Computational Super-resolution Microscopy: Theory and Practical Application

by

Jian Xing

B.Eng., The Hong Kong Polytechnic University, 2014

M.S., University of Colorado Boulder, 2016

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Electrical, Computer, and Energy Engineering

2021

Committee Members:

Prof. Carol Cogswell, Chair

Prof. Stephen Becker

Prof. Robert McLeod

Prof. Rafael Piestun

Prof. Wounjhang Park

Xing, Jian (Ph.D., Electrical Engineering)

Computational Super-resolution Microscopy: Theory and Practical Application

Thesis directed by Prof. Carol Cogswell

High-resolution fluorescence microscopy is an indispensable tool in biological studies. Due to the diffraction of light, images acquired using conventional fluorescence microscopy techniques exhibit limited resolution. This *diffraction limit* restricts the level of fine details accessible to the users of a fluorescence microscope, and bypassing this limit would enable the investigation of finer structures, and unlock additional insights for biological studies.

Over the past few decades, an array of super-resolution fluorescence microscopy techniques that bypass the diffraction limit have been proposed, developed, and popularized. These techniques typically exploit the physical properties of the fluorophores used to label the sample, and require the use of specialized illumination setups or photo-switchable fluorophores. These requirements can hinder the adoption of super-resolution microscopy by the larger biological research community, and it is therefore beneficial to attempt to achieve super-resolution while circumventing some or all of these special requirements. It has recently been shown that super-resolution can indeed be achieved by using non-switchable fluorophores and numerical post-processing. Despite being a simpler system, this avenue of achieving super-resolution has received less research effort. While some earlier proof-of-concept experiments have been performed in less practical settings, it is not until much more recently that a proof-of-concept experiment was demonstrated in a biological fluorescence microscopy setting.

Despite recent progress, there are open questions that need to be answered before this *computational* super-resolution approach can be adopted by a wider audience. In this thesis work, three questions will be addressed: 1) what is the most suitable (i.e., achieves the highest resolution) processing scheme for this super-resolution technique? 2) is the recovered image the only possible solution (i.e., is the solution unique)? and 3) given these promising super-resolution results, what

conclusions can be drawn in the analysis of the *information transfer* ability of a microscope?

For the first question, a careful examination of the overall imaging system is conducted. It is found that the non-negative least squares (NNLS), used so far in the proof-of-concept work, is not optimal in terms of achieving the highest resolution. This is due to the fact that NNLS cannot: 1) properly account for the dominant noise model in a typical biological fluorescence microscopy image, and 2) account for the prior information of object sparsity. A properly designed processing scheme based on the correct noise model, and that takes advantage of prior information, is shown to achieve improved super-resolution accuracy. It is first developed using numerical simulation, and then verified in experiment using a commercially-available calibration sample. A 60nm resolution is shown to be achieved, which is approximately four times beyond the resolution limit. Experimental results, obtained using this imaging technique with up-conversion nanoparticles (UCNPs) as the contrast agents, are also presented and compared with a ground-truth image obtained using transmission electron microscopy (TEM). It is shown that the super-resolved image are a close match to the ground-truth TEM image.

The second question is necessitated by the fact that the inverse problems that need to be solved to recover the super-resolved objects do not seem to have a unique solution, even in noiseless scenarios. This argument can be supported by either: 1) the fact that the microscope rejects some spatial frequencies completely, or 2) the fact that the matrices used in the recovery of super-resolved objects have non-empty null spaces. However, in the existing literature on computational super-resolution, it is generally an accepted conclusion that the solution *is* unique, at least in the noiseless case. This apparent contradiction hinders the wider adoption of this technique. In this thesis, it is shown that whether the solution is unique or not is not a clear-cut issue. Instead, it is dependent on the sparsity level of the true solution: 1) if the true solution is sparse, the solution is more likely to be unique, 2) if the true solution is somewhat sparse, the solution is non-unique, but the non-unique solutions do not interfere much with achievable super-resolution, and 3) if the true solution is dense, or non-sparse, the solution is non-unique, and no meaningful super-resolution can be realized. Two methods of determining if the solution is unique are proposed in this thesis, and

used to analyze the uniqueness of solution for a range of sparsity conditions.

Finally, the last question stems naturally from the fact that the images being processed are diffraction-limited. Because of this, conventional Fourier optics analysis suggests that no information beyond the diffraction limit can be recovered. This is again in contradiction with a wide array of successful and verifiable computational super-resolution results, presented in existing literature, as well as in this thesis. It is shown in this thesis that while Fourier optics is a good model for *image formation* in a microscope, it is not a good model for *information transfer* in a microscope. The main inadequacy of Fourier optics is that it fails to consider the inherent sparsity that is a result of the chemically-specific labeling in fluorescence microscopy. As a result, Fourier optics is not entirely appropriate in analyzing the information transfer in a microscope. It is shown that, if the object exhibits sparsity, *information* about sample fine details (that are beyond the diffraction limit) is still transmitted via the diffraction-limited images, and that this information can be used by a properly designed processing scheme to reconstruct the super-resolved image. Based on existing research in this topic, a singular value decomposition (SVD) based analysis, which can be shown to be a generalization of Fourier optics analysis, is revisited. By incorporating sparsity in this SVD-based analysis, a more general tool can be obtained that provides a more complete description of the information transfer in a microscope.

Dedication

To my fiancée, Simeng, for her unconditional support and encouragement throughout my study, for which I am eternally grateful.

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my advisor, Prof. Carol Cogswell. From taking a chance with me when I had no optics background, to helping me navigate the chaotic daily life of a PhD student, I am truly grateful for her encouragement and guidance on a daily basis.

I would also like to offer my special thanks to Prof. Stephen Becker, for his comments, guidance, and technical assistance at every step of this thesis. Without his vast expertise on numerical optimization, this thesis would have been completely different.

I am deeply grateful for the faculty members of the Department of Electrical, Computer, and Energy Engineering (ECEE) at University of Colorado Boulder. It is their deep technical knowledge and plentiful experience that made the past few years of my life a never-ending learning opportunity.

Finally, I would like to thank the graduate student community of ECEE for their camaraderie and support.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction and overview of super-resolution imaging

By enabling the observation of small biological structures, fluorescence microscopes have become an essential tool for biological research. The spatial resolution of optical instruments (e.g. microscopes) is limited by the diffraction of light. This *diffraction limit* is often quantified by the Rayleigh resolution distance, d_R [1, 2]. This distance is the minimum separation between two incoherent point sources for them to be "barely resolved" by the microscope, and is commonly estimated using $d_{\mathrm{R}} = 0.61\lambda/\mathrm{NA}$, where λ is the light wavelength and NA is the numerical aperture of the microscope objective. As a result, for high-NA (e.g., NA = 1.4) fluorescence microscopy conducted in the visible spectrum, due to the diffraction limit, only details that are approx. 200nm or larger in size can be resolved.

While this resolution of ≈200nm is often sufficient for biological research, there are many even smaller structures. Therefore, a higher resolution beyond the diffraction limit would enable observation of these fine structures and unlock additional insights in biological studies. As a result, in the past few decades, a collection of *super-resolution* fluorescence microscopy techniques have been proposed, developed, and popularized. These methods bypass the diffraction limit using various mechanisms, and some representative examples of these techniques include stimulated emission depletion (STED) microscopy [3, 4], photoactivated localization microscopy [5] (PALM), stochastic optical reconstruction microscopy [6] (STORM), super-resolution optical fluctuation imaging [7] (SOFI), and structured illumination microscopy [8] (SIM). These techniques generally take advantage of specialized optical processes (such as stimulated emission or on-off state transitions of

fluorophores) to achieve resolution beyond the diffraction limit. As a result, specialized illumination setups or photo-switchable fluorophores are generally required in these existing super-resolution microscope systems. Here, a brief literature overview of these existing super-resolution fluorescence microscopy techniques is provided.

1.1 Existing super-resolution microscopy techniques

Existing super-resolution microscopy techniques achieve imaging beyond the diffraction limit in different ways. These existing methods can be roughly divided into three categories: 1) stimulated emission depletion based; 2) localization based; and 3) structured illumination based. These categories of techniques differ in the mechanisms for resolution enhancement, and a brief overview of each category is provided below.

1.1.1 Stimulated emission depletion microscopy

Stimulated emission depletion (STED) microscopy achieves super-resolution by effectively decreasing the size of the illumination spot [3, 4]. At the heart of this technique is the specialized point-scanning illumination setup used to induce the spatially-selective stimulated emission depletion. This illumination setup consists of two lasers, an excitation laser, and a STED laser. The excitation laser is first applied, which is closely followed by the STED laser. Instead of a conventional diffraction-limited spot, the STED laser is shaped in such a way that a region exists in the center where the STED laser spot has very low (ideally zero) intensity. Because stimulated emission rate has a non-linear dependence on the intensity of the STED beam, in the area excited by the excitation laser, the fluorophores around the center will be "turned-off", or quenched, by the STED laser. This leaves a central region, still emitting fluorescence, that is much narrower than if the STED beam is not applied. As a result of this effective reduction of illumination spot size, a super-resolved image can be acquired as the combined excitation-STED spot is scanned across the sample. It should be noted that while the original proof-of-concept and early experiments of STED microscopy were conducted using pulsed lasers, continuous wave (CW) lasers have been shown to

be compatible with this technique [9].

Selectively quenching the outer area of the focal spot can be achieved with methods other than STED. For example, a technique similar to STED microscopy has been proposed: ground state depletion, or GSD microscopy [10]. In GSD microscopy, an illumination laser of appropriate wavelength is used to excite the fluorescent molecules to a long-living, non-emitting dark state. As a result, the outer regions of the focal spot can be selectively quenched, thus achieving the same type of resolution enhancement as seen in STED microscopy. In fact, STED and GSD microscopy have been shown to be different implementations of REversible Saturable Optical Fluorescence Transitions, or RESOLFT microscopy [11, 12].

1.1.2 Single molecule localization microscopy

Single molecule localization microscopy (SMLM) is a term that encompasses a family of super-resolution microscopy techniques based on the concept of isolating the emitters (spatially and temporally) and fitting the location of the emitters to an accuracy that is beyond the diffraction limit. Normally (without isolating the emitters), the size of the point spread function (PSF)[1] limits the achievable resolution. If a high degree of separation between the emitters can be achieved, such that for each patch of light in the collected image there is a high probability that only a single fluorescent molecule gives rise to it, the exact location of that molecule can be estimated by a properly designed post-processing step. If this localization process is repeated for a sufficiently large number of molecules in the sample (by acquiring a large number of images and processing them), a super-resolved image can be recovered from *diffraction-limited* raw images.

Many implementations of SMLM exist for different applications. In fluorescence microscopy, two representative variants, both based on on-off state transitions of fluorophores, are stochastic optical reconstruction microscopy (STORM), and photo-activated localization microscopy (PALM). In these techniques, multiple images are acquired sequentially, each containing only a very small

[1] In this thesis, PSF is defined as the image formed by a microscope when it is used to image an infinitesimal point.

subset of activated fluorophores. As a result of this sparse activation, there is a high probability that for each observed patch of light, only a single fluorophore molecule gives rise to it. After the collection of one raw image, the activated fluorophore molecules are then switched off before a different set of fluorophores are activated. In PALM, this is achieved through photo-bleaching [5], and in STORM, it is achieved through reversible switching of fluorophores [6].

1.1.3 Structured illumination microscopy

Structured illumination microscopy, or SIM, is another commonly used super-resolution microscopy technique [8]. SIM enhances resolution by producing a sinusoidal illumination pattern on the sample plane. This sinusoidal pattern effectively shifts the spatial frequencies that are initially beyond the diffraction limit into a position that is within the cut-off frequency of the optical microscope. The highest achievable cut-off is the sum of the unmodified optical system and the highest possible sinusoidal pattern frequency, both on the scale of $1/d_{\mathrm{R}}$. This means that SIM can achieve, at best, a two-fold resolution improvement. Because of this fundamental limit to achievable resolution improvement, SIM is only mentioned here, while the rest of this thesis will focus on techniques that can achieve *more than two-fold* resolution improvement. However, it should be noted that when combined with the use of on-off state transitions of flourophores, more than two-fold resolution improvement has been demonstrated [13].

1.2 An alternative approach: computationally achieving super-resolution

Besides these existing techniques, another avenue of achieving super-resolution has also been demonstrated: that of achieving super-resolution with numerical processing, *without* specialized illumination setups or on-off state transitions of fluorophores. This method of super-resolution fluorescence microscopy has received comparatively less research effort in the past. Because of this, a discussion on the motivation and the advantages of this super-resolution approach is first provided before its development throughout history is summarized.

1.2.1 Motivation and benefit of achieving super-resolution computationally

As mentioned before, to achieve more than two-fold resolution improvement (which excludes SIM in this discussion), existing super-resolution microscopy techniques generally require either specialized illumination setup (STED) or on-off state transitions of the fluorophores (SMLM). For STED, two lasers of different wavelengths need to be carefully aligned such that their respective focused spots can coincide spatially to accurately induce stimulated emission where it is desired. Besides the alignment requirement, an additional complication in STED microscopy is the PSF engineering, and the additional complexity it introduces, to shape the STED beam into a ring-like configuration to preserve the fluorescence in the central part of the illuminated area. Finally, for STED microscopy in particular, due to the non-linear dependence of the rate of stimulated emission on incident intensity, to achieve a high level of depletion (and therefore, a high level of super-resolution), a high power STED beam is generally used, which can lead to potential photo-damage of the sample [10].

For SMLM, the need to selectively switch on and off a sparse subset of all fluorophores limits the microscopist to solely use these switchable fluorophores in their sample preparation. This can hinder its use on a larger scale, where access to these switchable fluorophores could be limited or impractical. Furthermore, because in each SMLM raw image only a very sparse subset of fluorophores is emitting light, a large number of raw images need to be collected in order to activate enough fluorophores to yield a useful super-resolved image to the end user. Combined with the stochastic nature of the sparse activation, this leads to the long acquisition time of SMLM, which hinders its application in studying dynamic processes, or real-time fluorescence imaging.

As a result of these shortcomings of existing super-resolution methods, there is a need for a microscopy technique that can: 1) super-resolve more than two times beyond the diffraction limit to enable the study of small cellular structures; 2) be deployed using relatively simple, minimally modified optical setups; and 3) work with a large array of commonly used, *non-switchable* fluorophores. Identifying a purely *optical* process that simultaneously fulfills these requirements

is rather challenging. Instead, a more promising approach is to first acquire raw images using a minimally modified optical setup and non-switchable fluorophores, and later attempt to recover the image details that are beyond the diffraction limit. This *computational super-resolution* approach thus has the potential of becoming a powerful tool for those who wish to conduct super-resolution fluorescence microscopy using standard fluorophore sample labeling and a simple optical setup.

1.2.2 Computationally achieving super-resolution: a historic overview

Given its simplistic nature, it is not surprising that researchers throughout the past few decades have independently attempted to realize the idea of applying numerical post-processing to achieve super-resolution. The results from these early attempts can be summarized as follows: 1) computational super-resolution is real; but 2) computational super-resolution is only achievable under some strict conditions.

In one of the early works discussing this computational super-resolution phenomenon [14], Donoho cited examples of applying maximum entropy inversion to recover closely spaced spectral peaks that are blurred together in a nuclear magnetic resonance (NMR) spectrum. In this example, spectral peaks that are spaced too close together to be resolved by the NMR spectrometer can be later recovered computationally (and verified against the known ground-truth result) using an appropriately designed numerical processing scheme. In this 1992 paper, it was discovered that this super-resolution phenomenon is "real" but "delicate". In other words, while the super-resolution can be achieved with computation to a satisfactory and useful degree, the conditions under which it is achieved is rather restrictive. Specifically, Donoho concluded that non-negativity and sparsity (termed "near-blackness" in [14]) are two key properties an object needs to possess so that a properly designed processing scheme can successfully recover a super-resolved image from diffraction-limited noisy data. Note that these properties are also necessary for SMLM to successfully super-resolve an object, where non-negativity is fulfilled by the incoherent contrast mechanism, and sparsity is fulfilled by the switchable fluorophores used to label the sample.

In another early theoretical result by Sementilli et al. (1993) [15], it is shown that for a non-

negative, space-limited object, the achievable super-resolution is inversely related to the size of the object (i.e., positively related to the sparsity of the object) and the noise level. At high enough signal level (i.e., noiseless), exact recovery to any user-defined level of precision can be expected.

In another independent series of works, Bertero et al. (1982-89) applied singular value decomposition (SVD) to analyze this super-resolution phenomenon in detail. Scenarios such as coherent imaging [16], incoherent imaging [17], and imaging with non-uniform illumination [18] were investigated. Again, it was found that achievable super-resolution is related to the sparsity of the object in question and the signal to noise ratio (SNR). The above five citations represent a selection from a large body of work, spread out through several decades, centered about the common theme that achievable super-resolution is greatly affected by sparsity and SNR. As a result, a discussion of a "resolution limit" should incorporate consideration for these factors as well.

Given these early theoretical papers, one could expect that experimental application of computational super-resolution microscopy would have followed. However, as is well-known, the world of super-resolution microscopy has been dominated by the techniques summarized in the previous section, and the use of computation to achieve super-resolution has been largely reserved to limited fields of study, such as astronomy (see [19] and references therein). Here, and in the larger astronomy community in general, it is frequently shown that super-resolution is achievable computationally only if the underlying object is point-like (i.e., exhibits sparsity). If the object does not possess sparsity, all that numerical processing can achieve is to enhance the contrast of the image, without any meaningful enhancement of resolution.

At this point, it can be concluded that while computationally achieving super-resolution is feasible, it requires some rather restrictive conditions to be met, chief among which are non-negativity and sparsity. In astronomy, distant stars can be considered as point-like objects with a high degree of sparsity, which means that these conditions are easily met. In biological fluorescence microscopy, while the non-negativity condition is easily met with the incoherent contrast mechanism, the sparsity condition is much harder to meet, with the exception of the fluorophore manipulation used in SMLM. This is caused by the diverse nature of the samples observed in

fluorescence microscopy. Since users of the microscopes cannot control what type of samples are imaged, this poses a significant challenge. One method to artificially enhance sparsity (and achieve computational super-resolution) is to use a point scanning illumination setup with a standard widefield microscope. In [18], it is shown that such a microscope system, without considering object sparsity, can achieve at least a two-fold resolution improvement.[2] To image an area larger than the diffraction-limited spot, this illumination spot can be scanned laterally across the sample. At each scanning position (step), one can collect a small sub-image, then numerically process it, and finally reassemble all of the processed sub-images according to the scanning pattern to yield a final, super-resolved image. However, this proposed super-resolution microscope system is very difficult to use for practical imaging purposes due to the intensive computational resources needed to perform the numerical processing. As an example, to acquire an image with a modest field of view (FOV) of 200-by-200 pixels, a total of 40,000 sub-images are needed. While the acquisition of these sub-images is not difficult, the processing is where the practical challenge lies. If a single sub-image, which requires the solving of one inverse problem, takes 1s to process, the entire data set would take approximately half a day to process, which is too slow for practical applications.

1.3 Current state of art and barriers to wider use of computational super-resolution

In recent years, thanks to advances in GPU computing and large-scale numerical optimization techniques, this practical challenge is beginning to be overcome, with demonstrated experimental success by other members of this thesis author's research group [21]. These recent results provide further justification of why this avenue of super-resolution microscopy should receive renewed attention. Using a modified widefield microscope (with a point-scanning illumination setup [21]) and a large-scale numerical optimization routine [22], this work demonstrated that super-resolution can be achieved with illumination enhanced sparsity and numerical post-processing without photoswitchable fluorophores. A resolution improvement of three times was experimentally demonstrated

[2] This proposed system is equivalent to image scanning microscopy (ISM) [20].

and verified using a calibration sample.

Given the early theoretical results and the more recent experimental demonstration of computationally achieving super-resolution in fluorescence microscopy, it can be argued that this technique is now ready for wider adoption by the biological research community. However, there are three major questions that need to be answered before this can become reality: 1) what is the most suitable (i.e., the highest resolution) processing scheme for this super-resolution technique? 2) is the recovered image the only possible solution (i.e., is the solution unique)? and 3) given these promising super-resolution results, what conclusion can be drawn in the analysis of the *information transfer* ability of a microscope? This thesis aims to answer these three problems.

1.3.1 Problem of the processing scheme

So far, most works in this computational super-resolution technique has been focused on the simple fact that super-resolution is achieved and verified, and how and why the selection of the processing scheme affects the achievable resolution has been given comparatively less attention. In [14], maximum entropy inversion is used to recover the super-resolved details, and no alternatives are explored. In [16, 17, 18], a truncated-SVD based inversion scheme is used to recover super-resolved details in simulation. This processing scheme requires the accurate knowledge of the spatial extent of the object, meaning it cannot plausibly be used in a real-world experiment, where such knowledge is inaccessible. In [21], a large-scale non-negative least squares (NNLS) optimization routine is used to obtain proof-of-concept verification of this super-resolution approach. However, in order to fully understand how and why processing schemes affect the achievable super-resolution, a much more thorough study is needed.

To answer this problem, in this thesis work, a careful examination of the overall microscope imaging system is performed. It is found that the NNLS processing scheme used in the above proof-of-concept work [21] does not achieve the highest resolution. This is caused by two factors: 1) NNLS assumes a different noise model than what is the dominant noise source commonly found in a typical fluorescence microscopy image; 2) NNLS does not take advantage of the prior information

of sparsity and therefore cannot fully recover all the super-resolution information. A properly designed processing scheme that utilizes the dominant noise model and prior information can recover more information than NNLS and achieve higher super-resolution. Based on this, an improved processing scheme is proposed and is shown to be able to achieve approximately four times resolution improvement in experiment using a calibration sample [23]. This processing scheme is then used to process images obtained using a different contrast mechanism: up-conversion nanoparticles (UCNPs). The processed super-resolved image is verified against a ground-truth image obtained using transmission electron microscopy (TEM) and is shown to be accurate to within the size of the particles themselves.

1.3.2 Problem of unique solutions

In the published works on computational super-resolution, it is generally accepted that the solution to the computational problem is unique. However, because the matrices used to characterize the microscope systems do not have an empty null space, the solution can be shown to be non-unique. Alternatively, this can be seen from the fact that the microscope will reject spatial frequencies beyond its cut-off frequency, f_c. As a result, an apparent contradiction exists which hinders the wider spread of this technique.

This thesis reconciles this apparent contradiction. It is shown that whether the solution is unique is not a clear-cut issue, and instead depends on the sparsity level of the object. The sparser the object, the higher the probability the solution is unique. Due to this dependence, the uniqueness of the solution should be considered for each imaged object. In this thesis, two methods of *certifying* the uniqueness of the solution are proposed. Using the method that is better suited for a practical-sized problem, the uniqueness of a sparse, two-dot object and a less-sparse, two-line object are analyzed.

1.3.3 Problem of information transfer

The familiar Fourier optics analysis shows that the microscope system used to implement computational super-resolution should only be able to achieve diffraction-limited resolution. This is because the collected images are themselves diffraction limited, and can not transmit finer details. This is in contradiction with the verifiable super-resolution results shown in this thesis and in published papers. In this thesis, it is shown that while the diffraction-limited images do not contain details finer than the diffraction limit, in the case where the object being imaged exhibits sparsity, they do contain *information* about these fine details. This information can later be used by a properly designed processing scheme to reconstruct a super-resolved image. It is also concluded that because Fourier optics analysis cannot incorporate the sparsity in the object into its consideration, it does not provide a complete description of information transfer in this scenario. Building upon published work, a SVD-based analysis framework is proposed in this thesis, and is shown to be able to account for the sparsity in the object and provide a more complete description of information transfer in a microscope.

Chapter 2

Computationally achieving super-resolution: current technique

In this chapter, a through description of computational super-resolution microscopy is given. First, the experimental setup used to realize computational super-resolution microscopy is presented, and experimental procedures described. Next, the image formation and the noise model are presented. Finally, recovery of the super-resolved images is described.

2.1 Experimental setup and procedure

The experimental realization of computational super-resolution is achieved using a standard widefield fluorescence microscope with a custom-implemented point-scanning illumination system. Instead of the conventional illuminator, a laser beam is spatial-filtered and expanded to cover the entire back aperture of the microscope objective. This ensures the focused illumination spot is as small as possible to maximize the degree to which the sparsity is enhanced. This focused illumination spot is scanned across the sample via a pair of galvanometric mirror scanners, positioned at a plane conjugate to the objective back focal plane using a 4-f relay. As the mirrors rotate, the focused illumination spot moves laterally across the sample, thus achieving point illumination scanning. The fluorescence image is then recorded on a scientific complementary metal-oxide-semiconductor (sCMOS) camera and stored for later processing. This imaging system is shown in Figure 2.1.

In an actual experiment, the point-scanning illumination system generates a focused illumination spot and scans it (one step at a time) across the sample laterally, completing just one pass

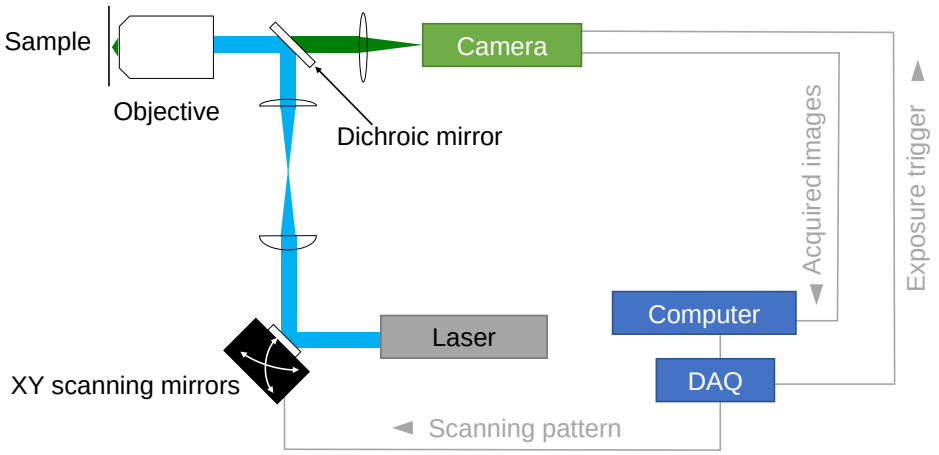


Figure 2.1: The experimental imaging setup used throughout this thesis. This microscope system is based on a conventional widefield microscope, but with its illumination system replaced by a point scanning setup. The computer generates the control signal for the scanning of the focused illumination spot and the triggering of the camera. The acquired images are stored in the computer for later processing.

in x and y. At each of these scanning steps, a small image is collected from the sCMOS camera. These small raw images are called *sub-images* in this thesis. After the acquisition is completed, these sub-images are processed using a numerical routine, called a *processing scheme* in this thesis. Each sub-image, after processing, represents a small portion of the final super-resolved image. Because the scanning pattern is pre-determined, these small parts can be reassembled to yield a final, super-resolved image covering the entire scanned field of view.

2.2 Image formation and noise model

The recovery of the super-resolved image is done by solving a series of inverse problems, one for each scanning location. As a result, an accurate image formation and noise model are needed in order for the solution to reflect physical reality. Here, these models are discussed in detail.

2.2.1 Image formation model

For a microscope described previously, assuming such a microscope system has a 1:1 magnification (without the loss of generality), the noiseless image generated by it can be formulated as:

$$I_{\text{noiseless}} = I_{\text{PSF}} \otimes [I_{\text{ill}} \cdot I_{\text{obj}}], \tag{2.1}$$

where I_{PSF} is the point spread function (PSF) of the microscope system, $\otimes$ the convolution operator, I_{ill} the illumination profile (here a diffraction-limited focused spot), $\cdot$ the element-wise multiplication operation, and I_{obj} the object being imaged.

Physically, Equation (2.1) describes a scenario where a biological sample is selectively labeled with fluorescence markers. This is described by I_{obj}. This sample is then excited with a focused spot (I_{ill}) generated by the custom illumination setup. The spot illumination will result in an illuminated object defined by $I_{\text{ill}} \cdot I_{\text{obj}}$, which, depending on how densely labeled the sample is, will be spatially limited to a maximum of the size of the illumination spot.[1] This illuminated object is then imaged by the microscope characterized by I_{PSF}. The imaging process is modeled by the convolution between the illuminated object ($I_{\text{ill}} \cdot I_{\text{obj}}$) and the PSF I_{PSF}.

Equation (2.1) is defined in the continuous domain. In order to implement Equation (2.1) numerically, it is re-formulated as:

$$I_{\text{noiseless}} = H_{\text{PSF}} \times [I_{\text{ill}} \cdot I_{\text{obj}}], \tag{2.2}$$

where $I_{\text{noiseless}}$, H_{PSF}, I_{ill}, and I_{obj} are discretized and vectorized versions of $I_{\text{noiseless}}$, I_{PSF}, I_{ill}, and I_{obj} in Equation (2.1) respectively. In this formulation, H_{PSF} is a matrix whose j-th column stores the (discretized and vectorized) PSF centered at location s_j (here, s_j indicates a point (x_j,y_j) on the two-dimensional image plane), and $\times$ represents a matrix multiplication. In this thesis, H_{PSF} is referred to as the *dictionary*.

