Research Article

Max Kontak* and Volker Michel

The Regularized Weak Functional Matching Pursuit for linear inverse problems

https://doi.org/10.1515/jiip-2018-0013
Received February 27, 2018; revised July 13, 2018; accepted August 27, 2018

Abstract: In this work, we present the so-called Regularized Weak Functional Matching Pursuit (RWFMP) algorithm, which is a weak greedy algorithm for linear ill-posed inverse problems. In comparison to the Regularized Functional Matching Pursuit (RFMP), on which it is based, the RWFMP possesses an improved theoretical analysis including the guaranteed existence of the iterates, the convergence of the algorithm for inverse problems in infinite-dimensional Hilbert spaces, and a convergence rate, which is also valid for the particular case of the RFMP. Another improvement is the cancellation of the previously required and difficult to verify semi-frame condition. Furthermore, we provide an a-priori parameter choice rule for the RWFMP, which yields a convergent regularization. Finally, we will give a numerical example, which shows that the "weak" approach is also beneficial from the computational point of view. By applying an improved search strategy in the algorithm, which is motivated by the weak approach, we can save up to 90 % of computation time in comparison to the RFMP, whereas the accuracy of the solution does not change as much.

Keywords: Convergence rate, greedy algorithm, ill-posed problem, inverse problem, non-linear approximation, Tikhonov regularization

MSC 2010: 65J22, 65R32, 35R30, 45Q05, 47A52, 65J20

1 Introduction

In this paper, we deal with the solution of the inverse problem

$$Tf = g, \tag{1.1}$$

where $T\colon \mathcal{X} \to \mathcal{Y}$ is a linear and bounded operator between two Hilbert spaces $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ and $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}})$, $g \in \mathcal{Y}$ represents given data, and $f \in \mathcal{X}$ is the desired solution.

Linear inverse problems like (1.1) arise in a variety of applications and are often ill-posed, which means that one of the following conditions is violated:

(a) a solution of (1.1) exists for every $g \in \mathcal{Y}$,
(b) there is at most one solution of (1.1) for every $g \in \mathcal{Y}$,
(c) the solution $f \in \mathcal{X}$ depends continuously on the data $g \in \mathcal{Y}$.

Often, when solving inverse problems, one does no longer consider "true" solutions but rather so-called *best-approximate solutions* using the Moore–Penrose inverse [22, 23]. Since such a best-approximate solution always exists and it is always unique, the well-posedness of the inverse problem can be reduced to a variation

***Corresponding author: Max Kontak,** Geomathematics Group, Department of Mathematics, University of Siegen, Walter-Flex-Str. 3, 57068 Siegen; and High-Performance Computing, Simulation and Software Technology, DLR German Aerospace Center, Linder Höhe, 51147 Köln, Germany, e-mail: max.kontak@dlr.de. http://orcid.org/0000-0003-3738-7483
Volker Michel, Geomathematics Group, Department of Mathematics, University of Siegen, Walter-Flex-Str. 3, 57068 Siegen, Germany, e-mail: michel@mathematik.uni-siegen.de. http://orcid.org/0000-0002-2551-0491

of condition (c) above. On this basis, a different characterization of well-posedness going back to Nashed is: the inverse problem (1.1) is well-posed if and only if ran T is closed in $\mathcal{Y}$ (cf. [3, Proposition 2.4]).

The ill-posedness of inverse problems plays an important role in the applications as well as in the implementation of solution algorithms. Small data errors, which may be introduced by measurements or by using floating point arithmetic on computers, may lead to large errors in the approximation of the solution. To overcome this difficulty, so-called regularization methods are applied, which replace the original ill-posed inverse problem by a well-posed problem, which is an approximation of the original problem in a certain sense. A standard technique is Tikhonov regularization, which means that instead of minimizing the squared data error

$$f \mapsto \|g - Tf\|_{\mathcal{Y}}^2 \tag{1.2}$$

the so-called Tikhonov functional

$$f \mapsto \|g - Tf\|_{\mathcal{Y}}^2 + \lambda \|f\|_{\mathcal{X}}^2 \tag{1.3}$$

is minimized, where $\lambda > 0$ is a so-called regularization parameter [28].

Based on [16], in [4, 5, 17], a greedy algorithm for the solution of linear inverse problems, the so-called *Regularized Functional Matching Pursuit* (RFMP), was derived, which incorporates such a Tikhonov regularization. In the unregularized and the regularized case, the algorithm is based on the idea of iteratively minimizing the squared data error (1.2) and the Tikhonov functional (1.3), respectively, by adding the optimal element from a so-called dictionary $\mathcal{D} \subseteq \mathcal{X}$ to the current solution. For more details, see Section 3.1 of this paper and the references above. In comparison to other regularization algorithms for inverse problems, the RFMP is able to combine very diverse types of basis functions to form an approximate solution of the inverse problem. For example, both localized and global functions can be used in geoscientific applications with a spherical domain (for examples, see [4–7, 19]). A modification of the algorithm, called ROFMP, was presented in [20, 26]. Furthermore, a greedy algorithm for use in industrial applications was given in [8].

So far, the theoretical analysis of the RFMP was restricted to a finite-dimensional data space $\mathcal{Y}$. This is a reasonable assumption in practical cases where the data are, for example, samples of an observable. On the other hand, there are also applications where the right-hand side is given as a function (e. g., derived as a model of some data). Furthermore, from a theoretical point of view, the range of a linear operator is closed if it is finite-dimensional [3, Chapter 2.2]. In consequence, the problems which were handled by the RFMP were actually well-posed in the sense of Nashed (but probably ill-conditioned).

In this paper, which is based on the PhD thesis [15] of the first author, we present the RWFMP, a generalization of the RFMP that can handle data which can be an element of an arbitrary (possibly infinite-dimensional) Hilbert space. We adopt the idea of the Weak Greedy Algorithm (WGA) from [27] and apply it to the RFMP algorithm. As the WGA is a modification of the Pure Greedy Algorithm (PGA, see [2]), the RWFMP is an analogous modification of the RFMP.

When considering infinite-dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, another difficulty arises in the analysis of the algorithm. As already mentioned, the RFMP is based on an iterative minimization of the Tikhonov functional. Unfortunately, it is not clear if there exists a minimizing dictionary element (which is characterized as the maximizer of a certain term). The novel RWFMP algorithm instead adds a dictionary element to the approximation, which is near to the optimum in a certain sense, such that this difficulty is surmounted. For more details, see Section 3.1 of this article.

We will furthermore provide a numerical proof of concept, which shows that the approach of the RWFMP is not only beneficial from the theoretical point of view. We provide a numerical example, which shows that the inclusion of a specific search strategy in the algorithm can save up to 90 % of the computation time.

The paper is structured as follows: After the introduction in Section 1, Section 2 is dedicated to the Weak Functional Matching Pursuit (WFMP), which can be applied to well-posed inverse problems of the form (1.1). We show weak and strong convergence of the residuals as well as convergence of the approximations in the domain to the best-approximate solution of the inverse problem. Finally, rates of convergence are derived. In Section 3, it turns out that for ill-posed problems a regularizing version of the algorithm, the RWFMP, can be derived by applying the WFMP to a modified inverse problem. It is shown that the algorithm converges to a minimizer of the Tikhonov functional and also in the regularized case, convergence rates will be proved.

Moreover, we prove that there exists an a-priori parameter choice method for the RWFMP that yields a convergent regularization. A numerical example in Section 4 shows the possibilities for an optimization of the RFMP algorithm by using the approach of a weak greedy algorithm. Finally, Section 5 sums up the presented results and an outlook is given on further research connected to the RWFMP.

2 The Weak Functional Matching Pursuit (WFMP)

In this section, we will present the Weak Functional Matching Pursuit (WFMP), which we obtain by applying the idea of the Weak Greedy Algorithm (WGA) from [27] to the Functional Matching Pursuit presented in [4, 5]. We will derive the convergence of the algorithm in the weak and the strong topology both in the range $\mathcal{Y}$ and the domain $\mathcal{X}$ of the operator T. It will be shown that, for given data $g \in \mathcal{Y}$, the algorithm converges to a solution f^+ of the normal equation

$$T^\star T f^+ = T^\star g, \tag{2.1}$$

where $T^\star : \mathcal{Y} \to \mathcal{X}$ is the adjoint operator of T. It is well known that the solution of the normal equation (2.1) is also a least-squares solution of the inverse problem (1.1) [3, Theorem 2.6].

In this section, we will assume the well-posedness of the inverse problem (1.1) in the sense of Nashed, that is $\overline{\operatorname{ran} T} = \operatorname{ran} T$. We will drop this constraint in Section 3 when regularization is applied to the ill-posed inverse problem.

2.1 The algorithm

The Functional Matching Pursuit (FMP) as presented in [4, 5] is based on the following concept: Let a so-called *dictionary* $\mathcal{D} \subseteq \mathcal{X}$ be given. Beginning with an initial approximation $f_0 \in \mathcal{X}$, we iteratively define

$$f_{n+1} := f_n + \alpha_{n+1} d_{n+1},$$

where $d_{n+1} \in \mathcal{D}$ is chosen such that

$$\left| \frac{\langle r_n, Td_{n+1} \rangle_\mathcal{Y}}{\| Td_{n+1} \|_\mathcal{Y}} \right| = \max_{d \in \mathcal{D}} \left| \frac{\langle r_n, Td \rangle_\mathcal{Y}}{\| Td \|_\mathcal{Y}} \right|, \tag{2.2}$$

and

$$\alpha_{n+1} := \frac{\langle r_n, Td_{n+1} \rangle_\mathcal{Y}}{\| Td_{n+1} \|_\mathcal{Y}^2},$$

where $r_n := g - Tf_n$ denotes the residual (or data misfit) in step n. A geometrical motivation for the previous equations is the following: For an arbitrary element d from the dictionary, the term

$$\frac{\langle r_n, Td \rangle_\mathcal{Y}}{\| Td \|_\mathcal{Y}} = \left\langle r_n, \frac{Td}{\| Td \|_\mathcal{Y}} \right\rangle_\mathcal{Y}$$

represents the “length” of the projection of the current residual r_n into the direction of Td. Equation (2.2) therefore characterizes Td_{n+1} as the direction, which is closest to being parallel to r_n. The image of the update

$$T(\alpha_{n+1} d_{n+1}) = \frac{\langle r_n, Td_{n+1} \rangle_\mathcal{Y}}{\| Td_{n+1} \|_\mathcal{Y}^2} Td_{n+1} = \left\langle r_n, \frac{Td_{n+1}}{\| Td_{n+1} \|_\mathcal{Y}} \right\rangle_\mathcal{Y} \frac{Td_{n+1}}{\| Td_{n+1} \|_\mathcal{Y}}$$

consequently is the projection of r_n onto the direction Td_{n+1} itself. This choice of d_{n+1} appears to be natural in the sense that Td_{n+1} is most suitable for covering large parts of the residual r_n.

The FMP possesses several drawbacks both in theory and in practice.

First, the range of the operator is assumed to be finite-dimensional. On the one hand, this is no problem in practice since an infinite-dimensional range is not realizable on computers and there is always only a finite amount of measured data. On the other hand, this is not the usual setting in the theoretical analysis of inverse problems since often the infinite-dimensionality of the range causes the ill-posedness of the problem and makes it, therefore, more interesting. Also, as already mentioned in the introduction, one might be interested in data that are given by a model function.

Secondly, in (2.2), one assumes that the maximum exists. It is trivial that this holds true if $\#\mathcal{D} < \infty$, which is the only case that can be realized on a computer in practice. However, if $\#\mathcal{D} = \infty$, which is necessary in theory to span the infinite-dimensional Hilbert space $\mathcal{X}$, it is not clear that the supremum is attained and the maximum exists. Nevertheless, if the dictionary is finite but very large, it may also be computationally expensive to find the maximum in practice, even if one can be sure that it exists.

Both drawbacks are fixed by applying the basic concept of the Weak Greedy Algorithm (WGA, see [27]) to the FMP, which yields the following algorithm. In contrast to the FMP, we use a normalization of the dictionary elements in the co-domain here, that is, we require $\|Td\|_{\mathcal{Y}} = 1$ for all $d \in \mathcal{D}$. This implies that $\ker T \cap \mathcal{D} = \emptyset$. In the theoretical analysis of the algorithm, it will turn out that this is no restriction at all.

Algorithm 2.1 (Weak Functional Matching Pursuit, WFMP). Let $\mathcal{X}$, $\mathcal{Y}$, T be given as for problem (1.1). Furthermore, let data $g \in \mathcal{Y}$, a weakness parameter $\varrho \in (0, 1]$ and the initial approximation $f_0 = 0 \in \mathcal{X}$ be given. Choose a dictionary $\mathcal{D} \subseteq \{d \in \mathcal{X} \mid \|Td\|_{\mathcal{Y}} = 1\} \subseteq \mathcal{X}$.

(1) Set $n := 0$, define the residual $r_0 := g - Tf_0 = g$ and choose a stopping criterion.

(2) Find an element $d_{n+1} \in \mathcal{D}$ which fulfills

$$|\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}}| \geq \varrho \sup_{d\in\mathcal{D}} |\langle r_n, Td\rangle_{\mathcal{Y}}| \tag{2.3}$$

Set

$$\alpha_{n+1} := \langle r_n, Td_{n+1}\rangle_{\mathcal{Y}}, \tag{2.4}$$

as well as $f_{n+1} := f_n + \alpha_{n+1} d_{n+1}$ and $r_{n+1} := g - Tf_{n+1} = r_n - \alpha_{n+1} Td_{n+1}$.

