A discrepancy principle for the Landweber iteration based on risk minimization

Federico Benvenuto[a,*], Cristina Campi[b]

[a]*Dipartimento di Matematica, Università degli Studi di Genova, Via Dodecaneso 35, 16146, Genova, Italy*
[b]*Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, Via Trieste 63, 35121, Padova, Italy*

Abstract

In this paper we propose a criterion based on risk minimization to stop the Landweber algorithm for estimating the solution of a linear system with noisy data. Under the hypothesis of white Gaussian noise, we provide an unbiased estimator of the risk and we use it for defining a variant of the classical discrepancy principle. Moreover, we prove that the proposed variant satisfies the regularization property in expectation. Finally, we perform some numerical simulations when the signal formation model is given by a convolution or a Radon transform, to show that the proposed method is numerically reliable and furnishes slightly better solutions than classical estimators based on the predictive risk, namely the Unbiased Predictive Risk Estimator and the Generalized Cross Validation.

Keywords: Landweber algorithm, Risk minimization, Stopping rule, Regularization, Inverse problems

1. Introduction

The Landweber iterative method was proposed in [1] for solving Fredholm integral equations of the first kind. This method received a lot of attention in the eighties because of its easy implementability and relatively low computational cost reaching a certain degree of popularity in inverse problems literature becoming a reference algorithm in many text books [2, 3, 4, 5, 6]. Usually, since data are discrete and corrupted by noise, the Landweber algorithm presents a semi-convergent behavior, i.e. the iterative process converges to the true solution until a certain iterate, and thereafter it diverges away from the true solution [2]. In the framework of discrete inverse problems, a vast literature exists on the semi-convergence property, and on how to early stop the iteration (see for example [7] and references therein). The available methods to stop the iteration can be subdivided according to the information needed to estimate the optimal iterate: among

---

*Corresponding author
Email addresses:* `benvenuto@dima.unige.it` (Federico Benvenuto), `cristina.campi@unipd.it` (Cristina Campi)

the methods requiring the knowledge of the data error norm within reasonable accuracy, we can recall the discrepancy principle (DP) [8], the Mallow's criterion [9] - also known as unbiased predictive risk estimator (UPRE) - and some heuristic parameter choice rules [10]; among the methods that do not require the knowledge of the data error norm, we mainly recall the generalized cross validation (GCV) method [11] and the L-curve criterion [12]. The performance of these methods strongly depends on the forward operator and, in general, no method works better than the others. From the theoretical point of view, each of these methods has its own interesting property: DP selects a solution during the convergent regime provided that the convergence is computed with respect to the ℓ_2 norm [13]. Nonetheless, this solution is often not optimal as performing the iteration process beyond the selected solution can lead to better estimates of the true solution. On the other hand, it has been proved that UPRE, as well as GCV, selects the solution on the basis of an estimate of the predictive risk. Therefore, the selected solutions are not necessarily optimal with respect to the ℓ_2 norm of the solution space.

In this paper we focus on the risk minimization and we compute ℓ_2 norm error between the iteratively reconstructed solutions and the true solution as a measure for evaluating semi-convergence. We show that, when the Landweber method is initialized with the 0's solution the difference between the 2-norm reconstruction error of two successive iterates can be written without the need of the true solution Moreover, assuming that the noise corrupting the data is white Gaussian, we prove that such function is an unbiased estimator of the discrete derivative of the risk computed between two successive Landweber iterates. Therefore, selecting the last iterate for which the discrete derivative of the risk is negative furnishes a discrepancy principle which is adaptive in the sense that the lower bound of the discrepancy is iteration dependent. Finally, we test the proposed discrepancy principle in the case of three different inverse problems: deblurring, deconvolution, and inverse Radon transform. This letter is organized as follows. In Section 2 we set up the notation and we present the theoretical results on the reconstruction error, the novel discrepancy principle for the Landweber algorithm and its regularization property. In Section 3 we show the application of the proposed method to three two-dimensional reconstruction problems. We conclude the paper in Section 4.