The dictionary H_{PSF} warrants further discussion here. In many existing works, the microscope imaging system is described using a single, square PSF image. In the context of this

[1] In this case of maximum size of illuminated object, the object is labeled everywhere.

technique, H_{PSF} does not need to be a square matrix, and is in general, not a square matrix. This is because H_{PSF} does not store the image of a single PSF, but instead a number of shifted PSFs, each corresponding to a spatial location. These PSFs are vectorized and stored in the columns of H_{PSF}. This way of modeling the imaging system (as opposed to using a single PSF) allows different discretization schemes to be used. If s_j has a fine discretization that is smaller than the camera pixel size, features with sizes smaller than a camera pixel can then be modeled using this discretization scheme [21]. For example, if d_{camera} represents the size of a camera pixel, and d_{dict} represents the distance between two adjacent PSFs stored in the columns of H_{PSF}, then by specifying d_{dict} to be smaller than d_{camera}, features smaller than d_{camera} can be modeled by the smaller d_{dict}. More importantly, the images formed by the microscope when imaging these fine features can also be modeled, and these fine features later recovered using a properly designed processing scheme. To achieve this, a series of PSFs, one for each particular sub-pixel position of the point source, is calculated. Then these sub-camera-pixel PSFs are shifted to form the entire H_{PSF} matrix. For example, if $d_{\text{dict}} = d_{\text{camera}}/4$, a series of PSFs, each representing one of a total of $4^2 = 16$ sub-camera-pixel positions, is calculated first. These sub-camera-pixel PSFs are then shifted by some integer number of pixels to form the complete dictionary H_{PSF}. This method is used throughout this thesis to generate H_{PSF}.

2.2.2 Noise model

In this thesis, a sCMOS camera is used as the detector. Such a camera has a quantum efficiency (QE) between 0 and 1, and will corrupt the collected image with a uniform additive Gaussian noise with a variance of σ^2. In addition, the noisy images will also be corrupted by photonic shot noise. These sources of degradation yield the following model for the acquired noisy image:

$$I_{\text{noisy}} = \text{Poiss}\left(I_{\text{noiseless}} \cdot \text{QE}\right) + \mathcal{N}(0, \sigma^2). \tag{2.3}$$

Here, I_{noisy} is the vector containing the intensity values of the noisy pixels collected from the camera, Poiss is an operation that generates Poisson random numbers with the mean set to its input, and

QE is the camera's quantum efficiency. $\mathcal{N}(0, \sigma^2)$ returns a vector of the same size as I_{noisy}, and its elements are Gaussian random numbers with zero mean and variance of σ^2. The photon shot noise is modeled by Poiss, and readout noise is modeled by $\mathcal{N}(0, \sigma^2)$.

2.3 Recovery of the super-resolved image

The recovery of the super-resolved images is accomplished by a processing scheme that numerically solves a series of inverse problems, one for each sub-image. As an example, the NNLS numerical optimization problem, used in [21], can recover the super-resolved object. The optimization problem is:

$$x_{\text{NNLS},m} = \underset{x}{\text{argmin}} \ \|H_{\text{PSF}} \times x - I_{\text{noisy},m}\|_2^2, \text{ subject to } x \geq 0, \tag{2.4}$$

where $x_{\text{NNLS},m}$ is the super-resolved object indexed by m, and m refers to the m-th raw image in the scanning sequence. Since the sub-images are independent from one another, this index m is left out in this thesis. The inequality is element-wise, i.e., the solution vector x_{NNLS} should contain only non-negative elements.

It should be noted NNLS is not the only processing scheme possible. NNLS makes specific assumptions about the dominant noise source in the image that may or may not be accurate. As a result, different processing schemes may recover additional information than NNLS. This is explored in greater detail in the next chapter.

Chapter 3

Properly designed processing scheme achieves the highest resolution

Having established the imaging system setup, the image formation and the noise model, and the recovery of the super-resolved object in Chapter 2, this chapter will go into greater detail regarding the factors that need to be considered when properly designing a processing scheme that achieves the highest resolution. Factors such as the dominant noise model and prior information will be investigated. It is found that the NNLS approach used in the proof-of-concept results shown in [21] cannot achieve the highest resolution. In this chapter, after accounting for the dominant noise model and prior information, an improved processing scheme is identified, implemented, and used to process noisy data in both simulation and in experiment. As a demonstration, in section 3.5, an improved resolution of 60nm (approximately four times beyond the diffraction limit) is shown to be achieved and verified experimentally using a commercially available calibration sample.

3.1 NNLS is unlikely to achieve the highest resolution

So far, NNLS has been the processing scheme used to recover the super-resolved image. Despite its simplistic nature and wide-spread use in different fields of scientific study, NNLS is unlikely to yield the highest resolution for this particular application. This is because NNLS does not take advantage of two powerful factors in biological fluorescence microscopy.

The first factor is that, for high-sensitivity fluorescence microscope systems, due to its low Gaussian readout noise level, photonic shot noise is usually the dominant noise source in a typical image. Photonic shot noise has a variance equal to its mean, indicating that for a typical image with

dark and bright pixels, these pixels will have different means and therefore, variances. However, NNLS is derived from maximum likelihood estimation (MLE) for additive, uniform and zero-mean Gaussian noise, which is not in agreement with the shot noise model.

The second factor is that NNLS does not incorporate any prior information except the solution vector must be positive. If a processing scheme incorporates additional prior information (should it be available), one would expect it to achieve higher resolution. In biological fluorescence microscopy, besides non-negativity, sparsity is a frequently occurring piece of prior information. This is because fluorescent samples typically consist of sub-micron structures labeled with fluorescent markers against a non-labeled, dark background. It is widely known that a suitable regularization term can take advantage of this sparsity [24, 25].

As a result of these two shortcomings, it is unlikely that NNLS would achieve the highest resolution. To show this, alternative processing schemes, based on the MLE for photonic shot noise and accounting for the prior information of sparsity, are proposed. These processing schemes are then used in simulation to attempt to super-resolve an incoherent two-dot test object to determine which achieves the highest resolution.

3.2 Simulation methodology

To investigate how the achievable resolution is affected by factors such as dominant noise source and sparsity, a simulation study is performed. The reason for this is that numerical simulation grants access to the ground-truth object, which enables fair and accurate comparisons between processing schemes.

3.2.1 Design of numerical experiment

In this numerical experiment, based on an examination of the imaging system, alternative processing schemes are proposed and their performances compared. To to this, the following numerical experiment procedures are proposed: 1) define a test object consisting of two incoherent dots separated by a distance smaller than the Rayleigh resolution distance; 2) simulate the image

formed by the microscope that is collected by the sCMOS camera, including blur (by the diffraction-limited microscope) and noise (by the noisy sCMOS camera); 3) recover the super-resolved image of the test object using the processing scheme under consideration; 4) evaluate processing scheme performance by calculating the probability for a processing scheme to successfully resolve two dots in a single trial. This numerical experiment procedure is depicted in Figure 3.1.

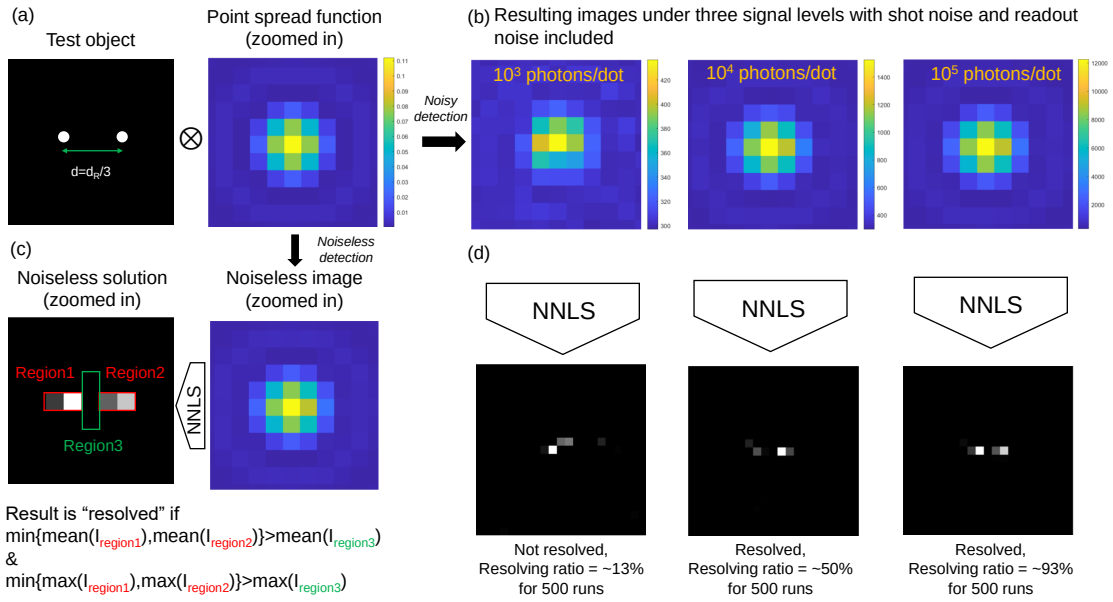


Figure 3.1: Design of the numerical experiment. (a) The test object and the point spread function (PSF). The test object consists of two incoherent point sources separated by a small distance d, which is approx. 1/3 of the Rayleigh resolution distance, d_{R}. PSFs are normalized such that the sum of all pixel values is 1. (b) Acquired noisy images under various signal levels. Images shown here are representative examples out of 500 runs for each signal level. (c) Criterion for determining if a processed image is "resolved" or "not resolved". By comparing the relative intensities between the three regions shown, a binary "resolved"/"not resolved" determination can be made. (d) Examples of successful and unsuccessful reconstructions and resolving ratios for three levels of the signal strength for 500 runs. NNLS was used as the processing scheme in this figure. All other processing schemes considered in this paper use the same procedures.

3.2.2 Test object and evaluation of processing scheme performance

The test object used in this chapter consists of two incoherent dots separated by a small distance. In this case the small distance is approx 1/3 of the Rayleigh resolution distance d_{R}. d_{camera} is selected such that the PSF spans approx. 3 pixels (see Figure 3.1). In this experiment, the

images are generated with a dictionary having $d_{\text{dict}} = d_{\text{camera}}/16$, and processed with a dictionary having $d_{\text{dict}} = d_{\text{camera}}/4$, respectively. These d_{dict} values are selected in a way such that 1) the small, sub-diffraction-limit distance of approximately $d_{\text{R}}/3$ can be accurately represented; and 2) the individual incoherent dots do not fall exactly at the center of a pixel in the processed image, which is the most common case in a real-world experiment.

To evaluate the performance of a given processing scheme, a signal level (in terms of number of photons per dot) is first specified, and the noisy image is simulated. Then, the processing scheme under consideration is used to recover a super-resolved image of the test object. Finally, a binary "resolved"/"not resolved" metric is used to evaluate the processed image. This metric is based on the relative intensities of pixels in different regions in the processed image, and has been used before [21]. These steps are then repeated 500 times, after which a "resolving ratio" metric is generated for a processing scheme and the particular signal level. This resolving ratio metric is defined as:

$$\text{resolving ratio} = \frac{\text{number of "resolved" runs}}{500}. \tag{3.1}$$

After this metric is evaluated for a signal level, the signal level is then changed and this entire process repeated. In this technique, the focused illumination spot is scanned across the sample *once*, and this resolving ratio metric is suitable for evaluating and comparing the performances of different processing schemes in this scenario. An optimal processing scheme will be the one that gives the highest resolving ratio using the least signal.

3.3 Simulation results and alternative processing schemes

Here, the properties of the imaging system, formed by the microscope and the sCMOS camera, are carefully examined. Based on the results of this examination, alternative processing schemes are proposed and their performances evaluated.

3.3.1 Noise model and dominant noise source

To establish a baseline performance of the resolving power of a processing scheme, in Figure 3.2, the performance of NNLS is plotted. It can be clearly seen that the resolving ratio for NNLS increases with increasing signal level, and eventually approaches 1. This result is expected because, as signal level increases, the signal-to-noise ratio (SNR) also increases. At very high signal levels, the noisy images will be approximately noiseless, which has been shown to allow for exact recovery of object details to any user-defined precision [15].

In NNLS, the following numerical optimization problem is solved:

$$x_{\text{NNLS}} = \underset{x}{\operatorname{argmin}} \, \|H_{\text{PSF}} \times x - I_{\text{noisy}}\|_2^2, \text{ subject to } x \geq 0. \tag{3.2}$$

An intuitive understanding for Equation (3.2) is that the total discrepancy energy between the acquired image (I_{noisy}) and the reconstructed image ($H_{\text{PSF}} \times x$) is minimized. Its solution is an object that when blurred by the microscope, creates an image (i.e., $H_{\text{PSF}} \times x$), that closely resembles the acquired image. However, this straightforward approach may not be suitable for the dominant noise source in the acquired image.

From the perspective of maximum likelihood estimations, NNLS is derived from the log-likelihood function for noise modeled by additive, zero-mean, and independent and identically distributed (i.i.d.) Gaussian random variables. A noise source present in modern sCMOS cameras does fit this model — the Gaussian readout noise [26]. However, the photonic shot noise that is inherent to all optical signals is decidedly Poissonian. For a Poisson random variable, its variance is equal to its mean. This means that, as signal level (i.e., mean) increases, shot noise power (i.e., variance) increases as well. For a sCMOS camera with fixed Gaussian readout noise, as signal level increases, at some point, shot noise will become the dominant noise in the acquired noisy image. For realistic levels of signal acquired with modern, low-readout-noise sCMOS cameras, shot noise almost always dominates. In this case, NNLS can no longer maximize the likelihood function because of this mismatch between the Gaussian readout noise assumed by NNLS and the Poissonian shot noise that dominates the acquired noisy image.

To demonstrate the effect this mismatch has on the performance of the processing scheme, an alternative processing scheme that is appropriate for the noise model is proposed: weighted non-negative least squares (WNNLS). This is a valid approximation for the MLE for shot noise because, for a Poission random variable with sufficiently large mean (which is equal to its variance), a Gaussian random variable with the same mean and variance can be used to accurately approximate the original Poission random variable. Under this approximation, the acquired noisy image is approximated by a series of Gaussian random variables, one for each pixel, with their individual (unique) means and variances. A maximum likelihood estimator can be obtained for such a random process with weighted least squares with weights set to the true variances [27]. In this numerical experiment, the true variances are accessible because they are simply the noiseless image (variances for shot noise) plus the variance for the Gaussian readout noise, both available as parameters for the simulation. To apply MLE for shot noise to real-world experimental data, where the true variances are not accessible, a different implementation is introduced in a later section.

For WNNLS the numerical optimization problem solved is:

$$x_{\text{WNNLS}} = \underset{x}{\operatorname{argmin}} \left\| W \times (H_{\text{PSF}} \times x - I_{\text{noisy}}) \right\|_2^2, \text{ subject to } x \geq 0, \tag{3.3}$$

where W is a diagonal matrix storing the weights for each pixel. The diagonal elements of W are calculated by:

$$W_{\text{diag}} = (I_{\text{noiseless}} + \sigma^2)^{\frac{-1}{2}}, \tag{3.4}$$

where σ^2 is the variance for the Gaussian readout noise of the simulated sCMOS camera.

WNNLS is used to super-resolve the two-dot test object, and its performance is plotted in Figure 3.2 along with that of NNLS. From the plot, the following conclusions can be drawn: 1) both NNLS and WNNLS show improved performance (as measured by the resolving ratio) as signal level increases, as expected; and 2) NNLS and WNNLS perform very similarly at extremely low signal levels, but as signal level increases, WNNLS shows a decidedly higher performance over NNLS.

This behavior is expected because, when the optical signal level is low, shot noise has a low variance and does not contribute much to the noise power. Under this signal regime, Gaussian

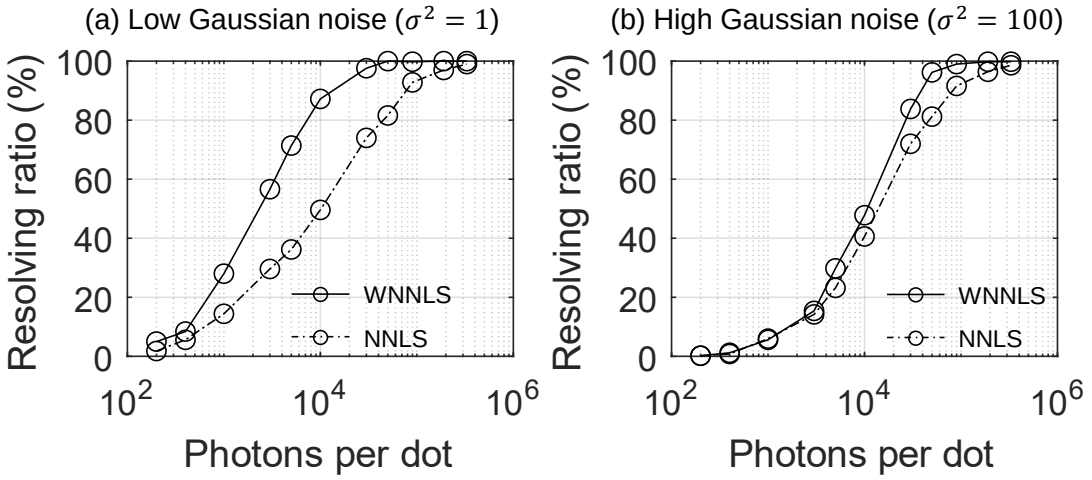


Figure 3.2: Performance of NNLS and WNNLS under different levels of Gaussian readout noise. (a) shows the performance with a QE of 0.6, and σ^2 of 1. (b) shows the performance with a QE of 0.6, and σ^2 of 100. (a) and (b) show similar trends: WNNLS performs roughly the same as NNLS before yielding better performance at some higher signal level. Note that in (b), with higher noise power (i.e., lower SNR), WNNLS gives a better resolving ratio than NNLS at a higher signal level.

readout noise contributes much of the noise power, meaning NNLS is the maximum likelihood estimator that matches the physical reality. However, WNNLS does not perform worse than NNLS under this signal regime. This is also expected because when signal level is low, $I_{\text{noiseless}} \approx 0$, and therefore $W_{\text{diag}} \approx \sigma^{-1} I$ (where I is an identity matrix), making Equation (3.3) nearly identical to Equation (3.2). As signal level increases, shot noise variance increases with it and eventually, shot noise becomes the dominant source of noise in the acquired image. Under this signal regime, WNNLS is the more appropriate processing scheme, and achieves higher resolving ratio compared with NNLS for the same signal level. To further illustrate this effect, in Figure 3.2, the same numerical experiment is repeated and performances of NNLS and WNNLS plotted, but with a higher (100-fold) Gaussian readout noise variance. The following conclusions can be drawn from this modified numerical experiment: 1) optical signal level being the same, the increased Gaussian readout noise level degrades the resolving ratios of both NNLS and WNNLS; 2) the performance of WNNLS stays roughly the same with NNLS for a longer interval at the lower end of the signal range, before yielding improved performance at a higher optical signal level than that in Equation

(3.2). This is because, now that the Gaussian readout noise is higher, shot noise needs to have an even higher variance (i.e., signal) to become the dominant noise source in the image.

From this numerical experiment, it can be concluded that performance improvement can be expected if the processing scheme used assumes the correct dominant noise model. In this particular instance, since shot noise is the dominant noise source, a processing scheme that generates the maximum likelihood estimator for shot noise gives improved performance.

3.3.2 Prior information and sparsity

Inverse problems such as Equation (3.2) and Equation (3.3) are well-known to be very ill-conditioned [14, 16, 28]. A direct effect of this is that even when the underlying object that is giving rise to the acquired noisy images remains unchanged (as is the case in Figure 3.2), the processed image will be different for successive trials. This is the reason why the resolving ratio is not 1 for the majority of the signal levels considered in this figure.

Prior information has been shown to be an effective method to alleviate the effect of this ill-conditioned inverse problem and improve the quality of the solution [14, 15, 24, 25, 29]. A typical way to incorporate prior information in the processing step is to first gather knowledge about the solution vector to the inverse problem, and then impose appropriate constraints or use a regularization term when solving it. In the context of computational super-resolution microscopy, constraints can often be derived from the physical properties of the imaging system. The non-negativity constraint, which is derived from the incoherent contrast mechanism, is used in NNLS and WNNLS, and has been shown to be an especially important piece of prior information [14, 15].

Sparsity is another piece of prior information that is crucial for super-resolution to occur. For example, it has been concluded that sparsity (termed near-blackness in [14]) is a key reason for super-resolution to occur. In the super-resolution microscopy technique of this thesis, sparsity comes from two sources: 1) from the fluorescently labeled structures in the sample; and 2) from the spatially constrained illumination profile generated by the focused illumination spot. In fluorescence microscopy, the sample often consists of structures that are much smaller than the size of the

illumination spot, and these structures are labeled with fluorescent markers against a non-labeled, dark background. When excited by the focused illumination spot, the illuminated region will be a mixture of light-emitting, labeled structures, and a non-emitting background. Because of this, the solution vector should contain many zeros, i.e., the solution vector is sparse with many zero elements. ℓ_1 regularization has been shown to be able to induce sparsity in the computed solution vector and improve its accuracy [30]. As a result, applying this sparsity-inducing ℓ_1 regularization term should further improve the performance of the processing schemes.

Based on this thinking, NNLS and WNNLS are each modified to incorporate the prior information of sparsity. This is done by simply adding an ℓ_1 regularization term to the original numerical problem. The modified processing schemes, r-NNLS and r-WNNLS (where r- indicates "regularized") are:

$$x_{\text{r-NNLS}} = \underset{x}{\operatorname{argmin}} \, \|H_{\text{PSF}} \times x - I_{\text{noisy}}\|_2^2 + \lambda \|x\|_1, \text{ subject to } x \geq 0, \tag{3.5}$$

and

$$x_{\text{r-WNNLS}} = \underset{x}{\operatorname{argmin}} \, \|W \times (H_{\text{PSF}} \times x - I_{\text{noisy}})\|_2^2 + \lambda \|x\|_1, \text{ subject to } x \geq 0. \tag{3.6}$$

In Equations (3.5) and (3.6), $\|x\|_1$ calculates the 1-norm of the vector x. λ is the regularization parameter, the selection of which is discussed in section 3.7.

In Figure 3.3, the performances of all four processing schemes considered so far (NNLS, WNNLS, r-NNLS and r-WNNLS)) are plotted. The processing schemes are all used to recover the two-dot test object from the same noisy image. From the plot, it can be concluded that, due to the existence of sparsity as the prior information, the addition of a sparsity-inducing ℓ_1 term improves the performances of both NNLS and WNNLS. Here, r-WNNLS exhibits the highest resolving ratio among all processing schemes for the signal levels considered in Figure 3.3. From the results of this numerical experiment, it can be concluded that improved performance can be expected if a processing scheme is appropriate for the dominant noise in the image, and utilizes the prior information of sparsity.

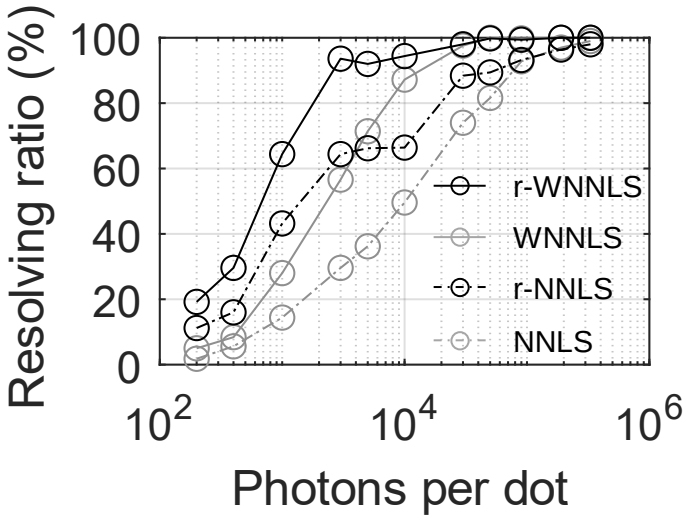


Figure 3.3: Performance of NNLS and WNNLS, as well as their regularized counterparts, with a QE of 0.6 and σ^2 of 1. WNNLS and NNLS curves are replicated from Fig. 3.2. The following observations are made: 1) for non-regularized processing schemes (NNLS and WNNLS), if the processing scheme under consideration is appropriate for the noise model (WNNLS), improved performance can be expected; 2) for regularized and non-regularized processing schemes (NNLS vs. r-NNLS, or WNNLS vs. r-WNNLS), the inclusion of the regularization term improves the performance of the processing scheme drastically; and 3) the best performance is achieved when the processing scheme is appropriate for the noise model, and includes a regularization term.

3.4 Applying optimized processing schemes to real-world data

From the results in section 3.3, it is obvious that r-WNNLS achieves the highest performance. However, its implementation requires the noiseless image (i.e., $I_{\text{noiseless}}$), which is inaccessible in a real-world experiment. In this section, the insight from section 3.3 is used to propose processing schemes that achieve improved performance and can be applied in a real-world experiment (i.e., without requiring the noiseless image).

3.4.1 Poisson maximum likelihood estimation

For shot noise dominated images, a processing scheme that is appropriate for the noise model, and does not require the noiseless image, is readily available. This processing scheme is derived from maximum likelihood estimation applied to Poisson random variables, which is the theoretical

model for shot noise. This processing scheme is referred to as Poisson-MLE, and is formulated as:

$$x_{\text{Poisson-MLE}} = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} [(H_{\text{PSF}} \times x)_i - I_{\text{noisy},i} \cdot \log(H_{\text{PSF}} \times x)_i], \text{ subject to } x \geq 0, \tag{3.7}$$

where i is the index pointing to the i-th pixel in the acquired or reconstructed image, and n is the total number of pixels in the acquired or reconstructed image. Note that Poisson-MLE does not require the noiseless image ($I_{\text{noiseless}}$) in its formulation.

In Figure 3.4, performances of Poisson-MLE and WNNLS, when used to process the same noisy images from section 3.3, are plotted. From this plot of resolving ratio, it can be seen that Poisson-MLE and WNNLS achieve nearly identical performance. This means that Poisson-MLE is a processing scheme that is suitable for the shot noise dominated images.

3.4.2 Variance stabilizing transform

The reason that NNLS delivers worse performance than WNNLS and Poisson-MLE is that NNLS assumes the variances of the pixels are uniform (i.e., the same regardless of signal level). As a result, if the shot noise dominated images can be transformed in a way such that the variances of the

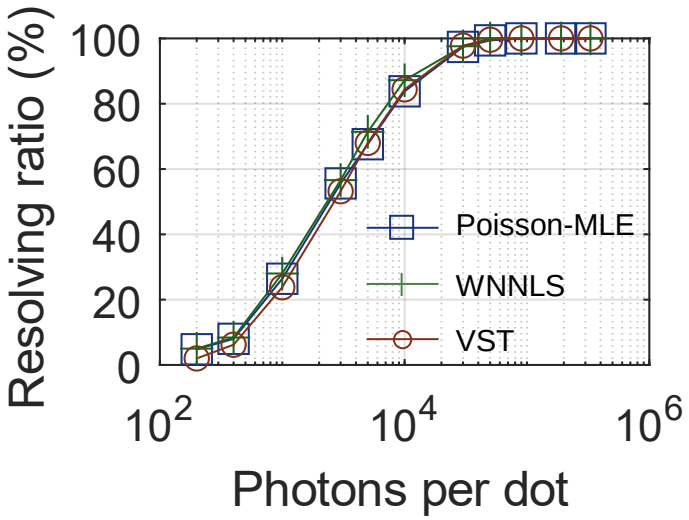


Figure 3.4: Plot of resolving ratios versus the signal level for WNNLS, Poisson-MLE, and VST. While Poisson-MLE and VST do not require the noiseless image, all three processing schemes yield roughly the same performance, and are hence all appropriate for the shot noise model.

pixels becomes uniform or approximately uniform, a NNLS-like processing scheme can then operate on the transformed images and should now be appropriate for the newly transformed images.

Such "variance stabilizing" transforms existing in prior research, one of them is the Anscombe transform [31]. Anscombe transform is described by:

$$z = 2\sqrt{p + 3/8}, \tag{3.8}$$

where p is the original Poisson random variable, and z is the transformed random variable, which has a uniform variance of 1 for p with a sufficiently large mean.

Variance stabilizing transforms, such as the Anscombe transform, operate on random variables whose variance is dependent on their mean (e.g., Poisson random variable), and generate "stabilized" random variables, with the new variance now *independent or nearly independent* of their mean. Because of this, one can use a NNLS-like processing scheme, operating on the newly transformed random variables, and expect improved performance over the original NNLS. This newly formulated processing scheme, termed VST for variance stabilizing transform, is formulated as:

$$\begin{aligned} z &= 2\sqrt{I_{\text{noisy}} + 3/8} \\ x_{\text{VST}} &= \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} [z_i - 2\sqrt{(H_{\text{PSF}} \times x + 3/8)_i}]^2, \quad \text{subject to } x \geq 0. \end{aligned} \tag{3.9}$$

It can be seen that NNLS and VST take on similar forms: both attempt to minimize the total squared error between the acquired image and reconstructed image. In NNLS, the original images are used. In VST, the Anscombe transformed images are used. The performance of VST is plotted in Figure 3.4, in addition to WNNLS and Poisson-MLE. It can be seen that despite its similarity with NNLS, VST achieves improved performance over NNLS and its performance is roughly the same as WNNLS and Poisson-MLE. From this, it can be concluded that *both* Poisson-MLE and VST are suitable for the shot noise dominated images, and can achieve improved performance without requiring the noiseless image. This equivalency between WNNLS, Poisson-MLE, and VST is discussed further in section 3.6.