(3) If the stopping criterion is fulfilled, then f_{n+1} is the output. Otherwise, increase n by 1 and return to step (2).

Remark. If $\varrho = 1$, then the WFMP is equivalent to the FMP (up to the normalization of the dictionary). Thus, all the following results also apply to the FMP, even with an infinite-dimensional range space $\mathcal{Y}$. Additionally, if $\varrho < 1$, the existence of d_{n+1} in (2.3) is guaranteed.

In the following sections, we will prove the convergence of the WFMP both in the data space as well as in the domain of the operator T. This is done in several steps: Firstly, we prove weak convergence of the residuals in the data space. Secondly, strong convergence in the data space is shown. Finally, we prove the convergence of the iteration also in the domain.

2.2 Weak convergence of the residuals

To prove that the residuals converge to zero in the weak sense, we first prove the convergence of the norm of the residuals. The first few lemmas are identical to the considerations in [4, 5] such that we omit the proofs.

Lemma 2.2. *Let $(r_n)_{n\in\mathbb{N}_0}$ be the sequence of the residuals arising in Algorithm 2.1. Then the following holds true:*

(a) *The sequence $(\|r_n\|_{\mathcal{Y}})_{n\in\mathbb{N}_0}$ is monotonically decreasing.*

(b) *The sequence $(\|r_n\|_{\mathcal{Y}})_{n\in\mathbb{N}_0}$ converges.*

Note that we do not know yet that the limit of the sequence of the norms of the residuals is 0; we only know that it exists. To prove weak convergence of the residuals to 0, we need the following lemma.

Lemma 2.3. *The sequence*

$$(\alpha_{n+1})_{n\in\mathbb{N}_0} = (\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}})_{n\in\mathbb{N}_0}$$

is square-summable.

Lemma 2.3 gives rise to the two following additional lemmas.

Lemma 2.4. *Since every square-summable sequence converges to zero, we have*

$$\lim_{n\to\infty} \alpha_{n+1} = \lim_{n\to\infty} \langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} = 0.$$

Lemma 2.5. *We have*

$$\lim_{n\to\infty} \langle r_n, Td\rangle_{\mathcal{Y}} = 0$$

for all $d \in \mathcal{D}$.

Proof. Because of (2.3), we obtain for every $d \in \mathcal{D}$ that

$$0 \le |\langle r_n, Td\rangle_{\mathcal{Y}}| \le \frac{1}{\varrho} |\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}}|.$$

Since $\varrho \in (0, 1]$ is fixed and the right-hand side tends to 0 for $n \to \infty$ by the preceding corollary, this proves our claim. □

Finally, we obtain weak convergence of the WFMP in the following theorems. For the sake of readability, we define $\mathcal{V} := \operatorname{span}\{Td \mid d \in \mathcal{D}\} \subseteq \mathcal{Y}$.

Theorem 2.6. *We have weak convergence of $(r_n)_{n\in\mathbb{N}_0}$ to zero in the space $\overline{\mathcal{V}}$, that is,*

$$\lim_{n\to\infty} \langle r_n, \overline{v}\rangle_{\mathcal{Y}} = 0$$

for all $\overline{v} \in \overline{\mathcal{V}}$.

Proof. By Lemma 2.5, we have $\lim_{n\to\infty}\langle r_n, v\rangle_{\mathcal{Y}} = 0$ for all $v \in \mathcal{V}$ due to the bilinearity of the inner product. Since $(r_n)_{n\in\mathbb{N}_0}$ is a bounded sequence, we obtain $\lim_{n\to\infty}\langle r_n, \overline{v}\rangle_{\mathcal{Y}} = 0$ for all $\overline{v} \in \overline{\mathcal{V}}$, too. □

Theorem 2.7. *Let the given data fulfill $g \in \overline{\mathcal{V}}$. Then $r_n \rightharpoonup 0$ in $\mathcal{Y}$ for $n \to \infty$.*

Proof. Let $z \in \mathcal{Y}$ be arbitrary. Since $\overline{\mathcal{V}}$ is a closed subspace of $\mathcal{Y}$, we obtain the decomposition $\mathcal{Y} = \overline{\mathcal{V}} \oplus \overline{\mathcal{V}}^{\perp}$, where $\overline{\mathcal{V}}^{\perp}$ denotes the orthogonal complement of $\overline{\mathcal{V}}$. Thus, there exist uniquely defined $z_\parallel \in \overline{\mathcal{V}}$, $z_\perp \in \overline{\mathcal{V}}^{\perp}$ such that $z = z_\parallel + z_\perp$. It follows that

$$\lim_{n\to\infty} \langle r_n, z\rangle_{\mathcal{Y}} = \lim_{n\to\infty} (\langle r_n, z_\parallel\rangle_{\mathcal{Y}} + \langle r_n, z_\perp\rangle_{\mathcal{Y}}).$$

Since $r_n = g - Tf_n \in \overline{\mathcal{V}}$ and $z_\perp \perp \overline{\mathcal{V}}$, the latter term vanishes, and

$$\lim_{n\to\infty} \langle r_n, z\rangle_{\mathcal{Y}} = \lim_{n\to\infty} \langle r_n, z_\parallel\rangle_{\mathcal{Y}} = 0$$

by Theorem 2.6 since $z_\parallel \in \overline{\mathcal{V}}$. □

Note that, in the case $g \in \overline{\mathcal{V}}$, we have proved so far that the sequence $(\|r_n\|_{\mathcal{Y}})_{n\in\mathbb{N}_0}$ is convergent and that $r_n \rightharpoonup 0$ $(n \to \infty)$ in $\mathcal{Y}$. Unfortunately, we cannot conclude convergence $r_n \to 0$ in $\mathcal{Y}$ in the strong sense from these facts since $\mathcal{Y}$ may be infinite-dimensional. This is different in the considerations in [4, 5], where it was assumed that $\dim \mathcal{Y} = \ell \in \mathbb{N}$. The next section is dedicated to the proof of strong convergence of the residuals, which requires a more complicated technique.

2.3 Strong convergence of the residuals

The following proofs are based on the technique introduced in [12] for *projection pursuit regression*, a variant of the WGA in statistics.

Lemma 2.8. *Let $(a_n)_{n\in\mathbb{N}_0}$ be a square-summable sequence, where $a_n \ge 0$ for all $n \in \mathbb{N}_0$. Then the identity*

$$\liminf_{n\to\infty} \left(a_n \sum_{k=1}^{n} a_k \right) = 0.$$

holds.

Proof. See [12, Lemma 2]. □

Corollary 2.9. *We have*

$$\liminf_{n\to\infty} |\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}}| \sum_{k=1}^{n} |\langle r_k, Td_{k+1}\rangle_{\mathcal{Y}}| = 0.$$

Proof. This follows from Lemma 2.3 in conjunction with Lemma 2.8. □

The preceding corollary is a crucial ingredient in the proof of the following theorem, where it is shown that the sequence of residuals converges strongly in $\mathcal{Y}$.

Theorem 2.10. *The sequence $(r_n)_{n\in\mathbb{N}_0}$ is a Cauchy sequence in $\mathcal{Y}$ and thus convergent.*

Proof. This proof is an extended version of a similar proof in [12, Section 2].

Assume that the sequence is not a Cauchy sequence in $\mathcal{Y}$. Then there exists $\varepsilon > 0$ such that, for all $N \in \mathbb{N}_0$, there exist $m, n \geq N$, which fulfill

$$\|r_m - r_n\|_{\mathcal{Y}} > \varepsilon. \tag{2.5}$$

Let $\gamma > 0$ be an arbitrary constant.

From Lemma 2.2 (b), we obtain the existence of $R := \lim_{k\to\infty}\|r_k\|_{\mathcal{Y}}$. Thus, there exists $N \in \mathbb{N}_0$ such that $\|r_N\|_{\mathcal{Y}}^2 < R^2 + \gamma$. Since $(\|r_k\|_{\mathcal{Y}})_{k\in\mathbb{N}_0}$ is monotonically decreasing due to Lemma 2.2 (a) and by (2.5), we obtain that there exist $m, n \geq N$, which fulfill

$$\begin{aligned} \|r_m - r_n\|_{\mathcal{Y}} &> \varepsilon, \\ \|r_m\|_{\mathcal{Y}}^2 &< R^2 + \gamma, \\ \|r_n\|_{\mathcal{Y}}^2 &< R^2 + \gamma. \end{aligned} \tag{2.6}$$

Furthermore, by Corollary 2.9, there exists $p > \max\{m, n\}$ such that

$$|\langle r_p, Td_{p+1}\rangle_{\mathcal{Y}}| \sum_{k=1}^{p} |\langle r_k, Td_{k+1}\rangle_{\mathcal{Y}}| < \gamma. \tag{2.7}$$

Since

$$\varepsilon < \|r_m - r_n\|_{\mathcal{Y}} \leq \|r_m - r_p\|_{\mathcal{Y}} + \|r_p - r_n\|_{\mathcal{Y}},$$

we have $\|r_m - r_p\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$ or $\|r_p - r_n\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$.

Without loss of generality, let $\|r_m - r_p\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$. We obtain

$$\begin{aligned} \|r_m - r_p\|_{\mathcal{Y}}^2 &= \|r_m\|_{\mathcal{Y}}^2 + \|r_p\|_{\mathcal{Y}}^2 - 2\langle r_m, r_p\rangle_{\mathcal{Y}} \\ &= \|r_m\|_{\mathcal{Y}}^2 + \|r_p\|_{\mathcal{Y}}^2 - 2\left\langle r_p + \sum_{k=m}^{p-1} \alpha_{k+1} Td_{k+1}, r_p \right\rangle_{\mathcal{Y}} \\ &\leq \|r_m\|_{\mathcal{Y}}^2 - \|r_p\|_{\mathcal{Y}}^2 + 2\sum_{k=m}^{p-1} |\alpha_{k+1}|\,|\langle Td_{k+1}, r_p\rangle_{\mathcal{Y}}| \\ &= \|r_m\|_{\mathcal{Y}}^2 - \|r_p\|_{\mathcal{Y}}^2 + 2\sum_{k=m}^{p-1} |\langle r_k, Td_{k+1}\rangle_{\mathcal{Y}}| \underbrace{|\langle r_p, Td_{k+1}\rangle_{\mathcal{Y}}|}_{\leq \frac{1}{\varrho}|\langle r_p, Td_{p+1}\rangle_{\mathcal{Y}}|} \\ &\leq \underbrace{\|r_m\|_{\mathcal{Y}}^2}_{\leq R^2+\gamma} \underbrace{- \|r_p\|_{\mathcal{Y}}^2}_{\leq -R^2} + \frac{2}{\varrho} \underbrace{|\langle r_p, Td_{p+1}\rangle_{\mathcal{Y}}| \sum_{k=m}^{p-1} |\langle r_k, Td_{k+1}\rangle_{\mathcal{Y}}|}_{<\gamma} \\ &\leq R^2 + \gamma - R^2 + \frac{2}{\varrho}\gamma = \left(1 + \frac{2}{\varrho}\right)\gamma, \end{aligned}$$

where (2.4), (2.3), (2.6), Lemma 2.2 (a) and (2.7) have been used in this order.

Since $\gamma > 0$ was arbitrary, one can choose γ small enough such that $(1 + \frac{2}{\varrho})\gamma < \frac{\varepsilon^2}{4}$, which yields a contradiction to $\|r_m - r_p\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$. □

Since $\mathcal{Y}$ is a Hilbert space, the sequence $(r_n)_{n\in\mathbb{N}_0}$ converges in $\mathcal{Y}$ in the strong sense. In the following, we will prove several properties of the limit of this sequence.

Theorem 2.11. *For $r_\infty := \lim_{n\to\infty} r_n$, we have $r_\infty \perp \overline{\mathcal{V}}$.*

Proof. Let $\overline{v} \in \overline{\mathcal{V}}$. Since the inner product is a continuous function of its arguments due to the Cauchy–Schwarz inequality, we may interchange the limit and the inner product to obtain

$$\langle r_\infty, \overline{v}\rangle_{\mathcal{Y}} = \langle \lim_{n\to\infty} r_n, \overline{v}\rangle_{\mathcal{Y}} = \lim_{n\to\infty} \langle r_n, \overline{v}\rangle_{\mathcal{Y}} = 0$$

by Theorem 2.6. □

Corollary 2.12. *If $g \in \overline{\mathcal{V}}$, we have $r_\infty = 0$ due to the fact that the weak limit is 0 according to Theorem 2.7, the strong limit has been proven to exist in Theorem 2.10 and the latter has to coincide with the weak limit.*

2.4 Convergence in the domain

So far, we have only considered the convergence of the WFMP in the image space $\mathcal{Y}$ of the operator T. To achieve the convergence also in the domain $\mathcal{X}$ of the operator, we can adopt the proof of the analogue statement for the FMP, which has first been stated in [4]. We will adhere to the improved version of the proof which has recently been given in [18]. Due to the normalization of the dictionary, we can omit the second condition that is required in the latter reference.

Theorem 2.13. *Let the assumptions of Algorithm 2.1 be fulfilled. Furthermore, let the dictionary $\mathcal{D} \subseteq \mathcal{X}$ fulfill the following condition:*

Semi-frame condition (SFC). *There exists a constant $c > 0$ and an integer $M \in \mathbb{N}$ such that, for all expansions $H = \sum_{k=1}^\infty \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k are not necessarily pairwise distinct, but $\#\{j \in \mathbb{N} \mid d_j = d_k\} \leq M$ for all $k \in \mathbb{N}$, the inequality*

$$c\|H\|^2_{\mathcal{X}} \leq \sum_{k=1}^{\infty} \beta_k^2$$

is valid.

