A FAST MULTIGRID METHOD FOR INVERTING LINEAR PARABOLIC PROBLEMS

ANDREI DRĂGĂNESCU*

Abstract. We present and analyse a multigrid algorithm for linear L^2-regularized inverse parabolic equations, formulated as an unconstrained optimization problem. The method requires only one iteration at the finest level in order to resolve the problem to optimal order. The analysis is based on a two-grid approximation for the inverse Hessian that uses the smoothing property of parabolic equations.

Key words. inverse problems, parabolic equations, multigrid methods, finite elements

AMS subject classifications. 65M32, 65M55, 65M60

This work started as a bid for a cost-efficient method for inverting time-dependent partial differential equations. The original motivation comes from the problem of comparing simulations with experimental results. Given that only partial information is available from experiments, we address the question of finding the best possible match by controlling initial conditions and/or parameters. Since we may be talking about large problems, for which the forward simulation may be taking many hours on a powerful machine, it becomes clear that the attached optimization problem should take no longer than a factor of ten of that time, a demand which is in conflict with the large size of the control space. Numerical experiments with two model problems – two dimensional gas dynamics and a one dimensional advection-reaction-diffusion equation – suggested that parabolic equations are harder to invert than hyperbolic ones, which motivated our decision to investigate the former from a theoretical viewpoint. The results below indicate that the optimization can be done in tis case with the needed efficiency. The work in this article is extended in [4] and [5].

1. Notation and problem formulation. Let $\Omega \subset \mathbb{R}^2$ be an open set. We shall denote by $L^2(\Omega)$ and $H^m(\Omega), H_0^m(\Omega)$ ($m > 0$, integer) the standard Lebesgue respectively Sobolev spaces, and by $\|\cdot\| = \|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_m$ (resp. $|\cdot|_m$) the corresponding norms (resp. seminorms); furthermore $\langle\cdot,\cdot\rangle$ (resp. $\langle\cdot,\cdot\rangle_m$) will be the standard inner product in $L^2(\Omega)$ (resp. $H^m(\Omega)$). Let $\widetilde{H}^{-m}(\Omega)$ be the dual (in L^2) of $H^m(\Omega) \cap H_0^1(\Omega)$ for $m > 0$. For $T \in \mathcal{L}(V_1, V_2)$ we denote the *operator norm* of T by $|\!|\!|T|\!|\!|_{V_1,V_2} = \sup_{u \in V_1 \setminus \{0\}} \|Tu\|/\|u\|$, and we write $|\!|\!|T|\!|\!|_V$ for $|\!|\!|T|\!|\!|_{V,V}$. L^2, H^m, etc., will stand for $L^2(\Omega)$, $H^m(\Omega)$, unless otherwise specified.

We consider the following initial value problem

$$\begin{cases} \partial_t u + A(t)u = 0 & , \quad \text{on } \Omega \times (0,\infty) \ , \\ u(x,t) = 0 & , \quad \text{on } \partial\Omega \times (0,\infty) \ , \\ u(x,0) = u_0(x) & , \quad \text{for } x \in \Omega \ , \end{cases} \tag{1.1}$$

where

$$A(t)u = -\sum_i \partial_i \left(\sum_j a_{ij}(x,t)\partial_j u + b_i(x,t)u \right) + c(x,t)u \ , \tag{1.2}$$

with $a_{ij}(x,t) = a_{ji}(x,t), b_i(x,t), c(x,t)$ being smooth functions with uniformly bounded derivatives of all orders on $\overline{\Omega} \times [0,\infty)$. We define the time-dependent bilinear form $a : (0,\infty) \times H_0^1 \times H_0^1 \to \mathbb{R}$ by

$$a(t;\phi,\psi) = \sum_{i,j} \langle a_{ij}\partial_j\phi, \partial_i\psi\rangle + \sum_i \langle b_i\phi, \partial_i\psi\rangle + \langle c\phi,\psi\rangle \ , \quad \text{for } \phi,\psi \in H_0^1 \ . \tag{1.3}$$

It is assumed that a is coercive, i.e. there exists a constant $c_1 > 0$ independent of t such that

$$a(t;\phi,\phi) \geq c_1 \|\phi\|_1^2, \quad \text{for } \phi \in H_0^1 \ , \tag{1.4}$$

*Department of Mathematics, University of Chicago, Chicago, Illinois 60637.

and that the following regularity condition holds:

$$\|\phi\|_2 \leq c_2 \|A(t)\phi\|, \quad \text{for } \phi \in H_0^1 \cap H^2 \ . \tag{1.5}$$

Cf. [6] and [7], for $u_0 \in L^2$ there exists a unique solution $u : (0, \infty) \to H_0^1$ to the *weak* formulation of (1.1):

$$\begin{cases} \langle \partial_t u, \phi \rangle + a(t; u, \phi) = 0, \quad \text{for all } \phi \in H_0^1, \ t > 0, \\ \lim_{t \to 0} \|u(t) - u_0\| = 0 \ . \end{cases} \tag{1.6}$$

We denote the solution operator by $\mathcal{S}(t)u_0 \stackrel{\text{def}}{=} u(\cdot, t)$.

By *inverse parabolic problem* we mean the following unconstrained minimization problem: given a final time $T > 0$, find

$$\min_{u \in L^2} \mathcal{J}_\epsilon(u) \ , \tag{1.7}$$

where $\mathcal{J}_\epsilon : L^2 \to \mathbb{R}$ is defined by

$$\mathcal{J}_\epsilon(u) = \epsilon^{-1} \underbrace{\frac{1}{2}\|\mathcal{S}(T)u - f\|^2}_{\mathcal{J}^1(u)} + \frac{1}{2}\|u\|^2 \ , \tag{1.8}$$

with $f \in L^2$ a given "target" function.

