A NEW PARAMETER-FREE REGULARIZATION METHOD FOR INVERSE PROBLEMS.

ABINASH NAYAK

ABSTRACT. Inverse problems arise in a wide spectrum of applications in fields ranging from engineering to scientific computation. Connected with the rise of interest in inverse problems is the development and analysis of regularization methods, such as TSVD, Tikhonov or iterative regularization methods (like Landweber), which are a necessity in most inverse problems due to their ill-posedness. TSVD can be used when dealing with (small) finite dimensional liner problems, but is computationally very expensive for large scale linear problems and not feasible for nonlinear problems. In such scenarios Tikhonov regularization is an attractive alternate, but it also comes with the price of calculating an optimal value for the relevant regularizing parameter; which is a non-trivial task. The best candidate in these situations turns out to be iterative regularization methods, such as the Landweber type iterations. In this paper we propose a new iterative regularization technique to solve inverse problems, without any dependence on external parameters and thus avoiding all the difficulties associated with their involvement. To boost the convergence rate of the iterative method different descent directions are provided, depending on the source conditions, which are based on specific aprior knowledge about the solution. In addition, we also provide a very efficient stopping strategy in the absence of noise information. This is very crucial since most of the regularization methods depends critically on the noise information (error norm) to determine the stopping rule, but for a real life data it is usually unknown; although there are few heuristic approach to determine the stopping criteria, but again they are not very efficient. To illustrate the effectiveness and computational efficiency of this method we apply this technique to numerically solve classical integral inverse problems, like Fredholm or Volterra (in particular, numerical differentiation) type integral equations, and compare the results with the Tikhonov and TSVD regularization methods.

1. Introduction

An inverse problem in general is a problem where the effect (output) is known but the source (input) is not, in contrast to a direct problem where we deduce the effect from the source. Mathematically, an inverse problem is often expressed as the problem of finding a \( \varphi \) (source) which satisfies the following operator equation:

\[
T \varphi = g,
\]

where \( g \) is the given data (effect) and \( T \) is some operator describing the underlying process\(^1\). Inverse problems are usually ill-posed, in the sense of not satisfying at least one of the following Hadamard conditions for well-posedness:

1. Existence of a solution
2. Uniqueness of the solution

\(^1\)the domain and range of the operator \( T \) varies depending on the problem.
Continuous dependence on the input data. Typically when solving an inverse problem, where the underlying mathematical model depicts a real life phenomenon, the condition (3) is violated due to the unboundedness of the inverse operator $T^{-1}$ (even if it exist\(^2\)) and the presence of noise in the measured data $g$. Thus for a slightly perturbed data $g_\delta$, such that $||g - g_\delta|| \leq \delta$ (small), the inverse recovery becomes unstable, $||\varphi - \varphi_\delta|| >>> \delta$ (very large), where the norms correspond to their respective spaces. To counter such instability or the ill-posed nature of inverse problems, regularization methods are employed. In the last few decades, several regularization methods have been established for linear as well as nonlinear inverse problems. In principle, there exist regularization methods that are based on Tikhonov’s approach (i.e. adding regularizing constraints), iteration methods and discretization approaches. For a comprehensive discussion on this subject we refer the reader to the rich literature, see \cite{2, 23}. For each of those methods abundant refinements/generalizations and different regularization parameter rules were established, e.g. the discrepancy principle \cite{25, 26, 24}, L-curve \cite{9, 10}, the monotone error rule \cite{27}, or, more recently, the Lepskii principle \cite{28, 29}, just to name a few. Tikhonov regularization, see \cite{2, 15}, is probably the most well known regularization method for linear as well as nonlinear inverse problems. However, when it comes to an implementation of Tikhonov regularization for nonlinear or even large scale linear problems, iterative methods are often or even have to be used for finding a minimizer of the Tikhonov functional. A more direct approach is offered by iterative methods, which have been investigated in the frame work of regularization of nonlinear problems more recently, see \cite{7, 16, 4, 17, 21}. Beside simple gradient-type iterations, like Landweber iteration \cite{19}, Newton-type methods seem to be especially attractive, due to their well known, fast convergence properties for well-posed problems. In order to take into account the ill-posed nature of the problem (1.1), several variants of Newton’s method have been proposed for the stable, iterative solution of inverse problems, e.g., the iteratively regularized Gauss-Newton method \cite{7, 17, 18, 21}, the Levenberg-Marquardt method \cite{16}, a Newton-CG algorithm \cite{22}, or a Newton-Landweber method \cite{17, 20}.

1.1. An overview of the regularization methods.
The general form of the minimizing functional in Tikhonov regularization method is given by

\begin{equation}
F(\psi; \alpha, \delta, L, \psi_0) = ||T\psi - g_\delta||^2 + \alpha||L(\psi - \psi_0)||,
\end{equation}

where $\alpha > 0$ is the called the regularization parameter, $\delta \geq ||g - g_\delta||$ is the error norm (with appropriate norms $|| \cdot ||$), $\psi_0$ is an initial guess and $L$ is the regularization operator\(^3\). Under the choice of $L = I$ and $\psi_0 \equiv 0$, known as the standard form of Tikhonov regularization, the (minimal norm) solution is given by

\begin{equation}
\varphi_{\alpha, \delta}^\dagger = \arg \min_{\psi} ||T\psi - g_\delta||^2 + \alpha||\psi||.
\end{equation}

The first term in equation (1.2) is known as the data fitting term, the second term is called the penalty (or regularization) term and the parameter $\alpha$ balances the trade-off between them. The fitting term ensures that the recovered solution fits well

\(^2\)It not, then we use the generalized inverse or Moore-Penrose pseudoinverse, $T^\dagger$, of $T$, which is also unbounded.

\(^3\)again the space for $\psi$ is defined appropriately.
with the given data, when applied by the forward operator, and the regularization term provides some level of smoothness to the inverse recovery. Many strategies have been proposed for the choice of both $L$ and $\alpha$, see [2, 13, 14]. In most cases $L$ is considered as a sum of differentiation operator of some order with different norms in the penalty term, depending on the smoothness of the solution. For example, with the usual $L^2$ or Sobolev norm, which impose strong smoothness, it is hard to recover solutions with sharp features. In such a situation the $L^1$ norm for a suitable space, like the space of functions of bounded variation, provides better recoveries, i.e., the total variation regularization.

For a linear operator $T$, when discretized to a matrix $[T] \in \mathbb{R}^{n \times m}$, it is often convenient to find the minimizer of the functional in (1.2) or (1.3) through the SVD analysis of the matrix $[T]$ and $L$, for small $n$ and $m$. However, for large scale linear system or nonlinear system the above method is computationally expensive or even infeasible and iterative regularization approaches are attractive alternatives to the ordinary Tikhonov regularization, and some of them, for instance, Landweber iteration [4] and the steepest descent method [5], have been suggested to solve nonlinear ill-posed problems. The minimizer of the functional in (1.3) can be approximate, when $T$ is linear, in an iterative way, known as \textit{iterated Tikhonov regularization}, see [2, 8], where at each iteration the minimizer is improved as follows

$$(1.4) \quad \varphi_{k+1}^\dagger = \varphi_k^\dagger - (T^*T + \alpha_k I)^{-1}T^*(T\varphi_k^\dagger - g_0), \quad k = 0, 1, 2, \cdots,$$

where if $\alpha_k = \alpha$, for all $k$, then it’s called \textit{stationary} iterative method, else it’s known as \textit{non-stationary} iteration, and the iterations can be terminated with the aid of the \textit{discrepancy principle}, [31]. Similarly, for a general linear penalty term $L$, in equation (1.2), the minimizer is approximated iteratively as

$$(1.5) \quad \varphi_{L,k+1}^\dagger = \varphi_{L,k}^\dagger - (T^*T + \alpha_k L^*L)^{-1}T^*(T\varphi_{L,k}^\dagger - g_0).$$

For nonlinear $T$, expressions (1.4) and (1.5) are extended to get the well-known Levenberg-Marquardt iterations, see [12, 17]. Later, Bakushinskii [7] proposed an iterative approach for nonlinear operator $T$, namely, the iteratively regularized Gauss-Newton method

$$(1.6) \quad \varphi_{k+1}^\dagger = \varphi_k^\dagger - (\alpha_k I + T'(\varphi_k^\dagger)^*T'(\varphi_k^\dagger))^{-1} \left( T'(\varphi_k^\dagger)^* (T(\varphi_k^\dagger) - g_0) + \alpha_k (\varphi_k^\dagger - \psi_0) \right),$$

to obtain the stable approximate solutions to nonlinear ill-posed problems, see [7, 11, 17, 12]. And for a non-trivial regularization term $L \neq I$, we have

$$(1.7) \quad \varphi_{k+1}^\dagger = \varphi_k^\dagger - (\alpha_k L^*L + T'(\varphi_k^\dagger)^*T'(\varphi_k^\dagger))^{-1} \left( T'(\varphi_k^\dagger)^* (T(\varphi_k^\dagger) - g_0) + \alpha_k L^*L(\varphi_k^\dagger - \psi_0) \right).$$

On the other hand Landweber iteration is a simpler version of the descent method (1.7), where the solution of the inverse problem is approximated iteratively using the gradient direction only

$$(1.8) \quad \varphi_{k+1}^\dagger = \varphi_k^\dagger - T'(\varphi_k^\dagger)^* (T(\varphi_k^\dagger) - g_0) + \alpha_k (\varphi_k^\dagger - \psi_0), \quad k = 0, 1, 2, \cdots,$$

where if $\alpha_k = 0$, for all $k$, then it’s called the \textit{classical Landweber iteration}, otherwise it’s known as the \textit{modified Landweber iteration}, see [4, 6]. An overview on the convergence (rates) results for different iterative methods is presented in [6]. The
simplicity of the Landweber iteration comes with the cost of a slower convergence rate, due to the absence of a damping regularization term.

