Electronic Transactions on Numerical Analysis.
Volume 25, pp. 67-100, 2006.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ERROR CONTROLLED REGULARIZATION BY PROJECTION*

WOLFGANG DAHMEN† AND MARKUS JÜRGENS†

Dedicated to Ed Saff on the occasion of his 60th birthday

Abstract. The paper is concerned with regularization concepts for the inversion of diffusion processes. The application of the involved evolution operators is based on Dunford integral representations combined with the adaptive application of resolvents using recent wavelet methods. In particular, this allows us to develop and realize numerically, to our knowledge for the first time in this context, an SVD projection method which is compared to several versions of Tikhonov-type schemes. The theoretical findings are complemented by numerical tests shedding some light on the quantitative performance of the schemes.

Key words. inverse problems, Dunford integrals, quadrature, Tikhonov method, projection methods, truncated SVD expansion, adaptive wavelet methods

AMS subject classifications. 47A52, 65J20, 65J22

1. Introduction. The notion of "inverse problem" serves as an important conceptual link between experimental and theoretical sciences. For example, unknown system parameters appearing in a model suggested by theory need to be retrieved from typically indirect measurements. "Given an answer one has to look for the question" perhaps best reflects the typical ill-posedness of inverse problems. An important class of such even "severely" ill-posed inverse problems arises in connection with evolution equations of the form

$$\dot{u} + Au = 0, \quad u(0) = u_0, \tag{1.1}$$

where u is a function in time taking values in some Hilbert space X and A is a positive definite operator on $\mathcal{D}(A;X) \subset X$. When $A = -\Delta$, the negative Laplacian, $X = L_2(\Omega)$, where $\Omega \subseteq \mathbb{R}^d$ is an open domain, this is the classical heat equation describing heat diffusion. The "solution" operator S that assigns to any initial data u_0 the solution $u(t)$ at later time $t > 0$ is given by the operator exponential $S = e^{-tA}$ and is known to define an analytic semigroup. When $A : \mathcal{D}(A;X) \to X$ has an unbounded spectrum S will be compact. A classical inverse problem consists in finding the initial data u_0 from the solution $S(u_0)(t) = e^{-tA}u_0$ at some fixed time $t > 0$. Thus one has to "undo diffusion" which is expected to be a delicate task.

A similar situation arises with the inhomogeneous problem

$$\dot{u} + Au = f, \quad u(0) = 0,$$

where the solution operator is given by

$$(Sf)(t) = \int_t^T e^{-(\tau - t)A} f(\tau) d\tau. \tag{1.2}$$

These are special instances of the following general situation. Let $S : X \to Y$ be a linear and compact operator between two Hilbert spaces X and Y. We aim at solving

$$Sx = y \tag{1.3}$$

*Received August 12, 2005. Accepted for publication February 1, 2006. Recommended by D. Lubinsky. This research was supported in part by the EEC Human Potential Programme under contract HPRN-CT-2002-00286, "Breaking Complexity", and the SFB 401, "Flow Modulation and Fluid-Structure Interaction at Airplane Wings", funded by the German Research Foundation.

†RWTH Aachen, Institut für Geometrie und Praktische Mathematik, Templergraben 55, D-52056 Aachen, Germany (dahmen@igpm.rwth-aachen.de).

ETNA
Kent State University
etna@mcs.kent.edu

for some given $y \in Y$. This equation has no solution unless $y \in \mathrm{Range}(S)$. Even if $y \in \mathrm{Range}(S)$, the solution is in general neither unique nor continuously depending on y. A way to restore uniqueness is to impose some additional selection criterion. A theoretical tool to solve (1.3) in this sense is the *Moore-Penrose generalized inverse* $S^\dagger : \mathcal{D}(S^\dagger) \subset Y \to X$, $\mathcal{D}(S^\dagger) = \mathrm{Range}(S) \oplus \mathrm{Range}(S)^\perp$, defined by

$$x^\dagger = S^\dagger y = \int_0^{\|S\|^2} \frac{1}{\lambda}\, dE_\lambda\, S^* y,$$

where $\{E_\lambda\}_{\lambda \geq 0}$ denotes the spectral family of the self-adjoint, compact operator S^*S. The integral converges if and only if y belongs to $\mathcal{D}(S^\dagger; Y)$. Thus when $\mathrm{Range}(S)$ is not closed $y \in Y$ may fail to belong to $\mathcal{D}(S^\dagger; Y)$. Equivalently, $x^\dagger$ can be defined as the *least-squares solution* of (1.3) with minimal norm, i.e.

$$\|Sx^\dagger - y\|_Y = \inf_{z \in X} \|Sz - y\|_Y$$

and $x^\dagger$ has minimal norm among all $z \in X$ for which the infimum is attained.

Still the unboundedness of the (generalized) inverse prevents a continuous dependence of the solution on the data so that *regularization* is necessary. Of course, this subject has been treated in numerous studies documented in the literature. The two most common regularization strategies are Tikhonov type schemes and projection methods. The numerical implementation of Tikhonov schemes requires the repeated approximate application of the solution operator and its dual which is typically done with the aid of time stepping schemes like the method of lines for the solution of the forward problem. This is computationally rather demanding while the so called "qualification" (limiting the attainable order of accuracy in terms of a decreasing noise level, see [9]) is only $\mu_0 = 1$. Among the class of projection methods a prominent role is played by truncated *singular value expansions* (SVD-expansions) which yield in some sense "optimal" projection methods, see e.g. [9]. Unfortunately the required system of singular basis functions is usually not numerically accessible at affordable cost.

The objective of this paper is to develop and analyze certain new variants of both strategies, Tikhonov and projection methods, for the inversion of diffusion processes. Since the essential algorithmic ingredients for the treatment of operators of the form $S = e^{-tA}$ can be carried over to those given by (1.2) (see [15]) we shall exemplify matters for $S = e^{-tA}$. In Tikhonov schemes the explicit regularization parameter affects the whole spectrum of the operator. Therefore one might expect improvements from penalizing only the small part of the spectrum which will lead to one of the variants. This will be facilitated through certain projections based on representations as contour integrals in the complex plane similar to analogous representations of the solution operator itself. A central issue in this paper is therefore the development and analysis of error controlled numerical schemes for the evaluation of such integral representations which will allow us to completely avoid costly time stepping schemes in either regularization method. Perhaps more importantly, this will allow us to realize, to our knowledge for the first time in an efficient way an (approximate) SVD-projection method for the inversion of (the full infinite dimensional) diffusion processes without ever having to determine the singular basis functions.

Our schemes are based on a suitable quadrature rule for the contour integrals combined with an error controlled application of the involved resolvents at the quadrature points. A computationally attractive feature is that the different resolvent calculations are completely independent which leads to a trivial but highly efficient parallelization of the schemes. In principle, the resolvent equations (which are well-posed) can be solved by any discretization that guarantees a given accuracy tolerance. Here we shall employ recent adaptive wavelet

ETNA
Kent State University
etna@mcs.kent.edu

schemes for the following reasons: Due to the induced norm equivalences, the wavelet setting allows us to realize conveniently the relevant topologies arising in this context and to control errors in the right norms. For instance, involved applications of resolvents can be performed with optimal complexity in the sense of [2] within desired error tolerances. Consequently, the application of the *true* operator exponential (not of e^{-tA_h} for some given discretization A_h of A and its inherited inaccuracy) can be given with rigorous accuracy bounds. In this sense the approach allows us to disentangle regularization and discretization of A in favor of transparent and accurate error bounds. Finally, the complexity results for the adaptive schemes from [2] will be seen to offer a first rough assessment of the computational complexity of the SVD-projection regularization. In this context one encounters a number of interesting questions that remain open but indicate in our opinion a promising potential of this approach.

It should be mentioned that a direct approximate application of the operator exponential can be also based on the *hierarchical matrix* concept, see e.g. [11]. However, this is an entirely discrete approach approximating e^{-tA_h} for a given matrix approximation A_h of A. Some comments on the resulting conceptual distinctions can be found at the end of §2.

The layout of the paper is as follows. In §2 we collect some basics from regularization theory and formulate the principal strategies. §3 is devoted to the concept of separating the spectrum of closed linear operators which our projection method will be based upon. We identify a contour integral representation of the boundedly invertible projection of S onto the eigenspace of A corresponding to a bounded part of its spectrum. This will serve later as the regularized approximation to S.

In §4 we present the main regularization algorithm in terms of several basic routines. At this stage these routines will be formulated merely in terms of abstract building blocks each of which describes a numerical task (such as the solution of certain operator equations or the application of the exponential of an operator) to be performed within a given accuracy tolerance. As an important ingredient in this context we analyze in addition the quadrature error arising from the discretization of the integrals and explain how to control this error when using the discrepancy principle.

In principle one could realize the above routines through any of the standard discretization concepts. Here we employ recent adaptive wavelet methods discussed in §5. More precisely, our approach is based on an adaptive paradigm developed in [2] for operators with sparse wavelet representations. The main reasons for this choice can be summarized as follows. The current understanding of adaptive wavelet concepts allows us to draw first conclusions on the complexity of such regularization tasks. Moreover, the norm equivalences characterizing relevant function spaces in terms of weighted sequence norms for wavelet coefficient arrays suggests, as a byproduct, a new regularization scheme employing different smoothness norms. The resulting computable algorithms for the above building blocks hinge on suitable wavelet representations of the involved operators and corresponding adaptive evaluation schemes for their application within any desired accuracy tolerance. These ingredients are also derived in this section.

Finally, in §6 the theoretical findings are quantified and illustrated by numerical experiments.

2. Regularization.

We begin with briefly recalling some facts from regularization theory that will guide the subsequent developments, see [9] for details and proofs of the quoted results.

One way to approximate $S^\dagger$ by a bounded function is to replace $1/\lambda$ by a family of functions $g_\alpha(\lambda)$, $\alpha > 0$, which are bounded and tend to $1/\lambda$ pointwise if α approaches zero.

Abbreviating $\|S\| := \|S\|_{\mathcal{L}(X,Y)}$, the regularized solution is then defined as

$$x_\alpha = g_\alpha(S^*S)S^*y = \int_0^{\|S\|^2} g_\alpha(\lambda)\, dE_\lambda\, S^*y,$$

where the integral converges for every $y \in Y$ as $g_\alpha(\lambda)$ is bounded on $[0, \|S\|^2]$. In particular, the operator $g_\alpha(S^*S)S^* : Y \to X$ is continuous with

$$\|g_\alpha(S^*S)S^*\| \le \sup\{\sqrt{\lambda}\, |g_\alpha(\lambda)| : \lambda \in [0, \|S\|^2]\}.$$

The residual can be explicitly calculated as

$$\begin{aligned} x^\dagger - x_\alpha = x^\dagger - g_\alpha(S^*S)S^*y &= (I - g_\alpha(S^*S)S^*S)x^\dagger \\ &= \int_0^{\|S\|^2} r_\alpha(\lambda)\, dE_\lambda\, x^\dagger \end{aligned}$$

with

$$r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda).$$

The approximation properties of this method are stated in the following theorem.

THEOREM 2.1. ([9], Theorem 4.1) *Let* $g_\alpha(\lambda) : [0, \|S\|^2] \to \mathbb{R}$ *fulfill the following assumptions for all* $\alpha > 0$*:* g_α *is piecewise continuous, there exists a constant* $C_g > 0$ *independent of* α *such that*

$$|\lambda g_\alpha(\lambda)| \le C_g \;\; \textit{and} \;\; \lim_{\alpha\to 0} g_\alpha(\lambda) = \frac{1}{\lambda} \;\; \forall\, \lambda \in (0, \|S\|^2]. \tag{2.1}$$

Then

$$\lim_{\alpha\to 0} g_\alpha(S^*S)S^*y \quad = \; S^\dagger y, \quad \textit{if } y \in \mathcal{D}(S^\dagger), \tag{2.2}$$

and

$$\lim_{\alpha\to 0} \|g_\alpha(S^*S)S^*y\|_X = \infty, \qquad \textit{if } y \notin \mathcal{D}(S^\dagger). \tag{2.3}$$

Concrete rates of convergence in (2.2) can only be expected when $x^\dagger$ belongs to some compact subset of X usually described in terms of "regularity" properties. A typical assumption leading to convergence rates is that $x^\dagger$ belongs to the range of some power of S^*S. Knowing such a convergence rate would be important for dealing with *noisy* data y^δ which are usually encountered in realistic applications. In fact, suppose that the noise level is $\delta > 0$, i.e.

$$\|y - y^\delta\|_Y \le \delta.$$

Accordingly, one obtains perturbed results

$$x_\alpha^\delta = g_\alpha(S^*S)S^*y^\delta.$$

If a rate for $\|x^\dagger - x_\alpha^\delta\|_X$ was known, in view of

$$\|x^\dagger - x_\alpha^\delta\|_X \le \|x^\dagger - x_\alpha\|_X + \|x_\alpha - x_\alpha^\delta\|_X,$$

it would be sufficient to bound the 'propagated error' $\|x_\alpha - x_\alpha^\delta\|_X$ in order to balance both terms and thereby obtain the right regularization.

THEOREM 2.2. ([9], Theorem 4.2) *Let the assumptions of Theorem 2.1 be fulfilled and set for $\alpha > 0$*

$$G_\alpha = \sup\left\{|g_\alpha(\lambda)| : \lambda \in [0, \|S\|^2]\right\}.$$

Then the estimates

$$\|Sx_\alpha - Sx_\alpha^\delta\|_Y \leq C_g\delta, \quad \textit{and} \quad \|x_\alpha - x_\alpha^\delta\|_X \leq \sqrt{C_g G_\alpha}\delta,$$

hold true with C_g from (2.1).

Thus, the propagated error stays proportional to the noise level δ and a bound for $\|x^\dagger - x_\alpha^\delta\|_X$ would tell us how to choose α.

In practice, the necessary information is typically not available and a reasonable parameter value $\alpha = \alpha(\delta, y^\delta) > 0$ has to be chosen for given data y^δ in a different way. A widely used method is the *discrepancy principle* which defines

$$\alpha = \alpha(\delta, y^\delta) = \sup\{\beta > 0 : \|Sx_\beta^\delta - y^\delta\| = \tau\delta\} \tag{2.4}$$

with some fixed parameter

$$\tau > \sup\left\{|r_\alpha(\lambda)| : \alpha > 0, \lambda \in [0, \|S\|^2]\right\}.$$

Under fairly general conditions on $g_\alpha(\lambda)$ convergence

$$x^\delta_{\alpha(\delta,y^\delta)} \to x^\dagger \quad \text{if } \delta \to 0,$$

of optimal order can be shown ([9], Theorem 4.17).

The strict condition (2.4) can be relaxed to finding the largest α such that

$$\|Sx_\alpha^\delta - y^\delta\|_Y \leq \tau\delta \leq \|Sx_{2\alpha}^\delta - y^\delta\|_Y$$

holds which eases the use of the discrepancy principle.

Clearly, x_α^δ is generally again the solution of an operator equation. Thus, in practice, neither x_α^δ nor the residual $\|Sx_\alpha^\delta - y^\delta\|_Y$ can be calculated exactly. Aiming at a rigorous error control we have to analyze the behavior of the discrepancy principle when both quantities are replaced by approximations $\widehat{x}_\alpha^\delta$ and Res_α.

LEMMA 2.3. *Let $\widehat{x}_\alpha^\delta$ be a numerical approximation to the regularized solution x_α^δ satisfying*

$$\|\widehat{x}_\alpha^\delta - x_\alpha^\delta\|_X \leq \tau_1\|S\|^{-1}\delta. \tag{2.5}$$

Moreover, suppose that the approximation to the true residual $S\widehat{x}_\alpha^\delta - y^\delta$ has norm Res_α satisfying

$$\left|\mathrm{Res}_\alpha - \|S\widehat{x}_\alpha^\delta - y^\delta\|_Y\right| \leq \tau_2\delta, \tag{2.6}$$

where the positive constants τ_1, τ_2 are chosen such that

$$\tau - \tau_1 - \tau_2 > \gamma := \sup\left\{|r_\alpha(\lambda)| : \alpha > 0, \lambda \in [0, \|S\|^2]\right\}. \tag{2.7}$$

Now let $\alpha = \alpha(\delta, y^\delta)$ be chosen such that

$$\mathrm{Res}_\alpha \leq \tau\delta \leq \mathrm{Res}_{2\alpha}. \tag{2.8}$$

ETNA
Kent State University
etna@mcs.kent.edu

Then the numerical approximations $\widehat{x}_\alpha^\delta$ to the regularized solutions x_α^δ corresponding to the method $(g_\alpha(S^*S)S^*, \alpha)$ converges in X to $x^\dagger$ for $\delta \to 0$ for all $y \in \mathrm{Range}(S)$. In fact, it has optimal order $\delta^{\frac{2\mu}{2\mu+1}}$ when $x^\dagger \in \mathrm{Range}\,(S^*S)^\mu$ for $\mu \in (0, \mu_0 - 1/2]$, where μ_0 is the "qualification" of the regularization method, see [9, Chapter 4.1].

