Journal of Physics: Conference Series

OPEN ACCESS

Inverse problems in machine learning: An application to brain activity interpretation

To cite this article: M Prato and L Zanni 2008 *J. Phys.: Conf. Ser.* **135** 012085

View the article online for updates and enhancements.

You may also like

- A new formulation of gradient boosting
Alex Wozniakowski, Jayne Thompson, Mile Gu et al.
- Direct estimation of gas holdup in gas–liquid bubble column reactors using ultrasonic transmission tomography and artificial neural processing
Jingyi Hu, Nan Li, Lina Wang et al.
- An accurate and reliable operational transfer path analysis for transfer path contribution evaluation based on Landweber iterative method
Wei Cheng, Yan Zhu, Jingbai Lu et al.

This content was downloaded from IP address 93.36.216.167 on 22/05/2023 at 12:58

Journal of Physics: Conference Series **135** (2008) 012085 doi:10.1088/1742-6596/135/1/012085

Inverse problems in machine learning: an application to brain activity interpretation

M Prato[1] and L Zanni[2]

[1] CNR-INFM LAMIA, Via Dodecaneso 33, Genova, I-16146, Italy
[2] Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via G. Campi 213/B, I-41100 Modena, Italy

E-mail: marco.prato@unimore.it

Abstract. In a typical machine learning problem one has to build a model from a finite training set which is able to generalize the properties characterizing the examples of the training set to new examples. The model has to reflect as much as possible the set of training examples but, especially in real-world problems in which the data are often corrupted by different sources of noise, it has to avoid a too strict dependence on the training examples themselves. Recent studies on the relationship between this kind of learning problem and the regularization theory for ill-posed inverse problems have given rise to new regularized learning algorithms. In this paper we recall some of these learning methods and we propose an accelerated version of the classical Landweber iterative scheme which results particularly efficient from the computational viewpoint. Finally, we compare the performances of these methods with the classical Support Vector Machines learning algorithm on a real-world experiment concerning brain activity interpretation through the analysis of functional magnetic resonance imaging data.

1. Introduction

Statistical inference, which means to build a probabilistic model based on a certain type of a priori knowledge and which is able to make predictions, is finding more and more space in real-life applications. In several cases, the amount of information at one's disposal is huge either because it is relatively easy to get a high number of data or because the size of the data themselves is big due to the elevate number of features which characterize the problem. Dealing with such a quantity of information can be very difficult and often the computational power one can have use of, despite it is in continuous growth, is not sufficient to obtain the desired results in a reasonable time. Therefore, it is necessary both to look for possible ways to accelerate the classical machine learning algorithms already existent in literature and to identify new methods which can compete with the traditional ones in speed, robustness and quality of the results.

Thanks to the strict connection between machine learning and regularization theory for linear inverse problems already described in many studies (see for example [6, 9] and references therein), it is possible to exploit many regularization methods originally proposed in the inverse problems framework in order to build machine learning methods which result very simple both from a purely intuitive point of view and with regard to the implementation on a computer machine [7, 14]. These algorithms have been theoretically studied in details but, as far as we know, they have been rarely used in applications with real data and their performances have been compared

© 2008 IOP Publishing Ltd

to the ones of classical machine learning methods mainly in the case of classification problems with benchmark datasets.

An opportunity to apply and test the methods above in the case of regression problems has been recently offered by the Experience Based Cognition (EBC) project of the University of Pittsburgh, Pennsylvania concerning the general problem of inferring subjective experience through the analysis of functional Magnetic Resonance Imaging (fMRI) data. The fMRI is a powerful non-invasive technique for the investigation of brain activity resulting from sensory stimulation or cognitive functions [2]. The focal objective of fMRI data analysis is to disclose the distributed patterns of activated brain areas involved in specific function and their interactions, by applying a variety of univariate or multivariate statistical approaches. However, in the last few years, a growing number of studies have taken a different approach, where the direction of analysis is reversed in order to probe whether the acquired fMRI signals can be used to *predict* perceptual or cognitive states [12]. To this end, many different methods are proposed in literature (Recurrent Neural Networks, Dynamic Gaussian Markov Random Field, Support Vector Machines for Regression, Relevance Vector Regression) but so far no one distinguished itself in such a way that it could be considered universally better than the others. The challenge issued from the EBC staff in 2006 through the Pittsburgh Brain Activity Interpretation Competition (http://www.ebc.pitt.edu/2006/competition.html) allows to advance the analysis of the different approaches on an interesting real-life problem. This competition provided high quality 3 Tesla Echo-Planar Imaging BOLD-fMRI data from three subjects acquired during passive viewing of three movie clips and, only for two movies, the rating data representing the subjective on going relevance of 13 features present in the videos (like for example food, music, amusement, etc.). The goal of the competition was to find, for each subject, a good prediction of the rating data relative to the third movie.