In all the choices of the regularization operator mentioned above, the solution of the inverse problem is not the minimizer of the regularization term, unless it is a trivial solution, and hence one needs to properly determine the choice of \( \alpha \) to balance between smoothing and fitting in the inverse recovery. In such parameter based regularization methods the ill-posed problem is first converted to a family of well-posed problems (depending on \( \alpha \)) and then, after choosing an appropriate \( \alpha_0 \), the solution is approximated by minimizing the corresponding functional. In this paper we propose an iterative regularization method where the regularization operator \( L \) is connected with the operator \( T \), in equation (1.2), without the involvement of any external parameters, and thus avoiding all the difficulties associated with its involvement. That is, for a given operator \( T \) we construct the regularizing operator \( L \) and the initial guess \( \psi_0 \), depending on \( g_\delta \), such that not only does \( L \) ensure the smoothness of the recovery but also it has the solution, \( \varphi \), as its minimizer.

The iteration is terminated using the discrepancy principle, when error norm \( \delta \) is known, and for unknown \( \delta \) we provide an efficient stopping strategy, which in fact can be coupled with discrepancy principle (when \( \delta \) known) to further improve the efficiency of the recovery.

Let us briefly outline our method. Here we consider the operator \( T : D_T \to \mathcal{H}_2 \) as a linear bounded (injective)\(^4\) operator, where \( D_T \subset \mathcal{H}_1 \) and \( g \in R(T) \)\(^5\) is a known function defined on \([a, b]\) \( \subset \mathbb{R} \), a bounded set. Here \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) are Hilbert spaces, that we take to be subsets of \( \mathcal{L}^2[a, b] \). First we reformulate problem (1.1) in a different way; for a given \( T \) and \( g \), find a \( \varphi \) that satisfies:

\[
T\varphi = -u'',
\]

where \(-u'' = g\) or equivalently

\[
u(x) = \int_a^b \int_{\eta}^{\eta} g(\xi)d\xi d\eta.
\]

The solution of (1.9) is approximated by the minimizer of the functional

\[
G(\psi) = ||T\psi - g||^2_{L^2} + ||u' - u'||^2_{L^2},
\]

where \( u_\psi \), for a given \( \psi \in \mathcal{L}^2[a, b] \), is the solution of the following bvp

\[
-u'' = T\psi,
\]

\[
u(a) = u(a), u'(b) = u(b).
\]

The first term in equation (1.11) is minimized by the solution \( \varphi \) and in [1] we proved that the second term also has the same minimizer, \( \varphi \). Hence for the given operator \( T \) and data \( g \), we are able to create the corresponding regularizing functional \( L \) and the initial guess \( \psi_0 \) such that

\[
\varphi = \arg \min_{\psi} G(\psi)
\]

\[
= \arg \min_{\psi} \{ ||T\psi - g||^2_{L^2} + ||L(\psi - \psi_0)||^2_{L^2} \},
\]

\(^4\)for simplicity, we work in an unique solution scenario. For a non-injective \( T \) we can still recover a regularized solution, the generalized or pseudo-inverse solution.

\(^5\)first we develop the theory with the exact \( g \) and later we prove the stability of the process when given a perturbed \( g_\delta = g + \epsilon g_\delta \), such that \( ||g - g_\delta||_{L^2} \leq \delta \).
where \( L\psi = u'_\psi \) and \( L\psi_0 = u' \).

The significance of \( \varphi \) being the minimizer of both the terms in (1.11) is that one can even minimize only the second term to recover \( \varphi \), which is very helpful in presence of extreme noise level, see [1], or when the first term is very sensitive to the noisy \( g_\delta \), see Example 7.9. For mild noise, with unknown smoothness of the source function, using both the terms leads to a stronger convergence rate, see Table 1 and proved in \( \S \) section 5.1. Here we follow an improved gradient descent algorithm for the inverse recovery of \( \varphi \), where we start with the normal \( L^2 \)-gradient of \( G \) and then upgrade it to various other gradients, including the \( H^1 \)-gradient, to enhance the descent rate and efficiency of the recovery. Using an appropriate gradient is very crucial in the optimization process as it helps in retrieving the features of \( \varphi \) more accurately, for example the \( H^1 \)-gradient not only smooths out the noisy \( L^2 \)-gradient but also helps in pred-conditioning certain desired boundary effects, depending on some prior information of the boundary data (Example 7.5), or going for characteristic gradients helps in preserving the sharp edges and discontinuities of \( \varphi \) (Examples 7.1 and 7.8); it’s discussed in detail in \( \S \) section 4.

During the descent process we generate a sequence of \( L^2 \)-functions \( \{ \psi_m \} \) that converge weakly to \( \varphi \) in \( L^2[a,b] \), and the corresponding sequence \( \{ g_m := T\psi_m \} \) in \( L^2[a,b] \) converges strongly to \( g \) in \( L^2[a,b] \), or converges weakly in \( L^2[a,b] \) if only the second term is present, proved in \( \S \) section 5.1.

Now let us define functionals \( G_1 \) and \( G_2 \) corresponding to the first and the second term of the functional \( G \),

\[
G_1(\psi) = ||T\psi - g||_{L^2}^2 \tag{1.15}
\]

\[
G_2(\psi) = ||u' - u'_\psi||_{L^2}^2 \tag{1.16}
\]

While in [1] a specific example of \( T \) was considered, the theory developed was for a general \( T \), and consequently we can state some of the properties of \( G_2 \) as follows:

- An equivalent form of \( G_2 \), for any \( \psi \in L^2[a,b] \), is

\[
G_2(\psi) = \int_a^b (u'^2 - (u'_\psi)^2) - 2(T\psi)(u - u_\psi) \, dx \tag{1.17}
\]

- For any two \( \psi_1, \psi_2 \in L^2[a,b] \), we have

\[
G_2(\psi_1) - G_2(\psi_2) = \int_a^b -2(T(\psi_1 - \psi_2))(u - \frac{u_{\psi_1} + u_{\psi_2}}{2}) \, dx \tag{1.18}
\]

- The first Gâteaux differential\(^6\), at \( \psi \in L^2[a,b] \), for \( G_2 \) is given by

\[
G'_2(\psi)[h] = \int_a^b (Th)(-2(u - u_\psi)) \, dx \tag{1.19}
\]

where \( h \in L^2[a,b] \). The \( L^2 \)-gradient of \( G_2 \), at \( \psi \), is given by

\[
\nabla^\psi_{L^2} G_2 = T^*(-2(u - u_\psi)), \tag{1.20}
\]

where the \( T^* \) is the adjoint of the operator \( T \).

- The second Gâteaux differential\(^7\), at any \( \psi \in L^2[a,b] \), of \( G_2 \) is given by

\[
G''_2(\psi)[h,k] = 2(-\Delta^{-1}(Th),(Tk))_{L^2} \tag{1.21}
\]

\(^6\)It can be further proved that it’s also the first Fréchet differential of \( G_2 \) at \( \psi \).

\(^7\)again it can be proved that it’s the second Fréchet differential of \( G_2 \) at \( \psi \).
where \( h, k \in \mathcal{L}^2[a, b] \) and \( \Delta = \frac{\partial^2}{\partial x^2} \). Hence for any \( \psi \in \mathcal{L}^2[a, b] \), \( G_2(\psi) \) is a positive definite quadratic form.

Thus \( G_2 \) is a strictly convex\(^8\) functional and has a unique global minimum which is attained at \( \varphi \).

In §section 3 we extend these properties to the functional \( G \). In §section 4 we provide a descent process to minimize the functional \( G \) by using different gradients depending on the scenarios, which is crucial for the minimization process. The convergence of the sequence of functions constructed during the descent process, the stability of the method, and an error analysis is discussed in §section 5. In §section 6 and 8 we provide two stopping criteria for the descent process, when the error norm \( \delta \) is known and when it is unknown, respectively. To corroborate the developed theory we perform numerical experiments on some of the standard inverse problems and compare the results with the results obtained using other regularization methods, like Tikhonov and TSVD; detailed in §section 7.