If the sequence $(f_n)_{n\in\mathbb{N}}$ is produced by the WFMP and no dictionary element is chosen more than M times, then $(f_n)_n$ converges in $\mathcal{X}$ to $f_\infty := \sum_{n=1}^\infty \alpha_n d_n$.

Proof. From Lemma 2.3, we obtain that $(\alpha_n)_{n\in\mathbb{N}}$ is square-summable. The latter and the semi-frame condition (SFC) give rise to the convergence of the series $\sum_{n=1}^\infty \alpha_n d_n$ in the strong sense in $\mathcal{X}$, and hence, $f_\infty \in \mathcal{X}$ as defined above exists. It is also clear that $f_\infty = \lim_{N\to\infty} f_N$ holds in the sense of $\mathcal{X}$ since

$$\lim_{N\to\infty} \|f_\infty - f_N\|^2_{\mathcal{X}} = \lim_{N\to\infty} \left\| \sum_{n=N+1}^{\infty} \alpha_n d_n \right\|^2_{\mathcal{X}} \leq \frac{1}{c} \lim_{N\to\infty} \sum_{n=N+1}^{\infty} \alpha_n^2 = 0,$$

due to the square-summability of the sequence of coefficients (Lemma 2.3). □

Corollary 2.14. *Now that we know that $(f_n)_{n\in\mathbb{N}_0}$ converges, it follows from Theorem 2.11 that $Tf_\infty = P_{\overline{\mathcal{V}}} g$, where $P_{\overline{\mathcal{V}}}$ denotes the orthogonal projection in $\mathcal{Y}$ onto $\overline{\mathcal{V}}$. In other words, Tf_∞ is the best approximation of g in $\overline{\mathcal{V}}$.*

After showing that the sequence of iterates $(f_n)_{n\in\mathbb{N}_0}$ converges, we will prove that the limit satisfies the associated normal equation (2.1).

Theorem 2.15. *If the assumptions of Theorem 2.13 are fulfilled and additionally $\overline{\operatorname{span} \mathcal{D}} = (\ker T)^\perp$ holds, then the limit f_∞ satisfies the normal equation*

$$T^* T f_\infty = T^* g. \tag{2.8}$$

Proof. Since the sequences $(r_n)_{n\in\mathbb{N}_0} = (g - Tf_n)_{n\in\mathbb{N}_0}$ and $(f_n)_{n\in\mathbb{N}_0}$ both converge and T is continuous, we have $r_\infty = g - Tf_\infty$.

By Theorem 2.11, we have $r_\infty \perp \overline{\mathcal{V}}$. Hence,

$$g - Tf_\infty \perp \overline{\operatorname{span}\{Td \mid d \in \mathcal{D}\}}$$

such that, for all $d \in \mathcal{D}$, we have

$$0 = \langle g - Tf_\infty, Td \rangle_{\mathcal{Y}} = \langle T^*(g - Tf_\infty), d \rangle_{\mathcal{X}}.$$

Furthermore, for all $d \in \ker T$, we also obtain

$$\langle T^*(g - Tf_\infty), d \rangle_{\mathcal{X}} = \langle g - Tf_\infty, Td \rangle_{\mathcal{X}} = \langle g - Tf_\infty, 0 \rangle_{\mathcal{X}} = 0.$$

Since $\operatorname{span} \mathcal{D} \oplus \ker T$ is dense in $\mathcal{X}$ and the inner product is a non-degenerate bilinear form, we obtain

$$T^* Tf_\infty = T^* g$$

as desired. □

Remark. We make the following three observations:
(a) We have $\overline{\mathcal{V}} = \overline{\operatorname{ran} T} = \operatorname{ran} T$ due to the condition $\overline{\operatorname{span} \mathcal{D}} = (\ker T)^\perp$ and the well-posedness of the inverse problem.
(b) Thus, as remarked earlier, $r_\infty = 0$ if $g \in \operatorname{ran} T$. From Theorem 2.13, we obtain that the limit $f_\infty \in \mathcal{X}$ exists, and in the preceding proof, we already employed that $r_\infty = g - Tf_\infty$ due to the continuity of T. Thus, f_∞ is a solution of the inverse problem (1.1) since $Tf_\infty = g$.
(c) It is well known that the solution of the normal equation (2.8) is also a least-squares solution of the inverse problem [3, Theorem 2.6]. That is, in our case,

$$\|g - Tf_\infty\|_{\mathcal{Y}} = \min_{f \in \mathcal{X}} \|g - Tf\|_{\mathcal{Y}}.$$

This means that the WFMP (as well as the FMP) has an interpretation as a minimization algorithm for the optimization problem

$$\|g - Tf\|_{\mathcal{Y}} \to \min! \quad \text{subject to} \quad f \in \mathcal{X}.$$

As a matter of fact, the FMP was originally motivated as an iterative minimization of exactly the functional $f \mapsto \|g - Tf\|^2_{\mathcal{Y}}$, and so things have come full circle.

2.5 Convergence rates

In analogy to the convergence result for the WGA in [27, Theorem 5.1], we can prove a convergence rate of the WFMP in the data space in the following.

We first state the following lemma, which is an analogy to [2, Lemma 3.4] and [27, Lemma 3.1].

Lemma 2.16. *Let $(a_n)_{n \in \mathbb{N}_0}$ be a sequence of non-negative numbers, which satisfies*

$$a_0 \leq 1, \quad a_{n+1} \leq a_n(1 - \varrho^2 a_n) \quad \text{for all } n \geq 0.$$

Then

$$a_n \leq \frac{1}{1 + n\varrho^2} \quad \text{for all } n \in \mathbb{N}_0. \tag{2.9}$$

Proof. We prove the claim by induction on n. For $n = 0$ the statement is obvious.

Assuming that the inequality in (2.9) is true for some $n \in \mathbb{N}_0$, we have to show that

$$a_{n+1} \leq \frac{1}{1 + (n+1)\varrho^2}.$$

The latter is clearly true if $a_{n+1} = 0$. If $a_{n+1} > 0$, then also $a_n > 0$, and we obtain

$$\begin{aligned} a_{n+1} \leq a_n(1 - \varrho^2 a_n) \leq a_n \frac{1}{1 + \varrho^2 a_n} &= \frac{1}{\frac{1}{a_n} + \varrho^2} \\ \leq \frac{1}{(1 + n\varrho^2) + \varrho^2} &= \frac{1}{1 + (n+1)\varrho^2} \end{aligned}$$

since $(1 - x) \leq (1 + x)^{-1}$ for $x \geq 0$. □

Next, we define the following norm on $\overline{\mathcal{V}}$.

Definition 2.17. Given a linear and bounded operator $T\colon \mathcal{X} \to \mathcal{Y}$ and a dictionary $\mathcal{D}$, for $z \in \overline{\mathcal{V}}$, we define

$$|z|_{T\mathcal{D}} := \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; z = \sum_{k=1}^{\infty} \bar{\beta}_k T\bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\}, \tag{2.10}$$

where the limit of the series is considered in the sense of $\mathcal{Y}$.

Furthermore, we define the set

$$\overline{\overline{\mathcal{V}}} := \{z \in \overline{\mathcal{V}} \mid |z|_{T\mathcal{D}} < \infty\}.$$

In general, an element $z \in \overline{\mathcal{V}}$ may be represented as a linear combination of images of dictionary elements in several different ways since the dictionary itself and its image do not need to form a basis of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Thus, the quantity $|z|_{T\mathcal{D}}$ can be read as a measure of how sparse the element z can be expressed as a linear combination of images of dictionary elements.

However, note that $|z|_{T\mathcal{D}}$ does not need to be finite. Even if $T\mathcal{D}$ was an orthonormal basis in $\mathcal{Y}$, it is not natural that the Fourier coefficients of some element z are absolutely summable.

Lemma 2.18. *Let $z \in \overline{\mathcal{V}}$. Then we have $\|z\|_{\mathcal{Y}} \le |z|_{T\mathcal{D}}$.*

Proof. If $z \in \overline{\mathcal{V}} \setminus \overline{\overline{\mathcal{V}}}$, then the inequality is clear since the right-hand side is infinite.

If $z \in \overline{\overline{\mathcal{V}}}$, then for all $\varepsilon > 0$, there exist $\bar{\beta}_k \in \mathbb{R}$ and $\bar{d}_k \in \mathcal{D}$ such that

$$z = \sum_{k=1}^{\infty} \bar{\beta}_k T\bar{d}_k \quad \text{and} \quad \sum_{k=1}^{\infty} |\bar{\beta}_k| \le |z|_{T\mathcal{D}} + \varepsilon.$$

Thus,

$$\|z\|_{\mathcal{Y}} = \left\| \sum_{k=1}^{\infty} \bar{\beta}_k T\bar{d}_k \right\|_{\mathcal{Y}} \le \sum_{k=1}^{\infty} |\bar{\beta}_k| \|T\bar{d}_k\|_{\mathcal{Y}} = \sum_{k=1}^{\infty} |\bar{\beta}_k| \le |z|_{T\mathcal{D}} + \varepsilon.$$

Since ε was chosen arbitrarily, this proves the inequality. □

Theorem 2.19. *The mapping $z \mapsto |z|_{T\mathcal{D}}$ is a norm on $\overline{\overline{\mathcal{V}}}$.*

Proof. We show the axioms for a norm one by one.

Definiteness. On the one hand, it is obvious that $|0|_{T\mathcal{D}} = 0$ since the zero sequence is a feasible choice in (2.10). Assume on the other hand that $|z|_{T\mathcal{D}} = 0$ for $z \in \overline{\overline{\mathcal{V}}}$. Then by Lemma 2.18, we obtain

$$0 \le \|z\|_{\mathcal{Y}} \le |z|_{T\mathcal{D}} = 0,$$

thus $z = 0$.

Absolute homogeneity. We have to prove that $|\lambda z|_{T\mathcal{D}} = |\lambda|\,|z|_{T\mathcal{D}}$ for $z \in \overline{\overline{\mathcal{V}}}$ and $\lambda \in \mathbb{R}$. For $\lambda = 0$, the absolute homogeneity follows from the definiteness of the norm. Let $\lambda \ne 0$ and $z = \sum_{k=1}^{\infty} \bar{\beta}_k T\bar{d}_k \in \overline{\overline{\mathcal{V}}}$. Thus, we have $\lambda z = \sum_{k=1}^{\infty} (\lambda\bar{\beta}_k) T\bar{d}_k$ such that $|\lambda z|_{T\mathcal{D}} \le |\lambda|\,|z|_{T\mathcal{D}}$ is guaranteed. It remains to show that there are no $\hat{\beta}_k$, $\hat{d}_k$ such that $\lambda z = \sum_{k=1}^{\infty} \hat{\beta}_k T\hat{d}_k$ and $\sum_{k=1}^{\infty} |\hat{\beta}_k| < |\lambda|\,|z|_{T\mathcal{D}}$. Assume the contrary, then consequently,

$$z = \sum_{k=1}^{\infty} \frac{\hat{\beta}_k}{\lambda} T\hat{d}_k \quad \text{and} \quad \sum_{k=1}^{\infty} \left| \frac{\hat{\beta}_k}{\lambda} \right| = \frac{1}{|\lambda|} \sum_{k=1}^{\infty} |\hat{\beta}_k| < |z|_{T\mathcal{D}},$$

which is a contradiction to the definition of $|z|_{T\mathcal{D}}$.

Triangle inequality. Let $z, w \in \overline{\overline{\mathcal{V}}}$ and $\varepsilon > 0$. Due to the definition of the norm, there exist $\bar{\beta}_k, \hat{\beta}_k \in \mathbb{R}$ and $\bar{d}_k, \hat{d}_k \in \mathcal{D}$ such that

$$z = \sum_{k=1}^{\infty} \bar{\beta}_k T\bar{d}_k, \quad \sum_{k=1}^{\infty} |\bar{\beta}_k| \le |z|_{T\mathcal{D}} + \frac{\varepsilon}{2},$$

$$w = \sum_{k=1}^{\infty} \hat{\beta}_k T\hat{d}_k, \quad \sum_{k=1}^{\infty} |\hat{\beta}_k| \le |w|_{T\mathcal{D}} + \frac{\varepsilon}{2}.$$

Set

$$\gamma_k := \begin{cases} \bar{\beta}_{k/2} & \text{for even } k, \\ \hat{\beta}_{(k+1)/2} & \text{else,} \end{cases} \qquad f_k := \begin{cases} \bar{d}_{k/2} & \text{for even } k, \\ \hat{d}_{(k+1)/2} & \text{else.} \end{cases}$$

Note that there is no problem with rearranging the series since the series is absolutely convergent (see the proof of Lemma 2.18) and thus unconditionally convergent since $\overline{\mathcal{V}}$ is complete with respect to $\|\cdot\|_{\mathcal{Y}}$ (cf. [13, Chapter 1, Section 3]). Consequently, $z + w = \sum_{k=1}^{\infty} \gamma_k T f_k$ and

$$\sum_{k=1}^{\infty} |\gamma_k| \leq \sum_{k=1}^{\infty} |\bar{\beta}_k| + \sum_{k=1}^{\infty} |\hat{\beta}_k| \leq |z|_{T\mathcal{D}} + |w|_{T\mathcal{D}} + \varepsilon.$$

Thus,

$$|z + w|_{T\mathcal{D}} \leq \sum_{k=1}^{\infty} |\gamma_k| \leq |z|_{T\mathcal{D}} + |w|_{T\mathcal{D}}$$

by the definition of the norm and the fact that ε was chosen arbitrarily. □

Theorem 2.20. *Let $(r_n)_{n \in \mathbb{N}_0}$ be the sequence of residuals generated by the WFMP, and let $g \in \overline{\mathcal{V}}$. Then*

$$\|r_n\|_{\mathcal{Y}} \leq |g|_{T\mathcal{D}} (1 + n\varrho^2)^{-\varrho/(4+2\varrho)}. \tag{2.11}$$

Proof. If $g \in \overline{\mathcal{V}} \setminus \hat{\overline{\mathcal{V}}}$, then the inequality is clear since the right-hand side is infinite. If, furthermore, $|g|_{T\mathcal{D}} = 0$, then both sides of the inequality are zero since $\|r_n\|_{\mathcal{Y}} = \|g\|_{\mathcal{Y}} = 0$.

