DENOISE METHODS IN SIGNAL PROCESS

GUO WEI; JIALE WONG; LUO TAO; WENCHAO ZHANG

Abstract. Our group utilize several denoise methods and achieve the goal from different perspectives. This article is aimed to demostrate and compare these methods.

Contents

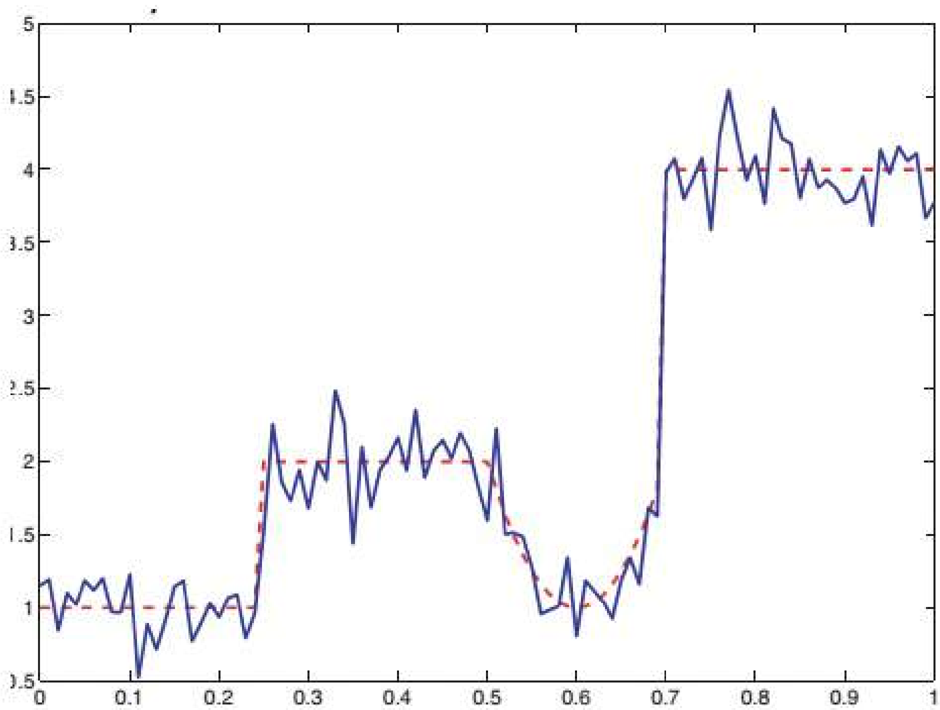
Part 1. Backgroud

1. mathematical expression

this problem is concerned with recovering a original signal $u(t)$ on time interval $t \in [0,1]$. While we measured $n+1$ discrete equidistance data, which contains both noise and signal, on this interval. assume we get data b_i at point $t_i = ih$, $i = 0, 1, 2.....n$, $h = 1/n$.

Date: 05/12/13.

2. GRAPHIC EXPLANATION



As shown above, the dashed red line is original signal, blue line is noise signal. we desire to recover the original signal from measured data.

Part 2. denoise methods

3. OPTIMIZATION PERSPECTIVE

3.1. Tikhonov Regularization.

3.1.1. *Origins of Tikhonov Regularization.* For a general linear least squares problem there may be infinitely many least squares solutions.If we consider that the data contain noise, and that there is no point in fitting such noise exactly,it becomes evident that there can be many solutions that adequately fit the data in the sense that $||Gm - d||_2$ is small enough.

In Tikhonov regularization, we consider all solutions with$||Gm - d||_2 \leqslant \delta$,and select the one that minimizes the norm of m:

$$\begin{aligned} &||Gm - d||_2 \leqslant \delta \\ &\min \|m\|_2 \end{aligned} \tag{3.1}$$

the minimization of $\|m\|_2$ should ensure that unneeded features will not appear in the regularized solution. Note that as δ increases, the set of feasible models expands, and the minimum value of $\|m\|_2$ decreases. We can thus trace out a curve of minimum values of $\|m\|_2$ versus δ.

We have got conditions condition 3.1 .It is also possible to trace out this curve by considering problems of the form:

$$\begin{aligned} &\min ||Gm - d||_2 \\ &\|m\|_2 \leqslant \varepsilon \end{aligned} \tag{3.2}$$

As ε decreases, the set of feasible solutions becomes smaller, and the minimum value of$||Gm - d||_2$ increases.

So to sum up above two situation,our final option is to consider the damped least squares problem

$$\min \left(J_\alpha\left(m\right)\right) = \min(||Gm - d||_2^2 + \alpha^2 \left\|m\right\|_2^2) \tag{3.3}$$

which arises when we apply the method of Lagrange multipliers to condition 3.1, where α is a regularization parameter.It can be shown that for appropriate choices of δ ,ε, and α ,the three methods (condition 3.1, 3.2 and 3.3)yield the same solution.

3.1.2. *The Method of Tikhonov Regularization.* Maybe it's easy to think as above direction, but we have got less information about it, i.e, now we should find a proper m which satisfies formula 3.3.

Now we recall the first equation, as we have done,if the system is discrete,we can rewrite it as an equation of the form

$$Gm = d \tag{3.4}$$

where G is a compact linear operator from a Hilbert space H_1 into a Hilbert space H_2.