We discretize the forward problem via the Galerkin method using continuous piecewise linear functions in space, and backward Euler in time. Let $\mathcal{T}_{h_0}$ be a triangulation of the domain Ω with $h_0 = \max\{\text{diam}(\tau) : \tau \in \mathcal{T}_{h_0}\}$, and let $\mathcal{T}_{h/2}$ be defined inductively as the Goursat refinement of $\mathcal{T}_h$ (each triangle in $T \in \mathcal{T}_h$ is cut along the three lines obtained by joining the midpoints of its edges), for $h \in I = \{h_0 2^{-i} : i \in \mathbb{N}\}$. We define the spaces

$$V_h = \{f \in \mathcal{C}(\overline{\Omega}) \ : \ \forall \tau \in \mathcal{T}_h \ \ f|_\tau \text{ linear, and } f|_{\partial\Omega} \equiv 0\}, \quad \text{for } h \in I, \tag{1.9}$$

and denote by π_h the L^2-projection onto V_h. Note that $V_{2h} \subset V_h \subset H_0^1$. Let $t_m = mk$ with $m = 0, 1, \ldots, M$ so that $t_M = T$. Since our approximation in time is only first order accurate we choose $k = k(h) = k_0 T^{-1} h^2$ and, consequently, $M = M(h) = M_0 T h^{-2}$. The backward difference approximation U^m to $u(t_m)$ is computed succesively by

$$\begin{cases} \langle d_t U^{m+1}, \phi \rangle + a(t_{m+1}; U^{m+1}, \phi) = 0, \quad \forall \phi \in V_h \ , \\ U^0 = \pi_h u_0 \ , \end{cases} \tag{1.10}$$

where $d_t U^{m+1} = k^{-1}(U^{m+1} - U^m)$. We will write for $u_0 \in L^2$, $\mathcal{S}^h(t_m)u_0 \stackrel{\text{def}}{=} U^m, \ m = 1, 2, \ldots, M$. We denote by $f_h = \pi_h(f) \in V_h$, and we consider the following discrete minimization problem associated to (1.7):

$$\min_{u \in V_h} \mathcal{J}_\epsilon^h(u), \tag{1.11}$$

with $\mathcal{J}_{\epsilon,h} : V_h \to \mathbb{R}$ given by

$$\mathcal{J}_{\epsilon,h}(u) = \epsilon^{-1} \underbrace{\frac{1}{2}\|\mathcal{S}^h(T)u - f_h\|^2}_{\mathcal{J}_h^1(u)} + \frac{1}{2}\|u\|^2 \ . \tag{1.12}$$

Since the functional $\mathcal{J}_{h,\epsilon}$ is quadratic, its Hessian form $b^h(v_1, v_2) = \epsilon^{-1} D^2 \mathcal{J}_h^1(u)(v_1, v_2) + \langle v_1, v_2 \rangle$ is independent of $u \in V_h$. We therefore define the discrete Hessian operator by

$$\langle H_\epsilon^h u, v \rangle = b^h(u, v), \quad \forall u, v \in V_h \ . \tag{1.13}$$

Note that $H_\epsilon^h = \epsilon^{-1}H^h + I$, where H^h is the Hessian operator corresponding to $\mathcal{J}_h^1$. The symmetry of b^h implies that H_ϵ^h is symmetric. From the linearity of equation (1.10) we obtain

$$\langle H^h u, u\rangle = \|S^h(T)u\|^2 \ , \tag{1.14}$$

hence, by taking (2.1) into account, we obtain

$$\|u\|^2 \le \langle H_\epsilon^h u, u\rangle \le \left(1 + \frac{C^2}{\epsilon}\right)\|u\|^2 \ , \text{ for } u \in V_h \ . \tag{1.15}$$

Similarly, $\mathcal{J}^1$ is quadratic and continuous (again by (2.1)), and we define its Hessian to be the unique element of $H \in \mathcal{L}(L^2)$ satisfying

$$\langle Hu_1, u_2\rangle = D^2\mathcal{J}^1(u)(u_1, u_2), \quad \forall u_1, u_2 \in L^2, \tag{1.16}$$

independently of $u \in L^2$. It follows that H satisfies

$$\langle Hu, u\rangle = \|S(T)u\|^2 = \langle S^*(T)\cdot S(T)u, u\rangle \ , \tag{1.17}$$

where $S^*(T)$ denotes the dual of $S(T)$. Therefore

$$\|u\|^2 \le \langle H_\epsilon u, u\rangle \le \left(1 + \frac{C^2}{\epsilon}\right)\|u\|^2 \ , \tag{1.18}$$

where $H_\epsilon = \epsilon^{-1}H + I$ is the Hessian operator associated to the entire cost functional $\mathcal{J}$. Note that (1.17) together with the symmetry of both H and $S^*(T)\cdot S(T)$ and the polarization identity imply that

$$H = S^*(T)\cdot S(T) \ . \tag{1.19}$$

In light of (1.15) and (1.18) we close the discussion with the following

REMARK 1.1. *Since the continuous and discrete Hessians are positive definite (everywhere), we conclude that the minimization problems* (1.7) *and* (1.11) *have unique solutions* $u^{\min}$ *and* $u_h^{\min}$ *respectively, namely the unique zeros of the gradients of the corresponding functionals. Since* $\nabla\mathcal{J}_\epsilon(u) = H_\epsilon u - \epsilon^{-1}S^*(T)f$, *we obtain*

$$u^{\min} = \epsilon^{-1}(H_\epsilon)^{-1}\, S^*(T)f \ , \tag{1.20}$$

and similarly

$$u_h^{\min} = \epsilon^{-1}(H_\epsilon^h)^{-1}\, \left(\mathcal{S}^h(T)\right)^* f_h \ . \tag{1.21}$$

The main goal of this article is the construction and analysis of an efficient algorithm for finding an approximation to $u_h^{\min}$, that is for solving (1.21).

