MOX-Report No. 54/2015

Adaptive Spectral Galerkin Methods with Dynamic Marking

Canuto, C.; Nochetto, R. H.; Stevenson, R.; Verani, M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Adaptive Spectral Galerkin Methods with Dynamic Marking

Claudio Canuto[a], Ricardo H. Nochetto[b], Rob Stevenson[c] and Marco Verani[d]

November 2, 2015

[a] Dipartimento di Scienze Matematiche, Politecnico di Torino
Corso Duca degli Abruzzi 24, I-10129 Torino, Italy
E-mail: claudio.canuto@polito.it

[b] Department of Mathematics and Institute for Physical Science and Technology
University of Maryland, College Park, MD, USA
E-mail: rhn@math.umd.edu

[c] Korteweg-de Vries Institute for Mathematics, University of Amsterdam
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
E-mail: r.p.stevenson@uva.nl

[d] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: marco.verani@polimi.it

Abstract

The convergence and optimality theory of adaptive Galerkin methods is almost exclusively based on the Dörfler marking. This entails a fixed parameter and leads to a contraction constant bounded below away from zero. For spectral Galerkin methods this is a severe limitation which affects performance. We present a dynamic marking strategy that allows for a super-linear relation between consecutive discretization errors, and show exponential convergence with linear computational complexity whenever the solution belongs to a Gevrey approximation class.

1 Introduction

The modern analysis of adaptive discretizations of partial differential equations aims at establishing rigorous results of *convergence* and *optimality*. The former results concern the convergence of the approximate solutions produced by the successive iterations of the adaptive algorithm towards the exact solution u, with an estimate of the error decay rate measured in an appropriate norm. On the other hand, optimality results compare the cardinality of the active set of basis functions used to expand the discrete solution to the minimal cardinality needed to approximate the exact solution with similar accuracy; this endeavor borrows ideas from Nonlinear Approximation theory. Confining ourselves in the sequel to second-order elliptic boundary value problems, such kind of analysis has been carried out first for wavelet discretizations [14, 18], then for h-type finite elements [19, 22, 13, 12, 15], [19] dealing just with convergence, and more recently for spectral-type methods [6, 7, 10]; we refer to the surveys [5, 11, 20, 24]. In contrast, the state of the art for hp-type finite elements is still in evolution; see [16, 3] and the more recent paper [8] which includes optimality estimates.

For all these cases, convergence is proven to be *linear*, i.e., a certain expression controlling the error (a norm, or a combination of norm and estimator) contracts with some fixed parameter $\rho < 1$ from one iteration to the next one, e.g., $\|u-u_{k+1}\| \leq \rho\|u-u_k\|$. This is typically achieved if the adaptation strategy is based on some form of *Dörfler marking* (or *bulk chasing*) with fixed parameter $\theta < 1$: assuming that $\sum_{i\in\mathcal{I}} \eta_i^2$ is some additive error estimator at iteration k, one identifies a minimal subset $\mathcal{I}' \subset \mathcal{I}$ such that

$$\sum_{i\in\mathcal{I}'} \eta_i^2 \geq \theta^2 \sum_{i\in\mathcal{I}} \eta_i^2$$

and utilizes $\mathcal{I}'$ for the construction of the new discretization at iteration $k+1$. For wavelet or h-type fem discretizations, optimality is guaranteed by performing cautious successive adaptations, i.e., by choosing a moderate value of θ, say $0 < \theta \leq \theta_{\max} < 1$ [22]. This avoids the need of cleaning-up the discrete solution from time to time, by subjecting it to a *coarsening* stage.

On the other hand, the resulting contraction factor $\rho = \rho(\theta)$ turns out to be bounded from below by a positive constant, say $0 < \rho_{\min} \leq \rho < 1$ (related to the 'condition number' of the exact problem), regardless of the choice of θ. This entails a limitation on the speed of convergence for infinite-order methods [6, 7], but is not restrictive for fixed-order methods [22, 13].

It has been shown in [6] that such an obstruction can be avoided if a specific property of the differential operator holds, namely the so-called *quasi-sparsity* of the inverse of the associated stiffness matrix. Upon exploiting this information, a more aggressive marking strategy can be adopted, which judiciously enlarges the set $\mathcal{I}'$ coming out of Dörfler's stage. The resulting contraction factor ρ can be now made arbitrarily close to 0 by choosing θ arbitrarily close to 1.

When a method of spectral type is used, one expects a fast (possibly, exponentially fast) decay of the discretization error for smooth solutions. In such a situation, a slow convergence of the iterations of the adaptive algorithm would endanger the overall performance of the method; from this perspective, it is useful to be able to make the contraction factor as close to 0 as desired. Yet, linear convergence of the adaptive iterations is not enough to guarantee the optimality of the method. Let us explain why this occurs, and why a *super-linear* convergence is preferable, using the following idealized setting.

As customary in Nonlinear Approximation, we consider the *best N-term approximation error* $E_N(u)$ of the exact solution u, in a suitable norm, using combinations of at most N functions taken from a chosen basis. We prescribe a decay law of $E_N(u)$ as N increases, which classically for fixed-order approximations is *algebraic* and reads

$$\sup_N N^s E_N(u) < \infty, \tag{1.1}$$

for some positive s. However, for infinite-order methods such as spectral approximations an *exponential* law is relevant that reads

$$\sup_N \mathrm{e}^{\eta N^\alpha} E_N(u) < \infty \tag{1.2}$$

for some $\eta > 0$ and $\alpha \in (0, 1]$, where $\alpha < 1$ accommodates the inclusion of C^∞-functions that are not analytic. This defines corresponding algebraic and exponential *sparsity classes* for the exact solution u. These classes are related to Besov and Gevrey regularity of u respectively.

We now assume the ideal situation that at each iteration of our adaptive algorithm [1]

$$\|u-u_k\| \eqsim N_k^{-s} \qquad \text{or} \qquad \|u-u_k\| \eqsim e^{-\eta N_k^\alpha}, \tag{1.3}$$

[1] Throughout the paper, we write $A_k \lesssim B_k$ to indicate that A_k can be bounded by a multiple of B_k, independently of the iteration counter k and other parameters which A_k and B_k may depend on; $A_k \eqsim B_k$ means $A_k \lesssim B_k$ and $B_k \lesssim A_k$.

where N_k is the cardinality of the discrete solution u_k, i.e., the dimension of the approximation space activated at iteration k. We assume in addition that the error decays linearly from one iteration to the next, i.e., it satisfies precisely

$$\|u - u_{k+1}\| = \rho \, \|u - u_k\|. \tag{1.4}$$

If u belongs to a sparsity class of algebraic type, then one easily gets $N_k \eqsim \rho^{-k/s}$, i.e., cardinalities grow exponentially fast and

$$\Delta N_k := N_{k+1} - N_k \eqsim N_k \eqsim \|u - u_k\|^{-1/s},$$

i.e., the increment of cardinality between consecutive iterations is proportional to the current cardinality as well as to the error raised to the power $-1/s$. The important message stemming from this ideal setting is that for a *practical* adaptive algorithm one should be able to derive the estimates $\|u - u_{k+1}\| \le \rho \, \|u - u_k\|$ and $\Delta N_k \lesssim \|u - u_k\|^{-1/s}$, because they yield

$$N_n = \sum_{k=0}^{n-1} \Delta N_k \lesssim \sum_{k=0}^{n-1} \|u - u_k\|^{-1/s} \le \|u - u_n\|^{-1/s} \sum_{k=0}^{n-1} \rho^{(n-k)/s} \lesssim \|u - u_n\|^{-1/s}.$$

This geometric-series argument is precisely the strategy used in [22, 13] and gives an estimate similar to (1.1). The performance of a practical adaptive algorithm is thus quasi-optimal.

If u belongs to a sparsity class of exponential type, instead, the situation changes radically. In fact, assuming (1.3) and (1.4), one has $e^{-\eta N_k^\alpha} \eqsim \rho^k$, and so

$$\lim_{k\to\infty} k^{-1/\alpha} N_k = \big(|\log \rho|/\eta\big)^{1/\alpha},$$

i.e., the cardinality N_k grows polynomially. For a practical adaptive algorithm, proving such a growth is very hard if not impossible. This obstruction has motivated the insertion of a coarsening stage in the adaptive algorithm presented in [6]. Coarsening removes the negligible components of the discrete solution possibly activated by the marking strategy and guarantees that the final cardinality is nearly optimal [14, 6], but it does not account for the workload to create u_k.

One of the key points of the present contribution is the observation that if the convergence of the adaptive algorithm is super-linear, then one is back to the simpler case of exponential growth of cardinalities which is ameanable to a sharper performance analysis. To see this, let us assume a super-linear relation between consecutive errors:

$$\|u - u_{k+1}\| = \|u - u_k\|^q \tag{1.5}$$

for some $q > 1$. If additionally u_k satisfies (1.3), then one infers that $\mathrm{e}^{-\eta N_{k+1}^\alpha} \eqsim \mathrm{e}^{-\eta q N_k^\alpha}$, whence

$$\lim_{k\to\infty} \frac{\Delta N_k}{N_k} = q^{1/\alpha} - 1, \qquad \lim_{k\to\infty} \frac{|\log \|u - u_k\||^{1/\alpha}}{N_k} = \eta^{1/\alpha},$$

the latter being just a consequence of (1.3). This suggests that the geometric-series argument may be invoked again in the optimality analysis of the adaptive algorithm.

This ideal setting does not apply directly to our *practical* adaptive algorithm. We will be able to prove estimates that are consistent with the preceding derivation to some extend, namely

$$\|u - u_{k+1}\| \le \|u - u_k\|^q, \qquad \Delta N_k \le Q |\log \|u - u_k\||^{1/\bar\alpha},$$

with constants $Q > 0$ and $\bar\alpha \in (0, \alpha]$. Invoking $\|u - u_n\| \le \|u - u_k\|^{q^{n-k}}$, we then realize that

$$N_n = \sum_{k=0}^{n-1} \Delta N_k \le Q \sum_{k=0}^{n-1} \big|\log \|u - u_k\|\,\big|^{1/\bar\alpha} \le \frac{Q q^{1/\bar\alpha}}{q^{1/\bar\alpha} - 1} \big|\log \|u - u_n\|\,\big|^{1/\bar\alpha}.$$

Setting $\bar\eta := \left(\frac{Qq^{1/\bar\alpha}}{q^{1/\bar\alpha}-1}\right)^{-\bar\alpha}$, we deduce the estimate

$$\sup_n \mathrm{e}^{\bar\eta N_n^{\bar\alpha}} \|u-u_n\| \le 1,$$

which is similar to (1.2), albeit with different class parameters. The most important parameter is $\bar\alpha$. Its possible degradation relative to α is mainly caused by the fact that the residual, the only computable quantity accessible to our practical algorithm, belongs to a sparsity class with a main parameter generally smaller than that of the solution u. This perhaps unexpected property is typical of the exponential class and has been elucidated in [6].

