SISSA preprint 20/2019/MATE

CONVERGENCE OF THE CONJUGATE GRADIENT METHOD WITH UNBOUNDED OPERATORS

NOE CARUSO AND ALESSANDRO MICHELANGELI

ABSTRACT. In the framework of inverse linear problems on infinite-dimensional Hilbert space, we prove the convergence of the conjugate gradient iterates to an exact solution to the inverse problem in the most general case where the self-adjoint, non-negative operator is unbounded and with minimal, technically unavoidable assumptions on the initial guess of the iterative algorithm. The convergence is proved to always hold in the Hilbert space norm (error convergence), as well as at other levels of regularity (energy norm, residual, etc.) depending on the regularity of the iterates. We also discuss, both analytically and through a selection of numerical tests, the main features and differences of our convergence result as compared to the case, already available in the literature, where the operator is bounded.

1. Introduction

We are concerned in this work with the rigorous proof of the convergence, in various meaningful senses, of a particular and well-known iterative algorithm for solving inverse linear problems, the celebrated conjugate gradient method, in the generalised setting of *unbounded* operators on Hilbert space.

In abstract terms, given a Hilbert space $\mathcal{H}$ over the (real or) complex field and a non-negative self-adjoint operator A on $\mathcal{H}$, we consider the *inverse linear problem*

$$Af \;=\; g\,, \qquad g \in \mathrm{ran}A \tag{1.1}$$

in the unknown $f \in \mathcal{H}$ with datum g – assuming $g \in \mathrm{ran}A$ makes the problem (1.1) solvable. A may be unbounded, in which case $\mathcal{H}$ has infinite dimension and the domain $\mathcal{D}(A)$ of A is only a dense subspace of $\mathcal{H}$. The positivity assumption on A reads $\langle \psi, A\psi\rangle \geqslant 0$ for all $\psi \in \mathcal{D}(A)$: here and in the following $\langle\,\cdot\,,\,\cdot\rangle$ is the scalar product in $\mathcal{H}$: if $\mathcal{H}$ is taken over the complex field, then $\langle\,\cdot\,,\,\cdot\rangle$ is assumed to be anti-linear in the first entry and linear in the second, and $\|\cdot\|$ is the corresponding norm.

This setting generalises the classical, *finite-dimensional* one where $\mathcal{H} = \mathbb{C}^d$ for some $d \in \mathbb{N}$ and A is a $d \times d$ positive semi-definite matrix (in which case (1.1) can be interpreted as a system of d linear equations), as well as the setting where A is a bounded (and everywhere-defined) self-adjoint operator on an *infinite-dimensional* Hilbert space.

Infinite dimensionality is also the framework where the phenomenon of ill-posedness may occur. Indeed, it is a standard fact that for a (not necessarily bounded) self-adjoint operator A on an infinite-dimensional Hilbert space $\mathcal{H}$ the properties

(i) the point 0 belongs to $\sigma(A)$ and is not isolated in $\sigma(A)$,
(ii) $\mathrm{ran}A$ is not closed,
(iii) on $\mathrm{ran}A$ the operator A has unbounded inverse,

Date: August 27, 2019.
Key words and phrases. inverse linear problems, infinite-dimensional Hilbert space, ill-posed problems, Krylov subspaces methods, conjugate gradient, self-adjoint operators, spectral measure, orthogonal polynomials.

are all equivalent (and could not occur if $\dim \mathcal{H} < +\infty$): when any such property is satisfied, the solution f fails to depend continuously on the datum g, as is evident from (iii), and the problem (1.1) is said to be *ill-posed*.

As opposite to that, if any among (i), (ii), and (iii) above fails to hold and in addition A is injective, and hence equivalently if A has an everywhere-defined bounded inverse, the problem (1.1) is *well-posed*: in this case the solution f exists, is unique, and depends continuously on the datum g.

A popular algorithm for the numerical solution to (1.1) in the above-mentioned classical framework is the method of conjugate gradients (also referred to as CG). It was first proposed in 1952 by Hestenes and Stiefel [9] and since then, together with its related derivatives (e.g., conjugate gradient method on the normal equations (CGNE), least-square QR method (LSQR), etc.), it has been widely studied in the finite-dimensional setting (see the monographs [18, 21, 12]) and also, though to a lesser extent, in the infinite-dimensional Hilbert space setting with bounded operators.

In order to describe the algorithm explicitly, let us introduce the solution manifold

$$\mathcal{S} \;:=\; \{f \in \mathcal{D}(A)\,|\,Af = g\} \tag{1.2}$$

relative to the problem (1.1). By assumption $\mathcal{S}$ is a convex, non-empty set in $\mathcal{H}$ which is also closed, owing to the fact that A, being self-adjoint, is in particular a closed operator. As a consequence, the projection map $P_{\mathcal{S}} : \mathcal{H} \to \mathcal{S}$ is unambiguously defined and produces, for generic $x \in \mathcal{H}$, the closest to x point in $\mathcal{S}$.

In its iterative implementation, the conjugate gradient algorithm starts with an initial guess $f^{[0]} \in \mathcal{H}$ and produces iterates $f^{[N]}$ according to a prescription that can be described in various equivalent ways [18, 12], the most convenient of which for our purposes is

$$f^{[N]} \;:=\; \underset{h\in\{f^{[0]}\}+\mathcal{K}_N(A,\mathfrak{R}_0)}{\arg\min} \|A^{1/2}(h - P_{\mathcal{S}}f^{[0]})\|\,, \qquad N \in \mathbb{N}\,. \tag{1.3}$$

More generally, we shall discuss conjugate gradient style algorithms with iterates given by

$$f^{[N]} \;:=\; \underset{h\in\{f^{[0]}\}+\mathcal{K}_N(A,\mathfrak{R}_0)}{\arg\min} \|A^{\theta/2}(h - P_{\mathcal{S}}f^{[0]})\|\,, \qquad N \in \mathbb{N} \tag{1.4}$$

for some parameter $\theta \geqslant 0$ (the case $\theta = 1$ being *the* conjugate gradient method). It will be then convenient to refer to such $f^{[N]}$'s as the θ-iterates.

In (1.3)-(1.4) the vector $\mathfrak{R}_0$ is the zero-th order of the *residuals* defined by

$$\mathfrak{R}_N \;:=\; Af^{[N]} - g\,, \qquad N \in \mathbb{N}_0 \tag{1.5}$$

in terms of each iterate, and the vector space

$$\mathcal{K}_N(A,\mathfrak{R}_0) \;:=\; \mathrm{span}\{\mathfrak{R}_0, A\mathfrak{R}_0, \dots A^{N-1}\mathfrak{R}_0\}\,, \qquad N \in \mathbb{N} \tag{1.6}$$

is the *N-th order Krylov subspace* associated to A and $\mathfrak{R}_0$.

For (1.6) and hence (1.3)-(1.4) to make sense for any N when A is *unbounded*, additional technical assumptions are needed in order to avoid possible domain issues. We shall discuss them in the general set-up of the problem presented in Section 2. Clearly, the above definitions are all well-posed if A is bounded.

As well known, for *finite-dimensional* inverse problems CG is an extremely popular, versatile, and efficient numerical scheme – it belongs, in particular, to the class of Krylov subspace methods, that are sometimes even counted among the 'Top 10 Algorithms' of the 20th century [6, 3] – and the convergence of $f^{[N]}$ to the exact solution f is by now a classical and deeply understood theory (see, e.g., the monographs [18, 12]).

The convergence theory of CG has been markedly less explored in the setting of *infinite-dimensional* $\mathcal{H}$, a line of investigation in which yet important works have been produced over the last five decades, both in the scenario where A is bounded with everywhere-defined bounded inverse [4, 5, 8], or at least with bounded inverse on its range [10], and in the scenario where A is bounded with possible unbounded inverse on its range [10, 14, 15, 13, 1].

In contrast, the scenario where A is *unbounded* has been only recently considered from special perspectives, in particular in view of existence [16] (for GMRES algorithms), or convergence when A is regularised and made invertible with everywhere-defined bounded inverse [7], whereas the general convergence theory (that is, including the case where (1.1) is ill-posed) is virtually unexplored.

In the present work we establish a class of convergence results for the conjugate gradient algorithm precisely in the most general setting where A is unbounded and the associated inverse problem (1.1) is possibly ill-posed. Our analysis consists of a non-trivial generalisation of the very subtle approach by Nemirovskiy and Polyak from their above-mentioned 1984 work [14] for bounded A.

In such work the full convergence estimates of the error $\|f^{[N]} - f\|$ and residual $\|Af^{[N]} - g\|$ were proved, and in the follow-up work [15] the results were shown to be *optimal* in the sense that for the entire class of bounded, ill-posed problems, one can do no better than the estimates provided. The boundedness of A was crucial in a two-fold way. First, it forced the blow-up of a suitable sum (δ_N, in their notation – see (3.21) below) of the reciprocals of the N zeroes of a polynomial that represents the minimisation (1.4): since no such zero can exceed $\|A\|$, the reciprocals cannot vanish and their sum necessarily diverges. As a consequence, the error and the residual, which in turn can be controlled by an inverse power of δ_N, are then shown to vanish as $N \to \infty$, thus establishing convergence. Second, boundedness of A was also determinant to quantify the convergence, as the latter was boiled down to a min-max procedure for polynomials on the *finite* spectral interval containing $\sigma(A)$, then on such interval (suitable modifications of) the Chebyshev polynomials are recognised to optimise the rate of convergence, and explicit properties of (the zeroes of) Chebyshev polynomials finally provide a quantitative version of the vanishing of error and residual.

In our approach we are able to bypass the restriction of the finiteness of $\|A\|$ as far as the convergence alone is concerned. As for the quantitative rate, the min-max strategy of [14] by no means can be adapted to polynomials over the whole $[0, +\infty)$ and in fact a careful analysis of the structure of the proof of [14, 15], as we shall comment in due time, seem to indicate that if A is unbounded with unbounded inverse on its range, then the convergence rate can be arbitrarily small.

The discussion is organised as follows. In Section 2 we introduce the rigorous set-up of the convergence problem for the conjugate gradient method and we state and comment our main result. In Section 3 we develop an amount of preparatory materials of algebraic and measure-theoretic nature, which are needed to finally prove our main theorem in Section 4. Last, in Section 5 we discuss a selection of numerical tests that confirm the main features of our convergence result and corroborate our intuition on certain relevant differences with respect to the bounded case.

2. Set-up and main results

Let us start with the rigorous formulation of all the notions needed for our convergence result. Here and in the following A is a non-negative, densely defined, self-adjoint operator on a Hilbert space $\mathcal{H}$, including the possibility that A be unbounded and with a non-trivial kernel.

