Non-local Methods for DNA Detection and Image Restoration

Yuewei Liu

Resubmitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Engineering and Physical Sciences

September 2018

ABSTRACT

Many methods have been developed for abnormal linear systems. However, these methods often assume locality for datasets; thus the features of the data cannot be extracted most efficiently to formulate models, which can lead to the loss of information and incorrect results. To overcome the disadvantages of locality methods, we develop new adaptive non-local data analysis methods for two kinds of linear systems and apply the methods to two applications: component detection in multiplexed Raman spectra and image restoration. The first application, which was a project sponsored by Renishaw, involves the detection of component dyes from an observed DNA spectrum. The second application is to approximate original scenes from observed images degraded by blurring and noise. In contrast to pointwise methods, our methods investigate patches or neighbourhoods in datasets and extract features from these patches for discrimination and denoising through correlation analysis of the similarity of patches.

In this thesis, our methods are developed by combining non-local data analysis with many popular traditional algorithms, such as model selection with the maximum p-value, conjugate gradient descent and the iteration method of higher convergence order; therefore, our methods are general and can be utilised for further applications to other image processing problems.

ACKNOWLEDGMENTS

This thesis cannot be completed if there is no support and help from many kind people. First, I would like to express my sincere gratitude to my supervisor: Dr. Weiping Lu. Weiping has been encouraging my free exploration and attempt to various methods for spectra analysis and image restoration with great patience and enthusiasm. He also spends much time in guiding me to improve writing from discussing construction of the thesis framework to logic of expression and revision of language. With his help, I finished my appraisal, project reports and presentations. In every discussion, Weiping presents profound knowledge, incisive views and enough experience, which not only directs my study but also benefits my research career in future. In addition, I would like to greatly thank Prof. Rory Duncan and Dr. Colin Rickman. They introduce me to biological knowledge which is important for me to make a better understanding of applications of my research. They also share their research resources with me. The high-performance servers help me computing complicated calculations. The fluorescence microscopy images are practical data that I can use for my research. I also thanks to the staffs in the school of EPS. I am an over-sea student and have inevitable difficulties in study and daily life. They always help me to deal with problems in time. Finally, I am grateful to my family and friends for their constant support and encouragement. When I suffer from loneliness, they accompany with me and chat with me. When I feel dispirited, they comfort me. It is clear that I cannot complete my study without their support and encouragement.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Yuewei Liu		
School:	School of Engineering and Physical Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	Doctor of Philosophy in Physics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*	Yuewei Liu	Date:	09/27/18

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses**)*			
Signature:		Date:	

Contents

ABSTRACT **I**

ACKNOWLEDGMENTS **II**

RESEARCH THESIS SUBMISSION FORM **III**

LIST OF FIGURES **XI**

LIST OF TABLES **XII**

1 Introduction **1**

1.1 Challenges to this study 2
1.2 Non-local methods 4
1.3 Objective of this study 11
1.4 Contributions 12
1.5 Thesis overview 13

2 Review of Related Work **15**

2.1 Model selection for DNA detection 15
2.1.1 AIC and BIC 15
2.1.2 Lack of Fitness 17
2.2 Methods review for image denoising 19
2.2.1 Patch-based methods 19
2.2.2 Transform-based methods 20
2.2.3 Block matching 3D collaborative filter 21
2.3 Methods review for image restoration 23
2.3.1 Iterative algorithms 24
2.3.2 Regularization methods 26
2.3.3 Classical solvers for regularization methods 28
2.4 Summary 34

3 Window-Adjusted Lack of Fitness for DNA Detection **35**

3.1 Description of Renishaw data sets 35
3.1.1 Component dyes and their variabilities 35

3.1.2 Generation of synthetic observations . . . 37
3.2 Window-Adjusted Lack of Fitness . . . 39
3.2.1 Why windows are needed? . . . 40
3.2.2 Determination of windows . . . 45
3.2.3 Naive Bayes classifier . . . 47
3.2.4 New criteria based on Lack of Fitness . . . 49
3.2.5 Optimisation based on naive Bayes . . . 54
3.3 Experiments and results . . . 58
3.4 Summary . . . 59

4 Adaptive Regularised Steepest Descent Method for Image Restoration 61
4.1 Image degradation and common point spread functions . . . 62
4.1.1 Image degradation . . . 62
4.1.2 Common point spread functions . . . 63
4.2 Ill-posed problem . . . 67
4.3 Adaptive regularised steepest descent method . . . 70
4.3.1 Gradient regularization . . . 70
4.3.2 Optimisation of the iterative cost functions . . . 72
4.3.3 Error analysis . . . 75
4.4 Experiments and results . . . 77
4.4.1 Method Implementation . . . 79
4.4.2 Restoration of popular test images . . . 81
4.4.3 Restoration of practical images . . . 86
4.4.4 Experiments for resolution improvement . . . 89
4.4.5 Experiments on the stability of Gaussian blurring . . . 91
4.4.6 Choice of parameters . . . 92
4.5 Summary . . . 94

5 Multi-Frame Image Restoration with ARSD 96
5.1 Multi-frame image restoration . . . 96
5.1.1 What is super-resolution? . . . 97
5.1.2 Method review of super-resolution . . . 99
5.2 Extending ARSD for multi-frame image restoration . . . 100
5.2.1 Restoration with neither shifts nor down-sampling . . . 100
5.2.2 Super-resolution with sub-pixel shifts . . . 101
5.3 Experiments and results . . . 103
5.3.1 Experiments with no shifts and no down-sampling . . . 104
5.3.2 Experiments with sub-pixel shifts . . . 110
5.4 Summary . . . 111

6 Conclusions and Future Research **112**
6.1 Conclusions 112
6.2 Novelty 114
6.3 Future work 115

References **115**

LIST OF FIGURES

1 A demonstration of the ill-posed problem solved by LS. The original image is generated by the MATLAB function "checkerboard". It is degraded by a Gaussian blurring and Gaussian noise, to forms the degraded image. The output image is the solution of the LS approach. 4

2 Demonstration of a detection with the LOF method by Renishaw. These lines represent synthetic observation (blue) and a fitting curve (red) with Blank and Cy3. The two dotted lines stand for the detected components Blank and Cy3. The marked line is the spectrum of ATTO520, which is rejected by the method. 6

3 (a) Residual curve fitted with two components, Blank and Cy3. (b) The missed ATTO520 spectrum by the LOF method. 7

4 Restored images of the cameraman. (a) The original image of the cameraman at a size of 256×256. (b) The blurred and noisy image. The blurring filter is $1/(1+x^2+y^2)$, at a size of 15 pixels. The noise variance is 2 in an 8-bit image. (c) Result via Wiener deconvolution with the signal-to-noise ratio $5.1458e-4$, which is the variance of noise divided by the variance of the original image. (d) Restored image via IDDBM3D. 8

5 Demonstration of windows defined in one-dimensional spectra of residual (a) and ATTO520 (b). 9

6 Demonstration of windows defined in a two-dimensional image of Barbara.jpg. 10

7 Flow chart of LOF method. 18

8 An example to illustrate the principle of transform-based methods.The left is original signal with a single frequency. However, this feature cannot be seen in the middle signal due to the noise. Therefore, the noisy signal is transformed into the frequency domain and the feature of single frequency can be observed. 21

9 Spectrum of Cy3.5 in the wavelength range from 300 to 3800. 36

10 Random spectra of thirteen components dyes. 37

11 Interpretation of variability by 103 spectra of component dye Cy3.5. . 37

12 Diagram of the construction of a synthetic observation spectrum. . . 39
13 Demonstration of a detection with the LOF method by Renishaw. These lines represent synthetic observation (blue), and a fitting curve (red) with Blank,TAMAR,Rhodamine Green and BODIPY_TMRX. The four dotted lines stand for the detected components Blank,TAMAR, Rhodamine and BODIPY_TMRX. The star line is the spectrum of JOE, which is rejected by LOF. 41
14 (a)Residual curve fitted with four components, Blank, TAMRA, Rhodamine Green and BODIPY_TMRX, (b) The spectrum of JOE, which is rejected by the LOF method. 42
15 The blue line represents a synthetic DNA spectrum containing three component dyes of Blank, ATTO520 and BODIPY_TMRX. The red line is a curve fitting to the three component dyes, which are plotted by dotted lines. The spectrum of MAX marked with stars is a false component dye detected by LOF. 42
16 Similarity between the spectrum of ATTO520 (dotted line) and MAX spectrum marked with stars. The marked peaks show positions at which they are close to each other. 43
17 (a) The residual curve generated by fitting the three component dyes of Blank, ATTO520 and BODIPY_TMRX. (b) The spectrum of ATTO520. (c) The residual curve by fitting generated by fitting the three component dyes and the false component dye MAX. (d) The spectrum of MAX, which is recognised as a false ingredient by the LOF method. 44
18 Variability between the fitted spectrum and the true spectrum of ATTO520. 44
19 Illustration of window determination for the window location in a spectrum. 45
20 Demonstration of separating sub-windows and further window region shrinkage. 46
21 Example of window location with the parameter N=7. A sub-peak is located at the X-axis position of 1364, with N1=5. 47
22 Three regions to call WALOF based on LOF. 51
23 Demonstration of choosing the window for WALOF analysis. 53
24 Flow chart of the WALOF method compared to the original LOF method. 55

25 Demonstration of blurring filters on the fragment of image "Cameraman.jpg". (a) The original scene, (b) a blurring by motion filter with $\theta = \pi/4$ and $vT = 9$ pixels, (c) a blurring by uniform filter in size 9×9, (d) a Gaussian type blurring with STD of 1.6, (e) a Gaussian type blurring with STD of 0.5, and (f) a X-ray blurring image with $\xi = 1$. 67
26 Noise amplification error measured by the L_2 norm. 69
27 Demonstration of algorithm (4.25). (a) The original scene and the position of the following fragment images, (b) the original scene, (c) the blurred image, (d) the blurred and noisy image, (e)-(h) D_k in (4.25b) for $k = (85, 425, 765, 1105)$, (i)-(l) I_k in (4.25a) for $k = (85, 425, 765, 1105)$, which are iterative images before applying the noise reduction filter R_N in (4.25b), (m)-(p) I_k in (4.25c) for $k = (85, 425, 765, 1105)$, which are iterative images after applying the noise reduction filter R_N in (4.25b). 80
28 Restored images via four methods. (a) The original scene, (b) the image blurred by scenario 3, (c) the restored image by L0-Abs, (d) the restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method. 83
29 Restored images via four methods. (a) The original scene, (b) the image blurred by scenario 2, (c) the image restored by L0-Abs, (d) the image restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method. 84
30 Restored images via three methods. (a) The full scene of the blurred and noisy image by scenario 5, (b) a fragment contained in the black box of the original scene, (c) the fragment of the blurred and noisy image, (d) the image restored by Fnnmm, (e) the image restored by IDDBM3D, (f) the image restored by our method. 85
31 Restored images via four methods for scenario 7. (a) The original scene, (b) the blurred and noisy image, (c) the image restored by L0-Abs, (d) the image restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method. 86
32 X-ray images of a grating with 4 different widths. 87
33 Restored images via our method. (a) and (b) are two images to be processed. (c) and (d) are the restored images for (a) and (b), respectively. 88
34 Restored images via four methods for the image of a grating with width $0.7\mu m$. 89

35 Restoration of the single bead image. From left to right are the following images: the original scene, the blurred image using Gaussian PSF with STD=2, the blurred image using Gaussian PSF with STD=4, and two restored images from the two blurred images. 90
36 Curves of the centre section of the beads. The black, solid blue, dotted blue, solid red and dotted red curves are the centre sections of images of the original, blurred image using a PSF with STD=2, blurred image using a PSF with STD=4, and the two results from the two blurred images, respectively. 91
37 PSNRs of reconstructions with different PSFs. 92
38 Demonstration of decoupling noise reduction from structure restoration. 93
39 The relationship between the parameter s and the noise strength. The blue curve represents the best value of s for five different noise. The red curve represents a fitting line with a exponent function of the noise STD. 94

40 Flow chart of our weighted model for multi-frame image restoration. . 101
41 A demonstration of shifts of 16 frames. 102
42 An illustration of reconstruction for a high-resolution image. 103
43 Restored images of Cameraman, Scenario 1. (a) Zoomed fragments of the original scene, (b) the blurred and noisy image, (c) the image restored using a single frame, (d) the image restored using the weighted method with 10 frames. 105
44 Restored images of Cameraman, Scenario 1. (a) Zoomed fragments of the original scene, (b) Zoomed fragments of the restored image via averaging method with 10 frames, (c)Zoomed fragments of the restored image via the weighted method with 10 frames. 106
45 Restored images of Cameraman blurred with PSF of Scenario 1 and a mixed noise. (a) The original scene, (b) the blurred image, (c) the blurred and noisy image, (d) the image with the mixed noise, whose type is mixed sparkle noise with a parameter of 0.001 and Gaussian with a standard deviation of 2, (e) the restored image using a single frame, and (f) the restored image using the weighted method with 10 frames. 107
46 The curve of PSNR for different frame numbers. 108
47 Restored images with different numbers of frames, Scenario 1. (a) The original scene, (b) one frame, (c) 10 frames, (d) 20 frames, (e) 30 frames, and (f) 40 frames. 109

48 Reconstruction of a high-resolution image, Scenario 1. The parameters (σ, s) in ARSD are set as $(7.5, 85)$. (a) The original scene (256×256) with the sampling interval 0.5 in $[0, 128]$, (b) the first frame of low-resolution images (128×128), (c) the restored image of the first frame (128×128) via ARSD, and (d) the reconstructed high-resolution image (256×256). 110
49 Reconstruction of a high resolution image from 16 low-resolution frames, Scenario 1. The parameters (σ, s) in ARSD are set as $(7.5, 85)$. (a) Original scene (512×512) with the sampling interval 0.25 in $[0, 128]$, (b) Reconstructed high-resolution image (512×512). 111

LIST OF TABLES

1 Names of the known component dyes 36
2 WALOF vs. LOF for training data. 59
3 WALOF vs. LOF for test data. 59

4 Maximum and minimum eigenvalues of several linear operators. . . . 70
5 PSF and noise variation for each scenario 79
6 PSNRs and SSIMs of the images during iterations in algorithm (4.25). 80
7 PSNRs of the methods in six scenarios 82
8 SSIMs of the four methods in six scenarios 82
9 PSNRs of the three methods in six scenarios 85
10 PSNRs of the four methods for scenario 7. 86

11 PSNRs of multi-frame model(10 frames) vs. single-frame image restoration. 104
12 Restored PSNRs for mixed noise 107
13 Restored PSNRs for different frame numbers 108

Chapter 1

Introduction

Systems of linear equation describe the linear relationships between considered factors and observed measurements [1]. In such a system, each factor is expressed as either a constant or as the product of a constant and an isolated variable. This linear relationship can explain many laws in the world, i.e., certain known component dyes sequences consisting of a compounded chemical via linear a sum [2] and image degradation [3]. Compared to non-linear systems, linear systems are simpler,meaning that there are applications showing complicated laws that are non-linear, but that can be approximated by linear systems [4].

The earliest method to solve linear equations is found in a Chinese text from over 2000 years ago. In 1809, the famous German mathematician Carl Friedrich Gauss recalled this method; subsequently, this method for solving linear system became known as Gauss-Jordan Elimination [1]. The French mathematician Adrien-Marie Legendre proposed the method of least squares (LS) in 1809 [5], which serves to identify a solution to a linear equation via minimising the sum of squares of residuals. This method is important for those applications of requiring data fitting, because data fitting is always an over-determined system in which the number of equations is considerably larger than that of variables.

These traditional methods mentioned above can be applied, but solving linear systems remains a challenge. These methods might fail to address certain linear systems when the linear operators in said systems are uncertain or have an abnormal condition number defined by the division of its maximum and minimal absolute eigenvalues. The uncertain operator does not permit the use such traditional methods because operators must be known when using these methods. The abnormal condition number will lead to complication that sufficiently many solutions will exist that we cannot identify a single solution as our expectation. For example, in the application of DNA detection described by a linear system, a mixed chemical consists of several components from a set of known component dyes, while the individual iden-

tities of components present in the mixture is unknown. Furthermore, intensities of component dyes are also unknown because they change randomly. Therefore, the linear operator is uncertain; such that the linear system cannot be directly solved by LS. Another example of a linear systems is image restoration. Acting as a blurring filter in the space domain, the linear operators in the systems usually have small eigenvalues, which means that the systems have large condition numbers and that they are ill-conditioned. When the inverse of the linear operators is calculated for a solution, a considerable amount of noise will be introduced into the solution [3].

1.1 Challenges to this study

In this study, two types of linear systems are considered. The first kind has a uncertain linear operator. The column vectors are randomly chosen from a set of known vectors, and their intensities also vary randomly. This first type represents the base of methodology for determining predictor variables, which models many important applications summarised into the problem of model selection [6]. The second type is a well-known discrete ill-posed system with an ill-conditioned linear operator [7]. This type is a common model for image restoration employed in many popular applications, such as applications in biomedical [8] and modern information sciences [9].

The two types of systems are both linear but they have different models. For a better understanding of the systems, we introduce them in the context of linking with two applications. For the first type, the problem of DNA detection is introduced, which is to determine the components of DNA Raman spectra through a linear regression model. The aim of the model is to discriminate among each component to demonstrate its presence versus absence in the DNA spectra. Mathematically, this problem can be formulated as

$$Y = SA + N, \tag{1.1}$$

where Y is an observed DNA spectrum, S is the matrix of all component dye spectra, A is the amplitude vector of all components and N is the white noise following normal distribution. In this model, each column of the linear operator S is one component dye. The observed DNA spectrum Y is a column vector that is the linear sum of several components among the set of all component dyes. Present components are indicated by positive amplitudes, whereas absent components are indicated by amplitudes of zero.

The linear operator of S is uncertain in model (1.1). Due to the complexity of imaging systems, the spectra of one component dye always exhibits wide variation

in term of its intensity in the spectra collection procedure. Moreover, the exact spectra of components present in an observed DNA spectrum is unknown for a given detection. Therefore, this approach leads to difficulties in determining the amplitudes of all present components with an unknown matrix S in the model.

For the second type of linear system the problem of image restoration is introduced. This system has a known and non-singular linear operator while the operator has small eigenvalues. The linear system can be applied to approximate many imaging procedures by which an observed image is generated from an original scene degraded by blurring and noise. For example, a camera shake will cause blurring in an output image such that the image of small things, (e.g., cells) often seems to be large discs due to the very small aperture used in a microscope. In addition to blurring, noise always appears as consequence of a certain factor, such as high temperature in transmission equipment or a dirty environment. Such degradation distorts the image of an original scene, which makes it difficult to decision making or further analysis.

Mathematically, a procedure for image degradation can be approximated by

$$J = PI + N, \tag{1.2}$$

where J is the observed image, P is a matrix acting as a blurring filter, I is the original scene, and N is the white noise following normal distribution. For a 2D image, it can be written as a vector J which is the concatenation of its columns. The blurring filter P can be constructed from a known point spread function (PSF) that describes the blurring. This model shows an imaging procedure containing blurring and independent noise.

The simplest way to estimate the real scene from the observation is by minimising the energy function

$$\min \|J - PI\|_2^2 . \tag{1.3}$$

However, this function is a well-known ill-posed problem [7]. When attempting to solve the above energy function, the LS method leads to false solutions because the noise in the output images is infinitely amplified due to the ill-conditioning of P. To explain this ill-posed problem, an example is demonstrated in Figure 1. In this figure, a grid image of size $128 * 128$ is generated on the left using the MATLAB function "checkerboard" as the ground truth. Next, the image is blurred with a Gaussian PSF with standard deviation of 2, and Gaussian noise (40db) is added in the middle image. The result of applying the least square method is presented in the right image. As seen from the output image on the right, no structures can be seen apart from noise. Therefore, further considerations based on LS should be taken

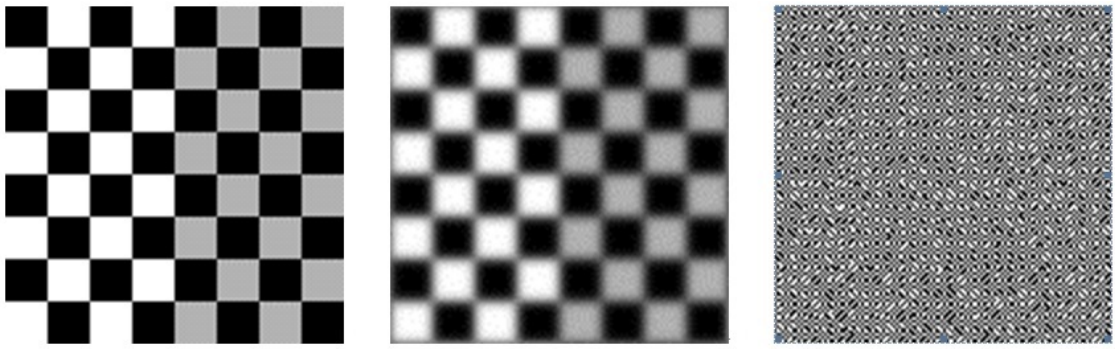


Figure 1: A demonstration of the ill-posed problem solved by LS. The original image is generated by the MATLAB function "checkerboard". It is degraded by a Gaussian blurring and Gaussian noise, to forms the degraded image. The output image is the solution of the LS approach.

into account to remove the illness of the problem. Currently, lots of regularization methods have been developed by introducing penalties to LS.

1.2 Non-local methods

Many methods are developed to solve these linear systems. For model (1.1), the popular methods are Akaike's information criterion (AIC)[10], the Bayesian information criterion (BIC)[11] and Lack of Fitness (LOF)[12]. AIC and BIC are methods based on model selection. Each construct a set of candidate models in which all components can be present or absent, and they subsequently attempt to choose the true model including only components present with information criteria. The LOF method constructs a regression model in which the candidate components are introduced in one by one. A measurement of LOF improvement is defined to evaluate the model. If the LOF improvement is smaller than a give threshold; next, we consider the true model to have been identified.

For the model of image restoration (1.2), regularization methods are most popular. In early research, some classical works include the Wiener filter [13], Tikhonov [14] regularization, total variation (TV) regularization [15], and wavelet shrinkage [16]. Later, non-local methods appeared, such as the non-local mean [17] and shape-adaptive discrete cosine transform [18], as well as BM3D frames and variational image deblurring [19]. These methods construct regularization terms as measurements in spatial domain or in transform domains. These terms act as penalties appended to the least square term and they are stable functional. Therefore, the new energy function constructed with least square and regularization terms can yield a reasonable solution.

However, these methods exhibits shortcomings. As discussed in [20], AIC and BIC methods have weak consistency, which means they are able to select out the true components at probability converging to one when the number of samplers is sufficiently large; however, they are not efficient because huge computations are needed when the candidate set is large. Furthermore, AIC and BIC methods select models by maximising the model information in which a log-likelihood function subtracts a penalty. This penalty depends on parameters in the model, i.e., dependent on the complexity of the model. It is difficult to define this penalty. AIC chooses the value of twice the number of parameter dimensions, and this approach is often blamed for over fitting through the choosing of more components. Because BIC multiplies the parameter dimension by the logarithm of the number of samplers, the penalty will increase quickly and leads to low model information measurements. Therefore, BIC is not amenable to complicated models with more components.

Compared AIC and BIC, the LOF method has high efficiency because it addresses a few candidates but all of the candidate models. However, LOF method often leads to more false detections for trivial components. LOF method constructs stepwise regression models by considering one component at time. During each step, this method chooses the best component in terms of LOF and measures the regression model with LOF improvement. This algorithm will stop until the LOF improvement is lower than a given threshold, at which point the model is thought to be optimised.

LOF improvement measurement does not work well for trivial and false components because similarity between residuals and an investigated component actually not present. When a small residual via a fitting is obtained, the algorithm often stops due to the small LOF improvement by small residual. Thus, the component that is actually present will be missed if it is minimal and contained in the residual. An example is demonstrated in details to show how a present component is missed by LOF. Figure 2 shows a result of detection. A synthetic observation spectrum is generated by three component dyes: Blank, ATTO520 and Cy3. When the LOF algorithm is applied with the parameter of LOF improvement limit valued at 0.10, the component Cy3 is detected successfully, whereas the component ATTO520 is missed. The blur line is a synthetic observation, and the red line is the fitting curve with the component dyes of Blank and Cy3. The detected component dyes are plotted as dotted lines and the missed component ATTO520 is marked with stars.

As seen in Figure 2, the fitting curve is very close to the synthetic spectrum, which shows the residual is already small before the investigated component dye ATTO520 added. The small residual suggests that the remaining dye would be a minimal component. Looking back at the procedure generating this observation, the random amplitudes for these component dyes are 1.0321 for Cy3 and 0.1166 for ATTO520. By the LOF method, Cy3 is detected because it leads to a significant LOF improve-

ments (0.9552). When ATTO520 is investigated, it leads to an LOF improvement of 0.0893, which is less than 0.10, the parameter of LOF improvement limit. Thus, the component ATTO520, which is actually present, is rejected by the LOF method.

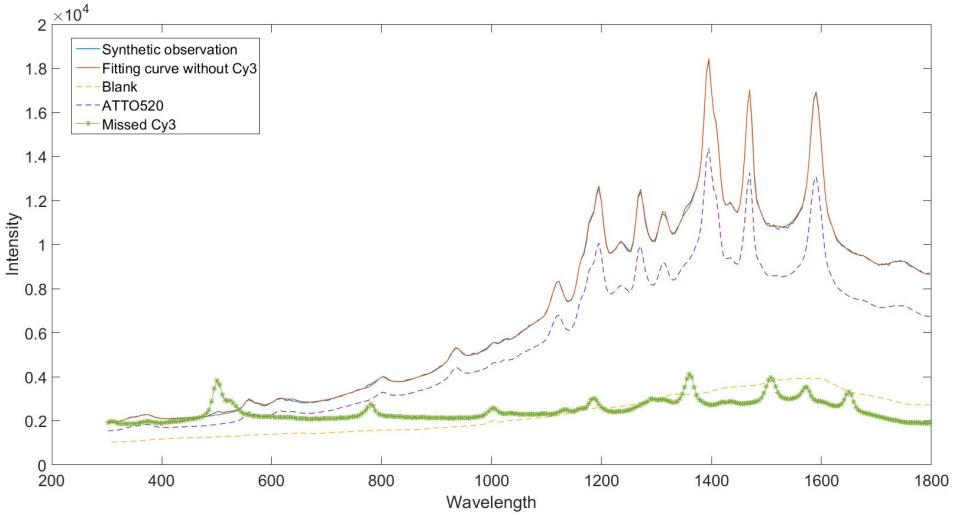


Figure 2: Demonstration of a detection with the LOF method by Renishaw. These lines represent synthetic observation (blue) and a fitting curve (red) with Blank and Cy3. The two dotted lines stand for the detected components Blank and Cy3. The marked line is the spectrum of ATTO520, which is rejected by the method.

Further analysis is implemented on the residual and the investigated component ATTO520 based on the LOF improvement. Figure 3 plots the residual curve and ATTO520 spectrum individually. At the marked locations of 500.5, 774.9 and 1355 wavelengths in this figure, the three peaks in plot (a) looks similar as those in plot (b). We can see the three pairs of peaks located at the very similar or identical wavelengths. Furthermore, the three pairs of peaks have similar shapes in spite of their significant differences in intensity. This information indicates that ATTO520 has a high probability of being a present component.

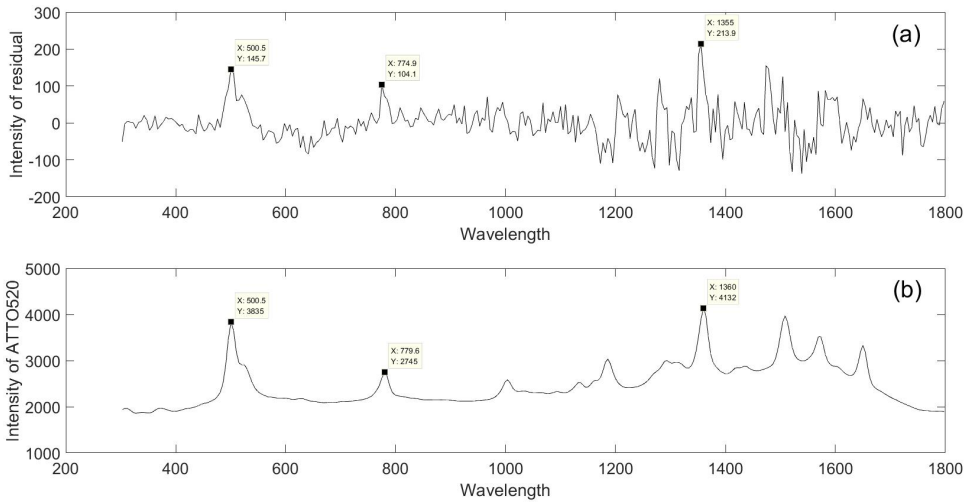
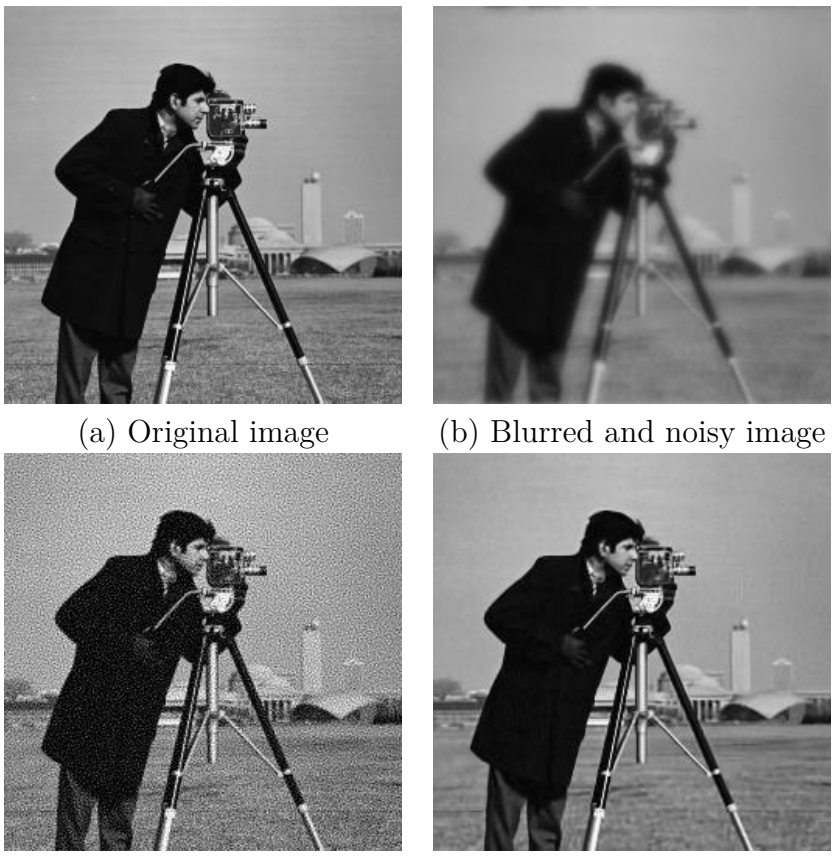


Figure 3: (a) Residual curve fitted with two components, Blank and Cy3. (b) The missed ATTO520 spectrum by the LOF method.

From the above example, the similarity between the residual curve and the investigated component should be considered in measurements. For components that lead to LOF improvements that are a slightly larger or smaller than the LOF improvement limit, we should compare them against residual curves in the wavelength ranges that contain features. This is a direct and efficient way to determine whether the investigated components are present or absent. To this end, the non-local idea would be introduced. However, the LOF method cannot undertake the task of non-local analysis because it models a pure regression point by point.

In the field of image restoration, the appearance of non-local methods yields great improvements in terms of visual effects and the peak signal to noise ratio (PSNR). From visual effects, restored images via non-local methods appear clearer, with less noise and with edges better preserved than those generated via local methods. As seen in Figure 4, both methods, i.e., Wiener deconvolution [21] and IDDBM3D [19], perform well at restoring several small structures, such as pillars in the building, whereas the noise in the output image via Wiener deconvolution is more obvious than that via IDDBM3D. Many local methods have been developed to improved their performance, but they do not work as well as non-local methods. Many results for comparisons can be found in [17], [22], [23], [24], and [19].



(a) Original image (b) Blurred and noisy image

(c) By Wiener deconvolution (d) By IDDBM3D

Figure 4: Restored images of the cameraman. (a) The original image of the cameraman at a size of 256×256. (b) The blurred and noisy image. The blurring filter is $1/(1 + x^2 + y^2)$, at a size of 15 pixels. The noise variance is 2 in an 8-bit image. (c) Result via Wiener deconvolution with the signal-to-noise ratio $5.1458e - 4$, which is the variance of noise divided by the variance of the original image. (d) Restored image via IDDBM3D.

Although non-local methods are popular, there are few definitions of the concept of the non-local operator. We introduce the non-local concept defined in [25]. A non-local operation means that any point can interact directly with any points in a spatial domain. To explain this concept in advance, we would like to use the term of "patch". A "patch" is defined as a neighbourhood centred at one point; thus, it may also be called a window. An example of a window defined in one-dimensional data is shown in Figure 5. In this figure, three windows are defined, respectively, as the intensities of the spectra of residual and ATTO520 in the wavelength ranges [480.9 539.5], [765.4 793.8] and [1347 1373], with window borders marked by red lines.

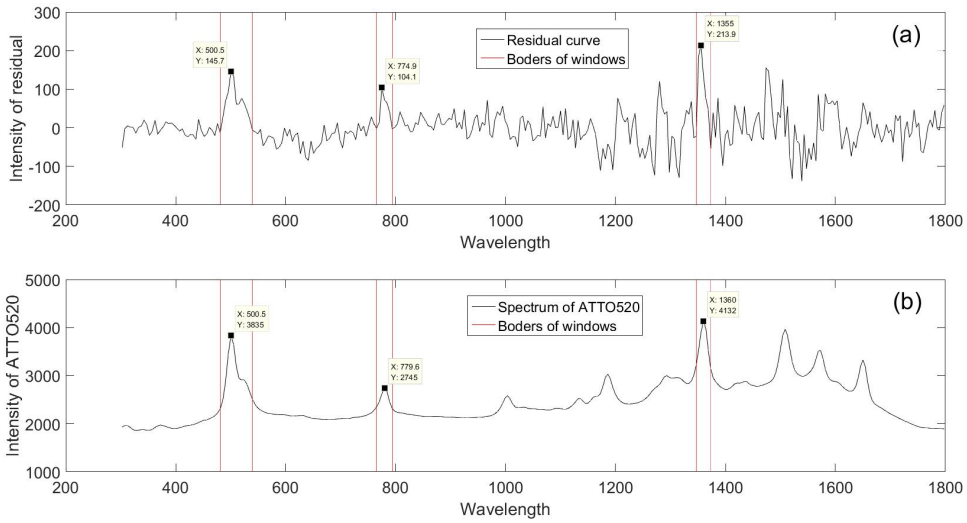


Figure 5: Demonstration of windows defined in one-dimensional spectra of residual (a) and ATTO520 (b).

As for two-dimensional data, such as an image, a patch usually is defined as a square neighbourhood centred at a pixel. Figure 6 shows a 512×512 image of Barbara.jpg. In this image, three patches are marked with black squares at a size of 23×23 pixels. The window size of 23 pixels is just to keep clear watch for patches. Practically, the size of patches is set to a smaller value, e.g., 7, due to the overloaded computation caused by a large patch size.

Figure 6: Demonstration of windows defined in a two-dimensional image of Barbara.jpg.

Based on our knowledge of patches, we explain how patches interact in non-local operators. Interaction between patches are operations that usually aim to measure the similarities among patches for special destinations, e.g., classification or denoising. For example, to determine whether the component ATTO520 is present or absent in a synthetic observation, a comparison is implemented via similarity analysis in the windows 5. Properties of the spectra in the three windows are extracted out, such as correlation coefficient and gradient, and these properties are then input into a system of learning and training to set up a criterion for determination.

An operation for image denoising is often executed through the weighted sum of similar patches. It is well-known that noise at a pixel can be reduced by adding weight to the intensities of several pixels, because the noise will cancel off in the summation. For the intensity of a pixel to be estimated, we choose the pixels centered at the points for which the patches are mostly similar to the patch centered at the pixel and then construct the intensity of this pixel around the weighted average intensities at the chosen pixels. The weights are commonly defined as the distances between the patches at the chosen pixels and the estimated pixel, so it shows similarity between the chosen pixels and the estimated pixel. Higher weight means stronger similarity of the pixels.

Compared to local methods, a great advantage of non-local methods is the improved

accuracy. One main reason is the similarity analysis on the data. Local methods always model using the information of neighbouring points while neglecting relationships between local structures contained in patches. This method is a simple but not powerful approach. In the example of DNA detection, the LOF method often fails to detect minimal component dyes. For image restoration, the local method [26] estimates the intensity of a pixel with the averaged intensity of those pixels whose intensities are equal to that of the estimated pixel. However, the restored image does not appear satisfactory, because of indistinct edges and spotted face. After all, equal intensity does not mean similar structure due to noise.

In summary, the idea behind non-local methods is patch-based analysis and most relevant regions assumed in the data. Those methods mentioned above, such as LOF, Wiener filter and TV, do not touch upon patches and even lack definitions of relevant regions in data; thus, they are local methods. In non-local method, characteristics of local structures are taken into account models as useful knowledge through analysing common features in those regions. To obtain such useful knowledge, many clever ideas are developed for similarity analysis, such as shape-adaptive, low-rank regularization. These ideas not only play important roles in the development of non-local methods, but also enrich theories in the fields of computer vision and pattern recognition.

1.3 Objective of this study

The motivation of this study is to develop non-local methods for two kinds of linear systems. These systems cannot be solved directly using common algorithms such as least square, because their linear operators are uncertain or ill-conditioned. Although there are many quick and efficient methods for these systems, they still exhibits defects. The LOF method is weak at predicting minimal factors in a linear operator, and is loose insofar as it identifies for false factors that are similar to true ones. Truncated VanCittert's algorithm inflates the noise in the solution such that there is a deviation from the solution to the real one. Regularization methods cannot allow error analysis.

To overcome these shortcomings, we develop and combine non-local techniques based on LOF and VanCittert's algorithm, and apply them to the applications of DNA detection and image restoration. In fact, the non-local techniques can be easily combined with other algorithms for linear systems, e.g., AIC,BIC, conjugate gradient (CG), because the non-local techniques are designed independently of those algorithms. Specifically, our non-local methods will accomplish the following tasks:

1. Improving the effectiveness of the LOF method for the detection of minimal

components in DNA spectra;

2. Allowing error analysis in image restoration; and
3. Preserving structures better in the restored images and using them conveniently.

Based on LOF for a linear system (1.1), the proposed method, Window-Adjusted Lack of Fitness (WALOF), focusses on the component leading to LOF improvements that is not significant enough to determine whether the component present or absent. Through comparisons between residual curves and the component spectrum in local areas, a Bayesian classifier is established. If comparisons have great similarity, the Bayesian classifier will deem this component as present at a high probability. On the other hand, if comparisons show no similarity, the classifier will reject this component at a high probability. Due to the further similarity analysis, WALOF leads to fewer false detections than LOF.

As discussed above, the linear system (1.2) has an ill-conditioned linear operator that introduces great noise in the solution when VanCittert's algorithm is applied, which means that the direction of gradient in VanCittert's algorithm fails and a correction should be required. The proposed method, Adaptive Regularized Steepest Descent (ARSD), corrects the distorted gradient directions in VanCittert's iteration by a non-local denoising filter and yields a solution sufficiently close to the real solution. VanCittert's algorithm is powerful in structure reconstruction, thus the result via ARSD often shows the highest Structural Similarity Index (SSIM) compared to many excellent regularization methods. Because of the special formation of VanCittert's iteration, this method allows error analysis, whereas regularization methods cannot. Furthermore, we demonstrate that our method guarantees a solution sufficiently close to the real solution with any denoising filters satisfying certain conditions.

