Application of the UPRE Method to Optimal Parameter Selection for Large Scale Regularization Problems

Youzuo Lin and Brendt Wohlberg

***Abstract*— Regularization is an important method for solving a wide variety of inverse problems in image processing. In order to optimize the reconstructed image, it is important to choose the optimal regularization parameter. The Unbiased Predictive Risk Estimator (UPRE) has been shown to give a very good estimate of this parameter. Applying the traditional UPRE is impractical, however, in the case of inverse problems such as deblurring, due to the large scale of the associated linear problem. We propose an approach to reducing the large scale problem to a small problem, significantly reducing computational requirements while providing a good approximation to the original problem.**

***Index Terms*— Parameter Selection, Large Scale Problem, Inverse Problem, Tikhonov Regularization, Total Variation Regularization**

I. Introduction

Many image restoration and related image processing tasks can be posed as linear inverse problems of the form $A\mathbf{x} = \mathbf{b} + \boldsymbol{\nu}$, where $\mathbf{b}$ represents the measured data, $\boldsymbol{\nu}$ represents noise, A is a linear transform (e.g. a convolution operator in the case of a deconvolution problem), and $\mathbf{x}$ represents the restored image. Regularization provides a method for controlling the noise and possible ill-posedness of the operator A, prominent examples being the classical Tikhonov regularization [1],

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

where A is the known blurring operator and B is a highpass filter in the case of a deconvolution problem, and the more recent Total Variation (TV) regularization [2],

$$\arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1 \right\}, \tag{2}$$

which is more difficult computationally, but usually provides superior results.

Effective application of these regularization methods depends critically on correct selection of the regularization parameter λ. The optimal choice of regularization parameter maximizes 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: (1) those requiring some knowledge of the noise $\boldsymbol{\nu}$, such as the *Discrepancy Principle* [3], and the *Unbiased Predictive Risk Estimator* (UPRE), and (2) those that do not, such as *Generalized Cross-Validation(GCV)* [4], [5] and the *L-Curve Method* [6], [7]. Optimal parameter selection for TV regularization, in contrast, has received surprisingly little attention.

In this paper, we first extend the traditional UPRE approach to large scale Tikhonov regularization, and then adapt it to TV regularization.

II. Unbiased Predictive Risk Estimator

The UPRE approach, also known as the C_L method, was first recommended [8] for regression problems. Vogel [9] applied this idea to the selection of the optimal parameter λ for Tikhonov problems, as in (1). The UPRE is based on minimization of the predictive error, which is defined by

$$\frac{1}{n}||P_\lambda||^2 = \frac{1}{n}||Ax_\lambda - Ax_{\text{true}}||^2, \tag{3}$$

where x_λ is the computed solution for parameter λ, x_{true} is the true solution, and we define the optimal parameter λ_{opt} to be such that

$$\lambda_{\text{opt}} = \arg\min_{\lambda} \left\{ \frac{1}{n}||P_\lambda||^2 \right\}. \tag{4}$$

Since x_{true}, is, in practice, unknown, it is necessary to define an unbiased estimator to estimate the value of (4). Vogel [9] proved that, if we have

$$\text{UPRE}_{\text{Tikh}}(\lambda) = \frac{1}{n}||r_\lambda||^2 + \frac{2\sigma^2}{n}\text{trace}(A_\lambda) - \sigma^2, \tag{5}$$

where the regularized residual $r_\lambda = Ax_\lambda - b$, and the influence matrix $A_\lambda = A(A^TA + \lambda I)^{-1}A^T$, then $E(\text{UPRE}_{\text{Tikh}}(\lambda)) = E(\frac{1}{n}||P_\lambda||^2)$, which is the desired estimator. Under this assumption, the problem of (4) becomes

$$\lambda_{\text{opt}} = \arg\min_{\lambda} \{\text{UPRE}_{\text{Tikh}}(\lambda)\}. \tag{6}$$

A. UPRE and Its Limitations

While the computation of (5) is straightforward if the SVD decomposition of A is available, in many large scale problems, it is too expensive to compute the SVD of A. Kilmer [10] mentioned a method for using the Lanczos procedure to approximate the eigenvalues of the large scale system matrix by a small matrix. Girard [11] introduced an alternative approximation of the $\text{trace}(A_\alpha)$, based on randomization and