Proof. We refer to the proof of Theorem 4.17 in [9] were the analogous statement is shown for the discrepancy principle with exact values of x_α^δ and $\|Sx_\alpha^\delta - y^\delta\|_Y$. A careful inspection reveals that the claim is reduced to verifying the two estimates

$$\|Sx_\alpha - y\|_Y \le c_1\delta \quad \text{and} \quad \|Sx_{2\alpha} - y\|_Y > c_2\delta \tag{2.9}$$

for some constants $c_1, c_2 > 0$.

We infer the first part of (2.9) from the assumption $\mathrm{Res}_\alpha \le \tau\delta$ as follows. One has

$$\begin{aligned}\|Sx_\alpha - y\|_Y &\le \|Sx_\alpha^\delta - y^\delta\|_Y + \gamma\delta \le \|S\widehat{x}_\alpha^\delta - y^\delta\|_Y + \|S\|\|x_\alpha^\delta - \widehat{x}_\alpha^\delta\|_X + \gamma\delta \\ &\le \mathrm{Res}_\alpha + \tau_2\delta + \tau_1\delta + \gamma\delta \le (\tau + \tau_1 + \tau_2 + \gamma)\delta,\end{aligned}$$

where the first inequality results from the properties of $r_\alpha(\lambda)$ and the definition of γ.

To see the second estimate of (2.9) we observe that

$$\begin{aligned}\|Sx_{2\alpha}^\delta - y^\delta\|_Y &\ge \|S\widehat{x}_{2\alpha}^\delta - y^\delta\|_Y - \|S\|\|x_{2\alpha}^\delta - \widehat{x}_{2\alpha}^\delta\|_X \le \mathrm{Res}_{2\alpha} - (\tau_1 + \tau_2)\delta \\ &> (\tau - \tau_1 - \tau_2)\delta.\end{aligned}$$

This implies $\|Sx_{2\alpha} - y\|_Y \ge (\tau - \tau_1 - \tau_2 - \gamma)\delta$, where $\tau - \tau_1 - \tau_2 - \gamma$ is positive due to the choice of τ, τ_1, and τ_2. $\square$

Loosely speaking, it is sufficient to keep the discretization errors proportional to the noise level. It should be noticed that this result is purely *asymptotic* for $\delta \to 0^+$. In particular, the constants τ, τ_1, and τ_2 which heavily influence the outcome of numerical experiments can be chosen freely as long as (2.7) is satisfied.

It remains to specify the regularization method g_α. Setting $g_\alpha(\lambda) = (\alpha + \lambda)^{-1}$ yields the well-known *Tikhonov regularization* which reads

$$(\alpha I + S^*S)x_\alpha^\delta = S^*y^\delta. \tag{2.10}$$

x_α^δ can equivalently be defined as solution of the minimization problem

$$\alpha\|x\|_X + \|Sx - y^\delta\|_Y \to \min_{x \in X},$$

which can be interpreted as compromise between the size of $\|x\|_X$ and the residual error $\|Sx - y^\delta\|_Y$. It is well-known that in this case the qualification is $\mu_0 = 1$.

The penalization imposed by the Tikhonov scheme affects the whole spectrum of S^*S. This raises the question whether the resulting bias could perhaps be diminished by making the penalization dependent on the spectrum. In fact, it should be active only on the small eigenvalues of S^*S. Thus, a first alternative is

$$g_\alpha(\lambda) := (\alpha\chi_{[0,\alpha]}(\lambda) + \lambda)^{-1}. \tag{2.11}$$

This gives rise to what may be called "modified" Tikhonov regularization.

The second alternative deviates in a more essential way from the Tikhonov concept by taking

$$g_\alpha(\lambda) := \begin{cases} \frac{1}{\lambda}, & \lambda \ge \alpha; \\ 0, & \lambda < \alpha; \end{cases} \tag{2.12}$$

ETNA
Kent State University
etna@mcs.kent.edu

simply clipping the small eigenvalues of S^*S. This corresponds to setting the small singular values of S to zero and thus to a *truncated singular value expansion* which is known to be in some sense an "optimal" regularization method by projection. Moreover, in this case one has $\mu_0 = \infty$, see e.g. [9, Example 4.8]. We refer to this regularization scheme as SVD-projection. But of course, in practice the issue is to realize this projection method at affordable cost.

Both strategies (2.11), (2.12) hinge on the ability to realize *projections* $Q_\alpha : X \to X$ taking X into an eigenspace corresponding to $\sigma(S^*S) \cap [0, \alpha)$. Then (2.11) gives rise to solving

$$(\alpha Q_\alpha + S^*S)x_\alpha^\delta = S^*y^\delta, \tag{2.13}$$

while (2.12) leads to

$$R_\alpha^* S^* S R_\alpha x_\alpha^\delta = R_\alpha^* S^* y^\delta, \quad R_\alpha := I - Q_\alpha. \tag{2.14}$$

We shall show below that for $S = e^{-tA}$, A as in §1, such projections can be efficiently realized numerically which to our knowledge is new. Before addressing this in a systematic fashion, a few comments are in order. Of course, discretizations of A lead to approximations A_h of A defined on a finite dimensional subspace of X being now bounded with a norm depending on the discretization parameter h. Hence the spectrum of $S_h = e^{-tA_h}$ is bounded away from zero so that solving

$$S_h^* S_h x_\alpha^\delta = S_h^* y^\delta \tag{2.15}$$

has a similar effect as (2.14) where e.g. the mesh size h of the underlying discretization acts as a regularization parameter. Solving (2.15) (which becomes increasingly ill-conditioned for $h \to 0$) iteratively, the termination of the iteration adds another bias. Each application of S_h or S_h^* (either by approximately solving evolution equations or by employing $\mathcal{H}$-matrix techniques, [11]) is inaccurate. For $\mathcal{H}$-matrix concepts the problem is posed now in Euclidean metric which hampers somewhat taking the topological requirements of the original problem into proper account. So several sources of inaccuracies are mixed. We shall therefore pursue here a different line *separating* regularization and discretization. In particular, this will allow us to apply $(R_\alpha^* S^* S R_\alpha)^{-1}$ for the "true" S within any desired tolerance. This will be based on two conceptual pillars: (a) the actual numerical realization of projections of the form Q_α as well as of e^{-tA} within any desired target accuracy; (b) the involved application of resolvents $(\gamma I - A)^{-1}$ appearing in the Dunford integral is based on adaptive wavelet schemes again realizing proper target accuracies without ever fixing any discretization of A beforehand.

We shall now fix our assumptions on the operator A which, of course, relates to the topologies of X and the image space Y under S. A will always be assumed to be *sectorial* and *normal* with spectrum in the right complex half plane. In particular, this covers selfadjoint positive definite operators. Moreover, it will be important to specify the mapping properties of A. To this end, we shall always assume to be given a Gelfand triple of Hilbert spaces H, V, H' satisfying

$$H \hookrightarrow V \hookrightarrow H', \tag{2.16}$$

where H' is the dual of H represented through the inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{V \times V}$ of the pivot space V. In these terms we require that there exist bounded positive constants c_A, C_A such that

$$c_A \|v\|_H \le \|Av\|_{H'} \le C_A \|v\|_H, \qquad u \in H, \tag{2.17}$$

i.e. $A : H \to H'$ is a norm-isomorphism from H onto H'. A typical case is $V = L_2(\Omega)$ and $H = H_0^1(\Omega), H' = H^{-1}(\Omega)$ which fits to the heat equation with $A = -\Delta$, say. This will serve as a guiding example in the selfadjoint positive definite case.

In most part of the subsequent developments we shall have

$$V = X = Y, \tag{2.18}$$

which accounts for the fact that noise should not be measured in too strong a norm. Nevertheless, an alternative would be $X = H'$, so that $\mathcal{D}(A; X) = H = Y$, and later in §5.5 we shall encounter intermediate cases.

REMARK 1. *An analogous mapping property to* (2.17) *does not hold for* $S = e^{-tA}$, *i.e. we cannot find a pair of spaces for which an analogous estimate like* (2.17) *holds. Therefore, the concepts used e.g. in* [3] *do not apply in the present situation.*

3. Separation of Spectrum.

We return to problems of the form (1.1). When X is a Hilbert space and A is self adjoint and positive definite there exists an orthonormal basis of eigenvectors $\{x_k\}_{k\in\mathbb{N}}$ of A with associated eigenvalues $\lambda_k > 0$, sorted by decreasing size. Then, the operator exponential can be written as

$$e^{-tA}u_0 = \sum_{k=1}^{\infty} e^{-t\lambda_k} \langle u_0, x_k\rangle_{X\times X} x_k, \qquad t \geq 0.$$

From this representation, it is immediately clear that

$$\|e^{-tA}u_0\|_X \leq e^{-t\lambda_1}\|u_0\|_X \ \text{ and } \ \sigma(e^{-tA}) \setminus \{0\} = e^{-t\sigma(A)}, \ \ t > 0.$$

Moreover, the eigenvector of e^{-tA} to the eigenvalue $e^{-t\lambda_k}$ coincides with the eigenvector of A to the eigenvalue λ_k. This close relationship can be carried over to a more general situation as stated in the following theorem. To make that more precise, let us recall the following facts relating the spectrum of A to that of e^{-tA} even for the wider class of *sectorial operators*.

THEOREM 3.1. (Spectral mapping theorem, [10, 14]) *Let A be a sectorial operator on a Banach space X. For every $t > 0$*

$$\sigma(e^{-tA}) \setminus \{0\} = e^{-t\sigma(A)} := \{e^{-t\lambda} : \lambda \in \sigma(A)\},$$

and

$$\sigma_p(e^{-tA}) = e^{-t\sigma_p(A)},$$

where σ_p denotes the point spectrum, i.e. the set of eigenvalues. Moreover, the eigenspaces of A and e^{-tA} are related by

$$\mathcal{N}(e^{-t\lambda}I - e^{-tA}) = \overline{\mathrm{lin}}_{k\in\mathbb{Z}}\, \mathcal{N}\Big(\Big(\lambda + \frac{2\pi k i}{t}\Big)I - A\Big)$$

for each $t > 0$ and $\lambda \in \mathbb{C}$.

A representation of projections needed in (2.13), (2.14) can be based on the following principle, known as *separation of the spectrum*, that projects A as well as e^{-tA} onto the eigenvector spaces of A associated with a bounded subset of $\sigma(A)$.

To this end, assume that $\sigma(A)$ is split into two disjoint, nonempty sets σ_1 and σ_2 such that σ_1 lies in the interior of some Jordan curve Γ_1 and σ_2 in the exterior part. Then,

$$P := P_{\Gamma_1} := P_{\Gamma_1}(A) := \frac{1}{2\pi i}\int_{\Gamma_1} (\gamma I - A)^{-1}\, d\gamma \tag{3.1}$$

defines a bounded linear operator. Its basic properties can be summarized as follows.

THEOREM 3.2. (Theorem III.6.17 of [12]; §2.3 of [14]) *The operator P is a projection with* $\mathrm{Range}(P) \subset \mathcal{D}(A^k; X)$ *for every* $k \in \mathbb{N}$. *Moreover, setting*

$$Y_1 = P(X), \quad Y_2 = (I - P)(X),$$

$$A_1 : Y_1 \to Y_1 : x \mapsto Ax \quad \textit{for all } x \in Y_1,$$

$$A_2 : \mathcal{D}(A_2) = \mathcal{D}(A; X) \cap Y_2 \to Y_2 : x \mapsto Ax \ \textit{ for all } x \in \mathcal{D}(A_2),$$

the operator $A_1 : Y_1 \to Y_1$ is linear and bounded, and

$$\begin{aligned} \sigma(A_1) = \sigma_1, \qquad & \sigma(A_2) = \sigma_2, \\ (\gamma I - A_1)^{-1} = (\gamma I - A)^{-1}|_{Y_1}, \qquad & \gamma \in \rho(A), \\ (\gamma I - A_2)^{-1} = (\gamma I - A)^{-1}|_{Y_2}, \qquad & \gamma \in \rho(A). \end{aligned}$$

If, in addition, X is a Hilbert space and A is normal, then P is the orthogonal projection onto Y_1.

Thus, A is split into a bounded part A_1 and an unbounded one A_2. As the projection P is defined by a curve integral similar to the definition of the operator exponential e^{-tA}, both are compatible in the following sense.

PROPOSITION 3.3. ([14], Prop. 2.3.3) *The projection P commutes with e^{-tA}, which implies $e^{-tA}(Y_1) \subset Y_1$ and $e^{-tA}(Y_2) \subset Y_2$, i.e. Y_1 and Y_2 are invariant subspaces of X with respect to the semigroup. A_1 and A_2 generate semigroups in Y_1 and Y_2, respectively, with*

$$\begin{aligned} e^{-tA_1} &= e^{-tA}|_{Y_1} = e^{-tA}P|_{Y_1}, \\ e^{-tA_2} &= e^{-tA}|_{Y_2} = e^{-tA}(I-P)|_{Y_2}. \end{aligned}$$

As A_1 is bounded, $\{e^{-tA_1}\}_{t \ge 0}$ extends to a group $\{e^{-tA_1}\}_{t\in\mathbb{R}}$. In particular, e^{-tA_1}, $t > 0$, is boundedly invertible with inverse

$$(e^{-tA_1})^{-1} = e^{tA_1} = \frac{1}{2\pi i} \int_{\Gamma_1} e^{t\gamma} (\gamma I - A)^{-1} \, d\gamma.$$

Moreover, for every $\sigma > \sup\{\Re(\gamma) : \gamma \in \sigma_1\}$ there exists a constant $M_\sigma \ge 1$ such that

$$\|e^{tA_1}\|_X \le M_\sigma e^{\sigma t}, \qquad t \ge 0. \tag{3.2}$$

To see how this relates to an SVD-projection for $S = e^{-tA}$, suppose that the Jordan curve Γ_1 encloses all (positive) eigenvalues of the operator $B = A^* + A$ that are smaller than $\left|\frac{1}{t}\log\frac{1}{\alpha}\right|$. Recall that when A is normal the product S^*S takes the form

$$e^{-tA^*} e^{-tA} = e^{-tB}, \quad B = A + A^*, \tag{3.3}$$

a single exponential. Since by Theorem 3.1, any $\xi \in \sigma(S^*S)$ has the form $\xi = e^{-t\lambda}$, λ an eigenvalue of B, we infer from Proposition 3.3 and (2.14) that with the projection $R_\alpha = P_{\Gamma_1} = P_{\Gamma_1}(B)$ with respect to B one has

$$x_\alpha^\delta = (P_{\Gamma_1} S^* S P_{\Gamma_1})^\dagger S^* y^\delta = e^{tB_1} e^{-tA^*} y^\delta. \tag{3.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

Hence the SVD-projection reduces here to the application of an exponential and a projected exponential.

Things simplify when A is selfadjoint and positive definite, so that singular values and eigenvalues coincide. Now one needs a curve enclosing all eigenvalues of A up to the size $\left|\frac{1}{2t}\log\frac{1}{\alpha}\right|$ and a projection $P_{\Gamma_1}(A)$ to obtain from (3.4), again in view of the properties stated in Proposition 3.3,

$$x_\alpha^\delta = e^{tA_1}y^\delta. \tag{3.5}$$

Thus regularization along this line amounts to just applying the inverse of a projected semigroup (and in the general case of an exponential) as opposed to solving operator equations of the type (2.10) or (2.13).

Of course, (3.5) can always be used in this form as long as the separation of spectrum is feasible, even when A is not selfadjoint. The regularization scheme (3.5) does then, however, not necessarily agree with an SVD-projection. For simplicity, we shall concentrate in what follows on (3.5), remarking that with the tools developed below, we could realize (3.4) as well.

We close this section with the following observation.

REMARK 2. *The projected semigroup provides also an efficient way for long time integration, i.e. to apply e^{-tA} for $t \gg 0$. Due to Proposition 3.3, one has*

$$\|e^{-tA}u_0 - e^{-tA}Pu_0\|_X = \|(I-P)e^{-tA}u_0\|_X \le e^{-t\inf\Re(\sigma_2)}\|u_0\|_X.$$

Thus, for $\inf\Re(\sigma_2) > 0$ and $t \to \infty$, the semigroup e^{-tA} is closely approximated by the projected semigroup and it is sufficient to apply $e^{-tA}P$ instead of e^{-tA}. Moreover, when Γ is a suitable open curve separating the spectrum of A from the rest of the complex plane we have formally the same Dunford integral representation

$$e^{-tA}v = e^{-tA}Pv = \frac{1}{2\pi i}\int_\Gamma e^{-t\gamma}(\gamma I - A)^{-1}d\gamma, \quad t \ge 0, \tag{3.6}$$

of the full exponential which is valid for sectorial operators.

4. Algorithmic Concepts.

In this section we shall formulate the main algorithm, which realizes an appropriate regularization scheme based on a version of the discrepancy principle that accounts for numerical errors in a theoretically sound way.

To this end, we shall first identify the main building blocks of this scheme which we formulate at this point as abstract routines that perform certain tasks within any prescribed accuracy tolerance.