3.4.3 Incorporating sparsity in Poisson-MLE and VST

From the results obtained in section 3.3, adding an ℓ_1 regularization term to Poisson-MLE and VST should further improve their performances. Their regularized counterparts, r-Poisson-MLE and r-VST, are formulated as:

$$x_{\text{r-Poisson-MLE}} = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} [(H_{\text{PSF}} \times x)_i - I_{\text{noisy},\, i} \cdot \log(H_{\text{PSF}} \times x)_i] + \lambda \|x\|_1, \text{ subject to } x \geq 0. \quad (3.10)$$

and

$$z = 2\sqrt{I_{\text{noisy}} + 3/8};\ x_{\text{r-VST}} = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} [z_i - 2\sqrt{(H_{\text{PSF}} \times x + 3/8)_i}]^2 + \lambda \|x\|_1, \quad \text{subject to } x \geq 0. \quad (3.11)$$

To show the effect that regularization has on these processing schemes, in Figure 3.5 the performances of r-WNNLS, r-VST, r-Poisson-MLE are plotted and compared with that of WNNLS. From the plot, it can again be concluded that regularized processing schemes can achieve improved performance over their non-regularized counterparts. Interestingly, making sure the processing scheme is appropriate for the dominant shot noise (e.g., NNLS vs WNNLS/Poisson-MLE/VST) and incorporating the prior information of sparsity (e.g., WNNLS vs r-WNNLS) have a significant effect on the resolving ratio of that processing scheme. Comparatively, the exact formulation (e.g., Poisson-MLE, VST and WNNLS and their regularized counterparts) has less effect. This phenomenon is explored further in section 3.6.

3.5 Experimental results and verification

Having developed these alternative processing schemes and verified their improved performances in simulation, their application in a real-world experiment is discussed here. In Table 3.1, a summary of these improved processing schemes is provided. For a processing scheme to be applicable in a real-world experiment and achieve improved performance over NNLS, it needs to fulfill all three of these criteria: 1) be appropriate for the shot noise dominated images; 2) incorporate the prior information of sparsity; and 3) does not require the noiseless image in its formulation.

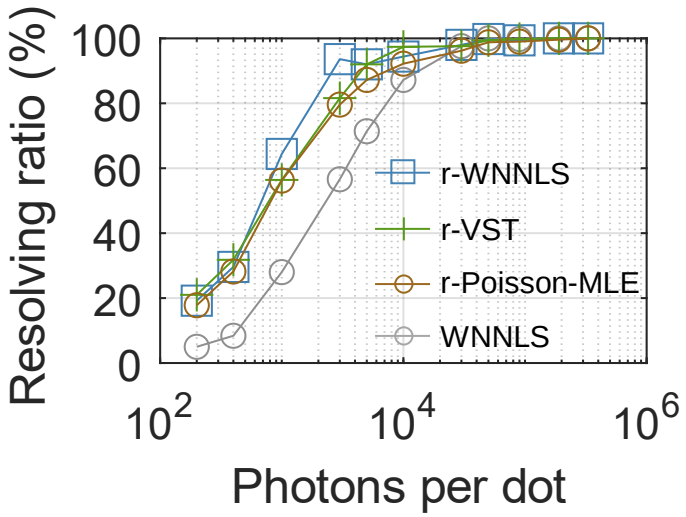


Figure 3.5: Plot of resolving ratios versus the signal level for Poisson-MLE, r-Poisson-MLE, r-VST, and r-WNNLS. Note that Poisson-MLE also represents the performance of VST and WNNLS (Fig.3.4). It can be seen that the addition of a regularization term can further improve the performance of Poisson-MLE and VST. Also note that r-VST and r-Poisson-MLE yield roughly the same performance, showing that these two processing schemes can potentially achieve the highest performance in a real-world experiment.

From Table 3.1, the only two processing schemes that satisfy all three criteria are r-Poisson-MLE and r-VST. From Figure 3.5, r-Poisson-MLE and r-VST achieve similar performance. From these comparisons, either r-Poisson-MLE or r-VST can be used in a real-world experiment and achieve improved performance over NNLS.

To obtain verifiable experimental results in a real-world system, an imaging experiment was performed on a commercially-available fluorescent resolution-test sample made by Argolight [32]. The result is presented in Figure 3.6. The calibration sample consists of successive fluorescently-labeled line-pairs separated by increasing distances from 0nm (no separation) to 270nm. A standard inverted fluorescence microscope, with a custom illumination setup to enable the scanning of the focused illumination spot, was used. This imaging system is illustrated in Figure 2.1. The excitation wavelength was 488nm and the fluorescence images were collected through a bandpass optical filter centered at 520nm, using a 1.4NA objective.

The results are shown in Figure 3.6. First, using the acquired raw images, an equivalent

Processing scheme	Account for shot noise?	Account for sparsity?	Apply to real data?
NNLS			✓
r-NNLS		✓	✓
WNNLS	✓		
r-WNNLS	✓	✓	
Poisson-MLE	✓		✓
r-Poisson-MLE	✓	✓	✓
VST	✓		✓
r-VST	✓	✓	✓

Table 3.1: A summary of the properties for the processing schemes considered. For a processing scheme to achieve the highest resolution, it needs to be appropriate for the shot noise model, take advantage of object sparsity, and be applicable in a real experiment. Here, only two processing schemes satisfy all three criteria: r-Poisson-MLE and r-VST.

widefield diffraction-limited image can be obtained by simply summing the acquired raw images. This diffraction-limited image is shown in the top row of Figure 3.6. From the diffraction-limited image, it can be observed that the 1.4NA objective achieves a resolution of approximately 240nm, as expected. The imaging experiment was conducted per the procedures described in section 2.1. These raw images were then processed using an alternative formulation of r-VST.[1] In this processed image, the 60nm line pair is successfully resolved. Since the conventional diffraction-limited resolution is approximately 240nm, this represents a resolution improvement of approximately four times. When compared with the processed image obtained using NNLS and VST, it can be seen that while NNLS and VST can achieve super-resolution (i.e., the 90nm line pair is resolved), the 60nm line pair is beyond the resolution achievable with NNLS or VST, and the images obtained using NNLS or VST have relatively worse visual quality. This can be assessed by the line profile in Fig. 3.6, or by simply inspecting the enlarged inset region of the processed image. The improved visual quality can be demonstrated by the enhanced contrast between the super-resolved line pair and the dark background exhibited by the r-VST image. This experiment confirms that the highest resolution can be achieved when the processing scheme used is 1) appropriate for the shot noise dominant images; and 2) able to account for the prior information of sparsity.

[1] The reason for this alternative formulation is due to processing speed and selection of a regularization parameter, and is described further in section 3.7.

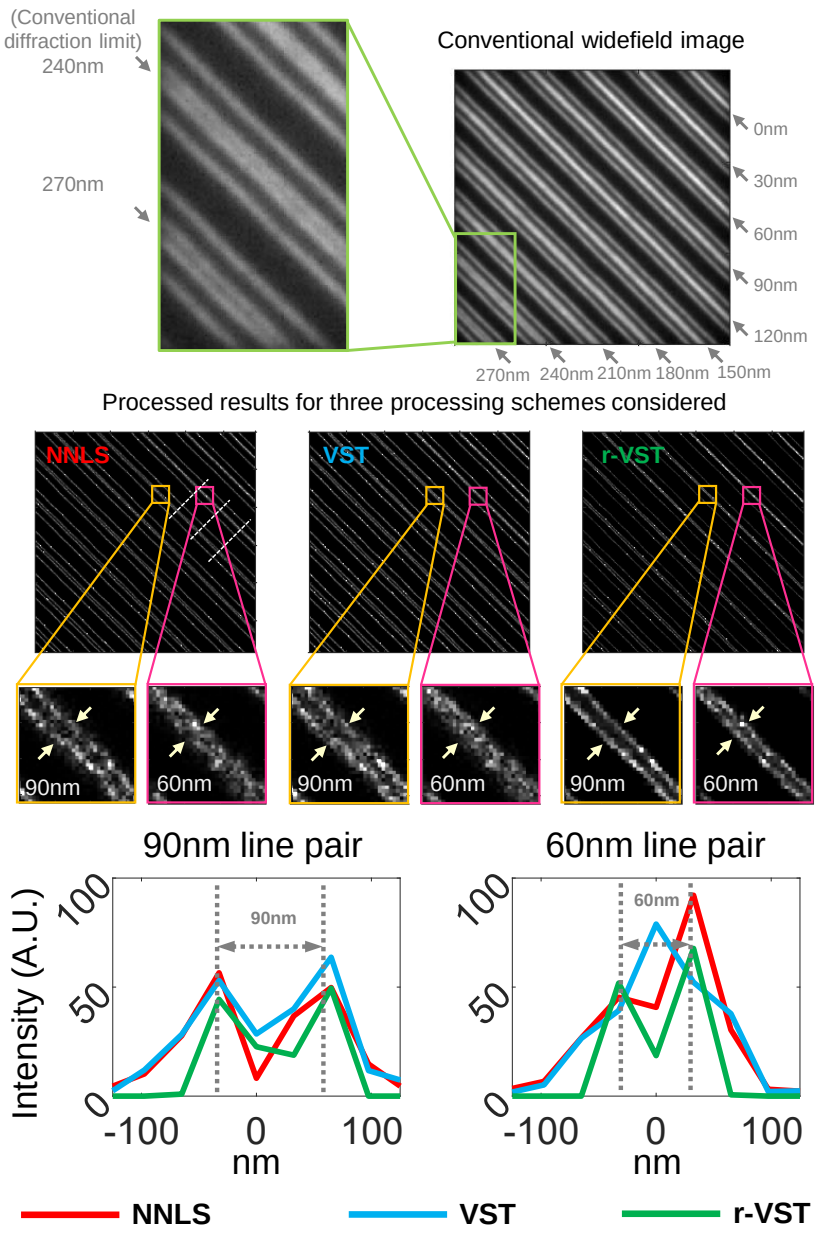


Figure 3.6: Processed Argolight experimental data versus conventional diffraction-limited resolution. Bottom row shows average of three cross-section profiles for the 90nm and 60nm line pairs obtained using NNLS, VST, and r-VST respectively. White dashed lines in the NNLS result indicate the location at which the cross-section profiles are taken (repeated for VST and r-VST results). NNLS, VST, and r-VST all demonstrate successful super-resolution and resolve the 90nm line pair. However, only r-VST resolves the 60nm line pair, and provides the best visual quality. This shows that the highest achievable resolution can be expected when the processing scheme used is 1) appropriate for the noise model, and is 2) able to take advantage of the object sparsity.

3.6 Equivalency between WNNLS, Poisson-MLE, and VST

It is intuitive to assume that the processing scheme performance depends on its exact formulation. However, it can be directly observed from Figure 3.4 that the performances of WNNLS, Poisson-MLE, and VST are all approximately the same. The reason for this, despite their vastly different formulations, is that these three processing schemes are developed to achieve the same goal: generating a maximum likelihood estimator for the shot noise dominated raw images. And as long as that goal is achieved, improved performance can be expected.

Poisson-MLE is derived directly from the log-likelihood function for Poisson random variables, and therefore accommodates the shot noise model natively. For WNNLS, this accommodation is achieved by approximating Poisson random variables with Gaussian random variables with non-zero mean and non-uniform variances. After this approximation is made, Equation (3.3) is appropriate for the dominant noise model because it generates a maximum likelihood estimator for Gaussian random variables with *non-uniform* variances [27]. On the other hand, by applying the Anscombe transform, VST stabilizes the variances of Poisson random variables, and makes them approximately uniform. As a result, an NNLS-like approach, which is appropriate for Gaussian random variables with *uniform* variances, can generate a maximum likelihood estimator, giving rise to Equation 3.9.

The approximations made in WNNLS and VST achieve the same goal: to ensure that the shot noise dominated images can be properly accounted for by the processing schemes. As long as this condition is met, improved performance can be expected. As a result, despite these approximations and different formulations, Poisson-MLE, WNNLS and VST achieve the same performance. This is significant for another reason: the approximations made in WNNLS and VST are only valid when Poisson random variables behave like Gaussian random variables. This occurs in the limit as the mean of the Poisson random variables (i.e., the signal level) goes to infinity. The fact that all three formulations achieve similar performances in Figure 3.4 means that these approximations are valid in the realistic photon levels plotted.

It should be noted that the shot noise dominant model that has been used so far does not

describe the complete picture — it is itself an approximation of the real-world conditions. While the shot noise dominant model is valid for most signal levels, Gaussian readout noise is still a noise source that is present in the acquired images. As a result, the pixels in the acquired noisy images are most completely represented by a compound Poisson-Gaussian random variable. If a processing scheme is developed based on the maximum likelihood estimation of this compound random variable, such as derived in [33], one should expect the performance to improve further. This is not pursued in this thesis work, because the performance improvement obtained this way can be expected to be relatively minor, and would only be apparent at extremely low signal levels or when low-sensitivity sCMOS cameras are used (i.e., high Gaussian readout noise variance). This is because in the context of computational super-resolution microscopy with non-switchable fluorophores, the contribution from Gaussian readout noise is comparatively much lower than that of the shot noise. In [33], this model for compound random variables is used in the context of single molecule localization microscopy. Due to the photo-switchable fluorophores at the center of that technique, the optical signal is produced by a single emitter, which limits the signal level greatly, thus making the contribution of Gaussian readout noise to the total noise much more significant.

3.7 Selection of a regularization parameter

A regularization parameter, λ, needs to be specified for regularized processing schemes such as Equations (3.5) and (3.6). This parameter specifies how sparse the solution vector is expected to be, based on the prior information available. As a result, its selection is crucial in obtaining good performance. This is demonstrated in Figure 3.7. If λ is too small, the regularization term has little effect on the outcome of the numerical problem, and the regularized processing schemes show no significant improvement compared with that of the non-regularized processing schemes. If λ is too large, the regularized processing schemes becomes biased towards an all-zero solution in successive trials. This is the expected behavior of this numerical problem, since if λ is too large,

the optimization problem effectively becomes:

$$x = \text{argmin}\|x\|_1, \text{ subject to } x \geq 0. \tag{3.12}$$

Equation (3.12) has a trivial solution of x being an all-zero vector. This effect of over-regularization can be seen in Figure 3.7: the processed images for this over-regularized case (bottom row) show pixels with extremely low intensities.

Because of this dependence of processing scheme performance on λ, its selection scheme needs to be carefully considered. A trivial scheme is to select a universal λ and use it for all sub-images. However, in the technique described in this thesis, as the focused illumination spot is scanned across the sample laterally, the illuminated part of the sample will change as the scanning proceeds and

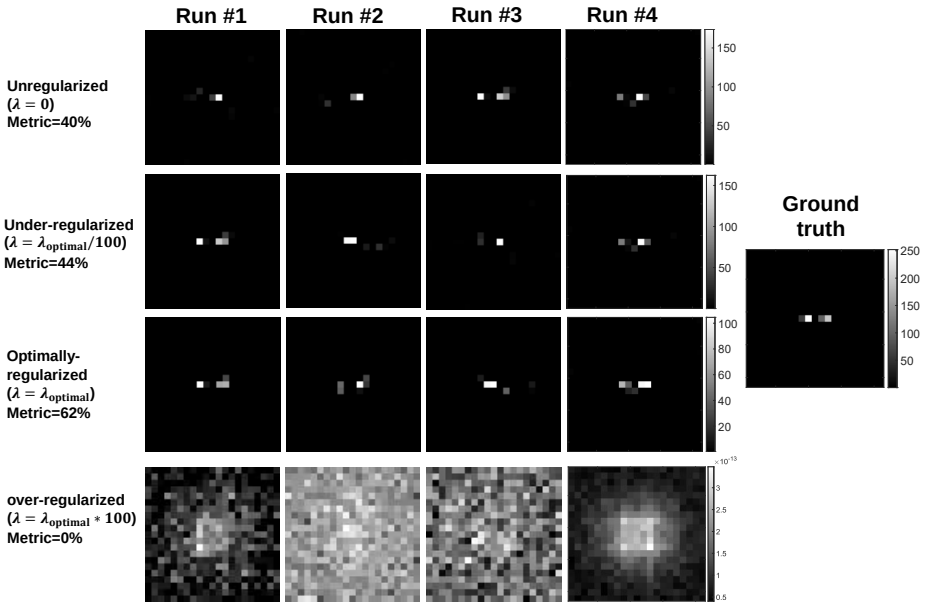


Figure 3.7: Effect of regularization parameter selection. Results from 4 out of 50 runs are presented here. Processing scheme used is r-NNLS. If the parameter is too small compared to optimal, it may still offer some performance benefit compared to no regularization. However, if the parameter is too large, it may have detrimental effect, causing the processing scheme to converge to the trivial all-zero solution regardless of input data. This can be seen from the extremely small (10^{-13}) pixel values in the images on the bottom row.

will thus exhibit different amounts of sparsity. As a result, a universal λ value is unlikely to yield optimal results. Another possible solution is to manually tune the regularization parameter, until the processed image is "visually pleasing". However, this selection scheme faces significant practical challenges. Firstly, tens of thousands of sub-images are typically acquired in an experiment. As a result, individually tuning for an optimal regularization parameter for each sub-image would be too time-consuming. Secondly, because for each sub-image only a very small region (i.e., the size of the tightly focused illumination spot) is illuminated, it is very difficult to determine if the processed image is "visually pleasing" because the processed result for a single sub-image only contains several bright pixels.

As a result of these factors, a selection scheme, ideally automatic, that uses acquired noisy data only, is needed to select an optimal λ. Such a selection scheme is not trivial to develop, and is outside of the scope of this thesis. Despite this, techniques like L-Curve [34, 35] and generalized cross-validation (GCV) [36] have been shown to be viable solutions to this problem. These methods utilize a "goodness-of-the-fit" measure and a "degree-of-freedom" measure to find an appropriate trade-off between under- and over-regularization. Both of these methods were used to generate the simulation results presented in this chapter. The selection of either L-Curve or GCV is based on a trial run of the numerical experiment, and the λ selection scheme that achieves the higher resolving ratio is selected. For the goodness-of-the-fit term, the value of the unregularized term from the regularized numerical problem is used. This term is sometimes referred to as the "fidelity term" in the literature. For the degree-of-freedom measure, the 1-norm of the solution for a given value of λ is used. For a concrete example, for Equation (3.5), the goodness-of-the-fit term, f_1, is defined as:

$$f_1 = \|H_{\text{PSF}} \times x_\lambda - I_{\text{noisy}}\|_2^2, \tag{3.13}$$

and the degree-of-freedom measure, f_2 is defined as:

$$f_2 = \|x_\lambda\|_1. \tag{3.14}$$

Here, x_λ is the solution to the numerical problem in question when the regularization parameter is set to λ.

To identify an optimal λ, a numerical search is conducted. First, the range of λs over which the search is conducted is determined. Next, a series of λs within that range is generated. For each of these λs, an inverse problem is solved with the regularization parameter set to its value. Using the solution, a goodness-of-the-fit measure and a degree-of-freedom measure are calculated and stored. These measures are then interpolated across across the entire range of λs in which the search is conducted. Finally, for GCV, its resulting curve is calculated and the λ value that gives the lowest GCV value is used. For L-Curve, an expression for the analytical curvature from [35] is adopted and the λ value that produces the greatest curvature is used. In the simulations shown in section 3.3, based on observed performance, L-Curve is used for r-NNLS, r-WNNLS, and r-Poisson-MLE. GCV is used for r-VST.

3.7.1 Selecting a good regularization parameter for use in a real-world experiment

The automatic λ selection schemes described before produce good results for simulation data, but they encountered additional challenges when being used to process the real-world experiment data in Figure 3.6. Firstly, the processing time is prohibitively long. To generate the processed results shown in Figure 3.6, 9×10^4 sub-images are collected. From the simulation data, it is found that 10 values in the λ search range usually yields good results. Since for each λ, an inverse problem needs to be solved, this means that there are in total, 9×10^5 inverse problems to be solved. Typically, solving one inverse problem requires approximately 1s of run time. This then corresponds to a total run time of 9×10^5 seconds, or about 10.4 days, which is prohibitively slow.

Furthermore, it is discovered that even if GCV or L-Curve is used, they produce λ values that converge to the two ends of the search space, leading to either under- or over-regularized results, instead of a good trade-off value seen in the simulation. This could be caused by the ever-present fluorescence background in a real-world experiment. For example, if a 10% background signal is added to the two-dot simulation, both GCV and L-Curve produce λ values that severely under-regularize the image, with no practical benefit to performance.

Because of these real-world complications, the images in Figure 3.6 are processed with the

following alternative optimization problem:

$$x_{\text{r-VST,alt}} = \underset{x}{\operatorname{argmin}} \|x\|_1, \text{ subject to } \sum_{i=1}^{n} [z_i - 2\sqrt{(H_{\text{PSF}} \times x + 3/8)_i}]^2 \leq \beta^2, \text{and } x \geq 0. \quad (3.15)$$

Here, instead of λ, a different parameter is used to specify the strength of applied regularization: β. It has been shown that for a properly selected set of λ and β, Equations (3.11) and (3.15) are equivalent to each other [37]. Using this alternative formulation, it is discovered that an effective regularization parameter can be found by simply specifying a universal β instead of conducting an individual search for each sub-image. For the experimental results shown in Figure 3.6, a series of β values were attempted, before the one that yielded the highest resolution was identified (by plotting a diagonal profile) and used to produce the figure. In practice, a human user will have to evaluate the final processed image produced by the processing scheme, and determine its scientific merits. In that case, the processing scheme can output images processed with several different values of β and the end user will decide what is the optimal one for their purpose.

Chapter 4

Imaging with alternative contrast agents

Here, additional experimental results utilizing computational super-resolution microscopy are presented. It is shown that because this technique relies on sparsity, if the sparsity present in the image can be enhanced, for example, via improved contrast agents that exhibit low background signal levels, super-resolution performance can be improved.

4.1 Up-conversion nano particles (UCNPs)

So far, standard, non-switchable fluorescence markers have been used as the contrast agents in this technique. However, they are not necessary for this technique to successfully achieve super-resolution. From the discussions in the previous chapters, it can be seen that this computational super-resolution technique does not place any specialized requirement on the contrast agent, such as being photo-switchable. The only criterion the contrast agent needs to fulfill is that it emits incoherent light. This is so that the imaging model, Equation (2.1), where the image is formed by the sum of a series of shifted and intensity-scaled versions of the PSF, is valid. As long as the contrast agent emits incoherent light, and the object labeled with the contrast agent exhibits some degree of sparsity, super-resolution can occur. This is true for fluorescence microscopy, where the contrast agents emit incoherent light, and the labeled objects contain sparsity due to the chemically-specific labeling.

However, fluorescence is not the only contrast mechanism that can work with this technique. For example, up-conversion nano-particles (UCNPs), can be shown to be effective contrast agents

for use with this technique. UCNPs are small (nano-scale) particles that absorb in the infrared wavelengths, and emit light in the visible or ultraviolet wavelengths. Because of these properties, when used as a contrast agent, images acquired with UCNPs as the contrast agent exhibit lower background signal and deeper optical penetration [38]. Additionally, in recent years, up-conversion signal strength enhancement techniques have been proposed [39], which should improve UCNP performance for computational super-resolution microscopy due to its dependency on signal strength.

Here, an experiment showing computational super-resolution with UCNPs is presented. It is seen that the UCNPs are a valid contrast agent for computational super-resolution. The recovered super-resolved image is verified against the transmission electron microscope (TEM) result and shown to be within close alignment (within the size of the UCNPs) with the ground-truth TEM results. As expected from Chapter 3, the processed result is dependent on the processing scheme used, and the improved processing scheme, identified in Chapter 3, is shown to be able to achieve the best result.

4.2 Experimental procedure

The UCNPs used in this experiment have a nominal diameter of 44nm. They are first deposited onto a TEM grid, and a ground-truth image is obtained using TEM. The concentration of UCNPS and the FOV are selected such that the imaged UCNPs are largely isolated single particles separated by sub-diffraction-limit distances. Then, the sample is mounted onto a glass slide, and imaged on a widefield microscope modified for the technique discussed in this thesis (see Figure 2.1), using a 63X/NA1.4 objective under 980nm excitation. The exact same FOV as that in the ground truth image is recorded. This is possible due to the use of a TEM grid, which contains unique geometric features that aids in finding the correct FOV in the optical microscope.

4.3 Results

First, the ground truth image and the unprocessed optical image are shown in Figure 4.1. Here, it can be seen that since the distances between the particles are smaller than the diffraction

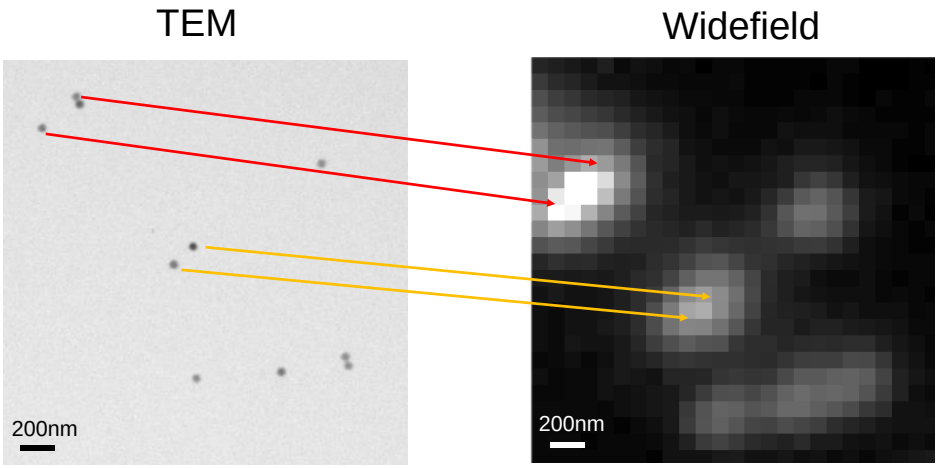
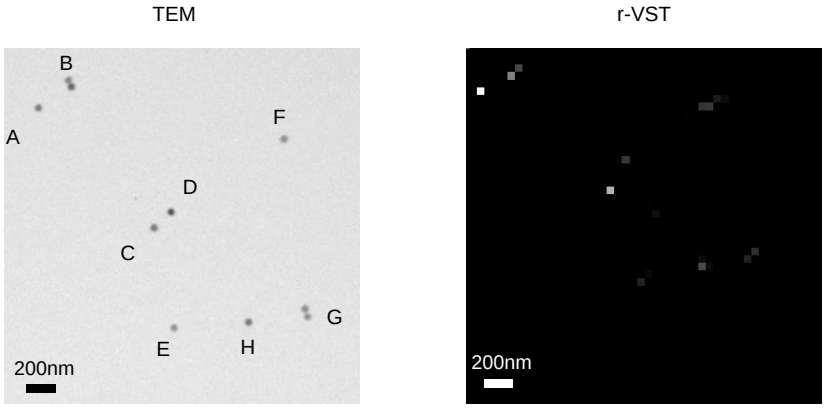
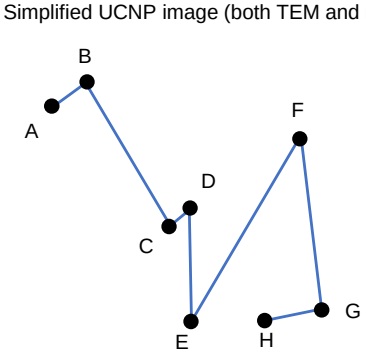


Figure 4.1: TEM and diffraction-limited widefield microscope images of 44nm UCNPs. The TEM image serves as the ground truth for this experiment. Because the distances between some particle pairs are smaller than the diffraction limit, the widefield microscope image cannot resolve these particle pairs. Red and yellow arrows are used to establish the relationship between the images obtained by the two imaging modalities. Contrast of the widefield image is adjusted for improved visualization.

limit, in the optical microscope image the particles are blurred together and cannot be resolved. In Figure 4.2, the processed image obtained using r-VST from Chapter 3 is shown. It can be seen that r-VST successfully super-resolves the details beyond the diffraction limit. Specifically, by treating each UCNP as an isolated point source (in the case where two UCNPs are in contact, they are treated as a single particle), a series of angle and distance measurements can be made and are shown in Figure 4.2 as well. It can be seen that r-VST successfully resolves the UCNPs to a precision that is within their diameters.

As a comparison, Figure 4.3 shows the processed image obtained using NNLS and VST respectively. Here, it can be seen that while NNLS and VST are both able to resolve some of the particle pairs, they both exhibit less accurate results than that obtained by r-VST. This is most clearly seen in the isolated particle in the top right portion of the FOV: while r-VST successfully





Measurements, TEM

Line segment	Length (nm)
AB	225-305
BC	1022-1158
CD	103-206
DE	710-805
EF	1384-1477
FG	1073-1224
GH	331-429

Vertex	Angle (°)
ABC	70-102
BCD	62-97
CDE	30-70
DEF	27-35
EFG	34-40
FGH	81-102

Measurements, r-VST

Line segment	Length (nm)
AB	271
BC	1054
CD	230
DE	845
EF	1345
FG	1085
GH	350

Vertex	Angle (°)
ABC	102
BCD	66
CDE	30
DEF	29
EFG	36
FGH	86

Figure 4.2: TEM image and widefield microscope image processed using r-VST. The sub-diffraction-limit details are successfully recovered by r-VST. By treating each UCNP as an isolated point source (in the case where two UCNPs are in contact with each other, such as point B, they are treated as a single particle), a "constellation" can be generated, and the distances and angles formed by lines between point sources can be measured. Due to the non-trivial size of the UCNPs, the measurements from the TEM image are shown as a range. The r-VST result matches closely with the TEM result, both visually and in terms of the distance and angle measurements.

