Refined Exponential Filter with Applications to Image Restoration and Interpolation

Yanlin Geng[1], Tong Lin[1], Zhouchen Lin[2], and Pengwei Hao[1,3]

[1] Center for Information Science, Peking Univ., Beijing 100871, China
{gengyanlin, lintong, phao}@cis.pku.edu.cn
[2] Microsoft Research Asia, Zhichun Road #49, Haidian, Beijing 100190, China
zhoulin@microsoft.com
[3] Dept. of Computer Science, Queen Mary, Univ. of London, London E1 4NS, UK
phao@dcs.qmul.ac.uk

Abstract. Ill-posed linear equations are pervasive in computer vision. A popular way to solve an ill-posed problem is regularization. In this paper, we propose a new criterion for designing the regularizing filter. This criterion reveals the implicit assumption made by regularizing filters. Then with the help of the discrete Picard condition, we refine the exponential filter using our criterion. The effectiveness of our method is demonstrated on image restoration and interpolation.

1 Introduction

Computer vision involves many ill-posed problems [1], such as image restoration, edge detection, optical flow, motion estimation, and surface reconstruction. According to [2], a well-posed problem has three properties: existence, uniqueness and stability of the solution; if any one of these properties does not hold, the problem is ill-posed.

Regularization is a prevailing method to solve ill-posed problems. Based on the methods used, there are mainly three approaches to regularization: optimization, filtering and iterative methods. The first method is actually Tikhonov method [3] and has a Bayesian interpretation; the second one utilizes the spectrum of the problem and devotes to tailoring a suitable filter; the third method settles the problem using an iterative process, and in fact the number of iterations plays the role of regularization. These three methods are closely related, especially under L_2 norm. In this paper we mainly focus on the filtering approach. Before that we would like to introduce Tikhonov regularization.

Many ill-posed problems come from a first kind Fredholm integral equation [4]

$$\int K(x,t)f(t)dt = g(x)\ . \tag{1}$$

And they can be discretized as linear equations of the form

$$Ax = b\ . \tag{2}$$

Based on the idea of balancing the residual and some apriori constraint on the solution, Tikhonov regularization [3] finds the solution by minimizing

$$J(x) := \|b - Ax\|^2 + \alpha[\Omega(x)]^2 \ . \tag{3}$$

The constraint $\Omega(x)$ ensures the stability of the solution, while the regularization parameter α controls the closeness between the original and the new equation. Thus two important issues in regularization are choosing proper constraints and finding the optimal parameters.

Basically speaking, a proper constraint should penalize what we do not want the solution to exhibit. And there has already been a lot of work on choosing the constraint $\Omega(x)$. For example, ordinary Tikhonov regularization (oTik) takes the constraint as $\|x\|_2$, which restricts the size of the solution. For image restoration, Phillips [5] proposed to use $\|Lx\|_2$, where L is the Laplace operator. This assumes small differences in luminance between neighboring pixels. As a variance, it was shown in [6] that using the total variation can preserve edges better than $\|Lx\|_2$. For sparse solution, lasso [7] suggests using $\|x\|_1$. Although each kind of term has a meaningful interpretation, an interesting question is that, how can we refine the constraint that is being used?

To facilitate the analysis, we consider the filtering approach, which makes use of the spectrum of A. Some exemplar filters include the exponential filter (Exp) [8], modified Tikhonov regularization (MTR) [9], spatial regularization [10], and so on. These methods usually design a filter heuristically: they just modify the filter to satisfy certain subjective request.

In this work, using backward error analysis, we propose a criterion for designing the regularizing filter. This criterion shows that there is a relationship between the constraint and the problem itself (*e.g.*, A and b). We further study the characteristic of a solvable problem, namely the Picard condition [11]; then we show how the Picard condition helps refine the filter for a specific constraint.

2 Designing the Regularizing Filter

In this section, we first introduce the regularizing filter, then we propose our criterion. To make use of the criterion, we consider the Picard condition and show how to refine the exponential filter. For the notation, throughout the paper, we use A_i as the i-th column of a matrix A and b_i as the i-th element of a vector b. Without special clarification, the norm used is the L_2 norm.

