Developments Towards Chromophore-Selectivity in Photoacoustic Remote Sensing Microscopy

by

Nicholas Pellegrino

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2022

Author's Decleration

This thesis consists of material all of which I either authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Nicholas Pellegrino is the sole author of all chapters except Chapter 5. All chapters were authored under the supervision of Dr. Parsin Haji Reza and Dr. Paul W. Fieguth and are not intended for publication.

Components of Chapter 5 were submitted to the 26th International Conference on Pattern Recognition (ICPR 2022) and Optica Optics Letters (see table below). Figures 5.6 and 5.9 were created for this thesis; however, they were submitted as part of [1]. Figures 5.10 and 5.13 were also created for this thesis; however, versions of these figures were submitted as part of [2].

Figure 3.14 in Chapter 3 is created based on data collected by Zohreh Hosseinaee and Nima Abbasi Firoozjah. Processing of the data was done by Nicholas Pellegrino and the figure was created for this thesis; however, it was also submitted as part of [3].

The raw PARS images used in Chapter 5 were collected by Benjamin R. Ecclestone. In particular these are Figures 5.1, 5.3, 5.6 and 5.10.

Chapter	Citation	Author's Contributions
5	N. Pellegrino, P. Fieguth, and P. Haji Reza, "K-Means for Noise-Insensitive Multi-Dimensional Feature Learning," 2022, Submitted to the 26th International Conference on Pattern Recognition. (available via arXiv [1])	Developed method. Primary author of manuscript.
5	N. Pellegrino, B. R. Ecclestone, P. Fieguth, and P. Haji Reza, "Time-domain feature extraction for target-specificity in Photoacoustic Remote Sensing Microscopy," 2022, Submitted to Optics Letters. (available via arXiv [2])	Primarily performed data analysis and jointly wrote manuscript with equal contribution to Benjamin R. Ecclestone.

Abstract

Medical imaging serves to diagnose and monitor illnesses in clinical settings. Several modalities such as X-Ray and Ultrasound have become ubiquitous due to their high utility. Photoacoustic microscopy, based on the photoacoustic effect originally discovered by Alexander Graham Bell, is sensitive to optical *absorption* contrast. This is highly useful in biomedical applications where the endogenous absorption contrast of tissue can directly be imaged, enabling label-free microscopy. Applications include histological assessment of tissue in support of cancerous tissue resection surgeries, as well as the functional imaging applications of blood oxygen saturation (sO_2) and metabolic rate (MRO_2) imaging. Recently, Photoacoustic Remote Sensing Microscopy (PARS), an all-optical implementation of photoacoustic microscopy, was pioneered by Parsin Haji Reza.

This thesis makes three main contributions supporting the development of PARS microscopy. The first is the creation of an inverse model designed to solve for the concentrations of individual chromophores when imaged using several excitation wavelengths. To achieve this, constraints and considerations specific to PARS are designed and employed. The inverse model allows for the concentrations of oxy- and deoxyhemoglobin to be solved for, enabling sO_2 estimation. This was performed *in-vivo* in an ocular setting, demonstrating the first non-contact photoacoustic measurement of sO_2 in the eye. The second contribution is an in-depth experimental study of Stimulated Raman Scattering (SRS) in single-mode optical fiber as a means of generating multi-wavelength light from a conventional single-wavelength laser source. Effects associated with several laser parameters and properties of optical fiber are studied. Results of this study will find use in the apparatus design for nearly any multi-wavelength PARS application, typically where imaged absorbers must be unmixed, and in particular, in sO_2 estimation. The final contribution is the development of a novel unsupervised time-domain feature-learning algorithm, designed to learn characteristic signal shapes. This allowed tissue sub-components to be discerned in PARS imagery of both unstained human breast tissue on slides and freshly resected murine brain tissue without the need to use multiple excitation wavelengths nor have any prior knowledge of the time-domain characteristics associated with individual components.

The contributions made in this thesis represent significant steps towards the use of PARS for a broad range of applications where unmixing, or more specifically, discerning underlying components of the imaged target is required — beyond sO_2 estimation or emulation of standard histological techniques. Furthermore, the improved understanding of how SRS can be used to generate additional excitation wavelengths opens the door to imaging an abundance of bio-molecules, thus broadening the scope and richness of the gamut of targets that PARS is capable of imaging.

Acknowledgements

I would like to thank my graduate supervisors, Professor Parsin Haji Reza and Professor Paul W. Fieguth for their ongoing support, guidance, and encouragement over the *three* years I have been their student. Most of all, I thank you for never giving up on me — instead, you continued to support and mentor me, allowing me to grow academically and finally complete my degree.

I would like to thank my thesis readers, Dr. Karim Sallaudin Karim and Dr. Nima Maftoon. In particular, I thank you both for your patience while I pursued additional research prior to finally writing my thesis, many months later than what was originally planned for and expected by you.

I would like to thank all my friends and colleagues at PhotoMedicine Labs for creating an amazing environment and team that I had the pleasure of being part of. Additionally, I truly appreciate the companionship of my pre-pandemic office-mates, Saad Abbasi, and Benjamin R. Ecclestone (and Kevan Bell, who regularly joined us). My special thanks go to Zohreh Hosseinaee, Nima Abbasi Firoozjah, and Benjamin R. Ecclestone for collecting images supporting my work during the COVID-19 pandemic. Additionally, I would like to thank Alkris Warren and Lyazzat Mukhangaliyeva for helping run my experiments, Hager Gaouda for helping in preparing biological samples, and again, Kevan Bell for his guidance.

This research relied upon clinicians in Alberta, Dr. John R. Mackey and Dr. Deepak Dinakaran, who provided tissue samples and helped with clinical interpretation, as well as support from the staff at University of Waterloo Central Animal Facility, especially Jean Flanagan, in procuring and preparing animal and tissue specimens.

I would like to thank Spencer Small and Engineering Counselling as a whole. You helped me navigate my troubled path and led me to a point where I could grow as a person, as a researcher, and complete my degree.

I would like to thank my family for supporting me throughout all of this. At the end of my undergraduate degree, when my parents expected I would be looking for a job and moving on from education, I instead applied to graduate school without telling them! In spite of this, they allowed me to continue living at home while doing my master's and were encouraging, even if they didn't really understand my motivation or quite what I did at school.

Finally, I would like to acknowledge all of our funding partners: Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation, Mitacs Accelerate, University of Waterloo Startup funds, Centre for Bioengineering and Biotechnology, illumiSonics Inc. and New frontiers in research fund — exploration.

Dedication

To my newborn niece, Isla.

Infant child, innocent and pure, you were born with your instincts,
knowing to breathe, knowing to eat only when hungry,
and not caring what anyone thinks of you or what you do.
Though life will try to change you, drag you one way then the next, create fears and doubts, please remember always to stay true to yourself, believe in yourself, and follow your heart.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

PAM	Photoacoustic Microscopy
AR-PAM	Acoustic-Resolution Photoacoustic Microscopy
OR-PAM	Optical-Resolution Photoacoustic Microscopy
PAT	Photoacoustic Tomography
PARS	Photoacoustic Remote Sensing
TD	Time-Domain
sO_2	Oxygen Saturation
HbO_2	Oxyhemoglobin
Hb	Deoxyhemoglobin
H&E	Hematoxylin and Eosin
OCT	Optical Coherence Tomography
SRS	Stimulated Raman Scattering
DNA	Deoxyribonucleic Acid
DAQ	Data Acquisition System or Card
UV	Ultraviolet
NIR	Near-infrared
IR	Infrared
SPM	Self-Phase Modulation
XPS	Cross-Phase Modulation
FWM	Four-Wave Mixing
PCA	Principal Component Analysis
FFPE	Formalin-Fixed Paraffin-Embedded
RGB	Red-Green-Blue (Colour Channels)

List of Symbols

Photoacoustic Remote Sensing

$p(x,t)$	Pressure at position x and time t
p_0	Initial photoacoustic pressure
Γ	Gruneisen parameter
ε_i	Extinction coefficient of chromophore i
μ_a	Optical absorption coefficient
n_i	Refractive index of medium i
R	Reflectance at interface with refractive index mismatch
η	Elasto-optic coefficient
ρ	Mass density
ν_a	Speed of sound in medium
I	Optical intensity
λ	Wavelength of light
ϕ_i	Optical fluence of beam i
d_c	Characteristic dimension of heated region
α_{th}	Thermal diffusivity
τ_{th}	Thermal relaxation time
τ_s	Stress relaxation time
w_0	Gaussian beam waist
f	Focal length of lens

Optical Effects in Optical Fibers

α	Attenuation constant (per km)
α_{dB}	Attenuation constant in dB (per km)
P_0	Power at proximal end of fiber
P_T	Power transmitted to distal end of fiber
n_i	Index of refraction of medium i
L	Fiber length
L_{eff}	Effective fiber length, reduced by attenuation
L_W	Walk-off length
a	Fiber core radius
z	Distance from the proximal end of the fiber
$I_i(z)$	Intensity of wave i at distance z along fiber
V	Normalized frequency
b	Normalized guide index
β	Propagation constant
η	Efficiency
ω	Angular frequency of light
Ω	Angular frequency shift
$g_R(\Omega)$	Raman gain spectrum as a function of frequency shift Ω
k_p	Polarization factor

Chromophore Unmixing

c_i	Concentration of chromophore i
σ	Singular value of a matrix
Φ	Matrix of fluences
E	Matrix of extinction coefficients
ν	Additive measurement noise
R	Measurement error covariance matrix
P	Estimation error covariance matrix

Time-Domain Signal Intelligent Feature Extraction

t	Time
$s(t)$	PARS time-domain signal
f_i	Basis function
β_i	Weight associated with basis function f_i
$d(\cdot,\cdot)$	Distance metric
S^n	n^{th} Dimensional sphere
ℓ	Cluster label
C_i	Cluster i
c_i	Cluster centroid of C_i

General Mathematical Notation

π	Archimedes constant
$\Delta\cdot$ or $\delta\cdot$	Change or difference in
$\widehat{\cdot}$	Estimate of
$\langle\cdot,\cdot\rangle$	Inner product
$\lvert\cdot\rvert$	Absolute value, magnitude, or cardinality of set
$\lVert\cdot\rVert$	Euclidean norm, equivalent to $\lVert\cdot\rVert_2$
$\lVert\cdot\rVert_p$	p-norm
$\mathrm{diag}(\cdot)$	Vector of diagonal entries of argument matrix
$\mathrm{Diag}(\cdot)$	Diagonal matrix, with diagonal entries determined by argument vector
$.^T$	Matrix transpose
$.^{-1}$	Matrix inverse
$.^+$	Moore–Penrose matrix pseudoinverse
$\mathrm{cond}(\cdot)$ or $\kappa(\cdot)$	Matrix condition number
$\sup S$	Supremum of set S
$\{\cdot\}$	Mathematical Set
$\overset{\propto}{=}$	Equal up to arbitrary proportionality
$\cdot * \cdot$	Convolution

Chapter 1

Introduction

1.1 Motivation

Medical imaging broadly serves as a set of powerful diagnostic tools, non-invasively giving medical care providers valuable information that cannot be obtained otherwise. Signs of many ailments and abnormalities are hidden by skin, layers of tissues, and bone on human or other animal patients [4, 5]. This necessitates medical imaging as a means of informing clinicians of the status of the patients they treat. For example, things as commonplace as monitoring the health and development of a fetus while in the womb are vastly aided by ultrasonic imaging [6, 7]. A more specialized example of medical imaging is the use of magnetic resonance imaging (MRI) to assess the brain and spinal cord for damage following trauma or due to disease [8–11].

Many medical imaging modalities already exist; however, there remains a need for continued research and development of new modalities and techniques. Several common modalities include ultrasound imaging [6, 12, 13], x-ray imaging [14, 15] (and computed tomography [16, 17]), and magnetic resonance imaging [18, 19]. Optical imaging, which uses light from the ultraviolet–visible–infrared range of the spectrum, avoids exposing patients to ionizing radiation (such as x-rays), providing a safer diagnostic tool [20]. Furthermore, because of the many ways in which different soft tissues interact with light (scattering and absorption), optical imaging techniques are especially well suited to biomedical applications [20]. Optical imaging modalities include optical coherence tomography [21–24], light microscopy / photography (e.g., bright-field [25], dark-field [26, 27], phase-contrast [28, 29], fundus photography [30–33]), confocal microscopy [34, 35], light sheet microscopy [36–38], fluorescence microscopy [39–41], Raman microscopy [42, 43], and lastly photoacoustic microscopy (PAM) [44, 45]. PAM is sensitive to *absorbed* light, a rather unique source of contrast not accessible by other modalities. Conventionally, PAM relies on contact-based detection via an ultrasonic transducer [45, 46], a significant disadvantage to the modality. Though PAM has shown much use in imaging the eye [46–49], it is not a perfect solution. Contact-based detection may increase the risk of infection, abrasion, or may cause patient discomfort. Additionally, the applied pressure to the eye has an influence on the ocular vasculature function [46].

Photoacoustic Remote Sensing Microscopy (PARS) [3, 50–63], is a novel biomedical imaging modality based on the photoacoustic effect (like the broader field of photoacoustic microscopy). PARS offers a non-contact method of imaging absorption contrast, providing the benefits of general PAM without its disadvantages. Additionally, PARS, and photoacoustic microscopy in general, allow for label-free microscopy, not requiring the use of exogenous dyes or labels to provide contrast. These characteristics make PARS an excellent candidate for clinical use.

Nonetheless, as a new modality, a need for much development remains. Main areas of improvement are in developments supporting chromophore-selectivity (allowing PARS to be usable as a clinical diagnostic tool) and feature extraction (improving PARS imaging contrast and extracting as much information from PARS images / time-domain signals as possible for potential improvements to chromophore-selectivity and for enriching the presentable informational content PARS microscopy affords).

1.2 Thesis Objectives

The overarching goal of this thesis is to research and make developments to PARS microscopy for the purpose of enabling chromophore-selective imaging. The ability to differentiate chromophores is one of the main advantages of photoacoustic microscopy compared to other modalities, and thus developing this attribute in PARS is a logical research step.

Several chromophores are of clinical importance. For example, the concentrations of oxy- and deoxyhemoglobin (HbO_2 and Hb) are necessary to estimate blood oxygen saturation (sO_2) [5]. The ability to quantitatively image HbO_2 and Hb and subsequently estimate blood oxygenation at the capillary-level is of great importance in ophthalmology as a diagnostic tool for several diseases [64], cancer-related research [65, 66], and pharmaceutical research [67]. Additionally, the ability to selectively image biological structures that would conventionally be stained by hematoxylin and eosin (H&E) [68], the gold standard in histopathology, for viewing on a bright-field microscope for cancer diagnosis and resection surgery, could eliminate time-consuming steps in resection surgeries and may allow PARS to be used as an intraoperative microscope. Indeed, much effort has been made in recent years towards achieving this goal in PARS [53, 55, 56, 58, 60, 69, 70].

The objective of enabling chromophore-selective imaging is achieved in three main ways:

1. Development of an unmixing algorithm with specific considerations for minimizing estimation error covariance, while addressing multiple effects inherent to PARS microscopy.
2. Comprehensive assessment and characterization of Stimulated Raman Scattering in optical fiber as a means for generating multi-wavelength light from a single-wavelength source. This multi-wavelength light source is necessary for efficient multi-wavelength imaging, offering wavelength tunability and high pulse repetition rates, allowing for future work on chromophore unmixing.
3. Development of a feature extraction technique for PARS microscopy, capable of labelling biologically significant structures in histological samples.

1.3 Thesis Organization

This thesis is organised as follows:

- **Chapter 2** gives background and literature review on the subjects of general photoacoustic microscopy, PARS microscopy, inverse problems, optical effects in optical fibers (including stimulated Raman scattering), techniques for extracting information from PARS time-domain signals, and finally dimensionality reduction & feature extraction in the context of signal processing.
- **Chapter 3** explores chromophore-selectivity, or unmixing, in PARS through multi-wavelength imaging. Firstly, unmixing is posed as an inverse problem, and a theoretical forward model is constructed. Considerations and constraints are added to the model to improve accuracy. A study seeking to optimally select wavelengths for given pairs of chromophores is conducted. Finally, initial unmixing results are presented, illustrating both the achievements and shortcomings of the proposed method.
- **Chapter 4** discusses experiments and simulations involving stimulated Raman scattering in optical fiber as a means of multi-wavelength light generation. This effect

is comprehensively studied in several fiber types. Measured results are compared to expectations based on theory as well as simulated results.

- **Chapter 5** explores techniques for extracting information beyond what the current state-of-the-art methods are able to from PARS time-domain signals. Imaging contrast is improved, revealing tissue structures that otherwise lacked definition. A feature extraction technique is developed, which enables differentiation and labelling of various tissue structures.
- **Chapter 6** concludes the thesis and provides thorough insights and discussion regarding the direction of future work.

Chapter 2

Background & Literature Review

This chapter gives background and literature review on a range of subjects including general photoacoustic microscopy, PARS microscopy, inverse problems, optical effects in optical fibers, existing techniques for extracting information from PARS time-domain signals, and finally dimensionality reduction & feature extraction in the context of signal processing.

Although this thesis is about developments towards chromophore-selectivity in PARS, each contribution relies on background in other areas. For context, a broad background on the photoacoustic effect, photoacoustic microscopy, and PARS is given before going into the detailed background required for each of the works presented in this thesis.

2.1 Photoacoustic Microscopy

This thesis mainly focuses on addressing problems of unmixing and feature extraction in Photoacoustic Remote Sensing (PARS) Microscopy, with applications of *in-vivo* blood oxygen saturation estimation and virtual labelling of unstained tissue, rather than the detailed physics underlying the photoacoustic effect and PARS. As such, this section seeks to give the reader a broad background on the field of photoacoustic microscopy and to present a model for PARS as a basis for understanding the measured PARS signals (as a function of the target) that will be worked with throughout the thesis.

The photoacoustic effect was discovered in 1880 by Alexander Graham Bell [71]. This discovery enabled the generation of acoustic waves from absorbed light. Based on this effect, the field of photoacoustic microscopy (PAM) emerged [44, 45]. The premise is that a modulated, highly intense light source is directed onto the target or tissue being imaged. Due to the intrinsic or extrinsic (added) absorption characteristics of the target, a portion of the light is absorbed. This results in a localized rapid rise in temperature. Thermo-elastic expansion ensues, resulting in the generation of ultrasonic waves. These waves propagate to the surface of the target where they are measured using an ultrasonic transducer [45, 72].

Generally, light from a pulsed laser is focused onto target in what is known as optical-resolution PAM (OR-PAM) [73–76], whereby the imaging resolution is limited by the focal spot size of the light beam. The case where the acoustic focus of the transducer is tighter than the focal spot of the light beam is known as acoustic-resolution PAM (AR-PAM) [77, 78]. To achieve greater penetration depth in the target, the modulated light source can be made quite diffuse to provide wide-field illumination. This configuration is known as photoacoustic tomography (PAT) [79–81]. Again, the resolution of the system is determined by the ultrasonic transducer. Note that in all cases, an ultrasonic transducer, physically coupled to the target through contact, is used to detect the induced sound waves. This coupling is required because the acoustic impedance mismatch between tissue and air is too great for effective sonic transmission in a non-contact manner.

The requirement for contact-based detection can be problematic. To accommodate this, targets are often immersed in a coupling fluid such as water, or an ultrasonic couplant gel is used. This is acceptable for certain situations — such as with smaller targets including rodents where either a part or the whole of the animal can be immersed in water — but is cumbersome at best for practical applications involving humans. Generally, PAM systems for imaging thick tissues require a reflection-mode arrangement, where both the excitation light and the acoustic detection system are placed on the *same* side of the tissue. This calls for complex ultrasonic transducers, such as ring arrays [82–84], allowing the excitation light to transmit through the central opening to the tissue. Furthermore, because the transducer is in contact with the tissue, bulk is added, causing difficulty in intra-operative usages.

Several non-contact detection schemes have been proposed for measuring photoacoustic pressures [50, 85–92]; however, the majority rely on interferometry to detect microscopic vibrations at the surface of the tissue, imposing a significant limitation to the architecture. The main limitation is a susceptibility to errors resulting from small changes in the optical path length. When imaging living targets, subtle movements resulting from breathing and even the pulse of the heart become significant. Photoacoustic Remote Sensing (PARS) [3, 50–63] pioneered by Parsin Haji Reza, is an all-optical, non-interferometric implementation of photoacoustic microscopy. Much information is given in the seminal article [50]. PARS offers a non-contact method of imaging absorption contrast, providing the benefits of PAM without its disadvantages.

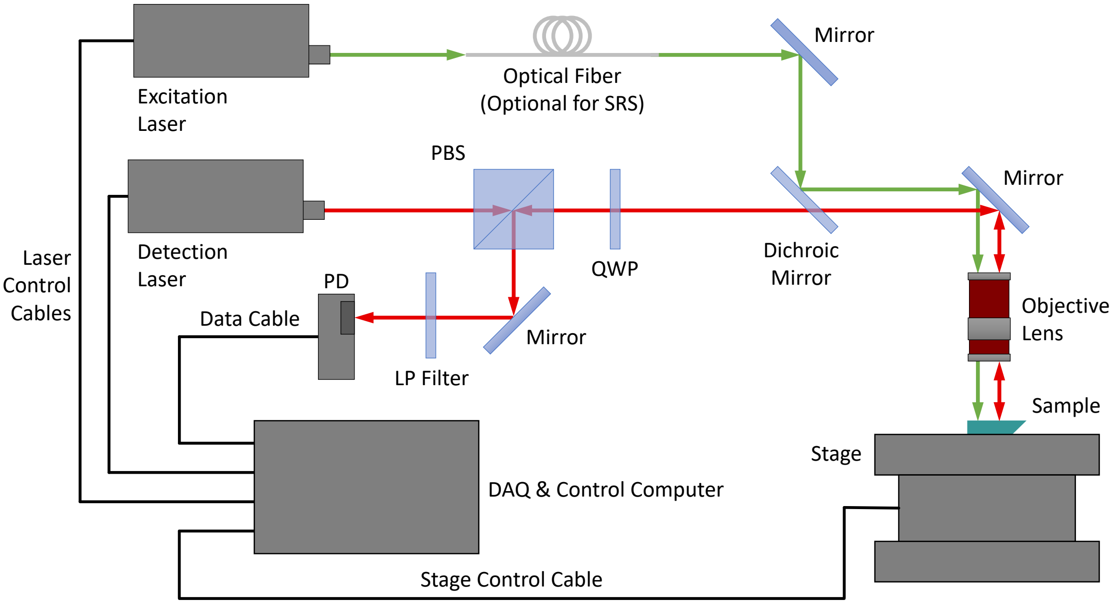


Figure 2.1: A simplified schematic of the PARS microscope is shown here. Abbreviated component labels are: Polarizing Beam Splitter (PBS), Quarter-wave Plate (QWP), Photodiode (PD), Long-pass Filter (LP Filter), and Data Acquisition Card / System (DAQ). Briefly, light from both the excitation and detection lasers are combined (via the dichroic mirror) before reaching the imaged sample. The back-reflected (notice direction of arrowheads) detection beam is directed to the photo-diode (using polarization, controlled by the polarizing beam splitter and the quarter-wave plate), where it is measured as a time-domain signal and recorded by the DAQ. Additional details are given in [50].

Briefly, PARS replaces the ultrasonic transducer of conventional PAM methods with a co-focused continuous-wave laser, known as the *detection* laser. For reference, a simplified schematic of the PARS microscope is shown in Figure 2.1. When excitation light is absorbed, localized heating and subsequent pressure generation via the photoacoustic effect occurs. The generated pressure and heat results in modulations to the local refractive index, thereby modulating the amount of back-reflected detection laser light. These modulations are detected by a photo-diode and are recorded digitally by a data acquisition (DAQ) card. Often, an in-line electronic band-pass filter is employed to remove high-frequency noise and zero the DC level of the signals. More details regarding signal formation are given in Section 2.1.1.

The absorption characteristics of biological chromophores (absorbers) vary by wavelength, and can be quite specific [45, 72]. This means that specific chromophores can be targeted by appropriately selecting the excitation wavelength. Examples of such chromophores / bio-media include melanin, hemoglobin, DNA and lipids. Beyond structural imaging applications, the ability to target specific chromophores enables functional imaging, revealing information on physiological activity / parameters such as blood oxygen saturation (sO_2) [5]. Further background on this is presented in Section 2.2. Again, selecting specific chromophores based on their absorption requires the use of a multi-wavelength light source. To produce multi-wavelength light, Stimulated Raman Scattering (SRS) [93, 94] can be used. This effect is introduced in Section 2.3.

2.1.1 Photoacoustic Remote Sensing Signal Formation

This section describes the formation of PARS signals. Understanding this mechanism is important for Chapter 5, where the characteristics of PARS signals are analysed.

A PARS time-domain signal begins with the interaction between the focused continuous-wave detection laser beam and the imaged target. Some of the light striking the target is reflected backwards, propagating back through the microscope objective and making its way to the detection photo-diode. The ratio of back-reflected light can be calculated based on the intensity reflection co-efficient, R. In the ideal case of normal incidence on the target,

$$R = \left| \frac{n_2 - n_1}{n_2 + n_1} \right|^2, \tag{2.1}$$

for a medium with refractive index n_1 above the target (air, oil, or water), and a target with refractive index n_2.

When the excitation laser pulses, light is directed to the target, where a portion of that light is absorbed, according to the absorption spectrum or absorption coefficient at the excitation wavelength, $\mu_a(\lambda_{ex})$. When light is absorbed, energy is transferred to the target, resulting in localized heating and subsequent pressure generation via the photoacoustic effect [71]. The generated pressure and heat results in modulations to the local refractive index, thereby causing modulations in the amount of back-reflected detection laser light. If the refractive index is modulated by δn, then the modulation in the intensity reflection co-efficient is

$$\Delta R = \left| \frac{n_2 + \delta n - n_1}{n_2 + \delta n + n_1} \right|^2 - \left| \frac{n_2 - n_1}{n_2 + n_1} \right|^2. \tag{2.2}$$

After some algebraic manipulation and approximations [50], this becomes

$$\Delta R \approx \delta n(n_2 - n_1). \tag{2.3}$$

When a refractive index step is present at the boundary of the absorber, then the back-reflected intensity modulations are approximately proportional to the modulations in the absorber's refractive index, δn, which itself is proportional to the generated initial pressure. However, in the case that a refractive index step is *not* present, i.e., $n_2 - n_1 = 0$, the intensity modulations become dependant on higher-order terms beyond what the approximation in equation Equation (2.3) shows [50].

The modulations in the back-reflected detection laser light are detected and electronically amplified by the detection photo-diode, before being sampled over time and recorded digitally by a data acquisition (DAQ) card. For the purposes of forming an image, these signals need to be projected to scalar values. On page 24 of Section 2.4, conventional projection methods are described.

In more detail, the change in refractive index, δn, is described by the elasto-optic relation [50, 95], and is dependent on several parameters including the elasto-optic coefficient, η (about 0.32 for water), pressure, $p(x, t)$, at position x and time t, mass density, ρ, and the speed of sound in the medium, ν_a. According to the elasto-optic relation,

$$\delta n = \frac{\eta n_1^2 p(x,t)}{2\rho\nu_a^2}. \tag{2.4}$$

Note that the initial photoacoustic pressure, p_0, is a function of the optical excitation fluence, ϕ_{ex}, the optical absorption coefficient, μ_a, and the material-dependant Gruneisen parameter, Γ, [5] such that

$$p_0 = \Gamma\phi_{ex}\mu_a. \tag{2.5}$$

If one considers optical absorption saturation — a non-linear effect, normally seen as a limitation within the field of PAM, but that has been taken advantage of to aid in blood oxygen saturating estimation [96] — the optical absorption coefficient can be modified to

$$\mu_a = \frac{\mu_{a0}}{1 + \frac{I}{I_{sat}}}, \tag{2.6}$$

for nominal absorption, μ_{a0}, optical intensity, I, and saturation intensity, I_{sat}.

Overall, Equations (2.3) to (2.6) can be combined to form a generalized model describing the modulations in the back-reflected detection laser light:

$$\Delta R \approx \frac{\eta n_1^2 \Gamma \phi_{ex} \left(\frac{\mu_{a0}}{1 + \frac{I}{I_{sat}}} \right)}{2\rho\nu_a^2}(n_1 - n_2). \tag{2.7}$$

While this model provides more detail than is often practically usable, given that many of the material properties are unknown in applications where PARS is used, it is provided here for completeness.

2.2 Inverse Problems and Unmixing

This section discusses the theory of forward and inverse processes, and specifically deriving inverse models for the purpose of estimation [97–102]. The theory discussed here is applied in Chapter 3 where chromophore unmixing for PARS is explored.

General Inverse Problems

Consider a physical process whereby some underlying information or characteristics determine or lead to other observable characteristics. This is known as a forward process [97]. The underlying information is often referred to as the *state*, represented here by x, and the *observations* or *measurements* are represented here by y. For example, consider a wooden bar on a xylophone. The bar has some characteristics — the size or shape, thickness, stiffness, density, etc. — which determine how it will vibrate when struck by a mallet. The forward process here is the manifestation of the characteristics of the bar in the sound the bar makes when played.

Forward problems are closely related to forward processes and perhaps differ only technically. Given a model, $f(x)$, a mathematical description of a forward process, the act of predicting the observable characteristics is a forward problem. Continuing with the same example, the forward problem would be the act of predicting the sound a xylophone bar would make when it is struck, based on the knowledge of its underlying characteristics (geometry, stiffness, etc.), x. Forward problems tend to be *relatively straightforward* in the sense that if an accurate model is available, predicting derived quantities (i.e., an estimate $\widehat{y}$, denoted by the hat symbol above) is as simple as direct evaluation of the model $\widehat{y} = f(x)$.