2. A stopping rule based on risk minimization

We consider a linear data formation model $Ax = y$ where $x \in \mathbb{R}^N$ is a vector of unknown parameters, $y \in \mathbb{R}^M$ is the vector of measured data and A is the matrix mapping an object in the corresponding data. A linear discrete inverse problem consists in finding an estimate of x starting from noisy data $y^\sigma \neq y$. **We consider that A is ill-conditioned, i.e. a small change in the data may result in a large variation of the parameters. Therefore, the generelized inverse solution $x^\dagger = A^\dagger y^\sigma$ is not physical being corrupted by the noise amplification. To overcome this issue we consider the Landweber method which** furnishes the following family of estimates

$$x_{k+1}^\sigma = x_k^\sigma - \tau A^T (A x_k^\sigma - y^\sigma). \tag{1}$$

Taking the limit with respect to k, x_k^σ converges to a solution of the least squares problem $\arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y^\sigma - Ax\|^2$ provided that $0 < \tau < 2\|A^T A\|^{-1}$. Moreover, if the initial point $x_0 = 0_N \in \mathbb{R}^N$ then the algorithm converges to the generalized solution $x^\dagger$. Under this assumption, each iterate can be written as $x_k^\sigma = L_k y^\sigma$ with $L_k = \tau \sum_{j=0}^{k-1} (I - \tau A^T A)^j A^T$. The first result of this work is that the difference between the reconstruction error at two successive iterations can be

2

written regardless x^* but only as a function of y^*. **This result is crucial to the following analysis and it is not restricted solely to the Lamdweber case, but it can be extended to all the regularization methods that provide estimated solutions belonging to $\ker(A)^\perp$.**

Theorem 1. Reconstruction error. *The difference between the reconstruction errors of two successive iterates of the Landweber algorithm can be written independently of x^*. In particular we have*

$$\|x_{k+1}^\sigma - x^*\|^2 - \|x_k^\sigma - x^*\|^2 = f_k(\sigma, y^\sigma, y^*) \tag{2}$$

where

$$f_k(\sigma, y^\sigma, y^*) = -2\tau \|r_k^\sigma\|^2 + \tau^2 \|A^T r_k^\sigma\|^2 + 2\tau \sum_{i=1}^{M} (y^* - y^\sigma)_i (r_k^\sigma)_i \tag{3}$$

where $r_k^\sigma := Ax_k^\sigma - y^\sigma$ is the residual.

PROOF. The reconstruction error of the $k+1$ iterate can be written as

$$\|x_{k+1}^\sigma - x^*\|^2 = \|x_k^\sigma - x^* - \tau A^T r_k^\sigma\|^2 = \|x_k^\sigma - x^*\|^2 + \tau^2 \|A^T r_k^\sigma\|^2 - 2\tau \sum_{i=1}^{N} (x_k^\sigma - x^*)_i (A^T r_k^\sigma)_i \tag{4}$$

where the last term of the r.h.s. takes the form

$$-2\tau \sum_{j=1}^{M} (x_k^\sigma - x^*)_j (A^T r_k^\sigma)_j \quad = \quad 2\tau \sum_{i=1}^{N} (y^* - Ax_k^\sigma)_i (r_k^\sigma)_i = 2\tau \sum_{i=1}^{N} (y^* - y^\sigma)_i (r_k^\sigma)_i - 2\tau \|r_k^\sigma\|^2 .$$

$\square$

We now consider that y^σ is the realization of a random variable Y with given unknown mean y^* and known (or estimated) covariance matrix $\sigma^2 I$. We can approximate the last term of equation 3 by means of the following results.

Lemma 2. *We have $\mathbb{E}\left(\sum_i (y^* - y^\sigma)_i (r_k^\sigma)_i\right) = -\sigma^2 \mathrm{tr}(R_k)$ **where $R_k(y^\sigma) := r_k^\sigma = Ax_k^\sigma - y^\sigma$ is the residual operator**.*

PROOF. The result is a straightforward consequence of the Trace Lemma (see Chapter 7 in [14]).

$\square$

Thanks to this result, we can state the following

Theorem 3. Unbiasedness. *The function*

$$\bar{f}_k(\sigma, y^\sigma) = -2\tau \|R_k y^\sigma\|^2 + \tau^2 \|A^T (R_k y^\sigma)\|^2 - 2\tau \sigma^2 \mathrm{tr}(R_k) \tag{5}$$

is an unbiased estimator of discrete derivative of the risk computed between two successive iterates, i.e. $\mathbb{E}(\bar{f}_k(\sigma, y^\sigma)) = \mathbb{E}(\|x_{k+1}^\sigma - x^*\|^2 - \|x_k^\sigma - x^*\|^2)$.