2.1 The Regularizing Filter

To solve $Ax = b$, the least squares method minimizes the residual $R(x) = \|b - Ax\|^2$ and the solution is $x = (A^T A)^\dagger A^T b$, where $\dagger$ is the Moore-Penrose pseudo inverse. Suppose the singular value decomposition (SVD) is $A = USV^T$, where U and V are unitary matrices, S is a diagonal matrix with its diagonal elements $s_i \geq 0$ called the singular values; then the solution can be expressed as

$$x = VS^\dagger U^T b =: VS^\dagger \beta = \sum\nolimits_i \beta_i s_i^{-1} V_i \ . \tag{4}$$

where $\beta := U^T b$ is called the *Fourier coefficients*.

However, when the small nonzero singular values of A decay gradually to zero, this solution can bias greatly from an acceptable one. This is because in practice, b is often contaminated by noise, thus a very small s_i tends to amplify the noise enormously. In this sense, the problem is ill-posed. To solve this ill-posedness, oTik minimizes $J(x) = \|b - Ax\|^2 + \lambda^2 \|x\|^2$, and the solution is

$$x = \sum s_i^2 (s_i^2 + \lambda^2)^{-1} \beta_i s_i^{-1} V_i =: \sum q_{\text{otik}}(\lambda, s_i) \beta_i s_i^{-1} V_i \ . \tag{5}$$

Compared to the least squares solution, this solution involves a low pass filter

$$q_{\text{otik}}(\lambda, s) = s^2 (s^2 + \lambda^2)^{-1} \ , \tag{6}$$

thus noises in high frequencies are restrained. That is why $q(\lambda, s)$ is called the *regularizing filter* [4]; and these $q_i = q(\lambda, s_i)$ are called the *filter factors*.

For a general constraint $\|Lx\|$, let $y = Lx$, we can transform the problem of minimizing $J(x) = \|b - Ax\|^2 + \lambda^2 \|Lx\|^2$ into oTik

$$\min \tilde{J}(y) = \|b - AL^\dagger y\|^2 + \lambda^2 \|y\|^2 \ . \tag{7}$$

To obtain the filter in this case, we need the generalized SVD of (A, L)

$$A = U \varXi X^{-1} \ , \quad L = VMX^{-1} \ , \tag{8}$$

where X is invertible, U and V are orthonormal, $\varXi$ and M are diagonal matrices with the diagonals being ξ and μ, respectively. So $AL^\dagger = U \varXi M^{-1} V^T =: USV^T$, where $S := \varXi M^{-1}$ is a diagonal matrix with its diagonal elements $s_i := \xi_i \mu_i^{-1}$ called the generalized singular values. According to Eqn.(5), we have the solution as $y = \sum q_{\text{otik}}(\lambda, s_i) \beta_i s_i^{-1} V_i$; and substitute this into $x = L^\dagger y$, we obtain

$$x = \sum q_{\text{otik}}(\lambda, s_i) \beta_i \xi_i^{-1} X_i \ . \tag{9}$$

In the solution above[4], $q_{\text{otik}}(\lambda, s)$ is also called the regularizing filter, where s_i are the generalized singular values.

2.2 Criterion for Designing the Filter

In practice b is often corrupted by noise η, thus we should not solve $Ax = b$ directly. To eliminate the noise, we introduce a perturbation term E and solve $(A + E)x = b + \eta$ instead. This is motivated by the method of backward error analysis in numerical analysis. As the true solution satisfies $Ax = b$, our goal is to find a proper E that is expected to satisfy $Ex = \eta$.

From Eqn.(4), the solution to the exact equation $Ax = b$ is $x = VS^\dagger \beta$, so we get $\eta = EVS^\dagger \beta$. Suppose the variance matrices of η and β are $\sigma^2 I$ and CC^T respectively, we have

$$\sigma^2 I = var(\eta) = var(EVS^\dagger \beta) = (EVS^\dagger C)(EVS^\dagger C)^T \ . \tag{10}$$

[4] If L is rank deficient, an extra $x_0 = \sum_{i > \text{rank}(L)} \beta_i X_i$ should be added to Eqn.(9).

Table 1. Comparison of several filters.