Let $g \in \hat{\overline{\mathcal{V}}}$ and $|g|_{T\mathcal{D}} > 0$. For $n \in \mathbb{N}_0$, define the sequence

$$b_n := |g|_{T\mathcal{D}} + \sum_{k=1}^{n} |\alpha_k|.$$

Then we obtain

$$\begin{aligned} |r_n|_{T\mathcal{D}} &= |g - T f_n|_{T\mathcal{D}} = \left| g - \sum_{k=1}^{n} \alpha_k T d_k \right|_{T\mathcal{D}} \\ &\leq |g|_{T\mathcal{D}} + \sum_{k=1}^{n} |\alpha_k T d_k|_{T\mathcal{D}} \leq |g|_{T\mathcal{D}} + \sum_{k=1}^{n} |\alpha_k| = b_n \end{aligned}$$

by using the triangle inequality, the absolute homogeneity of the norm and the fact that $|Td_j|_{T\mathcal{D}} \leq 1$.

For arbitrary $\varepsilon > 0$, there exist $\bar{\beta}_k \in \mathbb{R}$ and $\bar{d}_k \in \mathcal{D}$ such that $r_n = \sum_{k=1}^{\infty} \bar{\beta}_k T \bar{d}_k$ and $\sum_{k=1}^{\infty} |\bar{\beta}_k| \leq b_n + \varepsilon$. Thus,

$$\begin{aligned} \|r_n\|_{\mathcal{Y}}^2 &= |\langle r_n, r_n \rangle_{\mathcal{Y}}| = \left| \left\langle r_n, \sum_{k=1}^{\infty} \bar{\beta}_k T \bar{d}_k \right\rangle_{\mathcal{Y}} \right| \\ &\leq \sum_{k=1}^{\infty} |\bar{\beta}_k| \, |\langle r_n, T\bar{d}_k \rangle_{\mathcal{Y}}| \leq (b_n + \varepsilon) \sup_{d \in \mathcal{D}} |\langle r_n, Td \rangle_{\mathcal{Y}}|, \end{aligned}$$

and since ε was arbitrary, this yields

$$\sup_{d \in \mathcal{D}} |\langle r_n, Td \rangle_{\mathcal{Y}}| \geq \frac{\|r_n\|_{\mathcal{Y}}^2}{b_n}$$

and

$$|\alpha_{n+1}| \geq \frac{\varrho \|r_n\|_{\mathcal{Y}}^2}{b_n}$$

by the definition of α_{n+1} in (2.4).

Since $\|r_{n+1}\|_{\mathcal{Y}}^2 = \|r_n\|_{\mathcal{Y}}^2 - |\alpha_{n+1}|^2$, we obtain, on the one hand, that

$$\|r_{n+1}\|_{\mathcal{Y}}^2 \leq \|r_n\|_{\mathcal{Y}}^2 - \frac{\varrho^2 \|r_n\|_{\mathcal{Y}}^4}{b_n^2} = \|r_n\|_{\mathcal{Y}}^2 \left(1 - \frac{\varrho^2 \|r_n\|_{\mathcal{Y}}^2}{b_n^2}\right) \tag{2.12}$$

and, on the other hand,

$$\|r_{n+1}\|_{\mathcal{Y}}^2 \leq \|r_n\|_{\mathcal{Y}}^2 - |\alpha_{n+1}| \frac{\varrho \|r_n\|_{\mathcal{Y}}^2}{b_n} = \|r_n\|_{\mathcal{Y}}^2 \left(1 - \frac{\varrho |\alpha_{n+1}|}{b_n}\right). \tag{2.13}$$

Since $(b_n)_{n\in\mathbb{N}_0}$ is monotonically increasing, (2.12) gives

$$\frac{\|r_{n+1}\|_{\mathcal{Y}}^2}{b_{n+1}^2} \leq \frac{\|r_n\|_{\mathcal{Y}}^2}{b_n^2}\Big(1 - \frac{\varrho^2\|r_n\|_{\mathcal{Y}}^2}{b_n^2}\Big).$$

The application of Lemma 2.16 with $a_n := \frac{\|r_n\|_{\mathcal{Y}}^2}{b_n^2}$ (note that $a_0 = \frac{\|r_0\|_{\mathcal{Y}}^2}{b_0^2} = \frac{\|g\|_{\mathcal{Y}}^2}{b_0^2} \leq \frac{|g|_{\mathcal{TD}}^2}{|g|_{\mathcal{TD}}^2} = 1$) yields

$$\frac{\|r_n\|_{\mathcal{Y}}^2}{b_n^2} \leq (1 + n\varrho^2)^{-1}. \tag{2.14}$$

From the inequality in (2.13) and the fact that $b_{n+1} = b_n(1 + \frac{|\alpha_{n+1}|}{b_n})$ combined with the generalized Bernoulli inequality [21, Section 2.4]

$$(1 + x)^\alpha \leq 1 + \alpha x, \quad 0 \leq \alpha \leq 1,\ x \geq 0,$$

we obtain

$$\begin{aligned}
\|r_{n+1}\|_{\mathcal{Y}}^2 b_{n+1}^\varrho &= \|r_{n+1}\|_{\mathcal{Y}}^2 b_n^\varrho\Big(1 + \frac{|\alpha_{n+1}|}{b_n}\Big)^\varrho \\
&\leq \|r_{n+1}\|_{\mathcal{Y}}^2 b_n^\varrho\Big(1 + \varrho\frac{|\alpha_{n+1}|}{b_n}\Big) \\
&\leq \|r_n\|_{\mathcal{Y}}^2\Big(1 - \varrho\frac{|\alpha_{n+1}|}{b_n}\Big)b_n^\varrho\Big(1 + \varrho\frac{|\alpha_{n+1}|}{b_n}\Big) \\
&\leq \|r_n\|_{\mathcal{Y}}^2\Big(1 - \varrho^2\frac{|\alpha_{n+1}|^2}{b_n^2}\Big)b_n^\varrho \\
&\leq \|r_n\|_{\mathcal{Y}}^2 b_n^\varrho.
\end{aligned}$$

Thus, by induction, the following sequence of inequalities holds true for all $n \in \mathbb{N}_0$:

$$\|r_{n+1}\|_{\mathcal{Y}}^2 b_{n+1}^\varrho \leq \|r_n\|_{\mathcal{Y}}^2 b_n^\varrho \leq \cdots \leq \|r_1\|_{\mathcal{Y}}^2 b_1^\varrho \leq \|r_0\|_{\mathcal{Y}}^2 b_0^\varrho = \|g\|_{\mathcal{Y}}^2 |g|_{\mathcal{TD}}^\varrho \leq |g|_{\mathcal{TD}}^{2+\varrho},$$

where the last inequality is true due to Lemma 2.18.

Thus, using (2.14), we obtain

$$\|r_n\|_{\mathcal{Y}}^{4+2\varrho} = \|r_n\|_{\mathcal{Y}}^{2\varrho}\|r_n\|_{\mathcal{Y}}^4 \leq \|r_n\|_{\mathcal{Y}}^{2\varrho} b_n^{-2\varrho}|g|_{\mathcal{TD}}^{4+2\varrho} \leq |g|_{\mathcal{TD}}^{4+2\varrho}(1 + n\varrho^2)^{-\varrho},$$

which implies (2.11). □

As already mentioned, the constant $|g|_{\mathcal{TD}}$ in the rate of convergence is connected to how the dictionary matches the data, which is plausible. Also, it has already been laid out that the constant does not need to be finite such that the inequality (2.11) is meaningless if it is infinite. Nevertheless, the proof of convergence of the algorithm did not need the finiteness of $|g|_{\mathcal{TD}}$ such that only the rate of convergence depends on that property.

Corollary 2.21. *If it is not known whether $g \in \overline{\mathcal{V}}$, one still obtains*

$$\|r_n - r_\infty\|_{\mathcal{Y}} = \|P_{\overline{\mathcal{V}}}g - Tf_n\|_{\mathcal{Y}} \leq |P_{\overline{\mathcal{V}}}g|_{\mathcal{TD}}(1 + n\varrho^2)^{-\varrho/(4+2\varrho)},$$

where the constant $|P_{\overline{\mathcal{V}}}g|_{\mathcal{TD}}$ can alternatively be replaced by $|Tf_\infty|_{\mathcal{TD}}$ due to Corollary 2.14.

Proof. First, notice that

$$r_n - r_\infty = g - Tf_n - (g - Tf_\infty) = Tf_\infty - Tf_n = P_{\overline{\mathcal{V}}}g - Tf_n,$$

holds, where the last equality is due to Corollary 2.14. Thus, for arbitrary $d \in \mathcal{D}$, we have

$$\langle r_n - r_\infty, Td\rangle_{\mathcal{Y}} = \langle P_{\overline{\mathcal{V}}}g - Tf_n, Td\rangle_{\mathcal{Y}}.$$

Application of Theorem 2.11 yields $\langle r_\infty, Td\rangle_{\mathcal{Y}} = 0$ such that

$$\langle g - Tf_n, Td\rangle_{\mathcal{Y}} = \langle P_{\overline{\mathcal{V}}}g - Tf_n, Td\rangle_{\mathcal{Y}}. \tag{2.15}$$

We will now prove that the iterates $(f_n)_{n\in\mathbb{N}_0}$ provided by the WFMP when it is applied to the inverse problem

$$Tf = g \tag{2.16}$$

may also be generated by the WFMP applied to the inverse problem

$$Tf = P_{\overline{V}}g. \tag{2.17}$$

For this purpose, consider the characterization of d_{n+1} in (2.3) for both inverse problems. It turns out that, due to (2.15), every choice of d_{n+1} for the inverse problem (2.16) is also a valid choice for the inverse problem (2.17) and vice versa. Furthermore, the definition of α_{n+1} in (2.4) yields the same result in both settings if the same element d_{n+1} from the dictionary is chosen.

Thus, the sequence $(f_n)_{n\in\mathbb{N}_0}$ could also be generated by the WFMP when it is applied to (2.17). Since $P_{\overline{V}}g \in \overline{V}$, we can apply Theorem 2.20 and obtain the desired result. □

Remark. Note that, in general, the iterates of the WFMP for the inverse problems (2.16) and (2.17) are not identical due to the non-uniqueness of the choice of d_{n+1}. Nevertheless, the proof shows that the convergence rate is the same for both problems.

In consequence, for the FMP, that is, the case $\varrho = 1$, we obtain the following convergence rate.

Corollary 2.22. *For the sequence $(r_n)_{n\in\mathbb{N}_0}$ generated by the FMP and its limit r_∞, we have*

$$\|r_n - r_\infty\|_{\mathcal{Y}} = \|P_{\overline{V}}g - Tf_n\|_{\mathcal{Y}} \le |P_{\overline{V}}g|_{T\mathcal{D}}(1+n)^{-1/6}.$$

3 The Regularized Weak Functional Matching Pursuit (RWFMP)

Many inverse problems are ill-posed, that is, $\overline{\operatorname{ran} T} \neq \operatorname{ran} T$. In consequence, the WFMP cannot be applied to the inverse problem since the convergence was only proved for well-posed inverse problems in Section 2. Thus, a regularization technique has to be applied.

3.1 The algorithm

In [4, 5, 17], a regularization of the inverse problem was achieved by adding a penalty term, which is equivalent to the application of a Tikhonov regularization, which yields the so-called Regularized Functional Matching Pursuit (RFMP) algorithm. As for the FMP, we apply the strategy of the WGA to the RFMP to obtain the following algorithm.

Algorithm 3.1 (Regularized Weak Functional Matching Pursuit, RWFMP). Let $\mathcal{X}$, $\mathcal{Y}$, T be given as for problem (1.1). Furthermore, let data $g \in \mathcal{Y}$, a weakness parameter $\varrho \in (0,1]$, a regularization parameter $\lambda > 0$, and the initial approximation $f_0 = 0 \in \mathcal{X}$ be given. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$, whose elements $d \in \mathcal{D}$ satisfy $\|Td\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2 = 1$.

(i) Set $n := 0$, define the residual $r_0 := g - Tf_0 = g$ and choose a stopping criterion.

(ii) Find an element $d_{n+1} \in \mathcal{D}$, which fulfills

$$|\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} - \lambda\langle f_n, d_{n+1}\rangle_{\mathcal{X}}| \ge \varrho \sup_{d\in\mathcal{D}} |\langle r_n, Td\rangle_{\mathcal{Y}} - \lambda\langle f_n, d\rangle_{\mathcal{X}}|. \tag{3.1}$$

Set

$$\alpha_{n+1} := \langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} - \lambda\langle f_n, d_{n+1}\rangle_{\mathcal{X}}, \tag{3.2}$$

as well as $f_{n+1} := f_n + \alpha_{n+1}d_{n+1}$ and $r_{n+1} := g - Tf_{n+1} = r_n - \alpha_{n+1}Td_{n+1}$.

