Self-Regularization Method for Image Restoration

Jae-Hung Yoo

ABSTRACT

This paper suggests a new method of finding regularization parameter for image restoration problems. Wiener filter requires priori information such that power spectrums of original image and noise. Constrained least squares restoration also requires knowledge of the noise level. If the prior information is not available, separate optimization functions for Tikhonov regularization parameter are suggested in the literature such as generalized cross validation and L-curve criterion. In this paper, self-regularization method that connects bias term of augmented linear system and smoothing term of Tikhonov regularization is introduced in the frequency domain and applied to the image restoration problems. Experimental results show the effectiveness of the proposed method.

키워드

자기 정규화, 선형 최소 평균 자승 규칙, 티코노프 정규화, 영상 복원
\[ G(u,v) = H(u,v) F(u,v) + N(u,v) \]  \hspace{1cm} (2)

In the inverse filtering, we have
\[ \hat{F}(u,v) = \frac{G(u,v)}{H(u,v)} = F(u,v) + \frac{N(u,v)}{H(u,v)} \]  \hspace{1cm} (3)

This formula shows that if \( H(u,v) \) is zero or very small in the high frequency region and \( N(u,v) \) is still not vanished in the corresponding region, the second term \( \frac{N(u,v)}{H(u,v)} \) is amplified. Thus we need a remedy solving ill-posed inverse problem.

In the Wiener filter, we have
\[ \hat{F}(u,v) = \frac{H^*(u,v) G(u,v)}{|H(u,v)|^2 + S_f(u,v)/S_n(u,v)} \]  \hspace{1cm} (4)

where \( S_f \) and \( S_n \) denote power spectrums of \( f \) and \( \eta \) respectively\[1\]. The vertical bar denotes the magnitude of complex number or absolute value of real number.

In the constrained least square method or parametric Wiener filter, we have
\[ \hat{F}(u,v) = \frac{H^*(u,v) G(u,v)}{|H(u,v)|^2 + \lambda |P(u,v)|^2} \]  \hspace{1cm} (5)

where \( \lambda \) is the regularization parameter to be determined and \( P \) denotes the Fourier transform of smoothing functional usually given by the Laplacian operator\[1\],[3].

Wiener filter requires priori information such as power spectrums of original image and noise. If the noise level is known a priori, Morozov’s Discrepancy Principle\( (: \text{MDP}) \) can be applied for the determination of regularization parameter in the constrained least squares image restoration. Otherwise separate optimization functions such as Generalized Cross Validation\( (: \text{GCV}) \) function and \( L\)-curve criterion are suggested as alternative methods in the literature\[4-8\].

In this paper, we apply the self-regularization (SR) method\[9-10\] to the image restoration problems. In section II, regularization parameter selection methods in the frequency domain are introduced. In section III and IV, SR method is introduced both in the problem domain and in the frequency domain. Both augmented linear system embedding differential smoother and linear Least Mean Square estimation\( (: \text{LMS}) \) rule\[11-17\] in the frequency domain are suggested. The relationship between regularization term and bias term is established in the frequency domain. In section V, experimental results show the effectiveness of the proposed method followed by the conclusion and reference sections.

II. Regularization Parameter Selection in the Frequency Domain

Regularized estimation vector is defined as
\[ \hat{f}_\lambda = (H^T H + \lambda C^T C)^{-1} H^T g \]  \hspace{1cm} (6)

Here, \( H \) and \( C \) are the block circulant matrices of a PSF \( h \) and a smoothing operator \( p \) respectively. \( \lambda \) is the Tikhonov regularization parameter. \( g \) and \( \hat{f}_\lambda \) are the column vectors stacking columns of a degraded image \( g \) and a recovered image \( \hat{f}_\lambda \) respectively. Both \( H \) and \( C \) matrices have the dimension, \( MN \) by \( MN \). That is, \( m = n = MN \) with image dimension \( M \) by \( N \).