2. Preliminaries. In this section we first list (Theorem 2.1) some results on parabolic equations of type (1.1) and their numerical approximation. These results, mostly contained in [6] and [7], are needed for the multigrid analysis in §4 and §5. After that we prove the (optimal) estimates in Theorem 2.6.

THEOREM 2.1. *There exists* $C = C(\Omega)$ *such that the following hold for all* $T > 0, \epsilon > 0, h \in I$:

(a) L^2 *- stability of the solution operators (both continuous and discrete)*

$$\max\left(\|\mathcal{S}(T)\|_{L^2}, \|\mathcal{S}^h(T)\|_{V_h}\right) \le C \ ; \tag{2.1}$$

(b) smoothing[1] : $\mathcal{S}(T)u$ *is defined for* $u \in \widetilde{H}^{-2}$ *and we have*

$$\|\mathcal{S}(T)u\|_{\widetilde{H}^{-2},L^2} \le C\,T^{-1} \ ; \tag{2.2}$$

[1]this is an easy consequence of the following fact proved in [6]: $\|u(\cdot,t)\|_{H^p} \le Ct^{-p/2}\|u_0\|$, for $t > 0$ and $p \in \{0, 1, 2\}$.

(c) smoothing and approximation of the solution operator:

$$\|\mathcal{S}(T)u - \mathcal{S}^h(T)u\| \le CT^{-1}h^2\|u\|, \quad \forall u \in V_h \ . \tag{2.3}$$

We would like to remark here that the operators adjoint (with respect to the L^2-inner product) to $\mathcal{S}(T)$ (resp. $\mathcal{S}^h(T)$), denoted by $\mathcal{S}^*(T)$ (resp. $(\mathcal{S}^h)^*(T)$), are defined almost like $\mathcal{S}(T)$ (resp. $\mathcal{S}^h(T)$) in the sense that we only have to replace b_i with $-b_i$ in (1.2) (resp. (1.3); here we assume that the linear system in (1.10) is solved exactly at each step). Therefore Theorem 2.1 applies to $\mathcal{S}^*(T)$ and $(\mathcal{S}^h)^*(T)$ as well (see [4] for details).

We state without proof the following well known result:

LEMMA 2.2. *There exists a constant $K = K(\Omega)$ such that for $h \in I$*

$$\|(I - \pi_h)\|_{L^2, \widetilde{H}^{-2}} \le Kh^2 \ . \tag{2.4}$$

REMARK 2.3. *The stability property* (2.1) *applied to $\mathcal{S}^*(T)$ together with* (2.2) *and* (1.19) *imply that there is a $C = C(\Omega)$ such that*

$$\|H\|_{\widetilde{H}^{-2}, L^2} \le C\, T^{-1} \ . \tag{2.5}$$

REMARK 2.4. *The next inequality is a consequence of Lemma* 2.2 *and the smoothing property* (2.2)*:*

$$\max\left(\|\mathcal{S}(T)(I - \pi_h)\|_{L^2}, \|\mathcal{S}^*(T)(I - \pi_h)\|_{L^2}\right) \le CT^{-1}h^2 \tag{2.6}$$

for some $C = C(\Omega) > 0$.

LEMMA 2.5. *There exists a constant $C = C(\Omega)$ such that*

$$\|\pi_h(H^h - H)\|_{V_h} \le CT^{-1}h^2 \ . \tag{2.7}$$

Proof. For $u \in V_h$ we have

$$\begin{aligned} \langle \pi_h(H^h - H)u, u \rangle &= \|\mathcal{S}^h(T)u\|^2 - \|\mathcal{S}(T)u\|^2 \\ &\le \|\mathcal{S}^h(T)u - \mathcal{S}(T)u\| \cdot \left(\|\mathcal{S}^h(T)u\| + \|\mathcal{S}(T)u\|\right) \overset{(2.1),(2.3)}{\le} 2C^2T^{-1}h^2\|u\|^2 \ , \end{aligned}$$

and the symmetry of $\pi_h(H^h - H) \in \mathcal{L}(V_h)$ now implies (2.7). □

The following result shows that $u_h^{\min}$ approximates $u^{\min}$ to optimal order in the L^2-norm.

