An entropic Landweber method for linear ill-posed problems

View the article online for updates and enhancements.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

An entropic Landweber method for linear ill-posed problems

M. Burger[*]   E. Resmerita[†]   M. Benning[‡]

November 24, 2019

Abstract

The aim of this paper is to investigate the use of a Landweber-type method involving the Shannon entropy for the regularization of linear ill-posed problems. We derive a closed form solution for the iterates and analyze their convergence behaviour both in a case of reconstructing general nonnegative unknowns as well as for the sake of recovering probability distributions. Moreover, we discuss several variants of the algorithm and relations to other methods in the literature. The effectiveness of the approach is studied numerically in several examples.

1   Introduction

This work deals with linear ill-posed equations $Au = y$ with $A : X \to Y$ acting between a Banach space X and a Hilbert space Y, for which solutions with specific properties (such as positivity) are sought. In this respect, we consider iterative regularization methods of the following type

$$u_{k+1} \in \arg\min_u \left\{ \frac{1}{2}\|Au - y\|^2 + cd(u, u_k) - \frac{1}{2}\|Au - Au_k\|^2 \right\}, \quad k \in \mathbb{N}, \tag{1}$$

where $d = D_f$ denotes the Bregman distance [10] associated with a convex functional $f : X \to \mathbb{R} \cup \{+\infty\}$ which is nonnegative and c is some positive number. The term $d(u, u_k)$ acts as a penalty enforcing the desired features for the solutions.
Note that we can rewrite (1) as

$$u_{k+1} \in \arg\min_u \left\{ \langle Au - y, Au_k - y \rangle + cd(u, u_k) \right\}, \tag{2}$$

which shows that the scheme can be obtained also by linearizing the quadratic data-fitting term $\frac{1}{2}\|Au - y\|^2$ at the current iterate u_k. This class of methods incorporates several procedures that have been proposed so far in the literature. For instance, the classical case when f is quadratic in Hilbert spaces reduces to the Landweber method, as emphasized by [17] and as investigated for nonlinear operator equations by means of surrogate functionals in [28], see also the discussion in [31]. The case of quadratic f in reflexive Banach spaces has been studied by [32]. The setting when f is the total variation functional smoothed by a quadratic has been analyzed by [4], requiring fine analysis tools due to the bounded variation function space context. The case of ℓ^1-penalties has been treated in [33, 12], resulting in the so-called linearized Bregman algorithm. In all those cases however, some quadratic term had to be part of f to guarantee even well-definedness of the iterates and subsequently convergence.

---

[*]Department of Mathematics, Friedrich-Alexander Universität Erlangen-Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany (`martin.burger@fau.de`)

[†]Institute of Mathematics, Alpen-Adria Universität Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt, Austria (`elena.resmerita@aau.at`)

[‡]School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK (`m.benning@qmul.ac.uk`)

We are interested here in the Shannon entropy setting without any quadratic term, i.e.

$$f(u) = \int_\Omega u(t) \ln u(t) \ dt.$$

We mention that one can alternatively consider a linear shift to

$$\tilde{f}(u) = \int_\Omega \left[u(t) \ln u(t) - u(t) + 1 \right] \ dt,$$

which is a nonnegative functional inducing the same Bregman distance.

This raises challenges to analyze the problem in the L^1 setting without quadratic terms in the functional, but provides a simple closed iterative method with preserving the sign of the starting point (function) along the iterations. The latter formulation involves entropic projections, as one can see in the following section.

Moreover, we shall also be interested in the solution of inverse problems with unknowns being probability densities, i.e. we minimize on the domain of f subject to the constraint

$$\int_\Omega u(t) \ dt = 1, \tag{3}$$

which again results in a simple closed iterative form.

The advantages of using Bregman projections for solving variational problems with unknown probability densities have been exploited by several authors before, e.g. in optimal transport (cf. [8, 27]).

In order to write both problems in a closed form, we will use the equivalent formulation

$$u_{k+1} \in \arg\min_u \left\{ \langle Au - y, Au_k - y \rangle + cd(u, u_k) + \chi_m(u) \right\}, \quad m \in \{0, 1\}, \tag{4}$$

where $\chi_0 \equiv 0$ denotes the original problem without integral constraint, and

$$\chi_1(u) = \begin{cases} 0 & \text{if } \int_\Omega u(t) \ dt = 1, \\ +\infty & \text{else,} \end{cases}$$

is employed for enforcing probability densities. The minimization is taken here over the domain of the entropy functional.

One finds the above entropy based algorithm in the finite dimensional optimization literature, as well as in the machine learning one, under quite different names. It is precisely the mirror descent algorithm with the negative entropy as the mirror map - see [25], [7], as well as [16] and Chapter 9 in [6] for a detailed analysis. It is also called the exponentiated gradient descent method for linear predictions - see [24]. The work [22] investigated three versions of the so-called approximate (linearized) proximal point methods for optimization in combination with line search strategies. The reader is referred to Sections 6.6 - 6.9 in [14] for other iterative optimization methods employing the Shannon entropy.

The main contribution of our work is the convergence of the iterates (4) to a solution of the equation $Au = y$ in an infinite dimensional setting of such a nonquadratic penalty version, by stating also error estimates in the sense of a distance between the solution and the iterates, as opposed to the classical situation encountered in optimization, where the error for the objective function values is highlighted. This contribution will be emphasized in the next section (Subsection 2.3), where more details will be provided about the existing results on the mirror descent and other related methods.

The manuscript is organized as follows. Section 2 provides the necessary background on entropy functionals, as well as on well-definedness of the proposed iterative procedure. Section 3 analyzes (weak) convergence of the method when both a priori and a posteriori stopping rules are considered, while Section 4 deals with error estimates only for the former rule. Section 5 explores a version of the entropic Landweber method for nonquadratic data fidelity terms. The theoretical results are tested in Section 6 on several integral equation examples, in comparison with the Expectation-Maximization algorithm and the projected Landweber method - see [15] for an overview on regularization methods for nonnegative solutions of ill-posed equations.

2

2 Preliminaries

In the following we collect some basic results and assumptions needed for the analysis below. We start with properties of the entropy and then proceed to the operator A.

2.1 Entropy

Let Ω be an open and bounded subset of $\mathbb{R}^d$. *The negative of the Boltzmann-Shannon entropy is the function* $f : L^1(\Omega) \to (-\infty, +\infty]$, *given by*[1]

$$
f(u) = \begin{cases} \int_\Omega u(t) \ln u(t)\, dt & \text{if } u \geq 0 \text{ a.e. and } u \ln u \in L^1(\Omega), \\[2mm] +\infty & \text{otherwise.} \end{cases} \tag{5}
$$

Here and in what follows $L^p_+(\Omega)$, $p \in [1, \infty]$, stands for the set $\{u \in L^p(\Omega) : u(t) \geq 0 \text{ a.e.}\}$, while $\|\cdot\|_p$ denotes, as usual, the norm of the space $L^p(\Omega)$.

The Kullback-Leibler functional or the Bregman distance with respect to the Bolzmann-Shannon entropy can be defined as $d : \operatorname{dom} f \times \operatorname{dom} f \to [0, +\infty]$ by

$$
d(v, u) = f(v) - f(u) - f'(u, v - u) \tag{6}
$$

where $f'(u, \cdot)$ is the directional derivative at u. Here $\operatorname{dom} f = \{u \in L^1(\Omega) : f(u) < \infty\}$ denotes the domain of f. One can also write

$$
d(v, u) = \int_\Omega \left[v(t) \ln \frac{v(t)}{u(t)} - v(t) + u(t) \right] dt \tag{7}
$$

if $d(v, u)$ is finite, as one can see below.

Some properties of the entropy functionals f and d are recalled below (see, e.g., [29, 30]).

Lemma 2.1 The function defined by (5) has the following properties:

(i) The domain of the function f is strictly included in $L^1_+(\Omega)$.

(ii) The interior of the domain of the function f is empty.

(iii) The set $\partial f(u)$ is nonempty if and only if u belongs to $L^\infty_+(\Omega)$ and is bounded away from zero. Moreover, $\partial f(u) = \{1 + \ln u\}$.

(iv) The directional derivative of the function f is given by

$$
f'(u, v) = \int_\Omega v(t)[1 + \ln u(t)]\, dt,
$$

whenever it is finite.

(v) For any $u, v \in \operatorname{dom} f$, one has

$$
\|u - v\|_1^2 \leq \left(\frac{2}{3} \|v\|_1 + \frac{4}{3} \|u\|_1 \right) d(v, u). \tag{8}
$$

Based on Lemma 2.1 (iii) we define in the following

$$
\operatorname{dom} \partial f = \{u \in L^1(\Omega) : u \text{ bounded and bounded away from zero } a.e.\}. \tag{9}
$$

Lemma 2.2 The statements below hold true:

---
[1]We use the convention $0 \ln 0 = 0$.

(i) The function $(v, u) \mapsto d(v, u)$ is convex;

(ii) The function $d(\cdot, u^*)$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$, whenever $u^* \in \operatorname{dom} f$;

(iii) For any $C > 0$ and any nonnegative $u \in L^1(\Omega)$, the following sets are weakly compact in $L^1(\Omega)$:

$$\{x \in L^1(\Omega) : f(x) \le C\}, \qquad \{x \in L^1(\Omega) : d(x, u) \le C\}.$$

(iv) The set $\partial d(\cdot, u^*)(u)$ is nonempty for $u^* \in \operatorname{dom} f$ if and only if u belongs to $L^\infty_+(\Omega)$ and is bounded away from zero. Moreover, $\partial d(\cdot, u^*)(u) = \{\ln u - \ln u^*\}$.