The first ingredient is the solution of an operator equation

$$Bu = v, \tag{4.1}$$

where either $B : H \to H'$ or $B : V \to V$ is invertible and where in either case accuracy is to be controlled in V, recall (2.18). Thus we require a routine

$$w \leftarrow \textsc{Solve}(B,\eta,v), \quad \text{such that } \|w - B^{-1}v\|_V \le \eta. \tag{4.2}$$

In view of (2.10) and (2.13), we will have to use this routine for instance with

$$B = \alpha I + S^*S \quad \text{or} \quad B = \alpha Q_\alpha + S^*S, \quad v = S^*y^\delta,\ S = e^{-tA},\ t > 0. \tag{4.3}$$

ETNA
Kent State University
etna@mcs.kent.edu

Note that in both cases the right hand side is the result of the application of S^* while the regularization (3.5) requires the application of e^{tA_1}. In both cases the integrals (3.6) and (3.2) involve the application of resolvents which corresponds to (4.1) with

$$B = \gamma I - A, \quad v = u_0. \tag{4.4}$$

Thus the second main building block (which in turn will invoke the SOLVE-routine) is the approximate application of operators of the form e^{-tA}, $t > 0$, or e^{tA_1}, $t \in \mathbb{R}$. We shall formalize this as a routine

$$w \leftarrow \text{APPLY}(t, \Gamma, \eta, v) \quad \text{such that} \quad \|w - e^{-tA} P_\Gamma v\|_X \leq \eta. \tag{4.5}$$

Recall that when Γ separates the complete spectrum we have $e^{-tA}P_\Gamma = e^{-tA}$ and t must be positive.

To make our approach as transparent as possible, we shall indicate first how to realize (4.5) given (4.2) and how to realize the discrepancy principle given (4.5) postponing the concrete realization of (4.2) to the next section, see §5.

4.1. Applying Projections. We are now prepared to formulate algorithms for computing the application of operators of the form

$$e^{-tA}P_{\Gamma_1} v = \frac{1}{2\pi i}\int_{\Gamma_1} e^{-t\gamma}(\gamma I - A)^{-1} v \, d\gamma, \qquad t \in \mathbb{R},$$

or as the special case $t = 0$, of the projection

$$P_{\Gamma_1} v = \frac{1}{2\pi i}\int_{\Gamma_1} (\gamma I - A)^{-1} v \, d\gamma,$$

to some fixed $v \in X$. We emphasize that in connection with projected semigroups t may attain negative values as well. Aside from being able to apply resolvents within a given accuracy tolerance (see (4.4)) we shall need a suitable quadrature scheme.

Our convention will be that when writing $\Gamma = \Gamma_1$ we assume that the Jordan curve Γ_1 encloses some bounded subset of $\sigma(A)$, as described in §3, and that an analytic and 2π-periodic parameterization

$$s \mapsto \gamma(s), \qquad s \in [0, 2\pi),$$

is given. Then, the integrand

$$F(s) = -i\gamma'(s)e^{-t\gamma(s)}(\gamma(s)I - A)^{-1}v, \qquad s \in [0, 2\pi),$$

is also 2π-periodic and analytic in the topology of X so that the following theorem applies.

THEOREM 4.1. ([13], Theorem 12.6) *Let $F : \mathbb{R} \to X$ be analytic and 2π-periodic. Then, there exists $q > 0$ such that F extends to an analytic, bounded, and 2π-periodic function on the strip $\mathcal{D}_q = \{x + iy : x \in \mathbb{R}, |y| < q\}$. Moreover, the quadrature error of the trapezoidal rule*

$$Q_N(F) = \frac{1}{2N}\sum_{k=0}^{2N-1} F\Big(\frac{k\pi}{N}\Big), \qquad N \in \mathbb{N},$$

is bounded by

$$\Big\|\frac{1}{2\pi}\int_0^{2\pi} F(s)\, ds - Q_N(F)\Big\|_X \leq \|F\|_{L_\infty(\mathcal{D}_q, X)} \frac{1}{e^{2qN} - 1}. \tag{4.6}$$

ETNA
Kent State University
etna@mcs.kent.edu

Analogous statements hold when replacing Γ_1 by an open curve Γ to represent the full exponential (for $t > 0$). For corresponding details on suitable analytic parameterizations we refer e.g. to [16] and will make use of these results without further elaboration.

Since we are lacking any estimate for q and $\|F\|_{L_\infty(\mathcal{D}_q,X)}$ it seems reasonable to use

$$\|Q_N(F) - Q_{2N}(F)\|_X \approx \|Q_N(F) - e^{-tA}P_{\Gamma_1}v\|_X$$

as error estimator. This is certainly true when (4.6) holds with $\leq$ replaced by $\sim$. To ensure convergence of the upcoming algorithm we therefore make the following

ASSUMPTION 1. *There exist constants* $c_Q, C_Q > 0$ *and* $N_0 \geq 1$ *such that*

$$c_Q\|Q_N(F) - Q_{2N}(F)\|_X \leq \|Q_N(F) - e^{-tA}P_{\Gamma_1}v\|_X$$

and

$$\|Q_N(F) - e^{-tA}P_{\Gamma_1}v\|_X \leq C_Q\|Q_N(F) - Q_{2N}(F)\|_X$$

for all $N \geq N_0$.

To formulate the following algorithm, we define $N_l = 2^l N_0$, $l \in \mathbb{N}$. The quadrature points associated with N_l are $s_{l,k} = \frac{k\pi}{N_l}$, $k = 0, 1, \ldots, 2N_l - 1$. Obviously, the sequences of quadrature points are nested, i.e. $s_{l+1,2k} = s_{l,k}$. Therefore, $Q_{N_l}(F)$ can be improved to $Q_{2N_l}(F)$ by evaluating the integrand F at the nodes $s_{l+1,2k+1}$, $k = 0, 1, \ldots, N_l - 1$.

ALGORITHM 1. $w \leftarrow \text{APPLY}(t, \Gamma_1, A, \eta, v)$

- *Fix some* $0 < \varrho \leq \frac{1}{3C_Q}$
- *For* $k = 0, 1, \ldots, 2N_0 - 1$,

  $$w_{0,k} \leftarrow \text{SOLVE}(\gamma(s_{0,k})I - A, \frac{\eta \min\{1, \frac{3}{2}\varrho\}}{3|\gamma'(s_{0,k})\exp(-t\gamma(s_{0,k}))|}, v)$$
- $w_0 \leftarrow \frac{-i}{2N_0}\sum_{k=0}^{2N_0-1}\gamma'(s_{0,k})e^{-t\gamma(s_{0,k})}w_{0,k}$
- *For* $l = 1, 2, \ldots$
  - *for* $k = 1, 3, \ldots, 2N_l - 1$

    $$w_{l,k} \leftarrow \text{SOLVE}(\gamma(s_{l,k})I - A, \frac{\eta \min\{1, \frac{3}{2}\varrho\}}{3|\gamma'(s_{l,k})\exp(-t\gamma(s_{l,k}))|}, v)$$
  - $u_l \leftarrow \frac{-i}{2N_l}\sum_{k=0}^{N_l-1}\gamma'(s_{l,2k+1})e^{-t\gamma(s_{l,2k+1})}w_{l,2k+1}$
  - $d_l \leftarrow \frac{1}{2}w_{l-1} - u_l$
  - $w_l \leftarrow \frac{1}{2}(w_{l-1} + u_l)$
  - *If* $\|d_l\|_X \leq \varrho\eta$ *then return* $w = w_{l-1}$.

REMARK 3. *To remain consistent with the theory the algorithm returns* w_{l-1}. *In practice,* w_l *should be returned as it is the more accurate approximation.*

LEMMA 4.2. *If Assumption 1 holds the algorithm* $\text{APPLY}(t, \Gamma_1, A, \eta, v)$ *terminates after a finite number of steps for every* $\eta > 0$ *and its output* w *satisfies*

$$\|w - e^{-tA}P_{\Gamma_1}v\|_X \leq \eta.$$

Proof. Let $\widehat{w}_{l,k}$ denote the exact solution of

$$(\gamma(s_{l,k})I - A)\widehat{w}_{l,k} = v,$$

ETNA
Kent State University
etna@mcs.kent.edu

and define $\widehat{u}_l$, $\widehat{d}_l$ and $\widehat{w}_l$ as the corresponding quantities from the algorithm. According to the error bounds guaranteed by SOLVE$(\gamma(s_{l,k}), \eta, v)$ we have

$$\begin{aligned}\|w_{l-1} - \widehat{w}_{l-1}\|_X &\le \frac{1}{2N_{l-1}} \sum_{k=0}^{2N_{l-1}-1} |\gamma'(s_{l-1,k}) e^{-t\gamma(s_{l-1,k})}| \cdot \|w_{l-1,k} - \widehat{w}_{l-1,k}\|_X \\ &\le \frac{1}{2N_{l-1}} \sum_{k=0}^{2N_{l-1}-1} \frac{\eta}{3} \min\{1, \frac{3}{2}\varrho\} \le \frac{1}{3}\eta \min\{1, \frac{3}{2}\varrho\}.\end{aligned}$$

The same argument yields

$$\|u_l - \widehat{u}_l\|_X \le \frac{1}{6}\eta \min\{1, \frac{3}{2}\varrho\}, \quad \|d_l - \widehat{d}_l\|_X \le \frac{1}{3}\eta \min\{1, \frac{3}{2}\varrho\}.$$

First, the termination criterion is shown to be met for sufficiently large $l \in \mathbb{N}$. From Theorem 4.1 and Assumption 1 the estimate

$$\begin{aligned}\|d_l\|_X &\le \|\widehat{d}_l\|_X + \|d_l - \widehat{d}_l\|_{\ell_2(\mathcal{J})} \\ &\le (c_Q)^{-1} \|F\|_{L_\infty(\mathcal{D}_q, X)} \frac{1}{e^{2dN_l} - 1} + \frac{1}{2}\varrho\eta\end{aligned}$$

follows. Since the first summand tends to zero for $l \to \infty$ there exist $l_0 \in \mathbb{N}$ such that the termination criterion $\|d_{l_0}\|_X \le \varrho\eta$ is indeed met.

In a second step, the error bound is shown to hold for $w = w_{l_0-1}$. Obviously, one has

$$\begin{aligned}\|w_{l_0-1} - e^{-tA} P_{\Gamma_1} v\|_X &\le \|w_{l_0-1} - \widehat{w}_{l_0-1}\|_X + \|\widehat{w}_{l_0-1} - e^{-tA} P_{\Gamma_1} v\|_X \\ &\le \frac{1}{3}\eta + \|\widehat{w}_{l_0-1} - e^{-tA} P_{\Gamma_1} v\|_X.\end{aligned}$$

The second summand is bounded with the aid of Assumption 1,

$$\begin{aligned}\|\widehat{w}_{l_0-1} - e^{-tA} P_\Gamma v\|_X &\le \|Q_{N_{l_0-1}}(F) - e^{-tA} P_\Gamma v\|_X \\ &\le C_Q \|Q_{N_{l_0-1}}(F) - Q_{N_{l_0}}(F)\|_X \\ &\le C_Q \|\widehat{w}_{l_0-1} - \widehat{w}_{l_0}\|_X \\ &\le C_Q \left(\|d_{l_0}\|_X + \|\widehat{d}_{l_0} - d_{l_0}\|_X \right). \end{aligned} \tag{4.7}$$

Since the termination criterion is met the first summand in the parenthss is bounded by

$$\|d_{l_0}\|_X \le \rho\eta \le \frac{\eta}{3C_Q}.$$

The second one can be estimated by

$$\|\widehat{d}_{l_0-1} - d_{l_0-1}\|_X \le \frac{1}{2}\varrho\eta \le \frac{\eta}{4C_Q}.$$

Inserting both estimates in (4.7) yields

$$\|\widehat{w}_{l_0-1} - e^{-tA} P_\Gamma v\|_X \le \frac{2}{3}\eta.$$

Thus, the overall error of the output w is bounded by η. $\square$

It should be noted that the explicit knowledge of the constant c_Q is not needed in the algorithm. Its existence is merely needed to show that the algorithm terminates (i.e. it is actually an algorithm).

REMARK 4. *An analogous scheme* APPLY(t, Γ, A, η, v) *can be constructed when* Γ *is a suitable open curve separating the spectrum* $\sigma(A)$ *from the rest of the complex plane and* t *is positive, see* [16]. *We shall use the same notation for this version as well.*

Since the constants in the quadrature rule are not numerically accessible, one might think of extrapolation techniques to overcome this handicap. But extrapolation is useless for analytic periodic functions since all constants in the asymptotic error expansion of the trapezoidal rule vanish.

4.2. Approximate Regularized Inversion of e^{-tA}. We shall use next the tools developed in the preceding section in order to regularize the inversion of the operator $S = e^{-tA}$ for some fixed time $t > 0$. Since S is one-to-one and compact the operator fulfills the assumptions made in §2. We shall work in the setting (2.18).

The following algorithm comes in three variants each of which is based on Lemma 2.3. We assume that the three parameters $\tau, \tau_1, \tau_2 > 0$ are given such that $\tau - \tau_1 - \tau_2 > 1$ holds (cf. (2.7)). Moreover, we fix $c \in [1/2, 1)$ and test the stopping criterion

$$\|Sx_{\alpha_{k+1}} - y^\delta\|_V \le \tau\delta \le \|Sx^\delta_{\alpha_k} - y^\delta\|_V$$

for $\alpha_k = c^k\alpha_0$. Recall that we have here $V = X = Y$. The choice of c permits a certain tuning of the algorithm, as the search for the optimal α becomes more thorough and more expensive for c closer to one.

The three ways of determining a sequence of regularized solutions $x^\delta_{\alpha_k}$ are based on the two variants (2.10), (2.13) of the Tikhonov method or on the projection method (3.5). As a core ingredient of the main scheme we shall provide for each variant a routine

$$x \leftarrow \text{REG}(\alpha, \eta, y^\delta) \quad \text{satisfying} \quad \|x - x^\delta_\alpha\|_V \le \eta.$$

In the cases (2.10), (2.13) the realization of REG is given by the scheme SOLVE for solving the operator equations (4.1) corresponding to the choices (4.3).

The third method is quite different and uses the fact that, by Proposition 3.3, the projection of $S = e^{-tA}$ onto the eigenspaces associated with a bounded subset of $\sigma(A)$ becomes boundedly invertible.

To fix ideas we briefly specify the discussion in §3 and assume that $\sigma(A) = \{\lambda_k : k \in \mathbb{N}\}$ with $\Re(\lambda_k) \le \Re(\lambda_{k+1})$ and $\Re(\lambda_k) \to \infty$ for $k \to \infty$. Then, by the spectral mapping theorem, e^{-tA} has the eigenvalues $e^{-t\lambda_k}$ which tend to zero if $k \to \infty$. Thus, we discard the eigenvalues for large k by projection. Fix a $k_0 \in \mathbb{N}$ and set $\sigma_1 = \{\lambda_k : 1 \le k \le k_0\}$. According to the notation of §3, the projection onto the eigenspaces associated with σ_1 is denoted by P_{Γ_1}. The projected semigroup $P_{\Gamma_1}e^{-tA} = e^{-tA_1} : V \to V$ is boundedly invertible with inverse

$$e^{tA_1} = \frac{1}{2\pi i}\int_{\Gamma_1} e^{t\gamma}(\gamma I - A)^{-1}\, d\gamma$$

and

$$\|e^{tA_1}\|_{\mathcal{L}(X)} \le Ce^{t\Re(\lambda_{k_0})},$$

for some constant $C = C(k_0) \ge 1$. Here we take again $V = X = Y$. Thus, the regularized solution defined as

$$x^\delta_{k_0} = e^{tA_1}y^\delta,$$

depends continuously on y^δ, and $\alpha = 1/k_0$ plays the role of the regularization parameter, see (3.5). Clearly, if $k_0 \to \infty$ (i.e. $\alpha \to 0$) and y is given exactly the regularized solution x_{k_0} will tend to $x^\dagger$ if A has a complete system of eigenvectors.

When dealing with noisy data, k_0 has to be chosen properly. This can be done again with the aid of the discrepancy principle, i.e. we pick

$$k_0 = \inf\{k \in N : \|e^{-tA} x_k^\delta - y^\delta\|_V \le \tau\delta\}$$

for some fixed parameter $\tau \ge 1$.

In summary, the three variants of the main algorithm arise from the choice of the corresponding auxiliary routine $\textsc{Reg}(\alpha, \eta, y^\delta)$ which can be any one of the following three:

1. $\textsc{Solve}(\alpha I + S^*S, \eta, S^* y^\delta)$,
2. $\textsc{Solve}(\alpha Q_\alpha + S^*S, \eta, S^* y^\delta)$, c.f. (4.3),
3. $\textsc{Apply}(-t, \Gamma_1(\alpha), A, \eta, y^\delta)$.

Each of these routines returns a regularized solution of the ill-posed equation $Sx = y^\delta$.

Of course, embedded in the routines SOLVE is a way for evaluating the given data y^δ (depending on the format in which they are given). This is explicitly used when computing residuals. We refer to a corresponding routine

$$\hat{w} \leftarrow \textsc{Eval}(\eta, w) \quad \text{satisfying} \quad \|\hat{w} - w\|_V \le \eta.$$

The realization of the discrepancy principle for either of these three choices reads now as follows.

ALGORITHM 2. $x \leftarrow \textsc{DiscrepancyPrinciple}(y^\delta, \delta, \alpha_0, \textsc{Reg})$

For k = 0, 1, ...