(iii) If the stopping criterion is fulfilled, then f_{n+1} is the output. Otherwise, increase n by 1 and return to step (ii).

Similarly to the non-regularized case, the RWFMP algorithm coincides with the original RFMP if $\varrho = 1$, up to the normalization of the dictionary.

It now comes into play that we were able to show convergence of the WFMP for arbitrary (especially infinite-dimensional) Hilbert spaces $\mathcal{Y}$. The following strategy could not be pursued in the previous setting, where $\mathcal{Y} = \mathbb{R}^\ell$, or at least $\dim \mathcal{Y} < \infty$, was required. We will give an interpretation of the RWFMP as WFMP for a modified well-posed inverse problem

$$\tilde{T}_\lambda f = \tilde{g}, \tag{3.3}$$

where $\tilde{T}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$. Let us first equip the space $\mathcal{Y} \times \mathcal{X}$ with an inner product to obtain a Hilbert space.

Lemma 3.2. *Let* $(\mathcal{X}, \langle \cdot, \cdot \rangle_\mathcal{X})$ *and* $(\mathcal{Y}, \langle \cdot, \cdot \rangle_\mathcal{Y})$ *be Hilbert spaces. Then*

$$\left\langle \begin{pmatrix} g_1 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_2 \\ f_2 \end{pmatrix} \right\rangle_{\mathcal{Y}\times\mathcal{X}} := \langle g_1, g_2 \rangle_\mathcal{Y} + \langle f_1, f_2 \rangle_\mathcal{X}, \quad \begin{pmatrix} g_1 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_2 \\ f_2 \end{pmatrix} \in \mathcal{Y} \times \mathcal{X},$$

defines an inner product on $\mathcal{Y} \times \mathcal{X}$ *and that space is complete with respect to the given inner product.*

Furthermore, the associated norm is given by

$$\left\| \begin{pmatrix} g \\ f \end{pmatrix} \right\|_{\mathcal{Y}\times\mathcal{X}} := \sqrt{\|g\|_\mathcal{Y}^2 + \|f\|_\mathcal{X}^2}, \quad (g, f) \in \mathcal{Y} \times \mathcal{X}.$$

Using this topology on the space $\mathcal{Y} \times \mathcal{X}$, we obtain the following lemma.

Lemma 3.3. *When using the same data and weakness parameter, the RWFMP in Algorithm 3.1 produces iterates* $(f_n)_{n\in\mathbb{N}_0}$*, which are also valid iterates generated by the WFMP in Algorithm 2.1 if the latter is applied to the inverse problem* $\tilde{T}_\lambda f = \tilde{g}$*, where* $\tilde{T}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$, $\tilde{T}_\lambda f := (Tf, \sqrt{\lambda} f)^\mathrm{T}$ *and* $\tilde{g} := (g, 0)^\mathrm{T} \in \mathcal{Y} \times \mathcal{X}$.

Proof. Since all input parameters are the same, it remains to show that one obtains (3.1) and (3.2) if one inserts $\tilde{T}_\lambda$ and $\tilde{g}$ into (2.3) and (2.4), respectively.

Fortunately, this can easily be seen since, for $d \in \mathcal{D}$, we have

$$\begin{aligned} \langle \tilde{r}_n, \tilde{T}_\lambda d \rangle_{\mathcal{Y}\times\mathcal{X}} &= \langle \tilde{g} - \tilde{T}_\lambda f_n, \tilde{T}_\lambda d \rangle_{\mathcal{Y}\times\mathcal{X}} \\ &= \left\langle \begin{pmatrix} g - Tf_n \\ 0 - \sqrt{\lambda} f_n \end{pmatrix}, \begin{pmatrix} Td \\ \sqrt{\lambda} d \end{pmatrix} \right\rangle_{\mathcal{Y}\times\mathcal{X}} \\ &= \langle g - Tf_n, Td \rangle_\mathcal{Y} + \langle -\sqrt{\lambda} f_n, \sqrt{\lambda} d \rangle_\mathcal{X} \\ &= \langle r_n, Td \rangle_\mathcal{Y} - \lambda \langle f_n, d \rangle_\mathcal{X} \end{aligned}$$

and

$$\|\tilde{T}_\lambda d\|_{\mathcal{Y}\times\mathcal{X}}^2 = \left\| \begin{pmatrix} Td \\ \sqrt{\lambda} d \end{pmatrix} \right\|_{\mathcal{Y}\times\mathcal{X}}^2 = \|Td\|_\mathcal{Y}^2 + \|\sqrt{\lambda} d\|_\mathcal{X}^2 = \|Td\|_\mathcal{Y}^2 + \lambda \|d\|_\mathcal{X}^2.$$

Note that, in accordance to the other definitions, we used the notation $\tilde{r}_n := \tilde{g} - \tilde{T}_\lambda f_n$ and $r_n := g - Tf_n$ in the considerations above. □

Note that the remark that was stated after Corollary 2.21 is also true in this case. We cannot expect that the RWFMP for the original problem and the WFMP for the modified problem produce identical iterates. The important result is that the iterates of one algorithm fulfill the selection criterion of the other algorithm and could thus be chosen there.

In Section 2, we assumed well-posedness of the inverse problem. Since the idea of a regularization is to substitute an ill-posed problem by a related well-posed problem, it is well known that the minimization of the Tikhonov functional is well-posed. This can also be characterized in terms of the modified operator $\tilde{T}_\lambda$.

Lemma 3.4. *For the operator*

$$\tilde{T}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}, \quad \tilde{T}_\lambda f = \begin{pmatrix} Tf \\ \sqrt{\lambda} f \end{pmatrix},$$

we have $\overline{\operatorname{ran} \tilde{T}_\lambda} = \operatorname{ran} \tilde{T}_\lambda$ *such that the inverse problem* (3.3) *is well-posed in the sense of Nashed.*

Proof. Neglecting permutation of the components, for $\lambda > 0$, the set $\operatorname{ran} \tilde{T}_\lambda$ is the graph of the operator $\lambda^{-1/2} T$. The assertion follows since every continuous operator has a closed graph [25, Proposition 2.14]. □

3.2 Convergence results

By applying Lemma 3.3, the following results about the convergence of the RWFMP are direct consequences of the corresponding results for the WFMP. In the following, we always assume that the sequence $(f_n)_{n\in\mathbb{N}_0}$ is generated by the RWFMP.

Lemma 3.5. *The sequence* $(\|r_n\|^2_{\mathcal{Y}} + \lambda\|f_n\|^2_{\mathcal{X}})_{n\in\mathbb{N}_0}$ *is monotonically decreasing and convergent.*

Proof. Since $\|\tilde{r}_n\|^2_{\mathcal{Y}\times\mathcal{X}} = \|r_n\|^2_{\mathcal{Y}} + \lambda\|f_n\|^2_{\mathcal{X}}$, this is a consequence of Lemma 2.2. □

An immediate consequence of Theorem 2.10 is the following result.

Theorem 3.6. *The sequence* $(f_n)_{n\in\mathbb{N}_0}$ *converges in* $\mathcal{X}$.

Proof. By Theorem 2.10, we have that $(\tilde{r}_n)_{n\in\mathbb{N}_0}$ converges in $\mathcal{Y}\times\mathcal{X}$. Since

$$\tilde{r}_n = \begin{pmatrix} g - Tf_n \\ -\sqrt{\lambda}f_n \end{pmatrix},$$

we obtain the convergence of $(f_n)_{n\in\mathbb{N}_0}$ because $\lambda > 0$ and a vectorial sequence converges if and only if its components converge. □

Note that, in contrast to the WFMP, we do not need the semi-frame condition (SFC) to prove the convergence of the sequence of the approximations generated by the RWFMP in $\mathcal{X}$. This is desirable since the semi-frame condition cannot be easily verified for a given dictionary.

Theorem 3.7. *If* $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X}$, *the limit* f_∞ *fulfills the Tikhonov-regularized normal equation*

$$(T^\star T + \lambda\,\mathrm{id})f_\infty = T^\star g.$$

Proof. We will apply Theorem 2.15 here, which requires the assumptions of Theorem 2.13, but only for the implication of the convergence in the solution space $\mathcal{X}$. Since we have this property due to Theorem 3.6, we do not need the assumptions of Theorem 2.13 (including the semi-frame condition) here, either. First, observe that $\tilde{T}_\lambda\colon \mathcal{X}\to\mathcal{Y}\times\mathcal{X}$ is injective due to the identity operator in the second component (and since $\lambda > 0$). Hence, $\ker\tilde{T}_\lambda = \{0\}$ and $(\ker\tilde{T}_\lambda)^\perp = \mathcal{X}$. Thus, since $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X} = (\ker\tilde{T}_\lambda)^\perp$, we obtain from Theorem 2.15 that f_∞ fulfills

$$\tilde{T}_\lambda{}^\star\tilde{T}_\lambda f_\infty = \tilde{T}_\lambda{}^\star\tilde{g}. \tag{3.4}$$

The adjoint operator of $\tilde{T}_\lambda$ is given by

$$\tilde{T}_\lambda{}^\star\colon \mathcal{Y}\times\mathcal{X}\to\mathcal{X},\quad \tilde{T}_\lambda{}^\star\tilde{z} = T^\star z + \sqrt{\lambda}w,\quad \text{where } \tilde{z} = \begin{pmatrix} z \\ w \end{pmatrix}\in\mathcal{Y}\times\mathcal{X},$$

since, for $\tilde{z} = \binom{z}{w}\in\mathcal{Y}\times\mathcal{X}$ and $f\in\mathcal{X}$, we obtain

$$\begin{aligned}\langle\tilde{z},\tilde{T}_\lambda f\rangle_{\mathcal{Y}\times\mathcal{X}} &= \left\langle\begin{pmatrix} z \\ w \end{pmatrix},\begin{pmatrix} Tf \\ \sqrt{\lambda}f \end{pmatrix}\right\rangle_{\mathcal{Y}\times\mathcal{X}} = \langle z, Tf\rangle_{\mathcal{Y}} + \langle w,\sqrt{\lambda}f\rangle_{\mathcal{X}} \\ &= \langle T^\star z, f\rangle_{\mathcal{X}} + \langle\sqrt{\lambda}w, f\rangle_{\mathcal{X}} = \langle T^\star z + \sqrt{\lambda}w, f\rangle_{\mathcal{X}}.\end{aligned}$$

Finally, elaborating both sides of (3.4) as

$$\tilde{T}_\lambda{}^\star\tilde{T}_\lambda f_\infty = \tilde{T}_\lambda{}^\star\begin{pmatrix} Tf_\infty \\ \sqrt{\lambda}f_\infty \end{pmatrix} = T^\star Tf_\infty + \lambda f_\infty = (T^\star T + \lambda\,\mathrm{id})f_\infty$$

and

$$\tilde{T}_\lambda{}^\star\tilde{g} = \tilde{T}_\lambda{}^\star\begin{pmatrix} g \\ 0 \end{pmatrix} = T^\star g,$$

we have proved the assertion. □

Remark. From the theory of Tikhonov regularization, it is well known that a solution of the regularized normal equation is also the unique minimizer of the Tikhonov functional

$$f \mapsto \|g - Tf\|_{\mathcal{Y}}^2 + \lambda \|f\|_{\mathcal{X}}^2 \tag{3.5}$$

(see, e. g., [3, Theorem 5.1]). Thus, by Theorem 3.7, the limit f_∞ of the RWFMP iteration is the unique minimizer of (3.5). The RWFMP consequently has an interpretation of an iterative minimization algorithm for the Tikhonov functional. Actually, the RFMP was derived as such an algorithm in [4] such that this interpretation fits well.

From Theorem 2.20 and Corollary 2.21, we can also derive a convergence rate of the algorithm, measured in the Tikhonov functional. The following lemma states that the constant $|\tilde{T}_\lambda f_\infty|_{\tilde{T}_\lambda \mathcal{D}}$, which arises consequently, can be rewritten to represent how sparse the solution f_∞ can be expressed as a linear combination of dictionary elements.

Lemma 3.8. *Let $f \in \mathcal{X}$, then*

$$|\tilde{T}_\lambda f|_{\tilde{T}_\lambda \mathcal{D}} = |f|_{\mathcal{D}} := \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \,\middle|\, f = \sum_{k=1}^{\infty} \bar{\beta}_k \bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\}.$$

Proof. We have

$$\begin{aligned}
|\tilde{T}_\lambda f|_{\tilde{T}_\lambda \mathcal{D}} &= \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \,\middle|\, \tilde{T}_\lambda f = \sum_{k=1}^{\infty} \bar{\beta}_k \tilde{T}_\lambda \bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\} \\
&= \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \,\middle|\, \begin{pmatrix} Tf \\ \sqrt{\lambda} f \end{pmatrix} = \sum_{k=1}^{\infty} \bar{\beta}_k \begin{pmatrix} T\bar{d}_k \\ \sqrt{\lambda}\bar{d}_k \end{pmatrix},\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\} \\
&= \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \,\middle|\, \begin{pmatrix} Tf \\ f \end{pmatrix} = \sum_{k=1}^{\infty} \bar{\beta}_k \begin{pmatrix} T\bar{d}_k \\ \bar{d}_k \end{pmatrix},\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\} \\
&= \inf\left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \,\middle|\, f = \sum_{k=1}^{\infty} \bar{\beta}_k \bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\} = |f|_{\mathcal{D}},
\end{aligned}$$

where the second to last equality is true since the equation in the second component implies the equation in the first component. □

Consequently, we obtain the following convergence rate result.