Denote

$$\langle u, v \rangle = \int_\Omega u(t) v(t) \, dt,$$

for $u, v \in L^1(\Omega)$, when the integral exists.

A key observation for obtaining well-definedness of the iterative scheme as well as an explicit form for the iterates is the following result on the entropic projection.

Proposition 2.3 Let $\ell \in L^\infty(\Omega)$ and $v \in \operatorname{dom} \partial f$. Then the problem

$$\langle \ell, u \rangle + d(u, v) + \chi_m(u) \to \min_{u \in \operatorname{dom}(f + \chi_m)} \tag{10}$$

has a unique solution in the cases $m = 0$ and $m = 1$, respectively, given by

$$u_m = c_m v e^{-\ell}, \qquad c_m = \begin{cases} 1 & \text{if } m = 0, \\ \frac{1}{\int_\Omega v e^{-\ell} \, dt} & \text{if } m = 1, \end{cases} \tag{11}$$

which satisfies $u_m \in \operatorname{dom} \partial f$.

Proof: We simply rewrite the functional as

$$\langle \ell, u \rangle + d(u, v) + \chi_m(u) = \int_\Omega \left[u(t) \ln \frac{u(t)}{v(t)} - u(t) + v(t) + u(t)\ell(t) \right] \, dt + \chi_m(u)$$

$$= \int_\Omega \left[u(t) \ln \frac{u(t)}{u_m(t)} - u(t) + v(t) + u(t) \ln c_m \right] \, dt + \chi_m(u)$$

$$= d(u, u_m) + \ln c_m \left(\int_\Omega u(t) \, dt - 1 \right) + \chi_m(u) + C_m,$$

where $C_m = \ln c_m - \int_\Omega u_m(t) \, dt$ is a constant independent of u. It is straightforward to notice that

$$\ln c_m \left(\int_\Omega u(t) \, dt - 1 \right) + \chi_m(u) = \chi_m(u), \quad \forall u \in \operatorname{dom}(f + \chi_m).$$

Hence, the problem is equivalent to minimizing $d(u, u_m) + \chi_m(u)$. Since both terms are nonnegative and vanish for $u = u_m$, we see that u_m is indeed a minimizer in $\operatorname{dom} f$. Strict convexity of d implies the uniqueness and since u_m is the product of v with a function strictly bounded away from zero it also satisfies $u_m \in \operatorname{dom} \partial f$. $\quad\square$

2.2 Forward operators and entropy

In this paper we always assume that $A : L^1(\Omega) \to Y$ is a linear and bounded operator with Y being a Hilbert space. In addition to the norm boundedness of A, we assume a continuity property in terms of the Bregman distance. More precisely we assume that

$$\|Au - Av\| \le \gamma \sqrt{d(u, v)} \tag{12}$$

holds on $\operatorname{dom}(f + \chi_m)$ in the respective cases $m = 0$ or $m = 1$ for some positive number γ. It is easy to see that the latter is already implied by the boundedness of A in case $m = 1$:

4

Lemma 2.4 Let A be as above with $\|A\|$ denoting its operator norm, let $u, v \in \mathrm{dom}(f + \chi_1)$, and $v \in \mathrm{dom}\,\partial f$. Then (12) is satisfied with $\gamma = \sqrt{2}\|A\|$.

Proof: By the boundedness of A we have

$$\|Au - Av\|^2 \leq \|A\|^2 \, \|u - v\|^2_{L^1(\Omega)}.$$

Lemma 2.1 (v) further implies

$$\|Au - Av\|^2 \leq \|A\|^2 \, 2d(u, v),$$

which yields the assertion. $\square$

We define the nonlinear functional

$$D(u, v) = cd(u, v) - \frac{1}{2}\|Au - Av\|^2, \tag{13}$$

which will be useful for the further analysis. Note that $D(u, v) \geq 0$ for $u, v \in \mathrm{dom}(f + \chi_1)$, and $v \in \mathrm{dom}\,\partial f$, whenever $c \geq \frac{\gamma^2}{2}$ (cf. Lemma 2.4). In case $m = 0$, we restrict the analysis to the class of operators A for which $D(u, v) \geq 0$ for $u \in \mathrm{dom}f$ and $v \in \mathrm{dom}\,\partial f$.

2.3 The entropic Landweber method versus the mirror descent and other related methods

The mirror descent algorithm (MDA) was introduced in [25] for minimizing nonsmooth convex functions f over closed convex sets $X \subset \mathbb{R}^n$,

$$x^{k+1} \in \arg\min_{x \in X} \left\{ \langle x, f'(x^k) \rangle + \frac{1}{t_k} B_\psi(x, x^k) \right\},$$

where $f'(x^k)$ is a subgradient of f at x^k, the parameters t_k are positive and B_ψ is the Bregman distance associated with a well-behaving function ψ. It is known that the algorithm converges in the sense $f(x^k) \to f(x^*)$, where x^* is a minimizer of f in X. It holds also in infinite dimensional spaces - see, e.g. Chapter 9 in [6]. As shown in [16], even the iterates converge to a solution when the auxiliary function ψ is strongly convex, that is $B_\psi(z, x) \geq \frac{\mu}{2}\|z - x\|^2$, for some $\mu > 0$ and for any z, x belonging to a set which is large enough. This property is crucial in the analysis performed in spaces where the sublevel sets of the norms are (weakly) compact, implying boundedness of the iterates from the boundedness of the sequence with term $B_psi(x^*, x^k)$. It is however not helpful in infinite dimension for the entropy setting, since strong convexity of the Kullback-Leibler distance yields boundedness of the iterates with respect to the L^1 norm only, which is not powerful enough to guarantee convergence on subsequences. The entropic Landweber method is shown to converge by means of the L^1 weak compactness of the sublevel sets of the Shannon entropy - see, e.g. the use of Lemma 2.2 (iii) in the proof of Proposition 3.3 (iii). Another difference to the setting studied in our paper is that we use a constant parameter $t_k = 1/c$, which would not satisfy the assumption tipically employed for the mirror descent algorithm: $\sum_{k \in \mathbb{N}} t_k^2 < \infty$.

The first order methods analyzed by [5] in a finite dimensional setting are also closely related to the MDA und thus, to the finite dimensional counterpart of the entropic Landweber algorithm. A Lipschitz-like condition as in Subsection 2.2 becomes an essential tool in that study. However, the so-called assumption **H** requiring convergence $x^k \to x$ from a Bregman distance convergence of the form $D_h(x, x^k) \to 0$ does not hold in the entropy infinite dimensional framework, as mentioned in the previous paragraph.

Last but not least, we provide additionally error estimates for the iterates under a solution regularity condition.

3 Convergence of the Entropic Landweber Method

In the following we consider the iterative method defined by (4), where d is the Kullback-Leibler divergence given by (7).

Noticing that A^* maps to $L^\infty(\Omega)$, we can equivalently rewrite the minimization in (4) in the form of Proposition 2.3, which implies the following result.

Proposition 3.1 Let $u_0 \in \mathrm{dom}\, \partial f$. Then there exists a unique minimizer in (4) for any $k \geq 0$, given by $(\lambda = \frac{1}{c})$

$$u_{k+1} = u_k c_k^m e^{\lambda A^*(y-Au_k)}, \qquad c_k^m = \begin{cases} 1 & \text{if } m = 0, \\ \frac{1}{\int_\Omega u_k e^{\lambda A^*(y-Au_k)}\, dt} & \text{if } m = 1, \end{cases} \tag{14}$$

which further satisfies $u_{k+1} \in \mathrm{dom}\, \partial f$.

Note that, from pointwise manipulation of (14) we rigorously obtain the first-order optimality condition for the variational problem in each step, i.e.,

$$\ln u_{k+1} = \ln u_k + \ln c_k^m + \lambda A^*(y - Au_k), \tag{15}$$

where $\ln c_k^0 = 0$ and $\ln c_k^1$ are to be interpreted as a Lagrange multiplier for the integral constraint.

Remark 3.2 In the latter case, this constant term is orthogonal to all functions of the form $w - v$, since $\int_\Omega \ln c_k^m [w(t) - v(t)]\, dt = \ln c_k^m \int_\Omega [w(t) - v(t)]\, dt = 0$ when $\chi_1(w) = \chi_1(v) = 0$. Since most estimates below for iterates will be based on taking duality products of (15) with such functions, they can be carried out in the same way for $m = 0$ and $m = 1$.

The analysis of the above method ressembles the one for proximal point methods, which is apparent from rewriting (4) as

$$u_{k+1} \in \arg\min_u \left\{ \frac{1}{2}\|Au - y\|^2 + \chi_m(u) + D(u, u_k) \right\}. \tag{16}$$

However, the quantity D is neither a metric distance nor necessarily a Bregman distance of a convex function, rather a weighted difference of Bregman distances. This and the involved Kullback-Leibler divergence in an infinite dimensional setting require thus a careful investigation.