THEOREM 2.6. *There exists a constant $C = C(\Omega)$ such that for $h \le h_0(\epsilon, T, \Omega)$ we have the following stability and error estimates:*

$$\|u_h^{\min}\| \le C\left(\|u^{\min}\| + \frac{h^2}{\epsilon T}\|f\|\right) \ , \tag{2.8}$$

$$\|u_h^{\min} - u^{\min}\| \le \frac{C}{\epsilon T}h^2\left(\|f\| + \|u^{\min}\|\right) \ . \tag{2.9}$$

Proof. Denote by $e_h = u_h^{\min} - u^{\min}$. A simple computation yields

$$H_\epsilon e_h = \epsilon^{-1}\left\{\left((\mathcal{S}^h)^*(T) - \mathcal{S}^*(T)\right)f_h - \mathcal{S}^*(T)(I - \pi_h)f + (H - H^h)u_h^{\min}\right\} \ .$$

Now

$$\begin{aligned}
\epsilon\|e_h\|^2 &\overset{(1.18)}{\leq} \epsilon\left\langle H_\epsilon e_h, e_h\right\rangle \\
&= \left\langle \left((\mathcal{S}^h)^*(T) - \mathcal{S}^*(T)\right) f_h - \mathcal{S}^*(T)(I-\pi_h)f, e_h\right\rangle + \left\langle (H-H^h)u_h^{\min}, e_h\right\rangle \\
&\overset{(2.3),(2.6)}{\leq} CT^{-1}h^2\|f\|\,\|e_h\| + \left\langle (H-H^h)u_h^{\min}, e_h\right\rangle \\
&\overset{(I-\pi_h)e_h\perp V_h}{=} CT^{-1}h^2\|f\|\,\|e_h\| + \left\langle \pi_h(H-H^h)u_h^{\min}, \pi_h e_h\right\rangle + \left\langle Hu_h^{\min}, (I-\pi_h)e_h\right\rangle \\
&\overset{(2.7)}{\leq} CT^{-1}h^2\|e_h\|(\|f\| + \|u_h^{\min}\|) + \left\langle \mathcal{S}(T)u_h^{\min}, \mathcal{S}(T)(I-\pi_h)e_h\right\rangle \\
&\overset{(2.1),(2.6)}{\leq} CT^{-1}h^2\|e_h\|(\|f\| + \|u_h^{\min}\|)\ ,
\end{aligned}$$

for some $C = C(\Omega)$, hence

$$\|u_h^{\min} - u^{\min}\| \leq \frac{C}{\epsilon T}h^2\left(\|f\| + \|u_h^{\min}\|\right)\ . \tag{2.10}$$

The conclusions follow by a standard argument. □

3. Tools: the spectral distance between operators. In this section we introduce the *spectral distance* between (symmetric) positive definite operators and show how it can be used in evaluating the convergence of the simple iteration (3.2). For space economy we present the results only, and we refer the reader to [4] or [5] for detailed proofs. Throughout this section $(V, \langle\cdot,\cdot\rangle)$ denotes a real, finite dimensional Hilbert space, and $\mathcal{L}_+^s(V)$ is the space of symmetric positive definite operators. As usual, $\|u\| = \sqrt{\langle u, u\rangle}$ is the Hilbert space norm of $u \in V$; for $A \in \mathcal{L}_+^s(V)$ we write $\|u\|_A \overset{\text{def}}{=} \sqrt{\langle Au, u\rangle} = \|A^{\frac{1}{2}}u\|$. Let $H \in \mathcal{L}_+^s(V)$ and $b \in V$. Consider the equation

$$Hx = b\ . \tag{3.1}$$

It is known that the sequence x_n defined by the simple iteration

$$x_{n+1} = x_n + M(b - Hx_n) \tag{3.2}$$

converges to the solution x^* of (3.1) if $M \approx H^{-1}$ in some sense. If $M \in \mathcal{L}_+^s(V)$, the natural quantity for measuring the rate of convergence in (3.2) is $\|I - H^{\frac{1}{2}}MH^{\frac{1}{2}}\|_V$. For aesthetic reasons, but not only, we would prefer to use another measure instead, that is also a scale-free distance between M and H^{-1} (unlike $\|M - H^{-1}\|_V$, which **is** a distance, but not scale-free).

DEFINITION 3.1. *We define the* spectral distance *between two operators $T_1, T_2 \in \mathcal{L}_+^s(V)$ to be*

$$d_V(T_1, T_2) = \sup_{u\in V\setminus\{0\}}\left|\ln\frac{\langle T_1u, u\rangle}{\langle T_2u, u\rangle}\right| = \sup_{\|u\|=1}\left|\ln\left\langle T_2^{-\frac{1}{2}}T_1T_2^{-\frac{1}{2}}u, u\right\rangle\right|\ . \tag{3.3}$$

Definition 3.1 is extended in [4] to operators with positive definite symmetric part by replacing T_1, T_2 in (3.3) with their complexifications, allowing u to be complex as well, and by using the standard branch of the complex logarithm. This more general distance, a notion related to the *field of values* of an operator, is particularly useful if nonsymmetric preconditioners are needed, as is the case with inverse parabolic equations regularized with a multiple of the square H^1-norm instead of the L^2-norm used in this article.

PROPOSITION 3.2.

(i) d_V is a distance function;

(ii) for $T_1, T_2 \in \mathcal{L}_+^s(V), d_V(T_1, T_2) \approx \|I - T_1^{\frac{1}{2}}T_2^{-1}T_1^{\frac{1}{2}}\|$ as $d_V(T_1, T_2) \to 0$;

(iii) $d_V(T_1, T_2) = d_V((T_1)^{-1}, (T_2)^{-1})$.

In sections 4 and 5 we design and analyse an algorithm for solving (3.1) with $H = H_\epsilon^h$ **by means of finding a recursively-defined, multilevel preconditioner for** H_ϵ^h. In light of this we prefer to regard (3.2) an *iteration of preconditioners*. More precisely, we think of (3.2) as a sequence of **single** iterations with "improved" preconditioners:

$$x_n = x_0 + M_n(b - Hx_0) \ , \tag{3.4}$$

with M_n defined recursively by: $M_0 = 0,\ M_{n+1} = M + M_n - MHM_n$. A simple calculation shows that

$$M_n = H^{-1} - (I - MH)^n H^{-1} = H^{-\frac{1}{2}}\left(I - \left(I - H^{\frac{1}{2}} M H^{\frac{1}{2}}\right)^n\right) H^{-\frac{1}{2}} \ . \tag{3.5}$$

In particular $M_0 = 0$, $M_1 = M$ and $M_2 = 2M - MHM$.