Methods	Filter q	Coefficients $\|\beta\| \propto$
oTik [3]	$s^2(s^2+\lambda^2)^{-1}$	$s^2\lambda^{-2}$
Exp [8]	$1-\exp\{-s^2\lambda^{-2}\}$	$\exp\{s^2\lambda^{-2}\}-1$
MTR [9]	$s^2(s^{2\sigma}+\lambda^{2\sigma})^{-\frac{1}{\sigma}}$	$s^2\{(s^{2\sigma}+\lambda^{2\sigma})^{\frac{1}{\sigma}}-s^2\}^{-1}$

This leads to $E = \sigma W C^{\dagger} S V^T$, where W is an arbitrary orthonormal matrix. Due to the arbitrariness, we may set $W = U$ and obtain

$$E = \sigma U C^{\dagger} S V^T \ . \tag{11}$$

With this estimate of E, we are going to solve $U(I + \sigma C^{\dagger})SV^T x = b$. For a general β, suppose its elements are independent (thus C is diagonal), then the solution is

$$x = \sum \frac{1}{1+\sigma c_i^{-1}} \frac{\beta_i}{s_i} V_i \ , \tag{12}$$

where c is the diagonal of C. This solution suggests taking the filter as $q_i = (1+\sigma c_i^{-1})^{-1}$, which results in $c_i = \sigma q_i (1-q_i)^{-1}$. Notice that $var(\beta) = CC^T$, we arrive at *our criterion for designing regularizing filter*

$$|\beta_i| \approx \sigma q_i (1-q_i)^{-1} \propto q_i (1-q_i)^{-1} \ . \tag{13}$$

Our criterion suggests that the filter should be designed closely related to the Fourier coefficients $\beta = U^T b$. With this criterion, we can also analyze what a filter models β.

2.3 Using the Picard Condition

According to our criterion $|\beta_i| \approx \sigma q_i(1-q_i)^{-1}$, a filter q can be designed by modeling β. However, it is difficult to model a general β. Here we consider this problem in the viewpoint of the Picard condition, which is essential for solving an ill-posed problem [11].

The Picard Condition. Suppose the kernel K in Eqn.(1) has a singular value expansion $K(x,t) = \sum s_i u_i(x) v_i(t)$, and $\beta_i := \langle u_i, g\rangle$ are the coefficients. In order that the problem is solvable, the Picard condition requires that [11] $\sum_{i=1}^{\infty} (\beta_i s_i^{-1})^2 < \infty$. While discretized, the Picard condition desires that *the elements of β decay faster than the corresponding singular values on the average.*

In Table 1, we compare some existing filters, most of which assume that $|\beta_i| \propto s_i^2$. This ad hoc setting requires that β_i decays as fast as s_i^2; while the Picard condition desires that β_i decays faster than s_i.

Our Filter 'rExp'. Inspired by the exponential filter, we propose to model

$$|\beta_i| \approx \sigma(\exp\{s_i^{\rho}\lambda^{-\rho}\} - 1) \quad \text{with} \quad \rho > 1 \ , \tag{14}$$

$$q_{\mathrm{rexp}}(\lambda, s) = 1 - \exp\{-s^{\rho}\lambda^{-\rho}\} \ , \tag{15}$$

Algorithm 1: Choosing Parameters for rExp

Initialize $\rho = 2$ and estimate σ
Loop:
  Find the optimal λ for $x = \sum (1 - \exp\{-s_i^{\rho}\lambda^{-\rho}\})\beta_i s_i^{-1} V_i$
  Update $\rho \leftarrow$ mean $\left|\ln\{\ln(|\beta_i|\sigma^{-1} + 1)\}/(\ln s_i - \ln \lambda)\right|$
End of Loop
Find the optimal λ for $x = \sum (1 - \exp\{-s_i^{\rho}\lambda^{-\rho}\})\beta_i s_i^{-1} V_i$

and denote it as the refined exponential filter (rExp). Here a free parameter ρ is incorporated so that we can better model β; and rather than setting it as 2 for convenience, we just require $\rho > 1$ so that the Picard condition is satisfied. In the following paragraph, we also provide an algorithm for determining ρ.

Choosing the Parameters. It is a crucial problem to choose a suitable parameter λ for all regularization schemes. Fortunately there have been several robust and popular ways. For example, L-curve [12] and generalized cross-validation (GCV) [13]. If the noise level is predictable, Morozov discrepancy principle [14] can also be used. Here we also provide an iterative method to choose ρ. With an initial $\rho = 2$, we obtain λ from one of the methods mentioned above. Then from $|\beta_i| \approx \sigma(\exp\{s_i^{\rho}\lambda^{-\rho}\} - 1)$, we arrive at

$$\rho \approx \text{mean}\left|\frac{\ln\{\ln(|\beta_i|\sigma^{-1} + 1)\}}{\ln s_i - \ln \lambda}\right| . \tag{16}$$

This procedure can be performed repeatedly until we get a proper ρ. The algorithm is summarized in Algorithm 1.