- $\alpha_k := c^k \alpha_0$
- $x_k \leftarrow \textsc{Reg}(\alpha_k, \tau_1 \|S\|^{-1}\delta, y^\delta)$
- $w_k \leftarrow \textsc{Apply}(t, \Gamma, A + A^*, \frac{1}{2}\tau_2\delta, x_k)$
- $r_k \leftarrow w_k - \textsc{Eval}(\frac{1}{2}\tau_2\delta, y^\delta)$
- *If* $\|r_k\|_{\ell_2(\mathcal{J})} \le \tau\delta$ *then return* $x \leftarrow x_k$.

THEOREM 4.3. *Algorithm* DISCREPANCYPRINCIPLE *defines a regularization scheme of optimal order for* $0 < \mu \le 1/2$ *for the Tikhonov-type schemes (cases* 1. *and* 2.*) and for any positive* μ *for the projection scheme. More precisely, the output* x *of* DISCREPANCYPRINCIPLE $(y^\delta, \delta, \alpha_0)$ *satisfies*

$$\|x - x^\dagger\|_V \le C\delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \tag{4.8}$$

for all $y^\delta \in Y$ *with* $\|y^\delta - y\|_V \le \delta$ *and all* $x^\dagger = (S^*S)^\mu w$ *with* $\|w\|_V \le \rho$*, under the above restrictions on* μ*.*

Proof. From the target accuracy in REGSOLVE, APPLY, and EVALRHS it is immediate that the conditions (2.5) and (2.6) of Lemma 2.3 are satisfied for x_k and $\mathrm{Res}_\alpha = \|r_k\|_X$. The algorithm terminates if and only if (2.8) is satisfied. Therefore we infer from Lemma 2.3 the validity of (4.8), recalling that $\mu_0 = 1$ for the Tikhonov scheme. ☐

As pointed out at the end of §3 the version 3. becomes an SVD-projection when A is selfadjoint and positive definite. In this case the qualification is $\mu_0 = \infty$, i.e. the range of the smoothness index μ that can be exploited is now unbounded.

Of course, in principle, the projection method is a familiar concept in regularization theory. But, to our knowledge, it has not been implemented yet in the present context with a rigorous error control since the eigenvalues and eigenspaces are in general not numerically accessible. By the techniques developed in [15] and §4.1 the scheme becomes for the first time realizable in a fairly general setting.

A rough comparison already indicates at this point that, aside from the fact that higher order regularity can be exploited better, the projection method is computationally more efficient than the Tikhonov scheme. In our setting, for some fixed *positive* $t > 0$ the algorithm $\text{APPLY}(-t, \Gamma_1, A, \eta, y^\delta)$ will be called for some $\eta \ll \delta$ until the discrepancy principle is met. Each call will need N elliptic subproblems to be solved within a prescribed tolerance where we observed N to be well below 80 in our experiments. In contrast, the Tikhonov method requires several calls of $\text{SOLVE}(\alpha I + S^* S, \eta, S^* y^\delta)$. Any linear solver applied to this problem will perform an application of $\text{APPLY}(t, \Gamma, A + A^*, \eta, w)$ *once per step* and will need several steps to solve the linear problem within the desired tolerance. In fact, it will need more and more steps as α decreases. As $\text{APPLY}(-t, \Gamma_1, A, \eta, y^\delta)$ and $\text{APPLY}(t, \Gamma, A + A^*, \eta, w)$ need at least for small t roughly the same amount of work the projection method is significantly faster than the Tikhonov scheme.

Of course, it is crucial to come up with suitable curves Γ_1 separating the spectrum according to the needs of the discrepancy principle. Moreover, Assumption 1 needs to be validated. These issues will be addressed in §6 for a simple test case.

5. Realization of Computational Routines. So far we have left open how to realize the algorithmic building blocks SOLVE, APPLY as well as EVAL numerically. We shall address this issue now for the canonical choice $X = V$ (e.g. $L_2(\Omega)$). The main tool will be a suitably chosen wavelet basis. This leads to computable routines with guaranteed error bounds and some information on the computational complexity of SOLVE for (4.3) and (4.4).

Moreover, as an interesting byproduct, the fact that wavelet bases allow us to characterize the spaces $H^s(\Omega)$ by simple rescaling, a new regularization method which measures the residual $Sx_\alpha^\delta - y^\delta$ in the norm of $H^s(\Omega)$, $-1 \leq s \leq 1$, suggests itself that will later be compared with the above versions.

5.1. Wavelets for Operator Equations. We shall not give any details on specific realizations of wavelet bases but collect only their relevant properties. How to realize them has been discussed in the literature and we refer to references given e.g. in [7, 4]. As for some notational conventions, a wavelet basis will be denoted by $\boldsymbol{\Psi} = \{\psi_\lambda : \lambda \in \mathcal{J}\}$ where $\mathcal{J}$ is a countable index set whose elements typically encode the scale, the spatial location and the type of the wavelets ψ_λ. The wavelets will always be assumed to be *local*, i.e. $\operatorname{diam}(\operatorname{supp}\psi_\lambda) \sim 2^{-|\lambda|}$, where $|\lambda|$ denotes the scale, i.e. the dyadic refinement level on which the wavelet lives. Our first requirement is that $\boldsymbol{\Psi}$ be a *Riesz basis* for V. This means that every $v \in V$ has a *unique* expansion $v = \sum_{\lambda\in\mathcal{J}} \langle v, \tilde\psi_\lambda\rangle\, \psi_\lambda$ which is stable in the following sense. Denoting by $\mathbf{v} = (v_\lambda)_{\lambda\in\mathcal{J}}$ the array of wavelet coefficients of v, there exists bounded constants $c_R, C_R > 0$ such that

$$c_R \|\mathbf{v}\|_{\ell_2(\mathcal{J})} \leq \|v\|_V \leq C_R \|\mathbf{v}\|_{\ell_2(\mathcal{J})}, \qquad v \in V. \tag{5.1}$$

It is well known that this implies the existence of a *dual* Riesz basis $\widetilde{\boldsymbol{\Psi}} = \{\tilde\psi_\lambda : \lambda \in \mathcal{J}\} \subset V$ for V, i.e. $\langle \tilde\psi_\lambda, \psi_{\lambda'}\rangle = \delta_{\lambda,\lambda'}$, so that $v_\lambda = \langle v, \psi_\lambda\rangle$, and one has the alternate expansion $= \sum_{\lambda\in\mathcal{J}} \langle \psi_\lambda, v\rangle\, \tilde\psi_\lambda$, see e.g. [5]. Viewing $\boldsymbol{\Psi}, \widetilde{\boldsymbol{\Psi}}$ as column vectors with respect to some fixed ordering of the indices, $\langle v, \widetilde{\boldsymbol{\Psi}}\rangle$, $\langle \widetilde{\boldsymbol{\Psi}}, v\rangle$ are the arrays of wavelet coefficients in the first representation of v viewed as row and column vectors, respectively. The above expansions can then conveniently be rewritten as

$$v = \langle v, \widetilde{\boldsymbol{\Psi}}\rangle \boldsymbol{\Psi} = \langle v, \boldsymbol{\Psi}\rangle \widetilde{\boldsymbol{\Psi}},$$

ETNA
Kent State University
etna@mcs.kent.edu

where we formally view the expansions as "inner products" of the coefficient arrays with the bases. As a default we shall view sequences as column vectors.

Moreover, a duality argument shows that in these terms

$$\frac{1}{C_R}\|\langle \mathbf{\Psi}, v\rangle\|_{\ell_2(\mathcal{J})} \le \|v\|_V \le \frac{1}{c_R}\|\langle \mathbf{\Psi}, v\rangle\|_{\ell_2(\mathcal{J})}, \qquad v \in V.$$

For $V = L_2(\Omega)$ and a wide range of domains one can construct $\mathbf{\Psi}$ and $\widetilde{\mathbf{\Psi}}$ simultaneously in a way that the primal as well as dual wavelets are compactly supported. Moreover, these constructions actually offer more, namely that properly scaled versions of $\mathbf{\Psi}$ and $\widetilde{\mathbf{\Psi}}$ form Riesz bases for certain smoothness spaces H, H', respectively, appearing in (2.16). More precisely, for H from (2.16), we shall assume (in agreement with all known constructions) from now on that for the positive diagonal matrix

$$\mathbf{D} = \mathrm{diag}\{d_\lambda : \lambda \in \mathcal{J}\}, \quad d_\lambda := \max\{1, \|\psi_\lambda\|_H\}, \tag{5.2}$$

one has

$$v \in H \iff \mathbf{D}\mathbf{v} \in \ell_2(\mathcal{J}), \quad \text{where } \mathbf{v} = \langle \widetilde{\mathbf{\Psi}}, v\rangle, \text{ i.e. } v = \mathbf{v}^T\mathbf{\Psi},$$

and that the norm equivalence

$$c_H\|\mathbf{D}\mathbf{v}\|_{\ell_2(\mathcal{J})} \le \|v\|_H \le C_H\|\mathbf{D}\mathbf{v}\|_{\ell_2(\mathcal{J})}, \qquad v = \mathbf{v}^T\mathbf{\Psi} \in H, \tag{5.3}$$

holds with constants $c_H, C_H > 0$. Again it follows from biorthogonality and duality that

$$C_H^{-1}\|\mathbf{D}^{-1}\tilde{\mathbf{v}}\|_{\ell_2(\mathcal{J})} \le \|v\|_{H'} \le c_H^{-1}\|\mathbf{D}^{-1}\tilde{\mathbf{v}}\|_{\ell_2(\mathcal{J})}, \qquad v = \tilde{\mathbf{v}}^T\widetilde{\mathbf{\Psi}} \in H'.$$

In these terms the rescaled bases $\mathbf{D}^{-1}\mathbf{\Psi}$ and $\mathbf{D}\widetilde{\mathbf{\Psi}}$ are Riesz bases of H and H', respectively.

This has been realized for Sobolev spaces $H = H^t(\Omega)$ or $H_0^t(\Omega)$, $t < 3/2$. In particular, for $H = H_0^1(\Omega)$ the scaling weights are of the order $d_\lambda = 2^{|\lambda|}$. This suggests working with the rescaled bases exclusively, if one is only interested in the spaces H and H'. However, we are going to employ wavelet representations of operators from H onto H' (and even from $H(\gamma) \to H(\gamma)'$ for certain $V \subset H(\gamma) \subseteq H$) as well as of endomorphisms of V. Recording the dependence on $\mathbf{D}$ (and further scalings) explicitly will enable us to adapt to the relevant topologies. Therefore, expansion coefficients will always relate to the V-normalized bases.

With these preparations we can transform now any variational problem of the form: Find $u \in B$ such that for $f \in H'$

$$\langle v, Bu\rangle = \langle v, f\rangle, \quad v \in H, \tag{5.4}$$

into an equivalent one in the wavelet coordinate domain. To this end, it will be convenient to define for any two countable collections Θ, Ξ of functions, ordered in an unspecified but fixed way, the corresponding generalized Gramian by

$$\langle \Theta, \Xi\rangle := \big(\langle \theta, \xi\rangle\big)_{\theta\in\Theta, \xi\in\Xi},$$

i.e. the first entry is formally viewed as a column vector while the second one acts as a row vector which explains the subsequent formal manipulations. Using the scaled basis functions as test functions in (5.4), shows that the original variational problem (5.4) is equivalent to the infinite-dimensional system

$$\langle \mathbf{D}^{-1}\mathbf{\Psi}, Bu\rangle = \langle \mathbf{D}^{-1}\mathbf{\Psi}, f\rangle,$$

ETNA
Kent State University
etna@mcs.kent.edu

as $\mathbf{D}^{-1}\mathbf{\Psi}$ is complete in H. Inserting the representation $u = \mathbf{u}^T\mathbf{\Psi}$ into the previous equation, yields

$$\mathbf{D}^{-1}\underbrace{\langle \mathbf{\Psi}, B\mathbf{\Psi}\rangle}_{=\mathbf{B}}\mathbf{D}^{-1}(\mathbf{D}\mathbf{u}) = \mathbf{D}^{-1}\underbrace{\langle \mathbf{\Psi}, f\rangle}_{=\mathbf{f}}. \tag{5.5}$$

$\mathbf{B} = \langle \Psi, B\Psi\rangle$ is called the *(standard) wavelet representation* of B with respect to Ψ. It is well known that, when B satisfies a relation like (2.17) and the basis satisfies (5.3), this latter problem is now well-posed in $\ell_2(\mathcal{J})$, see e.g. [2]. The relevant properties may be summarized as follows.

REMARK 5. *For every bounded linear operator $B : H \to H'$ we have*

$$\|\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}\|_{\mathcal{L}(\ell_2(\mathcal{J}),\ell_2(\mathcal{J}))} \le C_H^2\|B\|_{\mathcal{L}(H,H')}.$$

If (2.17) holds for B with constants c_B, C_B, i.e. $B^{-1} \in \mathcal{L}(H', H)$ exists, we have

$$c_B c_H^2\|\mathbf{v}\|_{\ell_2(\mathcal{J})} \le \|\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}\,\mathbf{v}\|_{\ell_2(\mathcal{J})} \le C_B C_H^2\|\mathbf{v}\|_{\ell_2(\mathcal{J})}, \qquad \mathbf{v} \in \ell_2(\mathcal{J}). \tag{5.6}$$

i.e. the representation of B with respect to the scaled basis $\mathbf{D}^{-1}\mathbf{\Psi}$ for H. Likewise, for every continuous endomorphism $B : V \to V$ the estimates

$$\|\langle \mathbf{\Psi}, B\mathbf{\Psi}\rangle\|_{\mathcal{L}(\ell_2(\mathcal{J}),\ell_2(\mathcal{J}))} \le C_R^2\|B\|_{\mathcal{L}(V,V)}$$

and

$$\|\langle \widetilde{\mathbf{\Psi}}, B\widetilde{\mathbf{\Psi}}\rangle\|_{\mathcal{L}(\ell_2(\mathcal{J}),\ell_2(\mathcal{J}))} \le c_R^{-2}\|B\|_{\mathcal{L}(V,V)}$$

are valid.

The property (5.6) provides one of the main foundations of the recent adaptive paradigm developed in [2]. In fact, it suggests that one can find simple preconditioners $\mathbf{C}_n$ such that the idealized iteration

$$\mathbf{D}\mathbf{u}^{k+1} = \mathbf{D}\mathbf{u}^k + \mathbf{C}_n\mathbf{D}^{-1}(\mathbf{f} - \mathbf{B}\mathbf{u}^k), \qquad k = 0, 1, \ldots, \tag{5.7}$$

converges to the solution $u = \mathbf{u}^T\mathbf{\Psi}$ of (5.5) for any initial guess $\mathbf{u}^0$ with some fixed error reduction rate $\rho < 1$ per step. In fact, for symmetric positive definite B (implying that also $\mathbf{B}$ is symmetric positive definite) the choice $\mathbf{C}_n = \alpha\mathbf{I}$ would work for a suitable damping parameter depending on the constants in (5.6). In general, $\mathbf{C}_n := \alpha(\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1})^T$ would in principle do the job although the corresponding squaring of condition numbers might degrade the quantitative performance.

For a numerical realization of (5.7), one has to approximate for any finitely supported input vector $\mathbf{v}$ the infinite-dimensional vector $\mathbf{D}^{-1}\mathbf{B}\mathbf{v}$ by some finitely supported $\mathbf{w}$ that approximates the exact sequence $\mathbf{D}^{-1}\mathbf{B}\mathbf{v}$ in $\ell_2(\mathcal{J})$ within suitable dynamically updated tolerances. In fact, one actually applies compressed versions of $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$ to scaled vectors $\mathbf{D}\mathbf{v}$. These perturbations of the ideal iteration (5.7) can be shown to provide an adaptive solution scheme

$$\textbf{Solve}(\mathbf{D}, \mathbf{B}, \eta, \mathbf{f}) \to \mathbf{u}_\eta$$

that outputs for any target accuracy $\eta > 0$ an approximate solution $\mathbf{u}_\eta$ of $\mathbf{B}\mathbf{u} = \mathbf{f}$ such that

$$\|\mathbf{D}(\mathbf{u} - \mathbf{u}_\eta)\|_{\ell_2(\mathcal{J})} \le \eta, \tag{5.8}$$

ETNA
Kent State University
etna@mcs.kent.edu

which, in view of (5.3), means that $\|u - u_\eta\|_H \le C_H \eta$.

