# Convergence of a non-linear wavelet algorithm for the solution of PDE's

Silvia Bertoluzza * and Marco Verani †

February 13, 2001

**Abstract**

We prove convergence of an adaptive wavelet algorithm for the solution of elliptic PDE's, which combines Richardson type iterations with non linear projection steps.

**Key words:** Adaptivity, Non linear Approximation, Wavelets, Iterative schemes.

**AMS (MOS) Subject Classification:** 65F10, 65N12, 65N30

## 1 Introduction

One of the ultimate goals in the study of numerical methods for the solution of P.D.E's is to design methods that are optimal, in the sense that they allow either to get the best possible accuracy in a fixed amount of time or to the solution with a prescribed accuracy in the shortest computation time. If we consider for instance the problem of finding the solution $u$ to the partial differential equation

$$\operatorname{div}(\mathbf{a}\nabla u) = f \quad \text{in } \Omega \subset \mathbb{R}^d, \qquad\qquad u = 0 \quad \text{on } \partial\Omega, \tag{1.1}$$

what one would like is an algorithm allowing to fix a priori the number of degrees of freedom for representing the approximate solution, without fixing a priori the approximation space. The algorithm should be able to find the best possible approximation achievable, within all the approximation spaces (of a certain type) of the given dimension. Moreover, the computational cost should grow only linearly with the number $N$ of degrees of freedom.
Recently new developements on wavelets and non linear approximation have provided new tools, which allow in principle (at least in a wavelet framework) to build algorithms that have such characteristics. We recall that if we define, given a wavelet basis $\{\psi_\lambda,\ \lambda \in \Lambda\}$ for $L^2(\Omega)$, the non-linear space

$$\Sigma_N = \{u = \sum_\lambda c_\lambda \psi_\lambda : \quad \underline{c} = \{c_\lambda\}_{\lambda\in\Lambda} \in \sigma_N\} \qquad \text{with} \qquad \sigma_N = \{\underline{c} \in \ell^2(\Lambda) : \quad \#\{\lambda \in \Lambda : \ c_\lambda \neq 0\} \leq N\},$$

of functions in $L^2(\Omega)$ which can be expressed as the linear combination of at most $N$ wavelets, it is easy to define a non-linear projection operator $\mathbb{P}_N : H^s(\Omega) \to \Sigma_N$ which minimizes the error $\|u - \mathbb{P}_N u\|_{H^s(\Omega)}$ in an equivalent $H^s(\Omega)$ norm. In other words, $\mathbb{P}_N$ allows to get, given $u$, the best (in the equivalent $H^s$ norm) $N$ terms approximation, achievable within the subspaces spanned by subsets of the given basis. The rate of convergence of $\mathbb{P}_N u$ to $u$, as $N \to \infty$, is linked to the Besov regularity of $u$ (rather than to its Sobolev regularity as in the case of the classical linear approximation).
In order to obtain methods with the required features for solving (1.1), one possibility is to look for iterative approximation schemes in which by definition the iterates belong to the nonlinear space $\Sigma_N$. Convergence of algorithms of this type strongly relies on a key feature of wavelets, namely what is usually referred to as *wavelet preconditioning*: equation (1.1) can be rewritten in an equivalent form as infinite system:

$$\mathcal{A}\underline{u} = \underline{g}, \qquad \text{with } \ \mathcal{A}, \mathcal{A}^{-1} \in \mathcal{L}(\ell^2, \ell^2),$$

*Istituto di Analisi Numerica del C.N.R., v. Ferrata 1, 27100 Pavia (Italy), e-mail: aivlis@ian.pv.cnr.it, http://www.ian.pv.cnr.it/~aivlis

†Dip. di Matematica "F. Casorati", Univ. di Pavia, v. Ferrata 1, 27100 Pavia (Italy), e-mail: verani@dimat.unipv.it

where $\underline{u}$ is the infinite array of the coefficients $u_\lambda$ of the unknown solution $u = \sum_\lambda u_\lambda \check{\psi}_\lambda$, expressed with respect to the basis $\{\check{\psi}_\lambda\}_\lambda$ obtain by suitably renormalizing the basis $\{\psi_\lambda\}_\lambda$.
We are interested in schemes of the following type, where, by abuse of notation, we will also denote by $\mathbb{P}_N : \ell^2(\Lambda) \to \sigma_N$, the operator that associates to the coefficients $\underline{u} = \{u_\lambda\}_\lambda$ of a function $u$, the coefficients of its non linear projection $\mathbb{P}_N(u)$.