REMARK 3.3. *Note that M_2 is the first Newton iterate with initial guess $M_1 = M$ of the operator-function $X \mapsto X^{-1} - H$. We define the following matrix-operator: $\mathcal{N}_H(X) \stackrel{\text{def}}{=} 2X - XHX$. Performing* **one** *simple iteration with $\mathcal{N}_H(X)$ as preconditioner is equivalent to* **two** *simple iterations with X as preconditioner.*

THEOREM 3.4. *Let $M, H \in \mathcal{L}_+^s(V)$ and M_n be defined as in (3.5). If $d_V(M, H^{-1}) < \ln 2$ then*

$$d_V(M_n, H^{-1}) \leq \ g_n\left(d_V(M, H^{-1})\right) \ , \tag{3.6}$$

where $g_n(x) = |\ln(1 - (e^x - 1)^n)| = x^n + O(x^{n+1})$, hence $\lim_{n\to\infty} M_n = H^{-1}$ in the d_V-metric.

COROLLARY 3.5. *Under the conditions of Theorem 3.4, the sequence x_n defined by (3.2) converges to the solution x^*, and we have the following estimates:*

$$\|e_n\| \leq 2\ \mathrm{cond}(H^{\frac{1}{2}}) \cdot g\left(d_V(M_n, H^{-1})\right) \cdot \|e_0\| \ , \tag{3.7}$$

$$\|e_n\|_H \leq 2\ g\left(d_V(M_n, H^{-1})\right) \cdot \|e_0\|_H \ , \tag{3.8}$$

where $g(x) = e^x - 1 = x + o(x)$. We conclude with the following

REMARK 3.6. *In light of the asymptotic behaviour of the functions g_n and g above, we rewrite (3.7) and (3.8) in the following way: for every $C > 2$ there exists $\delta(C) > 0$ such that, if $d_V(M, H^{-1}) < \delta(C)$, then*

$$\|e_n\| \leq C\ \mathrm{cond}(H^{\frac{1}{2}}) \cdot \left(d_V(M, H^{-1})\right)^n \cdot \|e_0\| \ , \tag{3.9}$$

$$\|e_n\|_H \leq C\ \left(d_V(M, H^{-1})\right)^n \cdot \|e_0\|_H \ . \tag{3.10}$$

4. Two-level preconditioners for the discrete Hessian. In this section we construct a two-level, symmetric, preconditioner for the Hessian H_ϵ^h introduced in §1. Recall that $V_{2h} \subset V_h \subset H^1$. We shall refer to V_h as the *fine* space in relation to V_{2h}, which will be the *coarse* space. Denote $d_h = d_{V_h}$.

Construction of a preconditioner. We have

$$H_\epsilon^h = \underbrace{\pi_{2h} H_\epsilon^h \pi_{2h}}_{M_1} + \underbrace{(I - \pi_{2h}) H_\epsilon^h \pi_{2h}}_{M_2} + \underbrace{\pi_{2h} H_\epsilon^h (I - \pi_{2h})}_{M_3} + \underbrace{(I - \pi_{2h}) H_\epsilon^h (I - \pi_{2h})}_{M_4} \ .$$

By analogy with Fourier analysis we regard the coarse space V_{2h} as the "smooth" space (which is not exactly true), and its orthogonal complement in V_h, namely $\mathrm{Im}(I - \pi_{2h}) \cap V_h$, as the "rough" space. Since the operator H^h is "very" smoothing, we think that $H^h(I - \pi_{2h}) \approx 0$, hence $H_\epsilon^h(I - \pi_{2h}) = (\epsilon^{-1}H^h + I)(I - \pi_{2h}) \approx I - \pi_{2h}$. As a consequence $M_3 \approx \pi_{2h}(I - \pi_{2h}) = 0$, and $M_4 \approx (I - \pi_{2h})(I - \pi_{2h}) = (I - \pi_{2h})$. Another interpretation of the smoothing property of H^h is that H^h takes "smooth" functions into "smooth" functions, that is $H^h(V_{2h}) \subset V_{2h}$, which allows us to approximate $M_2 \approx 0$. The analysis of the large eigenvalues of H_ϵ^h versus the ones of H_ϵ^{2h} for the of the inverse linear heat equation on a uniform grid also points to the fact that for $h \ll 1$ we have $\pi_{2h} H_\epsilon^h \pi_{2h} \approx H_\epsilon^{2h} \pi_{2h}$. It remains that

$$H_\epsilon^h \approx M_\epsilon^h \stackrel{\text{def}}{=} H_\epsilon^{2h} \pi_{2h} + (I - \pi_{2h}) \ .$$

Since π_{2h} is a projection we have an explicit form for $(M_\epsilon^h)^{-1}$, namely

$$L_\epsilon^h \stackrel{\text{def}}{=} (M_\epsilon^h)^{-1} = (H_\epsilon^{2h})^{-1}\pi_{2h} + (I - \pi_{2h}) \ . \tag{4.1}$$

We propose the operator $L_\epsilon^h \in \mathcal{L}(V_h)$ as a candidate for a good approximation of $(H_\epsilon^h)^{-1}$, at least for the case when h is small. Note that the operators M_ϵ^h and L_ϵ^h are in $\mathcal{L}_+^s(V_h)$.