This unbiased estimator allows us to define a novel criterion to stop the Landweber iteration. The idea is to select the first iterate for which the risk is non-decreasing, i.e. $\overline{f} \geq 0$. We define the function

$$K(\sigma, y^\sigma) = \min_{k \in \mathbb{N}} \{k \mid \bar{f}_k(\sigma, y^\sigma) \geq 0\} \tag{6}$$

Therefore, the function $K(\sigma, y^\sigma)$ selects the iterate such that

$$\|R_k y^\sigma\|^2 \leq \frac{\tau}{2}\|A^T(R_k y^\sigma)\|^2 - \sigma^2 \mathrm{tr}(R_k). \tag{7}$$

Inequality 7 can be considered as a variant of the classical discrepancy principle, as it is verified when the discrepancy is smaller than an iteration-dependent quantity instead of a constant proportional to $M\sigma^2$. For this reason we refer to it as the Risk-based Discrepancy Principle (RDP). The numerical implementation of the RDP requires to compute the trace of the operator R_k at each iteration. This implies that the proposed method have the same overall computational burden as the GCV or the UPRE method.

We now prove the following

Theorem 4. Finiteness. $K(\sigma, y^\sigma)$ *is finite for each* σ *and* y^σ.

PROOF. Since
$$\|R_k y^\sigma\|^2 - \frac{\tau}{2}\|A^T(R_k y^\sigma)\|^2 = \langle R_k y^\sigma, (I - \frac{\tau}{2}AA^T)R_k y^\sigma\rangle,$$

inequality in equation 7 can be written as

$$\langle R_k y^\sigma, (I - \frac{\tau}{2}AA^T)R_k y^\sigma\rangle \leq -\sigma^2 \mathrm{tr} R_k \tag{8}$$

where we have that $\|R_k\| \to 0$ as $k \to \infty$. The l.h.s is non-negative and tends to zero of order 2 w.r.t. k, and the r.h.s. is positive with order 1 as R_k is negative definite for each $k \in \mathbb{N}$. Therefore the l.h.s goes to zero faster than the r.h.s. and then there exists a $\bar{k}$ such that equation 8 is satisfied. This proves that $K(\sigma, y^\sigma)$ is finite. $\qquad\square$

The RDP makes the Landweber iteration a regularization method:

Theorem 5. Regularization. *Let us consider the Landweber algorithm initialized with a zero object* $x_0 = 0_N$ *and suppose* $0 < \tau < 2\|A^T A\|^{-1}$. *Let* $x^\sigma_{K(\sigma, y^\sigma)}$ *be the RDP solution. Then, we have*

$$\lim_{\sigma \to 0} \mathbb{E}\|x^\sigma_{K(\sigma, y^\sigma)} - x^*\|^2 = 0. \tag{9}$$

PROOF. Let $x := \lim_{k \to \infty} L_k(y)$. Let $\{\sigma_n\}_{n=1}^\infty$ be a sequence converging to zero when $n \to \infty$. Denote with $k_n := K(\sigma_n, y^{\sigma_n})$ the stopping index determined by the RDP criterion applied to L_k. Let k be a finite accumulation point of k_n. From the stopping criterion we have:

$$\lim_{n \to \infty} \|R_k y^{\sigma_n}\|^2 - \frac{\tau}{2}\|A^T(R_k y^{\sigma_n})\|^2 + \sigma_n \, \mathrm{tr}(R_k) = \langle R_k y, (I - \frac{\tau}{2}AA^T)R_k y\rangle \leq 0$$

Since $I - \frac{\tau}{2}AA^T$ is positive definite, we have that $R_k y = 0$ and then x_k is such that $Ax_k = y$. Then, $x^{\sigma_n}_{k_n} \to x_k$ and equation 9 is satisfied.

If $k_n \to \infty$ as $n \to \infty$, we can assume that k_n increases monotonically with n. From the definition of $K(\sigma, y^\sigma)$ follows that $\mathbb{E}(\bar{f}_j) \leq 0$ for each $j < k_n$. Then, for $m < n$, we know that

$$\mathbb{E}\left(\|x^{\sigma_n}_{k_n} - x^*\|^2\right) - \mathbb{E}\left(\|x^{\sigma_n}_{k_m} - x^*\|^2\right) = \mathbb{E}\left(\sum_{j=k_m}^{k_n-1} \bar{f}_j\right) \leq 0. \tag{10}$$