2. Notations and Preliminaries

We adopt the following notations that are used throughout the paper. All functions are real-valued defined on a bounded closed domain \( [a, b] \subset \mathbb{R} \). For \( 1 \leq p < \infty \), \( \mathcal{L}^p[a, b] := (\mathcal{L}^p, |||\cdot|||_{\mathcal{L}^p}, [a, b]) \) denotes the usual Banach space of \( |f|^p \) integrable functions on \( [a, b] \) and the space \( \mathcal{L}^\infty[a, b] := (\mathcal{L}^\infty, |||\cdot|||_{\mathcal{L}^\infty}, [a, b]) \) contains the essentially bounded measurable functions. Likewise the Sobolev space \( \mathcal{H}^q[a, b] := (\mathcal{H}^q, |||\cdot|||_{\mathcal{H}^q}, [a, b]) \) contains all the functions for which \( f, f', \cdots, f^{(q)} \in \mathcal{L}^2[a, b] \) and the space \( \mathcal{H}_0^q[a, b] := \{ f \in \mathcal{H}^q : "f \text{ vanishes at the boundary}" \} \). The spaces \( \mathcal{L}^2[a, b] \) and \( \mathcal{H}^q \) are Hilbert spaces with inner-products denoted as \( (.,.)_{\mathcal{L}^2} \) and \( (.,.)_{\mathcal{H}^q} \), respectively.

As described in [1] instead of using \( u \in \mathcal{H}^2[a, b] \), defined in (1.10), we update it via the following transformation so that it belongs to \( \mathcal{H}_0^2[a, b] \):

\[
    u(x) = \int_x^b \int_a^\eta g(\xi) \, d\xi d\eta - \frac{b-x}{b-a} \int_a^b \int_a^\eta g(\xi) \, d\xi d\eta.
\]

The advantage of this transformation is that it enables the negative Laplacian operator \( -\Delta = -\frac{\partial^2}{\partial x^2} \) to be a positive operator in \( \mathcal{L}^2[a, b] \) on \( \mathcal{D}_\Delta = \mathcal{H}_0^2[a, b] \). By using the bounds for the \( G_2(\psi) \) from [1] we also have a bound for \( G(\psi) \), for any \( \psi \in \mathcal{L}^2[a, b] \), given by

\[
    (1 + \frac{1}{\lambda_1})^{-1} ||u - u_\psi||^2_{\mathcal{L}^2} \leq G(\psi) \leq ||u - u_\psi||^2_{\mathcal{H}_0^2},
\]

where \( \lambda_1 > 0 \) is the first eigenvalue of the positive operator \( -\Delta \).

Remark 2.1. Note that integrating the given (noisy) data \( g \in \mathcal{L}^2[a, b] \) twice puts the function \( u \) in \( \mathcal{H}^2[a, b] \subset \mathcal{L}^2[a, b] \). This is particularly very significant in the sense that we are able to upgrade the smoothness of the working data or information space from \( \mathcal{L}^2[a, b] \) to \( \mathcal{H}^2[a, b] \) and hence improve the stability and efficiency of the numerical computations.

\(^8\)the strict convexity follows if the operator \( T \) is injective, and if not, then \( G_2 \) is still a convex functional.
3. Convexity of the functional $G$

First we state an ancillary result related to $u_\psi$, from [1].

**Lemma 3.1.** For fixed $\psi$, $h \in L^2[a,b]$, we have

(3.1) \[ \lim_{\epsilon \to 0} u_{\psi + \epsilon h} = u_\psi , \]

in $H^1[a,b]$.

Now we state some of the important properties of the functional $G$.

**Theorem 3.2.**

1. For any two $\psi_1, \psi_2 \in L^2[a,b]$, we have

(3.2) \[ G(\psi_1) - G(\psi_2) = \int_a^b -2T(\psi_1 - \psi_2) \left[ u - \frac{u_{\psi_1} + u_{\psi_2}}{2} + g - \frac{T\psi_1 + T\psi_2}{2} \right] dx \]

2. The first Gâteaux differential, at $\psi \in L^2[a,b]$, for $G$ is given by

(3.3) \[ G'(\psi)[h] = \int_a^b (Th)(-2(u - u_\psi + g - T\psi)) dx \]

where $h \in L^2[a,b]$. The $L^2$-gradient of $G$, at $\psi$, is given by

(3.4) \[ \nabla_{L^2} G = -2T^*(u - u_\psi + g - T\psi) , \]

where $T^*$ is the adjoint of the operator $T$.

3. The second Gâteaux differential, at any $\psi \in L^2[a,b]$, of $G$ is given by

(3.5) \[ G''(\psi)[h,k] = 2(-\Delta^{-1}(Th) + Th,Tk)_{L^2} \]

where $h, k \in L^2[a,b]$. Hence for any $\psi \in L^2[a,b]$, $G''(\psi)$ is a positive definite quadratic form.

**Proof.** All the proofs follow directly from [1] and some additional algebra. \qed

4. The Descent Algorithm for $G$

In this section we discuss the problem of minimizing the functional $G$, via a descent method. One can guess a descent direction by looking at the truncated Taylor expansion of the functional $G$, which is

(4.1) \[ G(\psi - \alpha h) - G(\psi) = -\alpha G'(\psi)[h] , \]

for any $\psi, h \in L^2[a,b]$, and sufficiently small $\alpha > 0$. Thus $G$ is minimized at $\psi$ if the direction $h$ is chosen such that $G'(\psi)[h] > 0$ for an appropriate $\alpha$.

The followings are some descent directions that can make $G'(\psi)[h] > 0$.

4.1. The $L^2$-Gradient.

First, notice from Theorem 3.2 that at a given point $\psi \in L^2[a,b]$

(4.2) \[ G'(\psi)[h] = (h, \nabla_{L^2}^\psi G)_{L^2} , \]

so if we choose the direction $h = \nabla_{L^2}^\psi G$ at $\psi$, then $G'(\psi)[h] > 0$. Though this gradient works well in most situations, there are certain theoretical and numerical issues associated with using the $L^2$-gradient,

\[ \nabla_{L^2}^\psi G = -2T^*(u - u_\psi + g - T\psi) , \]
in the descent process. From a theoretical point of view, if \( T^*(\cdot)(x_0) = 0 \) for some \( x_0 \in [a, b] \) at every step of the descent process then \( \psi_{\text{initial}}(x_0) \) will be invariant during the descent process and if \( \psi_{\text{initial}}(x_0) \neq \varphi(x_0) \) then it will lead to severe decay or fluctuation near the point \( x_0 \) as \( \psi_m \rightarrow \varphi \) weakly in \( L^2[a, b] \), see \S \text{section 5}. For example, if \( T = \int_a^x \frac{d}{dt} \cdot dt \) is a Volterra operator then the \( L^2 \)-gradient at any \( \psi \) is always zero at the end point \( b \), since \( T^*(\cdot) = \int_a^b \cdot dt \), which implies that at the end point \( b \) the recovery will be invariant during the descent process. Hence for the initial choice of the descent function \( \psi_0 \) if \( \psi_0(b) \neq \varphi(b) \) then it will lead to large fluctuations at that end point (see example 7.5). From a numerical viewpoint, if the operator \( T^* \) is very sensitive to noise then it will make the inverse recovery unstable, as \( \nabla_{L^2} G \) involves the noisy \( g_0 \) directly. For example, if the operator \( T \) is a differential operator then the adjoint operator \( T^* \) is also a differential operator acting on the noisy \( g_0 \) and thus will greatly amplify the noise.

4.2. The \( H^1 \)-Gradient.

One can circumvent this problem by opting for the Sobolev or Neuberger gradient, \( \nabla_{L^2} \psi, G \), instead (see [1]). It is the solution of the following boundary value problem

\[
-\psi'' + g = \nabla_{L^2} \psi G
\]

(4.3)

This provides us a gradient, \( \nabla_{H^1} G = g \), at any \( \psi \), with considerably more flexibility at the boundary points \( \{a, b\} \). In particular one has

1. Dirichlet Neuberger gradient : \( g(a) = 0 \) and \( g(b) = 0 \).
2. Neumann Neuberger gradient : \( g'(a) = 0 \) and \( g'(b) = 0 \).
3. Robin or Mixed Neuberger gradient : \( g(a) = 0 \) and \( g'(b) = 0 \) or \( g'(a) = 0 \) and \( g(b) = 0 \).

In addition to the flexibility at the end points, it also enables the new gradient to be a preconditioned (smoothed) version of \( \nabla_{L^2} G \), as \( g = (I - \Delta)^{-1} \nabla_{L^2} G \), and hence gives a superior convergence in the steepest descent algorithms when recovering a smooth function. One can exploit the flexibility of the gradient at the end points according to some prior information (if known) of \( \varphi \) at the end points. For example, if prior knowledge of \( \varphi(a) \) and \( \varphi(b) \) are known, then one can define \( \varphi_{\text{initial}} \) as the straight line joining them and use the Dirichlet Neuberger gradient for the descent. Thus the boundary data is preserved in each of the evolving \( \psi_m \)'s during the descent process which leads to a more efficient and faster descent compared to the normal \( L^2 \)-gradient descent. Even when \( \varphi|_{\{a, b\}} \) is unknown, one can use the Neumann Neuberger gradient that allows free movements at the boundary points, Example 7.5. However, in certain situations (especially when \( T^* \) is a smooth operator) the \( \nabla_{L^2} \psi G \) has its own advantages. In addition to saving computational time, it also preserves sharp features or discontinuities in the recovery better than \( \nabla_{H^1} \psi, G \), as it is a cruder gradient.