Moreover, for operators B considered here one can contrive efficient schemes for the approximate application of $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$ so that the algorithm **Solve** can be shown to exhibit an overall asymptotically optimal work/accuracy balance in the following sense. To this end, let us denote for $r > 0$ by $\mathcal{A}^r = \mathcal{A}^r(H, \mathbf{\Psi})$ the class of those functions $v \in H$ such that the error $\sigma_{N,H}(v)$ of best N-term approximation of v in H, i.e., the accuracy obtainable by a linear combination of at most N wavelets ψ_λ, decays like N^{-r}. It is not hard to see that $|v|_{\mathcal{A}^r} := \sup_{N\in\mathbb{N}} N^r \sigma_{N,H}(v)$ defines a quasi-norm for $\mathcal{A}^r$. Moreover, when $H = H^t(\Omega)$ for some $t > 0$, the space $\mathcal{A}^r$ is almost characterized by the Besov space $B^{t+rd}_\tau(L_\tau(\Omega))$, where $1/\tau = r + 1/2$, see [2] for more details. Thus, an element $v \in \mathcal{A}^r$ can be recovered within accuracy η using only the order of $\eta^{-1/r}|v|^{1/r}_{\mathcal{A}^r}$ terms. It is important to note that, on account of (5.1), an element $v = \mathbf{v}^T\mathbf{\Psi} \in H$ belongs to $\mathcal{A}^r(H, \mathbf{\Psi})$ if and only if its scaled wavelet coefficient array $\mathbf{D}\mathbf{v}$ can also be approximated in $\ell_2(\mathcal{J})$ within accuracy N^{-r} by the order of N terms. Equivalently, approximating $\mathbf{D}\mathbf{v}$ in $\ell_2(\mathcal{J})$ within a tolerance $\eta > 0$, takes the order of $\eta^{-1/r}|v|^{1/r}_{\mathcal{A}^r}$ terms. Thus, in brief, (near-)best N-term approximation in H is directly linked to (near-)best N-term approximation in $\ell_2(\mathcal{J})$. Clearly, best N-term approximation in $\ell_2(\mathcal{J})$ is conceptually very simple since it boils down to retaining the N largest (in absolute value) terms of $\mathbf{D}\mathbf{v}$.

Now the main complexity estimate from [2] states that whenever the solution u of (5.4) belongs to $\mathcal{A}^r$ for some $r < r^*$, then the computational work required by **Solve** $(\mathbf{D}, \mathbf{B}, \eta, \mathbf{f})$ to output an η-accurate approximation $\mathbf{w}_\eta$ of $\mathbf{u}$ and hence, in particular, the number of nonzero terms in $\mathbf{w}_\eta$, is bounded uniformly by a constant multiple of $\eta^{-1/r}|v|^{1/r}_{\mathcal{A}^r}$. Moreover, when using spline wavelets of exactness order $m \in \mathbb{N}$ it has been shown in [17] that the optimality range covers the highest possible order, i.e. $r^* > m - t/d$. For this optimal complexity estimate it is crucial to have a relation like (5.6) which requires controlling the constants in (5.3) and those in a mapping property like (2.17).

We shall apply these concepts next to equations (5.4) for the choices (4.3) and (4.4).

5.2. Solution of the Resolvent Equations. To apply the wavelet technique to the resolvent equation

$$(\gamma I - A)v = f, \qquad \gamma \in \rho(A), \tag{5.9}$$

we will start from the following weak formulation: For any given $f \in H'$, find $u \in H$ such that

$$\langle A(\gamma)u, v\rangle = \langle \gamma^{-1} f, v\rangle \qquad \text{for all } v \in H, \tag{5.10}$$

where

$$A(\gamma) := I - \gamma^{-1} A.$$

Here and in the sequel we assume $\gamma \in \rho(A)$, i.e. for any $f \in H'$, equation (5.10) has a unique solution continuously depending on f. In particular, the norm equivalence

$$c_{A(\gamma)}\|v\|_H \le \|A(\gamma)v\|_{H'} \le C_{A(\gamma)}\|v\|_H, \qquad u \in H. \tag{5.11}$$

holds for some constants $c_{A(\gamma)}, C_{A(\gamma)} > 0$ that may depend on γ. The corresponding wavelet representations is denoted by $\mathbf{A}(\gamma) = \langle \mathbf{\Psi}, A(\gamma)\mathbf{\Psi}\rangle$. In principle, the comments from the previous section imply that, because of (5.11), the scheme **Solve** with the scaling $\mathbf{D}$ would indeed work for each fixed γ in the resolvent set with optimal complexity. However, we expect the ratios $C_{A(\gamma)}/c_{A(\gamma)}$ to become large when γ gets close to the spectrum. Moreover, when $|\gamma|$ grows $A(\gamma)$ acts more and more like the identity on lower scales so that the mapping properties of $A(\gamma)$ agree less and less with the topology of H. Thus the complexity bounds deteriorate due to the increasing ratio $C_{A(\gamma)}/c_{A(\gamma)}$.

This problem can be ameliorated by introducing appropriate γ-dependent norms on H. The lower constant $c_{A(\gamma)}$ will always depend on the *distance* of γ from the spectrum. To control this will be a matter of the choice of the curve Γ_1. So all one can hope for in addition is to reduce (or even avoid) the dependence of such stability constants on the *size* $|\gamma|$. To this end, we introduce next different norms on H and set

$$\mathbf{D}_\gamma = \operatorname{diag}\,(d_\lambda(\gamma))_{\lambda\in\mathcal{J}}, \quad d_\lambda(\gamma) := \max\{1, |\gamma|^{-1/2} d_\lambda\}, \tag{5.12}$$

(with d_λ from (5.2)) and endow H with the norm

$$\|v\|_{H(\gamma)} := \|\mathbf{D}_\gamma \mathbf{v}\|_{\ell_2(\mathcal{J})}, \quad v = \mathbf{v}^T\mathbf{\Psi}.$$

Clearly, by duality

$$\|w\|_{H(\gamma)'} = \|\mathbf{D}_\gamma^{-1}\langle \mathbf{\Psi}, w\rangle\|_{\ell_2(\mathcal{J})}.$$

The analog to (5.3) holds for $H(\gamma)$ trivially with constants $c_{H(\gamma)} = C_{H(\gamma)} = 1$. If we had in addition a mapping property

$$\hat{c}\|v\|_{H(\gamma)} \le \|A(\gamma)v\|_{H(\gamma)'} \le \hat{C}\|v\|_{H(\gamma)}, \quad v \in H(\gamma),$$

with constants independent of $|\gamma|$, we could conclude according to Remark 5 that

$$\operatorname{cond}_2(\mathbf{D}_\gamma^{-1}\mathbf{A}(\gamma)\mathbf{D}_\gamma^{-1}) \sim 1, \tag{5.13}$$

which is needed for **Solve** to perform with asymptotically optimal complexity. To this end, it is easy to see that for $v = \mathbf{v}^T\mathbf{\Psi}$

$$\|A(\gamma)v\|_{H(\gamma)'} = \|\mathbf{D}_\gamma^{-1}\langle \mathbf{\Psi}, A(\gamma)v\rangle\|_{\ell_2(\mathcal{J})} \le C\|\mathbf{D}_\gamma\mathbf{v}\|_{\ell_2(\mathcal{J})} = C\|v\|_{H(\gamma)},$$

where the constant C depends only on the constants in (5.1) and (5.3) and is thus independent of γ. Hence $A(\gamma)$ is uniformly bounded in $H(\gamma)$ for γ in the resolvent set. This means that the upper estimate in (5.6) holds for $\mathbf{D}_\gamma^{-1}\mathbf{A}(\gamma)\mathbf{D}_\gamma^{-1}$ with a constant independent of γ. The validity of a lower bound independent of $|\gamma|$ seems less clear. However, computational evidence shows clearly that scaling by $\mathbf{D}_\gamma$ is by far superior to scaling by $\mathbf{D}$.

Thus we shall employ **Solve** with respect to the scaling $\mathbf{D}_\gamma$. To bound the error with respect to $\|\cdot\|_V$ independent of γ we shall make frequent use of the following simple facts.

REMARK 6. *Note that for all the above scalings one has $\mathbf{D}_{\lambda,\lambda} \ge 1$, $(\mathbf{D}_\gamma)_{\lambda,\lambda} \ge 1$, for all $\lambda \in \mathcal{J}$. This implies that for $w = \mathbf{w}^T\mathbf{\Psi}$*

$$\|\mathbf{D}^{-1}\mathbf{w}\|_{\ell_2(\mathcal{J})} \le \|\mathbf{w}\|_{\ell_2(\mathcal{J})}, \quad \mathbf{w} \in \ell_2(\mathcal{J}), \tag{5.14}$$

which are the discrete analogues of the embedding inequalities

$$\|w\|_V \le \|w\|_H, \quad \|w\|_V \le \|w\|_{H(\gamma)}, \quad w \in H.$$

We are now prepared to formulate a realization of SOLVE$(\gamma I - A, \eta, u_0)$.

ALGORITHM 3. $w \leftarrow$ SOLVE$(\gamma I - A,\, \eta,\, u_0)$

- $\mathbf{u}_0 \leftarrow \langle \mathbf{\Psi}, u_0\rangle$

ETNA
Kent State University
etna@mcs.kent.edu

- $\mathbf{w} \leftarrow$ **Solve(**$\mathbf{D}_\gamma, \mathbf{A}(\gamma), C_R^{-1}\eta, \mathbf{u}_0$**)**
- $w \leftarrow \mathbf{w}^T\mathbf{\Psi}$

The requirement $\|(\gamma I - A)^{-1}u_0 - w\|_V \leq \eta$ is met thanks to the accuracy of **Solve**, c.f. (5.8), the discrete embedding inequality (5.14) and the right inequality in (5.1).

Concerning the specification of the scheme **Solve** in the above Algorithm 3, we can refer directly to [2] since the underlying equivalent formulation of the resolvent equation in wavelet coordinates is straightforward as given above. Moreover, the fast approximate application of $\mathbf{A}(\gamma)$ as well as the approximation of the data $\mathbf{u}_0$ follow here standard lines. One such ingredient is the approximation of given data in terms of wavelet coefficient series by sequences with shorter support. These approximations can be obtained (after appropriate preprocessing of the data) with the aid of a coarsening routine **Coarse** $(\eta, \mathbf{v}) \rightarrow \mathbf{v}_\eta$ that approximates $\mathbf{y}$ by a finitely supported vector such that $\|\mathbf{v} - \mathbf{v}_\eta\|_{\ell_2(\mathcal{J})} \leq \eta$ and $\mathbf{v}_\eta$ has (nearly) minimal support, see [2] for details concerning this routine. The routine EVAL appearing in Algorithm 2 looks as follows:

ALGORITHM 4. $w \leftarrow$ EVAL(η, v)

- $\mathbf{v} \quad \leftarrow \langle \widetilde{\mathbf{\Psi}}, v \rangle$,
- $\mathbf{w} \quad \leftarrow$ **Coarse**$(\eta, \mathbf{v})$,
- $w \quad \leftarrow \mathbf{w}^T\mathbf{\Psi}$.

The realization of SOLVE also for (4.3), i.e. in the context of the Tikhonov-type schemes is more involved since the application of the concepts from [2] require

(a) the wavelet representation of the underlying operators (in this case $B = \alpha I + S^*S$ or $B = \alpha Q_\alpha + S^*S$),
(b) their efficient application in wavelet coordinates as well as
(c) the less straightforward approximate evaluation of the right hand side in wavelet coordinates.

Therefore we turn now to providing these necessary ingredients in the subsequent subsections.

5.3. Application of (Projected) Semigroups in Wavelet Coordinates. The adaptive wavelet schemes hinge on the wavelet representation of the involved operators and data. Therefore we we shall derive next the wavelet representation of the projected semigroup and of the semigroup itself. Since the resolvents (5.9) apply to elements in H' which are naturally expanded in terms of the dual basis $\widetilde{\mathbf{\Psi}}$ we choose to represent $e^{-tA}P_{\Gamma_1}$ with respect to $\widetilde{\mathbf{\Psi}}$, i.e. we set

$$\mathbf{PE}(t) = \langle \widetilde{\mathbf{\Psi}}, e^{-tA}P_{\Gamma_1}\widetilde{\mathbf{\Psi}} \rangle, \quad \text{thus} \quad e^{-tA}P_{\Gamma_1}v = (\mathbf{PE}(t)\mathbf{v})^T\mathbf{\Psi}. \tag{5.15}$$

To apply this operator to a given vector $\mathbf{v}$ we invoke Algorithm 1. As we have already provided a realization of SOLVE$(\gamma I - A, \eta, u_0)$, Algorithm 1 and consequently the following algorithm is a computable scheme.

ALGORITHM 5. $\mathbf{w} \leftarrow$ **Apply(**$\mathbf{PE}(t), \eta, \mathbf{v}$**)**

- $w \leftarrow$ APPLY$(e^{-tA}P_{\Gamma_1}, c_R\eta, \mathbf{v}^T\widetilde{\mathbf{\Psi}})$
- $\mathbf{w} \leftarrow \langle \widetilde{\mathbf{\Psi}}, w \rangle$

The formulation of **Apply** is in principle only a formality which could be dispensed with. However, on one hand, it will be convenient in what follows and, on the other hand, reflects how the APPLY–routine actually works. In fact, since the resolvent equations are solved in the

ETNA
Kent State University
etna@mcs.kent.edu

wavelet coordinate domain, the intermediate quantities $w_{l,k}, u_l, d_l, w_l$ generated in Algorithm 1 are actually given in terms of their wavelet coefficient arrays.

Similarly to (5.15), the representation of $S = e^{-tA}$ in wavelet coordinates is given by

$$\mathbf{E}(t) := \langle \widetilde{\mathbf{\Psi}}, e^{-tA} \widetilde{\mathbf{\Psi}} \rangle.$$

We note that approximate evaluation schemes

$$\textbf{Apply}\,(\mathbf{E}(t), \bar{\mathbf{v}}, \eta), \tag{5.16}$$

can be constructed analogously based on the corresponding version of APPLY. For a detailed discussion and analysis of these evaluation schemes we refer to [15, 16, 11]. Alternatively, in view of Remark 2, we could realize the scheme **Apply** $(\mathbf{E}(t), \bar{\mathbf{v}}, \eta)$ with the aid of Algorithm 5 combined with a proper choice of the curve Γ_1. Therefore, we shall assume that the scheme in (5.16) is given without further elaborating on its specific features.

5.4. Tikhonov-Type Schemes. We shall now provide computable schemes for SOLVE $(\alpha I + S^*S, \eta, S^*y)$ and SOLVE$(\alpha Q_\alpha + S^*S, \eta, S^*y)$, see (4.3). To this end, we will apply the general solution concept of §5.1 to the operator $B = \alpha I + S^*S$ and $B = \alpha Q_\alpha + S^*S$, respectively. Both are norm-automorphisms of the space V such that it is sufficient to choose $\mathbf{D} = \mathbf{I}$, i.e. preconditioning is not necessary.

Recalling that A is assumed to be normal, the (modified) Tikhonov regularization requires, in view of (3.3), solving the operator equation

$$(\alpha Q_\alpha + e^{-tB})x = e^{-tA^*}y, \quad B := A^* + A, \tag{5.17}$$

Concerning an equivalent formulation in wavelet coordinates, recall that the resolvent application suggested to represent the evolution operators in the dual basis $\widetilde{\mathbf{\Psi}}$, i.e.

$$\mathbf{Q}_\alpha = \langle \widetilde{\mathbf{\Psi}}, Q_\alpha \widetilde{\mathbf{\Psi}} \rangle, \quad \mathbf{E}^B(t) = \langle \widetilde{\mathbf{\Psi}}, e^{-tB} \widetilde{\mathbf{\Psi}} \rangle,$$

where Q_α either denotes the identity or a projection, depending on the size of α. Here we use the superscript B to distinguish $\mathbf{E}^B(t)$ from the representation $\mathbf{E}(t)$ of the original semigroup e^{-tA}. If we were content with representing the solution x in dual coordinates $x = \widetilde{x}^T \widetilde{\mathbf{\Psi}}$, (5.17) is equivalent to

$$\big(\alpha \mathbf{Q}_\alpha + \mathbf{E}^B(t)\big)\, \widetilde{\mathbf{x}} = \langle \mathbf{\Psi}, e^{-tA^*}y \rangle = \mathbf{E}(t)^* \langle \mathbf{\Psi}, y \rangle. \tag{5.18}$$

But representing the solution in dual coordinates has several drawbacks. In practical constructions the primal basis $\mathbf{\Psi}$ is typically piecewise polynomial and comfortable routines are available for further manipulating or processing the approximate solutions. The dual basis $\widetilde{\mathbf{\Psi}}$ is typically less regular and less convenient to handle. Therefore, we indicate next briefly how to switch representations with the aid of numerically executable Riesz maps. Clearly, expanding the ψ_λ in terms of the $\widetilde{\psi}_\lambda$, yields

$$\mathbf{\Psi} = \langle \mathbf{\Psi}, \mathbf{\Psi} \rangle \widetilde{\mathbf{\Psi}} \quad \text{i.e.} \quad \mathbf{I} = \langle \mathbf{\Psi}, \mathbf{\Psi} \rangle \langle \widetilde{\mathbf{\Psi}}, \widetilde{\mathbf{\Psi}} \rangle.$$

Thus, defining the Gramian $\mathbf{G} = \langle \mathbf{\Psi}, \mathbf{\Psi} \rangle$, we have

$$\bar{\mathbf{v}}^T \mathbf{\Psi} = (\mathbf{G}\bar{\mathbf{v}})^T \widetilde{\mathbf{\Psi}},$$

i.e. application of the Gramian to primal coordinates yields the dual coordinates. Likewise, since $\mathbf{I} = \langle \mathbf{\Psi}, \widetilde{\mathbf{\Psi}} \rangle$, we obtain for any endomorphism B of V

$$\langle \mathbf{\Psi}, B\mathbf{\Psi} \rangle = \langle \mathbf{\Psi}, \mathbf{\Psi} \rangle \langle \widetilde{\mathbf{\Psi}}, B\widetilde{\mathbf{\Psi}} \rangle \langle \mathbf{\Psi}, \mathbf{\Psi} \rangle = \mathbf{G} \langle \widetilde{\mathbf{\Psi}}, B\widetilde{\mathbf{\Psi}} \rangle \mathbf{G}. \tag{5.19}$$

ETNA
Kent State University
etna@mcs.kent.edu

If, as argued above, we wish to represent $x = \mathbf{x}^T\boldsymbol{\Psi} \in V$ in primal coordinates we obtain, on account of (5.19), the alternate equivalent formulation

$$\mathbf{G}\left(\alpha\mathbf{Q}_\alpha + \mathbf{E}^B(t)\right)\mathbf{G}\mathbf{x} = \mathbf{G}\mathbf{E}^*(t)\langle\boldsymbol{\Psi}, y\rangle. \tag{5.20}$$

(5.18) would save two applications of $\mathbf{G}$ per iteration step in comparison with (5.20). Of course, one could transform at the end $\widetilde{\mathbf{x}}$ into primal coordinates by applying the inverse $\mathbf{G}^{-1} = \langle\widetilde{\boldsymbol{\Psi}}, \widetilde{\boldsymbol{\Psi}}\rangle$. But the latter one is usually less compressible rendering its application less efficient. Therefore we concentrate in what follows on (5.20).

