@math.unifi.it).

†Laboratorio di Matematica Applicata, Dipartimento di Ingegneria Civile e Ambientale, Università di Trento, Via Mesiano 77, 38050 Trento, Italy, (vincenzo.casulli@unitn.it).

T1 : T is a Stieltjes matrix, i.e., a symmetric M-matrix (see, *e.g.*, [9]), or
T2 : $\mathtt{null}(T) \equiv \mathtt{span}(\mathbf{v})$ with $\mathbf{v} > 0$ (componentwise), and $T + D$ is a Stieltjes matrix for all diagonal matrices D such that

$$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}, \qquad \sum_{i=1}^{n} d_i > 0, \qquad d_i \geq 0, \quad i = 1, 2, \ldots, n.$$

The efficient solution of system (1.1) is also of interest in numerical optimization because this system can be cast as a *linear complementarity problem* (see, *e.g.*, [5, 12]). In fact, by setting $\mathbf{s} = \max\{\mathbf{0}, \mathbf{x}\}$ and $\mathbf{y} = \mathbf{s} - \mathbf{x}$, system (1.1) can be formulated either as a horizontal linear complementarity problem

$$(I + T)\mathbf{s} = T\mathbf{y} + \mathbf{b}, \qquad \mathbf{s}^\top \mathbf{y} = 0, \qquad \mathbf{s}, \mathbf{y} \geq 0, \tag{1.2}$$

or, equivalently, as a standard linear complementarity problem

$$\mathbf{s} = M\mathbf{y} + \mathbf{q}, \qquad \mathbf{s}^\top \mathbf{y} = 0, \qquad \mathbf{s}, \mathbf{y} \geq 0, \tag{1.3}$$

where $\mathbf{q} = (I + T)^{-1}\mathbf{b}$ and $M = (I + T)^{-1}T$ is a symmetric, positive semidefinite matrix (in this case the resulting linear complementarity problem is also said to be *monotone* [2, 20]).

When the size of a linear complementarity problem is reasonably small, it can be solved by means of a (direct) pivoting method (see, e.g., [5, 6, 11]). For large and sparse problems, however, these methods suffer from unacceptable roundoff error accumulation and excessive storage requirement.

Linear complementarity problems can also be solved by iterative schemes such as interior-point type methods (see, e.g., [13, 14, 15, 20]). These methods are characterized by having a convergence which is only asymptotic, thus the exact solution is obtained only in the limit of an infinite number of iterations.

Alternatively, linear as well as nonlinear complementarity problems can be solved by means of nonsmooth/semismooth Newton methods (see, e.g., [7, 16, 17] and the numerous references contained therein). Among others, an interesting algorithm based on the inexact Newton method that applies to large-scale standard linear complementarity problems (1.3) has been investigated in Reference [10]. Here, the matrix M was restricted to be an M-matrix, thus nonsingular.

In the next section an efficient semi-iterative procedure for solving directly system (1.1) will be derived and its convergence in a *finite* number of iterations will be established. In Section 3, some numerical tests are provided to confirm the excellent convergence properties of the proposed algorithm. Finally, in Section 4, a few concluding remarks are given.

2. The Newton-type iteration. Some introductory results are stated first in order to derive an efficient iterative procedure for solving system (1.1) and prove its finite termination.

The following two results are rather straightforward and their proofs have been omitted.

LEMMA 2.1. *Let matrix T in system (1.1) satisfy either* **T1** *or* **T2**. *If T satisfies:*

- **T1** *then* $T^{-1} > 0$*;*
- **T2** *then* $(T + D)^{-1} > 0$.

LEMMA 2.2. *System (1.1) can also be written in the following equivalent form*

$$[P(\mathbf{x}) + T]\,\mathbf{x} = \mathbf{b}, \tag{2.1}$$

where $P(\mathbf{x})$ *is a diagonal matrix whose diagonal entries, for* $i = 1, 2, \ldots, n$*, are piecewise constant functions defined as*

$$p(x_i) = \begin{cases} 1 & \text{if } x_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

Because of the characterization (2.2) of system (2.1), this will be said to be a ***piecewise linear system***.