By contrast, inverse problems are the act of inferring the underlying characteristics of some physical system based on observations produced through a forward process [97]. This would be the act of estimating the underlying characteristics, x, of a xylophone bar based on an observed sound it made when struck, y. Inverse problems tend to be *relatively difficult*. Almost no forward processes are invertible; however, most inverse problems are still tractable and estimated solutions can be found. In general, models are only an approximation of the real world, neglecting (losing) some information. Mathematically, functions (forward models) are invertible if and only if they are injective and surjective (i.e., bijective), meaning that they are one-to-one on their domain. For example, if $f(x_i) = f(x_j)$ for $x_i \neq x_j$, then $f(x)$ is not invertible, and the information differentiating x_i from x_j is lost. Note, mathematical problems where a unique solution exists are referred to as being *well-posed*, whereas problems that either have no solution or have many (non-unique) solutions are termed *ill-posed* [97]. Additionally, noise, ν, is often present in measured observations, i.e., $y = f(x) = g(x) + \nu$, thus corrupting the observations to some extent and presenting an additional challenge. In any forward process that removes information (either because it is not bijective, or because noise is present), the forward model is not invertible. In these cases, the inverse problem is approached by *estimating* the solution.

Unmixing Problems

A class of forward processes exists, known as mixing problems, whereby several signals, perhaps represented as entries of a vector, $\vec{x} = [x_1, \ldots, x_n]^T$, are combined and are observed as a mixed signal: $y = f(\vec{x})$. A simple example of this would be several sounds occurring at once and being heard together. In this case, the forward process, $f()$, often referred to as a mixing model, would be well represented as a weighted sum of the individual sounds / signals,

$$f(\vec{x}) = \sum_{\forall i} \beta_i x_i, \tag{2.8}$$

with weights, β_i, determined, for example, by the proximity to the sound source. In fact, a classical example of this is The Cocktail Party Problem, defined initially by E. C. Cherry [103, 104] and studied by many others [105, 106]. Imagine that at a cocktail party there are many people standing around talking. Our ears pick-up everyone's voices at once; however, if we try to listen to any particular person, our brain is amazingly capable of ignoring the other voices and somewhat isolating the single speaker we wish to hear (i.e., isolating just one of x_i). Here, the forward (mixing) process is the sonic combination of everybody's speech, and the inverse problem is the isolation of a particular speaker's voice. The inverse problem of isolating components of a mixed signal is referred to as signal *unmixing*.

As was mentioned, the goal when solving inverse problems is to isolate the individual components of a mixed signal. More specifically, in the case that the characteristics of each of the constituent parts is known ahead of time (prior information), the goal becomes to estimate the *abundance*, or relative abundance, of each constituent part. In PARS, generally the concentration, expressed as c_i for absorber i, rather than abundance as such, is desired. Thus, in the context of PARS, we replace $\vec{x}$ with $\vec{c}$. For example, for oxyhemoglobin, $\mathrm{HbO_2}$, the concentration would be expressed as $c_{\mathrm{HbO_2}}$. Note that concentration is related to abundance through the volume of a defined region, assuming concentration is homogeneous in the region. For abundance a and volume V, the concentration is $c = a/V$. In the context of PARS, the region can be thought of as the excited region within a sample.

The type of information generally available beforehand when solving inverse problems is the *spectral* characteristics of the targets. In daily life, we can relate this concept to the colour of objects around us, and our ability to discern, separate, or unmix them; for example, telling ripe vs. unripe fruit apart based on their colour. Another excellent example is discerning the instruments heard in a song. Each instrument has unique spectral and time-varying characteristics, known as the *timbre*, largely determined by the relative strengths of the harmonics of the fundamental tone being played. In the context of photoacoustic microscopy, where the main contrast mechanism is based on optical absorption, the targets' absorption spectra are the main characteristic by which they may be discerned. By definition, for a single chromophore, the absorption coefficient, μ_a, is

$$\mu_a = \varepsilon c, \tag{2.9}$$

for optical extinction ε and concentration c. The absorption coefficient is often expressed as a function of wavelength, λ, thereby defining the absorption *spectrum*, $\mu_a(\lambda)$. The PARS microscope effectively samples the *combined* absorption of several chromophores spectrally at the excitation wavelength, λ_{ex}, constituting a forward (mixing) process. Two relevant absorption spectra are those of oxy- and deoxyhemoglobin, for which the molar extinction spectra, $\varepsilon(\lambda)$, are shown in Figure 2.2.

Due to the duality between spectral and temporal or spectral and spatial measurements, in many cases, either long integration-time measurements or multiple measurements are required to resolve spectral characteristics. In other words, generally speaking, the temporal and the spectral sensitivity of any instrument are inverses of each other, related to the fundamental limitation referred to as the uncertainty principle [108]. The PARS microscope has spectral sensitivity determined by the spectrum of the excitation laser used. Generally, a narrow-band excitation is used, thus allowing for precise sampling of the absorption spectra of targets. However, to reveal information from more of the absorption spectra, and to ultimately discern target absorbers, additional measurements at other excitation wavelengths are required. Chapter 4 studies Stimulated Raman scattering, a means of generating broad, comb-spectrum light from a narrow-band light source, namely a 532 nm laser. The peaks from the generated comb-spectra can be isolated via band-pass filtering for the purpose of spectral sampling in PARS and subsequent unmixing.

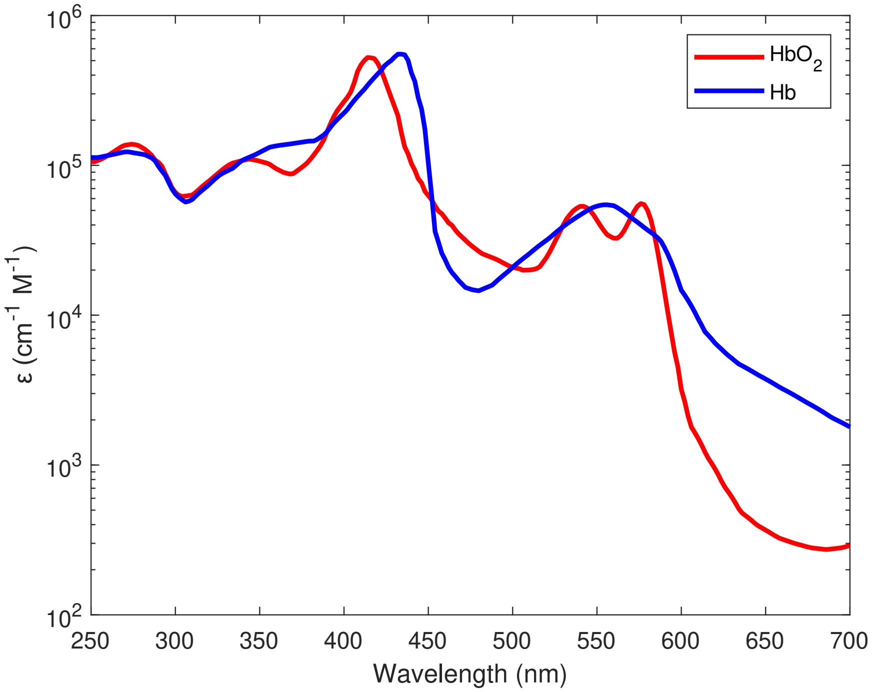


Figure 2.2: The molar extinction spectra of oxyhemoglobin (red) and deoxyhemoglobin (blue). Note that the absorption coefficient, μ_a, is related to extinction, ε, through concentration, c, via Equation (2.9). Data sourced from [107].

Unmixing with PARS

As was explained, the forward process in PARS is the measurement of optical absorption. Equation (2.9) defines the absorption coefficient for a *single* absorber; however, in practice (especially in situations where unmixing is desired), there are *several* absorbers present. To accommodate for that, Equation (2.9) can be modified to include the contributions of several absorbers via a sum, weighted by the concentration of each absorber. Thus, the combined absorption coefficient is

$$\mu_a = \sum_i \varepsilon_i c_i, \tag{2.10}$$

for absorbers indexed by i. Note that up to an arbitrary proportionality constant, α, wholly dependant on the characteristics of the specific PARS microscope used, the PARS signal amplitude, s, (left without concrete definition here), is equivalent to the combined absorption: $s = \alpha\mu_a \overset{\propto}{=} \mu_a$, where the notation, $\overset{\propto}{=}$, signifies equality for some arbitrary proportionality. A suitable definition for the PARS signal amplitude is not inherently obvious and is touched upon in Section 2.4 and further explored in Chapter 5 in the context of feature extraction and in Section 6.1 as future work. Equation (2.10) forms a basic forward model for PARS. Again, as was explained, to gain information encoded by the absorption spectra of the targets, imaging at multiple excitation wavelengths is necessary. Thus, to extend Equation (2.10) such that PARS measurements are made at multiple wavelengths, a matrix-based representation is used,

$$\vec{s} \overset{\propto}{=} \vec{\mu}_a = \begin{bmatrix} \mu_a(\lambda_1) \\ \vdots \\ \mu_a(\lambda_j) \\ \vdots \\ \mu_a(\lambda_m) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(\lambda_1) & \cdots & \varepsilon_i(\lambda_1) & \cdots & \varepsilon_n(\lambda_1) \\ \vdots & & & & \\ \varepsilon_1(\lambda_j) & & \ddots & & \vdots \\ \vdots & & & & \\ \varepsilon_1(\lambda_m) & & \cdots & & \varepsilon_n(\lambda_m) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_n \end{bmatrix}, \tag{2.11}$$

or more simply,

$$\vec{s} \overset{\propto}{=} \vec{\mu}_a = E\vec{c}. \tag{2.12}$$

E is a matrix of extinction coefficients; however, in the context of general unmixing problems, it is referred to as the *mixing matrix*, describing how the underlying sought-after variables (c_i) are mixed or combined in the observed measurements. Framing PARS as a general mixing / unmixing problem, we have the forward model defined as

$$\vec{s} = f(\vec{c}) \overset{\propto}{=} E\vec{c}. \tag{2.13}$$

To solve, if the forward model is invertible, the matrix inverse can directly be applied. In this case the estimated concentrations, $\widehat{\vec{c}}$, would be given by

$$\widehat{\vec{c}} = E^{-1}\vec{s}. \tag{2.14}$$

If the system is over-constrained (more measurements than unknowns), then the pseudo-inverse [109–111] can be applied to produce an estimate,

$$\widehat{\vec{c}} = E^{+}\vec{s}, \tag{2.15}$$

analogous to a best-fit solution. Note that the problem may be ill-posed, in the case that measurements are not consistent with the forward model (no solution exists); however, the pseudo-inverse, as in Equation (2.15), can be applied, yielding the closest solution that is consistent with the forward model. Another way the problem may be ill-posed is if the system is under-constrained — meaning that too few measurements are present relative to the degrees of freedom, and therefore the system has (many) non-unique solutions. Again, the pseudo-inverse can be applied in this case, yielding in the least-norm solution from the set of all solutions [97]. Alternatively, constraints based on knowledge of the problem could be included to arrive at an improved estimate. A solution via Tikhonov regularization can then be employed [112, 113] by reformulating as a least squares problem with added constraints. The solution would be found by minimizing an expression of the form

$$\widehat{\vec{c}} = \arg\min_{\vec{c}} \Big\{ \|\vec{s} - E\vec{c}\| + \sum_i w_i \Psi_i(\vec{c}) \Big\}, \tag{2.16}$$

for constraint weights w_i and functions $\Psi_i()$ imposing constraints on $\vec{c}$.

Real-world targets imaged by a PARS microscope, especially in future clinical / pre-clinical settings, are complex in the sense that they are composed of *many* chromophores, each with their own absorption characteristics. To extract clinically relevant information from the PARS images of such targets, chromophore-selectivity is necessary. In histological tissue samples, when imaged using a 266 nm (UV) excitation laser, the extreme contrast between the absorption peak of DNA at UV wavelengths and the absorption of other chromophores means that the images *dominantly* show DNA, or cell nuclei, contrast. This contrast resembles that of hematoxylin staining, which normally colours cell nuclei a purple-blue [68]. To mimic H&E staining, contrast from cytochromes in the extracellular matrix is required. This remains a challenge that may be solved through multi-wavelength imaging and subsequent unmixing. As mentioned in the introduction (Chapter 1), one of the main applications areas of PARS where unmixing is required is the estimation of blood oxygen saturation (sO_2) at the scale of micro-vasculature, and *in-vivo*. This has an abundance of applications in the field of ophthalmology [64, 114–117], as well as cancer-related [65, 66] and pharmaceutical research [67]. Although the act of calculating sO_2 is rather trivial,

$$sO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb}}, \tag{2.17}$$

the real challenge here is in estimating the concentrations of HbO_2 and Hb.

The task of estimating sO_2 has been approached for more than a decade by many research groups within the broad field of photoacoustic microscopy [3, 62, 63, 79, 96, 118–130]; however, nearly all reports use *contact-based* PAM. Indeed, many advancements have been made with PAM including *in-vivo* real-time functional sO_2 measurement [122] and highly detailed single red blood cell sO_2 imaging [121], able to resolve sub-cellular details. Nonetheless, reviews [79, 123] on the subjects of unmixing and functional photoacoustic microscopy explain that many challenges still remain, including spectral colouring (whereby excitation light is absorbed by chromophores other than hemoglobin, degrading measurements of blood oxygenation), noise from detection electronics, image reconstruction artefacts, laser power fluctuation, and animal motion. All of these challenges exist for sO_2 estimation with PARS and will need to be addressed as the technology is developed.

In contrast to the large quantity of work on sO_2 estimation using contact-based PAM, relatively few reports [3, 62, 63, 118–120] using PARS (non-contact) have been made. The author of this thesis was involved with most of these. Given that PARS was only recently pioneered, sO_2 estimation remains a challenge that requires further research and development. Chapter 3 broadly explores chromophore unmixing in PARS and in Section 3.3 applies the developments to *in-vivo* sO_2 estimation.

2.3 Optical Effects in Optical Fibers

This section discusses several optical effects which occur in optical fibers. The study of one particular effect, stimulated Raman scattering, is the primary subject of Chapter 4 due to its utility in producing multi-wavelength light, enabling multi-wavelength imaging. In Chapter 3, the use of stimulated Raman scattering for multi-wavelength imaging in PARS is discussed.

It is assumed that the reader has a basic familiarity and understanding of optics when reading this section; however, if not, [131, 132] are good starting points.

2.3.1 Optical Fiber

Optical fiber is a thin cylindrical conduit for light, which allows light to propagate from one end to the other based on the principle of total internal reflection. More specifically, optical fibers act as wave-guides for light. Optical fibers have an internal core at one refractive index (n_1), a cladding surrounding it at a smaller refractive index (n_2, where $n_1 \geq n_2$), and finally a protective sheathing or jacket on top of that. Fibers of this type are called step-index fibers, but graded-index fibers are also available, in which the refractive index from the core to the cladding changes gradually. In this thesis, only step-index fibers are used. Generally, optical fibers are made from fused silica with added dopants, used to alter the refractive index in the core and cladding.

As light propagates through optical fiber, some power is lost. These losses are characterized by the fiber attenuation. Fiber attenuation is mostly due to Rayleigh scattering caused by non-uniformities frozen into the fiber [94]. The Rayleigh cross section has $1/\lambda^4$ dependence [133], meaning that for short wavelengths Rayleigh scattering and attenuation is high, whereas for near-infrared (NIR) and infrared (IR) it is relatively low. This makes optical fiber more suitable for IR transmission, hence the widespread use of IR for digital optical communication. Interestingly enough, this is also the dominant reason that the sky appears blue: short-wavelength light such as UV and blue light are scattered, whereas other wavelengths are transmitted. Furthermore, sunrises and sunsets are reddish because we observe light that has transmitted through a large section of the atmosphere, causing short-wave components to be scattered (removed), leaving behind only the long-wave components we see as orange and red. Regarding attenuation in optical fiber, in addition to Rayleigh scattering, there are other factors such as molecular absorption with specific absorption bands associated with various bonds; however, these details are of little importance for the work presented in this thesis.

Formally, fiber attenuation is defined based on the following. For power, P_0, at the proximal end of a fiber of length L, the transmitted power (observed at the distal end), P_{T}, is given by

$$P_{\mathrm{T}} = P_0 \exp\left(-\alpha L\right), \tag{2.18}$$

where α is the attenuation constant of the fiber, encapsulating losses from all sources. It is common to report attenuation in units of dB/km, using

$$\alpha_{dB} = -\frac{10}{L}\log_{10}\left(\frac{P_{\mathrm{T}}}{P_0}\right) \approx 4.343\alpha, \tag{2.19}$$

to relate α_{dB} to α [94].

Another effect which occurs in optical fibers is chromatic dispersion: the separation of light by wavelength or frequency. This is most readily seen in white light, separating into its constituent wavelengths after transmission through a prism, or through droplets of water in the formation of rainbows. This is also a major effect in lenses where, generally, designers attempt to mitigate it for improved image quality. Essentially, the index of refraction has wavelength-dependence, meaning that light of different wavelengths travels at different speeds within any given medium and thereby splits. In optical fibers, this means that a multi-wavelength pulse of light travelling down the fiber will separate by wavelength, each arriving at the end of the fiber at slightly different times. In most media, longer-wavelength light travels faster than shorter-wavelength light. This is called normal chromatic dispersion. It is also possible for the opposite to occur, which is called anomalous chromatic dispersion. Chromatic dispersion in optical fibers tends to present itself as an issue for relatively short pulses travelling in relatively long fibers. As will be discussed in Chapter 4, this effect becomes significant for pico-second-scale pulses in fibers longer than 1 m, but it is generally not significant for nano-second-scale pulses in fibers as long as 100 m [94]. This effect is prominent in digital optical communication, where fibers are significantly longer than those used and discussed in this thesis.

Due to the wave-nature of light, propagation in optical fibers occurs in vibrational modes of the electro-magnetic field. These vibrational modes are much akin to those of a drum-head, an elastic membrane stretched over a circular rim. The motion of the membrane is constrained at the edge / rim, thus causing vibrational waves propagating outwards from the center to be reflected back inwards, leading to resonance and the emergence of characteristic vibrational modes based on the geometry and mechanical attributes of the membrane. Much like the rim of a drum, an optical fiber constrains the vibrations in the associated electro-magnetic field (light), thus confining the waves to the inside of the fiber and again leading to the emergence of characteristic vibrational modes determined by the wavelength of the propagating light, the geometry of the fiber, and the refractive indices of the core and cladding (akin to the mechanical impedance of a drum head and the rim). For sufficiently narrow (small radius) cores or correspondingly large wavelength light, only the fundamental mode will propagate. This is referred to as single-mode operation. Where multiple modes propagate is referred to as multi-mode operation.

2.3.2 Raman Scattering

Generally, refractive index is thought of as having real and imaginary parts, describing the speed and attenuation of a light wave as it travels though a given medium, and is dependant on the vacuum wavelength (λ) or equivalently, the angular frequency (ω). This is valid for linear scattering processes such as Rayleigh and more generally, Mie scattering. However, all dielectrics become non-linear for intense electromagnetic fields, and thus there exists an intensity-dependence of the refractive index. This is explained by higher-order (namely third order in the case of silica) electric susceptibility terms [134]. This non-linearity gives rise to non-linear scattering, where the scattered wavelength is not equal to the incident wavelength. In addition, *inelastic* scattering is also possible, where energy is transferred from the optical field to the medium, which is inherently a non-linear process (meaning the wavelength changes here too). Raman scattering is an example of inelastic scattering, whereby an incident photon (referred to as pump radiation) is annihilated to create a photon at a lower frequency (referred to as Stokes radiation) and an optical *phonon*, vibrational energy gained by the non-linear medium. This effect was discovered by C. V.

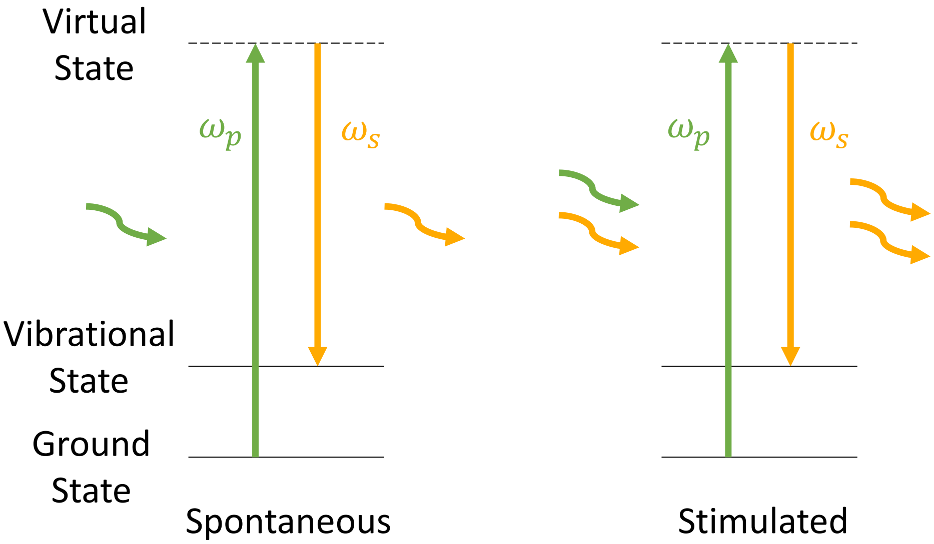


Figure 2.3: Jablonski diagrams of Spontaneous Raman Scattering (left) and Stimulated Raman Scattering (right) Processes. In the spontaneous case, a single pump photon interacts with the medium, resulting in a single Stokes photon being emitted. In the stimulated case, a pump and Stokes photon interact *simultaneously* with the medium, resulting in two Stokes photons being emitted.

Raman in 1928 [93]. This process is illustrated quantum-mechanically via a Jablonski diagram in the left panel of Figure 2.3.

The process of Raman scattering naturally occurs, in a spontaneous manner, scattering a small fraction of photons, corresponding to a factor of roughly 10^{-6} of the optical power [94]. In this case, it is referred to as *Spontaneous* Raman scattering. The Raman gain spectrum, $g_R(\Omega)$, describes the pump-to-Stokes conversion as a function of frequency shift from pump (incident) to Stokes (emitted) waves, $\Omega \equiv \omega_p - \omega_s$. Figure 2.4 shows the Raman gain spectrum for fused silica, the base-material found in optical fibers. The Raman gain spectrum of optical fibers varies depending on the specific dopants used [94].

The Raman gain spectrum of many media has relatively narrow-band peaks, enabling Raman scattering as a spectroscopic tool. This is known as Raman spectroscopy [136–138]. In contrast to this, the Raman gain spectrum of fused silica is fairly broad, as a result of the amorphous (rather than crystalline) structure of fused silica, which in turn causes the spreading of molecular vibration frequencies and a broad Raman gain spectrum [139].

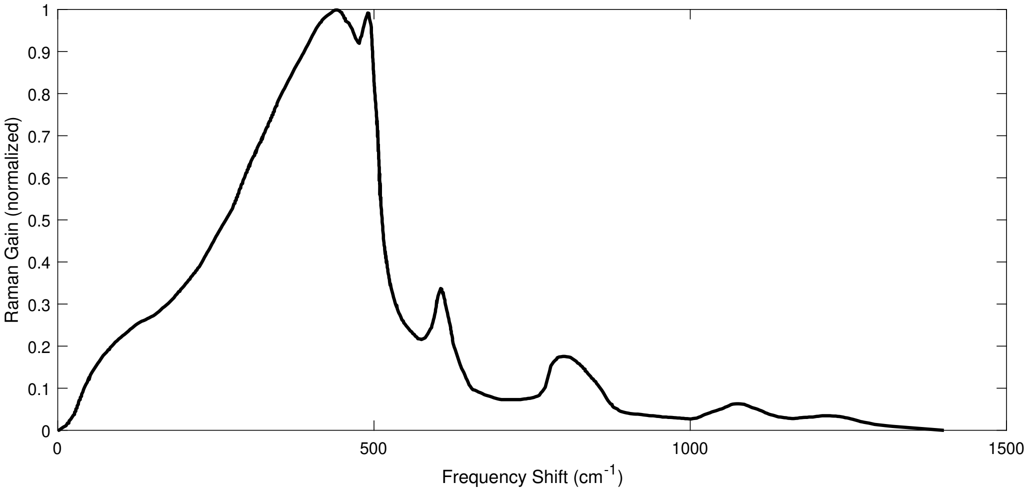


Figure 2.4: The Raman gain spectrum, $g_R(\Omega)$, of fused silica specifies the rate of conversion per unit length for specific frequency shifts, Ω. Plot adapted from [135]. This can be thought of as being analogous to the likelihood of specific frequency shifts occurring as a result of Raman scattering. Notice the dominant peaks at a shift of roughly 400 cm^{-1} or 13 THz. The gain spectrum is normalized such that $g_R = 1 \times 10^{-13}$ W/m at a pump wavelength of $\lambda_p = 1$ μm. Note that g_R scales inversely with λ_p.

In addition to the spontaneous case, stimulated emission is also possible, whereby a pump and a Stokes photon interact simultaneously, causing two Stokes photons to be emitted. In this case, the rate of pump-to-Stokes conversion is greatly increased compared to spontaneous Raman scattering [94]. This process is illustrated in the right panel of Figure 2.3 and is referred to as *Stimulated* Raman scattering (SRS).

The growth of the Stokes wave intensity, I_s, over distance, z, can be described by

$$\frac{\mathrm{d}I_s}{\mathrm{d}z} = g_R I_p I_s, \tag{2.20}$$

where I_p is the pump intensity. Additional effects of attenuation and the loss of pump intensity as a result of SRS conversion to the Stokes wave are at play however, and affect the SRS process. Attenuation has already been described in Equations (2.18) and (2.19). The decrease in pump intensity as a result of SRS conversion to the Stokes wave is equal and opposite to the intensity growth, of Equation (2.20), in the Stokes wave, adjusted for the change in frequency (noting that photon energy, $E = \hbar\omega$, is proportional to frequency, and thus for single photon changes in frequency, the change in energy and thus wave intensity must be proportional to the ratio of frequencies, before and after). By including these effects, Equation (2.20) becomes the coupled differential equation

$$\begin{aligned} \frac{\mathrm{d}I_s}{\mathrm{d}z} &= \qquad g_R I_p I_s - \alpha_s I_s, \\ \frac{\mathrm{d}I_p}{\mathrm{d}z} &= -\frac{\omega_p}{\omega_s} g_R I_p I_s - \alpha_p I_p, \end{aligned} \tag{2.21}$$

for attenuation constants α_p and α_s for pump and Stokes frequencies respectively.

In optical fibers, SRS can arise as a result of spontaneous Raman scattering. If high-intensity light causes even a few photons to be Raman shifted, those Stokes photons can then induce stimulated Raman scattering, continuing the process, creating more and more Stokes photons. Raman scattering, and even more so stimulated Raman scattering, is not noticed in every-day life because of the dependence on intensity for these non-linear effects to become prominent. Optical fibers however, create a fairly ideal scenario for Raman scattering and SRS to occur for two reasons:

1. light is confined to a small cross-sectional area (roughly the area of the core for single-mode operation), meaning that the intensity is greatly increased, and
2. fibers allow light to propagate over long distances, making for long interaction lengths and increased probability of initial Raman scattering and subsequent stimulated Raman scattering.

If the intensity of the Stokes wave becomes great enough, photons of the Stokes wave may themselves be Raman shifted, creating a second, higher-order Stokes wave. This creates a *cascading* effect, whereby each new Stokes wave acts as a pump, leading to the creation of additional higher-order Stokes and resulting in a *comb* spectrum.

As was already described, chromatic dispersion in optical fibers causes light of different wavelengths to separate. This can limit the cascading effect, because the SRS process requires both the pump and Stokes waves (different wavelengths) to overlap within the fiber. Chromatic dispersion manifests itself as a mismatch in group velocity between the pump and Stokes waves, v_{gp} and v_{gs}. The length over which the pump and Stokes waves overlap, called the walk-off length, L_W, can be calculated as

$$L_W = T_0 / |v_{gp}^{-1} - v_{gs}^{-1}|, \tag{2.22}$$

where T_0 is the temporal pulse width. Generally, pulse walk-off is not a strong effect except for the cases where long fibers (hundreds of meters) or short pulses (pico-second-scale) are used. For nano-second-scale pulses, walk-off lengths are generally greater than 100 m, longer than the fibers worked with in Chapter 4.

Other non-linear optical effects are known to compete with SRS. In particular, for pulses in the nano-second range or longer, Stimulated Brillouin Scattering (SBS) is the main competing effect, causing light conversion to backwards propagating waves [94, 140]. This effect is fairly insignificant for pulses near to 1 ns, likely having a minimal impact on the studies described in this thesis, where 1.5 ns and 3 ns pulses are used. For sub-nano-second pulses, the effects of SBS are effectively eliminated; however, other 3^{rd} order non-linearities such as self-phase modulation (SPM) [141], cross-phase modulation (XPM) [142], and four-wave mixing (FWM) [143] become the dominant competing effects [94]. These effects will become significant in Chapter 4 where the effects of laser parameters on multi-wavelength generation through SRS are explored.

SRS in single-mode optical fiber has been well-studied by the physics and optics community. Significant works and findings include:

- measurement of the Raman gain spectrum [144],
- the notion of *critical power* [145] (input pump power such that Stokes equals pump intensity),
- a reduction in critical power by up to a factor of 2 through by using polarization-maintaining fiber [146],
- coupled SRS simulation and experimental validation [147, 148]
- transient modelling of SRS, showing a femto-second time-response [149, 150],
- studies on the effects of temperature on SRS, showing colder temperature are generally yield more stable and higher-order Stokes waves [151, 152],
- high-order Stokes wave generation [153], and
- studies on pico-second SRS using infrared light, which show temporal pulse erosion and separation [154–156], an effect observed through simulation later in this thesis, in Section 4.3.2.