We wish to solve the system (5.20) within any desired target accuracy with the aid of the scheme **Solve** following the lines of [2]. To invoke the concepts developed there we need algorithms to apply the involved operators up to any prescribed accuracy. We provide these ingredients in the following subsections. The above representations involve, in particular, the (infinite) matrix $\mathbf{G}$. The vanishing moment property of wavelets allows one to apply $\mathbf{G}$ to any finitely supported vector within any prescribed accuracy tolerance by the schemes developed in [2, 1]. Moreover, within a certain basis dependent range such an application has asymptotically optimal complexity in the sense described earlier. Such a multiplication scheme is also a central building block in the scheme **Solve** and will be denoted by $\mathbf{Mult}\,(\mathbf{G}, \eta, \mathbf{v}) \to \mathbf{w}_\eta$ providing an approximation $\mathbf{w}_\eta$ to $\mathbf{G}\mathbf{v}$ such that $\|\mathbf{G}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{J})} \le \eta$.

It should be noted that the following building blocks are all acting in the wavelet domain. This is necessary since the scheme **Solve($\mathbf{D}, \mathbf{B}, \eta, \mathbf{v}$)** itself operates in the wavelet domain. Once all ingredients are gathered we shall formulate an algorithm acting in the continuous domain, see Algorithm 9 at the end of §5.4.3.

5.4.1. The Right Hand Side z.

Then the right hand side of (5.20) has the representation

$$\mathbf{z} = \langle\boldsymbol{\Psi}, e^{-tA^*}y\rangle = \langle\boldsymbol{\Psi}, \boldsymbol{\Psi}\rangle\,\langle\widetilde{\boldsymbol{\Psi}}, e^{-tA^*}\widetilde{\boldsymbol{\Psi}}\rangle\,\mathbf{y} = \mathbf{G}\mathbf{E}^*(t)\mathbf{y}.$$

Again since $\mathbf{y}$ has possibly infinitely many non-vanishing entries the following scheme starts formally with a coarsening step.

ALGORITHM 6. $\mathbf{w} \leftarrow$ **EvalRhs(η, y)**

- $\mathbf{w}_1 \leftarrow$ **Coarse**$(\eta_1, \mathbf{y})$, $\qquad \eta_1 = \frac{c_R^2}{3C_R^2\|e^{-tA^*}\|_{\mathcal{L}(V)}}\eta$;
- $\mathbf{w}_2 \leftarrow$ **Apply**$(\mathbf{E}^*(t), \eta_2, \mathbf{w}_1)$, $\qquad \eta_2 = \frac{\eta}{3C_R^2}$;
- $\mathbf{w} \leftarrow$ **Mult**$(\mathbf{G}, \eta_3, \mathbf{w}_2)$, $\qquad \eta_3 = \frac{\eta}{3}$.

LEMMA 5.1. *The output* $\mathbf{w}$ *of* **EvalRhs**$(\eta, \mathbf{y})$ *satisfies*

$$\|\langle\boldsymbol{\Psi}, e^{-tA^*}\boldsymbol{\Psi}\rangle\mathbf{y} - \mathbf{w}\|_{\ell_2(\mathcal{J})} \le \eta.$$

Proof. Repeated application of the triangle inequality yields

$$\begin{aligned}&\|\langle\boldsymbol{\Psi}, e^{-tA^*}y\rangle - \mathbf{w}\|_{\ell_2(\mathcal{J})}\\ &\le \|\mathbf{G}\|_{\mathcal{L}(\ell_2(\mathcal{J}))}\left(\|\mathbf{E}^*(t)\|_{\mathcal{L}(\ell_2(\mathcal{J}))}\cdot\eta_1 + \eta_2\right) + \eta_3.\end{aligned}$$

As, by Remark 5, $\|\mathbf{G}\|_{\mathcal{L}(\ell_2(\mathcal{J}))} \le C_R^2$ and $\|\mathbf{E}^*(t)\|_{\mathcal{L}(\ell_2(\mathcal{J}))} \le c_R^{-2}\|e^{-tA^*}\|_{\mathcal{L}(V)}$, the choice of η_1, η_2, and η_3 ensures that the error is bound by η. ∎

ETNA
Kent State University
etna@mcs.kent.edu

5.4.2. The Projection. Given a regularization parameter $\alpha > 0$ the orthogonal projection Q_α onto the eigenspace associated with the eigenvalues of $e^{-tA^*}e^{-tA} = e^{tB}$, in $[0,\alpha]$ is to be evaluated. Denote by $\{\lambda_k\}_{k\geq 1}$ the eigenvalues of $B = A + A^*$ sorted by increasing size, $\lambda_k \leq \lambda_{k+1}$. Due to the spectral mapping theorem 3.1 we have

$$\begin{aligned}\sigma(e^{-tB}) \cap [0,\alpha] &= \big(\{0\} \cup \{e^{-t\lambda_k} : k \geq 1\}\big) \cap [0,\alpha] \\ &= \{0\} \cup \{e^{-t\lambda_k} : \lambda_k \geq \frac{1}{t}\log(\frac{1}{\alpha})\}.\end{aligned}$$

By the results of §3 we are capable to evaluate projections onto the eigenspaces associated to any bounded subset of $\sigma(B)$. Thus, the representation $Q_\alpha = I - P_\alpha$ is used where $P_\alpha = P_{\Gamma_1}(B)$ and the Jordan curve Γ_1 encloses $\{\lambda_k : \lambda_k < \frac{1}{t}\log(\frac{1}{\alpha})\}$ (cf. §3). The representation of Q_α with respect to $\mathbf{\Psi}$ becomes

$$\langle \mathbf{\Psi}, Q_\alpha \mathbf{\Psi}\rangle = \mathbf{G} - \mathbf{G}\mathbf{P}_\alpha\mathbf{G},$$

where $\mathbf{P}_\alpha = \mathbf{PE}(0)$ is the projection in wavelet coordinates and can be evaluated by **Apply** $(\mathbf{PE}(0), \eta, \mathbf{v})$, see Algorithm 5. We do not define a **Apply**–routine for this operator, but use this representation in the upcoming Algorithm 8.

5.4.3. The Operator $e^{-tA^*}e^{-tA} = e^{-tB}$. The observations in §5.4, suggest the following algorithm to evaluate $\langle \mathbf{\Psi}, (\alpha I + e^{-tB})\mathbf{\Psi}\rangle \mathbf{v} = \mathbf{G}(\alpha\mathbf{I} + \mathbf{E}^B(t))\mathbf{G}\mathbf{v}$, where $\mathbf{v}$ are here the expansion coefficients with respect to the basis $\mathbf{\Psi}$ of V.

ALGORITHM 7. $\mathbf{w} \leftarrow$ **Apply(**$\langle \mathbf{\Psi}, (\alpha I + e^{-tB})\mathbf{\Psi}\rangle, \eta, \mathbf{v}$**)**

- $\mathbf{w}_1 \leftarrow \mathbf{Mult}(\mathbf{G}, \eta_1, \mathbf{v})$, $\quad \eta_1 = \frac{c_R^2}{3C_R^2\|e^{-tB}\|_{\mathcal{L}(V)}}\eta$;
- $\mathbf{w}_2 \leftarrow \mathbf{Apply}(\mathbf{E}^B(t), \eta_2, \mathbf{w}_1)$, $\quad \eta_2 = \frac{1}{3C_R^2}\eta$;
- $\mathbf{w} \leftarrow \mathbf{Mult}(\mathbf{G}, \eta_3, \alpha\mathbf{v} + \mathbf{w}_2)$, $\quad \eta_3 = \frac{1}{3}\eta$.

Clearly, in the third step we invoke the scheme **Apply** $(\mathbf{E}^B(t), \bar{\mathbf{v}}, \eta)$ from (5.16) for the operator $\mathbf{E}^B(t) = \langle \widetilde{\mathbf{\Psi}}, e^{-t(A+A^*)}\widetilde{\mathbf{\Psi}}\rangle$. An analogous reasoning as in the proof of Lemma 5.1 yields the following estimate.

LEMMA 5.2. *The output* $\mathbf{w}$ *of* **Apply**$(\langle \mathbf{\Psi}, (\alpha I + e^{-tB})\mathbf{\Psi}\rangle, \eta, \mathbf{v})$ *satisfies*

$$\|\langle \mathbf{\Psi}, (\alpha I + e^{-tB})\mathbf{\Psi}\rangle - \mathbf{w}\|_{\ell_2(\mathcal{J})} \leq \eta.$$

If the identity is replaced by a projection $Q_\alpha = I - P_\alpha$ we use an algorithm based on the decomposition

$$\langle \mathbf{\Psi}, (\alpha Q_\alpha + e^{-tB})\mathbf{\Psi}\rangle = \left[\alpha I + \mathbf{G}\left(\mathbf{E}^B(t) - \alpha\mathbf{P}_\alpha\right)\right]\mathbf{G}.$$

$\mathbf{P}_\alpha$ is chosen depending on α as discussed in the previous section.

ALGORITHM 8. $\mathbf{w} \leftarrow$ **Apply(**$\langle \mathbf{\Psi}, (\alpha Q_\alpha + e^{-tB})\mathbf{\Psi}\rangle, \eta, \mathbf{v}$**)**

- $\mathbf{w}_1 \leftarrow \mathbf{Mult}(\mathbf{G}, \eta_1, \mathbf{v})$, $\quad \eta_1 = \frac{c_R^2}{4C_R^2\left(\alpha + \|e^{-tB}\|_{\mathcal{L}(V)}\right)}\eta$;
- $\mathbf{w}_2 \leftarrow \mathbf{Apply}(\mathbf{E}^B(t), \eta_2, \mathbf{w}_1)$, $\quad \eta_2 = \frac{1}{4C_R^2}\eta$;
- $\mathbf{w}_3 \leftarrow \mathbf{Apply}(\langle \widetilde{\mathbf{\Psi}}, P_\alpha\widetilde{\mathbf{\Psi}}\rangle, \eta_3, \mathbf{w}_1)$, $\quad \eta_3 = \frac{1}{4\alpha C_R^2}\eta$;
- $\mathbf{w} \leftarrow \mathbf{Mult}(\mathbf{G}, \eta_4, \alpha(\mathbf{v} - \mathbf{w}_3) + \mathbf{w}_2)$, $\quad \eta_4 = \frac{1}{4}\eta$.

A similar proof as in Lemma 5.1 yields the following estimate.

LEMMA 5.3. *The output* $\mathbf{w}$ *of* **Apply**$(\langle \mathbf{\Psi}, (\alpha Q_\alpha + e^{-tB})\mathbf{\Psi}\rangle, \eta, \mathbf{v})$ *satisfies*

$$\|\langle \mathbf{\Psi}, (\alpha Q_\alpha + e^{-tB})\mathbf{\Psi}\rangle - \mathbf{w}\|_{\ell_2(\mathcal{J})} \leq \eta.$$

We have prepared now all the algorithmic ingredients needed to apply the scheme **Solve** from [2] to the operator equation (5.17) which will return an approximate solution x_η that approximates x with accuracy η. Thus we have the following implementation of the abstract routine $\textsc{Solve}(\alpha Q_\alpha + S^* S, \eta, S^* y)$.

ALGORITHM 9. $w \leftarrow \textsc{Solve}(\alpha Q_\alpha + S^* S, \eta, S^* y)$

- $\mathbf{y} \ \leftarrow \langle \mathbf{\Psi}, y\rangle$
- $\mathbf{z} \ \leftarrow$ **EvalRhs**$(\frac{1}{2}\alpha c_R^2 C_R^{-1}\eta, \mathbf{y})$
- $\mathbf{w} \leftarrow$ **Solve**$(\mathbf{I}, \mathbf{G}(\alpha\mathbf{Q} + \mathbf{E}^B(t))\mathbf{G}, \frac{1}{2}C_R^{-1}\eta, \mathbf{z})$
- $w \ \leftarrow \mathbf{w}^T\mathbf{\Psi}$

5.5. Regularization of the Transformed Problem. So far we have first regularized the continuous problem and then transformed it into a well-posed problem in $\ell_2(\mathcal{J})$ (whose condition, of course, depends on the regularization parameter). In this subsection, we will exchange the order of regularization and transformation. We shall exemplify this briefly only for the Tikhonov method, the modified version being similar. This gives rise to a scheme where norms are easily changed by scaling offering an additional way of influencing the regularization compromise. Accordingly X and Y are now different smoothness spaces and no longer coincide with V.

We consider the problem

$$\mathbf{E}_s(t)\mathbf{x} = \mathbf{y},$$

posed in $\ell_2(\mathcal{J})$ with the evolution operator given in wavelet coordinates as

$$\mathbf{E}_s(t) = \mathbf{D}^s\langle\widetilde{\mathbf{\Psi}}, e^{-tA}\widetilde{\mathbf{\Psi}}\rangle\mathbf{D}^s, \qquad s \in [-1, 1],$$

where now $x = \mathbf{x}^T\mathbf{D}^{-s}\widetilde{\mathbf{\Psi}}$. The power s of the diagonal elements of $\mathbf{D}$ steers the involved norms. For instance, when A is a second order elliptic operator, $\mathbf{E}_0(t) = \mathbf{E}(t)$ is a representation of the operator e^{-tA} regarded as a mapping $X = V = \mathsf{L}_2(\Omega) \to \mathsf{L}_2(\Omega) = Y$. In general, for $H = H_0^t(\Omega)$, it is well-known that the scaling by $\mathbf{D}^s$ induces equivalent norms on the interpolation spaces between $H^{-t}(\Omega)$ and $H_0^t(\Omega)$, see e.g. [5]. Let us assume here the standard case $t = 1$.

As e^{-tA} is compact the transformed operator has the same property as endomorphism of $\ell_2(\mathcal{J})$ and the resulting sequence equation is ill-posed, as well. Applying Tikhonov-regularization in $\ell_2(\mathcal{J})$ yields the equation

$$(\alpha I + \mathbf{E}_s^*(t)\mathbf{E}_s(t))\mathbf{x}_\alpha^\delta = \mathbf{E}_s^*(t)\mathbf{y}^\delta,$$

where $\mathbf{I}$ denotes again the identity in $\ell_2(\mathcal{J})$ and not the Gramian matrix. $\mathbf{y}^\delta$ and $\mathbf{x}_\alpha^\delta$ are given by

$$\mathbf{y}^\delta = \mathbf{D}^s\langle\mathbf{\Psi}, y^\delta\rangle, \qquad \mathbf{x}_\alpha^\delta = \mathbf{D}^{-s}\langle\mathbf{\Psi}, x_\alpha^\delta\rangle.$$

The regularized equation is equivalent to the solution of the minimization problem

$$\begin{aligned} J_s(\alpha, \mathbf{y}^\delta) &= \alpha\|\mathbf{x}_\alpha^\delta\|_{\ell_2(\mathcal{J})}^2 + \|\mathbf{E}_s(t)\mathbf{x}_\alpha^\delta - \mathbf{y}^\delta\|_{\ell_2(\mathcal{J})}^2 \\ &= \alpha\|\mathbf{D}^{-s}\langle\mathbf{\Psi}, x_\alpha^\delta\rangle\|_{\ell_2(\mathcal{J})}^2 + \|\mathbf{D}^s\langle\mathbf{\Psi}, e^{-tA}x_\alpha^\delta - y^\delta\rangle\|_{\ell_2(\mathcal{J})}^2 \\ &\to \min_{\mathbf{x}_\alpha^\delta \in \ell_2(\mathcal{J})}. \end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

By the norm equivalences of the wavelet basis one sees that

$$\|\mathbf{x}_\alpha^\delta\|_{\ell_2(\mathcal{J})} \sim \|x_\alpha^\delta\|_{H^{-s}(\Omega)} \quad \text{and} \quad \|\mathbf{y}^\delta\|_{\ell_2(\mathcal{J})} \sim \|y^\delta\|_{H^s(\Omega)},$$

i.e. we take here $X = H^{-s}(\Omega)$, $Y = H^s(\Omega)$. Thus, the smaller s is the stronger the norm becomes in which smoothness of x_α^δ is measured. In turn, the norm in which the residual is measured becomes weaker for smaller s. In practice, the range $s \in [-1, 0]$ is of interest as the solution x_α^δ should be smooth.