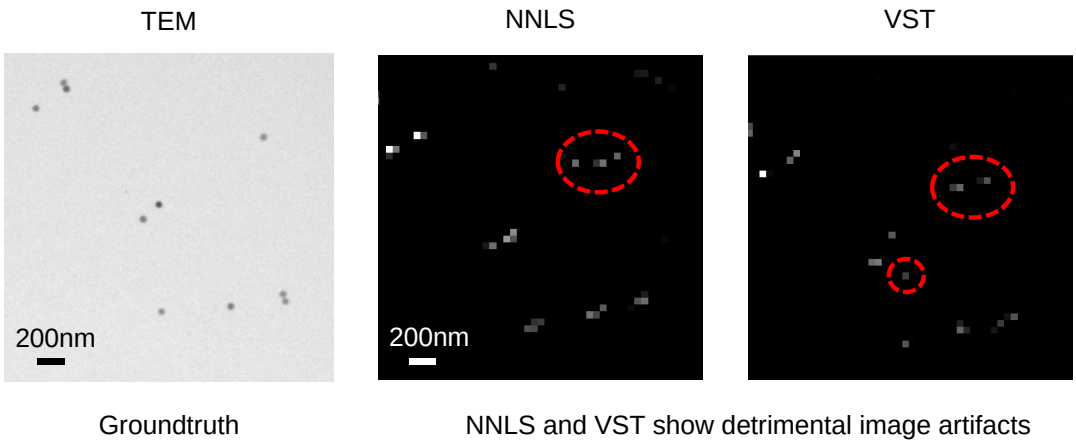


Figure 4.3: Processed images obtained using NNLS and VST, compared with the TEM ground-truth image. As established in Chapter 3, NNLS and VST offer comparatively worse super-resolution performance than r-VST.

identifies it as an isolated particle, both NNLS and VST erroneously identify it as a pair of particles. This again further supports the conclusion drawn in Chapter 3, namely, the highest resolution and the best visual quality can be achieved if the processing scheme is appropriate for the dominant noise model, and takes advantage of the existing prior information.

Another interesting conclusion can be draw from this experiment. It is clear that this super-resolution microscopy technique depends heavily on sparsity. In [21], it is shown that the inherent sparsity in the sample labeling is often not enough to achieve good super-resolution performance. As a result, the experimental results such as those presented in [21] and [23] are obtained with a focused illumination spot, which further enhances the sparsity present in the acquired image. However, the optical microscope image shown in Figure 4.1 is not acquired with scanning, but instead with a single exposure with a large illuminated area. While for a general object, such an illumination profile would not yield meaningful super-resolution, in this case, due to the fact that UCNP images have very low background signal level, coupled with the fact that UCNPs are distributed as isolated particle pairs, meaningful super-resolution is achieved despite the sub-

optimal illumination profile. This further supports the conclusion drawn throughout this thesis: that the sparsity in the illuminated object is the source of super-resolution, and a properly designed processing scheme can extract that super-resolution information.

Chapter 5

Uniqueness of Solutions in Noiseless Imaging

Oftentimes, an examination of a microscopy technique, especially a computation-based super-resolution microscopy technique, effectively becomes a discussion of its artifacts. In the super-resolution microscopy technique described in this thesis, two main sources of artifact exist. In the preceding chapter, the artifact caused by the random nature of the noisy detection is discussed. It is found that, due to the existence of random noise in the detection of microscope images, even when the underlying object remains unchanged, the recovered objects from successive noisy images will be different. To achieve the highest possible resolution, a properly designed processing scheme is needed.

There is another potential source of artifact that is not caused by the noisy detection. This source of artifact is the existence of possible non-unique solutions when solving the inverse problem. This problem is more difficult to mitigate as it exists even in the noiseless scenario. The potential for the solution to be non-unique would obviously make the results obtained using this technique somewhat questionable (in the minds of researchers trying to understand it) and hinder the wider adoption of the technique for microscopy. To address this, first, a brief description of the problem and its cause is given. Next, it is shown that the set of all objects can be further divided into two classes: 1) objects with *similar* images; and 2) objects with the *same* images. It is then shown that, while objects belonging to the second class cause this problem of non-uniqueness, these objects can be rejected by the optimization routine if: 1) the true solution is sparse; and 2) a non-negativity constraint, derived from the incoherent contrast mechanism, is used in the

optimization routine. After conceptually illustrating this relationship between an object's level of sparsity and the existence of a unique solution, two methods of *certifying* the uniqueness of the solution are proposed and explained. Using the method that is better suited for a practical-sized problem, the uniqueness of the solution is examined for: 1) a sparse, two-dot object; and 2) a less sparse, two-line object. It is shown that if the object possesses a high level of sparsity (i.e., the two-dot object), the solution is indeed unique. If the object possesses a slightly lower level of sparsity (i.e., the two-line object), the solution is non-unique. However, the effect of non-unique solutions in this particular instance is not to introduce significant artifacts that will render the recovered object useless to an end user. Instead, the existence of non-unique solutions simply introduces an intensity variation on the recovered object that is of little consequence in a real-world experiment when visualizing object feature intensities is the goal. Finally, some connections between the results obtained in this chapter and the results described in the existing literature are presented.

The significance of the work presented in this chapter is as follows. In the existing literature on computational super-resolution, it is generally accepted that if the signal level is high enough so that it approaches noiseless, a unique solution that enables exact recovery (i.e., unlimited resolution) is possible. For example, in [15], it is stated that:

> "...for noise-free imaging with a criterion of exact passband restoration...unlimited bandwidth extrapolation is possible."

This conclusion is also supported by the simulations conducted in section 3.3.1, where the resolving ratio approaches 1 as the signal level is increased. However, there is no denying that non-unique solutions do exist for this problem. This can be plainly seen from the fact that the matrix used to represent the microscope system (H_{PSF}) in the inverse problem has a non-empty null space. This chapter recognizes that whether the solution to the inverse problem is unique or not depends on the exact object being considered and its level of sparsity. This then enables the reconciliation of this apparent conflict between existing literature versus the inverse problem analysis. This has not been previously reported until now.

5.1 Description of the Problem

A microscope is readily modeled by its point spread function (PSF). The image formed by a microscope with a certain PSF can be calculated using a convolution operation. In the frequency space, Fourier optics is also often used to analyze a microscope. Under this framework, the convolution in the spatial domain is equivalent to the low-pass filter in the frequency domain. The microscope is modeled as a low-pass filter having a passband with a cut-off frequency, f_c. To model the image of an object, the object is first Fourier transformed to obtain the object spectrum. Then the object spectrum is low-pass filtered by the microscope to yield the image spectrum. Lastly, the image spectrum is inverse Fourier transformed to give a final image.

In the above-mentioned low-pass filtering operation, based on f_c, amplitudes of some spatial frequencies are set to zero. As a result of this loss of amplitude, the resultant image will exhibit limited resolution, especially for the higher spatial frequencies (i.e., spatial frequencies needed to represent finer details).

The fact that the low-pass filtering operation sets some spatial frequency amplitudes to **zero** is where potential non-unique solutions arise. Because the goal of computational super-resolution is to recover the amplitudes of these spatial frequencies that are beyond the cut-off frequency f_c, and because their amplitudes are set to zero by the microscope, it is not immediately clear how the processing schemes can recover this seemingly "lost" information. This problem is graphically illustrated in Figure 5.1.

As an example, consider a scenario where a least squares (LS) algorithm is used as the processing scheme to process a *noiseless* image.[1] The goal of LS is to minimize the total squared error between the reconstructed image and the acquired image. LS arrives at a solution x^* when the total squared error between these two terms are no longer decreasing. However, one can construct an alternative solution $x^* + \delta x$ that is as good as x^* in terms of total squared error. This is possible if δx is located *entirely* outside of f_c. Or equivalently, if δx is in the null space of H_{PSF}. When this

[1] Note that the following analysis also holds for NNLS, which is LS with a non-negativity constraint.

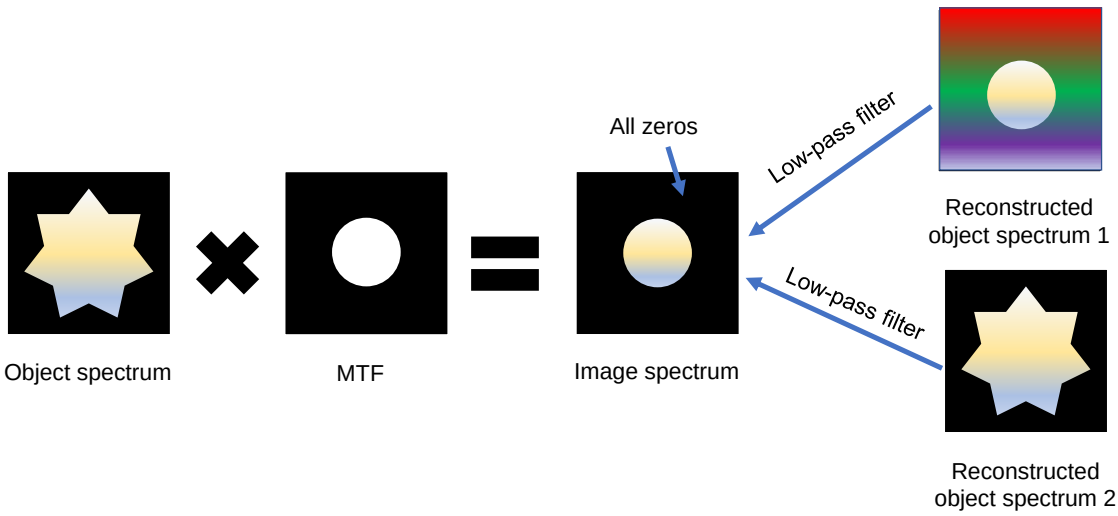


Figure 5.1: Description of the non-uniqueness problem. In a Fourier optics modeling of the microscope imaging process, the object spectrum is multiplied with the MTF to yield the image spectrum. Because the MTF of a microscope is zero outside of the passband, the reconstructed objects can be drastically different outside of the passband, as shown in Reconstructed object spectrum 1 and 2, yet still produce the exact same image spectrum after the low-pass filter operation performed by the microscope. In other words, the two spectra produce the exact same image despite their differences outside of the passband. The rainbow pattern in "Reconstructed object spectrum 1" represents the dramatically different spectrum for this object outside of the passband.

condition is met, because all of it amplitudes are set to zero, δx produces an image that is all zero. As a result, when the original solution x^* is added to δx, the cost function value (i.e., the total squared error between the newly reconstructed image and the acquired image) remains unchanged. Since the cost function value is the only metric available to LS to determine which solution is the better solution, both x^* and $x + \delta x$ can be returned as valid solutions (depending on the starting condition supplied to LS). As a result, there are at least two solutions, x^* and $x + \delta x$, that are equally good from the perspective of LS.

Because the image of δx is zero, this object is commonly referred to as an "invisible object" in existing literature. It is this type of invisible object that gives rise to the problem of non-unique solutions. There are, in fact, an infinite number of invisible objects, all equally good from the perspective of NNLS. This is because, as long as these invisible objects consist purely of spatial frequencies that are outside of the passband of the microscope, their images will be all-zeros. As

a result, there are an *infinite* number of potential solutions that provide the same cost function value, but only *one* true solution that correctly describes the object.

This problem can also be viewed through a linear algebra lens: as stated in section 2.2.1, the matrix H_{PSF}, which describes the imaging system and is used for the recovery of super-resolved images, need not be a square matrix, and is, in general, not a square matrix. In fact, in practice, H_{PSF} usually has more columns than rows. This is because the rows of H_{PSF} correspond to the physical pixels on the acquired images, which only need to cover a region that is, at most, on the scale of the illumination spot. In contrast, the columns of H_{PSF} need to cover the same space with a finer discretization to enable the modeling of small, super-resolved structures. As a result, more columns are needed. The fact that H_{PSF} has more columns than rows means that H_{PSF} is over-determined, which implies that there are an infinite number of solutions.

5.2 Two classes of objects

In section 5.1, the concept of an invisible object, that is, an object that results in an all-zeros when imaged by a microscope, is introduced. The concept of invisible objects, sometimes referred to as "null objects", is rather common in existing research in computational super-resolution (e.g., see [15] and [40]), and offers powerful insights when discussing this problem of non-uniqueness.

Intuitively, it is not difficult to imagine that when there is no noise (or when the SNR is extremely high), unique and exact recovery is possible, as has been shown in section 3.3.1 at the higher end of the SNR range. The same result has also been proven in a more rigorous manner, for example in [15] and [41]. However, this is at odds with the non-uniqueness problem introduced in the previous section. To address this, it is useful to separate all possible objects into two classes [42]: 1) objects with *similar* images; and 2) objects with the *same* images. Under this framework, for a given noiseless image, if there is a unique solution to the inverse problem, the underlying object falls under the first class; or if the solution is non-unique, the underlying object falls under the second class. This is because, if the images of two objects are only similar, they will still be ever so slightly different, and often different enough for a properly designed processing scheme to

distinguish them. In other words, objects with similar images have a one-to-one mapping from an object to an image, whereas objects with the same images have a many-to-one mapping. As a result of this classification, if the processing scheme is searching for the solution in a space that includes the second class of objects (i.e., objects with the same image), the solution will be not be unique because multiple objects can give rise to *exactly* the same image. Hence, to ensure that there is a unique solution to the inverse problem in this noiseless scenario, the goal is to limit the search to only the objects in the first class. To achieve this, a detailed investigation into the properties of invisible objects is provided below.

5.3 Properties of invisible objects

In this discussion, NNLS is chosen as the processing scheme. In the noiseless scenario, to "solve for the object" is equivalent to finding the inverse of the system matrix H_{PSF}. Since H_{PSF} need not be a square matrix, and is usually not a square matrix (see section 2.2.1), this is achieved by finding the pseudo-inverse of H_{PSF}, which is the same as finding the least squares solution. Since the object is known to be non-negative (from the incoherent contrast mechanism), an iterative solver (e.g., MATLAB `quadprog`) can be used to solve the NNLS problem, which allows a non-negativity constraint to be implemented. Here, it is shown that in the noiseless scenario, a unique solution is only possible when the solution is **1) known to be non-negative and is 2) limited in size spatially, or exhibits sparsity**.

For a given ground-truth object x_{true} that, when imaged by a microscope described by matrix H_{PSF}, gives rise to the noiseless image $I_{\text{noiseless}}$, the following NNLS problem is used to find an estimate for x_{true}, or x^*:

$$x^* = \underset{x}{\operatorname{argmin}} \, \|H_{\text{PSF}} \times x_{\text{true}} - I_{\text{noiseless}}\|_2^2, \text{ subject to } x \geq 0. \tag{5.1}$$

Since the same matrix H_{PSF} is used in the generation of $I_{\text{noiseless}}$ and the processing to find x^*, the error between the reconstructed image ($H_{\text{PSF}} \times x^*$) and the acquired image ($I_{\text{noiseless}}$) can be

eliminated completely if $x^* = x_{\text{true}}$, i.e.,

$$f(x_{\text{true}}) = \|H_{\text{PSF}} \times x_{\text{true}} - I_{\text{noiseless}}\|_2^2 = \|I_{\text{noiseless}} - I_{\text{noiseless}}\|_2^2 = 0. \tag{5.2}$$

Under this condition ($x^* = x_{\text{true}}$, or NNLS correctly recovers the super-resolved image), assuming that the solution to the numerical problem is non-unique, and an alternative solution $x^{'}$ can therefore be constructed using the above-mentioned method, by adding an invisible object represented by vector δx to the original estimate x^*, i.e., $x^{'} = x^* + \delta x$. Substituting $x^{'}$ in the the cost function for NNLS, the following expression for the total squared error for this alternative solution can be obtained as:

$$\begin{aligned} f(x^{'}) &= \|H_{\text{PSF}} \times x^{'} - I_{\text{noiseless}}\|_2^2 \\ &= \|H_{\text{PSF}} \times (x^* + \delta x) - I_{\text{noiseless}}\|_2^2 \\ &= \|H_{\text{PSF}} \times x^* + H_{\text{PSF}} \times \delta x - I_{\text{noiseless}}\|_2^2. \end{aligned} \tag{5.3}$$

Substituting Equation 5.2 into Equation 5.3, the final expression for the total squared error for the alternative solution $x^{'}$ is:

$$f(x^{'}) = \|H_{\text{PSF}} \times \delta x\|_2^2. \tag{5.4}$$

To investigate how the this final error term for the alternative solution $x^{'}$ affects the uniqueness of the problem, it is useful to divide the discussion into three scenarios based on the relative magnitude of the error term and zero. These scenarios are: 1) $f(x^{'}) < 0$; 2) $f(x^{'}) > 0$; and 3) $f(x^{'}) = 0$. The first two scenarios are trivial: because of the 2-norm in NNLS, $f(x^{'})$ cannot be negative; if $f(x^{'}) > 0$, that simply implies that NNLS has not converged yet. The more interesting $f(x^{'}) = 0$ scenario is examined in greater detail next.

As discussed in section 5.1, when $f(x^{'}) = 0$, $f(x^{'}) = f(x^*) = 0$. As a result, NNLS cannot distinguish between x^* and $x^{'}$, and there will be an infinite number of possible solutions that are equally good. From this, it can be concluded that: 1) the non-uniqueness of the solution is caused by invisible objects; 2) for such an invisible object, its image is a vector of all zeros; and 3) because microscopes form a positive image for a positive object (due to the incoherent contrast mechanism),

an invisible object must oscillate between positive and negative values for it to generate an all-zero image.

5.4 Sparsity in the object and non-negative processing leads to a unique solution

Here, it is shown that a combination of a non-negativity constraint and sparsity in the solution vector permits the existence of a unique solution to the numerical problem in the context of this super-resolution technique.

First, the concept of sparsity, and how it is achieved in this microscopy technique, is discussed. Sparsity usually refers to the fact that not all elements contained in the solution vector are non-zero. While there are many possible definitions for this ubiquitous concept, in this thesis, the following definition is adopted: if a vector possesses sparsity, this means that at least one element of that vector is zero. In the context of imaging, this means that some of the pixels show zero intensity, or are visually very dim.

In biological fluorescence microscopy, sparsity arises from two sources. In Chapter 3, specifically section 3.3.2, one source of sparsity is discussed. This source comes from the chemically-specific labeling used in fluorescence imaging. Whether the labeling is achieved extrinsically via antibodies or dyes that only bind to specific cellular structures, or intrinsically via genetic engineering of fluorescence proteins that are only expressed at specific sites, the labeled structures are placed against a non-labeled, dark (zero intensity) background. Because the microscope is imaging these labeled structures, even if the acquired images are blurry, the underlying object will always exhibit some degree of sparsity. This sparsity condition only becomes invalid if the object itself is composed of uniform intensity everywhere: in this case the object is effectively a fluorescence background, and is not of interest in an imaging experiment.

The exact structure (and sparsity) of the object being imaged is dependent on one more factor other than the labeling: the illumination profile. While the structure of the fluorescent samples may be beyond the control of the microscopist, the control of the illumination profile is entirely in

their hands. As discussed in section 2.2.1, when the illumination profile is taken into consideration, the actual object being imaged by the microscope is $I_{\text{ill}} \cdot I_{\text{obj}}$, where I_{ill} is the illumination profile and I_{obj} is the fluorescently labeled structures. If a widefield illumination profile (i.e., uniform illumination) is used (i.e., $I_{\text{ill}} = \mathbf{1}$), the object can potentially (depending on the labeling) stretch across entire field of view (FOV); conversely, if a focused illumination spot is used, it can be guaranteed that the object will be zero outside of the illumination spot. While some techniques (e.g., STED [3, 4]) take advantage of illumination-induced sparsity by utilizing artificially reduced-size PSFs to directly produce an image with improved resolution, it is shown in ([21, 23] and chapter 3) that the sparsity as a result of chemically-specific fluorescence sample labeling, when enhanced by the spot illumination, is often enough for a properly designed post-processing scheme to recover information that is significantly beyond the diffraction limit.

Next, the factor of non-negativity is discussed. In this technique, non-negativity means that the solution vector should only contain elements that are equal to or greater than zero. This is a result of the incoherent contrast mechanism used in fluorescence imaging, which forces the detected image to be of intensity, and therefore non-negative. In the processing schemes discussed in chapter 3, this non-negativity requirement is implemented by adding an inequality constraint in the formulation of the schemes.

By adding a non-negative constraint, the search space where the processing schemes can look for a solution is significantly reduced, and combined with the sparsity in the object, this leads to the potential for the solution to be unique. Specifically, the sparser the object, the more likely it is that there is a unique solution. This is because when the object exhibits sparsity, either inherent in the labeling, or enhanced by the focused spot illumination, the true solution vector will contain zero elements. As a result, when one attempts to construct an alternative solution that gives the same error term value, due to the oscillatory nature of the invisible objects, for regions where the true solution is zero, some elements in this alternative solution will be negative. Due to the inclusion of a non-negativity constraint, solutions such as this one will be rejected by the processing scheme, ensuring that the solution is unique. Conversely, if the true object is not sparse, i.e., all elements

contain non-zero intensities, then one can find an invisible object that, after being added to the true solution, still generates an alternative solution that is non-negative. Thus it cannot be rejected by the non-negative processing scheme. These two scenarios are illustrated graphically in Figure 5.2 and Figure 5.3.

In the context of the super-resolution microscopy approach described here, this shows the importance of focused spot illumination. If the object is not inherently sparse (e.g., if the object is a continuous structure, such as the line pairs in Chapter 3), it is still possible for some invisible objects to be admissible, thus creating the problem of non-uniqueness. By illuminating the sample using a spatially confined focused laser spot, pixels in the true solution vector will be zeros outside of the illumination spot, which is very small compared to the space spanned by the columns of H_{PSF}.

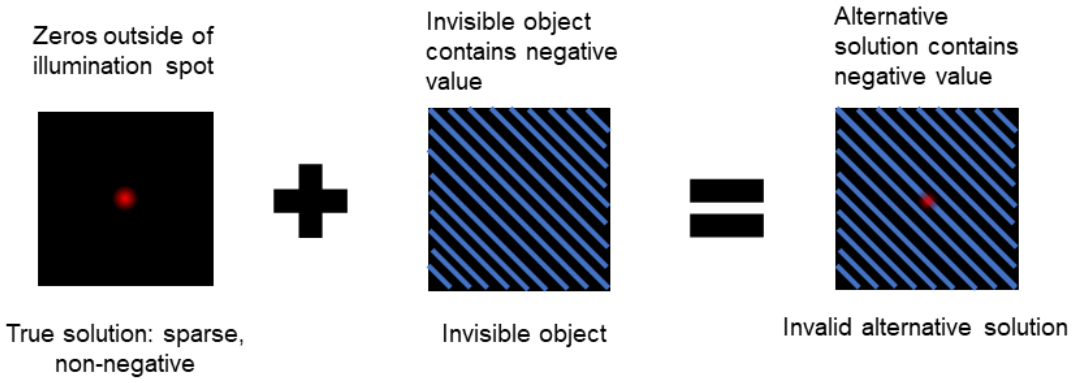


Figure 5.2: Constructing an alternative solution when the true solution is sparse and non-negative. A possible alternative solution can be constructed by adding an invisible object to the true solution. Since invisible objects produce a zero image and the microscope system does not subtract intensity (i.e., the PSFs stored in H_{PSF} do not contain negative values), an invisible object has to contain negative values. Since the true solution is zero outside of the illumination spot, the resultant alternative solution will contain negative values. If a non-negative optimization routine is used to process the image, such an alternative solution will be rejected.

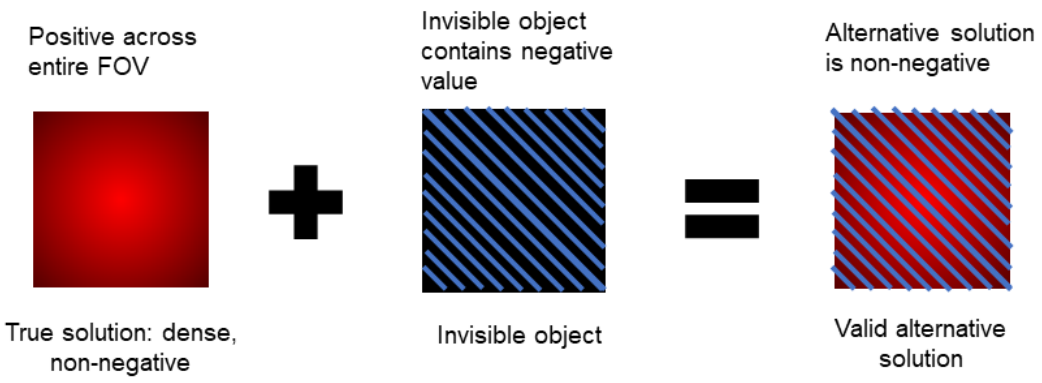


Figure 5.3: Constructing an alternative solution when the true solution is dense and positive. Similarly to Figure 5.2, an alternative solution is constructed by adding an invisible object to the true solution. However, in this case, since the true solution is dense and is positive across the entire FOV, by adding an invisible object, it is possible to find an invisible object with small enough negative elements such that the resultant alternative solution still satisfies the non-negative optimization routine used to process the image. As a a result, this alternative solution will not be rejected in the processing.

5.5 Certifying uniqueness of the solution quantitatively

Existing publications on computational super-resolution generally treat the noiseless recovery problem as having a unique solution all the time (for example, in [41] and [15]). However, since the MTF sets some spatial frequency amplitudes to zero, or equivalently from a linear algebra perspective, the matrix used in the inverse problem, H_{PSF}, has a non-empty null space, non-unique solutions do exist for this problem. This apparent contradiction has received less attention in existing literature. One reason it has received comparatively less research effort is that the existence of non-unique solutions only poses a significant change if H_{PSF} has a non-trivial null space. If H_{PSF} has smaller numbers of columns, this is unlikely to happen. However, in this microscopy technique, where H_{PSF} can be discretized on a very fine grid in order to enable the recovery of small details (see section 2.2.1), H_{PSF} can contain many columns. As a result, it is more likely to have a large non-trivial null space. Therefore, researchers using this computational super-resolution microscopy technique are more likely to encounter this non-uniqueness issue.

As shown in Figures 5.2 and 5.3, whether the solution is unique or not depends on the sparsity

level of the imaged object. The sparser the object, the more likely the solution is unique. while this result successfully reconciles the above-mentioned contradiction, it cannot determine the uniqueness of the solution for an arbitrary object. For a potential user of this microscopy technique, it is more beneficial to generate a quantitative "certification" regarding the uniqueness of the solution. This will allow the user of a microscope to evaluate if the sparsity level in their imaging experiment is sufficient to ensure a unique solution by first conducting this certification process. If the solution is determined to be non-unique, the user can then proceed to determine how detrimental the effect of non-uniqueness will be. This need for a quantitative certification of uniqueness of the solution motivates this section, where two methods of certification are proposed.

First, the conditions which a valid non-unique solution needs to satisfy are presented. In section 5.4, a method to construct a non-unique solution is described: by adding to the true solution an invisible object that is located entirely outside of the passband of the microscope. In other words, an alternative solution x_{alt} to the true solution x_{true} can be obtained by:

$$x_{\text{alt}} = x_{\text{true}} + \delta x. \tag{5.5}$$

In Equation (5.5), δx is the invisible object that is entirely outside of the passband, and it therefore produces a zero image. i.e., δx needs to satisfy:

$$H_{\text{PSF}} \times \delta x = \vec{0}. \tag{5.6}$$

In Equation (5.6), $\vec{0}$ is a vector of all zeros. If the alternative solution x_{alt} is a valid non-unique solution, it must satisfy the non-negativity constraint as well. That is, $x_{\text{alt}} \geq \vec{0}$, where the inequality is taken element-wise, then δx needs to additionally satisfy:

$$x_{\text{alt}} = x_{\text{true}} + \delta x \geq \vec{0}. \tag{5.7}$$

Or equivalently,

$$\delta x \geq -x_{\text{true}}. \tag{5.8}$$

Combining Equations (5.6) and (5.8), it can be seen that if the solution to the optimization problem is non-unique, and that if the non-unique solution x_{alt} is expressed in terms of the true solution,

x_{true}, as $x_{\text{alt}} = x_{\text{true}} + \delta x$, then δx must satisfy the following two conditions:

$$H_{\text{PSF}} \times \delta x = \vec{0}$$

$$\delta x \geq -x_{\text{true}}. \tag{5.9}$$

Here, any δx that satisfies Equation (5.9) is termed an ***admissible*** invisible object. It can be seen that whether an invisible object is admissible depends on the value of x_{true}. Equation (5.9) is a constrained linear programming problem. By simple inspection, two conclusions can be reached: 1) Equation (5.9) has a trivial solution of $\delta x = \vec{0}$. This is because if $\delta x = \vec{0}$, Equation (5.6) is always satisfied. Since x_{true} is non-negative, or $x_{\text{true}} \geq 0$, Equation (5.8) is also always satisfied for $\delta x = \vec{0}$; and 2) whether the solution to the optimization is unique depends on two factors: the microscope system, H_{PSF}, and the true object, x_{true}. For a particular combination of H_{PSF} and x_{true}, if a *non-trivial* solution to Equation (5.9) can be found, an admissible invisible object exists, and the solution is non-unique. Alternatively, if only the trivial all-zero solution exists for Equation (5.9), then the solution is unique. How this can be determined is discussed next.

5.5.1 Solution for a dense object is non-unique: proof

Here, a proof for the solution being non-unique is provided for the case of a dense object, i.e., Figure 5.3.

First, assume H_{PSF} has a non-trivial null space, which is valid in this microscopy technique. Also assume the true solution, x_{true}, is not sparse and all pixels in x_{true} are strictly positive, i.e., $x_{\text{true}} > 0$. Then for a vector $\vec{z}$ in the null space of H_{PSF}, $H_{\text{PSF}} \times (x_{\text{true}} + \epsilon\vec{z}) = I$, where $I = H_{\text{PSF}} \times x_{\text{true}}$. By choosing a sufficiently small ϵ, the non-negativity of the alternative solution can be guaranteed, i.e., there exists ϵ, such that $x_{\text{true}} + \epsilon\vec{z} \geq 0$. As a result, $x_{\text{true}} + \epsilon\vec{z}$ is a valid alternative solution. It should be noted that this is an extreme case where the object is completely non-sparse, and methods for determining the uniqueness of the solution for a more general case are shown next.