It is to be noted that the left-hand side of system (2.1) is not everywhere differentiable. Nevertheless, a *Newton-type method* for solving system (2.1) is taken to be

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \left(P^k + T\right)^{-1} \left[\left(P^k + T\right)\mathbf{x}^k - \mathbf{b}\right],$$

which simplifies to the following Picard iteration,

$$\left(P^k + T\right)\mathbf{x}^{k+1} = \mathbf{b}, \qquad k = 0, 1, 2, \ldots, \tag{2.3}$$

where the upper index k denotes the iteration step and $P^k = P(\mathbf{x}^k)$.

Note that (2.3) can be directly derived from (2.1) as a fixed point iteration.

THEOREM 2.3. *Let matrix* T *in system (2.1) satisfy either* **T1** *or* **T2**. *If* T *satisfies* **T2** *assume also that* $P^0 \neq 0$ *and* $\mathbf{v}^\top \mathbf{b} > 0$. *Then* $P^k + T$ *is a Stieltjes matrix and the iterations (2.3) are well defined for all* $k \geq 0$.

Proof. Since P^k is a nonnegative diagonal matrix, if T satisfies **T1** then $P^k + T$ is a Stieltjes matrix and hence the iterates (2.3) are well defined.

If T satisfies **T2** and $P^0 \neq 0$, then $P^0 + T$ is a Stieltjes matrix. Next, by induction, one assumes that for $k \geq 1$ one has $P^{k-1} \neq 0$. Therefore, the vector $\mathbf{x}^k$, satisfying

$$\left(P^{k-1} + T\right)\mathbf{x}^k = \mathbf{b},$$

is well defined. Then, since $\mathbf{v}^\top \mathbf{b} > 0$, one has

$$\mathbf{v}^\top \left(P^{k-1} + T\right)\mathbf{x}^k = \mathbf{v}^\top P^{k-1}\mathbf{x}^k = \mathbf{v}^T \mathbf{b} > 0.$$

This implies that at least one entry of $\mathbf{x}^k$ is strictly positive. Consequently, $P^k \neq 0$, $P^k + T$ is a Stieltjes matrix, and $\mathbf{x}^{k+1}$ is well defined. ∎

REMARK 1. *In practice, the determination of* $\mathbf{x}^{k+1}$ *from (2.3) can be accomplished quite efficiently by using a preconditioned conjugate gradient method (see, e.g., [8, 18]). This is particularly the case in applications where* T *is a sparse and very large matrix. To this purpose, in light of the convergence property that will be demonstrated*

later, $\mathbf{x}^k$ *is conveniently used as a starting point for the conjugate gradient method (the effectiveness of this choice has been confirmed by several numerical tests).*

The iteration (2.3) allows a very simple stopping criterion as provided by the following lemma.

LEMMA 2.4. *Under the assumptions of Theorem 2.3, if for some* $K \geq 0$ *one gets* $P^{K+1} = P^K$*, then* $\mathbf{x} = \mathbf{x}^{K+1}$ *is an exact solution of problem (2.1)-(2.2).*

Proof. Since $P^{K+1} = P^K$ one has

$$\left(P^K + T\right)\mathbf{x}^{K+1} = \left(P^{K+1} + T\right)\mathbf{x}^{K+1} = \mathbf{b}.$$

The thesis then follows from Lemma 2.2. □

In addition to the previous results, the iteration (2.3) is characterized by a remarkable finite termination property as shown by the following theorem.

THEOREM 2.5. *Let matrix* T *in system (2.1) satisfy either* **T1** *or* **T2**. *If* T *satisfies* **T2** *assume also that* $P^0 \neq 0$ *and* $\mathbf{v}^\top \mathbf{b} > 0$. *Then the iterations (2.3) converge to an exact solution of problem (2.1)-(2.2) in at most* $n+1$ *iterations.*