3 Experiments

In the experiments, we apply our method to image restoration and image interpolation. The test images shown in Figure 1 are the 24 Kodak Images[5].

3.1 Image Restoration

A blurred and noisy image can be modeled as $g = h * f + \eta$, where f is the original image, g is the observed image, h is the blurring kernel, $*$ denotes convolution and η is the additive noise. Image restoration is to recover the original image by solving $Hf = g$. In [5], Phillips proposed to minimize

$$\|g - Hf\|^2 + \lambda^2\|Lf\|^2 , \tag{17}$$

where L is the Laplacian operator. In [6], the authors suggested to minimize $\|g - Hf\|^2 + \lambda^2\|f\|_{TV}$, where $\|\cdot\|_{TV}$ denotes the total variation. With this kind

[5] http://r0k.us/graphics/kodak/

Fig. 1. The 24 Kodak Images used in our experiments.

of constraint, edges can be preserved. This method was developed as 'scalar TV' and further as 'adaptive TV' methods [15].

In practice, we often deal with the Toeplitz matrices. A *block-circulant-circulant-block* (BCCB) matrix can be diagonalized very efficiently using *fast Fourier transform* (FFT). Suppose H and L are BCCB matrices, then we have $H = F \Xi F^*, L = FMF^*$, where F is the unitary discrete Fourier transform matrix. Similar to Eqn.(9), the solution is

$$x = \sum q(\lambda, s_i) \beta_i \xi_i^{-1} F_i \ , \tag{18}$$

where $\beta = F^* b$, namely applying the inverse Fourier transform to b.

In the experiment, the images are first degraded by a 3×3 average filter, and then corrupted by white Gaussian noise with a standard deviation $\sigma = 10$. During the restoration, the blurring kernel h is estimated using the method in [16]; H and L are constructed as BCCB matrices so that FFT can be used. We apply rExp to restore the images, followed by a Wiener filter to further reduce the noise. We compare our method with oTik, Exp [8], Wiener filter, and total variation methods (Scalar TV and Adaptive TV) [15]. The results are reported using the *peak signal-to-noise ratio* (PSNR)

$$PSNR = 10 \cdot \log_{10}\{MAX_I^2 / MSE\} \ , \tag{19}$$

where MAX_I is the maximum possible pixel value for the image (255 for 8-bits images), and MSE is the mean square error for the original and restored images.

We show the PSNR on the restored images in Figure 2 and detail the average PSNR of each method in Table 2. Our method provides the highest average PSNR on the 24 images; and significant improvement is achieved compared with the exponential filter. We also plot the parameter ρ of our method in Figure 2, which illustrates the necessity of allowing ρ other than 2. For visual comparison, we show blowups of the restored images in Figure 3. It is clear that our method

Table 2. Average PSNR of the 24 restored Kodak images.

Methods	oTik	Exp	Wiener	Scalar TV	Adaptive TV	rExp
PSNR	25.29	26.51	26.83	26.87	27.28	27.75

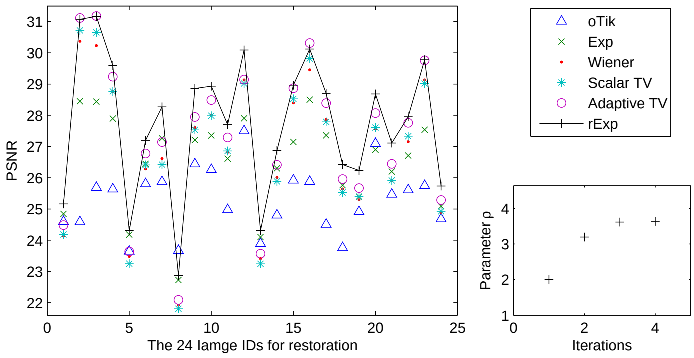

Fig. 2. (Left) PSNR of the 24 restored Kodak images. (Right bottom) The parameter ρ of rExp with respect to iterations on the 5-th Kodak image.

provides restored images visually comparable with Adaptive TV and better than other methods.