Theorem 3.9. *For the sequence $(f_n)_{n \in \mathbb{N}_0}$ generated by the RWFMP, we have*

$$\|r_n - r_\infty\|_{\mathcal{Y}}^2 + \lambda \|f_n - f_\infty\|_{\mathcal{X}}^2 \le |f_\infty|_{\mathcal{D}}^2 (1 + n\varrho^2)^{-\varrho/(2+\varrho)}.$$

Proof. From Corollary 2.21, we obtain for the RWFMP that

$$\|\tilde{r}_n - \tilde{r}_\infty\|_{\mathcal{Y} \times \mathcal{X}} \le |\tilde{T}_\lambda f_\infty|_{\tilde{T}_\lambda \mathcal{D}} (1 + n\varrho^2)^{-\varrho/(4+2\varrho)},$$

and consequently,

$$\|r_n - r_\infty\|_{\mathcal{Y}}^2 + \lambda \|f_n - f_\infty\|_{\mathcal{X}}^2 \le |\tilde{T}_\lambda f_\infty|_{\tilde{T}_\lambda \mathcal{D}}^2 (1 + n\varrho^2)^{-\varrho/(2+\varrho)}.$$

The application of the preceding lemma proves the assertion. □

For the RFMP, we obtain the following convergence rate by setting $\varrho = 1$.

Corollary 3.10. *For the sequences $(r_n)_{n \in \mathbb{N}_0}$ and $(f_n)_{n \in \mathbb{N}_0}$ generated by the RFMP and their limits r_∞ and f_∞, we have*

$$\|r_n - r_\infty\|_{\mathcal{Y}}^2 + \lambda \|f_n - f_\infty\|_{\mathcal{X}}^2 \le |f_\infty|_{\mathcal{D}}^2 (1 + n)^{-1/3}.$$

Note that the right-hand side does not seem to depend on the operator T and the regularization parameter λ. However, this is not true since f_∞ is the minimizer of the Tikhonov functional corresponding to T and λ such that there is an implicit dependence on both of them.

3.3 The RWFMP as a convergent regularization method

This section is dedicated to the analysis of the RWFMP as an iterative regularization algorithm. We will present a-priori parameter choices for the Tikhonov regularization parameter λ and the number of iterations n after which the RWFMP is stopped in dependence on the noise level δ of the data. For optimal choices of λ and n, we obtain a convergence rate of $\delta^{2/3}$ for $\delta \to 0+$, which is the optimal rate for Tikhonov regularization, on which the RWFMP is based.

For this purpose, by $(f^\delta_{\lambda,n})_{n\in\mathbb{N}_0}$, we denote the sequence of iterates of the RWFMP when the algorithm is applied to the inverse problem $Tf = g^\delta$ using the regularization parameter $\lambda > 0$. Here, $y^\delta \in \mathcal{Y}$ denotes noisy data, which fulfill $\|g^\delta - g\|_\mathcal{Y} \le \delta$.

Let $f^\delta_{\lambda,\infty}$ be the minimizer of the Tikhonov functional, and let $f^+ = T^+ g$ be the best-approximate solution of $Tf = g$, where T^+ denotes the Moore–Penrose pseudo-inverse of T. Then the following theorem is a well-known result for Tikhonov regularization.

Theorem 3.11 (cf. [14, Theorem 2.12]). *If $f^+ = T^* Th$ for some $h \in \mathcal{X}$ with $\|h\|_\mathcal{X} \le \tau$ and $\lambda(\delta) = m_1(\delta/\tau)^{2/3}$ for some constant $m_1 > 0$, then*

$$\|f^\delta_{\lambda(\delta),\infty} - f^+\|_\mathcal{X} \le \left(\frac{1}{2\sqrt{m_1}} + m_1\right)\tau^{1/3}\delta^{2/3} = C_1\delta^{2/3}$$

for all $\delta > 0$, where $C_1 := (\frac{1}{2\sqrt{m_1}} + m_1)\tau^{1/3}$.

Furthermore, from Theorem 3.9, we immediately obtain the following corollary.

Corollary 3.12. *Let $(f^\delta_{\lambda,n})_{n\in\mathbb{N}_0}$ be the sequence of iterates of the RWFMP with weakness parameter $\varrho \in (0,1]$ when the algorithm is applied to the inverse problem $Tf = g^\delta$ using the regularization parameter $\lambda > 0$. Furthermore, assume $|f^\delta_{\lambda,\infty}|_\mathcal{D} < \infty$. Then, for all $n \in \mathbb{N}$, we have*

$$\|f^\delta_{\lambda,n} - f^\delta_{\lambda,\infty}\|_\mathcal{X} \le \lambda^{-1/2}|f^\delta_{\lambda,\infty}|_\mathcal{D}\varrho^{-\varrho/(2+\varrho)}n^{-\varrho/(4+2\varrho)} = C_2(\lambda)\lambda^{-1/2}n^{-\varrho/(4+2\varrho)},$$

where $C_2(\lambda) := |f^\delta_{\lambda,\infty}|_\mathcal{D}\varrho^{-\varrho/(2+\varrho)}$.

Proof. From Theorem 3.9, we can conclude that

$$\lambda\|f_{\lambda,n} - f_{\lambda,\infty}\|^2_\mathcal{X} \le |f^\delta_{\lambda,\infty}|^2_\mathcal{D}(1 + n\varrho^2)^{-\varrho/(2+\varrho)}$$

and thus

$$\|f_{\lambda,n} - f_{\lambda,\infty}\|_\mathcal{X} \le \lambda^{-1/2}|f^\delta_{\lambda,\infty}|_\mathcal{D}(1 + n\varrho^2)^{-\varrho/(4+2\varrho)}.$$

Since the exponent of the last term is negative, we obtain

$$\begin{aligned}\|f_{\lambda,n} - f_{\lambda,\infty}\|_\mathcal{X} &\le \lambda^{-1/2}|f^\delta_{\lambda,\infty}|_\mathcal{D}(n\varrho^2)^{-\varrho/(4+2\varrho)}\\ &= \lambda^{-1/2}|f^\delta_{\lambda,\infty}|_\mathcal{D}\varrho^{-\varrho/(2+\varrho)}n^{-\varrho/(4+2\varrho)},\end{aligned}$$

from which the desired estimate follows immediately. □

The preceding theorem and the corollary enable us to prove the following result.

Theorem 3.13. *Let $(f^\delta_{\lambda,n})_{n\in\mathbb{N}_0}$ be the sequence of iterations of the RWFMP with weakness parameter $\varrho \in (0,1]$ when the algorithm is applied to the inverse problem $Tf = g^\delta$ using the regularization parameter $\lambda > 0$. Furthermore, we assume that there exists $C_3 > 0$ such that $C_2(\lambda) \le C_3$ for all $\lambda > 0$. Additionally, let $f^+ = T^* Th$ for some $h \in \mathcal{X}$ with $\|h\|_\mathcal{X} \le \tau$ and $\lambda(\delta) = m_1(\delta/\tau)^{2/3}$ and $n(\delta) = m_2\delta^{-(4+2\varrho)/\varrho}$ for some constants $m_1, m_2 > 0$. Then there exists $C > 0$ such that*

$$\|f^\delta_{\lambda(\delta),n(\delta)} - f^+\|_\mathcal{X} \le C\delta^{2/3}$$

for all $\delta > 0$.

Proof. Let $\delta > 0$ be chosen arbitrarily. Since all conditions of Theorem 3.11 are fulfilled, we obtain

$$\|f^{\delta}_{\lambda(\delta),\infty} - f^{+}\|_{\mathcal{X}} \leq C_1 \delta^{2/3},$$

where $f^{\delta}_{\lambda(\delta),\infty}$ is the minimizer of the Tikhonov functional with regularization parameter $\lambda(\delta)$.

Corollary 3.12 and the assumption $C_2(\lambda) \leq C_3$ yield

$$\begin{aligned}\|f^{\delta}_{\lambda(\delta),n(\delta)} - f^{\delta}_{\lambda(\delta),\infty}\|_{\mathcal{X}} &\leq C_2(\lambda(\delta))(\lambda(\delta))^{-1/2}(n(\delta))^{-\varrho/(4+2\varrho)} \\ &\leq C_3(\lambda(\delta))^{-1/2}(n(\delta))^{-\varrho/(4+2\varrho)}.\end{aligned}$$

Inserting the definition of $\lambda(\delta)$ and $n(\delta)$ gives

$$\begin{aligned}\|f^{\delta}_{\lambda(\delta),n(\delta)} - f^{\delta}_{\lambda(\delta),\infty}\|_{\mathcal{X}} &\leq C_3(m_1(\delta/\tau)^{2/3})^{-1/2}(m_2 \delta^{-(4+2\varrho)/\varrho})^{-\varrho/(4+2\varrho)} \\ &= C_4 \delta^{-1/3}\delta = C_4 \delta^{2/3},\end{aligned}$$

where $C_4 := C_3\, m_1^{-1/2}\, m_2^{-\varrho/(4+2\varrho)} \tau^{1/3}$.

In conclusion, we have by the triangle inequality that

$$\begin{aligned}\|f^{\delta}_{\lambda(\delta),n(\delta)} - f^{+}\|_{\mathcal{X}} &\leq \|f^{\delta}_{\lambda(\delta),n(\delta)} - f^{\delta}_{\lambda(\delta),\infty}\|_{\mathcal{X}} + \|f^{\delta}_{\lambda(\delta),\infty} - f^{+}\|_{\mathcal{X}} \\ &\leq C_4 \delta^{2/3} + C_1 \delta^{2/3} = (C_4 + C_1)\delta^{2/3},\end{aligned}$$

which proves the claim for $C := C_4 + C_1$. □

The preceding theorem is important since, in the implementation of the RWFMP algorithm, only a finite number of iterations can be realized, but results about the convergence of Tikhonov regularization can only be applied directly if we allow for an infinite number of iterations. The theorem shows that, by stopping the RWFMP iteration after $n(\delta) \sim \delta^{-(4+2\varrho)/\varrho}$ steps, one is able to conserve the convergence rate $\delta^{2/3}$ for $\delta \to 0+$ of Tikhonov regularization even with only a finite number of iterations.

Note that the conditions on the obtained limit of the iterative algorithm, which are stated in Theorems 3.12 and 3.13, namely,

$$|f^{\delta}_{\lambda,\infty}|_{\mathcal{D}} < \infty \quad \text{and} \quad C_2(\lambda) = |f^{\delta}_{\lambda,\infty}|_{\mathcal{D}}\, \varrho^{-\varrho/(2+\varrho)} \leq C_3,$$

need to be verified. Certainly, they depend on the precise problem and the precise data situation. In the case of real data, where the exact solution is unknown and where numerical limitations do not allow the calculation of an infinite number of iterations, it appears to be unlikely that these conditions can be verified of falsified. For cases with a more detailed knowledge about the operator and the data, the derivation of strategies for practically verifying the named conditions is postponed to future investigations.

As it was already stated, for example, in [10, Remark 3.14] and in [1], an a-priori parameter choice rule, as Theorem 3.13 provides, may not be useful in the implementation of regularization methods. This is the case because it may be difficult to determine the parameter τ, which is related to the source representer h in the source condition $f^{+} = T^{*}Th$. Furthermore, when dealing with real-world data, the noise level δ might not even be known. Additionally, since the a-priori parameter choice rule yields an asymptotic estimate for $\delta \to 0+$, it is not reasonable to apply it for a fixed noise level (as one has with real-world data), even if it is known.

The derivation and theoretical analysis of an a-posteriori parameter choice rule is therefore desirable. However, we postpone this question to our upcoming research.

Instead, we will provide a numerical example in the following section, where we will show that the approach of the RWFMP can be used to speed up the iteration of the RFMP without a big loss in accuracy.

4 Numerical example

In this section, we will present a proof of concept that the *weak* approach is not only advantageous from the theoretical perspective as already mentioned, but also offers the opportunity to speed up the iteration of the RFMP.

The main difference of the RFMP and the RWFMP is the selection criterion for the next dictionary element d_{n+1}. For the RFMP, it is given by

$$|\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} - \lambda\langle f_n, d_{n+1}\rangle_{\mathcal{X}}| = \sup_{d\in\mathcal{D}}|\langle r_n, Td\rangle_{\mathcal{Y}} - \lambda\langle f_n, d\rangle_{\mathcal{X}}|, \tag{4.1}$$

and for the RWFMP, the criterion is

$$|\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} - \lambda\langle f_n, d_{n+1}\rangle_{\mathcal{X}}| \geq \varrho \sup_{d\in\mathcal{D}}|\langle r_n, Td\rangle_{\mathcal{Y}} - \lambda\langle f_n, d\rangle_{\mathcal{X}}|. \tag{4.2}$$

Here, we adopted the normalization $\|Td\|^2_{\mathcal{Y}} + \lambda\|d\|^2_{\mathcal{X}} = 1$ for the dictionary elements $d \in \mathcal{D}$ from the RWFMP also for the RFMP. On a computer, we can only realize a finite dictionary, that is, $\#\mathcal{D} = N < \infty$, $N \in \mathbb{N}$. At first sight, it seems that the implementation of the RWFMP has no advantage over the RFMP since, for the determination of a d_{n+1} that fulfills (4.2), the supremum has to be known. That is, one would have to go through all of the finitely many dictionary elements to determine the maximum. The maximizer, which can of course be stored, fulfills (4.1) such that an application of the RWFMP would have no advantage.