We choose $u_0 \in \mathrm{dom}\, \partial f$ such that $\xi_0 := 1 + \ln u_0 \in \mathcal{R}(A^*)$, that is $\xi_0 = \lambda A^* w_0$ for some $w_0 \in Y$, and denote

$$v_k = w_0 + \sum_{j=0}^{k-1} (y - Au_j) \tag{17}$$

for $k \geq 1$. Then (15) can be expressed as $\xi_k = \xi_{k-1} + \ln c_{k-1}^m + \lambda A^*(y - Au_{k-1})$ which implies

$$\xi_k = \xi_0 + \sum_{j=0}^{k-1} \ln c_j^m + \lambda A^* \left(\sum_{j=0}^{k-1} (y - Au_j) \right) = \sum_{j=0}^{k-1} \ln c_j^m + \lambda A^* v_k. \tag{18}$$

For instance, one can consider $u_0 = e^{-1}$ (constant function), $\xi_0 = 0$ and $w_0 = 0$. We show next that the entropic Landweber method converges in the exact data case.

Proposition 3.3 Let $A : L^1(\Omega) \to Y$ be a bounded linear operator which satisfies (12) and such that the operator equation $Au = y$ has a positive solution z verifying $\chi_m(z) = 0$ if $m = 1$. Let $u_0 \in \mathrm{dom}\, \partial f$ be an arbitrary starting element such that $1 + \ln u_0 \in \mathcal{R}(A^*)$. Moreover, let $\chi_m(u_0) = 0$ if $m = 1$. Then the following statements are true:

(i) The residual $\|Au_k - y\|$ decreases monotonically.

(ii) The term $D(z, u_k)$ decreases monotonically.

(iii) The sequences $\{u_k\}_{k \in \mathbb{N}}$ generated by the iterative method (14) converge weakly on subsequences in $L^1(\Omega)$ to solutions of the equation $Au = y$, with $\chi_m(u) = 0$ if $m = 1$.

Proof: We will use the proximal point method techniques in order to prove the statements, by taking care of the fact that D is a nonnegative functional satisfying $D(u, u) = 0$ for any u in this function's domain.

(i) We have for all $k \in \mathbb{N}$,

$$\frac{1}{2}\|Au_{k+1} - y\|^2 + D(u_{k+1}, u_k) \le \frac{1}{2}\|Au_k - y\|^2,$$

which implies that the sequence $\{\|Au_k - y\|\}$ is nonincreasing, since $D(u_{k+1}, u_k) \ge 0$.

(ii) Consider first the case $m = 0$. Let z verify $Az = y$ and denote

$$a_k = D(z, u_{k+1}) + D(u_{k+1}, u_k) - D(z, u_k), \quad \xi_k = 1 + \ln u_k \in \partial f(u_k).$$

By using (15), one has for all $k \in \mathbb{N}$:

$$
\begin{aligned}
a_k &= cd(z, u_{k+1}) - \frac{1}{2}\|Au_{k+1} - y\|^2 + cd(u_{k+1}, u_k) - \frac{1}{2}\|Au_{k+1} - Au_k\|^2 \\
&\quad -cd(z, u_k) + \frac{1}{2}\|Au_k - y\|^2 \\
&= \langle A^*(Au_k - y), z - u_{k+1}\rangle + \langle Au_{k+1} - y, Au_k - Au_{k+1}\rangle \\
&= \langle Au_k - y, y - Au_{k+1}\rangle + \langle Au_{k+1} - y, Au_k - Au_{k+1}\rangle \\
&= -\|y - Au_{k+1}\|^2 \le 0.
\end{aligned}
$$

This implies the typical inequality for a proximal-like method:

$$\frac{1}{2}\|Au_{k+1} - y\|^2 + D(z, u_{k+1}) + D(u_{k+1}, u_k) \le D(z, u_k) \tag{19}$$

which yields the conclusion.

(iii) Let $m = 0$. Inequality (19) leads to

$$\sum_{j=0}^{k} \frac{1}{2}\|Au_{j+1} - y\|^2 + D(z, u_{k+1}) + \sum_{j=0}^{k} D(u_{j+1}, u_j) \le D(z, u_0), \tag{20}$$

which yields

$$\frac{k+1}{2}\|Au_{k+1} - y\|^2 \le \sum_{j=0}^{k} \frac{1}{2}\|Au_{j+1} - y\|^2, \tag{21}$$

since the sequence $\{\|Au_k - y\|^2\}_{k \in \mathbb{N}}$ is monotone. We show now that $\{f(u_k)\}_{k \in \mathbb{N}}$ is bounded. To this end, due to nonnegativity of $D(z, u_k)$, $k \in \mathbb{N}$, and to (18), one has

$$
\begin{aligned}
cf(u_k) + \frac{1}{2}\|Au_k - y\|^2 &\le cf(z) + c\langle \xi_k, u_k - z\rangle = cf(z) + \langle w_0 + \sum_{j=0}^{k-1}(y - Au_j), Au_k - y\rangle \\
&\le cf(z) + \|w_0\|\|Au_k - y\| + \sum_{j=0}^{k-1}\frac{1}{2}\|y - Au_j\|^2 + \frac{k}{2}\|Au_k - y\|^2.
\end{aligned}
$$

The right hand side is bounded by (20) and (21), thus ensuring boundedness of $\{f(u_k)\}_{k \in \mathbb{N}}$. Consequently, there exists a subsequence $\{u_l\}_{l \in \mathbb{N}}$ in $\text{dom}\,\partial f$ which is L^1-weakly convergent to some $u \in \text{dom}\, f$, cf. Lemma 2.2. Then one has $Au_l \to Au$ weakly in Y and moreover $Au_l \to y$ in the Y-norm since

$$\frac{1}{2}\|Au_{k+1} - y\|^2 \le D(z, u_k) - D(z, u_{k+1}) \to 0$$

7

due to inequality (19) and to monotonicity of $\{D(z, u_k)\}_{k \in \mathbb{N}}$. Hence, u satisfies $Au = y$.

The proof of the statements above for the case $m = 1$ is similar, the main difference being the optimality condition (15) with c_k^m satisfying $c(\xi_k - \xi_{k+1}) = A^*(Au_k - y) + c \ln c_k^m$. In more detail, the term $\langle c \ln c_k^m, z - u_{k+1} \rangle$ vanishes for $z, u_{k+1} \in dom\, \chi_1$ when evaluating a_k and does not influence further calculations, while other terms containing c_k^m behave similarly in the remaining argumentation (see also Remark 3.2). □

Let us consider now the iterative method based on the noisy data, that is

$$u_{k+1} \in \arg\min_u \left\{ \langle Au - y^\delta, Au_k - y^\delta \rangle + cd(u, u_k) + \chi_m(u) \right\}. \tag{22}$$

We propose first a discrepancy principle for stopping the algorithm in this case. Before detailing how it works, denote

$$v_k = w_0 + \sum_{j=0}^{k-1} (y^\delta - Au_j) \tag{23}$$

for $k \geq 1$. Then the optimality condition for (22) yields

$$\xi_k = \xi_0 + \sum_{j=0}^{k-1} \ln c_j^m + \lambda A^*\left(\sum_{j=0}^{k-1} (y^\delta - Au_j)\right) = \sum_{j=0}^{k-1} \ln c_j^m + \lambda A^* v_k. \tag{24}$$

Proposition 3.4 Assume that $A : L^1(\Omega) \to Y$ is a bounded linear operator which satisfies (12) and such that the operator equation $Au = y$ has a positive solution z verifying $\chi_m(z) = 0$ if $m = 1$. Let $y^\delta \in Y$ be noisy data satisfying $\|y - y^\delta\| \leq \delta$, for some noise level δ. Let $u_0 \in \text{dom}\, \partial f$ be an arbitrary starting element with the properties $1 + \ln u_0 \in \mathcal{R}(A^*)$ and $\chi_m(u_0) = 0$ if $m = 1$. Then

(i) The residual $\|Au_k - y^\delta\|$ decreases monotonically and the following inequalities hold

$$\frac{1}{2}\|y^\delta - Au_{k+1}\|^2 + D(z, u_{k+1}) + D(u_{k+1}, u_k) \leq \frac{\delta^2}{2} + D(z, u_k), \;\; k \in \mathbb{N}, \tag{25}$$

$$\|y^\delta - Au_k\|^2 \leq \delta^2 + \frac{2D(z, u_0)}{k}, \;\; k \geq 1. \tag{26}$$

(ii) The term $D(z, u_k)$ decreases as long as $\|y^\delta - Au_k\|^2 > \delta^2$.

(iii) The index $k_*(\delta)$ defined by

$$k_*(\delta) = \min\{k \in \mathbb{N} : \|Au_k - y^\delta\| < \sqrt{\tau}\delta\}, \;\; \text{for some } \tau > 1, \tag{27}$$

is finite.

(iv) There exists a weakly convergent subsequence of $\{u_{k_*(\delta)}\}_\delta$ in $L^1(\Omega)$. If $\{k_*(\delta)\}_\delta$ is unbounded, then each limit point is a solution of $Au = f$.

Proof: We consider only the case $m = 0$, since for $m = 1$ one can use similar arguments, as explained in Remark 3.2.