**Scheme A**

<u>**begin**</u>
  **Input** : $N, tol$
  $\underline{u}^{(0)} = \underline{0}$
  <u>**while**</u> $\|\underline{r}^{(n)}\|_{\ell^2} > tol$ <u>**do**</u>
    **construct** *an index set* $\Lambda^{(n)}$ *such that* $\#(\Lambda^{(n)}) < CN \log N$
      *and such that* $\mu \notin \Lambda^{(n)} \Longrightarrow u_\mu^{(n+1)} = 0$
      *(with* $\underline{u}^{(n+1)} = \mathbb{P}_N(\underline{u}^{(n)} + \theta(\underline{g} - \mathcal{A}\underline{u}^{(n)}))$ *not yet computed)*
    **compute** $\underline{r}^{(n)} = (r_\lambda)_{\lambda \in \Lambda^{(n)}}$
      *as* $r_\lambda^{(n)} = g_\lambda - (\mathcal{A}\underline{u}^{(n)})_\lambda$
    **update**
      $\underline{u}^{(n+1)} = \mathbb{P}_N(\underline{u}^{(n)} + \theta \underline{r}^{(n)}) \in \sigma_N$
  <u>**end**</u>
  **Output** : $u_N = \sum_\lambda u_\lambda^{(n+1)} \check{\psi}_\lambda$
<u>**end**</u>

The construction of a suitable set $\Lambda^{(n)}$, which is possible thanks to the good space frequency localization properties of wavelets, is a necessary step for an algorithm of this type to be pratically implemented, but we will not consider here such an issue. Our aim is rather to study the influence of the nonlinear projection step in the Richardson type algoritm and we will therefore concentrate on a model (non - computable) algorithm, namely

**Scheme B**

<u>**begin**</u>
  **Input** : $N, tol$
  $\underline{u}^{(0)} = \underline{0}$
  <u>**while**</u> $\|\underline{r}^{(n)}\|_{\ell^2} > tol$ <u>**do**</u>
    **compute** $\underline{r}^{(n)}$
      as $\underline{r}^{(n)} = \underline{g} - \mathcal{A}\underline{u}^{(n)}$
    **update**
      $\underline{u}^{(n+1)} = \mathbb{P}_N(\underline{u}^{(n)} + \theta \underline{r}^{(n)}) \in \sigma_N$
  <u>**end**</u>
  **Output** : $u_N = \sum_\lambda u_\lambda^{(n+1)} \check{\psi}_\lambda$
<u>**end**</u>

To be as general as possible we will analyse an algorithm of this type for a generic elliptic operator of order $s$. The aim of this paper is to give an analysis of such algorithm, showing convergence as well as an estimate of the error $\|u - u^{(n)}\|_{H^s(\Omega)}$.

## 2 The functional framework

In the following we will employ the notation $A \lesssim B$ to indicate that the quantity $A$ is bounded from above by a positive constant times the quantity $B$, while $A \simeq B$ will stand for $A \lesssim B \lesssim A$.

For simplicity let us fix the following functional setting: let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain, and suppose we are given a Riesz basis $\{\psi_\lambda\}_{\lambda \in \Lambda}$, $\Lambda = \cup_{j=0}^\infty \Lambda_j$, for $L^2(\Omega)$, such that the following norm equivalence for the Besov spaces $B^s_{p,q}(\Omega)$ holds for all $s, p, q$, $0 \le s \le S$, $0 < p < \infty$, $q > 0$:

$$\| \sum_\lambda u_\lambda \psi_\lambda \|^q_{B^s_{p,q}(\Omega)} \simeq \sum_j \rho^{q(s+\frac{n}{2}-\frac{n}{p})j} \left( \sum_{\lambda \in \Lambda_j} |u_\lambda|^p \right)^{q/p} . \tag{2.2}$$

The splitting of the index set $\Lambda$ as $\Lambda = \cup_{j=0}^\infty \Lambda_j$, corresponds to distinguishing functions "living" at different scales ($\lambda \in \Lambda_j \leftrightarrow (\mathrm{supp}\psi_\lambda) \sim 2^{-\mathrm{j}}$). It is beyond the goal of this paper to describe how and under which conditions on $\Omega$ such bases $\{\psi_\lambda\}_\lambda$ are constructed (see, among others, [**?**, **?**]).