Youzuo Lin is with the Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287 USA. Email: youzuo.lin@asu.edu, Tel: (+1 480) 727 8537

Brendt Wohlberg is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: brendt@t7.lanl.gov, Tel: (+1 505) 667 6886, Fax: (+1 505) 665 5757

Monte-Carlo techniques. Our idea is the combination of the above two; we first use a low dimension matrix to approximate the large scale matrix of A_α by the Lanczos method, and then, in order to compute the trace of the approximated matrix, we apply the the randomization technique to perform the second approximation. Similar ideas can be found in [12], which focuses on the Generalized Cross-Validation approach, and which does not provide any specific information on how to choose the number of samples to be used for the Monte-Carlo test. In our work, inspired by [13], we are able to control the size of the Monte Carlo test in terms of the required accuracy of trace value.

B. Extension of UPRE to Large Scale Tikhonov

In the computation of (5), the most expensive part is the trace of the influence matrix, $\mathrm{trace}\{A(A^TA+\lambda I)^{-1}A^T\}$ because we need to deal with an inverse first then find the trace value. Applying the approach of Hutchinson [14], we can approximate $\mathrm{trace}(f(M))$ by the unbiased trace estimator

$$\mathrm{E}(u^T f(M)u) \simeq \mathrm{trace}(f(M)), \tag{7}$$

where u is a discrete random variable which takes the values -1 and +1 each 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 in increasing order. Then, following [12], [13], it can be shown that,

$$\begin{aligned} u^T f(M)u &= \sum_{i=1}^{n} f(\lambda_i)\tilde{u}_i^2 \\ &= \int_a^b f(\lambda_i)d\mu(\lambda), \end{aligned} \tag{8}$$

where $\tilde{u} = (\tilde{u_i}) = Qu$, and the integral is the Riemann-Stieltjes integral. In order to solve (8), Gauss quadrature is used to calculate this integral as follows,

$$\begin{aligned} \int_a^b f(\lambda_i)d\mu(\lambda) &= I_G[f] \\ &= \sum_{i=1}^{k} \omega_i f(\theta_i), \end{aligned} \tag{9}$$

where the weights ω_i and the nodes θ_i are unknown. If we start the Lanczos procedure for matrix M with the initial vector $x^{(0)} = u/\|u\|_2$, we will obtain a tridiagonal matrix T_k. It can be shown that θ_i in (9) are the eigenvalues of T_k, and ω_i are the squares of the first components of the normalized eigenvectors of T_k. Further details of the procedure outlined above can be found in [12], [13]. This approximation works best for sparse matrices.

For our problem, it can be proved that the blurring matrix is reasonably sparse, and more importantly, the matrix of $A^TA + \lambda I$ is SPD. We can adapt the above algorithm by setting $f(x) = 1/x$ and deriving the trace of A_λ as

$$\begin{aligned} \mathrm{trace}(A_\lambda) &= u^T A((A^TA+\lambda I)^{-1})A^T u, \\ &= v^T((A^TA+\lambda I)^{-1})v, \end{aligned} \tag{10}$$

where $v = A^T u$. Instead of using $x^{(0)} = u/\|u\|_2$ as the initial vector for the Lanczos procedure for computing ω_i and θ_i in equation (9), we now need to utilize the corresponding initial vector for this case, i.e., $x^{(0)} = v/\|v\|_2$.

If provided with a confidence probability $p = 1-\alpha$ and the required accuracy δ of the approximated trace value, it has been shown [13] that we can find out the samples size, N, of the Monte Carlo test which needs to be processed to achieve this accuracy of δ. Specifically, we have

$$N \geq \left(\frac{\lambda_\alpha}{\delta}\right)^2 \left(\frac{\sigma}{\mathrm{trace}(A_\alpha)}\right)^2, \tag{11}$$

where σ denotes the standard deviation of the trace value.