4.3. The \( L^2 - H^1 \) Conjugate Gradient.

Now, one can make use of both the gradients by forming an average of them, or using them to compute the standard Polak-Ribière conjugate gradient scheme (see [1]) to further boost the descent rate, approximately by a factor of 2. Specifically, the initial search direction at \( \psi_0 \) is \( h_0 = g_0 = \nabla_{H^1} \psi, G \). At \( \psi_m \) one can use an exact
or inexact line search routine to minimize $G(\psi)$ in the direction of $h_m$ resulting in $\psi_{m+1}$. Then $g_{m+1} = \nabla_{H_1}^{\psi} G$ and $h_{m+1} = g_{m+1} + \gamma_m h_m$ where

\begin{equation}
\gamma_m = \frac{(g_{m+1} - g_m, g_{m+1})_{H_1}}{(g_m, g_m)_{H_1}} = \frac{(g_{m+1} - g_m, \nabla_{L^2}^{\psi} G)_{L^2}}{(g_m, \nabla_{L^2}^{\psi} G)_{L^2}}.
\end{equation}

Note that one can also use either $\nabla_{L^2}^{\psi} G$ or $\nabla_{H_1}^{\psi} G$ to form the conjugate gradient.

4.4. Some other gradients.

As mentioned above the Neuberger gradients, being smoothed versions of the $L^2$-gradient, has an advantage when recovering a smooth parameter, but is sometimes not quite effective when the parameters have discontinuities or sharp features in them. In these situations one can opt for the $L^2$-gradient, while noting that some other descent directions can provide even sharper recoveries than, depending on the scenarios. From (4.1) and (4.2), we need to find a direction $h_0$ such that $G'(\psi)[h_0] > 0$.

4.4.1. The $L^\infty$-gradient.

Here we seek a function $h_0 \in L^\infty(\Omega)$ with norm one that maximizes $G'(\psi)[h_0]$, see [3]. For a given $\nabla_{L^2}^{\psi} G$, a quick inspection provides us the following $L^\infty$-gradient

\begin{equation}
\nabla_{L^\infty}^{\psi} G(x) = \begin{cases}
1, & \nabla_{L^2}^{\psi} G(x) > 0, \\
0, & \nabla_{L^2}^{\psi} G(x) = 0, \\
-1, & \nabla_{L^2}^{\psi} G(x) < 0.
\end{cases}
\end{equation}

4.4.2. The $\chi$-gradient.

We define a characteristic gradient as

\begin{equation}
\nabla_{\chi}^{\psi} G = \sum_{k=1}^{n} \alpha_k \chi_{A_k},
\end{equation}

where $\alpha_k \in \mathbb{R}$ and $A_k \subset [a,b]$ are chosen such that $G'(\psi)[\nabla_{\chi}^{\psi} G] > 0$, and $\chi$ is the usual characteristic function

\begin{equation}
\chi_{A_k}(x) = \begin{cases}
1, & x \in A_k, \\
0, & x \notin A_k.
\end{cases}
\end{equation}

Thus we can see that the previously defined $\nabla_{L^\infty}^{\psi} G$ is a special case of the characteristic gradient. Though the $L^\infty$-gradient provides a discrete gradient for the descent process, but upon choosing $\alpha_k$ and $A_k$ appropriately we can obtain an even faster descent rate and better recoveries. One way to choose $\alpha_k$ and $A_k$ is by splitting $[a,b]$ into simple disjoint intervals, such that $A_k = (x_k, x_{k+1})$ and $\nabla_{L^2}^{\psi} G$ is either positive, negative or zero on $A_k$, and $\alpha_k = Avg_{A_k}(\nabla_{L^2}^{\psi} G)$, the average of $\nabla_{L^2}^{\psi} G$ on the interval $A_k$ given by

\begin{equation}
Avg_{A_k}(\nabla_{L^2}^{\psi} G) = \frac{\int_{A_k} \nabla_{L^2}^{\psi} G \, dx}{\int_{A_k} d x}.
\end{equation}

This is helpful because during the descent process we are descending discretely depending on the $\nabla_{L^2}^{\psi} G$ values, unlike in the $L^\infty$-gradient where we descend with constant values $\pm 1$ depending on the sign of $\nabla_{L^2}^{\psi} G$. 
We can further generalize the characteristic gradient by not restricting it to only the simple functions. It can be extended rather to a piece-wise defined function,

\[
\nabla^\psi \chi f G = \sum_{k=1}^{n} f_k \chi A_k ,
\]

for suitable choices of functions \( f_k \) and \( A_k \subset [a,b] \). A choice for \( A_k \) can be simple intervals where \( \nabla^\psi \chi G \) has a particular sign, as chosen for (4.6), but here we have \( f_k := \nabla^\psi H_1 G \cdot \chi A_k \) with either the Neumann conditions (if no knowledge on the boundary points of that particular sub-interval \( A_k \) is known) and Dirichlet or Mixed boundary conditions (depending on the given boundary information of \( A_k \)). This gradient has an advantage over \( \nabla^\chi G \) when the source function has both the smooth and sharp features, since using \( \nabla^\chi G \) results in the staircase formation (caused by the step-function) while recovering the smooth features. So when no information on the smoothness of the recovery function is given one should opt for the extended \( \chi \)-gradient rather than the normal \( \chi \)-gradient. Another feature of such a gradient arises when dealing with a non-injective operator \( T \); see Remark 4.1.

4.4.4. Alternating gradient.

Sometimes, it is better to have a descent direction that fluctuates between the smooth and the sharp descent directions, as it preserves both the smooth and the sharp features of the recovery. So we define a descent direction, denoted by \( \nabla^L_2 L_\infty G \), that alternates between the \( \nabla^L_2 G \) and \( \nabla^L_\infty G \), or define \( \nabla^L_2 \chi G \), that alternates between the \( \nabla^L_2 G \) and \( \nabla^\chi G \) during the descent process.

Remark 4.1. For all of the above defined descent directions we have performed computational testing to validate their numerical viability. We further comment on some promising alternatives.

(1) To use the characteristic direction \( \nabla^\psi m G \) at any particular stage of the descent process with optimal values of \( \alpha_k \), one can perform a multi-variable minimization:

\[
\text{minimize } F(\bar{\alpha}) = G(\psi_m) - (\nabla^\psi m G, \sum_{k=1}^{n} \alpha_k \chi A_k)_{L^2} ,
\]

for \( \bar{\alpha} = \{\alpha_k\}_{k=1}^{n} \in \mathbb{R}^n \) and appropriately defined \( A_k \subset [a,b] \), instead of performing a single variable minimization using the \( \chi \)-gradient as defined in (4.6),

\[
\text{minimize } F(\alpha) = G(\psi_m) - \alpha \|\nabla^\psi m G\|_{L^2}^2 , \text{ for } \alpha \in \mathbb{R}.
\]

(2) If some prior information is known about the parameter function to be recovered, then one can use the extended \( \chi \)-gradient (4.8) to improve the optimization process. For example, if \( \{\varphi(x_i)\}_{i=1}^{n} \) is known, then we can start with the initial guess \( \psi_0 \) as a linear interpolation of the given data points \( (x_i, \varphi(x_i))_{i=1}^{n} \), with \( A_k = [x_k, x_{k+1}] \) and \( f_k = \nabla^\psi H t G \) with Dirichlet boundary conditions (and mixed conditions for the initial and final sub intervals if \( x_1 \neq a \) or \( x_n \neq b \)). This technique is very useful in recovering \( \varphi \) almost uniquely even when the the operator \( T \) is not injective, as it uses the known information as a constraint and reduces the search space for \( \varphi \), when the null space is non-trivial.
4.5. **The line search method for the functional \( G \).**

At any given \( \psi_m \in L^2[a, b] \) the functional \( G \) is minimized in the gradient direction via the single variable function \( f_m(\alpha) = G(\psi_m + 1(\alpha)) \) and using a line search minimization, where \( \psi_m + 1(\alpha) = \psi_m - \alpha \nabla_{\psi_m} G \). To bracket the minimum the initial step size \( \alpha_0 \) is chosen by solving the quadratic approximation of the function \( f_m \), (see [1]), which is given by

\[
\alpha_0 = \frac{G'(\psi_m)[g_m]}{G''(\psi_m)[g_m, g_m]}. \tag{4.11}
\]

Hence in the descent process, starting from an initial guess \( \psi_0 \in L^2[a, b] \), we obtain a sequence of \( L^2 \)-functions \( \psi_m \) for which the sequence \( \{G(\psi_m)\} \) is strictly decreasing. In the next section, we discuss the convergence of \( \psi_m \) to \( \varphi \) and the stability of the recovery.