First part of (i) follows by the definition of the iterative procedure. For proving the remaining inequalities in (i) and (ii), we consider as in the previous proof

$$
\begin{aligned}
a_k &= cd(z, u_{k+1}) - \frac{1}{2}\|Au_{k+1} - y\|^2 + cd(u_{k+1}, u_k) - \frac{1}{2}\|Au_{k+1} - Au_k\|^2 \\
&\quad -cd(z, u_k) + \frac{1}{2}\|Au_k - y\|^2 \\
&= c\langle \xi_k - \xi_{k+1}, z - u_{k+1} \rangle + \langle Au_{k+1} - y, Au_k - Au_{k+1} \rangle \\
&= \langle Au_k - y^\delta, y - Au_{k+1} \rangle + \langle Au_{k+1} - y, Au_k - Au_{k+1} \rangle \\
&= \langle y - Au_{k+1}, Au_{k+1} - y^\delta \rangle = -\|y^\delta - Au_{k+1}\|^2 + \langle y - y^\delta, Au_{k+1} - y^\delta \rangle \\
&\leq -\|y^\delta - Au_{k+1}\|^2 + \frac{\delta^2}{2} + \frac{1}{2}\|y^\delta - Au_{k+1}\|^2 \\
&= -\frac{1}{2}\|y^\delta - Au_{k+1}\|^2 + \frac{\delta^2}{2}.
\end{aligned}
$$

8

Inequality (26) can be obtained by writing (25) for $k = 0, ..., n-1$ und calculating the telescope sum:

$$\frac{n\|y^\delta - Au_n\|^2}{2} \leq \frac{1}{2}\sum_{k=0}^{n-1}\|y^\delta - Au_{k+1}\|^2 + D(z, u_n) \leq D(z, u_0) + \frac{\delta^2 n}{2}.$$

Moreover, (ii) follows from (25) by neglecting the nonnegative term $D(u_{k+1}, u_k)$. (iii) follows from (26) and the definition of $k_*(\delta)$:

$$\frac{k_*(\delta)\tau\delta^2}{2} \leq \frac{1}{2}\sum_{k=0}^{k_*(\delta)-1}\|y^\delta - Au_k\|^2 + D(z, u_{k_*(\delta)}) \leq D(z, u_0) + \frac{\delta^2 k_*(\delta)}{2}, \tag{28}$$

which implies

$$k_*(\delta) \leq \frac{2D(z, u_0)}{(\tau-1)\delta^2}. \tag{29}$$

(iv) can be shown similarly to Proposition 3.3 (iii). Due to nonnegativity of $D(z, u_k)$ for any $k \in \mathbb{N}$ and to (24), one has

$$
\begin{aligned}
cf(u_k) + \frac{1}{2}\|Au_k - y\|^2 \ &\leq\ cf(z) + c\langle\xi_k, u_k - z\rangle \\
&=\ cf(z) + \langle w_0 + \sum_{j=0}^{k-1}(y^\delta - Au_j), Au_k - y\rangle \\
&\leq\ cf(z) + \|w_0\|\|Au_k - y\| + \sum_{j=0}^{k-1}\|y^\delta - Au_j\|\|Au_k - y\| \\
&\leq\ cf(z) + \|w_0\|\|Au_k - y^\delta\| + \delta\|w_0\| + \delta\sum_{j=0}^{k-1}\|y^\delta - Au_j\| \\
&\quad + \sum_{j=0}^{k-1}\|y^\delta - Au_j\|\|Au_k - y^\delta\| \\
&\leq\ cf(z) + \|w_0\|\|Au_k - y^\delta\| + \delta\|w_0\| + \sum_{j=0}^{k-1}\frac{1}{2}\|y - Au_j\|^2 + \frac{k\delta^2}{2} \\
&\quad + \sum_{j=0}^{k-1}\frac{1}{2}\|y - Au_j\|^2 + \frac{k}{2}\|Au_k - y^\delta\|^2.
\end{aligned}
$$

The right hand side written for $k = k_*(\delta)$ is bounded by (27), (28) the monotonicity of the residual and by (29), thus ensuring boundedness of $\{f(u_{k_*(\delta)})\}_{\delta>0}$ for δ small enough. The conclusion follows then as in the proof of Proposition 3.3 (iii). □

A convergence result can be established also in case of an a priori stopping rule with $k_*(\delta) \sim \frac{1}{\delta}$ by following the lines of Proposition 3.4 (iv).

Proposition 3.5 Assume that $A : L^1(\Omega) \to Y$ is a bounded linear operator which satisfies (12) and such that the operator equation $Au = y$ has a positive solution z verifying $\chi_m(z) = 0$ if $m = 1$. Let $y^\delta \in Y$ be noisy data satisfying $\|y - y^\delta\| \leq \delta$, for some noise level δ. Let $u_0 \in \text{dom}\,\partial f$ be an arbitrary starting element with the properties $1 + \ln u_0 \in \mathcal{R}(A^*)$ and $\chi_m(u_0) = 0$ if $m = 1$. Let the stopping index $k_*(\delta)$ be chosen of order $1/\delta$. Then $\{f(u_{k_*(\delta)})\}_\delta$ is bounded and hence, as $\delta \to 0$, there exists a weakly convergent subsequence $\{u_{k(\delta_n)}\}_n$ in $L^1(\Omega)$ whose limit is a solution of $Au = y$ in $dom\,\chi_m$ when $m = 1$. Moreover, if the solution of the equation is unique, then $\{u_{k_*(\delta)}\}_{\delta>0}$ converges weakly to the solution as $\delta \to 0$.

4 Error estimates

In this section we derive error estimates under a specific source condition (on a solution) for the entropy type penalty. We proceed first with the case of exact data on the right-hand side of the operator equation and then with the noisy data case, by employing an a priori rule for stopping the algorithm.

4.1 Exact data case

Proposition 4.1 Assume that $A : L^1(\Omega) \to Y$ is a bounded linear operator which satisfies (12) and such that the operator equation $Au = y$ has a positive solution z verifying $\chi_m(z) = 0$ if $m = 1$. Let $u_0 \in \operatorname{dom} \partial f$ be an arbitrary starting element with the properties $1 + \ln u_0 \in \mathcal{R}(A^*)$ and $\chi_m(u_0) = 0$ if $m = 1$. Additionally, let the following source condition hold:

$$1 + \ln z \in \mathcal{R}(A^*). \tag{30}$$

Then one has

$$d(z, u_k) = O(1/k). \tag{31}$$

Moreover, $\|u_k - z\|_1 = O(1/\sqrt{k})$ if $m = 1$.

Proof: We consider only the case $m = 0$ (similar arguments for the other case, cf. Remark 3.2).

First, we symmetrize D by considering $D^s(x,y) = D(x,y) + D(y,x)$. Let $\xi = 1 + \ln z = \lambda A^* v$ for some $v \in Y$.

One can use similar techniques as in [11] for deriving the announced error estimates, by carefully dealing with the setting of the D distance penalty. Based on (17), one has

$$
\begin{aligned}
D^s(u_k, z) &= c\langle \xi_k - \xi, u_k - z\rangle - \|Au_k - Az\|^2 \\
&= \langle A^* v_k - A^* v, u_k - z\rangle - \|Au_k - y\|^2 \\
&= \langle v_k - v, Au_k - y\rangle - \|Au_k - y\|^2 \\
&= \langle v_k - v, v_k - v_{k+1}\rangle - \|Au_k - y\|^2 \\
&= \frac{1}{2}\|v_k - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 + \frac{1}{2}\|v_{k+1} - v_k\|^2 - \|Au_k - y\|^2 \\
&= \frac{1}{2}\|v_k - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 - \frac{1}{2}\|Au_k - y\|^2.
\end{aligned}
$$

By writing the last inequality also for $k - 1, k - 2, ..., 1$, by summing up and by combining with monotonicity of $\{D(z, u_k)\}$, one obtains

$$kD(z, u_k) \le \sum_{j=1}^{k} D(z, u_j) \le \sum_{j=1}^{k} D^s(u_j, z) \le \frac{1}{2}\|v_1 - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 - \frac{1}{2}\sum_{j=1}^{k}\|Au_j - y\|^2$$

and thus, due to (21),

$$d(z, u_k) \le \frac{\lambda}{2k}\|v_1 - v\|^2$$

holds. The announced convergence rate in the L^1-norm holds in case $m = 1$ by Lemma 2.1 (v). □

4.2 Noisy data case

Proposition 4.2 Assume that $A : L^1(\Omega) \to Y$ is a bounded linear operator which satisfies (12) and such that the operator equation $Au = y$ has a positive solution z verifying $\chi_m(z) = 0$ if $m = 1$. Let $u_0 \in \operatorname{dom} \partial f$ be an arbitrary starting element with the properties $1 + \ln u_0 \in \mathcal{R}(A^*)$ and $\chi_m(u_0) = 0$ if $m = 1$. Let $y^\delta \in Y$ be noisy data satisfying $\|y - y^\delta\| \le \delta$, for some noise level

δ. Let the stopping index $k_*(\delta)$ be chosen of order $1/\delta$. and let the source condition (30) hold. Then one has

$$d(z, u_{k_*(\delta)}) = O(\delta). \tag{32}$$

Moreover, $\|u_{k_*(\delta)} - z\|_1 = O(\sqrt{\delta})$ if $m = 1$.

