School of Mathematical
and Physical Sciences

Department of Mathematics and Statistics

Preprint MPS-2012-07

01 March 2012

On Discrimination Algorithms for Ill-Posed Problems with an Application to Magnetic Tomography

by

Natalie Lowery, Roland Potthast, Maria Vahdati and William Holderbaum

On Discrimination Algorithms for Ill-Posed Problems with an Application to Magnetic Tomography

Natalie Lowery [1], **Roland Potthast**[1], **Maria Vahdati** and **William Holderbaum**

[1] ep en o M e c n e y o e d n en O Bo Be e AX

[2] ep en o Con c on M n e en n e y o e d n en O Bo Be e AX

[3] ep en o y e n nee n n e y o e d n en O Bo Be e AX

n.l.h.lowery@student.reading.ac.uk, r.w.e.potthast@reading.ac.uk, m.m.vahdati@reading.ac.uk and w.holderbaum@reading.ac.uk

A r c

n p pe e e po e c c on ec n e o po ed p o e . o c e e *ine r y sep r* e n o e e p ce X ey c n e ep ed y y e p ne e n e e *st e sep r i ity*, e e c e e e e e po e d nce e een o ep n ype p ne en e d e p ce Y ene ed y co p c ope o A pp ed o e y e e $\in X$, e o n ene e do no o n e ep y n Y e en e p e n X y ep n p c , e o o e c e e e non ne c c on ene ed o non con en y o ne c e n X

e pp y o e o e p o e o y con o o e ce e e e c y e ce cco d n o e e c ency e c n po en y c y e e n e e o e e e n e ed ne c e d o o e n e n c en o e e e c nno e e ec

of tumours [1]. In [4] the cancer area classification problem is investigated with voxels training patterns which are used for discrimination is high. We show that, in the case of data generated by a compact integral operator, it cannot stay stable when the number of measurement points tends to infinity.

r, we investigate the application of classification algorithms to classify study particular nonlinear classes which are obtained from linear classes as intersections of affine halfspaces U_ℓ, $\ell = 1, ..., n$, i.e.

$$C := \bigcap_{\ell=1}^{n} U_\ell. \tag{2.2}$$

In this case we can reduce the study of stability of the classification to the stability in the linear case. For smooth classes, i.e. where the boundary of C is a smooth manifold in X, we can locally approximate the general nonlinear classification by a linear classification, such that in this case the stability analysis can also be carried over from the linear to the nonlinear case.

The process of classification is given by the way to achieve a definition of a class C.

Definition 2.1 (Supervised Classification) ... samples $x_1^{(1)}, ..., x_1^{(N_1)} \in X$ ... $C_1 \subset X$ ... $x_1^{(\ )}$ $\omega = 1, ..., N_1$ ... C_1 ... $x^{(\ )}$ $\omega = 1, ..., N$ ... $C = X \setminus C_1$ ... target values ... supervised classification algorithm ...

when we study stable classifications in X and how they can be achieved on the images in Y without the inversion of the operator A.

A generic example. As a first step towards the clarification of the situation we first want to consider a special case. We use the singular value decomposition of the operator A, compare [6], i.e. we have an orthonormal basis $\{\varphi_\ell \in X,\ \ell \in \mathbb{N}\}$ in X and an orthonormal basis $\{g_\ell \in Y,\ \ell \in \mathbb{N}\}$ and a monotonously decreasing sequence $(\mu_\ell)_{\ell \in \mathbb{N}}$ of positive real values such that

$$A\varphi_\ell = \mu_\ell g_\ell, \quad A^* g_\ell = \mu_\ell \varphi_\ell \tag{2.3}$$

for all $\ell \in \mathbb{N}$. We define a linear stable classification by

$$C^{(1)}_{\ell,\rho} := \{x : \langle x, \varphi_\ell \rangle \geq \rho\}, \quad C^{(2)}_{\ell,\rho} := \{x : \langle x, \varphi_\ell \rangle \leq -\rho\}, \tag{2.4}$$

which we call a [illegible]. Clearly, the distance between $C^{(1)}_{\ell,\rho}$ and $C^{(2)}_{\ell,\rho}$ is $2\rho > 0$. For every pair $C^{(1)}_{\ell,\rho}, C^{(2)}_{\ell,\rho}$ from the sequence of classes the classification is stable uniformly with respect to $\ell \in \mathbb{N}$.