Since $H^s(\Omega) = B^s_{2,2}(\Omega)$, from equivalence (2.2) we deduce that for all $s$, $0 \le s \le S$ :

$$\| \sum_\lambda u_\lambda \check{\psi}_\lambda \|_{H^s(\Omega)} \simeq \| \underline{u} \|_{\ell^2}, \quad \text{with } \check{\psi}_\lambda = \rho^{-js} \psi_\lambda. \tag{2.3}$$

Moreover, when considering nonlinear approximation in $H^s$, the scale of Besov spaces $B^{r+s}_{\tau,\tau}(\Omega)$ – where $\tau = \tau(r)$ is defined by the relation $\frac{1}{\tau} = \frac{r}{d} + \frac{1}{2}$ – will naturally appear. For these spaces the norm equivalences in terms of wavelet coefficients are quite simple; indeed using again equivalence (2.2), we obtain:

$$\| \sum_\lambda u_\lambda \check{\psi}_\lambda \|_{B^{r+s}_{\tau,\tau}(\Omega)} \simeq \| \underline{u} \|_{\ell^\tau}, \quad \lambda \in \Lambda_j, \qquad \text{with } \frac{1}{\tau} = \frac{r}{d} + \frac{1}{2} \tag{2.4}$$

where again $\check{\psi}_\lambda = \rho^{-js}\psi_\lambda$.

In the following it will also be useful to consider the space of functions whose coefficients, with respect to the rescaled basis $\{\check{\psi}_\lambda\}_\lambda$, are in the *weak–$\ell^\tau$* space $\ell^\tau_w$, which can be defined as the space of sequences $\underline{u} = \{u_\lambda\}_\lambda$ for which there exists a constant $C$ such that

$$\#\{\lambda : \quad |u_\lambda| \ge \varepsilon\} \le C\varepsilon^{-\tau}, \tag{2.5}$$

the norm $\|\underline{u}\|^\tau_{\ell^\tau_w}$ being defined as the smallest $C$ which verifies relation (2.5). It is possible to prove that $\ell^\tau \subseteq \ell^\tau_w$, which implies that the coefficients $\{u_\lambda\}_\lambda$ of a function $u \in B^{r+s}_{\tau,\tau}$ verify $\{u_\lambda\}_\lambda \in \ell^\tau_w$.

The variational framework considered here corresponds to an elliptic Neumann BVP. on the bounded domain $\Omega$, however it is not difficult to check that the whole proof of the result obtained in the next section relies on the representation (3.9) of the Problem and on the norm equivalences (2.2) and (2.4). Therefore such results carry over to much more general situations (Dirichelet BVP, Integral equations) where the space $B^{r+s}_{\tau,\tau}$ is substituted by the space for which a representation of the form (2.4) holds

# 3 Analysis of the non linear Richardson scheme

Let $A : H^s(\Omega) \to H^{-s}(\Omega)$, $0 \le s \le S$, ($H^{-s}(\Omega)$ denoting here the dual of $H^s(\Omega)$) be a bounded linear operator and let the corresponding bilinear form $a : H^s(\Omega) \times H^s(\Omega) \to \mathbb{R}$ be defined as:

$$a(u, v) := < Au, v >, \qquad \forall u, v \in H^s(\Omega),$$

where $< .,. >$ denotes the duality pairing between $H^{-s}$ and $H^s$. We assume that the bilinear form $a$ is symmetric positive definite, that is $\forall u, v \in H^s(\Omega)$:

$$a(u, v) \le M \|u\|_{H^s(\Omega)} \|v\|_{H^s(\Omega)}, \qquad a(u, u) \ge \alpha \|u\|^2_{H^s(\Omega)}.$$

We consider the following problem: given $g \in H^{-s}(\Omega)$, find $u \in H^s(\Omega)$ such that:

$$Au = g. \tag{3.7}$$

Under our assumptions it is well known that for any $f \in H^{-s}(\Omega)$ equation (3.7) has a unique solution; this is also the unique solution of the equivalent variational problem: find $u \in H^s(\Omega)$ such that

$$a(u, v) = < g, v >, \qquad \forall v \in H^s(\Omega). \tag{3.8}$$

Expressing $u$ in terms of the rescaled basis $\{\check{\psi}_\lambda\}_\lambda$, we can rewrite (3.8) in terms of the Fourier coefficients $\underline{u} = \{u_\lambda\}_\lambda$ of the unknown solution