1.4 Contributions

The main contributions of the thesis are outlined as follows:

- As a supplement to the theory of statistical residual analysis, our method, WALOF, achieves a decrease (30%) in the number of false detections compared to the original LOF method proposed in [16].
- Our method extends inverse operator theory to image restoration in the presence of noise.
- Deblurring and denoising are decoupled in our algorithm, which leads to a performance similar to that of state-of-the-art methods and favourable to many

other methods.

- Errors can be estimated in our method, whereas doing so is extremely difficult in regularization methods.
- Our method can be generalized with any well-performing noise reduction filters and multi-frame image restoration models, which can lead to significant improvements of restored image. The parameter configuration is very simple and robust in terms of performance.

1.5 Thesis overview

The non-local methods (WALOF and ARSD) developed for DNA detection and image restoration are patch-based data analysis methods. These methods can improve accuracy in detections and preserve better structures for image restoration through similarity analysis on non-local patches because these patches provide more detailed information on structures than does a single point. In the WALOF method, a new algorithm for feature extraction from one-dimensional spectra is introduced into a regression model for model selection. Concerning image restoration, error estimation can be attained by our ARSD method based on the theory of inverse operators, whereas the error can hardly be analysed using popular regularization methods. The contents of the six chapters include the introduction, a review of related methods, our proposed methods and conclusions.

Chapter 1 introduces challenges in this study, which focuses on how linear systems that cannot be solved by LS. Next, we provide the reason why non-local methods are needed to address these challenges. We attempt to introduce the non-local concept and explain key words in the concept. The advantages of non-local methods are summarized. Section 3 gives our motivation of this study and lists tasks. The idea behind our methods (WALOF and ARSD) is briefly summarised for these tasks. Section 4 outlines the main contributions of the thesis.

Chapter 2 reviews methods related to our work in three sections. The first section introduces popular model selection methods for DNA detection, such as AIC, BIC and the LOF method proposed by Renishaw. The second section reviews methods for image denoising in three categories, patch-based methods, transform-based methods and hybrid methods. The principles of patch-based and transform-based methods are illustrated with simple formulations and examples; next, popular and classical methods are reviewed. In the third category, the BM3D method is introduced in details, which is applied in ARSD as the denoising filter. The last section introduces methods to reconstruct images degraded by blurring and noise. Popular methods of iterative algorithms and regularization are reviewed. Meantime, classical

solvers are introduced for complicated cost functions in regularization methods. The relationship between ASRD and regularization methods can be seen in the steps of these solvers.

Chapter 3 introduces a new non-local method, demonstrating how to detect component dyes in a multi-plex DNA Raman spectrum. We first introduce the project supported by Renishaw and demonstrate the data. Next, the linear system to be solved is reviewed and linked to the data with regard to the problem of DNA detection. For convenience, the LOF method proposed by Renishaw is briefly reviewed. Based on LOF, a new non-local method, called the window-adjusted LOF method (WALOF), is proposed. Compared to the LOF method, our method introduces the concept of 'peaks' as windows in spectra and constructs criteria of discrimination through correlation analysis on features of these windows. We present WALOF in four steps: 1) defining windows for peaks, 2) extracting features from the windows, 3) establishing a dictionary via learning with these features, and 4) calculating the posterior probability for the discriminant. Finally, numerical experiments are presented.

Chapter 4 proposes a new non-local method of ARSD for image restoration. In this method, an iterative energy function is first formulated to transform the ill-posed problem into a well-posed problem using regularized gradients. Then, we optimise this energy function via a combination of VanCittert's algorithm and the block matching 3D collaborative filter. Furthermore, we conclude that the filter can be replaced with any noise reduction filter if the filters satisfy certain conditions. Finally, we compare our method to some regularization methods, and present the numerical results in terms of efficiency and stability measurements.

In Chapter 5, our ARSD method is extended for multi-frame image restoration. At the beginning, super-resolution is introduced for image restoration from multiple low-resolution images, and popular methods are reviewed. Next, we discuss two cases, multiple frames without sub-pixel shift and multiple frames with sub-pixel shift. For the first case, a weighted method is proposed based on ARSD. Due to considerations of similarity of the solutions via ARSD iteration, the weighted method outperforms the simple averaged methods. For the second, we apply the ARSD method to each frame of the low resolution images separately in advance, and thus obtain an estimated low resolution frames with blurring and noise removed. Next, a high-resolution image is reconstructed by an iterative back projection. Finally, the experimental results are presented.

In the final chapter, we present the study's conclusions and also discuss a number of on-going efforts.

Chapter 2

Review of Related Work

2.1 Model selection for DNA detection

Model selection is a term for creating linear regression models to detect known components that consist of an observed DNA spectrum. To predict the ingredients of the DNA spectrum, many regression models are evaluated by combining several components into a guess. The best model is adopted to provide the prediction of the ingredients consisting of the DNA. There are many methods to evaluate guessed models. Several popular methods, AIC, BIC, and LOF, are introduced because they can combine with our non-local technique easily.

2.1.1 AIC and BIC

Akaike information criterion (AIC) is the measurement for model selection among a list of candidates. This criterion aims to measure complexity of a model and fitness with data. To estimate ingredients of a DNA with a suitable model, this method applies Akaike's information criterion to evaluate candidate models. The AIC measurement is defined as [10]

$$\mathrm{AIC}(M) = 2\log \mathcal{L}(\boldsymbol{\beta}) - 2\mathrm{length}(\boldsymbol{\beta}), \tag{2.1}$$

where $\mathcal{L}$ is the likelihood function and $\boldsymbol{\beta}$ is the parameter vector with length of $\mathrm{length}(\boldsymbol{\beta})$ in model M. From this definition, it is known that the AIC measurement is a logarithmic likelihood with a penalty for complexity of the model.

Applied to the model (1.1)

$$Y = SA + N, \tag{2.2}$$

with the assumption that noise distributions in all observations of Y are independent and identically distributed (i.i.d.) and follow a normal distribution with a zero mean

and a variance σ^2, the likelihood can be written as

$$\mathcal{L}(A) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-(Y-SA)^T(Y-SA)/2\sigma^2}. \tag{2.3}$$

The parameter σ, the standard deviation of the noise, can be given a priori or be estimated by the standard deviation of the model residual.

Bayesian methods provide efficient inference for parameters estimation using priori information. Such methods suppose a priori that the data Y have been sampled from a population with unknown parameters $\boldsymbol{\beta}$ which are specified by some random models M. For example, we estimate the distribution of the data with their empirical distribution under the condition of unknown parameters which are random variables from a multivariate normal model M. The priori model can be expressed mathematically as $\text{Prob}(Y|M, \boldsymbol{\beta})$. Based on the priori model, a risk or benefit function is proposed to establish an optimization problem for the determination of parameters in the priori model. For example, the most popular benefit function is the maximum posterior function which can be calculated with the Bayesian formula

$$\text{Prob}(\boldsymbol{\beta}|Y, M) = \frac{M(\boldsymbol{\beta})\text{Prob}(Y|\boldsymbol{\beta})}{\sum_{\boldsymbol{\beta}\in\Theta}\text{Prob}(Y|\boldsymbol{\beta})}, \tag{2.4}$$

where Θ is the space of the parameters.

However, the calculation of the posterior function is a very challenging problem because the denominator in (2.4) is often a complicated integral. When the priori distribution in the model is not assumed to be Gaussian, the maximum posterior function cannot be attained analytically and it is often approximated. The BIC method has been widely used to solve this problem. That method is an asymptotic expansion of the quantity, $\log\int_{\boldsymbol{\beta}\in\Theta}\text{Prob}(Y, \boldsymbol{\beta}|M)\text{d}\boldsymbol{\beta}$, to define the BIC [11],

$$\text{BIC}(M) = -2\log\text{Prob}(Y|M) \approx -2\text{Prob}(Y|\hat{\boldsymbol{\beta}}, M) - (\log n)\text{length}(\boldsymbol{\beta}), \tag{2.5}$$

where, $\hat{\boldsymbol{\beta}}$ is the estimation of the parameter vector and n is the size of sampling Y. Due to the complicated posterior function, the estimation is often calculated via Monte Carlo estimations based on the Gibbs sampler [27].

In addition to AIC and BIC, other methods based on likelihood are non-negative matrix factorization NMF [28] [29]. Since spectra always take positive values, the matrix of observed DNA spectra is called a "non-negative matrix". Different from model selection, the key idea behind this method is to decompose the matrix into two factors which represent the amplitude matrix and the ingredient matrix. The theory of NMF can be observed in the book [30]. However, the NMF method is seldom undertaken on practical tasks because most applications are interested in estimating the values of the individual A and the columns of S from a known set.

2.1.2 Lack of Fitness

Bayesian methods are powerful methods for component detection but they must use posterior probabilities, which are complicated to compute. Compared to Bayesian methods, other methods are simple and reliable. In one of the most important-namely, the multivariate linear regression models known as model selection, only one component is determined at a time with criteria for the model. Usually, minimising residuals or minimising intersection angles is chosen as the criteria for determining the best components in the model. A regression model, LOF analysis [12], was proposed by Renishaw to discriminate between the presence and absence of spectra in an observation based on minimizing residuals. The underlying technique is based on a direct classical least squares (DCLS) analysis of the unknown spectrum; therefore, the technique minimises the sum of the squared deviations in model (1.1).

In the method, the DCLS is applied iteratively to the observation spectrum to resolve the components one at a time in order of decreasing significance. All available components are considered, and separately added to the existing solution. The best component is selected such that it produces the greatest improvement in the fit, as measured by the lowest Lack of Fit (LOF), which is defined as follows:

$$LOF = \frac{\sum_{i=1}^{N}(Y_i - \sum_{k=1}^{M} S_{i,k}A_k)^2}{\sum_{i=1}^{N} Y_i^2}, \tag{2.6}$$

where Y_i is the ith element of the observed DNA spectrum Y, $S_{i,k}$ is the ith element of the kth component dye spectrum assumed present in the component dye matrix, A_k is the kth element of the amplitude vector A, N is the length of the DNA spectrum and M is the number of components assumed in the model. Known from (2.6), the value of LOF is defined as the relative mean square error.

After each iteration step, the solution is inspected to determine whether the iteration has terminated. Termination is considered to have occurred if one of the following conditions is satisfied:

1. The new component has a negative amplitude (the solution is non-physical).
2. The improvement in the LOF is below a pre-set limit, indicating that the improvement of the fit is not significant. The most effective implementation of this limit has been found to be as a proportional/percentage improvement in the LOF. The limit value is latter observed to be relatively independent of data quality or pre-processing options.
3. (Optional)The amplitude of the added component is below a pre-determined limit. This secondary limit is used to allow a limit to be imposed on the concentration of material present in the sample that shall be judged to be

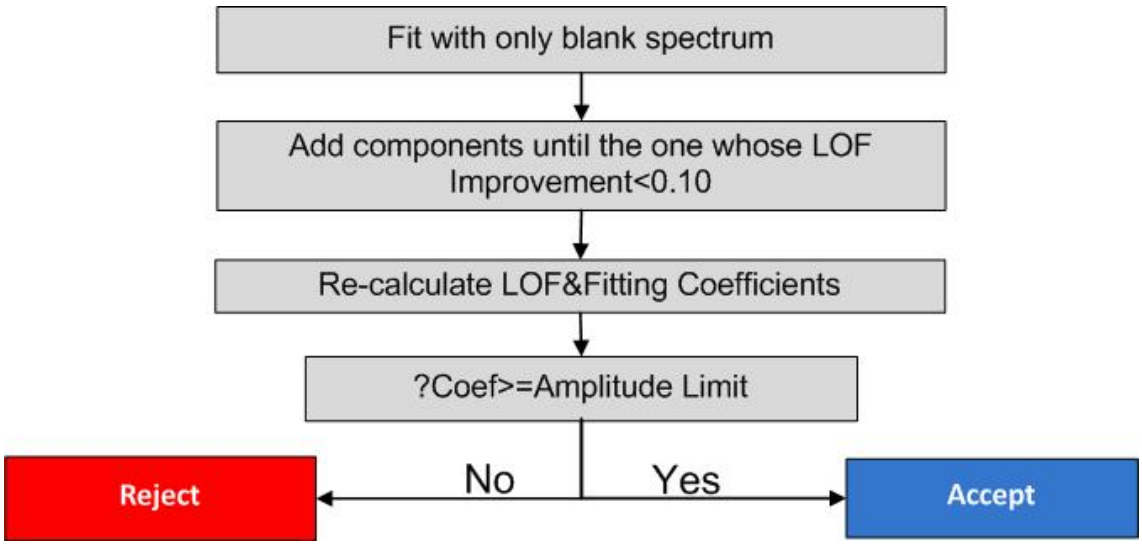


Figure 7: Flow chart of LOF method.

significant. This condition is not included in the code provided by Renishaw.

In each iteration, all of the components previously identified as significant to the solution, together with one additional candidate component, are refitted based on minimising the deviation. Figure 7 illustrates this method.

There are several advantages to this LOF method. First, DCLS is a simple and reliable method for minimising deviations. When seeded with all component spectra, the method will, in general, produce a non-zero solution for the amplitudes of each component. The dominant amplitudes in this solution correspond to components that are present in the sample. When all components are present at high concentrations, it is a simple matter to discriminate between components that are present and those that are not. Second, DCLS always returns the best solution in terms of mean square error. Non-zero amplitudes for minor components, which are not present in the sample, will be admitted at some level by noise on the spectra. Third, the amplitudes of such spurious minor components can be significant compared to the amplitudes of trace components that the technique is required to detect.

However, their effect on the LOF tends to be far less significant, i.e., a very small improvement in the LOF can lead to significant component amplitude. The reason for this is that, although the fit can be improved in some spectral regions, it will be worsened in others due to the component spectra being unique. A limit on the LOF therefore tends to be more effective than a limit on component amplitude in defining which components are actually present in a sample.

2.2 Methods review for image denoising

In the procedure of imaging, noise always plays a central role. A low level of noise is typical in output images in many fields such as cell molecular biology, medical science and geology. For example, to observe distinct structures in cells, noise appearing in output images must be kept at a very low level [31]. There are also similar standards of noise control required in medical science imaging [32]. Textures with special shapes on the surfaces of stones are studied to determine the locations of energy. If noise cannot be suppressed to a low level, the textures will become too vague to provide credible evidence for efficient exploration [33].

2.2.1 Patch-based methods

The basic principle for noise reduction in patch-based methods derives from statistics. The variance of an averaged sampler from a population is smaller than that of the population. More specifically, supposing that random variables $X_i, i = 1, 2, \ldots, n$ are i.i.d. and that they are sampled from a population with mean μ and variance σ^2, we have the following result:

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sigma^2 \tag{2.7}$$

This principle tells us that if a population has a variance of σ^2, then the average of i.i.d. samples from the population has a variance n times less than σ^2.

We discuss how to apply this principle to image denoising. With the assumption of additive noise, the imaging model is defined as

$$J(x, y) = I(x, y) + N \tag{2.8}$$

where J is the observed image and I is the original scene, and N denotes a noise independent of the original scene, with a mean of zero and a variance of σ^2. Applying model (2.8) to multi-frame images taken n times in a fixed position for the same scene, we obtain a series of output images $J_k(x, y) = I(x, y) + N_k$, for $k = 1, 2, \ldots, n$. The average of $J_k(x, y)$ will be the output

$$\frac{1}{n}\sum_{k=1}^{n} J_k(x, y) = I(x, y) + \frac{1}{n}\sum_{k=1}^{n} N_k. \tag{2.9}$$

Using the principle (2.7) we know that the variance of the noise in the averaged image is only σ^2/n. If n is sufficiently large, the noise in the output will be reduced to zero.

In practical situations multiple-frame images of the same scene cannot be obtained. For example, the scene in images taken by a spacecraft will be shifted and rotated because the spacecraft is always moving. However, principle (2.7) can be applied using similar methods known as neighbourhood methods. These methods add intensities via weights at pixels whose intensities most closely resemble that of the investigated pixel to reduce noise. However, because the intensities are summed at the pixels in different structures, this type of method always introduces discrepancies in the structures. A better method is to extend the idea of neighbourhood methods to patch-based methods which investigate patches and not pixels. Here, we discuss two classical patch-based methods: non-local means (NLM) filter.

The original research on NLM can be found in [17], which introduces a new conception of patches, namely, a neighbouring area centred at the local pixels. The NLM algorithm estimates the intensity of a local pixel with the total intensity of some pixels using weights as in neighbourhood methods, although the pixels and the determination of the weights are different. The pixels are centres of patches that are similar to the investigated patch centred at the local pixel, and the weights are valued with some similarity measurements. Usually, the greater the similarity between patches, the larger the weight that is adopted. The advantage of this method is that it can better preserve structures because a local patch can interpret structures in greater detail than a local pixel. Mathematically, the NLM filter can be represented as,

$$\hat{I}(x_p, y_p) = \sum_{(x,y)\in I} \exp\left(-\frac{\left\|P_{x,y} - P_{x_p,y_p}\right\|_2^2}{h^2}\right) J(x,y), \tag{2.10}$$

where $P_{x,y}$ represents the patch centred at pixel (x, y), and h is the parameter that describes the decay speed of the similarity between patches.

2.2.2 Transform-based methods

Transform-based methods were first developed in [16]. These methods first transform the observed image from the spatial (or time) domain into other domains such as the frequency, wavelet, and principal component Analysis (PCA) domains; and then, they reduce noise through a shrinkage function. Finally, they restore the image with the inverse transform. The potential principle of transform-based methods is that structures can be represented sparsely in certain domains in which random noise cannot. Mathematically, structures are linearly correlated with only some of bases in the transformed domain and in other bases the coefficients are almost zeros; thus, the energy becomes concentrated in a few bases. Due to energy conservation signal-to-noise ratio(SNR) will increase dramatically in these bases, leading to easy noise suppression via shrinkage. Figure 8 provides a simple illustration. In this example,

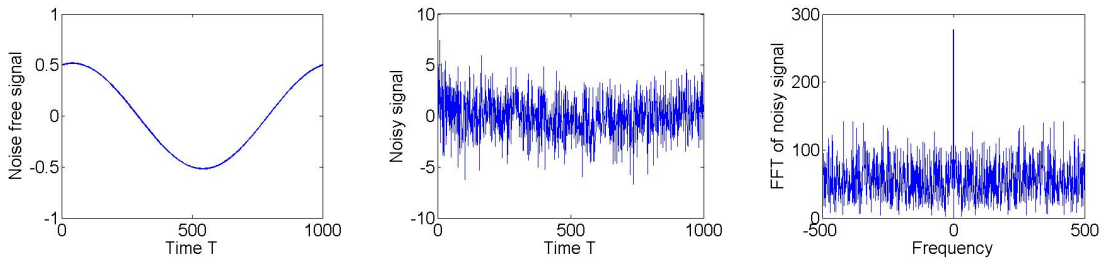


Figure 8: An example to illustrate the principle of transform-based methods.The left is original signal with a single frequency. However, this feature cannot be seen in the middle signal due to the noise. Therefore, the noisy signal is transformed into the frequency domain and the feature of single frequency can be observed.

the real signal (the left curve) is,

$$y = \sin\left(\frac{2\pi}{1000}T\right) + \cos\left(\frac{2\pi}{1000}T + \frac{\pi}{3}\right), \tag{2.11}$$

Gaussian noise with standard deviation of 0.6 is added (the middle curve). The right curve is a transformed signal in the frequency domain. As shown in Figure 1, there is a distinct difference between the real signal and the noisy signal due to the strong noise. The real signal cannot be guessed from the time domain. However, when noisy signal is represented in the frequency domain, we can see clearly the energy concentrating on the first frequency,whereas noise level far smaller. Therefore, noise reduction in the frequency domain becomes easier than in the time domain.

There are three steps included in transform-based methods. The first step is to transform the observed image from the spatial (or time) domain into a domain in which the image can be represented sparsely. Next, through the definition of a function, shrinking is applied to coefficients in this new domain. The last step is to restore the image as an estimation of the original scene with the inverse transform from the new domain into the spatial (or time) domain. The most important step of transform-based methods is to define rules for coefficients shrinking. Common shrinking rules, such as the RiskShrink and VisuShrink rule, are discussed with hard and soft thresholds in the initial research [16].

2.2.3 Block matching 3D collaborative filter

Recently, on the basis of transform-based methods, many hybrid methods have been developed to improve performance on noise reduction through combinations of patch-based and transform-based methods. Usually, these methods construct shrinking rules adaptively. The Bayesian framework and statistical models are popular for shrinking rules. In [34] an adaptive, data-driven threshold was proposed for image denoising via wavelet soft threshold. The threshold is derived within a

Bayesian framework, and the prior is assumed as the generalised Gaussian distribution (GGD); the author called their method BayesShrink. That method uses wavelet transforms and outperformed SureShrink. The article [35] presented an example of a statistical method that uses scale mixtures of Gaussians in the wavelet domain. That method is based on the coefficients of an over-complete multi-scale-oriented basis. In [36], PCA was applied for finding similar patches instead of the Euclidean norm. Three algorithms were introduced: local PCA, hierarchical PCA and global PCA. Because similarity-based methods apply photometric similarities that cannot represent actual geometrical similarities, [37] exploited a non-local patch-based Wiener filter, the PLOW filter, for both geometrically and photometrically similar patches in terms of optimal linear minimum mean square error. Patch-based methods often suffer from huge computational loads because they find similar patches in large images. The article [38] presented a fast patch-based denoising algorithm. That algorithm is based on patch geodesic paths, which treats image patches as nodes and patch differences as edge weights for computing the shortest paths. Other excellent research on denoising can be found in the works on FOE [39], K-SVD [40], NLSM [41] and [42]. Although, some of these methods are regarded as dictionary-based, they do combine transform-based and patch-based principles.

Among these hybrid methods, we introduce the algorithm of BM3D [22] because it is the state-of-the-art and is applied in our method. This method is based on an enhanced sparse representation in the transform domain, instead of the representation of the observed image itself. The observed image is regarded as a set of patches and all the 2D image patches are put into a 3D data structure based on their similarity. Next, a collaborative filter is defined to address these 3D data through three steps: 3D transformation, shrinkage of the transform spectrum and inverse 3D transformation. Finally, a final estimation of the true image is computed by aggregation, which averages the local intensities at overlap pixels in adjacent patches. We introduce the BM3D algorithm in detail.

In the first stage, a distance between patches in the observed image J is defined as

$$d(P_{J(\boldsymbol{x})}, P_{J(\boldsymbol{y})}) = \frac{\left\| P_{J(\boldsymbol{x})} - P_{J(\boldsymbol{y})} \right\|_2^2}{\omega^2} \tag{2.12}$$

to measure similarity of two blocks, where $P_{\cdot}$ represents the patch centred at the pixel $\cdot$, ω represents the size of the patch, and $\|\cdot\|_2^2$ represents the Frobenius norm. However, distance (2.12) is not suitable to measure similarities of noisy patches, because the probability densities of the distances are likely to overlap heavily for large noise or small patch size. To avoid this problem, a filter is performed on the distance to construct a new distance in advance via normalised linear transform T_N

and shrinking rule γ_μ on these similar patches. The new distance is defined as

$$d_{\text{new}}(P_{J(\boldsymbol{x})}, P_{J(\boldsymbol{y})}) = \frac{\left\|\gamma_\mu T_N P_{J(\boldsymbol{x})} - \gamma_\mu T_N P_{J(\boldsymbol{y})}\right\|_2^2}{\omega^2} \tag{2.13}$$

where γ_μ is a hard threshold operator with threshold μ and T_N is a normalised linear transform. With the concept of the new distance, a rule for patch grouping is given using a constant τ to collect the coordinates of similar patches in the set

$$S_{\boldsymbol{x}} = \{\boldsymbol{y} \in \Omega_{\boldsymbol{x}} : d_{\text{new}}(P_{J(\boldsymbol{x})}, P_{J(\boldsymbol{y})}) \leq \tau\}, \tag{2.14}$$

where $\Omega_{\boldsymbol{x}}$ denotes a neighbourhood around the pixel $\boldsymbol{x}$ in which the block-matching is implemented. The set $S_{\boldsymbol{x}}$ collects pixels centred whose patches are similar to the patch centred at pixel $\boldsymbol{x}$. Formula (2.14) is calculated for all pixels in the noisy image, which is the procedure of blocking-match. Known from (2.14), blocking-match leads to huge computation for images with large size. Therefore, this operation is often implemented in a part of the image instead of the whole image.

After blocking-match, all patches are collected into groups each of which contains similar patches measured with the distance (2.13). For all patches in a group $S(\boldsymbol{x})$, we denote them as $\mathcal{P}_{S(\boldsymbol{x})} = \{P_x : \ x \in S(\boldsymbol{x})\}$. These patches are stacked into a 3D array. Letting $\left|\mathcal{P}_{S(\boldsymbol{x})}\right|$ be the number of patches in the set $\mathcal{P}_{S(\boldsymbol{x})}$, the 3D array has a size $\omega \times \omega \times \left|\mathcal{P}_{S(\boldsymbol{x})}\right|$. As discussed in the paper, there is no order of patches in the Frobenius norm-stacked 3D array.

Having the 3D array, a transform will be applied to this 3D array, therein projecting this 3D array onto a set of bases, i.e., the discrete cosine transform (DCT). Then, Wiener shrinking is applied to the coefficients in the DCT domain and the inverse transform of the DCT is used to restore the patches in the estimated image. Finally, all patches are aggregated into an output image.

2.3 Methods review for image restoration

Blurring is another factor that decreases the quality of observed images. Blurring effects are generally a bandwidth limit on an image during imaging processes as a result of different causes. For example, due to optical diffraction, the image of a point source is diffused to an Airy disk via an optical microscope. When an image is recorded by a remote sensor through the atmosphere, blur can be caused by atmospheric turbulence. Moreover, blur can also be induced by the movement of objects or cameras, referred to as motion-induced blur. PSFs are commonly used to describe blurring effects. Although there are many causes of image blurring, their

effects can be approximated by the linear system

$$J = PI + N, \tag{2.15}$$

where I is the original scene, P is the PSF acting as a blurring filter, N is noise assumed to be normally distributed with a mean of zero and a variance of σ^2, and J is the observed image degraded by blurring and noise. As discussed in Chapter 1, the linear operator P is invertible, but it is ill-conditioned due to its small eigenvalues.

2.3.1 Iterative algorithms

Iterative techniques have been developed for image restoration in recent years [43, 44, 45, 46, 47, 48, 49]. The earliest work goes back to VanCittert's iteration in 1931 which was introduced in [3]. This algorithm has a long history as a simple and efficient approach for image restoration. The algorithm attempts to define the perfect inverse filter for the convolution filter. The algorithm is formulated for spatially invariant or variant restoration problems with neglect of noise contribution in (2.15). Originally, it is a steepest descent method to calculate the solution by

$$I_{k+1} = I_k + \beta(J - PI_k), \tag{2.16}$$

where β is the step parameter. For a certain β, the solution converges to the original scene. However, when the blur PSFs have both positive and negative eigenvalues there are no real values of the step parameter that can make the iteration converge. To overcome this shortcoming, an iterative procedure was proposed [50],

$$I_k = I_{k-1} + \beta P^T(J - PI_{k-1}), \tag{2.17}$$

where P^T is the transpose of P. This iterative solution converges to the ground truth only if the noise in an observation is negligible. When an observation includes noise, the VanCittert's iteration (2.17) can be expressed as[3][51][52]

$$\begin{aligned} I_k = &\sum_{u,v}(1 - (1 - \beta|\zeta_{uv}|^2)^k)\langle I, Z_{uv}\rangle Z_{uv} \\ &+ \sum_{u,v}\frac{1}{\zeta_{uv}}(1 - (1 - \beta|\zeta_{uv}|^2)^k)\langle N, Z_{uv}\rangle Z_{uv} \quad \text{for } k = 1, 2, \ldots, \end{aligned} \tag{2.18}$$

where $\langle\cdot,\cdot\rangle$ is the inner product, $\{\zeta_{uv} : u = 1, 2 \ldots, R, v = 1, 2, \ldots, C\}$ and Z_{uv} are the respective eigenvalues and eigenvectors of P, R and C the numbers of rows and columns of the image respectively, and (u, v) is the position of a pixel. Because the original scene I is a column vector of size $R \times C$, the linear operator P is a square matrix of size $RC \times RC$, and the number of all its eigenvalues is RC.

The first term involving I describes structure recovery, whereas the second term, involving N, shows noise amplification, so structures and noise are separated in (2.18). For a noisy observation, however, small eigenvalues ζ_{uv} can lead to significant noise amplification in the second term of (2.18), so the inverse problem becomes ill-posed. This explains why noise is amplified when LS methods are directly applied to solve the ill-conditioned linear system.

There are two ways for VanCittert's algorithm to solve the ill-conditioned linear system. One is to truncate the iteration when noise is acceptable. The truncated solution is often weak for PSFs, which leads to a slow speed of structure recovery. Another is to consider regularization in following iteration

$$I_{k+1} = (E - \lambda^2 \beta L^T L) I_k + \beta (J - P I_k), \tag{2.19}$$

where E is the identity matrix, L is a regularization operator and λ^2 is the regularization factor. Similar to (2.18), the iterative solution can be rewritten as

$$I_k = \sum_{u,v} \frac{\zeta_{uv}}{\zeta_{uv}^2 + \lambda^2 \gamma_{uv}^2} \left(1 - \left(1 - \beta(|\zeta_{uv}|^2 + \lambda^2 \left|\gamma_{uv}^2\right|)\right)^k\right) \langle J, Z_{uv} \rangle Z_{uv}, \tag{2.20}$$

where γ_{uv} are the eigenvalues of the regularization operator. Because the factor

$$\frac{\zeta_{uv}}{\zeta_{uv}^2 + \lambda^2 \gamma_{uv}^2}, \tag{2.21}$$

is always between $(-1, 1)$, noise in J is reduced. However, the structures are not as well recovered as in (2.18). The balance between noise suppression and structure restoration depends on the regularization factor λ^2. A smaller factor means more powerful noise suppression and poor structure restoration. In contrast, a bigger factor means poor noise suppression and better structure restoration.

The sufficient condition of convergence of iteration (2.18) is

$$\left|1 - \beta \left|\zeta_{uv}\right|^2\right| < 1, \quad \forall\ u, v, \tag{2.22}$$

which is equivalent to

$$0 < \beta < \frac{2}{\max_{u,v} |\zeta_{uv}|}. \tag{2.23}$$

For most of PSFs that act as low pass filters, the maximum eigenvalue is approximately 1, so the step parameter is easily set at $\beta = 1$.

With regard to the speed of convergence, two measurements are usually used to quantify an iterative algorithm: the convergence order and the convergence rate. If

the error during iteration can be given as

$$\|I_{k+1} - I_\infty\| \approx \kappa \|I_k - I_\infty\|^n, \quad n \geq 1, \tag{2.24}$$

then κ is called the convergence rate and n is the convergence order. For the case of $n = 1$, we say that the algorithm has a linear convergence speed. According to this definition, VanCittert's iteration leads to a linear convergence speed, while its convergence rate of $|1 - \beta\zeta_{\min}|$ depends on $\zeta_{\min}$, the smallest eigenvalue of the PSF. In general, smaller eigenvalues lead to lower convergence speeds.

2.3.2 Regularization methods

Regularization methods attempt to address ill-posed linear systems by introducing stable functional terms as penalties into the residual analysis. Compared to the two above mentioned methods, these methods avoid unbounded noise and present flexibility in the choice of regularization methods, such as BM3D and low-rank priories, which are highly complicated. However, in these complicated methods, determination of regularization factors mostly depends on experience because there is little theory underlying this challenging problem. Due to the complexity of the regularization factor, error analysis is difficult to apply.

In the last several decades, although there has been substantial work on regularization methods, it can be stated that the majority of work has been focused on two important factors: regularization factors and the form of regularizations. To determine good regularization factors in Tikhonov models, Golub, Heath and Wahba proposed a classical method, generalized cross-validation(GCV) [53], to balance structures and noise. Later, Hansen provided two improved criteria, the L-curve [54] and the normalized cumulative periodogram (NCP) [55], in 1992 and 2006, respectively.

There are numerous methods that model regularizations with local information. An early and well-known method is the standard Tikhonov regularization with the penalty term $\|EI\|_2^2$. In addition to the identity matrix, the gradient and Laplacian operators are also popular choices for the regularization operator. However, these operators are suitable only for smooth images because they force spatial smoothness. Considering the over-smoothed edges that are often caused by L_2 norm regularization which is defined as the square sum of the elements in a vector, TV [15] has been proposed as a classical method to better preserve edges; this regularization introduces a penalty with L_1 norm of the first derivative, where L_1 norm is defined as the sum of the absolute elements in a vector. However, L_1 norm regularization suffers from other issues: for one, the solution is not unique because the L_1 norm is not strictly convex. In this paper, a nonlinear-partial-differential-equations-based algorithm is applied to find a solution.

A well-known algorithm is developed to solve TV regularization problems in [56]. That method proposes a splitting algorithm based on Bregman iterations; thus, it is referred to as the Split Bregman method. Similar methods include the Douglas-Rachford splitting (DRS) algorithm [57] and the alternating direction of multiplier method (ADMM) [58]. One disadvantage to these methods is their time-consuming nature due to the nonlinear equations involved. To improve efficiency, some linearisation techniques [59] and primal-dual framework methods [60] have been developed. Very recently, a new class of fixed point algorithms [61] [62] were proposed and shown to be more efficient than operator splitting methods.

Although TV regularization methods can better preserve edges, they introduce undesired artefacts in what is known as the staircase effect. The staircase effect includes redundant information in term of piecewise-constants. To address this shortcoming, sparsity-based methods are widely used in many image restoration applications. These methods assume that the observed images have sparse representations in certain transformed domains such as the frequency or wavelet domains. Therefore, regularizations with the norm of transformed spectra are considered in energy functions in addition to TV terms. Through the sparse representation of observed images in the transformed domain, redundant information is eliminated for constructing sparse solutions. Compared to TV regularization, these methods can restore more details of the structures. Several excellent studies are described in [63][64][65].

Since NLM [17] was first developed for noise reduction in 2005, numerous non-local operators referred to as patch-based regularizations have also been applied to image restoration. A generalised derivative operator, which can be viewed as a non-local extension of TV, has been applied for regularization [66]. However, this method does not consider the dependencies between sparsely coded patches. To overcome this problem, the simultaneous sparse coding (SSC) method extends a non-local sparse representation to learning dictionaries, therein combining non-local mean approaches [41]. In this method, the observed image is first decomposed with features, and these features are grouped by exploiting self-similarities. Then, an energy function is proposed through a combination of sparse decomposition problems, which are formulated with dictionaries and non-local regularization terms corresponding to the groups. Another method is the classic IDDBM3D method, which considers similarities of patches with a simplified BM3D operator. In contrast to SSC method, IDDBM3D applies spatial residuals as a fidelity term, but the residuals are in the transformed domain.

Recent research on non-local regularization has focused on low-rank approaches. In [67], a method is proposed to remove mixed noise from videos with minimal assumptions on the statistical distributions of the noise. In this paper, the denoising problem is converted into the problem of a low-rank matrix completion, which can

be solved by minimising the nuclear norm of the matrix with linear constraints. The principle is that a similarly packed patch matrix is re-arranged, and the rank of this matrix should be lower if there is minimal noise. In [68], a low-rank prior was used for regularizations in terms of a pseudo-matrix norm. The authors of this paper also attempted to determine why non-local regularizations, such as SSC and IDDBM3D, perform better than do conventional sparse coding for the class of natural images. In their paper, the researchers noted the reason that the singular value decomposition of similar packed patches can be viewed as extracting both local and non-local information for estimating signal variances by a low-rank approach towards SSC and explained via bilateral variance estimation. Another application of low rank as regularization can be found in [69], which regularises penalties with the nuclear norm of low-rank patch matrix and its gradient and the L_2 norm of PSF matrix for blind deblurring. These non-local regularizations improve performances of the methods, although the optimization of the energy functions becomes more difficult with increasing numbers of regularizations and more complicated regularization forms.

2.3.3 Classical solvers for regularization methods

Wiener deconvolution

The Wiener filter attempts to minimise the residual, i.e., the mean square error between the original image and the restoration. This filter is a statistical method used to compute an estimation of the unknown image with the assumption of stable additive noise and is commonly used to reduce noise and deconvolution.

In mathematics, the Wiener deconvolution filter is applied in the frequency domain; therefore, it can be considered as a transform-based method. In the frequency domain, the SNR is given by regularization algorithms, and the original image is estimated with the minimum mean square error (MMSE). The Wiener filter is defined by minimising the following energy function:

$$F_w = \min_{\mathcal{F}} \|\mathcal{F}J - I\|_2^2 . \tag{2.25}$$

Using the rule of the Fourier transform $F(x \bigotimes y) = F(x)F(y)$, where $\bigotimes$ is a convolution operator, the solution of this energy function can be deduced as

$$F_w = F^{-1}\left(\frac{F^{conj}(P)S(F(I))}{|F(P)|^2 S(F(I)) + S(F(N))}F(J)\right), \tag{2.26}$$

where

- $F(I)$ is the Fourier transform of the original scene I, $F(N)$ is the Fourier

transform of noise and $F(P)$ is the Fourier transform of the blurring filter;

- $S(F(I))$ is the mean power of $F(I)$ and $S(F(N))$ is the mean power of $F(N)$;
- The notation *conj* denotes conjugation.

The solution of (2.26), contains a term $S(F(I))$, the mean powers of the Fourier transform of the original image and noise which are unknown. For this, we rewrite this solution to provide a better interpretation:

$$F_w = F^{-1}\left(\frac{F^{conj}(P)}{|F(P)|^2 + 1/SNR_F} F(J)\right), \tag{2.27}$$

where $SNR_F = S(F(I))/S(F(N))$ is the SNR in the frequency domain, which is the regularization.

Because the Wiener filter aims to find a minimum least square solution with regularization, its solution is close to the least square solution. Considering that there is a priori knowledge Γ for the SNR, this ill-posed problem is then regularised based on the mean square error criterion with a priori knowledge Γ, and the solution is [3]

$$\hat{I} = F_w J = (P^T P + \Sigma\Gamma^{-1})^{-1} P^T J, \tag{2.28}$$

where Σ is the correlation matrix of the noise. This solution is identical to the results of the MAP estimator when noise is assumed to take on a multivariable Gaussian distribution. If there is no noise contamination considered in the blurred image, which is to say that Σ is equal to zero, the solution reduces to the least square solution.

Solver for L_2 norm regularization

The famous regularization with the L_2 norm is Tikhonov regularization [14]. This method was first studied by Tikhonov and Arsenin, hence its being named for the author. The general energy function is defined as

$$E(I) = \|J - PI\|_2^2 + \lambda^2 \|LI\|_2^2, \tag{2.29}$$

where λ^2 is a regularization factor and L is a linear operator for regularization. This method regularises the ill-posed problem into a well-posed problem through the introduction of a penalty term measured with the norm. The linear operator is chosen as a stable functional for the penalty. For example, the identity matrix is referred to as standard Tikhonov regularization. Based on the Lagrange multiplier method, this method optimise $\|J - PI\|_2^2$ with the constraint

$$\|LI\|_2^2 \leq \epsilon(\lambda^2), \tag{2.30}$$

where ϵ is a real value depending on the regularization factor and the random noise. Therefore, Tikhonov regularization is actually a method for residual minimisation.

Because (2.29) is strictly convex, the minimisation of this energy function leads to the unique solution [70]

$$\hat{I} = (P^T P + \lambda^2 L^T L)^{-1} P^T J, \tag{2.31}$$

which is referred to as the Tikhonov-Miller regularised solution. From equation (2.28)), the Tikhonov-Miller solution (2.31) is actually a linear filter similar to the Wiener filter.

The direct calculation of this solution (2.31) is a challenge, because the inverse of the matrix is difficult to calculate when the image size is large. For example, when an observed image has a size of 512*512 (pixels), the PSF matrix and the regularization operator are both 262144*262144. Finding the inverse of such a massive matrix is impractical. To overcome this problem, we may rewrite the solution as a linear system, i.e., $AI = B$, where $A = P^T P + \lambda^2 L^T L$ and $B = P^T J$, and next, we solve this system with numerical methods such as conjugate gradient and Cholesky decomposition. However, there remains a question as to how such a huge matrix can be stored in a computer, given that approximately 60GB of memory is required to save the operator in the format of 'unit 8'. Although the PSF matrix and the regularization operator are often spare and although a majority of their elements are zero, the calculation also presents a heavy burden.