In this work, starting from a formulation of the regression problem as an inverse problem, we introduce some learning algorithms derived by well known regularization methods and we evaluate their effectiveness on the regression problems arising in the analysis of the fMRI data made available within the above competition. We consider classical regularization approaches, such as the Landweber Iterative algorithm [4] and the ν-method [8], and propose a generalization of the Landweber scheme derived by exploiting recent ideas on the adaptive steplength selections in gradient methods [21]. These regularization methods are evaluated in comparison with the state-of-the-art learning methodology Support Vector Machines (SVMs) [17]. From our experiments we may observe that the regularization techniques, despite their simplicity, are able to provide an accuracy in the reconstructed behavioral rating very well-comparable with that of the SVMs. Furthermore, both the ν-method and the generalized Landweber method largely reduce the iterations required by the standard Landweber algorithm and, in this application, they seem a valid alternative to the SVMs in terms of computational costs.

2. Learning as an inverse problem

The problem of learning from examples can be stated in this way: given a set of examples $\{(\boldsymbol{x}_i, y_i) \ : \ i = 1, \ldots, n\} \subseteq X \times [-M, M]$, with X compact subset of $\mathbb{R}^m$ and $M > 0$, find the solution f_λ of the minimization problem

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} V(y_i, f(\boldsymbol{x}_i)) + \lambda \|f\|_K^2, \tag{1}$$

where $V(\cdot, \cdot)$ is a positive loss function, $\| \cdot \|_K$ is a norm in a Reproducing Kernel Hilbert Space $\mathcal{H}$ defined by the positive definite kernel K and λ is a positive regularization parameter which handles the trade-off between complexity of the solution and error on the training set.

The particular choice

$$V(y_i, f(\boldsymbol{x}_i)) = |y_i - f(\boldsymbol{x}_i)|_\varepsilon := \max(|y_i - f(\boldsymbol{x}_i)| - \varepsilon, 0) \tag{2}$$

where $\varepsilon \geq 0$, leads to the so-called Support Vector Machines for Regression (SVMR) algorithm [9, 17]. The solution of (1) in the case of the loss function (2) can be found through the technique of Lagrange multipliers. The result is that the solution of the problem (1) can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) K(\boldsymbol{x}, \boldsymbol{x}_i) + b, \tag{3}$$

where $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}$ are the solution of the quadratic problem

$$\begin{array}{ll} \min & \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \varepsilon\sum_{i=1}^{n}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \\ \text{sub. to} & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0, \quad 0 \leq \alpha_i^*, \alpha_i \leq \frac{1}{2n\lambda}, \quad i = 1, \ldots, n, \end{array} \tag{4}$$

and their knowledge leads to the threshold b through the Kuhn-Tucker conditions.
SVMR represent a classical approach in machine learning and have been deeply studied in the last decades. In recent years it has been shown [6, 9] that the general machine learning problem can be restated as a linear inverse problem and several algorithms typically used to reconstruct a stable solution of a linear unstable inverse problem have been adapted to the machine learning framework. This connection can be easily seen by considering the quadratic loss function

$$V(y_i, f(\boldsymbol{x}_i)) = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2 \tag{5}$$

and by introducing the sampling operator $S_{\boldsymbol{x}} : \mathcal{H} \longrightarrow \mathbb{R}^n$ defined by $(S_{\boldsymbol{x}} f)_i = f(\boldsymbol{x}_i)$, $i = 1, ..., n$. From the relation