5.5.2 Certifying non-uniqueness by finding a non-unique solution

The certification of *non-uniqueness* is comparatively easier, and can be achieved simply by solving Equation (5.9) and obtaining a *single* non-trivial solution. Because there exists at least one admissible invisible object whose addition to the true solution does not affect the reconstructed image (i.e., $H_{\text{PSF}} \times x_{\text{true}} = H_{\text{PSF}} \times (x_{\text{true}} + \delta x)$), the solution to the optimization problem constructed by this particular combination of H_{PSF} and x_{true} is non-unique. Since Equation (5.9) is a constrained linear programming problem, one can use an established solver such as MATLAB `linprog` to solve it.

Here, an example of a non-unique solution for a dense object is shown. The object is a uniform object, or $x_{\text{true}} = \vec{1}$. This object possesses the lowest possible level of sparsity, as all of its elements are the same value. For such an object, as conceptually described in section 5.4, the solution is non-unique. This can be verified by solving Equation (5.9) using MATLAB `linprog` and the obtained solution, δx is shown in Figure 5.4. Here, it can be seen that the invisible object spans the entire FOV, and oscillates with high frequency between negative and positive values. As a result, the image of this δx will be zero, or in practice, due to numerical error in `linprog`, be very close to zero. Therefore, when this invisible object is added to the true object, the reconstructed image is unchanged.

The invisible object was obtained using the MATLAB `linprog` command with its default starting point. As stated in section 5.4, if there exists a single admissible invisible object, an infinite number of invisible objects exists. This is caused by two reasons. Firstly, one can scale the existing admissible invisible object $x_{\text{invis,1}}$, and obtain another admissible invisible object $x_{\text{invis,2}} = \alpha \cdot x_{\text{invis,1}}$, where α is a scalar, as long as the new invisible object $x_{\text{invis,2}}$ still satisfies $x_{\text{invis,2}} \geq -x_{\text{true}}$. Secondly, if multiple invisible objects, e.g., δx_1 and δx_2, can be found, as shown later in section 5.5.3.1, any convex combination between δx_1 and δx_2, i.e., $\beta \delta x_1 + (1 - \beta)\delta x_2$ is also a valid invisible object.

In solving Equation (5.9) for a dense x_{true}, if a different starting point is picked, the resultant admissible invisible object may also be different. The MATLAB linear programming routine used

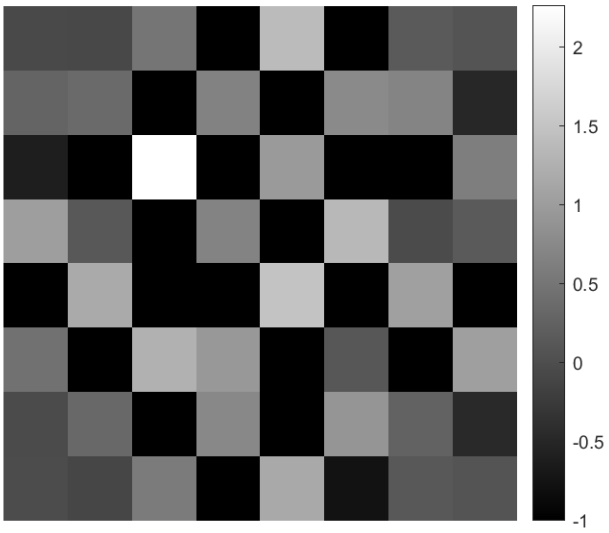


Figure 5.4: Image of an admissible invisible object for a uniform true object. This object spans the entire FOV and oscillates between positive and negative values rapidly. When this invisible object is added to the original true object, the output image is not changed. As a result, an infinite number of non-unique solutions can be found by scaling this invisible object and adding it to the true solution.

to generate Figure 5.4 implements a built-in starting point that is not accessible to its user. Using the MATLAB constrained optimization routine `fmincon`, the effect of different starting points can be investigated. In Figure 5.5, four admissible invisible objects are shown, each obtained using a different starting point. In Figure 5.5 (a), an all-zero starting point is used, and in Figure 5.5 (b-d), a randomized starting point is used for each panel.

From the figure, the following two observations can be made: 1) since Equation (5.9) has a trivial solution of all zeros, when an all-zero starting point is supplied to `fmincon`, it is returned as the optimal solution without any processing. This is shown in Figure 5.5 (a); 2) if the user-supplied starting point is not the trivial solution, `fmincon` will attempt to find an optimal point from that starting point. If the starting points are different from one another, as is the case in Figure 5.5 (b-d), the resultant admissible invisible objects will be different as well. Despite their differences, these invisible objects exhibit common properties, chief among which is that they all span the entire

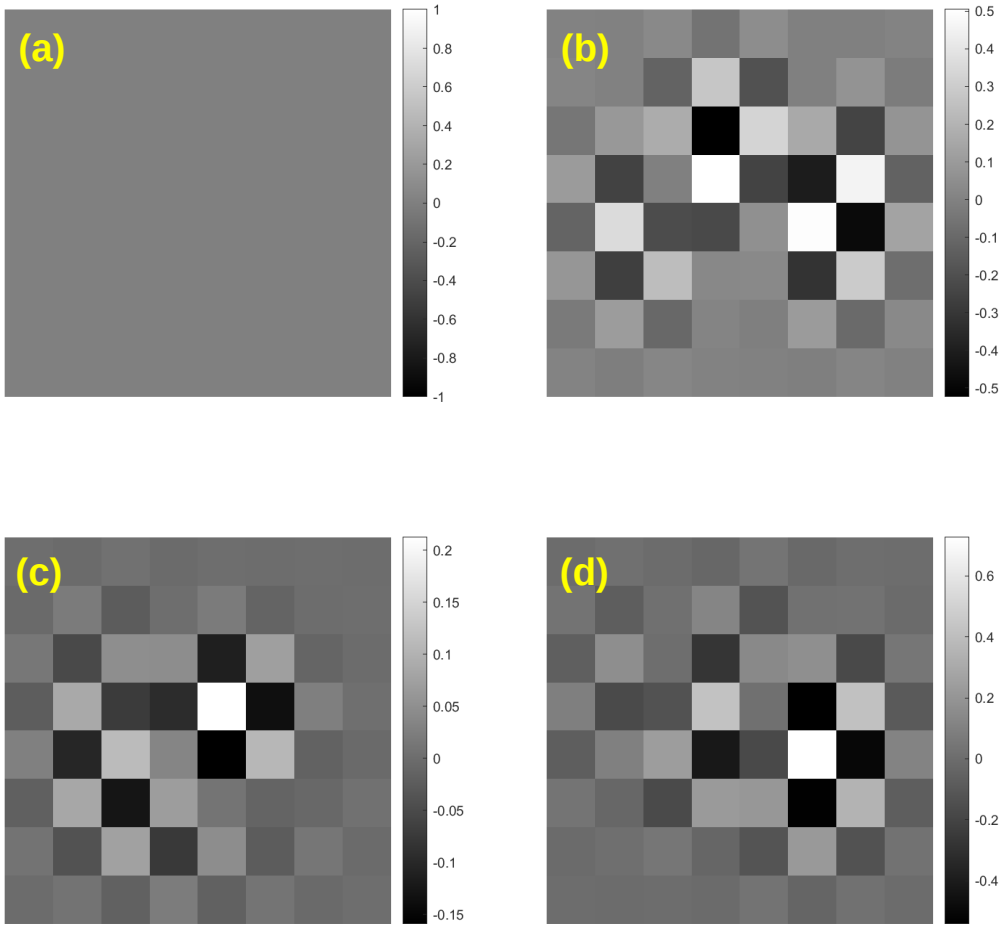


Figure 5.5: Effect of starting point on obtained admissible invisible objects. The same uniform true object from Figure 5.4 is used, and `fmincon` is used to find these invisible objects. Panel (a) shows the result when the trivial all-zero solution is used as the starting point. Panels (b)-(d) show the results when random starting points are used. As expected, since the starting point for (a) is itself a trivial solution, it is returned without further processing. In contrast, because there are an infinite number of admissible invisible objects, the results shown in (b) through (d) are all different from one another.

FOV and oscillate rapidly between positive and negative elements.

Through investigations in this section, it can be concluded that: 1) for a dense true object,

the solution to the optimization problem is non-unique; 2) this can be certified by simply finding a single non-trivial admissible invisible object; 3) an admissible invisible object, which can be used to construct alternative solutions, can be found by solving a linear programming problem (LP), Equation (5.9); 4) the LP in question has a trivial solution of all-zeros; 5) if the solution to the LP returned by a numerical optimization routine is not the trivial all-zero solution, then the solution to this particular inverse problem (as defined by H_{PSF} and x_{true}) is not unique; and 6) if the solution to the inverse problem is non-unique, there are infinitely many possible solutions.

5.5.3 Certifying uniqueness: two methods

Certifying the solution to the inverse problem is *unique* is a more difficult problem. This is because, even if the solution to LP, Equation (5.9), is the trivial all-zero solution, it does not guarantee the uniqueness of the inverse problem. Since an infinite many solutions may exist, and the all-zero solution is one of them, the fact that the solver returns the all-zero solution does not necessarily ensure the solution to the inverse problem is unique.[2] As a result, a different method is necessary to certify the uniqueness of the inverse problem. Here, two methods to certify the uniqueness of the solution to the inverse problem are investigated.

5.5.3.1 Finding vertices of a polyhedron

Here, it is shown that the conditions that admissible non-unique solutions need to satisfy can be geometrically represented as a polyhedron. By finding the vertices of the polyhedron, all non-unique solutions can be found. In addition, a certification regarding the uniqueness of the solution can be generated by observing these vertices.

As previously stated, an admissible invisible object needs to satisfy Equation (5.9). This equation, composed of a series of linear inequality and equality constraints, effectively describes a polyhedron in n-dimensional space, where n is the number of columns in H_{PSF}. Any point that falls within this polyhedron is an admissible invisible object. This concept is demonstrated in a

[2] In fact, a popular initialization point for many numerical optimization routines is the trivial all-zero vector. If the all-zero vector is used to initialize the LP, it will be returned immediately, since it is already an optimal point.

simplified 2-dimension example in Figure 5.6. Here, the following inequality constraints are placed on the variables, x_1 and x_2: $0 \leq x_1 \leq 3$, and $0 \leq x_2 \leq 2$. Each of these inequality constraints specify a half space in the 2-dimensional plane, and the intersection of these half spaces is the feasible region. In this example, the feasible region specified by the inequalities is a rectangle. Additionally, an equality constraint can be represented as a line in this graphical example. The equality constraint used here is a simple linear equation: $x_1 + \frac{3}{2}x_2 = 3$. The points that satisfy both the equality and inequality constraints can be represented by:

$$(x_1, x_2) \in \{(x_1, x_2) \ : \ 0 \leq x_1 \leq 3, \ 0 \leq x_2 \leq 2, \ x_1 + \frac{3}{2}x_2 = 3\} \tag{5.10}$$

Graphically, these points make up the line segment that is the intersection between the rectangle described by the inequality constraints and the line described by the equality constraint. As can be seen in Figure 5.6, in this current form, Equation (5.9) is specifying the edges of this polyhedron.[3] If a point falls within the polyhedron that resulted from Equation (5.9), it is an admissible invisible object.

Besides specifying a polyhedron by its edges, an alternative method of specifying a polyhedron exists using its vertices. For example, instead of $0 \leq x_1 \leq 3$, and $0 \leq x_2 \leq 2$, the feasible region described by these inequality constraints in Figure 5.10 can be described using the following set of four vertices: $\{(x_1, x_2) \ : \ (0,0), (3,0), (0,2), (3,2)\}$. Once the four vertices of the feasible rectangle are specified, all points within it can be specified by a convex combination of the vertices. For example, point $(1.5, 1)$ can be specified by $(1.5, 1) = 0.5(3,0) + 0.5(0,2) + 0(0,0) + 0(3,2)$. The fact that all points within the feasible region can be expressed by convex combinations of its vertices means that it is possible to effectively find *all* admissible invisible objects. If the vertices of the polyhedron can be found, then equivalently, by adjusting the coefficients of the convex combination, all admissible invisible objects can be specified. Because the coefficients for a convex combination need to add up to 1, this ensures that no other admissible invisible objects exist outside of those expressed as convex combinations of vertices.

[3] In this simplified 2-dimensional case, the "polyhedron" is a line segment. As a result, it only has a single edge.

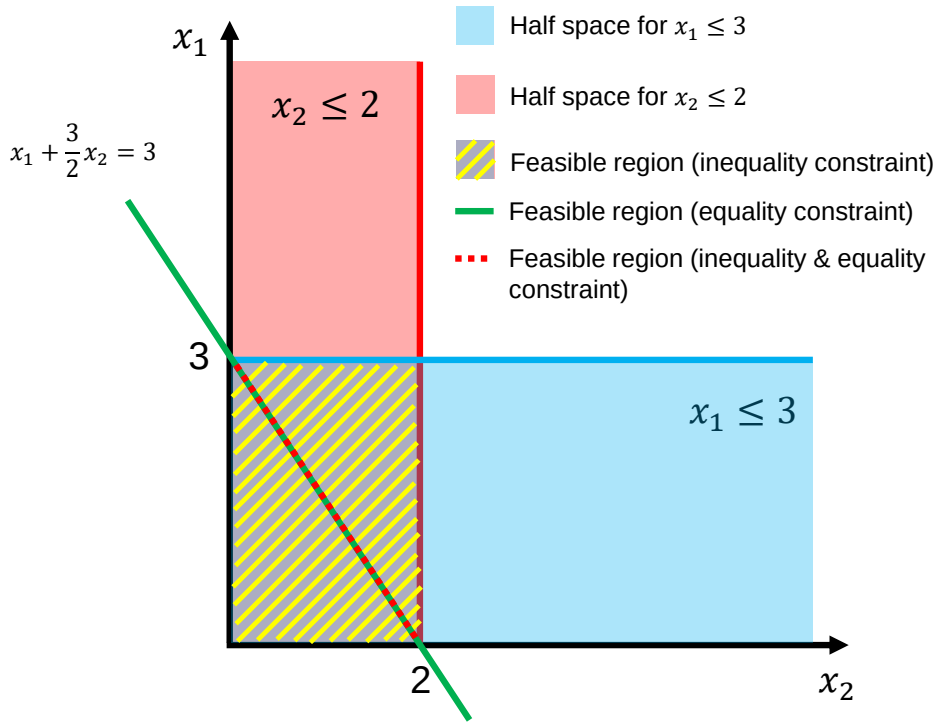


Figure 5.6: 2-dimensional example of constraint forming a polyhedron. In this 2-dimensional example, the feasible region formed by the inequality constraints is a rectangle. An equality constraint as specified in Equation (5.9) is a line in this representation. The feasible region described by the combined equality and inequality constraints is the line segment that is the intersection between the rectangle and the line.

Determining the vertices of a polyhedron is known as the "vertex enumeration " problem, and it has been the subject of extensive research [43, 44]. Because vertex enumeration is not the focus of this thesis, in the results shown in this chapter a third-party MATLAB script, `con2vert`, is used to generate vertices of a polyhedron from its equality and inequality specifications [45]. `con2vert` takes the inequality and equality constraints and returns the calculated vertices. Uniqueness of the solution to the inverse problem can be investigated using this function. Because the computational complexity of vertex enumeration increases dramatically for higher dimensional problems, the dimensionality of H_{PSF} is reduced by selecting a small subset of its columns.

First, the scenario where x_{true} is sparse is discussed. The sparse object, $x_{\text{true,sparse}}$ is shown in Figure 5.7, and is made up of only two non-zero pixels. As discussed before, due to its high level

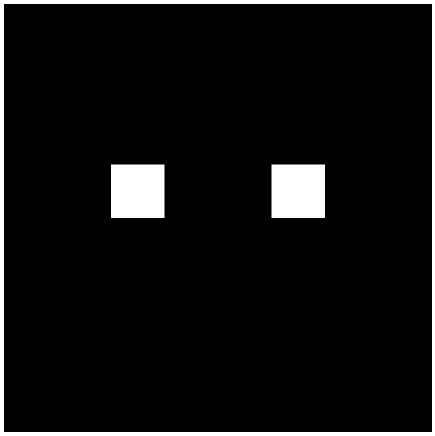


Figure 5.7: Both the sparse object and its true solution, $x_{\text{true,sparse}}$. This figure represents both the sparse object and its true solution, which are identical.

of sparsity, the inverse problem defined by H_{PSF} and this x_{true} should have a unique solution. For this object, using `linprog` to solve the following equation

$$H_{\text{PSF}} \times \delta x = \vec{0}$$
$$\delta x \geq -x_{\text{true,sparse}} \tag{5.11}$$

yields an all-zero vector. However, this does not guarantee the solution is indeed unique. Using `con2vert`, the polyhedron described by Equation (5.11) can be converted to vertex form. For this $x_{\text{true,sparse}}$, `con2vert` returns *two identical vertices, both all-zeros.*

Since the vertex form of the polyhedron defined by Equation (5.11) consists of two identical all-zero vectors, this means that the feasible region within which *any* admissible invisible object must reside is made up only of the all-zero vector. Or, to state this another way, only the trivial all-zero solution exists for the LP (Equation (5.9)) defined by H_{PSF} and $x_{\text{true,sparse}}$. As a result, for this object, the solution to the inverse problem is unique.

In the case of the second scenario, if the object is dense, for example, $x_{\text{true,dense}} = \vec{1}$, the the solution is not unique. Using `con2vert`, the polyhedron described by

$$H_{\text{PSF}} \times \delta x = \vec{0}$$
$$\delta x \geq -x_{\text{true,dense}} \tag{5.12}$$

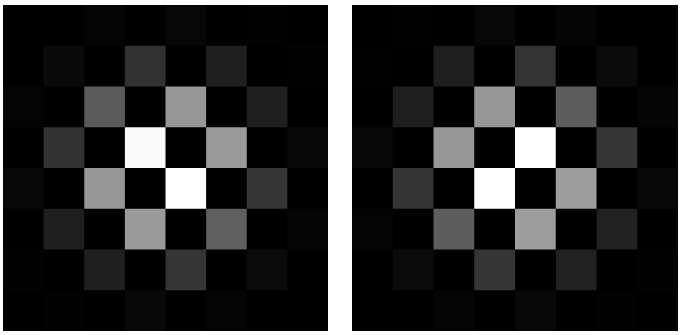


Figure 5.8: Returned two vertices for a dense object $x_{\text{true,dense}}$.

is converted to the vertex form. However, instead of returning the trivial solution, this time, the returned two vertices are not the trivial all-zero vectors (Figure 5.8). As a result, the solution to the inverse problem defined by H_{PSF} and x_{true} is not unique.

Summarizing the results in this section, one method to certify uniqueness of the solution to the inverse problem is shown. The conditions need to be satisfied by an admissible invisible object are Equation (5.9). This describes a polyhedron in n-dimensional space, where n is the number of columns of H_{PSF}. This polyhedron can be converted to vertex form, i.e., be specified by its vertices. For a given true solution x_{true}, a set of vertices can be found. If the vertices are all trivial solutions to Equation (5.9), i.e., the all-zero vector, the solution to the inverse problem is unique. If the vertices found contain a non-trivial solution to Equation (5.9), the solution is not unique.

5.5.3.2 Solving a pair of optimization problems

In the previous section, it is shown that the uniqueness of the solution to the inverse problem can be determined by converting the polyhedron described by Equation (5.9) to its vertex form, and observing whether the returned vertices are the trivial all-zero solution. While this method provides a definitive answer to the uniqueness of the solution to the inverse problem, it is impractical to use for larger problems. This is because, as the dimensionality of the problem increases, the number of vertices and, therefore, the computation time needed for `con2vert` to return the vertices increases

dramatically. As a concrete example, in Figure 5.7 and its associated example, the dimension of the object is 8-by-8 pixels, i.e., H_{PSF} only has 64 columns. In this small example, `con2vert` returns only two vertices. The reason this is valid is that the true object is artificially limited to be within an 8-by-8 region. In a more practical scenario, this is not possible due to 1) the size of the illumination spot; and 2) the labeling of the fluorescent sample. Because of this, to use `con2vert` in a more practical setting, more columns of H_{PSF} need to be included, and the computation becomes intractable.[4] As a result, an alternative method that can generate a similar uniqueness certification, while remaining computationally tractable for larger problems, is needed.

Here, inspired by [46], an alternative method to certify the uniqueness of the solution is shown. First, c is drawn once and used in the following two optimization problems, which can be solved to determine whether the solution to the inverse problem is unique:

$$\min_{\delta x} (\delta x)^T c, \text{such that } H_{\text{PSF}} \times \delta x = 0, \text{and } \delta x \geq -x_{\text{true}}, \tag{5.13}$$

and

$$\max_{\delta x} (\delta x)^T c, \text{such that } H_{\text{PSF}} \times \delta x = 0, \text{and } \delta x \geq -x_{\text{true}}. \tag{5.14}$$

In Equations (5.13) and (5.14), $c \sim \mathcal{N}(0, 1)$. The principle behind using these two equations is as follows. If the solution to the inverse problem defined by H_{PSF} and x_{true} is non-unique, an admissible invisible object that is not all-zero, δx_1 exists. If c is drawn randomly from a normal distribution, then the probability of $(\delta x_1)^T c = 0$ is very small, in fact with probability zero. As a result $(\delta x_1)^T c$ will be either a positive or a negative number, and can be found by one of either Equation (5.13) or Equation (5.14). In contrast, if the solution to the inverse problem is indeed unique, then the admissible invisible object in this case, δx_2, can only be the trivial all-zero solution. Since $\delta x_2 = \vec{0}$, regardless of c, Equation (5.13) and (5.14) will both return zero at their respective optimal point.

Therefore, the uniqueness of the solution to the inverse problem can be established by examining the solutions to Equations (5.13) and (5.14): if both solutions are the trivial all-zero vectors,

[4] In fact, `con2vert` will simply hang up when it is used to solve the vertex enumeration problem for a H_{PSF} with enough columns to cover a more realistic area.

the solution is unique; if one or both solutions are not the trivial all-zero vector, the solution is not unique. Two examples of this uniqueness certification method are provided below. The uniqueness of the solution is determined using Equations (5.13) and (5.14), and compared with that determined by `con2vert` in the previous section.

First the situation where the solution is unique is considered. Using the same sparse object as in Figure 5.7, $x_{\text{true,sparse}}$ is substituted in Equations (5.13) and (5.14). These optimization problems are then solved using MATLAB `fmincon`. Both of these problems return the trivial all-zero vector at its optimal point. Based on the previous arguments, this means that the solution to the inverse problem defined by H_{PSF} and $x_{\text{true,sparse}}$ is indeed unique, matching the conclusion drawn by `con2vert`.

In contrast, when the dense object that produces Figure 5.8 is used, i.e., $x_{\text{true,dense}} = \vec{1}$, Equations (5.13) and (5.14) both return non-trivial solutions at their respective optimal point (Figure 5.9). These non-trivial solutions are shown in Figure 5.9. It can be seen that the solutions returned by Equations (5.13) and (5.14) are very similar to the vertices returned by `con2vert` from section 5.5.3.1. This is expected and further illustrates the validity of the certification method proposed here. Since `con2vert` returns only two vertices for this dimensionally-reduced example, and because all admissible invisible objects are a convex combination of the vertices, all admissible invisible objects are situated along a line segment in higher-dimensional space, whose ends are the two vertices. As a result, it is expected that the solutions returned by Equations (5.13) and (5.14) should be scaled versions of the vertices returned by `con2vert`.

5.5.4 Summary

At this point, it is helpful to quickly sum up what this chapter has discussed. From the results and discussions presented so far, an interesting conclusion can be drawn: that the uniqueness of the solution to the inverse problem encountered in computational super-resolution microscopy is not a clear-cut issue. An apparent contradiction exists in published research on computational super-resolution microscopy and Fourier optics analysis. Specifically, Fourier optics analysis shows that

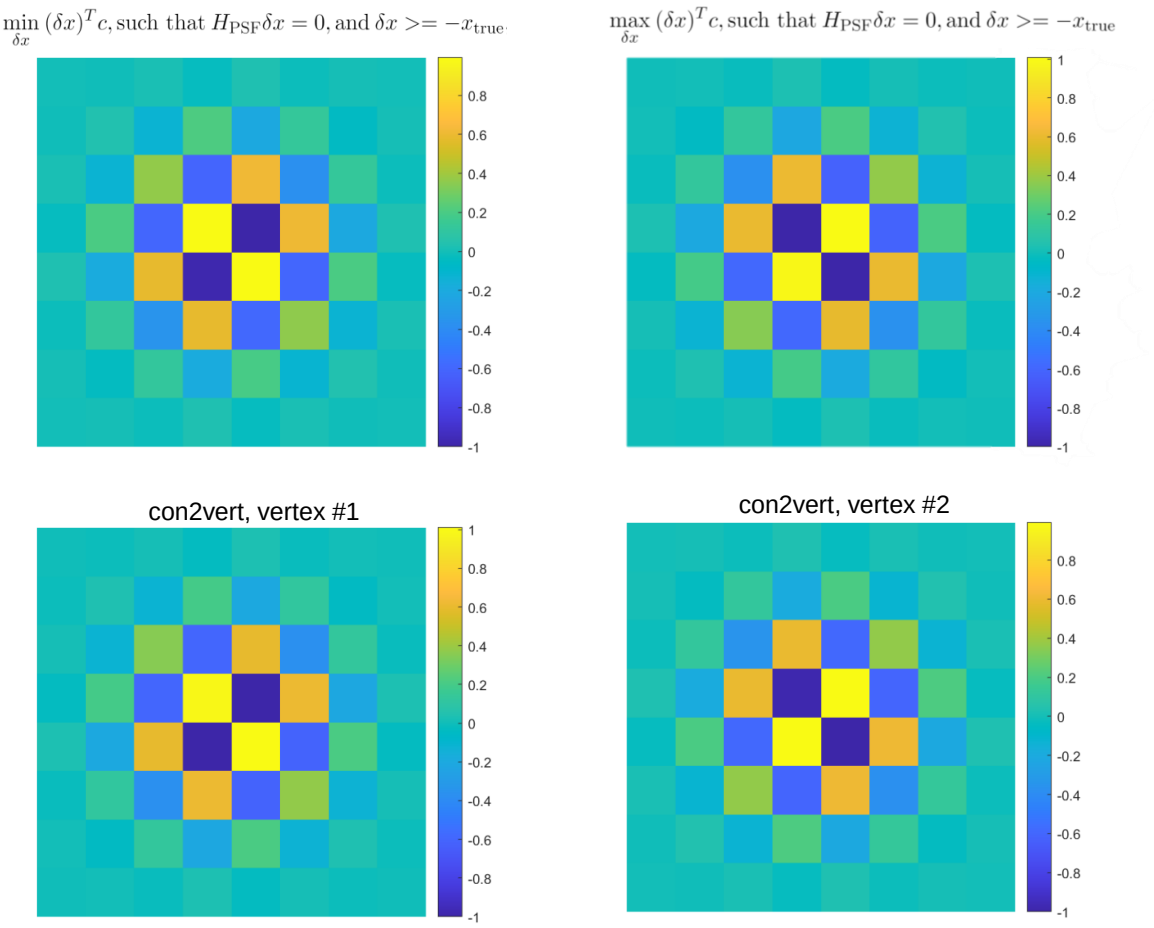


Figure 5.9: Non-trivial solutions returned by Equations (5.13) and (5.14) for the dense, uniform object. As a comparison, the bottom row shows the two vertices returned by `con2vert` in section 5.5.3.1. These two methods return extremely similar admissible invisible objects. This shows the validity of the alternative uniqueness certification methods proposed in section 5.5.3.2.

because the microscope sets some spatial frequency amplitudes to zero, the solution is not unique, even in noiseless recovery; however, it is generally accepted in existing literature on computational super-resolution microscopy that a unique solution does exist (for example, [41] and [15]).

In section 5.4, it is shown conceptually that this apparent contradiction can be reconciled by recognizing that: 1) there are invisible objects that can *potentially* lead to the existence of non-unique solutions; 2) for an invisible object to lead to non-unique solutions, it also needs to be *admissible* (i.e., satisfy Equation (5.9)); 3) whether an invisible object is admissible is determined by two factors, H_{PSF} and x_{true}, which implicitly depend on the discretization used for H_{PSF} (section 2.2.1); 4) for the same H_{PSF} (i.e., not changing the microscope objective), if x_{true} is sparse, it is possible to reject the entire class of invisible objects by applying a non-negativity constraint in solving the inverse problem, meaning the solution is unique; 5) if x_{true} is dense/non-sparse, there will often be an infinite number of admissible invisible objects, meaning the solution is non-unique.

Following this, section 5.5 proposes two methods of *certifying* the uniqueness of the solution. The first method, termed Method (1), shows how Equation (5.9), which describes a polyhedron, can be converted into its vertex form. This is the "gold-standard" method and provides a definitive answer to whether the solution is unique. Because all admissible invisible objects can be expressed as a convex combination of the vertices, by finding these vertices, *all* admissible invisible objects are effectively found. By inspecting the identified vertices, a certification of uniqueness can be generated. This method is not without its downsides: despite the definitive answer, it cannot be readily applied to larger, more practical problems. This is because when the dimensionality of the problem increases, the computational time for the vertex enumeration problem increases dramatically.

To bypass this practical constraint, a second uniqueness certification method (Method 2) is proposed. Its aim is to solve a pair of optimization problems, Equations (5.13) and (5.14), and observe the solutions. Due to the random vector $\vec{c}$ used in the cost functions of Equations (5.13) and (5.14), if there is a non-trivial and admissible invisible object, the probability of the cost function value being zero at the optimal point is zero, excluding round-off error. As a result, if

both Equations (5.13) and (5.14) return an all-zero vector, this means that the solution vector δx must be an all-zero vector (i.e., the only admissible invisible object is the trivial all-zero object). As a result, the solution is unique. If any of the solution vectors are not the trivial all-zero vector, the solution is non-unique.

5.6 Simulation of the effect of sparsity on uniqueness

In this section, Method (2) is used to examine the uniqueness of solution for two imaging scenarios: 1) the imaging of a sparse, two-dot object; and 2) the imaging of a less sparse, two-line object.