Split Bregman method

The problem of optimising a functional with L_1 regularizations can be solved by the Split Bregman method. The general form of such problems is

$$\min_I \|\Phi(I)\|_1 + H(I), \tag{2.32}$$

where $\|\cdot\|_1$ is the L_1 norm and $\Phi(\cdot)$ and $H(\cdot)$ are convex functions. For example, the Rudin-Osher-Fatemi (ROF) functional serves to model image denoising through regularization of the TV,

$$\min_I \|\nabla I\|_1 + \lambda \|I - J\|_2^2, \tag{2.33}$$

where ∇ is the gradient operator. Usually, $\|\nabla I\|_1$ is a bounded variation (BV). If the TV is applied for image restoration, model (2.33) becomes

$$\min_I \|\nabla I\|_1 + \lambda \|PI - J\|_2^2. \tag{2.34}$$

Because the BV is not strictly convex, finding an analytical solution of (2.34) is extremely difficult; thus, numerical methods are applied. However, many common numerical methods are also limited due to an ill-posed coefficient matrix or BV. For example, Newton-type methods require the derivative of the objective functions, but they are not differentiable. Conjugate gradient or Gauss-Seidel methods suffer from problems in terms of convergence if structure recovery is the focus. In [56], the authors proposed the Split Bregman algorithm, which is one of the most popular solvers for complicated regularizations.

The Split Bregman method considers the optimisation problem (2.34) with the constraint

$$\min_{I} \|d\|_1 + \lambda \|PI - J\|_2^2 \text{ such that } d = \nabla I, \tag{2.35}$$

which is equal to the unconstraint optimisation problem

$$\min_{I,d} \|d\|_1 + \lambda \|PI - J\|_2^2 + \beta \|d - \nabla I\|_2^2, \tag{2.36}$$

by the Lagrange multiplier method. By applying the Bregman iteration to (2.36), the Split Bregman iteration is given by [56]

$$(I_{k+1}, d_{k+1}) = \min_{I,d} \|d\|_1 + \lambda \|PI - J\|_2^2 + \beta \|d - \nabla I - X_k\|_2^2, \tag{2.37}$$

and

$$X_{k+1} = X_k + (\nabla I_{k+1} - d_{k+1}). \tag{2.38}$$

The energy function (2.37) contains two variables, (I, d). To find a local optimal solution, it can be solved by fixing variables one by one. First, d is fixed; therefore, (2.37) is written as

$$I_{k+1} = \min_{I} \lambda \|PI - J\|_2^2 + \beta \|d - \nabla I - X_k\|_2^2. \tag{2.39}$$

The problem (2.39) has a unique analytical solution because it is strictly convex, with only the L_2 norm included. Next, I is fixed in (2.37), yielding

$$d_k = \min_{d} \|d\|_1 + \beta \|d - \nabla I - X_k\|_2^2. \tag{2.40}$$

The problem (2.40) becomes simpler than (2.34) because there is no gradient operator inside the L_1 norm and this step can be solved by linearised Bregman iteration. The linearised Bregman iteration is defined as

$$\begin{cases} c^{(i+1)} = c^{(i)} + (\nabla I_{k+1} + X_k - d_k^{(i)}), \\ d_l^{(i+1)} = \delta \cdot shrink(c^{(i+1)}, \frac{1}{2\beta}), \end{cases} \tag{2.41}$$

where $c^{(i)} = 0$ and $d_k^{(i)} = 0$ are initial values for $i = 0$, and δ is a parameter.

Singular value threshold algorithm

Many non-local regularizations are applied for image restoration. A popular way of non-local regularization is via the low-rank prior method which assumes the matrix constructed by similar packed patches from the restored image has a small rank. This can be formulated as the nuclear norm optimisation (NNM)

$$\min \|X\|_* \text{ such that } X = M, \tag{2.42}$$

where M is the noisy image and $\|\cdot\|_*$ is the nuclear norm defined as the sum of all the singular values of M. The NNM problem is equal to the unconstrained minimisation

$$\min_X \|X\|_* + \lambda \|X - M\|_F, \tag{2.43}$$

where e $\|\cdot\|_F$ is the Frobenius norm. Because image denoising is often implemented in a transformed domain, (2.43) can also be written as

$$\min_X \|X\|_* + \lambda \|BX - BM\|_F \tag{2.44}$$

The solution of (2.44) is given by the singular value threshold iterative

$$\begin{cases} X_k = shrink(Y_{k-1}, \lambda), \\ Y_k = Y_{k-1} + \delta_k B(M - X_k). \end{cases} \tag{2.45}$$

Because this energy function is strictly convex, the iterative solution (2.45) converges to its unique analytical solution [71]

$$S_\lambda = U D_\lambda(\Sigma) V^T, \tag{2.46}$$

where U, Σ, V is the singular value decomposition of X and $D_\lambda = \text{diag}(\max(0, s_i))$ with the ith singular value s_i.

However, this method cannot be directly applied for non-local regularizations. One reason for this constraint is that the non-local regularization terms are always constructed using a similarly packed patch matrix rather than the image itself. The other reason is that features of structures are grouped such that there are many regularisation terms for these groups. To solve this problem, a weighted nuclear norm minimization method (WNNM) is developed in [72] by treating singular values differently.

We have discussed low-rank regularization for image denoising but image restora-

tion can include blurring. To apply the WNNM method we have to decouple the deblurring and denoising procedures for a complicated energy function with non-local regularization terms. A very useful technique can facilitate decoupling: the half-quadratic splitting technique. The key aspect of this technique is to constrain the objective function by introducing new auxiliary variables. Then, this constrained problem can be converted into an unconstrained one using the Lagrange multiplier method.

As a demonstration, we use the algorithm introduced in [69]. Because this algorithm is formulated with a matrix of a 2D image, rather than the 1D image vector in our method, we rewrite model (2.15) in matrix form:

$$J^{2D} = P^{2D} \bigotimes I^{2D} + N^{2D} \tag{2.47}$$

where I^{2D} and J^{2D} are the 2D matrices of the original scene and the observed image, respectively; N^{2D} is a 2D matrix representing noise; and P^{2D} is the known PSF blurring the original image by $\bigotimes$, the convolution operator. To solve (2.47), in the paper [69], a complicated energy function is proposed:

$$\hat{I}^{2D} = \arg\min_{I^{2D}} \left\| P^{2D} \bigotimes I^{2D} - J^{2D} \right\|_1 + \lambda \sum_i \left\| I_i^{2D} \right\|_{w,*} + \sigma \left\| \nabla I_i^{2D} \right\|_{w,*}, \tag{2.48}$$

where I_i^{2D} is the ith group of similar packed patches from the restored image, and $\|\cdot\|_{w,*}$ is a weighted nuclear norm defined in [72].

We will apply the half-quadratic splitting technique to decouple the deblurring and denoising procedures. First, three auxiliary variables, d, p and g and three constraints are introduced for the energy function (2.48):

$$\begin{cases} d = P^{2D} \bigotimes I^{2D} \\ p = I^{2D} \\ g = \nabla I^{2D}. \end{cases} \tag{2.49}$$

Using the Lagrange multiplier method the constrained optimisation problem (2.49) can be converted to

$$\begin{aligned} \hat{I}^{2D} = \arg\min_{I^{2D}} & \left\| P^{2D} \bigotimes I^{2D} - J^{2D} - d \right\|_1 + \beta \left\| I^{2D} - p \right\|_2^2 + \tau \left\| \nabla I^{2D} - g \right\|_2^2 \\ & + \eta \left\| d \right\|_1 + \lambda \sum_i \left\| I_i^{2D} \right\|_{w,*} + \sigma \left\| \nabla I_i^{2D} \right\|_{w,*}, \end{aligned} \tag{2.50}$$

which is an optimisation problem without constraints. Thus, the above problem can

be divided into four subproblems:

$$\hat{d} = \arg\min_{d} \left\| P^{2D} \bigotimes I^{2D} - J^{2D} - d \right\|_1 + \eta \left\| d \right\|_1 , \tag{2.51}$$

$$\hat{p} = \arg\min_{p} \left\| I^{2D} - P^{2D} \right\|_2^2 + \lambda \sum_i \left\| p_i \right\|_{w,*} , \tag{2.52}$$

$$\hat{g} = \arg\min_{g} \tau \left\| \nabla I^{2D} - g \right\|_2^2 + \sigma \sum_i \left\| g_i \right\|_{w,*} , \tag{2.53}$$

$$\hat{I} = \arg\min_{I} \left\| P^{2D} \bigotimes I^{2D} - J^{2D} - d \right\|_2 + \beta \left\| I^{2D} - p \right\|_2^2 + \tau \left\| \nabla I^{2D} - g \right\|_2^2 , \tag{2.54}$$

in which the variables, d, p, g and I^{2D} can be solved for separately in an iterative algorithm with a given initial I.

As seen in the four subproblems, the first and last subproblems attempt to recover structures, and the second and third subproblems attempt to reduce noise via non-local low-rank regularizations. The first subproblem (2.51) can be solved by linearised Bregman (2.41). Equations (2.52) and (2.53) can be solved by WNNM [72]. The last subproblem involves only the L_2 norm and the optimal solution is difficult to calculate through the derivative of (2.54) due to the gradient and convolution operators. A fast solution is achieved via a variational method and the fast Fourier transform.

2.4 Summary

Through the introduction of methods for DNA detection and image restoration, we can see that it is difficult to apply the non-local concept into AIC, BIC and LOF methods, because the measurements used in these methods will become very complicated if patches are applied to likelihood function or LOF. Regularization methods, such as IDDBM3D and low rank prior techniques can use non-local regularization terms, while they cause complex models which are difficult to be solved. Moreover, regularization factors are very important parameters in regularization methods, regardless of whether the regularizations are modelled with local or non-local information. As shown in both the Bregman method for TV regularization and WNNM for non-local regularization, the threshold is determined by the regularization factors and provides a balance between structure recovery and noise. If the factors are too large, the edges will be removed; if the factors are too small the noise will increase. However, regularization factors are so complicated that error analysis is difficult. In most of publications, the determination of regularization factors has mostly depended on experience.

Chapter 3

Window-Adjusted Lack of Fitness for DNA Detection

This chapter consists of a research project supported by Renishaw for the detection of component dyes in a multi-plex DNA spectrum. Renishaw developed the Lack of Fitness (LOF) method for detection purposes but wished to improve the detection results by further developing the LOF method. To this end, we study a non-local method that introduces a range of additional measurements to further investigate the results close to the criterion boundaries of the original LOF-based classification. These additional measurements are conducted in relation to the windows in the observations and reference components. Therefore, our method is called window-adjusted LOF or WALOF for short. The numerical results show that our method decreases the false detection rate by 30% compared to the LOF method.

3.1 Description of Renishaw data sets

3.1.1 Component dyes and their variabilities

In advance, the Renishaw data and model background are introduced. There are 13 known component dyes in the Renishaw data and the names of these components are listed in Table 1. The Blank component is an artificial dye, which is designed to guarantee positive intensity in fitness. The shape of all the known components can be found in Figure 10. Due to uncertain factors in the complex imaging system the spectra of one component vary greatly in intensity, although they have similar appearances with respect to shape. Therefore, the present components are random variables and are precisely unknown with regard to detection.To have a full view of these components, Renishaw provided variation sets for the components. Each dye has a variation set containing approximately one-hundred spectra, except for Blank,

which has a variation set of 48 spectra.

Table 1: Names of the known component dyes

Number	Name	Number	Name
1	Blank	8	TYE
2	Cy3.5	9	BODIPY_FL
3	Cy3	10	JOE
4	FAM	11	ATTO520
5	HEX	12	BODIPY_TMRX
6	Rhodamine Green	13	MAX
7	TAMRA		

In this section, all spectra and their variations are presented. These thirteen dye spectra are collected with the wavelengths in the range from 300 to 3800. Figure 9 demonstrates a random spectrum of Cy3.5. As seen in this plot, the spectrum has a long tail that contains no useful signals. Therefore, the wavelength range for spectra analysis is from 300 to 1800, with the tail removed. The random thirteen spectra are plotted in Figure 10.

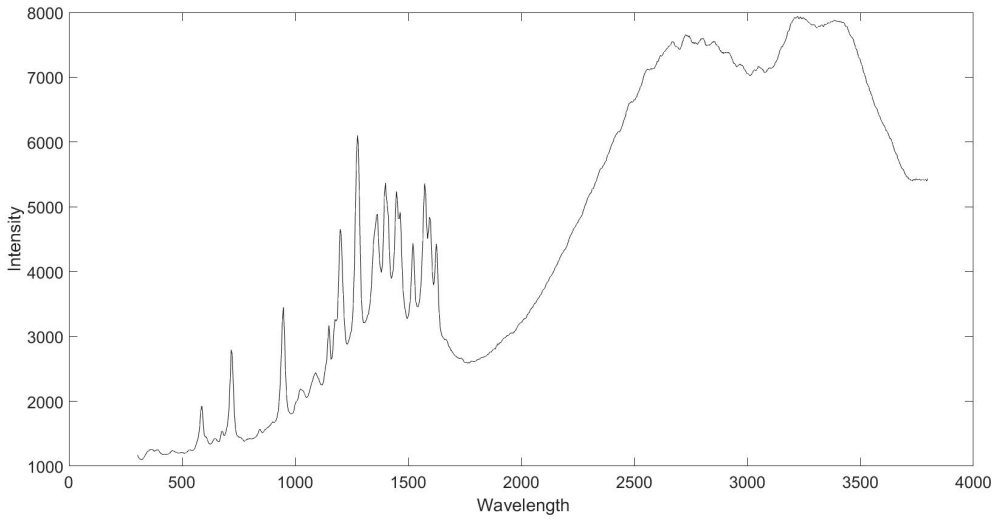


Figure 9: Spectrum of Cy3.5 in the wavelength range from 300 to 3800.

The reason for that the spectrum of one component dye is random is because it varies across diverse intensities due to the complexity of the imaging system. Despite the differences in intensity, a spectra of the same component dye present similar shapes, which is called the variability of a component dye. Figure 11 interprets the variability using the example of the component dye CY3.5. In this plot, the total number of spectra collected by Renishaw is 103. Seen in this figure, the intensities of these spectra are different but their shapes are very similar. Calculating the correlation coefficient of these 103 spectra, we obtain a minimum of 0.9853, which shows that

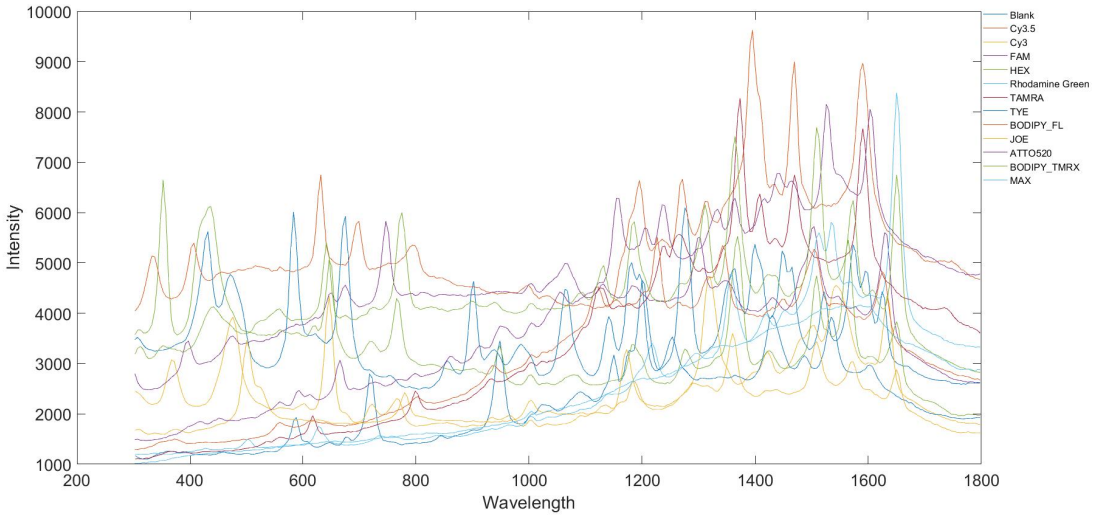


Figure 10: Random spectra of thirteen components dyes.

the spectral shapes do present the characteristics of one component, although they change randomly in intensity.

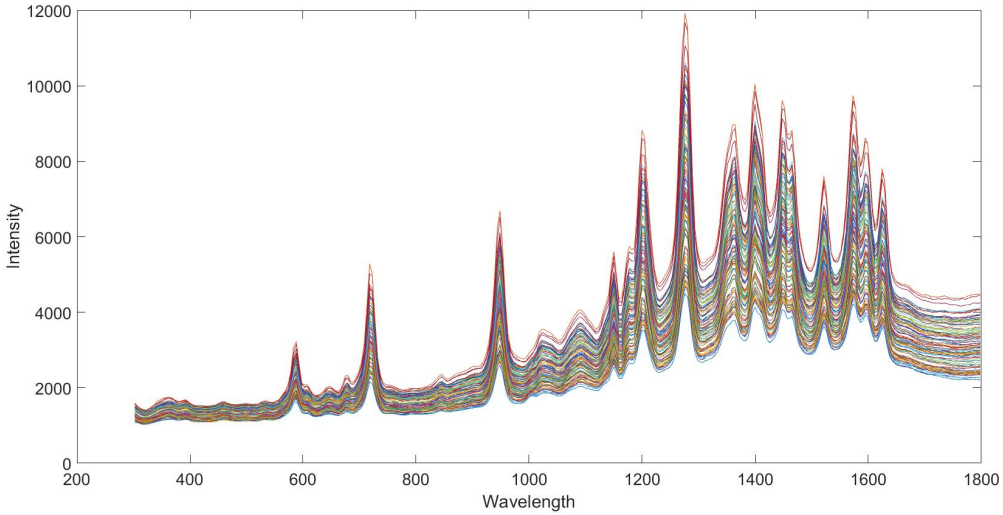


Figure 11: Interpretation of variability by 103 spectra of component dye Cy3.5.

3.1.2 Generation of synthetic observations

Evaluation of an algorithm using the real, observed spectra is impractical, because the actual components contained are unknown. Thus, Renishaw built up a platform to generate synthetic observations that simulates real DNA spectra, wherein the components contained in the synthetic observations are known. The following introduces the construction of a synthetic observation.

The construction of an observation contains five steps. 1: Determine how many

dyes are present in the observed spectrum. Here, we apply the experience criteria from Renishaw. The Blank is always presented. The number of presented dyes among the remaining 12 dyes is assumed to follow the distribution of individual points [0 37% 44% 19%]. The element of the vector indicates the probability of the event that a certain number of component dyes are present in the observation. Zero indicates that there are at least two dyes present. The other three values indicate that three, four and five dyes are present with the probabilities of 37%, 44% and 19%, respectively.

2: Determine the presented dyes and their real spectra in variation sets for the construction of an observation. At first, the presented dyes are chosen randomly among the 13 labelled dyes, with exception of Blank, which is always included. We assume equal chances of the presence of each dye, meaning that the presence distribution of dyes is uniform. Then, a random spectrum in the variation set of the dye is selected as the component spectrum. The presence distribution of the random spectrum is also assumed to be uniform.

3: Generate the amplitudes of the presented dyes. Because the Blank dye is artificial there are different two ways to generate amplitudes of the Blank and the other 12 dyes. The amplitudes of the 12 dyes are generated based on the absolute value of a normal distribution with mean 0.6 and variance 1. The amplitude of the Blank dye is one minus the sum of the amplitudes of all the presented dyes.

4: Construct a noise-free observation. Having the presented spectra and their corresponding amplitudes determined as described above, we construct an observation as the linear sum of the spectra weighted with their amplitudes.

5: Noise is added to the noise-free observation. The noise is assumed to follow normal distribution a mean of zero, while the standard deviation depends on the intensity. As seen in Figure 12, the two curves, representing a synthetic observation (blue) and noise-free observation (red), nearly overlap across the whole wavelength range. Therefore, the noise effects little influence on the observation. Here, we would not introduce the algorithm by Renishaw to generate noise.

The procedure for constructing a synthetic observation can be described using the model reviewed in Chapter 1,

$$Y = SA + N, \tag{3.1}$$

where the column vector Y is the synthetic observation, matrix S is the spectra of the component dyes, column vector A represents the amplitudes of the component dyes, and N is noise that follows a normal distribution. The spectra of the component dyes that are present are also represented as column vectors in matrix S, such that the number of component dyes present equals the number of columns of matrix S.

Figure 12 demonstrates the construction of the synthetic observation spectrum.

In the first step, a random variable of uniform distribution in $[0, 1]$ is sampled to determine the number of component dyes present in the synthetic spectrum to be generated. The number of dyes present is three, because of the sampled value 0.6210, which is larger than 0.37 and smaller than $0.81 = 0.37 + 0.44$. Second, four component dyes are selected from the thirteen. With the exception of Blank, which must be included, the other three, Cy3.5, BODIPY_FL and MAX, are selected randomly with equal probability from the remaining 12 component dyes. Having the four component dyes, we select a spectrum randomly for each dye from its variability set and value the columns of matrix S with the four selected spectra. Third, the amplitudes of the three component dyes are sampled from the distribution in step 3, and the amplitude of Blank is calculated by the subtracting the sum of the three amplitudes from one. Thus, we get the amplitude vector A. Finally, a synthetic spectrum Y is generated with Eq. (3.1).

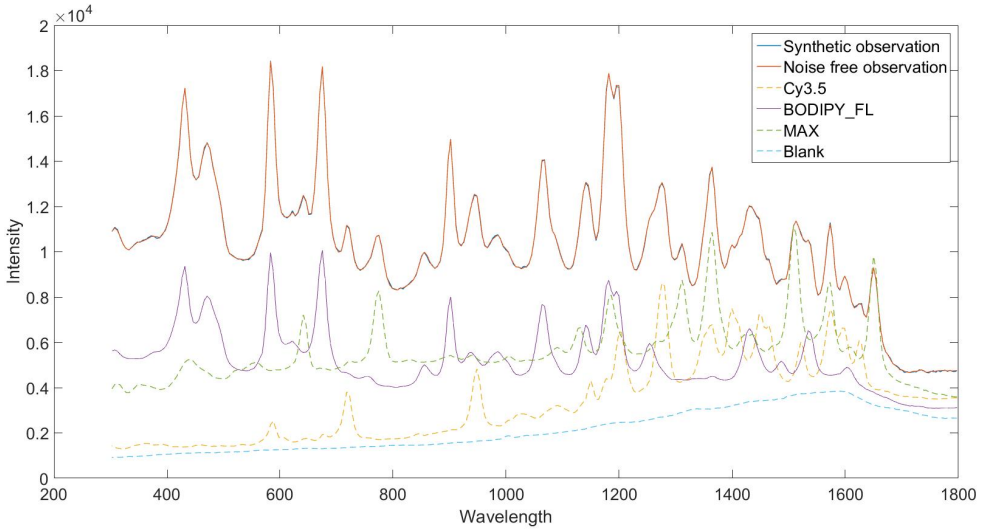


Figure 12: Diagram of the construction of a synthetic observation spectrum.

3.2 Window-Adjusted Lack of Fitness

In this section, we present a new approach to improve the original LOF-based classification method. Briefly, we introduce a range of additional measurements to further investigate the results close to the criterion boundaries of the original LOF-based classification. These additional measurements are conducted in terms of the peak regions (windows) in the observations and reference components, thereby representing it is a non-local method. For simplicity of description, the method is called the "Window-Adjusted LOF", or WALOF for short.

3.2.1 Why windows are needed?

The LOF method proposed by Renishaw attempts to determine the possible component dyes one by one with the criteria of minimum LOF, which is the square sum of the fitting residual divided by square sum of the observation. This algorithm stops when LOF improvement is less than a hard threshold. However, this algorithm does not work well to detect minimal components and to reject false components similar to present components. The reason is that the LOF measurement cannot sufficiently explain the relationship between the residual and the investigated component when the residual is small. Therefore, further analysis of the similarity between the residual curve and the investigated component is warrant.

The basic rationale behind our WALOF method is that features in peak regions can provide additional information that can be used to refine the classification results of the LOF-based method. Two examples will be demonstrated via LOF, one of which shows how a minimal component dye is missed, the other of which shows how a false component is selected. In section 1.2 in Chapter 1, there is an example demonstrating a missed component, which is simple to understand because the synthetic observation contains only two ingredients. Such cases exist, but they occur with low probabilities. Therefore, the first example here is demonstrated with a complicated synthetic observation.

Figure 13 demonstrates an example of missing a minimal component during a detection. A synthetic observation spectrum is generated with five component dyes: Blank, TAMRA, Rhodamine Green, JOE and BODIPY_TMRX. Applied with the LOF algorithm with the parameter of LOF improve limit valued 0.10, four component dyes are detected successfully, while the component JOE is missed. The blue line is a synthetic observation and the red line is the fitting curve with the four detected component dyes. The detected component dyes are plotted as dotted lines and the missed component JOE is marked with stars.

As seen in Figure 13, the fitting curve is very close to the synthetic spectrum, which shows the residual is already small before adding in the investigated component dye JOE. The small residual shows a component dye must be a minimal component if it is to be included. Looking back at the generation procedure for this observation, the random amplitudes for these component dyes are 0.3030 for TAMRA, 0.9093 for Rhodamine Green, 0.1081 for JOE and 0.5172 for BODIPY_TMRX. By the LOF method, Rhodamine is selected because it leads to a significant LOF improvement of 0.7282. Next, BODIPY_TMRX and TAMRA are detected in that order, leading to LOF improvements of 0.5378 and 0.7213, respectively. This result proves the LOF method is efficient for detection of components that are large or moderate in scale. However, when JOE is investigated, it leads to an LOF improvement of 0.0621

which is less than the parameter of LOF improvement limit set to be 0.10. Thus, the component JOE, despite actually being present, is rejected by the LOF method.

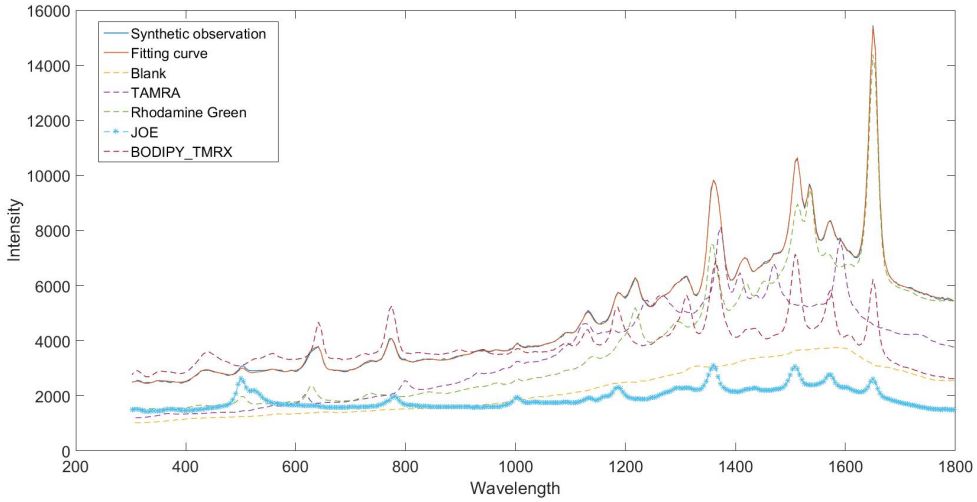


Figure 13: Demonstration of a detection with the LOF method by Renishaw. These lines represent synthetic observation (blue), and a fitting curve (red) with Blank,TAMAR,Rhodamine Green and BODIPY_TMRX. The four dotted lines stand for the detected components Blank,TAMAR, Rhodamine and BODIPY_TMRX. The star line is the spectrum of JOE, which is rejected by LOF.

We analyse the residual curve and the investigated component JOE. Figure 14 plots the residual curve and the spectrum of JOE individually. In this figure, there are three peaks marked at wavelengths of 500.5, 1360.0 and 1650 wavelength in plot (b). In plot (a), there are also three peaks located at positions close to those marked in (b). In particular, the peak at a wavelength of 505.4 is obvious. This finding indicates that there might be a minimal component dye remaining in the residual. The reason is that there would be little similarity between the residual and the investigated component dye if only noise remained in the residual without any information. Therefore, similarity between the residual curve and the investigated component dye JOE should be further analysed.

The similarity between the residual curve and the investigated component inspires the use of a non-local method. Based on the LOF method, this method focusses on the investigation of components that lead to LOF improvement somewhat larger or smaller than the parameter, aiming to reject false components or detect minimal components. In this method, windows are first defined to separate features of spectra from the residual. Then certain characteristics are extracted from these windows, such as correlation coefficients, peak height and width, peak locations and peak gradients. These characteristics are inputted into a learning system to construct a criterion for a classifier. Through the classifier, the investigated component is determined to be present or absent.

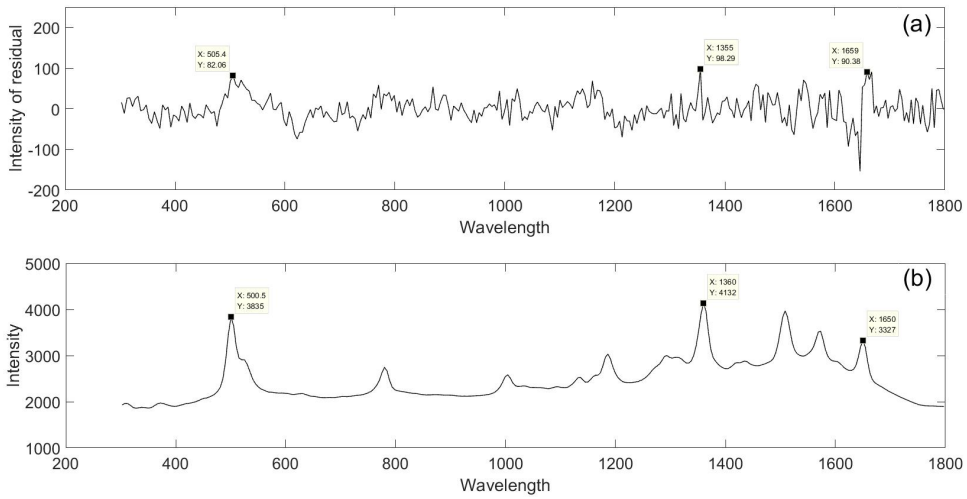


Figure 14: (a)Residual curve fitted with four components, Blank, TAMRA, Rhodamine Green and BODIPY_TMRX, (b) The spectrum of JOE, which is rejected by the LOF method.

The second example shows a false component dye detected via LOF in Figure 15. The synthetic observation (blue line) has three component dyes of Blank, ATTO520 and BODIPY_TMRX. The amplitude of ATTO520 is 1.0063 and is 0.8657 of BODIPY_TMRX, which means that the two component dyes support their inclusion as ingredients in the synthetic DNA spectrum. Via LOF, BODIPY_TMRX is first detected, with an LOF improvement of 0.2468; then, ATTO520 follows, with an LOF improvement of 0.8223. However, when the false component dye MAX is investigated, the LOF improvement is 0.1473, which is larger than the LOF improvement limit of 0.10. Thus, the false dye is accepted as an ingredient.

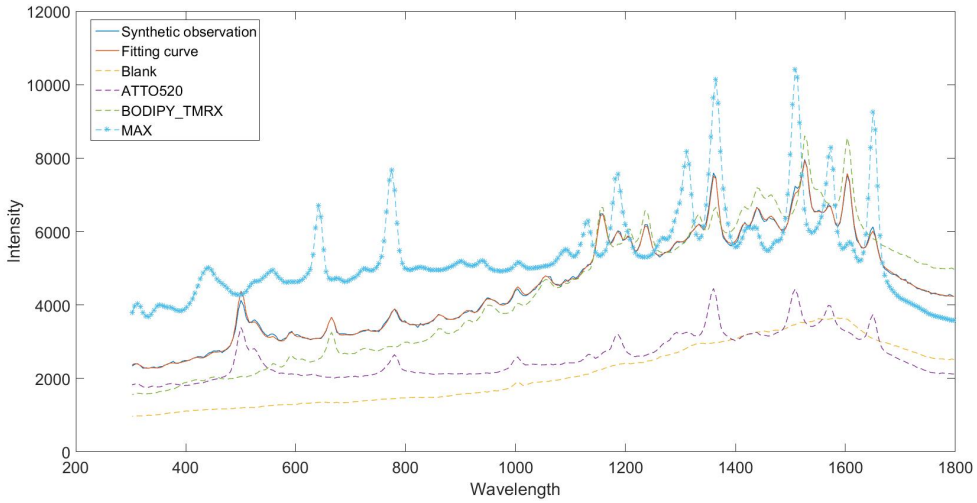


Figure 15: The blue line represents a synthetic DNA spectrum containing three component dyes of Blank, ATTO520 and BODIPY_TMRX. The red line is a curve fitting to the three component dyes, which are plotted by dotted lines. The spectrum of MAX marked with stars is a false component dye detected by LOF.

The reason for the false acceptance of dye MAX is the similarity existing between the false component and the present component. Removing the curves not relevant to MAX, only ATTO520 and MAX are plotted in Figure 16. As seen in this figure, the two spectra have similar shapes at the marked peaks. Their correlation coefficient is as high as 0.8152. Although, this value is not as high as the correlation coefficients of spectra in a variability set, it shows great similarity between MAX and ATTO520 which might mislead the LOF method.

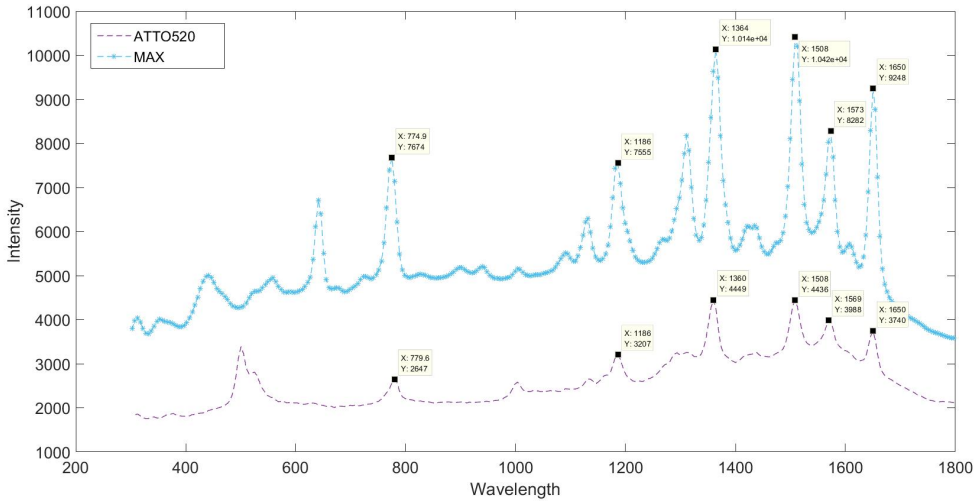


Figure 16: Similarity between the spectrum of ATTO520 (dotted line) and MAX spectrum marked with stars. The marked peaks show positions at which they are close to each other.

Further analysis of residual is shown in Figure 17. There are four curves plotted. As seen in this figure, the curve in plot (a) looks very similar to that in plot (c), which indicates that MAX has little effect on the residual when it is added to the model. So, what makes the false dye MAX appear as a present ingredient with a larger LOF improvement? The reason is variability. Seen in Figure 15, the fitting curve is very close to the synthetic observation when ATTO520 is added in, whereas the LOF improvement is still larger than the limit after adding in MAX, which is similar to ATTO520. This finding shows that most of the features of the true spectrum are present by the fitted spectrum, while the remaining features can be well explained by MAX similar to being explaned ATTO520. Therefore, there must be a larger variability between the true spectrum of ATTO520 and the fitted spectrum of ATTO520. Figure 18 plots the fitted spectrum (blue line) and the true spectrum (red line) of ATTO520. Known from this figure, the fitted spectrum has higher intensities than the true spectrum up to a wavelength of 1000, whereas its intensity is lower than that of the true spectrum above a wavelength of 1000. Their correlation coefficient, 0.9622, is obviously less than 0.9978, the averaged correlation coefficient of the variability set of ATTO520.

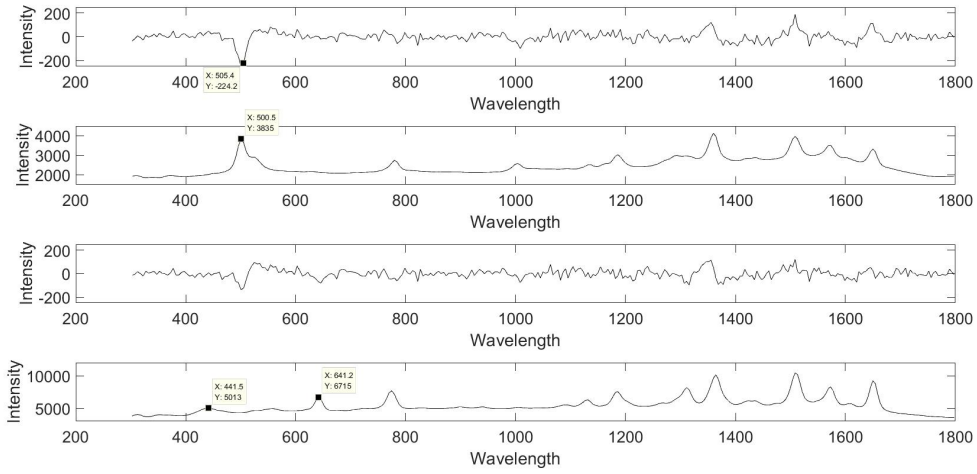


Figure 17: (a) The residual curve generated by fitting the three component dyes of Blank, ATTO520 and BODIPY_TMRX. (b) The spectrum of ATTO520. (c) The residual curve by fitting generated by fitting the three component dyes and the false component dye MAX. (d) The spectrum of MAX, which is recognised as a false ingredient by the LOF method.

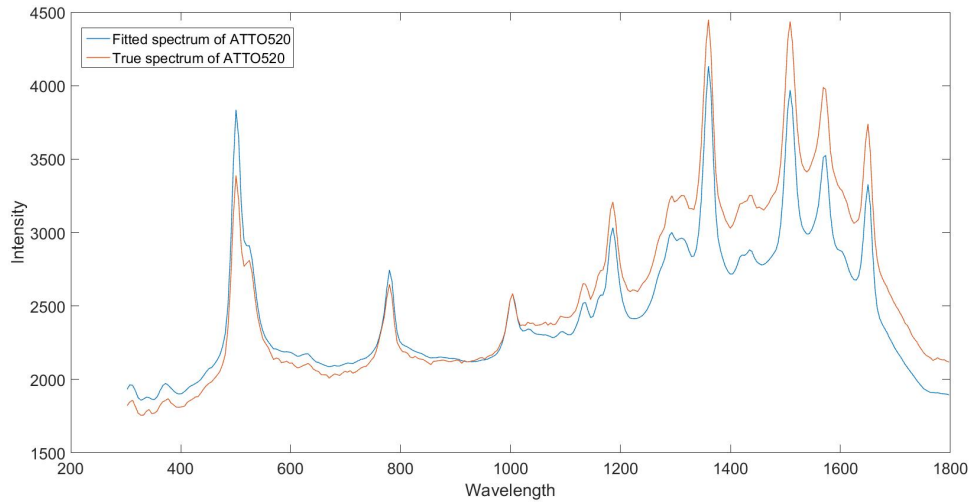


Figure 18: Variability between the fitted spectrum and the true spectrum of ATTO520.