The images of the classes $C^{(1)}_{\ell,\rho}, C^{(2)}_{\ell,\rho}$ under the application of the operator A is given by

$$\begin{aligned}
\tilde{C}^{(1)}_{\ell,\rho} &:= \{y = Ax : \langle x, \varphi_\ell \rangle \geq \rho\}, \\
&= \{y \in A(X) : \langle A^{-1}y, \varphi_\ell \rangle \geq \rho\}, \\
&= \{y \in A(X) : \langle y, (A^*)^{-1}\varphi_\ell \rangle \geq \rho\}, \\
&= \left\{y \in A(X) : \left\langle y, \tfrac{1}{\mu_\ell} g_\ell \right\rangle \geq \rho \right\}, \\
&= \{y \in A(X) : \langle y, g_\ell \rangle \geq \mu_\ell \rho\},
\end{aligned} \tag{2.5}$$

and

$$\tilde{C}^{(2)}_{\ell,\rho} := \{y : \langle y, g_\ell \rangle \leq -\mu_\ell \rho\}. \tag{2.6}$$

The distance between the classes $\tilde{C}^{(1)}_{\ell,\rho}$ and $\tilde{C}^{(2)}_{\ell,\rho}$ is $2\mu_\ell\rho$. The distance is depending on $\ell \in \mathbb{N}$ and since the singular values μ_ℓ tend to zero for $\ell \to \infty$, the stability of the separation of the pairs of classes is no longer uniform in ℓ. We summarize these basic but important observations in the following lemma.

Lemma 2.2 [illegible] A [illegible] X [illegible] Y [illegible] ρ [illegible] Y.

The general case. Next, we consider the general case of a sequence of linear classes $C_1, C_2, C_3, \ldots$. Let v_ℓ, $\ell \in \mathbb{N}$ be the corresponding vectors in X and ρ_ℓ, $\ell \in \mathbb{N}$ be the affine distances. Here, we also assume that the sequence (v_ℓ) does not have a convergent subsequence. Clearly, the classes are given by

$$C_\ell = \{x : \langle x, v_\ell \rangle \geq \rho_\ell\}, \quad \ell \in \mathbb{N}. \tag{2.7}$$

We use a calculation similar to (2.5) to show that if v_ℓ is in the range of A , then

$$\tilde{C}_\ell := AC_\ell = \{y : \langle y, (A\ )^{-1} v_\ell \rangle \geq \rho_\ell\}. \tag{2.8}$$

With the definition

$C_\ell := AC_\ell =$

$$\psi_\ell := (A\ )^{-1} v$$

$$n \quad r \quad n \quad n\,A \quad r \quad r \quad r \tag{8}$$

and with the same argument we also have the slightly more general form of this statement

Theorem 2.5 n r n r **C** n **X** n y n r r **v** n n **ρ** r n n **A** ρ n r ρ r r **A** : **X** → **Y** **v** ∉ **A** **Y** n n $\tilde{\mathbf{C}} = \mathbf{AC}$ r n r

Since the norm of $\frac{R_\alpha}{||R_\alpha||}$ is bounded by one, there is a weakly convergent subsequence for $\alpha \to 0$, for which we denote the regularization parameters by α_ℓ, $\ell \in \mathbb{N}$. We call the limit element $\psi \in \mathbf{Y}$. We note that for $j \in \mathbb{N}$ the element

$$\mathbf{y}_j := \psi \cdot 2\frac{\rho}{||\psi||\,||\mathbf{R}_j \mathbf{v}||} \tag{2.35}$$

satisfies

$$\left\langle \mathbf{y}_j, \frac{\mathbf{R}_\ell \mathbf{v}}{||\mathbf{R}_\ell \mathbf{v}||} \right\rangle = \frac{2\rho}{||\psi||\,||\mathbf{R}_j \mathbf{v}||}$$

3. Static Magnetic Tomography

Here, we collect basic notation and results on static magnetic tomography, for more details we refer to [