First, one needs to ensure that the conjugate gradient iterates are well-defined. As mentioned in the Introduction, one chooses a datum $g \in \mathrm{ran}A$ and an initial guess $f^{[0]} \in \mathcal{H}$, and for some $\theta \geqslant 0$ defines the θ-iterates

$$f^{[N]} := \underset{h\in\{f^{[0]}\}+\mathcal{K}_N(A,\mathfrak{R}_0)}{\arg\min} \|A^{\theta/2}(h-P_{\mathcal{S}}h)\|\,, \qquad N\in\mathbb{N} \tag{2.1}$$

with

$$\mathfrak{R}_N = Af^{[N]}-g\,, \qquad N\in\mathbb{N}_0\,, \tag{2.2}$$

$$\mathcal{K}_N(A,\mathfrak{R}_0) = \mathrm{span}\{\mathfrak{R}_0, A\mathfrak{R}_0, \dots A^{N-1}\mathfrak{R}_0\}\,, \qquad N\in\mathbb{N}\,. \tag{2.3}$$

In order to apply an arbitrary positive power of A to $Af^{[0]}-g$, we require that both g and $f^{[0]}$ be A-smooth vectors [17, Sect. X.6], meaning that they belong to the space

$$C^\infty(A) := \bigcap_{N\in\mathbb{N}} \mathcal{D}(A^N)\,. \tag{2.4}$$

In the applications where A is a differential operator, A-smoothness is a regularity requirement.

In turn, A-smoothness of g and $f^{[0]}$ implies $\mathcal{K}_N(A,\mathfrak{R}_0)\subset C^\infty(A)$, and obviously $P_{\mathcal{S}}h \in \mathcal{S} \subset C^\infty(A)$, whereas by interpolation $C^\infty(A)\subset\mathcal{D}(A^{\theta/2})$ for any $\theta\geqslant 0$. This guarantees that in the minimisation (2.1) one is allowed to apply $A^{\theta/2}$ to any vector $h-P_{\mathcal{S}}h$.

We have thus seen that under the assumptions

$$g \in \mathrm{ran}A\cap C^\infty(A)\,, \qquad f^{[0]} \in C^\infty(A) \tag{2.5}$$

the corresponding θ-iterates $f^{[N]}$ are unambiguously defined by (2.1)-(2.3) above for *any* $\theta\geqslant 0$. If A is bounded, (2.5) simply reduces to $g\in\mathrm{ran}A$.

Such iterates have three notable properties, whose proof is deferred to Section 3.

Proposition 2.1. *The θ-iterates $f^{[N]}$ defined for a given $\theta\geqslant 0$ by means of* (2.1)-(2.3) *under the assumption* (2.5) *satisfy*

$$f^{[N]}-P_{\mathcal{S}}f^{[N]} \in (\ker A)^\perp \qquad \forall N\in\mathbb{N}_0\,, \tag{2.6}$$

$$P_{\mathcal{S}}f^{[N]} = P_{\mathcal{S}}f^{[0]} \qquad \forall N\in\mathbb{N}\,, \tag{2.7}$$

$$f^{[N]}-P_{\mathcal{S}}f^{[N]} = p_N(A)(f^{[0]}-P_{\mathcal{S}}f^{[0]}) \qquad \forall N\in\mathbb{N}\,, \tag{2.8}$$

where $p_N(\lambda)$ is for each N a polynomial of degree up to N and such that $p_N(0)=1$.

As, by (2.7), all such $f^{[N]}$'s have the same projection onto the solution manifold $\mathcal{S}$, the approach of $f^{[N]}$ to $\mathcal{S}$ consists explicitly of a convergence $f^{[N]}\to P_{\mathcal{S}}f^{[0]}$. Let us now specify in which sense this convergence is to be monitored.

The underlying idea, as is clear in the typical applications where A is a differential operator on a L^2-space, is that $\|f^{[N]}-P_{\mathcal{S}}f^{[0]}\|_{(A)}\to 0$ in some A-dependent Sobolev norm. For this to make sense, clearly one needs enough A-regularity on $f^{[N]}-P_{\mathcal{S}}f^{[0]}$, which eventually is guaranteed by the regularity initially assumed on $f^{[0]}$. Thus, the general indicator of convergence has the form $\|A^{\sigma/2}(f^{[N]}-P_{\mathcal{S}}f^{[0]})\|$, but an extra care is needed if one wants to control the convergence in the abstract analogue of a low-regularity, negative-Sobolev norm, which would amount to formally consider $\sigma<0$, for in general A can have a kernel and hence is only invertible on its range.

Based on these considerations, and inspired by the analogous discussion in [14] for bounded A, let us introduce the class $\mathscr{C}_{A,g}(\theta)$ defined for generic $\theta \in \mathbb{R}$ as

$$\mathscr{C}_{A,g}(\theta) \;:=\; \begin{cases} \{x \in \mathcal{H} \,|\, x - P_{\mathcal{S}}x \in \mathcal{D}(A^{\frac{\theta}{2}})\}\,, & \theta \geqslant 0 \\ \{x \in \mathcal{H} \,|\, x - P_{\mathcal{S}}x \in \mathrm{ran}(A^{-\frac{\theta}{2}})\}\,, & \theta < 0\,. \end{cases} \tag{2.9}$$

(The dependence of $\mathscr{C}_{A,g}(\theta)$ on g is implicit through the solution manifold $\mathcal{S}$.) Distinguishing the two cases in (2.9) is needed whenever A has a non-trivial kernel. If instead A is injective, and so too is therefore $A^{-\frac{\theta}{2}}$ for $\theta < 0$, then $A^{-\frac{\theta}{2}}$ is a bijection between the two dense subspaces $\mathcal{D}(A^{-\frac{\theta}{2}}) = \mathrm{ran}(A^{\frac{\theta}{2}})$ and $\mathrm{ran}(A^{-\frac{\theta}{2}}) = \mathcal{D}(A^{\frac{\theta}{2}})$ of $\mathcal{H}$.

Related to the class $\mathscr{C}_{A,g}(\theta)$ we have two further useful notions. One, for fixed $\theta \in \mathbb{R}$ and $x \in \mathscr{C}_{A,g}(\theta)$, is the vector

$$u_\theta(x) \;:=\; \begin{cases} A^{\frac{\theta}{2}}(x - P_{\mathcal{S}}x)\,, & \theta \geqslant 0 \\ \text{the minimal norm solution } u \text{ to } A^{-\frac{\theta}{2}}u = x - P_{\mathcal{S}}x\,, & \theta < 0\,. \end{cases} \tag{2.10}$$

The other is the functional ρ_θ defined on the vectors $x \in \mathscr{C}_{A,g}(\theta)$ as

$$\rho_\theta(x) \;:=\; \|u_\theta(x)\|^2\,. \tag{2.11}$$

Thus,

$$\rho_\theta(x) \;=\; \begin{cases} \|A^{\frac{\theta}{2}}(x - P_{\mathcal{S}}x)\|^2\,, & \theta \geqslant 0\,, \\ \Big\| \Big(A^{-\frac{\theta}{2}}\Big|_{\mathrm{ran}\left(A^{-\frac{\theta}{2}}\right)}\Big)^{-1}(x - P_{\mathcal{S}}x)\Big\|^2\,, & \theta < 0\,, \end{cases} \tag{2.12}$$

with an innocent abuse of notation in (2.12) when $\theta < 0$, as the operator inverse is to be understood for a (self-adjoint, and positive-definite) operator on the Hilbert subspace $\overline{\mathrm{ran}A}$.

It is worth remarking that in the special case where A is bounded, the following interesting properties hold, whose proof is deferred to Section 3, which do not have a counterpart in the unbounded case except for the obvious identity $\mathscr{C}_{A,g}(0) = \mathcal{H}$.

Lemma 2.2. *If A (besides being self-adjoint and non-negative) is bounded, and if $g \in \mathrm{ran}A$, then:*

(i) $\mathscr{C}_{A,g}(\theta) = \mathcal{H}$ *whenever* $\theta \geqslant 0$*;*
(ii) $\mathscr{C}_{A,g}(\theta) \subset \mathscr{C}_{A,g}(\theta')$ *for* $\theta \leqslant \theta'$*;*
(iii) *for* $\theta \leqslant \theta'$ *and* $x \in \mathscr{C}_{A,g}(\theta)$ *one has* $u_{\theta'}(x) = A^{(\theta'-\theta)/2}u_\theta(x)$*, whence also* $\rho_{\theta'}(x) \leqslant \|A\|^{\theta'-\theta}\rho_\theta(x)$.

Back to the general case where A is unbounded, the goal is to evaluate certain ρ_σ-functionals along the sequence of the $f^{[N]}$'s. This may require an extra assumption on the initial guess $f^{[0]}$, as the following Lemma shows.

Lemma 2.3. *Consider the θ-iterates $f^{[N]}$ defined for a given $\theta \geqslant 0$ by means of* (2.1)-(2.3) *under the assumption* (2.5)*. Then:*

(i) $f^{[N]} \in \mathscr{C}_{A,g}(\sigma)$ $\forall \sigma \geqslant 0$*;*
(ii) $f^{[N]} \in \mathscr{C}_{A,g}(\sigma)$ *for any* $\sigma < 0$ *such that, additionally,* $f^{[0]} \in \mathscr{C}_{A,g}(\sigma)$*, in which case*

$$u_\sigma(f^{[N]}) \;=\; p_N(A)\,u_\sigma(f^{[0]})\,, \tag{2.13}$$

where $p_N(\lambda)$ is precisely the polynomial mentioned in Proposition 2.1.

Lemma 2.3 is a direct consequence of (2.8) in Proposition 2.1 above: for completeness we include its proof in Section 3.

It then makes sense to control the convergence $f^{[N]} \to \mathcal{P}_S f^{[0]}$ in the ρ_σ-sense, for σ positive or negative, with suitable assumptions on $f^{[0]}$. Explicitly,

$$\tag{2.14} \begin{aligned} \rho_\sigma(f^{[N]}) \;&=\; \|u_\sigma(f^{[N]})\|^2 \\ &= \begin{cases} \|A^{\frac{\sigma}{2}}(f^{[N]} - P_{\mathcal{S}}f^{[0]})\|^2\,, & \sigma \geqslant 0\,, \\ \Big\|\Big(A^{-\frac{\sigma}{2}}\Big|_{\mathrm{ran}\left(A^{-\frac{\sigma}{2}}\right)}\Big)^{-1}(f^{[N]} - P_{\mathcal{S}}f^{[0]})\Big\|^2\,, & \sigma < 0\,, \end{cases} \end{aligned}$$

having used (2.7). The most typical and meaningful choices in the applications are

$$\tag{2.15} \begin{aligned} \rho_0(f^{[N]}) \;&=\; \big\|f^{[N]} - \mathcal{P}_S f^{[0]}\big\|^2 \\ \rho_1(f^{[N]}) \;&=\; \big\langle f^{[N]} - \mathcal{P}_S f^{[0]}, A\,(f^{[N]} - \mathcal{P}_S f^{[0]})\big\rangle \\ \rho_2(f^{[N]}) \;&=\; \big\|A(f^{[N]} - \mathcal{P}_S f^{[0]})\big\|^2\,, \end{aligned}$$

that is, respectively, the norm of the error, the so-called 'energy' (semi-)norm, and the norm of the residual.

We are finally in the condition to formulate our main result.

