

APPROXIMATION OF PARABOLIC EQUATIONS USING THE WASSERSTEIN METRIC

DAVID KINDERLEHRER¹ AND NOEL J. WALKINGTON²

Abstract. We illustrate how some interesting new variational principles can be used for the numerical approximation of solutions to certain (possibly degenerate) parabolic partial differential equations. One remarkable feature of the algorithms presented here is that derivatives do not enter into the variational principles, so, for example, discontinuous approximations may be used for approximating the heat equation. We present formulae for computing a Wasserstein metric which enters into the variational formulations.

AMS Subject Classification. 65M60, 49R10.

Received: November 3, 1998.

1. INTRODUCTION

New variational principles have opened the possibility of realizing dynamical processes as gradient flows in the weak topology [2,5,9,10,12]. Our focus here is their implementation for the numerical approximation of their governing equations, generally parabolic of convection-diffusion type. Confining ourselves to a one dimensional setting, such a problem has the form

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x} (K(u)_x + \Psi' u) = f \quad \text{in } \Omega = (0, 1) \quad (1.1)$$

subject to the initial and boundary conditions

$$u|_{t=0} = u_0 \geq 0 \quad \text{and} \quad K(u)_x + \Psi' u = 0 \quad \text{on } \partial\Omega. \quad (1.2)$$

Here we suppose that K is monotone increasing, Ψ is a smooth potential, and $\Omega = (0, 1)$ is an interval. We allow K' to vanish on intervals, so equation (1.1) may be degenerate. It will turn out that for the cases we consider, $u \geq 0$ in $\Omega \times (0, \infty)$. Observe that this equation is not formally self adjoint; however, by working in the weak topology it can be realized as a gradient flow.

When we think of evolution, especially when we have in hand a solution that is smooth, we often neglect to recognize that in saying states are near each other, we are imposing an environment for the motion of the system. Or in many situations, we may find solutions by means of weak topology methods but may, indeed,

Keywords and phrases. Wasserstein metric, parabolic equations, numerical approximations.

¹ Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by ARO DAAH Grant 04 96 0060, NSF Grant DMS-9505078. e-mail: davidk@andrew.cmu.edu

² Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by NSF Grant DMS-9504492.

strive to prove convergence in some stronger sense, a norm for example. Even in bifurcation phenomena, we are accustomed to imposing the strongest possible topology for the study of the equation. The environment of the motion does not play a significant role in our thinking. Different considerations may prevail when the system arises as a coarse grained description of a finer scale phenomenon. The averaging procedure that delivers the coarse grained system can determine the context of the dynamics [13]. Important equations that we study here arise in this fashion.

To highlight this issue, if u is a solution of (1.1) with $f = 0$, then

$$\frac{d}{dt} \int_{\Omega} (\Phi(u) + \Psi u) dx \leq 0, \quad \text{where} \quad u\Phi'(u) - \Phi(u) = K(u),$$

but we cannot realize u as the gradient flow of the functional above in a conventional sense. We can introduce a Wasserstein metric for this purpose. This metric gives the weak* topology on suitable bounded sets of measures and a main preoccupation in this work is to provide a convenient way to compute with it. Instead of the standard semi-discrete approximation of (1.1) obtained with the implicit Euler scheme, we introduce a new semi-discrete approximation proposed in [12] where, for a given time step τ , a sequence $\{u^n\}$ is determined iteratively by $u^0 = u_0$ and u^n is the solution of

$$\frac{1}{2}d(u, u^{n-1})^2 + \tau \int_{\Omega} (\Phi(u) + \Psi u) dx = \min \tag{1.3}$$

in a suitable class. $d(\cdot, \cdot)$ is the metric we shall introduce (the Wasserstein metric). More generally, we shall consider the semi-discrete scheme with the functional

$$\frac{1}{2}d(u, v)^2 + \tau \int_{\Omega} (\Phi(u) + \Psi u) dx = \min \quad \text{with} \quad v = u^{n-1} + \frac{1}{\tau} \int_{(n-1)\tau}^{n\tau} f dx. \tag{1.4}$$

When $K(u) = \epsilon u$, equation (1.1) is the classical convection diffusion equation or, when $f = 0$, the Fokker-Planck equation (forward Kolmogorov equation). When $\Psi = 0$ and K has a “flat spot” as indicated in Figure 1.1, (1.1) is the classical Stefan problem with u the energy and $\theta = K(u)$ the temperature. These two problems are challenging from a computational perspective. The convection-diffusion equation exhibits sharp transition layers and the Stefan problem has a discontinuous solution, corresponding to release or absorption of latent heat at the phase transition. This lack of regularity limits the accuracy of numerical algorithms, and the convective diffusion equation, for example, requires very fine meshes (or some form of up-winding) near the transition layers for stability.

Since solutions of (1.1) may exhibit discontinuities it would be natural to use discontinuous functions to approximate them. However, the need to compute spatial derivatives precludes this possibility. Indeed, approximating derivatives by differencing across sharp transition regions is one source of instabilities that plague various algorithms. One remarkable feature of the algorithms based on the functional (1.3) is that derivatives do not enter into the variational scheme so discontinuous approximations may be adopted.

We emphasize that the variational principles are currently quite specific; we only know how to treat (1.1) and (1.2). It is not clear how to deal with essential (Dirichlet) boundary data nor with situations where solutions may be negative. However, it is possible to allow the non-homogeneous term f to depend upon u , $f = f(u)$.