Concerning the convergence of the regularization scheme for $\delta \to 0^+$, the properties of Tikhonov regularization read as follows. Suppose that

$$y = e^{-tA}x^\dagger \quad \text{with} \quad x^\dagger \in H^{-s}(\Omega)$$

and

$$\|y - y^\delta\|_{H^s(\Omega)} \leq \delta.$$

Then,

$$\|x_\alpha^\delta - x^\dagger\|_{H^{-s}(\Omega)} \to 0 \quad \text{for} \quad \delta \to 0^+,$$

if $\alpha = \alpha(\delta, y^\delta)$ is chosen properly, e.g. according to the discrepancy principle.

5.6. Some Comments on Computational Complexity. Not much seems to be known about the actual computational cost of determining the numerical approximation $\widehat{x}_\alpha^\delta$ for the regularized solution x_α^δ depending on the given noise level δ as δ tends to zero. We address this question now in a somewhat sketchy and partly heuristic way for the simplest scenario that A is a symmetric self adjoint second order elliptic operator so that $H = H_0^1(\Omega)$, $H' = H^{-1}(\Omega)$. Hence the singular values σ_k of $S = e^{-tA}$ are for fixed $t > 0$ given by $\sigma_k = e^{-t\lambda_k}$, where λ_k are the eigenvalues of A. A detailed rigorous study would be beyond the scope of the present paper and will be given elsewhere. Here we merely wish to bring out what the approach might offer at best and to identify some of the questions along that way.

Let Γ_1^K denote a suitable Jordan curve whose interior contains the first K eigenvalues $\lambda_k, k = 1, \ldots, K$, and let $P_K := P_{\Gamma_1^K}$ be again the projector defined in (3.1). As pointed out before, $x_K^\delta := e^{tA_1}y^\delta$ with $e^{-tA_1} = P_K e^{-tA} = e^{-tA}P_K$, corresponds to a truncated SVD expansion. Since by assumption $\|P_K(y - y^\delta)\|_V \leq C\delta$ we can invoke [9, Theorem 3.26] to conclude that

$$\|x^\dagger - x_K^\delta\|_V \leq C'\Big(\sigma_{K+1} + \frac{\delta}{\sigma_K}\Big) \tag{5.21}$$

for some constant C' independent of δ, K. Having some estimates for the λ_k and hence for the σ_k, a reasonable parameter choice is to balance both terms on the right hand side of (5.21). This suggests choosing

$$K = K(\delta) = \max\{k : \sigma_{k+1}\sigma_k \geq \delta\}.$$

In order to retain the accuracy offered by this projection method we compute

$$\hat{x}_K^\delta = \mathbf{x}^T\mathbf{\Psi}, \quad \mathbf{x} = \mathbf{Apply}\,(\mathbf{PE}(-t), \mathbf{y}^\delta, \eta), \quad \text{with } \eta := \delta/\sigma_K.$$

Recall that the latter routine is given in Algorithm 1. As before, we do not specify in which form the noisy data $y^\delta \in V$ are actually given but assume for simplicity that we can observe its wavelet coefficients $\mathbf{y}^\delta$ for the expansion $y^\delta = (\mathbf{y}^\delta)^T\widetilde{\mathbf{\Psi}}$. Note that by (5.1), we have then

$$\|\mathbf{y} - \mathbf{y}^\delta\|_{\ell_2(\mathcal{J})} \leq c_R^{-1}\delta. \tag{5.22}$$

ETNA
Kent State University
etna@mcs.kent.edu

In this scheme the computational cost is then dominated by the approximate applications of resolvents at quadrature points $\gamma(s_{l,k})$, $k = 0, \dots, N_l - 1$, $N_l := 2^l N_0$, on the curve Γ_1^K. By a geometric series argument the overall work is dominated by the quadrature at the terminating level L of quadrature resolution. For that L the scheme **Solve** is called $N_L = 2^L N_0$ times with target accuracies $\sim \eta/|e^{t\gamma(s_{L,k})}|$, $k = 0, \dots, N_L - 1$. Thus, the most stringent accuracy constraint is encountered when the real part of $\gamma(s_{L,k})$ is of the order of λ_K. Hence, a rough bound for the computational cost $\mathcal{C}_K$ of $\mathbf{Apply}\,(\mathbf{PE}(-t), \mathbf{y}^\delta, \eta)$ with $\eta := \delta/\sigma_K$ is given by

$$\mathcal{C}_K \lesssim 2^L N_0 \mathcal{C}\big(\mathbf{Solve}\,(\mathbf{D}_{\gamma(s_{L,N_L-1})}, \mathbf{A}(\gamma(s_{L,N_L-1})), c\delta, \gamma(s_{L,N_L-1})^{-1}\mathbf{y}^\delta)\big), \tag{5.23}$$

where we have used that $\sigma_K = e^{-t\lambda_K}$ (recall the scaling $\mathbf{D}_\gamma$ given by (5.12)). Here, $\mathcal{C}(\mathbf{Solve})$ denotes the computational cost of the routine **Solve** and $a \lesssim b$ means that a can be bounded by a fixed multiple of b which is independent of the parameters on which a and b may depend.

It remains now to estimate L and $\mathcal{C}(\mathbf{Solve})$ for a target accuracy of the order δ. We shall address this latter issue first. Recall that, according to (5.7), **Solve** endowed with the scaling $\mathbf{D}_\gamma$ operates on right hand side data of the form $\gamma^{-1}\mathbf{D}_\gamma^{-1}\mathbf{y}^\delta$ which are the expansion coefficients with respect to the $H(\gamma)'$-scaled basis $\mathbf{D}_\gamma\widetilde{\mathbf{\Psi}}$. Thus, the noisy input data $\mathbf{y}^\delta$ are damped in two ways, first by γ^{-1} when $|\gamma|$ is large, and second, the entries y_λ^δ for $|\lambda| \geq \frac{1}{2}\log_2|\gamma|$ are attenuated by the weights $(\mathbf{D}_\gamma)_{\lambda,\lambda}^{-1}$, see (5.12). By (5.22) and (5.14) we also know that

$$\|\gamma^{-1}\mathbf{D}_\gamma^{-1}(\mathbf{y} - \mathbf{y}^\delta)\|_{\ell_2(\mathcal{J})} \leq c_R^{-1}\delta \tag{5.24}$$

uniformly in $|\gamma| \geq 1$.

To obtain an impression of $\mathcal{C}(\mathbf{Solve})$ in this context, suppose now that the exact wavelet coefficient array $\mathbf{y}$ of $u(t) = y = \mathbf{y}^T\widetilde{\mathbf{\Psi}}$ is sparse in the sense that

$$y \in \mathcal{A}^r(V, \widetilde{\mathbf{\Psi}})$$

for some $r > 0$, i.e. the error of best N-term approximation in $\ell_2(\mathcal{J})$ decays as $N^{-r}|y|_{\mathcal{A}^r(V,\widetilde{\mathbf{\Psi}})}$. Since the expansion coefficients of y with respect to the $H(\gamma)$-basis $\mathbf{D}_{\gamma(s)}\widetilde{\mathbf{\Psi}}$ are given by $\mathbf{D}_{\gamma(s)}^{-1}\mathbf{y}$, (5.14) implies that also

$$y \in \mathcal{A}^r(H(\gamma)', \widetilde{\mathbf{\Psi}})$$

for that $r > 0$ and any point $\gamma(s)$ on Γ_1^K, uniformly in s and K. We shall explain later what this means in terms of the regularity of y. Now suppose further that the operator A and its shifted versions $A(\gamma(s))$ permit the following "regularity" result:

ASSUMPTION 2. *Given $\bar{u} \in \mathcal{A}^r(H(\gamma)', \widetilde{\mathbf{\Psi}})$ for some $r < r^*$ (with $r*$ the fixed compressibility limit depending on the wavelet bases (see [2, 17])) then the solution $v \in H(\gamma(s))$ of $A(\gamma(s))v = \bar{u}$ belongs to $\mathcal{A}^r(H(\gamma(s)), \mathbf{\Psi})$, i.e. sparseness is preserved under the inversion of $A(\gamma(s))$.*

Then, if (5.13) was true, the results in [2] would assert that

$$\mathcal{C}\big(\mathbf{Solve}\,(\mathbf{B}(\gamma(s_{L,N_L-1})), c\delta, \bar{\mathbf{y}})\big) \lesssim \delta^{-1/r}|v|_{\mathcal{A}^r}^{1/r}. \tag{5.25}$$

However, **Solve** "sees" only the perturbed data $\bar{\mathbf{y}}^\delta := \mathbf{D}_{\gamma(s)}^{-1}\mathbf{y}^\delta$. We dispense here with introducing a statistical model and try to construct an estimator reflecting the sparseness of

ETNA
Kent State University
etna@mcs.kent.edu

the exact data with high probability or in expectation. Instead we resort to a more heuristic reasoning why under the above assumptions one may still obtain a work count like (5.25). This requires a somewhat closer look at the way **Solve** works, see [2]. First of all the data $\bar{\mathbf{y}}^\delta$ can be considered as a finite array, since every coefficient $\bar{\mathbf{y}}^\delta_\lambda$ with $(\mathbf{D}_{\gamma(s)})^{-1}_{\lambda,\lambda} \leq \delta$ can be discarded which, in view of (5.22), introduces only an error of the order at most δ in $\ell_2(\mathcal{J})$ for the scaled right hand side (which is what matters for **Solve**). Let us call this finite array again $\bar{\mathbf{y}}^\delta$. Thus, by (5.24), we still have that $\|\bar{\mathbf{y}} - \bar{\mathbf{y}}^\delta\|_{\ell_2(\mathcal{J})} \leq \|\mathbf{y} - \mathbf{y}^\delta\|_{\ell_2(\mathcal{J})} \leq C\delta$ for some finite constant C. Now, given some target accuracy ε for **Solve**, the right hand side data are approximated first by a possibly small array within a tolerance $a\varepsilon$ where a is typically smaller than one. This approximation is done by the routine **Coarse** described e.g. in [2], see also §5.4.1. If we relax our target accuracy $c\delta$ in **Solve** by some constant factor larger than one, the coarsening of the right hand side data $\bar{\mathbf{y}}^\delta$ would have to be done with a target accuracy $bC\delta$ where b is now any fixed number *larger* than one. But then the Coarsening Lemma from [2] states that the resulting approximation $\mathbf{y}^*$ to $\bar{\mathbf{y}}^\delta$ now satisfies

$$\begin{aligned} \|\bar{\mathbf{y}} - \bar{\mathbf{y}}^*\|_{\ell_2(\mathcal{J})} &\leq (1+b)C\delta, \quad \#\mathrm{supp}\, \bar{\mathbf{y}}^* \lesssim \delta^{-1/r}|\bar{y}|^{1/r}_{\mathcal{A}^r(H(\gamma),\mathbf{\Psi})}, \\ |\bar{y}^*|_{\mathcal{A}^r(H(\gamma))} &\lesssim |\bar{y}|_{\mathcal{A}^r(H(\gamma))}. \end{aligned}$$

Thus, using a fixed but sufficiently large multiple of δ as target accuracy in **Solve** the corresponding coarsening of the data that "peels" off at this level of accuracy the sparsity of the exact array $\bar{\mathbf{y}}$ in the neighborhood. In that sense the routine **Solve** still "sees" $\bar{\mathbf{y}}$ and exploits its compressibility giving the work count (5.25).

Let us briefly comment now on the above assumptions. We wish to recover the initial data $x^\dagger = u(0) = (\mathbf{x}^\dagger)^T\mathbf{\Psi}$ from the data y^δ. Suppose that $u(0) \in B^r_P(L_p(\Omega))$ where $\frac{1}{p} = \frac{r}{d} + \frac{1}{2}$. The space $B^r_P(L_p(\Omega))$ is just embedded in $L_2(\Omega)$ and signifies in some sense the largest space of smoothness r contained in $L_2(\Omega)$. Note that the larger r the smaller p and the weaker the measure for smoothness. It is well known that $x^\dagger \in B^r_P(L_p(\Omega))$ implies $\mathbf{x}^\dagger \in \mathcal{A}^r$, see e.g. [8]. Moreover, it is known that then $y = u(t) = e^{-tA}u(0)$ belongs also to $B^r_P(L_p(\Omega))$, see [16]. Assumption 2 then says that the assumed gain in regularity through the application of $A(\gamma(s))^{-1}$ suffices to ensure that the solution array with respect to $\mathbf{D}^{-1}_{\gamma(s)}\mathbf{\Psi}$ is still in $\mathcal{A}^r$. For moderate $|\gamma(s)| \sim 1$ this is simply a regularity theorem for A itself. For A as above we would have $v = A(\gamma(s))^{-1}y \in B^{r+2}_p(L_p(\Omega))$. Note that already $v \in B^{r+1}_p(L_p(\Omega))$ would suffice to ensure that for $v = \mathbf{v}^T\mathbf{\Psi}$ the array $\mathbf{D}_{\gamma(s)}\mathbf{v}$ belongs to $\mathcal{A}^r(H(\gamma),\mathbf{\Psi})$. On the other hand, for large $|\gamma(s)|$ the operator $A(\gamma(s))$ gets closer to the identity (at least on low scales) and the norm becomes closer to the $L_2(\Omega)$-norm so that Assumption 2 becomes less demanding. Nevertheless, the validity of (5.13) as well as of Assumption 2 is crucial for the above reasoning.

Finally, let us quickly address the estimation of N_L. A qualitative estimate on the L_∞-norm of the integrand F is at best of the order $e^{t\lambda_K}$. Since the required quadrature accuracy in **Apply** is of the order $\eta = \delta/\sigma_K$ we have to ensure, in view of (4.6), that $\delta/\sigma_K \sim 1/(\sigma_K e^{2qN})$. Although we do not know q we see that $N \sim |\log\delta|$. Thus, by (5.23), the overall work count can at best be expected to be of the order $|\log\delta|\delta^{-1/r}$ which would almost correspond to the sparseness of $x^\dagger = u(0)$.

6. Numerical Experiments. For the numerical experiments, we choose the operator $A = -\Delta$ with homogeneous boundary conditions on the unit interval $\Omega = (0,1)$. The eigenvalues and corresponding eigenvectors of this operator are given by

$$\lambda_k = \pi^2k^2, \quad v_k = \sqrt{2}\sin(k\pi x), \qquad k = 1,2,\ldots$$

We will exploit this orthonormal basis of $L_2(\Omega)$ to obtain a reference solution in the following experiment. Furthermore, we use the wavelet basis constructed in [6] of order 3 on the primal

ETNA
Kent State University
etna@mcs.kent.edu

and dual side for the spatial discretization.. In all computations, a highest level of $J = 15$ is fixed.

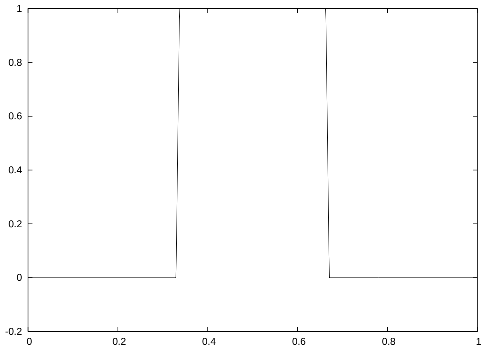

FIG. 6.1. *Initial value*

6.1. Application of Projections. First, the projection of a vector $u_0 \in L_2(\Omega)$ into the joint eigenspaces of A and e^{-tA} is tested. The main concern is to justify Assumption 1 which the adaptive projection algorithm is based upon.

For the value u_0 we choose the characteristic function of $[1/3, 2/3]$, see Figure 6.1. Its expansion into the eigenvector basis of A can easily be calculated and the Fourier coefficients with respect to this basis decay sufficiently slowly to pose a meaningful test.

To apply the projection $\mathbf{PE}(0)$ and the projected semigroup $\mathbf{PE}(t)$ we choose the path of integration as an ellipse

$$\gamma(s) = c + a\cos(s) + ib\sin(s), \qquad s \in [0, 2\pi),$$

enclosing the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_K$ for some fixed $K > 0$. The most satisfactory results are obtained by the choice

$$c = (\lambda_1 + \lambda_K)/2, \quad a = (\lambda_K - \lambda_1)/2 + 8K, \quad b = a/K.$$

To justify Assumption 1 we compare the true error

$$\mathrm{err}(N) := \|\langle \widetilde{\mathbf{\Psi}}, Q_N(F) - e^{-tA} P_\Gamma u_0 \rangle\|_{\ell_2(\mathcal{J})} \sim \|Q_N(F)u_0 - e^{-tA} P_\Gamma u_0\|_{L_2(\Omega)}$$

with the error estimator

$$[\mathrm{err}(N)] := \|\langle \widetilde{\mathbf{\Psi}}, Q_N(F) - Q_{2N}(F) \rangle\|_{\ell_2(\mathcal{J})} \sim \|Q_N(F)u_0 - Q_{2N}(F)u_0\|_{L_2(\Omega)}$$

for the three time steps $t = -0.01, 0$, and 0.01 and $K = 2, 4, 6, 8$. The arising elliptic equations are solved up to machine accuracy. This latter accuracy is chosen only in order to validate Assumption 1 and should not be confused with the adapted tolerances in **Solve** as part of the numerical solution process.