Theorem 2.4. *Let A be a non-negative self-adjoint operator on the Hilbert space $\mathcal{H}$ and let*

$$g \;\in\; \mathrm{ran}A \cap C^\infty(A) = \mathrm{ran}A \cap \bigcap_{N\in\mathbb{N}} \mathcal{D}(A^N)\,.$$

Consider the conjugate gradient algorithm associated with A and g where the initial guess vector $f^{[0]}$ satisfy

$$f^{[0]} \;\in\; C^\infty(A) \cap \mathscr{C}_{A,g}(\sigma^*)\,, \qquad \sigma^* = \min\{\sigma, 0\}$$

for a given $\sigma \in \mathbb{R}$, and where the iterates $f^{[N]}$, $N \in \mathbb{N}$, are constructed via (2.1) *with parameter $\theta = \xi \geqslant 0$ under the condition $\sigma \leqslant \xi$. Then*

$$\lim_{N\to\infty} \rho_\sigma(f^{[N]}) \;=\; 0\,.$$

In other words, the convergence holds at a given 'A-regularity level' σ for ξ-iterates built with *equal or higher* 'A-regularity level' $\xi \geqslant \sigma$, and with an initial guess $f^{[0]}$ that is A-smooth if $\sigma \geqslant 0$, and additionally belongs to the class $\mathscr{C}_{A,g}(\sigma)$ if $\sigma < 0$.

In particular, with no extra assumption on $f^{[0]}$ but its A-smoothness, the ξ-iterates with $\xi \geqslant 0$ automatically converge in the sense of the error ($\sigma = 0$, see (2.15) above), the ξ-iterates with $\xi \geqslant 1$ automatically converge in the sense of the error and of the energy norm ($\sigma = 1$), the ξ-iterates with $\xi \geqslant 2$ automatically converge in the sense of the error, energy norm, and residual ($\sigma = 2$).

Remark 2.5. If, for a finite N, $\rho_\sigma(f^{(N)}) = 0$, then the very iterate $f^{(N)}$ *is* a solution to the linear problem $Af = g$, and one says that the algorithm 'has come to convergence' in a finite number (N) of steps. Indeed, $\rho_\sigma(f^{(N)}) = 0$ is the same as $A^{\frac{\sigma}{2}}(f^{[N]} - P_{\mathcal{S}}f^{[0]}) = 0$ if $\sigma \geqslant 0$, i.e., $f^{[N]} - P_{\mathcal{S}}f^{[0]} \in \ker A^{\frac{\sigma}{2}} = \ker A$; this, combined with $f^{[N]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp$ (see (2.6)-(2.7) above), implies that $f^{[N]} = P_{\mathcal{S}}f^{[0]} \in \mathcal{S}$. On the other hand, $\rho_\sigma(f^{(N)}) = 0$ is the same as $u_\sigma(f^{[N]}) = 0$ with $A^{-\frac{\sigma}{2}}u_\sigma(f^{[N]}) = f^{[N]} - P_{\mathcal{S}}f^{[0]}$ if $\sigma < 0$, whence again $f^{[N]} = P_{\mathcal{S}}f^{[0]} \in \mathcal{S}$.

Remark 2.6.

(i) In the special scenario where A is (everywhere-defined and) bounded, A-smoothness is automatically guaranteed, so one only needs to assume that $g \in \mathrm{ran}A$ and $f^{[0]} \in \mathscr{C}_{A,g}(\sigma^*)$ for some $\sigma \in \mathbb{R}$ ($\sigma^* = \min\{\sigma, 0\}$) in order for the convergence of the ξ-iterates ($\xi \geqslant \sigma$) to hold in the sense $\rho_\sigma(f^{[N]}) \to 0$. Then, owing to Lemma 2.2, one automatically has also $\rho_{\sigma'}(f^{[N]}) \to 0$ for any $\sigma' \geqslant \sigma$. This is precisely the form of the convergence result originally established by Nemirovskiy and Polyak [14].

(ii) Thus, in the bounded-case scenario, if σ is the minimum level of convergence chosen, then not only are the ξ-iterates with $\xi \geqslant \sigma$ proved to ρ_σ-converge, but in addition the *same* ξ-iterates also $\rho_{\sigma'}$-converge at any other level $\sigma' \geqslant \sigma$, with no upper bound on σ'. In particular, it is shown in [14] that

$$\rho_{\sigma'}(f^{[N]}) \leqslant C(\sigma'-\sigma)\,(2N+1)^{-2(\sigma'-\sigma)}\,\rho_\sigma(f^{[0]})\,, \qquad \sigma < \sigma' \leqslant \xi\,, \tag{2.16}$$

for some constant $C(\theta)$, thus providing an explicit *rate of convergence* of the ξ-iterates in a generic $\rho_{\sigma'}$-sense such that $\sigma' \in (\sigma, \xi]$.

(iii) In the general unbounded-case scenario, instead, the ρ_σ-convergence guaranteed by Theorem 2.4 is not exportable to $\rho_{\sigma'}$-convergence with $\sigma' > \sigma$.

Remark 2.7. When, in the unbounded case, A has an everywhere-defined bounded inverse, one has $\mathscr{C}_{A,g}(\sigma) = \mathcal{H}$ for any $\sigma \leqslant 0$. Therefore, Theorem 2.4 guarantees the ρ_σ-convergence of the ξ-iterates for any $\sigma \leqslant 0$, provided that g and $f^{[0]}$ are A-smooth. Such 'weaker' convergence can be still informative in many contexts. For instance, choosing

$$\begin{aligned} \mathcal{H} &= L^2(\mathbb{R}^d) \\ A &= -\Delta + \mathbb{1} \quad \text{with} \quad \mathcal{D}(A) = H^2(\mathbb{R}^d) \qquad (\mathrm{ran} A = \mathcal{H}) \\ g, f^{[0]} &\in C^\infty(\mathbb{R}^d)\,, \end{aligned}$$

we see that the θ-iterates defined by (1.4) with the above data converge to the unique solution f to the inverse problem $-\Delta f + f = g$ in any negative Sobolev space $H^\sigma(\mathbb{R}^d)$, $\sigma < 0$; in particular, $f^{[N]}(x) \to f(x)$ point-wise almost everywhere.

3. Intermediate technical facts

We discuss in this Section an amount of technical properties that are needed for the proof of the main Theorem 2.4.

For convenience, let us set for each $N \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}([0,+\infty]); &:= \{\text{ polynomials } p(\lambda),\ \lambda \in [0,+\infty)\ \} \\ \mathbb{P}_N &:= \{\, p \in \mathbb{P}([0,+\infty]) \,|\, \deg p \leqslant N \,\} \\ \mathbb{P}_N^{(1)} &:= \{\, p \in \mathbb{P}_N \,|\, p(0) = 1 \,\}\,. \end{aligned} \tag{3.1}$$

Let us start with the proof of those stated in Section 2. The proof of Proposition 2.1 requires the following elementary property.

Lemma 3.1. *Let $z \in \mathcal{H}$. For a point $y \in \mathcal{S}$ these conditions are equivalent:*

(i) $y = P_{\mathcal{S}} z$,

(ii) $z - y \in (\ker A)^\perp$.

Proof. By linearity of A, $\mathcal{S} = \{y\} + \ker A$. If $z - y \in (\ker A)^\perp$, then for any $x \in \ker A$, and hence for a generic point $y + x \in \mathcal{S}$, one has

$$\|z - (y + x)\|^2 = \|z - y\|^2 + \|x\|^2 \geqslant \|z - y\|^2\,,$$

therefore y is necessarily the closest to z among all points in $\mathcal{S}$, i.e., $y = P_{\mathcal{S}} z$. This proves that (ii) $\Rightarrow$ (i). Conversely, if $y = P_{\mathcal{S}} z$, and if by contradiction $z - y$ does not belong to $(\ker A)^\perp$, then $\langle x_0, z - y\rangle > 0$ for some $x_0 \in \ker A$. In this case, let us consider the polynomial

$$p(t) := \|z - y - t x_0\|^2 = \|x_0\|^2 t^2 - 2\,\langle x_0, z - y\rangle\, t + \|z - y\|^2\,.$$

Clearly, $t = 0$ is not a point of minimum for $p(t)$, as for $t > 0$ and small enough one has $p(t) \leqslant p(0)$. This shows that there are points $y + t x_0 \in \mathcal{S}$ for which

$\|z-(y+tx_0)\| \leqslant \|z-y\|$, thus contradicting the assumption that y is the closest to z among all points in $\mathcal{S}$. Then necessarily $z-y \in (\ker A)^\perp$, which proves that (i) $\Rightarrow$ (ii). □

Proof of Proposition 2.1. In the minimisation (2.1)

$$h - f^{[0]} \;=\; q_{N-1}(A)(Af^{[0]}-g) \;=\; q_{N-1}(A)\,A\,(f^{[0]}-P_{\mathcal{S}}f^{[0]})$$

for some polynomial $q_{N-1} \in \mathbb{P}_{N-1}$, whence also

$$h - P_{\mathcal{S}}f^{[0]} \;=\; q_{N-1}(A)\,A\,(f^{[0]}-P_{\mathcal{S}}f^{[0]}) + (f^{[0]}-P_{\mathcal{S}}f^{[0]})\,.$$

This implies, upon setting $p_N(\lambda) := \lambda\, q_{N-1}(\lambda)+1$, that

$$(*) \qquad f^{[N]} - P_{\mathcal{S}}f^{[0]} \;=\; p_N(A)(f^{[0]}-P_{\mathcal{S}}f^{[0]}) \qquad \forall N \in \mathbb{N}\,,$$

where $p_N \in \mathbb{P}_N^{(1)}$.

Moreover, $f^{[N]} - P_{\mathcal{S}}f^{[N]} \in (\ker A)^\perp$, as a consequence of Lemma 3.1 applied to the choice $z = f^{[N]}$ and $y = P_{\mathcal{S}}f^{[N]}$. With an analogous argument, also $f^{[0]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp$. Thus, (2.6) is proved.

Owing to (2.5) and (2.6), $f^{[0]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp \cap C^\infty(A)$. Now, $(\ker A)^\perp \cap C^\infty(A)$ is invariant under the action of polynomials of A, and therefore owing to (*) we deduce that $f^{[N]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp$.

Next, let us split

$$P_{\mathcal{S}}f^{[N]} - P_{\mathcal{S}}f^{[0]} \;=\; (f^{[N]} - P_{\mathcal{S}}f^{[0]}) - (f^{[N]} - P_{\mathcal{S}}f^{[N]})\,.$$

Obviously, $P_{\mathcal{S}}f^{[N]} - P_{\mathcal{S}}f^{[0]} \in \ker A$. But in the right-hand side, as just shown, both $f^{[N]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp$ and $f^{[N]} - P_{\mathcal{S}}f^{[N]} \in (\ker A)^\perp$. So $P_{\mathcal{S}}f^{[N]} - P_{\mathcal{S}}f^{[0]} \in (\ker A)^\perp$. The conclusion is necessarily $P_{\mathcal{S}}f^{[N]} - P_{\mathcal{S}}f^{[0]} = 0$.

This establishes (2.7), by means of which formula (*) above takes also the form of (2.8). □

Let us now prove Lemmas 2.2 and 2.3.

Proof of Lemma 2.2. Part (i) is evident from the fact that $\mathcal{D}(A^{\frac{\theta}{2}}) = \mathcal{H}$ for any $\theta \geqslant 0$, as A is (everywhere-defined and) bounded and non-negative.