2. WASSERSTEIN METRIC

In this section we introduce the Wasserstein metric and describe several of its properties. Let $\mathcal{M}(\Omega)$ denote the Borel measures on Ω and

$$\mathcal{M}_M = \{\mu \in \mathcal{M}(\Omega) : \mu(\Omega) = M\}.$$

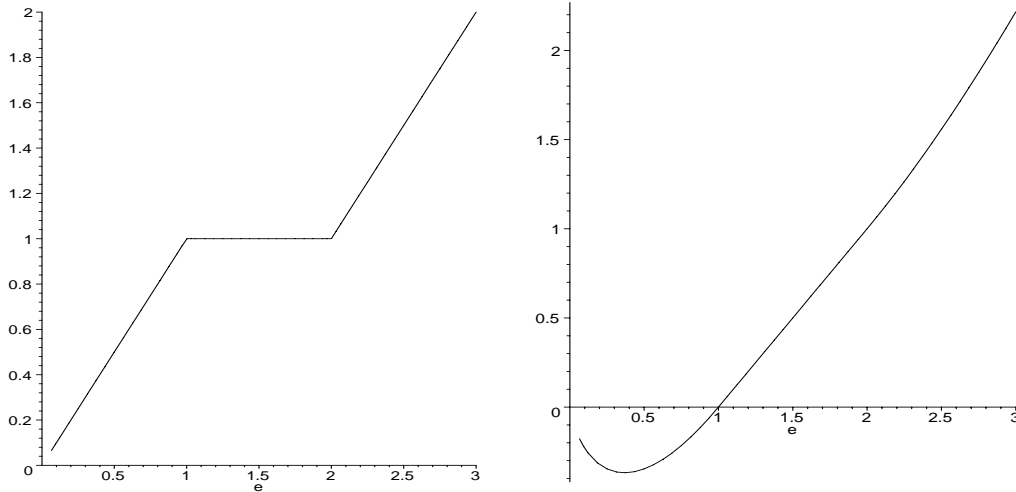


FIGURE 1.1. $K(u)$ and $\Phi(u)$ for the Stefan problem.

For $\mu, \nu \in \mathcal{M}_M$, we define the Wasserstein distance between them by

$$d(\mu, \nu)^2 = \inf_{p \in \mathcal{M}(\mu, \nu)} \iint_{\Omega \times \Omega} (x - y)^2 dp(x, y) \tag{2.1}$$

where $\mathcal{M}(\mu, \nu)$ is the set of all Borel measures p in $\mathcal{M}(\Omega \times \Omega)$ with marginals μ and ν . Recall that p has marginals μ and ν provided

$$\mu(A) = p(A \times \Omega) \text{ and } \nu(B) = p(\Omega \times B)$$

for all Borel sets $A, B \subset \Omega$.

The distance (2.1) is essentially a cost of transferring mass distributed by μ to mass distributed by ν . When $\mu = u dx$ and $\nu = v dx$ for functions $u, v \in L^1(\Omega)$, we write $d(u, v)$ for $d(\mu, \nu)$. It turns out that $d(\cdot, \cdot)$ is equivalent to a weak* metric on \mathcal{M}_M .

Frechet [6] recognized that the Wasserstein metric is realized by an increasing function, or more generally in arbitrary dimension, by the gradient of a convex function, see Brenier [3], Gangbo *et al.* [7, 8] and the exposition of Otto in [12]. This means that if $p \in \mathcal{M}(u dx, v dx)$ realizes the infimum in (2.1),

$$d(u, v)^2 = \iint_{\Omega \times \Omega} (x - y)^2 dp(x, y),$$

then there is an increasing function ϕ on Ω , satisfying $\phi(0) = 0$ and $\phi(1) = 1$, such that for all $\xi \in C_0^\infty(\Omega \times \Omega)$

$$\iint_{\Omega \times \Omega} \xi(x, y) dp(x, y) = \int_{\Omega} \xi(x, \phi(x))u(x) dx \tag{2.2}$$

and in particular

$$\int_{\phi(E)} v(y) dy = \int_E u(x) dx \quad \text{and} \quad d(u, v)^2 = \int_{\Omega} (x - \phi(x))^2 u(x) dx. \tag{2.3}$$

In fact, by density, there is a dense set of measures for which (2.2) holds for some function ϕ . It then follows by rearrangement that ϕ is increasing, see our Appendix. The limit of increasing functions is increasing, from which we conclude that (2.2) holds for all u and v .

Equation (2.3) allows us to identify ϕ . Setting $E = (0, x]$, $0 < x < 1$, and differentiating with respect to x gives

$$\phi'(x) = \frac{u(x)}{v \circ \phi(x)} \quad x \in \Omega \quad (2.4)$$

provided that $u, v > 0$. Thus if U and V are the distribution functions (indefinite integrals) of u and v respectively, we obtain the convenient formula

$$\phi(x) = V^{-1} \circ U(x) \quad (2.5)$$

and for the metric

$$d(u, v)^2 = \int_{\Omega} (x - V^{-1} \circ U(x))^2 u(x) dx \quad (2.6)$$

which, upon the change of variable $\eta = U(x)$, can be written in the more symmetric form

$$d(u, v)^2 = \int_0^M (U^{-1}(\eta) - V^{-1}(\eta))^2 d\eta.$$

Equations (2.5, 2.6) were stated by Frechet in [6].

Let us check that (2.5, 2.6) assure us that $d(\cdot, \cdot)$ generates the weak* topology on densities using our formulas.

Let $u_k dx, u dx \in \mathcal{M}_M$, and suppose initially that $u_k > \delta, u > \delta$ for some $\delta > 0$.

If $u_k dx$ converges to $u dx$, then the distribution functions U_k and U have the property that $U^k \rightarrow U$ uniformly in $\bar{\Omega}$, and by our positivity hypothesis, $(U_k)^{-1} \rightarrow U^{-1}$ uniformly. Then $\phi_k(x) = U_k^{-1} \circ U(x) \rightarrow x$ uniformly in $\bar{\Omega}$, and from (2.6), $d(u_k, u) \rightarrow 0$.

If u_k and u are not strictly positive, we may approximate them by a sequence that converges both in weak* topology and in the metric, *e.g.*, given $u \geq 0, u dx \in \mathcal{M}_M$,

$$u_{\epsilon} dx = \frac{M}{M + \epsilon} (u + \epsilon) dx \rightarrow^* u dx$$

and¹

$$dp_{\epsilon}(x, y) = \frac{1}{M + \epsilon} (M\delta(x - y)u(x) dx + \epsilon u(x) dx dy)$$

has marginals u and u_{ϵ} , so that

$$d(u, u_{\epsilon})^2 \leq \iint_{\Omega \times \Omega} (x - y)^2 dp_{\epsilon}(x, y) \leq \epsilon.$$

¹ $\delta()$ is the Dirac distribution.