In order for the marking strategy to guarantee super-linear convergence, one needs to adopt a dynamic choice of Dörfler's parameter θ, which pushes its value towards 1 as the iterations proceed. We accomplish this requirement by equating the quantity $1-\theta_k^2$ to some function of the dual norm of the residual r_k, which is monotonically increasing and vanishing at the origin. This defines our *dynamic marking strategy*. The order of the root at the origin dictates the exponent q in the super-linear convergence estimate of our adaptive algorithm.

The paper is organized as follows. In Sect. 2 we introduce the model elliptic problem and its spectral Galerkin approximation based on either multi-dimensional Fourier or (modified) Legendre expansions. In particular, we highlight properties of the resulting stiffness matrix that will be fundamental in the sequel. We present the adaptive algorithm in Sect. 3, first for the static marking (θ fixed) and later for the dynamic marking (θ tending towards 1); super-linear convergence is proven. With the optimality analysis in mind, we next recall in Sect. 4 the definition and crucial properties of a family of sparsity classes of exponential type, related to Gevrey regularity of the solution, and we investigate how the sparsity class of the Galerkin residual deteriorates relative to that of the exact solution. Finally, in Sect. 5 we relate the cardinality of the adaptive discrete solutions, as well as the workload needed to compute them, to the expected accuracy of the approximation. Our analysis confirms that the proposed dynamic marking strategy avoids any form of coarsening, while providing exponential convergence with linear computational complexity, assuming optimal linear solvers.

2 Model Elliptic Problem and Galerkin Methods

Let $d \ge 1$ and consider the following elliptic PDE in a d-dimensional rectangular domain Ω with periodic or homogeneous Dirichlet boundary conditions:

$$Lu = -\nabla\cdot(\nu\nabla u) + \sigma u = f \qquad \text{in } \Omega, \tag{2.1}$$

where ν and σ are sufficiently smooth real coefficients satisfying $0 < \nu_* \le \nu(x) \le \nu^* < \infty$ and $0 < \sigma_* \le \sigma(x) \le \sigma^* < \infty$ in Ω; let us set

$$\alpha_* = \min(\nu_*, \sigma_*) \qquad \text{and} \qquad \alpha^* = \max(\nu^*, \sigma^*)\ .$$

Let V be equal to $H_0^1(\Omega)$ or $H_p^1(\Omega)$ depending on the boundary conditions and denote by V^* its dual space. We formulate (2.1) variationally as

$$u \in V \quad : \quad a(u,v) = \langle f, v\rangle \qquad \forall v \in V\ , \tag{2.2}$$

where $a(u,v) = \int_\Omega \nu\nabla u\cdot\nabla\bar v + \int_\Omega \sigma u\bar v$ (bar indicating as usual complex conjugate). We denote by $|\!|\!| v |\!|\!| = \sqrt{a(v,v)}$ the energy norm of any $v \in V$, which satisfies

$$\sqrt{\alpha_*}\|v\|_V \le |\!|\!| v |\!|\!| \le \sqrt{\alpha^*}\|v\|_V\ . \tag{2.3}$$

2.1 Riesz Basis

We start with an abstract formulation which encompasses the two examples of interest: trigonometric functions and Legendre polynomials. Let $\phi = \{\phi_k : k \in \mathcal{K}\}$ be a Riesz basis of V. Thus, we assume the following relation between a function $v = \sum_{k\in\mathcal{K}} \hat{v}_k \phi_k \in V$ and its coefficients:

$$\|v\|_V^2 \simeq \sum_{k\in\mathcal{K}} |\hat{v}_k|^2 d_k =: \|v\|_\phi^2 , \tag{2.4}$$

for suitable weights $d_k > 0$. Correspondingly, any element $f \in V^*$ can be expanded along the *dual basis* $\phi^* = \{\phi_k^*\}$ as $f = \sum_{k\in\mathcal{K}} \hat{f}_k \phi_k^*$, with $\hat{f}_k = \langle f, \phi_k\rangle$, yielding the dual norm representation

$$\|f\|_{V^*}^2 \simeq \sum_{k\in\mathcal{K}} |\hat{f}_k|^2 d_k^{-1} =: \|v\|_{\phi*}^2 . \tag{2.5}$$

For future reference, we introduce the vectors $\mathbf{v} = (\hat{v}_k d_k^{1/2})_{k\in\mathcal{K}}$ and $\mathbf{f} = (\hat{f}_k d_k^{-1/2})_{k\in\mathcal{K}}$ as well as the constants $\beta_* \le \beta^*$ of the norm equivalence in (2.4)

$$\beta_* \|v\|_V \le \|v\|_\phi = \|\mathbf{v}\|_{\ell^2} \le \beta^* \|v\|_V \qquad \forall v \in V . \tag{2.6}$$

This implies

$$\frac{1}{\beta^*}\|f\|_{V^*} \le \|f\|_{\phi^*} = \|\mathbf{f}\|_{\ell^2} \le \frac{1}{\beta_*}\|f\|_{V^*} \qquad \forall f \in V^* . \tag{2.7}$$

The two key examples to keep in mind are trigonometric basis and tensor products of Babuška-Shen basis. We discuss them briefly below.

Trigonometric basis. Let $\Omega = (0, 2\pi)^d$ and the trigonometric basis be $\phi_k(x) = \frac{1}{(2\pi)^{d/2}} \mathrm{e}^{ik\cdot x}$ for any $k \in \mathcal{K} = \mathbb{Z}^d$ and $x \in \Omega$. Any function $v \in L^2(\Omega)$ can be expanded in terms of $\{\phi_k\}_{k\in\mathbb{Z}^d}$ as follows:

$$v = \sum_k \hat{v}_k \phi_k , \qquad \hat{v}_k = \langle v, \phi_k\rangle , \qquad \|v\|_{L^2(\Omega)}^2 = \sum_k |\hat{v}_k|^2 . \tag{2.8}$$

The space $V := H_p^1(\Omega)$ of periodic functions with square integrable weak gradient can now be easily characterized as the subspace of those $v \in L^2(\Omega)$ for which

$$\|v\|_V^2 = \|v\|_{H_p^1(\Omega)}^2 = \sum_k |\hat{V}_k|^2 < \infty \qquad (\text{where } \hat{V}_k := \hat{v}_k d_k^{1/2}, \text{ with } d_k := 1 + |k|^2).$$

This induces an *isomorphism* between $H_p^1(\Omega)$ and $\ell^2(\mathbb{Z}^d)$: for each $v \in H_p^1(\Omega)$ let $\mathbf{v} = (\hat{V}_k)_{k\in\mathcal{K}} \in \ell^2(\mathbb{Z}^d)$ and note that $\|v\|_{H_p^1(\Omega)} = \|\mathbf{v}\|_{\ell^2}$. Likewise, the dual space $H_p^{-1}(\Omega) = (H_p^1(\Omega))'$ is characterized as the space of those functionals f for which

$$\|f\|_{V^*}^2 = \|f\|_{H_p^{-1}(\Omega)}^2 = \sum_k |\hat{F}_k|^2 \qquad \text{with} \quad \hat{F}_k := \hat{f}_k d_k^{-1/2}.$$

We also have an isomorphism between $H_p^{-1}(\Omega)$ and $\ell^2(\mathbb{Z}^d)$ upon setting $\mathbf{f} = (\hat{F}_k)_{k\in\mathcal{K}}$ for $f \in H_p^{-1}(\Omega)$ and realizing that $\|f\|_{H_p^{-1}(\Omega)} = \|\mathbf{f}\|_{\ell^2}$.

Babuška-Shen basis. Let us start with the one-dimensional case $d = 1$. Set $I = (-1, 1)$, $V := H_0^1(I)$, and let $L_k(x)$, $k \ge 0$, stand for the k-th Legendre orthogonal polynomial in I, which satisfies $\deg L_k = k$, $L_k(1) = 1$ and

$$\int_I L_k(x) L_m(x)\, dx = \frac{2}{2k+1}\, \delta_{km} , \qquad m \ge 0 . \tag{2.9}$$

The natural modal basis in $H_0^1(I)$ is the *Babuška-Shen basis* (BS basis), whose elements are defined as

$$\eta_k(x) = \sqrt{\frac{2k-1}{2}} \int_x^1 L_{k-1}(s)\, ds = \frac{1}{\sqrt{4k-2}} \big(L_{k-2}(x) - L_k(x)\big) \ , \qquad k \geq 2 \ . \tag{2.10}$$

The basis elements satisfy $\deg \eta_k = k$ and

$$(\eta_k, \eta_m)_{H_0^1(I)} = \int_I \eta_k'(x) \eta_m'(x)\, dx = \delta_{km} \ , \qquad k, m \geq 2 \ , \tag{2.11}$$

i.e., they form an orthonormal basis for the $H_0^1(I)$-inner product. Equivalently, the (semi-infinite) stiffness matrix S_η of the Babuška-Shen basis with respect to this inner product is the identity matrix.

