Signal Processing 90 (2010) 2546–2551

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

UPRE method for total variation parameter selection

Youzuo Lin [a,*], Brendt Wohlberg [b,1], Hongbin Guo [a]

[a] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287, USA
[b] T-5, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Article history:
Received 10 November 2009
Received in revised form
29 January 2010
Accepted 25 February 2010
Available online 3 March 2010

Keywords:
Parameter selection
Total variation regularization
Large scale problem
Inverse problem

ABSTRACT

Total variation (TV) regularization is a popular method for solving a wide variety of inverse problems in image processing. In order to optimize the reconstructed image, it is important to choose a good regularization parameter. The unbiased predictive risk estimator (UPRE) has been shown to give a good estimate of this parameter for Tikhonov regularization. In this paper we propose an extension of the UPRE method to the TV problem. Since direct computation of the extended UPRE is impractical in the case of inverse problems such as deblurring, due to the large scale of the associated linear problem, we also propose a method which provides a good approximation of this large scale problem, while significantly reducing computational requirements.

1. Introduction

Many image restoration tasks can be posed as linear inverse problems of the form

$$K\mathbf{x} = \mathbf{b} + \boldsymbol{\nu}, \tag{1}$$

where $\mathbf{b}$ represents the measured data, $\boldsymbol{\nu}$ represents noise, K is a linear transform (e.g. a convolution operator in the case of a deconvolution problem, and the identity in the case of denoising), and $\mathbf{x}$ represents the vectorized form of the recovered image. Regularization provides a method for controlling the noise and possible poor-conditioning of the operator K, prominent examples being the classical Tikhonov regularization [1],

$$\arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \|K\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\lambda}{2} \|H\mathbf{x}\|_2^2 \right\}, \tag{2}$$

where the matrix H is usually defined as a high-pass filtering operator, or identity matrix. The more recent TV regularization [2],

$$\arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \|K\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_{TV} \right\}, \tag{3}$$

where the TV norm $\|\mathbf{x}\|_{TV}$ is defined as $\|\sqrt{(D_x\mathbf{x})^2 + (D_y\mathbf{x})^2}\|_1$, with scalar operations applied to a vector considered to be applied element-wise, and the horizontal and vertical derivative operators written as D_x and D_y, respectively. These two methods differ in the regularization term; TV regularization is more difficult to compute, but usually provides superior results.

Effective application of these regularization methods depends critically on correct selection of the regularization parameter λ. While it is common practice for the user to simply try various values until the solution looks reasonable, the preferred approach is to estimate the λ value which optimizes some objective measure of image quality, such as the signal to noise ratio (SNR) of the reconstructed image with respect to the original undegraded image. There are several existing parameter selection methods for Tikhonov regularization [3,4]: (1) those requiring some knowledge of the noise $\boldsymbol{\nu}$, such as the *discrepancy principle* [5], and the *UPRE* [4], and (2) those that do not, such as *generalized cross-validation*

* Corresponding author.
E-mail addresses: youzuo.lin@asu.edu (Y. Lin), brendt@lanl.gov (B. Wohlberg), hguo1@asu.edu (H. Guo).
[1] This research was supported by the U.S. Department of Energy through the LANL/LDRD Program.

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.02.025

(GCV) [6,7] and the *L-curve method* [8]. Optimal parameter selection for TV regularization, in contrast, has received surprisingly little attention. To the best of our knowledge, there are very few papers discussing this issue under the TV framework [9–13].

We chose to extend the UPRE method to TV regularization, based on its good performance in the Tikhonov case [14], as well as the conceptual simplicity of the extension. Since the direct extension is only able to deal with relatively small-scale problems, we also discuss how to bypass this obstacle by using a Krylov subspace method. Experimental results are provided to demonstrate the efficacy of our approach.

2. Unbiased predictive risk estimator

The UPRE approach, also known as the C_L method, was first proposed [15] for regression problems, and then extended [4] to optimal parameter selection for Tikhonov problems. Define the *predictive error* $\mathbf{p}_\lambda = K\mathbf{x}_\lambda - K\mathbf{x}_{\text{true}}$, where $\mathbf{x}_\lambda \in \mathbb{R}^n$ is the computed solution for parameter λ, and $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$ is the ground truth solution. According to the UPRE method, the optimal parameter λ as the minimizer of the *predictive risk* $(1/n)\|\mathbf{p}_\lambda\|^2$, which is statistically estimated since $\mathbf{x}_{\text{true}}$ is, in general, unknown. The full derivation [4, Section 7.1], which is too lengthy to reproduce here, depends on the ability to express the regularized solution as having linear dependence on the data, $\mathbf{x}_\lambda = R_{\text{TK},\lambda}\mathbf{b}$, where the *regularization matrix* $R_{\text{TK},\lambda} = (K^TK+\lambda I)^{-1}K^T$. Defining the *regularized residual* $\mathbf{r}_\lambda = K\mathbf{x}_\lambda - \mathbf{b}$, and the *influence matrix* $A_{\text{TK},\lambda} = K(K^TK+\lambda I)^{-1}K^T$, the optimal parameter λ is the minimizer of

$$\text{UPRE}_{\text{TK}}(\lambda) = \frac{1}{n}\|\mathbf{r}_\lambda\|^2 + \frac{2\sigma^2}{n}\text{trace}(A_{\text{TK},\lambda}) - \sigma^2, \tag{4}$$

where σ^2 is the noise variance. The primary computational cost of evaluating the function UPRE_{TK} at λ consists of solving the Tikhonov problem at λ to obtain $\mathbf{x}_\lambda$, from which $\mathbf{r}_\lambda$ is obtained, and, more significantly the computation of $\text{trace}(A_{\text{TK},\lambda})$.