For this reason, we pursue a different approach. We observe that

$$\begin{aligned}
\sup_{d\in\mathcal{D}}|\langle r_n, Td\rangle_{\mathcal{Y}} - \lambda\langle f_n, d\rangle_{\mathcal{X}}| &= \sup_{d\in\mathcal{D}}\left|\left\langle\begin{pmatrix} r_n \\ -\sqrt{\lambda}f_n\end{pmatrix}, \begin{pmatrix} Td \\ \sqrt{\lambda}d\end{pmatrix}\right\rangle_{\mathcal{Y}\times\mathcal{X}}\right| \\
&\leq \left\|\begin{pmatrix} r_n \\ -\sqrt{\lambda}f_n\end{pmatrix}\right\|_{\mathcal{Y}\times\mathcal{X}} \sup_{d\in\mathcal{D}}\left\|\begin{pmatrix} Td \\ \sqrt{\lambda}d\end{pmatrix}\right\|_{\mathcal{Y}\times\mathcal{X}} \\
&= \sqrt{\|r_n\|^2_{\mathcal{Y}} + \lambda\|f_n\|^2_{\mathcal{X}}}\sup_{d\in\mathcal{D}}\sqrt{\|Td\|^2_{\mathcal{Y}} + \lambda\|d\|^2_{\mathcal{X}}} \\
&= \sqrt{\|r_n\|^2_{\mathcal{Y}} + \lambda\|f_n\|^2_{\mathcal{X}}}
\end{aligned}$$

by the Cauchy–Schwarz inequality and the normalization of the dictionary.

In consequence, if we choose $d_{n+1} \in \mathcal{D}$ such that

$$|\langle r_n, Td_{n+1}\rangle_{\mathcal{Y}} - \lambda\langle f_n, d_{n+1}\rangle_{\mathcal{X}}| \geq \varrho\sqrt{\|r_n\|^2_{\mathcal{Y}} + \lambda\|f_n\|^2_{\mathcal{X}}}, \tag{4.3}$$

then it also fulfills (4.2) and thus can be chosen by the RWFMP. Since the term on the right-hand side does no longer depend on $d \in \mathcal{D}$, it can be computed without going through the whole dictionary.

This enables us to improve the search procedure for a next dictionary element. Instead of going through all of the dictionary elements and storing the index of the maximizer, we can abort this search as soon as a dictionary element fulfills (4.3). Since, in most of the cases, this will not be the last element in the dictionary, there arises an improvement in computation time from this procedure. This will, in particular, be of importance if an a-posteriori parameter choice method is applied, where the algorithm has to be applied several times for a sequence of regularization parameters. Of course, it is possible that there exists no dictionary element that fulfills (4.3). In this case, one would simply store the dictionary element for which the term on the left-hand side of (4.3) is maximal and thus perform an RFMP iteration.

In the following section, we will consider a one-dimensional model problem, for which we will present numerical results later on.

4.1 A one-dimensional model problem

To give a proof of concept for the improved computation time that the RWFMP offers, we restrict to a simple one-dimensional model problem, which we adopt from [24, Beispiel 3.2.2].

Consider the boundary value problem

$$\begin{aligned}
-g''(x) &= f(x) \quad \text{for all } x \in (0, 1), \\
g(0) &= g(1) = 0
\end{aligned} \tag{4.4}$$

for given $f \in \mathrm{C}^{(0)}((0,1))$, where (4.4) is the one-dimensional Poisson equation. By the Lax–Milgram theorem, it is well known that the weak formulation

$$\int_0^1 g'(x)\varphi'(x)\,\mathrm{d}x = \int_0^1 f(x)\varphi(x)\,\mathrm{d}x$$

for all $\varphi \in \mathrm{H}_0^1((0,1))$ possesses a unique solution $g \in \mathrm{H}_0^1((0,1))$. This solution is also continuous due to the Sobolev embedding theorem.

By applying the concept of Green's function, we can deduce the integral operator

$$T\colon \mathrm{L}^2((0,1)) \to \mathrm{H}_0^1((0,1)), \quad (Tf)(x) = \int_0^1 k(x,y)f(y)\,\mathrm{d}y,$$

where

$$k(x,y) = \begin{cases} x(1-y), & x \le y, \\ y(1-x), & x > y, \end{cases}$$

such that $Tf = g$ if and only if g fulfills the weak formulation above.

In our implementation, we assume that we are given evaluations

$$\big((Tf)(x_j)\big)_{j=1,\dots,J} \quad \text{for } X = (x_1,\dots,x_J) \in [0,1]^J,$$

which are denoted by $(g_j)_{j=1,\dots,J} \in \mathbb{R}^J$. We assume that $x_1 < x_2 < \dots < x_J$. We denote the corresponding operator by $T_X\colon \mathrm{L}^2([0,1]) \to \mathbb{R}^J$. Note that this operator is well-defined due to the continuity of Tf.

We will consider two different dictionaries $\mathcal{D}_1, \mathcal{D}_2 \subseteq \mathrm{L}^2([0,1])$.

The first dictionary is motivated by the singular-value decomposition of the operator. In [24], it is stated that a singular system is given by $(1/(\pi n)^2, u_n, u_n)_{n\in\mathbb{N}}$, where $u_n(x) = \sqrt{2}\sin(\pi n x)$. We therefore first define

$$\tilde{\mathcal{D}}_1 := \{\sin(\pi n\,\cdot\,) \mid n = 1,\dots,N\}$$

for some $N \in \mathbb{N}$ to obtain the dictionary

$$\mathcal{D}_1^\lambda := \left\{ \frac{d}{\sqrt{\|T_X d\|_{\mathbb{R}^J}^2 + \lambda \|d\|_{\mathrm{L}^2([0,1])}^2}} \;\middle|\; d \in \tilde{\mathcal{D}}_1 \right\}$$

if the regularization parameter is $\lambda > 0$ to obtain the correct normalization of the elements from the dictionary.

The second dictionary is motivated by one-dimensional finite element methods (see, e.g., [11, Section 1.2]). For this purpose, given a vector $Y = (y_1,\dots,y_S) \in [0,1]^S$ of nodes, which fulfill $y_1 < y_2 < \dots < y_S$, we define the functions

$$\varphi_s(y) = \begin{cases} 0, & y \le y_{s-1}, \\ \frac{y-y_{s-1}}{y_s-y_{s-1}}, & y_{s-1} < y \le y_s, \\ \frac{y_{s+1}-y}{y_{s+1}-y_s}, & y_s < y \le y_{s+1}, \\ 0, & y_{s+1} < y, \end{cases}$$

for $s = 1,\dots,S$ and $y \in [0,1]$. These functions are *hat functions*, that is, they are piece-wise linear between the nodes in Y and they fulfill $\varphi_s(y_j) = \delta_{sj}$. Using these functions, we first define

$$\tilde{\mathcal{D}}_2 := \{\varphi_s \mid s = 1,\dots,S\}$$

and obtain, for the regularization parameter $\lambda > 0$, the dictionary

$$\mathcal{D}_2^\lambda := \left\{ \frac{d}{\sqrt{\|T_X d\|_{\mathbb{R}^J}^2 + \lambda \|d\|_{\mathrm{L}^2([0,1])}^2}} \;\middle|\; d \in \tilde{\mathcal{D}}_2 \right\}$$

by normalization.

4.2 Numerical results

To test the improvement of the computation time, when applying the search strategy presented above, we prescribed the solution $f(y) = 1$ for all $y \in [0, 1]$ such that $(Tf)(x) = \frac{1}{2}x(1-x) = g(x)$, which can be easily checked using the boundary value problem or by integration.

We set $J = 1000$ to sample g at 1000 evenly distributed data points in $[0, 1]$ and apply 1 % of deterministic noise. Furthermore, we chose $N = S = 400$ to obtain 400 elements in both of the dictionaries $\mathcal{D}_1^\lambda$, $\mathcal{D}_2^\lambda$.

We implemented a simple a-posteriori parameter choice method, where we applied 100 iterations of the RWFMP using 100 different regularization parameters and chose the regularized solution where the approximation error is the smallest. The regularization parameters were chosen equally spaced on a logarithmic scale between 1 and 10^{-14}. Of course, a different parameter choice method must be used if the solution to the problem is not known. For the RFMP, several of these methods were compared in [9]. For all of them, one needs to execute the algorithm for many regularization parameters such that the saving in computation time will also apply for other parameter choice methods.

Moreover, to show the power of the algorithm for arbitrary dictionaries, we did not use the dictionaries in the order in which they were defined, but instead used 20 random permutations of the dictionary elements to eliminate the bias of dictionaries whose order is by chance very well suited or very badly suited to the problem at hand. Finally, we tested the algorithm for 10 different values of the weakness parameter ϱ.

The results for the dictionary $\mathcal{D}_1^\lambda$ consisting of sine functions are given in Table 1. For the dictionary $\mathcal{D}_2^\lambda$ of hat functions, the results can be found in Table 2.

ϱ	λ	RMSE	Computation time/s	/% of first line
1.000000	$1.49 \times 10^{-4} \pm 5.40 \times 10^{-4}$	$6.33 \times 10^{-1} \pm 3.84 \times 10^{-1}$	$2.29 \times 10^{2} \pm 1.24 \times 10^{1}$	100.0
0.464159	$4.54 \times 10^{-7} \pm 2.91 \times 10^{-7}$	$2.02 \times 10^{-1} \pm 3.65 \times 10^{-2}$	$2.25 \times 10^{2} \pm 1.18 \times 10^{1}$	98.2
0.215443	$4.32 \times 10^{-7} \pm 2.06 \times 10^{-7}$	$1.94 \times 10^{-1} \pm 2.40 \times 10^{-2}$	$2.25 \times 10^{2} \pm 1.04 \times 10^{1}$	98.0
0.100000	$2.70 \times 10^{-7} \pm 1.39 \times 10^{-7}$	$1.74 \times 10^{-1} \pm 1.07 \times 10^{-2}$	$2.22 \times 10^{2} \pm 1.23 \times 10^{1}$	97.1
0.046416	$2.20 \times 10^{-7} \pm 1.08 \times 10^{-7}$	$1.70 \times 10^{-1} \pm 9.34 \times 10^{-3}$	$2.01 \times 10^{2} \pm 8.57 \times 10^{0}$	87.6
0.021544	$2.70 \times 10^{-7} \pm 9.41 \times 10^{-8}$	$\mathbf{1.62 \times 10^{-1} \pm 5.28 \times 10^{-3}}$	$1.72 \times 10^{2} \pm 8.58 \times 10^{0}$	75.2
0.010000	$2.70 \times 10^{-7} \pm 9.41 \times 10^{-8}$	$1.63 \times 10^{-1} \pm 1.60 \times 10^{-3}$	$1.59 \times 10^{2} \pm 7.74 \times 10^{0}$	69.6
0.004642	$2.98 \times 10^{-7} \pm 9.56 \times 10^{-8}$	$1.62 \times 10^{-1} \pm 1.03 \times 10^{-3}$	$1.47 \times 10^{2} \pm 7.70 \times 10^{0}$	64.2
0.002154	$3.83 \times 10^{-7} \pm 5.43 \times 10^{-23}$	$1.63 \times 10^{-1} \pm 4.90 \times 10^{-4}$	$1.39 \times 10^{2} \pm 5.97 \times 10^{0}$	60.9
0.001000	$3.83 \times 10^{-7} \pm 5.43 \times 10^{-23}$	$1.63 \times 10^{-1} \pm 1.70 \times 10^{-4}$	$\mathbf{1.33 \times 10^{2} \pm 5.98 \times 10^{0}}$	**58.2**

Table 1: Results for the dictionary $\mathcal{D}_1^\lambda$ of sine functions. The columns are: the weakness parameter ϱ, the mean optimal regularization parameter λ and its standard deviation, the average RMSE and its standard deviation, the mean and standard deviation of the computation time for one execution of the RWFMP, and the percentage of the computation time with respect to the first line. The minimal RMSE and computation time are set in a bold font.

ϱ	λ	RMSE	Computation time/s	/% of first line
1.000000	$1.08 \times 10^{-2} \pm 3.56 \times 10^{-3}$	$8.55 \times 10^{-1} \pm 2.59 \times 10^{-2}$	$2.25 \times 10^{2} \pm 6.55 \times 10^{0}$	100.0
0.464159	$9.95 \times 10^{-3} \pm 2.67 \times 10^{-3}$	$8.53 \times 10^{-1} \pm 2.25 \times 10^{-2}$	$2.23 \times 10^{2} \pm 6.36 \times 10^{0}$	98.8
0.215443	$1.08 \times 10^{-2} \pm 3.56 \times 10^{-3}$	$8.41 \times 10^{-1} \pm 1.40 \times 10^{-2}$	$1.94 \times 10^{2} \pm 1.07 \times 10^{1}$	85.9
0.100000	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.48 \times 10^{-1} \pm 1.64 \times 10^{-2}$	$6.64 \times 10^{1} \pm 9.95 \times 10^{0}$	29.4
0.046416	$1.51 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$\mathbf{8.32 \times 10^{-1} \pm 4.41 \times 10^{-3}}$	$3.98 \times 10^{1} \pm 2.49 \times 10^{0}$	17.6
0.021544	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.38 \times 10^{-1} \pm 6.39 \times 10^{-3}$	$3.05 \times 10^{1} \pm 2.70 \times 10^{0}$	13.4
0.010000	$9.95 \times 10^{-3} \pm 2.67 \times 10^{-3}$	$8.48 \times 10^{-1} \pm 1.25 \times 10^{-2}$	$2.01 \times 10^{1} \pm 2.95 \times 10^{0}$	8.9
0.004642	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.44 \times 10^{-1} \pm 1.48 \times 10^{-2}$	$1.64 \times 10^{1} \pm 3.37 \times 10^{0}$	7.2
0.002154	$1.16 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$8.46 \times 10^{-1} \pm 1.43 \times 10^{-2}$	$1.64 \times 10^{1} \pm 3.40 \times 10^{0}$	7.2
0.001000	$1.16 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$8.50 \times 10^{-1} \pm 1.35 \times 10^{-2}$	$\mathbf{1.53 \times 10^{1} \pm 3.23 \times 10^{0}}$	**6.8**

Table 2: Results for the dictionary $\mathcal{D}_2^\lambda$ of hat functions. The columns are: the weakness parameter ϱ, the mean optimal regularization parameter λ and its standard deviation, the average RMSE and its standard deviation, the mean and standard deviation of the computation time for one execution of the RWFMP, and the percentage of the computation time with respect to the first line. The minimal RMSE and computation time are set in a bold font.