We now consider, for simplicity, the two-dimensional case $d = 2$ since the case $d > 2$ is similar. Let $\Omega = (-1, 1)^2$, $V = H_0^1(\Omega)$, and consider the *tensorized Babuška-Shen basis*, whose elements are defined as

$$\eta_k(x) = \eta_{k_1}(x_1) \eta_{k_2}(x_2) \ , \qquad k_1, k_2 \geq 2 \ , \tag{2.12}$$

where we set $k = (k_1, k_2)$ and $x = (x_1, x_2)$; indices vary in the set $\mathcal{K} = \{k \in \mathbb{N}^2 \, : \, k_i \geq 2 \text{ for } i = 1, 2\}$, which is ordered 'a la Cantor' by increasing total degree $k_{\rm tot} = k_1 + k_2$ and, for the same total degree, by increasing k_1. The tensorized BS basis is no longer orthogonal, since

$$(\eta_k, \eta_m)_{H_0^1(\Omega)} = (\eta_{k_1}, \eta_{m_1})_{H_0^1(I)} (\eta_{k_2}, \eta_{m_2})_{L^2(I)} + (\eta_{k_1}, \eta_{m_1})_{L^2(I)} (\eta_{k_2}, \eta_{m_2})_{H_0^1(I)} \ , \tag{2.13}$$

whence $(\eta_k, \eta_m)_{H_0^1(\Omega)} \neq 0$ if and only if $k_1 = m_1$ and $k_2 - m_2 \in \{-2, 0, 2\}$, or $k_2 = m_2$ and $k_1 - m_1 \in \{-2, 0, 2\}$. Obviously, we cannot have a Parseval representation of the $H_0^1(\Omega)$-norm of $v = \sum_{k \in \mathcal{K}} \hat{v}_k \eta_k$ in terms of the coefficients $\hat{v}_k$. With the aim of getting (2.4), we follow [10] and we first perform the orthonormalization of the BS basis via a Gram-Schmidt procedure. This allows us to build a sequence of functions

$$\Phi_k = \sum_{m \leq k} g_{mk} \eta_m \ , \tag{2.14}$$

such that $g_{kk} \neq 0$ and

$$(\Phi_k, \Phi_m)_{H_0^1(\Omega)} = \delta_{km} \qquad \forall \ k, m \in \mathcal{K} \ .$$

We will refer to the collection $\Phi := \{\Phi_k : \ k \in \mathcal{K}\}$ as the *orthonormal Babuška-Shen basis* (OBS basis), for which the associated stiffness matrix S_Φ with respect to the $H_0^1(\Omega)$-inner product is the identity matrix. Equivalently, if $G = (g_{mk})$ is the upper triangular matrix which collects the coefficients generated by the Gram-Schmidt algorithm above, one has

$$G^T S_\eta G = S_\Phi = I \ , \tag{2.15}$$

that is the validity of (2.6) with $d_k = 1$. However, unlike S_η, which is very sparse, the upper triangular matrix G is full; in view of this, we next apply a thresholding procedure to wipe-out a significant portion of the non-zero entries sitting in the leftmost columns of G. This leads to a modified basis whose computational efficiency is quantitatively improved, without significantly deteriorating the properties of the OBS basis. To be more precise, we use the following notation: G_t indicates the matrix obtained from G by setting to zero a certain finite set of off-diagonal entries, so that in particular $\mathrm{diag}(G_t) = \mathrm{diag}(G)$; correspondingly, $E := G_t - G$ is the matrix measuring the truncation quality, for which $\mathrm{diag}(E) = 0$. Finally, we introduce the matrix

$$S_\phi = G_t^T S_\eta G_t \tag{2.16}$$

which we interpret as the stiffness matrix associated to the modified BS basis defined in analogy to (2.14) as

$$\phi_k = \sum_{m \in \mathcal{M}_t(k)} g_{mk} \eta_m \tag{2.17}$$

where $\mathcal{M}_t(k) = \{m \leq k : E_{mk} = 0\}$. This forms a new basis in $H_0^1(\Omega)$ (because $k \in \mathcal{M}_t(k)$ and $g_{kk} \neq 0$). We will term it a *nearly-orthonormal Babuška-Shen basis* (NOBS basis). Note that only the basis functions ϕ_k having total degree not exceeding a certain value, say p, may be affected by the compression, while all the others coincide with the corresponding orthonormal basis functions Φ_k defined in (2.14).

If $D_\phi = \operatorname{diag} S_\phi$, then for any value of p there are strategies to build G_t (depending on p)such that the eigenvalues λ of

$$S_\phi x = \lambda D_\phi x \tag{2.18}$$

are close to one and bounded from above and away from 0, independently of p [10]. This guarantees the validity for NOBS basis of (2.4) with d_k equal to the diagonal elements of the matrix D_ϕ and (2.6) for suitable choice of constants β_*, β^* depending on the eigenvalues of (2.18) (see [10] for more details).

2.2 Infinite Dimensional Algebraic Problem

Let us identify the solution $u = \sum_k \hat{u}_k \phi_k$ of Problem (2.1) with the vector $\mathbf{u} = (\hat{u}_k)_{k \in \mathcal{K}}$ of its coefficients w.r.t. the basis $\{\phi_k\}_{k \in \mathcal{K}}$. Similarly, let us identify the right-hand side f with the vector $\mathbf{f} = (\hat{f}_\ell)_{\ell \in \mathcal{K}}$ of its dual coefficients. Finally, let us introduce the bi-infinite, symmetric and positive-definite stiffness matrix

$$\mathbf{A} = (a_{\ell,k})_{\ell,k \in \mathcal{K}} \qquad \text{with} \qquad a_{\ell,k} = a(\phi_k, \phi_\ell) \ . \tag{2.19}$$

Then, Problem (2.1) can be equivalently written as

$$\mathbf{A}\mathbf{u} = \mathbf{f} \ , \tag{2.20}$$

where, thanks to (2.3) and the norm equivalences (2.6)-(2.7), $\mathbf{A}$ defines a bounded invertible operator in $\ell^2(\mathcal{K})$.

Decay Properties of $\mathbf{A}$ and $\mathbf{A}^{-1}$. The decay of the entries of $\mathbf{A}$ away from the diagonal depends on the regularity of the coefficients ν and σ of L. If ν and σ are real analytic in a neighborhood of Ω, then $a_{k,m}$ decays exponentially away from the diagonal [6, 7, 10]: there exist parameters $c_L, \eta_L > 0$ such that

$$|a_{k,m}| \leq \ c_L \exp(-\eta_L |k - m|) \quad \forall k, m \in \mathcal{K}; \tag{2.21}$$

we then say that $\mathbf{A}$ belongs to the exponential class $\mathcal{D}_e(\eta_L)$, in particular $\mathbf{A}$ is quasi-sparse. This justifies the *symmetric truncation* $\mathbf{A}_J$ of $\mathbf{A}$ with *parameter* J, defined as $(\mathbf{A}_J)_{\ell,k} = a_{\ell,k}$ if $|\ell - k| \leq J$ and $(\mathbf{A}_J)_{\ell,k} = 0$ otherwise, which satisfies [6, 7, 10]

$$\|\mathbf{A} - \mathbf{A}_J\| \leq C_{\mathbf{A}} (J+1)^{d-1} \mathrm{e}^{-\eta_L J} \tag{2.22}$$

for some $C_{\mathbf{A}} > 0$ depending only on c_L.

Most notably, the inverse matrix $\mathbf{A}^{-1}$ is also quasi-sparse [6, 7, 10]. Precisely $\mathbf{A}^{-1} \in \mathcal{D}_e(\bar{\eta}_L)$ for some $\bar{\eta}_L \in (0, \eta_L]$ and $\bar{c}_L$ only dependent on c_L and η_L. Thus, there exists an explicit constant $C_{\mathbf{A}^{-1}}$ (depending only on c_L and η_L) such that the symmetric truncation $(\mathbf{A}^{-1})_J$ of $\mathbf{A}^{-1}$ satisfies

$$\|\mathbf{A}^{-1} - (\mathbf{A}^{-1})_J\| \le C_{\mathbf{A}^{-1}}(J+1)^{d-1}\mathrm{e}^{-\bar{\eta}_L J} \le C_{\mathbf{A}^{-1}}\mathrm{e}^{-\tilde{\eta}_L J} \ , \tag{2.23}$$

for a suitable exponent $\widetilde{\eta}_L < \bar{\eta}_L$.

Galerkin Method. Given any finite index set $\Lambda \subset \mathcal{K}$, we define the subspace $V_\Lambda = \mathrm{span}\,\{\phi_k \,|\, k \in \Lambda\}$ of V; we set $|\Lambda| = \mathrm{card}\,\Lambda$, so that $\dim \mathrm{V}_\Lambda = |\Lambda|$. If $v \in V$ admits the expansion $v = \sum_{k\in\mathcal{K}} \hat{v}_k\phi_k$, then we define its projection $P_\Lambda v$ upon V_Λ by setting $P_\Lambda v := \sum_{k\in\Lambda} \hat{v}_k\phi_k$. Similarly, we define the subspace $V_\Lambda^* = \mathrm{span}\,\{\phi_k^* \,|\, k \in \Lambda\}$ of V^*. If f admits an expansion $f = \sum_{k\in\mathcal{K}} \hat{f}_k\phi_k^*$, then we define its projection $P_\Lambda^* f$ onto V_Λ^* upon setting $P_\Lambda^* f := \sum_{k\in\Lambda} \hat{f}_k\phi_k^*$.

Given any finite $\Lambda \subset \mathcal{K}$, the Galerkin approximation of (2.1) is defined as

$$u_\Lambda \in V_\Lambda \quad : \quad a(u_\Lambda, v_\Lambda) = \langle f, v_\Lambda\rangle \qquad \forall v_\Lambda \in V_\Lambda \ . \tag{2.24}$$

Let $\mathbf{u}_\Lambda$ be the vector collecting the coefficients of u_Λ indexed in Λ; let $\mathbf{f}_\Lambda$ be the analogous restriction for the vector of the coefficients of f. Finally, denote by $\mathbf{R}_\Lambda$ the matrix that restricts a vector indexed in $\mathcal{K}$ to the portion indexed in Λ, so that $\mathbf{R}_\Lambda^H$ is the corresponding extension matrix. If

$$\mathbf{A}_\Lambda := \mathbf{R}_\Lambda\mathbf{A}\mathbf{R}_\Lambda^H \ , \tag{2.25}$$

then problem (2.24) can be equivalently written as

$$\mathbf{A}_\Lambda\mathbf{u}_\Lambda = \mathbf{f}_\Lambda \ . \tag{2.26}$$

For any $w \in V_\Lambda$, we define the residual $r(w) \in V^*$ as

$$r(w) = f - Lw = \sum_{k\in\mathcal{K}} \hat{r}_k(w)\phi_k^* \ , \qquad \text{where} \qquad \hat{r}_k(w) = \langle f - Lw, \phi_k\rangle = \langle f, \phi_k\rangle - a(w, \phi_k) \ .$$

The definition (2.24) of u_Λ is equivalent to the condition $P_\Lambda^* r(u_\Lambda) = 0$, i.e., $\hat{r}_k(u_\Lambda) = 0$ for every $k \in \Lambda$. By the continuity and coercivity of the bilinear form, one has

$$\frac{1}{\alpha^*}\|r(u_\Lambda)\|_{V^*} \le \|u - u_\Lambda\|_V \le \frac{1}{\alpha_*}\|r(u_\Lambda)\|_{V^*} \ , \tag{2.27}$$

which in view of (2.3) and (2.7) can be rephrased as

$$\frac{\beta_*}{\sqrt{\alpha^*}}\|r(u_\Lambda)\|_{\phi^*} \le |\!|\!| u - u_\Lambda |\!|\!| \le \frac{\beta^*}{\sqrt{\alpha_*}}\|r(u_\Lambda)\|_{\phi^*} \ . \tag{2.28}$$

Therefore, if $(\hat{r}_k(u_\Lambda))_{k\in\mathcal{K}}$ are the coefficients of $r(u_\Lambda)$ with respect to the dual basis ϕ^*, the quantity

$$\|r(u_\Lambda)\|_{\phi^*} = \left(\sum_{k\notin\Lambda} |\hat{R}_k(u_\Lambda)|^2\right)^{1/2} \qquad \text{with} \quad \hat{R}_k(u_\Lambda) = \hat{r}_k(u_\Lambda)d_k^{-1/2}$$

is an *error estimator* from above and from below. However, this quantity is not computable because it involves infinitely many terms. We discuss *feasible versions* in [6, 7, 10] but not here.