Wannert et.al. In [

There are two basic options to approach the classification problem from magnetic

Example. As a two-dimensional example for our nonlinear class we consider the space $\mathbb{R}$ with two classes $\mathbf{C}_1, \mathbf{C}$ defined by

$$\mathbf{C}_1 := \{\mathbf{j} \in \mathbb{R} \;:\; \mathbf{j}_1 \geq \rho \text{ c.2]15]4e44=48(t)26}$$

for every $L \in \mathbb{N}$. With the area or volume $\left|\Omega_\ell^{(L)}\right|$, respectively, of some set $\Omega_\ell^{(L)}$ we define

$$\chi_\ell^{(L)}(x) := \begin{cases} \frac{1}{\left|\Omega_\ell^{(L)}\right|^{1/2}}, & x \in \Omega_\ell^{(L)}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.12}$$

the functions $\chi_\ell^{(L)}$ are in $L^2(\Omega)$ for $\ell = 1, ..., n_L$ and $L \in \mathbb{N}$. Then we have

$$\left\|\chi_\ell^{(L)}\right\|_{L^2(\Omega)} = 1 \tag{4.13}$$

We now collect all vectors $\chi_\ell^{(L)}$ for $\ell = 1, ..., n_L$ and $L = 1, 2, 3, ...$ into one sequence, for which we use the letter v_k, $k \in \mathbb{N}$.

For the fuel cell application nonlinear classes will naturally appear when the flow through the cell membrane is monitored. For example, the vectors $\chi_\ell^{(L)}$ can be chosen to be the special basis used for current reconstructions by Wannert and Potthast [22]. Here, the class $C_\ell^{(L)}$ is the set of all currents which have a component larger than $\rho_\ell^{(L)}$ along the direction $\chi_\ell^{(L)} \in L^2(\Omega)$. Good cells are those where we have a homogeneous distribution of the current, which means that all components are larger than some threshold ρ. This corresponds to the nonlinear class C defined in (4.6), which is composed of a sequence of linear classes.

We may choose a hierarchy of finer and finer discretizations to test the homogeneity

eel, whehs 245057(ss)]TΩR6511.95518(h)-339..9391TdΩ[(yh)-06(a)0.12969olaizatioedeto

5. On the Ill-Posedness of Fisher’s Linear Discriminant for Remote Data

So far we have studied the ill-posedness of classification problems which can be based on linear classification. We have shown that in general linear compact operators map stably separable problems into classifications which are no longer stably separable. However, we have not yet studied a particular algorithm for such classifications.

The task of this section is to investigate a well-known scheme for supervised classification 2.1 known as r L n r r n n. We will show that the method is also ill-posed in the sense that for an increased number of measurement points the norm of the inverse operators employed by the method become unbounded. As a particular application, we will apply the method to the problem of fuel cell classification and investigate the relationship between di erent ways to regularize the problem.

r L n r r n n n j n H

Fisher’s linear discriminant is not strictly speaking a discriminant but rather a method of r n n n y of the input space in such a way that we have maximum class separation in the new 0.147034(h)-0.312447(e)-313.123(i)0.21(a)0.245057(m7-0.147034(h)-n)-0.31

$n\rho \quad r \quad n \quad nA \quad r \quad r \quad \rho \quad \rho r$

We will complete this section by a rigorous proof showing that in this case the norm of the inverse from which the statement follows. □

Since $W : (C(\Omega)) \to (C(\Lambda))$ is compact and $W : (C(\Lambda)) \to (C(\Omega))$ is compact as well, the operator $S_F^{(\ )}$ is compact in $(C(\Lambda))$.

A discretization of this operator is achieved by using numerical quadrature for all three of its factors. With nodes z, $\kappa = 1, \ldots, M$ in Ω and quadrature weights s we discretize x via z and y via z. Then, the operator $S_F^{(\ )}$ is approximated via the matri79(m)-0.0898541 uniformly for all $\tilde{\mathbf{M}} \in \mathbb{N}$, such that (5.20) is satisfied. We denote an interpolation operator on $\mathbf{\Lambda}$ by $\mathbf{Q}^{(M)}$ and assume to have

$$||\mathbf{Q}^{(M)}\psi|| \quad \leq \mathbf{c}||\psi|| \tag{5.20}$$

with some constant c uniformly for $\tilde{\mathbf{M}} \in \mathbb{N}$ and a result analogous to (5.18).

The continuous form of the Biot-Savart operator W is given by (3.1). A discretization of W via standard quadrature or via finite integration technique leads to some matrix W

3.4]TJ/F7 11.9551 Tf 103118444.24 [(a)0.2450C26(e)38995 11.9551 Tf 8.437 390..33788

Lemma 5.6 L **W : X** $\to$ **X** n r r r r **X** n
L **W**$_N$ **: X** $\to$ **X** y r r r n r n **X**$_N \subset$ **X**
n n **N** **W**$_N$ n **W** n n
W $-$ **NtnAR**

The goal of this part is to work out the analysis to compare the two approaches to the classification problem, i.e.