In these tables, the first column corresponds to the chosen weakness parameter ϱ. We chose 10 values of ϱ, which are logarithmically distributed between 1 and 0.001. The second column shows the regularization parameter λ, which had been chosen from the 100 prescribed parameters to minimize the approximation error. The values shown are the mean value of λ over all 20 random permutations of the dictionary and the corresponding standard deviation. The mean and the standard deviation are also shown in the third column for the approximation error, which we computed in the form of the *root mean squared error* (RMSE) by the evaluation on a grid with 1000 evenly distributed points in $[0, 1]$ in this case. For the fourth column, we computed the average computation time for 100 iterations of the RWFMP and its standard deviation. Finally, the last column shows the ratio of the average computation time to the time needed in the case $\varrho = 1$, that is, the RFMP.

We will first discuss the results for the dictionary consisting of sine functions $\mathcal{D}_1^\lambda$. Concerning the regularization parameter λ, we observe that the chosen optimal parameter (apart from the case $\varrho = 1$) always has the same order of magnitude. This is what we would expect since all of the algorithms converge to the minimizer of the same Tikhonov functional such that in theory the same parameter should be chosen for all ϱ. The choice of the parameter might be affected by the fixed number of iterations, which explains the differences in the chosen parameters. The results for the RMSE are surprising. From the proved convergence rate, we would expect that the approximation error for a fixed number of iterations is worse for smaller values of ϱ. From the given data, this is obviously not the case. Looking at the subsequent decimal places, we obtain the minimal approximation error for $\varrho = 0.021544$, but the RMSE values are nearly all the same for $\varrho \leq 0.1$. Concerning the computation time, we observe that lower values of ϱ result in a lower computation time, as one would expect due to the optimization of the search strategy.

For the dictionary $\mathcal{D}_1^\lambda$, we can conclude: On average, we can save over 40 % of computation time by applying the RWFMP with the parameter $\varrho = 0.001$ instead of the RFMP. We even obtain a smaller RMSE value if we do so. We have to admit that this example is special since we are using an orthogonal basis as dictionary and, additionally, the sine functions are arising in the singular system of the considered operator. Furthermore, we try to approximate a constant function by sine functions, which might lead to additional errors induced by Gibb's phenomenon (Figure 1). In this figure, we plotted the exact solution together with the approximations generated by the RWFMP using exact and noisy data, respectively. In the case with exact data, we can clearly observe the typical boundary behavior induced by Gibb's phenomenon. Additionally, with noisy data, the errors at the boundary of the interval are even worse. We conjecture that the results are affected by these facts. Therefore, we will also discuss the results that we achieved by using the dictionary $\mathcal{D}_2^\lambda$, which is not specifically connected to the inverse problem at hand.

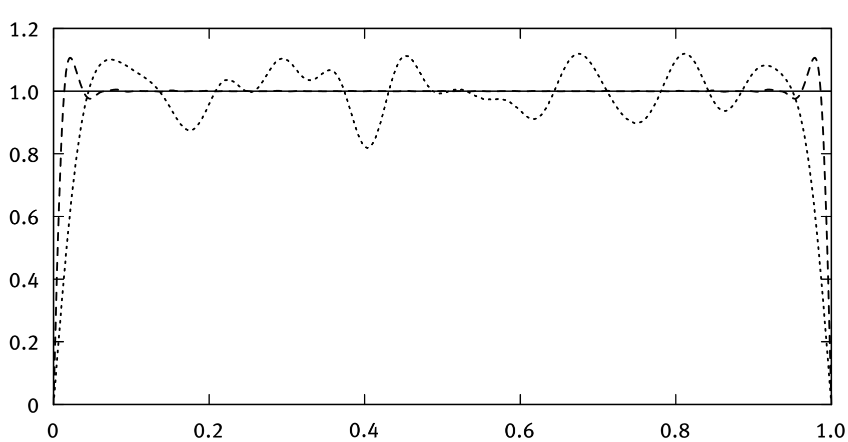

Figure 1: Gibb's phenomenon influences the result of the computation in the first example: the solid line corresponds to the exact solution, the dashed line corresponds to the approximation by the RWFMP using exact data, and the dotted line corresponds to the approximation by the RWFMP using data with 1 % noise.

The results for the chosen regularization parameter when using $\mathcal{D}_2^\lambda$ in Table 2 are similar to the results presented above. In this case, even all of the average chosen parameters are nearly the same, as we would expect. However, the regularization parameters are much larger than in the previous example. In theory, we consider the same inverse problem for both examples and would assume that choosing the same regularization parameter would be natural. In practice, both dictionaries span different subspaces of $L^2([0, 1])$ such that we solve the projections of the inverse problem onto different sets, which is a reason for the difference in the regularization parameter. The need for a stronger regularization in the second example is also the reason for higher values of the RMSE in this case. In this case, the optimal approximation error is achieved for $\varrho = 0.046416$, but all of the errors are not larger than 2.7 % of this minimal error. In comparison to the numbers presented for $\mathcal{D}_1^\lambda$, the trend in the computation times is very different. For large values of ϱ, the times are very similar to the first example. Starting with $\varrho = 0.1$, the weak approach shows its potential for improving the performance of the algorithm. Although the approximation error does not change dramatically, the computation time drops from 225 s for $\varrho = 1$ down to 15.3 s for $\varrho = 0.001$.

We can thus conclude for the dictionary $\mathcal{D}_2^\lambda$ of hat functions that the RWFMP has the potential to outperform the RFMP drastically if the presented search strategy for the next dictionary element is used. For a fixed number of iterations, saving over 90 % of computation time is possible leading to nearly no change in the approximation error.

As a general conclusion for both of the given dictionaries, we can say that the implementation of a weak greedy algorithm for inverse problems may give a large improvement in the efficiency of the algorithm. As already stated above, one can save up to 90 % of computation time without losing the accuracy of the approximation. For problems from the geoscientific applications, where one may have more than 10 000 data points and a dictionary consisting of thousands of functions, this speed-up is very promising. The improvement of computation time makes it possible to even put more different kinds of functions into the dictionary, which might be better adapted to the solution. This may lead to an improved approximation quality, while one can obtain the same overall computation time as one has for the RFMP with a smaller dictionary.

5 Conclusions and outlook

In this paper, we proposed a generalization of the Regularized Functional Matching Pursuit algorithm (RFMP) for linear inverse problems, called the Regularized Weak Functional Matching Pursuit (RWFMP). In comparison to the RFMP, the RWFMP has an improved theoretical justification for the application of greedy algorithms to inverse problems in arbitrary Hilbert spaces, whereas the RFMP was only considered for a finite-dimensional data space before. Moreover, the weak version guarantees the existence of the approximation in the next iterative step without sacrificing the convergence of the algorithm. Indeed, it was shown that the non-regularized version of the algorithm converges to a solution of the associated normal equation if the inverse problem is well-posed. In this case, a rate of convergence for the data misfit was derived.

If the inverse problem is ill-posed, it was shown that a regularized version of the algorithm converges to the solution of the Tikhonov-regularized normal equation. In addition, there is a benefit also for the RFMP since we managed to overcome the previously required and impracticable semi-frame condition. Also, for the regularized algorithm, a rate of convergence was proved, which measures the convergence in a special metric on the Cartesian product of the data space and the domain, which is motivated by the Tikhonov functional. Furthermore, we were able to provide an a-priori parameter choice rule for the regularization parameter and the number of iterations of the RWFMP such that this yields a convergent regularization method.

We also considered the practical implications of the application of a weak greedy algorithm. We were able to show that this approach gives rise to an improved search strategy inside the algorithm, which yields an improvement of the computation time of up to 90 % without a great change in the accuracy of the method.

In further research, we want to apply the idea of a weak greedy algorithm to the so-called ROFMP [20, 26], which is an orthogonal version of the RFMP. Since a greedy algorithm for the approximation problem exists, which is both “weak” and “orthogonal”, called the WOGA [27], this should be possible. Since both the ROFMP and the RWFMP algorithm themselves have certain advantages over the RFMP, we assume that a combination of both, an RWOFMP, may yield even better results. Furthermore, we would like to apply the RWFMP to the geoscientific problems to which the RFMP has already been applied before. The results might yield an additional proof that the approach that we have taken in this paper is beneficial for the numerical solution of inverse problems in real applications.

Acknowledgment: We would like to thank the anonymous reviewers for their valuable comments, which helped to improve the paper.

Funding: The authors gratefully acknowledge the financial support by the School of Science and Technology of the University of Siegen, Germany

References

[1] T. Bonesky, Morozov's discrepancy principle and Tikhonov-type functionals, *Inverse Problems* **25** (2009), no. 1, Article ID 015015.
[2] R. A. DeVore and V. N. Temlyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.* **5** (1996), no. 2–3, 173–187.
[3] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. 375, Kluwer Academic, Dordrecht, 1996.
[4] D. Fischer, *Sparse regularization of a joint inversion of gravitational data and normal mode anomalies*, PhD thesis, Geomathematics Group, University of Siegen, 2011, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2012/544/; Dr. Hut, Munich, 2011.
[5] D. Fischer and V. Michel, Sparse regularization of inverse gravimetry—case study: Spatial and temporal mass variations in South America, *Inverse Problems* **28** (2012), no. 6, Article ID 065012.
[6] D. Fischer and V. Michel, Automatic best-basis selection for geophysical tomographic inverse problems, *Geophys. J. Int.* **193** (2013), 1291–1299.
[7] D. Fischer and V. Michel, Inverting GRACE gravity data for local climate effects, *J. Geod. Sci.* **3** (2013), 151–162.
[8] S. Gramsch, M. Kontak and V. Michel, Three-dimensional simulation of nonwoven fabrics using a greedy approximation of the distribution of fiber directions, *ZAMM Z. Angew. Math. Mech.* **98** (2018), no. 2, 277–288.
[9] M. Gutting, B. Kretz, V. Michel and R. Telschow, Study on parameter choice methods for the RFMP with respect to downward continuation, *Front. Appl. Math. Stat.* (2017), DOI 10.3389/fams.2017.00010.
[10] K. Ito and B. Jin, *Inverse Problems. Tikhonov Theory and Algorithms*, Ser. Appl. Math. 22, World Scientific, Hackensack, 2015.
[11] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Dover, Mineola, 2009.
[12] L. K. Jones, On a conjecture of Huber concerning the convergence of projection pursuit regression, *Ann. Statist.* **15** (1987), no. 2, 880–882.
[13] M. I. Kadets and V. M. Kadets, *Series in Banach Spaces. Conditional and Unconditional Convergence*, Oper. Theory Adv. Appl. 94, Birkhäuser, Basel, 1997.
[14] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, 2nd ed., Appl. Math. Sci. 120, Springer, New York, 2011.
[15] M. Kontak, *Novel algorithms of greedy-type for probability density estimation as well as linear and nonlinear inverse problems*, PhD thesis, Geomathematics Group, University of Siegen, 2018, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1316/.
[16] S. G. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* **41** (1993), 3397–3415.
[17] V. Michel, RFMP: An iterative best basis algorithm for inverse problems in the geosciences, in: *Handbook of Geomathematics*, Springer, Berlin (2015), 2121–2147.
[18] V. Michel and S. Orzlowski, On the convergence theorem for the regularized functional matching pursuit (RFMP) algorithm, *GEM Int. J. Geomath.* **8** (2017), no. 2, 183–190.
[19] V. Michel and R. Telschow, A non-linear approximation method on the sphere, *GEM Int. J. Geomath.* **5** (2014), no. 2, 195–224.
[20] V. Michel and R. Telschow, The regularized orthogonal functional matching pursuit for ill-posed inverse problems, *SIAM J. Numer. Anal.* **54** (2016), no. 1, 262–287.
[21] D. S. Mitrinović, *Analytic Inequalities*, Grundlehren Math. Wiss. 165, Springer, New York, 1970,
[22] E. H. Moore, On the reciprocal of the general algebraic matrix, *Bull. Amer. Math. Soc.* **26** (1920), 394–395.

[23] R. Penrose, A generalized inverse for matrices, *Math. Proc. Cambridge Philos. Soc.* **51** (1955), 406–413.
[24] A. Rieder, *Keine Probleme mit inversen Problemen. Eine Einführung in ihre stabile Lösung*, Friedr. Vieweg & Sohn, Braunschweig, 2003.
[25] W. Rudin, *Functional Analysis*, 2nd ed., McGraw-Hill, New York, 1991.
[26] R. Telschow, *An orthogonal matching pursuit for the regularization of spherical inverse problems*, PhD thesis, Geomathematics Group, University of Siegen, 2014; Dr. Hut, Munich, 2015.
[27] V. N. Temlyakov, Weak greedy algorithms, *Adv. Comput. Math.* **12** (2000), no. 2–3, 213–227.
[28] A. N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math.* **4** (1963), 1035–1038.