Equivalent Formulation of the Galerkin Problem. For future reference, we now rewrite the Galerkin problem (2.24) in an equivalent (infinite-dimensional) manner. Let

$$\mathbf{P}_\Lambda : \ell^2(\mathcal{K}) \to \ell^2(\mathcal{K})$$

be the projector operator defined as

$$(\mathbf{P}_\Lambda \mathbf{v})_\lambda := \begin{cases} v_\lambda & \text{if } \lambda \in \Lambda\ , \\ 0 & \text{if } \lambda \notin \Lambda\ . \end{cases}$$

Note that $\mathbf{P}_\Lambda$ can be represented as a diagonal bi-infinite matrix whose diagonal elements are 1 for indexes belonging to Λ, and zero otherwise. We set $\mathbf{Q}_\Lambda := \mathbf{I} - \mathbf{P}_\Lambda$ and introduce the bi-infinite matrix $\widehat{\mathbf{A}}_\Lambda := \mathbf{P}_\Lambda \mathbf{A} \mathbf{P}_\Lambda + \mathbf{Q}_\Lambda$ which is equal to $\mathbf{A}_\Lambda$ for indexes in Λ and to the identity matrix, otherwise. The definitions of the projectors $\mathbf{P}_\Lambda$ and $\mathbf{Q}_\Lambda$ yield the following property:

$$\textit{If } \mathbf{A} \textit{ is invertible with } \mathbf{A} \in \mathcal{D}_e(\eta_L), \textit{ then the same holds for } \widehat{\mathbf{A}}_\Lambda. \tag{2.29}$$

Furthermore, the constants $C_{\widehat{\mathbf{A}}_\Lambda}$ and $C_{(\widehat{\mathbf{A}}_\Lambda)^{-1}}$ which appear in the inequalities (2.22) and (2.23) for $\widehat{\mathbf{A}}_\Lambda$ can be bounded uniformly in Λ, since in turn they can be bounded in terms of η_L and c_L, respectively.
Now, let us consider the following extended Galerkin problem: find $\hat{\mathbf{u}} \in \ell^2$ such that

$$\widehat{\mathbf{A}}_\Lambda \hat{\mathbf{u}} = \mathbf{P}_\Lambda \mathbf{f}\ . \tag{2.30}$$

Let $\mathbf{u}_\Lambda$ be the Galerkin solution to (2.26); then, it is easy to check that $\hat{\mathbf{u}} = \mathbf{R}_\Lambda^H \mathbf{u}_\Lambda$.

3 Adaptive Spectral Galerkin Method

In this section we present our adaptive spectral Galerkin method, named **DYN-GAL**, that is based on a new notion of marking strategy, namely a *dynamic marking*. In Section 3.1 we recall the enriched Dörfler marking strategy, introduced in [6], which represents an enhancement of the classic Dörfler marking strategy. In Section 3.2 we introduce the dynamic marking strategy, present **DYN-GAL** and prove its quadratic convergence.

3.1 Static Dörfler Marking

Fix any $\theta \in (0,1)$ and set $\Lambda_0 = \emptyset$, $u_{\Lambda_0} = 0$. For $n = 0, 1, \ldots$, assume that Λ_n and $u_n := u_{\Lambda_n} \in V_{\Lambda_n}$ and $r_n := r(u_n) = Lu_n - f$ are already computed and choose $\Lambda_{n+1} := \Lambda_n \cup \partial\Lambda_n$ where the set $\partial\Lambda_n$ is built by a two-step procedure that we call **E-DÖRFLER** for *enriched Dörfler*:

$$\begin{aligned} \partial\Lambda_n &= \textbf{E-DÖRFLER}\,(\Lambda_n, \theta) \\ \widetilde{\partial\Lambda}_n &= \textbf{DORFLER}\,(r_n, \theta) \\ \partial\Lambda_n &= \textbf{ENRICH}\,(\widetilde{\partial\Lambda}_n, J) \end{aligned}$$

The first step is the usual *Dörfler's marking* with parameter θ:

$$\|P^*_{\widetilde{\partial\Lambda}_n} r_n\|_{\ell^*} = \|P^*_{\widetilde{\Lambda}_{n+1}} r_n\|_{\ell^*} \geq \theta \|r_n\|_{\ell^*} \quad \text{or} \quad \sum_{k \in \widetilde{\partial\Lambda}_n} |\hat{R}_k(u_n)|^2 \ \geq\ \theta^2 \sum_{k \in \mathcal{K}} |\hat{R}_k(u_n)|^2\ , \tag{3.1}$$

with $\widetilde{\Lambda}_{n+1} = \Lambda_n \cup \widetilde{\partial}\Lambda_n$. This also reads

$$\|r_n - P^*_{\widetilde{\Lambda}_{n+1}} r_n\|_{\ell^*} \leq \sqrt{1-\theta^2}\, \|r_n\|_{\ell^*} \tag{3.2}$$

and can be implemented by rearranging the coefficients $\hat{R}_k(u_n)$ in decreasing order of modulus and picking the largest ones (*greedy approach*). However, this is only an idealized algorithm because the number of coefficients $\hat{R}_k(u_n)$ is infinite. This marking is known to yield a contraction property between u_n and the Galerkin solution $\widetilde{u}_{n+1} \in V_{\widetilde{\Lambda}_{n+1}}$ of the form

$$|||u - \widetilde{u}_{n+1}||| \leq \rho(\theta) |||u - u_n|||\ ,$$

with $\rho(\theta) = \sqrt{1 - \frac{\alpha_*}{\alpha^*}\theta^2}$ [6, 7]. When $\alpha_* < \alpha^*$ we see that in contrast to (3.2), $\rho(\theta)$ is bounded below away from 0 by $\sqrt{1 - \frac{\alpha_*}{\alpha^*}}$.

The second step of **E-DÖRFLER** is meant to remedy this situation and hinges on the a priori structure of $\mathbf{A}^{-1}$ already alluded to in §2.2. The goal is to augment the set $\widetilde{\partial\Lambda}_n$ to $\partial\Lambda_n$ judiciously. This is contained in the following proposition (see [6]) whose proof is reported here for completeness.

Proposition 3.1 (enrichment)**.** *Let* $\widetilde{\partial\Lambda}_n = \mathbf{DORFLER}\,(r_n, \theta)$*, and let* $J = J(\theta) > 0$ *satisfy*

$$C_{\mathbf{A}^{-1}}\, e^{-\widetilde{\eta}_L J} \leq \sqrt{\frac{1-\theta^2}{\alpha_*\alpha^*}}\ , \tag{3.3}$$

where $C_{\mathbf{A}^{-1}}$ *and* $\widetilde{\eta}_L$ *are defined in* (2.23)*. Let* $\partial\Lambda_n = \mathbf{ENRICH}\,(\widetilde{\partial\Lambda}_n, J)$ *be built as follows*

$$\partial\Lambda_n := \big\{k \in \mathcal{K} : \quad \textit{there exists } \ell \in \widetilde{\partial\Lambda}_n \textit{ such that } |k-\ell| \leq J\big\}.$$

Then for $\Lambda_{n+1} = \Lambda_n \cup \partial\Lambda_n$*, the Galerkin solution* $u_{n+1} \in V_{\Lambda_{n+1}}$ *satisfies*

$$|\!|\!| u - u_{n+1} |\!|\!| \leq \bar{\rho}(\theta) |\!|\!| u - u_n |\!|\!| \tag{3.4}$$

with

$$\bar{\rho}(\theta) = 2\frac{\beta^*\sqrt{\alpha^*}}{\beta_*\sqrt{\alpha_*}}\sqrt{1-\theta^2}. \tag{3.5}$$

Proof. Let $g_n := P^*_{\widetilde{\partial\Lambda}_n} r_n = P^*_{\widetilde{\Lambda}_{n+1}} r_n$ which, according to (3.2), satisfies

$$\|r_n - g_n\|_{\phi^*} \leq \sqrt{1-\theta^2}\|r_n\|_{\phi^*}\ .$$

Let $w_n \in V$ be the solution of $Lw_n = g_n$, which in general will have infinitely many components, and let us split it as

$$w_n = P_{\Lambda_{n+1}} w_n + P_{\Lambda^c_{n+1}} w_n =: y_n + z_n \in V_{\Lambda_{n+1}} \oplus V_{\Lambda^c_{n+1}}\ .$$

The minimality property in the energy norm of the Galerkin solution u_{n+1} over the set Λ_{n+1} yet to be defined, in conjunction with (2.3) and (2.28), implies

$$\begin{aligned} |\!|\!| u - u_{n+1} |\!|\!| &\leq |\!|\!| u - (u_n + y_n) |\!|\!| \leq |\!|\!| u - u_n - w_n + z_n |\!|\!| \\ &\leq \frac{1}{\sqrt{\alpha_*}}\|L(u - u_n - w_n)\| + \sqrt{\alpha^*}\|z_n\| = \frac{\beta^*}{\sqrt{\alpha_*}}\|r_n - g_n\|_{\phi^*} + \sqrt{\alpha^*}\|z_n\|\ , \end{aligned}$$

whence

$$|\!|\!| u - u_{n+1} |\!|\!| \leq \frac{\beta^*}{\sqrt{\alpha_*}}\sqrt{1-\theta^2}\,\|r_n\|_{\phi^*} + \sqrt{\alpha^*}\|z_n\|\ .$$