- (i) : first reconstruct the current densities j and then carry out a classification on the reconstructed current densities,
- (ii) : Apply the classification directly to the magnetic field data.

We will first look at the unregularized problem and then study their relation when Tikhonov regularization is applied. We will find that the unregularized approaches are equivalent, but of course they are not practically applicable since the ill-posedness needs to be taken care of. We prove that the regularized versions cannot be equivalent.

Theorem 5.7 B

For the classification task we start with the samples $\mathbf{H}^{(\omega)}$. When we carry out the reconstruction by a numerical method, the corresponding currents are linked to these by

$$\mathbf{H}^{(\omega)} = \mathbf{B}\beta^{(\omega)} = \mathbf{WJ}\left(\beta^{(\omega)}\right), \tag{5.34}$$

where $\mathbf{W}$ is a discretized version of the Biot-Savart operator and $\mathbf{J}$ and $\mathbf{B}$ are discretized current and magnetic field matrices respectively. Then, the scatter matrix for the approach in the image space is

$$\begin{aligned}\mathbf{S}_F^{(\mathbf{H})} &= \sum_{\xi=1} \sum_{\mathbf{H}^{(\omega)} \in C_\xi} \left(\mathbf{H}^{(\omega)} - \mathbf{m}_\xi^{(\mathbf{H})}\right)\left(\mathbf{H}^{(\omega)} - \mathbf{m}_\xi^{(\mathbf{H})}\right)^T \\ &= \sum_{\xi=1} \sum_{\beta^{(\omega)} \in C_\xi} \left(\mathbf{WJ}\beta^{(\omega)} - \mathbf{WJm}_\xi^{(\beta)}\right)\left(\mathbf{WJ}\beta^{(\omega)} - \mathbf{WJm}_\xi^{(\beta)}\right)^T \end{aligned} \tag{5.35}$$

where $\mathbf{m}_\xi^{(\mathbf{H})}$ and $\mathbf{m}_\xi^{(\beta)}$ represent the means of the magnetic field and basis function coefficient classes respectively. Therefore

$$\begin{aligned}\mathbf{S}_F^{(\mathbf{H})} &= \mathbf{WJS}_F^{(\beta)}\mathbf{J}^T\mathbf{W}^T \\ &= \mathbf{BS}_F^{(\beta)}\mathbf{B}^T. \end{aligned} \tag{5.36}$$

If we substitute (5.36) into (5.7) we find that the classification vector $\mathbf{w}^{(\mathbf{H})}$ in the image space is given by

$$\begin{aligned}\mathbf{w}^{(\mathbf{H})} &\propto \left(\mathbf{S}_F^{(\mathbf{H})}\right)^{-1}\left(\mathbf{m}^{(\mathbf{H})} - \mathbf{m}_1^{(\mathbf{H})}\right) \\ &= \left(\mathbf{BS}_F^{(\beta)}\mathbf{B}^T\right)^{-1}\left(\mathbf{m}^{(\mathbf{H})} - \mathbf{m}_1^{(\mathbf{H})}\right), \end{aligned} \tag{5.37}$$

where $\mathbf{w}^{(\mathbf{H})}$ represents the weight vector found by applying Fisher's linear discriminant to $\mathbf{H}$. Then

$$\mathbf{w}^{(\mathbf{H})} \propto \left(\mathbf{B}^\top\right)^{-1}\left(\mathbf{S}_F^{(\beta)}\right)^{-1}\mathbf{B}^{-1}\left(\mathbf{WJm}^{(\beta)} - \mathbf{WJm}_1^{(\beta)}\right)$$
$$= \left(\mathbf{B}^\top\right)^{-1}\left(\mathbf{S}_F^{(\beta)}\right)^{-1}\left(\mathbf{m}^{(\beta)} - \mathbf{m}_1^{(\beta)}\right). \tag{5.38}$$