In Figures 6.2, 6.3, and 6.4 we have plotted the true error and the error estimator as functions of N. Since the true error and the error estimator are too close in the graphical displays the ratio between true error and error estimator is additionally depicted as dotted line.

We observe an exponential convergence with respect to N as predicted by Theorem 4.1. The oscillating behavior of the error seems to be caused by the symmetry of the integrand

ETNA
Kent State University
etna@mcs.kent.edu

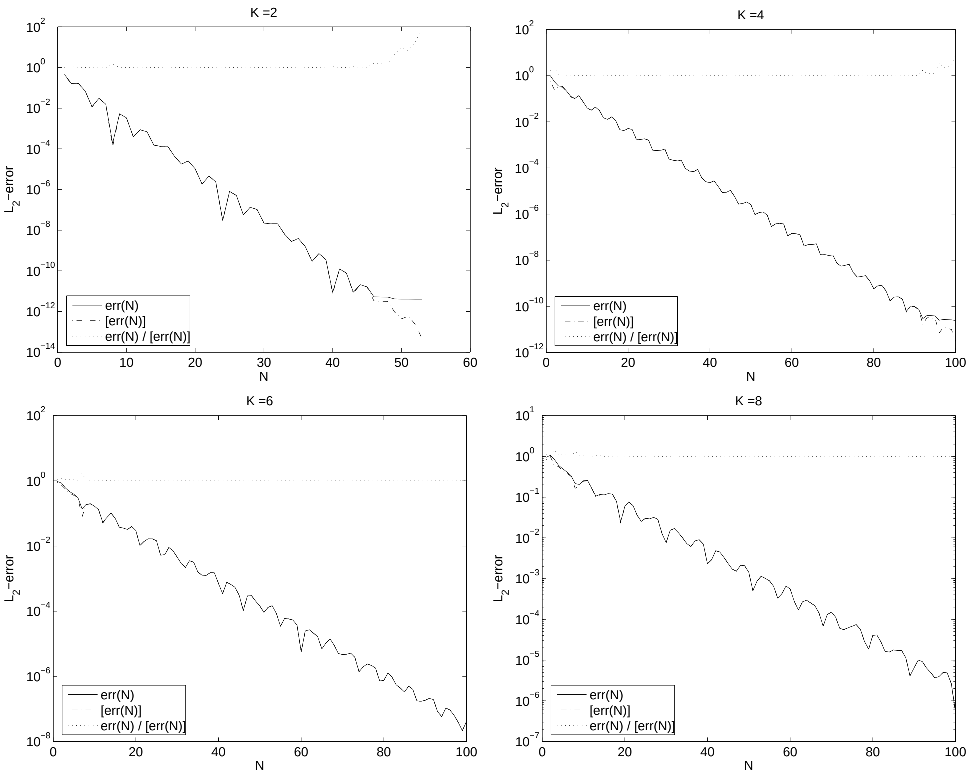

FIG. 6.2. *True error compared to error estimator for projection* $P_K u_0$

with respect to the real axis which was not taken into account by the analysis. It should be noted that the error estimator shows the same oscillatory behavior.

The speed of convergence becomes slower for increasing values of K. This indicates that $q > 0$, the width of the strip $\mathcal{D}_q$ to which γ extends analytically, decreases for increasing K. As for large K, the parameterization γ introduces a strong distortion of the complex plane, this effect is not surprising. It could be avoided by splitting Γ into several curves each enclosing one or a few neighboring eigenvalues.

When solving the backward problem, i.e. $t = -0.01$, for $K = 8$ the absolute error is substantially larger than for the other test cases. This effect is caused by the ill-posedness of the backward problem. For $t = -0.01$ the projected semigroup $e^{-tA}P_\Gamma$ has the highest eigenvalue $e^{-t\lambda_8} = 553.6\ldots$. Thus, some of the Fourier modes of u_0 are amplified substantially which results in a large norm of $e^{-tA}P_\Gamma u_0$ and an increased absolute error.

Concerning the quality of the error estimator we obtain a ratio of almost one between the true error and the estimated one. This close relation only breaks down for large N when the quadrature error drops below 10^{-8} which corresponds to the discretization error in $H_0^1(\Omega)$ when the wavelet basis is truncated at $J = 15$. Thus, the breakdown of the error estimator for large values of N seems to be caused by discretization errors. Thus, Assumption 1 is justified in our test case.

As the error estimator is almost equal to the true error already for small N, we choose the constants from the *adaptive algorithm* to be $N_0 = 5$ and $\varrho = \frac{1}{3}$ and test its reliability. In this experiment we observe that the actual error stays well below the target accuracy $\eta > 0$. Moreover, the actual error is roughly proportional to η which indicates that the analysis of

ETNA
Kent State University
etna@mcs.kent.edu

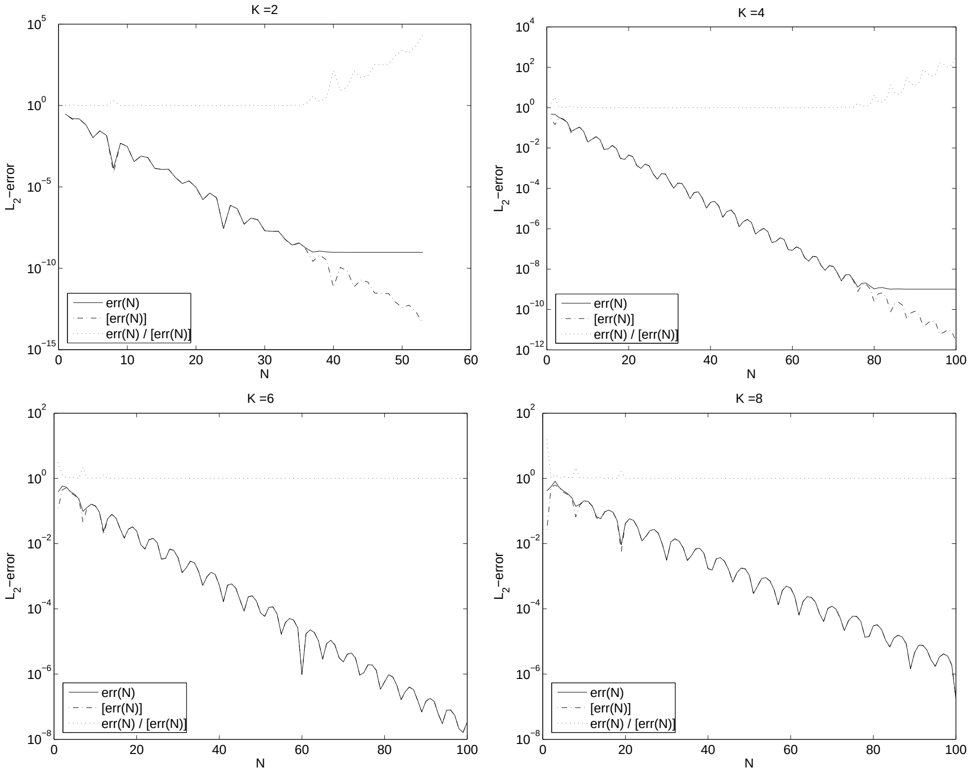

FIG. 6.3. *True error compared to error estimator for projected semigroup* $e^{-0.01A}P_K u_0$

Lemma 4.2 is not too pessimistic. We do not provide a graphical display of this results because there is no further insight to be gained from the plot.

6.2. Comparison of the four schemes. We strive to recover $x^\dagger = u_0$ from approximations y^δ to $y = u(t) = e^{-tA}u_0$, $t = 0.01$, where the initial value is given by $u_0(x) = \exp(\sin(2\pi x)) - 1$, $x \in (0,1)$. The perturbation $y^\delta - y$ is created by adding white noise to the wavelet coefficients (normalized in $L_2(0,1)$) such that error in the discrete $L_2(0,1)$ norm becomes exactly δ.

The four algorithms from §4.2 will be compared by measuring the difference of the regularized solution and the true one in the discrete $L_2(\Omega)$-norm, i.e. $\|x^\delta_\alpha - x^\dagger\|_{L_2(\Omega)}$ when the perturbation $\delta > 0$ tends to zero. The regularization parameter α will be chosen according to the discrepancy principle.

The results are reported in Figure 6.5. First, we observe that the conventional Tikhonov scheme and the modified one yield the same errors. This is somewhat disappointing as the modified Tikhonov scheme seems to be superior to the usual one by its construction. Nevertheless, the experiments indicate that the additional effort for applying the projection does not pay off.

Second, the regularization in the discrete setting with $s = 0$ produce an error close to the error of the Tikhonov regularization, but the methods with $s = -1/2$ and $s = -1$ yield more accurate approximations to $x^\dagger$ for large data errors δ. This effect was by no means clear in advance, as these methods regularize in norms different from $\|\cdot\|_{L_2(0,1)}$, in which the error was measured.

ETNA
Kent State University
etna@mcs.kent.edu

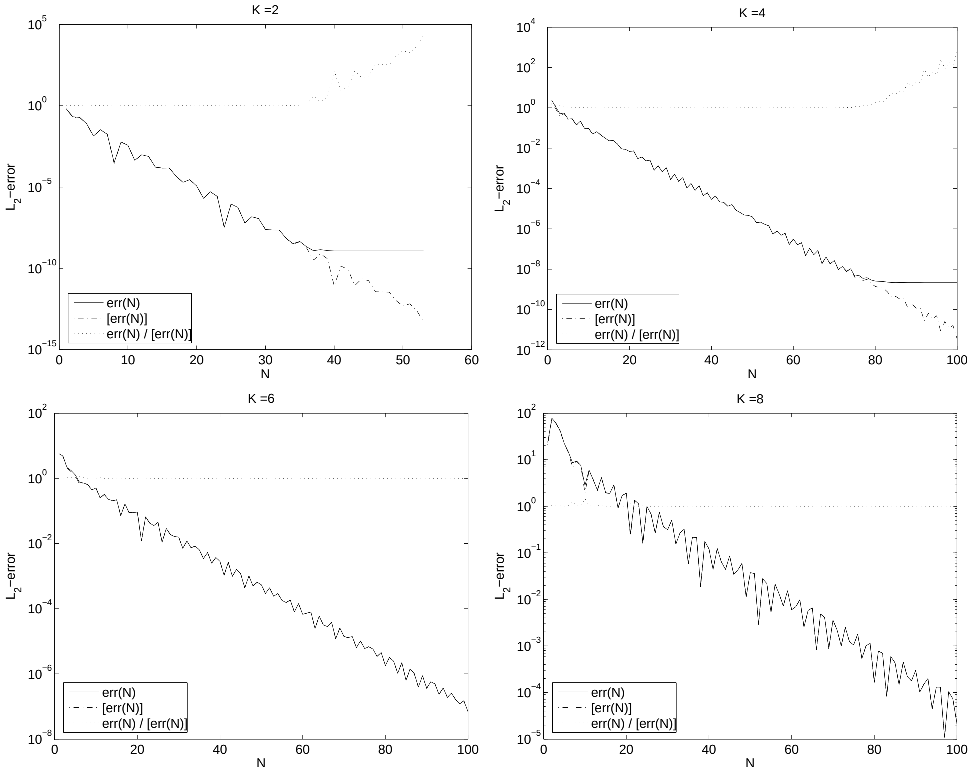

FIG. 6.4. *True error compared to error estimator for projected semigroup* $e^{0.01A} P_K u_0$

Third, we notice that the projection method yields errors $\|x_\alpha^\delta - x\|_{L_2(0,1)}$ which are much smaller than those of the other three methods. Since this method requires also less computational effort than the other methods as discussed in §4.2 it seems to be preferable over the other three methods.

As far as can be judged from Figure 6.5 none of the four methods converges with a substantially higher rate than the other ones. As the inversion of the operator exponential is a severely ill-posed problem (see [9] for a definition) there is perhaps no chance to improve on the quality of the standard methods but merely on their efficiency.

7. Conclusions. We have developed and implemented regularization schemes for the severely ill-posed inversion of diffusion processes. These schemes, including Tikhonov-type as well as projection methods, avoid any time stepping PDE solvers for the forward problems but rely on the error controlled approximate evaluation of certain Dunford integrals which are based on appropriate quadrature techniques and the error controlled solution of resolvent systems for well-posed operators. This latter task is trivially parallelized. In particular, this allows us to realize, to our knowledge for the first time in this context, approximate SVD-projections without actually having to compute or manipulate corresponding singular basis functions. Of course, the computational complexity is not proportional to the number of recovered modes but the scheme works also in situations where the SVD basis functions are not known and difficult to obtain. Moreover, the regularization is separated from the discretization of the diffusion operator yielding error bounds for the true infinite dimensional problem. Finally, we present a first discussion of the complexity of the projection scheme.

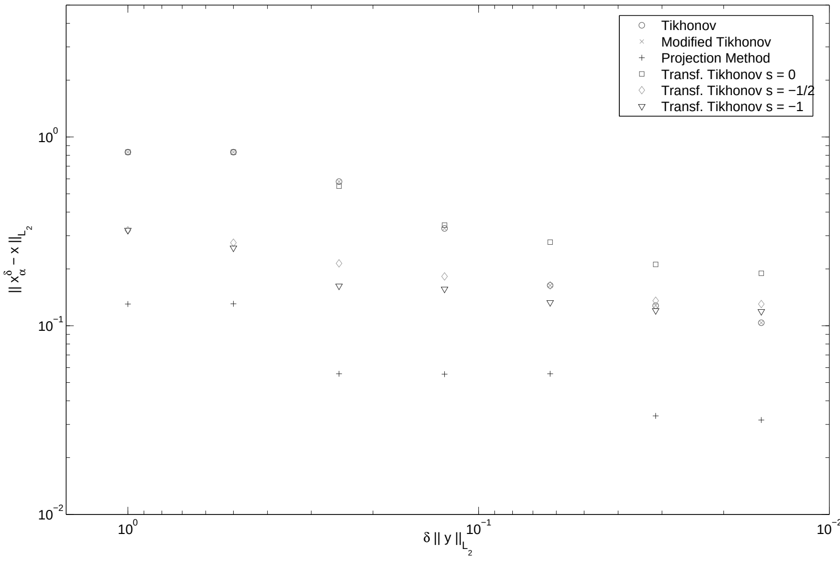

FIG. 6.5. *Comparison of Different Methods for Approximate Inversion of the Operator Exponential*

This together with the numerical experiments support its superiority over the Tikhonov-type schemes. This superiority is expected to grow with an increasing order of the wavelet basis since the qualification of the underlying projection scheme is (in contrast to the Tikhonov-type schemes) unbounded.

REFERENCES

[1] A. BARINKA, W. DAHMEN, AND R. SCHNEIDER, *Fast Computation of Adaptive Wavelet Expansions*, IGPM Report, RWTH Aachen, August 2004.
[2] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive wavelet methods II – Beyond the elliptic case*, Found. Comput. Math., 2 (2002), pp. 203–245.
[3] A. COHEN, M. HOFFMANN, AND M. REISS, *Adaptive wavelet-galerkin methods for inverse problems*, SIAM J. Numer. Anal., 42 (2004), 1479-1501.
[4] W. DAHMEN, *Wavelet and multiscale methods for operator equations*, Acta Numer., vol. 6, Cambridge University Press, 1997, pp. 55–228.
[5] W. DAHMEN, *Multiscale and wavelet methods for operator equations*, in Multiscale Problems and Methods in Numerical Simulation, C.I.M.E. Lecture Notes in Mathematics, vol. 1825, Springer, Heidelberg, 2003, pp. 31–96.
[6] W. DAHMEN, A. KUNOTH, AND K. URBAN, *Biorthogonal spline wavelets on the interval – stability and moment conditions*, Appl. Comput. Harmon. Anal., 6 (1999), pp. 132–196.
[7] W. DAHMEN AND R. SCHNEIDER, *Composite wavelet bases for operator equations*, Math. Comput., 68 (1999), pp. 1533–1567.
[8] R. DEVORE, *Nonlinear approximation*, Acta Numer., vol. 7, Cambridge University Press, 1998, pp. 51–150.
[9] H.W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 1996.
[10] K.-J. ENGEL AND R. NAGEL, *One-parameter Semigroups for Linear Evolution Equations*, Springer, 2000.
[11] I. P. GAVRILYUK, W. HACKBUSCH, AND B. N. KHOROMSKIJ, *$\mathcal{H}$-matrix approximation for the operator exponential with applications*, Numer. Math., 92 (2002), pp. 83–111 (2002).
[12] T. KATO, *Perturbation Theory for Linear Operators*, Corr. printing of the 2nd ed., Springer, 1980.
[13] R. KRESS, *Linear Integral Equations,* Springer, Berlin, 1999.
[14] A. LUNARDY, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser, 1995.

[15] M. JÜRGENS, *A Semigroup Approach to the Numerical Solution of Parabolic Differential Equations*, PhD thesis, RWTH Aachen, 2005.

[16] M. JÜRGENS, *Adaptive application of the operator exponential*, submitted to J. Numer. Math., special issue on Breaking Complexity: Multiscale Methods for Efficient PDE Solvers.

[17] R. STEVENSON, *On the compressibility of operators in wavelet coordinates*, SIAM J. Math. Anal., 35 (2004), pp. 1110–1132.