Within the field of photoacoustic microscopy, SRS has found significant use in multi-wavelength imaging applications. The first use in PAM [157] proved effective by enabling coloured ink to be differentiated by using multiple excitation wavelengths. Later, the same research group increased the range of generated wavelengths by employing a large mode-area photonic crystal fiber [158]. SRS has also been used to generate several wavelengths for the purpose of multi-focus imaging [159], taking advantage of chromatic aberration in refractive objective lenses. Most prominently, SRS is used in PAM for functional *in-vivo* imaging applications [122, 160, 161]. Indeed, within this thesis, Chapter 3 explores the use of multi-wavelength imaging in PARS to enable chromophore-specificity and functional imaging, and Chapter 4 comprehensively studies SRS for the purpose of creating such a multi-wavelength excitation source for PARS or other imaging modalities.

2.4 PARS Time-Domain Signal Information Extraction and Dimensionality Reduction Techniques

While Section 2.1.1 described the formation of PARS signals (observed as time-domain signals), this section discusses the state-of-the-art methods for extracting information from these signals. Understanding this mechanism becomes critical in Chapter 5 where the shape and characteristics of the time-domain signals are analysed. Additionally, the subjects of feature extraction and dimensionality reduction [162–166] are discussed and become the foundation for the methods developed in Chapter 5.

Before introducing the standard methods of extracting information from PARS time-domain signals, some notation is established to give the broader problems of feature extraction and dimensionality reduction context. For given measurements / data, $\vec{x}_i \in \mathbb{R}^n$ (i.e., n-dimensional data), it is desirable to extract meaningful information $\vec{y}_i \in \mathbb{R}^m$ from it. Abstractly, this is achieved through some transformation, $g()$, such that

$$\vec{y}_i = g(\vec{x}_i). \tag{2.23}$$

Nearly always, the extracted information, $\vec{y}_i$ is of a lower dimension than the original measurements, i.e., $m < n$. In this case the process is referred to as *dimensionality reduction.*

As part of the PARS mechanism (schematic shown in Figure 2.1), optical signals in the back-reflected light of the detection laser carry information to a photo-diode, which converts this signal to an electrical one that is then recorded via a data acquisition (DAQ) card. This constitutes the measurement of PARS time-domain signals. Once this is done, the next step in the image formation pipeline is to extract information from these TD signals. Converting the time-domain signals to scalar values is referred to as time-domain signal projection [50] and is a very extreme case of dimensionality reduction. In the context of the abstract information extraction formulation of Equation (2.23), the dimensionality of the extracted information is $m = 1$. Conventionally, in prior reports of PARS [50, 70, 167], projection has been achieved using either one of two methods:

1. Hilbert Transform, or
2. Maximum minus Minimum of the signal.

Both of these approaches seek to produce values that are proportional to the amount of absorbed energy — a logical objective, given that imaging contrast in PARS is based on optical absorption.

The first method, the Hilbert Transform [168], has been used in a variety of other modalities before its use in PARS, in cases where a signal *envelope* is desired. Some such modalities include ultrasound imaging [169–171], optical-coherence-tomography (OCT) [172–175], and OR-PAM [176–178]. From the Hilbert Transform, the amplitude of the envelope can be extracted by means of evaluating the maximum minus the minimum. Explicitly, for PARS TD signal, $s(t)$, and Hilbert Transform represented by $H\{\cdot\}$, the projection is

$$Proj_{\text{Hilbert}}(s(t)) := \max\Big(H\left\{s(t)\right\}\Big) - \min\Big(H\left\{s(t)\right\}\Big). \tag{2.24}$$

Computing the Hilbert Transform, especially as many as several millions of times to reconstruct a PARS image (which may have 20 million time-domain signals, one associated with each pixel) is a computationally expensive task and tends to yield results fairly similar to those of the second method, Maximum minus Minimum.

The Maximum minus Minimum of the signal is an intuitive method for finding the amplitude. The projection is defined explicitly as

$$Proj_{\text{Max - Min}}(s(t)) := \max\big(s(t)\big) - \min\big(s(t)\big). \tag{2.25}$$

Computationally, it is much less expensive than the Hilbert Transform method, making it more suitable for real-time imaging applications. Nonetheless, this method is inherently sensitive to outliers and noise (the extreme values of a signal could *very easily* be outliers or simply a result of noise), giving rise to some negative aspects of its performance.

The most fundamental shortcoming of both the Hilbert Transform method and the Maximum minus Minimum method is that they reduce entire (high-dimensional) TD signals simply to scalar representations. This is very extreme dimensionality reduction, surely discarding much of the information present in TD signals. Additionally, with only a scalar representation, almost no insight or specific information about the target can be identified after projection. Therefore, there is an unmet need for improved methods for extracting information from PARS TD signals that reduce the dimensionality while retaining the potentially very useful information contained within the TD signals.

State-of-the-art methods [179–181] in the broader field of photoacoustic microscopy seek to extract additional information beyond simply the signal amplitude. Continuing with the abstractly defined information extraction formulation in Equation (2.23), the dimensionality of the extracted information can be $m > 1$. Doing this allows for some level of target specificity post-imaging. These methods are mainly based upon frequency selectivity and principal component analysis (a broadly applicable technique in signal / feature analysis which will be touched upon later in this section on page 27). Article [179] demonstrates a frequency-selective method for contact-based photoacoustic microscopy and tomography. The results show that the spectral power distribution varies with the size of the absorber. Smaller targets (micro-beads, vessels, etc.) tend to exhibit signals of higher frequency content, whereas larger targets tend to exhibit signals of lower frequency content. The method was applied *in-vivo* on a zebrafish larva to select vessels by diameter. Articles [180, 181] propose a feature extraction method for PARS whereby a filter-bank of 20 1-MHz bandwidth band-pass filters are applied to the measured PARS time-domain signals. The filtered signals are then projected via the Hilbert transform projection, defined in Equation (2.24). This results in a feature vector, 20 elements long, for each time-domain signal. Principal component analysis is then applied to the set of feature vectors. The top two principal components are extracted, and the claim is made that the first component corresponds to parts of the image that would be stained by eosin (from H&E staining) in histological tissue samples, and that the second component corresponds to parts of the image that would be stained by hematoxylin. Through this method, PARS images of tissue can be false-coloured to resemble the results of H&E staining.

For the remainder of this section, the general subjects of dimensionality reduction and feature extraction are discussed, providing the reader with a basis of understanding for the methods applied in Chapter 5.

2.4.1 Dimensionality Reduction

Dimensionality reduction [165, 166] is a useful set of techniques for analysing and interpreting high-dimensional data: tasks which are difficult in the original high-dimensional space. To better understand what that means, consider the contrived example of a data analysis problem where a measured signal exists approximately on a plane (with added noise), embedded in 3-dimensional space. Although it may be possible to view and interpret the measured data through the use of 3D plotting tools, it would be much easier to directly view it in 2-dimensions, the underlying dimensionality of the information present. Thus, dimensionality reduction allows relatively complicated data to be reduced to a smaller set of representative variables.

Using the same notation as in Equation (2.23), for data points, $\vec{x}_i \in \mathbb{R}^n$, in high-dimensional space, dimensionality reduction techniques seek to find a low-dimensional representation, $\vec{y}_i \in \mathbb{R}^m$, such that $m < n$. One common approach for dimensionality reduction is feature learning and subsequent feature extraction. This method encompasses regression-based analysis techniques [165, 166]. Perhaps the best known and simplest version of this is linear regression analysis [182, 183] seeking to mathematically relate a dependant variable (often experimentally measured) to several independent variables (often enforced or set during experimental data collection). In the context of the formulation in Equation (2.23), the dependent variable would be $\vec{y}$ whereas the independent variable is $\vec{x}$. For simplicity in explanation, the dependant variable will be considered to be scalar (i.e., the dimensionality of $\vec{y}$ is 1, and thus can be denoted $y = \vec{y}$); however, this is not necessary for more general cases. The independent variable, $\vec{x}$, can be thought of as being composed of *many* individual independent variables, such that $\vec{x} = [x_1, \ldots, x_n]^T$. Here, a model is assumed — for the simple case of linear regression, it is of the form $y = \beta_0 + \sum_{j=1}^{n} \beta_j x_j + \varepsilon$, where β_j are model parameters and ε is an error term, which must be included in the model to account for any error due to noise between the values of the measured data and the relationship imposed by the model (which is chosen to accurately approximate the underlying relationship of the two variables). Note, for completeness, if $m > 1$ were desired, then a matrix implementation of the model would be used and would be of the form $\vec{y} = \vec{\beta}_0 + B\vec{x} + \vec{\varepsilon}$, for offset term, $\vec{\beta}_0$, matrix of weights, B, and error, $\vec{\varepsilon}$.

Estimates of the model parameters, $\widehat{\beta}_j$, can be learned or statistically inferred using a variety of approaches — the most common and well-known of which is the least squares error regression [184, 185] — assuming training data are available in the form of *paired* data of the form $(y_i, \vec{x}_i)$. Finally, once estimated values for the model parameters have been deduced, any data point comprised of the independent variables, $\vec{x} = [x_1, \ldots, x_n]^T$, can be concisely represented by a single scalar value $\widehat{y} = g(\vec{x}) = \widehat{\beta}_0 + \sum_{j=1}^{n} \widehat{\beta}_j x_j$. This means that information in the data, previously represented through n variables / dimensions, is reduced to a representation by a single scalar variable. Thus, the dimensionality of the data is reduced through this representation.

Tasks of feature learning and feature extraction are often more complex than this [162–166]; however, the underlying concept is always that for an imposed model, learning the model parameters, whether they be cluster centroids in the case of an unsupervised clustering problem or regression coefficients in the case of a supervised learning problem, is referred to as *feature learning*. After this step is complete, the act of actually representing high-dimensionality data in a lower dimensional space, whether that be evaluating cluster membership or evaluating a linear regression model, is referred to as *feature extraction*.

2.4.2 Dimensionality Reduction via Feature Elimination

Another class of dimensionality reduction techniques involves feature elimination. This type of strategy is often used when relatively few dimensions in the data are significant, eliminating (i.e., omitting) the other less significant dimensions.

A prime example of this approach is Principal Component Analysis (PCA) [186–191]. PCA effectively works by evaluating the covariance of a set of data points. Due to the symmetry of the covariance matrix, an alternate rotated basis for the data is implied. This basis has directions aligned with the directions of maximal covariance in the data. These are referred to as the *principal components* of the data. The principal components are then ranked by covariance, with the underlying assumption that dimensions of greater covariance contain more valuable information as opposed to directions of lower covariance which are thought to contain mostly noise. The top few principal components are chosen, and the data are projected onto those directions, using them as a new basis. Note that the number of principal components selected to be kept is left to the discretion of the analyst; although several well-established techniques are available for selecting a suitable number, including various statistical tests, the Akaike information criterion [192], or simply selection based on a threshold of total variance [191, 193, 194]. The other principal component directions are thus eliminated, leaving the data in only the directions of greatest covariance, creating a lower dimensional representation of the data set.

Note that the principal components are in no way representative of the information *underlying* the data. Suppose a forward process, $f()$, exists whereby underlying information, $\vec{z}$, is represented as observed data, $\vec{x} = f(\vec{z})$. For a set of observations, $\{\vec{x}_i\}$, PCA can be performed; however, the directions of the principal components (basis vectors onto which the data, $\vec{x}_i$, is projected to produce the lower-dimensional representation, $\vec{y}_i$) do *not* directly relate components of $\vec{y}$ to components of $\vec{z}$, the underlying information. Instead, components of $\vec{y}$ are simply mixtures of the data, $\vec{x}_i$, along the directions associated with maximal variance. Also note that by reducing the dimensionality of the data, information will almost certainly be lost, and there may be no way to recover the underlying information, $\vec{z}$, from $\vec{y}_i$ even if the forward process, $\vec{x} = f(\vec{z})$, is invertible.

In more detail, PCA evaluates the covariance of a set of p centered data points, $\{\vec{x}_i\}_{i=1}^{p}$, arranged into an $p \times n$ data-matrix, $X = \left[\vec{x}_1 \ldots \vec{x}_p\right]^T$, for data points, in n-dimensional space. The covariance can be thought of as being represented by a hyper-ellipsoid [195], where the size of the hyper-ellipsoid in any given direction represents the covariance in that direction. The covariance is computed in the form of a matrix, $C = \frac{1}{p-1}XX^T$. By performing an eigen-value decomposition, the covariance matrix can be written as $C = W\Lambda W^T$ [189] (orthogonal diagonalization is possible here due to symmetry of C), where the columns of W, the eigen-vectors, describe the direction of the principal axes of the hyper-ellipsoid, or in other words, the direction of the principal components of the data, and where the square-roots of the diagonal entries of $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, correspond to the lengths of the hyper-ellipsoid along the principal axes.

In practice, to avoid numerical loss-of-precision errors that can arise when multiplying XX^T, a singular value decomposition of X is used[1]: $X = U\Sigma V^T$. Here, the right singular vectors, columns of V, are the principal directions, and the diagonal entries of Σ, the singular values, σ_i, are related to the lengths of the hyper-ellipsoid by $\sigma_i/\sqrt{p-1}$.

After applying either the eigen-value decomposition or the singular value decomposition, the principal directions are ranked according to their covariance, the length of the ellipsoid in that direction. Based on the details of the problem, the top few principle directions are chosen, and the data are projected onto those directions, using them as a new reduced-dimensional basis.

2.4.3 Engineered Features vs. Learned Features

Rather than learning features via statistical methods or other algorithms, another option is to *engineer* features. Engineered features are designed / hand-crafted to achieve a specific task in such a way that their meaning is readily explainable and understood [199, 200], in contrast to learned features, which arise through training a model. This is one of the main advantages of engineered features. Compared to learned features, engineered features may not be as effective as learned features, where sometimes enormous amounts of training data are used to learn features, making a highly-accurate and comprehensive model. Conversely, the necessity for training data, especially in exceedingly high quantities, can be seen as a major shortcoming of learning features, and in fact to much of the field of network-based machine learning overall.

In this context, one example of an engineered feature is the amplitude of the PARS time-domain signals, implemented as the maximum minus the minimum value, as shown in Equation (2.25). This feature is engineered specifically for this application, based on the theoretical backing that the signal amplitude is proportional to the amount of absorbed light: the underlying quantity sought-after in PARS microscopy [50].

---

[1] Briefly, the condition number [97, 196–198] of a matrix, $\mathrm{cond}(X) \geq 1$, relates to the numerical stability associated with computations involving the matrix X, where higher condition numbers imply less stability. $\mathrm{cond}(XX^T) = \mathrm{cond}(X)\mathrm{cond}(X^T) = \mathrm{cond}(X)^2 \geq \mathrm{cond}(X)$, therefore, it is advantageous to avoid the multiplication XX^T.

2.4.4 Feature Learning via Clustering

Clustering algorithms can be a method of feature learning. Broadly, clustering is a class of unsupervised learning [201–204], for cases where labelled data are not available. This is in contrast to the situation in which regression was discussed, on page 25. There, paired training data of the form $(y_i, \vec{x}_i)$ were available; however, here, *only* $\vec{x}_i$ is available, the variable representing measured data, introduced in the formulation in Equation (2.23). Clustering provides a means of grouping similar data together, based on a given distance metric, $d()$, yielding representative cluster centroids, $\vec{c}_j$, in addition to the cluster labels, y_i (inherently discrete-valued scalar information), associated with the data points, $\vec{x}_i$. The cluster centroids can be thought of as the learned features here. These can later be used to evaluate the cluster membership of additional, or new / previously unseen, data points.

In more detail, clustering algorithms seek to create clusters where all data points within any given cluster are as similar as possible. This means the *intra*-cluster variance should be low. Additionally, clusters should be as distinct from one another as possible, meaning their centroids should be relatively distant, or the *inter*-cluster distance should be high.

K-Means [205–208] is a clustering algorithm that seeks to group the input data into a pre-specified number (K) of clusters, where the objective is to minimize the intra-cluster variance. Each cluster has an associated representative centroid, $\vec{c}_j$ for $j = 1, \ldots, K$, calculated as the mean of the data points within the cluster. The basic algorithm is as follows:

1. Initialization: Select (randomly, or through other methods) K data points to act as centroids.
2. Assignment: Assign each data point to the cluster of the nearest centroid.
3. Update: Recalculate cluster centroids by taking the mean of all data points within each given cluster.
4. Continue repeating the Assignment and Update steps until no further points change their cluster assignment, or equivalently, when the centroids cease to change.

Other initialization and stopping conditions are possible for variations of this algorithm. The choice of stopping condition is important given that the algorithm is not guaranteed to converge [208]. For this reason, it can be important to set a maximum number allowable of iterations.

It is important to note that while clustering reduces dimensionality, by virtue of representing data simply by its associated cluster label (a one-dimensional representation), it differs significantly from feature elimination methods, such as PCA, discussed on page 27. Recall that PCA projects the data onto a reduced-dimensional basis, thus forming a representation based on combinations of the data based on directions in which the data set exhibits maximal variance. These basis directions, maximal principal components, do not directly relate components of $\vec{y}$ to components of $\vec{z}$, the information underlying the data. In contrast, clustering represents the data (through labeling) in a way where the representation *does* relate specifically to an estimate of the underlying "type" or classification of the data, whether this estimate is actually in any way *correct* or not.

To solidify this concept, consider a contrived example whereby two types of fruit, say apples and bananas, are characterized through measurements, $\vec{x}_i$, of various attributes such as colour, texture, weight, shape, etc. (such that all characteristics have a real number representation for the purpose of this example). The underlying information, the type of fruit, $z \in \{\text{apple}, \text{banana}\}$, is observed via a forward measurement process, $\vec{x} = f(z)$. Looking only at $\vec{x}_i$ does not *directly* inform which fruit a given data point is associated with. PCA is capable of forming a reduced dimensional representation, $\vec{y}_i$; however, there is no direct relationship between the components of this representation and the underlying type of fruit, z. Clustering, via K-Means for example, would result in cluster labels associated with each data point. Assuming it is known ahead of time that only two types of fruit are represented in the data, K would be set to 2, and two clusters would be formed. Although the abstract labels would not specify *which* fruit each cluster is associated with (by name, perhaps), by virtue of grouping data points of similar characteristics together, it is likely that data points of each cluster are directly associated with one of the fruits, the underlying information, z, represented by the data, $\vec{x}_i$. Again, it is worthwhile noting that the cluster labels are only an estimate of the underlying information, and there is no guarantee that the estimate is correct. For example, the wrong number of clusters could be chosen, resulting in spurious cluster labels associated with the data that do not relate to the *true* underlying information.

Chapter 3

PARS Chromophore Unmixing

Chromophore unmixing is valuable in PARS [3, 50–63] because it allows visually similar targets of clinical relevance to be discerned. Without unmixing, images are effectively "gray-scale", and contain only *structural* information. There is no way to tell, from the image, what material / bio-media the various parts of the image are, aside from analysing their structure. By unmixing, various materials and bio-media can be directly discerned and isolated in the image – displayed in isolation, or together in the same image via false-colouring.

In addition, functional information can be extracted through unmixing. This is information relating to physiological activities currently taking place. The main driving application of unmixing in PARS is for functional imaging of blood oxygenation (sO_2). The ability to measure sO_2 at the capillary-level has numerous direct applications in the fields of ophthalmology [64, 114–117], cancer research [65, 66], and pharmaceutical research [67], as was discussed in Section 2.2.

3.1 Selection of Excitation Wavelengths

Section 2.2 in the thesis background introduces the subject of inverse problems and unmixing. It was discussed that a weighted sum of contributions from all absorbers, according to their concentration, c, and absorption, $\mu_a(\lambda)$, at the excitation wavelength $\lambda = \lambda_{\text{ex}}$, is the basis for the PARS signal amplitude s. As a reminder to the reader, Equation (2.12) is repeated here.

$$\vec{s} \overset{\propto}{=} \vec{\mu}_a = E\vec{c}. \tag{3.1}$$

Note that the scalar quantities c, μ_a, and s are shown as vector quantities, representing PARS measurements of mixed chromophores taken at several excitation wavelengths. This equation relates the absorption coefficients, μ_a, to the underlying chromophore concentrations, $\vec{c}$, through their molar extinction coefficients, ε_i, represented in matrix-form as E. Note that the excitation wavelengths are implicit in this form, with each row of E associated with one of the excitation wavelengths. PARS measurements, $\vec{s}$, are equivalent to the absorption coefficients up to arbitrary proportionality ($\overset{\propto}{=}$) determined by the characteristics of the specific PARS system.

This section explores the problem of selecting excitation wavelengths to best unmix a given set of chromophores. In particular, a simplified yet still applicable example of unmixing the chromophores oxy- and deoxyhemoglobin ($\mathrm{HbO_2}$ and Hb, respectively) is explored. Unmixing these chromophores is necessary for subsequent $\mathrm{sO_2}$ estimation. This example includes only two chromophores — or rather, two unknown concentrations, $c_{\mathrm{HbO_2}}$ and c_{Hb} respectively — and as such, as in any other system of equations, a minimum of two measurements are required to solve for these two unknowns (under the usual assumptions of linear independence). Although more measurements can be used to improve the quality of the estimated solution, initially only two are considered, and later considerations for additional measurements will be made.

The mixing matrix, E, from Equation (3.1), becomes the 2×2 matrix

$$E = \begin{bmatrix} \varepsilon_{\mathrm{HbO_2}}(\lambda_{ex,1}) & \varepsilon_{\mathrm{Hb}}(\lambda_{ex,1}) \\ \varepsilon_{\mathrm{HbO_2}}(\lambda_{ex,2}) & \varepsilon_{\mathrm{Hb}}(\lambda_{ex,2}) \end{bmatrix}. \tag{3.2}$$

What remains to be done is to select the excitation wavelengths, $\lambda_{ex,1}$ and $\lambda_{ex,2}$. The formation of matrix E can be thought of as sampling the absorption / extinction spectra of oxy- and deoxyhemoglobin. Figure 3.1 shows the spectra, similarly to Figure 2.2; however, two additional lines at $\lambda = \lambda_{ex,1}, \lambda_{ex,2}$, placed arbitrarily, are included to indicate the sampled wavelengths. The points of intersection between these lines and the spectra make up the entries of E. One can imagine "sliding" the vertical lines to other positions in the spectra, by changing $\lambda_{ex,1}$ and $\lambda_{ex,2}$, and considering the effects this would have on the values of E. The pairs of intersection points must be sufficiently different to avoid cases of linear dependence, or *near* linear dependence in the rows of E.

The problem of selecting $\lambda_{ex,1}$ and $\lambda_{ex,2}$ will be approached by considering the following three metrics:

1. The condition number of a model of the forward process.

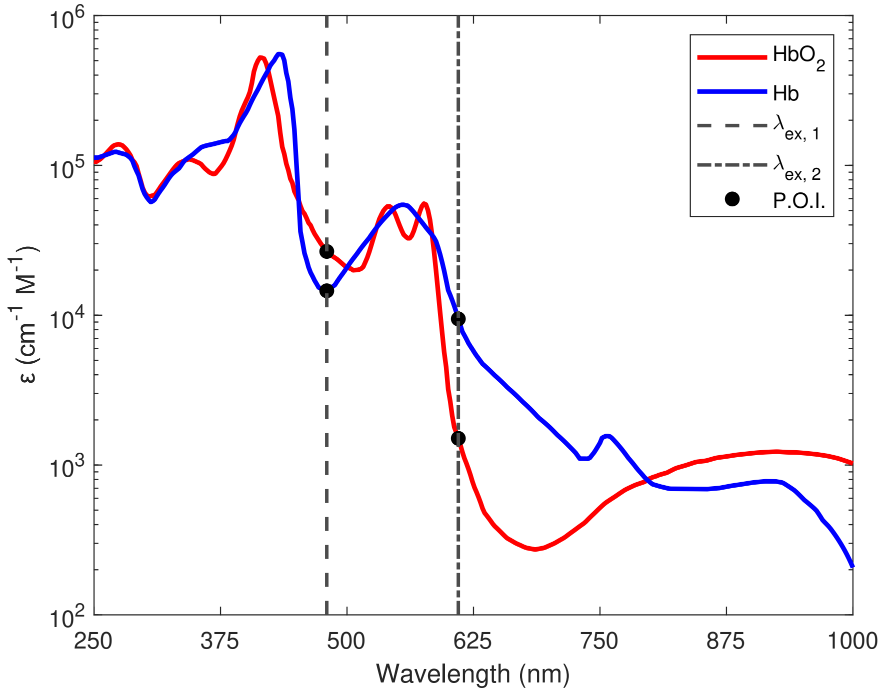


Figure 3.1: Building upon Figure 2.2, two vertical lines at $\lambda = \lambda_{ex,1}, \lambda_{ex,2}$, positioned arbitrarily, are overlaid upon the molar extinction spectra of oxyhemoglobin (red) and deoxyhemoglobin (blue). The points of intersection (P.O.I., labeled as black dots) between these vertical lines and the spectra determine the entries of E, seen in Equation (3.2). By altering $\lambda_{ex,1}$ and $\lambda_{ex,2}$, E can be adjusted to best unmix the given chromophores. Data sourced from [107].

2. The condition number normalized by the measurement magnitude.

3. The estimation error covariance.

These metrics will be developed throughout this section, in somewhat of an iterative and exploratory manner. In each case, the excitation wavelengths are selected to minimize the given metric, with the intention of eventually arriving at a potentially optimal, or at least suitable, set of wavelengths for unmixing.

3.1.1 Minimal Condition Number

A common method used to evaluate sensitivity to error (noise or perturbations) in a matrix model or in the measurements is the *condition number* [196, 197]. For a given matrix-vector equation (or equivalently a system of equations),

$$M\vec{x} = \vec{b}, \tag{3.3}$$

there may be error in one or both of $\vec{b}$ or M. Assuming one were solving for unknown $\vec{x}$ here, it is desirable to understand how errors in $\vec{b}$ or M will propagate to create error in the solution. The impact of error is not necessarily proportional to the magnitude of the error in the measurement [198]. That is, for a small relative error of roughly 10% in $\vec{b}$, for example, there is not necessarily 10% relative error in $\vec{x}$: it may in fact be substantially (up to infinity, in the case of non-invertible systems) larger. Not only is error in the solution not necessarily proportional to the error in the measurement, it is also dependent on the *direction* of the error (in the sense of vector direction). This clearly warrants investigation to avoid such pitfalls of selecting excitation wavelengths such that E leads to an especially sensitive system.

To help give the reader a sense of how this effect takes place, consider the matrix, M, acting as a linear mapping to transform the unit circle. If $M = aI$, any multiple of the identity matrix, the unit circle is essentially unchanged in shape; it is only scaled. However, for $M \neq aI$, the unit circle is deformed through a combination of rotation, reflection, and skewing. The degree of this deformation can be quite extreme, as is illustrated in Figure 3.2, where the transformed unit circle is a highly elongated ellipse.

To understand how error is propagated from measurements, $\vec{b}$, to the solution, $\vec{x}$, consider a small perturbation in $\vec{b}$. Figure 3.3 shows $\vec{b} = [1, 1]^T$ with a small circle around it, representing the set of points with 10% relative error compared to $\vec{b}$. The pre-image of the small circle is a highly elongated ellipse around the point $\vec{x} = [1, -1]^T$. This demonstrates how small errors in $\vec{b}$, perhaps as a result of measurement noise, manifest themselves as enormous directionally-dependent errors in the solution.

The *condition number*, cond() or $\kappa()$, of a matrix or system of equations is defined [97, 198] based on the matrix-norm, such that

$$\text{cond}(M) \equiv \kappa(M) := ||M|| \cdot ||M^{-1}||. \tag{3.4}$$

Note that here, $||\cdot||$ is the induced matrix norm, i.e.,

$$||M|| := \sup\left\{\frac{||M\vec{x}||}{||\vec{x}||} : \vec{x} \in \mathbb{R}^n, \text{with } \vec{x} \neq 0\right\}, \tag{3.5}$$

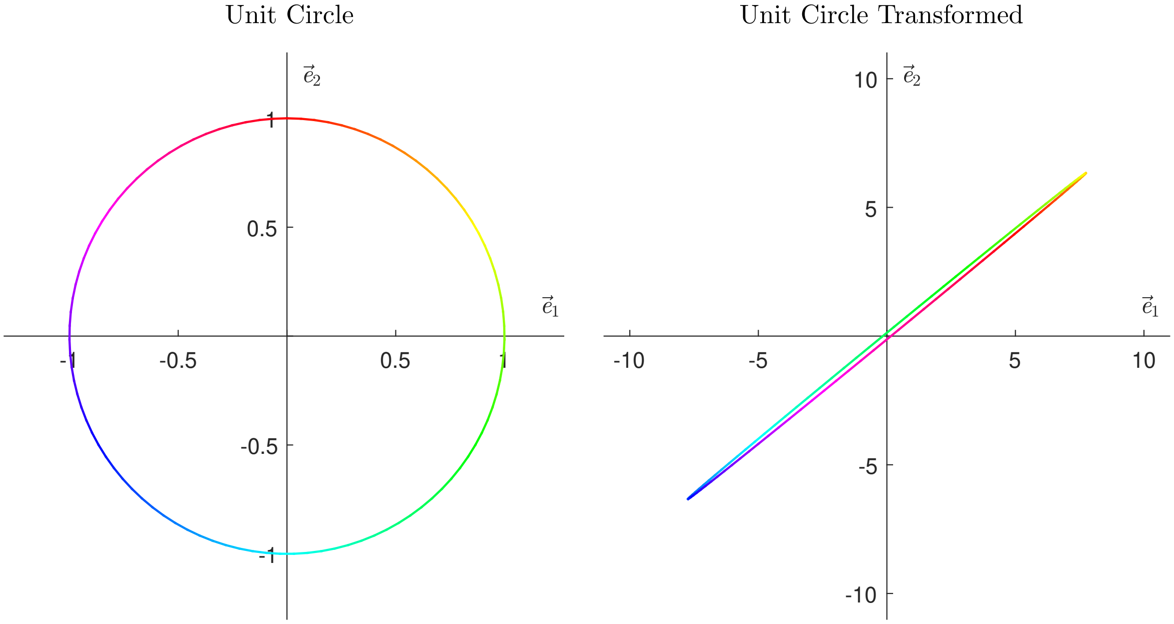


Figure 3.2: The unit circle (left) is mapped to a highly elongated ellipse (right) through the linear transformation defined by an ill-conditioned matrix, M. The axes are defined in terms of the standard basis vectors, $\vec{e}_i$. Note that if M were singular (zero-determinant), the unit circle would be mapped to either a line or a point; however, pictured here, $\det(M) = 1$, exactly the same determinant as that of the identity matrix, and still the unit circle is *nearly* mapped to a line (in fact, the ellipse could be made arbitrarily narrow). Thus, it is prudent to note that the determinant cannot be used to judge the conditioning of a matrix transformation. Image adapted from [198].

for vector norm $||\cdot||$ defined in vector space $\mathbb{R}^n$. If the usual 2-norm (or equivalently, the Frobenius norm [209]) in $\mathbb{R}^n$ is used, then $\kappa()$ can be specified in terms of the singular values, σ_i, of the matrix [97], as

$$\kappa(M) = \frac{\sigma_{max}(M)}{\sigma_{min}(M)}. \tag{3.6}$$

The maximal and minimal singular values represent the degree of “stretching” along the major and minor axes of the ellipse, from the analogy shown in Figure 3.2. In the best case, the unit circle is not unevenly stretched / skewed, and thus $\sigma_{max}(M) = \sigma_{min}(M)$ and $\kappa(M) = 1$. In poor cases, such as what is shown in Figure 3.2, $\kappa(M) \gg 1$.