Suppose on the other hand that $d(u_k, u) \rightarrow 0$. Then, without using our formulas, cf. [9, 12], if p_k is a joint distribution realizing the metric and ξ is a smooth function,

$$\begin{aligned} \left| \int_{\Omega} \xi u_k \, dy - \int_{\Omega} \xi u \, dx \right| &= \left| \iint_{\Omega \times \Omega} (\xi(y) - \xi(x)) \, dp_k(x, y) \right| \\ &\leq \sup |\xi'| \iint_{\Omega \times \Omega} |x - y| \, dp_k(x, y) \\ &\leq \sup |\xi'| \left(\iint_{\Omega \times \Omega} |x - y|^2 \, dp_k(x, y) \right)^{1/2} \left(\iint_{\Omega \times \Omega} dp_k(x, y) \right)^{1/2} \\ &\leq \sqrt{M} \sup |\xi'| d(u, u_k) \\ &\rightarrow 0. \end{aligned}$$

The result now follows since smooth functions are dense in continuous functions.

3. VARIATIONAL PRINCIPLES

We introduce the semi-discrete approximation where a sequence $\{u^n\}$ is determined iteratively by $u^0 = u_0$ and u^n is the minimizer of

$$I(u) = \frac{1}{2} d(u, v)^2 + \tau \int_{\Omega} (\Phi(u) + \Psi u) \, dx, \quad u \in X^n \tag{3.1}$$

with $v = u^n + \tau f^{n-1/2}$, τ is a given time step,

$$X^n = \left\{ u \in L^1(\Omega) : u \geq 0, \text{ and } \int_{\Omega} u \, dx = \int_{\Omega} v \, dx \right\},$$

and $f^{n-1/2}$ is an approximation of the average value of f over the interval $((n - 1)\tau, n\tau)$.

We compute the Euler equation of the variational principle in two ways. We first review the computation of [9] to show that the Euler Equation of (1.3) is (1.1). This is not especially obvious and employs variation of domain. Secondly we compute the Euler equation in the more common direct fashion to establish the boundary conditions on $\partial\Omega$, which involves the calculation of the variation of (2.6).

3.1. Variation of domain

This argument is due to Otto [12], see also [9]. Let ξ be smooth with compact support in Ω and define the variation $y = \psi(x, \epsilon) = \psi_{\epsilon}(x)$ by

$$\frac{dy}{d\epsilon} = \xi(y), \quad y|_{\epsilon=0} = x,$$

and the “push forward” density u_{ϵ} by

$$\int_{\Omega} \xi u_{\epsilon} \, dx = \int_{\Omega} \xi(\psi_{\epsilon}(x)) u(x) \, dx = \int_{\Omega} \xi(y) \frac{u \circ \psi_{\epsilon}^{-1}(y)}{\frac{d\psi_{\epsilon}}{dx} \circ \psi_{\epsilon}^{-1}(y)} \, dy$$

or

$$u_{\epsilon}(y) = \frac{u \circ \psi_{\epsilon}^{-1}(y)}{\frac{d\psi_{\epsilon}}{dx} \circ \psi_{\epsilon}^{-1}(y)}.$$

Let $u = u^n$ denote the solution of the implicit scheme at time step n , and u_ϵ be defined as above. We must compute

$$\frac{d}{d\epsilon} I(u_\epsilon) = 0 \quad \text{at } \epsilon = 0. \quad (3.2)$$

Substitute u_ϵ into the energy

$$\begin{aligned} E(u_\epsilon) &= \int_{\Omega} (\Phi(u_\epsilon) + \Psi u_\epsilon) \, dx \\ &= \int_{\Omega} \left(\frac{\Phi(u_\epsilon)}{u_\epsilon} + \Psi \right) u_\epsilon \, dx \\ &= \int_{\Omega} \left(\frac{\Phi(u_\epsilon \circ \psi_\epsilon(x))}{u_\epsilon \circ \psi_\epsilon(x)} + \Psi \circ \psi_\epsilon(x) \right) u(x) \, dx \end{aligned}$$

and first calculate $(d/d\epsilon)E(u_\epsilon)$ at $\epsilon = 0$. Now

$$u_\epsilon \circ \psi_\epsilon(x) = \frac{u(x)}{\frac{d\psi_\epsilon}{dx}(x)} \quad \text{so that} \quad \frac{d}{d\epsilon} \Big|_{\epsilon=0} u_\epsilon \circ \psi_\epsilon(x) = -\xi'(x) u(x).$$

From these formulas we find, for example,

$$\frac{d}{d\epsilon} \Big|_{\epsilon=0} \frac{\Phi(u_\epsilon \circ \psi_\epsilon)}{u_\epsilon \circ \psi_\epsilon} u = -(u\Phi'(u) - \Phi(u))\xi' = -K(u)\xi'.$$

After some additional manipulation we obtain

$$\frac{d}{d\epsilon} E(u_\epsilon) \Big|_{\epsilon=0} = \int_{\Omega} (K(u)_x + \Psi' u) \xi \, dx, \quad (3.3)$$

thus the functional $E(\cdot)$ has the correct Euler equation.