Residual vector is defined as
\[ \xi_\lambda = g - H \hat{f}_\lambda \]  \hspace{1cm} (7)

In the MDP method, the regularization parameter is adjusted by the constraint that a given noise level is equal to a norm of residual, \( \| \eta \| = \| \xi \| \). In the frequency domain, square residual is defined as
\[ \parallel \xi_{\lambda} \parallel^2 = \sum_{i=1}^{n} \frac{\lambda^2 |s_i|^4}{(|s_i|^2 + \lambda |t_i|^2)^2} |G_i|^2 \]  

and square noise level is defined as

\[ \parallel \eta \parallel^2 = \sum_{i=1}^{n} |N_i|^2 \].

Here, \( s_i \) and \( t_i \) are the diagonal elements of Fourier transform of the matrices \( H \) and \( C \) respectively. \( G_i \) and \( N_i \) are the elements of column vector stacking columns of 2D Fourier transform of the degraded image \( g \) and noise \( \eta \).

In the GCV method, we minimize following object

\[ gcv(\lambda) = m \frac{\parallel (I - HH^*)g \parallel^2}{\text{trace}((I - HH^*)^2)} \].

Here \( H^* \) is a pseudo-inverse defined as

\[ H^* = (H^T H + \lambda C^T C)^{-1} H^T \].

In the frequency domain, we have

\[ gcv(\lambda) = m \frac{\sum_{i=1}^{n} \frac{|s_i|^4}{(|s_i|^2 + \lambda |t_i|^2)^2} |G_i|^2}{\left( \sum_{i=1}^{n} \frac{|s_i|^2 |t_i|^2}{(|s_i|^2 + \lambda |t_i|^2)^2} \right)^2} \].

We note that equivalent expression has been reported using the complicated Kronecker products of the matrices \( H \) and \( C \) on the generalized singular value decomposition configuration contrast to using the simple spectral transform[8].

In the L-curve criterion, we find the corner of L-curve having maximum curvature. L-curve is the full log scale plot of residual versus smoothing term. Smoothing term is defined as

\[ \parallel \hat{f} \parallel^2 = \sum_{i=1}^{n} \frac{|s_i|^2 |t_i|^2}{(|s_i|^2 + \lambda |t_i|^2)^2} |G_i|^2 \].

Square curvature is defined in a full log scale as

\[ Lc(\lambda) = \frac{u' v'' v - v'^2}{v^2} - \frac{v' u'' u - u'^2}{u^2} \left( \frac{(u' u'' - v' v'' - 2 v'^2)}{v^2} \right)^{1.5} . \]

Here, we identify \( u \) with \( \parallel \hat{f} \parallel^2 \) and \( v \) with \( \parallel \xi_{\lambda} \parallel^2 \). Prime and double prime denote first and second order derivatives about \( \lambda \) respectively.

### III. Self-Regularization Method

We set up the cost equation for \( \hat{f} \) with regularization term

\[ J(\hat{f}) = \frac{1}{2} \parallel g - H \hat{f} \parallel^2 + \lambda \parallel C \hat{f} \parallel^2 \].

Then gradient equation \( \partial J(\hat{f})/\partial \hat{f} = 0 \) leads

\[ H^T (g - H \hat{f}) = \lambda C^T \hat{f} \].

In the linear system[11-12], we can use the augmented system for better performance as following

\[ J(\hat{f}_c) = \frac{1}{2} \parallel g - H_c \hat{f}_c \parallel^2 \]

with

\[ \hat{f}_c = [f_0; \hat{f}] \]

and

\[ H_c = [1, H] \]

where \( 1 \) denotes an appropriate column vector ones function of Matlab or Octave, \( \text{ones}(m,1) \). Then gradient equation leads

\[ H_c^T (g - H_c \hat{f}_c) = 0 \].
Thus we can express the bias or threshold term as the mean of error vector elements

$$f_0 = \frac{1}{m} \mathbf{1}^T (\mathbf{g} - \mathbf{Hf}) = \bar{\xi}$$  (21)

and the regularization parameter as a function of $$\bar{f}_e$$

$$\lambda = \frac{\| \mathbf{1}^T \mathbf{H}^T \mathbf{f} \|_{\text{op}}}{\| \mathbf{1}^T \mathbf{C}^T \mathbf{C} \bar{f}_e \|}.$$  (22)