Proof: Note that $c\xi - A^*Az = A^*q$ for $q = v - A^*z$. With this notation, one can show the following estimate as in Theorem 4.3 in [11]:

$$D(z, u_k) \leq \frac{\|q\|^2}{2k} + \delta\|q\| + \delta^2 k, \ \ \forall k \in \mathbb{N}.$$

Then one has

$$cd(z, u_k) \leq \frac{\|q\|^2}{2k} + \delta\|q\| + \delta^2 k + \frac{\|Au_k - y\|^2}{2} \leq \frac{\|q\|^2}{2k} + \delta\|q\| + \delta^2 k + \|Au_k - y^\delta\|^2 + \delta^2$$

which yields (32) when written for $k = k_*(\delta)$, due to (26). $\quad\square$

Establishing convergence rates by means of a discrepancy rule remains an open issue.

5 General data fidelities

Before we conclude with numerical examples, we want to emphasize that Problem (1) can easily be generalised to

$$u_{k+1} \in \arg\min_u \left\{ F_{y^\delta}(Au) + cd(u, u_k) - g(u, u_k)) + \chi_m(u) \right\}. \tag{33}$$

Here $F_{y^\delta} : Y \to [0, +\infty)$ is a more general data fidelity term that is assumed to be convex and Fréchet-differentiable and $g : L^1(\Omega) \times L^1(\Omega) \to [0, +\infty)$ is the Bregman distance with respect to the function F_{y^δ}, i.e.

$$g(u, v) := F_{y^\delta}(Au) - F_{y^\delta}(Av) - \langle A^* F'_{y^\delta}(Av), u - v \rangle.$$

Note that (33) is an instance of the Bregman proximal method [13, 20]. The update for (33) can be written, in analogy to (14), as

$$u_{k+1}^m = u_k^m c_k^m e^{-\lambda A^* F'_{y^\delta}(Au_k)}, \tag{34}$$

for $\lambda = 1/c$ and $m \in \{0, 1\}$. We want to emphasise that a more general data fidelity term that satisfies the assumptions mentioned above together with

$$cd(u, v) - g(u, v) \geq 0, \tag{35}$$

for all $u, v \in dom f$, is no restriction in terms of Fejér-monotonicity. In analogy to [9, Lemma 6.11] we can conclude

$$D(z, u_{k+1}) \leq D(z, u_k)$$

for all $k < k_*(\delta)$, with $k_*(\delta)$ chosen according to a modified version of (27) that reads as

$$k_*(\delta) = \min\{k \in \mathbb{N} : F_{y^\delta}(Au_k) < \delta\}. \tag{36}$$

However, we can also derive a monotonicity result for d directly. First of all we observe that (35) implies

$$-\langle A^* F'_{y^\delta}(Au_k), u_{k+1} - u_k \rangle - cd(u_{k+1}, u_k) \leq F_{y^\delta}(Au_k) - F_{y^\delta}(Au_{k+1}).$$

11

Inserting (34) into the inequality above then yields

$$c\langle \ln u_{k+1} - \ln u_k, u_{k+1} - u_k \rangle - cd(u_{k+1}, u_k) \leq F_{y^\delta}(Au_k) - F_{y^\delta}(Au_{k+1}) + c\langle \ln c_k^m, u_{k+1} - u_k \rangle.$$

As mentioned earlier in Section 3, we either have $\ln(c_k^m) = 0$ for $m = 0$, or orthogonality of $\ln(c_k^m)$ to all functions of the form $w - v$ with $\chi_m(w) = \chi_m(v) = 0$ for $m = 1$. Since $d(u_{k+1}, u_k) + d(u_k, u_{k+1}) = \langle \ln u_{k+1} - \ln u_k, u_{k+1} - u_k \rangle$, we therefore estimate

$$cd(u_k, u_{k+1}) \leq F_{y^\delta}(Au_k) - F_{y^\delta}(Au_{k+1}).$$

With the three-point identity we then observe

$$
\begin{aligned}
d(z, u_{k+1}) - d(z, u_k) &= -\langle \ln u_{k+1} - \ln u_k, z - u_{k+1} \rangle - d(u_{k+1}, u_k) \\
&= -\langle \ln u_{k+1} - \ln u_k, z - u_k \rangle + \langle \ln u_{k+1} - \ln u_k, u_{k+1} - u_k \rangle \\
&\quad - d(u_{k+1}, u_k) \\
&= -\langle \ln u_{k+1} - \ln(u_k), z - u_k \rangle + d(u_k, u_{k+1}) \\
&= \langle \lambda A^* F'_{y^\delta}(Au_k), z - u_k \rangle + d(u_k, u_{k+1}) \\
&\leq \lambda \left(F_{y^\delta}(Az) - F_{y^\delta}(Au_k) \right) + d(u_k, u_{k+1}) \\
&\leq \lambda \left(F_{y^\delta}(Az) - F_{y^\delta}(Au_{k+1}) \right).
\end{aligned}
$$

Together with (36) we can then conclude

$$d(z, u_{k+1}) < d(z, u_k)$$

for $k < k^*(\delta)$.

6 Example Problems

We finally discuss several types of problems that satisfy the conditions used in the analysis and present numerical illustrations for some of these situations.

6.1 Integral Equations

Let $\Omega \subset \mathbb{R}^d$ and $\tilde{\Omega} \subset \mathbb{R}^{\tilde{d}}$ be open and bounded sets and let $k \in L^\infty(\tilde{\Omega} \times \Omega)$. Then the integral operator

$$A : L^1(\Omega) \to L^2(\tilde{\Omega}), \qquad u \mapsto \int_\Omega k(\cdot, y) u(y) \, dy \tag{37}$$

is a well-defined and bounded linear operator. Thus, the convergence analysis is applicable due to Lemma 2.4.

We mention that in the case of k being a nonnegative function, and hence A and A^* preserving nonnegativity, standard schemes preserving nonnegativity are available. In particular for k including negative entries, the entropic Landweber scheme offers a straightforward alternative, since it does not depend on the positivity preservation of A respectively its adjoint. For comparison we consider the EM-Algorithm

$$u_{k+1} = \frac{u_k}{A^* 1} A^* \left(\frac{y}{Au_k} \right)$$

and the projected Landweber iteration

$$u_{k+1} = (u_k - \tau A^*(Au_k - y))_+.$$

We implement the forward operator by discretization of u on a uniform grid and a trapezoidal rule for integration. We use the following examples of kernels and initial values, all on $\Omega = (0, 1)$, the first two being standard test examples used in the literature on maximum entropy methods (cf. [1])
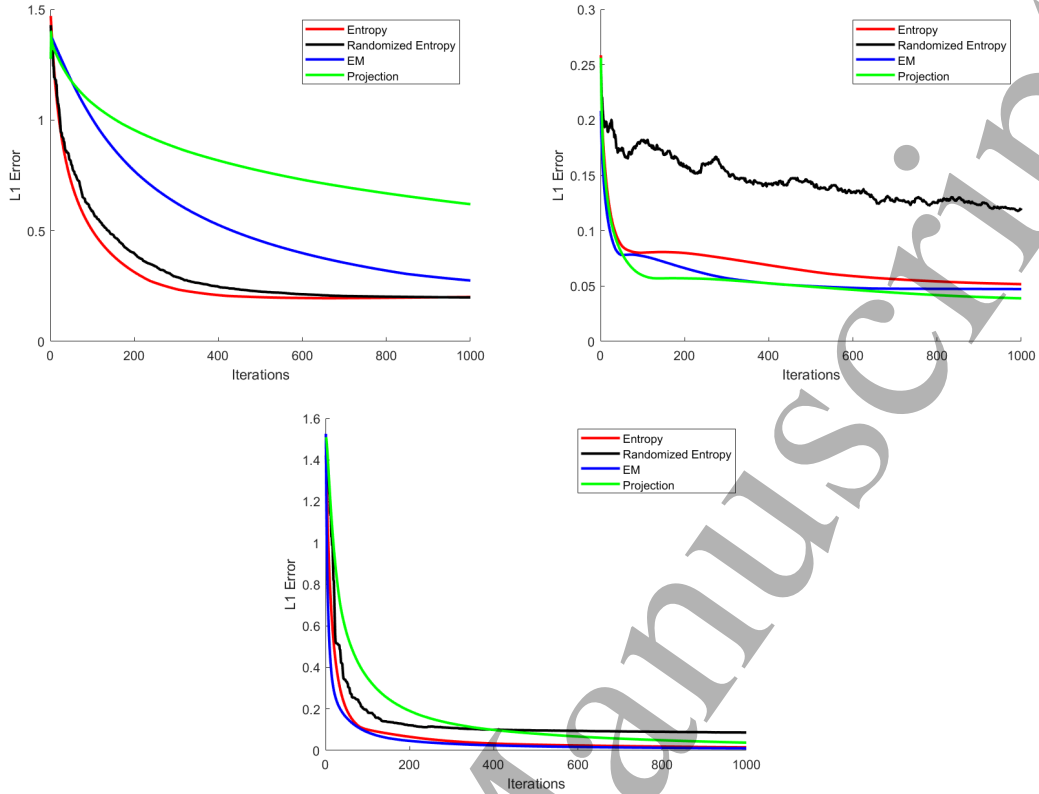
12

Figure 1: Results of the three test cases, L^1 error plotted vs. iteration number.