Since $z_n = \big(P_{\Lambda^c_{n+1}} L^{-1} P^*_{\widetilde{\partial\Lambda}_n}\big) r_n$, we now construct Λ^c_{n+1} to control $\|z_n\|$. If

$$k \in \Lambda^c_{n+1} \quad \text{and} \quad \ell \in \widetilde{\partial\Lambda}_n \qquad \Rightarrow \qquad |k-\ell| > J\ ,$$

then we have

$$\|P_{\Lambda^c_{n+1}} L^{-1} P^*_{\widetilde{\partial\Lambda}_n}\| \leq \|\mathbf{A}^{-1} - (\mathbf{A}^{-1})_J\| \leq C_{\mathbf{A}^{-1}} \mathrm{e}^{-\widetilde{\eta}_L J}\ ,$$

where we have used (2.23). We now choose $J = J(\theta) > 0$ to satisfy (3.3), and we exploit that $\|z_n\| \leq C_{\mathbf{A}^{-1}} e^{-\widetilde{\eta}_L J}\|r_n\|$ to obtain

$$|\!|\!| u - u_{n+1} |\!|\!| \leq 2\frac{\beta^*}{\sqrt{\alpha_*}}\sqrt{1-\theta^2}\,\|r_n\|_{\phi^*} \leq 2\frac{\beta^*\sqrt{\alpha^*}}{\beta_*\sqrt{\alpha_*}}\sqrt{1-\theta^2}\,|\!|\!| u - u_n |\!|\!|\ , \tag{3.6}$$

as asserted. □

We observe that, as desired, the new error reduction rate

$$\bar{\rho}(\theta) = 2\frac{\beta^*\sqrt{\alpha^*}}{\beta_*\sqrt{\alpha_*}}\sqrt{1-\theta^2} \tag{3.7}$$

can be made arbitrarily small by choosing θ suitably close to 1. This observation was already made in [6, 7], but we improve it in Section 3.2 upon choosing θ dynamically.

Remark 3.2 (Cardinality of $\partial\Lambda_n$). Since we add a ball of radius J around each point of $\widetilde{\partial\Lambda}_n$ we get a crude estimate

$$|\partial\Lambda_n| \le |B_d(0,J)\cap\mathbb{Z}^d|\ |\widetilde{\partial\Lambda}_n| \approx \omega_d J^d |\widetilde{\partial\Lambda}_n|, \tag{3.8}$$

where ω_d is the measure of the d-dimensional Euclidean unit ball $B(0,1)$ centered at the origin.

3.2 Dynamic Dörfler Marking and Adaptive Spectral Algorithm

In this section we improve on the above marking strategy upon making the choice of θ dynamic. At each iteration n let us select the Dörfler parameter θ_n such that

$$\sqrt{1-\theta_n^2} = C_0\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}} \tag{3.9}$$

for a proper choice of the positive constant C_0 that will be made precise later. This implies

$$J(\theta_n) = -\frac{1}{\widetilde{\eta}_L}\log\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}} + K_1 \tag{3.10}$$

according to (3.3), where $K_1 := -\frac{1}{\widetilde{\eta}_L}\log\left(\frac{1}{\sqrt{\alpha_*\alpha^*}}\frac{C_0}{C_{\mathbf{A}^{-1}}}\right) + \delta_n$ and $\delta_n \in [0,1)$.

We thus have the following adaptive spectral Galerkin method with dynamic choice (3.9) of the marking parameter $\theta_n = (1 - C_0^2\|r_n\|_{\phi^*}^2/\|r_0\|_{\phi^*}^2)^{1/2}$:

DYN-GAL(ε)
set $r_0 := f$, $\Lambda_0 := \emptyset$, $n = -1$
do
  $n \leftarrow n+1$
  $\partial\Lambda_n :=$ **E-DÖRFLER** $\left(\Lambda_n, (1 - C_0^2\|r_n\|_{\phi^*}^2/\|r_0\|_{\phi^*}^2)^{1/2}\right)$
  $\Lambda_{n+1} := \Lambda_n \cup \partial\Lambda_n$
  $u_{n+1} :=$ **GAL** (Λ_{n+1})
  $r_{n+1} :=$ **RES** (u_{n+1})
while $\|r_{n+1}\|_{\phi^*} > \varepsilon\ \|r_0\|_{\phi^*}$

where **GAL** computes the Galerkin solution and **RES** the residual. The following result shows the quadratic convergence of **DYN-GAL**.

Theorem 3.3 (quadratic convergence)**.** *Let the constant C_0 of (3.9) satisfy $C_0 \le \frac{1}{4}\sqrt{\frac{\alpha_*}{\alpha^*}}\frac{\beta_*}{\beta^*}$ and $C_1 := \frac{\sqrt{\alpha^*}}{2\beta_*\|f\|_{\phi^*}}$. Then the residual r_n of* **DYN-GAL** *satisfies*

$$\frac{\|r_{n+1}\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \le \left(\frac{\|r_n\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right)^2 \quad \forall n \ge 0, \tag{3.11}$$

and the algorithm terminates in finite steps for any tolerance ε. In addition, two consecutive solutions of **DYN-GAL** *satisfy*

$$|\!|\!| u - u_{n+1} |\!|\!| \le C_1 |\!|\!| u - u_n |\!|\!|^2 \quad \forall n \ge 0. \tag{3.12}$$

Proof. Invoke (3.6) and (3.9) to figure out that

$$\begin{aligned}\frac{\|r_{n+1}\|_{\phi^*}}{\|r_0\|_{\phi^*}} &\le \frac{\sqrt{\alpha^*}}{\beta_*}\frac{|||u-u_{n+1}|||}{\|r_0\|_{\phi^*}} \le 2\sqrt{\frac{\alpha^*}{\alpha_*}}\frac{\beta^*}{\beta_*}\sqrt{1-\theta_n^2}\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}} \\ &\le 2C_0\sqrt{\frac{\alpha^*}{\alpha_*}}\frac{\beta^*}{\beta_*}\left(\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}}\right)^2 \le \frac{1}{2}\left(\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}}\right)^2,\end{aligned} \tag{3.13}$$

which implies (3.11). We thus realize that **DYN-GAL** converges quadratically and terminates in finite steps for any tolerance ε. Finally, combining (2.28) with (3.13), we readily obtain (3.12) upon using $r_0 = f$. □

Remark 3.4 (super-linear rate). If the dynamic marking parameter θ_n is chosen so that $\sqrt{1-\theta_n^2} = C_0\left(\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}}\right)^\sigma$ for some $\sigma > 0$, then we arrive at the rate $|||u-u_{n+1}||| \le C_1|||u-u_n|||^{1+\sigma}$.

It seems to us that the quadratic rate (3.12) is the first one in adaptivity theory. The relation (3.11) reads equivalently

$$\frac{\|r_{n+1}\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \le \left(\frac{\|r_{n+1-k}\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right)^{2^k} \qquad 0 \le k \le n+1, \tag{3.14}$$

and implies that $\|r_n\|_{\phi^*}/\|r_0\|_{\phi^*}$ is within machine precision in about $n = 6$ iterations. This fast decay is consistent with spectral methods. Upon termination, we obtain the relative error

$$|||u-u_n||| \le \frac{\beta^*\sqrt{\alpha^*}}{\beta_*\sqrt{\alpha_*}}|||u||| \ \varepsilon,$$

because $\|f\|_{\phi^*} \le \frac{\sqrt{\alpha^*}}{\beta_*}|||u|||$.

The algorithm **DYN-GAL** entails exact computation of the residual r_n, which in general has infinitely many terms. We do not dwell here with inexact or feasible versions of **DYN-GAL** and refer to [6, 7, 10] for a full discussion which extends to our present setting.

4 Nonlinear Approximation and Gevrey Sparsity Classes

Given any $v \in V$ we define its *best N-term approximation error* as

$$E_N(v) = \inf_{\Lambda\subset\mathcal{K},\ |\Lambda|=N} \|v - P_\Lambda v\|_\phi \ .$$

We are interested in classifying functions v according to the decay law of $E_N(v)$ as $N \to \infty$, i.e., according to the "sparsity" of their expansions in terms of the basis $\{\phi_k\}_{k\in\mathcal{K}}$. Of special interest to us is the following exponential Gevrey class.

Definition 4.1 (exponential class of functions)**.** For $\eta > 0$ and $0 < t \le d$, we denote by $\mathcal{A}_G^{\eta,t}$ the subset of V defined as

$$\mathcal{A}_G^{\eta,t} := \left\{ v \in V \ : \ \|v\|_{\mathcal{A}_G^{\eta,t}} := \sup_{N\ge 0}\left(E_N(v)\exp\left(\eta\omega_d^{-t/d}N^{t/d}\right)\right) < +\infty\right\}$$

where ω_d is the measure of the d-dimensional Euclidean unit ball $B_d(0,1)$ centered at the origin.

Definition 4.2 (exponential class of sequences)**.** Let $\ell_G^{\eta,t}(\mathcal{K})$ be the subset of sequences $\mathbf{v} \in \ell^2(\mathcal{K})$ so that

$$\|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathcal{K})} := \sup_{n\ge 1}\left(n^{(1-t/d)/2}\exp\left(\eta\omega_d^{-t/d}n^{t/d}\right)|v_n^*|\right) < +\infty \ ,$$

where $\mathbf{v}^* = (v_n^*)_{n=1}^\infty$ is the non-increasing rearrangement of $\mathbf{v}$.

The relationship between $\mathcal{A}_G^{\eta,t}$ and $\ell_G^{\eta,t}(\mathcal{K})$ is stated in the following [6, Proposition 4.2].