For we can not always find the inverse of G,so the generally equation 3.4 does not have a unique solution,therefore we seek a particular generalized solution,namely the least square solution of minimum norm. Changed form haven't dealt with trivial case, our question comes to be a well-posed question until we assume that $d \in \mathcal{D}\left(G^\dagger\right)$ and find the approximation of $G^\dagger d$.

The generalized solution $m = G^\dagger d$ of equation 3.4 is a least squares solution and therefore it satisfies the normal equations

$$G^* G m = G^* d \tag{3.5}$$

where G^* is the adjoint of G.Now the self-adjoint operator G^*G has non-negative eigenvalues and therefore for none zero number α,the operator $G^*G+\alpha^2 I$,where I is identity operator on H_1,has strictly positive eigenvalues.In particular,the operator $G^*G + \alpha^2 I$ has a bounded inverse, that is, the problem of solving the equation

$$\left(G^* G + \alpha^2 I\right) m_\alpha = G^* d \tag{3.6}$$

is well-posed.The second kond of equation 3.6 is called a regularized form of equation 3.6 and its unique solution

$$m_\alpha = \left(G^* G + \alpha^2 I\right)^{-1} G^* d \tag{3.7}$$

is called the Tikhonov approximation to $G^\dagger y$,the minimum norm solution of equation 3.5.

This can be accomplished conveniently in terms of a singular system $\{v_j, u_j; \mu_j\}$ for $G.m_\alpha$ can be written as following,

$$m_\alpha = \sum_{j=1}^{\infty} \frac{\mu_j^2}{\mu_j^2 + \alpha^2} \left\langle d, u_j \right\rangle v_j \tag{3.8}$$

The true minimum norm least squares solution is ,according to the following equation:

$$G^{\dagger} d = \sum_{j=1}^{\infty} \frac{1}{\mu_j} \langle d, u_j \rangle v_j \tag{3.9}$$

The following theorem demostrates the solution of equation 3.6 satisfying condition 3.3.

Theorem 1. *Let G:d $\longrightarrow$ m be a linear and bounded operator between Hilbert spaces and $\alpha \neq 0$.Then the Tikhonov functional J_α has a unique minimum $m_\alpha \in m$.This minimum m_α is the unique solution of the normal equation*

$$\left(G^* G + \alpha^2 I\right) m_\alpha = G^* d \tag{3.10}$$

3.1.3. *Apply Tikhonov Regularization into Our Project.* In our Project, We are concerned with recovering a function $u(t)$ on the interval $[0,1]$ given noisy data b_i at points $t_i = ih, i = 0, 1, \ldots, N$,with $N = 1/h$ .Because the data values are noisy, we cannot simply set $u(t_i) = u_i = b_i$:knowing that $u(t)$ should be piecewise smooth,we add a Tikhonov regularization term to penalize excessive roughness in u .For the unknown vector $\mathbf{u} = (u_0, u_1, \ldots, u_N)^T$ we therefore solve

$$\min \phi(u) = \frac{h}{2} \sum_{i=1}^{N} \frac{1}{2}[(u_i - b_i)^2 + (u_{i-1} - b_{i-1})^2] + \frac{\beta h}{2} \sum_{i=1}^{N} \left(\frac{u_i - u_{i-1}}{h}\right)^2 \tag{3.11}$$

3.1.4. *Determination of Regularization Parameter α.* First of all, we suppose there is no error in our observed data d. Then the ill-poseness is because G has no inverse.For this case, we will lost some information of estimation. We'd better assume continuity or some other properties to find out the solution of singular points. This case, however, we haven't seen in our facing question, so we don't want to research it.

The second case,which is our most important and difficult one,is that we have some data which have errors. For this case,we know that this question is ill-posed, for it doesn't satisfy stability condition whatever.

We have some conclusions in this case.

The best we can hope for is some estimate d^δ of d satisfying

$$\left\|d^\delta - d\right\| \leqslant \delta \tag{3.12}$$

where δ is a known bound on the measurement error.Then we have

$$\left\|m_\alpha^\delta - m_\alpha\right\| \leqslant \frac{\delta}{\alpha} \tag{3.13}$$

According to condition 3.13, any a priori choice $\alpha = \alpha(\delta)$ of the regularization parameter satisfying $\frac{\delta^2}{\alpha^2(\delta)} \longrightarrow 0$ as $\delta \longrightarrow 0$ leads to a regular algorithm for the solution of $Gm = d$. Although this asymptotic result may be theoretically satisfying, it would seem that a choice of the regularization parameter that is based on the actual computations performed, that is, an a posteriori choice of the regularization parameter would be more effective in practice.

There are many ways to find out regularization parameter α, We have some useful methods as below.