C. General Algorithm

For Monte Carlo test sample size N, the trace is approximated by

$$\mathrm{Trace} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AppTR}_i,$$

where the value of AppTR_i computed as follows:

Setup the tolerance, TOL;
Setup the initial vector, $x^{(0)} = v_i/\|v_i\|_2$ as in (10);
while ($\mathrm{RelErr} > \mathrm{TOL}$);
  Construct T by one step of Lanczos procedure on $(A^TA+\lambda I)^{-1}$;
  Compute the eigenvalues θ_i and the first elements ω_i of eigenvectors of T;
  Compute $I = \sum \omega_i f(\theta_i)$ by using Gauss quadrature as described in equation (9);
  Update the relative error RelErr from two successive approximations of I;
end
Evaluate $\mathrm{AppTR}_i = \|v_i\|_2^2 I$;

D. Extension of UPRE to Total Variation Regularization

In the Total Variation case, the difficulty is that the penalty functional in the TV term is non-quadratic. The idea is to approximate the non-quadratic equation by a quadratic equation, using the interpolation formula,

$$\begin{aligned} f(x+p) &\approx f + p^T\nabla f + \frac{1}{2}p^T\nabla^2 fp \\ &\approx \frac{1}{2}p^T\nabla^2 fp. \end{aligned}$$

We see that we can use Hessian matrix to approximate our TV term by a quadratic term. Another feature of this approximation is that, by defining the influence matrix

$$A_\lambda = A(A^TA + \lambda\mathrm{Hessian})^{-1}A^T, \tag{12}$$

we can prove that A_λ is symmetric, which means the UPRE for TV methods shares the same expression as the Tikhonov method.

The primary difficulty is again the computation of (12), but we are now faced with the additional problem of dealing with the Hessian matrix. For the results presented here, we have used the direct computation of the UPRE, which poses severe computation time and memory constraints on the size of the problem that can be addressed, but we are currently developing an approximate method corresponding to that for Tikhonov regularization described above.

III. Computational Results

In this section, we present three numerical examples, all of which are implemented in Matlab. The first two are Tikhonov problems, and the third one is a TV problem. The blurring kernel is chosen to be Gaussian, with additive Gaussian white noise. Since we are using synthetic data, we can find out the true optimal parameter λ by sweeping over a wide enough range of possible numbers, $[\lambda_{\min}, \lambda_{\max}]$. Similarly, we can also compute the true predictive error as given in (3). In order to show the effectiveness of our approximated algorithm, we compute the exact UPRE value at the same time by exactly constructing the explicit matrix A in (5). We evaluate the performance of our approximated UPRE approach by comparing to the above three values.

A. Test Problem 1 for Tikhonov: Satellite Image with Size 32 × 32

This is a very small test problem, so that we can compute both the direct method, by using the explicit matrix to compute the UPRE value in (6), and the approximated method proposed above. We compare our approximated UPRE with the True Predictive Risk (4), the Exact UPRE and the actual result using the optimal parameter. The test image is a 32 × 32 region cut from the original 256 × 256 "satellite" image frequently used for deconvolution tests. The test image is blurred using an 11×11 Gaussian kernel before adding zero-mean white noise with a variance of 50. The corresponding blurring matrix has size 1024 × 1024. Table (I) shows the results of estimated parameter λ, the estimated SNR value of the recovered image and the run time cost of the direct and approximated methods described above.

B. Test Problem 2 for Tikhonov: Satellite Image with Size 256 × 256

In this test problem, we use the whole satellite image of size 256 × 256, again adding Gaussian white noise after blurring with a Gaussian kernel. The blurring matrix under this setup has size 65536 × 65536, which is is too large to be able to compute the SVD, so the traditional UPRE is impractical due to memory limitations. However by using our approximated algorithm, we can still process this image and the numerical results in Table II show the good accuracy of our approximated algorithm. Fig. 1 plots the "True Predictive Error" against the "UPRE Value", from which we can see that the minima of those two curves are located close together. However it is still worth pointing out that, as an unbiased estimator, it does not require that UPRE has exactly the same minimum value as the true predictive error, but the location of their minimum should be close. In Fig. (2), the dot is the SNR value achieved by using the UPRE approach, which is close to the maximum of the SNR value.