$$\|S_{\boldsymbol{x}} f - \boldsymbol{y}\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2, \tag{6}$$

where $\|\cdot\|_n$ is $1/n$ times the Euclidean norm in $\mathbb{R}^n$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, we can notice that the solution f_λ of the minimization problem (1) with the loss function (6) is equal to the Tikhonov regularized solution of the linear inverse problem $S_{\boldsymbol{x}} f = \boldsymbol{y}$ [16].
The explicit form of the solution f_λ can be obtained from the representer theorem [18], for which

$$f_\lambda(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i), \quad \boldsymbol{\alpha} = (\boldsymbol{K} + n\lambda \boldsymbol{I})^{-1} \boldsymbol{y}, \tag{7}$$

where $\boldsymbol{K}$ is the kernel matrix defined by $(\boldsymbol{K})_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.
We observe from (7) that the role of the positive parameter λ is to stabilize the matrix inversion problem which in general is numerically unstable. Generalizations of this approach are given in [7] by considering suitable functions $g_\lambda : M_n(\mathbb{R}) \longrightarrow M_n(\mathbb{R})$ (where $M_n(\mathbb{R})$ denotes the set of square matrices of order n with entries in $\mathbb{R}$) which define corresponding algorithms by means of

$$f_\lambda(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i), \quad \boldsymbol{\alpha} = \frac{1}{n} g_\lambda\left(\frac{\boldsymbol{K}}{n}\right) \boldsymbol{y}. \tag{8}$$

For instance, two possible functions g_λ are the following:

- $g_\lambda\left(\frac{\boldsymbol{K}}{n}\right) = \left(\frac{\boldsymbol{K}}{n} + \lambda \boldsymbol{I}\right)^{-1}$ (*Tikhonov regularization algorithm*);
- $g_t\left(\frac{\boldsymbol{K}}{n}\right) = \tau \sum_{i=0}^{t-1} \left(\boldsymbol{I} - \tau \frac{\boldsymbol{K}}{n}\right)^i$ (*Landweber iterative algorithm* [4]).

  Here $\tau = (\sup_{\boldsymbol{x} \in X} K(\boldsymbol{x}, \boldsymbol{x}))^{-1}$ and the role of the regularization parameter is played by the iteration number t ($\lambda = t^{-1}$). To compute the vector $\boldsymbol{\alpha}$, we observe that this algorithm can be rewritten in the iterative form

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i-1} + \frac{\tau}{n}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}_{i-1}), \quad i = 1, \ldots, t, \quad \boldsymbol{\alpha}_0 = \boldsymbol{0}. \tag{9}$$

These regularized learning algorithms have recently received an increasing interest due to both theoretical and computational motivations. On one hand, new theoretical results have been obtained by better understanding the connection between learning algorithms and regularization methods in inverse problems [1, 6, 7, 20]. On the other hand, despite their simplicity and low computational requests, the regularized algorithms have already shown promising performances in comparison with state-of-the-art approaches, such as SVMs, on both benchmark datasets and real-life test problems [14, 15]. This scenario strongly encourages to further investigate such kind of approaches in order to overcome possible disadvantages and design improved schemes. For instance, it could be useful to consider accelerated versions of the Landweber iterative algorithm, that often exhibits a very slow convergence rate. To this end, we describe two iterative approaches that are based on generalizations of the standard Landweber iteration.
The first approach is the so called *ν–method* that can be stated as follows:

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i-1} + u_i(\boldsymbol{\alpha}_{i-1} - \boldsymbol{\alpha}_{i-2}) + \frac{\omega_i}{n}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}_{i-1}), \quad i = 1, \ldots, t, \quad \boldsymbol{\alpha}_0 = \boldsymbol{0}, \tag{10}$$

where, given $\nu > 0$,

$$u_i = \frac{(i-1)(2i-3)(2i+2\nu-1)}{(i+2\nu-1)(2i+4\nu-1)(2i+2\nu-3)}, \qquad \omega_i = 4\frac{(2i+2\nu-1)(i+\nu-1)}{(i+2\nu-1)(2i+4\nu-1)}.$$