Part (ii) is therefore obvious if $\theta' \geqslant 0$. If, instead, $\theta \leqslant \theta' < 0$, then $\mathrm{ran}(A^{-\frac{\theta}{2}}) \subset \mathrm{ran}(A^{-\frac{\theta'}{2}})$, owing again to the boundedness and non-negativity of A, so part (ii) is actually valid in general.

If $0 \leqslant \theta \leqslant \theta'$, then

$$u_{\theta'}(x) \;=\; A^{\theta'/2}(x - P_{\mathcal{S}}x) \;=\; A^{(\theta'-\theta)/2}A^{\theta/2}(x-P_{\mathcal{S}}x) \;=\; A^{(\theta'-\theta)/2}u_\theta(x)\,.$$

If instead $\theta < 0 \leqslant \theta'$, then $u_{\theta'}(x) = A^{\theta'/2}(x-P_{\mathcal{S}}x)$ and $A^{-\theta/2}u_\theta(x) = x - P_{\mathcal{S}}x$, whence

$$A^{(\theta'-\theta)/2}u_\theta(x) \;=\; A^{\theta'/2}(x-P_{\mathcal{S}}x) \;=\; u_{\theta'}(x)\,.$$

Last, if $\theta \leqslant \theta' < 0$, then $A^{-\xi/2}u_\xi(x) = x - P_{\mathcal{S}}x$ for both $\xi = \theta$ and $\xi = \theta'$, therefore from

$$x - P_{\mathcal{S}}x \;=\; A^{-\theta/2}u_\theta(x) \;=\; A^{-\theta'/2}A^{(\theta'-\theta)/2}u_\theta(x) \quad \text{and} \quad A^{-\theta'/2}u_{\theta'}(x) \;=\; x - P_{\mathcal{S}}x$$

one deduces that $u_{\theta'}(x) = A^{(\theta'-\theta)/2}u_\theta(x)$. In all possible cases such an identity is therefore proved. The inequality $\rho_{\theta'}(x) \leqslant \|A\|^{\theta'-\theta}\rho_\theta(x)$ then follows at once from (2.11). This completes the proof of part (iii). □

Proof of Lemma 2.3. Owing to (2.8) and to the A-smoothness of g and $f^{[0]}$, $f^{[N]} - P_{\mathcal{S}}f^{[N]} \in C^\infty(A)$, which by interpolation means in particular that $f^{[N]} - P_{\mathcal{S}}f^{[N]} \in \mathcal{D}(A^{\frac{\sigma}{2}})$ $\forall \sigma \geqslant 0$. This proves part (i) of the Lemma.

Assume now that $f^{[0]} \in \mathscr{C}_{A,g}(\sigma)$ for some $\sigma < 0$. In this case (2.8) reads

$$f^{[N]} - P_{\mathcal{S}} f^{[N]} \;=\; p_N(A)(f^{[0]} - P_{\mathcal{S}} f^{[0]}) \;=\; p_N(A)\, A^{-\frac{\sigma}{2}} u_\sigma(f^{[0]})\,,$$

thanks to the definition (2.10) of $u_\sigma(f^{[0]})$. Therefore $f^{[N]} - P_{\mathcal{S}} f^{[N]} \in \mathrm{ran}(A^{-\frac{\sigma}{2}})$ and, again by (2.10), $u_\sigma(f^{[N]}) = p_N(A)\, u_\sigma(f^{[0]})$. This proves part (ii). □

Next, let us establish an amount of important results that are measure-theoretic in nature. To this aim, with customary notation [19], let E^A denotes the projection-valued measure associated with the self-adjoint operator A, and let $\mathrm{d}\langle x, E^A(\lambda)x\rangle$ denotes the corresponding scalar measure associated with a vector $x \in \mathcal{H}$. Such measures are supported on $\sigma(A) \subset [0, +\infty)$.

A special role is going to be played by the measure

$$\mathrm{d}\mu_\sigma(\lambda) \;:=\; \mathrm{d}\langle u_\sigma(f^{[0]}), E^A(\lambda) u_\sigma(f^{[0]})\rangle \tag{3.2}$$

defined under the assumption that $f^{[0]} \in \mathscr{C}_{A,g}(\sigma)$ for a given $\sigma \in \mathbb{R}$. Clearly, by definition, μ_σ is a *finite* measure with

$$\int_{[0,+\infty)} \mathrm{d}\mu_\sigma(\lambda) \;=\; \|u_\sigma(f^{[0]})\|^2\,. \tag{3.3}$$

Two relevant properties of μ_σ are the following.

Proposition 3.2. *For the given self-adjoint and non-negative operator A on $\mathcal{H}$, and for given $g \in C^\infty(A)$, $\sigma \in \mathbb{R}$, $f^{[0]} \in C^\infty(A) \cap \mathscr{C}_{A,g}(\sigma)$, consider the measure μ_σ defined by* (3.2)*. Then:*

(i) *one has*

$$\mathrm{d}\mu_\sigma(\lambda) \;=\; \lambda^\sigma\, \mathrm{d}\langle f^{[0]} - P_{\mathcal{S}} f^{[0]}, E^A(\lambda)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\rangle\,; \tag{3.4}$$

(ii) *the spectral value $\lambda = 0$ is not an atom for μ_σ, i.e.,*

$$\mu_\sigma(\{0\}) \;=\; 0\,. \tag{3.5}$$

Proof. The identity (3.4) when $\sigma \geqslant 0$ follows immediately from the definition (3.2) of $\mathrm{d}\mu_\sigma$ and from the definition (2.10) of $u_\sigma(f^{[0]}) = A^{\frac{\sigma}{2}}(f^{[0]} - P_{\mathcal{S}} f^{[0]})$, owing to the property

$$\mathrm{d}\langle A^\alpha \psi, E^A(\lambda) A^\alpha \psi\rangle \;=\; \lambda^{2\alpha} \mathrm{d}\langle \psi, E^A(\lambda)\psi\rangle\,, \qquad \alpha \geqslant 0\,, \qquad \psi \in \mathcal{D}(A^\alpha)\,.$$

If instead $\sigma < 0$, let us consider the auxiliary measures

$$\mathrm{d}\widetilde{\mu}_\sigma(\lambda) \;:=\; \lambda^{-\sigma} \mathrm{d}\mu_\sigma(\lambda)\,, \qquad \mathrm{d}\widehat{\mu}_\sigma(\lambda) \;:=\; \mathrm{d}\langle f^{[0]} - P_{\mathcal{S}} f^{[0]}, E^A(\lambda)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\rangle\,.$$

On an arbitrary Borel subset $\Omega \subset [0, +\infty)$ one then has

$$\begin{aligned} \widetilde{\mu}_\sigma(\Omega) \;&=\; \int_\Omega \lambda^{-\sigma} \mathrm{d}\mu_\sigma(\lambda) \;=\; \|E^A(\Omega) A^{-\frac{\sigma}{2}} u_\sigma(f^{[0]})\|^2 \\ &=\; \|E(\Omega)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\|^2 \;=\; \int_\Omega \mathrm{d}\widehat{\mu}_\sigma(\lambda) \;=\; \widehat{\mu}_\sigma(\Omega)\,, \end{aligned}$$

having used the definition (2.10) in the form $A^{-\frac{\sigma}{2}} u_\sigma(f^{[0]}) = f^{[0]} - P_{\mathcal{S}} f^{[0]}$. This shows that $\mathrm{d}\widetilde{\mu}_\sigma(\lambda) = \mathrm{d}\widehat{\mu}_\sigma(\lambda)$, whence again (3.4). Part (i) is proved.

Concerning part (ii), let us recall from (2.6) that $f^{[0]} - P_{\mathcal{S}} f^{[0]} \in (\ker A)^\perp$. Therefore, $\widehat{\mu}_\sigma(\{0\}) = 0$. Thus, (3.4) implies that also $\mu_\sigma(\{0\}) = 0$. □

In turn, Proposition 3.2 allows us to discuss one further set of technical ingredients for the proof of Theorem 2.4. They concern the polynomial p_N, in the expression (2.8) of the ξ-iterates $f^{[N]}$, that corresponds to the actual minimisation (2.1).

Proposition 3.3. *For the given self-adjoint and non-negative operator A on $\mathcal{H}$, and for given $g \in C^\infty(A)$, $\sigma \in \mathbb{R}$, $f^{[0]} \in C^\infty(A) \cap \mathscr{C}_{A,g}(\sigma)$, and $\xi \geqslant 0$ let $f^{[N]}$ be the N-th ξ-iterate defined by (2.1) with initial guess $f^{[0]}$ and parameter $\theta = \xi$, and let*

$$s_N \;:=\; \underset{p_N \in \mathbb{P}_N^{(1)}}{\arg\min} \int_{[0,+\infty)} \lambda^\xi \, p_N^2(\lambda) \, \mathrm{d}\langle f^{[0]} - P_{\mathcal{S}} f^{[0]}, E^A(\lambda)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\rangle \tag{3.6}$$

for each $N \in \mathbb{N}$., Then the following properties hold.

(i) *One has*

$$f^{[N]} - P_{\mathcal{S}} f^{[N]} \;=\; s_N(A)(f^{[0]} - P_{\mathcal{S}} f^{[0]}) \qquad \forall N \in \mathbb{N}\,. \tag{3.7}$$

(ii) *The family $(s_N)_{N\in\mathbb{N}}$ is a set of orthogonal polynomials on $[0,+\infty)$ with respect to the measure*

$$\begin{aligned} \mathrm{d}\nu_\xi(\lambda) \;&:=\; \lambda^{\xi-\sigma+1}\, \mathrm{d}\mu_\sigma(\lambda) \\ &=\; \lambda^{\xi+1}\, \mathrm{d}\langle f^{[0]} - P_{\mathcal{S}} f^{[0]}, E^A(\lambda)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\rangle \end{aligned} \tag{3.8}$$

and satisfying

$$\deg s_N \;=\; N\,, \qquad s_N(0) \;=\; 1 \qquad \forall N \in \mathbb{N}\,. \tag{3.9}$$

(iii) *One has*

$$\rho_\sigma(f^{[N]}) \;=\; \int_{[0,+\infty)} s_N^2(\lambda)\, \mathrm{d}\mu_\sigma(\lambda) \qquad \forall N \in \mathbb{N}\,. \tag{3.10}$$

Proof. Denote temporarily by $\widetilde{s}_N \in \mathbb{P}_N^{(1)}$ the polynomial that qualifies the iterate $f^{[N]}$ in (2.8) by means of the minimisation (2.1) with $\theta = \xi$. Then

$$\begin{aligned} \min_{h \in \{f^{[0]}\} + \mathcal{K}_N(A, \mathfrak{R}_0)} \|A^{\xi/2}(h - P_{\mathcal{S}} h)\|^2 \;&=\; \|A^{\xi/2}(f^{[N]} - P_{\mathcal{S}} f^{[N]})\|^2 \\ &=\; \|A^{\xi/2} \widetilde{s}_N(A)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\|^2 \\ &=\; \int_{[0,+\infty)} \lambda^\xi\, \widetilde{s}_N^{\,2}(\lambda)\, \mathrm{d}\langle f^{[0]} - P_{\mathcal{S}} f^{[0]}, E^A(\lambda)(f^{[0]} - P_{\mathcal{S}} f^{[0]})\rangle\,. \end{aligned}$$

Comparing the above identity with (3.6) we see that $\widetilde{s}_N$ must be precisely the polynomial s_N. Therefore, (2.8) takes the form (3.7). This proves part (i).