5.6.1 Simulation setup

The simulation scenario is similar to that used in section 3.2.2. Namely, H_{PSF} used here has $d_{\text{dict}} = d_{\text{camera}}/4$. To simplify the discussion, unlike in section 3.2.2, the same d_{dict} is also used to generate the images as well. Since this section focuses on the different levels of sparsity as a result of changing I_{obj}, different objects are proposed and shown as this section progresses. The uniqueness of the solution for an inverse problem defined by x_{true} and H_{PSF} is determined using Method (2) since it is much faster to run for a practical problem. As a concrete example, H_{PSF} used in section 3.2.2 has 441 columns, and covers a 21-by-21 pixel region, with each pixel having a size that is 1/4 of a camera pixel. In section 5.5.3.1, where Method (1) is used, to allow `con2vert` to finish executing within an acceptable amount of time, only the central 8-by-8 region is selected. Having first used the small-scale example to validate Method (2) against the results of Method (1), this section will subsequently use Method (2) to investigate the scenario where all 441 columns of H_{PSF} are used.

5.6.2 Sparse, two dot object

Here, a sparse, two-dot object example is shown. A real-world counterpart to this example object is the DNA Origami test sample used in [21]. As established in section 5.5.3, in this case,

the solution to the inverse problem is unique. Using the entire 441 columns of H_{PSF}, solving Equations (5.13) and (5.14) yields the following results shown in Table 5.1. Here, both equations

Optimization problem	Cost function value @ optimal point (average of 5 runs)
Equation (5.13), minimization	$1.3496^{-309} \approx 0$
Equation (5.14), maximization	$-9.4263^{-311} \approx 0$

Table 5.1: Cost function values for Equations (5.13) and (5.14) at their respective optimal points for a sparse, two-dot object. Five different random vectors are used, and their cost function values are averaged. Both Equations (5.13) and (5.14) return a cost function value of zero, meaning that the only admissible invisible object is the trivial, all-zero vector, and the solution is unique.

yield extremely small[5] cost function values at their respective optimal points. This mean that the only solutions to Equations (5.13) and (5.14), δx_1 and δx_2, are the trivial, all-zero vectors. Because of this, for the sparse, two-dot object, the solution is unique. To further illustrate this point, in Figure 5.10, δx_1 and δx_2 are shown. It can be seen that both of these solutions (i.e., admissible invisible objects) are uniform with extremely small values, i.e., both δx_1 and δx_2 are the trivial, all-zero solution.

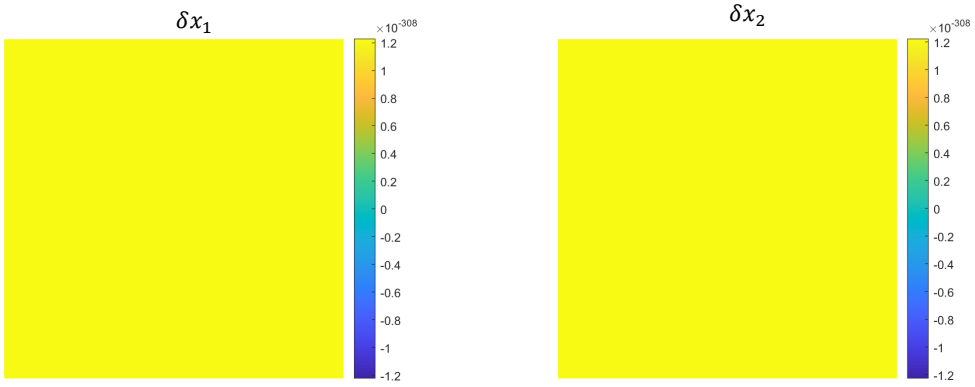


Figure 5.10: Solutions to Equations (5.13) and (5.14) for a sparse, two-dot object. Both of these solutions, δx_1 and δx_2 contain all zeros. This means that the only admissible invisible object for the inverse problem defined by x_{true} is the trivial, all-zero vector. As a result, the solution is unique.

[5] The fact that it is not exactly 0 is due to floating point roundoff error.

5.6.3 Less sparse, two line object

A two-dot object, such as the one used in the previous section, is extremely sparse. A slightly less sparse object is a two-line object, such as the Argolight test sample shown in Figure 3.6. Here, the uniqueness of the solution to the inverse problem when the object is a two-line object is investigated. Figure 5.11 shows the two-line object (comprised of two vertical white lines) and its image.

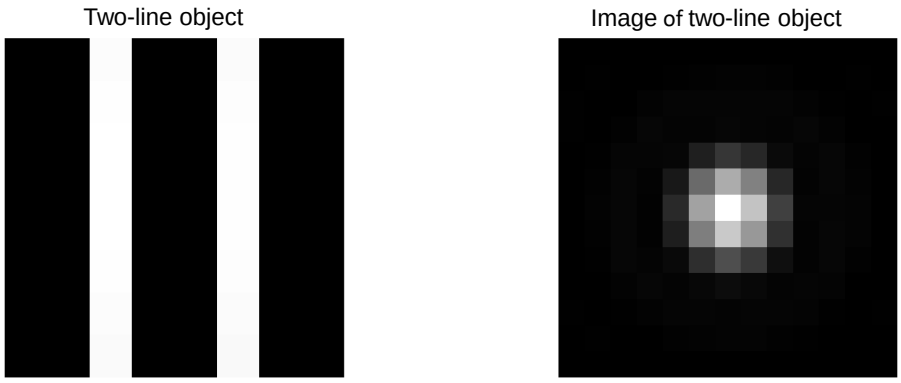


Figure 5.11: The less sparse, two-white-line object and its image. The two lines are separated by about 1/3 of the Rayleigh resolution distance, and as expected, in the image of the two-line object, the two lines are blurred together and cannot be distinguished.

Some interesting observations can be made regarding the uniqueness of solution in this scenario. Using Method (2), a table similar to Table 5.1 can be generated and it is shown in Table 5.2. Here, it can be seen that the solution is no longer unique for this two-line object because the

Optimization problem	Cost function value @ optimal point (average of 5 runs)
Equation (5.13), minimization	-7.8762
Equation (5.14), maximization	4.3288

Table 5.2: Cost function values for Equations (5.13) and (5.14) at their respective optimal points for a less sparse, two-line object. Five different random vector cs are used, and their cost function value are averaged. Both Equations (5.13) and (5.14) return a non-zero value, meaning there exist admissible invisible objects, and the solution is not unique.

cost function values of Equations (5.13) and (5.14) are no longer zero. In other words, there are non-zero solutions to Equation (5.9).

While this is not ideal, here it can be shown that the existence of non-unique solutions does not affect the achievement of super-resolution. This is because the admissible invisible objects are themselves mostly supported along the direction of the lines. Since these invisible objects are highly oscillatory, they affect the relative intensity values of the pixels in the recovered super-resolved image only. In other words, if a pixel in the true solution x_{true} is non-zero, then the existence of non-uniqueness solutions in this case will only affect how bright that pixel appears, not where that non-zero pixel is located (with respect to the true solution). As a result, in the processed super-resolved image for a two-line object, the locations of the lines, which is often the most important information the user wants to extract, will not be altered by the existence of the non-unique solution. As a result of non-unique solutions, the lines may appear "grainy", or textured, due to the existence of these highly oscillatory admissible invisible objects. However, this adverse effect is reduced even further in a real-world experiment, such as the one shown in Figure 3.6. This is because in such an experiment the illumination spot is scanned across the sample, and while for each scan position the solution might be textured, as the results are summed together to form the final image, these oscillations tend to average out due to their extremely high frequency.

To illustrate this, Figure 5.12 shows two admissible invisible objects, δx_1 and δx_2 for this two-line object. Compared with x_{true}, it can be seen that the admissible invisible objects have their non-zero pixels concentrated in the area where x_{true} is non-zero. This shows that the effect of non-unique solutions here is to introduce uncertainty in the relative pixel values in the processed image, but not to change where non-zero pixels are located with respect to the true solution.

To show how this can be further mitigated in a real-world experiment such as the one shown in Figure 3.6, Figure 5.13 shows four of the non-unique solutions for the inverse problem. These solutions are obtained by supplying different starting points to MATLAB `fmincon`. Because the solution in this case is non-unique, and because different starting points are used to solve the inverse problem, all four solutions are different from one another. However, because non-unique solutions

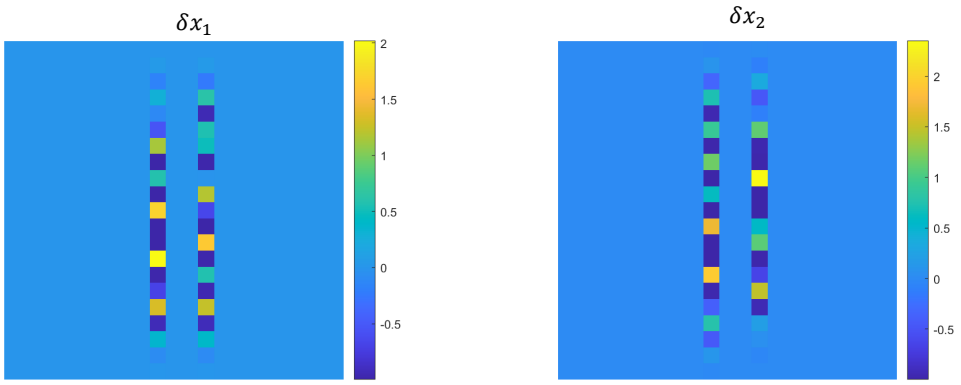


Figure 5.12: Solutions to Equations (5.13) and (5.14) for the less sparse, two-line object in Figure 5.11. Since the cost function values for Equations (5.13) and (5.14) are not zero at their respective optimal points, and by simple inspection of δx_1 and δx_2, it can be concluded that the solution is not unique. However, since the object is still somewhat sparse, the admissible invisible objects δx_1 and δx_2 mostly have their non-zero pixels concentrated in regions where x_{true} is non-zero. As a result, the effect of non-unique solutions is to introduce a relative intensity change along the direction of the line (see Figure 5.13), and not to change the location of the line.

only affect the relative intensities of the non-zero pixels, and not the locations of these pixels, these four solutions differ from one another only in the appearance of the lines, not the location of the lines. In a real-world experiment where an object like the one discussed in this subsection is imaged, because the scanning spot is overlapped between adjacent scanning positions, when the solutions from these scanning positions are summed to yield the final image, the high-frequency oscillations are "averaged out". This averaged image is also shown in Figure 5.13. To obtain this image, the four non-unique solutions are shifted by 4 pixels from one another before being summed together. The reason for shifting 4 pixels is that, in a real experiment, the scanning illumination spot is moved a distance equal to a single *camera* pixel between adjacent scanning positions. In the simulation conducted in this section $d_{\text{dict}} = \frac{d_{\text{camera}}}{4}$, as a result, a shift of 4 pixels is used. As can been seen in the averaged image, the effect of the non-unique solutions is greatly reduced, and the appearance of the lines are much less "grainy", or textured.

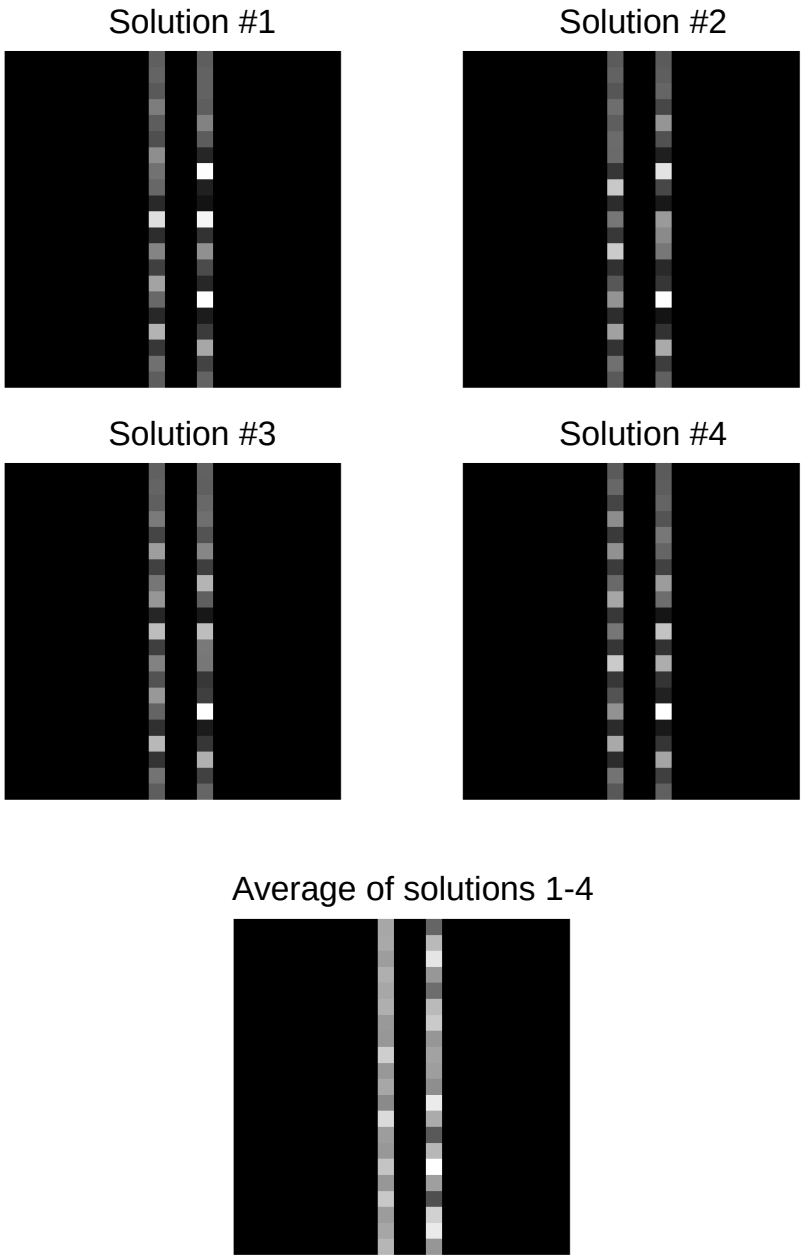


Figure 5.13: Non-unique solutions for the two-line object. As shown in Figure 5.12, the solution in this case is not unique. As a result, for four different starting points, four different solutions are obtained. Since the effect of non-unique solutions is to introduce a relative intensity variation along the direction of the lines, the four solutions shown appear "grainy", or textured. However, as established in Figure 5.12, the location of the lines are not affected by the non-unique solutions, and the textured appearance can be mitigated in a real-world experiment, where scanning and averaging occurs. This is shown in the bottom panel, where the the average of solutions 1 through 4 appears much less textured.

5.7 Connection to existing work

The discussion regarding sparsity and uniqueness presented in this chapter is not an exhaustive one. It should be viewed as one possible explanation best suited for this particular computational super-resolution microscopy technique, where a focused illumination spot is used to enhance object sparsity. The problem of *extrapolating* (i.e., super-resolving) data from band-limited measurements (both noisy and noiseless), where the underlying object is sparse, is ubiquitous in many fields of study; and throughout history, many different scenarios regarding the uniqueness of the problem have been discussed. The explanation in this chapter resonates strongly with published research. For example, in one of the earliest works on computational super-resolution from the 1960s [41], it is shown that

> "For objects of finite angular dimensions, knowledge of the spatial frequency spectrum within the passband of any imaging system implies knowledge of the spatial frequency spectrum over the entire frequency domain, and hence implies complete knowledge of the object."

In relation to the super-resolution technique currently under discussion, "finite angular dimensions" is equivalent to sparsity, or limited span in the spatial domain. If this condition is true, [41] again shows the following can be proven:

> "Two distinctly different objects of finite angular size cannot have identical images, so that the ambiguous image does not exist for realizable objects."

Similar to what is shown in this chapter, [41] shows that, while it is true that the solution to the inverse problem seems to be non-unique, these non-unique solutions can be excluded in the solving of the inverse problem under certain conditions. Namely, an incoherent (i.e., non-negativity) contrast mechanism and sparsity (i.e., a spatially-limited object).

This is also widely known in the field of astronomy, where the object can be sparse (e.g., a collection of distant stars, which can be viewed as point sources) or non-sparse (e.g., the surface of a planet or the Sun). For example, such a super-resolution phenomenon has been explored in works such as [47] and [48]. In an especially compelling example [19], it is found that

> "...super-resolution, i.e. resolution beyond the telescopic diffraction limit, is possible for point-like objects like stars... However, superresolution cannot be achieved for extended objects which are not band-limited."

The drastically different results obtained for sparse and dense objects, despite the fact that the same processing scheme is used for both types of objects, showcases the important role played by sparsity, and adds further support for the super-resolution microscopy approach discussed in this thesis.

What is missing so far from the published research is how to reconcile the apparent conflict between the published results and the irrefutable fact that, due to the limited microscope passband f_c, there must be non-unique solutions. This apparent conflict is successfully reconciled in this chapter by noting that, as the object's sparsity level increases, the adverse effect of non-unique solutions is reduced; and that eventually, if the object is sufficiently sparse, a unique solution exists.

Since whether the solution is unique depends on the true object, x_{true}, a universal statement regarding the uniqueness of the solution cannot be made without specific knowledge of x_{true}. To evaluate the uniqueness of the solution, two certification methods are proposed in this chapter. Converting to vertex form, i.e., Method (1), offers a definitive answer, but it is too slow for practical applications. Solving a pair of optimization problems, i.e., Method (2), provides a faster alternative. Both methods are shown to be powerful tools in assessing the uniqueness of the solution for computational super-resolution microscopy. In addition, the results in this chapter further support how the object sparsity relates to the uniqueness of the solution.

Chapter 6

Information transfer in a microscope

So far, this thesis has investigated two important aspects of this emerging super-resolution microscopy technique: 1) what is the effect of the processing scheme on the achievable super-resolution? and 2) what is the relationship between the object sparsity and the uniqueness of the solution? From the investigation conducted up to this point, the following can be concluded: 1) the achievable super-resolution is affected by whether the processing scheme is appropriate for the dominant noise model in the acquired noisy images, and whether it takes advantage of the prior information of sparsity; and 2) the uniqueness of the solution in this microscopy technique is not a clear-cut issue, and it is related to the level of sparsity in the object.

However, one major question still remains: despite the clear, comprehensive theoretical and experimental results showing not only the validity, but also the significant potential of computationally achieving super-resolution, such resolution enhancement nonetheless cannot be predicted using a Fourier optics based analysis. This is plainly clear from super-resolution results obtained from *diffraction-limited* data, not only from earlier chapters of this thesis, but also from a wide array of works on single molecule localization microscope (SMLM) [5, 6, 29]. This contradiction between the Fourier optics analysis and the observed simulation and experimental results is the focus of this chapter. It is shown that while Fourier optics analysis is an invaluable tool in analyzing the *image formation* in a microscope, it does not adequately model the *information transfer* in a microscope if there is an additional post-processing step. As a result, Fourier optics fails to predict the computational super-resolution phenomena obtained using diffraction-limited images

and post-processing, observed in earlier chapters of this thesis, as well as in the field of SMLM.

Before a comprehensive discussion is presented, it should be noted that despite this deficiency, Fourier-based analysis can explain certain super-resolution phenomena, such as structured illumination microscopy (SIM) and image scanning microscopy (ISM). For example, SIM is developed using Fourier analysis as the guiding principle and can be explained using it [8, 13]. In this chapter, it is shown that, while a Fourier optics based analysis can account for super-resolution achieved with a structured illumination profile, such as SIM and ISM, it cannot incorporate a scenario where the object exhibits sparsity. Because object sparsity is not included in this analysis, it fails to predict the super-resolution achieved by the combined effect of sparsity in the imaged object and computational post-processing, such as the technique covered in this thesis and SMLM. To address this, expanding upon work first published in [16], an alternative analysis method is proposed here. This method is based on singular value decomposition (SVD) and it is able to successfully predict the computational super-resolution phenomena. A discussion of the reason why the SVD analysis is able to provide a more complete description of the information transfer in a microscope (as compared to Fourier optics) is provided, and it is used to analyze some simple scenarios in the noiseless setting.

6.1 Fourier optics analysis and eigenvector decomposition

In a Fourier optics based analysis of a microscope, the input object is Fourier transformed. Here, the Fourier transform functions as a decomposition operation of the input object into a series of spatial frequencies [49]. The microscope is modeled by a low-pass filter that modifies the amplitudes of these spatial frequencies but leaves the spatial frequencies themselves unchanged. Finally, the resultant image spectrum is inverse transformed to yield the output image. In other words, under this theoretical framework, the microscope modifies how these spatial frequencies are assembled to yield a final image by setting the *amplitudes* of spatial frequencies, without modifying the spatial frequencies themselves. In this process, which views the microscope imaging as a linear operation, the spatial frequencies, or Fourier basis, function in the same way as eigenvectors,

where the effect of the microscope on a certain spatial frequency (an eigenvector) is to modify its magnitude only. In other words: $H_{\text{PSF}} \times \vec{v} = \lambda\vec{v}$, where $\vec{v}$ is a vector containing a spatial frequency(represented as a sinusoidal), and λ is the associated eigenvalue, or the amplitude of that spatial frequency after the microscope's low-pass filtering operation.[1] As a concrete example, if $\vec{v}$ is the DC Fourier component, then $\lambda = 1$; If $\vec{v}$ is a Fourier component outside of the microscope passband, f_c, then $\lambda = 0$; and if $\vec{v}$ is a Fourier component between DC and f_c, then $\lambda \in (0, 1)$, with its exact value determined by the modulation transfer function (MTF) of the microscope.

Besides viewing the spatial frequencies as eigenvectors in this analysis, another viewpoint originating from Fourier optics is that a spatial frequency can be seen as a plane wave propagating along a certain direction [49]. For such a plane wave, depending on the direction of travel, its projection onto a plane that is perpendicular to the optical axis will be sinusoidals of different orientations and periods, i.e., Fourier components of different orientations and periods. In other words, the eigenvectors used in a Fourier optics analysis are plane waves traveling in different directions. Here lies a major departure of Fourier optics analysis from physical reality: plane waves are the eigenvectors of *free-space* propagation, not propagation through a space-limited and band-limited optical instrument such as a microscope. In other words, after a plane wave is propagated, the only change it can undergo is a change in amplitude. As a result, for the use of plane waves to be valid in this analysis, the microscope has to take in a plane wave, and output the same plane wave with *only* a modified amplitude. This is clearly not in agreement with physical reality: in addition to the low-pass filtering operation, the microscope performs a truncation operation stemming from either the fact that a microscope will always have a limited field of view (FOV), or the fact that, in many microscope setups such as the one discussed in this thesis, an illumination spot is used to further restrict the FOV. This is in addition to the fact that the plane wave itself is an idealistic abstraction and has no real physical counterpart. This means that, while plane waves as a basis set are suitable for modeling the *image formation* of a microscope, when it comes to modeling the information transfer through a microscope, plane waves show deficiencies.

[1] i.e., the magnitude of the modulation transfer function.

6.1.1 Eigenvectors for a microscope

The series of arguments in this subsection are inspired by the paper of Bertero et al. [16]. At the time of this paper, it had become widely recognized that when the information transfer ability of a band-limited imaging system is analyzed, one needs to employ the proper eigenvectors specific to that system. For a microscope, the input is first truncated, and then low-pass filtered. For such a system, its eigenvectors are the prolate spheroidal wave functions, which are a set of band-limited functions that are the eigenvectors of a time-limiting operation (or cropping in microscope imaging), followed by a low-pass filtering operation [50, 51]. In this analysis, these wave functions can be regarded as "elements" of information, which remain unchanged except for their amplitudes after being imaged. This is because they are the true eigenvectors of the microscope. In other words, if u_k describes the k-th prolate spheroidal wave function, and λ_k describes its associated eigenvalue, the following is true:

$$H_{\mathrm{PSF}} u_k = \lambda_k u_k. \tag{6.1}$$

This equation illustrates the same approach as Fourier optics analysis, but uses the correct eigenvectors. The value of λ_k determines how efficiently the k-th prolate spheroidal wave function is transmitted through the microscope. The larger the λ_k, the more efficient is the transmission. By studying how the microscope modifies the eigenvalue spectrum, one can determine the information transfer capacity of that microscope.

6.1.2 Eigenvector decomposition is no longer applicable when the object is reduced in size

Using the eigenvector analysis approach with the correct eigenvector for a microscope, Bertero et al. first arrived at the same conclusion as when Fourier optics analysis is used: namely, for a general widefield microscope, the resolution it can achieve, even after processing from diffraction-limited data, is on the scale of the diffraction limit itself. In other words, for a widefield microscope where a *uniform* illumination profile is used, no meaningful resolution improvement can be obtained

from post-processing.[2] Up until this point in [16], all the analysis is done using an eigenvector approach, meaning that the elements of information remain *unchanged* between the input and output plane.[3] In [16], the authors argue that as the size of the object is reduced, this eigenfunction-based analysis approach is no longer valid, even if the correct eigenvectors are used. This is because, after accounting for the effect of diffraction for an extended object, the difference between the input and output will be smaller than that for a space-limited object. This is graphically illustrated in Figure 6.1. For a dot-like object that is smaller or on the scale of the diffraction limit, its image will be vastly different from itself. However, for an extended object that is much larger than the diffraction limit, its image will be visually similar to itself, save for a blur.

As a result of this, when the object is reduced in size — for example, with the use of a focused illumination spot and, in many cases, with the chemically-specific labeling in fluorescence microscopy — an alternative analysis approach that incorporates the effect of diffraction on the basis set used in the analysis should be employed to model the information transfer of the microscope. Authors of [16] argue that singular value decomposition (SVD) could be such a method. Some interesting theoretical results are derived in [16], which suggest that the achievable resolution, or the amount of information transmitted through a microscope, is related to the signal to noise ratio and size of the object. Later, the same analysis was extended to incoherent imaging [17], and more importantly, illumination with a non-uniform profile. In [18], the same SVD analysis was applied to a *widefield microscope with a focused spot illumination* setup, the same as the experimental setup used by the microscopy technique discussed in this thesis. It was shown that such a setup can achieve a maximum of two-fold resolution improvement beyond the diffraction limit.

Obviously, the two-fold resolution improvement is not in agreement with the approximately 4 times improvement that is observed and verified in the previous chapter. The reason is that the analysis conducted in [18] only accounted for the sparsity induced by the focused spot illumination profile, and not the implicit sparsity that is a result of the chemically-specific labeling performed

[2] Note that this is a remarkably similar result to the results presented in [19].

[3] In fact, because of this, authors of [16] refer to this particular analysis method as "geometric optics", since theses eigenvectors, or "information pieces", do not undergo diffraction.

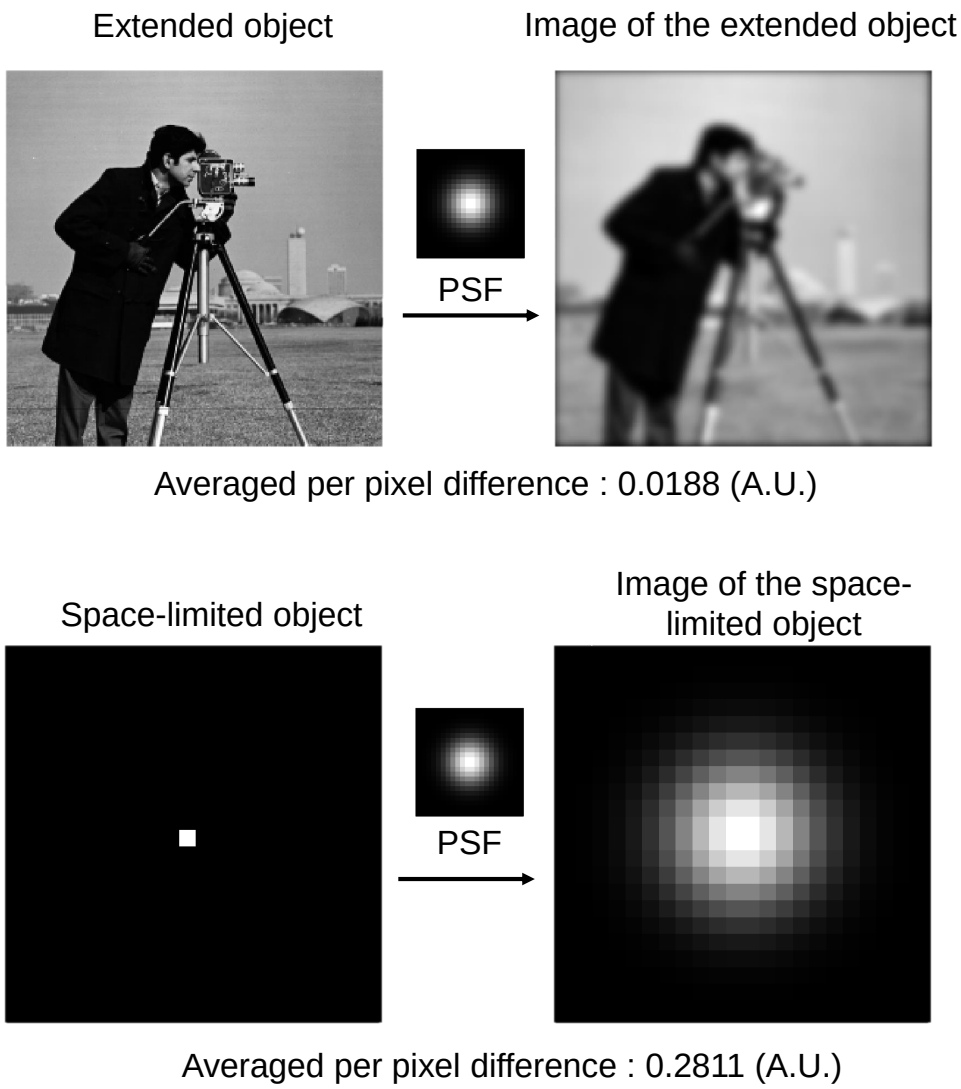


Figure 6.1: Difference between the object and its image for an extended object and a space-limited object. Top row shows an extended object (the cameraman test image) and its image. For an extended object, its image will be visually similar to the original object, save for a blur. However, for the space-limited object shown in the bottom row, its image is very different from the original object. This can be seen from the averaged per pixel difference metric. For each object, the original object and its image are both normalized to have a maximum value of 1. A Gaussian filter with $\sigma^2 = 3$ is used as the PSF.

in fluorescence microscopy. As a result, while the SVD-based analysis was valid, it was only able to predict a resolution improvement by a factor of 2. In [18], this was explicitly mentioned:

> "...greater increases in resolution are possible...if an object support of this dimension is defined by its self-luminosity or fluorescence..."