4

Consider $x_{k_m} = L_{k_m} y$. By applying the Cauchy-Schwartz inequality to the second term of the l.h.s, in equation 10 we have

$$\mathbb{E}(\|x_{k_n}^{\sigma_n} - x^*\|^2) \le \mathbb{E}(\|x_{k_m}^{\sigma_n} - x^*\|^2) \le C_{n,m} + B_m + 2\sqrt{B_m C_{n,m}} \tag{11}$$

where $C_{n,m} := \mathbb{E}(\|x_{k_m}^{\sigma_n} - x_{k_m}\|^2)$ and $B_m := \|x_{k_m} - x^*\|^2$. By using again the Trace Lemma (see [14]),

$$C_{n,m} = \|(L_{k_m} - I)y^*\|^2 + \sigma_n^2 \frac{1}{M} \mathrm{tr}(L_{k_m}^2)$$

and therefore, given $\varepsilon > 0$, we can fix m so large that $C_{n,m} \le \varepsilon + \sigma_n^2$, and $B_m \le \varepsilon$. Moreover, by taking an n large enough, we can bound $\sigma_n^2 \le \varepsilon$. Thus, we have $\mathbb{E}(\|x_{k_n}^{\sigma_n} - x^*\|^2) \le (3 + 2\sqrt{2})\varepsilon$. This shows that $x_{k_n}^{\sigma_n}$ converges in expectation to x^* as σ goes to 0. $\qquad\square$

3. Experimental results

In this section we consider three image reconstruction problems with the aim of testing the performance of the proposed RDP. The first is a deblurring problem (Motion Blur - MB) where the data is a blurred natural image and the forward operator represents the distortion due to the straight motion of a photo-camera. The second one is a deconvolution problem (Radial Blur - RB) where the data is a galaxy with a central large smooth signal and some small sharp features blurred with a radial point spread function representing the optical effect due to the acquisition. In these cases we have $N = M = 256^2$. The third problem is the inversion of a sinogram obtained by the Radon Transform (RT) of the Sheep-Logan phantom with $N = 128^2$ and $M = 185 \times 180$. For each one of these problems, we generate $N_{runs} = 1000$ replicates by adding to the exact data y^* as many noisy samples drawn from a white Gaussian distribution with a given variance σ^2. We fix the signal to noise ratio to 20 dB for MB, 30 dB for RB and 40 dB for RT problem where the signal to noise ratio is $snr = 20 \, \log_{10} \frac{\|y^\sigma\|}{\|y^\sigma - y^*\|}$. We solve each replicate by means of the Landweber algorithm and we select the best iterate according to the RDP and to the standard UPRE and GCV criteria, in order to compare their performances.

As a benchmark to evaluate the performances, we select the stopping index k^* minimizing the ℓ_2 reconstruction error $k^* = \arg\min_k \|x_k^\sigma - x^*\|$ and we refer to $x_{k^*}^\sigma$ as the "oracle solution". In particular, for all the replicates of each problem, we compute the relative error with respect to k^* as follows:

$$\mathscr{E}(k_{<>}) = \frac{k_{<>} - k_*}{k_*} \tag{12}$$

where $k_{<>}$ can be k_{RDP}, k_{GCV}, k_{UPRE}. Note that $\mathscr{E}(k_{<>}) = 0$ is the best achievable error. Moreover we compute the error on the corresponding solutions normalized over the error provided by the oracle solution:

$$\mathscr{E}(x_{k_{<>}}^\sigma) = \frac{\|x_{k_{<>}}^\sigma - x^*\|_2}{\|x_{k_*}^\sigma - x^*\|_2} \ . \tag{13}$$