Now that we have (3.3), we would like to understand how the Wasserstein metric provides an approximation to the difference quotient in time. Suppose that p is the solution of the minimization problem (2.1) defining the distance between u and v and with ψ_ϵ as above, let p_ϵ be defined by

$$\iint_{\Omega \times \Omega} \xi(x, y) \, dp_\epsilon(x, y) = \iint_{\Omega \times \Omega} \xi(x, \psi_\epsilon(y)) \, dp(x, y).$$

Then p_ϵ is a competitor for the determination of the Wasserstein distance between v and u_ϵ so

$$\frac{1}{2\epsilon} (d(u_\epsilon, v)^2 - d(u, v)^2) \leq \frac{1}{2\epsilon} \iint_{\Omega \times \Omega} ((x - \psi_\epsilon(y))^2 - (x - y)^2) \, dp(x, y)$$

and

$$\limsup_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} (d(u_\epsilon, v)^2 - d(u, v)^2) \leq \iint_{\Omega \times \Omega} (y - x) \xi(y) \, dp(x, y).$$

Combining this with (3.2, 3.3) we find

$$0 \leq \iint_{\Omega \times \Omega} (y - x) \xi(y) \, dp(x, y) + \int_{\Omega} (K(u)_x + \Psi' u) \xi \, dx.$$

Since we may replace ξ by $-\xi$ it follows that

$$0 = \iint_{\Omega \times \Omega} (y-x)\xi(y) dp(x,y) + \int_{\Omega} (K(u)_x + \Psi'u)\xi dx. \tag{3.4}$$

Putting $\xi = \zeta'$, an application of Taylor's theorem to the term $\zeta'(y)(y-x)$ gives

$$\begin{aligned} \iint_{\Omega \times \Omega} (y-x)\xi(y) dp(x,y) &= \iint_{\Omega \times \Omega} \left(\zeta(y) - \zeta(x) + \frac{1}{2}D^2\zeta(\cdot)(y-x)^2 \right) dp(x,y) \\ &= \int_{\Omega} (u-v)\zeta dx + O(\|D^2\zeta\|_{L^\infty(\Omega)}d(u,v)^2), \end{aligned}$$

where $v = u^{n-1} - \tau f^{n-1/2}$. Putting this back in (3.4), we obtain

$$\int_{\Omega} (u-v)\zeta dx - \tau \int_{\Omega} \frac{d}{dx} (K(u)_x + \psi'u)\zeta dx = O(\|D^2\zeta\|_{L^\infty(\Omega)}d(u,v)^2),$$

as required. In [9] it is shown that as $\tau \rightarrow 0$, we obtain a solution of (1.1).

3.2. Explicit variation

First we compute the variation of the metric as derived in (2.5, 2.6). Given the positive densities u and v with distributions U and V ,

$$d(u,v)^2 = \int_{\Omega} (x - \phi(x))^2 u(x) dx = \int_{\Omega} (x - V^{-1} \circ U(x))^2 u(x) dx$$

and $u(x) = \phi'(x)v(\phi(x))$. Consider a variation $u_\epsilon = u + \epsilon\xi$, with $\xi = \zeta'$ and $U_\epsilon = U + \epsilon\zeta$, where $\zeta(0) = \zeta(1) = 0$. So here $\phi_\epsilon(x) = V^{-1} \circ U_\epsilon(x)$ and the formula we need to know is that

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \phi_\epsilon(x)u(x) = \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} V^{-1} \circ U_\epsilon(x)\zeta(x) = \frac{u(x)}{v(\phi(x))}\zeta(x) = \phi'(x)\zeta(x).$$

If we let $id(x) = x$ denote the identity function, we now compute

$$\begin{aligned} \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} d(u_\epsilon, v)^2 &= \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \int_{\Omega} (id - \phi_\epsilon)^2 u_\epsilon dx \\ &= \int_{\Omega} \left(-2(id - \phi) \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \phi_\epsilon u + (id - \phi)^2 \xi \right) dx \\ &= \int_{\Omega} (-2(id - \phi)\phi'\zeta + (id - \phi)^2 \xi) dx \\ &= \int_{\Omega} (-2(id - \phi)\zeta + 2(id - \phi)(1 - \phi')\zeta + (id - \phi)^2 \xi) dx \\ (*) &= \int_{\Omega} \left(-2(id - \phi)\zeta + \frac{d}{dx} ((id - \phi)^2 \zeta) \right) dx \\ &= -2 \int_{\Omega} (id - \phi)\zeta dx \\ &= -2 \int_{\Omega} \left(\int_x^1 (s - \phi(s)) ds \right) \xi(x) dx. \end{aligned}$$

Note that for this formula to be valid, ζ need not vanish at 0 and 1; $\zeta(0) = \zeta(1)$ is sufficient to satisfy the constraint. The integrated derivative in (*) vanishes because $x = \phi(x)$ for $x = 0, 1$.

Thus when (3.1) holds,

$$-\int_{\Omega} \left(\int_x^1 (s - \phi(s)) ds \right) \xi(x) dx + \tau \int_{\Omega} (\Phi'(u(x)) + \Psi(x)) \xi(x) dx = 0, \quad (3.5)$$

whenever ξ has average zero. This means that

$$-\frac{d}{dx} (\Phi'(u(x)) + \Psi(x)) = \frac{1}{\tau} (x - \phi(x)) \quad (3.6)$$

and in particular

$$-\frac{d}{dx} (\Phi'(u(x)) + \Psi(x)) = 0 \quad \text{at } x = 0, 1.$$

Expressing this in terms of K gives that

$$\frac{1}{u} \left(\frac{d}{dx} K(u) + \Psi' u \right) = 0 \quad \text{at } x = 0 \text{ and } x = 1.$$

Note also that if Φ is well behaved, namely, $\Phi'' \geq \text{const.} > 0$, then (3.6) leads to an a priori estimate for d^2u/dx^2 which depends on τ as well as lower derivatives.

4. NUMERICAL APPROXIMATIONS

4.1. Approximation of the variational problem

We approximate the variational problem (1.4) in one dimension using piecewise constant functions. All of the data (Ψ , u_0 and f) were approximated using piecewise constant functions (Ψ_h , u_{0h} and f_h) that agreed at the mid-points of the constant intervals. The discrete space of admissible functions is

$$X_h = \{u \in L^1(\Omega) \mid u \geq 0, u|_{(x_i, x_{i+1})} = u_i \in \mathbb{R}, \sum_{i=0}^{N-1} u_i = \sum_{i=0}^{N-1} \tau f_i^{n+1/2} + u_i^n\}$$

where $x_i = ih$, $h = 1/N$, and $f_i^{n+1/2} = f((i+1/2)h, (n+1/2)\tau)$. By appealing to formula in (2.6), the approximate energy

$$I(u) = \frac{1}{2} d(u, \tau f_h^{n+1/2} + u^n)^2 + \tau \int_0^1 (\Phi(u) + \Psi_h u) dx$$

can be computed exactly using Simpson's rule on appropriate intervals when $u \in X_h$.