1. Kernel $k_1(x, y) = e^{xy}$, exact solution $z_1(x) = e^{-\frac{x^2}{2\sigma^2}}$

2. Kernel $k_2(x, y) = 3e^{-\frac{(x-y)^2}{0.04}}$, exact solution

$$z_2(x) = 1 - 0.9e^{-\frac{(x-0.1)^2}{2\sigma^2}} - 0.3e^{-\frac{(x-0.3)^2}{2\sigma^2}} - 0.5e^{-\frac{(x-0.5)^2}{2\sigma^2}} - 0.2e^{-\frac{(x-0.7)^2}{2\sigma^2}} - 0.7e^{-\frac{(x-0.9)^2}{2\sigma^2}}.$$

3. Kernel $k_3(x, y) = 1$ if $x \geq y$ and $k_3(x, y) = 0$ else, exact solution $z_3(x) = e^{-\frac{x^2}{2\sigma^2}}$.

In all examples, we have chosen $\sigma^2 = 0.01$ and a constant intial value u^0. In order to illustrate the behaviour of the iteration methods we plot the the error $\|u^k - z\|_{L^1}$ vs. the iteration number k in Figure 1.

We observe that the entropic Landweber is at least competitive to the other schemes in all examples, it outperforms the EM and projection method in the first example, which is a combination of severe ill-posedness with an exact solution having many entries close to zero (which is a particularly difficult case for the EM algorithm).

In the second case, again severely ill-posed, the projected Landweber iteration performs better, mainly due to the strong initial decrease when the solution is positive and no projection is applied.

The third case corresponding to numerical differentiation, i.e. a very mildly ill-posed problem, is characterized by fast convergence of the schemes, but again the projection method converges significantly slower. For comparison we also include the stochastic version of the entropic Landweber method, with only one equation used in each iteration step, hence a highly efficient computation. That is, the operator A is divided in M blocks $A = (A_1, A_2, \ldots, A_M)^T$, and the data y are partitioned in the same way: $y = (y_1, y_2, \ldots, y_M)$. With $J(k)$ a discrete uniform random variable in $\{1, \ldots, M\}$, we compute the iterates

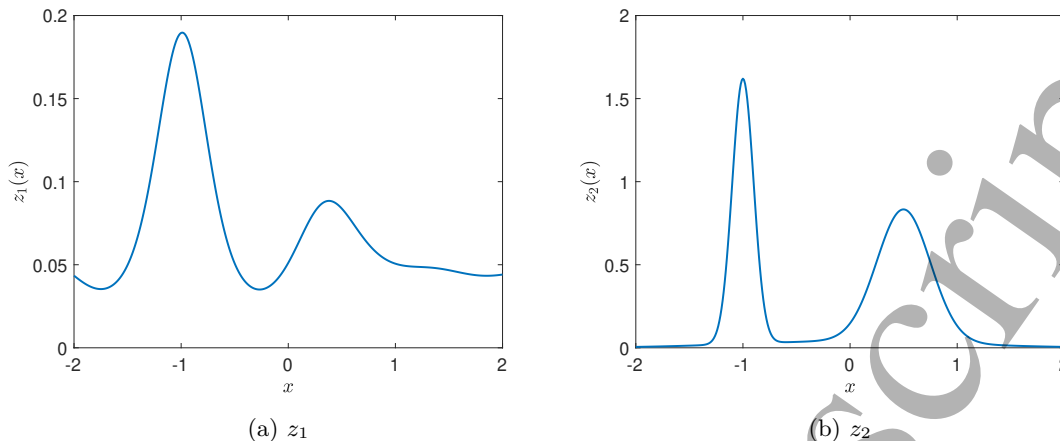$$u_{k+1} = u_k c_k^m e^{\lambda M A_{J(k)}^* (y_{J(k)} - A_{J(k)} u_k)}, \quad k \in \mathbb{N}.$$

13

(a) z_1

(b) z_2

Figure 2: The two ground truth functions z_1 and z_2 as defined in Section 6.2. Note that z_1 by construction satisfies (30) while z_2 does not satisfy (30).

The initial convergence curve is similar to the other method, with much lower computational effort, then the asymptotic convergence close to the exact solution becomes significantly slower. Hence, it might be very attractive to use the stochastic version at least for the first phase of the reconstruction.

6.2 Discrete sampling of continuous probability densities

Suppose that our forward operator is the Fourier integral of a real-valued function evaluated at discrete samples $\xi_1, \ldots, \xi_n$ on a compact domain $\Omega \subset I\!\!R^d$, i.e.

$$A : L^1(\Omega) \to \mathbb{C}^n, \qquad u \mapsto \left((2\pi)^{-\frac{d}{2}} \int_\Omega u(x) \, e^{-ix \cdot \xi_j} \, dx \right)_{1 \le j \le n}.$$

Then the adjoint operator A^* that satisfies $\sum_{j=1}^n (Au)_j \overline{v_j} = \int_\Omega u(x) \overline{(A^*v)(x)} dx = \int_\Omega u(x)(A^*v)(x) dx$ is given as

$$A^* : \mathbb{C}^n \to L^\infty(\Omega), \qquad v \mapsto \mathrm{Re} \left((2\pi)^{-\frac{d}{2}} \sum_{j=1}^n v_j \, e^{ix \cdot \xi_j} \right),$$

where Re denotes the real part of a complex function. For this choice of A the iterates of (4) read

$$u_{k+1}(x) = c_k^m u_k(x) e^{\lambda \mathrm{Re} \left((2\pi)^{-\frac{d}{2}} \sum_{j=1}^n \left(y_j - (2\pi)^{-\frac{d}{2}} \int_\Omega u_k(t) e^{-it \cdot \xi_j} \, dt \right) e^{ix \cdot \xi_j} \right)},$$

where c_k^m is defined as in (14). Note that this update can also be written as

$$u_k(x) = c_{k-1}^m \ldots c_0^m u_0(x) e^{\lambda \mathrm{Re} \left((2\pi)^{-\frac{d}{2}} \sum_{j=1}^n (k \, y_j - \tilde{y}_{jk}) \, e^{ix \cdot \xi_j} \right)},$$

for

$$\tilde{y}_{jk} := (2\pi)^{-\frac{d}{2}} \sum_{l=0}^{k-1} \int_\Omega u_l(t) e^{-it \cdot \xi_j} \, dt.$$

14

(a) u^{k^*}

(b) Energy decrease of objective

(c) u^{k^*}
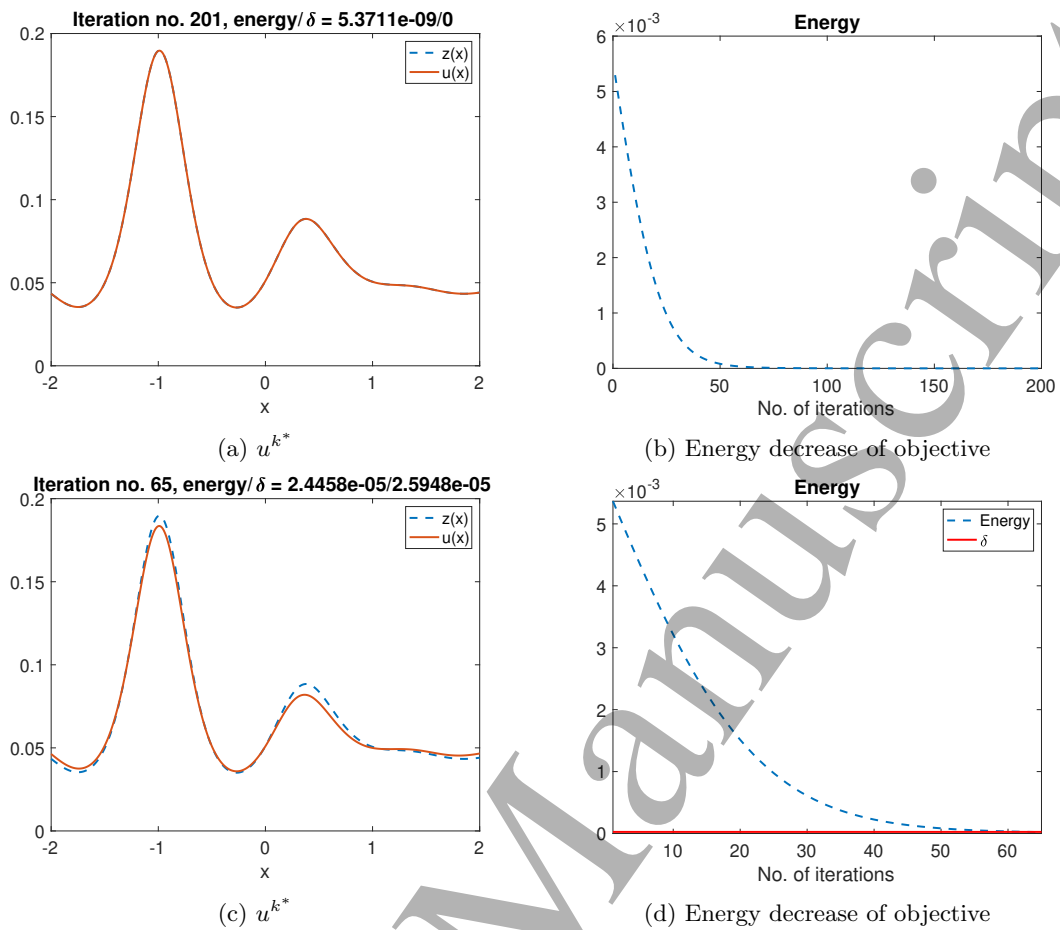
(d) Energy decrease of objective

Figure 3: Results of algorithm (4) for data measurements of the form (38) based on the underlying function z_1 as visualised in Figure 2a. Figure 3a shows the ground truth function z_1 and u^{k^*} under the assumption of $\sigma = 0$ in (38), whereas Figure 3b visualises the monotonic energy decrease over the course of the iterations. Figure 3c and Figure 3d show the same results under the assumption of $\sigma = 1/50$ in (38).