The above discussion provides an answer for that why a larger LOF improvement may be yielded after the inclusion of a false component dye, but this finding is not evidence to prove the component absent. It is known that if a component dye is present, there must be a response from the residual curve in the area of its features. For example, ATTO520 has a peak at a wavelength of 500.5, whereas Blank and BODIPY_TMRX are flat around this point. When ATTO520 is investigated before the model update, there must be a response in the residual around this point. Seen in Figure 17, the residual shows a peak at a wavelength of 505.4, which is a reliable

indicator that ATTO520 is present. Applying this principle for MAX, we can see two significant peaks at wavelengths of 441.5 and 641.2, at which points the other dyes in the model are flat. However, no responses can be detected in plot (a) of the residual at the positions at wavelengths of 441.5 and 641.2. Therefore, the component dye MAX might not be present in the synthetic observation.

3.2.2 Determination of windows

The first task in the WALOF method is to define windows to separate the features (peaks) in a spectrum. Locating peaks manually can be very accurate but time consuming and inconvenient for users. Therefore, an intelligent method is introduced to define peak windows in a spectrum. This method includes five steps, as follows.

Step 1. Locating the peak and bottom positions: Given a spectrum segment, a peak is defined as a point whose value is higher than N points on its left and right sides. The bottom is thought to be the lowest value between two peaks. For example, when $N = 7$ is set, as shown in Figure 19, P1 is a peak but P3 is not a peak because intensity of P3 is less than that in P1 and only $4 < N = 7$ points are located between them. For the same reason, P2 is also not a peak. B1 and B2 are bottom points. Thus, the window can be denoted as $\mathcal{W}_{B_1P_1B_2}$.

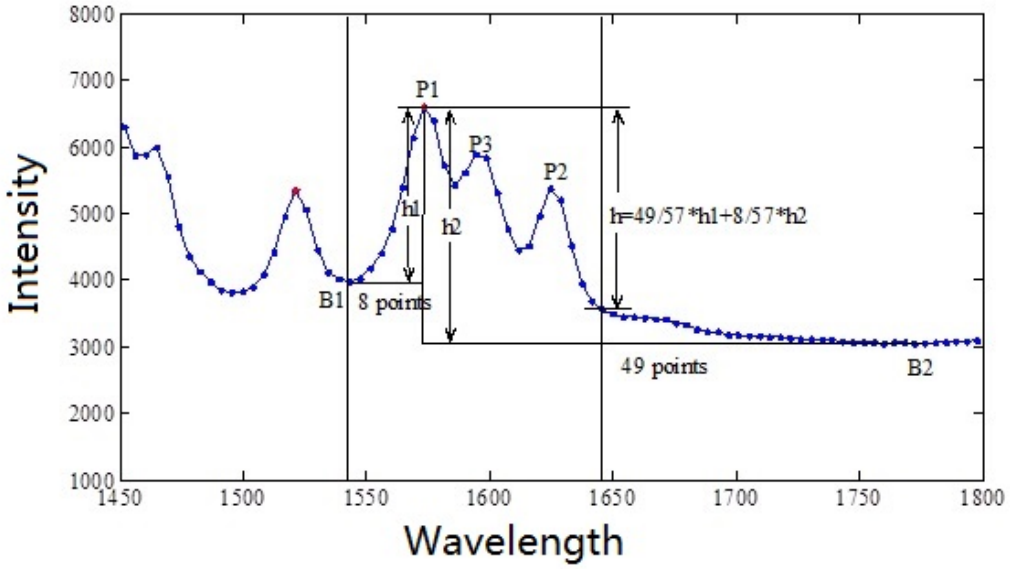


Figure 19: Illustration of window determination for the window location in a spectrum.

Step 2. Define the window height: It is found that there is a long tail $\mathcal{W}_{B_1P_1B_2}$ defined in step 1 in the right half peak $\mathcal{W}_{P_1B_2}$. This long tail is useless for similarity analysis due to containing no information. To cut off the long tail, we define the window height via the weighed sum of the left and right heights of the window. As shown in

Figure 19, the left height of window $\mathcal{W}_{B_1P_1}$ is $h1$ which is the absolute difference of intensities at P1 and B1, and its right height is $h2$ which is the absolute difference of intensities at P1 and B2. The window height is defined as $h = w * h1 + (1 - w) * h2$, where the weight w is obtained as $w = w2/(w1 + w2)$ and $w1$ and $w2$ are the left and right widths from the peak to the bottoms. The widths can be measured by the number of points because of the nearly equal wavelength interval between any two adjacent points. As depicted in the figure, $w1 = 8$ points and $w2 = 49$ points, thus $w = 0.86$. This weight indicates the greater importance of the left height and the weight $1 - w = 0.14$ is used to penalise the long tail to the right. Thus, the window height is obtained by the weighted sum; then, the long tail is cut off by picking out all points whose distances to P1 are less than h in terms of intensity.

Step 3. Separating sub-window: Because noise can also cause peaks in spectra of residuals and component dyes, it is important to distinguish between windows due to noise versus features through when similarity analysis implemented via comparisons. This step gives a criterion for the detection of sub-peaks in a spectrum. Choose a parameter, N_1, smaller than N, and detect any sub-windows using the same method described in Step 1. If both the left and right heights of a sub-window exceed a certain percentage (60% being used in WALOF) of the main window height, we consider this sub-window to be a new one. An example of this is given in Figure 20 (with $N = 7$ and $N_1 = 5$), where $\mathcal{W}_{B_1P_1B_2}$ is a main window and $\mathcal{W}_{B_1P_2B_3}$ is the sub-window separated from the main window.

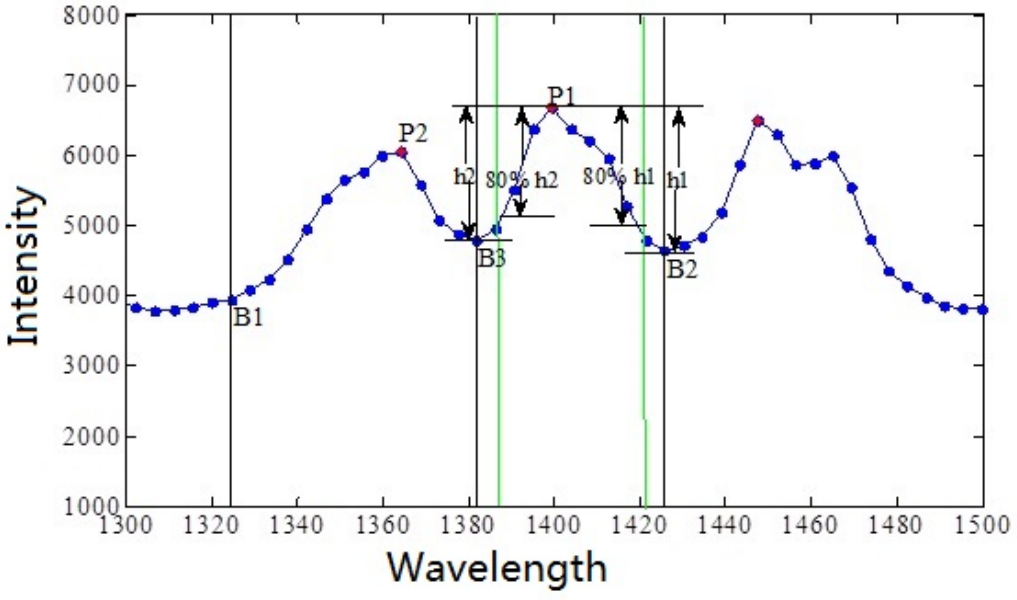


Figure 20: Demonstration of separating sub-windows and further window region shrinkage.

Step 4. Further shrinkage for window regions: To improve the sensitivity for feature

comparison between residuals and investigated component dyes, we further shrink the window region determined in Step 1-3. With the example shown in Figure 20, the region of the window $\mathcal{W}_{B_3P_1B_2}$ is reduced to a narrow one whose borders are marked with orange lines. Though the shrinking rule, the narrow window keeps 80% of the height of the window $\mathcal{W}_{B_3P_1B_2}$.

Step 5. Filtering non-significant peaks: Any peaks with a peak height of less than 10% of the maximum peak height in a spectrum are considered insignificant and are not included. To ensure the robustness in using the information within the windows for further classification, we require that a window has a minimum width (8 points).

Figure 21 shows an example of defining windows in a spectrum of CY3.5 plotted with a blue line. In this figure, the points marked with red stars are peaks for windows. The red vertical lines are borders of windows. The two points at 362.3 and 461.2 are also peaks for windows; however, their heights are sufficiently small that they are removed from consideration by Step 5.

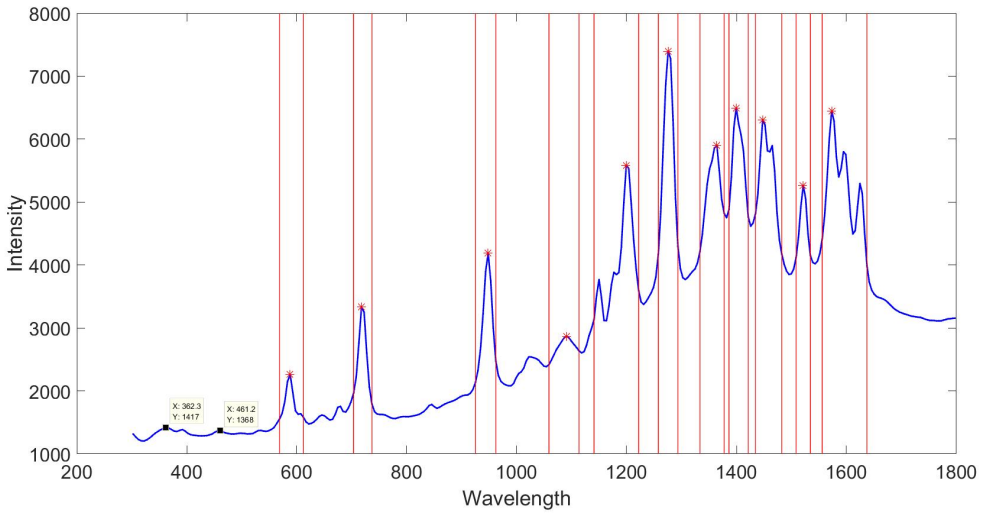


Figure 21: Example of window location with the parameter N=7. A sub-peak is located at the X-axis position of 1364, with N1=5.

3.2.3 Naive Bayes classifier

The task of distinguishing a component dye as being present or absent is actually a problem of classification. There are many famous classifiers in the fields of machine learning, such as decision tree classifiers [73], logistic regression classifiers [74], Bayes classifiers [75], neural networks [76], and support vector machines [77]. Among these classifiers, the decision tree classifier prefers to features which take discrete values, such as rank and gender. Logistic regression models feature as hyperplanes, an approach that is too simple to represent the complexity of our problem. Neural

networks and support vector machines are powerful for the classification for objects with complicated features, although they yield a high cost of computation and a long time for running, which is not suitable for detections stemming from thousands of observations. To avoid these disadvantages, we prefer the Bayes classifier which outputs a posterior probability of classification.

The naive Bayes classifier [78] is a fundamental method proposed by Langley in 1992. Based on this method, three important improvements have been proffered. Friedman proposed a tree augmented naive Bayes (TAN) in [79], which improved this method by generalising the naive Bayes classifier and representing the assumption of independence. However, TAN fixes the number of edges of the training network and pays no attention to training data. This problem in TAN is solved in [80] with a boosted augmented naive Bayes (BAN). Considering the order of nodes in the training network, two algorithms are proposed in [81] for general Bayes networks (GBN). As seen in [82], GBN and BAN show the best performance for prediction, while they are time consuming. TAN outperforms naive Bayes and but it is complicated for coding. The naive Bayes classifier is competent for detection, and it can be implemented with MATLAB.

The naive Bayes classifier assumes that the features are independent random variables for classification. Denoting $x_1, x_2, \ldots, x_n$ as the n features, and $G_k,\ k = 0, 1,$ as the categories, where G_0 means a component dye is absent and G_1 means a component dye present, the conditional probability is obtained such that

$$P_{post} = \text{Prob}(G_k|x_1, x_2, \ldots, x_n), \tag{3.2}$$

with the input data of the features. Applying Bayes' rule, the conditional probability can be written as

$$\text{Prob}(G_k|x_1, x_2, \ldots, x_n) = \frac{\text{Prob}(G_k)\text{Prob}(x_1, x_2, \ldots, x_n|G_k)}{\text{Prob}(x_1, x_2, \ldots, x_n)}, \tag{3.3}$$

where the probability $\text{Prob}(G_k)$ is known as a prior in training data. The conditional probability $\text{Prob}(x_1, x_2, \ldots, x_n|G_k)$ represents the joint distribution of all features under the condition that the category of an observation is known as G_k. This conditional joint distribution is often complicated to calculate because the domain of features is a high dimension space. With the assumption of conditional independence of features in naive Bayes classification, the conditional joint probability can be calculated as

$$\text{Prob}(x_1, x_2, \ldots, x_n|G_k) = \Pi_{i=1}^{n}\text{Prob}(x_i|G_k), \tag{3.4}$$

where $\text{Prob}(x_i|G_k)$ represents the conditional distribution of the ith feature when the category of an observation is known. This distribution can be estimated, e.g.,

by the empirical distribution function from training data. The denominator in (3.3) is the joint distribution of all the features. By the law of total probability, this joint distribution is calculated as

$$\text{Prob}(x_1, x_2, \ldots, x_n) = \sum_{k=0}^{1} (\Pi_{i=1}^{n} \text{Prob}(x_i|G_k)). \tag{3.5}$$

With (3.4) and (3.5), the posterior probability (3.2) is

$$P_{post} = \frac{\Pi_{i=1}^{n} \text{Prob}(x_i|G_k)\text{Prob}(G_k)}{\sum_{k=0}^{1} (\Pi_{i=1}^{n} \text{Prob}(x_i|G_k))}. \tag{3.6}$$

The denominator is the total probability so it is a constant for a given set of data. Therefore, the naive Bayes classifier calculates the probability with the maximum posterior law

$$P_{post}^{Max} = \max_{k \in \{0,1\}} \{\text{Prob}(G_k) \Pi_{i=1}^{n} \text{Prob}(x_i|G_k)\}. \tag{3.7}$$

From equation (3.7), it is known that the naive Bayes classifier is a method for carrying out a maximum a posterior probability (MAP) estimation.

3.2.4 New criteria based on Lack of Fitness

As presented upon introduction of LOF in Chapter 2, this method is weak at detecting minimally successful component dyes and easily introduces dyes as false ingredients dyes. When investigating a minimal component dye or a false dye similar to the present dyes during a detection, there is often a contradiction that the LOF improvement is large (small) despite the fitted amplitude's being small (large). In the first example in Figure 13, the LOF improvement of the present dye JOE is only 0.0621, while the fitted amplitude is 0.2531, a significant percentage. In the second example in Figure 15, the false component dye MAX leads to a LOF improvement of 0.1473, larger than the given LOF improvement limit of 0.10, whereas its fitted amplitude is only 0.0362, which is a notable small value. This result is confusing. Upon observing a large LOF improvement, we prefer to accept a component, whereas we trend to reject it upon observing a small fitted amplitude. The LOF method stops here and simply applies the criterion of LOF improvement. Therefore, to improve upon this method, the contradiction discussed above should be further analysed and a new criterion is required based on the LOF improvement.

WALOF analysis regions

To evaluate the amplitude, a parameter of amplitude limit is given as a hard threshold. If an amplitude of a component dye is larger than this threshold in a model

fitting, whether this component dye is present or absent should be further analysed even an LOF improvement occurs with a smaller value than the LOF improvement limit. In addition, it should be double-checked that an amplitude of a component dye is smaller than this threshold while the LOF improvement is larger than the LOF improvement limit.

With the two parameters of the amplitude limit (e.g. 5%) and the LOF improvement limit (e.g. 10%), a coordinate system is set up with the origin at the two limits,with a horizontal axis that is LOF improvement and with a vertical axis that is amplitude in a fitting, as shown in Figure 22. Thus, the whole plane is split into four parts. Part 1: LOF$<$10% and amplitude$>$5%; Part 2: LOF$\geq$10% and amplitude$>$5%; Part 3: LOF$\geq$10% and amplitude$\leq$5%; and Part 4: LOF$<$10% and amplitude$\leq$5%. Among the four parts, by the criteria of LOF improvement, Part 1 and Part 4 are the two regions for which an investigated component is rejected, and Part 2 and Part 3 are regions for which the component dye is accepted. As discussed, this is a coarse criterion.

In view of the different tendencies of an amplitude and an LOF improvement taking place in a fitting, our method of WALOF continues its analysis for an investigated component dye when its amplitude and LOF improvement fall in the coloured regions, as shown in Figure 22. The orange region shows a smaller LOF improvement in $[5\%, 10\%)$ but a bigger amplitude than 5%, which means the investigated component dye might be a true ingredient. However, LOF rejects it, without further considerations. Therefore, WALOF tries to draw back these true components for the orange region.

The blue region shows a larger LOF improvement but a small amplitude, which indicates that it represents a true component dye as a minimal ingredient or a false one similar to present component dyes. If the investigated component is a true ingredient, LOF yields a correct detection, whereas if it is false, then LOF makes an incorrect decision. Therefore, WALOF attempts to remove false component dyes based on LOF for this region.

The green region seems to show consistency of an amplitude and an LOF improvement caused by an investigated component dye during a detection. However, the two parameters of these limits are merely treated simply as fixed values referred for making decisions.In fact, the amplitude and LOF improvement caused by an investigated component might be so complicated that it is difficult to distinguish the component as being present or absent by only two simple limits. Therefore, WALOF is also implemented for the green region in a conservative way. One reason is that the number of false ingredients must be strictly controlled in DNA spectra for medicines, which follows a suggestion from Renishaw.

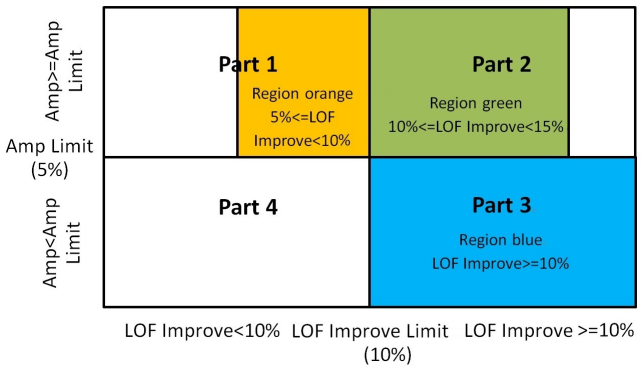


Figure 22: Three regions to call WALOF based on LOF.

WALOF analysis procedure

WALOF aims to improve LOF through four steps. LOF updates the detection model by adding component dyes one by one. When a component dye is added, a pair of values for the two parameters-the amplitude of the added component dye and the LOF improvement of the updated model are obtained after fitting. The procedure of LOF continues until an LOF improvement than 10% occurs in the updated model. During the procedure of LOF, WALOF merely records the fitted amplitudes and the LOF improvements for all component dyes added into the model, and the component dyes which leads to a pair of parameters valued in the colour regions as shown in Figure 22.

Step 1: Investigated component dyes are labelled. There might be several component dyes that lead to pairs of parameters, i.e., the fitted amplitude and the LOF improvement, falling into the same coloured (blue or green) region during a detection. However, the LOF method applies least square fitting which yields the minimal residual, which means that the first component dye reflects the best fitness. If this component dye is rejected as a false ingredient, it is reasonable that the other component dyes should be out of consideration because one task of WALOF is to rejected out the false ingredients with larger LOF improvements. Therefore, only a component dye which leads to a pair of parameters satisfying

- in the region orange, the component with 5% <= LOF improvement < 10% and a fitting coefficient larger than the amplitude limit in the LOF method simulation;
- in the region blue, the first component with LOF improvement >= 10% and a fitting coefficient less than the amplitude limit in the LOF method simulation;
- in the region green, the first component with 10% <= LOF improvement <= 15% and a fitting coefficient larger than the amplitude limit in the LOF method simulation;

being labelled for further investigation in WALOF in the order of their inclusion.

Step 2: WALOF starts its analysis after the determination of LOF. In advance, it identifies peak windows for an observation and the component dyes that are awaiting investigation. This step is implemented by the method of window determination introduced in the last section.

Step 3: Try to find a wavelength range for further analysis to determine whether the investigated component dye is present or absent. To do this, it is certain that the best choice for the range is that in which there is a unique and significant feature of the investigated component dye. For example, WALOF chooses the wavelength range in which the investigated dye alone shows a peak in its spectrum. This means that the chosen range is outside any ranges of windows in other component dyes added to the model. By this criteria of choice, it is easy to understand that the investigated component dye might be a present ingredient if there is a similar window appearing in the chosen range window in the observation; otherwise, it might be false.

Figure 23 shows an example of determination of investigated component dyes and the choice of the wavelength range for WALOF analysis. As seen in this figure, the synthetic observation has the three ingredients Blank, BODIPY_TMRX and ATTO520. Via LOF, BODIPY_TMRX is added to the model; its fitted amplitude is 0.5632 which is larger than the amplitude limit and the LOF improvement is 0.2468 which is also larger than the LOF improvement limit. Therefore, there is no component dye under investigation. LOF goes on adding the next component dye of ATTO520. The two fitted amplitudes are 0.6982 for BODIPY_TMRX, and 0.9428 for ATTO520, both larger than the amplitude limit. The LOF improvement is 0.8223 great bigger than the LOF improvement limit. Therefore, no component dyes are under investigated. The third component dye added in is MAX. The LOF improvement is 0.1473, whereas the fitted amplitude for MAX is 0.0362, which is smaller than the amplitude limit of 5%. Therefore, MAX is a component dye under investigation because $(0.1473, 0.0362)$ falling into the blue region in Figure 22.

Upon comparing the spectra of all component dyes, we find there are two windows ranged with vertical lines, in whose wavelength ranges, MAX shows peaks in its spectrum while the other three are flat in their spectra. Between the two windows, the one between the red lines is the highest. Therefore, the wavelength range between the two red vertical lines is applied for further analysis.

Step 4: Output the probabilities for the presence or absence of an investigated component dye via the naive Bayes classifier. To distinguish a component dye under investigation as present or absent in an observed DNA spectrum is a problem similar to classification. Following the principle of WALOF described in Step 3, it is easy to understand. However, the evaluation of similarity between the residual and the

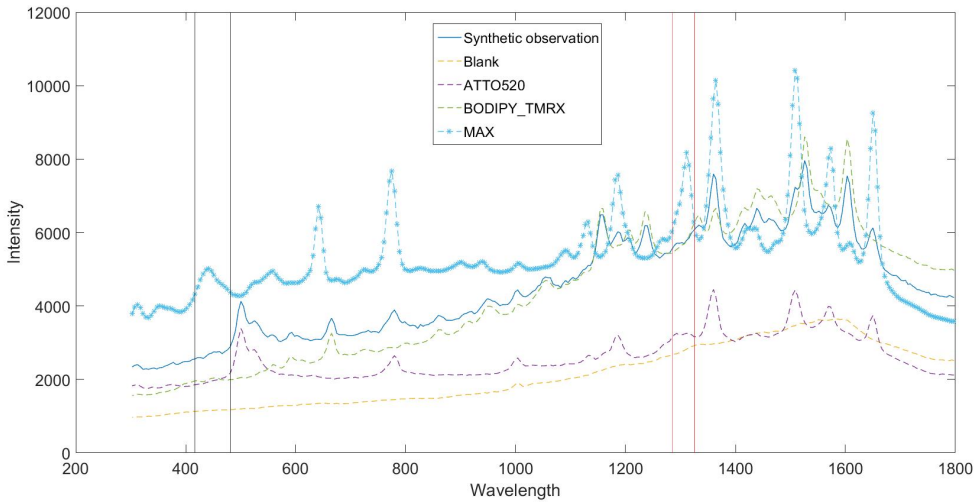


Figure 23: Demonstration of choosing the window for WALOF analysis.

investigated component dye is complicated due to disturbances resulting from noise and variability between spectra of the same ingredient. The naive Bayes Classifier [75], introduced in the last section, is applied in WALOF to output the probability of whether an investigated component dye is present or absent.

Five features are considered as the inputs to the naive Bayes:

- the LOF improvement due to the investigated component dye;
- the LOF improvement within the analysis range due to the investigated component (referred to as window LOF);
- the shape correlation (SC), where SC is measured by the shape correlation coefficient between the residual and the investigated component in the analysis range;
- the deferential Shape Correlation (DSC), which is the same as SC but uses the first derivative of the spectra in the analysis range; and
- the ratio of the fitted amplitude of the investigated component dye, to the amplitude limit.

Step 5: Optimisation based on the naive Bayes classifier. Knowing the posterior probability of presence or absence, a simple method is sampling to determine whether the investigated component present or absent. However, this is a coarse method regardless of differences in the training data with respect to the different analysis regions, orange, green and blue in Figure 22. For example, for an investigated component leading to an LOF improvement and an amplitude falling in region orange, we aim to accept a component via a strict check such the posterior probability of presence should be considered at a high value. However, for one in region blue, we might accept a finding with a lower posterior probability because of the

significant LOF improvement. Therefore, different limits are set for the posterior probabilities for rejections or acceptances.

A new method is proposed to make a decision based on these posterior probabilities outputted by Naive Bayes. Three parameters are introduced as probability limits, (p_O, p_B, p_G), for three analysis regions, respectively. For a component dye that leads to an amplitude and an LOF improvement falling into a region, it is determined to be present if its posterior probability larger than the probability limit for this region. An assessment is constructed as a function of the three parameters, and it is optimised to estimate the best parameters with the training data. Then, a decision is made regarding to the presence or absence of the investigated component dye by the estimated parameters and the posterior probabilities. The details of optimisation are discussed in the next subsection.

Finally, the flow chart of WALOF is given in Figure 24, where grey boxes represent the original LOF method. The orange box shows the WALOF analysis for the orange region, the blue box for the blue region and green boxes for the green region. As seen in this figure, there might be several component dyes under investigation during a detection.

3.2.5 Optimisation based on naive Bayes

In this subsection, a function of three parameters as probability limits is proposed to construct an optimal model for a determination that an investigated component dye is present or absent in an observed DNA spectrum. In advance, several concepts on classification are introduced for the function:

- True Positive (TP): a component dye is a true ingredient, and it is determined to be present;
- True Negative (TN): a component dye is a false ingredient, and it is determined to be absent;
- False Positive (FP): a component dye is a false ingredient, but it is accepted; and
- False Negative (FN): a component dye is a true ingredient, but it is rejected.

Based on the four concepts, two measurements for evaluation of a classifier are commonly used, sensitivity and specificity. Sensitivity is defined as the true positive ratio (TPR)

$$\mathrm{TPR} = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}}, \tag{3.8}$$

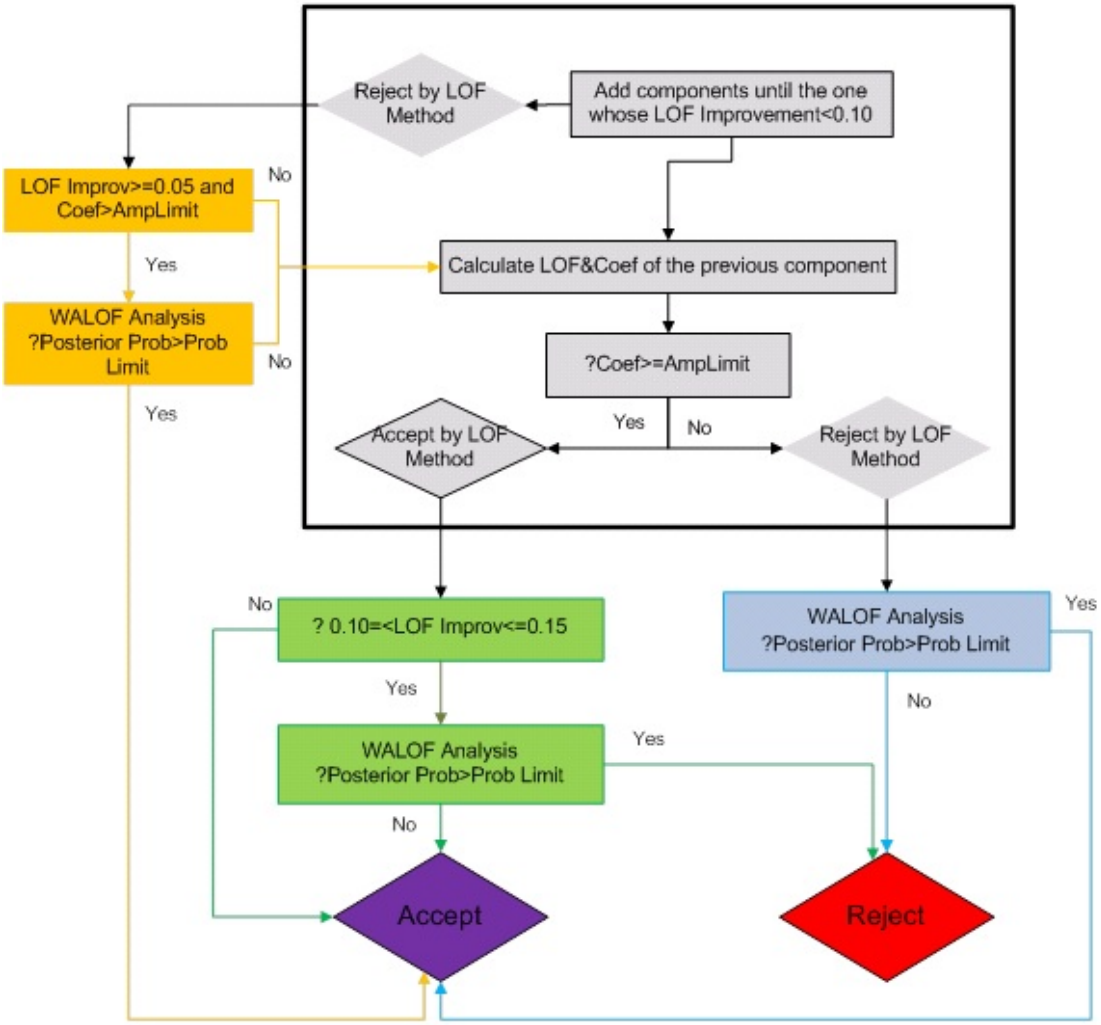


Figure 24: Flow chart of the WALOF method compared to the original LOF method.

and specificity is defined as the true negative ratio (TNR)

$$\mathrm{TNR} = \frac{N_{\mathrm{TN}}}{N_{\mathrm{TN}} + N_{\mathrm{FP}}}, \tag{3.9}$$

where $N_{\mathrm{TP}}, N_{\mathrm{TN}}, N_{\mathrm{FP}}, N_{\mathrm{FN}}$ represent the numbers of TP, TN, FP and FN findings in a classification. As seen, sensitivity is the probability of successful decisions in the positive group whereas specificity is the probability of successful cases in the negative group. Therefore, high sensitivity means few FN findings and high specificity indicates few FP findings.

Although the naive Bayes classifier is applied in our method, the evaluation by TPR and TNR cannot be applied directly for LOF and our method because DNA detection is not completely a problem of classification. First, what the investigated component is depends on the LOF limit and the amplitude limit. Therefore, the numbers of positive and negative cases are variables. Second, in a detection for ingredients in a DNA spectrum, there are often several investigated component dyes under consideration. For example, there are two false component dyes accepted by the LOF method and they are under investigation in WALOF. If they are both rejected by our method, then it is reasonable to evaluate by $N_{\mathrm{TN}} = 2$ but not $N_{\mathrm{TN}} = 1$. Third, in a problem of classification, it is impossible that the FP and FN findings occur for the classification of an object. However, this case might take place in a detection, in which a false component dye is accepted and a true one is rejected. This case does not occur in our experiment, but there is not evidence to prove that it is impossible.

We denote that

- GO is the group of component dyes that lead to a fitted amplitude and an LOF improvement in region orange in Figure 22;
- GB is the group of component dyes that lead to a fitted amplitude and an LOF improvement in region blue in Figure 22;
- GG is the group of component dyes that lead to a fitted amplitude and an LOF improvement in region green in Figure 22;
- G represents all component dyes under investigation;
- G_1 is the set of component dyes present as true ingredients, that is known for the training data;
- G_0 is the set of component dyes actually absent, that is known for the training data.

For a component dye O under investigation, it must be at least in one set among the three groups, GO, GB and GG. The component dye is accepted when the posterior

probability of acceptance exceeds the probability limit, $\text{Prob}(O \in G_1) > p$. In contrast, it is rejected when the posterior probability $\text{Prob}(O \in G_1) \leq p$. Therefore, the TPR is calculated to be

$$\text{TPR}(p_1, p_2, p_3) = \frac{\sum_{O \in G_1 \bigcap GO} \delta(P^{G_1}(O) > p_O) + \sum_{O \in G_1 \bigcap GB} \delta(P^{G_1}(O) > p_B) + \sum_{O \in G_1 \bigcap GG} (P^{G_1}(O) > 1 - p_G)}{\text{Number of members in } G_1}, \tag{3.10}$$

and TNR is calculated to be

$$\text{TNR}(p_1, p_2, p_3) = \frac{\sum_{O \in G_0 \bigcap GO} \delta(P^{G_1}(O) \leq p_O) + \sum_{O \in G_0 \bigcap GB} \delta(P^{G_1}(O) \leq p_B) + \sum_{O \in G_0 \bigcap GG} \delta(P^{G_1}(O) \leq 1 - p_G)}{\text{Number of members in } G_0}, \tag{3.11}$$

where $\delta(\cdot)$ is the characteristic function taking on a value of 0 or 1, and $P^{G_1}(O) = \text{Prob}(O \in G_1)$, the posterior probability of presence of component dye O.

Up to now, we have evaluated a method for detection with the following function

$$\mathcal{A}(p_O, p_B, p_G) = \mathit{Balance} * \text{TPR} + (1 - \mathit{Balance}) * \text{TNR}, \tag{3.12}$$

where *Balance* is a subjective parameter depending on the preference for TPR or TNR. It is fixed to adjust the numbers of FP and FN findings. A small value of *Balance* provides a preference for low FP findings or high TNR and a large one provides a preference for low FN findings or high TPR. As seen from equation (3.12), *Balance* indicates the relative importance of the two error classes, FN and FP. For example, when *Balance* is 50%, we can increase TPR by 1% at the price of a 1% reduction in TNR; if the *Balance* equals to 75%, an increase of 1% in TPR leads to a 3% decrease in TNR. According the definition of TPR and TNR, a smaller parameter results in fewer FP and more FN detections. In the case of ingredient detections in medicines, Renishaw prefers rejections of true minimal ingredients, rather than acceptances of false ingredients. Thus, a smaller parameter will be fit to Renishaw's choice.

By substituting the TPR (3.10) and TNR (3.11 into equation (3.12), we have

$$(p_O^*, p_B^*, p_G^*) = \underset{0.5 \leq p_O, p_B, p_G \leq 0.95}{\arg\max} \mathcal{A}(p_O, p_B, p_G). \tag{3.13}$$

This optimal problem is nonlinear. A grid method is applied to yield an estimation of probability limits. In the orange and blue region, the fitted amplitude and LOF improvement indicate opposite decisions against each other. To strictly control the

number of false ingredients included, the investigated component dye is accepted by a posterior probability greater or equal to 0.5. In the green region, the fitted amplitude and LOF improvement are both moderate, so the probability limit is lower. Therefore, we put cubic grids with a size of 0.01 in a three-dimensional space whose three axes range from 0.5 to 0.95. The vertex, (p_O^*, p_B^*, p_G^*), that leads to the maximum of $\mathcal{A}(p_O, p_B, p_G)$, is the estimation for the best parameters. The estimated parameter will be used for any test dataset.

3.3 Experiments and results

Five data sets, one for training and four for testing, are generated to test the stability of WALOF. Each data set contains 20,000 random observations. The configurations of parameters in all five data are the same. In the LOF method, the LOF improvement limit is 10%, which is a value suggested by Renishaw. In WALOF, the amplitude limit is 5% and the LOF improvement limit is 10%, the same as in LOF. Three values, 0.10,0.50 and 0.75, are tried for the parameter to balance the numbers of FN and FP. The result is adopted by Renishaw when $Balance$ is set to be 0.1, because they prefer low FP findings due to higher risk caused by FP findings than that by FN findings. The other two values of $Balance$ lead to results with no preference and a higher preference for FN findings, respectively. The requirement of accuracy from Renishaw is that no more than 100 false cases occur in per 20,000 detections. Therefore, only the numbers FN and FP findings are listed.

As seen in Table 2, the False Positive and False Negative numbers change in general with the "Balance" parameter. However, when the balance reaches to a certain value (0.5 in this case), the two numbers remain the same. The numbers of FP and FN findings via WALOF with a smaller $Balance = 0.1$ are both less than those via LOF. However, the FP number of 49 via WALOF becomes larger than that of 41 via LOF when the parameter increases to 0.5, but the total number is still smaller because of a further decrease in FN findings from 51 to 40.

With the optimal parameters of probability limits, WALOF yields the results of detection for the four test data, as shown in Table 3. Comparing the numbers for test data with the corresponding data for training, there are no obvious fluctuations; thus, we can draw the conclusion that both LOF and WALOF show good stability. By the total numbers of false detections, WALOF leads to a reduction of an approximately 25% compared to LOF.

Table 2: WALOF vs. LOF for training data.

Methods	FN Numbers	FP Numbers	Total
LOF(LOF improv=10%)	65	41	106
WALOF(Bal=0.10)	51	31	82
WALOF(Bal=0.50)	40	49	89
WALOF(Bal=0.75)	40	49	89

Table 3: WALOF vs. LOF for test data.

	Methods	FN Numbers	FP Numbers	Total
Test data one	LOF(LOF improv=10%)	68	50	118
	WALOF(Bal=0.10)	47	45	92
	WALOF(Bal=0.50)	40	56	96
	WALOF(Bal=0.75)	40	56	96
Test data two	LOF	62	48	110
	WALOF(Bal=0.10)	40	35	75
	WALOF(Bal=0.50)	34	50	84
	WALOF(Bal=0.75)	34	50	84
Test data three	LOF	62	43	103
	WALOF(Bal=0.10)	42	31	73
	WALOF(Bal=0.50)	31	56	87
	WALOF(Bal=0.75)	31	56	87
Test data three	LOF	71	35	106
	WALOF(Bal=0.10)	53	35	88
	WALOF(Bal=0.50)	33	44	77
	WALOF(Bal=0.75)	33	44	77

3.4 Summary

This chapter introduces the principles and the procedure of WALOF. In certain ways WALOF solves the problem of determining of minimal and false ingredients during a detection of a DNA spectrum. Although the LOF method offers advantages in the form of a quick and efficient method for significant ingredients, it is weak at dealing with minimal components and false components that are similar to true ingredients due to its simple criteria of the LOF improvement limit. From the results listed in the last section, the number of false detections via LOF is reduced by approximate 25% with the adoption of WALOF, should one be content with the accuracy requirements established by Renishaw. In addition, WALOF shows good stability as does LOF.

The reason for the obvious improvement of WALOF compared to the LOF method is that our algorithm performs stricter detections for minimal ingredients and false ingredients through similarity analysis on their unique features in individual windows. For a true ingredient, there must be responses to these unique features from

the residuals, even the LOF improvement is lower than the pre-set limit. In contrast, receiving no responses from the residual to the features indicates that the component dye might not be present despite higher LOF improvement. WALOF grasps the unique features of an investigated component dyes; therefore, it can refine a more elaborate criterion for DNA detection based on LOF.

In addition to offering improved accuracy and good stability, WALOF offers there several advantages:

1. A smart method developed for features segmentation for component dyes, which saves time and is convenient for practical usage;
2. A quick and efficient method whose time consumption is almost as same as that of LOF; and
3. A method showing a powerful flexibility. This method can be combined with other detection algorithms, such as AIC and BIC. Due to noise and variability, by using maximum AIC or BIC alone, it is possible to select an inaccurate model that is similar to the true one. Just as with LOF improvement criteria, AIC and BIC are simple criteria which may not be sufficient for reveal the complicated features. Thus, WALOF can be applied for those models for which AIC or BIC might falling into a possible range containing the true model.

WALOF also has several disadvantages. For example, a few false detections via LOF can be corrected by manual comparisons to the residual, an investigated component dye and the observed DNA spectra, whereas WALOF might fail. This shows that the similarity analysis in WALOF is not enough intelligent for complicated recognitions. To improve the performance of WALOF, more measurements may be considered in a higher space.