THEOREM 4.1. *For $h < h_0(T, \epsilon, \Omega)$ there exists a constant $C_1 = C_1(\Omega)$ such that*

$$d_h(H_\epsilon^h, M_\epsilon^h) \le C_1 \frac{h^2}{\epsilon T} \ . \tag{4.2}$$

Proof. We have the following relation on V_h:

$$\epsilon\left(M_\epsilon^h - H_\epsilon^h\right) = \underbrace{\pi_{2h}\left(H^{2h} - H\right)\pi_{2h}}_{A} + \underbrace{(\pi_{2h}H\pi_{2h} - \pi_h H\pi_h)}_{B} + \underbrace{(\pi_h H \pi_h - H^h)}_{C} \ .$$

Note that all operators in the above sum are symmetric in $\mathcal{L}(V_h)$. We analyse the products $\langle Au, u\rangle$, $\langle Bu, u\rangle$, $\langle Cu, u\rangle$ for $u \in V_h$:

$$\begin{aligned} |\langle Au, u\rangle| &= \left|\left\langle \pi_{2h}\left(H^{2h} - H\right)\pi_{2h}u, u\right\rangle\right| = \left|\left\langle \left(H^{2h} - \pi_{2h}H\right)\pi_{2h}u, \pi_{2h}u\right\rangle\right| \\ &\le \|H^{2h} - \pi_{2h}H\pi_{2h}\|_{V_{2h}} \cdot \|\pi_{2h}u\|^2 \stackrel{\text{Lemma 2.5}}{\le} CT^{-1}(2h)^p\|u\|^2 \ . \end{aligned}$$

The same Lemma 2.5 implies $|\langle Cu, u\rangle| \le CT^{-1}h^2\|u\|^2$. For the middle term we have

$$\begin{aligned} |\langle Bu, u\rangle| &= |(\langle \pi_{2h}H\pi_{2h} - \pi_h H\pi_h)u, u\rangle| \stackrel{u\in V_h}{=} |\langle H\pi_{2h}u, \pi_{2h}u\rangle - \langle Hu, u\rangle| \\ &= \left|\|\mathcal{S}(T)\pi_{2h}u\|^2 - \|\mathcal{S}(T)u\|^2\right| = \|\mathcal{S}(T)(I - \pi_{2h})u\| \cdot (\|\mathcal{S}(T)\pi_{2h}u\| + \|\mathcal{S}(T)u\|) \\ &\stackrel{(2.1),(2.6)}{\le} 2C^2T^{-1}(2h)^2\|u\|^2 \ . \end{aligned}$$

Putting the above estimates together we get

$$\left|\frac{\langle M_\epsilon^h u, u\rangle}{\langle H_\epsilon^h u, u\rangle} - 1\right| \le C\frac{h^2}{\epsilon T} \cdot \frac{\|u\|^2}{\epsilon^{-1}\langle H^h u, u\rangle + \|u\|^2} \le C\frac{h^2}{\epsilon T} \ . \tag{4.3}$$

Assuming $C(\epsilon T)^{-1}h_0^2 = \alpha < 1$ and $0 < h \le h_0$ we obtain

$$\sup_{u\in V_h\setminus\{0\}} \left|\ln\frac{\langle M_\epsilon^h u, u\rangle}{\langle H_\epsilon^h u, u\rangle}\right| \le \frac{|\ln(1-\alpha)|}{\alpha} \sup_{u\in V_h\setminus\{0\}} \left|\frac{\langle M_\epsilon^h u, u\rangle}{\langle H_\epsilon^h u, u\rangle} - 1\right| \le \underbrace{C\,\frac{|\ln(1-\alpha)|}{\alpha}}_{C_1} \cdot \frac{h^2}{\epsilon T} \ .$$

For the last inequality we have that for $\alpha \in (0, 1)$ and $x \in [1-\alpha, 1+\alpha]$ we have

$$\frac{\ln(1+\alpha)}{\alpha}|1 - x| \le |\ln x| \le \frac{|\ln(1-\alpha)|}{\alpha}|1 - x| \ .$$

□

In conjunction with Proposition (3.2) (*iii*), Theorem 4.1 has the following corollary, which legitimizes the use of L_ϵ^h as a preconditioner for the Hessian:

COROLLARY 4.2. *There exists a constant $C_1 = C_1(\Omega)$ such that*

$$d_h\left(L_\epsilon^h, \left(H_\epsilon^h\right)^{-1}\right) \le C_1\frac{h^2}{\epsilon T} \ . \tag{4.4}$$

5. The multigrid algorithm.

As pointed in §3, the multigrid iteration will be the simple iteration (3.2) with $H = H_\epsilon^h$ and a recursively defined preconditioner denoted K_ϵ^h. In light of (4.1) we define the affine operator-function $\mathcal{G} : \mathcal{L}(V_{2h}) \to \mathcal{L}(V_h)$ by

$$\mathcal{G}(T) = T\pi_{2h} + (I - \pi_{2h}) \ . \tag{5.1}$$

LEMMA 5.1. *The function $\mathcal{G}$ is non-expanding, i.e. for $T_1, T_2 \in \mathcal{L}_+^s(V_{2h})$*

$$d_h\left(\mathcal{G}(T_1), \mathcal{G}(T_2)\right) \le d_{2h}\left(T_1, T_2\right) \ . \tag{5.2}$$