Regrettably, this important concept was left unexplored. As a result, it seemed at the time that the resolution enhancement afforded by post-processing of diffraction-limited images would be limited to a maximum of two-fold beyond the diffraction limit. This thesis makes two important modifications on these early results: 1) it explicitly considers the sparsity resulting from the chemically-specific labeling in fluorescence microscopy; and 2) it uses SVD to analyze the information transfer ability of a microscope, instead of using SVD to recover the super-resolved information, as is done in earlier work such as [16, 17, 18]. By making these modifications, SVD is now shown (in this thesis) to be a powerful tool in studying information transfer in a widefield microscope when there is sparsity present; and, furthermore, SVD is able to successfully predict the improved resolving power beyond the diffraction limit in such a microscope. The details of this analysis are shown in the remainder of this chapter.

6.2 Incorporating illumination profile

The fundamental idea in applying SVD instead of eigenvector analysis in determining the information transfer in an optical microscope is that SVD allows for object space and image space to each have their own set of basis functions instead of sharing a common one. As a result, the problem encountered by the eigenvector analysis, described in the previous section, is bypassed. This is what originally motivated the use of SVD in [16]. Under this analysis framework, similar to Equation (2.2), the image formation is modeled as:

$$I_{\text{noiseless}} = H_{\text{PSF}} \times [I_{\text{ill}} \cdot I_{\text{obj}}] \,. \tag{6.2}$$

Here, $I_{\text{ill}} \cdot I_{\text{obj}}$, which represents the illuminated object, can be represented using a single vector $\vec{p} = I_{\text{ill}} \cdot I_{\text{obj}}$, where $\vec{p}$ is a called a "profile function". It can be seen that the microscope being analyzed is defined by a combination of its PSF (H_{PSF}) and the profile function $\vec{p}$. As a result, the PSF describes the information transfer ability of a microscope in terms of its optical characteristics, such

as the numerical aperture of the objective, and the profile function describes how that information transfer ability is *modified* by either the illumination or object labeling. Because of this modification of information transfer ability, following notation from [18], the modified microscope system can be represented as:

$$H_{\text{modified}} = H_{\text{PSF}} \cdot \text{repmat}(\vec{p}). \tag{6.3}$$

In Equation (6.3), which is a discrete and generalized version of Equation (1.1) of [18], $\vec{p}$ is a vector containing the profile function, $\cdot$ the element-wise multiplication (instead of the matrix multiplication used in Equation (6.2)), and repmat a special operation where vector $\vec{p}$ is expanded to have the same dimensions as H_{modified} and H_{PSF}. The effect of Equation (6.3) is to scale the intensity of each PSF stored in the columns of H_{PSF} by the corresponding value in $\vec{p}$. As a concrete example, if H_{PSF} is a matrix of size m-by-n, then $\vec{p}$ will be a n-by-1 vector. repmat will return a matrix of size m-by-n, where $\vec{p}$ is copied m times. In other words, the result of repmat($\vec{p}$) is a matrix of size m-by-n, where each column contains a constant that is equal to the corresponding value in $\vec{p}$. As the result of this series of operations, the resultant H_{modified} from Equation (6.3) will contain PSFs shifted to different locations, with their intensities scaled by $\vec{p}$. This H_{modified} now represents the modified microscope system when used to image $\vec{p}$.

By adopting different $\vec{p}$, H_{modified} can describe a multitude of microscope systems with different illumination setups, as well as object configurations. Setting aside the effect of the object for now (i.e., $I_{\text{obj}} = \vec{1}$; the case where $I_{\text{obj}} \neq \vec{1}$ is discussed in a later section), the effect of the illumination profile can be studied. For example, for a conventional widefield microscope system, where a uniform illumination profile is used, $\vec{p}_{\text{widefield}} = \vec{1}$.

After H_{modified} is calculated based on the type of simulated microscope system, a SVD operation is then performed on H_{modified}. Formally, this operation is defined by the following equation:

$$\text{SVD}(H_{\text{modified}}) = U\Sigma V^*, \tag{6.4}$$

where, H_{modified} is an m-by-n matrix, U is an m-by-m matrix, Σ is an m-by-n matrix, and V^* is an n-by-n matrix. The SVD operation decomposes the original H_{modified} into two components: the

image-space matrix U, and the object-space V. Specifically, U and V contain basis functions (or singular vectors) that are orthogonal to each other and thus can be used to represent an arbitrary function in image and object space respectively. How image space and object space are related to each other is determined by entries in matrix Σ, whose diagonal contains the singular values of each singular vector. Formally, for the k-th singular vector in the object space v_k, the following is true:

$$H_{\text{modified}} \times v_k = \sigma_k u_k, \tag{6.5}$$

where v_k and u_k are the k-th singular vectors in object and image space respectively, and σ_k is the k-th diagonal element of Σ.

By examining Equation (6.5), it is clear that in contrast to a Fourier optics based analysis, SVD allows the information elements, in this case the singular vectors, to be different between image and object space, with an additional scaling operation described by σ_k. Conceptually, this can be thought of as an imaging operation where the image of an object-space singular vector v_k is a scaled image-space singular vector, $\sigma_k u_k$. This capability of SVD, when combined with a proper modeling of the profile functions, $\vec{p}$, allows the super-resolution phenomenon discussed in this thesis to be adequately modeled.

6.3 Using SVD to analyze an imaging system

In [16] and the works that follow it, SVD was used as the processing scheme to recover the super-resolved image. From the analysis in the previous section of this thesis, it can be seen that SVD is a much more suitable tool for the characterization of information transfer in a microscope, rather than using it as the processing scheme for recovering the super-resolved image. This is because, to be used effectively as a processing scheme, the singular vectors in both image and object space need to be known, which implies the *exact* values of the profile function $\vec{p}$ must be known *a priori*. This is only possible in a numerical simulation.[4] However, in a numerical simulation, it is

[4] In other words, if the exact knowledge of the profile function is available, this means that the original object is known exactly as well.

possible to use SVD to analyze the *information transfer* ability of a microscope in a similar fashion as Fourier optics. To do this, a profile function, based on the exact microscope system setup, is first generated. Next, a corresponding H_{modified} is calculated based on Equation (6.3), and SVD is performed on the resultant H_{modified}. This yields a set of singular vectors, v_k, in the object space. These vectors play a similar role to that played by the Fourier basis (spatial frequencies) in a Fourier optics analysis, and represent a "piece" of information under this analysis framework. The efficiency with which these pieces of information are transmitted through the microscope is described by the singular values, σ_k. To provide a framework for evaluating the amount of information transmitted, these attenuated object-space singular vectors can be Fourier transformed, and their spectra energy summed, to evaluate how much information is transmitted using the familiar Fourier basis. Doing so yields an alternative modulation transfer function (MTF) for comparing with Fourier analysis. This SVD analysis approach is graphically illustrated in Figure 6.2.

6.4 Validating SVD analysis of imaging techniques

Here, using the methodology described in Figure 6.2, SVD is used to analyze two microscopy techniques that can also be analyzed using Fourier optics, and is shown to generate results that are in agreement with those obtained using Fourier optics. The reason behind this effort is that, since this SVD-based analysis is capable of incorporating the consideration of sparsity in the imaged object, it should also accommodate a scenario where there is no such sparsity consideration (i.e., $I_{\text{obj}} = \vec{1}$, and $\vec{p}$ consists only of the illumination profile). Here, using numerical simulation, SVD is used to analyze the information transfer ability of two microscope setups where sparsity is not explicitly considered: 1) conventional widefield, and 2) image scanning microscopy (ISM). The reason for the selection of these microscopy techniques is two-fold. Firstly, it demonstrates the general nature of a SVD analysis for information transfer, in that it can be used to analyze an arbitrary microscope system, provided that the profile function $\vec{p}$ is properly modeled. Secondly, and more importantly, these two microscopy systems are readily analyzed using Fourier optics techniques. Therefore, by comparing the results obtained using Fourier optics and SVD, it confirms the validity of this

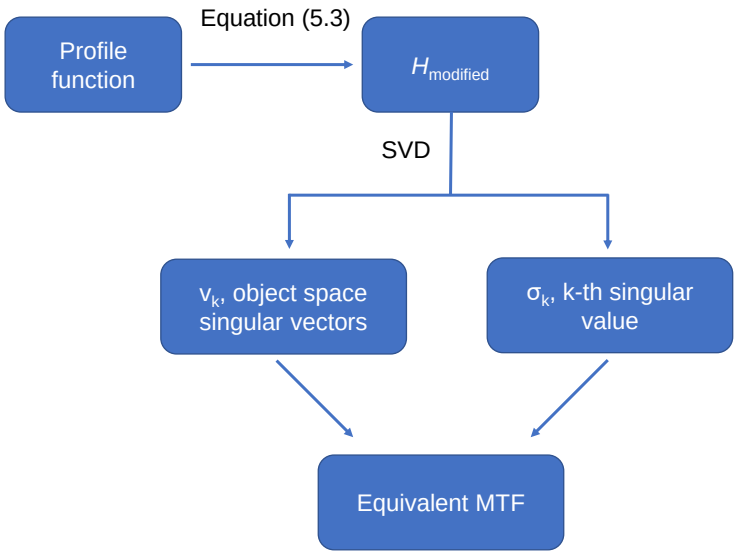


Figure 6.2: Calculation of an equivalent MTF. First, based on the illumination setup and the object structure, a profile function is generated. Then an H_{modified} matrix can be calculated based on the profile function. H_{modified} is then SVD-ed to yield the object space singular vectors, v_k, and their corresponding singular values, σ_k. A singular vector, v_k, is attenuated by a factor corresponding to its associated singular value, σ_k, in this process, and the attenuated singular vectors' energy is then summed to yield the final equivalent MTF.

alternative analysis approach.

6.4.1 Conventional widefield microscopy

In a conventional widefield microscope, a uniform illumination profile is applied. In other words, $\vec{p}_{\text{widefield}} = \vec{1}$. Here, an NA1.4 objective is used to simulate imaging at a wavelength of 520nm. For such a system, diffraction-limited resolution can be estimated using the Rayleigh criterion, which leads to a resolution of $0.61\frac{\lambda}{\text{NA}} \approx 230\text{nm}$. This system can also be analyzed by plotting the MTF, which is given by taking the amplitude of the optical transfer function, or OTF:

$$\text{MTF} = |\text{OTF}|. \tag{6.6}$$

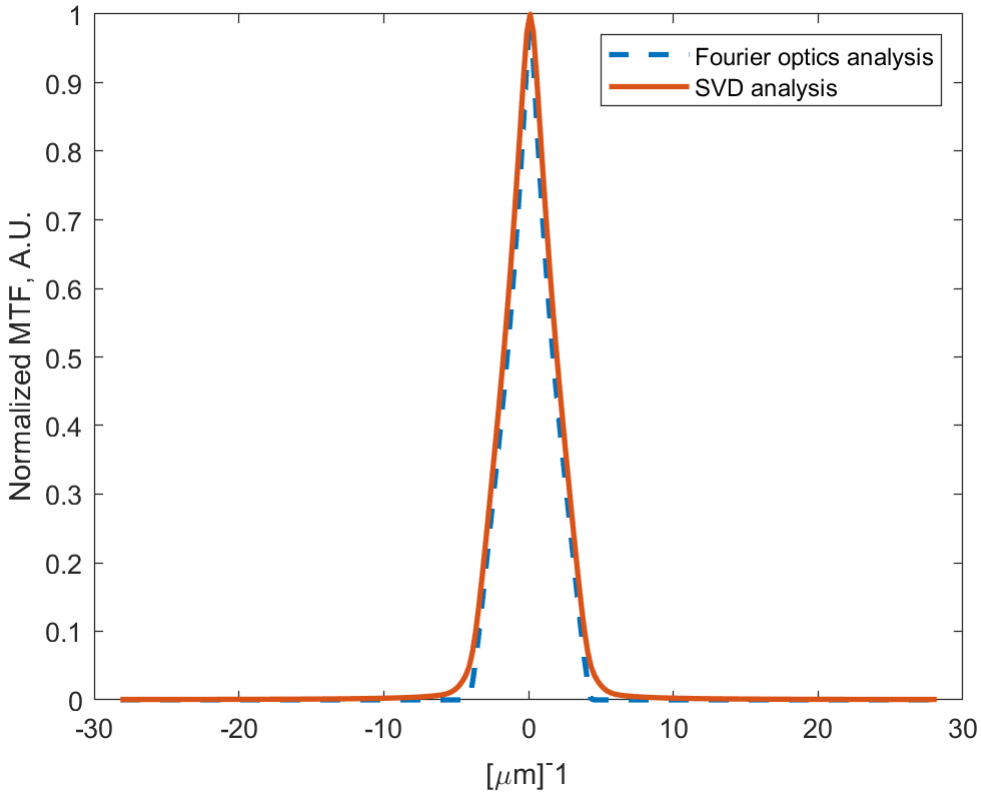


Figure 6.3: MTFs obtained using Fourier optics and SVD for a widefield microscope. The equivalent MTF generated from the SVD analysis closely resembles the MTF generated from Fourier optics analysis.

In Equation (6.6), the OTF is obtained by taking the Fourier transform of the PSF of the system. In Figure 6.3, the MTF of a conventional widefield microscope is plotted (i.e., Fourier optics analysis, blue dashed line). The resolution can be determined by finding the point where the normalized amplitude of the MTF decreases to zero. From the plot, the point at which this occurs is approximately 4.1 μm^{-1}, which is $\frac{1}{4.1\mu\text{m}^{-1}} \approx 240\text{nm}$, and is in agreement with the Rayleigh criterion.

An alternative analysis of information transfer can be conducted using SVD. Following the procedures outline in Figure 6.2, an alternative MTF is obtained, and also plotted in Figure 6.3. It can be seen that the MTFs obtained using Fourier optics and SVD closely resemble each other.

6.4.2 Image scanning microscopy (ISM)

In image scanning microscopy (ISM), a focused illumination spot is used as the illumination profile [20, 52]. This leads to a roughly two-fold improvement in resolution after appropriate processing. The reason that this particular technique is chosen here is that an effective PSF can be derived for ISM, which can then be used to evaluate its information transfer ability. The effective PSF of ISM can be obtained by first convolving the detection PSF and illumination PSF (which are at two wavelengths), and then rescaling the result to half its size [20]. Using this effective PSF, an MTF is obtained and shown in Figure 6.4. From the plot, as expected, the point at which the normalized amplitude of the MTF decreases to zero is approximately 8.3 μm^{-1}, which is approximately two times greater than that of a conventional widefield microscope.

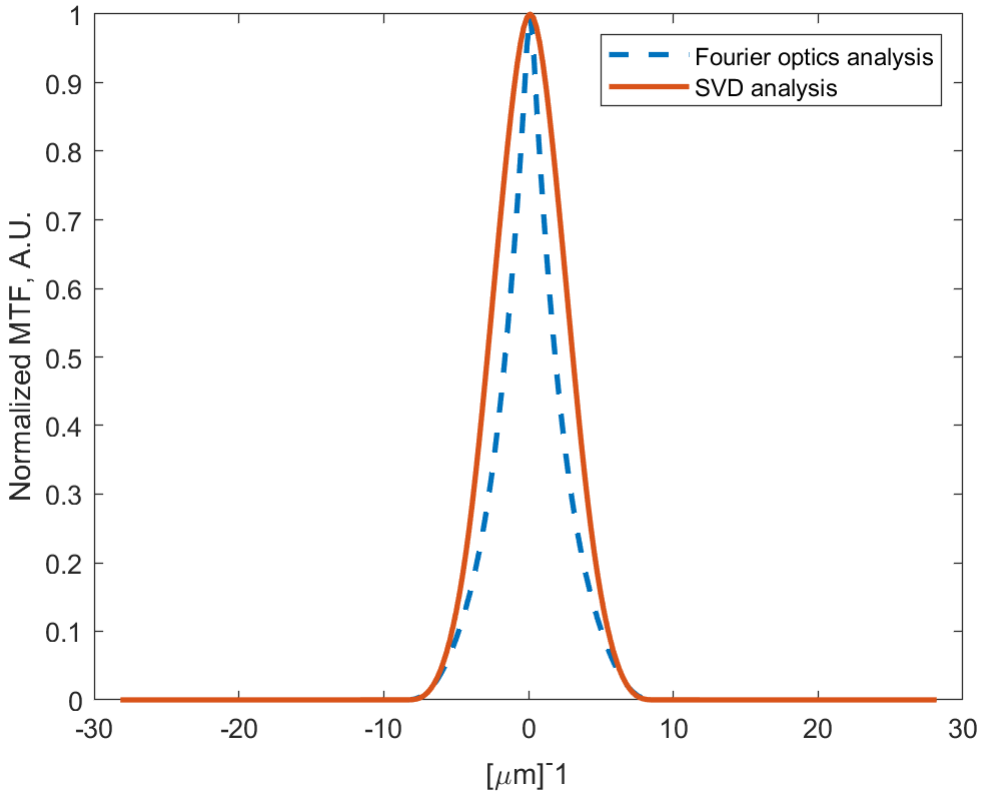


Figure 6.4: MTFs obtained using Fourier optics and SVD for an image scanning microscope (ISM). The equivalent MTF generated from the SVD analysis closely resembles the MTF generated from Fourier optics analysis.

Using SVD analysis, an MTF for ISM can be obtained for the ISM system. This is also plotted in Figure 6.4. Again, it can be seen that the MTFs obtained using Fourier optics and SVD closely resemble each other and both show an MTF cut-off that is approximately two times higher than that of the conventional widefield microscope.

6.5 Summary and discussion

These results show that this alternative SVD-based analysis agrees with Fourier optics based analysis when the object being imaged does not exhibit sparsity, i.e., $I_{\text{object}} = \vec{1}$. Namely, both analysis approaches show that, for a conventional widefield microscope, its information transfer ability is close to the familiar diffraction limit, and, for an image scanning microscope, its information transfer ability is approximately twice that of the conventional widefield microscope.

However, as is shown in the next section, if the object exhibits some degree of sparsity, Fourier optics analysis no longer provides the complete picture of information transfer. Before such a result is shown, this section quickly summarizes the similarities between SVD and Fourier optics analysis, and provides an example from a published paper illustrating why Fourier optics is insufficient in describing information transfer when there is sparsity present.

6.5.1 Similarities between SVD and Fourier optics

Here, a brief summary on the similarities between the SVD analysis and the Fourier analysis is provided. They are similar in a sense that both methods first decompose the object into a series of basis functions, which are transferred through the microscope with varying degrees of efficiency, before being re-combined to yield the final image. In both of these analysis approaches, information is represented in "pieces". In the familiar Fourier optics approach, these are the spatial frequencies, or sinusoidals of varying frequencies. A microscope system can pass some spatial frequencies through, and reject others. If more high spatial frequencies are allowed to pass through the microscope, more information about the fine structures is passed through, resulting in a higher resolution.

Similarly, in the SVD-based analysis, these information "pieces" are the singular vectors in object space, v_k. The reason the object space singular vectors are used for the analysis is that, when SVD is used as an analysis tool rather than a recovery tool, what is of interest is how much information about the original object can pass through the microscope. Hence, using the object space singular vectors is more appropriate for this purpose. Similar to the spatial frequencies used in Fourier optics analysis, these singular vectors are also oscillatory with varying degrees of "frequencies". As an example, in Figure 6.5, the first nine singular vectors, $v_{1\text{-}9}$, are shown for the conventional widefield microscope. These singular vectors are obtained by performing SVD on the H_{modified} appropriate for a widefield microscope. For each singular vector, it is reshaped into a two-dimensional image and plotted. It can be seen that as the indices of singular vectors increase, they become more oscillatory, and as a result, their Fourier transforms contain more magnitude in the higher spatial frequency regions.[5] This means that these higher-index singular vectors carry more information about the finer details of the object

Similar to the passband concept used in Fourier optics analysis, in the SVD-based analysis, as the information is transferred from object to image space, these information pieces are modified *in magnitude* by their corresponding singular values, and *in form* by being transformed into image space singular vectors. Hence, the singular values in a SVD analysis are similar in function to the role played by the MTF in a Fourier optics analysis. Therefore, the approach described in section 6.3 takes the total energy contained by the Fourier transform of magnitude-modified object singular vectors as a representation of the information transfer ability of a microscope.

6.5.2 Differences

Here, a brief summary on the differences between the SVD analysis and the Fourier analysis is provided. The most notable difference is that the information "pieces" used in the SVD analysis are more suitable for a discussion of information transfer, whereas the information "pieces" used in the Fourier optics analysis are more suitable for a discussion of image *formation*. A second

[5] A similar conclusion is also drawn in [16].

difference between these methods is that the SVD analysis allows the incorporation of the sparsity that is the result of chemically-specific labeling into the consideration of information transfer.

In a Fourier optics modeling of a microscope, as previously stated, the effect of the modeled microscope is to reject some of the spatial frequencies. This is valid and correctly models the image formation in a microscope because, due to the limited numerical aperture, for a point source that is emitting light in all directions, only a subset of emitted light (up to the light traveling at an angle defined by the numerical aperture) can be accepted by the microscope objective. However, while this correctly models the images formed by a microscope, the information that is embedded in the acquired image is not described by a Fourier optics analysis, as shown throughout this thesis.

It should be noted that this thesis is not the only work that describes this deficiency of

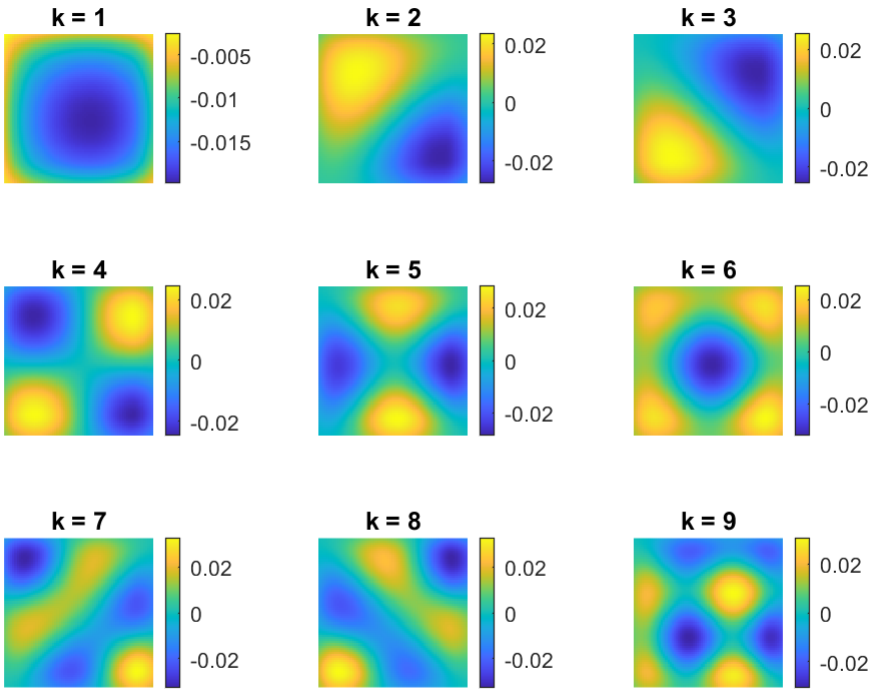


Figure 6.5: Plot of the first 9 object space singular vectors for the widefield microscope. As index k increases, the singular vectors become increasingly oscillatory, and contain higher and higher spatial frequencies, i.e., information about the finer details of the object.

Fourier optics. For example, in the field of astronomy [19], Puschmann et al. show that when the object exhibits some degree of sparsity, information regarding the fine details that are beyond the passband of the telescope can be transmitted via the diffraction-limited image. In this case, only *information* about the fine details are passed, not the fine details themselves. To recover a super-resolved image containing those fine details beyond the passband of the telescope or microscope, an additional processing step is needed to interpret the information and recover the super-resolved details. The example used by authors of [19] is briefly summarized below.

Here, imaging of a point source is considered. Under such a scenario, the image of a single star, i_{star}, is a scaled and shifted version of the PSF:

$$i_{\text{star}} = a \cdot \text{PSF}(x - \Delta x, y - \Delta y), \tag{6.7}$$

where a is the total intensity collected by the telescope, PSF the function that describes the PSF of the telescope, and Δx and Δy represent the location of the star. The image spectrum, I_{star}, is obtained by Fourier transform:

$$I_{\text{star}} = a \cdot \text{MTF}(u, v) \cdot e^{2\pi j(u\Delta x + v\Delta y)}, \tag{6.8}$$

where MTF is the function that describes the MTF of the telescope, and it has a cut-off frequency, f_c. Despite this, it can be seen that information about objects at spatial frequencies higher than f_c is persevered in the image in the term $e^{2\pi j(u\Delta x + v\Delta y)}$, even though the image itself does not contain spatial frequencies beyond f_c. As a concrete example, for two stars with equal intensities, and located at the same distance Δ away from the center of the FOV along the x axis, the image spectrum of the two stars will be:

$$I_1 + I_2 = \text{MTF}(u, v) \cdot 2a \cos 2\pi u\Delta. \tag{6.9}$$

It can be seen that in the case where the object exhibits sparsity,[6] while the MTF still limits the highest possible spatial frequency present in the image spectrum to the cut-off frequency f_c, the image spectrum is modified due to the presence of sparsity. From Equation (6.9), it can be

[6] This two-star object can be thought of as being the same as the two-dot object used in [21].

seen that, in the image spectrum, information regarding the fine details is preserved in the term $2a\cos 2\pi u\Delta$. While the details themselves cannot be transmitted through the imaging system, the information about these details allows a properly designed processing scheme to successfully reconstruct the fine details, i.e., to recover a super-resolved image.

This example shows that while Fourier optics can successfully model the image formation of an imaging system, it cannot provide the full picture regarding the information transfer in that imaging system if the object exhibits some level of sparsity. This is because even if the acquired image is diffraction-limited, it will still contain information about the fine details beyond the diffraction limit. That information can later be recovered using an appropriately designed processing step, and the assertion made by a Fourier optics analysis, which states that the cut-off frequency, f_c, fundamentally limits the amount of information that can be transferred through the imaging system, is not correct in this scenario.

6.6 Incorporating object sparsity in the analysis

Here, an analysis of information transfer for a two-dot object (as described in section 6.5.2) is presented. It is shown that contrary to a Fourier optics analysis, which shows that the information transmitted through such a system is *diffraction-limited even in the noiseless case*, the SVD analysis shows that the exact recovery of information beyond the diffraction limit is possible in this noiseless scenario.

In the results shown in section 6.4, the object is emitting fluorescence everywhere, in other words, the profile function $\vec{p} = \vec{1}$. Now, the object is modified such that it consists of two dots. This object, its image, and the processed image, are shown in Figure 6.6. Since the distance separating the dots is smaller than the diffraction limit, the image of the two dots shows a single patch of light, and the two dots can no longer be distinguished. However, solving the NNLS problem for that image shows that the diffraction-limited image does contain information that is beyond the diffraction limit, since the two dots are successfully resolved in the processed image.

This result suggests that information about details beyond the diffraction limit is contained

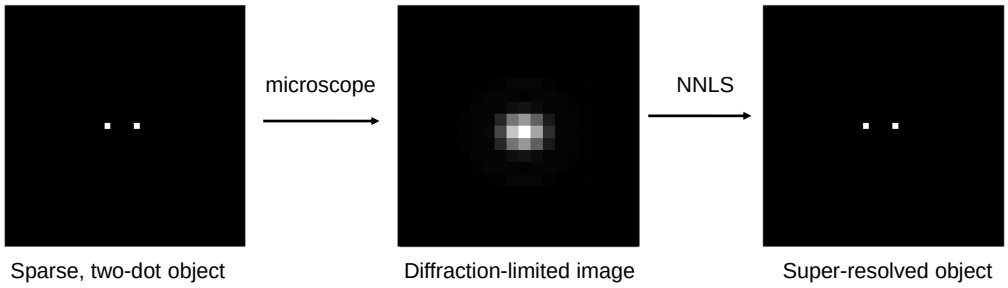


Figure 6.6: A two-dot object, its diffraction-limited image, and its super-resolved image. This object is used as an example to show that when there is sparsity present in the object, a diffraction-limited image collected by a microscope objective can contain information about the details beyond the diffraction limit even while rejecting those details in the displayed image. This information can be recovered using a processing scheme (such as NNLS shown here) to produce a super-resolved image. The sparse, two-dot object, the image that is formed by a diffraction-limited microscope, and the super-resolved object obtained using NNLS is shown.

in the diffraction-limited image, and this is not predicted using a Fourier optics analysis. Using the same SVD analysis method shown in section 6.3, an equivalent MTF for this particular imaging system described by H_{modified} can be generated. The MTF generated this way is unity for all spatial frequencies.[7] While this implies that all spatial frequencies can be transmitted through the imaging system without loss, it is not entirely appropriate as an analysis. This is because in the case where the object is sparse, the profile function $\vec{p}$ is itself the illuminated object, or the true solution. As a result, the modified optical system H_{modified} contains information about the true solution in the first place. While the subsequent result showing exact recovery in noiseless scenario is correct, this incorporation of the true solution in the analysis itself is not entirely rigorous.