3.2 Image Interpolation

Image interpolation is used to render high-resolution images from low-resolution images. A low-resolution image can be modeled as $g = DHf + \eta$, where f and g are the lexicographic order of high-resolution images F and low-resolution images G, respectively. D and H are the matrices that model the decimation and the blurring processes, respectively.

An interesting interpolation algorithm is proposed in [17]. The main idea is to solve the problem using the Tikhonov regularization. Considering the huge sizes of H and D, the authors assume that these matrices are separable:

$$H = H_1 \otimes H_2 \ , \quad D = D_1 \otimes D_2 \ , \tag{20}$$

where $\otimes$ represents the Kronecker product. Thus the model is equivalent to

$$G = (D_2H_2)F(D_1H_1)^T + \eta \ . \tag{21}$$

Then with the aid of the Kronecker product and SVD, the computation cost can be reduced greatly. For example, if the decimation factor is 2, then we have

$$D_1 = \begin{bmatrix} 1\,0\,0\,0\,0 \ldots \\ 0\,0\,1\,0\,0 \ldots \\ 0\,0\,0\,0\,1 \ldots \\ \vdots\,\vdots\,\vdots\,\vdots\,\vdots\,\ddots \end{bmatrix} , \quad H_1 = \begin{bmatrix} v_0 & v_1 & \ldots & v_{-1} \\ v_{-1} & v_0 & \ldots & v_{-2} \\ \vdots & \vdots & \ddots & \vdots \\ v_1 & v_2 & \ldots & v_0 \end{bmatrix} , \tag{22}$$

Fig. 3. The restored images using different methods. From left to right: (Top) Original image, Degraded image, oTik, Exp; (Bottom) Wiener, Scalar TV, Adaptive TV, rExp.

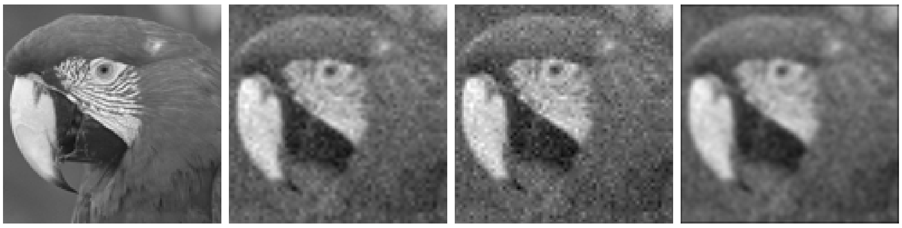

Fig. 4. The interpolated images using different methods. From left to right: Original image, Bicubic, oTik, rExp.

where $v = (v_{-k}, ..., v_{-1}, v_0, v_1, ..., v_k)^T$, and $h = uv^T$ is the blurring kernel. For a 3×3 mask, it is often assumed that $u = v = (a, 1-2a, a)^T$. Without any apriori information, we may set $a = 0.25$.

However, it is important to notice that, under the assumption of separability and with the selection of u and v above, the singular values of D_1H_1 range from $|1-2a|$ to $\sqrt{(1-2a)^2 + 4a^2}$ (see Appendix A). Thus if a is not near 0.5, we can use iterative methods such as the steepest descent or the conjugate gradient to solve Eqn.(21). So we propose our method as follows. First we restore the noisy low-resolution image g using the method we have introduced in Section 3.1, then we employ the separability and solve the normal equation of Eqn.(21) to obtain the high-resolution image F.

In the experiment, we first blur F with a 3×3 average filter. Then we subsample the blurred image and add Gaussian noise with $\sigma = 10$ to construct G. We use our method and the method in [17] to compute image F, respectively. For more comparison, we also resize the image G using bicubic interpolation.

Table 3. Average PSNR of the 24 interpolated Kodak images.

Methods	Bicubic	oTik	rExp
PSNR	25.04	24.82	25.85

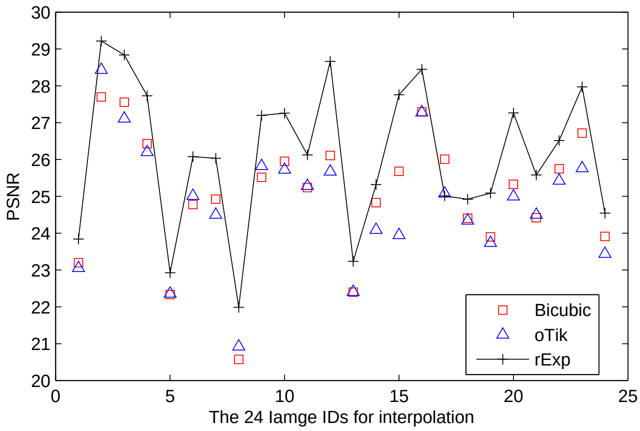

Fig. 5. PSNR of the 24 interpolated Kodak images.