Proof. The iterative scheme (2.3) implies the following equality

$$\left(P^k + T\right)\mathbf{x}^{k+1} = \left(P^{k-1} + T\right)\mathbf{x}^k = \mathbf{b}, \quad k = 1, 2, \ldots$$

from which it follows that

$$\left(P^k + T\right)\mathbf{x}^{k+1} = \left(P^k + T\right)\mathbf{x}^k - \boldsymbol{\xi}^k \tag{2.4}$$

where $\boldsymbol{\xi}^k \equiv \left(P^k - P^{k-1}\right)\mathbf{x}^k \geq 0$. In fact, by denoting hereafter by p_i^k the ith diagonal entry of P^k, one has

$$p_i^k - p_i^{k-1} \neq 0 \quad \Rightarrow \quad \begin{cases} p_i^k = 1 \quad \text{and} \quad p_i^{k-1} = 0 \quad \Rightarrow \quad x_i^k > 0, \\ \qquad\qquad \text{or} \\ p_i^k = 0 \quad \text{and} \quad p_i^{k-1} = 1 \quad \Rightarrow \quad x_i^k \leq 0. \end{cases}$$

Now, since $\left(P^k + T\right)^{-1} > 0$ and $\boldsymbol{\xi}^k \geq 0$, equation (2.4) implies $\mathbf{x}^{k+1} \leq \mathbf{x}^k$ and, consequently, $P^{k+1} \leq P^k$ for all $k = 1, 2, \ldots$.

Finally, from Lemma 2.4, it follows that if $P^{k+1} = P^k$ then $\mathbf{x}^{k+1}$ is an exact solution of system (2.1). Conversely, one obtains $P^{k+1} \neq P^k$ and, since $P^{k+1} \geq 0$, this may occur at most $n - m + 1$ times where (see (2.2))

$$m = \sum_{i=1}^{n} p(x_i).$$

□

REMARK 2. *Note that* $P^{k+1} \neq P^k$ *would occur* $n+1$ *times only in the hypothetical case that* $m = 0$ *and* $\sum_{i=1}^n p_i^k = n - k + 1, k = 1, 2, \ldots, n$. *In practice, several test cases have shown that convergence to the exact solution can be obtained in just a few iterations. Of course, if one can guess* $P^0 = P(\mathbf{x})$ *then convergence to the exact solution is obtained in just one step.*

Note that the diagonal of P^k can be considered as a binary string. The number of ones never increases when the iteration starts with all ones. In other words, the Hamming distances to the string corresponding to the exact solution never increases (except possibly in the first step for certain initial guesses). Consequently, the iteration (2.3) can be considered as an iteration on points of a finite lattice, as opposed to $\mathbb{R}^n$.

THEOREM 2.6. *Let matrix T in system (2.1) satisfy either* **T1** *or* **T2**. *If T satisfies* **T2** *assume also that $\mathbf{v}^\top \mathbf{b} > 0$. Then the solution of problem (2.1)-(2.2) exists and is unique.*

Proof. The existence of a solution has been established constructively by the previous Theorem 2.5. Regarding its uniqueness, note first that, with reference to (2.2), for any two vectors $\mathbf{x}$ and $\mathbf{y}$ one has

$$P(\mathbf{x})\mathbf{x} - P(\mathbf{y})\mathbf{y} = Q\,(\mathbf{x} - \mathbf{y}),$$

where Q is a suitable diagonal matrix whose diagonal entries q_i satisfy the inequalities $0 \le q_i \le 1$, $i = 1, 2, \ldots, n$. In fact, either one of the following four cases occurs:

1. $x_i, y_i > 0 \quad \Rightarrow \quad p(x_i) = p(y_i) = 1 \quad \Rightarrow \quad q_i = 1;$
2. $x_i, y_i \le 0 \quad \Rightarrow \quad p(x_i) = p(y_i) = 0 \quad \Rightarrow \quad q_i = 0;$
3. $x_i > 0 \ge y_i \quad \Rightarrow \quad p(x_i) = 1,\ p(y_i) = 0 \quad \Rightarrow \quad 0 < q_i \le 1;$
4. $x_i \le 0 < y_i \quad \Rightarrow \quad p(x_i) = 0,\ p(y_i) = 1 \quad \Rightarrow \quad 0 < q_i \le 1.$

Assume now that $\mathbf{x}$ and $\mathbf{y}$ are both solutions of system (2.1) so that

$$[P(\mathbf{x}) + T]\,\mathbf{x} = \mathbf{b}, \qquad [P(\mathbf{y}) + T]\,\mathbf{y} = \mathbf{b}.$$