It is well known that the gradient descent algorithm always get a solution satisfying the projection theorem regardless of whether or not $$\mathbf{H}^T \mathbf{H}$$ is singular [11]. We notice that this is the self-regularization or auto-regularization property of the minimum squared error descent procedure contrast to direct method such as pseudo-inverse. We note that early termination of iterative process has an regularizing effect with a proper stopping rule [2]. The augmented system has a smaller residual for over determined case or rank deficient case. Naturally we propose the SR learning scheme as in the Table 1.

### Table 1. SR algorithm

1. Given training data $$\mathbf{C}, \mathbf{H}, \mathbf{g}$$ construct augmented data $$\mathbf{H}_e, \mathbf{g}_e$$.
2. Find augmented solution $$\hat{\mathbf{f}}_e$$ using the Widrow_Hoff’s LMS rule.
3. Find $$\lambda$$ using Eq. (22).
4. Find $$\mathbf{f}_\lambda$$ using Eq. (6).

### IV. A New Image Restoration Method using Self-Regularization

Block circulant matrices $$\mathbf{C}$$ and $$\mathbf{H}$$ are diagonalized by the Fourier transform. We denote them $$\mathbf{T}$$ and $$\mathbf{S}$$ respectively. Let $$\mathbf{z}$$ and $$\mathbf{b}$$ be the column vector stacking columns of the 2-D Fourier transform of $$f$$ and $$g$$ respectively. We set up the cost equation for $$\hat{\mathbf{z}}$$ with regularization term

$$J(\hat{\mathbf{z}}) = \frac{1}{2} (\| \mathbf{b} - \mathbf{S} \hat{\mathbf{z}} \|^2 + \lambda \| \mathbf{T} \hat{\mathbf{z}} \|^2).$$  (23)

Then gradient equation $$\partial J(\hat{\mathbf{z}})/\partial \hat{\mathbf{z}} = 0$$ leads

$$\mathbf{S}^H(\mathbf{b} - \mathbf{S} \hat{\mathbf{z}}) = \lambda \mathbf{T}^H \mathbf{T} \hat{\mathbf{z}}$$  (24)

where superscript $$\mathbf{H}$$ denotes the Hermitian or conjugate transpose.

The augmented systems embedding differential smoother in the frequency domain can be represented as following

$$J(\hat{\mathbf{z}}) = \frac{1}{2} \| \mathbf{b} - \mathbf{S} \hat{\mathbf{z}} \|^2$$  (25)

with augmented column vector including bias

$$\hat{\mathbf{z}}_e = [\mathbf{z}_0; \hat{\mathbf{z}}]$$  (26)

and augmented matrix with smoothing vector

$$\mathbf{S}_e = [\mathbf{t}; \mathbf{S}]$$  (27)

where $$\mathbf{t}$$ is a column vector using diag($$\mathbf{T}$$). Then gradient equation leads

$$\mathbf{S}^H_e(\mathbf{b} - \mathbf{S}_e \hat{\mathbf{z}}_e) = 0$$  (28)

Next we can express the bias or threshold term

$$z_0 = \frac{1}{\mathbf{t}^H \mathbf{t}} \mathbf{t}^H (\mathbf{b} - \mathbf{S} \hat{\mathbf{z}})$$  (29)

and the regularization parameter

$$\lambda = \frac{\| \mathbf{1}^T \mathbf{S}^H \mathbf{t} \|_{\text{op}} \| z_0 \|}{\| \mathbf{1}^T \mathbf{T}^H \mathbf{T} \hat{\mathbf{z}} \|_{\text{op}}} = \frac{\sum_{k=1}^m \mathbf{S}^\ast_k \mathbf{t}_k \| z_0 \|}{\| \sum_{k=1}^m \mathbf{t}_k^\ast \mathbf{z}_k \|}.$$  (30)
Next we describe the update rule for the gradient decent procedure

$$\Delta \hat{z}_e = \eta S_e^H (b - S_e \hat{z}_e) .$$

(31)