By means of (3.4) we may re-write (3.6) as

$$s_N \;=\; \underset{p_N \in \mathbb{P}_N^{(1)}}{\arg\min} \int_{[0,+\infty)} \lambda^{\xi-\sigma}\, p_N^2(\lambda) \mathrm{d}\mu_\sigma(\lambda)\,.$$

The latter minimising property of s_N implies

$$\begin{aligned} 0 \;&=\; \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \int_{[0,+\infty)} \lambda^{\xi-\sigma}\, (s_N(\lambda) + \varepsilon\lambda\, q_{N-1}(\lambda))^2\, \mathrm{d}\mu_\sigma(\lambda) \\ &=\; 2\int_{[0,+\infty)} \lambda^{\xi-\sigma+1}\, s_N(\lambda)\, q_{N-1}(\lambda)\, \mathrm{d}\mu_\sigma(\lambda) \end{aligned}$$

for any $q_{N-1} \in \mathbb{P}_{N-1}$ (indeed, $s_N + \varepsilon\lambda\, q_{N-1} \in \mathbb{P}_N^{(1)}$). Equivalently, owing to (3.8),

$$\int_{[0,+\infty)} s_N(\lambda)\, q_{N-1}(\lambda)\, \mathrm{d}\nu_\xi(\lambda) \;=\; 0 \qquad \forall q_{N-1} \in \mathbb{P}_{N-1}\,.$$

Such a condition is valid for each $N \in \mathbb{N}$ and, as well known [22, 2, 11], this amounts to saying that $(s_N)_{N\in\mathbb{N}}$ is a set of orthogonal polynomials on $[0,+\infty)$ with respect to the measure $\mathrm{d}\nu_\xi$. The fact that $s_N(0) = 1$ was already demonstrated in Proposition 2.1. Part (ii) is thus proved.

If $\sigma \geqslant 0$, then (2.7), (2.14), (3.4), and (3.7) yield

$$\begin{aligned}\rho_\sigma(f^{[N]}) \;&=\; \big\|A^{\frac{\sigma}{2}}(f^{[N]}-P_{\mathcal{S}}f^{[N]})\big\|^2 \;=\; \big\|A^{\frac{\sigma}{2}}s_N(A)(f^{[0]}-P_{\mathcal{S}}f^{[0]})\big\|^2 \\ &=\; \int_{[0,+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda)\,.\end{aligned}$$

If instead $\sigma < 0$, then owing to (3.7) the identity (2.13) reads

$$u_\sigma(f^{[N]}) \;=\; s_N(A)\,u_\sigma(f^{[0]})\,.$$

The latter identity, together with (2.14) and (3.2), yield

$$\rho_\sigma(f^{[N]}) \;=\; \big\|u_\sigma(f^{[N]})\big\|^2 \;=\; \big\|s_N(A)\,u_\sigma(f^{[0]})\big\|^2 \;=\; \int_{[0,+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda)\,.$$

In either case (3.10) is established. This proves part (iii). $\square$

Remark 3.4. The measure ν_ξ too is finite, with

$$\int_{[0,+\infty)} \mathrm{d}\nu_\xi \;=\; \big\|A^{\frac{\xi+1}{2}}(f^{[0]}-P_{\mathcal{S}}f^{[0]})\big\|^2\,, \tag{3.11}$$

as is evident from (3.8). In fact, one could define ν_ξ for arbitrary $\xi \geqslant -1$: we keep the restriction to $\xi \geqslant 0$ because ξ here is the parameter $\theta = \xi$ required in the definition (2.1) of the ξ-iterates, and as such must therefore be non-negative.

Remark 3.5. There is an implicit dependence on ξ in each s_N, as is clear from (3.6), analogously to the fact that the iterates $f^{[N]}$'s depend on the choice of the parameter ξ. We simply omit such a dependence from the notation s_N.

We thus see from Proposition 3.3(iii) that the control of the convergence of the $f^{[N]}$'s in the ρ_σ-sense is boiled down to monitoring a precise spectral integral. For an efficient estimate of the right-hand side of (3.10) we shall make use of properties of the polynomials s_N and of the measure ν_ξ that we are going to discuss in the remaining part of this Section.

Here and for later reference it is convenient to denote by $(\widehat{s}_N)_{N\in\mathbb{N}}$ the *monic* system of polynomials corresponding to $(s_N)_{N\in\mathbb{N}}$, i.e.,

$$\widehat{s}_N(\lambda) \;:=\; \Big(\frac{1}{N!}\,\frac{\mathrm{d}^N}{\mathrm{d}\lambda^N}\Big|_{\lambda=0} s_N(\lambda)\Big)^{-1} s_N(\lambda)\,. \tag{3.12}$$

Thus, each $\widehat{s}_N$ is the multiple of s_N with unit coefficient of the λ^N-power.

Proposition 3.6. *Consider the set $(s_N)_{N\in\mathbb{N}}$ of orthogonal polynomials on $[0,+\infty)$ with respect to the measure ν_ξ, as defined in* (3.6) *and* (3.8) *under the assumptions of Proposition 3.3.*

(i) *For each $N\in\mathbb{N}$, either $s_N(\lambda)=0$ ν_ξ-almost everywhere, or s_N has exactly N simple zeroes, all located in $(0,+\infty)$.*

Assume now the s_N's are all non-vanishing with respect to the ν_ξ-measure, and denote by $\lambda_k^{(N)}$ the k-th zero of s_N, ordering the zeros as

$$0<\lambda_1^{(N)} \;<\; \lambda_2^{(N)} \;<\;\cdots\;<\; \lambda_N^{(N)}\,. \tag{3.13}$$

(ii) *(Separation.) One has*

$$\lambda_k^{(N+1)} \;<\; \lambda_k^{(N)} \;<\; \lambda_{k+1}^{(N+1)} \qquad \forall k\in\{1,2,\dots,N-1\}\,, \tag{3.14}$$

that is, the zeroes of s_N and s_{N+1} mutually separate each other.

(iii) *(Monotonicity.) For each integer* $k \geqslant 1$,

$$\begin{aligned} &(\lambda_k^{(N)})_{N=k}^\infty \quad \textit{is a decresing sequence,} \\ &(\lambda_{N-k+1}^{(N)})_{N=k}^\infty \quad \textit{is an incresing sequence.} \end{aligned} \tag{3.15}$$

In particular, the limits

$$\lambda_1 \;:=\; \lim_{N\to\infty} \lambda_1^{(N)}\,, \qquad \lambda_\infty \;:=\; \lim_{N\to\infty} \lambda_N^{(N)} \tag{3.16}$$

exist in $[0,+\infty)\cup\{+\infty\}$.

(iv) *(Representation.) The measure* ν_ξ *is actually supported only on the so-called 'true interval of orthogonality'* $[\lambda_1,\lambda_\infty]$, *and* λ_1 *is not an atom for* ν_ξ, *namely*

$$\nu_\xi(\{\lambda_1\}) \;=\; 0\,. \tag{3.17}$$

Here and in the following, the symbol $[\lambda_1,\lambda_\infty]$ *is understood as the closure of* $(\lambda_1,+\infty)$.

(v) *(Orthogonality.) One has*

$$\int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda_1^{(N)}-\lambda}\,\mathrm{d}\nu_\xi(\lambda) \;=\; \int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda-\lambda_1^{(N)}}\,\mathrm{d}\nu_\xi(\lambda) \tag{3.18}$$

for any $N\in\mathbb{N}$.

Proof. Part (i) is standard from the theory of orthogonal polynomials (see, e.g., [22, Theorem 3.3.1] or [2, Theorem 5.2]), owing to the fact that the map

$$\mathbb{P}([0,+\infty])\ni p\;\longmapsto\;\int_{[0,+\infty)} p(\lambda)\,\mathrm{d}\nu_\xi(\lambda)$$

is a positive-definite functional on $\mathbb{P}([0,+\infty])$.

Part (ii) is another standard fact in the theory of orthogonal polynomials (see, e.g., [22, Theorem 3.3.2] or [2, Theorem I.5.3]). Part (iii), in turn, is an immediate corollary of part (ii).

For Part (iv) let us first recall [2, Definition I.5.2] that the true interval of orthogonality $[\lambda_1,\lambda_\infty]$ is the smallest closed interval containing all the zeroes $\lambda_k^{(N)}$, and moreover [2, Theorem II.3.1] there exists a measure η on $[0,+\infty)$ supported only on $[\lambda_1,\lambda_\infty]$ such that the s_N's remain orthogonal with respect to η too and

$$\mu_k \;:=\; \int_{[0,+\infty)} \lambda^k\,\mathrm{d}\nu_\xi(\lambda) \;=\; \int_{[\lambda_1,\lambda_\infty]} \lambda^k\,\mathrm{d}\eta(\lambda)\,, \qquad \forall k\in\mathbb{N}_0\,.$$

The η-measure is actually a Stieltjes measure associated with a bounded, non-decreasing function ψ obtained as point-wise limit of a subsequence of $(\psi_N)_{N\in\mathbb{N}}$, where

$$\psi_N(\lambda) \;:=\; \begin{cases} 0\,, & \lambda<\lambda_1^{(N)}\,,\\ A_1^{(N)}+\cdots+A_p^{(N)}\,, & \lambda\in[\lambda_p^{(N)},\lambda_{p+1}^{(N)}) \quad \text{for} \quad p\in\{1,\dots,n-1\}\,,\\ \mu_0\,, & \lambda\geqslant\lambda_N^{(N)} \end{cases}$$

and $A_1^{(n)},\dots,A_N^{(n)}$ are positive numbers determined by the Gauss quadrature formula

$$\mu_k \;=\; \sum_{p=1}^N A_p^{(N)}(\lambda_p^{(N)})^k\,, \qquad \forall k\in\{0,1,\dots,2N-1\}\,.$$

Therefore,

$$\eta(\{\lambda_1\}) \;=\; \psi(\lambda_1)-\lim_{\lambda\to\lambda_1^-}\psi(\lambda) \;=\; 0\,,$$

because by part (ii) and (iii) $\lambda_1 < \lambda_1^N$ $\forall N \in \mathbb{N}$, whence $\psi(\lambda_1) = \lim_{N\to\infty} \psi_N(\lambda_1) = 0$ and $\psi(\lambda) = \lim_{N\to\infty} \psi_N(\lambda) = 0$ for $\lambda < \lambda_1$.