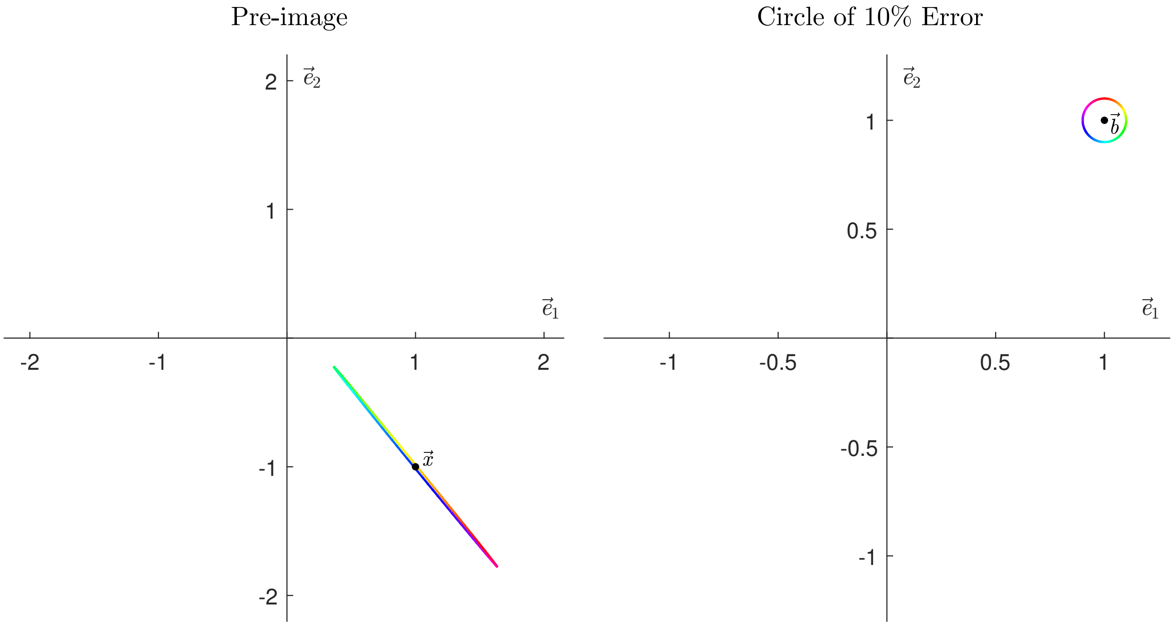


Figure 3.3: Following from Figure 3.2, a small circle surrounding $\vec{b} = [1, 1]^T$ (right), representing the set of points with 10% relative error compared to $\vec{b}$, has a pre-image (left) that is a highly elongated ellipse around the point $\vec{x} = [1, -1]^T$. Again, the axes are defined in terms of the standard basis vectors, $\vec{e}_i$. It is observed that small, 10%, errors in $\vec{b}$ can result in *much* larger errors in the solution. Image adapted from [198].

To understand how the selection of excitation wavelength affects the noise-sensitivity of the system, the condition number is evaluated for all wavelength pairs across the absorption spectra of oxy- and deoxyhemoglobin. Figure 3.4 shows the log-condition number associated with each wavelength pair. Note that log() is applied for the purpose of improving visualization, compressing the scale for greater condition numbers (which are of little interest) and broadening the scale for smaller condition numbers (which are of interest, since small condition numbers are associated with systems that are less sensitive to measurement error).

Notice, in Figure 3.4, that the wavelength pairs with the lowest log-condition number are those at roughly 700 nm and 1000 nm. However, these are very low points in the absorption spectra (Figure 3.1). Although numerically, these wavelengths may yield optimal condition numbers, in practice, they will not work very well at all. This is because

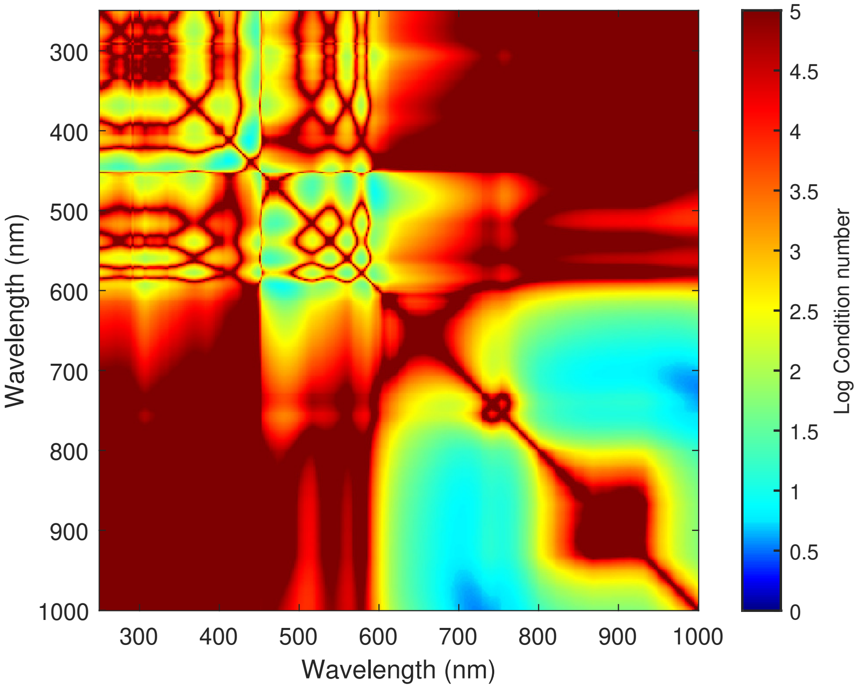


Figure 3.4: The log-condition number is evaluated for wavelength pairs according to the absorption spectra of oxy- and deoxyhemoglobin. Notice the diagonal symmetry of the plot. This arises as a result of the system's invariance to the *ordering* of excitation wavelengths, i.e., permuting the system of equations (rows of E) has no effect on solution nor the system error-sensitivity. Also note that wavelength pairs along the diagonal constitute linearly *dependent* measurements, resulting in a non-invertible system. Low (better) log-condition numbers tend to occur at wavelengths pairs of high contrast between the absorption of the two chromophores, with a minimum occurring at roughly 700 nm and 1000 nm.

sufficient optical energy absorption is required to create an easily measurable PARS response above the noise floor, and the low absorption at these wavelengths would make this incredibly challenging. This prompts reconsideration of the metric used to optimally select wavelengths for unmixing.

Condition number is used to quantify the estimate / solution sensitivity in proportion to the measurements themselves. If the target absorption spectrum is low at a particular wavelength, then measurements at that wavelength will also be low; however, the condition number does not take the magnitude of the measurement into consideration, merely the estimate sensitivity in proportion to the measurement is considered. Because condition number is defined as the ratio of maximum-to-minimum singular values, the overall scaling of the mixing matrix, or measurements, does not impact the condition number. This property may be beneficial in some scenarios; however, when additive noise is present in measurements, the magnitude of the measurement is extremely important and must be large relative to the noise floor to be easily interpreted.

3.1.2 Minimal Normalized Condition Number

To address the issue with condition number not taking into account the magnitude of the measurement (recall that large measurements relative to the noise floor are preferred), we consider a simple modification: the condition number divided by the norm (magnitude) of the measurement, $\kappa(E)/||\vec{s}||$. This metric will be referred to as the *normalized* condition number. Although this approach is a very rough way of penalizing small measurements as well as high condition numbers, it does give meaningful and insightful results as will be seen in Figure 3.5. Given that the noise floor is entirely dependant on the specific characteristics of the given PARS system, the required maximum allowable estimation error is also system-dependent, and thus it is impossible to draw strict universal quantitative conclusions from this analysis. However, if these quantities were known, weights or a cost associated with the magnitude of the measurement and the estimation sensitivity could be assigned to somewhat reshape these results. None the less, the overall trend would be similar.

The previous analysis was re-run, this time using the normalized condition number as the metric. In this case, the measurement is

$$\vec{s} = E \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{3.7}$$

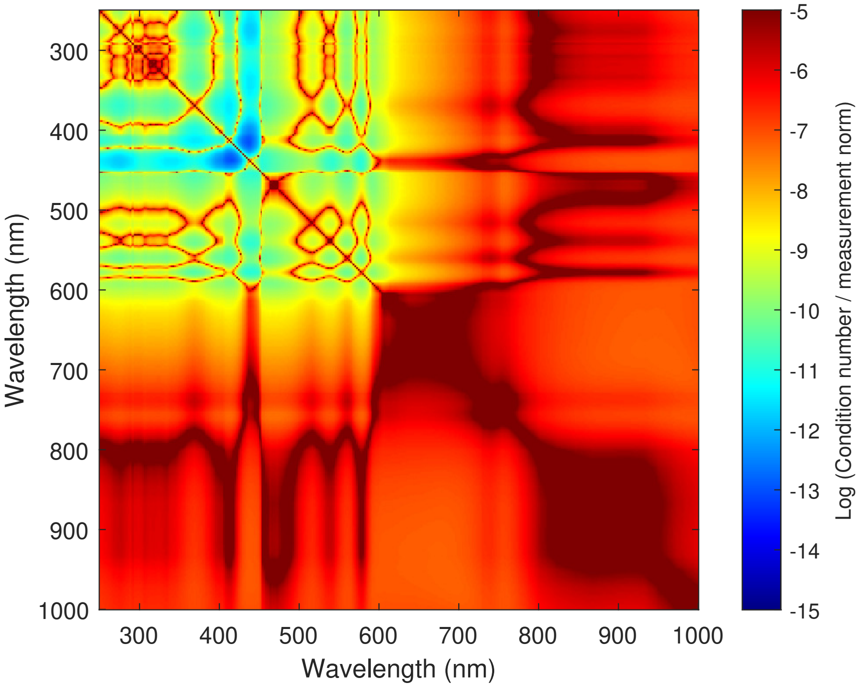


Figure 3.5: The log of the normalized condition number is evaluated for wavelength pairs according to the absorption spectra of oxy- and deoxyhemoglobin. This plot bears similarity to Figure 3.4, showing the log-condition number; however here, wavelengths from the weaker parts of the absorption spectra, from roughly 600 nm upwards, have very poor (large) normalized condition numbers, as expected. The optimal wavelength pairs occur at 412 nm and 436 nm, near the absorption peaks of the two spectra.

according to Equation (3.1), where the unknown concentrations are in equal proportion (1 and 1) up to arbitrary scaling. Figure 3.5 shows the result of this analysis: the log of the normalized condition number for excitation wavelength pairs.

Now, the optimal wavelengths are at 412 nm and 436 nm, roughly the absorption peaks of the two spectra (see Figure 3.1). This seems to be a more practical set of wavelengths, matching intuitions relating to both the ability to discern the targets based on the relative strength of the measurements as well as the magnitude of the measurements themselves. Again, this method is not necessarily the *correct* one to use; however, it does address the shortcomings of only using the condition number and thus is a worthwhile iterative improvement on that approach

3.1.3 Minimal Estimation Error Covariance

The next incremental step is to look at this as a maximum likelihood problem and select wavelengths to minimize the estimation error covariance. If we consider measurement error (i.e., noise) covariance, R, then estimation error covariance, P, is calculated [97] as

$$P = (E^T R^{-1} E)^{-1}. \tag{3.8}$$

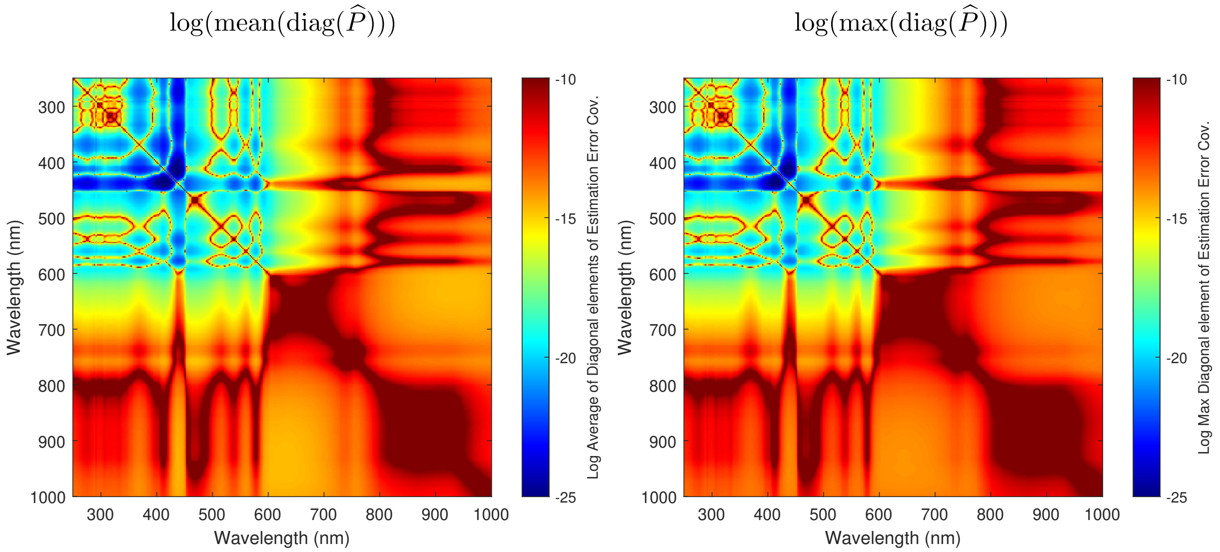
If we assume R is of the form $R = \sigma I$, where I is the identity matrix (i.e., no off-diagonal terms), then $P = \frac{1}{\sigma}(E^T E)^{-1}$.

This assumption implies that measurement error covariance is *equal for all excitation wavelengths*; however realistically, this may not be the case and would depend on the excitation laser characteristics at the specific excitation wavelengths. The most obvious factor would be the stability (i.e., levels of power fluctuations over time) of the source, something that cannot be taken for granted, especially when using an SRS-based multi-wavelength source, where the stochastic nature of initial Stokes wave growth leads to inherent temporal instability [94]. Other factors might include the polarization state of the excitation beams, or the beam quality, relating to how close the intensity profile is to being Gaussian, creating a limitation to the focused beam spot size and shape.

A result of this assumption is that, up to scaling, proportional error covariance, $\widehat{P} \propto P$, can simply be calculated as $\widehat{P} = (E^T E)^{-1}$. Although a selection of wavelength pairs to minimize the estimation error covariance is desired, given that $\widehat{P}$ is a matrix, what *precisely* is desired is not obvious: $\widehat{P}$ has several entries, not all of which will inherently be minimized simultaneously. The diagonal of $\widehat{P}$ contains error variances for each concentration estimate and the off-diagonal terms have information about the relationship between the concentration estimates. The variances relate to the expected magnitude of the error.

To address this, the metrics of either the average of the diagonal elements, $\text{mean}(\text{diag}(\widehat{P}))$, or the maximum diagonal element, $\text{max}(\text{diag}(\widehat{P}))$, can be used in order to search for wavelength pairs that yield a system approximating one of minimal estimation error covariance. Similar to before, wavelength pairs across the absorption spectra of oxy- and deoxyhemoglobin were compared under these metrics.

Figure 3.6 shows the evaluation of $\text{mean}(\text{diag}(\widehat{P}))$ and $\text{max}(\text{diag}(\widehat{P}))$ in its left and right panels respectively (again, with log applied). Both appear visually similar to each other, giving confidence that essentially the same objective is achieved by the two metrics. Like the previous analysis, using normalized condition number, the optimal wavelength pairs are 412 nm and 436 nm, roughly at the absorption peaks, although variation is seen throughout the rest of the wavelength range.



$\log(\mathrm{mean}(\mathrm{diag}(\widehat{P})))$

$\log(\max(\mathrm{diag}(\widehat{P})))$

Figure 3.6: The log of the average (left) and maximum (right) of the diagonal entries of the estimation error covariance matrix are evaluated for wavelength pairs according to the absorption spectra of oxy- and deoxyhemoglobin. The results here are very similar to those of Figure 3.5, in that the optimal wavelengths (identical in both cases), at 412 nm and 436 nm, coincide with the absorption peaks; however, there is variation throughout the rest of the wavelength range. The results of both metrics (mean and max) are extremely similar, meaning that the same objective may be achieved using either of the two metrics.

The results of this analysis are nice in that they favour wavelengths of strong absorption without the requirement for any assumption of a “measurement” or predetermined approximate concentrations, as was the case in the analysis based on the normalized condition number. Additionally, these metrics more closely align with the true goal of selecting suitable wavelengths for unmixing with – to have minimal error in the *solution*.

3.1.4 Discussion of Excitation Wavelength Selection

Several metrics for selecting excitation wavelengths were explored: condition number, normalized-condition number, and estimation error covariance. Each brought the selection problem closer to a suitable solution, taking into account sensitivity to perturbation (error) in either the model or the measurements through the condition number, the “strength” of the measurements based on the absorption at the selected wavelengths, and finally the magnitude of expected error in the solution based on estimation error covariance.

While there remains more to be done to properly adapt any of these solutions to a PARS microscope, the work presented here provides the groundwork required to optimally select excitation wavelengths for improved, noise-insensitive, unmixing in PARS microscopy. Additional steps include broadening the analysis to incorporate additional measurements (trading-off estimation sensitivity for system complexity), and considering the spectral profile of the excitation, rather than simply the assumed infinitesimal line-widths used here, which is especially necessary when using an SRS-based multi-wavelength source. Additionally, considerations based on how the SRS-based multi-wavelength source is implemented in the PARS system (i.e., single vs. multiple fibers, filtered vs. unfiltered SRS spectra used for excitation) and how this would impact the optimal selection of excitation wavelengths / SRS spectra are required. These next steps are discussed in Section 6.1.

With the task of excitation wavelength selection addressed, formalizing and adding constraints to the unmixing solution is undertaken next in Section 3.2, allowing practical unmixing estimates to be made later in Section 3.3.

3.2 A Formulated Unmixing Solution

Section 2.2 introduced a simplified forward model on page 14, repeated in Equation (3.1), as well as several simple solutions, yielding the estimates, $\widehat{\vec{c}}$, of the chromophore concentrations. For the most straightforward case, where the forward process is accurately modeled by Equation (3.1), Equations (2.14) and (2.15), the matrix inverse (or pseudo-inverse [109–111]) can be applied to find the solution. This is applicable in theoretical cases where measurement error / noise is not present and there is *no modeling error*; i.e., the measurements, $\vec{s}$ are in the span of the columns of the mixing matrix, E. Unfortunately, this is unrealistic and generally this approach will fail.

In reality, modeling and measurement error / noise will be present, necessitated the analysis done in Section 3.1 regarding selecting excitation wavelengths. To account for this when formulating an unmixing solution, the forward model is adjusted to explicitly include additive Gaussian noise, $\vec{\nu}$, with covariance, R:

$$\begin{aligned} \vec{s} &\overset{\propto}{=} E\vec{c} + \vec{\nu}, \\ \vec{\nu} &\sim \mathcal{N}(R). \end{aligned} \tag{3.9}$$

In this case, the optimal (in terms of least squared error) estimate can be obtained directly [97] through

$$\widehat{\vec{c}} = \left(E^T R^{-1} E\right)^{-1} E^T R^{-1} \vec{s}. \tag{3.10}$$

This formulation was the basis for the estimation error covariance of Equation (3.8), used for optimal wavelength selection. Though there is utility in this solution, additional knowledge of the problem, asserted through constraints, can be used to further improve the estimation accuracy.

Equation (2.16), slightly modified to incorporate the measurement error covariance,

$$\widehat{\vec{c}} = \arg\min_{\vec{c}} \Big\{ \|\vec{s} - E\vec{c}\|_{R^{-1}} + \sum_i w_i \Psi_i(\vec{c}) \Big\}, \tag{3.11}$$

shows the solution formulated as a general minimization problem including additional arbitrary constraints, $\Psi_i()$, and associated weights, w_i. Note that $\|\vec{s}-E\vec{c}\|_{R^{-1}}$ is shorthand for

$$\|\vec{s} - E\vec{c}\|_{R^{-1}} = (\vec{s} - E\vec{c})^{-1} R^{-1} (\vec{s} - E\vec{c}), \tag{3.12}$$

where a quadratic penalty is used for error. This formulation is more general than the above and may require an algorithmic optimizer to solve.

3.2.1 Constraints

Constraints are included in the estimation formulation to improve solution accuracy through improving the model accuracy. This is because prior information / knowledge about the problem is asserted through these constraints. This way, both prior information and measurements are taken into account, making for better estimates.

Firstly, there is a desire to constrain the unknown concentrations to be *non-negative*. This makes intuitive sense, given that by definition, concentration cannot be negative; however, for relatively low true concentrations or relatively high noise levels, it is possible that the least squares solution of Equation (3.10) could yield negative estimates. This constraint is asserted through

$$\Psi_{\text{nn}}(\vec{c}) = \sum_{\forall i} \big(|c_i| - c_i \big) \text{, for elements } c_i \text{ of } \vec{c}. \tag{3.13}$$

Note that this is a *soft* constraint, rather than a hard constraint, meaning that negative concentrations are technically allowed still in the solution; however, they are penalized according to the associated weight, w_{nn}, based on how negative they are. Positive concentrations are not penalized. What this constraint looks like for single concentration, c_i, is a downwards ramp for negative values and a constant zero-value for positive values. Figure 3.7 visualizes this constraint. The effective slope of this function is determined

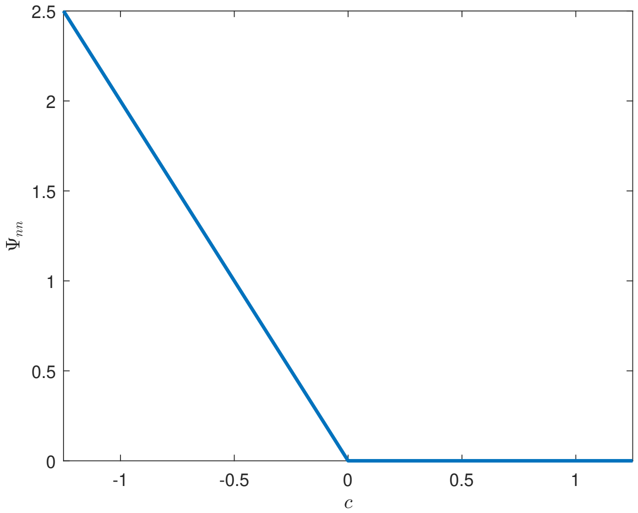


Figure 3.7: The non-negativity constraint of Equation (3.13), $\Psi_{\text{nn}}(\vec{c})$, is plotted in one dimension for $\vec{c} = [c]$. Only negative values for concentration are penalized, with linearly increasing penalty the more negative the value is.

by the weighting variable w_{nn}. If a limit is taken where $w_{\text{nn}} \to \infty$, then $w_{\text{nn}}\Psi_{\text{nn}}()$ tends towards a function which is infinite for values less than 0, and is 0 otherwise, which is the ideal penalty as a hard constraint. This definition has the benefit of being continuous for finite values of w_{nn}, which is simpler for optimization compared to discontinuous objective functions.

Note that while this formulation does not impose upper limits on concentrations, if *relative* concentrations were desired, an additional constraint could be included to enforce further limitations on concentrations. Firstly, relative concentrations should be within the interval $[0, 1]$; however, if multiple chromophores are present, and only the ratio of their concentrations is desired, the relative concentrations must *sum* to one. This can be imposed via

$$\Psi_{\sum c_i = 1}(\vec{c}) = \sum |c_i - 1| \,. \tag{3.14}$$

Again, the degree of penalization is controlled by the associated weight, $lambda_{\sum c_i = 1}$. This constraint is visualized in Figure 3.8 for two concentrations. Notice the “trough” along the line $\sum c_i = 1$, drawing the solution towards one where relative concentrations sum to 1. The combination of this and the non-negativity constraint of Equation (3.13), visualized in Figure 3.9, ensures relative concentrations are within the interval $[0, 1]$.

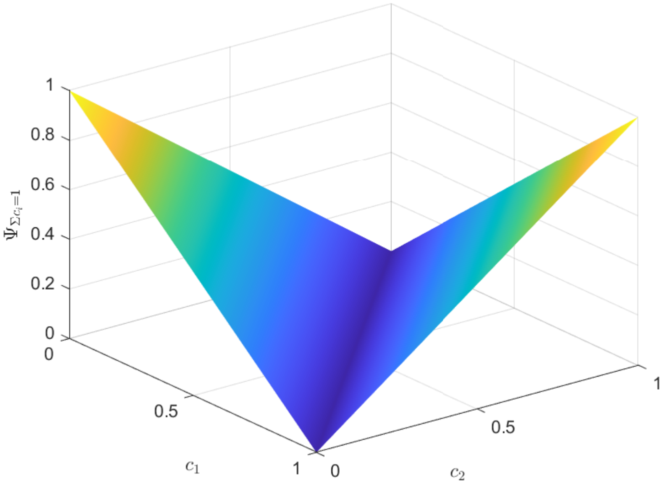


Figure 3.8: The sum-to-one constraint of Equation (3.14), $\Psi_{\sum c_i=1}(\vec{c})$, is plotted for $\vec{c} = [c_1, c_2]^T$. A "trough" along the line $\sum c_i = 1$ is evident. This constraint is applicable when *relative* concentrations are desired.

Next, a constraint preferring *sparse* solutions in introduced. Sparsity, here, relates to the number of non-zero estimated concentrations. Enforced solution sparsity is desirable in cases where it is known ahead of time that relatively few chromophores will dominate in any singular location of the target, corresponding to a single measurement. Incorporating this constraint allows for the absorption spectra of many chromophores to be included in the model; however, for any measurement, the estimated concentrations will be non-zero for relatively few chromophores, matching the expectation. This constraint can be imposed via the zero-norm, $\|\cdot\|_0$, which effectively counts the number of non-zero elements of a vector. Thus, the imposed constraint is asserted through

$$\Psi_{\text{sparsity}}(\vec{c}) = \max\left(\beta, \|\vec{c}\|_0\right), \tag{3.15}$$

for parameter β, set based on the number of anticipated dominant chromophores. Again, the degree of penalization for insufficiently sparse solutions is controlled by the associated weight, $lambda_{\text{sparsity}}$. The optimal solution we desire is not inherently as sparse as possible, rather it is only as sparse as required based on the assumed imaging target. For example, if it is assumed that there are three most prominent chromophores present, the optimal sparsity would be three. So, to avoid over-penalizing cases where $\widehat{\vec{c}}$ is more sparse than desired, the parameter β is introduced, and the maximum of either $\|\widehat{\vec{c}}\|_0$ or β is taken. This

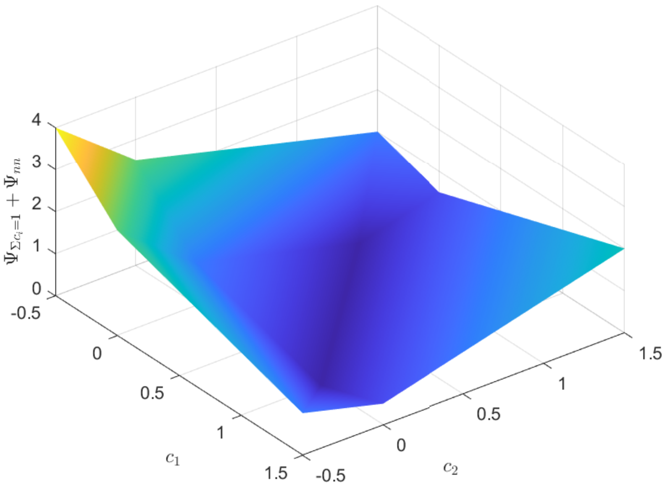


Figure 3.9: The sum-to-one constraint of Equation (3.14), $\Psi_{\sum c_i=1}(\vec{c})$, and the non-negativity constraint of Equation (3.13), $\Psi_{\text{nn}}(\vec{c})$, are summed together and plotted for $\vec{c} = [c_1, c_2]^T$. Note, for the purpose of this figure, the associated weights for the two constraints are assumed to be equal; however, in practice they would be separately tuned such that reasonable estimates are produced. Building upon Figure 3.8, which showed only a simple linear "trough", here a "basin" is formed, guiding solutions to be confined within the interval $[0, 1]$, with a preference towards solutions along the line $\sum c_i = 1$.

asserts a constant preference for values less than or equal to β. For the example where there are three most prominent chromophores, β is set to $\beta = 3$.