The good convergence rate of this method is well known in inverse problems [8]: it needs the square root of the number of iterations required by standard Landweber algorithm. Recently, numerical experiments on classification problems have shown its effectiveness also as learning algorithm [14].
The second approach is derived by observing that iterations (9) can be seen as the steps of a standard gradient method, with constant steplength τ, applied to the minimization of the quadratic function

$$\varphi(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T \frac{\boldsymbol{K}}{n}\boldsymbol{\alpha} - \frac{\boldsymbol{y}^T}{n}\boldsymbol{\alpha}.$$

Very effective strategies are nowadays available to accelerate standard gradient methods, mainly based on an appropriate rules for updating the steplength parameter at each iteration (see for example [5, 11, 21] and references therein). Following these acceleration strategies, we may generalize the Landweber iterative scheme by substituting the fixed parameter τ with a variable parameter:

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i-1} + \frac{\tau_{i-1}}{n}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}_{i-1}), \quad i = 1, \ldots, t, \quad \boldsymbol{\alpha}_0 = \boldsymbol{0}. \tag{11}$$

According to this revisal, the coefficient $\boldsymbol{\alpha}$ in (8) becomes

$$\boldsymbol{\alpha} = \frac{1}{n}\left(\tau_0 + \sum_{i=1}^{t-1} \tau_i \left(\prod_{j=0}^{i-1}\left(\boldsymbol{I} - \tau_j \frac{\boldsymbol{K}}{n}\right)\right)\right)\boldsymbol{y}.$$

Gradient methods with adaptive steplength selection have shown promising results on several inverse problem applications [3, 10, 19]; however, to our knowledge they have not been exploited in the framework of the regularization learning algorithms. Thus, a deep analysis should be carried out to identify the selection rules able to ensure an improved convergence rate and good generalization performances. Nevertheless, this goal is beyond the aims of this work and we simply show the interesting behaviour of one of these updating rules on the regression problems considered in our computational study. We focus on the steplength selection exploited in the monotone gradient algorithm called Adaptive Steepest Descent (ASD) method [21]:

$$\begin{cases} \tau_i = \tau_i^{\mathrm{MG}} & \text{if } \tau_i^{\mathrm{MG}}/\tau_i^{\mathrm{SD}} > \gamma\,, \\ \tau_i = \tau_i^{\mathrm{SD}} - 0.5\tau_i^{\mathrm{MG}} & \text{otherwise,} \end{cases} \tag{12}$$

where $\gamma \in (0,1)$ and

$$\tau_i^{\mathrm{SD}} = \underset{\tau\in\mathbb{R}}{\operatorname{argmin}}\ \varphi(\boldsymbol{\alpha}_i - \tau \boldsymbol{g}_i) = \frac{\boldsymbol{g}_i^T \boldsymbol{g}_i}{\boldsymbol{g}_i^T \left(\frac{\boldsymbol{K}}{n}\right) \boldsymbol{g}_i}, \qquad \tau_i^{\mathrm{MG}} = \underset{\tau\in\mathbb{R}}{\operatorname{argmin}}\ \|g(\boldsymbol{\alpha}_i - \tau \boldsymbol{g}_i)\| = \frac{\boldsymbol{g}_i^T \left(\frac{\boldsymbol{K}}{n}\right) \boldsymbol{g}_i}{\boldsymbol{g}_i^T \left(\frac{\boldsymbol{K}}{n}\right)^2 \boldsymbol{g}_i},$$

with $\boldsymbol{g}_i = g(\boldsymbol{\alpha}_i) = \nabla\varphi(\boldsymbol{\alpha}_i)$. The values τ_i^{SD} and τ_i^{MG} are the steplengths used by the well known Steepest Descent (SD) and the Minimal Gradient (MG) methods, respectively, for which the inequality $\tau_i^{\mathrm{MG}} \leq \tau_i^{\mathrm{SD}}$ holds. By equipping the iterative scheme (11) with the steplength selection (12), we obtain a generalization of the standard Landweber method, denoted in the following as L-ASD.