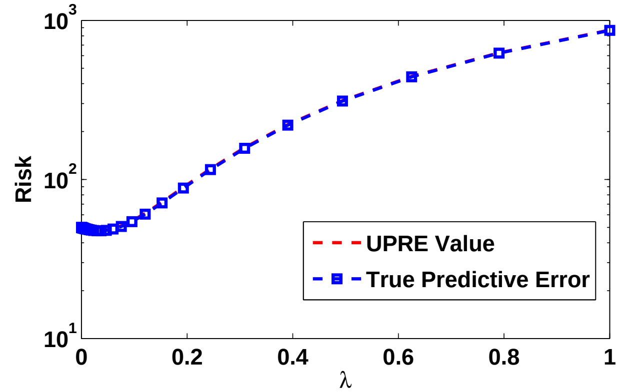

Fig. 1. Comparison of True Predictive Error and Approximate UPRE for the 256 × 256 test. Note that curves are so close that they are virtually indistinguishable.

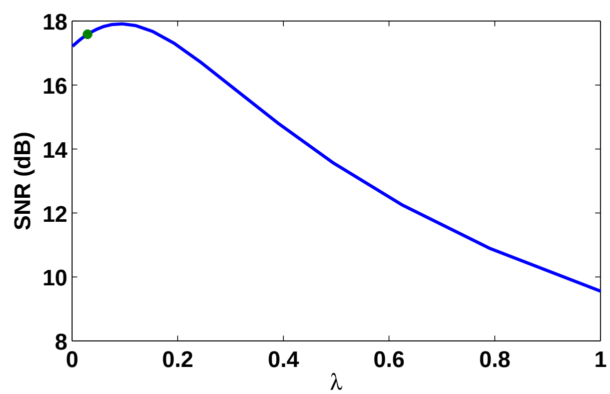

Fig. 2. Variation of reconstruction SNR (with respect to known ground truth image) against λ. Note the proximity of the maximum SNR region to the minimum of the Approximate UPRE in Fig. 1.

C. Test Problem 3 for Total Variation: Satellite Image with Size 64 × 64

In this test problem, we show the feasibility of (12) by implementing UPRE for a TV problem over a small image of size 64 × 64, with blurring and noise added as before. The minima of the "True Predictive Error" and "UPRE Value" curves in Fig. (3) are located close together, and the plot of SNR in Fig. (4) shows that the UPRE estimate gives a reconstructed SNR close to that of the optimum λ value.

IV. Conclusions

Our method for computing the UPRE for Tikhonov regularization, via the approximated trace, gives a good approximation to the UPRE computed using the exact trace value, while very significantly reducing computational requirements. The approximated value may be computed without explicit representation of very large matrices, and therefore also avoids memory limitations which prevent the application of the direct computation to problems involving images of practical size. We have also extended the UPRE method to TV regularization, and are currently developing an efficient algorithm for computing it for large problems, based on a similar approach to that applied for Tikhonov regularization.

	Optimal Value	True Pred. Error	Exact UPRE	Approx. UPRE
Estimated SNR(dB)	18.07	18.06	18.07	18.07
Estimated λ	0.0036	0.0014	0.0028	0.0028
Run Time Per Iteration (sec)	—-	—-	2.49	0.62

TABLE I

PERFORMANCE RESULTS FOR TEST PROBLEM 1. IMAGE SIZE: 32×32; BLURRING MATRIX SIZE: 1024×1024, CONSISTING OF 11×11 GAUSSIAN KERNEL WITH $\sigma = 0.41$; GAUSSIAN NOISE WITH 0 MEAN AND VARIANCE 50; SNR OF THE ORIGINAL NOISY IMAGE: 5.06 dB; ACCURACY OF THE TRACE APPROXIMATION IN (11): $\delta = 0.02$.

	Optimal Value	True Pred. Error	Exact UPRE	Approx. UPRE
Estimated SNR(dB)	17.91	17.67	Insufficient memory	17.60
Estimated λ	0.095	0.037	Insufficient memory	0.030
Run Time Per Iteration (sec)	—-	—-	Insufficient memory	1.81

TABLE II

PERFORMANCE RESULTS FOR TEST PROBLEM 2. IMAGE SIZE: 256×256; BLURRING MATRIX SIZE: 65536×65536, CONSISTING OF 21×21 GAUSSIAN KERNEL WITH $\sigma = 1$; GAUSSIAN NOISE WITH 0 MEAN, VARIANCE 50; SNR OF THE ORIGINAL NOISY IMAGE: 4.48 dB; ACCURACY OF THE TRACE APPROXIMATION IN (11): $\delta = 0.02$.