Using complex LMS rule of Widrow et. al.[13-14], we have

$$\Delta \hat{z}_e = \eta \frac{(b_k - \hat{z}_e^T s_e^k)}{\| s_e^k \|^2} (s_e^k)^* .$$

(32)

with column vector $s_e^k$ stacking two components as

$$s_e^k = [s_k^i, s_k^o] .$$

(33)

Then we have complex LMS update rule as

$$\Delta \hat{z}_0 = \eta \sum_{i=1}^{m} \frac{(b_i - \hat{z}_0^T t_i - \hat{z}_0^T s_i)}{(|t_i|^2 + |s_i|^2)} t_i^* .$$

(34)

$$\Delta \hat{z}_e = \eta \frac{(b_k - \hat{z}_e^T s_k - \hat{z}_e^T s_k^o)}{(|s_k|^2 + |s_k^o|^2)} s_k^o .$$

(35)

Stopping criteria can be the error level if available as priori information or user specified quantities such as the conventional one - relative change of weight vector

$$\| \Delta \hat{z}_e \|_2 / \| \hat{z}_e \|_2 \leq \epsilon$$

(36)

After learning $\hat{z}_e$ by using LMS rule, we find the parameter $\lambda$ using Eq. (30). We obtain $\hat{z}_\lambda$ using

$$\hat{z}_\lambda = (S^H S + \lambda T^H T)^{-1} S^H b$$

(37)

Reshaping the column vector $\hat{z}_\lambda$ into 2D matrix to obtain $\hat{F}(u,v)$, we finally get the estimation of original image by taking inverse 2D FFT.

We propose the SR learning scheme in the frequency domain in the Table 1.

| Table 2. SR algorithm in the frequency domain |
|---------------------------------------------|
| 1. Given training data $T, S, b$          |
| construct augmented data $S_e, b$.        |
| 2. Find augmented solution $\hat{z}_e$ using |
| Widrow_Hoff’s LMS rule.                   |
| 3. Find $\lambda$ using Eq. (30).         |
| 4. Find $\hat{z}_\lambda$ using Eq. (37). |

V. Experimental Results

We report the experiments with the new SR method proposed in the previous two sections. Figure 1 shows the satellite image data from the USAF Phillips Laboratory, Laser and Imaging Directorate, Kirtland AFB, NM[18].

![Fig. 1 Satellite image data.](image)

We compare the SR method with the conventional techniques MDP, GCV, L-curve and Wiener filter. Results are depicted in the figure 2 and 3. First row shows the restored image data having negative pixel components and the second row further processed by using projection with non-negativity constraint. Visual inspection shows that MDP and L-curve results in an over smoothed estimation with Laplacian smoother and GCV depicts under smoothing with identity matrix. SR1 and SR2 denote the results of step 2 of Table 2.
with learning coefficient $\eta$ setting to 1.0 but with different error limits. Termination of iteration takes 10 and 79 epochs with Laplacian and 169 and 1389 epochs with identity smoother respectively. SRH1 and SRH2 denote the results of step 4 of Table 2.

![Image of restored images](a)

![Image of restored images](b)

Fig. 2 (a) Restored image data with laplacian smoother with wiener filter as benchmark. (b) Restored image data with laplacian smoother.

In summary, an old SR method SR2 applying only the LMS rule takes a long time to get the wanted results and a new SR method SRH1 that combines early stopping of LMS rule and Tikhonov regularization takes a relatively short time with comparable results.

![Image of restored images](a)

![Image of restored images](b)

Fig. 3 (a) Restored image data with identity smoother. (b) Restored image data with identity smoother.