3. Application to fMRI data

In this section we want to investigate the effectiveness of the above regularized learning algorithms on the regression problems arising in the real-life experiment provided by the Pittsburgh Brain Activity Interpretation Competition.

We considered only the data relative to a single subject. They consisted of:

- three matrices (corresponding to three different movie clips), each one with dimensions of order $10^3 \times 10^5$. The number of columns of the matrix represents the number of voxels (i.e., volumetric pixels) in which the brain is sampled, and the values of each column gives the discretized temporal evolution of the fMRI signal in a single voxel. Therefore, each row provides a sort of 3-D picture of the brain in each single time bin;
- two sets of 13 arrays containing the feature ratings (values between 0 and 1) of 13 different categories assigned by the subject during the vision of the first two clips, lasting approximately 20-minute each.

The lack of the behavioral ratings relative to the third clip leads us to consider in this paper only the first and the second clips to be able to estimate the quality of our results. In particular, we train the algorithms with the first clip and test the model on the second clip; an inversion of the roles between training and test sets provided similar results.

In the first step of our approach an accurate pre-processing of the data had to be done in order to substantially decrease the computational cost due to the huge dimensions of the examples and the effects of noise. In fact, one of the most recognized problems of fMRI data is the low signal-to-noise ratio due to the presence of several sources of noise and artifacts with respect to the amplitude of the fMRI response. This pre-processing phase essentially included a voxel selection on the basis of a minimum threshold for the signal followed by the application of a low-pass filter to remove high frequencies due to noise. Furthermore, the voxels were re-sampled in cubes of 27 voxels contiguous in the space; the value of each cube was defined by averaging the values of its voxels. In this way, the size of the examples associated to the time bins was

reduced to the order 10^3.
At this point we applied the learning algorithms to the pre-processed data by using a Gaussian kernel. The variance σ of the Gaussian and the intrinsic parameters of the methods (i.e., the number of iterations t for the iterative schemes and the real positive numbers λ and ϵ for the SVMR algorithm) are chosen via a 10-fold cross validation applied to the training set. The parameters $\nu = 1$ and $\gamma = 0.55$ are used in the ν-method and the L-ASD algorithm, respectively. All the experiments are carried out on a computer equipped with an 1.8GHz AMD Sempron 3100+ in a Windows XP environment; the regularized algorithms are implemented in Matlab 7.0.4 while the binaries of the SVM^{light} code (ver. 6.01) [13] are used for the SVMs approach.
The predictions of the temporal evolutions for the category "Faces" of the first subject are presented in figure 1 together with the theoretical ones.

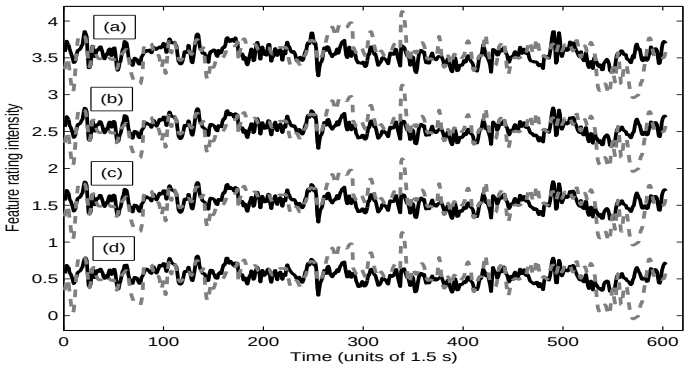

Figure 1. Real (dashed grey lines) and predicted (solid black lines) temporal evolutions of the feature rating "Faces" by using: (a) the SVMR algorithm; (b) the ν-method; (c) the L-ASD method; (d) the Landweber iterative scheme. The values in (a), (b) and (c) have been suitable shifted (of 3, 2 and 1 unit respectively) for the sake of graphic presentation.