Next, one sees that $\nu_\xi = \eta$, i.e., that the Hamburger moment problem that guarantees that $(s_N)_{N\in\mathbb{N}}$ is an orthogonal system on $[0,+\infty)$ is *uniquely* solved with the measure ν_ξ. In fact, this follows from the classical criterion [20, Theorem 2.9] for the uniqueness of the orthogonality measure (we refer to [11, Theorem 8.3] for a more modern discussion): such a measure is unique if and only if $\omega(z) = 0$ for some $z \in \mathbb{C}$, where

$$\omega(z) \;:=\; \Big(\sum_{N\in\mathbb{N}} |\widehat{s}_N(z)|^2\Big)^{-1}$$

and $(\widehat{s}_N)_{N\in\mathbb{N}}$ is the monic system obtained from $(s_N)_{N\in\mathbb{N}}$ (see (3.12) above). This is precisely the case, as $\omega(-1) = 0$: indeed, owing to (3.12) and (3.13), $\widehat{s}_N(\lambda) = \prod_{k=1}^N(\lambda - \lambda_k^{(N)})$, whence

$$\widehat{s}_N^{\,2}(-1) \;=\; \prod_{k=1}^N (-1-\lambda_k^{(N)})^2 \;>\; 1\,.$$

This shows that $\nu_\xi = \eta$, thus proving that ν_ξ is supported only on $[\lambda_1, \lambda_\infty]$ with $\nu_\xi(\{\lambda_1\}) = 0$.

Part (v) follows from the identity

$$\int_{[0,+\infty)} s_N(\lambda)\, q_{N-1}(\lambda)\, \mathrm{d}\nu_\xi(\lambda) \;=\; 0 \qquad \forall q_{N-1} \in \mathbb{P}_{N-1}$$

(already considered in the proof of Proposition 3.3, as a consequence of the orthogonality of the s_N's), when the explicit choice

$$q_{N-1}(\lambda) \;:=\; \frac{\lambda_1^{(N)}\, s_N(\lambda)}{\lambda_1^{(N)} - \lambda}$$

is made. □

Remark 3.7. Analogously to what already observed in Remark 3.5, there is an implicit dependence on ξ of all the zeroes $\lambda_k^{[N]}$. For a more compact notation, such a dependence is omitted.

In view of Proposition 3.6(i), when the s_N's are not identically zero we can explicitly represent

$$s_N(\lambda) \;=\; \prod_{k=1}^N \Big(1 - \frac{\lambda}{\lambda_k^{(N)}}\Big)\,, \qquad \widehat{s}_N(\lambda) \;=\; \prod_{k=1}^N (\lambda - \lambda_k^{(N)})\,. \tag{3.19}$$

The integral (3.18) is going to play a central role in the main proof, so the next technical result we need is the following efficient estimate of such a quantity.

Lemma 3.8. *Consider the set $(s_N)_{N\in\mathbb{N}}$ of orthogonal polynomials on $[0,+\infty)$ with respect to the measure ν_ξ, as defined in (3.6) and (3.8) under the assumptions of Proposition 3.3 and with the further restriction $\xi - \sigma + 1 \geqslant 0$. Assume that the s_N's are non-zero polynomials with respect to the measure ν_ξ. Then, for any $N \in \mathbb{N}$,*

$$\int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\, \frac{\lambda_1^{(N)}}{\lambda_1^{(N)} - \lambda}\, \mathrm{d}\nu_\xi(\lambda) \;\leqslant\; \mu_\sigma([0,\lambda_1^{(N)}))\, \Big(\frac{\xi-\sigma+1}{\delta_N}\Big)^{\xi-\sigma+1}, \tag{3.20}$$

where

$$\delta_N \;:=\; \frac{1}{\lambda_1^{(N)}} + 2\sum_{k=2}^N \frac{1}{\lambda_k^{(N)}}\,. \tag{3.21}$$

Remark 3.9. Estimate (3.20) provides a (ξ,σ)-dependent bound on a quantity that is ξ-dependent only. This is only possible for a constrained range of σ, namely $\sigma \leqslant \xi+1$.

Proof of Lemma 3.8. For each $N\in\mathbb{N}$ let us consider the function

$$[0,\lambda_1^{(N)}] \ni \lambda \longmapsto a_N(\lambda) := \frac{\lambda_1^{(N)}\lambda^{\xi-\sigma+1}s_N^2(\lambda)}{\lambda_1^{(N)}-\lambda} = \lambda^{\xi-\sigma+1}\Big(1-\frac{\lambda}{\lambda_1^{(N)}}\Big)\prod_{k=2}^{N}\Big(1-\frac{\lambda}{\lambda_k^{(N)}}\Big)^2$$

(where we used the representation (3.19) for s_N), which is non-negative, smooth, and such that $a_N(0)=a_N(\lambda_1^{(N)})=0$. Let $\lambda_N^*\in(0,\lambda_1^{(N)})$ be the point of maximum for h_N. Then $a_N'(\lambda_N^*)=0$, which after straightforward computations yields

$$\xi-\sigma+1 \geqslant \lambda_N^*\Big(\frac{1}{\lambda_1^{(N)}}+2\sum_{k=2}^{N}\frac{1}{\lambda_k^{(N)}}\Big) = \lambda_N^*\delta_N\,,$$

whence also

$$\lambda_N^* \leqslant \frac{\xi-\sigma+1}{\delta_N}\,.$$

Moreover, $0\leqslant 1-\lambda/\lambda_k^{(N)}\leqslant 1$ for $\lambda\in[0,\lambda_1^{(N)}]$ and for all $k\in\{1,\dots,N\}$, as $\lambda_1^{(N)}$ is the smallest zero of s_N. Therefore,

$$a_N(\lambda) \leqslant a_N(\lambda_N^*) \leqslant (\lambda_N^*)^{\xi-\sigma+1} \leqslant \Big(\frac{\xi-\sigma+1}{\delta_N}\Big)^{\xi-\sigma+1}, \qquad \lambda\in[0,\lambda_1^{(N)}]\,.$$

We then conclude

$$\int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda_1^{(N)}-\lambda}\,\mathrm{d}\nu_\xi(\lambda) = \int_{[0,\lambda_1^{(N)})} a_N(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) \leqslant \mu_\sigma([0,\lambda_1^{(N)}))\,\Big(\frac{\xi-\sigma+1}{\delta_N}\Big)^{\xi-\sigma+1},$$

which completes the proof. □

4. Proof of Theorem 2.4

Let us present in this Section the proof of our main statement, Theorem 2.4, based on the intermediate results established in the previous Section.

Owing to Proposition 3.3, we have to control the behaviour for large N of the quantity

$$\rho_\sigma(f^{[N]}) = \int_{[0,+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda)\,.$$

Obviously, in the following we assume that none of the polynomials s_N vanish with respect to the measure ν_ξ previously introduced in (3.8), for otherwise for some N one would have $\rho_\sigma(f^{[N]})=0$ and therefore $f^{[N]}=P_{\mathcal{S}}f^{[0]}\in\mathcal{S}$ (see Remark 2.5, or also (3.7)), meaning that the conjugate gradient algorithm has come to convergence in a finite number of steps. The conclusion of Theorem 2.4 would then be trivially true.

Let us first observe, from the relation (3.8) between the measures μ_σ and ν_ξ and from the fact that the latter is supported on the true interval of orthogonality $[\lambda_1,\lambda_\infty]$ with no atom at $\lambda=\lambda_1$ (Proposition 3.6(iv)), that the measure μ_σ too is supported on such an interval and $\mu_\sigma(\{\lambda_1\})=0$. Thus, in practice,

$$\rho_\sigma(f^{[N]}) = \int_{[\lambda_1,\lambda_\infty]} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda)\,.$$

(Let us recall that $[\lambda_1, \lambda_\infty]$ is a short-cut for the closure of $(\lambda_1, \lambda_\infty)$, even when $\lambda_\infty = \infty$.)

First of all, it is convenient to split

$$\begin{aligned}\int_{[0,+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) &= \int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) + \int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) \\ &\leqslant \mu_\sigma([0,\lambda_1^{(N)})) + \int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda)\,.\end{aligned} \tag{4.1}$$

Here we used the bound $s_N^2(\lambda) \leqslant 1$, $\lambda \in [0, \lambda_1^{(N)})$, that is obvious from (3.19).

Next, let us show that

$$\int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) \leqslant \frac{1}{(\lambda_1^{(N)})^{\xi-\sigma+1}} \int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda_1^{(N)}-\lambda}\,\mathrm{d}\nu_\xi(\lambda)\,. \tag{4.2}$$

In fact, (4.2) is a consequence of the properties of s_N discussed in Section 3. To see that, let us consider the inequality

$$\begin{aligned}1 \leqslant \Big(\frac{\lambda}{\lambda_1^{(N)}}\Big)^{\xi-\sigma} &= \frac{1}{(\lambda_1^{(N)})^{\xi-\sigma+1}}\cdot\frac{\lambda_1^{(N)}}{\lambda}\cdot\lambda^{\xi-\sigma+1} \\ &\leqslant \frac{1}{(\lambda_1^{(N)})^{\xi-\sigma+1}}\cdot\frac{\lambda_1^{(N)}}{\lambda-\lambda_1^{(N)}}\cdot\lambda^{\xi-\sigma+1} \qquad (\lambda \geqslant \lambda_1^{(N)})\,,\end{aligned} \tag{4.3}$$

which is valid owing to the constraint $\xi - \sigma \geqslant 0$. Then,

$$\begin{aligned}\int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) &\leqslant \frac{1}{(\lambda_1^{(N)})^{\xi-\sigma+1}} \int_{[\lambda_1^{(N)},+\infty)} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda-\lambda_1^{(N)}}\,\mathrm{d}\nu_\xi(\lambda) \\ &= \frac{1}{(\lambda_1^{(N)})^{\xi-\sigma+1}} \int_{[0,\lambda_1^{(N)})} s_N^2(\lambda)\,\frac{\lambda_1^{(N)}}{\lambda_1^{(N)}-\lambda}\,\mathrm{d}\nu_\xi(\lambda)\,,\end{aligned}$$

having used (3.8) and (4.3) in the first step, and the orthogonality property (3.18) in the second. Estimate (4.2) is thus proved.

Now, plugging (4.2) into (4.1) and applying Lemma 3.8, yields

$$\begin{aligned}\rho_\sigma(f^{[N]}) &= \int_{[0,+\infty)} s_N^2(\lambda)\,\mathrm{d}\mu_\sigma(\lambda) \\ &\leqslant \mu_\sigma([0,\lambda_1^{(N)})) + \frac{\mu_\sigma([0,\lambda_1^{(N)}))}{(\lambda_1^{(N)})^{\xi-\sigma+1}}\Big(\frac{\xi-\sigma+1}{\delta_N}\Big)^{\xi-\sigma+1}\,.\end{aligned} \tag{4.4}$$

The second summand in the right-hand side of (4.4) is estimated as

$$\begin{aligned}\frac{\mu_\sigma([0,\lambda_1^{(N)}))}{(\lambda_1^{(N)})^{\xi-\sigma+1}}\Big(\frac{\xi-\sigma+1}{\delta_N}\Big)^{\xi-\sigma+1} &= (\xi-\sigma+1)^{\xi-\sigma+1}\,\frac{\mu_\sigma([0,\lambda_1^{(N)}))}{(\lambda_1^{(N)}\delta_N)^{\xi-\sigma+1}} \\ &\leqslant (\xi-\sigma+1)^{\xi-\sigma+1}\,\mu_\sigma([0,\lambda_1^{(N)}))\,,\end{aligned} \tag{4.5}$$

because $\lambda_1^{(N)}\delta_N \geqslant 1$ (as is seen from (3.21)) and the exponent $\xi - \sigma + 1$ is positive. Thus,

$$\rho_\sigma(f^{[N]}) \leqslant \big(1 + (\xi-\sigma+1)^{\xi-\sigma+1}\big)\,\mu_\sigma([0,\lambda_1^{(N)}))\,. \tag{4.6}$$

Let us recall once again that $\mu_\sigma([0,\lambda_1^{(N)})) = \mu_\sigma([\lambda_1,\lambda_1^{(N)}))$.