Chapter 4

Adaptive Regularised Steepest Descent Method for Image Restoration

As discussed in Chapter 2, it is difficult to undertake error analysis in regularization methods due to the complexity of regularization factors. To address this problem, we present a new image restoration method based on an adaptive regularised gradient. Our approach decouples between deblurring and denoising during the restoration process through an iterative process; therefore, the approach allows any well-established noise reduction operator to be implemented in our model. Such an approach has led to an analytic expression for the error estimation of restored images in our method as well as simple parameter setting for real applications, both of which are difficult to achieve in many regularization-based methods. Numerical experiments show that our method can achieve a good balance between structure recovery and noise reduction, and the method performs similarly to state-of-the-art methods and favourably compared to other methods.

This chapter is organised into five sections. In the first section, image degradation and common point spread function (PSFs) are introduced. Next, the ill-posed problem mentioned in Chapter 1 is explained in mathematical terms. In the third section, our method is proposed and the error is estimated. The last two sections describe experiments and provide a summary.

4.1 Image degradation and common point spread functions

4.1.1 Image degradation

Images present useful information or original scenes that cannot be observed with the human eye; therefore, they play important roles in society. Many imaging techniques have been developed to present objects under investigation in different forms for applications in many different areas. For example, in biology and medicine [83], images are generated for researching the structures and functions of living cells in the study of life and for curing diseases. The fluorescence tagging of objects of interest is a general method that has been widely applied in biotechnology. In this method, fluorescence is attached to the objects as labels and is excited using light; the illumination of the objects is then imaged through an optical microscope. Green fluorescent protein (GFP), which is extracted from the light-producing organ of a jelly fish, is a fluorescent material widely used by biologists.

However, the quality of images can be far from perfect. There are two main reasons for image degradation: blurring and noise. Blurring effects are generally a bandwidth limit on an image during imaging processes as a result of different causes. For example, due to optical diffraction, the image of a point source is diffused to an Airy disk in an optical microscope. When an image is recorded by a remote sensor through the atmosphere, blur can be caused by atmospheric turbulence. Moreover, blur can also be induced by the movement of objects or cameras, referred to as motion-induced blur. PSFs are commonly used to describe blurring effects. Although there are many sources for image blurring, as discussed above, these effects can largely be approximated by a linear system.

Noise plays an important role in poor image quality. A low level of noise is typical in output images in many fields such as cell molecular biology, medical science and geology. For example, to observe distinct structures in cells, noise appearing in output images must be kept at a very low level [84]. There are also similar standards of noise control required in medical science imaging [85].

Noise originates from many sources. It can be generated during the process of transmission due to the medium or during recording due to the sensor. The limitations of measurements and technology also cause noise. For instance, instruments have limited precision, and digitalisation algorithms for image retrieval always produce noise. Concerning noise generated by instruments, we can reduce such noise by improving performance of machines, e.g., improving the purity of medium materials to reduce noise during transmission. Moreover, human factors and environments, such as incorrect operation and photographing in high-temperature environments,

are also noise sources.

For noise caused by human factors, improvements of instruments are not effective. One method to solve this problem is image processing with developed algorithms following imaging procedures, which is less expensive than improving instruments. Because there are so many different sources of noise, systems of noise are complicated and possess uncertain factors. Therefore, they are usually described in terms of probability distributions. In current algorithms, there are three types of common distributions assumed in imaging systems: Gaussian, Poisson and salt-and-pepper noise.

Reducing the level of image degradation is an important task for achieving high image quality. In the space programmes of both the United States and the former Soviet Union in the 1950s, many incredible images of the Earth and our solar system were generated. However, the images obtained from the various planetary missions were subject to substantial photographic degradation because of the imaging environments, i.e., vibrations in machinery and the spinning of the spacecraft. The degradation reduces the scientific values of such images, which is a serious problem because space exploration is always expensive. For example, the 22 pictures produced during the Mariner IV flight to Mars in 1964 were later estimated to cost almost 10 million dollars in terms of the number of bits transmitted alone.

Many methods have been developed to reduce image degradation and improve image resolution, for example, in the development of confocal microscopes [86]. However, precise instruments are always expensive. An alternative method to achieving high image quality is to process observed (low-quality) images through computational algorithms. In this thesis, we focus on the latter through the development of novel image restoration methods.

4.1.2 Common point spread functions

A PSF is the optical field distribution produced by a lens that transforms a plane wave front into a converging spherical wave. The PSF usually plays the dominant role in image deblurring, but in many practical situations it is extremely difficult to generate the exact PSF of an imaging system. Although the measurement of a PSF is not our focus in this research (because PSF is assumed to be known), we would like to provide a simple review. One way to measure the PSF is through testing with interferometer-based equipment; however, this is not an easy task and is relatively expensive [87]. Another method is image identification, which attempts to estimate the PSF from images with blind algorithms such as maximum likelihood (ML) [88][89], Bayesian blind deconvolution [90][91], and sparse representation [92].

Because of the difficulty in accurately measuring a PSF, powerful PSF models have

been developed to estimate PSFs that are frequently used in various fields. We present these PSFs in their continuous form because the blurring of images is a continuous process. As a result of blurring in images, it is reasonable that the energy in the original scene should be preserved because the blurring process does not lead to any change in energy. Therefore, for any PSF, its total integral should be equal to 1.

Linear motion blur

In the following, four common PSFs will be introduced. The first PSF is linear motion blur. A popular application in the traffic field is vehicle license plate recognition, in which distinguishing the translation blur is important. Let T be the time duration of the exposure. Supposing that a vehicle is travelling along a clockwise direction at an angle θ with the horizontal axis y and at a velocity v, the PSF is defined as [93]

$$f_M(x,y) = \begin{cases} \frac{1}{vT}, & \text{for } \sqrt{x^2+y^2} \leq \frac{vT}{2} \text{ and } x/y = -\tan(\theta), \\ 0, & \text{else.} \end{cases} \tag{4.1}$$

where x, y are vertical and horizontal axis.

As seen in this model, a distortion is generated in one dimension in the direction of motion. Moreover, the PSF is space-invariant. In addition to translation, there are many other types of motion blur, such as rotation. The identification of other types of motion blur can be found in [94].

Uniform defocus blur

When a three-dimensional object is imaged onto a two-dimensional plane, some parts of the objects will be out of focus due to the depth of the lens. The image of a point through circular apertures is a small disk, known as the circle of confusion. The diameter R of the circle of confusion is determined from the focal length L and aperture number η of the lens and the distance between the camera and the object d, which is given by [93]

$$R = \begin{cases} \frac{Lf}{\eta(f-L) - \frac{L^2 f}{\eta d(f-L)} - \frac{L}{\eta}} \text{ for } f < d < \infty \\ -\frac{Lf}{\eta(f-L) + \frac{L^2 f}{\eta d(f-L)} + \frac{L}{\eta}} \text{ for } L < d < f \end{cases} \tag{4.2}$$

where f is the focus of the lens. The intensity distribution within the circle of confusion is uniform when the degree of defocusing is far larger than the wavelengths

[95], which indicates that the PSF of a uniform out-of-focus blur can be defined as

$$f_U(x,y) = \begin{cases} \frac{1}{\pi(R/2)^2}, \text{ for } \sqrt{x^2+y^2} \leq \frac{R}{2}. \\ 0, \quad \text{else.} \end{cases} \tag{4.3}$$

Gaussian blur

Gaussian blur is generally assumed for numerous applications such as biological imaging through microscopy, weather prediction and emergency disaster management. To control the image size, biological imaging always utilises small pinholes, which leads to Gaussian blur. In weather prediction and emergency disaster management, the techniques of remote sensing and aerial imaging are often applied for tracking so that changes can be clearly observed in a timely manner. Images generated by these techniques are always inevitably blurred due to atmospheric turbulence. Although the blur introduced by atmospheric turbulence depends on a variety of factors, such as temperature, wind speed, and exposure time, the PSF can be well represented by a Gaussian function [96], defined as

$$f_G(x,y) = C \exp\{-\frac{x^2+y^2}{2\sigma^2}\} \tag{4.4}$$

where C is a uniform constant and σ is the standard deviation (STD) of the Gaussian function, which describes the severity of the blurring. A smaller STD results in a larger blur disk.

X-ray blur

X-ray imaging plays an essential role in the fields of medicine, safety and non-destructive detection due to its penetrability. When X-rays pass through an object, the remaining energy is different because the energy is not uniformly absorbed by different parts of the object, thereby yield different intensity distributions which details about the object can be determined. However, X-ray quanta are scattered from their paths, resulting in blur. The PSF can be accurately established in the following symmetric form [97]:

$$f_X(x,y) = \frac{C}{(\xi^2+x^2+y^2)^{3/2}}, \tag{4.5}$$

where C is a factor to make the integral of $f_X(x,y)$ be 1 in x-y plane and ξ is the parameter describing the degree of blurring.

PSF discretisation

Although PSFs are continuous in theory, digital images that we record through an imaging system are usually presented in discrete form, usually as a matrix of numbers. Therefore, it is necessary to discretise the PSF from its continuous form for image modelling. A simple way is to use a square sensor array to sample the PSF, which can be denoted as

$$f_d(s,t) = \int_{\Omega} f(x,y)\mathrm{d}x\mathrm{d}y, \tag{4.6}$$

where Ω is a square area centred at pixel position (s,t) and of size 1. As the impulse for the pixel intensity, the values of the PSFs cannot be negative or complex values. In addition, PSFs conserve energy; thus, a discrete PSF should be constrained by

$$\sum_{s=-\infty}^{+\infty}\sum_{t=-\infty}^{+\infty} f_d(s,t) = 1. \tag{4.7}$$

Obviously, we cannot choose a pixel in an infinite position and have to truncate an infinite PSF as a finite support.

Demonstration of different blurring

These blurring filters mentioned above will be applied as the linear operators in the ill-posed problem introduced in the next section. Figure 25 demonstrates the blurring filters. As seen in this figure, the blurring by motion shows an obvious direction along 45° in (b). The uniform blur (c) seems to be blurred seriously, with main features disappeared, e.g, the eyes of the man. The Gaussian blurred image (d) with a bigger STD preserves the main feature but is blurred for edges and small structures, e.g., bottoms on the camera. The Gaussian blurred image (e) with a smaller STD exhibits little blurring; it merely makes the original image coarse. The X-ray blurred image (f) seems to be similar to the Gaussian blurred image (d). Although the blurred by Gaussian image with a large STD seems to better preserves the features of the original scene than the other types, the restored image often seems inferior compared to those restored from the other types of blur. The reason for this is that Gaussian filter with a moderate STD has a very small eigenvalues, which causes strong amplification of noise. This phenomenon will be described in the next section.

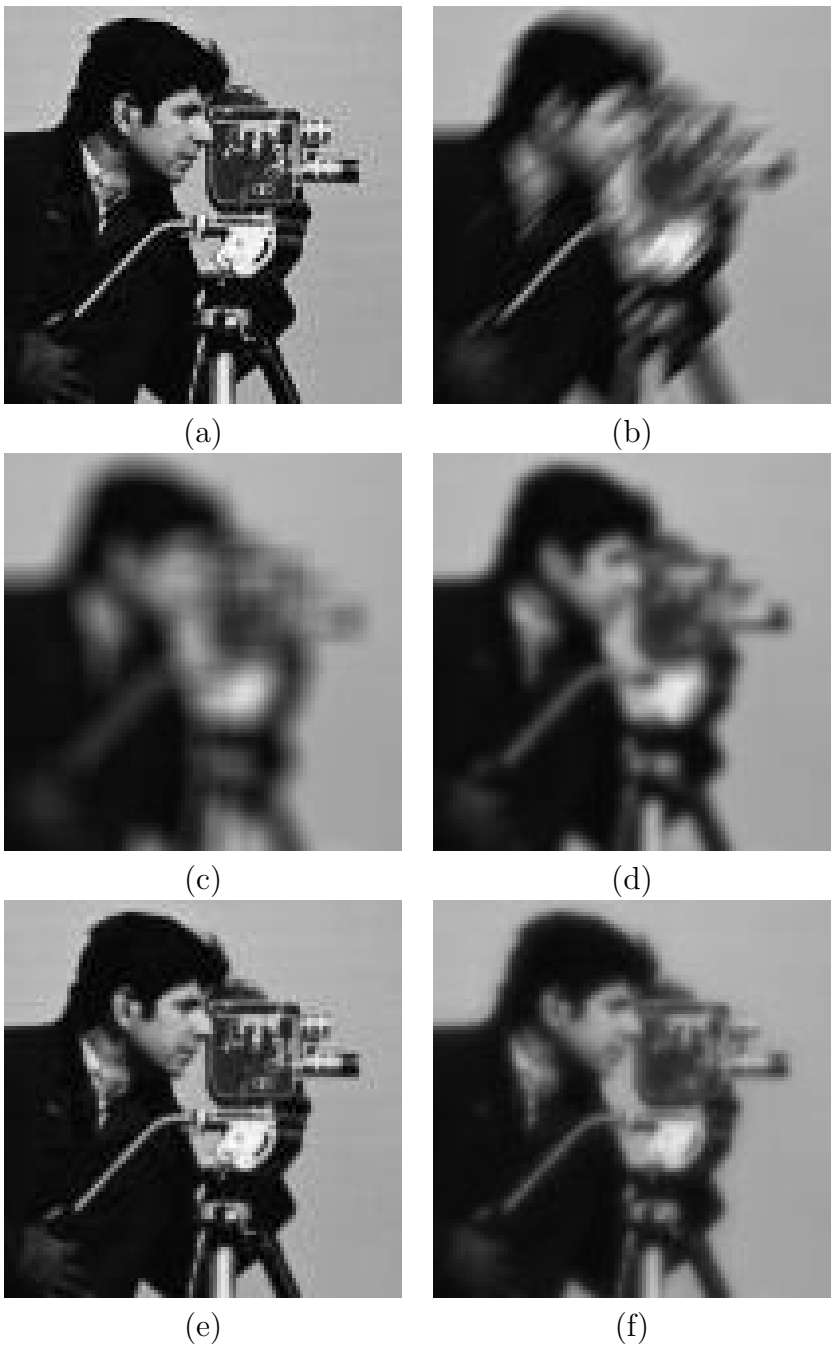


Figure 25: Demonstration of blurring filters on the fragment of image "Cameraman.jpg". (a) The original scene, (b) a blurring by motion filter with $\theta = \pi/4$ and $vT = 9$ pixels, (c) a blurring by uniform filter in size 9×9, (d) a Gaussian type blurring with STD of 1.6, (e) a Gaussian type blurring with STD of 0.5, and (f) a X-ray blurring image with $\xi = 1$.

4.2 Ill-posed problem

In this section we will review the ill-posed problem introduced in Chapter 1 and provide a mathematical explanation. The reviewed imaging model is

$$J = PI + N, \tag{4.8}$$

where I and J are the real scene and the corresponding observation, respectively; P is the PSF; and N is noise which is assumed to be independent of the real scene. In this model, the PSF P is assumed to be a circulant matrix and invertible. The simplest way to estimate the real scene from the observation is to minimising the energy function by the LS method

$$\min_I \|J - PI\|_2^2 . \tag{4.9}$$

However, this is a well-known ill-posed problem [98]. When attempting to solve the above energy function, the Least Squares (LS) algorithm leads to false solutions because the noise in the output images is infinitely amplified due to the ill-conditioning. This phenomenon can be explained based on the theory of inverse operators.

It is known that the LS solution is equal to that obtained by the direct inversion $P^{-1}J$ analytically. We substitute J with $PI + N$ and obtain

$$P^{-1}J = P^{-1}PI + P^{-1}N = I + P^{-1}N. \tag{4.10}$$

A noise term with an amplification factor P^{-1} can be obtained from the above equation. Because the PSF matrix P is often referred to as a "low-pass filer", it has very small eigenvalues. Because of these small eigenvalues, the noise is infinitely enlarged; thus, the LS solution fails. An explanation will be given along with mathematical deductions.

In Chapter 2, VanCittert's algorithm was introduced to restore an image via regularization of prior information, which guarantees a well-posed problem to be solved. Here, we will explain how the small eigenvalues of lead to noise amplification through VanCittert's iteration when neglecting regularization. Without considering regularization for noise reduction, VanCittert's iteration can be formulated as follows:

$$\begin{cases} I_0 = 0, \\ I_{k+1} = I_k + \beta(E - PI_k), \text{ for } k = 1, 2, \ldots . \end{cases} \tag{4.11}$$

The solution of (4.11) is calculated by enumeration as

$$I_{k+1} = \beta \sum_{i=0}^{k} (E - \beta P)^i J, \tag{4.12}$$

where E is the identity matrix and k is the step parameter used to control the convergence of the iterations. This solution can also be written as

$$\begin{aligned} I_{k+1} &= \beta(E - (E - \beta P))^{-1}(E - (E - \beta P)^{k+1})J \\ &= P^{-1}(E - E(\beta P)^{k+1})J. \end{aligned} \tag{4.13}$$

If β is assumed to satisfy the sufficient condition for convergence

$$\lim_{k\to\infty}(E-\beta P)^{k+1}=0, \tag{4.14}$$

the limiting solution is equal to

$$I_\infty = P^{-1}J. \tag{4.15}$$

This shows that the VanCittert's iteration converges to the inverse filter.

Letting Z_{uv} denote the eigenvectors of the matrix P and letting ζ_{uv} denote the corresponding eigenvalues, where u, v are pixel numbers denoting the image size, the iterative solution can be expanded in terms of the following eigenvectors:

$$\begin{aligned} I_k = &\sum_{u,v}(1-(1-\beta|\zeta_{uv}|^2)^k)(I,Z_{uv})Z_{uv} \\ &+\sum_{u,v}\frac{1}{\zeta_{uv}}(1-(1-\beta|\zeta_{uv}|^2)^k)(N,Z_{uv})Z_{uv} \quad \text{for } k=1,2,\dots \end{aligned} \tag{4.16}$$

Because P is supposed to be a low-pass filter, which has small eigenvalues, the noise is amplified by the factor $1/\zeta_{uv}$. Therefore, this problem is ill-posed and cannot be solved by the LS method.

Here, we offer an example to explain the noise amplification. The error of the noise amplification in (4.16) is measured by the L_2 norm for the "checkerboard" images in Figure 26. There are 1000 iterations as showed in this figure. As seen in the plot, the error increases through all the iterations. The error is approximately 0.11 at the beginning and increases to an initial level of 0.58 at the 22nd iteration. Finally, the error is 4.4, approximately 7.5 times the initial error.

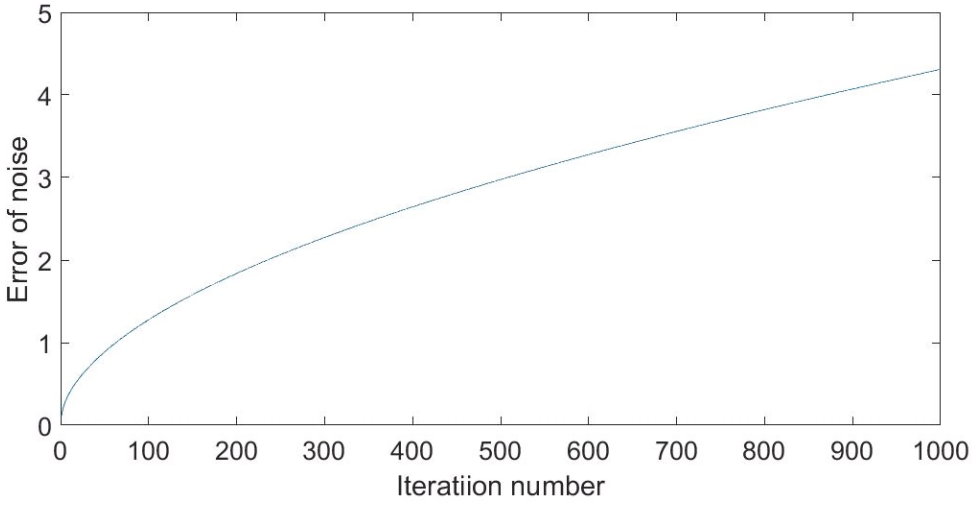


Figure 26: Noise amplification error measured by the L_2 norm.

It is known that the factor responsible for the ill-posed problem is the set of small eigenvalues of the linear operators. Table 4.2 lists out the maximum and minimum eigenvalues of the blurring filters introduced in section 1. As seen in the table, all of the filters have the same maximum eigenvalue of 1 while their minimum eigenvalues are different. The Gaussian blur with a larger STD has the smallest eigenvalue, $6.4944e-10$, which means that noise inflates at a very fast speed during the restored procedure if there is no control.

Table 4: Maximum and minimum eigenvalues of several linear operators.

Blurring type	Maximum	Minimum
Motion	1.0000	-1.2845e-4
Uniform	1.0000	7.6593e-6
Gaussian(STD=1.6)	1.0000	6.4944e-10
Gaussian(STD=0.5)	1.0000	0.3297
X-Ray	1.0000	0.0535

4.3 Adaptive regularised steepest descent method

As discussed in Chapter 2, it is difficult to undertake error analysis in regularization methods due to the complexity of regularization factors. To address this problem, we present a new method, called adaptive regularized steepest descent method (ARSD), to restore the original scene based on an adaptive regularized gradient. Our approach decouples deblurring and denoising during the restoration process through an iterative process; therefore, the approach allows any well-established noise reduction operator to be implemented in our model. Such an approach has led to an analytic expression for the error estimation of restored images in our method as well as simple parameter setting for real applications, both of which are difficult to achieve in many regularization-based methods. Numerical experiments show that our method can achieve a good balance between structure recovery and noise reduction, and the method performs similarly to state-of-the-art methods and compared favourably to other methods.

4.3.1 Gradient regularization

The method of ARSD implements image restoration based on the theory of inverse operators via gradient regularization. This method introduces an iterative energy function that forces the gradient corrupted by noise back into the regularized direction through noise reduction. A solver is designed for this energy function through combining VanCittert's iteration with a non-local noise reduction filter. The former recovers structures but amplifies noise, whereas the latter reduces noise through

collaborative filtering. Therefore, the combination of the two operations can restore the real scene with minimal noise. Mathematically, it can also be understood that the ill-posed problem is transformed into a pure noise reduction problem, which is a well-posed problem.

We present the ARSD method for the restoration of images degraded by blur and noise. The key to our approach is an iterative energy function proposed to regularize the corrupted gradient in the steepest descent by adaptively adjusting to the current state in the iterative process so that the iterative solutions converge to the original scene despite noise contamination in the observation. This approach is general and can be implemented by combining VanCittert's iterative algorithm with a well-established noise reduction method.

As discussed in the last section, the simplest way to estimate the real scene from the observation is minimising the energy function (4.9), which can be solved by an iterative algorithm through the steepest descent, known as VanCittert's algorithm. Theoretically, VanCittert's algorithm yields the inverse operator PSF matrix; thus, its iterative solutions converge to $P^{-1}J = I + P^{-1}N$. Therefore, when the observation is noise free, the iterative solutions converge to the real scene, whereas when is contaminated by noise, the gradient is corrupted, leading to noise amplification during the iterations [3]. As discussed in Chapter 2, many regularization methods have been developed to address this problem. By introducing a penalty (regularization term) into the energy function 4.9, the ill-posed problem becomes well-posed, and the latter can be solved by various approaches such as Wiener filter deconvolution (WFD), maximum-likelihood (ML), maximum a posteriori (MAP) and Tikhonov regularization. However, the introduction of penalties will invariably accompany complicated regularization, leading to difficulties in finding the optimal factors and in error analysis. This is a drawback of the regularization methods.

To avoid the complexity of regularization factors, we develop a robust iterative algorithm based on the theory of inverse operators. To this end, we propose a set of iterative energy functions,

$$I_k = \min_{I_r} \|I_r - D_k\|_2^2, \text{ for } k = 1, 2, \ldots, \tag{4.17}$$

where D_k is defined as

$$D_k = \sum_{u,v} (1 - (1 - \beta|\zeta_{uv}|^2)^k)\langle I, Z_{uv}\rangle Z_{uv}. \tag{4.18}$$

In this definition, $\{\zeta_{uv} : u = 1, 2, \ldots, R, v = 1, 2, \ldots, C\}$ and Z_{uv} are the eigenvalues and eigenvectors of the PSF matrix P; R and C represent the image size; and u, v are the indices of the image pixels. As seen in (4.18), D_k converges to the real scene,

and the energy functions restores the real scene via a series of estimations of D_k for $k = 1, 2, \ldots$

4.3.2 Optimisation of the iterative cost functions

Because the real scene is unknown, we optimise the energy functions (4.17) through a combination of VanCittert's algorithm and a non-local noise reduction filter. We first explain the energy functions by introducing VanCittert's algorithm. This algorithm has a long history as a simple and efficient approach to image restoration. It is formulated for spatially invariant or variant restoration problems when neglecting noise contribution in (4.8). Originally, it was a steepest descent method where the solution did not converge if the step parameter was assumed to take on real values. To overcome this problem, an iterative procedure was proposed [50],

$$I_k = I_{k-1} + \beta P^T (J - P I_{k-1}), \tag{4.19}$$

which searches a solution along the gradient of (4.9), with β being the iteration step. When noise is present in an observed image J, VanCittert's algorithm (4.19) can be expressed as [3][51][52]

$$\begin{aligned} I_k &= I_{k-1} + \beta P^T (J - P I_{k-1}) \\ &= I_{k-1} + \beta P^T (J - P I_{k-1} - N) + \beta P^T N \\ &\sum_{u,v} (1 - (1 - \beta |\zeta_{uv}|^2)^k) \langle I, Z_{uv} \rangle Z_{uv} + \sum_{u,v} \frac{1}{\zeta_{uv}} (1 - (1 - \beta |\zeta_{uv}|^2)^k) \langle N, Z_{uv} \rangle Z_{uv} \\ &= D_k + \sum_{u,v} \frac{1}{\zeta_{uv}} (1 - (1 - \beta |\zeta_{uv}|^2)^k) \langle N, Z_{uv} \rangle Z_{uv} \end{aligned} \tag{4.20}$$

The last expression in (4.20) shows that iterative solutions can be expressed as the sum of the noise-free iterative solutions and the noise contribution. That finding explains why the energy functions (4.17) are proposed for image restoration. Additionally, the noise is amplified by the factor $1/\zeta_{uv}$ for $k \to \infty$ for small values of ζ_{uv}, which explains why (4.9) is ill-posed. Equation (4.20) implies that if we can suppress noise at each iterative step, the iterative solutions can be maintained in the vicinity of the noise-free iterative solutions; therefore, the iterative solutions converge to the real scene. From the last expression in (4.20), a sufficient condition for the iterative process to converge is that $\left|1 - \beta \left|\zeta_{uv}\right|^2\right|$ must fall within (0,1) for all the eigenvalues; it is easy to choose β such that

$$0 < \beta \leq \frac{2}{\max_{u,v} \left|\zeta_{uv}\right|^2}. \tag{4.21}$$

The energy functions (4.17) can also be explained by gradient theory. The second term in (4.19), $G_{k-1} = P^T(J - PI_{k-1})$, is the gradient of (4.9). This gradient becomes corrupted for a noisy observation because (4.9) is ill-posed. Known from (4.20), D_k, the iterative solution from a noise-free observation, is given by

$$D_k = I_{k-1} + \beta P^T(J - PI_{k-1} - N). \tag{4.22}$$

Substituting (4.22) into (4.17), the energy function can be rewritten as

$$\begin{aligned} I_k &= \min_{I_r} \|I_r - D_k\|_2^2 \\ &= \min_{I_r} \left\|I_r - (I_{k-1} + \beta P^T(J - PI_{k-1} - N))\right\|_2^2 \\ &= \min_{I_r} \left\|I_r - I_{k-1} - \beta P^T(J - PI_{k-1}) + \beta P^T N\right\|_2^2. \end{aligned} \tag{4.23}$$

In this equation, the term $\beta P^T N$ is employed to correct the corrupted gradient to be the gradient of (4.23). Therefore, by adaptively adjusting itself to the current solution during the iterative process, the noisy image is restored by minimising the iterative energy functions. In contrast to (4.9), minimisation of (4.17) is a well-posed denoising problem and can be solved by well-established noise reduction methods, such as the collaborative filter [22].

Now, we introduce a noise reduction operator, Ψ , which minimises the estimation error of a cost function. Letting I be the real scene, N be white noise and $V = I + N$ we define the cost function for the noise reduction operator,

$$\mathcal{C}(\mathbf{\Psi}, I) = \mathrm{E}\{\|I - \mathbf{\Psi}(V)\|_2^2\}, \tag{4.24}$$

where $E\{\cdot\}$ is the expectation taken over the noise distribution. The error is measured by the L_2 norm and averaged over the noise distribution. To minimise the energy functions (4.17), we propose the following algorithm:

$$\begin{cases} I_k = I_{k-1} + \beta P^T(J - PI_{k-1}) & \text{(4.25a)} \\ R_N = \arg\min_{\mathbf{\Psi}} \mathcal{C}(\mathbf{\Psi}, D_k), & \text{(4.25b)} \\ I_k = R_N(I_k), \quad \text{for } k = 1, 2, \ldots, & \text{(4.25c)} \end{cases}$$

where (4.25a) recovers structures that are the same as D_k and forces I_k to be close to D_k by searching a solution along the corrupted gradient. Equations (4.25b) and (4.25c) constitute an operator R_N for reducing the amplified noise due to the corrupted gradient. Through the restoration of continuous structures and noise reduction, the iterative solution I_k will tend to D_k and thus converge to the real scene.

Concerning the noise reduction in (4.25b), our method does not expect an ideal

operator to remove all noise [16]. Instead, any denoising algorithm can be used as long as the operator satisfies the following condition:

$$R_N(I+N) = I + o(N) \text{ and } \mathrm{Var}(o(N)) \sim \Delta\sigma^2, \tag{4.26}$$

where $\sigma^2 = \mathrm{Var}(N)$ is the variance of the noise N and Δ is a noise reduction factor much smaller than 1. Condition (4.26) implies that the remaining noise $o(N)$ has a variance much smaller than the initial noise N after applying the operator R_N to a noisy image. We will show in the error analysis below that when condition (4.26) is satisfied, the iterative solutions of (4.25) converge to the real scene with a small higher order noise term, i.e., $\lim_{k\to\infty} I_k = I + o(N.)$

To implement the method (4.25), we can apply any well-established noise reduction algorithms in combination with VanCittert's iteration, for example, the wavelet domain shrinking filter [16] $T_{SW}(V,\delta) = \hat{I} = Ww$ where $\hat{I}$ is the estimated image, W is a group of wavelet bases, and w is a vector of shrinking coefficients depending on the smoothing parameter δ. The smoothing parameter can be determined in a similar form as (4.25b) by

$$\arg\min_{\delta} C(T_{SW}(V,\delta), I), \tag{4.27}$$

which has a noise shrinkage strength of

$$\delta = \frac{(2\log RC + 1)(\log RC + 1)}{RC}, \tag{4.28}$$

where R and C represent the image size. For images of modest size, δ is much smaller than 1, and thus, the wavelet algorithm satisfies (4.26).

Another popular denoising method is the state-of-the-art BM3D method [22]. The BM3D method improves upon wavelet domain shrinking by incorporating the concept of image patches and non-local means [17] into a transformed domain and has demonstrated the highest PSNR compared to the wavelet domain and other algorithms. Moreover, BM3D includes simple parameter setting and is easy to use. Mathematically, BM3D can be expressed as

$$O_{BM3D} = \mathcal{A}\mathcal{T}_{3D}{}^{-1} W_{wie} \mathcal{T}_{3D} Z, \tag{4.29}$$

where Z is the operator of stacked noisy blocks, T_{3D} is the transformation from the spatial domain to the frequency domain with discrete cosine bases, W_{wie} is the Wiener shrinking operator, and A is the aggregation operator, all of which are defined in [22]. In light of the advantages and excellent performance of BM3D, we choose the operator $R_N = O_{BM3D}$ in our method (4.25b) for the numerical experiments below.

We note that our method decouples deblurring with denoising in (4.25). Compared with other decoupled approaches such as [19] [63], our method has two advantages. First, parameters configuration is simple for easy usage; whereas the regularization factors for optimal deblurring in other methods are substantially complex to set. This leads to the overall simple parameter setting of our method compared to many regularization methods. Second, structure and noise can be separated analytically in (4.25), which allows us to perform error analysis on a restored image; in contrast, error analysis under regularization methods is generally difficult due to the complexity of regularization factors.

4.3.3 Error analysis

In this section, we will prove that our method (4.25) can determine an approximation close to the real scene I with a distance of no greater than $o(N)$. The error analysis is started with the following lemma.

Lemma 1. *Let $F_k = 1/\zeta_{uv}(1-(1-\beta|\zeta_{uv}|^2)^k)$ be the noise amplification factor in equation (4.16). Then, $1 < F_k/F_{k-1} < 2$ for β satisfying (4.21) and $k \geq 2$.*

Proof. For convenience we set $a = (1-\beta|\zeta_{uv}|^2)$, thus $F_k = 1/\zeta_{uv}(1-a^k)$. Because of condition (4.21) a falls in $(0,1)$. The ratio

$$\begin{aligned} \frac{F_k}{F_{k-1}} &= \frac{1-a^k}{1-a^{k-1}} = \frac{1+a+a^2+\cdots+a^{k-1}}{1+a+a^2+\cdots+a^{k-2}} \\ &= 1+\frac{a^{k-1}}{1+a+a^2+\cdots+a^{k-2}}, \end{aligned} \tag{4.30}$$

so $1 < F_k/F_{k-1} < 2$ for any $a \in (0,1)$ when $k \geq 2$. □

Theorem 1. *For any operator R_N satisfying condition (4.26) and β satisfying (4.21), the iterative solution of model (4.25) leads to*

$$\lim_{k\to\infty} I_k = I + o(N), \tag{4.31}$$

where I is the original scene.

Proof. Let $F_k = 1/\zeta_{uv}(1-(1-\beta|\zeta_{uv}|^2)^k)$ be the noise amplification factors for $k = 1, 2\ldots$. Known from lemma (1), F_k is infinite so there must be a minimum number k satisfying

$$\mathrm{Var}\left(\sum_{u,v}\langle\frac{N}{F_k}, Z_{uv}\rangle Z_{uv}\right) < \mathrm{Var}(N) \tag{4.32}$$

when noise N is bounded. Here, we suppose $k = 1$ for convenience, though this num-

ber depends on the eigenvalues of P. Thus, we start with the iteration solution(4.20) for $k = 1$,

$$\begin{aligned} I_1 &= I_0 + \beta P^T (J - P I_0) \\ &= \sum_{u,v} (1 - (1 - \beta|\zeta_{uv}|^2)) \langle I, Z_{uv} \rangle Z_{uv} \\ &+ \sum_{u,v} \frac{1}{\zeta_{uv}} (1 - (1 - \beta|\zeta_{uv}|^2)) \langle N, Z_{uv} \rangle Z_{uv} \\ &= D_1 + N_1, \end{aligned} \tag{4.33}$$

where I_0 is the initial image and D_1 is the first sum on right side. The first term in (4.33) is a noise-free iterative solution whereas the second term is noise contribution with the factor $F_1 = \beta\zeta_{uv}$. Because most of the eigenvalues have absolute values smaller than 1, $\mathrm{Var}(N_1) \propto \mathrm{Var}(N)$. Then we apply the filter(4.26) to (4.33) and have

$$I_1 = R_N(I_1) = R_N(D_1 + N_1) = D_1 + o(N_1). \tag{4.34}$$

The noise is now reduced by a factor $\Delta \ll 1$ according to (4.26), i.e., $\mathrm{Var}(o(N_1)) \propto \Delta\mathrm{Var}(N_1) \propto \Delta\sigma^2$, where σ^2 is the variance of noise N.

The intensity of image I_1 can be rewritten as

$$\begin{aligned} I_1 &= D_1 + o(N_1) \\ &= \sum_{u,v} (1 - (1 - \beta|\zeta_{uv}|^2)) \langle I, Z_{uv} \rangle Z_{uv} \\ &+ \sum_{u,v} \frac{1}{\zeta_{uv}} (1 - (1 - \beta|\zeta_{uv}|^2)) \langle \frac{o(N_1)}{F_1}, Z_{uv} \rangle Z_{uv}, \end{aligned} \tag{4.35}$$

where $o(N_1) = \sum_{uv} (o(N_1), Z_{uv}) Z_{uv} = \sum_{uv} F_1 (o(N_1)/F_1, Z_{uv}) Z_{uv}$ is used and F_1 is defined at the beginning of the proof.

From (4.35) and (4.20), we give the iteration solution for $k = 2$

$$\begin{aligned}
I_2 &= I_1 + \beta P^T(J - PI_1) \\
&= I_1 + \beta P^T\left(PI + \sum_{u,v}\langle\frac{o(N_1)}{F_1}, Z_{uv}\rangle Z_{uv} - PI_1\right) + \beta P^T\left(N - \sum_{u,v}\langle\frac{o(N_1)}{F_1}, Z_{uv}\rangle Z_{uv}\right) \\
&= \sum_{u,v}(1-(1-\beta|\zeta_{uv}|^2)^2)\langle I, Z_{uv}\rangle Z_{uv} \\
&+ \sum_{u,v}\frac{1}{\zeta_{uv}}(1-(1-\beta|\zeta_{uv}|^2)^2)\langle\frac{o(N_1)}{F_1}, Z_{uv}\rangle Z_{uv} + \beta P^T\left(N - \sum_{u,v}\langle\frac{o(N_1)}{F_1}, Z_{uv}\rangle Z_{uv}\right) \\
&= D_2 + \sum_{u,v}\frac{F_2}{F_1}\langle o(N_1), Z_{uv}\rangle Z_{uv} + \beta P^T\left(N - \sum_{u,v}\langle\frac{o(N_1)}{F_1}, Z_{uv}\rangle Z_{uv}\right) \\
&= D_2 + N_2.
\end{aligned} \tag{4.36}$$

where N_2 is the sum of the two noise terms on the right side. By (4.32) and lemma (1), $1 < F_2/F_1 < 2$, $\mathrm{Var}(N_2) \propto \mathrm{Var}(N) = \sigma^2$ is obtained.

Repeating in the operation (4.34), we have

$$I_2 = R_N(I_2) = R_N(D_2 + N_2) = D_2 + o(N_2), \tag{4.37}$$

where the variance of $o(N_2)$ is far less than that of N_2, and far less than that of N.

In general, we have

$$I_k = D_k + N_k,$$

where

$$N_k = \sum_{uv}\frac{F_k}{F_{k-1}}\langle o(N_{k-1}), Z_{uv}\rangle Z_{uv} + \beta P^T\left(N - \sum_{uv}\langle\frac{o(N_{k-1})}{F_{k-1}}, Z_{uv}\rangle Z_{uv}\right), \tag{4.38}$$

and

$$I_k = R_N(I_k) = D_k + o(N_k), \tag{4.39}$$

where $\mathrm{Var}(o(N_k)) \propto \Delta\mathrm{Var}(N) = \Delta\sigma^2$. Therefore, the iterative solutions converge to the real scene and the noise is controlled to the order of $\Delta\sigma^2$ in the iterative process. □

4.4 Experiments and results

This section includes four tasks. The first task is to compare our methods with existing state-of-art methods. Then, the efficiency and power of our method in

restoring an X-ray image of a grating are demonstrated. Through the comparisons, our method shows a better stability for different types of blurring filters. Third, we design an experiment to prove the stability of the resolution improvement via our method for the same kind of PSFs but with different parameters. Finally, the stability of our method is tested under Gaussian PSFs with various variances, which shows that our method is still valid if unknown PSFs can be estimated by a Gaussian type.