Thus,

$$[P(\mathbf{x}) + T]\,\mathbf{x} - [P(\mathbf{y}) + T]\,\mathbf{y} = (Q + T)\,(\mathbf{x} - \mathbf{y}) = \mathbf{0}. \tag{2.5}$$

Therefore, if T satisfies **T1**, then $Q+T$ is certainly a Stieltjes matrix and hence $\mathbf{x} = \mathbf{y}$. When T satisfies **T2**, since $\mathbf{v}^\top \mathbf{b} > 0$, one has

$$\mathbf{v}^\top\,[P(\mathbf{x}) + T]\,\mathbf{x} = \mathbf{v}^\top P(\mathbf{x})\mathbf{x} = \mathbf{v}^\top \mathbf{b} > 0,$$

$$\mathbf{v}^\top\,[P(\mathbf{y}) + T]\,\mathbf{y} = \mathbf{v}^\top P(\mathbf{y})\mathbf{y} = \mathbf{v}^\top \mathbf{b} > 0.$$

Consequently, $P(\mathbf{x}) \neq 0$, $P(\mathbf{y}) \neq 0$ and hence at least one of the diagonal entries of Q is strictly positive. Thus, $Q + T$ is a Stieltjes matrix and uniqueness ($\mathbf{x} = \mathbf{y}$) follows directly from (2.5). ∎

Although it may not be as interesting in practical applications, the next result completes the framework in the case when T satisfies **T2**.

COROLLARY 2.7. *Consider problem (2.1)-(2.2) where matrix T satisfies* **T2**. *Then:*

- *if $\mathbf{v}^\top \mathbf{b} = 0$, then a solution exists but is not unique;*
- *if $\mathbf{v}^\top \mathbf{b} < 0$, then the problem has no solution.*

Proof. Assume that $\mathbf{v}^\top \mathbf{b} = 0$. In this case $\mathbf{b}$ belongs to the range of T, thus a vector $\mathbf{u}$ exists such that $T\mathbf{u} = \mathbf{b}$. The thesis follows by observing that, if u_i and v_i denote the ith entries of the vectors $\mathbf{u}$ and $\mathbf{v}$, respectively, then for all $\alpha \geq \max_i u_i/v_i$ the vector

$$\mathbf{x}(\alpha) = \mathbf{u} - \alpha \mathbf{v}$$

satisfies

$$\mathbf{x}(\alpha) \leq 0, \qquad T\mathbf{x}(\alpha) = \mathbf{b}.$$

Consequently, $\mathbf{x}(\alpha)$ is solution of problem (2.1)-(2.2).

Assume now that $\mathbf{v}^\top \mathbf{b} < 0$. If a solution $\mathbf{x}$ would exist, then from (2.1) one has

$$\mathbf{v}^\top \left[P(\mathbf{x}) + T\right] \mathbf{x} = \mathbf{v}^\top P(\mathbf{x})\mathbf{x} = \mathbf{v}^\top \mathbf{b} < 0,$$

which is impossible because $\mathbf{v} > 0$ and $P(\mathbf{x})\mathbf{x} \geq 0$ for all $\mathbf{x}$. ∎

3. Numerical tests. Consider problem (2.1)-(2.2) where T is defined as

$$T = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}_{n\times n}.$$

Clearly, in this case T satisfies **T1**. Thus, given $\mathbf{b}$, problem (2.1)-(2.2) has a unique solution and convergence of the iterative scheme (2.3) is assured for any choice of the initial guess. The right-hand side in equation (2.1) is chosen in such a way that the exact solution is given by

$$x_i = e^{6\frac{i-1}{n-1}-5} - 1, \qquad i = 1, 2, \ldots, n.$$

The initial guess is taken to be $\mathbf{x}^0 \equiv \mathbf{1}$, so that $P^0 = I_n$. In Table 3.1 the required iterations (K) are listed for increasing values of n along with the corresponding Hamming distance (δ_k) between two subsequent iterations

$$\delta_k = \sum_{i=1}^{n} p_i^{k-1} - p_i^k \geq 0, \qquad k = 1, \ldots, K.$$

Due to the choice of $\mathbf{x}^0$, the monotony property $P^{k+1} \leq P^k$ holds for all $k \geq 0$. Accordingly, the Hamming distance of the initial guess P^0 from $P(\mathbf{x})$ is given by

$$\Delta \equiv \sum_{i=1}^{n} p_i^0 - p(x_i) = \sum_{k=1}^{K} \delta_k.$$

TABLE 3.1
Required iterations and Hamming distances.