The simple proof (see [4]) is based on the fact that $\mathcal{G}(\mathcal{L}_+^s(V_{2h})) \subset \mathcal{L}_+^s(V_{2h})$. Since, by Corollary 4.2, $\mathcal{G}((H_\epsilon^{2h})^{-1})$ approximates well $(H_\epsilon^h)^{-1}$, and $(H_\epsilon^{2h})^{-1} \approx K_\epsilon^{2h}$ holds by assumption, Lemma 5.1 would imply that $\mathcal{G}(K_\epsilon^{2h}) \approx H_\epsilon^h$ as well. However, errors may be adding up (see proof of Theorem 5.2), hence we prefer to take K_ϵ^h to be the *first Newton iterate* (see Remark 3.3) starting from $\mathcal{G}(K_\epsilon^{2h})$, that is:

$$K_\epsilon^h \stackrel{\text{def}}{=} \mathcal{N}_{H_\epsilon^h}\left(\mathcal{G}(K_\epsilon^{2h})\right) = 2\mathcal{G}(K_\epsilon^{2h}) - \mathcal{G}(K_\epsilon^{2h})H_\epsilon^h\mathcal{G}(K_\epsilon^{2h}) \ . \tag{5.3}$$

This multigrid iteration resembles, and will be referred to as an $\mathcal{W}$-cycle:

Algorithm 1.(*multigrid iteration, $\mathcal{W}$-cycle*)

- **Input:** u_0, h, b
- **Output:** $u \leftarrow u_0 + K_\epsilon^h(b - H_\epsilon^h u_0)$
- $MG(u_0, h, b)$

if $h \ge h_0$
  then
      $u \leftarrow u_0 + K_\epsilon^{h_0}(b - H_\epsilon^h u_0)$
      return u
  else
    for $i = 1:2$
      $r \leftarrow b - H_\epsilon^h u_{i-1}$
      $r_c \leftarrow \pi_{2h} r$ // restriction of residual to coarse grid
      $u_c = MG(0, 2h, r_c)$ // $u_c \leftarrow K_\epsilon^{2h} r_c$
      $u_i \leftarrow u_{i-1} + x_c + r - r_c$
    end for
return u_2

We assume for simplicity that $K_\epsilon^{h_0} = (H_\epsilon^{h_0})^{-1}$, i.e. at the coarsest level we perform an exact solve. It should be noted that the choice of the coarsest level h_0 depends on ϵ and T: it is the level where the approximation in Theorem 4.1 starts to deteriorate, namely when $h^2 \approx \epsilon T$. We have the following error estimate:

THEOREM 5.2. *Assume that $h_0^2 \le 2^{-5} C_1^{-1} \cdot \epsilon T$, where C_1 is the constant from Theorem 4.1, and that $K_\epsilon^{h_0} = (H_\epsilon^{h_0})^{-1}$. Then*

$$d_h\left(K_\epsilon^h, \left(H_\epsilon^h\right)^{-1}\right) \le 8C_1 \frac{h^4}{(\epsilon T)^2} \ , \quad \text{for } h \in I. \tag{5.4}$$

We give a sketch of the proof and refer the reader to [4] for details.

Proof. Theorem 3.4 implies that, given operators $M, H \in \mathcal{L}_+^s(V_h)$ with $d_h(M, H^{-1}) < 0.4$, we have

$$d_h(\mathcal{N}_H(M), H^{-1}) \le 2\, d_h(M, H^{-1})^2 \ , \tag{5.5}$$

because $|\ln(1-(e^x-1)^2)| \le 2\,x^2$ for $x \in [0, 0.4]$. If $e_{2h} \le 0.2$ and $C_1(\epsilon T)^{-1}h^2 \le 0.1$ then

$$\begin{aligned} e_h \overset{\text{def}}{=} d_h\left(\mathcal{G}(K_\epsilon^{2h}), \left(H_\epsilon^h\right)^{-1}\right) &\overset{\text{triangle ineq.}}{\le} d_h\left(\mathcal{G}(K_\epsilon^{2h}), L_\epsilon^h\right) + d_h\left(L_\epsilon^h, \left(H_\epsilon^h\right)^{-1}\right) \\ &\overset{\text{Cor. 4.2}}{\le} d_h\left(\mathcal{G}(K_\epsilon^{2h}), \mathcal{G}\left(\left(H_\epsilon^{2h}\right)^{-1}\right)\right) + C_1\frac{h^2}{\epsilon T} \\ &\overset{\text{Lemma 5.1}}{\le} \underbrace{d_{2h}\left(K_\epsilon^{2h}, \left(H_\epsilon^{2h}\right)^{-1}\right)}_{e_{2h}} + C_1\frac{h^2}{\epsilon T}\ , \end{aligned}$$

hence by (5.5) we obtain

$$e_h \le 2\left(e_{2h} + C_1\frac{h^2}{\epsilon T}\right)^2\ . \tag{5.6}$$

It is the $\mathcal{W}$-cycle that puts the square on the sum between e_{2h} and the small term $C_1\frac{h^2}{\epsilon T}$. In particular $e_h < 2 * 0.3^2 = 0.18 < 0.2$ and a standard inductive argument will finish the proof. □

The bounds (1.15) imply that cond $\left((H_\epsilon^h)^{\frac{1}{2}}\right) \le C\epsilon^{\frac{1}{2}}$, with C independent of h, hence by Remark 3.6 and Theorem 5.2 we have the following

Corollary 5.3. *Under the assumptions from Theorem 5.2 we have*

$$\|u_h^{\min} - MG(u_0, h, b)\| \le C'\frac{h^4}{\epsilon^{2.5}T^2}\|u_h^{\min} - u_0\|, \quad \text{for } h \in I. \tag{5.7}$$

A practical implementation of Algorithm 1. would probably be a mixture between a $\mathcal{V}$- and a $\mathcal{W}$-cycle by letting the loop on line 6 to run two times at coarse levels and one time at fine levels, with an adaptive way of deciding when to switch. This would lead to an error bound with half the powers in the right-hand side of (5.7).