Two measurements are applied to evaluate how an image is restored to a condition close to the original scene. The first is peak signal-to-noise ratio (PSNR), and the second is the structural similarity index (SSIM). For an estimated image, PSNR is defined as the logarithmic ratio of the maximum intensity of the original image and the mean square error (MSE). Higher PSNR means a smaller MSE, which yields a better restoration. The mathematical definition is formulated as

$$\text{PSNR} = 10 \times \log_{10} \frac{I_{Max}^2}{RC \left\| I - \hat{I} \right\|_2^2}, \tag{4.40}$$

where I is the original image, I_{Max} is the maximum intensity of I, $\hat{I}$ is the estimated image of I, and R and C represent the row and column numbers of I.

Different from PSNR which is based on distance in terms of intensity, SSIM measures the similarity between a restored image and the original scene by correlation coefficient. This measurement prefers to similarity of structures due to less influence from noise. Mathematically, it is defined as [99]

$$\text{SSIM} = \frac{(2\mu(I)\mu(\hat{I}) + c_1)(2\text{Cov}(I, \hat{I}) + c_2)}{(\mu^2(I) + \mu^2(\hat{I}) + c_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + c_2)}, \tag{4.41}$$

where $\mu(\cdot)$ is the mean function, $\text{Cov}(\cdot, \cdot)$ is the covariance function, σ_I^2 is the variance of I, and c_1 and c_2 are the constants determined by the intensity range. From its definition, SSIM is always in the area $[0, 1]$. Higher SSIM signifies high similarity. If SSIM is equal to 1, then I and $\hat{I}$ have same shapes and intensity.

Seven scenarios listed in Table 5 will be applied for experiments. Most of common PSFs are contained. In the first two scenarios the same X-ray type filter is used for blurring, but with different noise. In the third scenario, a uniform filter is used. In the fourth scenario, a synthetic filter is constructed by a vector. The blurring filter in scenarios five and six are Gaussian types with different STDs. The last is motion blurring. Of the sever scenarios, scenarios four and six have large amounts of noise but small blurring. Scenario three presents a severe blur but a small amounts of noise. The other scenarios reflect moderate amounts of noise.

Table 5: PSF and noise variation for each scenario

Scenario	Blur PSF	Noise variance
1	$1/(1+x^2+y^2), x, y = -7, \ldots, 7$	2
2	$1/(1+x^2+y^2), x, y = -7, \ldots, 7$	8
3	9×9 uniform	0.3
4	$[1\ 4\ 6\ 4\ 1]^T[1\ 4\ 6\ 4\ 1]/256$	49
5	Gaussian with $std = 1.6$	4
6	Gaussian with $std = 0.4$	64
7	Motion blurring	2

4.4.1 Method Implementation

An example is demonstrated for implementation of algorithm (4.25). The cameraman image is degraded by scenario 1 in Table 5 and the degraded image is processed by the algorithm. Noise doe not increase significantly after one iteration because the rate of noise amplification factors of two adjacent iteration falls in (1,2), seeing subsection of "error analysis". Therefore, to improve efficiency, the noise reduction filter R_N needs not to be called at every iteration. In the example, the filter is called after each 84 iterations which is a parameter in our algorithm. The choice of parameter is discussed in the last subsection. After 1190 iterations, the algorithm determinates. Figure 27 and Table 6 show the results of D_k and I_k at iterations (85,425,765,1105) when the filter is called. In this figure, plot(a) shows the full original scene and the position of fragment images from (b) to (p). The four images in the second row are D_k in (4.25b) which represents the restored structures during iterations. The third and forth rows are iterative images before and after applying the filter, respectively.

As seen in Figure 27 and Table 6, the main structures have been restored after 85 iterations and there is a great improvement in PSNR and SSIM. However, edges and small bottoms are not clear in image D_{85}. When $k \geq 425$ the structure are well reconstructed and the corresponding PSNRs keep increasing despite that no obvious changes are observed in image D_k. The images in the third row show obvious noise compared to those in the forth row. Their PSNRs are also smaller due to noise while their SSIMs are similar. This result shows that the structures are restored well after several hundred iterations, but noise fluctuates during iterations.

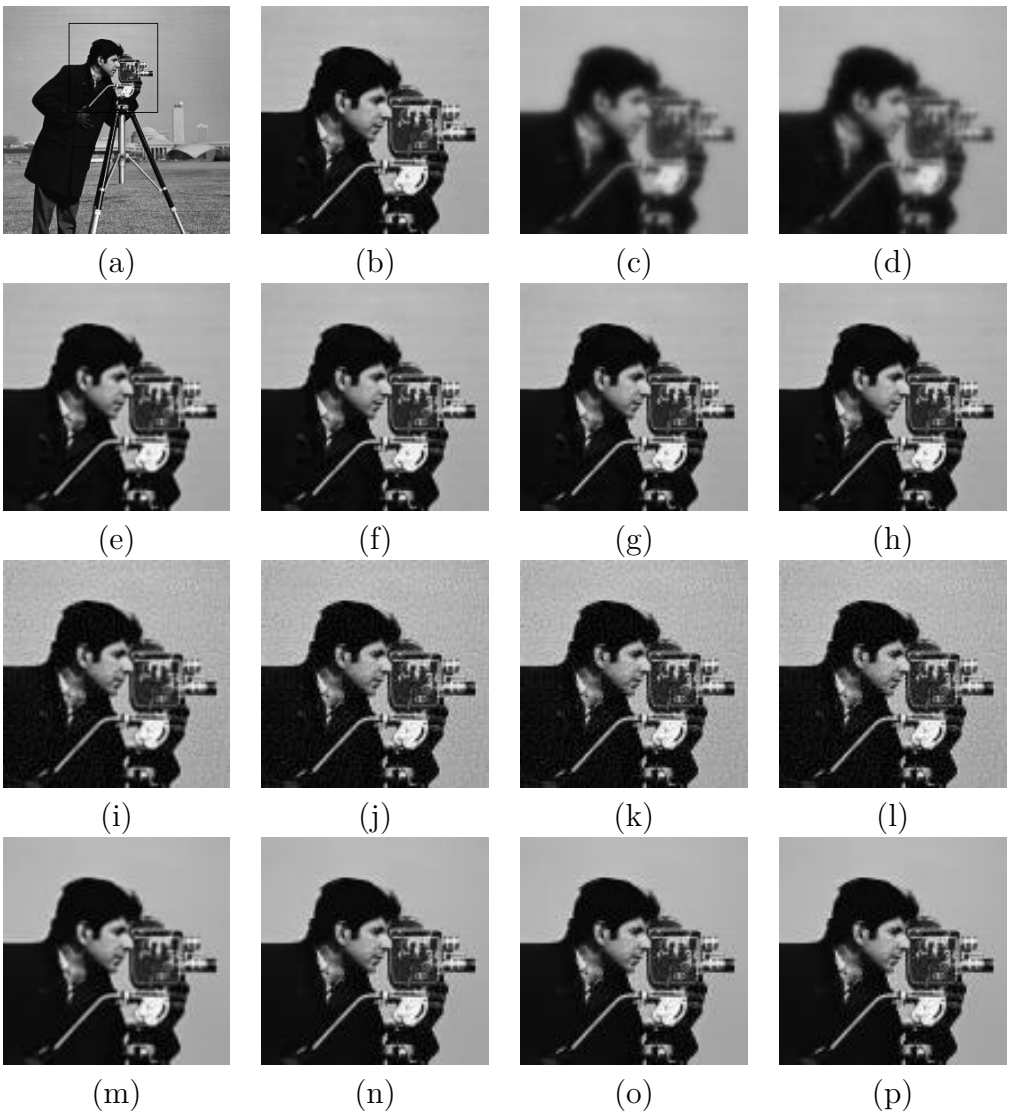


Figure 27: Demonstration of algorithm (4.25). (a) The original scene and the position of the following fragment images, (b) the original scene, (c) the blurred image, (d) the blurred and noisy image, (e)-(h) D_k in (4.25b) for $k = (85, 425, 765, 1105)$, (i)-(l) I_k in (4.25a) for $k = (85, 425, 765, 1105)$, which are iterative images before applying the noise reduction filter R_N in (4.25b), (m)-(p) I_k in (4.25c) for $k = (85, 425, 765, 1105)$, which are iterative images after applying the noise reduction filter R_N in (4.25b).

Table 6: PSNRs and SSIMs of the images during iterations in algorithm (4.25).

	PSNR			SSIM		
	Input PSNR=22.2284			Input SSIM=0.9446		
Iteration Number	D_k	I_k^{B}	I_k^{A}	D_k	I_k^{B}	I_k^{A}
$k = 85$	28.3406	27.3853	27.6587	0.9876	0.9846	0.9854
$k = 425$	33.1066	29.1500	30.0034	0.9959	0.9898	0.9916
$k = 765$	35.6485	29.4774	30.4790	0.9977	0.9906	0.9925
$k = 1105$	37.6548	29.5714	30.6335	0.9986	0.9908	0.9928

Note: I_k^{B} and I_k^{A} are images before and after applying the filter R_N, respectively.

4.4.2 Restoration of popular test images

We undertake two experiments to test our method and compare the results with those of some state-of-the-art methods. The first experiment contains two images, Cameraman256.png and Lena512.tif, which are commonly used for measuring the efficiency of algorithms for structure restoration because they contain elaborate structures, such as lines, buttons and textures. These images were the subjects of a recent extensive investigation using an iterative decoupled deblurring BM3D algorithm (IDD-BM3D) [19], which was formulated based on the Nash equilibrium balance of two objective functions applying separate denoising and deblurring operations. IDD-BM3D demonstrated state-of-the-art restoration performance compared to seven other methods: Fourier-Wavelet regularised deconvolution (ForWaRD) [100], space-variant Gaussian scale mixtures (SVGCM) [101], shape-adaptive discrete cosine transform(SA-DCT) [18], BM3D deblurring (BM3DDEB) [102], analysis-based sparsity (L0-Abs) [103], adaptive total variation image deblurring by a majorisation minimisation approach (TVMM) [104], and a method based on spatially weighted TV (CGMK) [105]. We test the same six scenarios as in [19], which have different PSF shapes and blurring strengths as well as noise levels as listed in Table 5. Comparisons of all eight methods are made quantitatively through PSNR measurements in Table 7.

In our method, there are two input parameters (σ, s), where σ is the input for the denoising filter. The parameter s is the interval between two adjacent call of denoising filter R_N during iterations. Unlike the description of the algorithm $s = 1$ in the previous section, we choose s as a larger value to improve efficiency. The parameter setting are (7.5, 85) for scenario 1, (7.5, 25) for scenario 2, (7.5, 550) for scenario 3, (7.5,10) for scenario 4, (7.5,50) for scenario 5, (7.5, 5) for scenario 6, (7.5, 85) for scenario 7. From the parameter configuration, we can see that the input for the denoising filter can be fixed and just barely changes the other parameter. However, how to choose the optimal parameters is a complicated problem, which will be discussed based on experiments in the later sub-section.

As seen in Table 7, IDD-BM3D outperforms other methods, as evidenced by its yielding the majority of the highest PSNRs, as marked with bold font. The PSNRs of the restored image via our method are close to those obtained via IDD-BM3D, which is obviously higher than those obtained via the other methods.

Regarding the measurement of SSIM, we list the results of four methods, L0-Abs, BM3DDEB, IDDBM3D and ours in Table 8. As seen in this table, our method outperforms all the methods because it leads to the majority of the highest observed SSIMs. This shows that our method is powerful in structure restoration.

Several of the restored images are listed in Figure 28. In this figure, we list the

Table 7: PSNRs of the methods in six scenarios

	Scenarios						Scenarios					
	1	2	3	4	5	6	1	2	3	4	5	6
Methods	Cameraman(256×256)						Lena(512×512)					
Input PSNR	22.23	22.16	20.76	24.62	23.36	29.82	25.61	25.46	24.11	28.06	27.81	29.98
ForWaRD	28.99	27.24	28.10	27.02	26.50	33.74	33.30	31.94	32.81	31.74	32.66	35.45
SV-GSM	29.68	27.71	28.09	27.35	26.61	34.01	-	-	-	-	-	-
SA-DCT	30.34	28.49	29.31	27.99	27.08	34.53	34.8	33.14	33.63	33.3	33.24	35.87
BM3DDEB	30.42	28.56	29.1	27.96	27.09	34.52	35.2	33.57	33.81	33.62	33.53	36.43
L0-Abs	29.93	27.93	29.72	27.61	26.94	33.21	33.91	32.75	33.63	32.9	33.38	31.96
TVMM	29.64	27.33	29.3	27.19	26.72	31.12	33.61	32.02	33.31	32.33	32.77	32.82
CGMK	30.03	27.65	29.91	27.42	26.9	33.15	34.01	32.41	33.7	32.3	33.09	34.49
IDD-BM3D	**31.08**	**29.28**	**31.21**	**28.60**	**27.67**	**34.71**	35.22	**33.65**	**34.75**	**33.78**	**34.01**	36.37
Ours	30.73	28.71	30.57	28.25	27.43	34.65	**35.29**	33.41	34.53	33.70	33.91	**36.46**

Table 8: SSIMs of the four methods in six scenarios

	Scenarios					
	1	2	3	4	5	6
Methods	Cameraman(256×256)					
Input SSIM	0.93	0.93	0.92	0.96	0.95	0.98
L0-Abs	0.99	0.98	0.99	0.98	0.98	0.99
BM3DDEB	0.99	0.99	0.99	**0.98**	0.98	**0.99**
IDD-BM3D	0.96	0.94	0.95	0.96	0.96	0.99
Ours(BM3D)	**0.99**	**0.99**	**0.99**	0.98	**0.98**	0.99

restored images via four methods-L0-Abs, BM3DDEB, IDDBM3D and ours-for scenario three. As seen, IDDBM3D (e) and ours (f) looks better than L0-Abs (c) and BM3DDEB (d). First, the output images (e) and (f) seem to be clearer with shaper edges. Second, the small structures, such as the lens and bottoms, are well reconstructed in (e) and (f), whereas they are poorly restored in (c) and (d). Compared to IDDBM3D, the result via our method shows more noise in the plain area, which is the reason that our method leads to a lower PSNR than IDDBM3D.

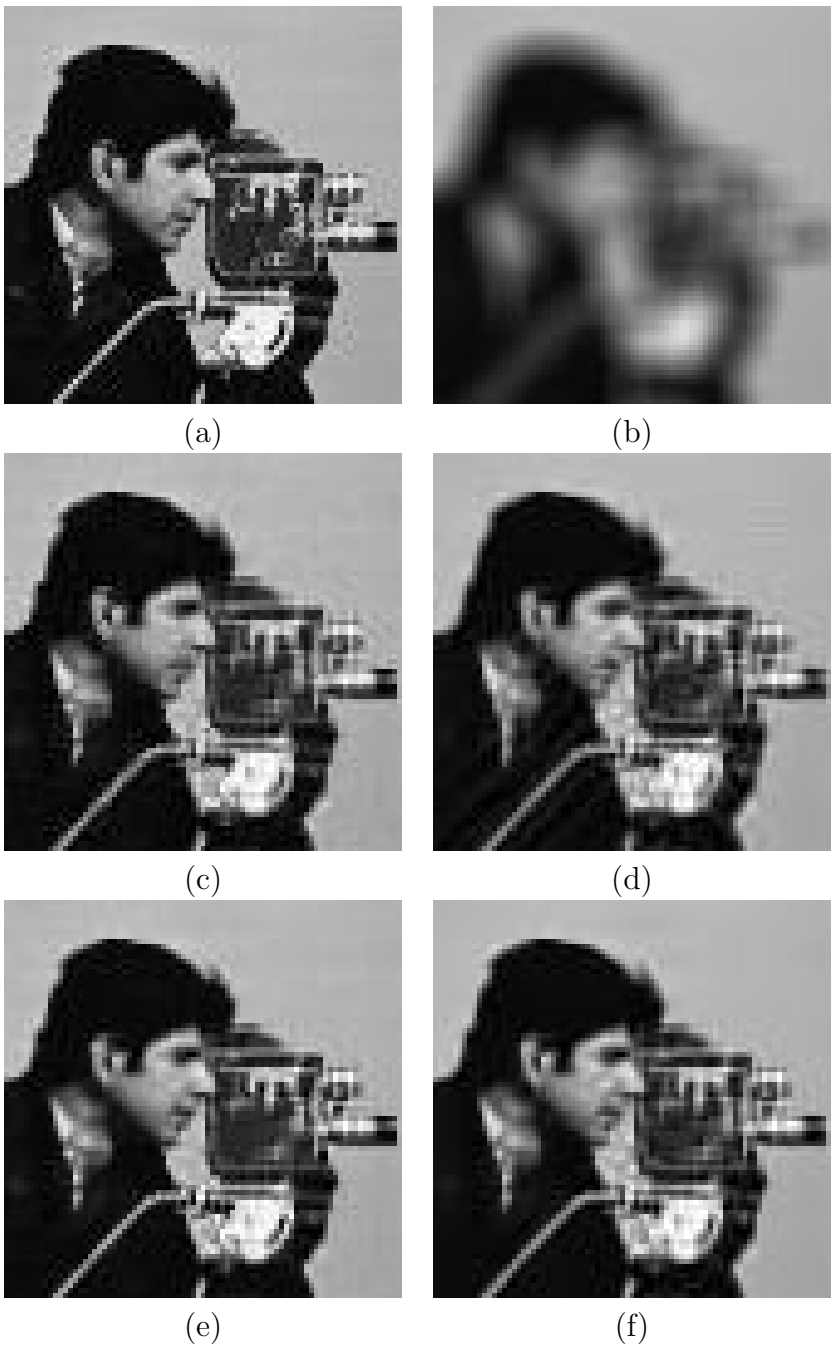


Figure 28: Restored images via four methods. (a) The original scene, (b) the image blurred by scenario 3, (c) the restored image by L0-Abs, (d) the restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method.

For the Lena image, the restored images from scenario two are shown as in Figure 29. As seen in this figure, the pupil in the eye can be well restored by BM3DDEB, IDDBM3D and ours, whereas it cannot be recovered via L0-Abs. As known from PSNRs in Table 7, the PSNR for L0-Abs is 32.75, while all the PSNR for the other three methods are greater than 33.50.

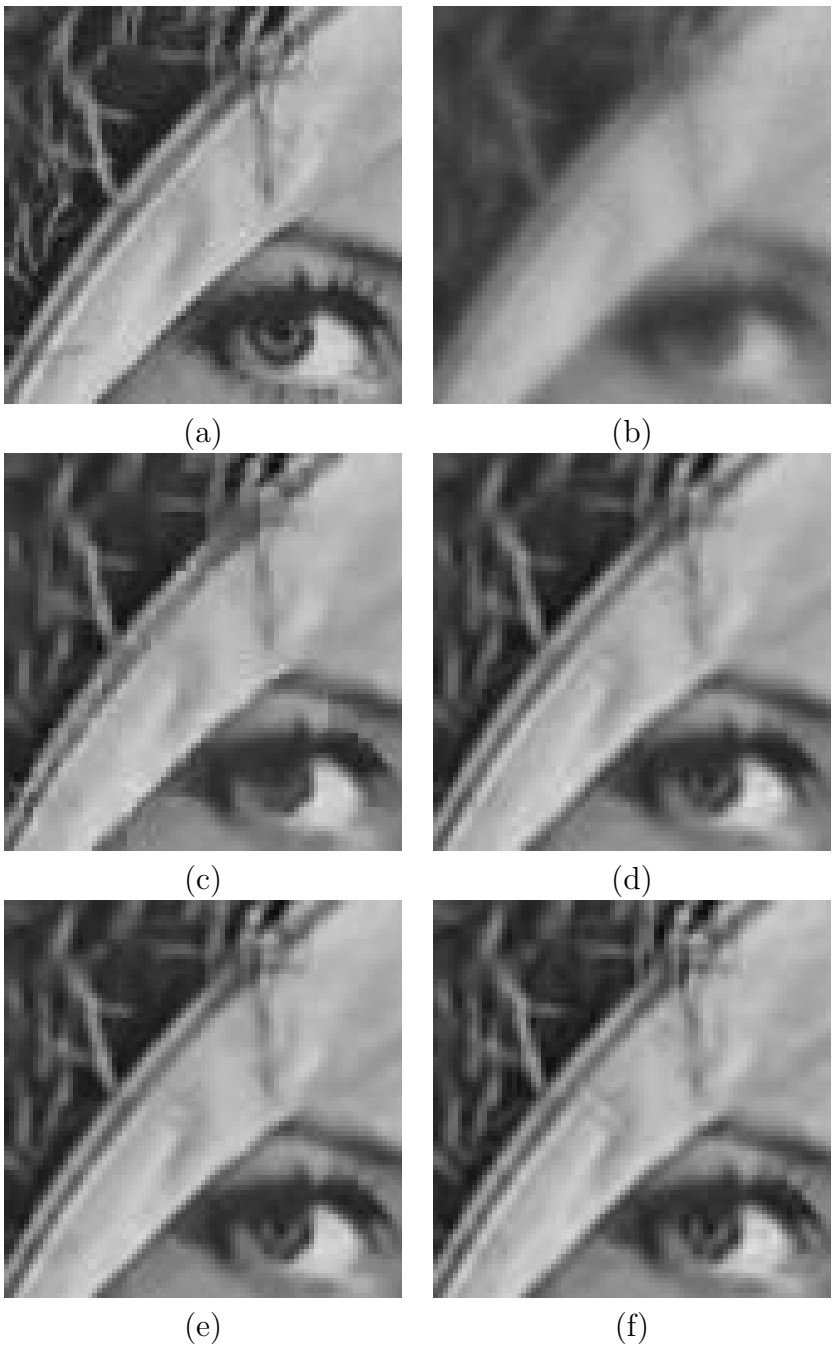


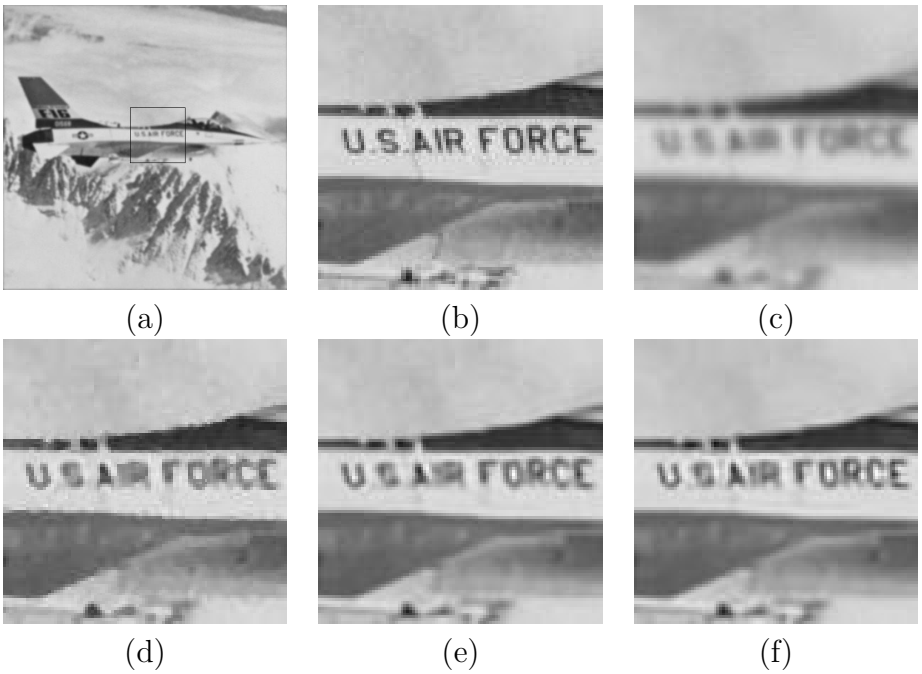
Figure 29: Restored images via four methods. (a) The original scene, (b) the image blurred by scenario 2, (c) the image restored by L0-Abs, (d) the image restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method.

The second experiment applies another popular test image, "JetPlane.png", for the first six scenarios. In this experiment, we compare with the fast nonconvex nonsmooth method (Fnnmm) [106], which is a classical method based on partial differential equations (PDEs). The code of Fnnmm can be downloaded at `http://www.math.hkbu.hk/mng/imging-software.html`. The PSNRs of the restored images via the three methods, Fnnmm, IDDBM3D and ours, are listed in Table 9, and the restored images are shown in Figure 30. As seen in the table, IDDBM3D shows the highest PSNRs except scenario six. The restored image via Fnnmm shows

shaper edges while there is greater noise causing the image to seem coarse. Although IDDBM3D presents a higher PSNR, the result via our method seems better for some elaborative structures, such as the text "U.S.AIR FORCE".

Table 9: PSNRs of the three methods in six scenarios

	Scenarios					
	1	2	3	4	5	6
Methods	Jetplane(512×512)					
Input PSNR	24.98	24.84	23.43	27.31	26.84	29.85
Fnnmm	33.10	30.89	32.97	30.76	30.60	34.91
IDD-BM3D	**35.43**	**33.41**	**34.58**	**32.63**	**32.12**	36.60
Ours	35.03	32.84	34.57	32.91	32.04	**36.65**



(a) (b) (c)

(d) (e) (f)

Figure 30: Restored images via three methods. (a) The full scene of the blurred and noisy image by scenario 5, (b) a fragment contained in the black box of the original scene, (c) the fragment of the blurred and noisy image, (d) the image restored by Fnnmm, (e) the image restored by IDDBM3D, (f) the image restored by our method.

In the third experiment, we compare the four methods-L0-Abs, BM3DDEB, IDDBM3D and ours-through the restorations from the last scenario with motion blurring. The PSNRs are listed in Table 10, and the restored images are shown in Figure 31. As seen in this table, BM3DDEB leads to the lowest PSNR, and IDDBM3D is still the superior method. From the output images, it is found that BM3DDEB causes severe artefacts along the right edge. As an improvement, IDDBM3D suppresses these artefacts, but they are still obvious. L0-Abs and our method avoid these artefacts but the restored images seem coarse due to noise.

Table 10: PSNRs of the four methods for scenario 7.

Methods	L0-Abs	BM3DDEB	IDDBM3D	Ours
PSNR	30.2427	29.9164	31.7879	31.5947

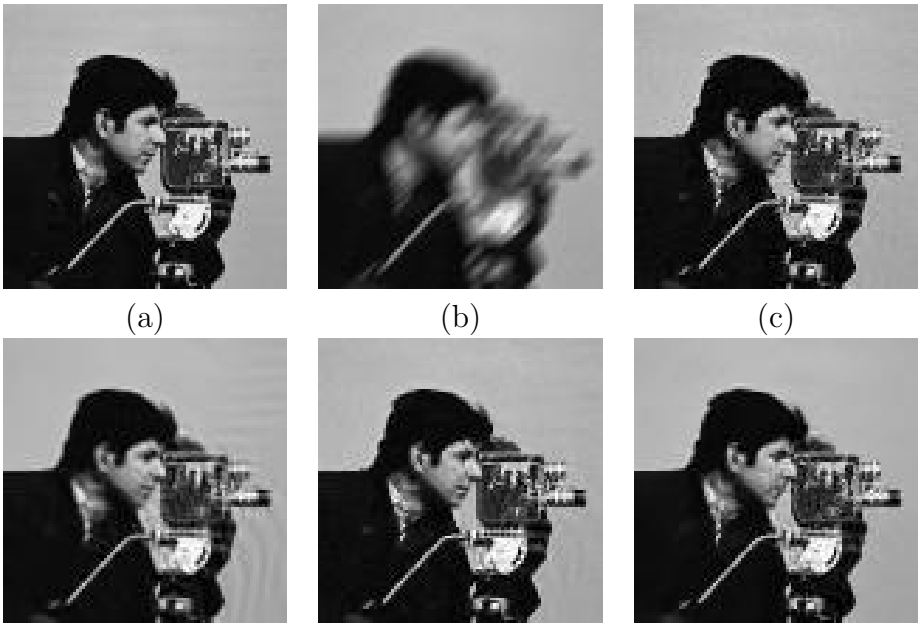
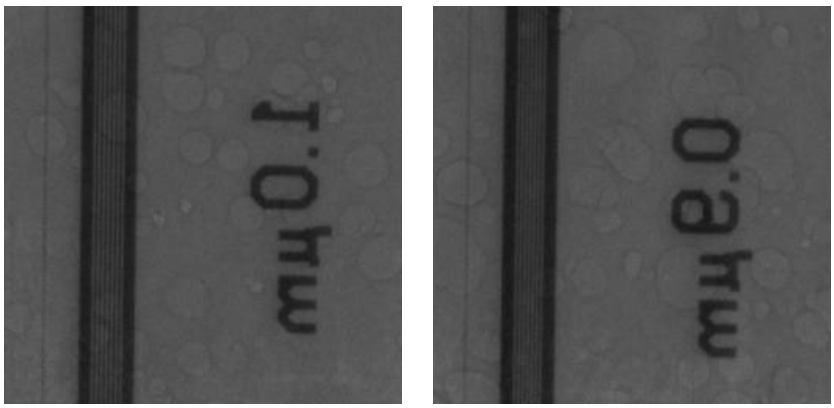


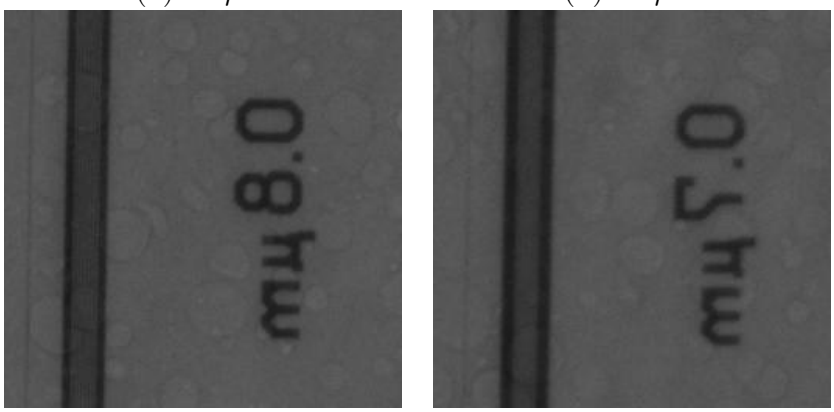
Figure 31: Restored images via four methods for scenario 7. (a) The original scene, (b) the blurred and noisy image, (c) the image restored by L0-Abs, (d) the image restored by BM3DDEB, (e) the image restored by IDDBM3D, and (f) the image restored by our method.

4.4.3 Restoration of practical images

We will now apply our method to a practical problem. Figure 32 shows X-ray images of four gratings with different widths. From the images, we can see that the slits are degraded due to diffraction that occurs when light passes through the narrow slits. When the width of the slits is $1.0\mu m$, we can see parallel slits presenting fine structures, whereas when the width of slits is reduced to 0.9 or $0.8\mu m$, the slits can still be seen, although they appear broken in certain areas. However, the silts cannot be observed when their widths decrease to $0.7\mu m$ because their intensity distribution looks flat.



(a) $1.0\mu m$ (b) $0.9\mu m$



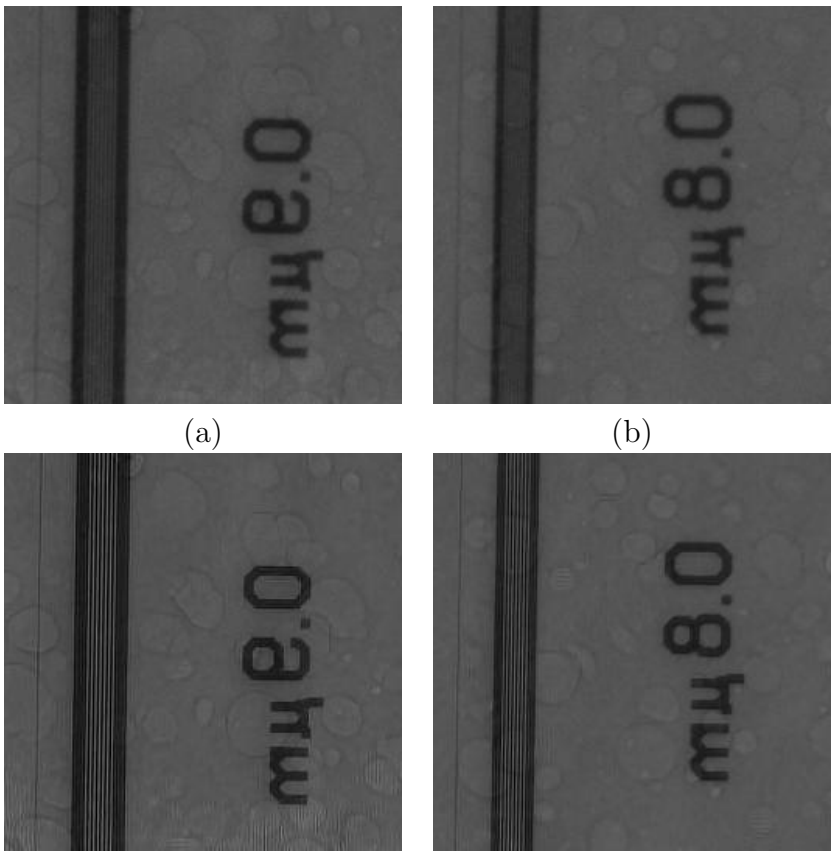
(c) $0.8\mu m$ (d) $0.7\mu m$

Figure 32: X-ray images of a grating with 4 different widths.

Four methods-L0-Abs, BM3DDEB, IDDBM3D and ours-are applied to restore the images of gratings whose widths are less than $1.0\mu m$. The PSF is estimated as the density distribution of 2D normal function with the covariance matrix

$$\Sigma = \begin{pmatrix} 1.344 & 0 \\ 0 & 1.508 \end{pmatrix} \tag{4.42}$$

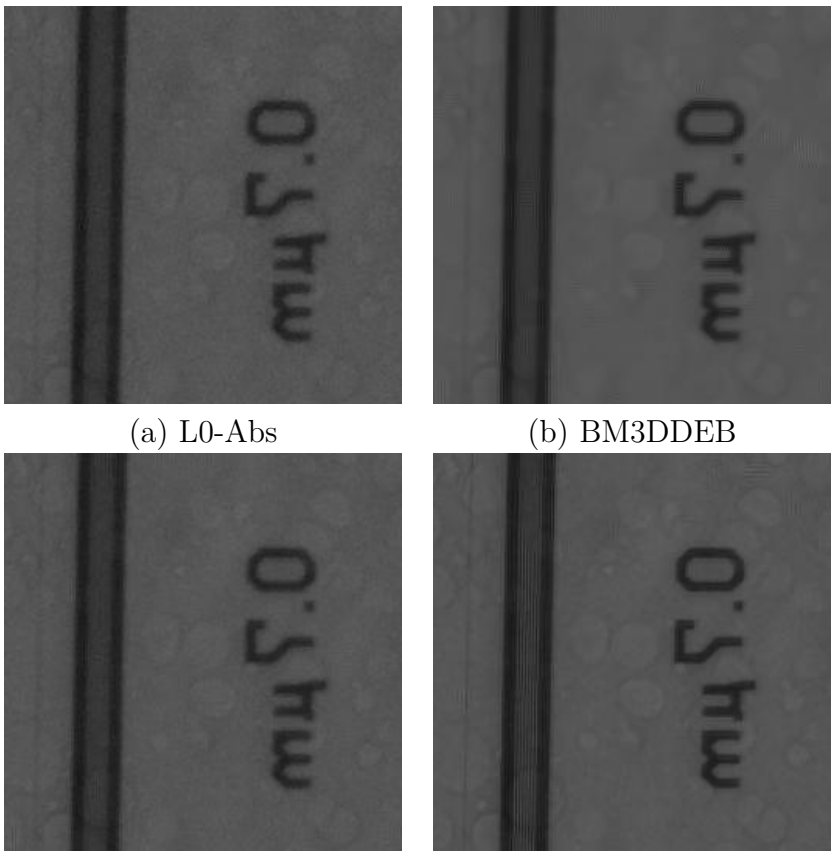
and a mean of zero. The STD of noise is estimated as 2. The parameters $(\sigma, s) = (7.5, 50)$. The results are shown in Figure 33 and Figure 34. As seen in Figure 33, the slits are restored well, with clear edges, compared to the slits with widths of 0.9 and $0.8\mu m$. As seen in Figure 34, the first three methods cannot restore the degraded slit with a width of $0.7\mu m$, whereas our method can. However, the restored slits seems broken in some areas and neither clear nor sharp compared to the results of 0.9 and $0.8\mu m$. These results show that our method has the strong ability to restore elaborative structures.



(a) (b)

(c) (d)

Figure 33: Restored images via our method. (a) and (b) are two images to be processed. (c) and (d) are the restored images for (a) and (b), respectively.



(a) L0-Abs (b) BM3DDEB

(c) IDDBM3D (d) Ours

Figure 34: Restored images via four methods for the image of a grating with width $0.7\mu m$.

4.4.4 Experiments for resolution improvement

As seen in the last two subsections, our method shows a better stability for PSFs which belong to different types. The goal of this experiment is to prove the stability of the resolution improvement obtained by our method for different PSFs which belong to the same type but have different parameters. In this experiment, PSNR cannot be applied as a measurement because the same type of PSFs with various parameters lead to different PSNR of restorations. For example, restorations from a Gaussian blurring with larger variance presents lower PSNR than those from a Gaussian blurring with smaller variance. Therefore, another of measurement is applied in terms of resolution improvement. We use an image containing a single bead for the experiment and measure the resolution in terms of the full width at half maximum.

The observations are generated in the following four steps:

- Plot a single bead on an image; the bead shape follows a Gaussian distribution with a standard deviation of 1;
- Two different PSFs are used to blur the single bead image, one of which has

a Gaussian shape with a standard deviation of 2 and the other of which has the same shape with a standard deviation of 4;

- Gaussian noise with a standard deviation of 2 is added to the two blurred images; and
- The fixed parameter algorithm $(\sigma, s) = (2, 100)$ is applied to restore the single bead.

Following the above steps, we obtained the results in Figure 35.

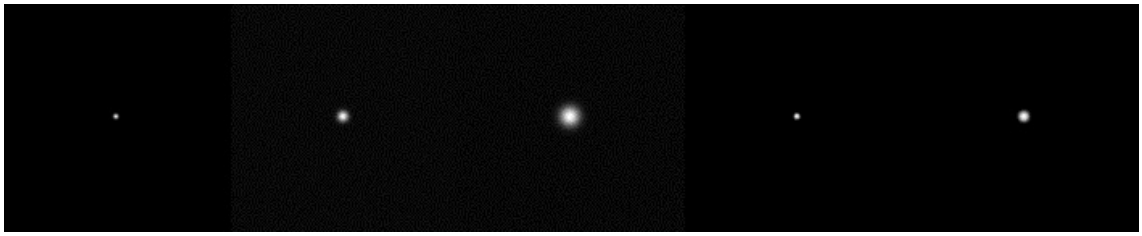


Figure 35: Restoration of the single bead image. From left to right are the following images: the original scene, the blurred image using Gaussian PSF with STD=2, the blurred image using Gaussian PSF with STD=4, and two restored images from the two blurred images.

Next, we plot the curve of the centre section of the beads in Figure 35 and fit these curves with a Gaussian filter in Figure 36. As seen in this figure, the black at the centre section of the original bead is the smallest, with an estimated STD=1.0011, which is approximately equal to the idea value of 1. The two blue curves have larger standard deviations, demonstrating that the blurred beads have larger diameters. The two red curves have moderate standard deviations that are less than those of the blue curves, demonstrating that the resolution is improved in the resulting images compared to the blurred images. We notice that $3.6964/1.5706 \approx 5.1092/2.28 \approx 2.3$, which means that our method achieves a stable improvement of the resolution by a factor of 2.3.