Moreover, in table 1 we give two different ways to estimate the prediction error, i.e., by calculating the number $\rho_2 = \|f_\lambda - \boldsymbol{y}\|_2/\|\boldsymbol{y}\|_2$ and the Pearson correlation coefficient $\rho_P(f_\lambda, \boldsymbol{y})$ defined by

$$\rho_P(f_\lambda, \boldsymbol{y}) = \frac{\sum_{i=1}^n \big((f_\lambda)_i - \bar{f_\lambda}\big)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n \big((f_\lambda)_i - \bar{f_\lambda}\big)^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{13}$$

where $\bar{f_\lambda}$ and $\bar{y}$ denote the arithmetic means of f_λ and $\boldsymbol{y}$, respectively.

Looking at the reconstructions errors, as well as at the comparison between the form of the predictions and the real feature ratings, we can make the following considerations:

- the regularized learning algorithms introduced in the paper well compare state-of-the-art SVM methods and the reconstructions obtained are essentially the same. This fact is coherent with the results obtained in the classification framework [14];
- the errors shown in table 1 are consistent with the ones published in the competition web page (http://www.ebc.pitt.edu/2006/competition.html). We remark that we have used just the first movie as the training set in order to evaluate the performance of the methods on the second clip. A training set containing both the first and the second clip would probably lead to a more accurate model.

The behaviour of the three regularized learning algorithms is compared in terms of the number of iterations in table 2, where for each method and for each test problem we report the number

Table 1. Reconstruction errors for the Landweber iterative scheme, the L-ASD method, the ν-method and the SVMR algorithm on five different categories.

		Amusement	Faces	Language	Motion	Sounds
Landweber	ρ_2	0.7555	0.3041	0.3382	0.3934	0.6829
	ρ_P	0.4356	0.4002	0.6866	0.5361	0.4775
L-ASD	ρ_2	0.7571	0.2993	0.3328	0.3954	0.6910
	ρ_P	0.4413	0.4300	0.6973	0.5383	0.4762
ν-method	ρ_2	0.7626	0.2991	0.3326	0.3969	0.6993
	ρ_P	0.4455	0.4309	0.6972	0.5358	0.4690
SVMR	ρ_2	0.7627	0.3083	0.3214	0.3998	0.7389
	ρ_P	0.4521	0.3909	0.7222	0.5263	0.4543

of iterations that in average provides the best reconstruction error in the 10-fold cross validation process (that is, the values assigned to the parameter t to obtain the results in figure 1 and table 1). The accelerated versions seem to behave in a similar way and to imply a very interesting gain in the number of iterations with respect to the standard Landweber scheme (that, in our experiment, has been stopped to 100000 iterations; increasing the number of iterations does not provide better results). This means that the training of the accelerated algorithms is a much less expensive task than the training of the Landweber algorithm.

Table 2. Optimal number of iterations for the Landweber iterative scheme, the L-ASD method and the ν-method for the five categories of table 1. The asterisk means that the maximum number of iterations has been reached.

	Amusement	Faces	Language	Motion	Sounds
Landweber	100000*	100000*	100000*	100000*	100000*
L-ASD	1114	2350	2666	1669	1472
ν-method	1225	2000	1853	1189	2193

Finally, we shortly discuss the computational costs of the regularized algorithms with respect to the SVMs. To this end, we need to take into account both the cost to prefix the parameters of the methods by means of the 10-fold cross validation and the training cost for computing the coefficients that define the predictors (3) and (8). Concerning the parameter choices, we remark that in the SVMs approach the additional parameter ε needs to be prefixed, with a significant increase of the computational time for the cross validation phase. In the training phase, the standard Landweber algorithm is largely penalized by its slow convergence rate while for the ν-method and the L-ASD method no remarkable differences with the SVMs approach are observed (this last sentence could not hold in case of learning problems with many more examples, where the management of the large kernel matrix could become an expensive task in the regularized methods, while specialized software is available for a long time to face these situations in the SVMs framework).
Summarizing, the simplicity, the reconstruction accuracy and the good convergence rate make the two accelerated regularized learning methods very well suited for the considered application.