Now,

$$\mu_\sigma([0,\lambda_1^{(N)})) \xrightarrow{N\to\infty} 0 \tag{4.7}$$

because $\lambda_1^{(N)} \downarrow \lambda_1$ and $\mu_\sigma([0,\lambda_1)) = 0$. This covers also the case when $\lambda_1 = 0$, as $\mu_\sigma(\{0\}) = 0$ (see (3.5) in Proposition 3.2(ii)).

Plugging (4.7) into (4.6) finally shows that $\rho_\sigma(f^{(N)}) \to 0$, thus completing the proof of Theorem 2.4.

Remark 4.1. In retrospect, the assumption $\xi \geqslant \sigma$ was necessary to establish the bound (4.2) – more precisely, the inequality (4.3). In the other steps, namely in (4.4) (which is an application of Lemma 3.8) and (4.5), only the less restrictive assumption $\xi \geqslant \sigma - 1$ was needed.

Remark 4.2. Where exactly the true interval of orthogonality lies within $[0, +\infty)$ depends on the behaviour of the zeroes of the s_N's. In particular, in terms of the quantity δ_N defined in (3.21) we distinguish two alternative scenarios:

Case I: $\delta_N \to \infty$ as $N \to \infty$;
Case II: δ_N remains uniformly bounded, strictly above 0, in N.

If the operator A is bounded, then we are surely in Case I: indeed the orthogonal polynomials s_N are defined on $\sigma(A) \subset [0, \|A\|]$, and their zeroes cannot exceed $\|A\|$: this forces δ_N to blow up with N. Moreover, $\lambda_\infty = \lim_{N\to\infty} \lambda_N^{(N)} < +\infty$.

If instead A is unbounded, the $\lambda_k^{(N)}$'s fall in $[0, +\infty)$ and depending on their rate of possible accumulation at infinity δ_N may still diverge as $N \to \infty$ or stay bounded.

Clearly in Case II one has $\lambda_1 > 0$ and $\lambda_N = \infty$, for otherwise the condition $\lambda_1 = \lim_{N\to\infty} \lambda_1^{(N)} = 0$ or $\lambda_\infty = \lim_{N\to\infty} \lambda_N^{(N)} < +\infty$ would necessarily imply $\delta_N \to \infty$. Thus, in Case II the true interval of orthogonality is $[\lambda_1, +\infty)$ and it is separated from zero.

Remark 4.3. Estimate (4.6) in the proof shows that the vanishing rate of $\rho_\sigma(f^{(N)})$ is actually controlled by the vanishing rate of $\mu_\sigma([\lambda_1, \lambda_1^{(N)}))$. It is however unclear how to possibly quantify the latter. Let us recall (see Remark 2.6 and (2.16) in particular) that the Nemirovskiy-Polyak analysis [14] for the bounded-A case provides an explicit vanishing rate for $\rho_{\sigma'}(f^{(N)})$ for any $\sigma' \in (\sigma, \xi]$, based on a polynomial min-max argument that relies crucially on the finiteness of the interval where the orthogonal polynomials s_N are supported on (i.e., it relies on the boundedness of $\sigma(A)$). Therefore, there is certainly no room for applying the same argument to the present setting. In fact, we find it reasonable to expect that for generic (unbounded) A the quantity $\rho_\sigma(f^{[N]})$ vanishes with arbitrarily slow pace depending on the choice of the initial guess $f^{[0]}$. An indication in this sense comes from the numerical tests discussed in Section 5.

Remark 4.4. Still with reference to the Nemirovskiy-Polyak analysis [14] for bounded A, it is worth pointing out a subtle improvement of our proof. In [14] one does not make use of the very useful property that μ_σ is only supported on $[\lambda_1, \lambda_\infty]$ with $\mu_\sigma(\{\lambda_1\}) = 0$. The sole measure-theoretic information used therein is that $\lambda = 0$ is not an atom for μ_σ. Then, instead of naturally splitting the integration as in (4.1) above, in [14] one separates the small and the large spectral values at a threshold $\gamma_N = \min\{\lambda_1^{(N)}, \delta_N^{-1/2}\}$. Clearly $\gamma_N \to 0$, because $\delta_N \to \infty$ since A is bounded (see Remark 4.2 above), and through a somewhat lengthy analysis of the integration for $\lambda < \gamma_N$ and $\lambda \geqslant \gamma_N$ one reduces both integrations to one over $[0, \gamma_N)$. This finally pulls out the upper bound $\mu_\sigma([0, \gamma_N))$, which vanishes as $N \to \infty$ precisely because μ_σ is atom-less at $\lambda = 0$. Our proof shortens the overall argument and applies both to the bounded and to the unbounded case, with no need to introduce the γ_N cut-off.

5. Numerical tests

Let us present in this Section a selection of numerical tests that confirm the main features of our convergence result and corroborate our intuition on certain relevant differences with respect to the bounded case.

We choose $\mathcal{H} = L^2(\mathbb{R})$ and

$$\begin{array}{llll} \text{test-1a:} & A = -\dfrac{\mathrm{d}^2}{\mathrm{d}x^2} + 1\,, & \mathcal{D}(A) = H^2(\mathbb{R})\,, & f(x) = e^{-x^2}\,, \\ \text{test-1b:} & A = -\dfrac{\mathrm{d}^2}{\mathrm{d}x^2} + 1\,, & \mathcal{D}(A) = H^2(\mathbb{R})\,, & f(x) = (1+x^2)^{-1}\,, \\ \text{test-2a:} & A = -\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\,, & \mathcal{D}(A) = H^2(\mathbb{R})\,, & f(x) = e^{-x^2}\,, \\ \text{test-2b:} & A = -\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\,, & \mathcal{D}(A) = H^2(\mathbb{R})\,, & f(x) = (1+x^2)^{-1}\,, \end{array} \tag{5.1}$$

where H^2 denotes the usual Sobolev space of second order. In either case A is an unbounded, injective, non-negative, self-adjoint operator on $\mathcal{H}$, but only in test-1 does A^{-1} exist as an everywhere defined bounded operator.

We then consider the inverse linear problem $Af = g$ with the datum $g \in \mathrm{ran}A$ given by the above explicit choice of the solution $f \in H^2(\mathbb{R}) \cap C^\infty(\mathbb{R})$, and we construct conjugate gradient approximate solutions $f^{[N]}$ to f, namely ξ-iterates with $\xi = 1$, with initial guess $f^{[0]} = 0$ (the zero function on $\mathbb{R}$). Thus, each $f^{[N]}$ is searched over the Krylov subspace $\mathcal{K}_N(A, g) = \mathrm{span}\{g, Ag, \dots, A^{N-1}g\}$. $f^{[0]}$ is trivially smooth, and provided that f is smooth too the 1-iterates are all well-defined.

Of course in practice we replace the minimisation (2.1) with the standard, equivalent algebraic construction for the $f^{[N]}$'s [18, 12], so as to implement it as a routine in a symbolic computation software.

Iteratively we evaluate

$$\begin{aligned} \rho_0(f^{[N]}) &= \big\| f^{[N]} - f \big\|^2 \\ \rho_1(f^{[N]}) &= \big\langle f^{[N]} - f, A\,(f^{[N]} - f) \big\rangle \\ \rho_2(f^{[N]}) &= \big\| A f^{[N]} - g \big\|^2\,, \end{aligned} \tag{5.2}$$

(see (2.15) above) and we monitor the behaviour of such three quantities as N increases. Obviously $f^{[0]} \in \mathscr{C}_{A,g}(\sigma)$ $\forall \sigma \geqslant 0$: then Theorem 2.4 ensures that $\rho_\sigma(f^{[N]}) \to 0$ for any $\sigma \in [0, 1]$. In particular, both the error (ρ_0) and the energy norm (ρ_1) are predicted to vanish as $N \to \infty$. In the bounded case also the residual (ρ_2) would automatically vanish (Remark 2.6), however this is not a priori certain any longer in the unbounded case.

A fourth meaningful quantity to monitor is $N^2 \rho_1(f^{[N]})$. Recall indeed that *if A was bounded* the energy norm would be predicted to vanish not slower than a rate of order N^{-2} (as given by (2.16) with $\sigma = 0$ and $\sigma' = 1$). Thus, detecting now the possible failure of $N^2 \rho_1(f^{[N]})$ to stay bounded uniformly in N is an immediate signature of the fact that one cannot apply to the unbounded-A scenario the 'classical' quantitative convergence rate predicted by Nemirovskiy and Polyak for the bounded-A scenario [14], which in fact was also proved to be optimal in that case [15].

The results of our tests are shown in Figures 1 to 4.

Test 1a (Figure 1) reveals that the iterates not only converge in the sense of the error and of the energy norm as expected, but also in the residual sense, and the classical Nemirovskiy-Polyak convergence rate for the energy norm is not violated.

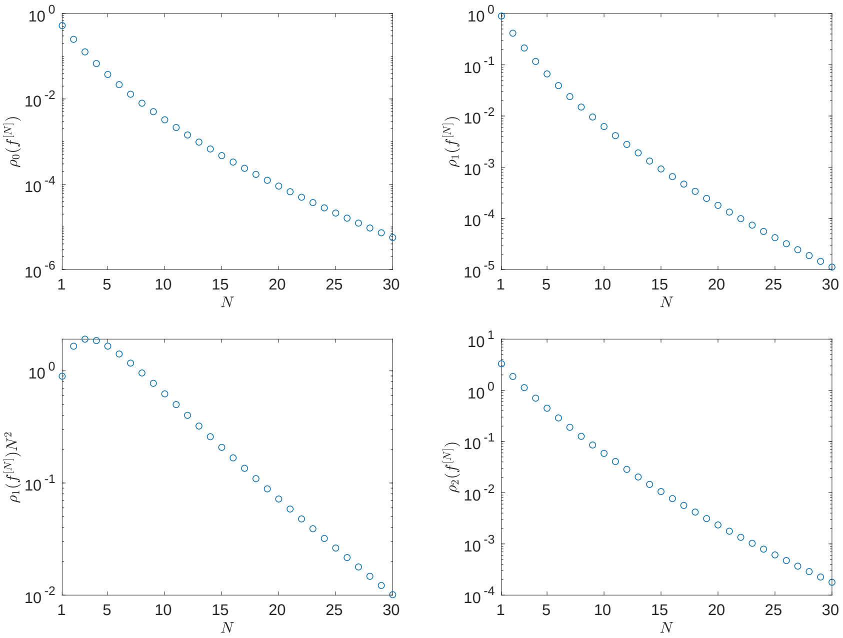

FIGURE 1. Numerical experiments for test-1a. From top left: $\rho_0(f^{[N]})$, $\rho_1(f^{[N]})$, $\rho_1(f^{[N]})N^2$, and $\rho_2(f^{[N]})$ indicators of convergence vs N.