We considered two algorithms to minimize the energy over X_h . The first algorithm was a simple relaxation algorithm:

1. Let $\epsilon = 1$ and $u = \alpha u^n$ where $\alpha = \|\tau f_h^{n+1/2} + u_h^n\|_{\ell^1} / \|u_h^n\|_{\ell^1}$.
2. While $\epsilon > 10^{-6}$ repeat Steps 3-4.
3. For each $i = 0, 1, \dots, N-1$, set $e_i = 1$ and $e_j = 0$ for $j \neq i$.
 - Let $v = \Pi(u + \epsilon e)$ and if $I(v) < I(u)$ assign $u = v$.
 - Let $v = \Pi(u - \epsilon e)$ and if $I(v) < I(u)$ assign $u = v$.
4. Repeat Step 3 until u doesn't change and then assign $\epsilon = \epsilon/2$.
5. Assign $u^{n+1} = u$.

In the above, $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the projection of \mathbb{R}^n onto (the closed convex subset²) X_h . While this algorithm is rather slow it is extremely robust; especially when there are small components in the solution, $u_i \simeq 0$, where derivatives of Φ are unbounded (recall that typically $\Phi(\rho) = \rho \ln(\rho)$).

For most of the experiments reported below we used the projected gradient method with (essentially) fixed step length.

1. Let $\rho = 1$ and $u = \alpha u^n$ where $\alpha = \|\tau f_h^{n+1/2} + u_h^n\|_{\ell^1} / \|u_h^n\|_{\ell^1}$.
2. Repeat until convergence (or fail):
 - Let $v = \Pi(u - \rho \nabla I(u))$.
 - If $I(v) < I(u)$ assign $u = v$, otherwise set $\rho = \rho/2$.
 - If ρ is too small the algorithm fails.
3. Assign $u^{n+1} = u$.

The elementary algorithm to determine ρ is a practical way of finding an appropriate step size, and for strictly convex problems with bounded derivatives will not fail (see for example [4]). The gradient is computed using the formulae in Section 3.2, and the details are discussed below. Since this algorithm uses gradient information, it works much faster for solutions bounded away from zero. Of course the number of iterations required for convergence depends upon the condition number, which varies like τ/h^2 , so these algorithms can converge slowly when $\tau \sim h \ll 1$.

4.2. Computing the gradient of the energy

The projected gradient algorithm requires the gradient of the energy to be computed at each step. From equation (3.5) we deduce that

$$\delta I(u)\xi = - \int_{\Omega} \left(\int_x^1 (s - V^{-1} \circ U(s)) ds \right) \xi(x) dx + \tau \int_{\Omega} (\Phi'(u(x)) + \Psi(x))\xi(x) dx.$$

When $u = u_h$ and the variation $\xi = \delta u_h$ are piecewise constant, computation of the second term is elementary:

$$\nabla \left(\int_{\Omega} (\Phi(u) + \Psi u) dx \right)_i = \Phi'(u_i) + \Psi_i.$$

The first term is the gradient of the Wasserstein metric, and for piecewise constant functions evaluates to

$$\nabla \left(\frac{1}{2} d(u, v)^2 \right)_i = - \int_{x_i}^{x_{i+1}} \left(\int_x^1 (y - V^{-1} \circ U(y)) dy \right) dx$$

where (x_i, x_{i+1}) is an interval over which u is constant. The inner integral is piecewise linear so can be evaluated exactly using the mid-point rule. In the numerical experiments we first evaluated the inner integral over each interval, and then approximated the gradient by:

$$\begin{aligned} \nabla \frac{1}{2} (d(u, v)^2)_i &= - \int_{x_i}^{x_{i+1}} \left(\int_x^{x_{i+1}} (y - V^{-1} \circ U(y)) dy \right) dx - (x_{i+1} - x_i) \int_{x_{i+1}}^1 (y - V^{-1} \circ U(y)) dy \\ &\simeq -(x_{i+1} - x_i) \left(\frac{1}{2} \int_{x_i}^{x_{i+1}} (y - V^{-1} \circ U(y)) dy + \int_{x_{i+1}}^1 (y - V^{-1} \circ U(y)) dy \right). \end{aligned}$$

²Recall that $v = \Pi(u)$ if and only if $(u - v, x - v) \leq 0$ for all $x \in X_h$, where (\cdot, \cdot) is the natural inner product in \mathbb{R}^n .

TABLE 4.1. Errors for Example 1: Non-homogeneous heat equation.

$h = 1/64$			$\tau = 1/4096$		
τ	$\ u(1) - u_h(1)\ _{\ell^1}$	$\ u - u_h\ _{\ell^1(\ell^1)}$	h	$\ u(1) - u_h(1)\ _{\ell^1}$	$\ u - u_h\ _{\ell^1(\ell^1)}$
1/8	0.00865447	0.00501116	1/8	0.00109435	0.00043957
1/16	0.00442464	0.00245183	1/16	0.00033866	0.00012566
1/32	0.00223227	0.00121354	1/32	0.00013911	0.00004534
1/64	0.00112579	0.00060531	1/64	0.00008540	0.00002702
1/128	0.00056703	0.00030483	1/128	0.00007892	0.00002387
1/256	0.00029193	0.00015626	1/256	0.00007803	0.00002370
1/512	0.00016001	0.00008231	1/512	0.00007785	0.00002349
Norm	0.250	0.225	Norm	0.250	0.225

4.3. Numerical examples

4.3.1. Example 1: non-homogeneous heat equation

In order to observe rates of convergence we consider a smooth solution of the heat equation:

$$u_t - u_{xx} = f, \quad u_x|_{\partial\Omega} = 0,$$

with u and f given by

$$u(x, t) = (2 + t(1/2 + \cos(\pi x))) / 10, \quad f(x, t) = (1/2 + (1 + \pi^2 t) \cos(\pi x)) / 10.$$

In order to estimate rates of convergence, one of the parameters h or τ was fixed, and the other varied. Discrete $L^1(\Omega)$ errors were computed according to the formulae:

$$\|e\|_{\ell^1} = \sum_{i=0}^{N-1} h|e_i|, \quad \|e\|_{\ell^1(\ell^1)} = \sum_{m=1}^M \tau \|e^m\|_{\ell^1}.$$

In the formula $e_i = u((i + 1/2)h) - u_{hi}$ where u is the exact solution and u_h is the computed solution and the superscript indicates the time at which the solution was computed, $e^m = e(m\tau)$.