Here, a more detailed analysis of this two-dot scenario is provided. The goal is to show that the true solution is also the solution that reconstructs the noiseless image *exactly*, and can be found using a convex optimization routine. In this case, the true solution has two non-zero elements, i.e., the two dots. Therefore, for H_{PSF} with a dimension of m-by-n, there are $\binom{n}{2}$ possibilities of where these two dots can be located. For this example, the true solution x_{true} is the same as the one

[7] Not shown due to its simplicity.

shown in Figure 6.6, and it gives rise to a noiseless image, $I_{\text{noiseless}}$, as shown in Figure 6.6. If the information about x_{true} is indeed transmitted through the system via the diffraction-limited image, then x_{true} should be the only solution vector that reconstructs $I_{\text{noiseless}}$ exactly. This is shown in the following numerical experiment.

First, a sufficient number of columns and rows are deleted from H_{PSF} such that it becomes a small square matrix $H_{\text{PSF, square}}$. This is valid because as long as the columns of $H_{\text{PSF, square}}$ span a large enough physical region to cover the two dots, the result will not be changed. Similarly, because each column stores an image that is a shifted version of the PSF, which largely contains zero-valued pixels except for the main lobe of the PSF, these zero-valued pixels can also be deleted to reduce the number of rows of H_{PSF}. For reference, H_{PSF} nominally has a dimension of 400-by-1681 in this chapter. Here, after the appropriate columns and rows are deleted, $H_{\text{PSF, square}}$ is 100-by-100. Next, the true solution containing the two-dot object is defined, and the corresponding noiseless image $I_{\text{noiseless}}$ is generated.

To verify that only the original true solution x_{true} reconstructs $I_{\text{noiseless}}$ exactly, the following procedure is used. First, two columns of $H_{\text{PSF, square}}$ are selected to form $H_{\text{PSF, small}}$, which is a 100-by-2 matrix. Then the following problem is solved using MATLAB `pinv`:[8]

$$H_{\text{PSF, small}} \times x = I_{\text{noiseless}}, \text{ subject to } x \geq 0. \tag{6.10}$$

The solution x and the corresponding norm of error, $\|H_{\text{PSF, small}} \times x - I_{\text{noiseless}}\|_2$, are then recorded for later analysis, and a different $H_{\text{PSF, small}}$ is generated until all possible combinations are exhausted.

The main conclusion that can be drawn through this endeavour is that the original true solution x_{true} is the only solution to Equation (6.10). When Equation (6.10) is solved by `pinv` in MATLAB, a least squares problem is used, and a vector that minimizes the squared error is returned. However, the equality condition in Equation (6.10) means the solution must reproduce the noiseless image exactly, i.e., the squared error of the solution must be zero. Therefore, after a

[8] i.e., $x = \text{pinv}(H_{\text{PSF, small}}) \times I_{\text{noiseless}}$, and then it is checked to determine if $H_{\text{PSF, small}} \times x = I_{\text{noiseless}}$ and if $x \geq 0$

vector is returned by `pinv`, it is checked by plugging it back into Equation (6.10) and observing if the equality is satisfied. If the vector does not produce a norm-of-error value of zero, it is not a solution to Equation (6.10). It is found that the solution only exists if both of the two selected columns are the non-zero elements in x_{true}. As expected, the norm-of-error value in this case is 0 (excluding rounding error). For all other choices of $H_{\text{PSF, small}}$, while a "solution" is returned by `pinv`, none of them satisfy the equality condition, and therefore do not solve Equation (6.10). The norm-of-error values in these cases range from ≈ 0.016 to ≈ 0.389. For the 100-by-100 $H_{\text{PSF, square}}$ used in this section, there are $\binom{100}{2} = 4950$ possible choices of $H_{\text{PSF, small}}$, and only one solves Equation (6.10). This in turn shows that information about the details beyond the diffraction limit is indeed preserved and transferred through the imaging system via the diffraction-limited image.

To further support the earlier claim that information about the object's fine details is present in the diffraction-limited image, even though the details themselves are rejected by the microscope's limited passband, the following example is presented. Figure 6.7 shows the spectrum of the (unprocessed) image of the two-dot object and compares it against the MTF of the microscope system. It can be seen that the two spectra are indeed different from one other, and using only the MTF to describe the information transfer ability of a microscope, without consideration of the actual object being observed, does not provide the complete picture regarding the information transfer ability of the microscope. While the image of the two dot object is diffraction-limited, it still contains information about the fine details that are beyond the diffraction limit, and that information can be used to recover the original, super-resolved object. In this case, the difference manifests itself as an intensity modulation of the microscope MTF, as shown in Equation 6.9. As the distance between the two dots, Δ, decreases, term $2a\cos 2\pi u\Delta$ will be closer to unity, and the difference between the spectrum of the image of the two-dot object and the MTF will decrease. As a result, as the feature size decreases, its recovery becomes more and more difficult.

6.7 Connection to existing research

This chapter presents an analysis of the achievable resolution of an imaging system in terms of *information transfer* instead of the more conventional analysis approach (which is based on image formation). Such an analysis is necessitated by the recent wide-spread use of a computational post-processing step after acquiring the image to further enhance it. As a result of this post-processing, the information content of an image is no longer restricted by the diffraction limit, in contrast to conventional analysis. As shown in this chapter, if the object possesses some level of sparsity, conventional analysis predicts only diffraction-limited resolution, whereas the more suitable, SVD-based analysis predicts the improved resolution beyond the diffraction limit, as shown throughout this thesis.

Here, it should be noted that the analysis proposed in this chapter, which itself is a modified analysis framework based on existing research [16, 17, 18], is not the only result that discusses

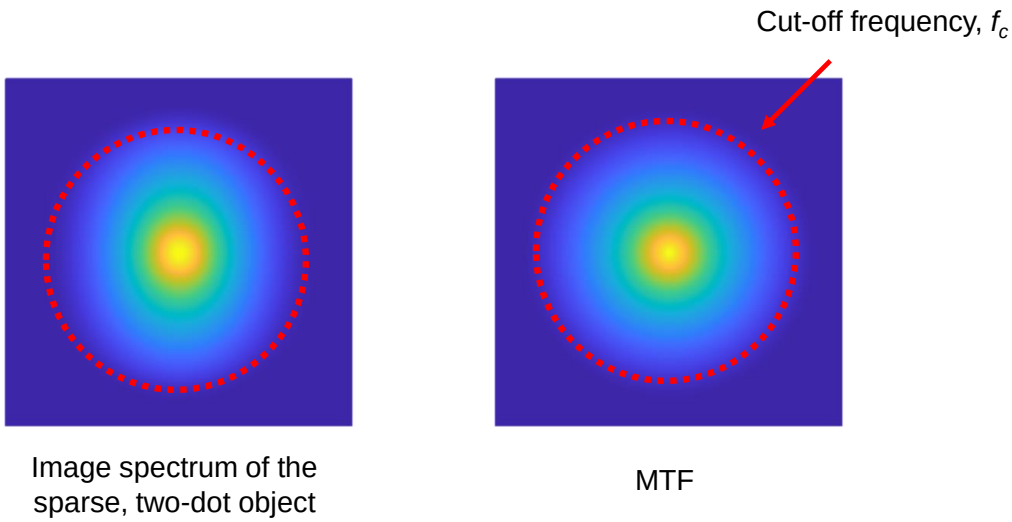


Figure 6.7: Spectrum of the image of a sparse object vs. the microscope MTF. The red dashed circle indicates the cut-off frequency, f_c. While both images do not contain spatial frequencies higher than the cut-off, the image spectrum of the sparse, two-dot object is still different (within f_c) from the MTF. This shows that the information about the fine details beyond the diffraction limit can be transmitted in a diffraction-limited image.

the information transfer in an imaging system, and such an analysis can be conducted from other perspectives. For example, a detailed treatment of information transfer, analyzed from a degrees of freedom perspective, can be found in [53]. Its connections to the results presented in this chapter are discussed below.

In [53], the degrees of freedom of an *arbitrary* optical system are analyzed. In this work, "degrees of freedom" can be regarded as independent communication channels used in digital communications, and analyzing degrees of freedom will yield a model of information transfer of an arbitrary optical system. Under this analysis framework, information is transmitted with EM waves between a transmitting domain and a receiving domain.[9]

The main result of [53] is a "rigorous formalism for defining, evaluating, and optimizing the degrees of freedom of an optical system". Using this result, the authors show that

> "...although in principle there is an infinity of degrees of freedom, the effective number is finite owing to the presence of noise."

This conclusion is strikingly similar to one reached in this chapter, namely, under the noiseless scenario, exact recovery of information through a diffraction-limited image is possible; and in practice, due to the noisy detection process, the amount of super-resolution achievable in a real-world experiment is limited, and highly dependent on the signal level (see section 3.3.1).

Additionally, using the developed formalism, the information transfer of an arbitrary optical system can be analyzed in terms of a series of transmitting and receiving functions that form a set of orthonormal bases in the transmitting and receiving domain respectively.[10] The authors of [53] then showed that "the best transmitting and receiving functions are the solutions of well-defined eigenvalue equations." This (solving the eigenvalue problem) is the exact approach used by Bertero et al. to obtain the singular vectors in [16]. In fact, the authors of [53] also proposed an "alternative and equivalent", *SVD-based* approach to determine the transmitting and receiving functions.

From this comparison between [53] and [16], it can been seen that the theoretical results

[9] Note the similarity between the transmitting versus receiving domain in [53] and the object versus image space proposed in this chapter.

[10] This is analogous to the object and image space singular vectors discussed in this chapter.

developed in [53] are a generalization of those developed in [16] and, by extension, of the results contained in this chapter. The difference being that this chapter and [16] focus more narrowly on analyzing imaging systems such as microscopes, whereas [53] focuses more broadly on the information transfer using EM waves for arbitrary optical systems. Both analyses draw the same set of conclusions in terms of information transfer, which are best summarized by the following quotes from [53]:

> "...information can be regarded as an entity that, in the absence of noise, is propagated from the input to the output of the system without destruction."

and

> "...The DOF [degrees of freedom] are defined by those source functions s_n that, after diffraction in space, generate field functions ϵ_n that contain the same information..."

In other words, in noiseless imaging, as have been shown in published papers such as [15] and in this chapter, exact recovery can be achieved. In the analysis of information transfer, the information can be represented in small "pieces". In [16], the information is represented as singular vectors in object and image space. In [53], the information is represented as transmitting and receiving functions. Both analysis approaches address the main deficiency of the conventional Fourier optics analysis, and enable a complete description of information transfer to be generated.

6.8 Conclusion

In this chapter, information transfer in a microscope is discussed. The conventional tool used for this analysis is the familiar Fourier optics, which states that the finest detail that can be transmitted by a microscope is determined by the cut-off frequency of the microscope, f_c. Under this analysis, any details composed of spatial frequencies higher than f_c are rejected by the microscope and cannot be recovered. However, it is shown throughout this thesis and in earlier works, such super-resolution is in fact, possible. Here, it is shown that while Fourier optics can model the *image formation* of a microscope correctly, it cannot provide the complete picture in terms of

information transfer. This is because the Fourier optics analysis does not take into consideration if the object is inherently sparse, such as is the case in chemically-specific fluorescent labeling that occurs in fluorescence microscopy. When such sparsity exists in the sample, while the image is still diffraction limited (i.e., it does not actually image details finer than f_c), it does contain *information* about the fine details beyond f_c which can be recovered in a post-processing step.

An analysis method, based on SVD, that incorporates such sparsity in its implementation, is first proposed in [16] and expanded in this chapter. It is shown that by using SVD, which in turn enables two different basis sets for the object and image space to be used in the analysis, a more complete description can be obtained regarding the information transfer that occurs in a microscope when the object exhibits sparsity.

Chapter 7

Discussion, conclusion, and future work

7.1 Similarity to single molecule localization microscopy

It is undeniable that the super-resolution microscopy technique presented in this thesis shares many similarities with single molecule localization microscopy (SMLM). Both techniques acquire diffraction-limited images using a camera, and both techniques do not achieve super-resolution directly, instead apply a numerical post-processing step to recover the super-resolved images. Additionally, techniques such as ℓ_1 regularized maximum likelihood estimation (i.e., Equation (3.11)) have been used to achieve improved results in both methods, for example, in [23] for the technique discussed in this thesis, and [29] for SMLM. Despite these experimental and numerical similarities, it should be noted that the technique discussed here differs from SMLM in one key aspect, i.e., not requiring photo-switchable fluorophores.

In [29], the fluorescent label density in the image is approximately 10 μm^{-2}, which is only achievable through the on-off state manipulation commonly used in SMLM. This is in contrast with imaging using non-switchable fluorophores, where the typical labeling density of common samples is estimated to be much higher than 1000 μm^{-2} [21]. As discussed in section 3.7, in SMLM, the level of sparsity present in the detected image is roughly the same for different acquisitions, since for any given acquired image, roughly the same number of fluorescent molecules are activated and emitting light. This is not the case for the technique discussed in this thesis, as different levels of sparsity may be present as the focused illumination spot is scanned across the sample. The fact that meaningful super-resolution can be achieved with ℓ_1 regularized maximum likelihood estimation without the

need for on-off state manipulation with photo-switchable fluorophores suggests that the level of sparsity already present in the image is often enough to achieve computational super-resolution.

Another related example in the field of *three-dimensional* SMLM can be found in [54]. Similar to [29], this work addresses the slow acquisition speed by allowing the fluorescent molecules to overlap[1] during the acquisition and attempts to super-resolve them after the acquisition with sophisticated post-processing.

Two main processing schemes are investigated in [54], convex optimization (CO), and matching pursuit (MP). Numerical simulation shows that CO achieves higher accuracy while MP is faster to execute but achieves lower accuracy. This is in agreement with the results presented in Chapter 3, where the achievable resolution (for the same noisy image) can be different if different processing schemes are used. The fact that CO achieves higher accuracy is of great interest, this is because the CO implementation in [54] is equivalent to r-NNLS in Chapter 3, which only accounts for the sparsity prior information, but not the shot noise dominant model. Based on the results presented in that chapter, it is reasonable to expect that even higher accuracy can be achieved if a processing scheme that accounts for both sparsity and noise model is used, such as r-Poisson-MLE or r-VST.

Additionally, as stated in its discussion section, [54] further supports the assertions made in the latter part of this thesis, namely, achieving super-resolution from diffraction-limited images is possible, and it is enabled by the sparsity in the underlying object. The encoded super-resolution information is an intrinsic property of the diffraction-limited images. The function of the processing schemes, such as CO and MP in [54], and the ones presented in this thesis, is to recover that information.

7.2 Conclusion and future work

The idea of improving fluorescence microscopy images is not a novel one, and improving the achievable resolution, especially improving the resolution beyond that allowed by the diffraction limit, can be beneficial for biological studies utilizing these images. Despite the extensive research

[1] Note that on-off state manipulation is still required.

effort this topic has received over the past decade, existing super-resolution microscopy techniques generally require specialized illumination setups or photo-switchable fluorophores to achieve more than two-fold resolution improvement beyond the conventional widefield diffraction-limited resolution. Developing an imaging technique that can achieve more than two-fold resolution improvement, while not requiring specialized illumination setups or photo-switchable fluorophores, could bring super-resolution microscopy to a much wider audience and range of applications, which would unlock additional insights in biological studies.

This thesis explores several important aspects of a promising computational super-resolution fluorescence microscopy approach. In this method, a diffraction-limited image is acquired. While this image does not contain details beyond the diffraction limit directly, it does contain *information* about those details that can later be used by a properly designed processing scheme to reconstruct a super-resolved image. While the theoretical proof-of-concept experiment of this technique was completed earlier on, due to its intensive computational resource requirement it is not until very recently that a biological proof-of-concept experiment was completed.

Despite this recent development, there are multiple problems that need solving before this technique can be adopted by a wider audience. Therefore, this thesis aims to answer three problems: 1) what are the properties a processing scheme must fulfill in order to achieve the highest resolution? 2) what is the effect of non-unique solutions on this technique? and 3) what is the implication of the successful demonstration and verification of this technique on the analysis of the *information transfer* ability of a microscope?

To answer question 1), a careful examination of the overall imaging system is conducted. It is found that the processing scheme used in the previous biological proof-of-concept experiment, NNLS, is not optimal in terms of maximizing the achievable super-resolution. In order for the achievable super-resolution to be maximized, the processing scheme needs to 1) be appropriate for the dominant noise model in the image, which in a typical fluorescence microscopy image is the photonic shot noise; and 2) be able to take advantage of the prior information of sparsity in the acquired image, which is a combined effect of focused spot illumination and the chemically specific

labeling used in fluorescence microscopy. Based on these findings, an improved processing scheme, r-VST, is proposed. Using numerical simulation, its improved performance over NNLS is validated. r-VST is then used to process data from a real-world experiment, and the result is compared with those obtained using NNLS and VST. It is shown that in agreement with the previous findings, because r-VST is appropriate for the dominant noise model in the image, and is able to take advantage of the prior information of sparsity, images processed with r-VST achieve the highest resolution. A verified resolution of 60nm is shown to be achieved, which is approximately four times beyond the diffraction limit.

Additionally, because the only assumption made by this technique is the acquired image is made up of incoherent light, this method can be readily extended to other imaging modalities. For example, [55] demonstrates the use of this super-resolution approach in conjunction with two-photon microscopy. Here, an extension of this technique in conjunction with UCNPs is shown. UCNPs are nano-scale particles that, when used as contrast agents for imaging, show very low background signal levels. This effectively increases sparsity, making them a prime candidates for this sparsity-dependent computational super-resolution microscopy technique. A proof-of-concept experiment using UCNPs is performed, and the results are verified against a TEM ground-truth image. It is again shown that super-resolution is verifiably achieved, and r-VST produces the image with the highest precision and visual quality.

To answer question 2), an investigation of the uniqueness of the solution to the inverse problem is conducted. An apparent conflict between linear algebra analysis and published work on computational super-resolution microscopy is discovered. From the linear algebra analysis, the solution is not unique, even in a noiseless scenario. However, published works on computational super-resolution generally treat the solution as unique in the noiseless scenario. This conflict is successfully reconciled in this thesis. It is shown that instead of having a clear-cut answer, the problem of whether the solution is unique depends on the level of sparsity present in the imaged object. Generally speaking, the sparser the object, the higher the probability of a unique solution. Two methods of certifying the uniqueness of the solution are proposed in this thesis. Utilizing

these methods, the uniqueness of the solution to the inverse problem is investigated for a sparse, two-dot object, and a less sparse, two-line object. It is determined that the solution for the sparse two-dot object is unique. While the solution to the less sparse, two-line object is non-unique, it is found that the effect of non-unique solutions is greatly reduced in a real-world experiment, and most importantly, does not affect the locations of the two lines. These two uniqueness certification methods can help the potential users of this technique to determine if the solution in their particular imaging scenario is unique or not.

To answer question 3), a literature survey on analyzing information transfer in a microscope is conducted. It is found that while the familiar Fourier optics analysis successfully models the *image formation* in a microscope, *information transfer* is better modeled by a SVD based analysis. If the object is sparse, while its image is still diffraction limited, the image spectrum is modified in such a way that it contains information about the fine details beyond the diffraction limit. As a result, that information can be utilized by a properly designed processing scheme to reconstruct a super-resolved final image. Because Fourier optics-based analysis cannot directly incorporate the consideration of sparsity, it fails to predict the existence of super-resolution in this scenario. Such a consideration can be achieved using SVD-based analysis. Based on previously published work, this thesis modifies that analysis, and shows that for scenarios where Fourier optics can successfully predict super-resolution, SVD can provide the same result by assuming the object is not sparse. However, in scenarios where Fourier optics fails to predict the existence of super-resolution, SVD successfully predicts such super-resolution by incorporating the object sparsity into the analysis.

The answers to the three problems presented in this thesis allow the potential users of this emerging technique to achieve the highest possible resolution, to evaluate if the solution is unique for their particular imaging application, and to understand the reason why the familiar Fourier optics analysis does not seem to suggest such resolution improvement is possible. By providing these answers, this technique has reached a previously unseen level of maturity, and is much closer to being adopted by a wider audience in a wider range of applications.

The main challenge of this technique remains the intensive computational resource load. As a

final post-processing step, most users of microscopes expect a timescale of seconds, perhaps minutes. However, this technique usually requires hours of processing using specialized large-scale numerical solvers, and sometimes multi-GPU workstations. While recent advances in large-scale numerical optimization and lowering cost of GPU computation has made this technique much more user friendly, further improvement is still highly needed. One possible avenue for such improvement is the potential use of deep neural networks for solving inverse problems, such as in [56]. Such application of deep neural networks can dramatically reduce the necessary processing time required by this technique, and may prove to be the final step in transforming this new method from laboratory novelty to a standard tool for biological investigations.

Bibliography

[1] L. Rayleigh, "On the theory of optical images, with special reference to the microscope," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 42, no. 255, pp. 167–195, 1896.

[2] J. W. Goodman, Introduction to Fourier optics, ch. 6.5.2. Roberts and Company Publishers, 2005.

[3] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Optics letters, vol. 19, no. 11, pp. 780–782, 1994.

[4] T. A. Klar and S. W. Hell, "Subdiffraction resolution in far-field fluorescence microscopy," Optics letters, vol. 24, no. 14, pp. 954–956, 1999.

[5] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science, vol. 313, no. 5793, pp. 1642–1645, 2006.

[6] M. J. Rust, M. Bates, and X. Zhuang, "Stochastic optical reconstruction microscopy (STORM) provides sub-diffraction-limit image resolution," Nature Methods, vol. 3, no. 10, p. 793, 2006.

[7] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, "Fast, background-free, 3d super-resolution optical fluctuation imaging (SOFI)," Proceedings of the National Academy of Sciences, vol. 106, no. 52, pp. 22287–22292, 2009.

[8] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," Journal of microscopy, vol. 198, no. 2, pp. 82–87, 2000.

[9] K. I. Willig, B. Harke, R. Medda, and S. W. Hell, "Sted microscopy with continuous wave beams," Nature methods, vol. 4, no. 11, pp. 915–918, 2007.

[10] S. W. Hell and M. Kroug, "Ground-state-depletion fluorscence microscopy: A concept for breaking the diffraction resolution limit," Applied Physics B, vol. 60, no. 5, pp. 495–497, 1995.

[11] S. W. Hell, S. Jakobs, and L. Kastrup, "Imaging and writing at the nanoscale with focused visible light through saturable optical transitions," Applied Physics A, vol. 77, no. 7, pp. 859–860, 2003.

[12] S. W. Hell, A. Schönle, and A. Van den Bos, "Nanoscale resolution in far-field fluorescence microscopy," in Science of microscopy, pp. 790–834, Springer, 2007.

[13] E. H. Rego, L. Shao, J. J. Macklin, L. Winoto, G. A. Johansson, N. Kamps-Hughes, M. W. Davidson, and M. G. Gustafsson, "Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution," Proceedings of the National Academy of Sciences, vol. 109, no. 3, pp. E135–E143, 2012.

[14] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," Journal of the Royal Statistical Society: Series B (Methodological), vol. 54, no. 1, pp. 41–67, 1992.

[15] P. J. Sementilli, B. R. Hunt, and M. S. Nadar, "Analysis of the limit to superresolution in incoherent imaging," J. Opt. Soc. Am. A, vol. 10, pp. 2265–2276, Nov 1993.

[16] M. Bertero and E. Pike, "Resolution in diffraction-limited imaging, a singular value analysis," Optica Acta: International Journal of Optics, vol. 29, no. 6, pp. 727–746, 1982.

[17] M. Bertero, P. Boccacci, M. Defrise, C. D. Mol, and E. R. Pike, "Super-resolution in confocal scanning microscopy: II. the incoherent case," Inverse Problems, vol. 5, pp. 441–461, aug 1989.

[18] M. Bertero, C. D. Mol, E. Pike, and J. Walker, "Resolution in diffraction-limited imaging, a singular value analysis," Optica Acta: International Journal of Optics, vol. 31, no. 8, pp. 923–946, 1984.

[19] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging," Astronomy & Astrophysics, vol. 436, no. 1, pp. 373–378, 2005.

[20] C. B. Müller and J. Enderlein, "Image scanning microscopy," Physical review letters, vol. 104, no. 19, p. 198101, 2010.

[21] J.-Y. Yu, S. R. Becker, J. Folberth, B. F. Wallin, S. Chen, and C. J. Cogswell, "Achieving superresolution with illumination-enhanced sparsity," Optics express, vol. 26, no. 8, pp. 9850–9865, 2018.

[22] S. Becker, E. Candes, and M. Grant, "TFOCS: flexible first-order methods for rank minimization," in Low-rank Matrix Optimization Symposium, SIAM Conference on Optimization, 2011.

[23] J. Xing, S. Chen, S. Becker, J.-Y. Yu, and C. Cogswell, "ℓ_1-regularized maximum likelihood estimation with focused-spot illumination quadruples the diffraction-limited resolution in fluorescence microscopy," Opt. Express, vol. 28, pp. 39413–39429, Dec 2020.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[25] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Transactions on information theory, vol. 52, no. 2, pp. 489–509, 2006.

[26] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," IEEE Transactions on Image Processing, vol. 17, no. 10, pp. 1737–1754, 2008.

[27] A. Charnes, E. L. Frome, and P.-L. Yu, "The equivalence of generalized least squares and maximum likelihood estimates in the exponential family," Journal of the American Statistical Association, vol. 71, no. 353, pp. 169–171, 1976.

[28] S. M. Riad, "The deconvolution problem: An overview," Proceedings of the IEEE, vol. 74, pp. 82–85, Jan 1986.

[29] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," Nature methods, vol. 9, no. 7, pp. 721–723, 2012.

[30] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," Acta Numerica, vol. 25, pp. 161–319, 2016.

[31] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," Biometrika, vol. 35, no. 3/4, pp. 246–254, 1948.

[32] "Argo-check resolution monitoring slide." `http://argolight.com/products/argo-check-resolution/`. Accessed: 2020-07-07.

[33] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, et al., "Video-rate nanoscopy using scmos camera–specific single-molecule localization algorithms," Nature methods, vol. 10, no. 7, pp. 653–658, 2013.

[34] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," SIAM Journal on Scientific Computing, vol. 14, no. 6, pp. 1487–1503, 1993.

[35] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," SIAM Review, vol. 34, no. 4, pp. 561–580, 1992.

[36] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics, vol. 21, no. 2, pp. 215–223, 1979.

[37] E. Van Den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," SIAM Journal on Scientific Computing, vol. 31, no. 2, pp. 890–912, 2009.

[38] X. Wu, G. Chen, J. Shen, Z. Li, Y. Zhang, and G. Han, "Upconversion nanoparticles: a versatile solution to multiscale biological imaging," Bioconjugate chemistry, vol. 26, no. 2, pp. 166–175, 2015.

[39] A. Das, C. Mao, S. Cho, K. Kim, and W. Park, "Over 1000-fold enhancement of upconversion luminescence using water-dispersible metal-insulator-metal nanostructures," Nature communications, vol. 9, no. 1, pp. 1–11, 2018.

[40] M. Bertero and P. Boccacci, Introduction to inverse problems in imaging. CRC press, 2020.

[41] J. L. Harris, "Diffraction and resolving power," JOSA, vol. 54, no. 7, pp. 931–936, 1964.

[42] M. Bertero and P. Boccacci, Introduction to inverse problems in imaging, ch. 2. CRC press, 2020.

[43] D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," Discrete & Computational Geometry, vol. 8, no. 3, pp. 295–313, 1992.

[44] L. Khachiyan, E. Boros, K. Borys, V. Gurvich, and K. Elbassioni, "Generating all vertices of a polyhedron is hard," in Twentieth Anniversary Volume:, pp. 1–17, Springer, 2009.

[45] "Analyze n-dimensional convex polyhedra." `https://www.mathworks.com/matlabcentral/fileexchange/30892-analyze-n-dimensional-convex-polyhedra`. Retrieved April 11, 2021.

[46] L. Laird, R. C. Tillquist, S. Becker, and M. E. Lladser, "Resolvability of Hamming graphs," SIAM Journal on Discrete Mathematics, vol. 34, no. 4, pp. 2063–2081, 2020.

[47] L. B. Lucy, "Resolution limits for deconvolved images," The Astronomical Journal, vol. 104, pp. 1260–1265, 1992.

[48] P. Magain, F. Courbin, and S. Sohy, "Deconvolution with correct sampling," The Astrophysical Journal, vol. 494, no. 1, p. 472, 1998.

[49] J. W. Goodman, Introduction to Fourier optics, ch. 2.1.2. Roberts and Company Publishers, 2005.

[50] B. R. Frieden, "Evaluation, design and extrapolation methods for optical signals, based on use of the prolate functions," in Progress in optics, vol. 9, pp. 311–407, Elsevier, 1971.

[51] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," Bell System Technical Journal, vol. 40, no. 1, pp. 43–63, 1961.

[52] J.-Y. Yu, V. Narumanchi, S. Chen, J. Xing, S. R. Becker, and C. J. Cogswell, "Analyzing the super-resolution characteristics of focused-spot illumination approaches," Journal of Biomedical Optics, vol. 25, no. 5, p. 056501, 2020.

[53] R. Piestun and D. A. Miller, "Electromagnetic degrees of freedom of an optical system," JOSA A, vol. 17, no. 5, pp. 892–902, 2000.

[54] A. Barsic, G. Grover, and R. Piestun, "Three-dimensional super-resolution and localization of dense clusters of single molecules," Scientific reports, vol. 4, no. 1, pp. 1–8, 2014.

[55] S. Chen, "Three-dimensional super-resolution microscopy using laser scanning and non-linear optimization," Ph. D. Thesis, 2020.

[56] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 20–36, 2018.