Image restoration performance is measured by the figure-of-merit functions such as relative error (RE), signal to noise ratio (SNR), peak SNR (PSNR) and improvement of SNR (ISNR), defined as following

$$RE = \frac{\| f - \hat{f} \|_2}{\| f \|_2},$$

$$SNR = 10 \log_{10} \frac{\| f \|_2^2}{\| f - \hat{f} \|_2^2},$$

$$PSNR = 10 \log_{10} \frac{m \| \max f \|_2^2}{\| f - \hat{f} \|_2^2},$$

and

$$ISNR = 10 \log_{10} \frac{\| f - g \|_2^2}{\| f - \hat{f} \|_2^2}.$$  

Table 3 and 4 show the image restoration performance with remarking value of regularization.
parameter $\lambda$ or error limit $\epsilon$. Results of Wiener filter is included for comparison benchmark. Here, we report only the restored image data projected with non-negativity constraint.

Table 3. Performance results with laplacian smoother with wiener filter as benchmark.

| Measure Method | RE ↓ | SNR ↑ | PSNR | ISNR | Remark $\lambda$ | $\epsilon$ |
|----------------|------|-------|------|------|------------------|----------|
| MDP            | 0.3803 | 8.307 | 22.03 | 5.356 | 8.29e−2         |          |
| GCV            | 0.3449 | 9.247 | 22.87 | 6.206 | 4.45e−3         |          |
| L-curve        | 0.4509 | 6.917 | 20.35 | 3.876 | 1.00e−0         |          |
| SR1            | 0.3768 | 8.479 | 22.11 | 5.438 | 1e−2            |          |
| SRH1           | 0.3425 | 9.252 | 22.59 | 6.241 | 3.65e−3         |          |
| SR2            | 0.3512 | 9.088 | 22.72 | 6.047 | 1e−3            |          |
| SRH2           | 0.3720 | 8.859 | 22.22 | 5.549 | 3.33e−4         |          |
| Wiener         | 0.3243 | 9.781 | 23.41 | 6.740 | N/A              |          |

Table 4. Performance results with identity smoothing operator.

| Measure Method | RE ↓ | SNR ↑ | PSNR | ISNR | Remark $\lambda$ | $\epsilon$ |
|----------------|------|-------|------|------|------------------|----------|
| MDP            | 0.3648 | 8.758 | 22.39 | 5.717 | 4.75e−4         |          |
| GCV            | 0.3755 | 8.935 | 22.56 | 5.895 | 9.80e−5         |          |
| L-curve        | 0.3623 | 8.818 | 22.45 | 5.777 | 4.38e−4         |          |
| SR1            | 0.4003 | 6.740 | 20.37 | 3.699 | $\epsilon=1e−3$ |          |
| SRH1           | 0.4235 | 7.463 | 21.19 | 4.423 | 1.59e−3         |          |
| SR2            | 0.3486 | 9.154 | 22.78 | 6.113 | $\epsilon=1e−4$ |          |
| SRH2           | 0.3504 | 9.108 | 22.74 | 6.067 | 2.30e−4         |          |

Table 3 shows that a new SR method(SRH1) gives good results. Table 4 shows that an old SR method(SR2) has comparable results.

V. Conclusions

In this paper, SR method in the frequency domain is suggested and applied to the image restoration problems. SR method combines gradient descent and Tikhonov regularization. The relationship between regularization term and bias term is established in the frequency domain. Both augmented linear system embedding differential smoother and LMS rule in the frequency domain are established. Experimental results show the comparable performance and robustness of the new SR method. Further research can be done on the method that performs the repeated calculation bias term and regulation parameter after the first few epochs of LMS iteration.