5. **Convergence, Stability and Error Analysis**

5.1. **Convergence.** First we see that for the sequence of \( L^2 \)-functions \( \{\psi_m\} \), obtained during the descent process, the corresponding sequence of real numbers \( \{G(\psi_m)\} \) converges to zero. The following theorems describe the convergence of the sequence \( \{\psi_m\} \) to \( \varphi \) in \( L^2[a, b] \).

If the functional \( G \) in (1.11) has only the first term then from [1] we have the convergence result for the functional \( G_2 \), which is

**Theorem 5.1.** Suppose that \( \{\psi_m\} \) is any sequence of \( L^2 \)-functions such that the sequence \( \{G_2(\psi_m)\} \) tends to zero. Then \( \{\psi_m\} \) converges weakly to \( \varphi \) in \( L^2[a, b] \) and the corresponding sequence \( \{g_m := T\psi_m\} \) also converges weakly to \( g \) in \( L^2[a, b] \), and the sequence \( \{u_m \psi_m\} \) converges strongly to \( u \) in \( H^1[a, b] \).

But if the functional \( G \) has both the terms, the functional \( G_1 \) as well as \( G_2 \), then we have a stronger convergence.

**Theorem 5.2.** Suppose that \( \{\psi_m\} \) is any sequence of \( L^2 \)-functions such that the sequence \( \{G(\psi_m)\} \) tends to zero. Then \( \{\psi_m\} \) converges weakly to \( \varphi \) in \( L^2[a, b] \) and \( \{u_m \psi_m\} \) converges strongly to \( u \) in \( H^1[a, b] \). Moreover, the sequence \( \{g_m := T\psi_m\} \) converges strongly to \( g \) in \( L^2[a, b] \).

**Proof.** The strong convergence of \( u_m \psi_m \rightarrow u \) in \( H^1[a, b] \) and \( g_m \rightarrow g \) in \( L^2[a, b] \) follows from the bounds on \( G(\psi) \) in equation (2.2), which gives

\[
||u - u_m||^2_{L^2} \leq \left( 1 + \frac{1}{\lambda_1} \right) G(\psi_m). \tag{5.1}
\]

The proof of the weak convergence of \( \{\psi_m\} \) to \( \varphi \) in \( L^2[a, b] \) is very similar to the proof in theorem 5.1, provided we can show that the range of \( T^* \) is dense in \( L^2[a, b] \) in the \( L^2 \)-norm. This is true because our operator \( T \) is an injective bounded linear operator. \( \square \)

5.2. **Stability.** Here we prove that regarding the problem of solving equation (1.1) as that of finding the minimizer of the functional \( G \), is a conditionally well posed problem. We see from subsection 5.1 that for a given \( g \) we are able to construct a sequence of functions \( \{\psi_m\} \) in \( L^2[a, b] \) such that \( G(\psi_m) \rightarrow 0 \) and \( \psi_m \rightarrow \varphi \) weakly in \( L^2[a, b] \), where \( T\varphi = g \). Now if we are given \( g_\delta \) instead, where \( ||g - g_\delta||_{L^2} \leq \delta \), then we show that the sequence \( \psi_m \) will still approach \( \varphi \), modulo some conditions.
First we examine the value of the functional $G_\delta$, formed based on $g_\delta$, at the point $\varphi \in L^2[a, b]$.

**Theorem 5.3.** Suppose given $g$ and $g_\delta$ are such that $\|g - g_\delta\|_{L^2} \leq \delta$ and lets denote their corresponding functional as $G$ and $G_\delta$, respectively,

$$
G(\psi) = \|T\psi - g\|_{L^2}^2 + \|u_\psi' - u_\varphi'|_{L^2}^2,
$$

(5.2)

$$
G_\delta(\psi) = \|T\psi - g_\delta\|_{L^2}^2 + \|u_\psi' - u_\varphi'|_{L^2}^2,
$$

(5.3)

where $-u_\varphi'' = g$, with $u_\varphi(a) = 0 = u_\varphi(b)$ and $-u_\varphi'' = g_\delta$, with $u_\varphi(a) = 0 = u_\varphi(b)$. Then the minimizer $\varphi$ of functional $G$, such that $G(\varphi) = 0$, satisfies the following upper bound

$$
G_\delta(\varphi) \leq C\delta,
$$

for some constant $C$.

**Proof.** Since $T\varphi = g$ and $T\varphi_\delta = g_\delta$ we have

$$
G_\delta(\varphi) = \|T\varphi - g_\delta\|_{L^2}^2 + \|u_\varphi' - u_\varphi'|_{L^2}^2
$$

$$
= \|g - g_\delta\|_{L^2}^2 + \|u_\varphi' - u_\varphi'|_{L^2}^2
$$

$$
\leq C\delta.
$$

$\square$

In the next theorem we prove that for a sequence of functions $\{\psi_m\}$ converging weakly to $\varphi_\delta$ in $L^2[a, b]$, that is, $G_\delta(\psi_m)$ is small for some large $m$, then $G(\psi_m)$ is also small for large $m$.

**Theorem 5.4.** If for a sequence of functions $\{\psi_m\} \subset L^2[a, b]$, the sequence $G_\delta(\psi_m) \to 0$, then there exists a positive number $M(\delta)$ such that for all $m \geq M(\delta)$, $G(\psi_m) \leq C\delta$ for some constant $C$.

**Proof.** For the functional $G_1$ we have

$$
G_1(\psi_m) = \|g - g_\delta\|_{L^2}^2 + \|T\psi_m - g_\delta\|_{L^2}^2 + 2(g_\delta - g, T\psi_m - g)_{L^2}
$$

and since $T\psi_m \to g_\delta$ in $L^2[a, b]$, we get $G_1(\psi_m) \leq c_1\delta$ for some constant $c_1$ and large $m$. As for $G_2$, it is proved in [1] that there exists an $M_2(\delta) \in \mathbb{N}$ such that $G_2(\psi_m) \leq c_2\delta$ for all $m \geq M_2(\delta)$, and the result follows. $\square$

Many operators, such as integral operators, are defined through a kernel function and sometimes these kernel functions are also not exactly known. That is, in certain situations we may not even have the exact operator $T$ but an approximation of it, $T_\delta$, such that the operator norm of the difference is small, $\|T - T_\delta\|_2 \leq \delta_1$. Even in these scenarios, the above inequalities continues to apply.

**Theorem 5.5.** If for a sequence of functions $\{\psi_m\}$ in $L^2[a, b]$, the corresponding sequence $G_\delta(\psi_m) \to 0$, where $G_\delta$ is the functional defined based on $g_\delta$ and $T_\delta$, then there exists a $M(\delta, \delta_1) \in \mathbb{N}$ such that $G(\psi_m) \leq C\delta + C_1\delta_1$ for all $m > M(\delta, \delta_1)$, some constants $C$ and $C_1$. 

5.3. **Conditional well-posedness.** For an exact \( g \) (or equivalently, an exact \( u \)) we have \( G(\varphi) = 0 \), but for a given noisy \( g_\delta \) (or \( u_\delta \)) and the functional based on it, we have \( G_\delta(\varphi) \neq 0 \) or \( G_\delta(\varphi) > 0 \), due to the strict convexity of \( G_\delta \) and \( G_\delta(\varphi_\delta) = 0 \). So if we construct the sequence of functions \( \psi_\delta m \in L^2[a,b] \), using the descent algorithm, such that \( G_\delta(\psi_\delta m) \to 0 \) then (from theorems 5.1 or 5.2) we will have \( \psi_\delta^m \overset{w}{\to} \varphi_\delta \) which implies \( \|\psi_\delta^m - \varphi\|_L^2 \to 0 \) due to the semi-convergence nature as \( G_\delta(\varphi) > 0 \) and \( G_\delta(\psi_\delta^m) \to 0 \). This is a typical behavior of any ill-posed problem and is managed by stopping the descent process at certain iteration such that \( G_\delta(\psi_\delta^m) > 0 \) but close to zero (which is attained from the stability theorems 5.3, 5.4 and 5.5). Following similar arguments as in (2.2) we have a lower bound for \( G_\delta(\varphi) \).

**Theorem 5.6.** Given \( u,u_\delta \in H^2_0[a,b] \) with their respective functionals \( G,G_\delta \) (as defined in 5.2) and their inverse recoveries \( \varphi,\varphi_\delta \) (such that \( G(\varphi) = 0 \) and \( G_\delta(\varphi_\delta) = 0 \)), we have the following lower bound for \( G_\delta(\varphi) \)

\[
G_\delta(\varphi) \geq \left( 1 + \frac{1}{\lambda_1} \right)^{-1} \|u-u_\delta\|^2_{H^2}.
\]

Therefore, combining theorems 5.3, 5.4 and 5.6 we have the following two-sided bound for \( G_\delta(\varphi) \)

\[
C_1\|u-u_\delta\|_{H^1} \leq G_\delta(\varphi) \leq C_2\|u-u_\delta\|_{H^1}.
\]

Thus, when \( \delta \to 0 \) we have \( G_\delta(\varphi) \to 0 \) which implies \( \varphi_\delta \to \varphi \) in \( L^2[a,b] \).