It is clear that our method outperforms other methods in both PSNR (Table 3 and Figure 5) and visual aspect (Figure 4).We believe that this benefits from the flexibility of ρ in our method.

4 Conclusions

In this paper, we suggest a criterion for designing the regularizing filter. By incorporating the Picard condition, we propose to refine the exponential filter. Our scheme works effectively for ill-posed problems, which has been demonstrated on image restoration and image interpolation.

References

1. Bertero, M., Poggio, T.A., Torre, V.: Ill-posed problems in early vision. Proc. of the IEEE **76**(8) (1988) 869–889
2. Hadamard, J.: Lectures on Cauchy's problem in linear partial differential equations. Courier Dover Publications (2003)
3. Tikhonov, A.N., Arsenin, V.Y.: Solutions of ill-posed problems. New York: Wiley (1977)
4. Kirsch, A.: An introduction to the mathematical theory of inverse problems. Springer (1996)
5. Phillips, D.L.: A technique for the numerical solution of certain integral equations of the first kind. Journal of the ACM **9**(1) (1962) 84–97

6. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
7. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B **58**(1) (1996) 267–288
8. Calvetti, D., Lewis, B., Reichel, L.: Smooth or abrupt: a comparison of regularization methods. In: Proc. SPIE. Volume 3461. (1998) 286–295
9. Li, G., Nashed, Z.: A modified Tikhonov regularization for linear operator equations. Numerical Functional Analysis and Optimization **26** (2005) 543–563
10. Velipasaoglu, E.O., Sun, H., Zhang, F., Berrier, K.L., Khoury, D.S.: Spatial regularization of the electrocardiographic inverse problem and its application to endocardial mapping. IEEE Transactions on Biomedical Engineering **47**(3) (2000) 327–337
11. Hansen, P.C.: Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. SIAM, Philadelphia (1999)
12. Hanke, M., Hansen, P.C.: Regularized methods for large scale problems. Surv. Math. Ind. **3** (1993) 253–315
13. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics (1979) 215–223
14. Groetsch, C.W.: The theory of Tikhonov regularization for Fredholm equations of the first kind. Pitman, Boston (1984)
15. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Texture preserving variational denoising using an adaptive fidelity term. In: Proc. VLSM. (2003) 137–144
16. Biggs, D.S.C., Andrews, M.: Acceleration of iterative image restoration algorithms. Appl. Opt. **36**(8) (1997) 1766–1775
17. Chen, L., Yap, K.H.: Regularized interpolation using Kronecker product for still images. In: Proc. ICIP. Volume 2. (2005) 1014–1017

Appendix

A Spectrum of Decimated Toeplitz Matrices

Property. If A is the odd rows of a circulant Toeplitz matrix H

$$H = \begin{bmatrix} b & a & 0 & \dots & 0 & a \\ a & b & a & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a & 0 & 0 & \dots & a & b \end{bmatrix} =: \mathrm{Toep}[b, a, 0, ..., 0, a] \ , \tag{23}$$

then the singular values $\sigma(A) \subseteq [|b|, \sqrt{b^2 + 4a^2}]$.

Proof. With a proper permutation matrix P, we have $B := AP = [A_0\ A_1]$, where $A_0 = bI$, $A_1 = aJ$ with $J = \mathrm{Toep}[1, 0, ..., 0, 1]$. Then $BB^T = b^2 I + a^2 JJ^T$. Notice $\|JJ^T\|_1 \leq \|J\|_1 \|J^T\|_1 = 4$ and the maximum eigen-value $\lambda_{\max}(M) \leq \|M\|_p$ for any $p \geq 1$, we get $\lambda(JJ^T) \subseteq [0, 4]$, which leads to $\lambda(BB^T) \subseteq [b^2, b^2 + 4a^2]$. Immediately we obtain that the singular values of B (also of A) range between $|b|$ and $\sqrt{b^2 + 4a^2}$.