3.1.5. *Discrepancy Principle of Morozov.* One such a posteriori strategy is the discrepancy principle of Morozov. The idea of the strategy is to choose the regularization parameter so that the size of the residual $||Gm_\alpha^\delta - d^\delta||$ is the same as the error level in the data:

$$\left\|Gm_\alpha^\delta - d^\delta\right\| = \delta \tag{3.14}$$

3.1.6. *Engl's Criterion.* Engl's idea[1]is to find an α which make the following formula minimum.

$$\varphi(\alpha) = \frac{\left\|Gm_\alpha^\delta - d^\delta\right\|}{\alpha} \tag{3.15}$$

3.1.7. *Arcangeli Criterion.* Arcangeli Criterion[3]is that αsatisfying

$$\left\|Gm_\alpha^\delta - d^\delta\right\| - \frac{\delta}{\alpha} = 0 \tag{3.16}$$

3.1.8. *Tikhonov Criterion.* Tikhonov claims[4] that α has optimal value satisfying

$$\alpha_{opt} = \min_{\alpha \neq 0}\left\{\left\|\alpha\frac{dm_\alpha}{d\alpha}\right\|\right\} \tag{3.17}$$

3.1.9. *Hanke's L-Curves.* Let $\rho = \log\|Gm_\alpha - d\|, \theta = \log\|m_\alpha\|$,then the curvature of $\rho - \theta$ curve is defined

$$c(\alpha) = \frac{\rho'\theta'' - \rho''\theta'}{\left((\rho')^2 + (\theta')^2\right)^{3/2}} \tag{3.18}$$

Hanke uses the αwhich has maximum $c(\alpha)$ as the optimal one[2].

3.2. **Total Variation Regularization.** Total variation (TV) regularization is one alternative to Tikhonov regularization which will be introduced in this section. Total variation uses a regularization term based on the 1-norm of the model gradient which does not penalize model discontinuities. A variant of TV allows for prescribing the number of discontinuities in a piecewise-constant solution.[5]

3.2.1. *Origins of TV Regularization.* The total variation (TV) regularization function is appropriate for problems where we expect there to be discontinuous jumps in the model. In the one-dimensional case, the TV regularization function is

$$TV(m) = \sum_{i=1}^{n-1}|m_{i+1} - m_i| = \|Lm\|_1 \tag{3.19}$$

where total Variation Regularization

Total variation (TV) regularization is one alternative to Tikhonov regularization which will be introduced in this section. Total variation uses a regularization term based on the 1-norm of the model gradient which does not penalize model discontinuities. A variant of TV allows for prescribing the number of discontinuities in a piecewise-constant solution.[5]

$$L = \begin{pmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \cdots & \cdots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{pmatrix}$$

In higher-dimensional problems,L is a discretization of the gradient operator. In first- and second-order Tikhonov regularization, discontinuities in the model are smoothed out and do not show up well in the inverse solution. This is because smooth transitions are penalized less by the regularization term than sharp transitions. The particular advantage of TV regularization is that the regularization term does not penalize discontinuous transitions in the model any more than smooth transitions.
This approach has seen wide use in the problem of denoising a model. The denoising problem is a linear inverse problem in which G = I . In denoising, the general goal is to take a noisy data set and remove the noise while still retaining long term trends and even sharp discontinuities in the model.
We could insert the TV regularization term in place of in $\|Lm\|_2^2$ the Tikhonov regularization optimization problem to obtain

$$\min \|Gm - d\|_2^2 + \alpha \|Lm\|_1 \tag{3.20}$$

3.2.2. *The Method of TV Regularization.* This is no longer a least squares problem, and the techniques for solving such problems such as the SVD will no longer be applicable. In fact, 3.20 is a non-differentiable optimization problem because of the absolute values in $\|Lm\|_1$. One simple technique for dealing with this difficulty is to approximate the absolute value with a smooth function that removes the derivative discontinuity, such as

$$|x| \approx \sqrt{x^2 + \epsilon} \tag{3.21}$$

where ϵ is a small positive parameter.

3.2.3. *Apply TV Regularization into our Project.* Similar to Tikhonov case, we get the similar equation as that by giving same conditions,
(3.22)

$$\min \phi_2(\mathbf{u}) = \frac{h}{2} \sum_{i=1}^{N} \frac{1}{2}[(u_i - b_i)^2 + (u_{i-1} - b_{i-1})^2] + \frac{\beta h}{2} \sum_{i=1}^{N} \sqrt{\left(\frac{u_i - u_{i-1}}{h}\right)^2 + \epsilon}$$

3.2.4. *2D TV Reuglarization.* A classical variational denoising algorithm is the total variation (TV) minimizing process of Rudin-Osher-Fatemi. This algorithm seeks an equilibrium state (minimal energy) of an energy functional comprised of the TV norm of the image I and the delity of this image to the noisy input image I_0:

$$E = \int_\Omega (|\nabla I| + \frac{1}{2}\lambda(I - I_0))dxdy \tag{3.23}$$

This is further generalized by the Φ-formulation with the functional:

$$E_\Phi = \int_\Omega (\Phi\,|\nabla I| + \frac{1}{2}\lambda(I - I_0))dxdy \tag{3.24}$$

The Euler-Lagrange (E-L) equation is:

$$F = div(\Phi' \frac{\nabla I}{|\nabla I|}) + \lambda(I_0 - I) = 0 \tag{3.25}$$

where $\lambda \in \mathbb{R}$ is a scalar controlling the ?delity of the solution to the input image (inversely proportional to the measure of denoising). Neumann boundary conditions are assumed. The solution is usually found by a steepest descent method:

$$I_t = F,\ I|_{t=0} = I_0$$

When the noise is approximated by an additive white Gaussian process of standard deviation σ, the problem can be formulated as ?nding

$$\begin{aligned} &\min_I \int_\Omega \Phi(|\nabla I|)dxdy \\ &subject\ to\ \frac{1}{|\Omega|}\int_\Omega (I - I_0)dxdy = \sigma^2 \end{aligned} \tag{3.26}$$