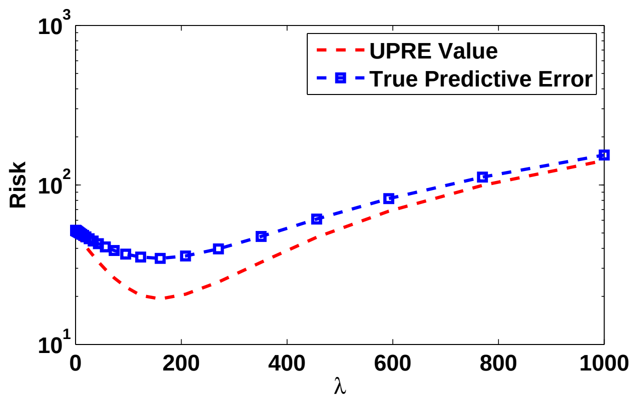

Fig. 3. Comparison of True Predictive Error and Approximate UPRE for the 64×64 test in Total Variation.

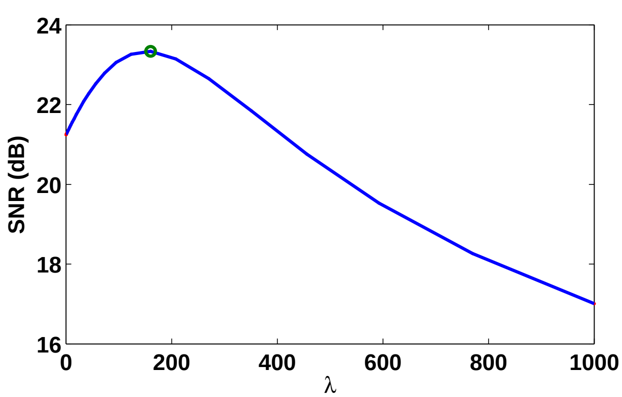

Fig. 4. Variation of reconstruction SNR (with respect to known ground truth image) against λ. Note the proximity of the maximum SNR region to the minimum of the Approximate UPRE in Fig. 3.

V. ACKNOWLEDGMENT

We thank Prof. Guo of Arizona State University for his suggestions on how to use the Lanczos procedure to solve for the trace problem.

REFERENCES

[1] A. N. Tikhonov, "Regularization of incorrectly posed problems," *Soviet Math. Dokl.*, vol. 4, pp. 1624–1627, 1963.
[2] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, 60*, pp. 259–268, 1992.
[3] V. A. Morozov, "Methods for solving incorrectly posed problems," *Springer-Verlag, New York*, 1984.
[4] G. H. Golub, M. T. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.
[5] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, vol. 14, pp. 651–667, 1977.
[6] P. C. Hansen, "Rank-deficient and discrete ill-posed problems," *SIAM*, 1997.
[7] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, pp. 1487–1503, 1993.
[8] C. L. Mallows, "Some comments on C_p," *Technometrics*, vol. 15, no. 4, pp. 661–675, November 1973.
[9] C. R. Vogel, *Computational Methods for Inverse Problems*. SIAM, 2002.
[10] M. E. Kilmer and D. P. O'Leary, "Choosing regularization parametres in iterative methods for ill-posed problems," *Journal on Matrix Analysis and Applications*, vol. 22, no. 4, pp. 1204–1221, 2001.
[11] D. A. Girard, "The fast Monte-Carlo cross-validation and C_L procedures: Comments, new results, and application to image recovery problems," *Computational Statistics*, vol. 10, pp. 205–231, 1995.
[12] G. H. Golub and U. von Matt, "Generalized cross-validation for large-scale problems," *Journal of Computational and Graphical Statistics*, vol. 6, pp. 1–34, 1997.
[13] H. Guo, "Computing trace of function of matrix," *Numerical Mathematics*, vol. 2, pp. 204–215, 2000.
[14] M. F. Hutchinson, "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines," *Communications in Statistics, Simulation, and Computation*, vol. 19, pp. 433–450, 1990.