Table 4.1 tabulates the error norms for $h = 1/64$ as τ varies and for $\tau = 1/4096$ as h varies. Observation of the table with $h = 1/64$ shows that errors vary proportionally with τ , indicating a first order rate of convergence with respect to this variable. It is also clear that, for this example, the errors due to the temporal discretization are larger than those of the spatial discretization, since even with the smallest time step the limiting error due to the spatial discretization hadn't been attained. One reason for this is indicated in the adjacent table where τ is held constant and h varied. This table exhibits a "super convergence" phenomena where the rate of convergence of approximately³ $\log_2(3)$ until the limiting value is attained. Of course if the piecewise constant functions u_h were exactly integrated over the whole interval (instead using the mid-point rule which is a higher order quadrature formula) the rate could not be greater than unity.

4.3.2. Example 2: Stefan problem

As a second example we consider a solution of the Stefan problem

$$u_t - K(u)_{xx} = f, \quad K(u)_x|_{\partial\Omega} = 0,$$

³The error decreased by about a factor of 3 when h decreased by a factor of 2.

TABLE 4.2. Errors for Example 2: Stefan problem.

$h = \tau$	$\ u(1) - u_h(1)\ _{\ell^1}$	$\ u - u_h\ _{\ell^1(\ell^1)}$
1/16	0.00683647	0.0145502
1/32	0.00230178	0.0097611
1/64	0.00098246	0.0055700
1/128	0.00046830	0.0029667
1/256	0.00023325	0.0015180
1/512	0.00011935	0.0007630
Norm	1.78646	1.426515

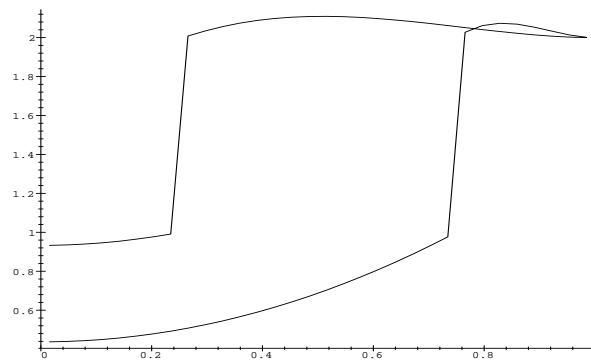


FIGURE 4.1. Solution of the Stefan problem at $t = 0$ and $t = 1$, $h = \tau = 1/32$.

where K is given by (see Fig. 1.1)

$$K(u) = \begin{cases} u & u < 1 \\ 1 & 1 \leq u \leq 2 \\ u - 1 & 2 < u \end{cases}$$

A solution with a free boundary moving from $x = 3/4$ at $t = 0$ to $x = 1/4$ at $t = 1$ was considered. Specifically, if $s(t) = 3/4 - t/4$ is the location of the front,

$$u(x, t) = \begin{cases} -2(s\dot{s} + 1) & x < s \\ \left[(1-x)^2\dot{s}(2(x-s) - (2s-\dot{s})(1-2(x-s)/(1-s))) \right. \\ \left. + (2s-\dot{s})(4(1-x) - 2(x-s)) \right] / (1-s)^2 & s < x. \end{cases}$$

A tedious computation shows that this function satisfies the correct jump conditions across the interface, and boundary condition at $x = 0$ and $x = 1$. f may be determined by substituting u into the equation. The discrete $L^1(\Omega)$ errors for this problem are tabulated in Table 4.2, and again a first order rate of convergence can be observed. A plot of the solution at $t = 0$ and $t = 1$ computed with $h = \tau = 1/32$ is shown in Figure 4.1.

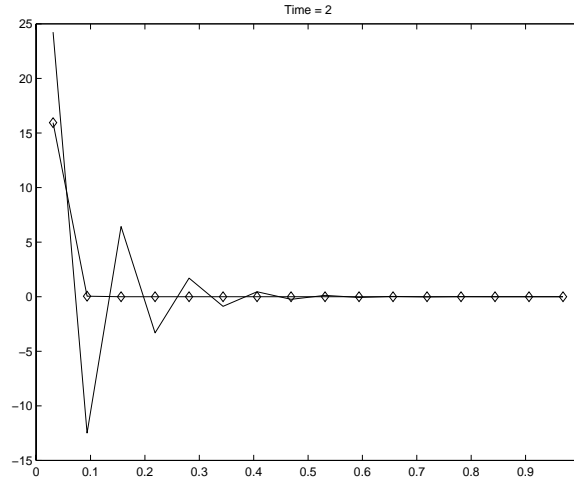


FIGURE 4.2. Example 3: Finite difference approximation —, variational algorithm —◇—.

4.3.3. Example 3: Convection diffusion problem

The convection diffusion problem:

$$u_t - \frac{\partial}{\partial x}(\nu u_x + \mathbf{U}u) = 0, \quad (\nu u_x + \mathbf{U}u)|_{\partial\Omega} = 0$$

differs from the previous two examples in that solutions can not be trivially translated away from zero, and this could potentially cause problems with our numerical scheme since the principle term has a logarithmic singularity at zero ($\Phi(u) = \nu u \ln(u)$). Here ν is a (small) diffusion coefficient and $\mathbf{U} = \Psi'$ is a mean flow convecting u .