Under natural assumptions we can prove:

Proposition 5.4 (Work estimates for multigrid iteration). *Let W_i^{res} be the work for computing the residual (line 7 of Algorithm 1.) on V_{h_i} ($h_i = h_0 2^{-i}$, $0 \le i \le N$), and W_i be the work for $MG(\cdot, h_i, \cdot)$. Then*

$$W_N \le \frac{2}{1-2^{-3}}W_N^{\text{res}} + \frac{h_0}{h_N}W_0\ . \tag{5.8}$$

Essentially what we see is that the work of the multigrid iteration is slightly larger than the cost of *two* residual computations plus the added costs of all coarse-grid exact solves, in number of h_0/h_N.

The *full multigrid* algorithm consists, as described in standard texts such as [2] or [3], in using the value obtained from a coarse-mesh calculation as an initial guess for the finer-grid calculation, and then performing a sequence of multigrid iterations. For the standard multigrid algorithm for elliptic problems it is shown that the multigrid iteration gives a resolution-independent error reduction, hence a fixed number of multigrid iterations has to be performed on each grid. In our situation, by Corollary 5.3, the error reduction on V_{h_i} takes the form $\rho_i = C'h_i^4/(\epsilon^{2.5}T^2)$ (hence is **decreasing** as h^4), therefore at fine resolutions only **one** multigrid iteration is needed, and the error reduction will be drastic.

Algorithm 2.(*full multigrid*)

1. $\hat{u}_{h_0} = u_{h_0}^{\min} = \epsilon^{-1}\left(H_\epsilon^{h_0}\right)^{-1}\left(\mathcal{S}^{h_0}(T)\right)^* f_{h_0}$ // use a direct method
2. for $i = 1 : N$
3. $\quad \hat{u}_{h_i} \leftarrow MG(\hat{u}_{h_{i-1}}, h_i, f_{h_i})$ // one iteration at each level
4. end for
5. return $\hat{u}_{h_N}$

THEOREM 5.5 (error and work estimates for the full multigrid algorithm). *Suppose the conditions from Theorem* 4.1 *hold and that* $\rho_{\lfloor \frac{N}{2} \rfloor} < 2^{-2}$. *Then*

$$\|\hat{u}_{h_N} - u_{h_N}^{\min}\| \le C \left(1 - 2^2 \rho_{\lfloor \frac{N}{2} \rfloor}\right)^{-1} \frac{\rho_N h_N^2}{\epsilon T} \left(\|f\| + \|u^{\min}\|\right) , \tag{5.9}$$

with $C = C(\Omega)$. *For the work* $\widehat{W}_N$ *we have the following estimate:*

$$\widehat{W}_N \le \underbrace{\frac{2}{(1-2^{-3})(1-2^{-4})}}_{\approx 2.44} W_N^{\text{res}} + \frac{h_0}{h_N} W_0 \ . \tag{5.10}$$

We would like to note that the full multigrid method solves the system (1.21) much better than needed if h_N is very small; an error factor of $(\epsilon T)^{-1} h_N^2$ would have been enough, but the expression on the right-hand side of (5.9) has an extra ρ_N factor at the numerator. The proof follows closely the idea from [1] and is omitted here.

6. Conclusions and future work. We have shown that the regularized inverse linear parabolic problem (1.11) can be solved at a cost slighly larger than four times the cost of a forward solve (Hessian-vector multiplication costs two forward solves). This work is extended in a few directions. The use of a different regularizer – a multiple of the H^1-norm as opposed to L^2-norm – introduces a Hessian that is the sum of a smoothing and a roughening operator. The preconditioner we use in the H^1-case is defined similarly as for the L^2-case, but it is slightly non-symmetric. The results, however, are similar with the ones discussed in the present paper. We are also hoping to extend these results to mildy nonlinear parabolic equations by adding an $O(1)$ reaction term.

Acknowledgment. The author thanks Todd Dupont for helpful comments in preparing this paper. This work was supported by the ASCI Flash Center at the University of Chicago under DOE contract B532820, and by the MRSEC Program of the National Science Foundation under award DMR-0213745.

REFERENCES

[1] R. E. BANK AND T. DUPONT, *An optimal order process for solving finite element equations*, Math. Comp., 36 (1981), pp. 35–51.

[2] S. C. BRENNER AND L. R. SCOTT, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer-Verlag, New York, second ed., 2002.

[3] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A multigrid tutorial*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2000.

[4] A. DRĂGĂNESCU, PhD thesis. University of Chicago, June 2004 (anticipated).

[5] A. DRĂGĂNESCU AND T. F. DUPONT, *A multigrid algorithm for L^2-regularized inverse parabolic problems*. in preparation.

[6] M. LUSKIN AND R. RANNACHER, *On the smoothing property of the Galerkin method for parabolic equations*, SIAM J. Numer. Anal., 19 (1982), pp. 93–113.

[7] ———, *On the smoothing property of the Crank-Nicolson scheme*, Applicable Anal., 14 (1982/83), pp. 117–135.