In this formulation, λ can be considered as a Lagrange multiplier, computed by:

$$\lambda = \frac{1}{\sigma^2 |\Omega|}\int_\Omega div(\Phi' \frac{\nabla I}{|\nabla I|})(I - I_0)dxdy \tag{3.27}$$

And in this application, we first using the differential to approximate the derivatives, second we compute the flows and iterate the λ and evolve image I using the gradient descent method.

3.3. **Second Order Regularization.** the regularization term in Tikhonov and TV both contain first order differential quotient, the regularization terms measure the roughness of both original signal and noise. We substitute it by the second order different quotient, and expect can amplify the effect of noise term under the assumption that each noise term is independent identical distributed random variable.

3.4. **optimization with constrains.**

3.4.1. *general model.* the non-integer optimization problem can be generalized as the following expression:

$$\begin{array}{rll} \min & f(x), & x \in \mathbb{R}^n \\ st & c_i(x) = 0 & i \in E\{1, 2, ..., l\}, \\ & c_i(x) \le 0 & i \in I\{l+1, l+2, ..., l+m\}. \end{array}$$

3.4.2. *first order necessary conditions of constrain problem :Karush-Kuhn-Tucjer(KKT) condition.*

Theorem 2. *assume $f(x), c_i(x)$have continuous first order derivative, and$x^*$is the local solution , and $\nabla c_i(x)(i \in (E \cup I^*))$are linear independent, then exist constant $\lambda^* = (\lambda_1^*, \lambda_2^*, ..., \lambda_{l+m}^*)$s.t.*

$$\begin{array}{rcl} \nabla_x L(x^*, \lambda^*) & = & \nabla f(x^*) + \sum_{i=1}^{l+m} \lambda_i^* \nabla c_i(x^*) = 0 \\ c_i(x^*) = 0 & , & i \in E \\ c_i(x^*) \le 0 & , & i \in I \\ \lambda_i^* \ge 0 & , & i \in I \\ \lambda_i^* c_i(x^*) = 0 & , & i \in I \end{array}$$

3.4.3. *application in denoise case.* assume there is a function $N(x) : x \to n \ x \in \mathbb{R}^n, n \geq 0$ which measures the noisiness of data x, our problem becomes

$$\begin{aligned} \min \quad & \|u-b\|_2 \quad u \in \mathbb{R}^n \\ st \quad & N(u) \leq \varepsilon \end{aligned}$$

we naturally define noiseness by second order differential quotient $N(x)$:

$$N_1(u) = \textstyle\sum_{i=1}^{n} (u_{i+2} - 2u_i + u_{i-1})^2$$

the physical meaning of λensures that λ^* here must be greater than zero, hence $N(x) = \varepsilon$.KKT condition become:

$$\nabla_{u,\lambda^*} L(u, \lambda^*) = \nabla(\|u-b\|_2 + \sum_{i=1}^{l+m} \lambda_i^*(N(u) - \varepsilon)) = 0$$

from above condition we can solve u.

4. FOURIOR TRANSFORMATION PERSPECTIVE

4.1. Discrete Fourier Transformation.

Definition 3. tirgonometric polynomials of degree less than or equal to n:

$$S_n(x) = \frac{a_o}{2} + a_n \cos nx + \sum_{k=1}^{n-1} (a_k \cos kx + b_k \sin kx)$$

for function $f \in C(-\pi, \pi)$,we can find the continuous least square approximation such that minimize the error term

$$E(S_m) = \sum_{j=0}^{2m-1} (y_i - S_m(x_j))^2$$

it can be calculated the coefficient is

$$a_k = \frac{1}{m} \sum_{j=0}^{2m-1} y_j \cos kx_j$$

$$k = 0.1....., m,$$

$$b_k = \frac{1}{m} \sum_{j=0}^{2m-1} y_j \sin kx_j$$

$$k = 0.1....., m-1,$$

these coefficient can be calculated efficiently by FFT algorithm, which we will not touch here. noticing the signal is the weighted sum of several trignometric signals with different frequency, and noise always contribute to the high frequency signal. So we simply discard these high frequency coefficients and recover the signal, this method have some flaws, since the origina signal may have high frequency component, this method will affect the original signal too.