If $\mathbf{U} = \Psi'$ is constant the steady state solution is

$$u_s(x) = \left(\int_0^1 u_0(x) dx \right) \frac{\mathbf{U}}{\nu} \frac{\exp(-\mathbf{U}x/\nu)}{1 - \exp(-\mathbf{U}/\nu)}.$$

When \mathbf{U}/ν is large the steady state solution is of order \mathbf{U}/ν at $x = 0$ and decays exponentially for $x > 0$, and traditional numerical schemes have problems resolving such boundary layers. For example, the classical cell centered finite difference scheme for the convection diffusion equation is

$$-\left(\frac{\nu\tau}{h^2} - \frac{\mathbf{U}\tau}{2h} \right) u_{i-1}^{n+1} + \left(1 + 2\frac{\nu\tau}{h^2} \right) u_i^{n+1} - \left(\frac{\nu\tau}{h^2} + \frac{\mathbf{U}\tau}{2h} \right) u_{i+1}^{n+1} = u_i^n.$$

This system of linear equations will satisfy a maximum principle when the off diagonal terms are negative, that is when $h \leq 2\nu/\mathbf{U}$, and for larger values of h spurious oscillations typically occur. Figure 4.2 illustrates this phenomena for $\mathbf{U} = 1$, $\nu = 1/100$ and $h = 1/16$ with initial data $u_0(x) = 1$. The solution obtained using our scheme didn't exhibit any oscillations. Table 4.3 tabulates the non-zero values of u_h and the average values of u_s over the cells $(ih, (i+1)h)$, and it is clear that the boundary layer is well approximated by the discrete solution. This solution was obtained using the relaxation algorithm since the projected gradient method fails at places where $u = 0$.

TABLE 4.3. Example 3: Convection diffusion problem, $t = 2$, $h = 1/16$, $\tau = 1/64$.

i	0	1	2	3	4–15
u_i	15.948961	0.050891	0.000143	0.000004	$< 10^{-6}$
$\int_{(x_i, x_{i+1})} u_s(x) dx$	15.969112	0.030827	0.000059	0.0000001	$< 10^{-6}$

5. SUMMARY

This work illustrates that new variational principles developed to study partial differential equations whose solution exhibit certain patterns can be used to construct approximations of certain classical problems. This approach has several remarkable properties which are very attractive numerically:

- The ability to optimize directly in the weak topology permits us to reduce oscillations, even though a weak neighborhood may contain very rough functions.
- Derivatives of the approximating functions do not enter the variational principle. This allows conforming approximations to be constructed from discontinuous functions which is very appealing for problems exhibiting steep gradients.
- These methods provide (convex) variational principles for some problems, such as the convective diffusion equation, whose usual formulation is not self adjoint. Algorithms based upon variational principles tend to be very stable since they typically inherit energy minimizing properties of the original problem.

However, currently there are still many open theoretical and practical issues that need to be addressed. These include:

- Proofs of convergence of the semi-discrete Euler approximations for a broad class of problems.
- A better understanding of how to implement the various standard boundary conditions on bounded domains is needed.
- Variational principles for signed measures are needed if solutions having arbitrary sign are to be accommodated.
- Efficient algorithms to compute the Wasserstein distance in multiple dimensions need to be developed. Benamou and Brenier [1] develop a novel algorithm for computing the distance between two periodic functions on \mathbb{R}^n which poses the functions as initial and final conditions for an evolution problem. It is not clear how errors associated with approximating the metric would impact the stability of the variational problems of interest here. Recall that the Wasserstein metric is multiplied by a factor of $1/\tau$, and this may exacerbate such errors.
- If these variational principles are to be used in practice, effective minimization algorithms need to be developed.

This work has benefited from many comments and suggestions from our colleagues R. Jordan, M. Kowalczyk, F. Otto, and B. Perthame, whose assistance we gratefully acknowledge.

APPENDIX A. PROOF OF MONOTONICITY

In this section we show that if

$$J(\phi) = \int_{\Omega} (x - \phi(x))^2 u(x) dx$$

is minimized subject to the constraint $\phi : [0, 1] \rightarrow [0, 1]$,

$$\int_{\phi(x) \in B} u(x) dx = \int_B v(x) dx$$

then the minimizer can be chosen to be monotone increasing.

Definition A.1. The monotone rearrangements of Borel sets $B \subset [0, 1]$ and functions $\phi : [0, 1] \rightarrow [0, 1]$ with respect to the measure $\mu_u = u \, dx$ are defined by:

- $B^* = (\beta, 1]$ where $\mu_u(B) = \mu_u(\beta, 1]$, i.e.

$$\int_B u(x) \, dx = \int_\beta^1 u(x) \, dx$$

- and

$$\phi^*(x) = \int_0^1 \chi_{\{\phi > s\}^*}(x) \, d\mu_u(x).$$

Let us collect a few elementary properties of ϕ^* .

- Since $\phi(x) \leq 1$ it follows that $\{\phi > s\} = \emptyset$ for $s \geq 1$ so that $0 \leq \phi^*(x) \leq 1$. In fact this argument shows that $\|\phi^*\|_{L^\infty(\Omega)} = \|\phi\|_{L^\infty(\Omega)}$.
- ϕ^* is monotone increasing and $\phi^*(0) = 0$.
- We claim that $\{\phi^* > t\} = \{\phi > t\}^*$.

Proof. The key ingredient is to observe that

$$\{\phi > t\} = \bigcup_{s>t} \{\phi > s\}$$

and the same holds for the “stared” sets. Then

$$x \in \{\phi^* > t\} \Leftrightarrow x \in \{\phi > s\}^* \text{ for some } s > t \Leftrightarrow x \in \{\phi > t\}^*.$$

□

- If ϕ is monotone increasing and right continuous, then $\phi^* = \phi$.
- We also claim that for Borel sets in $[0, 1]$ that

$$\int_{\phi(x) \in B} u(x) \, dx = \int_{\phi^*(x) \in B} u(x) \, dx.$$

Thus if the left hand side is equal to $\int_B v$ then so too does the right hand side. It follows that ϕ^* is a candidate for minimizing J whenever ϕ is.