Due to $\tilde{y}_{j(k+1)} = \tilde{y}_{jk} + (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} u_k(t) e^{-it \cdot \xi_j} \, dt$, this formulation has the advantage that the numerical costs for evaluating the integrals remains constant.

In the following we consider a one-dimensional setting ($d = 1$) with $\Omega = [-a, a]$ for $a = 10$, where we measure $n = 16$ samples $\{y_j\}_{j=1}^n$ of the Fourier integral for coordinates $\xi_j = (2\pi(j-1))/n$, $j = 1, \ldots, n$. We assume that these measurements are of the form

$$y_j = \frac{1}{\sqrt{2\pi}} \int_{-10}^{10} z(t) \, e^{-it\xi_j} \, dt + n_j \,, \tag{38}$$

for a function $z \in L^1([-10, 10])$ and where $n_j \in \mathcal{N}(0, \sigma^2)$ are normal-distributed random variables with mean zero and variance σ^2, for all $j \in \{1, \ldots, n\}$. We consider numerical experiments for two choices of z. The first choice is the following Gaußian-mixture model,

$$\tilde{z}(x) := \sum_{l=1}^{3} c_l \, g(x, \mu_l, \sigma_l) \,,$$

15

(a) u^{k^*}

(b) Energy decrease of objective
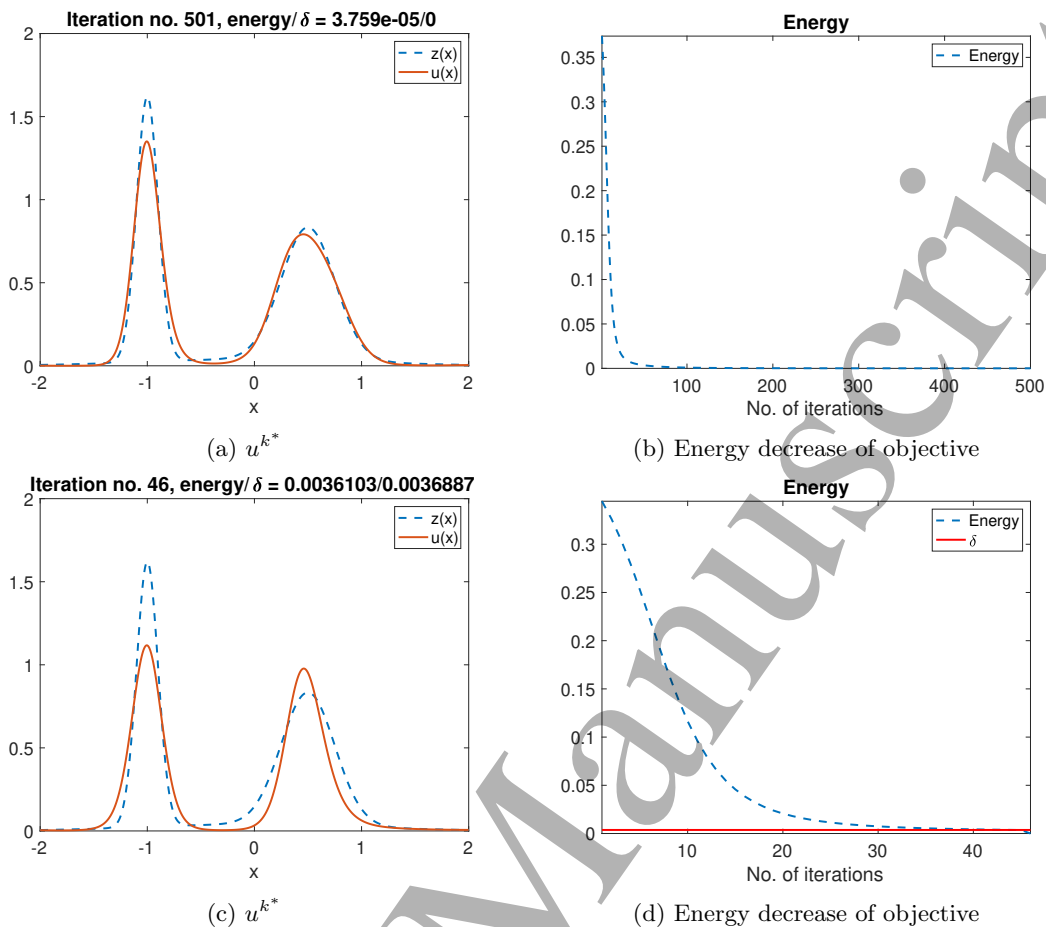

(c) u^{k^*}

(d) Energy decrease of objective

Figure 4: Results of algorithm (4) for data measurements of the form (38) based on the underlying function z_2 as visualised in Figure 2b. Figure 4a shows the ground truth function z_2 and u^{k^*} under the assumption of $\sigma = 0$ in (38), whereas Figure 4b visualises the monotonic energy decrease over the course of the iterations. Figure 4c and Figure 4d show the same results under the assumption of $\sigma = 1/500$ in (38).

that is constructed as a linear combination of three normalised Gaußians, i.e.

$$g(x, \mu, \sigma) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Note that $\tilde{z}$ does not satisfy the source condition (30), which is why we design a second function

$$z(x) := c \operatorname{Re}\left(\frac{1}{\sqrt{2}} \sum_{j=1}^{n} \left(\int_{\Omega} \tilde{z}(t) e^{-it\xi_j} \, dt \right) e^{ix\xi_j} \right),$$

where c is chosen to ensure $\int_{-10}^{10} z(x) \, dx = 1$, which by construction satisfies (30). We design two functions z_1 and z_2; z_1 is defined as $z_1 := z$ for $\tilde{z}$ with means $\mu_1 = 0, \mu_2 = -1$ and $\mu_3 = 1/2$, standard deviations $\sigma_1 = 1, \sigma_2 = 1/10$ and $\sigma_3 = 1/4$, and coefficients $c_1 = 1/10, c_2 = 3/5$ and $c_3 = 3/10$. The function $z_2 := \tilde{z}$ has the same means and standard deviations as $\tilde{z}$ in the previous example, but coefficients $c_1 = 1/10, c_2 = 2/5$ and $c_3 = 1/2$ instead. Both functions are visualised in Figure 2. Subsequently we create data samples via (38) with noise levels $\sigma = 0$ and $\sigma = 1/500$ for z_1, respectively $\sigma = 0$ and $\sigma = 1/50$ for z_2.
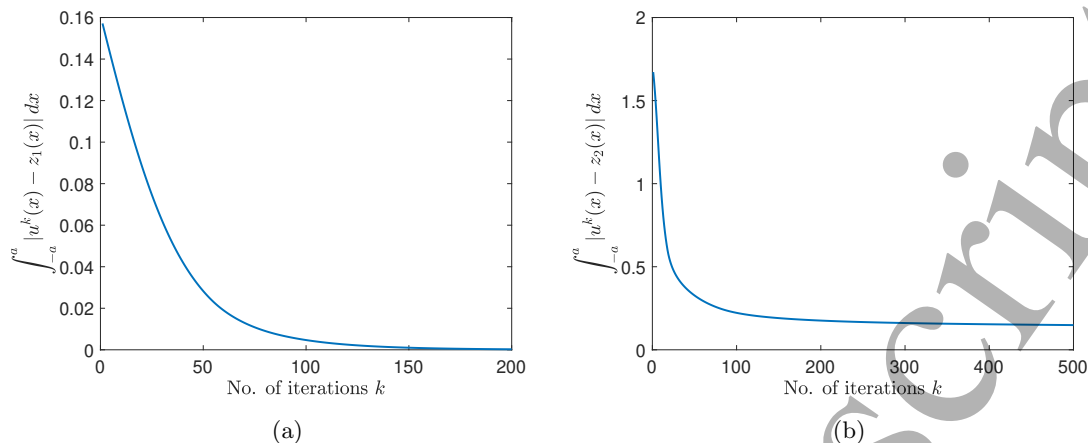
16

Figure 5: The L^1-norm of the difference of the iterates u^k and z_1 (Fig. 5a) and z_2, respectively (Fig. 5b), for the case of exact data ($\sigma = 0$).

In the following we run the entropic Landweber method (14) for $j = 1$, $\tau = 9/(10\sqrt{2\pi})$ and with the initial function

$$u_0(x) := \frac{1}{20} \begin{cases} 1 & x \in [-10, 10] \\ 0 & \text{else} \end{cases},$$

either until the discrepancy principle (27) is violated (for $\tau = 1$) or until we reach a certain maximum number of iterations. We first investigate the algorithm for the function z_1 as seen in Figure 2a, for perfect data ($\sigma = 0$) and for noisy data $\sigma = 1/500$. For perfect data we run the algorithm for 201 iterations and observe that we are converging towards z_1 as can be seen in Figure 3a as well as in Figure 5a, which is a numerical confirmation of Proposition 4.1. For the non-trivial noise-level $\sigma = 1/500$ the algorithm stops after 65 iterations according to the discrepancy principle (Figure 3c).