2.1. Extension of UPRE to total variation regularization

Extension of the UPRE to TV regularization is complicated by the absence of a linear equation $\mathbf{x}_\lambda = R_{\text{TV},\lambda}\mathbf{b}$ for the solution in terms of the data. Following part of the derivation of the lagged diffusivity algorithm [4, Section 8.2], we approximate the TV term $\|\mathbf{x}\|_{\text{TV}}$ by $\|\psi((D_x\mathbf{x})^2+(D_y\mathbf{x})^2)\|_1$, where $\psi(x)=\sqrt{x+\beta^2}$ provides differentiability at the origin. Correspondingly, the gradient of the TV term, $\nabla(\|\mathbf{x}\|_{\text{TV}})$, at x_λ can be written as $L(\mathbf{x}_\lambda)\mathbf{x}$, where

$$L(\mathbf{x}_\lambda) = D_x^T\text{diag}(\psi'(\mathbf{x}_\lambda))D_x + D_y^T\text{diag}(\psi'(\mathbf{x}_\lambda))D_y,$$

allowing one to define

$$R_{\text{TV},\lambda} = (K^TK+\lambda L(\mathbf{x}_\lambda))^{-1}K^T,$$

which is in the required form except for the dependence of matrix $L(\mathbf{x}_\lambda)$ on $\mathbf{x}_\lambda$.

Followed this idea, the influence matrix in the TV case can be written as

$$A_{\text{TV},\lambda} = K(K^TK+\lambda L(x_\lambda))^{-1}K^T. \tag{5}$$

The derivation (which is too lengthy to reproduce here, please refer to [4] for more details) of $\text{UPRE}_{\text{TK}}(\lambda)$ depends on the symmetry of $A_{\text{TK},\lambda}$ and the Trace Lemma [4]. Since $A_{\text{TV},\lambda}$ is also symmetric, the functional for $\text{UPRE}_{\text{TV}}(\lambda)$ can be derived in a similar way, and it can be shown that the UPRE for TV method shares the same form of expression as the Tikhonov method, with UPRE functional

$$\text{UPRE}_{\text{TV}}(\lambda) = \frac{1}{n}\|\mathbf{r}_\lambda\|^2 + \frac{2\sigma^2}{n}\text{trace}(A_{\text{TV},\lambda}) - \sigma^2. \tag{6}$$

2.2. Computational limitations

In the Tikhonov case the computation of $\text{trace}(A_{\text{TK},\lambda})$ in (4) is straightforward if the Singular Value Decomposition (SVD) of A is available, but in many large scale problems it is too expensive to compute the SVD of A. In [16] an approximation method is proposed to approximate the value of $\text{trace}(A_{\text{TK},\lambda})$ and related work can be found in [17–20].

The primary difficulty in implementing the UPRE in the TV case is the computation of $\text{trace}(A_{\text{TV},\lambda})$ in (6), since the linear approximation of the regularization term in the TV case further complicates the computation of UPRE in comparison with the Tikhonov case. Direct computation of the UPRE imposes very severe limits on the problem size due to computation time and memory requirements. In the following sections, we will introduce an algorithm which computes an approximation of the UPRE with vastly reduced computational cost, allowing application of this method to standard image sizes. In implementing this approximation, an enormous reduction in memory requirements is achieved by avoiding explicit construction of matrices such as A, D_x, D_x^T, D_y and D_y^T, the algorithm implementation requiring only matrix-vector products involving these matrices.

2.3. Extension of UPRE to large scale total variation

In the computation of (6), the most expensive part, as mentioned above, is the trace of the influence matrix, $\text{trace}\{K(K^TK+\lambda L(x_\lambda))^{-1}K^T\}$, since we need to deal with an inverse first then find the trace value. Applying the approach of Hutchinson [21], we can approximate trace $(f(M))$ by the unbiased trace estimator

$$E(\mathbf{u}^Tf(M)\mathbf{u}) \simeq \text{trace}(f(M)), \tag{7}$$

where $\mathbf{u}$ is a discrete multivariate random variable, which takes each entry the values -1 and $+1$ with probability 0.5, and the matrix M is symmetric positive definite (SPD).

Define the eigenvalue decomposition of M as $M = Q^T\Lambda Q$, where Q is an orthogonal matrix and Λ is a diagonal matrix of eigenvalues ρ_i in increasing order. Then, following [19,20], it can be shown that

$$\mathbf{u}^Tf(M)\mathbf{u} = \sum_{i=1}^{n} f(\rho_i)\tilde{u}_i^2 = \int_a^b f(\rho)\,d\mu(\rho), \tag{8}$$