5. AVERAGE SMOOTHING METHOD

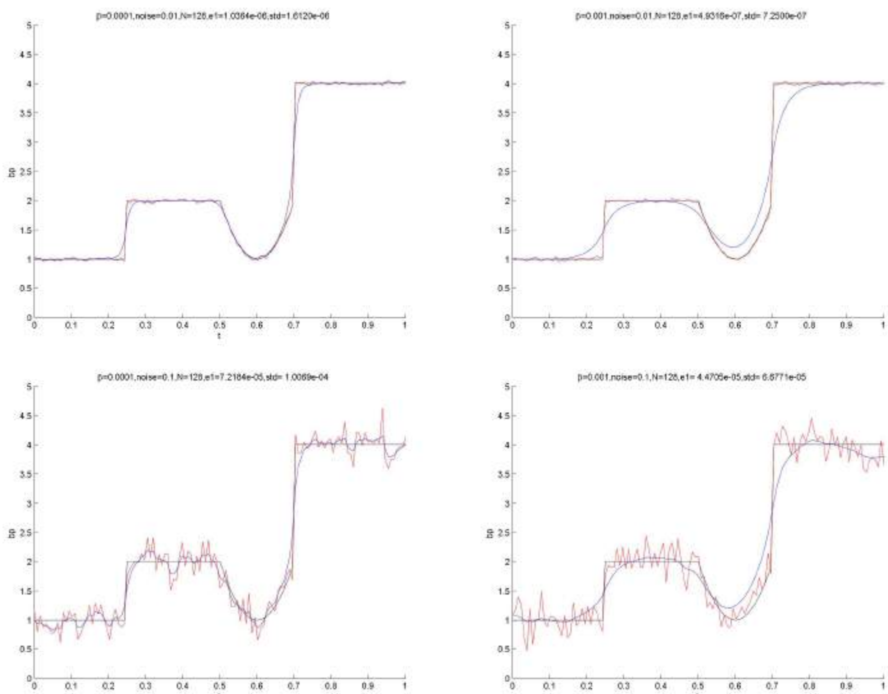
the idea of this method is very simple: because the data jumps intensivly, we just take the average of data at the neighbourhood of one signal data point, then displace this central data with average value. this can smooth the signal because if the noise are independent identity distributed(idd) random variables, the average operation can help us reduce the varience of the random term.

Part 3. result and disscusion

The original signal is shown in the Part I, we use two normally distributed noise with different varience, and the results are aompared below. we use the mean of the square of error term(MSE) to measure the effect of thoese method. Since the noise is random, the MSE also varies every time, we calculate the average MSE of each fixed parameter as well as its standard deviation. problems seeking for the proper parameter in order to minimize expectation of MSE is stochastic optimization which is rather higher than our current level.

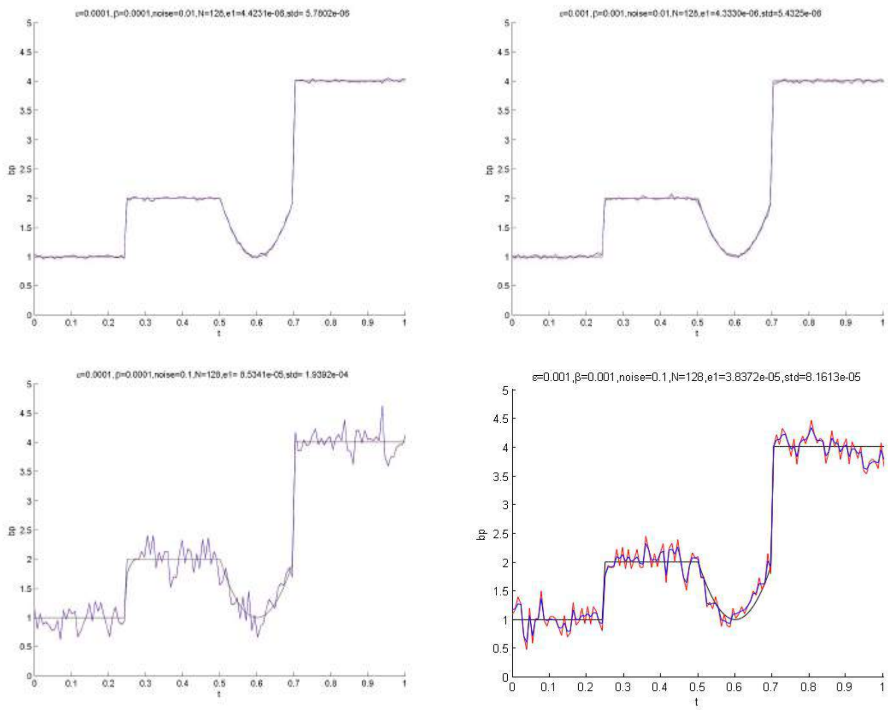
6. Tikhonov Regularization and Total Variation Regularization

We used 128 sample points, the Tikhonove regurization's results are shown here:

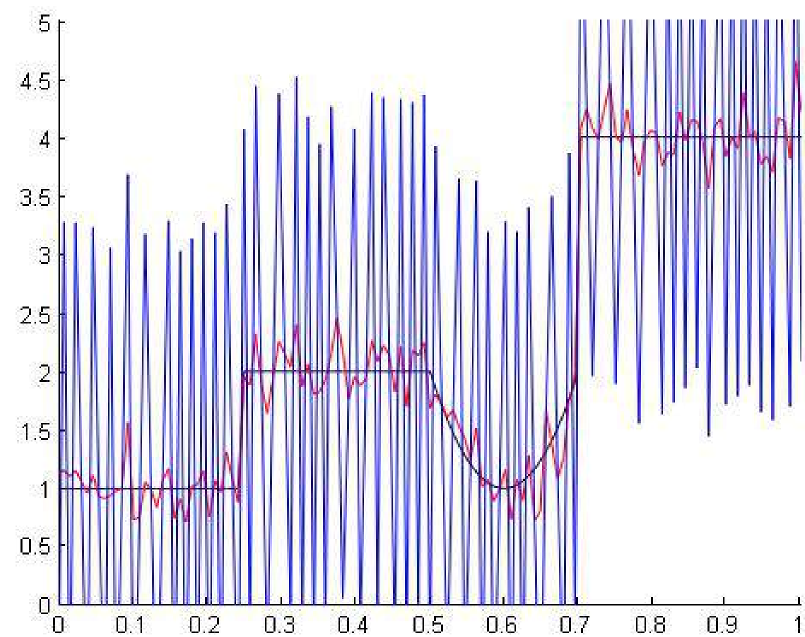


the greater β is the smooth the signal will be after denoising process, and as we will see this phenomenon holds for the most of the following regularization methods. another observation is the MSE of processed signal and its standard deviation usually have the same order. If the noise increases, we have to choose the greater β to suppress the noise.

the TV regurization's results are shown below:



We found the proper ϵ value should be set around β too large or too small is not good. since the signal is not very smooth, we increase the β hoping to get smooth signal, but the interesting thing is β has a limit beyond which the signal will jump intensively, even more intensive than noised signal. as we can see below. so our conclusion is TV regularization is not as good as Tikhonov Regularization.



(A) Original picture

(B) Noisy picture

(C) TV denoising picture eps=0.01

FIGURE 7.1. The pictures before and after the TV minimization

7. 2D TV REGULARIZATION

1. For time step $dt = 0.15$, we can get the result:

```
kk =
   86
```

which means that the total iteration number is 86, and we shall see later that the iteration number has a local minimum which is an important feature of gradient descent method.

And the image result is:

It is easy to discover in the pictures that after the TV minimization the figure become much smoother than before. Here the upper bound of the mean absolute error value(eps) here is 0.01. And it is obvious that, the smaller the eps is, the smoother the picture would be.

This result is showed below:

which is quite obvious that the picture on the left is much smoother than the right one.

2. For the case that time step $dt = 0.2$, there is:

(A) TV denoising picture eps=0.01

(B) TV denoising picture eps=0.1

FIGURE 7.2. The compare of two results with different eps

```
kk =
67
```

which is much less than the case $dt = 0.15$

3. For the case that time step $dt = 0.23$, there is :

```
kk =
63
```

while

```
kk =
60
```

for $dt = 0.24$, and

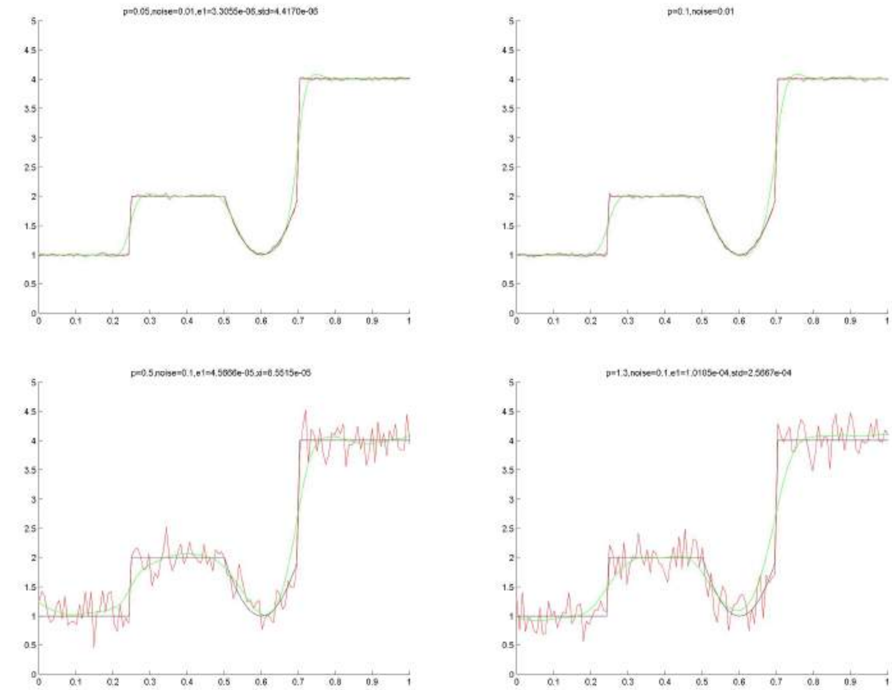
```
kk =
63
```

for $dt = 0.25$.

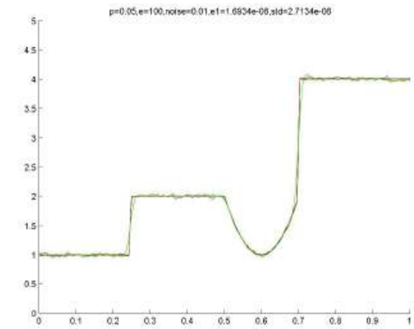
Thus we can find that it is a dt such that at which the total iteration number goes to a minimum, and that is quite reasonable due to the physical property of the gradient descent method.