Note that $\mathscr{E}(x_{k_{<>}}^\sigma) \ge 1$ and the best achievable error is 1. As a result, for each problem we have a distribution of N_{runs} samples of the error on the stopping index and the corresponding distribution of the error on the solution, both normalized with respect to the oracle index and error. In the first row of Figure 1 we plot the distribution of the stopping index error (equation 12) while in the second row we plot the distribution of the reconstruction error (equation 13). From the first row
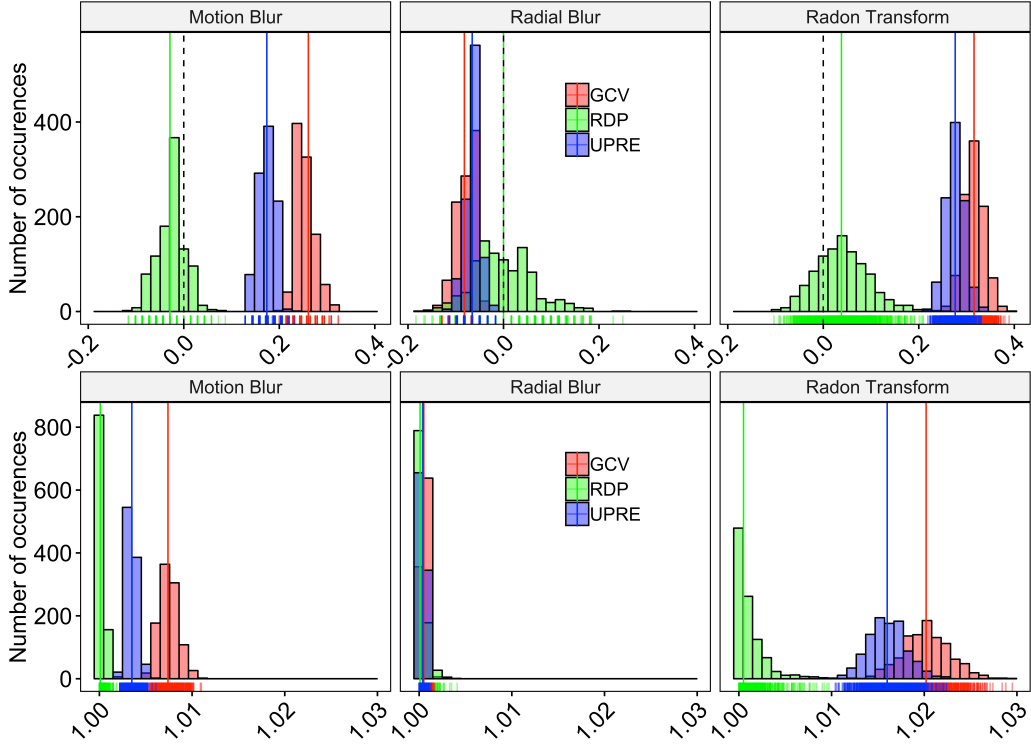
Figure 1: First row: distributions of $\mathscr{E}(k_{RDP})$ (green), $\mathscr{E}(k_{GCV})$ (red), and $\mathscr{E}(k_{UPRE})$ (blue) computed over N_{runs} noise realizations for the MB, RB, and RT problems. The vertical colored lines represent the medians of the distributions. The dashed black line represents the best achievable error 0. Second row: distributions of $\mathscr{E}(x^\sigma_{k_{RDP}})$ (green), $\mathscr{E}(x^\sigma_{k_{GCV}})$ (red), and $\mathscr{E}(x^\sigma_{k_{UPRE}})$ (blue) computed over N_{runs} noise realizations for the MB, RB, and RT problems. The vertical colored lines represent the medians of the distributions.

of Figure 1 it is evident that the RDP distribution is centered around 0 for all the problems, while GCV and UPRE provide significantly biased distributions. Although the latter two distributions show a smaller variance, the corresponding distributions of the reconstruction error (second row of Figure 1) show slightly worse results than the one obtained using RDP.

4. Conclusion

In this work, we introduced a novel stopping rule for the Landweber method in the case of data corrupted by white Gaussian noise. We proved that this stopping rule is an unbiased estimator of the ideal risk and we showed that it provides very good estimates of the "oracle" stopping index and their corresponding "oracle" solutions.

References

[1] L. Landweber, An iteration formula for fredholm integral equations of the first kind, American journal of mathematics 73 (3) (1951) 615–624.

[2] M. Bertero, P. Boccacci, Introduction to Inverse Problems in Imaging, CRC Press, 1998.

[3] H. W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Springer Science & Business Media, 1996.

[4] A. Björck, Numerical Methods for Least Squares Problems, SIAM, 1996.

[5] W. Hackbusch, Iterative solution of large sparse systems of equations, Springer International Publishing, 2016.

[6] R. A. Kress, Linear Integral Equations, Springer New York, 2014.

[7] U. Hämarik, R. Palm, T. Raus, On minimization strategies for choice of the regularization parameter in ill-posed problems, Numerical Functional Analysis and Optimization 30 (9-10) (2009) 924–950.

[8] H. W. Engl, Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates, Journal of Optimization Theory and Applications 52 (2) (1987) 209–215.

[9] C. L. Mallows, Some Comments on CP, Technometrics 15 (4) (1973) 661–675.

[10] U. Hämarik, U. Tautenhahn, On the monotone error rule for parameter choice in iterative and continuous regularization methods, BIT Numerical Mathematics 41 (5) (2001) 1029–1038.

[11] G. H. Golub, M. Heath, G. Wahba, Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter, Technometrics 21 (2) (1979) 215–223.

[12] P. Hansen, D. OLeary, The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, SIAM Journal on Scientific Computing 14 (6) (1993) 1487–1503.

[13] M. Defrise, C. De Mol, A note on stopping rules for iterative regularisation methods and filtered svd, Inverse Problems : An Interdisciplinary Study (P. C. Sabatier, ed.), Supplement 19 in the Series "Advances in Electronics and Electron Physics" (1987) 261–268.

[14] C. R. Vogel, Computational methods for inverse problems, SIAM, 2002.