n	K	δ_1	δ_2	δ_3	δ_4	δ_5	Δ
1000	5	828	3	1	1	0	833
2000	4	1661	3	2	0	–	1666
3000	5	2494	3	2	1	0	2500
4000	5	3327	3	2	1	0	3333
5000	4	4160	4	2	0	–	4166
6000	5	4993	4	2	1	0	5000
7000	5	5826	4	2	1	0	5833
8000	4	6660	4	2	0	–	6666
9000	5	7493	4	2	1	0	7500
10000	5	8326	4	2	1	0	8333

As indicated in Table 3.1, convergence to the exact solution is obtained in at most 5 iterations for all n.

The second test problem is derived from the mathematical modelling of a two-dimensional flow in a homogeneous phreatic aquifer. The governing differential equation, often called the Boussinesq equation (see Reference [1] for details), is given by

$$\varepsilon\eta_t = [\kappa(h+\eta)\eta_x]_x + [\kappa(h+\eta)\eta_y]_y + \varphi, \quad (x,y) \in \Omega(t), \quad t > 0, \tag{3.1}$$

where x and y are coordinates in a horizontal reference frame; t is the time; ε and κ are the *porosity* and the *hydraulic conductivity*, respectively; $h(x,y)$ is the prescribed aquifer's bottom and $\eta(x,y,t)$ is the unknown free-surface elevation (see Figure 3.1). Thus,

$$H(x,y,t) = h(x,y) + \eta(x,y,t), \quad (x,y) \in \Omega(t) \tag{3.2}$$

represents the aquifer thickness measuring the distance of the phreatic surface from the bottom. The time dependent domain is $\Omega(t) = \{(x,y) : H(x,y,t) > 0\}$. Obviously, one can assume $H(x,y,t) = 0$ for $(x,y) \notin \Omega(t)$. Finally, $\varphi(x,y,t)$ represents the prescribed source or sink.

In the present test the aquifer's bottom is described by a paraboloid of revolution given by

$$h(x,y) = h_0\left(1 - \frac{x^2+y^2}{L^2}\right),$$

where h_0 and L are given positive constants. Here, for simplicity only, porosity and permeability are assumed to be constants. With an initially flat phreatic surface $\eta(x,y,0) = 0$ one has $\Omega(0) = \{(x,y) : x^2 + y^2 < L^2\}$.

A square of side $2L$, centered at the origin and containing $\Omega(0)$ is covered by a grid having size $\Delta x = \Delta y = L/N$. Then, a consistent semi-implicit finite difference discretization of Equation (3.1) is taken to be (see, *e.g.*, [3, 4])