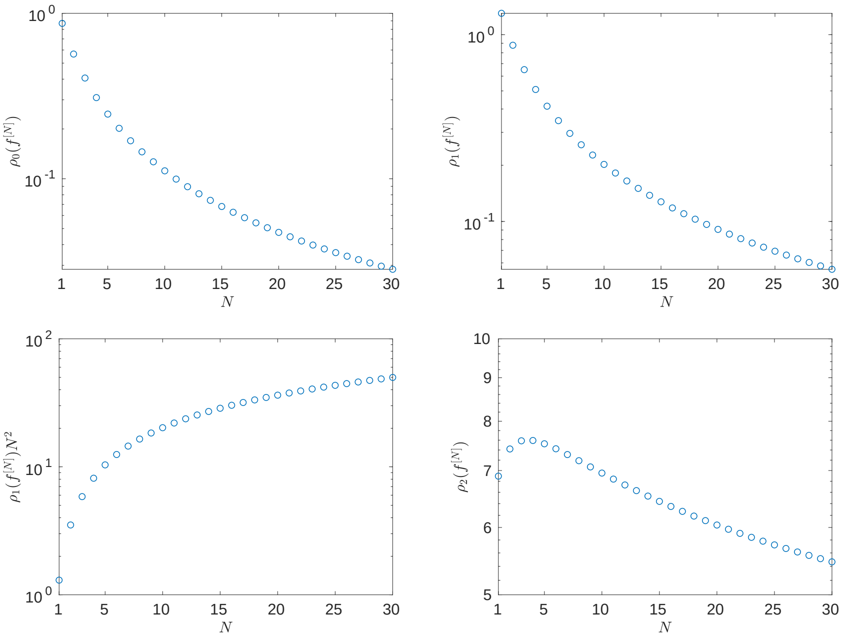

FIGURE 2. Numerical experiments for test-1b. From top left: $\rho_0(f^{[N]})$, $\rho_1(f^{[N]})$, $\rho_1(f^{[N]})N^2$, and $\rho_2(f^{[N]})$ indicators of convergence vs N.

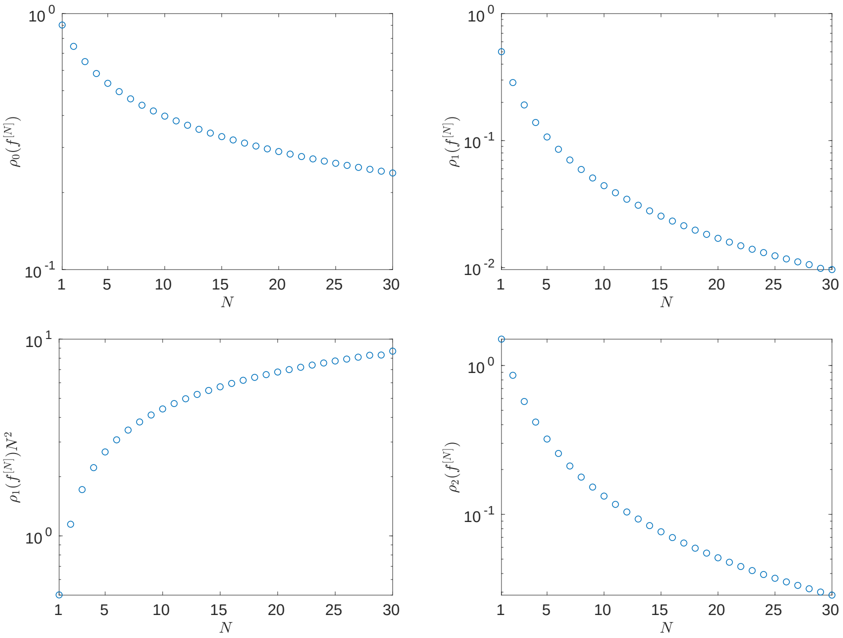

FIGURE 3. Numerical experiments for test-2a. From top left: $\rho_0(f^{[N]})$, $\rho_1(f^{[N]})$, $\rho_1(f^{[N]})N^2$, and $\rho_2(f^{[N]})$ indicators of convergence vs N.

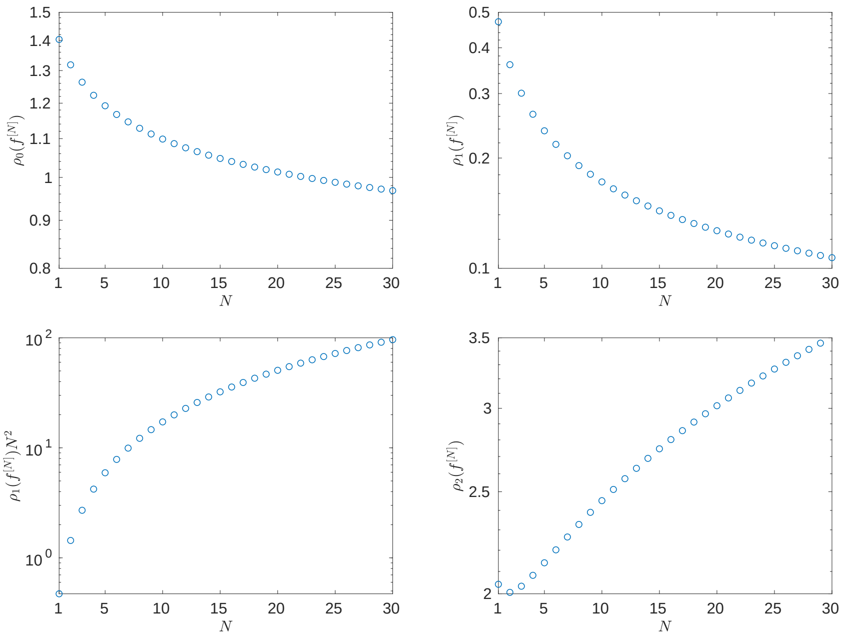

FIGURE 4. Numerical experiments for test-2b. From top left: $\rho_0(f^{[N]})$, $\rho_1(f^{[N]})$, $\rho_1(f^{[N]})N^2$, and $\rho_2(f^{[N]})$ indicators of convergence vs N.

Of course in retrospect the error vanishing is consistent with the residual vanishing, owing to the boundedness of A^{-1}: indeed, obviously,

$$\|f^{[N]} - f\| \leqslant \|A^{-1}\|_{\mathrm{op}}\|Af^{[N]} - g\|\,.$$

In test 2a the the invertibility of A over the whole $\mathcal{H}$ is removed: the residual (as well as the error and the energy norm) keeps approaching zero as N increases, however the Nemirovskiy-Polyak convergence rate for the energy norm is manifestly violated (Figure 3).

As opposite to tests 1a and 2a, in tests 1b and 2b the conjugate gradient algorithm is performed to approximate a solution f that is not localised as a Gaussian, but has instead a long tail at large distances. As the intuition suggests, this feature affects the convergence at higher regularity levels.

More precisely, test 1b (Figure 2) shows the vanishing of error, energy norm, and residual, however with an evident violation of the Nemirovskiy-Polyak vanishing rate for the energy norm.

Test 2b (Figure 4) confirms the vanishing of error and energy norm, as is expected in general, but the latter fails to behave according to the Nemirovskiy-Polyak vanishing rate, and also the iterates fail to converge in the sense of the residual.

Acknowledgements

We warmly thank Prof Nemirovskiy for providing us with very useful comments and clarifications on his work [14].

References

[1] H. Brakhage, *On ill-posed problems and the method of conjugate gradients*, in Inverse and Ill-Posed Problems, H. W. Engl and C. Groetsch, eds., Academic Press, 1987, pp. 165–175.
[2] T. Chihara, *An Introduction to Orthogonal Polynomials*, Dover Publications, New York, 1978.
[3] B. A. Cipra, *The best of the 20th century: Editors name top 10 algorithms*, SIAM News, 33 (2005).
[4] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), pp. 10–26.
[5] ———, *A correction concerning the convergence rate for the conjugate gradient method*, SIAM Journal on Numerical Analysis, 7 (1970), pp. 10–26.
[6] J. Dongarra and F. Sullivan, *The Top 10 Algorithms (Guest editors' intruduction)*, Comput. Sci. Eng., 2 (2000), pp. 22–23.
[7] M. Gilles and A. Townsend, *Continuous Analogues of Krylov Subspace Methods for Differential Operators*, SIAM Journal on Numerical Analysis, 57 (2019), pp. 899–924.
[8] R. Herzog and E. Sachs, *Superlinear convergence of Krylov subspace methods for self-adjoint problems in Hilbert space*, SIAM J. Numer. Anal., 53 (2015), pp. 1304–1324.
[9] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).
[10] W. J. Kammerer and M. Z. Nashed, *On the convergence of the conjugate gradient method for singular linear operator equations*, SIAM J. Numer. Anal., 9 (1972), pp. 165–181.
[11] T. H. Koornwinder, *Orthogonal polynomials*, in Computer algebra in quantum field theory, Texts Monogr. Symbol. Comput., Springer, Vienna, 2013, pp. 145–170.
[12] J. Liesen and Z. e. Strakoš, *Krylov subspace methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013. Principles and analysis.
[13] A. K. Louis, *Convergence of the conjugate gradient method for compact operators*, in Inverse and Ill-Posed Problems, H. W. Engl and C. Groetsch, eds., Academic Press, 1987, pp. 177–183.
[14] A. S. Nemirovskiy and B. T. Polyak, *Iterative methods for solving linear ill-posed problems under precise information. I*, Izv. Akad. Nauk SSSR Tekhn. Kibernet., (1984), pp. 13–25, 203.
[15] ———, *Iterative methods for solving linear ill-posed problems under precise information. II*, Engineering Cybernetics, 22 (1984), pp. 50–57.

[16] S. Olver, *GMRES for the Differentiation Operator*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 3359–3373.
[17] M. Reed and B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.
[18] Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.
[19] K. Schmüdgen, *Unbounded self-adjoint operators on Hilbert space*, vol. 265 of Graduate Texts in Mathematics, Springer, Dordrecht, 2012.
[20] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society Mathematical surveys, vol. I, American Mathematical Society, New York, 1943.
[21] V. Simoncini and D. B. Szyld, *Recent computational developments in Krylov subspace methods for linear systems*, Numerical Linear Algebra with Applications, 14 (2007), pp. 1–59.
[22] G. Szegő, *Orthogonal polynomials*, American Mathematical Society, Providence, R.I., fourth ed., 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.

(N. Caruso) International School for Advanced Studies – SISSA, via Bonomea 265, 34136 Trieste (Italy).
Email address: `ncaruso@sissa.it`

(A. Michelangeli) International School for Advanced Studies – SISSA, via Bonomea 265, 34136 Trieste (Italy).