It is linked to the classification vector $\mathbf{w}^{(\beta)}$ in the input space by

$$\mathbf{w}^{(\mathbf{H})} \propto \left(\mathbf{B}^\top\right)^{-1}\mathbf{w}^{(\beta)}, \tag{5.39}$$

which is a discrete version of (2.9). Classification in the image space given some data $\mathbf{H}$ is carried out by calculating $(\mathbf{w}^{(\mathbf{H})})^\top \mathbf{H}$. In the state space it is given by

$$\left(\mathbf{w}^{(\beta)}\right)^\top \beta = \left(\mathbf{w}^{(\beta)}\right)^\top \mathbf{B}^{-1}\mathbf{H} = \left(\mathbf{w}^{(\mathbf{H})}\right)^\top \mathbf{H}, \tag{5.40}$$

which proves the theorem. □

Finally, we need to study the relation between the two regularized versions of Fisher's linear discriminant. The first version uses Tikhonov regularization directly applied to invert $\mathbf{S}_F^{(\mathbf{H})}$, i.e. we calculate

$$\mathbf{R}^{(\mathbf{H})} := \left(\alpha\mathbf{I} + \left(\mathbf{S}_F^{(\mathbf{H})}\right)^\top \mathbf{S}_F^{(\mathbf{H})}\right)^{-1}\left(\mathbf{S}_F^{(\mathbf{H})}\right)^\top, \quad \alpha > 0. \tag{5.41}$$

The regularized version of (5.37) is thus given by

$$\mathbf{w}^{(\mathbf{H})} := \mathbf{R}^{(\mathbf{H})}\left(\mathbf{m}^{(\mathbf{H})} - \mathbf{m}_1^{(\mathbf{H})}\right). \tag{5.42}$$

The second version applies the discrimination algorithm to the reconstructed coefficients, i.e. it uses

$$\beta^{(\omega)} := \left(\alpha\mathbf{I} + \mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top\mathbf{H}^{(\omega)}, \quad \alpha > 0. \tag{5.43}$$

We define $\mathbf{m}_\xi$ as the mean of the $\beta^{(\omega)}$ for C_ξ and

$$\mathbf{S}_\xi^{(\beta)} := \sum_{\beta^{(\omega)} \in C_\xi}\left(\beta^{(\omega)} - \mathbf{m}_\xi\right)\left(\beta^{(\omega)} - \mathbf{m}_\xi\right)^\top \tag{5.44}$$

for $\xi = 1, 2$ and $\mathbf{S}_F^{(\beta)} = \mathbf{S}_1^{(\beta)} + \mathbf{S}_2^{(\beta)}$ as usual. Then, we calculate

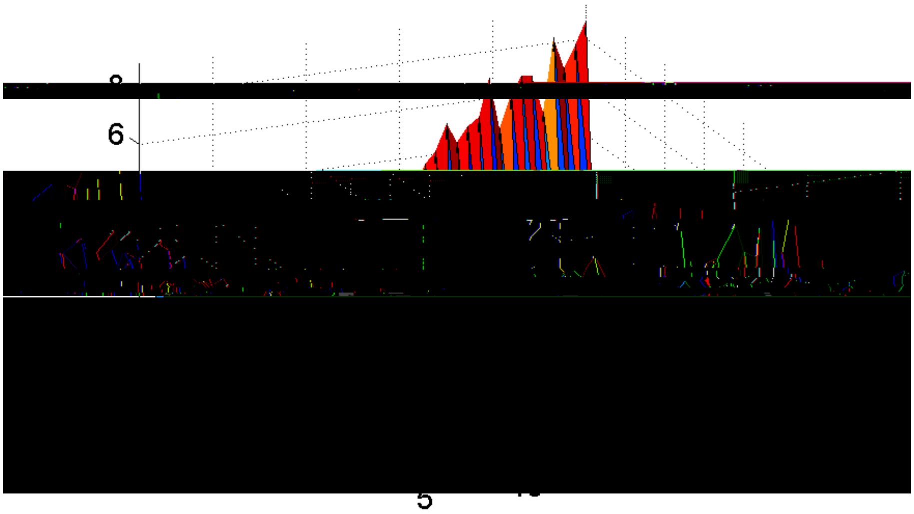
$$\mathbf{S}_F^{(\beta)} = \left(\alpha\mathbf{I} + \mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top\mathbf{S}_F^{(\mathbf{H})}\mathbf{B}\left(\alpha\mathbf{I} + \mathbf{B}^\top\mathbf{B}\right)^{-1}. \tag{5.45}$$

Now, the second version calculates a regularized verson of the discrimination vector $\mathbf{w}^{(\beta)}$ by

$$\mathbf{w}^{(\beta)} := \left(\mathbf{S}_F^{(\beta)}\right)^{-1}\left(\mathbf{m}^{(\beta)} - \mathbf{m}_1^{(\beta)}\right). \tag{5.46}$$

Lemma 5.8

F g re e ne d c n n pe o ed on ec o $\boldsymbol{\beta}^{(\omega)}$ o nc on



F g re