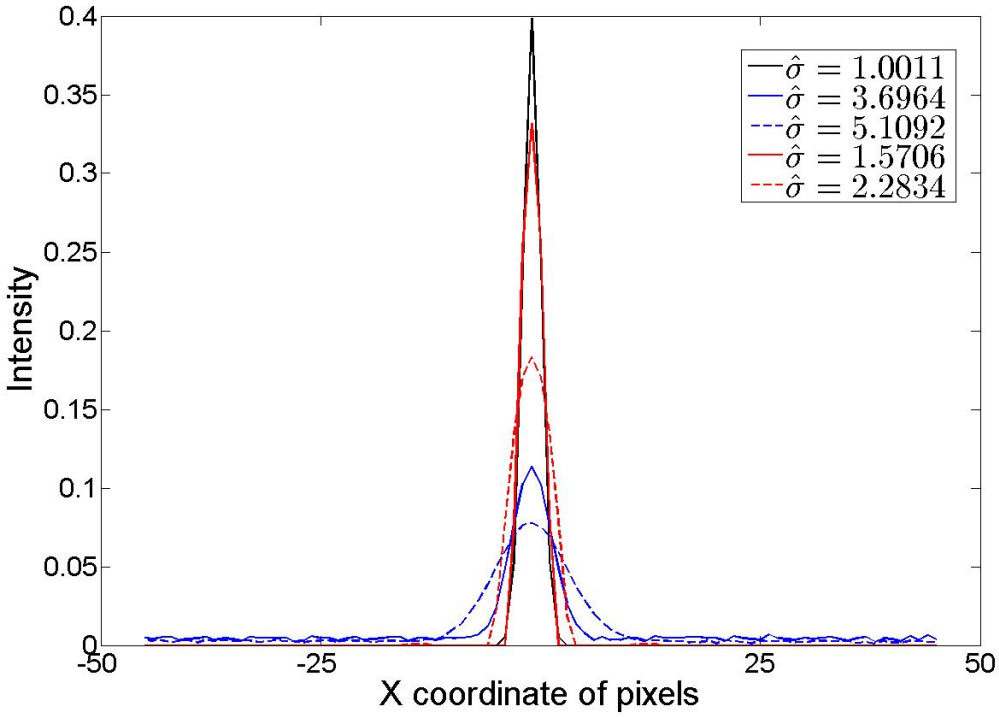


Figure 36: Curves of the centre section of the beads. The black, solid blue, dotted blue, solid red and dotted red curves are the centre sections of images of the original, blurred image using a PSF with STD=2, blurred image using a PSF with STD=4, and the two results from the two blurred images, respectively.

4.4.5 Experiments on the stability of Gaussian blurring

Finally, we test the robustness of ARSD, our method, against fluctuations in the size of the PSF in the model because the exact value is usually unknown in practice and is therefore estimated. For this, we undertake experiments on the Jetplane.png image, which is blurred using a Gaussian PSF with a standard deviation of $\sigma = 2$ and a noise level of 40 dB. We measure the PSNRs of the restored images while varying σ by 10% from the exact value. As seen in Figure 37, the PSNR decreases from its peak value of 29.5 on both sides, therein decreasing faster when σ is larger due to the more noticeable artefacts of the hard shoulder and plunge effects around an edge in this situation. As a result, the restored image appears to have an artificially higher contrast when it is larger than the exact value. Overall, the reductions in the PSNR are 2% and 4% when varying the standard deviation to level 10% smaller or larger than the exact value, respectively.

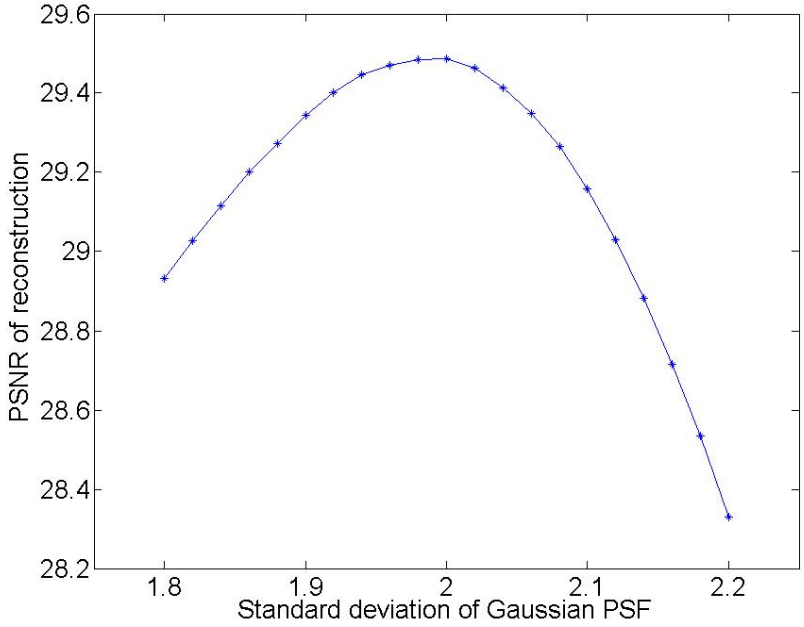


Figure 37: PSNRs of reconstructions with different PSFs.

The stability of our method is also represented in the parameters. We design experiments in which different images with different noise levels are tested by ARSD. The results show that a large area of parameter space can lead to good results under the fixed parameter method.

4.4.6 Choice of parameters

Playing important roles in energy functions, the parameters of regularization factors balance structure restoration (SR) and noise reduction (NR) in regularization methods. However, the determinations of factors to obtain an optimal balance is so challenging that most of them are based on experience, which presents an inconvenience to their use. In light of this, ARSD decouples NR from SR through the combination of VanCittert's algorithm and a noise reduction filter. One advantage of the decoupling method is that only parameters are required for the NR filter, which makes the configuration of parameters easy and convenient in practical applications.

Figure 38 demonstrates a decoupling and how to estimate the parameters (σ, s). The image of the cameraman is blurred by the PSF of scenario one in Table 5 and a Gaussian noise with a variance of 2 added. ARSD is applied to this blurred and noisy image with 1200 iterations. Two parameters (σ, s) are set to $(7.5, 85)$. In this figure, plot (a) shows two curves, one of which is the mean square error (MSE) of SR and the other of which is the MSE of SR and NR. The MSE of SR is measured by the original scene and D_k in (4.18), and the MSE of NR is measured by the original

scene and I_k in (4.17). Although the original scene cannot be known in practical applications, the theory of ARSD (4.36) proves the separation of NR from SR.

As seen in plot (a), MSE of SR converges to zero, which means that the structures are almost restored after 600 iterations. The MSE of noise reduction fluctuates, but it is controlled in a small range through all iterations. These two curves indicates that there is only a small amount of noise between the estimated image and the original scene. Plot (b) is the curve of the STD of noise in a uint 8 image. The first peak, located at iteration 84, shows a noise with an STD of approximately 7, which says that the initial noise with an STD of 1.414 inflates as big as one with an STD of 7 after 84 iterations. Therefore, for a fixed parameter $\sigma = 7.5$, the STD of noise, we choose $s = 85$, which shows that the noise reduction filter is called after 85 iterations. As an important review of (4.20) and (4.36), the calculation of noise is independent on original scene through all iterations. Therefore, parameter s, as an interval upon which to call the noise reduction filter, can be estimated without any information about the original scene.

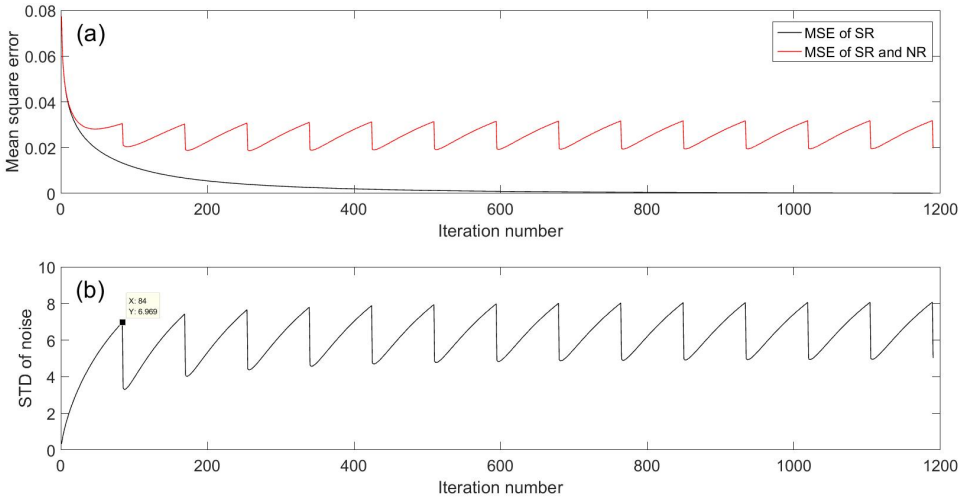


Figure 38: Demonstration of decoupling noise reduction from structure restoration.

Although there is a large region of parameters that can output qualified images, we would like to make as optimal of a choice as possible. Although it is difficult to find the optimum by theory, we attempt to curve the parameter s for various noise. We design an experiment in which the blurring filter is Gaussian and the noise level varies. Figure 39 shows the best choice for different noise levels for the observed image. As seen in this figure, the blue line represents the parameter values calculated from experiments at various noise levels. The red line represents the fitting curve of the blue data given by the function $\alpha e^{\beta x} + \gamma$ with $(\alpha, \beta, \gamma) = (1000, -1.32, 2.54)$. Additionally, the two parameters, the number of iterations between two adjacent adjustments to the corrupted gradient and the noise level, are usually dependent; higher noise levels produce smaller iteration intervals. In the automated method, the

only parameter that can be chosen is approximately 3, which is very convenient in practice. Due to the complexity involved in determining of the optimal parameters, we cannot currently formulate them mathematically and have to describe them through numerical measurements.

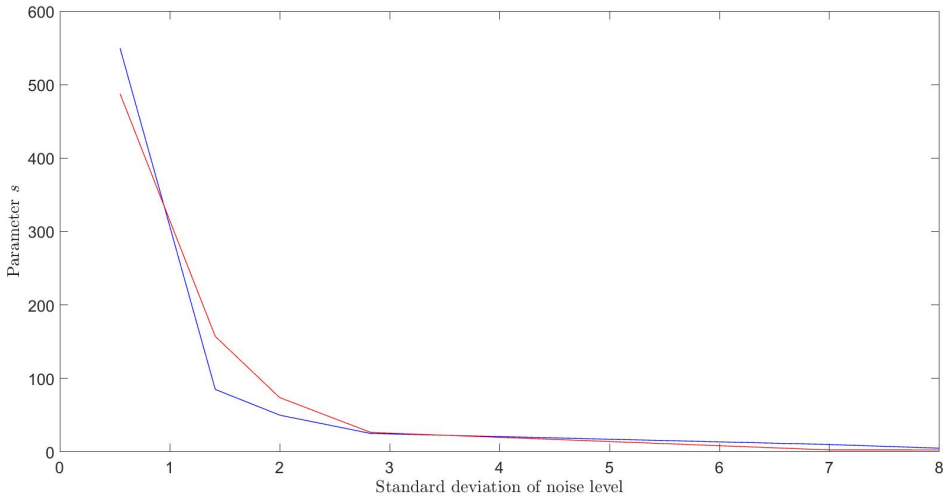


Figure 39: The relationship between the parameter s and the noise strength. The blue curve represents the best value of s for five different noise. The red curve represents a fitting line with a exponent function of the noise STD.

4.5 Summary

In summary, we have developed a new robust iterative method for image restoration in which an iterative energy function is utilised to regularize the gradient along the steepest descent by adaptively adjusting to the current state in the iterative process. We show that the iterative solutions converge to the real scene despite noise contamination in an observation and that the restoration error can be controlled to be substantially smaller than the noise level in the observation. In contrast to the well-established regularization methods, which introduce a penalty to solve the ill-posed problem, we directly apply residual optimisation through minimising the iterative energy function. This becomes a denoising problem with an iterative characteristic, and noise can be removed judiciously by applying any noise reduction method. We have undertaken two numerical experiments to investigate the performance of this method and compared the results to current regularization methods. We have shown that our method performs at a level close to the best performance of the other methods in terms of recovering elaborate structures while reducing noise and favourably compared to the most methods.

As seen in the comparison of motion blurring in 4.4.1, although IDDBM3D leads to the highest PSNR, the output image shows obvious artefacts along the edge while

our method does not. This indicates that IDDBM3D is not effective for motion blurring. Further, for the restoration of the practical image of a grating with 0.7um width in 4.4.2, IDDBM3D cannot restore the slits in its result while our method does. These results shows that our method is more stable for different types of blurring filter. In 4.4.3, the results shows that our method is stable for PSFs which are belong to Gaussian type but with various variance. At last, the result in 4.4.3 shows that our method is still effective if unknown PSFs can be estimated by Gaussian type. Moreover, our method requires simple parameter setting, particularly in the second experiment, where a single parameter is estimated from the observation. This property may represent a great advantage for real-world applications.

Chapter 5

Multi-Frame Image Restoration with ARSD

Multi-frame image restoration reconstructs one high-resolution image from a set of low-resolution images. There are several advantages to restoring images using multi-frame observations. First, multi-frame techniques perform better than single-frame models in terms of PSNR because the amount of data is greater than the amount of information required in the restored image [107]. Second, tiny features may be recovered because of sub-pixel accurate shifts between low-resolution images. Third, Gibbs phenomena can be well suppressed at the edges [108].

We extend ARSD (a single-frame model) to multi-frame models which can obtain significantly improved PSNRs and visual effects compared to single-frame models. We show that our multi-frame models perform well even when noise is not additive and depends on the structures, whereas single-frame model obtains poor results. In this chapter, we first present a review of multi-frame image restoration strategies, and then, we propose our multi-frame model. Finally, numerical results are presented and discussed.

5.1 Multi-frame image restoration

Multiple images can be acquired without many difficulties in many fields such as medical imaging, surveillance video, and satellite imagery; therefore, applications of multi-frame image restoration have become popular in these areas. For example, remote sensing images are processed into higher resolution images using several images of the same original scene. Another example is freeze-frame video, in which the quality of a single signal is too poor to be analysed; therefore, reconstruction is needed to improve the quality using several successive images. In the medical imaging techniques, such as ultrasound, CT, and MRI, due to the limited quality of

output images, clear and neat images are needed and can be obtained by restoring multi-frame low-resolution versions.

5.1.1 What is super-resolution?

As discussed in Chapter 4, imaging systems always introduce blur and noise in output images. If two points are very close, their images cannot be distinguished due to blur. Therefore, resolution is defined as the minimal distance between two distinguishable points in the evaluation of imaging systems. For example, the resolution of a microscope is often restricted by the diffraction limit and is given by the Abbe diffraction criterion of $\lambda/(2NA)$, where λ is the wavelength of light and NA is the numerical aperture of the lens. Usually, high resolution means a shorter distance between two distinguishable points. In other words, for the same ground truth, an imaging system with a higher resolution can provide greater information through output images than can a system with a lower resolution because more points can be distinguished.

Because of the resolution limit of high-frequency imaging systems, the details of objects cannot be seen in images. To discover this information, it is necessary to improve the resolution of an imaging system. Super-resolution (SR) imaging describes a class of techniques used to undertake this task. One type of technique is to develop advanced imaging systems. For example, stimulated emission deletion (STED) [109] reduced the size of PSFs by using a second excitation light beam, whereas structured illumination microscopy (SIM) [110] reduces the size of blur filters by employing optical patterning of excitations and a nonlinear response of samples.

Another type of technique for enhancing the resolution of an imaging system is SR reconstruction. SR uses signal processing techniques to restore a high-resolution (HR) image from multiple observed images referred to as low-resolution (LR) images. In many articles, the term "SR reconstruction" refers to a signal-processing approach towards resolution enhancement, because the word "super" represents very well the characteristics of the techniques for overcoming the inherent resolution limitations of LR imaging systems [111]. The advantage of SR reconstruction is its diminished cost, and existing LR imaging systems, such as confocal and widefield microscopies, can still be utilised.

Although single-frame image restoration can improve the resolution and yield high-quality images, SR reconstruction often outperforms single-frame methods in terms of improvements in resolution and better noise suppression in the restoration process [107]. This is because two types of new information are introduced by LR image sequences, in which a small difference is represented by a small shift in the same

scene [112]. First, frames with sub-pixel shifts can carry new information about structures for SR restoration. If integer pixel shifts are between the LR frames, then each frame contains the same structure information. Therefore, there is no new information about the structures-only more noise is introduced.

Second, multiple frames contain more noise patches than does a single frame if the down-sampling ratio is not too small. For example, a restored image of size $L \times L$ is reconstructed from a single frame that also has the same size $L \times L$ whereas multiple frames can provide $n \times M \times M$ samplers, where n is the total number of frames and M is the size of each frame via down sampling. For a moderate down-sampling ratio M/L, the data size of $n \times M \times M$ is greater than $L \times L$, which is the information required for the restored image. The necessary condition of a possible motionless SR restoration is formulated in [107] through discussions of the relations characterising the frame number, down-sampling ratio and kernel size.

The mathematical model of multi-frame imaging can be defined as [117]

$$J_k = D_k P_k C_k I + N_k, \text{ for } k = 1, 2, \ldots, m, \tag{5.1}$$

where I is the real scene, and J_k is the kth observation among the total m low-resolution frames. The matrix D_k represents down-sampling, P_k is the PSF of the kth frame, and C_k is the shift matrix referring to the reference image to be restored. N_k is the noise in the kth frame, which is often supposed to follow a normal distribution with a mean of zero and a variance of σ_k^2.

A common way to estimate the original scene I is to minimise the following energy function:

$$\sum_{k=1}^{m} \|J_k - D_k P_k C_k I\|_2^2 + \Phi(I), \tag{5.2}$$

where $\Phi(I)$ is the regularization.

The multi-frame model (5.2) reduces to the single-frame model,

$$\|J - PI\|_2^2 + \Phi(I), \tag{5.3}$$

when only one frame is considered. There are two differences between the two models. First, the single-frame model cannot be solved using traditional methods without the regularization term $\Phi(I)$ due to the ill-posed blur filter P. However, this conclusion may not be true for the multi-frame model,

$$\sum_{k=1}^{m} \|J_k - D_k P_k C_k I\|_2^2 \tag{5.4}$$

because more constraints from different PSFs may remove the ill-posed problem

caused by the singularity. Second, multi-frame models often yield solutions with higher resolution and less noise [107] than do single-frame models. Furthermore, it is noted that increasing the number of frames cannot increase the restored resolution beyond the upper bound posed by the kernel size; however, an increased number of frames with the same output resolution means that an improved noise reduction.

5.1.2 Method review of super-resolution

The first discussion of multi-frame reconstruction appeared in [113] by Tsay and Huang, in which a frequency-domain approach was used to reconstruct an SR image from several down-sampled noise-free observations. Other frequency-domain recursive algorithms can be found in [114]. Frequency-domain methods have the advantage of efficient computation, although they are sensitive to model errors and face difficulties in addressing complex motions in models. Due to these limitations, many methods in the spatial domain have been developed. Peleg proposed an approach [115] towards an SR restoration algorithm by introducing an iterative back projection (IBP) to analyse the simulation of LR images. One method based on theoretical restoration is POCS [116], in which images are projected onto a series of bounding ellipsoids as a priori knowledge containing the restored image. Then, the intersection vector that contains the restored image is estimated by an iterative algorithm. This method can be referred to as extended Miller's regularization. The IBP and POCS methods perform a reconstruction in a straightforward manner, although they do not produce a unique solution and cannot be applied to complex motions such as rotation. Regularization methods have become popular for multi-frame image restoration, as they do not have the shortcomings of the above methods. As in single-frame image restoration, these methods employ penalty terms on the total residual of all input frames. Some research on regularizations can be found in [117][118][119]. Based on the concept of non-local means, certain methods construct regularization terms with non-local information. Protter proposed a non-local fuzzy registration scheme-based super resolution reconstruction framework [120]. Gao proposed a novel Zernike moment-based method by improving non-local similarity matching based on the Zernike moment feature similarity [121]. Learning-based methods have received substantial attention. These methods estimate the missing high-frequency details by learning the relationships between patches in LR images and the corresponding HR patches from a training set, as in, for example, [122][123][124]. One main disadvantage of learning-based methods is their huge computational workloads from learning from training data.

5.2 Extending ARSD for multi-frame image restoration

In this section, two cases are considered: low-resolution images without shifts and down-sampling, and low-resolution images with subpixel accurate shifts and down-sampling. For the first case, we propose a weighted model that generates a resolution converging to the real scene by the weighted sum of the outputs in each iteration of ARSD. For the second case, we remove blurring and noise for each low-resolution frame by ARSD in advance and then reconstruct a high-resolution image from the restored images by an interpolation method.

We attempt to present benefits obtained by multi-frame image restoration compared to single-frame image restoration. Additionally, we would like to answer the question whether more information of noise is useful for a restoration and how the frame number improves the restoration in terms of PSNR.

5.2.1 Restoration with neither shifts nor down-sampling

Because low-resolution images contain neither shifts nor down-sampling, the structure recovery is the same between single-frame and multi-frame. The only difference is noise. Multiple frames contain more noisy patches than a single frame. The simplest way to apply ARSD method for multi-frame without shifts and down-sampling is to generate an input image by averaging all low-resolution images. However, the similarity to multi-framing is not considered during the procedure of structure recovery, which removes small structures, such as spots, and makes the output image over smooth. Therefore, we introduce a better model in which the processed image during each iteration of ARSD is constructed through weights designed by the similarity of the multi-frame. This is a weighted model and leads to further improvement over the simple averaging method.

For convenience, we denote $IterVan(J, I)$ as a step in VanCittert's iteration, and $R_N(\cdot)$ denotes the noise reduction filter. The model can be formulated in the following steps for iterations $i = 1, 2, \ldots, n$,

Deblurring: $I^*_{i+1,k} = IterVan(J_k, I_k) = I_k + P^T(J_k - PI_{ik})$, for $k = 1, 2, \ldots, m$,

Denoising: $I^{**}_{i+1,k} = R_N(I^*_{i+1,k})$, for $k = 1, 2, \ldots, m$,

Aggregation: $I_{i+1} = \frac{1}{\sum_{k=1}^{m} w_{i+1,k}} \sum_{l=1}^{m} w_{i+1,k} I^{**}_{i+1,k}$,

where $w_{i+1,k}$ is the weight defined as

$$w_{i+1,k} = \frac{1}{1 + (I_{i+1,k} - I_{i+1,l})^2.} \tag{5.5}$$

Because there is no shift assumed between multiple frames, the number l can be any integer between 1 and m. The definition of the weights (5.5) can mitigate disturbances caused by random noise of a high intensity. Figure 40 presents our weighted model.

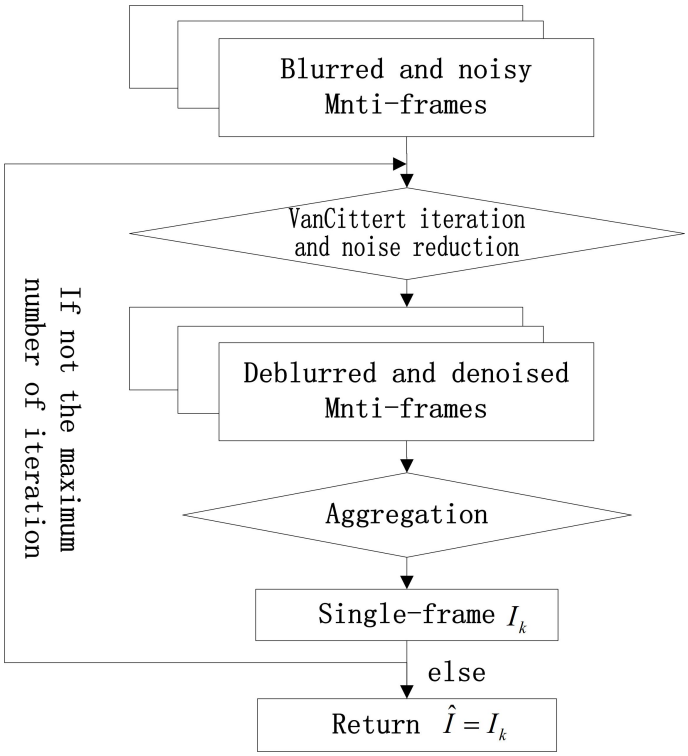


Figure 40: Flow chart of our weighted model for multi-frame image restoration.

As seen in this figure, there are three steps in this model. The first step recovers structures through VanCittert's iterative algorithm. The second step reduces the noise amplified in the first step. The last step updates the iterative solutions with the weighted sum of the image sequence output from step two. In this step, the similarity between all frames is applied to reduce the noise remaining in the iterative solutions. There are two reasons to apply the weighted sum. First, this technique considers further noise reduction for the iterative solutions while preserving structures; therefore, an input with higher quality is generated for the next iteration. The second reason is that the weighted sum can force the iterative solutions to converge by equalising $I_{i+1,k},\ k = 1, 2, \ldots, m$.

5.2.2 Super-resolution with sub-pixel shifts

As for an image sequence contained a set of low-resolution images with sub-pixel shift between each one, ARSD method is applied to each frame in the sequence at first. Thus the deblurred and denoised low-resolution sequence is obtained for the reconstruction of a high resolution image. Usually, the shifts between frames

are assumed to be small for image registration which is very important for super-resolution [125]. However, shifts are assumed to be known in our model, so this problem is not under consideration.

For convenience, we assume an equal distance of the two adjacent movements of the specimen (the original scene), and the direction of each movement is horizontal or vertical in $x - y$ plane. For example, Figure 41 demonstrates shifts between 16 frames. The first specimen is located at $(0, 0)$, then it moves along the directions indicated by the arrows. The spacial distance is 0.3 pixels for each movement. In this plot, the maximum shift is one pixel. Actually, when the maximum shift is larger than one pixel, e.g, 2.3 pixels to the right, the observed low-resolution image is almost the same as the one whose specimen locates at 0.3 pixels. The reason is that shift by integral number of pixels for a specimen cannot change the observed image except the borders due to finite scopes of lens.

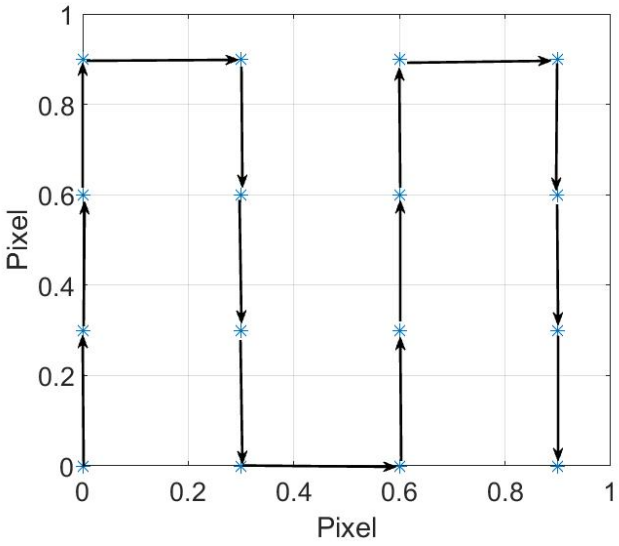


Figure 41: A demonstration of shifts of 16 frames.

A high-resolution image can be reconstructed via the deblurred and denoised low-resolution sequence. Figure 42 represents the method to reconstruct a high-resolution image. There are four low-resolution images in the left rectangle marked with five-pointed star, circle, star and triangle. The arrows show the direction of movement of specimen. The image in the right is the high-resolution image reconstructed from the four low-resolution images.

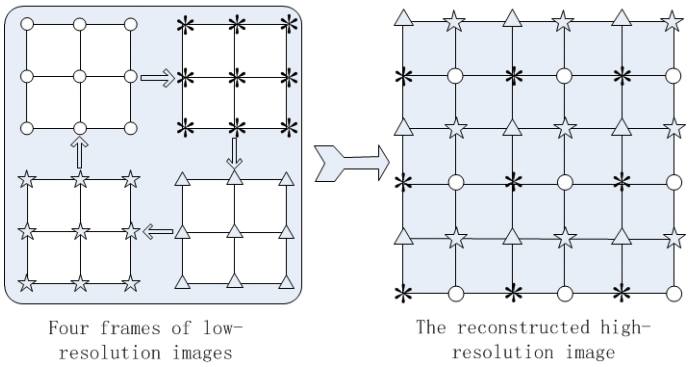


Figure 42: An illustration of reconstruction for a high-resolution image.

In the above-mentioned method of reconstruction, the displacement of adjacent two shifts are assumed to be equal. Although this assumption is not satisfied sometime, the images with equal displacement can be estimated by interpolation algorithms or Taylor expansion, which is another question not discussed in this study.

5.3 Experiments and results

Four experiments are designed to present the improvement of multi-frame image restoration and the relationship between the frame number and PSNR. In the first three, neither shifts nor down-sampling are considered within the low-resolution images; thus, these experiments test how multiple frames affect noise in the restoration process. The PSNR of multi-frame restoration is shown compared to single-frame restoration for Gaussian additive noise in the first experiment. In the second, a mixed and structure dependent noise is added, which is not consistent with our assumption. In the third experiment, PSNRs are calculated for outputs by the weighted model using 10,20,30 and 40 frames. In the last experiment, a high-resolution image is reconstructed from low-resolution images with sub-pixel accurate shifts and down-sampling by ARSD.

According the model (5.1), the PSFs and noise distributions in multi-frames might be different. However, in our experiments for multi-frame image restoration, the PSFs and noise distributions are assumed to be same. This assumption can be applied to many applications, such as nature pictures, and indoor photographing. In the procedure of imaging, the camera does not change so PSFs are the same. Also, the environment of photographing does not change much so the noise distributions can be assumed to the same.

5.3.1 Experiments with no shifts and no down-sampling

Compared to the single-frame model

This subsection attempts to present the level of improvement obtained by the multi-frame models. We analyse five scenarios to compare our multi-frame models with the single-frame model. The PSFs and variances of the noise of all scenarios are listed in Table 5, as in scenarios 1-5 in the numerical experiments in Chapter 4. Scenario 6 is not considered in this section because the PSF matrix in this scenario is nearly an identity matrix and provides minimal contribution for blurring, whereas there is heavy noise with a variance of 64. Due to the small variance of the Gaussian-shaped PSF, noise amplification during structure recovery is greatly limited, and image restoration in scenario 6 is simply a noise reduction problem.

To use ARSD with more convenience, an automatic noise estimation is applied in ARSD, which makes only one parameter in ARSD. The parameter σ_{thr} is the limit of noise strength. When the noise strength estimated in the procedure of restoration is larger than the limit, then the noise reduction filter is called. In noise estimation, we assume that there are lots of blank backgrounds which contain no information, just like the image of cameraman. The parameter can be fixed at 3 for small and moderate noise (e.g., noise in scenarios 1,2,3 and 5), 6 for large noise (e.g., noise in scenarios 4 and 6).

Table 11: PSNRs of multi-frame model(10 frames) vs. single-frame image restoration.

	Scenarios				
	1	2	3	4	5
Methods	Cameraman(256×256)				
Single-frame	30.2226	28.3734	30.3886	28.2259	27.2718
Averaging method	32.9472	31.4747	31.9246	30.0214	27.8829
Weighted method	**32.9691**	**31.4893**	**31.9285**	**30.0399**	**27.8839**

From Table 11, compared to the single-frame model, multi-frame models greatly improve the PSNR to obvious degrees in the five scenarios, e.g., PSNR rises from 30.2226 to 32.9691 for scenario 1. As shown in Figure 43, the restored images by the multi-frame models are obviously visually clearer and have sharper edges than do those by the single-frame model. Some small holes in the camera can be restored by the multi-frame models, whereas they cannot be seen in the results of the single-frame model.

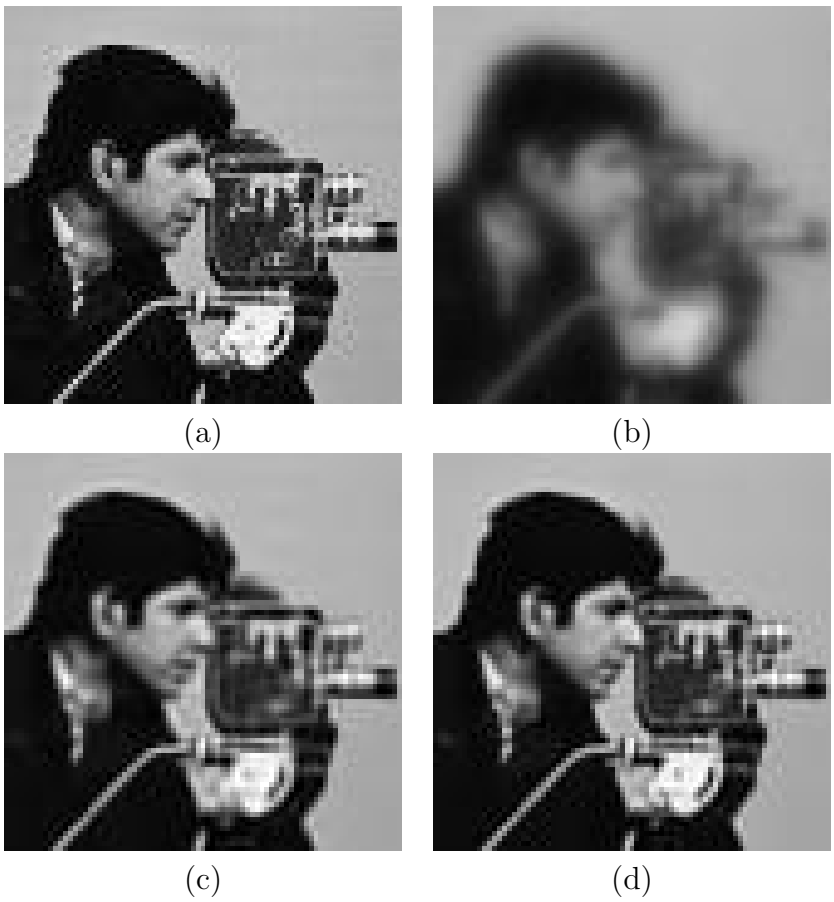


Figure 43: Restored images of Cameraman, Scenario 1. (a) Zoomed fragments of the original scene, (b) the blurred and noisy image, (c) the image restored using a single frame, (d) the image restored using the weighted method with 10 frames.

As seen from Table 11, the PSNRs of averaging and weighted methods are similar. However, the visual effect of their results are different. The output image via averaging method seems more smooth than that via weighted method and the original scene. In the original scene (a), there are spots looking like noise in the face, the black coat and the edges. Most of these spots can be restored in (c) by weighted method, while they are removed as noise in (b) by averaging method. This result shows that averaging method is preferred to smooth images due to its less computations. However, if there are small structures in images, such as spots, this method cannot distinguish this is noise or structures, and removes these structure as noise, which makes the original scene over smooth. The weighted method analyses the similarity of the iterative solutions through all the VanCittert's iteration, and thus, it can restored most of these small structures. Certainly, the price is higher computations and time consuming.

(a) (b) (c)

Figure 44: Restored images of Cameraman, Scenario 1. (a) Zoomed fragments of the original scene, (b) Zoomed fragments of the restored image via averaging method with 10 frames, (c)Zoomed fragments of the restored image via the weighted method with 10 frames.

Experiment with mixed noise

In our method we have so far assumed that the noise is independent of the structures. This assumption might not be true under certain circumstances. For example, fluorescence imaging using a microscope often contains a mixture of Poison and Gaussian noise. Although a study of mixed noise is beyond the scope of this thesis, we would like to attempt our method when the noise assumption in our model is not satisfied.

In the experiments, the image of the cameraman is blurred by the PSF of scenario 1 in Table 5, and mixed noise is added to the blurred image. The noise type is mixed sparkle noise with a parameter of 0.001 and Gaussian with a standard deviation of 2. Applying our single-frame method and multi-frame model, we obtain the output images. The original, degraded, noise and restored images are all shown in Figure 45, and PSNRs are listed in Table 12.

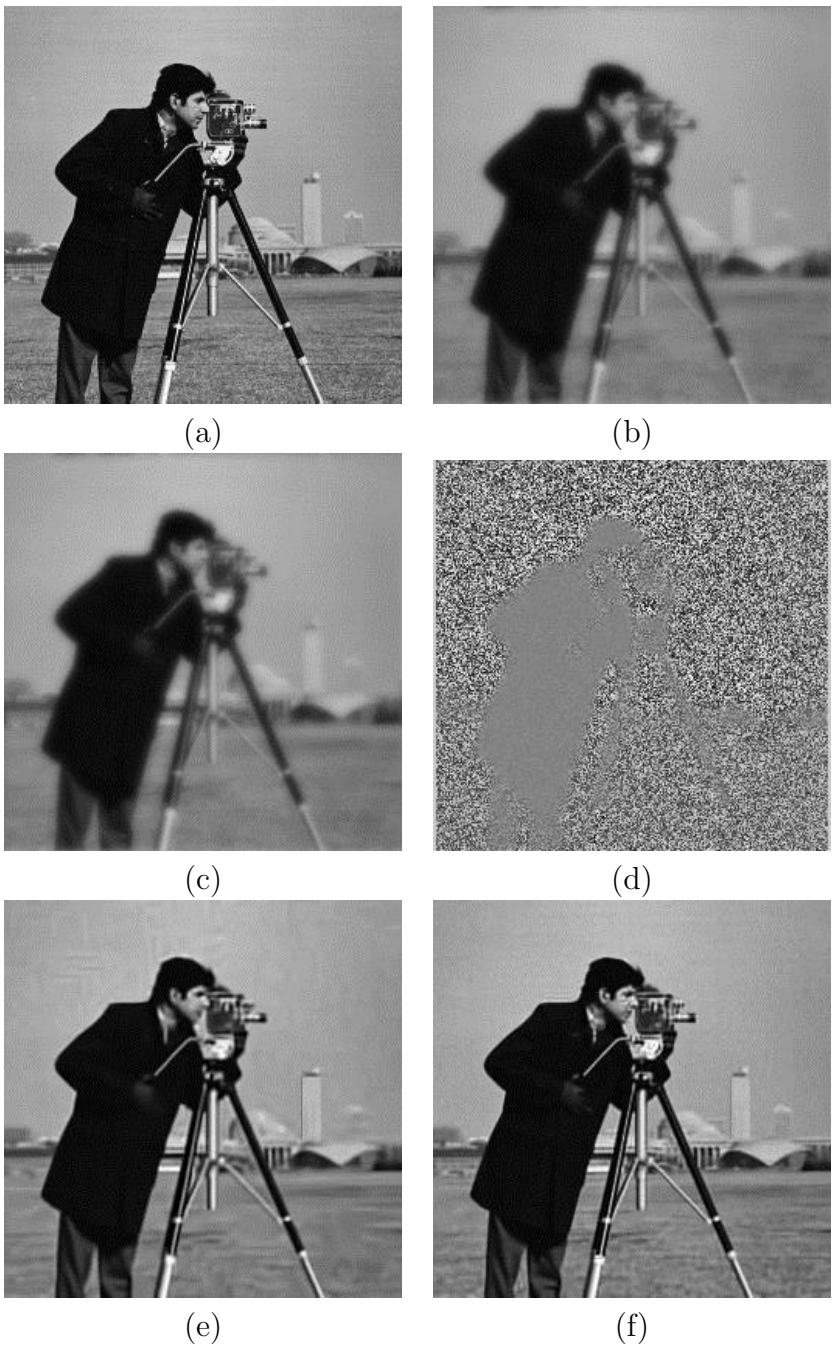


Figure 45: Restored images of Cameraman blurred with PSF of Scenario 1 and a mixed noise. (a) The original scene, (b) the blurred image, (c) the blurred and noisy image, (d) the image with the mixed noise, whose type is mixed sparkle noise with a parameter of 0.001 and Gaussian with a standard deviation of 2, (e) the restored image using a single frame, and (f) the restored image using the weighted method with 10 frames.

Table 12: Restored PSNRs for mixed noise

Methods	PSNR
Single-frame	27.1914
Weighted model	30.0823

Experiment with various frame numbers

Multi-frame models can produce an obvious improvement for image restoration, although they require numerous computations. Furthermore, more frames do not necessarily provide greater improvement. There should be a reasonable threshold for the number of frames over which the improvement of restoration saturates.

Here, we apply scenario 1, listed in Table 5, to the Cameraman image as the real scene. A total of 10, 20, 30 and 40 frames are used. The ARSD with an automated parameter setting is implemented to restore the real scene, leading to the PSNRs in Table 13. The PSNR plot is given in Figure 46, and the output images are shown in Figure 47. As seen from Table 13 and Figure 46, there is an obvious jump from the single frame to ten frames. From 10 frames to 20 frames, the PSNR increases moderately, and when the number of frames is greater than 20, the improvement of PSNR is slight. Therefore, the reasonable number of frames might be chosen to be approximately 20 in our multi-frame models. Although limited improvement of the PSNR is attained after 20 frames, we can find that some small structures (e.g., holes in the camera) are better restored in the results using 30 and 40 frames.