Rather than the zero-norm, as was used in Equation (3.15), it makes sense to consider using the 1-norm. It has been found that the 1-norm solution to optimization problems is often equivalent to the zero-norm solution [210]; however, it is far easier computationally, due to the continuity and convexity of the objective function [211, 212]. The zero-norm has zero gradient (i.e., $\frac{\mathrm{d}\|x\|_0}{\mathrm{d}x} = 0$ everywhere the derivative exists), creating a challenge for the optimizer since gradient-based methods cannot be applied. If instead, the 1-norm is used, the sparsity constraint of Equation (3.15) becomes

$$\Psi_{\text{sparsity}}(\vec{c}) = \max\left(\beta, \|\vec{c}\|_1\right). \tag{3.16}$$

Figure 3.10 illustrates both the zero-norm and 1-norm implementations of the sparsity constraint for comparison purposes. Notice that the locations of local and global minima

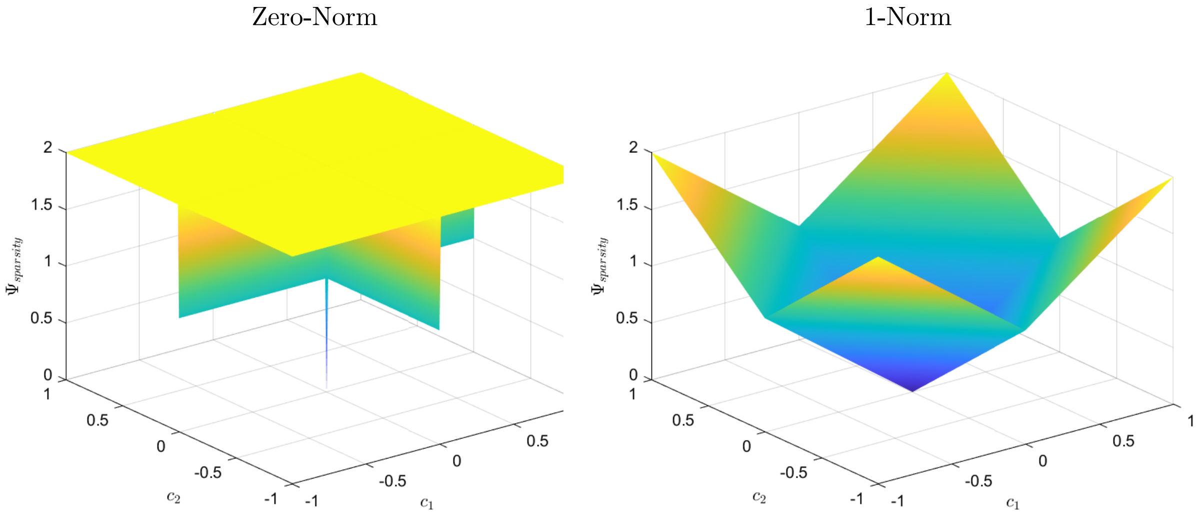


Figure 3.10: The sparsity constraint of Equation (3.16), $\Psi_{\text{sparsity}}(\vec{c})$, is plotted for $\vec{c} = [c_1, c_2]^T$, using the zero-norm implementation (left) and the 1-norm implementation (right). Note that both implementations share the same locations for local and global minima, a desirable characteristic when interchanging the 1-norm for the zero-norm implementation. The 1-norm implementation affords a continuous objective function, whereas the zero-norm is discontinuous and furthermore has zero gradient, creating a challenge for the optimizer that can be alleviated by using the 1-norm instead.

are shared between the two cases. Intuitively, one can understand that this is an important characteristic when substituting the 1-norm implementation for the zero-norm implementation, as this means the 1-norm implementation would "guide" the optimizer towards the same minima that would be imposed by the zero-norm implementation.

Though not a constraint on the estimates, the penalty associated with error can be thought of as a constraint related to the measurements. Thus, Equation (3.11) can be further generalized to

$$\widehat{\vec{c}} = \arg\min_{\vec{c}} \Big\{ \text{Measurement Constraints} + \text{Estimate Constraints} \Big\}. \tag{3.17}$$

Previously, in Equations (3.11) and (3.12), Gaussian measurement noise was assumed, and a quadratic penalty was applied; however, this assumption isn't necessarily realistic. Instead, a likelihood term could be introduced for the noise residual, $\vec{s} - E\vec{c}$, being observed based on a prior model of the noise distribution, $p_\nu()$. The prior model would be defined based on an empirical approximation and may be both system and target dependant. Including this in the formulation results in

$$\widehat{\vec{c}} = \arg\max_{\vec{c}} \Big\{ p_{\nu}(\vec{s} - E\vec{c}) - \sum_{i} w_i \Psi_i(\vec{c}) \Big\}. \tag{3.18}$$

Notice the change from the previous formulations: here, *maximization* is used because maximizing the prior likelihood term for the noise residual is desired. By simply inverting the sign of the constraints, relative to what is shown in Equation (3.11), this formulation is made consistent with the previous ones. Because no noise model is presented here in this thesis, this task is left as future work and is further discussed in Section 6.1.1.

3.2.2 Additional Considerations Related to PARS Microscopy

In the previous section, Equation (3.9) served as the simplified forward model for the PARS microscope. A few small – yet still important – details were omitted there but are introduced here, prior to Section 3.3, where *in vivo* sO_2 estimation is performed. These two details are

- optical fluence variation by wavelength, and
- the targets' diffuse reflectance at the *detection* wavelength.

Both of these effects must be compensated for, or rather included in the forward model, to achieve accurate unmixing in PARS microscopy.

First, optical fluence variation by wavelength is considered. Optical fluence, ϕ, is defined as the quantity of radiant energy per unit area, often given in units of $\mathrm{J/m^2}$ or similar. When imaging there is no rule nor guarantee that all excitation wavelengths must be of the same pulse energy. Additionally, assuming that the same pulse energy was used across all excitation wavelengths, the fluence would still vary between wavelengths. The excitation beam is focused onto / into the target when imaging, concentrating the optical energy of the beam to a small focal spot. Assuming a diffraction-limited focal spot[1], the diameter of the spot is proportional to the wavelength, λ [131]. Thus, the effective area that the light is focused to is related to λ^2, and therefore the fluence, ϕ_{λ_i}, at a given wavelength, λ_i, is proportional to that wavelength squared: $\phi_{\lambda_i} \propto \lambda_i^{-2}$.

---

[1] Ideally the excitation beam would have diffraction-limited focus (with the intention of improving resolution and fluence); however, due to limitations of instrumentation as well as the variable geometry of imaged targets, the focus may only be *close* to diffraction limited. Nonetheless, the same principle broadly applies whereby the wavelength imposes a restriction on the focal spot size, even for only approximately diffraction-limited focus.

This effect can nicely be encapsulated by a single diagonal matrix of fluences, Φ, for excitation wavelengths $\lambda_1, \ldots, \lambda_n$:

$$\Phi = \mathrm{Diag}([\phi_{\lambda_1}, \ldots, \phi_{\lambda_n}]) \propto \mathrm{Diag}([\lambda_1^{-2}, \ldots, \lambda_n^{-2}]). \tag{3.19}$$

Thus, the forward model of Equation (3.9) can be updated with the inclusion of Φ as

$$\vec{s} \overset{\propto}{=} \Phi E \vec{c} + \vec{\nu}. \tag{3.20}$$

Next, the diffuse reflectance of the target at the *detection* wavelength is considered. As a brief rationale, consider the interaction of the detection beam with the target: some portion of the light will be scattered / reflected while the remainder will be transmitted (or absorbed). The fraction of reflected light varies from target to target, and thus, for a non-homogeneous target, variation in back-scattered light is observed. This is the basis for scattered light microscopy [213].

Suspensions of red blood cells with different concentrations of oxy- and deoxyhemoglobin exhibit different diffuse reflectances, $r_{\mathrm{HbO_2}}(\lambda)$ and $r_{\mathrm{Hb}}(\lambda)$, [214]. An empirical model describing the relationship [214, 215],

$$r(\lambda) = \beta_1 + \frac{\mu_s'(\lambda)}{\beta_2 \times \mu_a(\lambda) + \beta_3}, \tag{3.21}$$

is fundamentally based on two characteristics of reflectance: inverse dependence on the absorption coefficient, μ_a, and linear dependence on the reduced scattering coefficient, μ_s', [216]. The additional parameters, β_i, are empirical and based on the optical properties of the system. Due to the strong inverse dependence on the absorption coefficient, Equation (3.21) can be further approximated (for the purpose of simplicity in creating an approximate "proof of concept" model) simply as being proportional to the inverse of the absorption coefficient or more generally, the extinction coefficient:

$$r(\lambda) \propto \frac{1}{\mu_a(\lambda)} \propto \frac{1}{\varepsilon(\lambda)}. \tag{3.22}$$

This approximation allows for a very direct way of incorporating the dominant effects of diffuse reflectance into the forward model, based simply on the extinction spectra of the targets at the detection wavelength.

The diffuse reflectance for the two absorbers examined here, oxy- and deoxyhemoglobin, can then be encapsulated into a single matrix, as

$$R = \begin{bmatrix} r_{\mathrm{HbO_2}}(\lambda_{det}) & 0 \\ 0 & r_{\mathrm{Hb}}(\lambda_{det}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\varepsilon_{\mathrm{HbO_2}}(\lambda_{det})} & 0 \\ 0 & \frac{1}{\varepsilon_{\mathrm{Hb}}(\lambda_{det})} \end{bmatrix}, \tag{3.23}$$

which can be included in the forward model, modifying Equation (3.20), to become

$$\vec{s} \overset{\propto}{=} \Phi E R \vec{c} + \vec{\nu} = H \vec{c} + \vec{\nu}, \tag{3.24}$$

for combined mixing matrix $H = \Phi E R$.

While previously, in Section 3.1, low intensity regions of the absorption / extinction spectra were avoided when selecting *excitation* wavelengths (where strong absorption is beneficial), here having low absorption at the *detection* wavelength is beneficial to have strongly back-scattered light, through which the PARS response is measured. Thus, the detection wavelength could also be tuned to increase the amount of back-scattered light, though other effects, especially in complex targets, must also be considered, such as transmissivity in sub-surface imaging applications. However, in this work, the effects on diffuse reflectance are simply evaluated for the purpose of more accurately modelling the forward process in an attempt to improve unmixing accuracy.

Finally, a suitable forward model, taking into account the details of PARS microscopy, is available to be used for unmixing. Next, in Section 3.3, the model developed here is applied in an *in-vivo* sO_2 estimation problem. Note that with these inclusions to the forward model, the excitation wavelength selection results of Section 3.1 would change to reflect this. The analysis could be repeated once a detection wavelength is chosen.

3.3 Unmixing Application: sO_2 Estimation

This section applies the theory and model developed in the preceding sections to achieve *in-vivo* blood oxygen saturation estimation. Multi-wavelength images of retinal vasculature of *rattus* (common name: rat) will be processed and unmixed.

Details of the experimental procedures (regarding in-lab imaging) are disclosed in article [3]. Note that all experimental procedures involving animals were conducted according to the laboratory animal protocol approved by the Research Ethics Committee at the University of Waterloo (Animal Utilization Project Protocol number 40149). Albino rats (Charles River, MA, USA) were imaged to explore the ability to perform *in-vivo* blood oxygen saturation estimation in PARS. The animal was anesthetized using ketamine during the experiment. One drop of 0.5% proparacaine hydrochloride (topical anesthetic; Alcaine, Alcon, Mississauga, ON, Canada) was applied to the eye to reduce eye movement, followed by one drop of 0.5% tropicamide (pupillary dilator; Alcon). A custom-made animal holder was used to restrain the animal. The base of the animal holder was lined with a thermal pad to maintain the body temperature of the animal between 36 °C and 38 °C. Artificial tears were applied frequently (approximately every 2 minutes) to keep the cornea hydrated. Vital signs of respiration rate, heart rate and body temperature were monitored during the experiment.

A dominant challenge in performing unmixing with retinal images stems from poor (shallow) depth of focus and changes in focus as a function of wavelength due to chromatic aberration. The shallow depth of focus causes unevenness in signal strength and the formation of localized in-focus regions where the vasculature is resolved. Figure 3.11 shows an example of this effect, whereby only some of the vasculature within the field of view is resolved. The left panels of Figure 3.13 show more complete pictures of the vasculature within the field of view. Focal shift by wavelength causes differences in the in-focus region between images of different wavelengths; a fundamental issue for unmixing, which requires observations at the *same* location. More detail surrounding this issue and possible solutions is discussed under Future Work (Section 6.1). Additionally, small animal movements can result in captured images with different fields of view, even with topical anesthetics applied to the eye to reduce movement. The issue necessitates image co-registration, but may also result in focal differences, similar to those caused by chromatic aberration.

Fortunately, oxygen saturation is known to be approximately constant within any given vessel, transitioning predominantly in capillaries, where oxygen is transferred to the tissue. This fact can be used to enable estimation in spite of the challenges of varying focus. Vasculature can be segmented into individual vessels based on any contrast available in the PARS (and scattering) images, allowing for subsequent sO_2 estimation on the level of vessels, rather than per pixel. The per-pixel measurements can be averaged within each segment to create per-segment measurements. This approach also has the benefit of being far more robust to noise, since many (noisy) pixel measurements are averaged to form a single combined measurement.

In this experiment, images were captured at 532 nm and 558 nm excitation wavelengths, produced via an SRS-based multi-wavelength source pumped with a 532 nm laser. While in Section 3.1, *all* wavelength pairs were considered when selecting optimal wavelengths for unmixing; however, here, as a consequence of using SRS to generate additional wavelengths (Stokes waves; see Section 2.3.2 for background), *only specific* wavelengths can be made available. Figure 3.12 shows the available SRS wavelength bands overlaid on the plot of estimation error covariance from Figure 3.6. Though it is not shown in Figure 3.12, lower-order (i.e., lower wavelength) Stokes waves are more readily generated and generally have greater optical power, making them somewhat preferable to higher-order Stokes waves. With this in mind, 523 nm and 558 nm were selected given their relatively low (better) associated estimation error covariance and ease of generation. Two images were captured

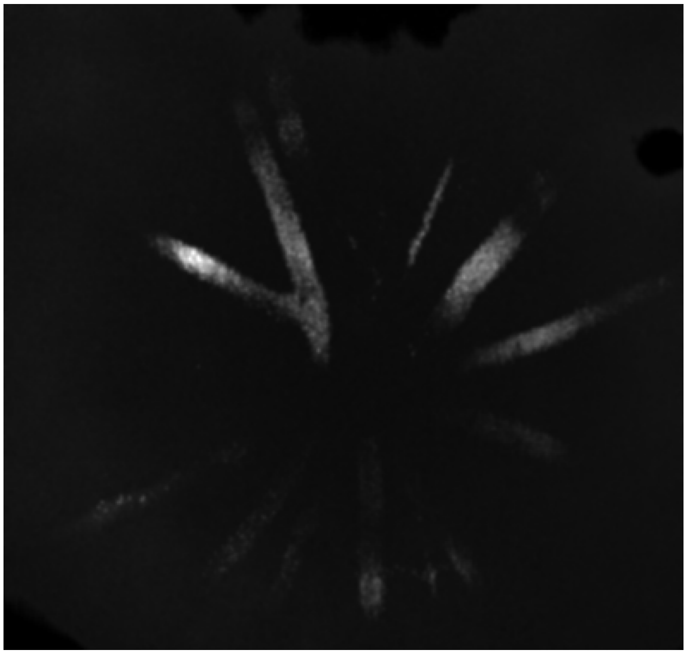


Figure 3.11: An *in-vivo* PARS image of *Rattus* retina exhibits localized regions of in-focus vasculature (white / grey structures, where brightness indicates signal strength). Note that the entire rectangular area shown here was imaged by the PARS microscope. Although vasculature is present throughout the imaged area, much of it is out of focus and did not produce a PARS response (and is thus not visible in PARS image). This effect, whereby only parts of retinal PARS images are in-focus and resolved, presents a challenge for unmixing and subsequent sO_2 estimation. The overall structure of the retinal vasculature is clarified in the left panels of Figure 3.13, composed of multiple co-registered and combined images.

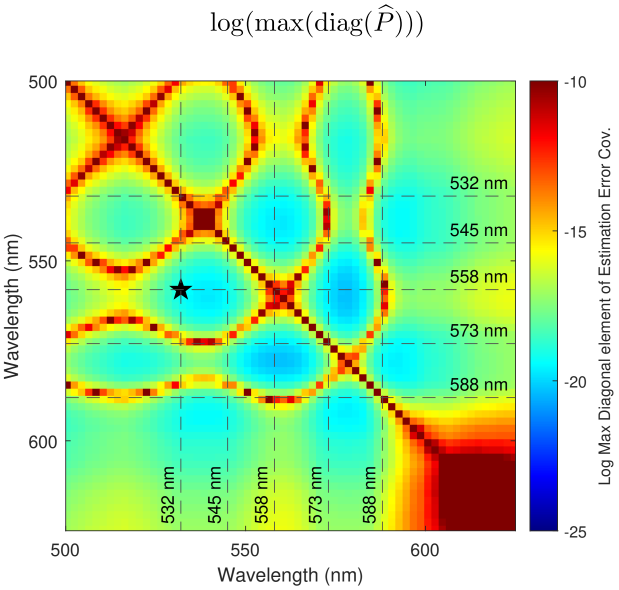


Figure 3.12: Building on Figure 3.6, which shows the log of maximum of the diagonal entries of the estimation error covariance, evaluated for wavelength pairs according to the absorption spectra of oxy- and deoxyhemoglobin, SRS wavelength bands available for excitation are overlaid as dashed lines, illustrating the combinations of wavelengths that can be used. The pair of wavelengths, 532 nm and 558 nm, used for excitation in the *in-vivo* imaging experiment, is marked with a star and has relatively low (better) associated estimation error covariance, especially compared to other low-order SRS-generated wavelength pairs.

at 532 nm, and five at 558 nm. Each image has variation in the resolved regions, and by combining them, a more full view of the retina is made available for both wavelengths.

Figure 3.13 outlines the process used for sO_2 estimation here. Firstly, the images from each excitation wavelength are co-registered and combined by taking the maximum intensity at each pixel. This way, resolved regions with strong PARS signals are kept, whereas dark unresolved regions containing only noise are rejected. Note that the combined 532 nm and 558 nm images are also co-registered. Following this, the individual vessels are segmented, resulting in a *segment mask*. Based on the background noise level, the images are thresholded and then averaged within each segmented vessel. Finally, the unmixing formulation described in the preceding sections is applied to each segmented vessel, resulting in relative estimates of concentrations of oxy- and deoxyhemoglobin, which are then used to compute sO_2 via Equation (2.17), introduced in the background and

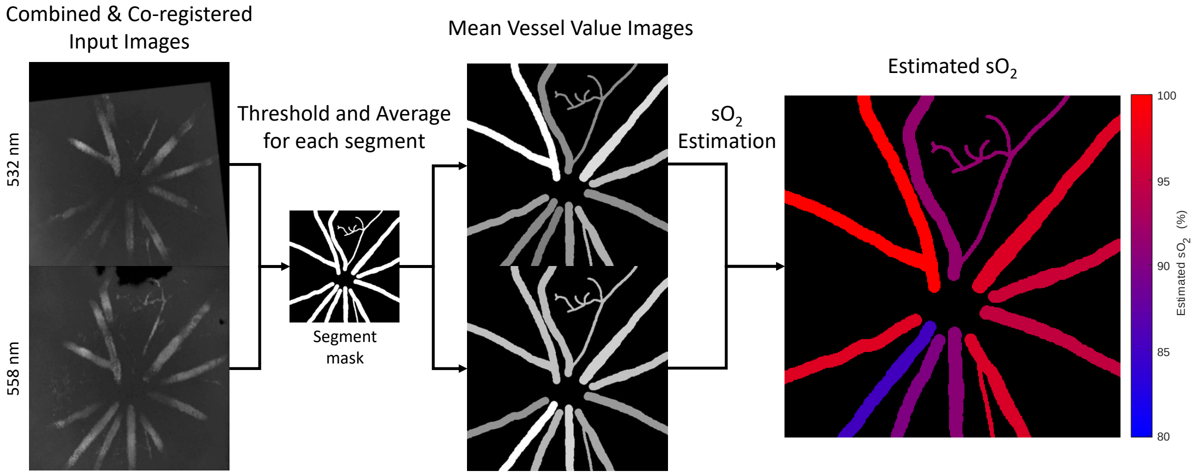


Figure 3.13: Process Diagram for Segment-based sO_2 Estimation: Images from each excitation wavelength (left) are co-registered and combined by taking the maximum intensity at each pixel (the combined 532 nm and 558 nm images are also co-registered). Next, the individual vessels are segmented, resulting in a *segment mask* (second from left, small panel). Averages are taken within each segmented vessel, ignoring values lower than the noise level threshold (center). Finally, on a per-vessel basis, the images are unmixed to estimate the relative concentrations of oxy- and deoxyhemoglobin and finally to compute estimates of sO_2 (right).

repeated here for convenience in reading:

$$\mathrm{sO_2} = \frac{c_{\mathrm{HbO_2}}}{c_{\mathrm{HbO_2}} + c_{\mathrm{Hb}}}. \tag{3.25}$$

The work presented in this section represents a proof-of-concept for the method more so than a finalized process, and as such, segmentation and image co-registration was performed manually; however, in future work these steps may be automated.

Figure 3.14 shows the final sO_2 estimates overlaid in false colour, for improved visualization, on top of the scattering image. To the best of the author's knowledge, this is the first time a non-contact photoacoustic imaging technique has been employed for *in-vivo* blood oxygen saturation measurement in the retina [3]. This result represents a significant

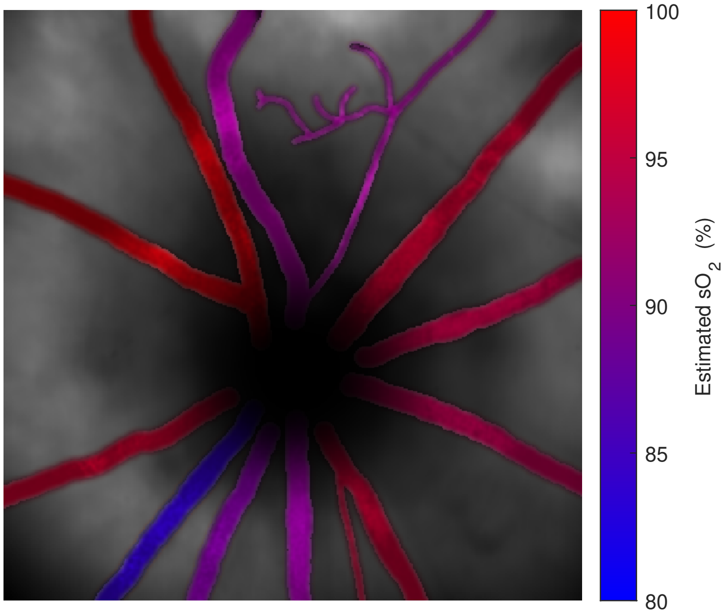


Figure 3.14: Estimated retinal blood oxygen saturation, measured *in-vivo* in a rat eye, is shown in false colour overlaid upon a scattering image of the same field of view for improved visualization. Variation in sO_2 between vessels is indicative of whether they are arteries or veins. The optic nerve head is central in this image. This image was created for this thesis but also submitted as part of article [3].

step towards the clinical use of PARS as a diagnostic tool for many ophthalmic diseases, through the measurement of sO_2.

3.4 Chapter Conclusion

This chapter covered several developments towards chromophore unmixing in PARS via multi-wavelength imaging.

Firstly, the question of how best to select excitation wavelengths was addressed. Several metrics were explored, incrementally improving the suitability of the solution while considering the sensitivity to perturbation (error) in either the model or measurements, the relative amplitude of measurements based on absorption, and the magnitude of expected error in the solution based on estimation error covariance.

Next, a solution to the unmixing problem was formulated, resulting in a constrained least-squares optimization problem. Constraints enforcing the non-negativity of chromophore concentrations, requiring concentrations to sum to one (for the case where relative concentrations are desired), and enforcing a sparse solution were developed and incorporated into the formulated solution. Additional considerations for optical fluence variation by *excitation* wavelength and diffuse reflectance at the *detection* wavelength were made, making accurate unmixing possible with PARS.

Finally, the theory and inverse model developed in the preceding sections of the chapter were used to estimate blood oxygen saturation, *in-vivo*, in the retina of a rat. The limited depth of focus of the PARS microscope imposed a challenge in adequately imaging the retinal vasculature. In any given image, limited regions are resolved. To address this, several images were acquired at both excitation wavelengths (532 nm and 558 nm), making minor focal adjustments between each. These images were co-registered and combined to produce composite images, suitably capturing the vasculature throughout the imaged area. Individual vessels were segmented, allowing per-vessel estimates of oxy- and deoxyhemoglobin to be made, and finally estimated sO_2 to be calculated. From this, a compelling false colour image, indicating sO_2 in the vasculature, was produced. This was the first non-contact *in-vivo* photoacoustic measurement of blood oxygen saturation in the retina, representing a significant step towards the clinical use of PARS as a diagnostic tool for many ophthalmic diseases, through the measurement of sO_2.

Chapter 4

Stimulated Raman Scattering in Optical Fiber as a Multi-wavelength Source

SRS is a non-linear optical effect [93, 94, 147, 148] introduced in Section 2.3.2. The effect can be very prevalent in optical fibers in the presence of high-intensity light and can be used for the purpose of generating multi-wavelength light from a single-wavelength light source simply by coupling the light into a suitable optical fiber and adjusting the input light intensity to vary the resulting spectrum.

A multi-wavelength light source is necessary for multi-wavelength imaging and unmixing. This is the true purpose in researching SRS in the context of this thesis – to optimize the SRS process for multi-wavelength imaging and unmixing in PARS microscopy [3, 50–63], thus improving the chromophore-selectivity of PARS. However, in the field of photoacoustic microscopy, relatively little is known about how best to select optical fiber for this purpose. To address this, and perhaps even to provide some form of a guide to the community, a comprehensive characterization experiment is conducted whereby individually, many fiber types (and over a range of lengths for each fiber type) are coupled into an optical test apparatus and measurements of the output SRS spectrum are made while also varying laser parameters.

Fiber Type	Operating Wavelength (nm)	Cutoff Wavelength (nm)	Mode Field Diameter (μm)	Attenuation (dB/km)	Polarization Maintaining
SM400†	405–532	305–400	2.5–3.4 @ 405 nm	$\leq$30 @ 532 nm	No
S405-XP†	400–680	360–400	2.7–3.8 @ 405 nm	$\leq$30 @ 488 nm	No
460HP†	450–600	410–450	3.0–4.0 @ 515 nm	$\leq$30 @ 515 nm	No
PM460-HP†	460–700	390–450	2.7–3.8 @ 515 nm	$\leq$100 @ 488 nm	Yes
630HP†	600–770	540–600	3.5–4.5 @ 630 nm	$\leq$12 @ 630 nm	No
HB450*	488–633	350–470	3.0–4.1 @ 488 nm	$\leq$100 @ 488 nm	Yes
HB600*	633–780	500–600	2.8–3.7 @ 633 nm	$\leq$15 @ 633 nm	Yes
HB750*	780–830	610–750	3.5–4.6 @ 780 nm	$\leq$8 @ 780 nm	Yes

$\dagger$= Thorlabs Inc., * = Fibercore Ltd.

Table 4.1: Summary of optical fiber types explored and their key parameters. Notice that the operating wavelength range varies for each fiber type. Outside of this range, the attenuation is unspecified and is generally substantially higher than the specified nominal value. Additionally, not all fibers maintain the polarization of the input light. These properties will be shown to strongly impact the efficacy in generating additional wavelengths through SRS.

4.1 Experimental Method

To understand the characteristics of a multitude of optical fibers and the impacts on SRS generation as a result of pump-laser parameters, a broad characterization experiment is conducted. A variety of optical fiber types are tried, comprising both polarization maintaining and non-polarization maintaining fibers, variations in the operating wavelength range, and variations in the (unfortunately unknown) dopants present in the fibers. The unique set of characteristics of each fiber essentially combine to determine (in some complex unknown way) the non-linear behaviour of the fiber and its efficiency in the SRS process [94]. All of these characteristics affect the propagation of light through the fiber. Table 4.1 outlines the fiber types explored and some key characteristics of each.

In addition to exploring a range of fiber types, parameters of the pump laser were varied so as to understand the impacts they each have. Table 4.2 concisely summarizes these, as well as the ranges over which they were varied. To understand the effect of fiber length (effectively the interaction length over which SRS may occur), each fiber was tested over a range of lengths from 25 m all the way down to 1 m.