$$u = \sum_\lambda u_\lambda \check{\psi}_\lambda,$$

as an infinite dimensional system of linear equations:

$$\mathcal{A}\underline{u} = \underline{g} \tag{3.9}$$

where

$$\mathcal{A} = (a_{\mu,\lambda})_{\mu,\lambda\in\Lambda}, \quad a_{\mu,\lambda} =< A\check{\psi}_\lambda, \check{\psi}_\mu >, \qquad \underline{g} = \{g_\mu\}_{\mu\in\Lambda}, \quad g_\mu =< g, \check{\psi}_\mu >,$$

are a bi-infinite matrix and an infinite array respectively. It is not difficult to check that $\mathcal{A} \in \mathcal{L}(\ell^2, \ell^2)$ and that it is boundedly invertible; that is:

$$\|\mathcal{A}\|_{\mathcal{L}(\ell^2,\ell^2)} \lesssim C_1, \qquad \|\mathcal{A}^{-1}\|_{\mathcal{L}(\ell^2,\ell^2)} \lesssim C_2.$$

Moreover we assume that the basis $\{\check{\psi}_\lambda\}_\lambda$ and the operator $\mathcal{A}$ are such that it holds for some $\tau_0 < 2$:

$$\mathcal{A} \in \mathcal{L}(\ell^{\tau_0}_w, \ell^{\tau_0}_w). \tag{3.10}$$

We remark that under suitable space-frequency localization properties of the wavelet basis $\{\check{\psi}_\lambda\}_\lambda$, condition (3.10) holds for a wide class of differential and pseudo-differential operators ([?]).

Let us now recall that $\Sigma_N \subseteq V$,

$$\Sigma_N = \{u = \sum_\lambda c_\lambda \psi_\lambda : \quad \underline{c} = \{c_\lambda\}_{\lambda\in\Lambda} \in \sigma_N\} \qquad \text{with} \qquad \sigma_N = \{\underline{c} \in \ell^2(\Lambda) : \quad \#\{\lambda \in \Lambda : \ c_\lambda \neq 0\} \leq N\},$$

is defined as the non linear space of functions in $L^2(\Omega)$ which can be represented as the linear combination of at most $N$ elements of the basis $\{\check{\psi}_\lambda\}_\lambda$. A non linear projector $\mathbb{P}_N : L^2(\Omega) \to \Sigma_N$ can be defined as follows: given $u = \sum_\lambda u_\lambda \check{\psi}_\lambda$, let us introduce a decreasing rearrangement $\{|u_{\lambda(n)}|\}_{n\in\mathbb{N}}$ of the sequence $\{|u_\lambda|\}_{\lambda\in\Lambda}$, where the application $n \in \mathbb{N} \longrightarrow \lambda(n) \in \Lambda$ is bijective and verifies $n < m \Longrightarrow |u_{\lambda(n)}| \geq |u_{\lambda(m)}|$; $\mathbb{P}_N(u)$ is then defined by:

$$\mathbb{P}_N(u) = \sum_{n=1}^N u_{\lambda(n)} \check{\psi}_{\lambda(n)},$$

that is only the $N$ greatest (in absolute value) coefficients of $u$ are retained. We recall that by abuse of notation we will also indicate by $\mathbb{P}_N : \ell^2 \to \sigma_N$ the operator associating to the sequence $\underline{u}$ the coefficients of the function $\mathbb{P}_N(\sum_\lambda u_\lambda \psi_\lambda)$. The accuracy of the corresponding approximation is directly related to $\ell^\tau_w$ regularity of the sequence of coefficients of $u$, as stated by the following theorem [?, ?].

**Theorem 3.1** *Let $u = \sum_{\lambda\in\Lambda} u_\lambda \check{\psi}_\lambda$, with $\check{\psi}_\lambda = \rho^{-js}\psi_\lambda$, $s \leq S$. If $\{u_\lambda\}_\lambda \in \ell^\tau_w$ then*

$$\inf_{w\in\Sigma_N} \|u - w\|_{H^s} \lesssim \|u - \mathbb{P}_N u\|_{H^s} \lesssim N^{-(\frac{1}{\tau}+\frac{1}{2})} \|\underline{u}\|_{\ell^w_\tau}$$