Proposition 4.3 (equivalence of exponential classes)**.** *Given a function $v \in V$ and the sequence $\mathbf{v} = (\hat{v}_k\sqrt{d_k})_{k\in\mathcal{K}}$ of its coefficients, one has $v \in \mathcal{A}_G^{\eta,t}$ if and only if $\mathbf{v} \in \ell_G^{\eta,t}(\mathcal{K})$, with*

$$\|v\|_{\mathcal{A}_G^{\eta,t}} \lesssim \|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathcal{K})} \lesssim \|v\|_{\mathcal{A}_G^{\eta,t}} .$$

For functions v in $\mathcal{A}_G^{\eta,t}$ one can estimate the minimal cardinality of a set Λ such that $\|v - P_\Lambda v\|_\phi \leq \varepsilon$ as follows: since $\|v - P_{\widetilde{\Lambda}} v\|_\phi > \varepsilon$ for any set $\widetilde{\Lambda}$ with cardinality $|\widetilde{\Lambda}| = |\Lambda| - 1$, we deduce

$$|\Lambda| \leq \omega_d \left(\frac{1}{\eta} \log \frac{\|v\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon} \right)^{d/t} + 1. \tag{4.1}$$

For the analysis of the optimality of our algorithm it is important to investigate the sparsity class of the image Lv for the operator L defined in (2.1), when the function v belongs to the sparsity class $\mathcal{A}_G^{\eta,t}$. Sparsity classes of exponential type for functionals $f \in V^*$ can be defined analogously as above, using now the best N-term approximation error in V^*

$$E_N^*(f) = \inf_{\Lambda\subset\mathcal{K},\ |\Lambda|=N} \|f - P_\Lambda^* f\|_{\phi^*} .$$

The following result is based on [6, Proposition 5.2].

Proposition 4.4 (continuity of L in $\mathcal{A}_G^{\eta,t}$)**.** *Let L be such that the associated stiffness matrix $\mathbf{A}$ satisfies the decay condition* (2.21)*. Given $\eta > 0$ and $t \in (0, d]$, there exist $\bar{\eta} > 0$, $\bar{t} \in (0, t]$ and a constant $C_L \geq 1$ such that*

$$\|Lv\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} \leq C_L \|v\|_{\mathcal{A}_G^{\eta,t}} \qquad \forall v \in \mathcal{A}_G^{\eta,t} . \tag{4.2}$$

Proof. Let $\mathbf{A}$ be the stiffness matrix associated with the operator L. In [6] it is proven that if $\mathbf{A}$ is banded with $2p+1$ non-zero diagonals, then the result holds with $\bar{\eta} = \frac{\eta}{(2p+1)^{t/d}}$ and $\bar{t} = t$; on the other hand, if $\mathbf{A} \in \mathcal{D}_e(\eta_L)$ is dense, but the coefficients η_L and η satisfy the inequality $\eta < \eta_L \omega_d^{t/d}$, then the result holds with $\bar{\eta} = \zeta(t)\eta$ and $\bar{t} = \frac{t}{1+t}$, where $\zeta(t) = \left(\frac{1+t}{2^d\,\omega_d^{1+t}}\right)^{\frac{t}{d(1+t)}}$. Finally, if $\eta \geq \eta_L \omega_d^{t/d}$, we introduce an arbitrary $\hat{\eta} > 0$ satisfying $\hat{\eta} < \eta_L \omega_d^{t/d}$; then the result holds with $\bar{\eta} = \zeta(t)\hat{\eta}$ and $\bar{t} = \frac{t}{1+t}$, since $\|v\|_{\mathcal{A}_G^{\hat{\eta},t}} \leq \|v\|_{\mathcal{A}_G^{\eta,t}}$. □

Keeping into account that $\zeta(t) \leq 1$ for $1 \leq d \leq 10$ (see again [6]), this result indicates that the residual is expected to belong to a less favorable sparsity class than the one of the solution. Counterexamples in [6] show that (4.2) cannot be improved.

Finally, we discuss the sparsity class of the residual $r = r(u_\Lambda)$ for any Galerkin solution u_Λ.

Proposition 4.5 (sparsity class of the residual)**.** *Let $\mathbf{A} \in \mathcal{D}_e(\eta_L)$ and $\mathbf{A}^{-1} \in \mathcal{D}_e(\bar{\eta}_L)$, for constants $\eta_L > 0$ and $\bar{\eta}_L \in (0, \eta_L]$ so that* (2.22) *and* (2.23) *hold. If $u \in \mathcal{A}_G^{\eta,t}$ for some $\eta > 0$ and $t \in (0, d]$, then there exist suitable positive constants $\tilde{\eta} \leq \eta$ and $\tilde{t} \leq t$ such that $r(u_\Lambda) \in \mathcal{A}_G^{\tilde{\eta},\tilde{t}}$ for any index set Λ and*

$$\|r(u_\Lambda)\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} \lesssim \|u\|_{\mathcal{A}_G^{\eta,t}} .$$

Proof. Proposition 4.4 yields the existence of $\bar{\eta} > 0$ and $\bar{t} \in (0, t]$ such that

$$\|f\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} = \|Lu\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} \lesssim \|u\|_{\mathcal{A}_G^{\eta,t}} . \tag{4.3}$$

In order to bound $\|r(u_\Lambda)\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}}$ in terms of $\|f\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}}$, let us write

$$\mathbf{r}_\Lambda = \mathbf{A}(\mathbf{u} - \mathbf{u}_\Lambda) = \mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda ,$$

then use $\mathbf{u}_\Lambda = (\widehat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f})$ from (2.30) to get

$$\mathbf{r}_\Lambda = \mathbf{f} - \mathbf{A}(\widehat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f}).$$

Now, assuming just for simplicity that the indices in Λ come first (this can be realized by a permutation), we have

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_\Lambda & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \qquad \text{and} \qquad \widehat{\mathbf{A}}_\Lambda = \begin{pmatrix} \mathbf{A}_\Lambda & \mathbf{O} \\ \mathbf{O}^T & \mathbf{I} \end{pmatrix}, \qquad \text{whence} \qquad (\widehat{\mathbf{A}}_\Lambda)^{-1} = \begin{pmatrix} (\mathbf{A}_\Lambda)^{-1} & \mathbf{O} \\ \mathbf{O}^T & \mathbf{I} \end{pmatrix}$$

Setting $\mathbf{f} = (\mathbf{f}_\Lambda \;\; \mathbf{f}_{\Lambda^c})^T$, so that $\mathbf{P}_\Lambda \mathbf{f} = (\mathbf{f}_\Lambda \;\; \mathbf{0})^T$, we have

$$\begin{aligned} \mathbf{A}(\widehat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f}) &= \begin{pmatrix} \mathbf{A}_\Lambda & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \begin{pmatrix} (\mathbf{A}_\Lambda)^{-1} & \mathbf{O} \\ \mathbf{O}^T & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{f}_\Lambda \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_\Lambda & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \begin{pmatrix} (\mathbf{A}_\Lambda)^{-1}\mathbf{f}_\Lambda \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_\Lambda \\ \mathbf{B}^T(\mathbf{A}_\Lambda)^{-1}\mathbf{f}_\Lambda \end{pmatrix} \end{aligned}$$

Then,

$$\mathbf{r}_\Lambda = \begin{pmatrix} \mathbf{0} \\ \mathbf{f}_{\Lambda^c} - \mathbf{B}^\mathbf{T}(\mathbf{A}_\Lambda)^{-1}\mathbf{f}_\Lambda \end{pmatrix} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ -\mathbf{B}^T(\mathbf{A}_\Lambda)^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{f}_\Lambda \\ \mathbf{f}_{\Lambda^c} \end{pmatrix} =: \mathbf{R}\mathbf{f} \, .$$

Now, since $\mathbf{A} \in \mathcal{D}_e(\eta_L)$ and $(\mathbf{A}_\Lambda)^{-1} \in \mathcal{D}_e(\bar{\eta}_L)$, it is easily seen that $\mathbf{R} \in \mathcal{D}_e(\tilde{\eta}_L)$ with $\tilde{\eta}_L = \bar{\eta}_L$ if $\bar{\eta}_L < \eta_L$, or $\tilde{\eta}_L < \eta_L$ arbitrary if $\bar{\eta}_L = \eta_L$.

Finally, we apply Proposition 4.4 to the operator R defined by the matrix $\mathbf{R}$, obtaining the existence of constants $\tilde{\eta} > 0$ and $\tilde{t} \in (0, \bar{t}\,]$ such that

$$\|r(u_\Lambda)\|_{\mathcal{A}_G^{\tilde{\eta},\tilde{t}}} \lesssim \|f\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} \, ,$$

whence the result. □

5 Optimality Properties of DYN-GAL

In this section we derive an exponential rate of convergence for $|\!|\!| u - u_n |\!|\!|$ in terms of the number of degrees of freedom Λ_n activated by **DYN-GAL** and assess the computational work necessary to achieve this rate. This is made precise in the following theorem.

Theorem 5.1 (exponential convergence rate)**.** *Let $u \in \mathcal{A}_G^{\eta,t}$ where the Gevrey class $\mathcal{A}_G^{\eta,t}$ is introduced in Definition 4.1. Upon termination of* **DYN-GAL***, the iterate $u_{n+1} \in V_{\Lambda_{n+1}}$ and set of active coefficients Λ_{n+1} satisfy $|\!|\!| u - u_{n+1} |\!|\!| \leq \frac{\beta^*}{\sqrt{\alpha_*}} \|f\|_{\phi^*} \varepsilon$ and*

$$|\Lambda_{n+1}| \leq \omega_d \left(\frac{1}{\eta_*} \log \frac{C_* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon} \right)^{d/t_*} , \tag{5.1}$$

with parameters $C_* > 0$, $\eta_* < \eta$ and $t_* < t$. Moreover, if the number of arithmetic operations needed to solve a linear system scales linearly with its dimension, then the workload $\mathcal{W}_\varepsilon$ of **DYN-GAL** *upon completion satisfies*

$$\mathcal{W}_\varepsilon \leq \omega_d \left(\frac{1}{\eta^*} \log \frac{C^* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon^4 |\log|\log \varepsilon||^{-1}}, \right)^{d/t_*} \tag{5.2}$$

where $\eta^* < \eta_*$ and $C^* > C_*$ but t_* remains the same as in (5.1)*.*

Proof. We proceed in several steps.