6. **Stopping criterion I**

When the error norm \( \delta = \|g-g_\delta\|_{L^2} \) is known then *Morozov’s discrepancy principle*, [31], can serve as a stopping criterion for the iteration process, that is, terminate the iteration when

\[
\|T\psi_m - g_\delta\|_{L^2} \leq \tau\delta
\]

for an appropriate \( \tau > 1^9 \). For the convergence and stability of the process see [4]. We can further improve the termination rule by coupling the above condition with Stopping criterion II (section 8), which is applicable even in the absence of the error norm \( \delta \).

7. **Numerical Results**

We follow a similar algorithm as given in [1]. We perform inverse recoveries on the classical integral type equations. A MATLAB code was written to test the numerical viability of the method and the results obtained is then compared with the standard regularization methods, like truncated SVD, Tikhonov regularization with \( L = \mu I \) and Tikhonov regularization with a \( L \) as defined in [30]. The grid spacing depends on the problem and is specified in the respective problem settings.

---

9In our experiments, we considered \( \tau = 1 \) and the termination condition as \( \|T\psi_m - g_\delta\|_{L^2} < \delta \).
First we consider functional $G$ to obtain the recoveries, where the iterations are terminated using stopping criterion I (section 6), and then we repeat the process for functional $G_2$ and compare the results.

7.1. **Numerical Differentiation.** In paper [1], we used the functional $G_2$ to solve the inverse problem of numerical differentiation. The gradient used in that paper was either the $\nabla_{L^2} G$ or the $\nabla_{H^1} G$ and we observed that the recoveries obtained using this method outperformed its counterparts (like Tikhonov regularization, total variation, smoothing spline, mollification method and least square polynomial approximation) under the same assumption, that is, $g \in H^1[a, b]$; see [1].

**Example 7.1.** Here we compare the recoveries obtained using $\psi \chi G$ gradient vs (i) $L^2 - H^1$ conjugate gradient from [1], (ii) total variation regularization [33] and (iii) adaptive step-size regularization [32], when the derivative has a discontinuous jump. For this purpose we perturb the following test function, given as Example 7.5 in [1],

$$y_2(t) = \begin{cases} 1 - t, & t \in [0, 0.5], \\ t, & t \in (0.5, 1], \end{cases}$$

and use it numerically compute the first derivative of $y_2, \delta$. Figure 1 shows the recoveries using different gradients. The relative error of the recovery using $L^2 - H^1$ conjugate gradient is 0.1684 and the number of iterations needed to achieve it is more than 1000; whereas the relative error of the recovery using $\psi \chi G$ gradient is 0.0038 and the number of iterations needed is 5.

**Remark 7.2.** One gets result very similar to Figure 1a when using $L^\infty$-gradient, to be precise the relative error of the recovery is 0.0079.

7.2. **Fredholm Integral Equation of the First Kind.** A Fredholm integral equation of the first kind is an integral operator equation, depending on a kernel function $K$, and is very similar to a Volterra type equation but with constant limits of integration:

$$T\varphi := \int_a^b K(s, t)\varphi(t)dt = g(s), \text{ for } s \in [c, d].$$

Such integral operator equations are classical inverse problems and can be quite ill-posed. An important case of these type of equation is when the kernel function is a function depending only on the difference of its arguments, namely $K(s, t) = K(s - t)$, and the limits of integration are $\pm \infty$, then the integral can be considered as the convolution of the kernel function and the source function, $T\varphi = K \ast \varphi$, and the corresponding inverse problem is called as the deconvolution problems, see Examples 7.8 and 7.9. Examples 7.4, 7.3, 7.5 and 7.6 are obtained by discretizing the corresponding Fredholm integral (7.1), where the discretizations are carried out by either Galerkin or Nyström methods. This yield a linear discrete ill-posed problem

$$Ax = b,$$

where the matrix $A \in \mathbb{R}^{m \times n}$ is the discretized representation of the operator $T$ and the vectors $x \in \mathbb{R}^n$, the $b \in \mathbb{R}^m$ are the discretized source and effect functions. MATLAB functions in [34] determine discretizations $A \in \mathbb{R}^{n \times n}$, and scaled discrete
A NEW PARAMETER-FREE REGULARIZATION METHOD FOR INVERSE PROBLEMS. 15

Figure 1. Numerical differentiation using different regularization.

approximations \( x \in \mathbb{R}^n \) and \( b := Ax \in \mathbb{R}^m \), where \( n = m = 200 \), unless otherwise stated. To test the stability of the method, we add a normally distributed (with mean zero) error vector \( \epsilon \delta \in \mathbb{R}^m \) to the original \( b \) to get a perturbed vector \( b \delta \in \mathbb{R}^m \), such that (the relative error norm) \[ \frac{||\epsilon \delta||}{||b||} \leq 10\%. \] In particular, when using stopping criterion I to terminate the descent process we assume \( \delta = ||\epsilon \delta|| \) to be known, but when applying stopping criterion II we consider it as unknown. We compare our results, in Table 1, with the results obtained using Standard Tikhonov, TVSD and Tikhonov with a regularization matrix \( L \) as specified in [30].

**Example 7.3. Shaw Test [Image Restoration].**

One dimensional image restoration can be modeled by a Fredholm equation with the following kernel function

\[
K(s, t) = (\cos(s) + \cos(t))^2 \left( \frac{\sin(u)}{u} \right)^2,
\]

where

\[
u(s, t) = \pi(\sin(s) + \sin(t)).
\]

The effect function \( g \) can be obtained from the following source function

\[
\varphi(t) = a_1e^{-c_1(t-t_1)^2} + a_2e^{-c_2(t-t_2)^2}.
\]

From [34], we take the parameter values for \( a_1 = 2, a_2 = 1, c_1 = 6, c_2 = 2, t_1 = 0.8 \) and \( t_2 = -0.5 \), which give the source term \( \varphi \) two distinct peaks, with the integration limits as \( [-\frac{\pi}{2}, \frac{\pi}{2}] \) for both \([a, b]\) and \([c, d]\). Figures 2a and 2b show the recoveries of
the solution using the functional $G$ and $G_2$, respectively, and considering the $L^2$-conjugate gradient direction during the descent. The relative error in the recoveries are 16.20% and 16.24% using the functional $G$ and $G_2$, respectively.

Figure 2. Shaw Test problem

**Example 7.4. Phillips problem.**

In this example we consider the Phillips problem from [34], where the source, the kernel and the effect functions are given by

\[(7.6) \quad \varphi(t) = \begin{cases} 1 + \cos\left(\frac{\pi t}{3}\right), & \text{for } |t| \leq 3, \\ 0, & \text{otherwise}, \end{cases} \]

\[(7.7) \quad K(s, t) = \varphi(|s - t|), \]

\[(7.8) \quad g(s) = (6 - |s|) \left( 1 + \frac{1}{2} \cos\left(\frac{\pi s}{3}\right) \right) + \frac{9}{2\pi} \sin\left(\frac{\pi |s|}{3}\right), \]

with the limits of integration $[a, b] = [c, d] = [-6, 6]$. Figures 3a and 3b show the recoveries of the solution using the functional $G$ and $G_2$, respectively, and considering the $L^2$-conjugate gradient direction during the descent. The relative error in the recoveries are 8.80% and 9.38% using the functional $G$ and $G_2$, respectively.

Figure 3. Phillips Test problem
Example 7.5. Deriv2 problem:
In this example we numerically compute the second derivative of a given function. Just as the solution of a Volterra equation gives us the first derivative of a function, see [1], the solution of a Fredholm equation with the corresponding Green’s function as the kernel gives us the second derivative of the function. Here, the following kernel function
\begin{equation}
K(s, t) = \begin{cases} s(t - b), & \text{for } s < t \\ t(s - d), & \text{for } s \geq t, \end{cases}
\end{equation}
helps us to compute the second derivative of the function \( g(s) = \frac{s^3 - s}{3} \) on \([a, b] = [c, d] = [0, 1] \). We use the perturbed function \( g_\delta \) instead and compare the computed \( g'' \) with the original \( g'' \). Figures 4a and 4b show the recoveries of the solution using the functional \( G \) and \( G_2 \), respectively, and considering the \( L^2 \) gradient direction during the descent. The relative error in the recoveries are 34.62\% and 34.40\% using the functional \( G \) and \( G_2 \), respectively. From the Figure 4 one can suspect that at the boundary point ‘\( b \)’ the recovery using \( \nabla_{L^2} G \)-conjugate gradient is invariant during the descent process, which in fact can be verified by examining the kernel function (7.9). Hence in this situation one can implement the Neumann Neuberger gradient, to have floating boundary values during the descent process. Figures 5a and 5b show the recoveries of the solution using the functional \( G \) and \( G_2 \), respectively, and considering the Neumann \( \nabla_{H^1} G \) gradient direction during the descent. The relative errors in the recoveries are 8.81\% and 10.66\% using the functional \( G \) and \( G_2 \), respectively, and considering the \( \nabla_{H^1} G - \nabla_{L^2} G \) conjugate-gradient direction.