*where the implicit constants in the bounds depend only on $\tau$.*

Our aim is to find an approximate solution to problem (3.9) in $\Sigma_N$. In order to do that we consider the abstract non-linear Richardson Scheme B; as already stated in the introduction this is not a computable numerical scheme since it involves operations on infinite matrices and vectors. Such scheme will eventually be coupled with suitable compression steps applied both to the operator $\mathcal{A}$ and to the right hand side $\underline{g}$, which will allow to actually implement it efficiently. Nevertheless it is interesting to consider such a scheme in order to analyze the influence of the nonlinear operator $\mathbb{P}_N$. In particular the main result of this paper is the following Theorem:

**Theorem 3.2** *Let $\mathcal{A} \in \mathcal{L}(\ell^{\tau_0}_w, \ell^{\tau_0}_w) \cap \mathcal{L}(\ell^2, \ell^2)$ for some $\tau_0 < 2$. Then there exists a $\tilde{\tau} < 2$ and a $\theta_0 > 0$ such that, for all $\theta$, $0 < \theta < \theta_0$, it holds (the implicit constants in the inequalities depending on $\theta$)*

**(i)** *stability: if $\underline{g} \in \ell^2$, we have*

$$\|\underline{u}^n\|_{\ell^2} \lesssim \|\underline{g}\|_{\ell^2}, \quad \forall n \in \mathbb{N}; \tag{3.11}$$

**(ii)** *approximation error estimate: if $\underline{g} \in \ell^\tau_w$, $\tilde{\tau} < \tau \le 2$ then, setting $\underline{e}^n = \underline{u}^n - \underline{u}$, it holds for some $\mu < 1$:*

$$\|\underline{e}^n\|_{\ell^2} \lesssim \mu^n \|e_o\|_{\ell^2} + \frac{1}{1-\mu} N^{-(\frac{1}{\tau}+\frac{1}{2})}; \tag{3.12}$$

In order to prove theorem 3.2 we need to recall the following lemma [?].

**Lemma 3.1** *Let $\mathcal{A} \in \mathcal{L}(\ell^{\tau_0}_w, \ell^{\tau_0}_w) \cap \mathcal{L}(\ell^2, \ell^2)$, for some $\tau_0 < 2$. Then there exists a costant $\theta_0 > 0$ and a $\tilde{\tau}$, $\tau_0 \le \tilde{\tau} < 2$, such that $\forall \theta$ with $0 \le \theta \le \theta_0$ it holds:*

$$\|I - \theta \mathcal{A}\|_{\mathcal{L}(\ell^2,\ell^2)} \le \mu < 1, \tag{3.13}$$

$$\|I - \theta \mathcal{A}\|_{\mathcal{L}(\ell^\tau_w,\ell^\tau_w)} \le \gamma < 1, \qquad \textit{for all } \tau, \quad \tilde{\tau} \le \tau \le 2 \tag{3.14}$$

**Proof of Theorem 3.2:** First of all we observe that $\mathbb{P}_N$ is $\ell^2$-contractive. Then, since $\underline{u}^n \in \ell^2$ (only $N$ coefficients are non zero by definition), using Lemma 3.1, one has (3.13):

$$\|\underline{u}^{n+1}\|_{\ell^2} = \|P_N(\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n))\|_{\ell^2} \le \|\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n)\|_{\ell^2} \le \|\theta \underline{g}\|_{\ell^2} + \mu \|\underline{u}^n\|_{\ell^2}.$$

By iterating this bound we obtain:

$$\|\underline{u}^{n+1}\|_{\ell^2} \le \left(\sum_{i=0}^{n} \mu^i\right) \|\theta \underline{g}\|_{\ell^2} + \mu^{n+1}\|\underline{u}^0\|_{\ell^2},$$

which, since $\mu < 1$ and $\underline{u}^0 = 0$, yields (3.11).
Now let

$$\underline{\varepsilon}^n = P_N(\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n)) - (\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n)).$$

A simple calculation yields:

$$\underline{e}^{n+1} - \underline{e}^n + \theta \mathcal{A} \underline{e}^n = \underline{\varepsilon}^n,$$

which, taking the $\ell^2$ norm and using (3.13), yields

$$\|\underline{e}^{n+1}\|_{\ell^2} \le \mu \|\underline{e}^n\|_{\ell^2} + \|\underline{\varepsilon}^n\|_{\ell^2}. \tag{3.15}$$

Iterating (3.15), we then obtain:

$$\|\underline{e}^{n+1}\|_{\ell^2} = \sum_{k=0}^{n} \mu^{n-k} \|\underline{\varepsilon}^k\|_{\ell^2} + \mu^{n+1}\|e_0\|_{\ell^2} \le \left(\max_{0\le k \le n} \|\underline{\varepsilon}^k\|_{\ell^2}\right) \sum_{i=0}^{n} \mu^i + \mu^{n+1}\|e_0\|_{\ell^2}. \tag{3.16}$$