1. *Expression of $J(\theta_k)$:* Our first task is to simplify the expression (3.10) for $J(\theta_k)$, namely to absorb the term K_1: there is $C_2 > 1$ such that

$$J(\theta_k) \leq \frac{C_2}{\widetilde{\eta}_L} \Big| \log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \Big|. \tag{5.3}$$

In fact, if C_2 is given by $C_2 = 1 + \frac{\widetilde{\eta}_L \max(0,K_1)}{\log 2}$, then

$$K_1 \leq \frac{C_2 - 1}{\widetilde{\eta}_L} \log 2 \leq \frac{C_2 - 1}{\widetilde{\eta}_L} \left| \log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \right|$$

because $\|r_k\|_{\phi^*} \leq \|r_0\|_{\phi^*}$ for all $k \geq 0$ according to (3.13). This in turn implies (5.3)

$$J(\theta_k) \leq \frac{1}{\widetilde{\eta}_L} \left| \log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \right| + \frac{C_2 - 1}{\widetilde{\eta}_L} \left| \log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \right| = \frac{C_2}{\widetilde{\eta}_L} \left| \log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}} \right|.$$

2. *Active set Λ_k:* We now examine the output Λ_k of **E-DÖRFLER**. Employing the minimality of Dörfler marking and (3.2), we deduce

$$E^*_{|\widetilde{\partial\Lambda_k}|}(r_k) = \|r_k - P^*_{\widetilde{\partial\Lambda_k}} r_k\|_{\phi^*} \leq \sqrt{1-\theta_k^2}\|r_k\|_{\phi^*}, \tag{5.4}$$

which clearly implies $E^*_{|\widetilde{\partial\Lambda_k}|-1} > \sqrt{1-\theta_k^2}\|r_k\|_{\phi^*}$. The latter inequality, together with the Definition 4.2 of $\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}}$, yields

$$\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}} > \sqrt{1-\theta_k^2}\|r_k\|_{\phi^*} \exp\left(\bar\eta \omega_d^{-\bar t/d} (|\widetilde{\partial\Lambda_k}| - 1)^{\bar t/d}\right).$$

We note that this, along with $|\widetilde{\partial\Lambda_k}| \geq 1$, ensures $\frac{\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}}}{\sqrt{1-\theta_k^2}\|r_k\|_{\phi^*}} > 1$ whence

$$|\widetilde{\partial\Lambda_k}| \leq \omega_d \left(\frac{1}{\bar\eta} \log \frac{\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}}}{\sqrt{1-\theta_k^2}\|r_k\|_{\phi^*}} \right)^{d/\bar t} + 1.$$

We now recall the membership of the residual $r_k := r(u_k)$ to the Gevrey class $\mathcal{A}_G^{\bar\eta,\bar t}$ for all $k \geq 0$, established in Proposition 4.5: there exists $C_3 > 0$ independent of k and u such that

$$\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}} \leq C_3 \|u\|_{\mathcal{A}_G^{\eta,t}}.$$

Combining this with the dynamic marking (3.9) implies

$$\frac{\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}}}{\sqrt{1-\theta_k^2}\|r_k\|_{\phi^*}} = \frac{\|f\|_{\phi^*}\|r_k\|_{\mathcal{A}_G^{\bar\eta,\bar t}}}{C_0\|r_k\|_{\phi^*}^2} \leq \frac{C_4\|f\|_{\phi^*}\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}^2},$$

with $C_4 = C_3/C_0$, whence

$$|\widetilde{\partial\Lambda_k}| \leq \omega_d \left(\frac{1}{\bar\eta} \log \frac{C_4\|f\|_{\phi^*}\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}^2} \right)^{d/\bar t} + 1.$$

We let C_5 satisfy $1 = \omega_d \big(\frac{1}{\bar\eta} \log C_5\big)^{d/\bar t}$, and use that $\bar t < t \leq d$, to obtain the simpler expression

$$|\widetilde{\partial\Lambda_k}| \leq \omega_d \left(\frac{1}{\bar\eta} \log \frac{C_4C_5\|f\|_{\phi^*}\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}^2} \right)^{d/\bar t}. \tag{5.5}$$

On the other hand, in view of (5.3), the enrichment step (3.8) of **E-DÖRFLER** yields

$$|\partial\Lambda_k| \le C_6\omega_d J(\theta_k)^d |\widetilde{\partial\Lambda}_k| \le \frac{C_6 C_2^d \omega_d^2}{\widetilde{\eta}_L^d \bar{\eta}^{d/\bar{t}}} \left|\log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right|^d \left(\log \frac{C_4C_5\|f\|_{\phi^*}\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}^2}\right)^{d/\bar{t}}. \tag{5.6}$$

We exploit the quadratic convergence (3.11) to write for $k \le n$

$$\left|\log \frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right|^d \le 2^{d(k-n)} \left|\log \frac{\|r_n\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right|^d = 2^{d(k-n)} \left(\log \frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}}\right)^d. \tag{5.7}$$

Using the bound $\|f\|_{\phi^*} \le \|f\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} = \|r_0\|_{\mathcal{A}_G^{\bar{\eta},\bar{t}}} \le C_3\|u\|_{\mathcal{A}_G^{\eta,t}}$ in the two previous inequalities yields

$$\log \frac{C_4C_5\|f\|_{\phi^*}\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}^2} \le 2\log \frac{C_7\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_k\|_{\phi^*}} \qquad \text{and} \qquad \log \frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}} \le \log \frac{2C_3\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_n\|_{\phi^*}},$$

with $C_7 = (C_3C_4C_5)^{1/2}$. Introducing the constants

$$C_8 = \max(2C_3, C_7), \qquad \frac{1}{\hat{\eta}} = \frac{2^{d/\bar{t}} C_6 C_2^d \omega_d}{\widetilde{\eta}_L^d \bar{\eta}^{d/\bar{t}}}, \qquad t_* = \frac{\bar{t}}{1+\bar{t}},$$

the derived upper bound for $|\partial\Lambda_k|$ can be simplified as follows:

$$|\partial\Lambda_k| \le 2^{d(k-n)} \frac{\omega_d}{\hat{\eta}} \left(\log \frac{C_8\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_n\|_{\phi^*}}\right)^{d/t_*}.$$

Recalling now that $|\Lambda_0| = 0$ and for $n \ge 0$

$$|\Lambda_{n+1}| = \sum_{k=0}^{n} |\partial\Lambda_k|$$

we have

$$|\Lambda_{n+1}| \le \frac{\omega_d}{\hat{\eta}} \left(\sum_{k=0}^{n} 2^{d(k-n)}\right) \left(\log \frac{C_8\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_n\|_{\phi^*}}\right)^{d/t_*}.$$

This can be written equivalently as

$$|\Lambda_{n+1}| \le \omega_d \left(\frac{1}{\eta_*} \log \frac{C_8\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|r_n\|_{\phi^*}}\right)^{d/t_*} \tag{5.8}$$

with $\eta_* = \left(\frac{\hat{\eta}}{\sum_{k=1}^{\infty} 2^{-dk}}\right)^{t_*/d}$. At last, we make use of $\|r_n\|_{\phi^*} > \varepsilon\|f\|_{\phi^*}$ to get the desired estimate (5.1) with $C_* = C_8$.

3. *Computational work:* Let us finally discuss the total computational work $\mathcal{W}_\varepsilon$ of **DYN-GAL**. We start with some useful notation. We set $\delta_n := \frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}}$ for $n \ge 0$ and $\varepsilon_{\ell+1} = \varepsilon_\ell^2$ for $\ell \ge 1$ with $\varepsilon_1 = \frac{1}{2}$. We note that there exists an integer $L > 0$ such that $\varepsilon_{L+1} < \varepsilon \le \varepsilon_L$ with ε being the tolerance of **DYN-GAL**. In addition, we observe that for every iteration $n > 0$ of **DYN-GAL** there exists $\ell > 0$ such that $\delta_n \in I_\ell := (\varepsilon_{\ell+1}, \varepsilon_\ell]$ and that for each interval I_ℓ there exists at most one $\delta_n \in I_\ell$ because $\delta_{n+1} \le \frac{1}{2}\delta_n^2$ according to (3.11). Finally, to each interval I_ℓ we associate the following computational work W_ℓ to find and store u_{n+1}

$$W_\ell = \begin{cases} C_\#|\Lambda_{n+1}| & \text{if there exists } n \text{ such that } \delta_n \in I_\ell \\ 0 & \text{otherwise .} \end{cases}$$

This assumes that the number of arithmetic operations needed to solve the linear system for u_n scales linearly with its dimension and $C_\#$ is an absolute constant that may depend on the specific solver. The total computational work of **DYN-GAL** is bounded by

$$\mathcal{W}_\varepsilon = \sum_{\ell=1}^{L} W_\ell.$$

We now get a bound for $\mathcal{W}_\varepsilon$. In view of (5.1) we have

$$W_\ell \leq C_\# \omega_d \left(\frac{1}{\eta_*} \log \frac{C_* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon_{\ell+1}} \right)^{d/t_*} = C_\# \omega_d \left(\frac{1}{\eta_*} \log \frac{C_* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon_1^{2^\ell}} \right)^{d/t_*}.$$

Therefore, upon adding over ℓ and using that $d/t_* \geq 1$, we obtain

$$\mathcal{W}_\varepsilon \leq \frac{C_\# \omega_d}{\eta_*^{d/t_*}} \left(\sum_{\ell=1}^{L} \log C_* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}} + \sum_{\ell=1}^{L} \log \varepsilon_1^{-2^\ell} \right)^{d/t_*} \leq \frac{C_\# \omega_d}{\eta_*^{d/t_*}} \left(\log \frac{L C_* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon_1^{2^{L+1}}} \right)^{d/t_*}.$$

Since $\varepsilon \leq \varepsilon_L = \varepsilon_1^{2^{L-1}}$, we deduce $\varepsilon^4 \leq \varepsilon_1^{2^{L+1}}$ and $L \leq \frac{\log \frac{|\log \varepsilon|}{\log 2}}{\log 2} + 1 \leq C_9 \log |\log \varepsilon|$. Inserting this bound in the preceding expression yields

$$\mathcal{W}_\varepsilon \leq \omega_d \left(\frac{1}{\eta^*} \log \frac{C^* \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|f\|_{\phi^*}}}{\varepsilon^4 |\log |\log \varepsilon||^{-1}} \right)^{d/t_*}. \tag{5.9}$$

with

$$\eta^* = \frac{\eta_*}{C_\#^{t_*/d}}, \quad C^* = C_9 C_*,$$

which is the asserted estimate (5.2). The proof is thus complete. □

Remark 5.2. Note that the bound on the workload, given in (5.2), is at most an absolute multiple of the bound, given in (5.1), on the number of active coefficients.

Remark 5.3 (super-linear convergence). If $\sqrt{1-\theta_n^2} = C_0 \left(\frac{\|r_n\|_{\phi^*}}{\|r_0\|_{\phi^*}} \right)^\sigma$ with $\sigma > 0$, then (5.1) still holds with the same parameters η_*, t_*.