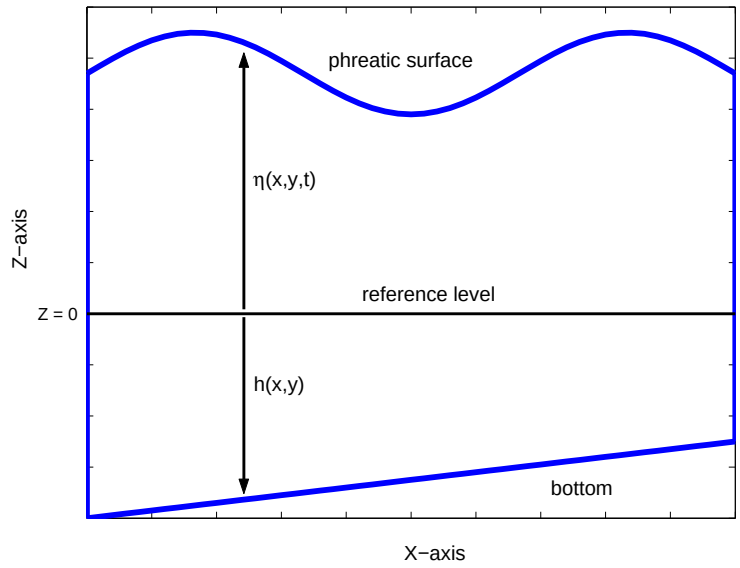

FIG. 3.1. *$h(x,y)$ and $\eta(x,y,t)$ on a vertical cross section.*

$$\varepsilon\frac{H_{ij}^{\ell+1}-H_{ij}^{\ell}}{\Delta t}=\kappa\frac{H_{i+\frac{1}{2},j}^{\ell}(\eta_{i+1,j}^{\ell+1}-\eta_{ij}^{\ell+1})-H_{i-\frac{1}{2},j}^{\ell}(\eta_{ij}^{\ell+1}-\eta_{i-1,j}^{\ell+1})}{\Delta x^2}$$
$$+\kappa\frac{H_{i,j+\frac{1}{2}}^{\ell}(\eta_{i,j+1}^{\ell+1}-\eta_{ij}^{\ell+1})-H_{i,j-\frac{1}{2}}^{\ell}(\eta_{ij}^{\ell+1}-\eta_{i,j-1}^{\ell+1})}{\Delta y^2}+\varphi_{ij}^{\ell},$$
$$i,j=-N,-N+1,\ldots,N, \tag{3.3}$$

where Δt denotes the time step size and η_{ij}^{ℓ} and H_{ij}^{ℓ} are the discrete free-surface elevation and the aquifer's thickness at $t_\ell=\ell\Delta t$, respectively. Here ℓ denotes the time level not to be confused with the iteration index k. Finally, between grid points, the aquifer thicknesses $H_{i\pm1/2,j}^{\ell}$ and $H_{i,j\pm1/2}^{\ell}$ are defined as averages from the nearest grid values.

In order to avoid the development of un-physical negative values of the aquifer's thickness, $H_{ij}^{\ell+1}$ in (3.3) is defined as

$$H_{ij}^{\ell+1}=\max\{0,h_{ij}+\eta_{ij}^{\ell+1}\},$$

which is consistent with (3.2) in the wet area, and gives $H_{ij}^{\ell+1}=0$ in the dry area. It is to be noted that for those grid points (i,j) where $H_{i\pm1/2,j}^{\ell}=0$ and $H_{i,j\pm1/2}^{\ell}=0$, equation (3.3) trivially implies $H_{i,j}^{\ell+1}=H_{i,j}^{\ell}$. In this case equation (3.3) does not contribute to the system that is being formulated.

The remaining set of equations can be assembled into a single *piecewise linear system*. In fact, upon multiplication of each term in (3.3) by $\Delta t/\varepsilon$, this system (which has to be solved *at every time step*) can be written in the form (1.1). The resulting matrix T is irreducible, sparse, symmetric, positive semidefinite, and of time dependent size $n_\ell\times n_\ell$, with n_ℓ being the number of grid points (i,j) where at least

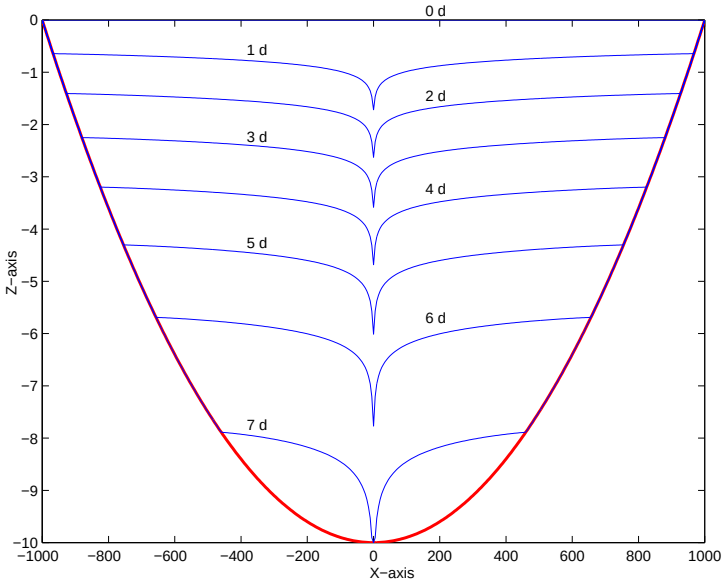

FIG. 3.2. *Computed free surfaces at the cross section* $y = 0$.