(a) Recovery using functional \( G \). (b) Recovery using functional \( G_2 \).

Figure 4. Deriv2 Test problem, using \( L^2 \)-conjugate gradient.

Example 7.6. Baart problem.
In this example we consider the Baart problem from [34], that is, with the following source, kernel and effect functions:
\begin{align}
\varphi(t) &= \sin(t), \\
K(s, t) &= e^{s \cos(t)}, \\
g(s) &= 2\frac{\sin(s)}{s},
\end{align}
where the integral intervals are \([0, \frac{\pi}{2}] \) for \( s \) and \([0, \pi] \) for \( t \). Figures 6a and 6b show the recoveries of the solution using the functional \( G \) and \( G_2 \), respectively, and
considering the $\mathcal{L}^2$-conjugate gradient direction during the descent. The relative error in the recoveries are 21.60% and 33.19% using the functional $G$ and $G_2$, respectively.

(a) Recovery using functional $G$.  
(b) Recovery using functional $G_2$.

**Figure 5.** Deriv2 Test problem, using $\mathcal{L}^2 - \mathcal{H}^1$-conjugate gradient.

(a) Recovery using functional $G$.  
(b) Recovery using functional $G_2$.

**Figure 6.** Baart Test problem

| Methods                | Shaw | Phillips | Deriv2 | Baart |
|------------------------|------|----------|--------|-------|
| Tikhonov ($L = \mu I$) | 0.17 | 0.05     | 0.35   | 0.26  |
| Tikhonov ($L$, in [30])| 0.16 | 0.02     | 0.32   | 0.21  |
| TVSD                   | 0.16 | 0.04     | 0.33   | 0.17  |
| $G$ (Stopping criteria I) | 0.16 | 0.09     | 0.34 (−∇$\mathcal{L}^2 G$) | 0.21 |
|                        |      |          | 0.09 (−∇$\mathcal{H}^1 G$) |       |
| $G$ (Stopping criteria II) | 0.10 | 0.07    | 0.34 (−∇$\mathcal{L}^2 G$) | 0.22 |
|                        |      |          | 0.09 (−∇$\mathcal{H}^1 G$) |       |
| $G_2$ (Stopping criteria I) | 0.16 | 0.09    | 0.34 (−∇$\mathcal{L}^2 G$) | 0.33 |
|                        |      |          | 0.10 (−∇$\mathcal{H}^1 G$) |       |
| $G_2$ (Stopping criteria II) | 0.16 | 0.04   | 0.34 (−∇$\mathcal{L}^2 G$) | 0.33 |
|                        |      |          | 0.10 (−∇$\mathcal{H}^1 G$) |       |

**Table 1**
Example 7.7. Large scale linear system.
Here we reconsider the Phillips problem, Example 7.4, but with $A \in \mathbb{R}^{1200 \times 1200}$. The relative error in the recoveries are $4.21\%$ and $2.34\%$ using stopping criteria I and II, respectively, and minimizing the functional $G$. The computational time for the descent process, when $n = 1200$, only exceeds $3.605$ times the computational time taken for the descent process when $n = 200$, in Example 7.4.

7.3. Deconvolution [Image Deblurring]:
Deconvolution has many practical applications, primarily used in image deblurring. It arises in connection with the degradation of digital images by atmospheric turbulence blur, modeled by a Gaussian point-spread function. The one dimensional kernel function is given by

$$K(s, t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(s-t)^2}{2\sigma^2}},$$

for $-\infty < t < \infty$ and $s \in [c, d]$. One of the most concerning and difficult aspect of such problems is to recover the sharp features of $\varphi$, if it has any, since the Gaussian kernel tends to smear them and usually it is quite hard to inversely recover them using smooth gradients. This is where all the different coarser gradients that we have discussed in Section 4 comes in handy.

Example 7.8. Here we consider the following discontinuous source function

$$\varphi = 1 \cdot \chi(-0.5,0.5) + 0 \cdot \chi(-0.5,0.5)^c,$$

on $[-1, 1]$ and the corresponding truncated Gaussian kernel

$$K_1(s, t) = K(s, t) \cdot \chi([-1, 1]),$$

where the std. deviation ($\sigma$) of the Gaussian kernel function $K$ is taken as 1. The integral domain for $t$ is $[a, b] = [-1, 1]$ and for $s$ is $[c, d] = [-2, 2]$, and the grid spacing in both the intervals is 0.01. Here we used the MATLAB functions $\text{conv}(..., \text{"full"})$ for the forward operation $T\psi_m$ and $\text{conv}(..., \text{"same"})$\(^{11}\) for the adjoint operation $T^* (u - u\psi_m + g - T\psi_m)$, to compute $\nabla_{L^2} G$. Figures 7a and 7b show the recoveries of the solution, using the $\nabla_{L^2} G$ and $\nabla_{L^2} G$, respectively, during the descent process and minimizing the functional $G$. The relative error in the recoveries are $27.74\%$ and $18.45\%$ using the functional $G$ and $G_2$, respectively.

![Figure 7](image_url)

**Figure 7.** Example 7.8 (Deconvolution problem).

\(^{11}\)for matrix dimension compatibility during the descent process, as $\psi_{m+1} = \psi_m - \alpha \nabla_{L^2} G$. 

Example 7.9. In the previous example we observed that using an appropriate gradient significantly affects the recovery of the solution. Here we show that using an appropriate functional, $G$ or $G_2$, also has a great impact on the recovery. We considered three cases, when (i) $\sigma = 1$ (Example 7.8), (ii) $\sigma = 0.5$ and (iii) $\sigma = 0.1$. We will see that for larger $\sigma$ using the functional $G$ provides better recovery than the functional $G_2$, and its vice-versa for smaller $\sigma$, that is, the functional $G_2$ is gives better recovery than the functional $G$ for smaller $\sigma$. Figures 8a and 8b show the recoveries of the solutions using the functional $G$ and $G_2$, respectively, considering the $\nabla \chi$ direction during the descent. The relative errors in the recoveries are 18.45%, 18.64% and 25.36% for $\sigma = 1$, 0.5, 0.1, respectively, using the functional $G$ and 30.59%, 23.46% and 16.29% using the functional $G_2$, for the respective $\sigma$.

![Recoveries using $G$ and $G_2$.](image)

(a) Recoveries using $G$. (b) Recovery using $G_2$.  

Figure 8. Example 7.9 (Deconvolution problem).

Remark 7.10. Example 7.9 reflects the importance of the functional $G_2$. Since the minimizer of the functional $G_2$ is same as the functional $G$, it provides us an alternate approach to recover the solution, when the functional $G$ is very sensitive to the noisy $g_\delta$.

8. Stopping Criterion II

In this section we provide a stopping criteria when the error norm $\delta$ is unknown, or even when $\delta$ is known one can combine this with stopping criteria I (§section 6) to have an enhanced recovery. Here we see the stabilizing effect of functional $G_2$ and make use of it to recover $\varphi$, even in the absence of error information. It is a two step process, where in the first step we terminate the iteration process using Algorithm 1. The purpose of counter $i_1$ is to provide some flexibility for wiggling during the descent of $G$, which is controlled by the value of max1, and counter $i_2$ handles the descent of $G$, that is, there can be some fluctuations in the descent process but the minimum over the set of iterations needs to be decreasing over time; and counter $i_3^{12}$ specifies the maximum number of iterations to be performed, if the descent process doesn’t attain the minimum before it. In the second step, once the descent process is terminated, one analyzes the graph of the solution norms $||\psi_m||_{L^2}$, the error norms $||T\psi_m - g_\delta||_{L^2}$ and the descent of functional $G_2$, over

\[12\text{In all our experiments we used the following bounds for the counters, } max1 = 5, max2 = 10 \text{ and } iter\_max = 50.\]
the set of iterations, to identify certain pivotal points; where we call a point to be pivotal if any one of the graphs changes its nature suddenly at that point\textsuperscript{13}. For example, Figures 9b, 9a and 11a show the graphs of the solution norms $||\psi_m||_{L^2}$, the error norm $||T\psi_m - g||_{L^2}$ and functional $G_2(\psi_m)$, respectively, for the Shaw test problem, Example 7.3. The pivotal points are plotted in the graphs, which are $m = 6, 13, 15, 20, 34$ and $43$. We call the first pivotal point as the reference point and it is the first point where the error norm $||T\psi_m - g||_{L^2}$ attains a stable state; for example in Figure 9a iteration 6 can be considered as the reference point. The choice of the reference point is very crucial, as we compare the results obtained at other pivotal points to the reference point result, that is, we compare $\psi_m$’s to $\psi_6$ for $m = 13, 15, 20, 34, 43$. The largest $m$ for which $||\psi_m||_{L^2}$ is close to $||\psi_6||_{L^2}$ and has the least value in $G_2(\psi_m)$ is considered as the recovered solution. Figure 12a shows the recoveries at different pivotal points, where we see that the graph of $\psi_{15}$ is close to $\psi_6$ and then the ill-posedness creeps in as $m$ increases. The relative error norm of $\psi_{15}$ is 9.16% and the least relative error norm is attained by $\psi_{14}$ (9.02%) during the descent process. The relative errors in the recoveries at the pivotal points 6, 13, 20, 34, 43 are 16.20%, 10.61%, 32.66%, 38.28% and 123%, respectively. Similarly, Figures 11b and 10b show the graph of functional $G_2$ descent and the $||\psi_m||_{L^2}$ graph, together with the pivotal points, for the Baart test problem, Example 7.6. Figure 12b shows the different recoveries at the pivotal points, where $\psi_7$ is considered as the recovered solution. The relative errors in the recoveries at the pivotal points $m = 6, 7, 9, 10, 16, 30, 33, 36, 37, 40, 41$ are 20.39%, 20.32%, 22.35%, 41.5%, 39.34%, 39.26%, 65.07%, 76.40%, 185%, 189%, 224%, respectively, where the least relative error norm (20.32%) is attained at $m = 7$.