The experimental apparatus for the study is relatively simple. The key components are a 532 nm pulsed laser, a fiber-launch (device to couple a free-space beam into an optical fiber), a spectrometer (CCS200, Extended Range: 200–1000 nm, Thorlabs Inc.), and the optical fiber to be tested. In more detail, Figure 4.1 diagrammatically illustrates

Parameter	Range
Fiber Length (m)	1–25
Input Pulse Energy (nJ)	20–19,000
Pulse Repetition Rate (kHz)	1–300
Pulse Width	2 ps, 1.5–50 ns

Table 4.2: Parameters explored in SRS characterization and their associated ranges. Each parameter is explored over a range of at least one order of magnitude, allowing the impacts on the SRS process to adequately be observed and studied.

the apparatus. Three lasers were used to explore the full range of pulse widths. In all cases, ytterbium-doped fiber lasers were used. For 3 ns to 50 ns pulse widths, VPFL-G-10-HE (Spectra-Physics, Inc.) was used. For 1.5 ns, GLPM-10 (IPG Photonics, Inc.) was used. For 2 ps, YLPP-25-3-50-R (IPG Photonics, Inc.) was used. In the 2 ps case, frequency doubling / second harmonic generation was employed to convert 1030 nm light from the laser to 515 nm light via a non-linear optical crystal (Lithium triborate / LBO). Following this, a dichroic filter was used to eliminate any unconverted 1030 nm light. Although the 515 nm light is not the same wavelength as the 532 nm light used throughout the rest of the characterization experiment, the wavelength difference is not significant relative to the far more dominant effects with this source which arise as a result of the almost *1000 times* shorter pulse width: 2 ps vs. 1.5+ ns.

The basic method or process used to evaluate the fibers is as follows:

1. For each fiber type, couple 25 m into the system.
2. *Calibration step*: At low laser intensity, measure the power before and after the fiber to calculate the coupling efficiency. Note that coupling in the range of $\sim 50\,\%$ is considered fairly high / good. Next, using the spectrometer, record the output spectrum. The area under the curve of the output spectrum is proportional to the optical power. Thus, an equivalence is made between the spectrometer readings and the already recorded power after the fiber.
3. Incrementally (and systematically) vary the laser parameters, and record the output spectrum. This is done for every combination of laser parameters.
4. Once this is done, the fiber is then cut-down in length (generally in increments of 2 m), and the process of calibration and varying the laser parameters is repeated.

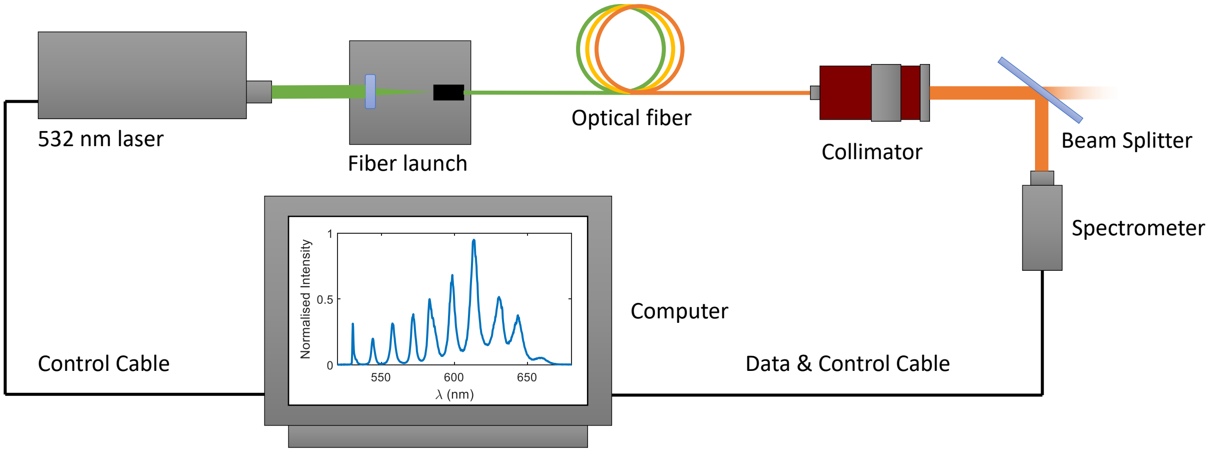


Figure 4.1: The experimental apparatus used to characterize SRS in optical fibers is illustrated here. Light from the laser source is coupled into optical fiber, where SRS occurs, generating additional wavelengths. The spectrum of the light after exiting the fiber is measured and recorded.

These steps are repeated for all fibers, ultimately yielding thousands of experimentally measured spectra associated with parameter combinations.

4.2 Experimental Results

The measured spectra comprise a data-set with dimension associated with the parameters of fiber type, fiber length, input pulse energy, pulse repetition rate, and pulse width. Rather than attempting to visualize and interpret the entire data-set at once (which would be cumbersome due to its dimensionality), representative "slices" of the measured data are shown in the form of graphical plots, examining one specific effect or parameter while holding all the other parameters constant. For each of the explored parameters, outlined in Table 4.2 (as well as fiber type), the generalized trend in findings is discussed.

4.2.1 Effect of Fiber Length

The first parameter explored is the fiber length. More specifically, adjusting the fiber length alters the interaction length over which SRS can occur. As anticipated, with longer fibers (all else held constant), higher-order Stokes waves are generated. The cascading effect (introduced in Chapter 2 on page 21) requires both sufficient length and input intensity to occur. Thus, for sufficient input intensity, length can be a limiting factor. Repeated from Chapter 2, Equation (2.21) is shown again here for reference:

$$\begin{aligned} \frac{\mathrm{d}I_s}{\mathrm{d}z} &= \qquad g_R I_p I_s - \alpha_s I_s \\ \frac{\mathrm{d}I_p}{\mathrm{d}z} &= -\frac{\omega_p}{\omega_s} g_R I_p I_s - \alpha_p I_p \end{aligned} \tag{4.1}$$

In accordance with the governing set of differential equations, both the intensity of the pump, I_p, (and Stokes, I_s) and an adequate fiber length is required for the cascading effect to occur. As length is increased, one can think of the differential equations being integrated over a longer range (z). This of course means further Stokes wave generation and subsequently cascaded higher-order Stokes waves as well.

Figure 4.2 shows how the resulting SRS spectra change as fiber length is increased. One can readily observe that as fiber length is increased, higher-order Stokes waves are generated, attenuation occurs (decreasing the overall power and gradually preventing additional Stokes wave generation), and spectral broadening of the SRS peaks occurs.

4.2.2 Effect of Input Intensity

The next parameter to examine is the effect that the input pulse intensity has on SRS generation. Again, following clearly from the theory of Stokes wave intensity growth from Equation (4.1), SRS occurs at a faster rate and higher-order Stokes waves are generated with greater intensity light. Both the rate of spontaneous and stimulated Raman scattering depend on the intensity of light, or rather, the abundance of photons available to interact with the medium. Therefore, it is no surprise that in Figure 4.3, one can clearly see that higher-order Stokes waves are generated as intensity (gain) is increased. Additionally, some anti-Stokes wave generation occurs for higher intensity. Spectral broadening for higher-order Stokes waves is seen here due to the cascading / convolution-like nature of SRS [148]. Subsequent peaks are at least as broad as the previous ones, and tend to be even broader because of the broadness of the Raman gain spectrum of fused silica itself.

4.2.3 Effect of Fiber Type

Here, a comparison between a selection of fiber types (summarized in Table 4.1) is made. Major differences between examined fiber types include the use of dopants, the fiber core

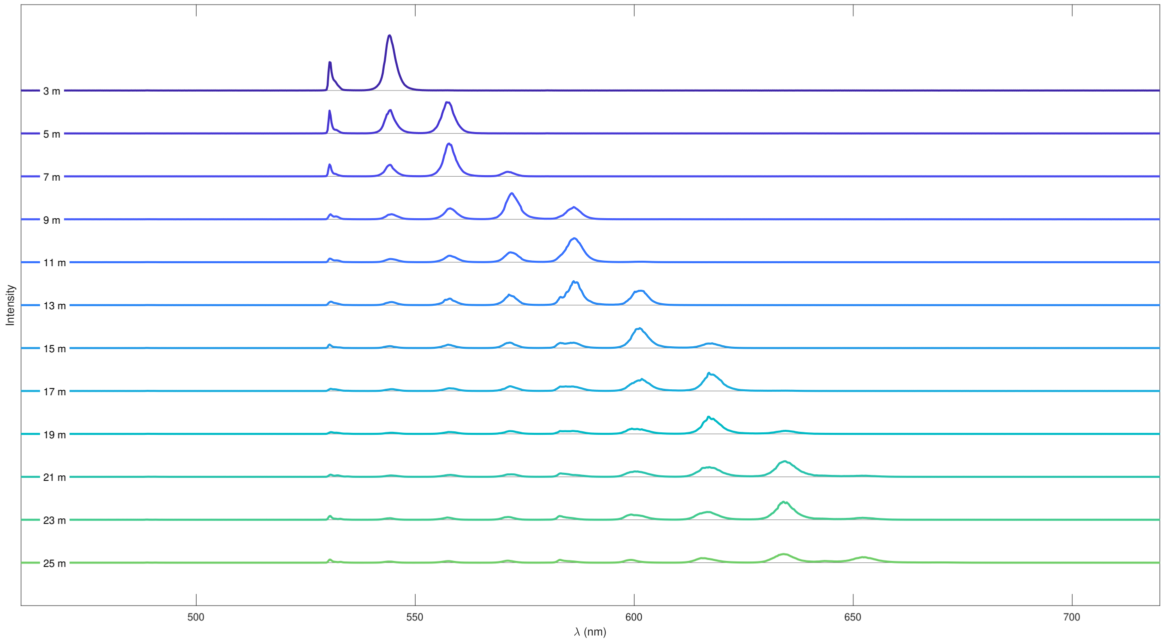


Figure 4.2: For a fixed input intensity and fiber type, observe that increasing the fiber length leads to higher-order Stokes wave generation, greater attenuation, and increased spectral broadening. Spectra are plotted in order of increasing fiber length, from top to bottom. Intensity is plotted on a linear scale.

diameter (affecting the wave-guiding ability for certain wavelengths and the intensity of the light as a result of the effective cross-sectional area), and how well the fiber maintains the polarization of light (birefringence). In general, the explored fibers had fairly comparable mode-field diameters (related to the fiber core diameter and the effective cross-sectional area) and no trend was seen with mode-field diameter. The dominant fiber characteristics affecting high-order Stokes wave generation were attenuation and how well the fiber maintains polarization. Figure 4.4 shows the resulting SRS spectra for a selection of four fibers, illustrating the effects that attenuation and maintenance of polarization have.

Intuitively, low attenuation is favorable for the SRS process, leaving greater intensity for Stokes wave generation. Additionally, it has been shown that the use of polarization-maintaining fiber can decrease the critical power (power at which Stokes wave intensity equals that of the pump wave, for a given fiber length) by a factor of two [146], meaning

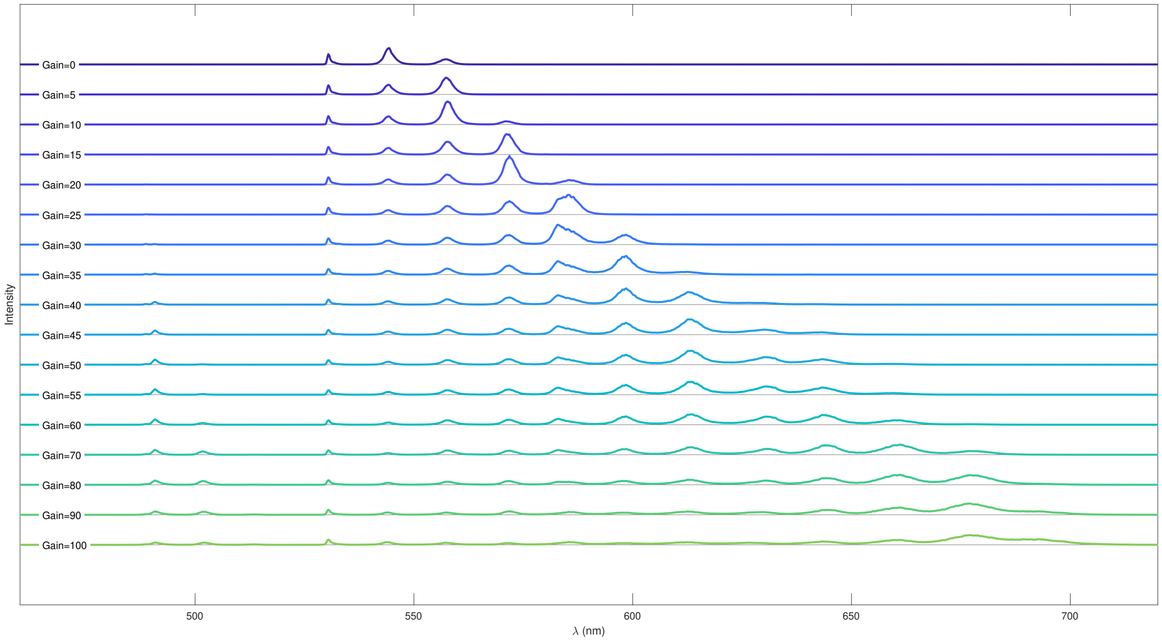


Figure 4.3: In contrast to the spectra shown in Figure 4.2, where fiber length was varied, here spectra are shown for variable input intensity. For a fixed fiber length and fiber type, when increasing the input intensity (gain), higher-order Stokes wave generation, attenuation at high wavelengths, and spectral broadening occur. Input intensity is lowest for the spectrum at the top of this figure (gain = 0), and increases moving downwards (finally to gain = 100). Again, intensity is plotted on a linear scale. Additionally, while not shown explicitly in this figure, burning the proximal end of the fiber occurs at sufficiently high input intensity. This imposes a fundamental upper bound on the intensity of light which may be coupled into the fiber.

that higher-order Stokes waves can be generated than with a similar non-polarization-maintaining fiber. This effect is a result of the fact that the Raman gain spectrum is sensitive to the polarization difference between the pump and Stokes waves [144, 146]. The Raman gain spectrum shown in Figure 2.4 of Section 2.3.2 is only representative for co-polarized pump and Stokes waves. For the orthogonally polarized case, the gain is significantly lower, requiring greater pump wave intensity for similar Stokes wave generation. Although Stokes photons maintain the polarization state of the pump photon that lead to their creation, in non-polarization-maintaining fiber, polarization is scrambled over the length of the fiber, meaning that it is less likely for pump photons to interact with Stokes waves of the same polarization state, ultimately leading to a reduction in the generation of higher-order Stokes waves. Highly birefringent fibers maintain polarization especially well, and thus exhibit excellent high-order Stokes wave generation. For HB600 fiber, shown in the bottom panel of Figure 4.4, Stokes waves up to approximately the 19th order are generated, although other effects are prominent, especially beyond 700 nm.

4.2.4 Effect of Pulse Repetition Rate

The pulse repetition rate has no impact on the SRS process for the range explored here (up to 300 kHz). This is because the temporal response is on the order of femto-seconds, orders of magnitude shorter then the time between pulses [149, 150]. Since there is no temporal overlap between excited locations in the medium (fused silica fiber) and subsequent pulses, altering the pulse repetition rate has no influence on the resultant SRS spectra. Figure 4.5 shows four resultant spectra at pulse repetition rates ranging from 40 kHz all the way to 300 kHz, with no significant difference. Although minor variations are present, these are a result of variations in the laser intensity. Precisely maintaining laser intensity over a range of pulse repetition rates was a challenge in this experiment, and more broadly, because Q-switched lasers generally operate at constant power settings, meaning that intensity is inversely proportional to pulse repetition rate.

4.2.5 Effect of Pulse Width

Experimentally, varying the pulse width had negligible impact throughout the nano-second range and high-order Stokes wave generation was possible. However, with the pico-second source, pulse walk-off and higher-order effects of self-phase modulation (SPM) [141], cross-phase modulation (XPM) [142], and four-wave mixing (FWM) [143] dominate and compete with SRS [94], making spectrally-distinct generation of *any* Stokes wave impossible. Figure 4.6 demonstrates the resultant spectra from the 2 ps source. Walk-off lengths of roughly

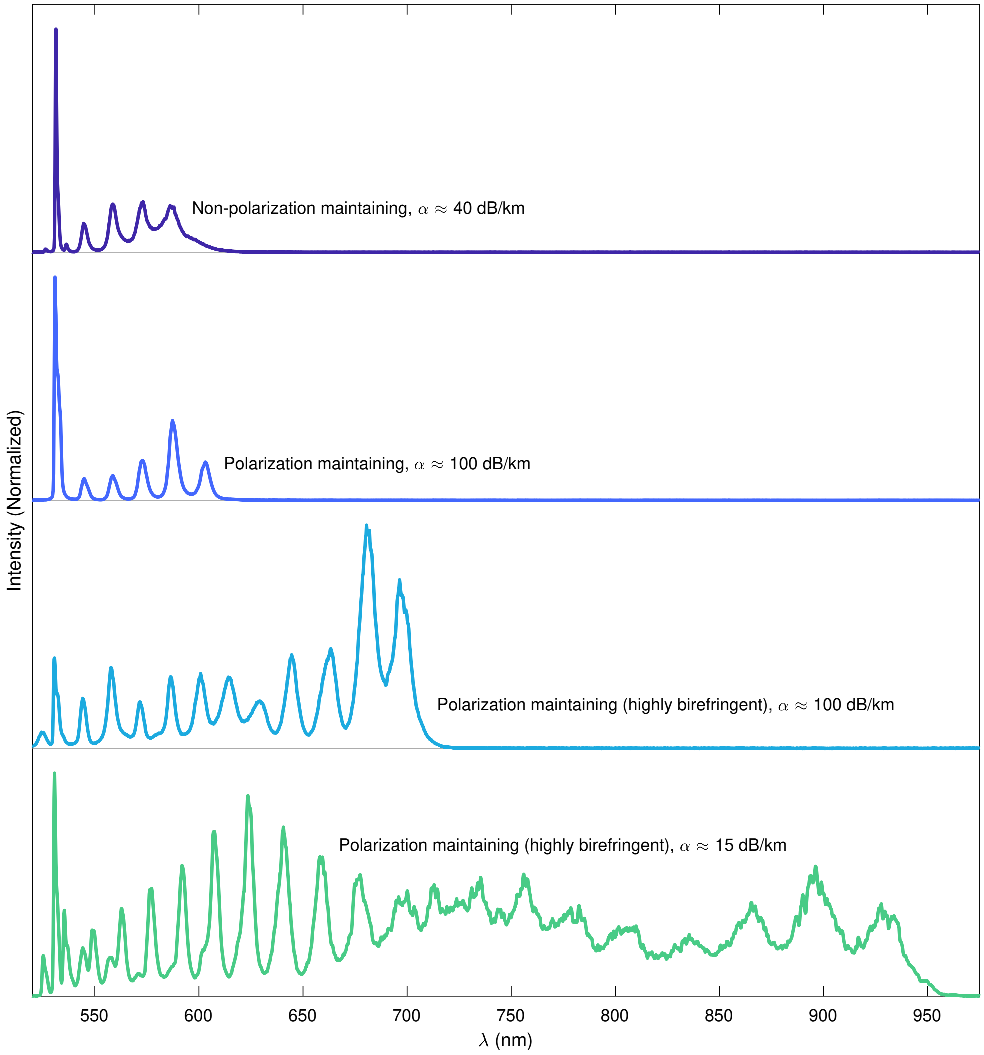


Figure 4.4: The choice of fiber influences the SRS generation ability through two main effects: optical attenuation and the maintenance of polarization. Low attenuation and high birefringence (yielding a high degree of polarization maintaining ability) are conducive to SRS generation. Shown here, in order from top to bottom, are spectra from 630HP, PM460-HP, HB450, and HB600 fiber. Note that while the PM460-HP fiber is considered to be polarization maintaining, its birefringence is less than that of HB450. Again, intensity is plotted on a linear scale.

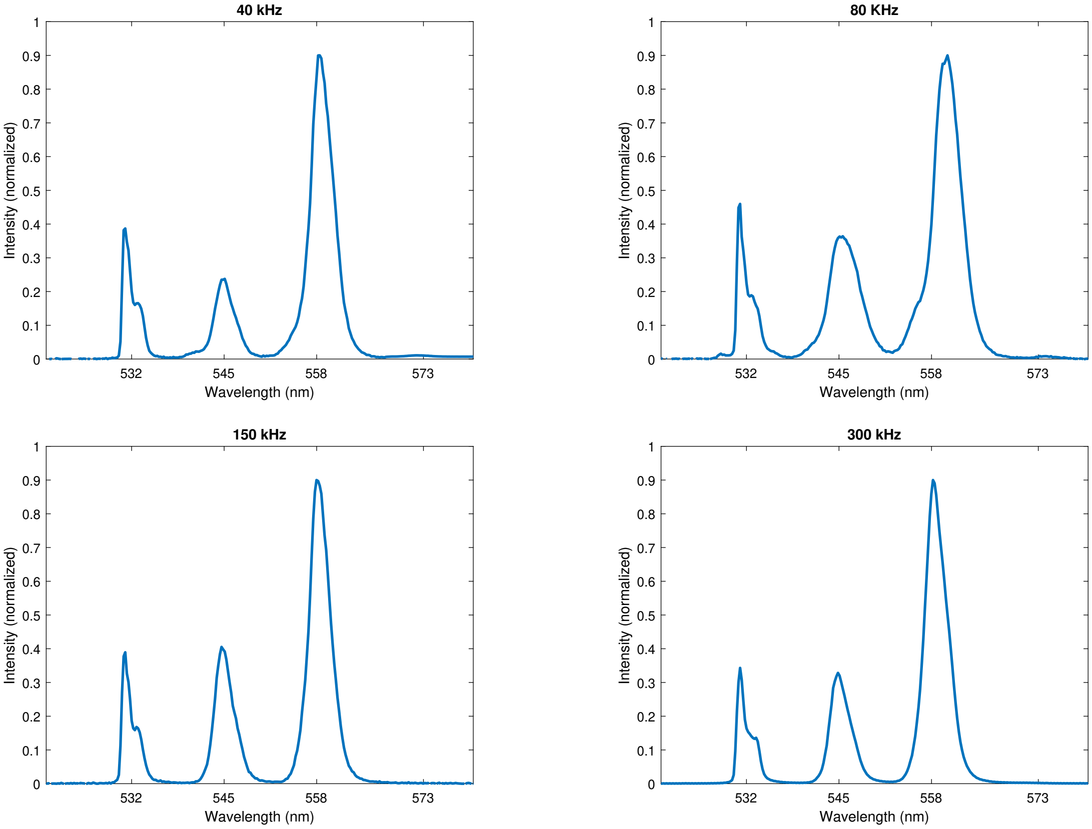


Figure 4.5: Resultant spectra at four pulse repetition rates are plotted here. The pulse repetition rate does not affect the SRS generation process. Since the effective relaxation time for the SRS effect is in the femto-second range in the scattering medium, at pulse repetition rates within the range explored, there is no spatial overlap between excited locations in the medium and subsequent pulses. Though small variation is seen between the four spectra shown, these differences are a result of variations in the laser intensity.

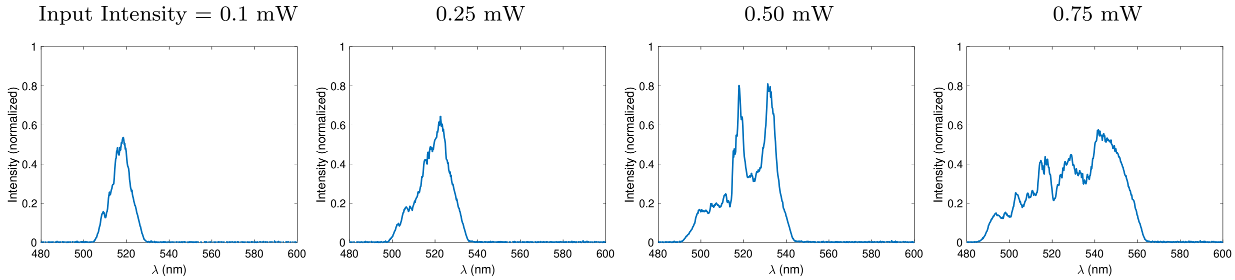


Figure 4.6: For short pulse widths, for example in the pico-second range (2 ps shown here), effects of pulse walk-off and higher-order effects of self-phase modulation, cross-phase modulation, and four-wave mixing dominate and compete with SRS [94]. The result is that high-order Stokes wave generation is impossible and spectral broadening is extreme. Plotted here, in order of increasing input intensity, are the resultant SRS spectra generated from the 2 ps source.

20 cm can be expected for 2 ps pulse widths [94], meaning that the use of fibers longer than this is futile and only leads to increased spectral broadening as a result of the competing higher-order effects.

4.3 A Simulation of SRS in Single-mode Fiber

To aid in understanding the theory governing SRS and to create correspondence with experimentally measured results, a simulation of the SRS process was created. In essence, the simulation solves the coupled differential equation in Equation (2.21) over a specified length. The simulation includes a model of the Raman gain spectrum, $g_R(\Omega)$, of fused silica, shown in Figure 2.4, a reasonable estimate for the true Raman gain spectrum of any given fiber (possibly containing additional unknown dopants). Additionally, parameters specific to the fiber including the mode-field diameter (incorporated as the effective area, A_{eff}, for use in computing optical intensity), and the attenuation spectrum, $\alpha(\lambda)$, where available, were included.

For improved accuracy, the simulation is implemented in such a way that no assumptions on the center wavelengths of the pump or Stokes wave are made. A simpler implementation might have assumed each Stokes wave was simply shifted by Ω = 440 cm^{-1}; however, this neglects spectral broadening as a result of the convolution-like nature of SRS. Instead, the spectral domain is discretized in steps of 0.25 nm, and the Raman gain spectrum is interpolated such that Raman gain at each wavelength step is evaluated. This

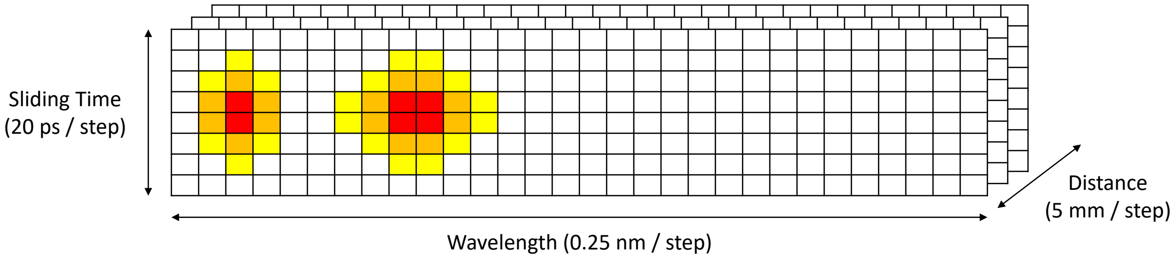


Figure 4.7: Discretization to simulate SRS propagation down the length of an optical fiber was done over wavelength (in steps of 0.25 nm), distance down the fiber (in steps of 5 mm), and over a sliding time-window following the pulse as it travels down the fiber (in steps of 20 ps, which correspond to a propagation distance of approximately 0.4 mm). The coloured pixels represent the optical intensity of an example pump (left) and first Stokes wave (right).

also allows the spectral profile, $I_p(\lambda)$, of the pump wave to be specified. Additionally, the temporal domain is discretized in 20 ps steps (which corresponds to a propagation distance of approximately 0.4 mm), allowing for arbitrary input pulse profiles, $I_p(t)$, to be used and for the pulse profiles of the resulting Stokes waves, $I_{s,i}(t)$, to be examined. For a reasonable representation of the lasers used experimentally, Gaussian profiles were used with their full width at half maximum (FWHM) [94] defined by the nominal pulse width. Note, for a Gaussian with standard deviation σ, this is given by $\mathrm{FWHM} = 2\sqrt{2\ln(2)}\sigma$. The temporal domain is implemented as a sliding time-window, following the pulse as it travels down the fiber, such that the center of the pulse remains at the same shifted time throughout the simulation. Lastly, the differential equation is evaluated in distance steps of 5 mm, yielding high-resolution simulation results over distance, time, and wavelength. Figure 4.7 offers a visual representation of the discretization scheme used for the simulation.

The simulation generally yields results similar to what was measured experimentally and gives thought-provoking insights into the fine details of the SRS process. One such detail, the evolution of the pulse, is explored here.

4.3.1 Comparison between Simulated and Experimentally Measured SRS Spectra

The SRS spectrum was simulated for HB450 fiber, for which Fibercore Ltd. provided the attenuation spectrum, $\alpha(\lambda)$, (see Figure 4.8), thereby aiding in producing an accurate

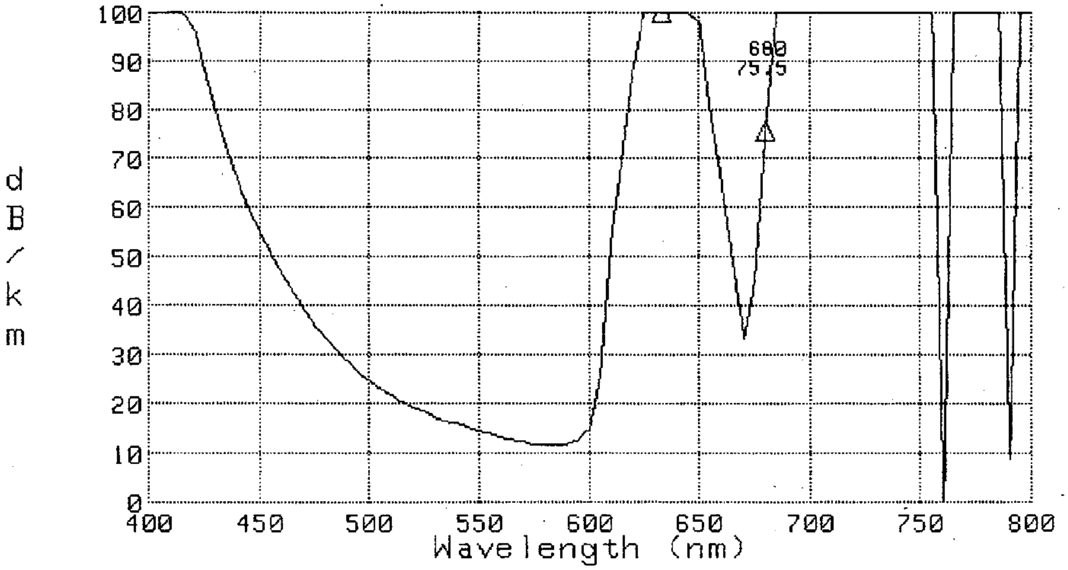


Figure 4.8: The attenuation spectrum, $\alpha(\lambda)$, of HB450 fiber, provided by Fibercore Inc. The structure and details shown demonstrate the difference between the nominal attenuation constant ($\leq$100 dB/km @ 488 nm for HB450) and the true attenuation characteristics. Fibers designed for NIR and IR generally have much smoother attenuation spectra, with predominantly Rayleigh scattering causing attenuation. Here, attenuation is relatively low until approximately 625 nm, and becomes extremely erratic beyond approximately 680 nm, where the fiber ceases to guide well. This matches the observed SRS spectra, shown in Figure 4.2, Figure 4.3, the third panel of Figure 4.4, and later, in Figure 4.10, where few Stokes waves above 680 nm are generated.

result. In principle, other fiber types could also have been simulated; however, HB450 was the only fiber type for which an adequate attenuation spectrum (rather than simple nominal attenuation) was available[1]. The resulting SRS spectrum is shown over a length of 25 m, with a spectral domain ranging from 530 nm to 930 nm, in Figure 4.9. Several intuition-matching insights can be drawn from this result.