one of $H^{\ell}_{i\pm1/2,j}$ and $H^{\ell}_{i,j\pm1/2}$ is strictly positive. This matrix satisfies **T2** and its null space is spanned by the vector $\mathbf{v} = (1, \ldots, 1)^{\top} \in \mathbb{R}^{n_\ell}$. Thus, according to Theorems 2.3–2.6, the sign of $\mathbf{v}^{\top}\mathbf{b}$ and the choice of the initial guess are crucial for the existence and uniqueness of the solution $H^{\ell+1}_{ij}$ and to obtain convergence of the iterations (2.3).

For the present simulations the chosen parameters are $\varepsilon = 0.4$, $\kappa = 1\,m/s$, $h_0 = 10\,m$ and $L = 10^3\,m$. The flow is then driven by an idealized pointwise sink, located at the origin, that pumps water out of the aquifer at a constant rate $q = 10\,m^3/s$. Thus, by setting $\varphi^{\ell}_{ij} = 0$, except $\varphi^{\ell}_{00} = -\frac{q}{\Delta x \Delta y}$, the expected domain $\Omega(t)$ is the area confined by concentric circles of decreasing radius and the new water volume at time $t_{\ell+1}$ is given by

$$V^{\ell+1} = \varepsilon \Delta x \Delta y \sum_{ij} H^{\ell+1}_{ij} = V^{\ell} - q\Delta t.$$

On the other hand, since $V^{\ell+1} = \varepsilon \Delta x \Delta y\, \mathbf{v}^{\top}\mathbf{b}$, the inequality $V^{\ell+1} > 0$ represents a necessary condition for the existence and the uniqueness of $H^{\ell+1}_{ij}$.

A numerical simulation has been carried out for seven *days* using a relatively large time step size $\Delta t = 1$ *day*. As initial guess for $H^{\ell+1}_{ij}$ the values of the water depths have been taken from the previous time step so that the condition $P^0 \neq 0$, required by Theorems 2.3 and 2.5, is certainly satisfied provided that $V^{\ell} > 0$. The linear systems defined by each iteration (2.3) have been solved by the conjugate gradient method after diagonal scaling and red/black reordering. Figure 3.2 shows the resulting free-surface elevation η^{ℓ}_{i0} at the cross section $y = 0$, for $\ell = 0, 1, \ldots, 7$.

For specified $N = 50$, 100, and 200, Table 3.2 shows the size of the resulting *piecewise linear system*, the required iterations and the computed water volume at each time step. As expected by a Newton-type scheme, the number of iterations turns out to be remarkably small and insensitive to grid resolution, thus confirming the usefulness of the proposed algorithm for real world applications.

TABLE 3.2
System size, required iterations and computed water volume.

		$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=5$	$\ell=6$	$\ell=7$
N=50	n_ℓ	8109	7629	7025	6345	5605	4701	3577
	K^ℓ	3	3	3	3	3	3	4
	V^ℓ	5419110	4555110	3691110	2827110	1963110	1099110	235110
N=100	n_ℓ	31965	29925	27549	24845	21853	18333	13905
	K^ℓ	3	3	3	3	3	3	4
	V^ℓ	5419173	4555173	3691173	2827173	1963173	1099173	235173
N=200	n_ℓ	126741	118693	109085	98369	86393	72449	54933
	K^ℓ	3	3	4	3	3	4	5
	V^ℓ	5419182	4555182	3691182	2827182	1963182	1099182	235182

Table 3.2 also shows that for all $\ell = 1, 2, \ldots, 7$ the water volumes are strictly positive and linearly decreasing at a constant rate q. In fact, the volume difference between two subsequent time levels is correctly given by $V^{\ell+1} - V^\ell = -864,000\, m^3$. Thus, any attempt to extend the simulation beyond *day* 7 would produce a *physically unrealistic* negative water volume $V^8 < 0$ implying $\mathbf{v}^\top \mathbf{b} < 0$. Consequently, when $\ell = 7$ Corollary 2.7 applies indicating that problem (3.3) does not have a solution. Moreover, the assumption $\mathbf{v}^\top \mathbf{b} > 0$ required by Theorems 2.3 and 2.5 is violated and the iterative scheme (2.3) would fail to converge in this case. This is an interesting example demonstrating that the proposed algorithm does not permit artificial overdrainage.