4. Conclusions and future work

In this work we mesh the clear connection between regularization theory for inverse problems and statistical learning with recent acceleration techniques for classical gradient methods to introduce a new learning algorithm, that can be seen as an accelerated version of the standard Landweber

iterative algorithm. Then, we apply the proposed algorithm together with other schemes derived from the inverse problems framework, the Landweber algorithm and the ν-method, to an actual regression problem concerning brain activity interpretation through functional Magnetic Resonance Imaging and compare them with the SVMs learning algorithm. The numerical results show that the proposed method behaves similarly to the ν-method, providing essentially the same accuracy of the SVMs and a much faster convergence rate with respect to the Landweber algorithm; furthermore, it preserves the simplicity and the low computational cost of the regularized learning algorithms. Thus, both the new accelerated scheme and the ν-method appear valid alternative to the standard SVMs approach for the considered regression problem. Further improvements have to be done both on the algorithms (different accelerations can be tried and their regularization properties must be deeply investigated) and the applications (different data pre-processing phases could be considered including a more specific voxel selection).

References

[1] Bauer F, Pereverzev S and Rosasco L 2007 On regularization algorithms in learning theory, *J. Complex.* **23** 52-72
[2] Belliveau J, Kennedy D, McKinstry R, Buchbinder B, Weisskoff R, Cohen M, Vevea J, Brady T and Rosen B 1991 Functional mapping of the human visual cortex by magnetic resonance imaging *Science* **254** 716-19
[3] Bertero M, Lanteri H and Zanni L 2008 Iterative image reconstruction: a point of view *Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy* (IMRT) Censor Y, Jiang M and Louis A K Editors Edizioni della Normale Pisa Italy (to appear)
[4] Bialy H 1959 Iterative behandlung linearer funktionalgleichungen *Arch. Rat. Mech. Anal.* **4** 166-76
[5] Dai Y H, Hager W W, Schittkowski K and Zhang H 2006 The cyclic Barzilai-Borwein method for unconstrained optimization *IMA J. Numer. Anal.* **26** 604-27
[6] De Vito E, Rosasco L, Caponnetto A, De Giovannini U and Odone F 2005 Learning from examples as an inverse problem *J. Mach. Learn. Res.* **6** 883-904
[7] De Vito E, Rosasco L and Verri A 2005 Spectral methods for regularization in learning theory *DISI Technical Report* DISI-TR-05-18 Università di Genova Italy
[8] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht:Kluwer Academic Publishers Group)
[9] Evgeniou T, Pontil M and Poggio T 2000 Regularization networks and support vector machines *Adv. Comput. Math.* **13** 1-50
[10] Figueiredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing* **1** 586-97
[11] Frassoldati G, Zanni L and Zanghirati G 2008 New adaptive stepsize selections in gradient methods *J. Indust. Manag. Optim.* (to appear - available at <http://www.optimization-online.org>)
[12] Haynes J D and Rees G 2006 Decoding mental states from brain activity in humans *Nature Rev. Neurosci.* **7** 523-34
[13] Joachims T 1999 Making large-Scale SVM learning practical in *Advances in Kernel Methods - Support Vector Learning* Schlkopf B, Burges C and Smola A eds. MIT-Press
[14] Lo Gerfo L, Rosasco L, Odone F, De Vito E and Verri A 2008 Spectral algorithms for supervised learning *Neural Comput.* (to appear)
[15] Prato M, Zanni L and Zanghirati G 2007 On recent machine learning algorithms for brain activity interpretation *23rd* Annual Review of Progress in Applied Computational Electromagnetics March 19-23 2007 Verona Italy pp 1939-46
[16] Tikhonov A N and Arsenin V Y 1977 *Solution of Ill Posed Problems* (Washington D.C.: Winston W H)
[17] Vapnik V N 1998 *Statistical Learning Theory* (New York: John Wiley and Sons)
[18] Wahba G 1990 Spline models for observational data *Series in Applied Mathematics* vol 59 (Philadelphia:SIAM)
[19] Wuang Y and Ma S 2007 Projected Barzilai-Borwein method for large-scale nonnegative image restoration *Inverse Probl. Sci. En.* **15** 559-83
[20] Wu Q, Ying Y and Zhou D-X 2006 Learning rates of least-square regularized regression *Found. Comput. Math.* **6** 171-92
[21] Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35** 69-86