**Result:** Inverse recovery of $\varphi$

\begin{verbatim}
while $i1 < maxi1 \& \& i2 < maxi2 \& \& i3 \leq iter_max$ do
    The descent process;
    ...
    ...
    if $G(\psi_m) > G(\psi_{m-1})$ then
        $i1 = i1 + 1$;
    else
        $i1 = 0$;
    end
    if $G(\psi_m) > \min[G(\psi_0) to G(\psi_{m-1})]$ then
        $i2 = i2 + 1$;
    else
        $i2 = 0$;
    end
    $i3 = i3 + 1$;
end

Algorithm 1: Termination condition for the descent process
\end{verbatim}

\textsuperscript{13}excluding the first few iterations, that is, we focus on the graphs after the first few sharp descents.
Remark 8.1. Note that when \( \delta \) is known, then the iteration number obtained using stopping criterion I (§section 6) can serve as the reference point. In fact one can observe that the first pivotal point obtained heuristically (when \( \delta \) unknown) often coincides with the reference point obtained using the stopping criterion I (when \( \delta \) is known). Hence, one doesn’t need to critically depend on the noise information and can still have a good (or even better) recovery, using stopping criterion II, provided the reference point is estimated appropriately.

Remark 8.2. Note that the ill-posedness of the problem is only reflected in the descent of functional \( G_2 \). Figures 13a and 13b show the descent of functional \( G_1 \) for the Shaw and Baart examples, respectively, which looks very similar to their respective error norms graph \( \|T\psi_m - g\|_{L^2} \), but one can not find the hiccups as seen in the \( G_2 \) descent, figures 11a and 11b.

9. Conclusion and future research

Here we provide a regularization method that does not require any external parameter and there by avoids the associated problems. We present various descent directions for the minimization processes depending on the applications, which not only speed up the rate of descent but also improve the efficiency of the recovery. The stopping criteria also plays a very important role in the descent process of
an inverse recovery. In most real world problems the error norm is not known, and even if it is known then sometimes following the simple stopping criterion I, in Section 6, does not result in optimal recovery. Hence one can combine it with the stopping criterion II, in Section 8, to further improve the recovery or when $\delta$ is unknown.

(a) $G_2$-descent for Shaw problem, Example 7.3.  
(b) $G_2$-descent for Baart problem, Example 7.6.

**Figure 11.** Functional $G_2$-descent for Shaw and Baart problem.

(a) Shaw test problem.  
(b) Baart test problem.

**Figure 12.** Recoveries at the pivotal points.

(a) $G_1$-descent for Shaw problem, Example 7.3.  
(b) $G_1$-descent for Baart problem, Example 7.6.

**Figure 13.** Functional $G_1$-descent for Shaw and Baart problem.
In a follow up paper we are going to extend this regularization method to inverse problems in higher dimensions. We are also trying to develop a similar theory using the $L^1$-norm for $G_2$, similar to total variation regularization, and generalize the approach to non-linear inverse problems in Banach spaces.

ACKNOWLEDGMENT

I am very grateful to Prof. Ian Knowles for his support, encouragement and stimulating discussions throughout the preparation of this paper.

REFERENCES

1. Nayak, Abinash. The Inverse Problem of Numerical Differentiation., in arXiv, (submitted to Applied Mathematics and Computation, Dec 2018).
2. Engl HW, Hanke M, Neubauer A., Regularization of inverse problems, vol. 375. Springer Science & Business Media, 1996.
3. Knowles, Ian W., Descent methods for inverse problems, Nonlinear Analysis, 2001.
4. M. Hanke, A. Neubauer and O. Scherzer, A convergence analysis of Landweber iteration for nonlinear ill-posed problems, Numer. Math., 72:21-37, 1995.
5. O. Scherzer, A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems, Numer. Funct. Anal. Optimiz., 1996.
6. O. Scherzer, A Modified Landweber Iteration for Solving Parameter Estimation Problems, Appl Math Optim., 1998.
7. A. B. Bakushinskii, The problems of the convergence of the iteratively regularized Gauss-Newton method, Comput. Math. Math. Phys., 32:1353-1359, 1992.
8. Alessandro Buccini, Marco Donatelli, and Lothar Reichel, Iterated Tikhonov regularization with a general penalty term, Numer. Linear Algebra Appl., (2017).
9. P.C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Rev., 1992.
10. C.L. Lawson and R.J.Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
11. Jin Qi-Nian, On the iteratively regularized Gauss-Newton method for solving nonlinear ill-posed problems, Math. of Computation, 2000.
12. Blaschke, A. Neubauer and O. Scherzer, On convergence rates for the iteratively regularized Gauss-Newton method, IMA J. Numer. Anal., 1997.
13. Martin Benning and Martin Burger, Modern regularization methods for inverse problems, Acta Numerica, Vol. 27, 2018.
14. F. Lenzen and O. Scherzer Tikhonov type regularization methods: history and recent progress, ECCOMAS - 2004.
15. C.W.Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, Boston, 1984.
16. M.Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse problems, 13:79-95, 1997.
17. B. Kaltenbacher, Some Newton type methods for the regularization of nonlinear ill-posed problems, Inverse Problems, 14:729-753,1997.
18. B. Kaltenbacher, A. Neubauer, and A.G.Ramm. Convergence rates of the continuous regularized Gauss-Newton method, J. Inv.Ill-Posed Problems, 10:261-280, 2002.
19. L. Landweber. An iteration formula for Fredholm integral equations of the first kind, Amer.J. Math., 73:615-624, 1951.
20. B. Kaltenbacher. A posteriori parameter choice strategies for some Newton type methods for the regularization of nonlinear ill-posed problems, Numer.Math., 79:501-528, 1998.
21. B. Kaltenbacher, A. Neubauer, and O. Scherzer. Iterative Regularization Methods for Non-linear Problems., Radon Series on Computational and Applied Mathematics.
22. M.Hanke. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems, Numer. funct. anal. optim., 18:165-183, 1997.
23. Louis A K, Inverse und Schlecht Gestellte Probleme, 1989 (Stuttgart: Teubner).
24. Gfrerer H, An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates, Math. Comput., 49 50722, 1987.
25. Morozov V A, On the solution of functional equations by the method of regularization, Sov. Math. Dokl. 7 4147, 1966.
26. Vainikko G M, The principle of the residual for a class of regularization methods, USSR Comput. Math. Math. Phys. 22 119, 1982.
27. Tautenhahn U and Hämärik, The use of monotonicity for choosing the regularization parameter in ill-posed problems, Inverse Problems, 15 1487505, 1999.
28. Bauer F and Hohage T, A Lepskij-type stopping rule for regularized newton methods, Inverse Problems 21 197591, 2005.
29. Mathe P, The lepskii principle revisited. Inverse Problems, 21 L115, 2005.
30. Martin Fuhry and Lothar Reichel, A new Tikhonov regularization method, Numer. Algor., 59:433-445, 2012.
31. V.A.Morozov, Methods for Solving Incorrectly Posed Problems, Springer-Verlag, NewYork, 1984.
32. Shuai Lu and Sergei V. Pereverzev Numerical differentiation from a viewpoint of Regularization Theory, Mathematics of Computation, 2006.
33. Ian Knowles, Robert J. Renka. Methods for numerical differentiation of noisy data, Proceedings of the Variational and Topological Methods: Theory, Applications, Numerical Simulations, and Open Problems, 235246, Electron. J. Differ. Equ. Conf., 21, Texas State Univ., San Marcos, TX, 2014.
34. Hansen, P. C., Regularization tools version 4.0 for MATLAB 7.3, Numer. Algor. 46, 189-194, 2004.
35. R Ramlau, G Teschke and M Zhariy A compressive Landweber iteration for solving ill-posed inverse problems., Inverse Problems 24(2008) 065013 (26pp)
36. H. Egger, Accelerated Newton-Landweber Iterations for Regularizing Nonlinear Inverse Problems, RICAM-Report 2005-01

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALABAMA AT BIRMINGHAM, CAMPBELL HALL, RM. 452, 1300 UNIVERSITY BLVD., BIRMINGHAM, AL 35233
E-mail address: nash101@uab.edu