To conclude, we run the same numerical experiments for z_2 as shown in Figure 2b. As we mentioned earlier, (30) is violated and even for perfect data (i.e. $\sigma = 0$) we cannot expect the results of Proposition 4.1 to hold true. It can be seen in Figure 4a that u^k does not seem to converge towards z_2 despite a decrease of the objective to values in the order of 10^{-5}. In fact, if we compare the L^1-norm of the difference $u^k - z_2$, we also see in Figure 5b that u^k does not seem to converge towards z_2. For noisy data with $\sigma = 1/500$ the discrepancy principle is violated after 46 iterations, with its result being visualised in Figure 4c.

6.3 Initial Densities for Stochastic Differential Equations

An interesting problem in several applications, e.g. in data assimilation scenarios (cf. e.g [21]), is the reconstruction of the initial density for a system evolving via stochastic differential equations with drift b and volatility a. The density evolves via the Fokker-Planck equation (cf. [19])

$$\partial_t \rho(x, t) + \nabla \cdot (\rho(x, t)b(x, t)) = \frac{1}{2}\Delta(a(x, t)^2 \rho(x, t)) \tag{39}$$

in $\Omega \times (0, T)$ with no-flux boundary conditions. Under appropriate smoothness conditions on a and b as well as positivity of a is is well-known that the Fokker-Planck equation has a unique nonnegative solution $\rho \in C(0, T; L^1(\Omega))$ for nonnegative initial values $u \in L^1(\Omega)$ such that $\int_\Omega u \ln u \, dx < \infty$. In problems related to reconstructing u it is hence rather natural to use methods penalizing its entropy.

17

The forward operator A maps the initial density to indirect measurements of the density ρ over time, e.g. moments or local integrals. Parametrizing the measurements by values σ in a bounded set σ we obtain

$$A : L^1(\Omega) \to L^2((0,T) \times \Sigma), \qquad u \mapsto \int_\Omega k(\sigma, y) \rho(y,t) \, dy. \tag{40}$$

It is well-known that Fokker-Planck equations satisfy an L^1-contractivity property on the domain of the entropy functional (cf. [23]) i.e. for ρ_i denoting the solution with initial value

$$\|\rho_1(t) - \rho_2(t)\|_{L^1} \le \|u_1 - u_2\| \tag{41}$$

for almost all $t \in (0,T)$. Thus, the map $u \mapsto \rho$ is Lipschitz continuous with unit modulus when considered as a map into $L^\infty(0,T; L^1(\Omega))$ on the domain of the entropy. Hence, if $k \in L^\infty(\Sigma \times \Omega)$ we can easily verify that the operator A satisfies (12).

We finally mention that in the case of stationary coefficients a and b, the Fokker-Planck equation has a unique stationary solution ρ_∞ among nonnegative functions with unit mass (cf. [18]), to which it converges with exponential speed in the relative entropy (cf. [2, 3]), i.e. the Bregman distance related to the entropy functional. Hence, it is natural to use ρ_∞ as an initial value for the reconstruction of u, since we may expect them to be close in particular in the relative entropy.

7 Conclusions and remarks

In this study we have investigated a multiplicative entropic type method for ill-posed equations, which preserves nonnegativity of the iterates. Historically, the underlying strategy has spreading roots in the inverse problems literature: Landweber iterates, surrogate functionals and linearized Bregman, to quote a few approaches. In parallel, this has been treated in different contexts in finite dimensional optimization, e.g., as a mirror descent or as a steepest descent (linearized proximal) algorithm with generalized distances, or in machine learning - as an exponentiated gradient descent algorithm for online prediction via linear models.

The closed form algorithm is shown to converge weakly in L^1 to a solution of the ill-posed problem and convergence rates are obtained by means of the Kullback-Leibler (KL) distance. All the results are quite naturally established when imposing "mean one" restriction to the unknown, while the case without restrictions relies on a norm combined with KL distance based Lipschitz condition, in which case operators satisfying it remain to be found.

Methods of this type involving other interesting fidelity terms, nonlinear operators and eventually stochastic versions and line search strategies might be considered in more detail for future research.

8 Acknowledgements

References

[1] U. Amato, W. Hughes, Maximum entropy regularization of Fredholm integral equations of the first kind, Inverse Problems 7 (1991), 793.

18

[2] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, Communications in Partial Differential Equations, 26 (2001), 43-100.

[3] A. Arnold, E. Carlen, Q. Ju, Large-time behavior of non-symmetric Fokker-Planck type equations, Communications on Stochastic Analysis 2 (2008), 11.

[4] M. Bachmayr and M. Burger, Iterative total variation schemes for nonlinear inverse problems, Inverse Problems 25 105004, 2009.

[5] H. Bauschke, J. Bolte and M. Teboulle, A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications, Mathematics of Operations Research, INFORMS (2016).

[6] A. Beck, First-Order Methods in Optimization, SIAM, 2017.

[7] A. Beck, M.Teboulle, Mirror descent and nonlinear projected subgradient methods for convex optimization, Operations Research Letters, 31, 167-175, 2003.

[8] J.D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyré, Iterative Bregman projections for regularized transportation problems, SIAM Journal on Scientific Computing, 37 (2015), A1111-A1138.

[9] M. Benning, M. Burger, Modern regularization methods for inverse problems, Acta Numerica 27 (2018), pp. 1–111.

[10] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR computational mathematics and mathematical physics, 7.3 (1967), pp. 200–217.

[11] M. Burger, E. Resmerita, L. He, Error estimation for bregman iterations and inverse scale space methods in image restoration, Computing, 81 (2007), pp. 109135.

[12] J.-F. Cai, S. Osher, Z. Shen, Convergence of the linearized Bregman iteration for l1-norm minimization, Math. Comp., 78 (2009), pp. 21272136.

[13] Y. Censor, S.A. Zenios, Proximal minimization algorithm with d-functions, Journal of Optimization Theory and Applications 73.3 (1992), pp. 451–464.

[14] Y. Censor, S.A. Zenios, Parallel Optimization: Theory, Algorithms, and Applications, Oxford University Press, New York, NY, USA, 1997.

[15] C. Clason, B. Kaltenbacher, E. Resmerita, Regularization of ill-posed problems with non-negative solutions, Splitting Algorithms, Modern Operator Theory and Applications, H. Bauschke, R. Burachik, R. Luke (eds.), Springer, to appear; arXiv:1805.01722

[16] T.T. Doan, S. Bose , D.H. Nguyen, C.L. Beck, Convergence of the iterates in mirror descent methods, IEEE Control Systems Letters, 3 (2019), 114–119.

[17] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math, 57 (11), pp.1413-1457, 2004.

[18] J. Droniou, J.L. Vazquez, Noncoercive convectiondiffusion elliptic problems with Neumann boundary conditions. Calculus of Variations and Partial Differential Equations, 34 (2009), 413-434.

[19] C. Gardiner, Stochastic Methods (Vol. 4). Springer, Berlin (2009).

[20] D.H.Gutman, J.F.Pena, A unified framework for Bregman proximal methods: subgradient, gradient, and accelerated gradient schemes. arXiv preprint arXiv:1812.10198.

[21] M. Hairer, A. Stuart, J. Voss, A Bayesian approach to data assimilation. Physica D (2005).

[22] A.N. Iusem, Steepest descent methods with generalized distances for constrained optimization, Acta Applicandae Mathematicae 46, 225-246, 1997.

[23] K.H. Karlsen, N.H. Risebro, On the uniqueness and stability of entropy solutions of nonlinear degenerate parabolic equations with rough coefficients, Discrete Contin. Dyn. Syst., 9 (2003), 10811104.

[24] J. Kivinen and M.K. Warmuth, Additive versus exponentiated gradient updates for linear prediction, Information and Computation, 132, 164, 1997.

[25] A. Nemirovski and D. Yudin, Problem Complexity and Method Efficiency in Optimization, Wiley-Intersci. Ser. Discrete Math. 15, John Wiley, New York, 1983.

[26] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Modeling and Simulation 4 (2), 460-489.

[27] G. Peyré, M. Cuturi, Computational Optimal Transport: With Applications to Data Science, Foundations and Trends in Machine Learning: 11(2019), 355-607

[28] R. Ramlau and G. Teschke, Tikhonov Replacement Functionals for Iteratively Solving Non-linear Operator Equations, Inverse Problems Vol. 21 (5): 1571-1592, 2005.

[29] E. Resmerita, Regularization of ill-posed problems in Banach spaces: convergence rates, Inverse Problems 21 (2005) 1303-1314.

[30] E. Resmerita and R. Anderssen, Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems, Mathematical Methods in the Applied Sciences, 30(13) (2007) 1527-1544.

[31] O. Scherzer, Convergence criteria of iterative methods based on Landweber iteration for solving nonlinear problems, Journal of Mathematical Analysis and Applications, 194 (1995), 911-933.

[32] F. Schöpfer, T. Schuster, A. K. Louis, An iterative regularization method for the solution of the split feasibility problem in Banach spaces. Inverse Problems, 24(5):20pp, 2008.

[33] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (2008), no. 1, 143-168.