The sum on the right hand side of (3.16) converges ($\mu < 1$) and then we can write:

$$\|\underline{e}^{n+1}\|_{\ell^2} \le \frac{1}{1-\mu} \max_k \|\underline{\varepsilon}^k\|_{\ell^2} + \mu^{n+1}\|e_0\|_{\ell^2}. \tag{3.17}$$

To conclude, we only need to give an uniform bound on $\|\varepsilon_k\|_{\ell^2}$. Using Lemma 3.1 it is not difficult to show that $\underline{g} \in \ell^\tau_w$ implies $\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n) \in \ell^\tau_w$ with:

$$\|\underline{u}^n + \theta(\underline{g} - \mathcal{A}\ \underline{u}^n)\|_{\ell^\tau_w} \le C, \tag{3.18}$$

uniformly in $n$. Indeed $\|I - \theta\mathcal{A}\|_{\mathcal{L}(\ell^\tau_w,\ell^\tau_w)} \le \gamma < 1$ and since $\mathbb{P}_N$ is $\ell^\tau_w$ contractive, setting

$$\underline{w}^{n+1} = \underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n)$$

we can write:

$$\begin{aligned} \|\underline{w}^{n+1}\|_{\ell^\tau_w} &= \|(I - \theta\mathcal{A})\underline{u}^n + \theta \underline{g}\|_{\ell^\tau_w} \le \|(I - \theta\mathcal{A})\underline{u}^n\|_{\ell^\tau_w} + \|\theta \underline{g}\|_{\ell^\tau_w} \le \gamma \|\underline{u}^n\|_{\ell^\tau_w} + \|\theta \underline{g}\|_{\ell^\tau_w} \\ &= \gamma \|\mathbb{P}_N(\underline{w}^n)\|_{\ell^\tau_w} + \|\theta \underline{g}\|_{\ell^\tau_w} \le \|\theta \underline{g}\|_{\ell^\tau_w} + \gamma \|\underline{w}^n\|_{\ell^\tau_w}. \end{aligned}$$

By iterating this bound, we obtain:

$$\|\underline{w}^{n+1}\|_{\ell^\tau_w} \le \left(\sum_{i=0}^{n} \gamma^i\right) \|\theta \underline{g}\|_{\ell^\tau_w} + \gamma^n \|\underline{w}^1\|_{\ell^\tau_w} \le \frac{\theta}{1-\gamma}\|\underline{g}\|_{\ell^\tau_w}, \quad \forall n,$$

which yields (3.18). Thanks to (3.18), by applying Theorem 3.1, we have that:

$$\max_n \|\underline{\varepsilon}^n\|_{\ell^2} \lesssim N^{-(\frac{1}{\tau}+\frac{1}{2})} \max_n \|\underline{u}^n + \theta(\underline{g} - \mathcal{A}\underline{u}^n)\|_{\ell^\tau_w} \le N^{-(\frac{1}{\tau}+\frac{1}{2})} C(\mathbf{g}).$$

Combining such bound with (3.17), implies the thesis. □

Using norm equivalence (2.4), this yields the following corollary.

**Corollary 3.1** *Let $u$ be the solution of (3.8) and let $g$ belong to $B^{r+s}_{\tau,\tau}(\Omega)$, with $r$, $0 < r+s \le \min\{S, d/\tilde{\tau} - d/2\}$ and with $\tau$ given by $1/\tau = r/d + 1/2$. If $u_N^{(n)} = \sum_\lambda u_\lambda^{(n)} \check{\psi}_\lambda$ is the non linear approximation of $u$ at step $n$ given by the non-linear Richardson scheme with $\theta < \theta_0$, then it holds, for some $\mu < 1$:*

$$\|u - u_N^{(n)}\|_{H^s(\Omega)} \le \mu^n \|u - u_N^{(0)}\|_{H^s(\Omega)} + \frac{C}{1-\mu} N^{-\frac{r}{d}}, \quad \forall n \in \mathbb{N}.$$

**Remark 3.1** *Though for simplicity we set $\underline{u}^0 = 0$ throughout the paper, it is not difficult to realize that the result holds unchanged also for any initial guess in $\Sigma_N$.*

# References