Proof. Notice that if we define

$$\mu(B) = \int_{\phi(x) \in B} u(x) \, dx = \mu_u\{\phi \in B\}$$

and

$$\mu^*(B) = \int_{\phi^*(x) \in B} u(x) \, dx = \mu_u\{\phi^* \in B\}.$$

Then μ and μ^* are Borel measures on $[0, 1]$. If we set $B = (t, 1]$ then

$$\mu(B) = \mu_u\{\phi > t\} = \mu_u\{\phi > t\}^* = \mu_u\{\phi^* > t\} = \mu^*(B).$$

Since sets of the form $(t, 1]$ generate the Borel sigma algebra under the Caratheodory construction it follows that $\mu = \mu^*$. □

We next need to show that substituting ϕ^* for ϕ doesn't increase J . We essentially follow the arguments in Leib and Loss [11]. The key ingredient is the following lemma which may be found in [11].

Lemma A.1. *Let ν be a Borel measure on⁴ \mathbb{R}_0^+ and assume that $g(t) \equiv \nu[0, t)$ is finite for all $t \in \mathbb{R}^+$. If (Ω, Σ, μ) is a measure space and $\phi : \Omega \rightarrow \mathbb{R}_0^+$ is measurable then*

$$\int_{\Omega} g \circ \phi \, d\mu = \int_0^{\infty} \mu\{\phi > s\} \, d\nu(s).$$

In particular, if we select $\mu = \delta_x$ and $\nu = dt$ then

$$\phi(x) = \int_0^{\infty} \delta_x\{\phi > s\} \, ds = \int_0^{\infty} \chi_{\{\phi > s\}}(x) \, ds.$$

Lemma A.2. *Let $\phi : [0, 1] \rightarrow [0, 1]$ be Borel and ϕ^* be it's monotone increasing rearrangement (respect to the measure μ_u). Then*

- *If $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ vanishes at zero, is monotone increasing and left continuous then*

$$\int_0^1 g \circ \phi \, d\mu_u = \int_0^1 g \circ \phi^* \, d\mu_u.$$

- *If $\phi, \psi : [0, 1] \rightarrow [0, 1]$ then*

$$\int_0^1 \phi \psi \, d\mu_u \leq \int_0^1 \phi^* \psi^* \, d\mu_u.$$

Given this lemma it follows that $J(\phi^*) \leq J(\phi)$ since

$$\begin{aligned} \int_0^1 (x - \phi(x))^2 u(x) \, dx &= \int_0^1 (x - \phi(x))^2 \, d\mu_u(x) \\ &= \int_0^1 (x^2 - 2x\phi(x) + \phi(x)^2) \, d\mu_u(x) \\ &\geq \int_0^1 (x^2 - 2x\phi^*(x) + \phi^*(x)^2) \, d\mu_u(x) \\ &\geq \int_0^1 (x - \phi^*(x))^2 u(x) \, dx. \end{aligned}$$

Proof. The hypotheses on g guarantee that it is the distribution function of a measure ν on \mathfrak{R}_0^+ . In the layer cake lemma select $\mu = \mu_u$ on $[0, 1]$ to get

$$\begin{aligned} \int_0^1 g \circ \phi \, d\mu_u &= \int_0^{\infty} \mu_u\{\phi > s\} \, d\nu(s) \\ &= \int_0^{\infty} \mu_u\{\phi > s\}^* \, d\nu(s) \\ &= \int_0^{\infty} \mu_u\{\phi^* > s\} \, d\nu(s) \\ &= \int_0^1 g \circ \phi^* \, d\mu_u. \end{aligned}$$

⁴ $\mathbb{R}_0^+ = \{x \in \mathbb{R} : x \geq 0\}$

To establish the second statement we use the second statement of the layer cake theorem and Fubini's theorem.

$$\int_0^1 \phi(x)\psi(x) \, d\mu_u(x) = \int_0^\infty \int_0^\infty \int_\Omega \chi_{\{\phi>s\}}(x)\chi_{\{\psi>t\}}(x) \, d\mu_u(x) \, ds \, dt.$$

For s and t fixed, let $A = \{\phi > s\}$ and $B = \{\psi > t\}$. Then the inner integral on the right evaluates to $\mu_u(A \cap B)$. Suppose that $\mu_u(A) \geq \mu_u(B)$; then $B^* \subset A^*$ so that

$$\mu_u(A^* \cap B^*) = \mu_u(B^*) = \mu_u(B) \geq \mu_u(A \cap B).$$

Since $A^* = \{\phi > s\}^* = \{\phi^* > s\}$ it follows that

$$\int_0^1 \phi\psi \, d\mu_u \leq \int_0^\infty \int_0^\infty \int_\Omega \chi_{\{\phi^*>s\}}(x)\chi_{\{\psi^*>t\}}(x) \, d\mu_u(x) \, ds \, dt = \int_0^1 \phi^*\psi^* \, d\mu_u. \quad \square$$

REFERENCES

- [1] J. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Preprint (1998).
- [2] J. Benamou and Y. Brenier, Weak existence for the semigeostrophic equations formulated as a coupled Monge–Ampere transport problem. *SIAM J. Appl. Math.* **58** (1998) 1450–1461.
- [3] Y. Brenier, Polar factorization and monotone rearrangement of vector–valued functions. *Comm. Pure Appl. Math.* **44** (1991) 375–417.
- [4] P.G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge University Press, Cambridge (1988).
- [5] W.E and F. Otto, Thermodynamically driven incompressible fluid mixtures. *J. Chem. Phys.* **107** (1998) 10177–10184.
- [6] M. Frechet, Sur la distance de deux lois de probabilité. *C.R. Acad. Sci.* **244** (1957) 689–692.
- [7] W. Gangbo and A. SWEICH, Optimal maps for the multidimensional Monge–Kantorovich problem. *CPAM* **51** (1998) 23–45.
- [8] W. Gango and R.J. McCann, The geometry of optimal transportation. *Acta Math.* **177** (1996) 113–161.
- [9] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.
- [10] R. Jordan, D. Kinderlehrer and F. Otto, Dynamics of the Fokker–Planck equation. *Phase Transitions* (to appear).
- [11] E.H. Lieb and M. Loss, *Analysis*, Vol. 14 of Graduate Studies in Mathematics. AMS (1997).
- [12] F. Otto, Dynamics of labyrinthine pattern formation in magnetic fluids. *Arch. Rational Mech. Anal.* **141** (1998) 63–103.
- [13] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*. North–Holland (1981).