References

[1] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Reading, MA: Addison-Wesley, 1992.
[2] H. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems. Dordrecht: Kluwer Academic Publishers, 1996.
[3] S. Kim, “An image denoising algorithm for the mobile phone cameras,” J. of the Korea Institute of Electronic Communication Sciences, vol. 9, no. 5, 2014, pp. 601-608.
[4] F. Bauer and M. Lukas, “Comparing parameter choice methods for regularization of ill-posed problems,” Mathematics and Computers in Simulation vol. 81, no. 9, 2011, pp. 1795-1841.
[5] G. Golub, M. Heath, and G. Wahba, “Generalized cross-validation as a method for choosing a good ridge parameter,” Technometrics, vol. 21, no. 2, 1979, pp. 215-223.
[6] P. Hansen and D. O’Leary, “The use of the L-curve in the regularization of discrete ill-posed problems,” Society for Industrial and Applied Mathematics J. on Scientific Computing, vol. 14, no. 6, 1993, pp. 1487-1503.
[7] V. Morozov, Methods for Solving Incorrectly Posed Problems. New York: Springer-Verlag, 1984.
[8] A. Bouhamidi and K. Jbilou, “Sylvester Tikhonov-regularization methods in image restoration,” J. of Computational and Applied Mathematics, vol. 206, no. 1, 2007, pp. 86 - 98.
[9] J. Jeong, “A new learning methodology for support vector machine and regularization
RBF neural network,” Master’s Thesis, Chonnam National University, Feb. 2002.

[10] J. Yoo, “Automation of model selection through neural networks learning,” Proc. Korea Fuzzy Logic and Intelligent Systems, Fall Conf. vol. 14, no. 2, Changwon, Korea, Oct., 2004, pp. 314-316.

[11] R. Duda and P. Hart, Pattern Classification and Scene Analysis. New York: John Wiley & Sons, 1973.

[12] A. Papoulis, Probability, Random Variables, and Stochastic Processes. 3rd edition, New York: McGraw-Hill, 1991.

[13] S. Haykin, Adaptive Filter Theory. 3rd edition, Upper Saddle River, NJ: Prentice Hall, 1996.

[14] B. Widrow and S. Stearns, Adaptive Signal Processing. Upper Saddle River, NJ: Prentice-Hall, 1985.

[15] O. Kwon, “The bi-directional least mean square algorithm and its application to echo cancellation,” J. of the Korea Institute of Electronic Communication Sciences, vol. 9, no. 12, 2014, pp. 1337-1344.

[16] Y. Han, “Interference cancellation system in repeater using adaptive algorithm with step sizes,” J. of the Korea Institute of Electronic Communication Sciences, vol. 9, no. 5, 2014, pp. 549-554.

[17] S. Hwang, “Channel estimation based on LMS algorithm for MIMO-OFDM system,” J. of the Korea Institute of Electronic Communication Sciences, vol. 7, no. 6, 2012, pp. 1455-1461.

[18] J. Nagy, K. Palmer, and L. Perrone, “Iterative methods for image deblurring: a Matlab object oriented approach,” Numerical Algorithms, vol. 36, no. 1, 2004, pp. 73-93.

저자 소개

류재흥(Jae-Hung Yoo)

1981년 한양대학교 기계공학과 졸업(공학사) (BE in Mechanical Engineering from Hanyang Univ. in 1981)
1986년 디트로이트 대학교 대학원 전산학과 졸업 (MA in Computer Science from Univ. of Detroit in 1986)
1993년 웨인주립 대학교 대학원 전산학과 졸업 (PhD in Computer Science from Wayne State Univ. in 1993)
1994년 여수대학교 컴퓨터공학과 교수 (Joined as a faculty member in the Dept. of Computer Engineering, Yosu Nat. Univ. in 1994)
2006년∼현재 : 전남대학교 컴퓨터공학과 교수 (Became a faculty member in the Dept. of Computer Engineering, Chonnam Nat. Univ. in 2006)

※ 관심분야 : 인공신경망, 패턴인식, 기계학습, 영상처리 및 컴퓨터비전 (Main research areas: Artificial Neural Networks, Pattern Recognition, Machine Learning, Image Processing and Computer Vision)