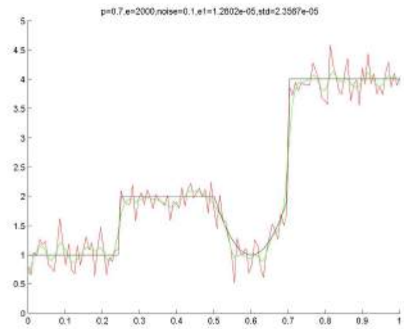
8. 2nd order regularization

2nd order Tikhonov method

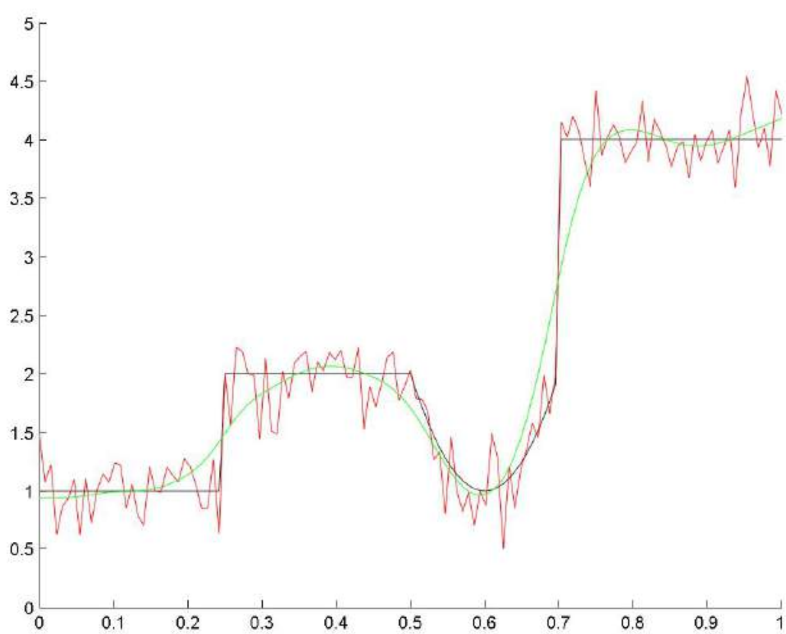


2nd order TV method

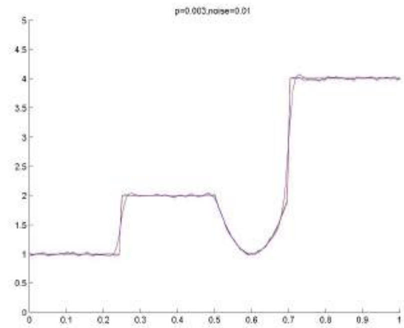


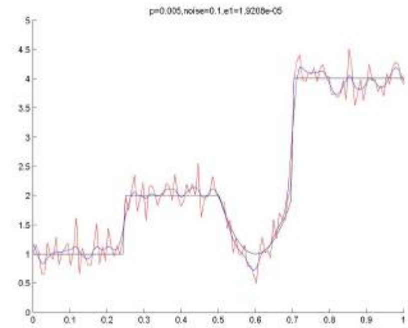


We can see this method can alse denoise the signal; but comparing the MSEs of this method with previous two methods, our 2nd order method doesn't give us a surprise, its performance is similar to other method. one difference is that the parameter β don't have limit in 2nd order TV method, we can get very smooth curve by increasing β. as you can see below:



9. OPTIMIZATION WITH CONSTRAINS

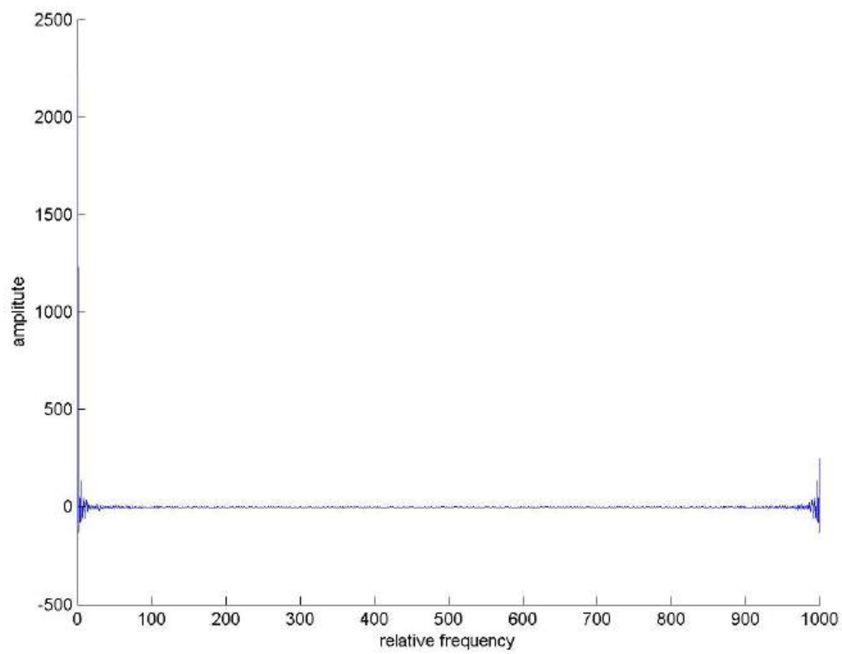




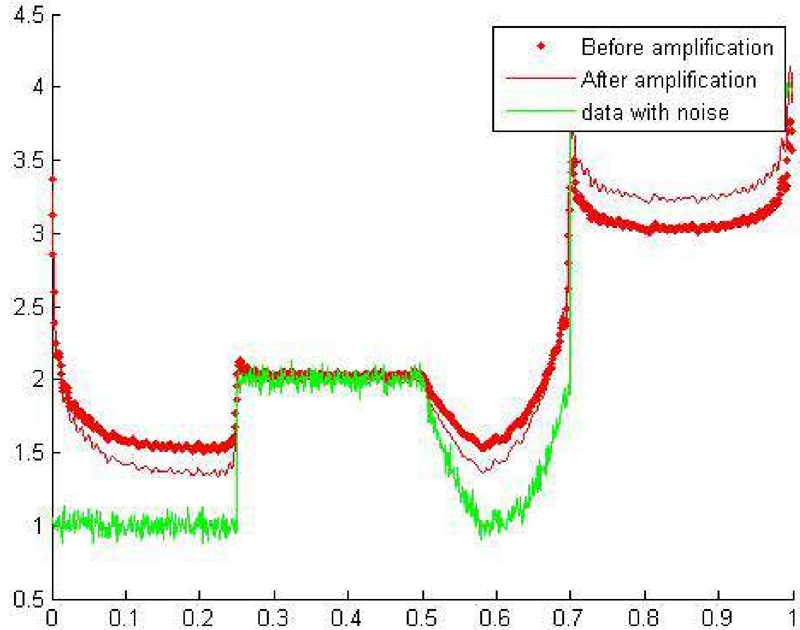
we found this methods' average MSE is hard to measure, because sometimes it works properly with the result as good as other method. sometimes it generate a list of Nans. Since the solution is not so stable, we don't recommend this method

10. Discrete Fourier Transformation

To implement this method we first do the FFT, the signal is transformed into intensities of different frequencies in frequency domain:

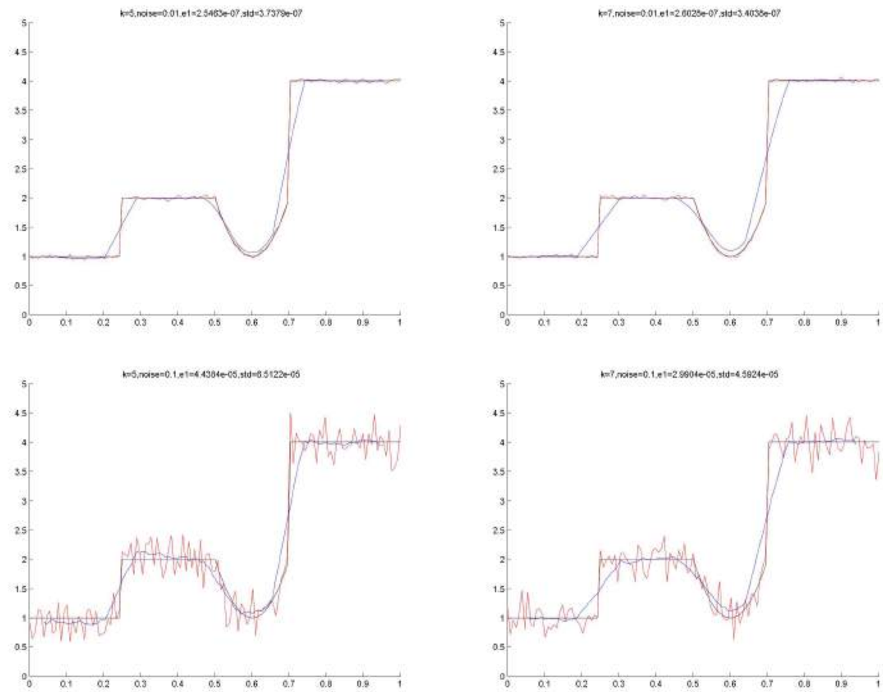


then we choose the cutoff frequency discard the high frequency region. after the recovering the signal, some amplification process is needed cause fourier filter always decreases the amplitude of original signal. here is the graph.



in the experiment the MSE of noised signal and denoised signal are very close. which means Fourier filter method almost don't denoise the signal from MSE point of view! from the graph we also can feel the obvious weakness of Fourier filter method in this case.

11. AVERAGE SMOOTHING METHOD



If the signal is very smooth, this method can depress the noise signigicantly. But at the step of signal there is a big distortion. and two ends of signal will lose information too.

12. CONCLUSION

Some methods have flaws, but there is no evidence one method having apparent excellent advantage.

References

[1] And A. Neubauer H.W.Engl, M.Hanke. *Regularization on Inverse Problem*. Kluwer Academic Publishers, 1996.

[2] Vogel C R. Non-convergence of the l-curve regularization parameter selection method. *Inverse Problems*, :535–547, 1996.

[3] Arsenin V Y Tikhonov A N.*Solutions of Ill-Posed Problems*. John Wiley and Sons, 1977.

[4] Groesch C W. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman Advanced Publishing Program, 1984.

[5] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, 2005.