Firstly, higher-order Stokes waves are seen for increasing distance along the fiber; however, the spacing, in distance (z), between subsequent Stokes waves subtly increases. This

---

[1] Attenuation spectra were available for HB450 and HB600 fibers; however, for HB600, attenuation information was only available for wavelengths greater than 580 nm, making it unsuitable for simulating SRS with a 532 nm pump wavelength.

is expected as a result of several related causes. For each additional Stokes wave, the total power is more spread-out spectrally, meaning there is less available intensity in the preceding Stokes wave acting as a pump, therefore longer and longer distance is required before the next Stokes wave emerges. Additionally, as the pulse travels down the fiber it is attenuated, once again leaving less and less intensity available for subsequent Stokes wave generation.

Secondly, each Stokes wave is spectrally broader than the one before it. This is a direct result of the convolution-like nature of SRS, and is entirely expected [148]. Effectively, the pump-laser spectrum, $I_p(\lambda)$, is convolved with the Raman gain spectrum, $g_R(\Omega)$, to yield the spectrum of the first Stokes wave, $I_{s,1}(\lambda)$. In the spectral domain, the result of convolution can be thought of as effectively having bandwidth approximately equal to the sum of the bandwidths of the two convolved spectra. Thus, for each subsequent Stokes wave (convolution of preceding Stokes wave spectrum with the Raman gain spectrum), the bandwidth, or spectral broadness, increases.

For visual comparison, experimentally measured SRS spectra from the HB450 fiber are displayed in a similar fashion in Figure 4.10. Because measurements were made at far coarser increments in fiber length, linear interpolation is used to fill in the gaps. A glaring difference exists between the spectra shown in Figure 4.9 and Figure 4.10, beyond wavelengths of approximately 680 nm. This is a result of the lack of detailed attenuation information available in this range (see Figure 4.8). It is known that HB450 does not guide well beyond 680 nm. Therefore, to observe better correspondence between the simulation and the experimental measurements, in principle, the simulation should be re-run with a more realistic attenuation spectrum.

For the preceding wavelength range, there is good correspondence between the simulated and measured spectra, at least at a qualitative level. Stokes waves emerge at roughly the same positions along the fiber, and at roughly the same central wavelengths (expected based on the main peak of the Raman gain spectrum). The largest difference is in the spectral broadness of the Stokes waves, $I_{s,i}(\lambda)$. The broadening may be a result of higher-order effects, namely four-wave mixing. Nonetheless, the correspondence is reasonable enough that the simulation may be useful as a guide when selecting fibers and laser parameters for use in a multi-wavelength imaging system.

4.3.2 Temporal Pulse Profile Evolution through SRS Process

As alluded to earlier in this section, on page 68, the simulation revealed interesting details of how the temporal pulse profiles of the pump, $I_p(t)$, and Stokes waves, $I_{s,i}(t)$, evolve

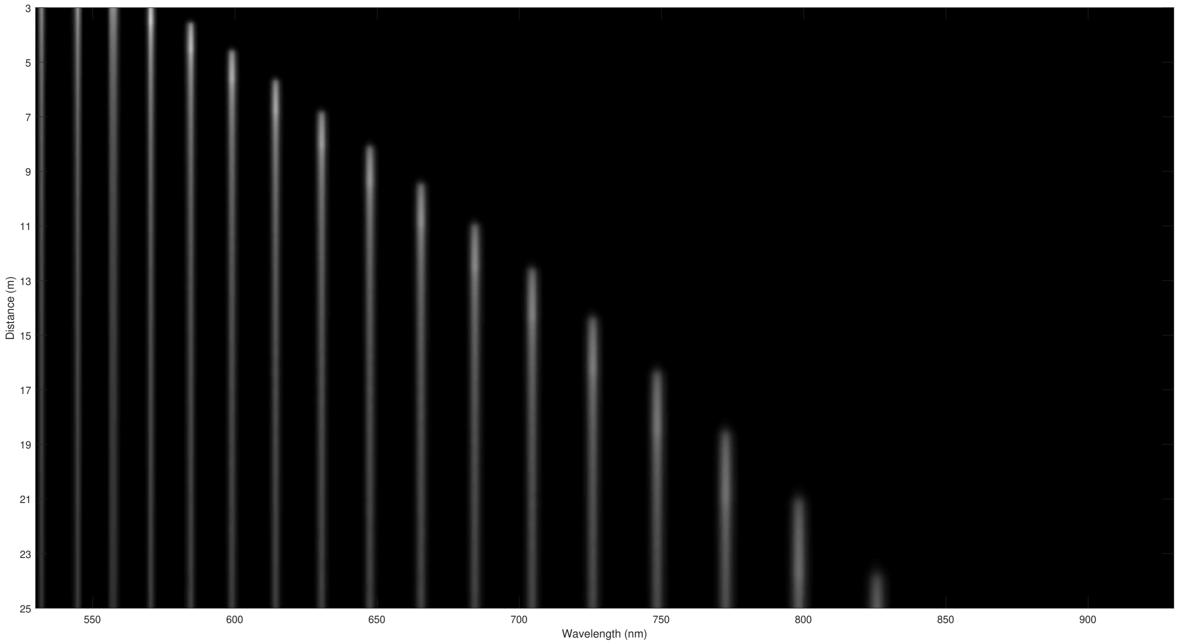


Figure 4.9: The SRS spectrum is simulated throughout the length of an optical fiber. Generation of higher-order Stokes waves are seen for increasing distance down the fiber. A gamma correction is applied to aid in visualizing the spectrum towards the end of the fiber where the intensity is quite low relative to at the start of the fiber due to the power being spread-out spectrally and attenuated by the fiber.

throughout the SRS process. Based on Equation (2.21), it is known that the rate of Stokes wave intensity growth is proportional to the intensity of the pump wave; however, the pump pulse does not have a uniform temporal profile. This means that more intense parts of the temporal pump pulse profile will begin Stokes wave generation at shorter travelled distances than the less intense parts. For a Gaussian pulse, this corresponds to the intense central region versus the weaker tails. Figure 4.11 illustrates this effect through simulation, at 1 m increments along the fiber. Correspondingly, Figure 4.12 shows the evolution of the first Stokes wave at the same positional increments.

Because it is most intense, power from the central region of the pump profile is converted first, contributing to the growth of the central region of the first Stokes wave. The result is a central erosion of the temporal pump pulse profile. The weaker tails of the pump profile

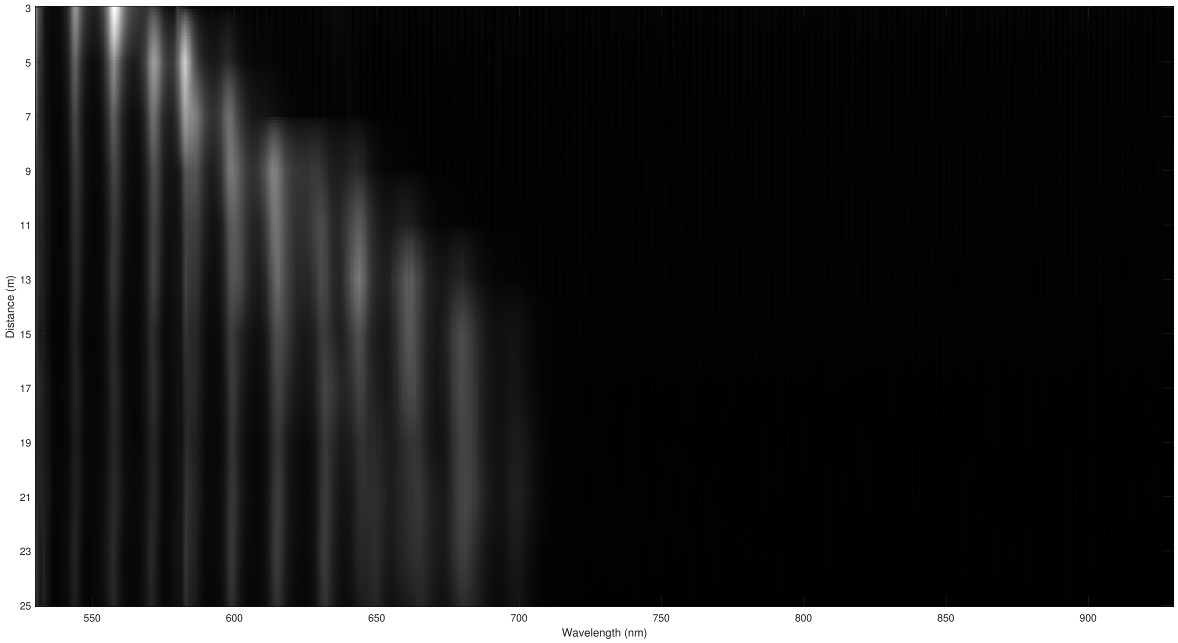


Figure 4.10: For comparison to Figure 4.9, the SRS spectrum is experimentally measured at increments in optical fiber length. The resulting experimentally measured spectra are visually similar to the simulated ones for wavelengths less than 680 nm. Higher-order Stokes wave generation is seen for increasing fiber length. Linear interpolation is used to fill in gaps between measured lengths, and a gamma correction is applied to aid in visualizing the spectrum towards the end of the fiber where intensity is quite low relative to at the start of the fiber due to the power being spread-out spectrally and attenuated by the fiber.

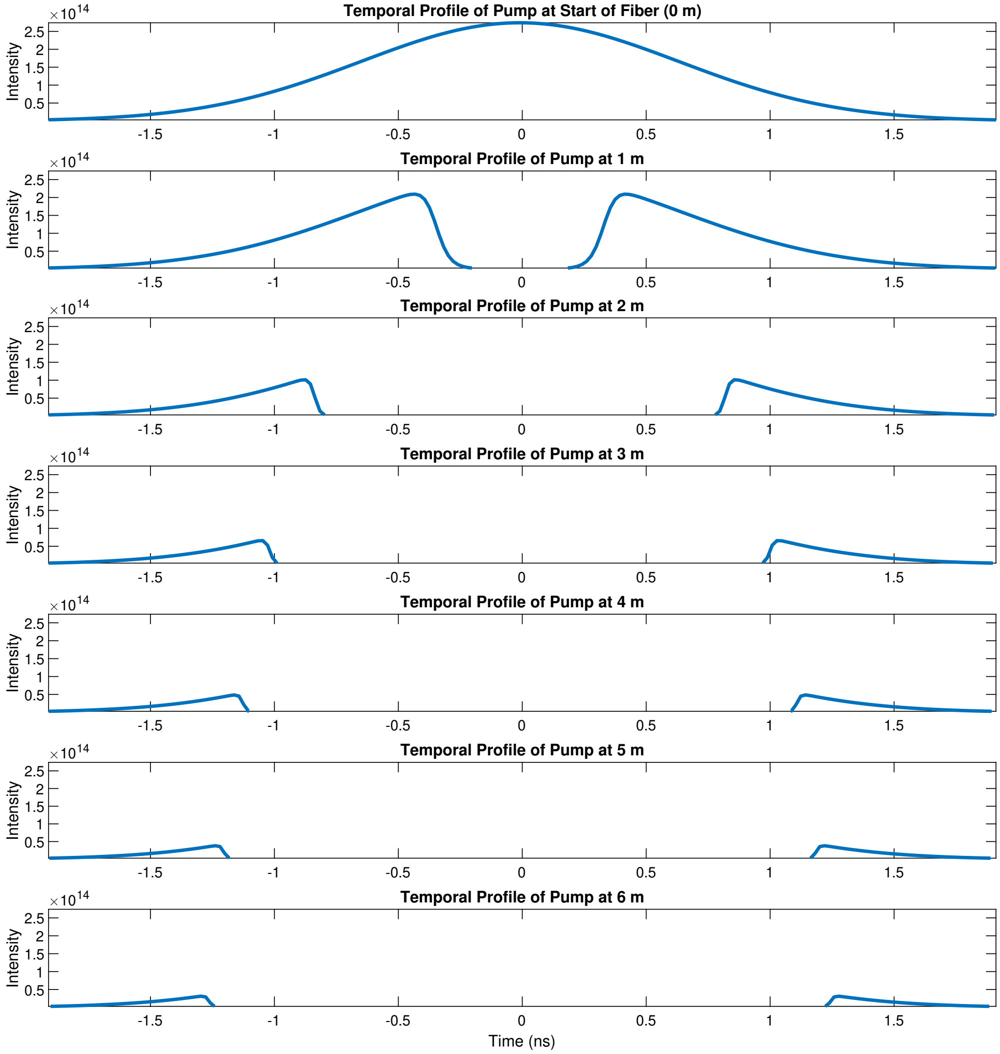


Figure 4.11: The simulated temporal profile of the pump pulse, $I_p(t)$, evolves due to SRS as it travels down the fiber. Initially, the profile is Gaussian; however, the highly intense central region undergoes SRS conversion, generating the next Stokes wave, in a shorter distance than the less intense outer regions. This causes a central erosion effect.

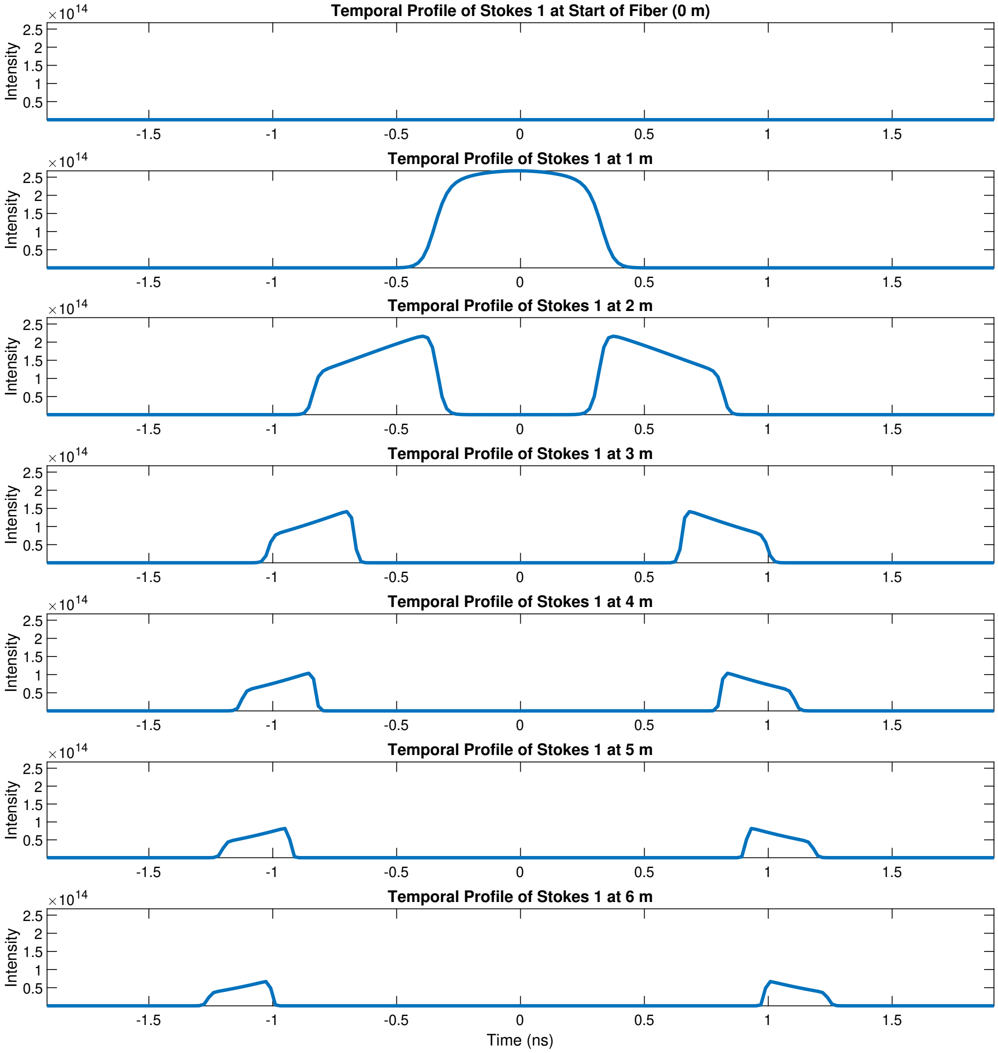


Figure 4.12: Corresponding with Figure 4.11, this figure shows the simulated temporal profile of the first *Stokes* pulse, $I_{s,1}(t)$, as it evolves over the length of the fiber. Initially, no Stokes photons are present (zero intensity), but through SRS, photons from the highly intense central temporal region of the pump pulse are converted to Stokes photons. This process continues through the less intense tails of the pump pulse, temporally broadening the Stokes pulse; however, simultaneously, the newly formed intense central region of the first Stokes pulse begins act as a pump for the second Stokes pulse, resulting in the same central erosion effect as is seen in the original pump pulse.

require more distance before sufficient spontaneous Raman scattering occurs, such that rapid conversion of power from the pump to the Stokes wave ensues. The same effect is at play in the Stokes waves. Initially the central region of a Stokes wave grows; however, once this becomes intense enough, the process repeats, leading to the central erosion of the Stokes wave profile and the growth of the central region of the next Stokes wave.

After just 3 m, the pump profile has a central gap of approximately 2 ns, which is *very* significant for a pulse characterized nominally by its full width at half maximum duration of 1.5 ns. The remaining tails are far less intense than the original pulse, but nonetheless, the pump pulse has effectively become two pulses. The impacts that this may have on PARS imaging are explored as Future Work in Section 6.1.

In more detail, Figure 4.13 shows the emergence of the central erosion effect taking place over a length of just 10 cm. This occurs at the point when sufficient Stokes photons are present for rapid conversion of power from the pump to the Stokes wave.

Note that for a perfectly uniform temporal pump pulse profile, the erosion effect would also be uniform, resulting in the *entire* profile being converted to the Stokes wave. That would mean that rather than obtaining a comb-spectrum, a spectrum of only one peak would arise – except for brief points where both a pumping wave and the subsequent Stokes wave are present, before the pump wave is depleted. Although this effect was not experimentally validated by observing the pulse profiles, the comb-spectra experimentally measured are evidence of this effect, which has been studied by others [156, 217].

4.4 Chapter Conclusion

In this chapter, the effects associated with parameters of fiber type, fiber length, input pulse energy, pulse repetition rate, and pulse width, on multi-wavelength light generation through SRS were comprehensively studied. Additionally, a simulation of the SRS effect in optical fiber was developed and used to aid in understanding SRS from a theoretical standpoint and to create correspondence with the experimentally measured results.

Generally, longer fiber lengths and greater input intensities lead to higher-order Stokes wave generation; however, attenuation over the length of the fiber and challenges with coupling a high-power pump beam into the fiber (burning the fiber tip) are limiting factors. Additionally, the choice of fiber type can have enormous effects on the ability to generate high-order Stokes waves. The two most important properties of the fiber are its operating wavelength range and ability to maintain polarization. Severe attenuation occurs for wavelengths outside of the operating range, preventing additional Stokes wave generation.

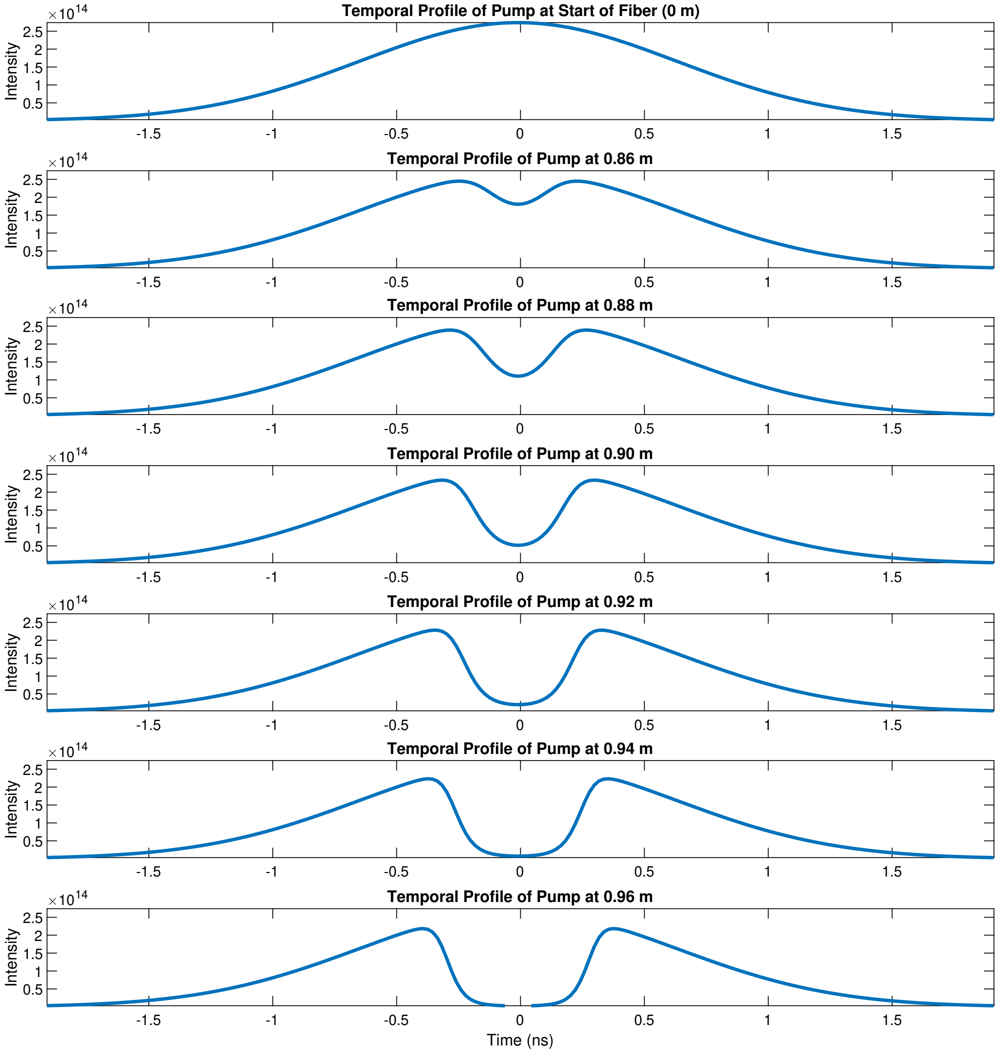


Figure 4.13: The same physical process is plotted here as in Figure 4.11, but on a much shorter distance-scale (10 cm) where it is possible to observe the initial onset of the central erosion effect.

Highly birefringent fiber, which maintains polarization over greater lengths, was found to dramatically extend the range of wavelengths that could be generated through SRS.

The pulse repetition rate was found not to affect the SRS process. This is expected, given that the effective relaxation time for the SRS effect is in the femto-second range, whereas for the pulse repetition rates explored, the time between consecutive pulses was orders of magnitude longer than that relaxation time. As such, consecutive pulses do not interact and thus the pulse repetition rate has no effect. This is beneficial in high-speed imaging applications, where imaging rates depend on the pulse repetition rate, allowing the pulse repetition rate to be set arbitrarily without concerns of impacting the SRS multi-wavelength light generation process.

The pulse width did not affect the SRS process throughout the nano-second range; however, orders of magnitude lower, at the pico-second range, pulse walk-off and higher-order optical effects dominated, competing with the SRS process and severely limiting the generation of Stokes waves. Thus, for applications involving nano-second pulses, SRS remains an effective means of generating multi-wavelength light.

By simulating the SRS process, several effects corresponding well with experimentally measured results were observed. These include the appearance of higher-order Stokes waves for increasing distance along the fiber and spectral broadening of Stokes waves. Simulated spectra showed excellent correspondence to experimentally measured spectra, with Stokes waves emerging at approximately the same positions along the fiber. Substantial differences arose only as a result of missing fiber attenuation information beyond a certain wavelength.

The simulation also revealed the temporal pulse profile evolution though the SRS process. Of interest was the central erosion effect, whereby the highly intense central portion of a pulse is converted to the next Stokes wave, leaving behind the less intense tails of the pulse, separated temporally by as much as approximately 2 ps over distances of only 3 m. The effects of using centrally eroded pulses in imaging applications such as PARS remain unknown.

To conclude, by understanding the effect each parameter has, informed decisions regarding parameter selection can made in future work involving SRS as a means of creating a multi-wavelength light source. Additionally, assuming adequate information on fibers of interest is available, simulation may prove to be useful as a guide for selecting parameters.

Chapter 5

PARS Time-Domain Signal Intelligent Feature Extraction

This chapter focuses on methods to extract information from PARS [3, 50–63] time-domain (TD) signals. PARS TD signals contain relatively large amounts of data, represented as arrays of samples, measured and recorded by a data acquisition card. The length of the signals are set by the PARS microscope operator; however, lengths of 256 up to 1024 samples per pixel are commonplace. During image formation, TD signals conventionally are projected to scalar-value representations. Section 2.4 introduced conventional methods for extracting amplitude information from PARS TD signals, two state-of-the-art feature extraction methods for photoacoustic microscopy more broadly [179, 180], and finally a primer on feature extraction. Because of the quantity of data that are collected during PARS imaging, the TD signals are difficult to interpret directly; however, information beyond what can be represented by a scalar (as in conventional methods) is desired, thus necessitating concise information extraction. Key motivating factors for the work presented in this chapter include the desires to improve imaging contrast and to achieve target specificity though improved signal processing.

To recap somewhat from Section 2.4, conventionally, PARS TD signals are projected to scalar values that can directly be used as pixel intensities in reconstructed PARS images. This is achieved though one of two main methods: one, based on the Hilbert transform, which extracts the amplitude of the envelope of the signal, defined in Equation (2.24), and the other, based simply on the Maximum minus Minimum amplitude of the TD signal itself, defined in Equation (2.25).

While these methods have been effective in works thus far, it would be beneficial to extract far more information from the PARS TD signals, beyond single scalar values. Improving imaging contrast through signal processing would be excellent, since it could mean that less excitation laser fluence may be required during imaging, thus improving safety for eventual clinical usage of PARS microscopy. Furthermore, tissue-specific / clinically relevant information may lay hidden within the PARS TD signals. Extracting this would obviously be beneficial for clinical uses, would provide insights beyond simply quantifying optical absorption, and would advance the technology of PARS microscopy considerably. Additionally, with current methods, obtaining tissue-specific information requires multi-wavelength imaging (multiple scans over the target) and unmixing; however, if this information were obtained directly from PARS TD signals produced from a *single* excitation wavelength (only one scan over the target), imaging times and the PARS system complexity could both be reduced.

Section 5.1 outlines a preliminary attempt to address this problem, showing success in improving imaging contrast, then in Section 5.2, a feature learning approach based on K-means is developed and applied, enabling some level of target specificity in both slides of human breast tissue and freshly resected murine brain tissue.

5.1 Preliminary Work

By incorporating information from *entire* PARS TD signals, it is proposed that more information may be extracted than what conventional projection methods (such as Maximum minus Minimum, which only directly uses information from *two* points) produce, which may lead to improved image contrast, measurement accuracy, and new visual information not ever present with current imaging methods. This section primarily explores the use of matched filtering [218] to improve PARS imaging contrast.

A matched filter is a linear filter capable of optimally extracting a known signal shape in the presence of additive noise [219]. Essentially, the filter is the correlation of the noisy measured signal with an "ideal" (noiseless) version of the expected signal shape.

Suppose that there exists an ideal PARS signal, $s_{\text{ideal}}(t)$, such that a measured TD signal, $s(t)$, is simply a scaled version of this ideal signal with added zero-mean noise, $\nu(t)$:

$$s(t) = a \cdot s_{\text{ideal}}(t) + \nu(t), \tag{5.1}$$

for scalar amplitude $a \in \mathbb{R}$. The matched filter then takes the form:

$$s(t)_{\text{filtered}} = s(t) * s_{\text{ideal}}(-t). \tag{5.2}$$

Notice the time-reversal of $s_{\text{ideal}}(-t)$, necessary to perform correlation via the convolution operator, $*$.

The next step is to extract the amplitude information from this. Assuming that PARS TD signals are time-aligned — a reasonable assumption given that the acquisition system is triggered using the same trigger signal as is used to trigger the excitation laser pulse — rather than using correlation, simply multiplying the ideal signal with the measured signal and integrating extracts the signal amplitude, a:

$$a = \int_0^T s(t) s_{\text{ideal}}(t) \, \mathrm{d}t. \tag{5.3}$$

Note that this relies on the ideal signal being normalized, such that

$$\int_0^T s_{\text{ideal}}(t) \, \mathrm{d}t = 1, \tag{5.4}$$

over the domain $t \in [0, T]$, determined by the acquisition length used.