Table 13: Restored PSNRs for different frame numbers

Frame Number	1	10	20	30	40
PSNR	30.2226	32.9691	33.4973	33.7276	33.9467

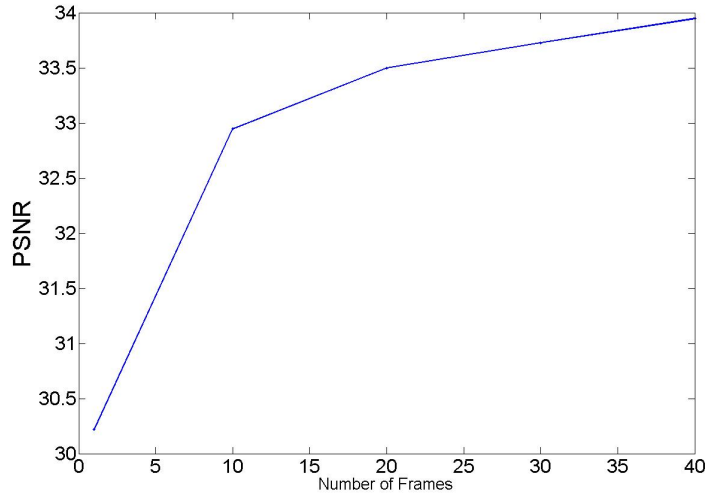


Figure 46: The curve of PSNR for different frame numbers.

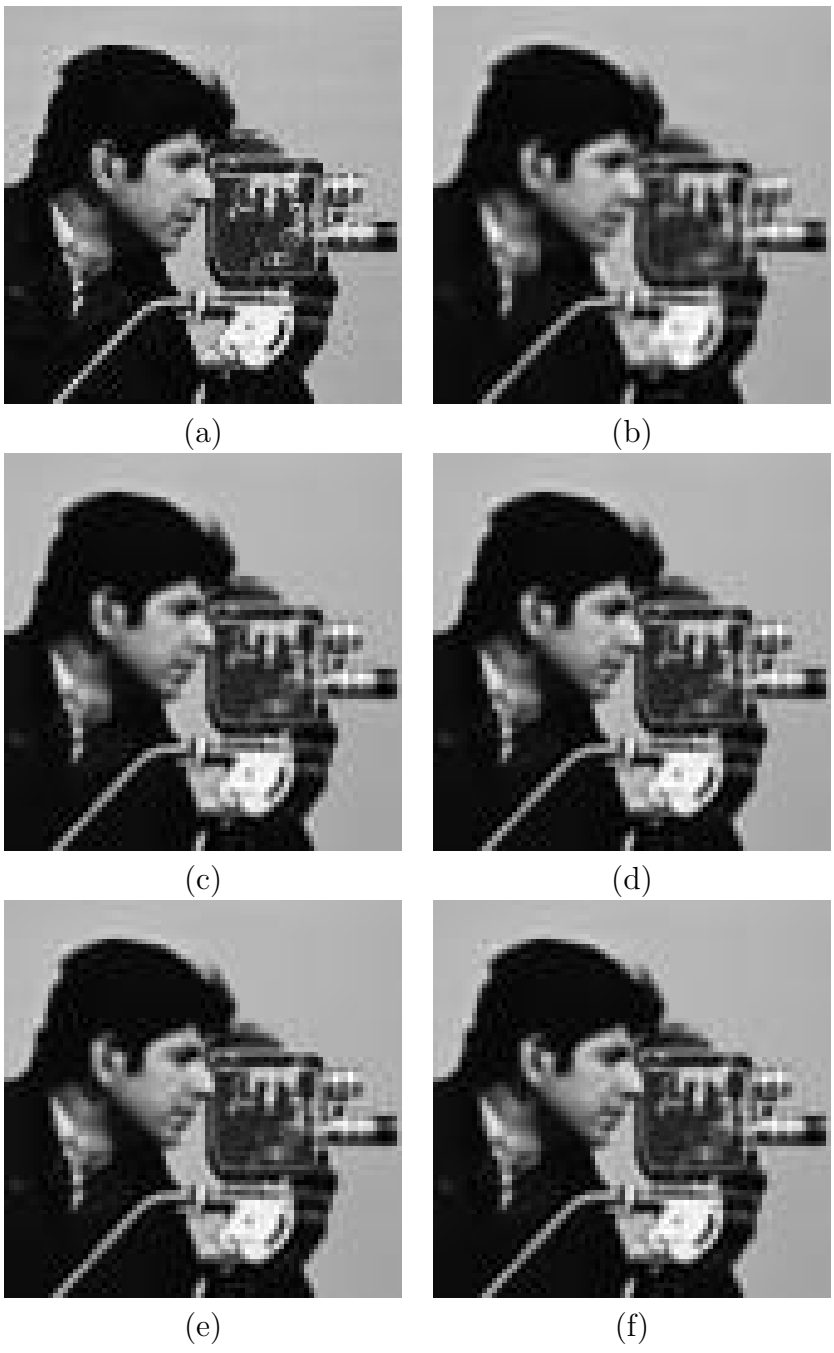


Figure 47: Restored images with different numbers of frames, Scenario 1. (a) The original scene, (b) one frame, (c) 10 frames, (d) 20 frames, (e) 30 frames, and (f) 40 frames.

5.3.2 Experiments with sub-pixel shifts

To present the advantages of multi-frame restoration from a low-resolution with sub-pixel accurate shifts, a continuous image with elaborate structures is used. The function of the image is $(1-\sin(xy))/2$. This image contains slim and sharp stripes, as shown in Figure 48 (a) and Figure 49 (a). As seen in this figure, the low-resolution image looks different from the original image; thus, it is impossible to find an estimation similar to the original with only one frame. However, the reconstructed image with four frames can express the main feature of stripes in the original image.

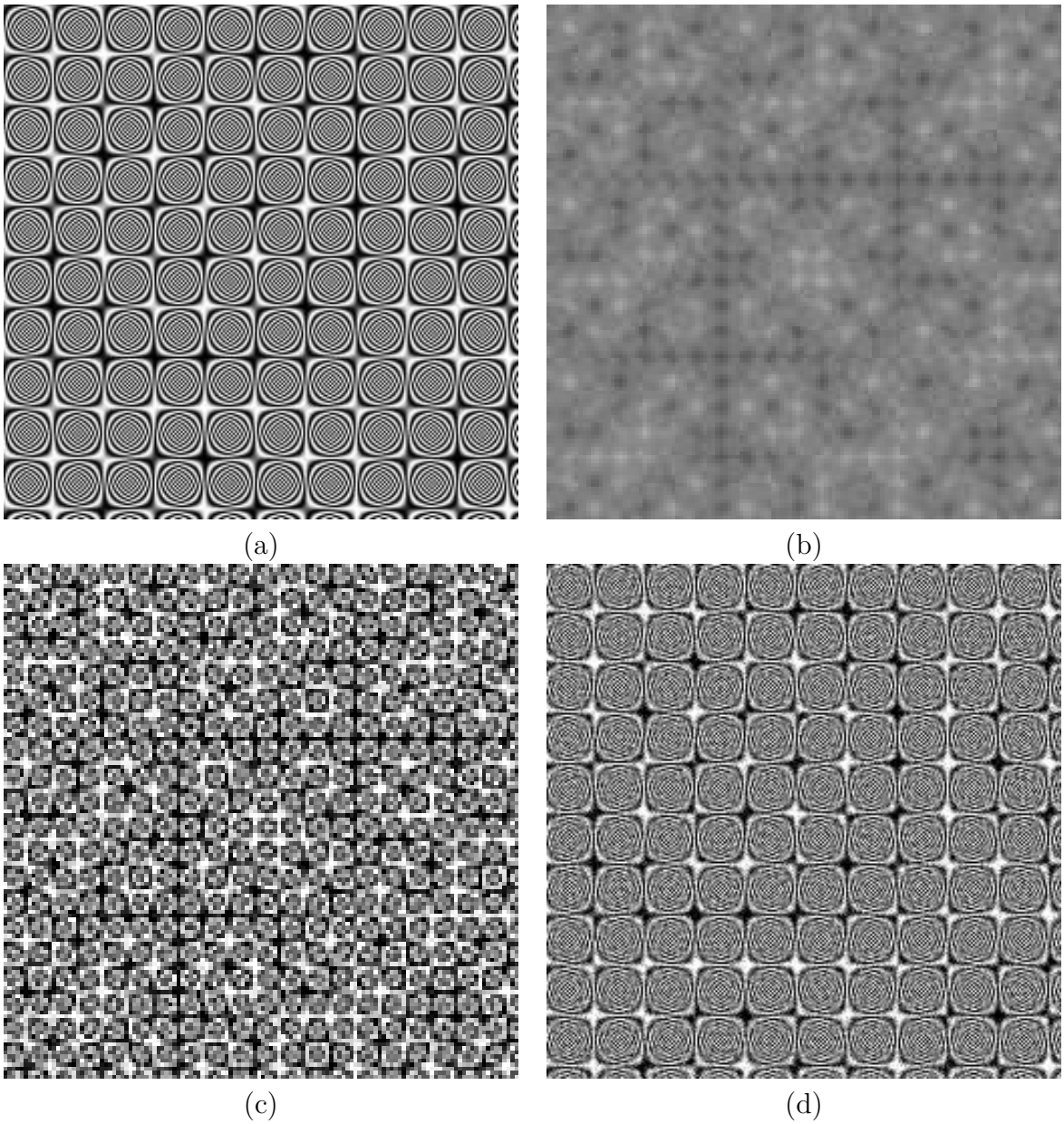
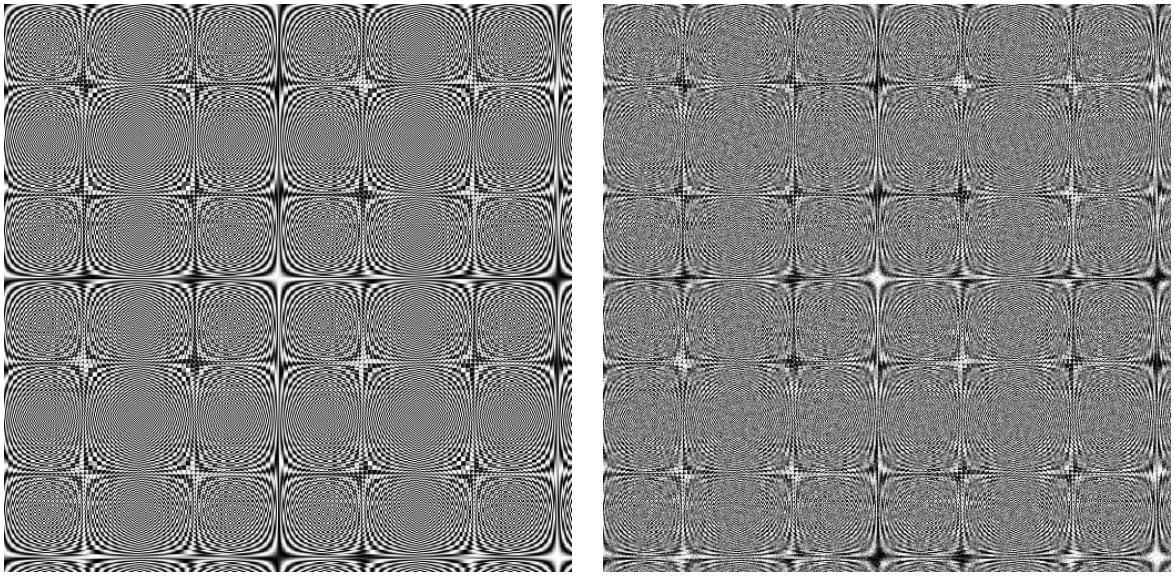


Figure 48: Reconstruction of a high-resolution image, Scenario 1. The parameters (σ, s) in ARSD are set as $(7.5, 85)$. (a) The original scene (256×256) with the sampling interval 0.5 in $[0, 128]$, (b) the first frame of low-resolution images (128×128), (c) the restored image of the first frame (128×128) via ARSD, and (d) the reconstructed high-resolution image (256×256).

In the original scene, Figure 48 (a), a shape of ♯ can be seen while it cannot be restored in the output, Figure 48 (d). Now, the sampling rate increases to 4 times and 16 low-resolution frames are obtained. The first frame is the same as Figure 48 (b). As seen in Figure 49 (b), the shape ♯ has been restored clearly.



(a) (b)

Figure 49: Reconstruction of a high resolution image from 16 low-resolution frames, Scenario 1. The parameters (σ, s) in ARSD are set as $(7.5, 85)$. (a) Original scene (512×512) with the sampling interval 0.25 in $[0, 128]$, (b) Reconstructed high-resolution image (512×512).

5.4 Summary

In this section, we can summarise the following points:

1. With the assumptions, known PSFs and additive noise independent of the structures, our multi-frame models can significantly improve the magnitudes of results in scenario 1, 2, and 3. For Gaussian PSFs there are also improvements, but they are limited. A total of 20 frames represent the reasonable number of frames for noise reduction.

2. When noise is not additive and is dependent on the structures, our single-frame model can restore the main structures, but the quality of the restored image is poor. However, multi-frame models perform well in terms of noise reduction. Therefore, the additional information about noise provided by the multiple frames can improve noise reduction.

3. For low-resolution sequence contained sub-pixel shifts, our method presents strong ability to reconstruct a high-resolution image with excellent structure preserving. High sampling rate will lead to accurate structures.

Chapter 6

Conclusions and Future Research

6.1 Conclusions

In this thesis, we present two new adaptive non-local methods, WALOF and ARSD, for component dye detection and image restoration. By utilising residual analysis and patch-based correlation analysis in these methods, significant improvements in terms of detection accuracy and structure preservation have been obtained in comparison with various existing methods.

Based on the LOF method, which is a powerful method for residual analysis in component dye detection, we propose a new method: WALOF. This non-local method designs criteria for windows located in spectra and performs stricter detections on minor components and false components through analysis-extracted features from component spectra and DNA spectra in the windows. Compared with the original LOF, WALOF can reduce the false detection ratio by 20%-30%. In addition to improving the detection accuracy, our method can also adjust the balance between two error rates (false positive and false negative) by properly setting the Balance parameter.

Concerning the image restoration, we introduce a novel iterative algorithm, ARSD, and then generalise it to multi-frame image restoration. The algorithm estimates iterative solutions by optimising a series of energy functions and forces these estimations to converge to the original scene. VanCittert's iteration and a noise reduction filter are combined for the optimization, in which the iteration restores structures based on the corrupted gradient, and the filter regularizes the corrupted gradient to obtain iterative solutions through noise reduction. Two principles, fixed parameter and automated parameter settings, are suggested for the parameter determination. There are two parameters in the fixed parameter setting, the fixed noise strength as input to the noise reduction operator of BM3D and s as the iteration interval between two regularizations for the gradients. Usually, the threshold can be fixed

at $\sigma = 7.5$, and s depends on the noise level of the observed image. A larger interval should be chosen for a lower noise level. The automated parameter setting involves only one parameter, which is a threshold for limiting the noise amplification during iterations. When the automatically estimated noise amplification is larger than this threshold, the noise reduction operator will be called with the input threshold. For most images, the threshold can be fixed at 3. The two principles are established based on enormous numerical measurements, which show that our method achieves a stable performance in a large parameter space. We undertake experiments with synthetic data, and the results show that our method performs better than most existing methods in terms of recovering elaborate structures and reducing noise.

Additionally, we apply our method to practical X-ray images of gratings with slit widths less than $1.0\mu m$. Due to the resolution limits of imaging instruments, the images of the gratings severely degrade when the slit width is less than $1.0\mu m$. When the slit width decreases to $0.7\mu m$, slits cannot be recognised in the images because their intensity distribution is flat. The results of our method show that these slits can be recovered well with clear edges and obvious structures, even when their width is less than $0.85\mu m$ the resolution limit of the instruments.

The efficiency and stability of the ARSD algorithm are measured with different Gaussian PSFs. The results of the efficiency tests show that our method can achieve an improvement of 2.3 times the resolution of the observed images; in other words, a shrinkage ratio of 0.43 is obtained by comparing output beads with beads in the observed images. Stability tests provide evidence that our method allows the applied PSF to have a disturbance of 10% of the actual Gaussian PSF in the variance. This property is important for practical applications in which the variance of the PSF cannot be measured accurately.

Based on the single-frame model, we extend the ARSD algorithm to a multi-frame image restoration model. In the model, each iterative solution is generated as a sum of all frames with weights determined based on the similarity of these frames. Because of the additional noise information provided and the similarity of these frames when considered in multiple frames, the model yields a higher PSNR than does the single-frame model, even when the single frame is the average of all frames. The improvement in terms of PSNR is also measured when increasing the number of frames. The results show that PSNR improves significantly when the number of utilised frames is less than 20, while a few improvements are obtained when the number is more than 20. Therefore, 20 frames are the proper number of frames for multi-frame image restoration. Furthermore, the results show that the majority of the improvement in the PSNR is caused by noise reduction; the deblurring contribution is rather limited.

After all, noise in practical images looks more complicated and it might not satisfy

our assumptions in the theoretical model. For this, finally, we attempt to apply the ARSD algorithm to real bead images compared to synthetic images with mixed noise depending on the structures. The results show that our multi-frame method can perform well for such complicated noise, whereas the single-frame model cannot output high-quality images.

Briefly, we developed non-local methods and combined them with traditional algorithms for data analysis. These methods achieve better performance than most existing methods such as LOF, ForWaRD, and SV-GSM. Moreover, our methods present advantages in terms of convenience of use, computational efficiencies and easy generalisation to models with other popular algorithms.

6.2 Novelty

This thesis yields two innovations:

- A novel approach to choosing non-local areas (windows) for feature extraction in spectra analysis.
- A different approach to presenting popular regularization methods. This new approach restores structures and reconstructs an original scene through the regularised gradient, thus allowing error analysis, which is difficult to attain in regularization methods. The main contributions of the thesis are outlined as follows:
- As a supplement to the theory of statistical residual analysis, our method, WALOF, achieves a decrease (30%) in the number of false detections compared to the original LOF method proposed by Renishaw.
- Our method extends inverse operator theory to image restoration in the presence of noise.
- Deblurring and denoising are decoupled in our algorithm, which leads to a performance similar to that of state-of-the-art methods and favourable compared to many other methods.
- Errors can be estimated in our method, whereas this is extremely difficult in regularization methods.
- Our method can be generalised with any well-performing noise reduction filters and multi-frame image restoration models, which can lead to significant improvements in the restored image. The parameter configuration is very simple and robust in terms of performance.

6.3 Future work

In this thesis, we develop the WALOF and the ARSD methods for use in complicated applications, and these new non-local methods may be extended into the following fields in the future.

1. With increasing the number of present dyes, a more complicated criterion should be developed to determine the terminated LOF, because the probability that multiple minor components are present in a DNA spectrum will increase. Thus the current criterion, in comparison to hard thresholds, faces difficulties in determining the terminated LOF.

2. Due to the side effect of the variability of dyes on the algorithm, other denoising algorithms (e.g., NLM) can be applied to replace the current algorithm denoising using a simple average.

3. In the WALOF method, the features of windows can be extracted in a high-dimensional space in addition to the Pearson correlation coefficient. For example, the structure similarity (SSIM) of a window can be applied to improve the detection accuracy.

4. For certain PSFs, the iterative solutions converge slowly, i.e., they produce a large, uniform PSF. An improvement of the algorithms in terms of speeding up the convergence under these PSFs will be considered.

5. Local PCA analysis will be combined with spatial shrinkage to achieve a better performance for noise reduction.

6. A priori information can be introduced in VanCittert's iteration to restore images when the PSFs are unknown.

References

[1] O. Bretscher. *Linear Equations*. Pearson Publisher, 2008.

[2] M. Zhong, M. Girolami, K. Faulds, and D. Graham. *Bayesian methods to detect dye-labelled dna oligonucleotides in multiplexed raman spectra.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 60(2):187–206, 2011.

[3] R.L. Lagendijk and J. Biemond. *Iterative Identification and Restoration of Images*. Kluwer Academic Publishers, 1991.

[4] Wikipedia. *Linear approximation.* `https://en.wikipedia.org/wiki/Linear_approximation`. Access date: 16 July, 2018.

[5] Wikipedia. *Least squares.* `https://en.wikipedia.org/wiki/Least_squares`. Access date: 16 July, 2018.

[6] G. Claeskens and N.L. Hjort. *Model Selection and Model Averaging*. Cambridge University Press,Cambridge, 2008.

[7] P.C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.

[8] F. Rooms, B. Goossens, A. Pižurica, and W. Philips. *Image restoration and application in biomedical processing.* International Journal of Climatology, 32(3):347–363, 2011.

[9] S. Wang, Y. Cao, and Y. Huang. *Advances in Neural Networks.* Third International symposium on Neural Networks, Chengdu, China, May 28 - June 1, 2006, Part II, 569–574, 2006.

[10] H. Akaike. *Information theory and an extension of the maximum likelihood principle.* Second International Symposium on Information Theory, 267-281, 1973.

[11] G. Schwarz. *Estimating the dimension of a model.* Annals of Statistics, 6(2), 15–18, 1978.

[12] I. Bell and J. Filik. *A method for the detection of known spectral components down to trace concentrations.* Report from Renishaw.

[13] N. Wiener. *Extrapolation Interpolation and Smoothing of Stationary Time Series.* Wiley, New York, 1949.

[14] A.N. Tikhonov and V.I.A. Arsenin. *Solutions of ill-posed problems.* Winston, 1977.

[15] L.I. Rudin, S. Osher, and E. Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D: Nonlinear Phenomena, 60(1), 259–268, 1992.

[16] D.L. Donoho and J.M. Johnstone. *Ideal spatial adaptation by wavelet shrinkage.* Biometrika, 81(3), 425–455, 1994.

[17] A. Buades, B. Coll, and J. Morel. *A review of image denoising algorithms, with a new one.* Multiscale Modeling & Simulation, 4(2), 490–530, 2005.

[18] A. Foi, K. Dabov, and K. Egiazarian. *Shape-adaptive dct for denoising and image reconstruction.* Proceedings of Spie the International Society for Optical Engineering, 6064:203–214, 2006.

[19] A. Danielyan, V. Katkovnik, and K. Egiazarian. *Bm3d frames and variational image deblurring.* IEEE Transactions on Image Processing, 21(4), 1715–1728, 2012.

[20] C.Y.Sin and H. White. *Information criteria for selecting possibly misspecified parametric models.* Journal of Econometrics, 71(1–2), 207–225, 1996.

[21] R. Woods R. Gonzalez and S. Eddins. *Digital Image Processing Using Matlab.* Prentice Hall, 2003.

[22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. *Image denoising by sparse 3-d transform-domain collaborative filtering.* IEEE Transactions on Image Processing, 16(8), 2080–2095, 2007.

[23] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, et al. *Bm3d image denoising with shape-adaptive principal component analysis.* Signal Processing with Adaptive Sparse Structured Representations, 2009.

[24] L. Zhang, W. Dong, D. Zhang, and G. Shi. *Two-stage image denoising by principal component analysis with local pixel grouping.* Pattern Recognition, 43(4), 1531–1549, 2010.

[25] G. Gilboa and S. Osher. *Nonlocal operators with applications to image processing.* Multiscale Modeling & Simulation, 7(3), 1005–1028, 2009.

[26] L. Yaroslavsky. *Digital picture processing: an introduction.* Springer-Verlag, 1985.

[27] M. Zhong, M. Girolami, K. Faulds, and D. Graham. *Bayesian methods to detect dye-labelled dna oligonucleotides in multiplexed raman spectra.* Journal of the Royal Statistical Society, 60(2), 187–206, 2011.

[28] D.D. Lee and H. Sebastian Seung. *Algorithms for non-negative matrix factorization.* In International Conference on Neural Information Processing Systems, 535–541, 2000.

[29] P. Sajda, S. Du, and L.C. Parra. *Recovery of constituent spectra using non-negative matrix factorization.* Wavelets: Applications in Signal and Image Processing X, International Society for Optics and Photonics, 5207, 321–331, 2003.

[30] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences.* Society for Industrial and Applied Mathematics, 1994.

[31] T. Furukawa. *Biological Imaging and Sensing.* Springer Science & Business Media, 2004.

[32] I. Bankman. *Handbook of medical image processing and analysis.* Elsevier, 2008.

[33] J.F. Claerbout. *Imaging the Earth's Interior.* Blackwell Scientific Publications, Inc., Cambridge, MA, USA, 1985.

[34] S.G. Chang, B. Yu, and M. Vetterli. *Spatially adaptive wavelet thresholding with context modeling for image denoising.* IEEE Transactions on image Processing, 9(9), 1522–1531, 2000.

[35] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. *Image denoising using scale mixtures of gaussians in the wavelet domain.* IEEE Transactions on Image Processing, 12(11), 1338, 2003.

[36] C.A. Deledalle. *Image denoising with patch based pca : Local versus global.* British Machine Vision Conference, 2011.

[37] P. Chatterjee and P. Milanfar. *Patch-based near-optimal image denoising.* IEEE Transactions on Image Processing, 21(4):1635–1649, 2012.

[38] X. Chen, S.B. Kang, J. Yang, and J. Yu. *Fast patch-based denoising using approximated patch geodesic paths.* IEEE Conference on Computer Vision and Pattern Recognition, 1211–1218, 2013.

[39] S. Roth and M.J. Black. *Fields of experts: A framework for learning image priors.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 860–867, 2005.

[40] M. Elad and M. Aharon. *Image denoising via sparse and redundant representations over learned dictionaries.* IEEE Transactions on Image Processing, 15(12), 3736-3745, 2006.

[41] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. *Non-local sparse models for image restoration.* IIEEE International Conference on Computer Vision, 2272–2279, 2010.

[42] L. Zhang, W. Dong, D. Zhang, and G. Shi. *Two-stage image denoising by principal component analysis with local pixel grouping.* Pattern Recognition, 43(4), 1531–1549, 2010.

[43] H.J. Trussell and M. Civanlar. *The feasible solution in signal restoration.* IEEE Transactions on Acoustics Speech & Signal Processing, 32(2), 201–212, 1984.

[44] R. Lagendijk, R. Mersereau, and J. Biemond. *On increasing the convergence rate of regularized iterative image restoration algorithms.* IEEE International Conference on Acoustics, Speech, and Signal Processing, 12, 1183–1186, 1987.

[45] T. Sarkar, F. Tseng, S. Dianat, and B. Hollmann. *Deconvolution by the conjugate gradient method.* IEEE International Conference on Acoustics, Speech, and Signal Processing, 10, 445–448, 1985.

[46] A.K. Katsaggelos, J. Biemond, R.W. Schafer, and R.M. Mersereau. *A regularized iterative image restoration algorithm.* IEEE Transactions on Signal Processing, 39(39), 914–929, 1991.

[47] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. *An iterative regularization method for total variation-based image restoration.* Multiscale Modeling & Simulation, 4(2), 460–489, 2005.

[48] F Sroubek and J Flusser. *Multichannel blind iterative image restoration.* IEEE Transactions on Image Processing, 12(9),1094–1106, 2003.

[49] J. M. Bioucas-Dias and M. A. T. Figueiredo. *A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image Processing*, 16(12), 2992-3004, 2007.

[50] S. Kawata and Y. Ichioka. *Iterative image restoration for linearly degraded images. ii. reblurring procedure.* Journal of the Optical Society of America, 70(7), 768–772, 1980.

[51] M.Z. Nashed. *Aspects of generalized inverses in analysis and regularization.* Generalized Inverses & Applications, 193–244, 1976.

[52] M.Z. Nashed. *Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory.* IEEE Transactions on Antennas & Propagation, 29(2), 220–231, 1981.

[53] G.H. Golub, M. Heath, and G. Wahba. *Generalized cross-validation as a method for choosing a good ridge parameter.* Technometrics, 21(2), 215–223, 1979.

[54] P.C. Hansen. *Analysis of discrete ill-posed problems by means of the l-curve.* Siam Review, 34(4), 561–580, 1992.

[55] M.E. Kilmer, P.C. Hansen and R.H. Kjeldsen. *Exploiting residual information in the parameter choice for discrete ill-posed problems.* Bit Numerical Mathematics, 46(1), 41–59, 2006.

[56] T. Goldstein and S. Osher. *The Split Bregman Method for L1-Regularized Problems.* Society for Industrial and Applied Mathematics, 2009.

[57] P.L. Combettes and J.C. Pesquet. *A douglas–rachford splitting approach to nonsmooth convex variational signal recovery.* IEEE Journal of Selected Topics in Signal Processing, 1(4), 564–574, 2007.

[58] E. Esser. *Applications of Lagrangian-based alternating direction methods and connections to split Bregman.* Cam Report, 2009.

[59] D. Chen. *Regularized generalized inverse accelerating linearized alternating minimization algorithm for frame-based poissonian image deblurring.* Siam Journal on Imaging Sciences, 7(1), 716–739, 2014.

[60] A. Chambolle. *An algorithm for total variation minimization and applications.* Journal of Mathematical Imaging and Vision, 20(1), 89–97, 2004.

[61] C.A. Micchelli, L. Shen, Y. Xu, and X. Zeng. *Proximity algorithms for the l1/tv image denoising model.* Advances in Computational Mathematics, 38(2), 401–426, 2013.

[62] P. Chen, J. Huang, and X. Zhang. *A primal dual fixed point algorithm for convex separable minimization with applications to image restoration.* Inverse Problems, 29(2), 025011, 2013.

[63] Y.W. Wen, M.K. Ng, and W.K. Ching. *Iterative algorithms based on decoupling of deblurring and denoising for image restoration.* Siam Journal on Scientific Computing, 30(5), 2655–2674, 2007.

[64] J.F. Cai, S. Osher, and Z. Shen. *Split bregman methods and frame based image restoration.* Siam Journal on Multiscale Modeling & Simulation, 8(2), 337–369, 2009.

[65] X.G. Lv, Y.Z. Song, and F. Li. *An efficient nonconvex regularization for wavelet frame and total variation based image restoration.* Journal of Computational and Applied Mathematics, 290, 553 – 566, 2015.

[66] G. Peyré, S. Bougleux, and Laurent Cohen. *Non-local regularization of inverse problems.* European Conference on Computer Vision, 57–68, 2008.

[67] H. Ji, C. Liu, Z. Shen, and Y. Xu. *Robust video denoising using low rank matrix completion.* Computer Vision and Pattern Recognition, 1791–1798, 2010.

[68] W. Dong, G. Shi, and X. Li. *Nonlocal image restoration with bilateral variance estimation: a low-rank approach.* IEEE Transactions on Image Processing, 22(2), 700–711, 2013.

[69] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.H. Yang. *Image deblurring via enhanced low-rank prior.* IEEE Transactions on Image Processing, 25(7), 3426–3437, 2016.

[70] K. Miller. *Least squares methods for ill-posed problems with a prescribed bound.* Siam Journal on Mathematical Analysis, 1(1), 52–74, 2012.

[71] J.F. Cai, E.J. Candés, and Z. Shen. *A singular value thresholding algorithm for matrix completion.* Siam Journal on Optimization, 20(4), 1956–1982, 2008.

[72] S. Gu, L. Zhang, W. Zuo, and X. Feng. *Weighted nuclear norm minimization with application to image denoising.* Computer Vision and Pattern Recognition, 2862–2869, 2014.

[73] S.R. Safavian and D. Landgrebe. *A survey of decision tree classifier methodology.* IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660–674, 1991.

[74] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, and C.J. Lin. *Liblinear: A library for large linear classification.* Journal of machine learning research, 9, 1871–1874, June 2008.

[75] T.M. Mitchell. *Machine Learning.* McGraw-Hill Science/Engineering/Math, 1997.

[76] A. Krizhevsky, I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks.* Proceedings of the 25th International Conference on Neural Information Processing Systems, 1, 1097–1105, USA, 2012.

[77] J.A.K. Suykens and J. Vandewalle. *Least squares support vector machine classifiers.* Neural Processing Letters, 9(3), 293–300, 1999.

[78] P. Langley, W. Iba, and K. Thompson. *An analysis of bayesian classifiers.* Proceedings of the Tenth National Conference on Artificial Intelligence, 223–228, 1992.

[79] N. Friedman, D. Geiger, and M. Goldszmidt. *Bayesian network classifiers.* Machine Learning, 29(2), 131–163, 1997.

[80] Y.s. Jing, V. Pavlović, and J.M. Rehg. *Boosted bayesian network classifiers.* Machine Learning, 73(2), 155–184, 2008.

[81] J. Cheng, D.A. Bell, and W. Liu. *Learning belief networks from data:an information theory based approach.* Artificial Intelligence, 137(1–2), 43–90, 2002.

[82] J. Cheng and R. Greiner. *Comparing Bayesian network classifiers.* Fifteenth Conference on Uncertainty in Artificial Intelligence, 101-108, 1999.

[83] G. D. Luker and K. E. Luker. *Optical imaging: current applications and future directions.* Journal of Nuclear Medicine Official Publication Society of Nuclear Medicine, 49(1), 1–4, 2008.

[84] T. Furukawa. *Biological Imaging and Sensing.* Springer Berlin Heidelberg, 2004.

[85] N. Bankman. *Handbook of Medical Image Processing and Analysis (Second Edition).* Elsevier„ 2009.

[86] M. Minsky. *Memoir on inventing the confocal scanning microscope.* Scanning, 10(4), 128–138, 1988.

[87] P. Török and F.J. Kao. *Optical imaging and microscopy:Techniques and Advanced Systems.* Springer Verlag, 2003.

[88] R.L. Lagendijk and A.M. Tekalp. *Maximum likelihood image and blur identification: a unifying approach.* Optical Engineering, 29(29), 422–435, 1990.

[89] A.K. Katsaggelos and K.T. Lay. *Maximum likelihood blur identification and image restoration using the em algorithm.* Signal Processing IEEE Transactions on, 39(3), 729–733, 2016.

[90] S.D. Babacan, R. Molina, M.N. Do, and A.K. Katsaggelos. *Bayesian blind deconvolution with general sparse image priors.* European Conference on Computer Vision, 341–355, 2012.

[91] C.L. Likas and N.P. Galatsanos. *A variational approach for bayesian blind image deconvolution.* IEEE Transactions on Signal Processing, 52(8), 2222–2233, 2004.

[92] M.M. Bronstein, A.M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. *Blind deconvolution of images using optimal sparse representations.* IEEE Transactions on Image Processing, 14(6), 726–736, 2005.

[93] M. Potmesil and I. Chakravarty. *A lens and aperture camera model for synthetic image generation.* Association for Computing Machinery, 15, 1832-1837, 1981.

[94] D.L. Angwin and H. Kaufman. *Image restoration using reduced order models.* Signal Processing, 16(1), 21–28, 1989.

[95] P.A. Stokseth. *Properties of a defocused optical system.* Journal of the Optical Society of America, 59(10), 1314–1321, 1969.

[96] B.L. Mcglamery. *Restoration of turbulence-degraded images.* Journal of the Optical Society of America, 57(3), 293–296, 1967.

[97] F.C. Wagner, A. Macovski, and D.G. Nishimura. *A characterization of the scatter point-spread-function in terms of air gaps.* IEEE Transactions on Medical Imaging, 7(4), 337–44, 1988.

[98] J. Hadamard and P.M. Morse. *Lectures on Cauchy's Problem in Linear Partial Differential Equations.* Dover, 1952.

[99] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. *Image quality assessment: from error visibility to structural similarity.* IEEE Transactions on Image Processing, 13(4), 600, 2004.

[100] R. Neelamani, H. Choi, and R. Baraniuk. *Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems.* IEEE Transactions on Signal Processing, 52(2), 418–433, 2002.

[101] J.A. Guerrero-Colon, L. Mancera, and J. Portilla. *Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids.* IEEE Transactions on Image Processing, 17(1), 27–41, 2008.

[102] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. *Image restoration by sparse 3d transform-domain collaborative filtering.* Image Processing: Algorithms and Systems VI, 6812(8), 681207, 2008.

[103] J. Portilla. *Image restoration through l_0 analysis-based sparse optimization in tight frames.* IEEE International Conference on Image Processing, 3909–3912, 2010.

[104] J.P. Oliveira, J.M. Bioucas-Dias, and M.A.T. Figueiredo. *Review: Adaptive total variation image deblurring: A majorization-minimization approach.* Signal Processing, 89(9), 1683–1693, 2009.

[105] G. Chantas, N.P. Galatsanos, R. Molina, and A.K. Katsaggelos. *Variational Bayesian image restoration with a product of spatially weighted total variation image priors.* IEEE Transactions on Image Processing, 19(2), 351–362, 2010.

[106] M. Nikolova, M.K. Ng, and C.P. Tam. *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction.* IEEE Transactions on Image Processing, 19(12), 3073-3088, 2010.

[107] M. Elad and A. Feuer. *Restoration of single super-resolution image from several blurred.* IEEE Transactions on Image Processing, 6(12), 1646 - 1658, 1997.

[108] D.D. Muresan and T.W. Parks. *Adaptive principal components and image denoising.* International Conference on Image Processing, 1, I-101-4, 2003.

[109] S.W. Hell. *Microscopy and its focal switch.* Nature Methods, 6(1), 24–32, 2009.

[110] R. Heintzmann and M.G.L. Gustafsson. *Subdiffraction resolution in continuous samples.* Nature Photonics, 3(7), 362–364, 2009.

[111] S.C. Park, M.K. Park, and M.G. Kang. *Super-resolution image reconstruction: a technical overview.* Nature Photonics, 20(7), 21–36, 2003.

[112] S. Chaudhuri. *Super-Resolution Imaging.* Kluwer Academic Publishers, 2001.

[113] R.Y. Tsai and T.S. Huang. *Multiframe image restoration and registration.* Advances in Computer Vision and Image Processing, 1(2), 317-339, 1984.

[114] S.P. Kim, N.K. Bose, and H.M. Valenzuela. *Recursive reconstruction of high resolution image from noisy undersampled multiframes.* IEEE Transactions on Acoustics Speech & Signal Processing , 38(6), 1013–1027, 1990.

[115] S. Peleg, D. Keren, and L. Schweitzer. *Improving image resolution using subpixel motion.* Pattern Recognition Letters, 5(3), 223–226, 1987.

[116] A.K. Katsaggelos. *A multiple input image restoration approach.* Journal of Visual Communication & Image Representation, 1(1), 93–103, 1990.

[117] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar. *Fast and robust multiframe super resolution.* IEEE Transactions on Image Processing, 13(10), 1327–1344, 2004.

[118] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning. *A multi-frame image super-resolution method.* Signal Processing, 90(2), 405–414, 2010.

[119] H. Takeda, S. Farsiu, and P. Milanfar. *Kernel regression for image processing and reconstruction.* IEEE Transactions on Image Processing, 16(2), 349–66, 2007.

[120] M. Protter, M. Elad, H. Takeda, and P. Milanfar. *Generalizing the nonlocal-means to super-resolution reconstruction.* IEEE Transactions on Image Processing, 18(1), 36–51, 2009.

[121] X. Gao, Q. Wang, X. Li, D. Tao, and K. Zhang. *Zernike-moment-based image super resolution.* IEEE Transactions on Image Processing, 20(10), 2738–2747, 2011.

[122] C.Y. Yang, C. Ma, and M.H. Yang. *Single-image super-resolution: A benchmark.* European Conference on Computer Vision, 372–386, 2014.

[123] M. Liang, J. Du, and L. Li. *Video super-resolution reconstruction based on correlation learning and spatio-temporal nonlocal similarity.* Multimedia Tools & Applications, 75(17), 10241–10269, 2016.

[124] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T.S. Huang. *Coupled dictionary training for image super-resolution.* IEEE Transactions on Image Processing, 21(8), 3467, 2012.

[125] M. Irani and S. Peleg. *Improving resolution by image registration.* Graphical models and image processing, 53(3), 231–239, 1991.