4. Conclusions. A simple Newton-type iterative procedure for solving certain *piecewise linear systems* that arise from numerical modelling of free-surface hydrodynamics has been derived and investigated.

It is shown that, under rather general assumptions, the iterates are well defined and converge to the exact solution of the given system in a finite number of steps.

Existence and uniqueness of the solution has been established under the same assumptions for which convergence is assured.

The present algorithm efficiently applies to very large systems that are often encountered in other applications governed by partial differential equations and in optimization.

Simple, and yet nontrivial numerical tests have confirmed the efficiency, the robustness and the usefulness of the proposed algorithm.

REFERENCES

[1] J. BEAR AND A. VERRUIJT, *Modeling Groundwater Flow and Pollution*, D. Reidel Publ. Co., Dordrecht, Holland, 1987.

[2] S. C. BILLUPS AND K. G. MURTY, *Complementarity problems*, Jour. Comput. Appl. Math., 124 (2000), pp. 303–318.

[3] V. CASULLI, *Semi-implicit finite difference methods for the two-dimensional shallow water equations*, Jour. of Computational Physics, 86 (1990), pp. 56–74.

[4] V. CASULLI AND P. ZANOLLI, *Semi-implicit modeling of nonhydrostatic free-surface flows for environmental problems*, Mathematical and Computer Modelling, 36 (2002), pp. 1131–1149.

[5] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, Academic Press, San Diego, CA, 1992.

[6] B. C. EAVES, *The linear complementarity problem*, Management Science, 17 (1971), pp. 612–634.

[7] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementary problems*, Mathematical Programming, 75 (1996), pp. 407–439.
[8] G. H. Golub and C. F. van Loan, *Matrix Computations,* 3rd *ed.*, The Johns Hopkins University Press, Baltimore, MD, 1996.
[9] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, New York, NY, 1991.
[10] C. Kanzow, *Inexact semismooth Newton methods for large-scale complementary problems*, Optimization Methods and Software, 19 (2004), pp. 309–325.
[11] C. E. Lemke and J. T. Howson, *Equilibrium points of bimatrix games*, SIAM J. Appl. Math., 12 (1964), pp. 413–423.
[12] K. G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann Verlag, Berlin, 1988 (`http://ioe.engin.umich.edu/people/fac/books/murty/linear_complementarity_webbook/`).
[13] F. A. Potra, *A superlinearly convergent predictor-corrector method for degenerate LCP in a wide neighborhood of the central path with $O(\sqrt{n}L)$-iteration complexity*, Mathematical Programming, Ser. A, 100 (2004), pp. 317–337.
[14] F. A. Potra and X. Liu, *Corrector-predictor methods for sufficient linear complementarity problems in a wide neighborhood of the central path*, SIAM Journal on Optimization, 17 (2006), pp. 871–890.
[15] F. A. Potra and S. J. Wright, *Interior point methods*, Jour. Comput. Appl. Math., 124, (2000), pp. 281–302.
[16] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.
[17] L. Qi and J. Sun, *A nonsmooth version of Newton method*, Mathematical Programming, 58 (1993), pp. 353–367.
[18] Y. Saad, *Iterative Methods for Sparse Linear Systems,* 2nd *ed.*, SIAM, Philadelphia, PA, 2003.
[19] G. S. Stelling and S. P. A. Duynmeyer, *A staggered conservative scheme for every Froude number in rapidly varied shallow water flows*, Int. J. for Numerical Methods in Fluids, 43 (2003), pp. 1329–1354.
[20] S. J. Wright, *Primal-Dual Interior Point Methods*, SIAM, Philadelphia, PA, 1997.