Remark 5.4 (algebraic class). Let us consider the case when u belongs to the algebraic class

$$\mathcal{A}_B^s := \Big\{ v \in V \ : \ \|v\|_{\mathcal{A}_B^s} := \sup_{N \geq 0} E_N(v) \big(N+1\big)^{s/d} < +\infty \Big\},$$

which is related to Besov regularity. We can distinguish two cases:

1. **A** belongs to an exponential class but the residuals belong to an algebraic class;
2. **A** belongs to an algebraic class $\mathcal{D}_a(\eta_L)$, i.e. there exists a constant $c_L > 0$ such that its elements satisfy $|a_{\ell,k}| \leq c_L (1 + |\ell - k|)^{-\eta_L}$, and the residual belongs to an algebraic class.

We now study the optimality properties of **DYN-GAL** for these two cases. Let us first observe that whenever the residuals belong to the algebraic class $\mathcal{A}_B^s$, the bound (5.5) becomes

$$|\widetilde{\partial \Lambda}_k| \lesssim \|u - u_k\|^{-2d/s}. \tag{5.10}$$

This results from the dynamic marking (3.9) together with (5.4) in the algebraic case.

Let us start with Case 1. Using (5.10) and (5.3), which is still valid here as we assume that **A** belongs to an exponential class, the bound (5.6) is replaced by

$$|\partial\Lambda_k| \lesssim \|u-u_k\|^{-2d/s}\left|\log\frac{\|r_k\|_{\phi^*}}{2\|r_0\|_{\phi^*}}\right|^d \lesssim \|u-u_k\|^{-2d/s}2^{d(k-n)}\left(\log\frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}}\right)^d, \tag{5.11}$$

where in the last inequality we have employed (5.7). Hence, we have

$$\begin{aligned}
|\Lambda_{n+1}| &= \sum_{k=0}^{n}|\partial\Lambda_k| \lesssim \left(\log\frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}}\right)^d \sum_{k=0}^{n}\|u-u_n\|^{-2\frac{d}{s}2^{k-n}}2^{d(k-n)} \\
&\lesssim \left(\log\frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}}\right)^d \|u-u_n\|^{-2\frac{d}{s}}\sum_{k=0}^{n}2^{d(k-n)} \\
&\lesssim \left(\log\frac{2\|f\|_{\phi^*}}{\|r_n\|_{\phi^*}}\right)^d \|u-u_n\|^{-2\frac{d}{s}} \lesssim \left(\log\frac{2\|f\|_{\phi^*}}{\|u-u_n\|}\right)^d \|u-u_{n+1}\|^{-\frac{d}{s}}
\end{aligned}$$

where in the last inequality we have employed the quadratic convergence of **DYN-GAL**. The above result implies that **DYN-GAL** is optimal for Case 1 (up to a logarithmic factor).

Let us now consider Case 2. Since **A** belongs to an algebraic class, (5.3) is replaced by

$$J(\theta_k) \eqsim \|u-u_k\|^{-1/s}.$$

Thus, the bound (5.6) is replaced by

$$|\partial\Lambda_k| \lesssim \|u-u_k\|^{-3d/s},$$

which implies

$$|\Lambda_{n+1}| = \sum_{k=0}^{n}|\partial\Lambda_k| \lesssim \|u-u_n\|^{-3\frac{d}{s}} \lesssim \|u-u_{n+1}\|^{-\frac{3}{2}\frac{d}{s}}$$

where in the last inequality we have again employed the quadratic convergence of **DYN-GAL**. This result is *not* optimal for Case 2, due to the factor 2/3 multiplying s in the exponent. We recall that the algorithm **FA-ADFOUR** of [6] is similar to **DYN-GAL** but with static marking parameter θ. The theory of **FA-ADFOUR** requires neither a restriction on θ nor coarsening and is proven to be optimal in the algebraic case; see [6, Theorem 7.2].

Acknowledgements. The first and the fourth authors are partially supported by the Italian research grant *Prin 2012* 2012HBLYE4 "Metodologie innovative nella modellistica differenziale numerica". The first author is also partially supported by Progetto INdAM-GNCS 2015 "Tecniche di riduzione computazionale per problemi di fluidodinamica e interazione fluido-struttura". The second author is partially supported by NSF grants DMS-1109325 and DMS-1411808. The fourth author is also partially supported by Progetto INdAM - GNCS 2015 "Non-standard numerical methods for geophysics".

References

[1] P. Binev, W. Dahmen, and R. DeVore. *Adaptive finite element methods with convergence rates.* Numer. Math., 97(2):219–268, 2004.

[2] P. Binev and R. DeVore. *Fast computation in adaptive tree approximation.* Numer. Math., 97(2):193–217, 2004.

[3] M. Bürg and W. Dörfler. *Convergence of an adaptive hp finite element strategy in higher space-dimensions.* Appl. Numer. Math., 61(11):1132–1146, 2011.

[4] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral methods. Fundamentals in single domains.* Scientific Computation. Springer-Verlag, Berlin, 2006.

[5] C. Canuto, R. H. Nochetto, R. Stevenson, and M. Verani. *High-order adaptive Galerkin methods.* Proceedings of ICOSAHOM 2014 (to appear).

[6] C. Canuto, R. H. Nochetto, and M. Verani. *Adaptive Fourier-Galerkin methods.* Math. Comp., 83(288):1645–1687, 2014.

[7] C. Canuto, R. H. Nochetto, and M. Verani. *Contraction and optimality properties of adaptive Legendre-Galerkin methods: the one-dimensional case.* Comput. Math. Appl., 67(4):752–770, 2014.

[8] C. Canuto, R. H. Nochetto, R. Stevenson, and M. Verani. *Convergence and Optimality of hp-AFEM.* arXiv:1503.03996.

[9] C. Canuto, V. Simoncini, and M. Verani. *On the decay of the inverse of matrices that are sum of Kronecker products.* Linear Algebra Appl., 452:21–39, 2014.

[10] C. Canuto, V. Simoncini, and M. Verani. *Contraction and optimality properties of an adaptive Legendre-Galerkin method: the multi-dimensional case.* J. Sci. Comput., doi: 10.1007/s10915-014-9912-3.

[11] C. Canuto and M. Verani. *On the numerical analysis of adaptive spectral/hp methods for elliptic problems.* In Analysis and numerics of partial differential equations, volume 4 of Springer INdAM Ser., pages 165–192. Springer, Milan, 2013.

[12] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. *Axioms of adaptivity.* Comput. Math. Appl. 67(6):1195-1253, 2014.

[13] J. M. Cascón, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. *Quasi-optimal convergence rate for an adaptive finite element method.* SIAM J. Numer. Anal., 46(5):2524–2550, 2008.

[14] A. Cohen, W. Dahmen, and R. DeVore. *Adaptive wavelet methods for elliptic operator equations – convergence rates.* Math. Comp, 70:27–75, 1998.

[15] L. Diening, Ch. Kreuzer, and R. Stevenson, *Instance optimality of the adaptive maximum strategy.* Found. Comp. Math., 1–36, 2015.

[16] W. Dörfler and V. Heuveline. *Convergence of an adaptive hp finite element strategy in one space dimension.* Appl. Numer. Math., 57(10):1108–1124, 2007.

[17] G. B. Folland, *Real analysis*, second ed., John Wiley & Sons Inc., New York, 1999.

[18] T. Gantumur, H. Harbrecht, and R. Stevenson. *An optimal adaptive wavelet method without coarsening of the iterands.* Math. Comp., 76(258):615–629, 2007.

[19] P. Morin, R.H. Nochetto, and K. G. Siebert. *Data oscillation and convergence of adaptive FEM.* SIAM J. Numer. Anal., 38(2):466–488 (electronic), 2000.

[20] R. H. Nochetto, K. G. Siebert, and A. Veeser. *Theory of adaptive finite element methods: an introduction.* In Multiscale, nonlinear and adaptive approximation, pages 409–542. Springer, Berlin, 2009.

[21] Ch. Schwab. *p- and hp-finite element methods.* Numerical Mathematics and Scientific Computation. The Clarendon Press, Oxford University Press, New York, 1998. Theory and applications in solid and fluid mechanics.

[22] R. Stevenson. *Optimality of a standard adaptive finite element method.* Found. Comput. Math., 7(2):245–269, 2007.

[23] R. Stevenson. *The completion of locally refined simplicial partitions created by bisection.* Math. Comp., 77:227–241, 2008.

[24] R. Stevenson. *Adaptive wavelet methods for solving operator equations: An overview.* In Multiscale, nonlinear and adaptive approximation, pages 543-597. Springer, Berlin, 2009.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

53/2015 Menafoglio, A; Grujic, O.; Caers, J.
Universal kriging of functional data: trace-variography vs cross-variography? Application to forecasting in unconventional shales

49/2015 Ghiglietti, A.; Ieva, F.; Paganoni, A.M.
Statistical inference for stochastic processes: two sample hypothesis tests

50/2015 Grillo, A.; Guaily, A.; Giverso, C.; Federico, S.
Non-Linear Model for Compression Tests on Articular Cartilage

51/2015 Ballarin, F.; Faggiano, E.; Ippolito, S.; Manzoni, A.; Quarteroni, A.; Rozza, G.; Scrofani, R.
Fast simulations of patient-specific haemodynamics of coronary artery bypass grafts based on a POD–Galerkin method and a vascular shape parametrization

52/2015 Giverso, C.; Scianna, M.; Grillo, A.
Growing Avascular Tumours as Elasto-Plastic Bodies by the Theory of Evolving Natural Configurations

47/2015 Colombo, M. C.; Giverso, C.; Faggiano, E.; Boffano,C.; Acerbi, F.; Ciarletta, P.
Towards the personalized treatment of glioblastoma: integrating patient-specific clinical data in a continuous mechanical model

48/2015 Ambrosi, D.; Pettinati, V.; Ciarletta, P.
Active stress as a local regulator of global size in morphogenesis

46/2015 Giverso, C.; Verani, M.; Ciarletta P.
Emerging morphologies in round bacterial colonies: comparing volumetric versus chemotactic expansion

45/2015 Lange, M.; Palamara, S.; Lassila, T.; Vergara, C.; Quarteroni, A.; Frangi, A.F.
Improved hybrid/GPU algorithm for solving cardiac electrophysiology problems on Purkinje networks

44/2015 Antonietti, P.F.; Houston, P.; Smears, I.
A note on optimal spectral bounds for nonoverlapping domain decomposition preconditioners for hp-Version Discontinuous Galerkin methods