The last remaining step before being able to apply this method to PARS images is to concretely define the ideal PARS TD signal. Given that relatively little is understood regarding the exact process signal formation, a simple method to *estimate* the ideal noise-free PARS TD is appropriate. The estimated ideal PARS TD, $\widehat{s_{\text{ideal}}}(t)$, is calculated by simply averaging the TD signals within a given image, I:

$$\widehat{s_{\text{ideal}}}(t) = \frac{1}{|I|} \sum_{\forall s(t) \in I} s(t). \tag{5.5}$$

Note that normalization, according to Equation (5.4), is still required.

This method is tested on a PARS image of a slide of formalin-fixed paraffin-embedded (FFPE) human breast tissue, shown in Figure 5.1. This image was captured and provided by Benjamin R. Ecclestone with gratitude from the author. This image was captured with 266 nm excitation and 1310 nm detection wavelengths. For the purpose of analysis, 500 μm by 500 μm regions are examined in closer detail; however, the estimated ideal signal for use in filtering is calculated by averaging over the entire image.

Figure 5.2 shows the results of matched filtering for the two sub-sections of the image. Visually it can be seen that contrast is improved in both dark and light regions of the matched filtered images. More specifically, there is greater clarity in the details, and the continuity of fine structures is greatly improved.

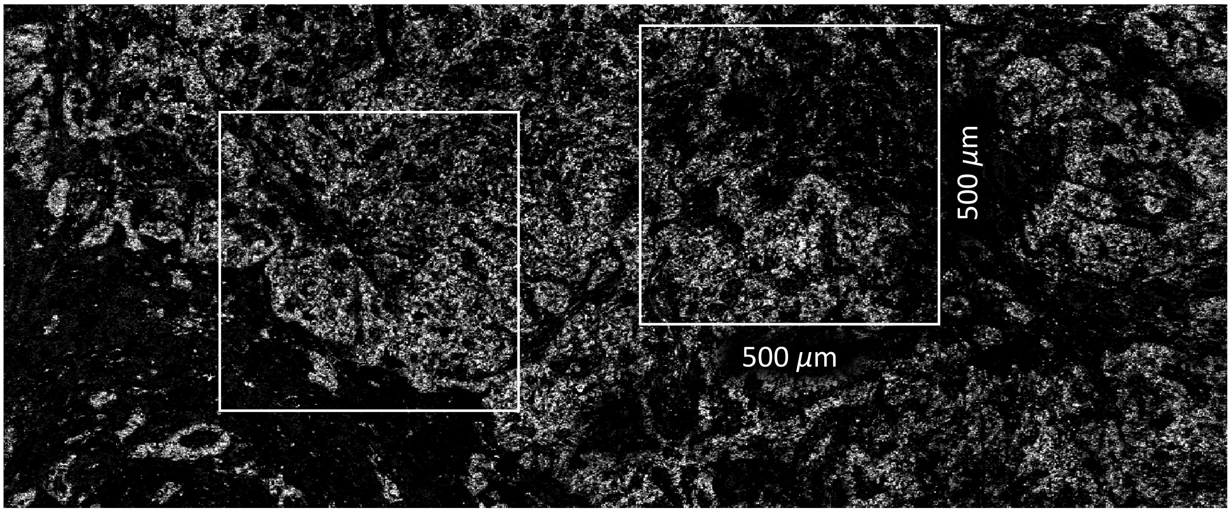


Figure 5.1: A standard TD signal projection PARS image of unstained human breast tissue on a slide, captured at 266 nm excitation and 1310 nm detection wavelengths. 500 μm by 500 μm regions are shown boxed-in and are later examined in Figure 5.2, where matched filtering is applied. Image captured by Benjamin R. Ecclestone.

Note that this method, matched filtering, uses only one characteristic signal shape and is thus limited by that. As was explained in the introduction to this chapter, it would be excellent to extract more information from the TD signals, perhaps information corresponding to tissue / target information. Having several characteristic signal shapes could enable this; however, this begs the question of *which* signal shapes to use and *where* to find them.

One solution is to use a lock-in amplifier / filter [220], or equivalently, a digital implementation of it. A lock-in filter enables frequency-based signal mixing, i.e., measured signals are correlated with sinusoids at specified reference frequencies, f_{ref}. The method is very similar to matched filtering whereby a sinusoid is used rather than an ideal signal. Using lock-in filtering in quadrature mode, whereby amplitudes from both $\sin(t)$ and $\cos(t)$ are combined via the Pythagorean theorem, yields a phase-insensitive amplitude associated only with the reference frequency. Effectively, a filter bank can be generated by specifying reference frequencies of interest.

This approach was explored by the author and presented in [221]. It was found that the filter did yield variations in contrast with reference frequency; however, rather than highlighting different tissue structures, the filter seemed only to broadly modulate contrast of the whole image, and furthermore did not exceed the contrast provided by the matched

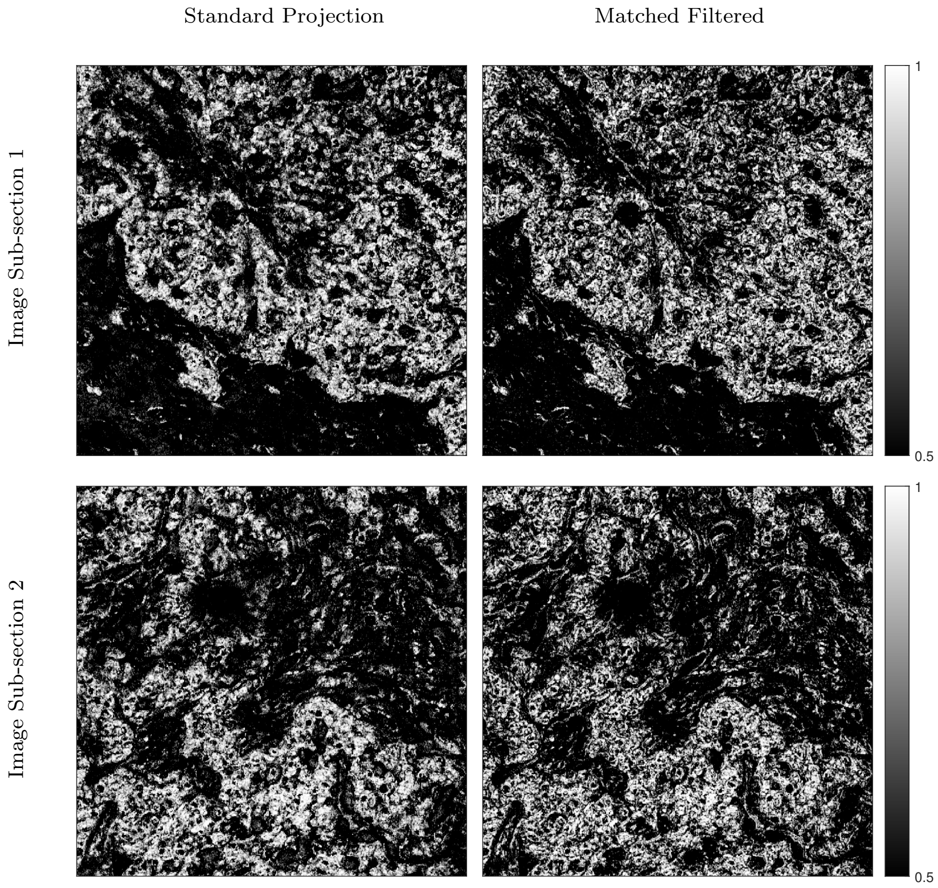


Figure 5.2: The standard projection method (left column) is compared to the matched filtered amplitude (right) for two sub-sections (defined in Figure 5.1) of a PARS image of unstained human breast tissue on a slide. Contrast is improved in the matched filtered images and the continuity of fine structures is greatly improved.

filter. This behaviour is largely explained by the fact that optimal contrast was achieved at a reference frequency corresponding to a sinusoid very similar in shape to the ideal PARS signal as was used in matched filtering. Lock-in filtering is conventionally used in situations where a signal of interested is modulated on top of a carrier waveform of known reference frequency [220]. Thus, this filtering method was not deemed successful in bringing forth additional information from the PARS TD signals.

Rather than using externally generated signal shapes, as is done in lock-in filtering, it would be advantageous to derive characteristic signals shapes directly from measured PARS TD signals. The next section explores this in detail and uses a modified K-Means clustering algorithm to extract characteristic signal shapes, representative of the signals present within the processed PARS image.

5.2 A Modified K-Means Algorithm

The previous section concluded by recognising the need for a method to derive, or learn, characteristic signal shapes from a given PARS image. Though matched filtering proved effective in improving image contrast and structural continuity, it was inherently limited by the fact that only one characteristic signal shape was used. In any given PARS image, there may be a wide range of prevalent signal shapes present, based on the properties of the target / bio-media that are imaged.

To address this, a method is proposed in this section to learn characteristic signal shapes, $f_i(t)$, from a given PARS image. This method is chosen based on the desire to extract signal shapes that are representative of pure or isolated components. Consider that any TD signal may include the PARS response associated with a variety of absorbers. What is desired is to learn signal shapes associated with isolated components, or absorbers. This would serve to enable evaluation of the chromophore composition associated with any given TD signal.

Section 2.4 introduced both principal component analysis (PCA) and clustering as methods for dimensionality reduction and feature extraction. On page 30, PCA and clustering (specifically K-Means) are compared and contrasted. While both methods serve to reduce dimensionality, they work in very different ways and produce very different features as a result. Recall that PCA projects the measurement data (in n dimensions), arranged into an $p \times n$ data-matrix, $X = \left[\vec{x}_1 \, \ldots \, \vec{x}_p\right]^T$, onto a reduced-dimensional basis, encapsulated as columns of V (right singular vectors, when SVD is used), thus forming a representation, $Y = XV$, based on combinations of the data based on directions in which the measurement data set exhibits maximal variance. These basis directions, maximal principal components, do not directly relate to the underlying information in the data (i.e., the local characteristics of the imaged target). In contrast, clustering represents the data (through labeling) in a way where the representation *does* relate specifically to an estimate of the underlying "type" or classification of the data, whether this estimate is actually in any way *correct* or not.

Note that measured TD signals are *not* externally band-pass filtered for use with the methods of this section, in contrast with the conventional approach described on page 7. This detail is *highly* central, and possibly essential, to this analysis. The act of band-pass filtering PARS TD signals in a conventional acquisition removes important signal content, substantially altering the shape of the PARS TD signals. The method developed here is based on the principle that the signal shape is indicative of the underlying characteristics of the target. Therefore, altering the signal shape or removing signal content by filtering is counterproductive and is thus not done.

5.2.1 Methods

Here, both feature extraction and feature learning methods are described. In brief, PARS TD signals, $s(t)$, are thought of as vectors in the space $\mathbb{R}^n$, where the dimension, n, of the space is simply the length of the PARS TD signal in discrete samples over time. Thus, the equivalence $s(t) \equiv \vec{s}$ is made. If a suitable set of basis vectors $\mathcal{F} = \{\vec{f}_i\}$ is available (or found through feature learning and selection), then a change of basis, $B_{\mathcal{F}}$, can be applied to the PARS TD signals, $\vec{\alpha}_j = B_{\mathcal{F}} s_j(t)$, resulting in a reduced-dimensional, feature-based representation, as described in Section 2.4. This reduced-dimensional representation lends itself well to interpretation and to generating colour-mapped images, which visually illustrate how the TD signals vary throughout imaged tissue.

Feature learning is used to discover characteristic signal shapes that are representative of pure or isolated target components and can be used to form a reduced-size basis. The method proposed is based on K-Means clustering and clusters based on the shape of the PARS TD signals. The motivation for clustering this way is based on the premise that TD signals from different targets / tissue components have different shapes. This phenomenon is visualized in Figure 5.3, where PARS signals are shown at three different points, corresponding with various tissue components, within a PARS image of unstained human breast tissue. The shapes of the three TD signals are distinct, demonstrating that signal shapes vary across tissue components. This image, again, was captured and provided by Benjamin R. Ecclestone with gratitude from the author. Thus, characteristic signals shapes may be discoverable for various bio-media of interest (nuclear regions, cytochromes, etc.). TD signals from a PARS image are clustered, resulting in a set of K centroids, $\mathscr{C} = \{\vec{c}_i\}$, for $i = 1, ..., K$. A subset of these centroids can then be used as basis vectors for feature extraction.

Feature Extraction

As discussed briefly above, feature extraction is achieved by performing a change of basis on the PARS TD signals. Given the set of basis feature vectors, $\mathcal{F} = \{\vec{f}_i\}$, the PARS TD signals, $s_j(t)$, can be expressed as a weighted sum of the feature vectors, $\vec{f}_i$, plus a residual term, $\vec{r}_j$:

$$s_j(t) = \sum_{\forall i} \alpha_{i,j} \vec{f}_i + \vec{r}_j. \tag{5.6}$$

Arranging the feature vectors as matrix columns, forming a matrix of features,

$$F = \begin{bmatrix} | & | & \\ \vec{f}_1 & \vec{f}_2 & ... \\ | & | & \end{bmatrix}, \tag{5.7}$$

the preceding expression then becomes

$$s_j(t) = F\vec{\alpha}_j + \vec{r}_j. \tag{5.8}$$

Generally, there are far fewer feature vectors than the dimension, n, of the PARS TD signals, allowing for dimensionality reduction.

To solve for the vector of feature weights, $\vec{\alpha}$, the pseudo-inverse [109–111] of F is used, such that the residual, $\|\vec{r}_j\| = \|s_j(t) - F\vec{\alpha}_j\|$, is minimized. Thus,

$$\vec{\alpha}_j = F^+ s_j(t). \tag{5.9}$$

Feature Learning: "K-Means-style" Clustering

Feature learning is achieved by clustering PARS TD signals based on their signal shape, and using the resulting selected cluster centroids, $\vec{c}_i$, as basis feature vectors. Because TD signals are treated as Cartesian vectors, the signal shape is then analogous to the vector *angle*. Any vector, $\vec{v}$, can be expressed as the product of a scalar magnitude, m, and a

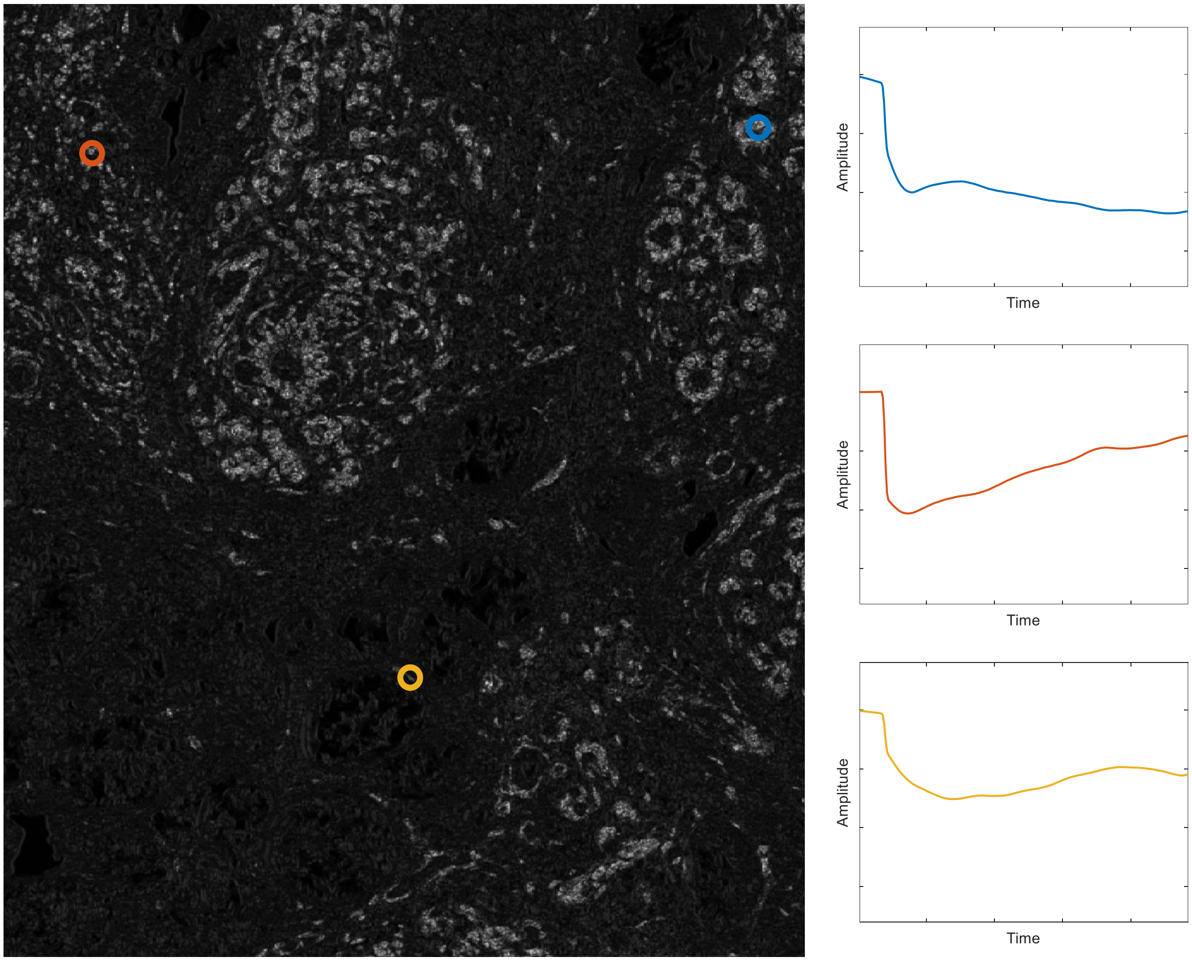


Figure 5.3: The left panel shows a standard TD signal projection PARS image of unstained human breast tissue on a slide, captured with 266 nm excitation and 637 nm detection wavelengths. Three points are marked in colour (blue, red, yellow), for which the corresponding filtered (for visual clarity) PARS TD signals are shown in the right panels. The blue point corresponds to a nuclear structure (absorption based on DNA) in a densely-packed region surrounding a duct, the red point corresponds to a nuclear structure in a lower-density region of the tissue, and the yellow point corresponds to part of an extra-cellular region, likely containing cytochromes. The shapes of the three TD signals are distinct, demonstrating that signal shapes vary across tissue components. Please note that this figure is purely shown to demonstrate the effect of signal shapes varying across tissue components and that the signals shown here are not necessarily representative of those that can be captured on more recent PARS system architectures. PARS image captured by Benjamin R. Ecclestone.

co-linear unit-vector, $\vec{u} = \vec{v}/\|\vec{v}\|$. Thus $\vec{v} = m\vec{u}$, and it is clear that the angle is encoded by the unit-vector, and signals can be clustered based on their angle.

Consider an arbitrary PARS TD signal, $\vec{s} = m\vec{u}$. The *negative* of this signal, $-\vec{s} = (-m)\vec{u}$, shares the same direction, $\vec{u}$, however the sign is opposite. These two signals are described as having opposite *polarity* (i.e., they are antipodal). Although the angle between $\vec{s}$ and $-\vec{s}$ is π radians, they share the exact same signal shape, $\vec{u}$, associated with the same underlying imaged target, and should be clustered together. Therefore, the clustering algorithm must be polarity-agnostic.

Distance Metric

Before defining the clustering algorithm, the distance metric to be used must first be defined. To cluster TD signals based on their shape (or associated vector angle), we require the distance metric to be

1. angle-based, and
2. polarity-agnostic.

Continuing with the example of antipodal signals $\vec{s}$ and $-\vec{s}$, a distance metric that is polarity-agnostic would measure $\vec{s}$ and $-\vec{s}$ as having zero distance, and would enable the two to be clustered together.

Formally, a distance metric [222] on a set $\mathcal{M}$ is a function, $d : \mathcal{M} \times \mathcal{M} \to [0, \infty)$, that for any $x, y, z \in \mathcal{M}$ satisfies the following three axioms:

1. $d(x, y) = 0 \iff x = y$ (identity of indiscernibles),
2. $d(x, y) = d(y, x)$ (symmetry), and
3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

The distance metric will be based on angle, so the angle between two vectors must be defined. For simplicity, and based on the symmetry axiom, all angles will be considered to be positive. The angle, ϑ, between two vectors, $\vec{v}_1$ and $\vec{v}_2$, is defined as

$$\vartheta := \angle\left(\vec{v}_1, \vec{v}_2\right) = \arccos\left(\frac{\langle \vec{v}_1, \vec{v}_2 \rangle}{\|\vec{v}_1\|\|\vec{v}_2\|}\right), \tag{5.10}$$

for $\arccos(\cdot)$ defined abstractly as $x \mapsto \arccos(x)$, $[-1, 1] \to [0, \pi]$.

The proposed distance metric is

$$d\left(\vec{v}_1, \vec{v}_2\right) = \sin(\vartheta), \tag{5.11}$$

for $\sin(\cdot)$ defined on the interval $[0, \pi]$, constraining the range of d to $[0, 1]$. Note that the function composition

$$\sin\left(\arccos(x)\right) = \sqrt{1 - x^2}, \tag{5.12}$$

an identity verifiable via the Pythagorean Theorem. Therefore, the distance metric may be simplified to

$$d(\vec{v}_1, \vec{v}_2) = \sqrt{1 - \left(\frac{\langle \vec{v}_1, \vec{v}_2 \rangle}{\|\vec{v}_1\|\|\vec{v}_2\|}\right)^2}. \tag{5.13}$$

Although the proposed metric does not satisfy the first axiom (identity of indiscernibles), since $d(\vec{v}, m\vec{v}) = 0$, thereby making d a *pseudo-metric* [223], it does satisfy the other two axioms and is still suitable for this application given that, by design, the clustering algorithm need only discern signals by shape (angle), not amplitude.

Algorithm

With the distance metric defined now, the clustering algorithm, detailed in Algorithm 1, follows fairly naturally from conventional K-Means.

The key difference from the conventional algorithm is in the *Centroid Update* step, shown on line 17 of Algorithm 1. The usual approach, in conventional K-Means, of taking the mean of all data points within a given cluster, S_i, to update the centroid, $\vec{c}_i$, does not work here due to polarity-agnosticism, whereby antipodal TD signals are clustered together; antipodal signals will largely cancel (negate each other) when averaged. Instead, the direction of greatest variance can be used as the cluster centroid. To contend again with polarity-agnosticism and the possibility that clusters are not balanced in polarity, when computing the direction of greatest variance, the union of the cluster, S_i, with its negative (i.e., set where all points are flipped across origin, by multiplying by -1) is used: $S_i^{\pm} = S_i \bigcup (-S_i)$. The direction of greatest variance is in fact the first principal component, and can be calculated via a singular value decomposition (SVD), $\vec{c}_i \leftarrow \mathrm{SVD}_1\left(S_i^{\pm}\right)$, shown on line 20. Following this, the centroid must be normalized, $\vec{c}_i \leftarrow \vec{c}_i / \|\vec{c}_i\|$, such that it falls on the unit-hypersphere.

The algorithm stops iterating when

1. sufficiently few data points change clusters between iterations, or

2. the difference in the mean residual is sufficiently small between iterations.

The user must specify criteria, *MovesCriterion* and *DifferenceCriterion*, based on performance requirements; however generally, setting $\mathit{MovesCriterion} = 1\%$ of the total data points and $\mathit{DifferenceCriterion} = 10^{-10}$ seems to perform adequately.

5.2.2 Results

In this section, the proposed algorithm is first validated using synthetically generated data in low dimensions for ease of interpretation before its robustness to several key parameters is evaluated, again on synthetically generated data. Finally, the algorithm is used to perform feature extraction on experimentally obtained PARS images of human breast tissue slides and freshly resected tissue murine brain tissue.

Results on Synthetic Data

Synthetic data are generated in both 2D and 3D to validate that the algorithm is performing as desired. Only two *true* classes are generated, with a third class comprised solely of low-intensity zero-mean noise. The low-intensity noise is included because tolerance to background noise is one of the main goals of this algorithm. All clusters contained variation, defined by some covariance matrix, to simulate noise that would normally be present in measured signals. In both the 2D and 3D cases, the first class is *mono-polar*, meaning that all data points are of the same polarity, or in other words, are on the same side of the origin. The second class however, is *bi-polar*, meaning that there are data points on both sides of the origin. This was done intentionally to test the polarity-agnosticism of the algorithm.

Classes are generated such that signals have non-zero amplitude and the classes are separated angularly. In 2D, the classes are centered (i.e., position of mean value(s)) arbitrarily at $[4, 4]^T$ and $\pm[5.5, 0]^T$, respectively. Similarly, in 3D, the classes are centered at $[4, 4, 4]^T$ and $\pm[4, -4, -4]^T$, respectively. These central locations are indicative of the true *directions* of the classes that should be learned by the algorithm. In 2D, the clusters are separated by an angle of $45°$, and in 3D, by $70°$.

Figure 5.4 shows the results of this validation study. The results from the 2D analysis (top row) are more easily visualized than those in 3D (bottom row), which is slightly more complicated yet still reasonable to visualize. In both cases, the algorithm does an excellent job of separating the input classes, even in the presence of noise. Additionally, it can be

Algorithm 1: K-Means-style Clustering Algorithm for PARS

```
Input   : Set of PARS TD signals, $S = \{s_j(t)\}$, acting as data points to be
          clustered.
          Number of desired clusters, $K$.
          Minimum allowable moves criterion, $MovesCriterion$.
          Difference in mean residual criterion, $DifferenceCriterion$.
Output: Set of cluster labels, $\mathscr{L} = \{\ell\}$, associated with each TD signal.
          Set of cluster centroids, $\mathscr{C} = \{\vec{c}_i\}$, for $i = 1, ..., K$.

Initialization:
Randomly select $K$ data points as initial centroids.
for $i = 1, ..., K$ do
    $\vec{c}_i \overset{Random}{\leftarrow} s(t) \in S$
end
Set previous value of mean residual to 0.
$\mu_r^{prev} \leftarrow 0$

Main loop:
repeat
    Set number of changed cluster labels to 0.
    $moves \leftarrow 0$

    Membership Update:  Determine cluster membership by finding nearest centroid
    to each point.
    foreach $s_j(t) \in S$ do
        $\ell_j \leftarrow \arg\min_{i\in\{1,...,K\}}\{d\left(s_j(t), \vec{c}_i\right)\}$
        Increment $moves$ if cluster membership changes.
        if $\ell_j$ changed this iteration then
            $moves \leftarrow moves + 1$
        end
    end

    Evaluate mean residual (objective).
    $\mu_r \leftarrow \frac{1}{\|S\|} \sum_{s_j(t)\in S} d\left(s_j(t), \vec{c}_{\ell_j}\right)$
    $\Delta\mu_r \leftarrow \mu_r - \mu_r^{prev}$
    $\mu_r^{prev} \leftarrow \mu_r$

Continued on next page...
```

Algorithm 1: K-Means-style Clustering Algorithm for PARS (Continued)

```
Resumed from previous page...
    Centroid Update: Use data points within clusters to update the centroids.
    for $i = 1, ..., K$ do
        Get set of data points within cluster.
        $S_i \leftarrow \left\{ s_j(t) \,\middle|\, \ell_j = i \right\}$
        Take union of set with its negative.
        $S_i^{\pm} \leftarrow S_i \bigcup (-S_i)$
        Compute first principal component via SVD. Assign to centroid.
        $\vec{c}_i \leftarrow \mathrm{SVD}_1\left(S_i^{\pm}\right)$
        Normalize centroid to fall on unit-hypersphere
        $\vec{c}_i \leftarrow \vec{c}_i / \|\vec{c}_i\|$
    end
until $\Delta\mu_r \leq DifferenceCriterion$ OR $moves \leq MovesCriterion$
```

seen that suitable cluster centroids, $\vec{c}_i$, are learned (observable plotted in a similar form as TD signals would be, in the right column). The learned centroids are approximately identical to the true cluster centers / directions, up to proportionality, given that the learned centroids are normalized.

Next, the performance of the algorithm is evaluated over a range of key parameter settings. The goal of this analysis is to observe the capabilities of the algorithm, measured in terms of clustering accuracy and centroid similarity to ground truth, in both favourable and adverse conditions. The three parameters tested here are

1. the angular separation of classes,
2. the background noise level, and
3. the fraction of the data set comprised of background noise.

Each parameter is varied independently, while holding the others constant at their default value for this experiment. The default values used are an angular class separation of 35°, a background noise level of 2/3 (relative to the nominal mean signal level), and a fraction of background noise points of 0.5. This experiment is run in 30-dimensional space, more closely mimicking the high-dimensionality of PARS TD signals, with randomly generated synthetic data composed of *two* classes plus background noise, similar to what is shown in Figure 5.4. Clustering accuracy is defined as the percent of correctly clustered data points (background noise is ignored). Centroid similarity is measured by cosine similarity to the ground truth class prototypes.

Figure 5.5 shows the results of this analysis. For each of the tested parameters (one per row), the clustering accuracy and centroid similarity are shown side-by-side. The results match intuition, in that the algorithm performs best at high class separation angles, low noise levels, and for low fractions of background noise. Performance tapers-off for the opposite cases. Notice that for small separation angles, the clustering accuracy suffers, likely as a result of class overlap. In spite of this, the centroid similarity remains high. Both clustering accuracy and centroid similarity suffer as the background noise exceeds that of the signals. This corresponds to a situation where signals are completely "buried" in noise, existing below the noise floor. Lastly, as the fraction of background noise begins to dominate, both metrics suffer as noise increasingly influences the learned clusters. Notice that in many cases, even when clustering accuracy (left column) is poor, the centroid similarity (right column) remains fairly high. Given that this algorithm was designed for use in *feature learning*, rather than for producing clustered points, these results are promising and indicate that the algorithm performs well, achieving its design objective. Though it is not shown in this analysis, for more favourable parameter default settings, the algorithm becomes substantially more robust to adverse conditions in any individual parameter.

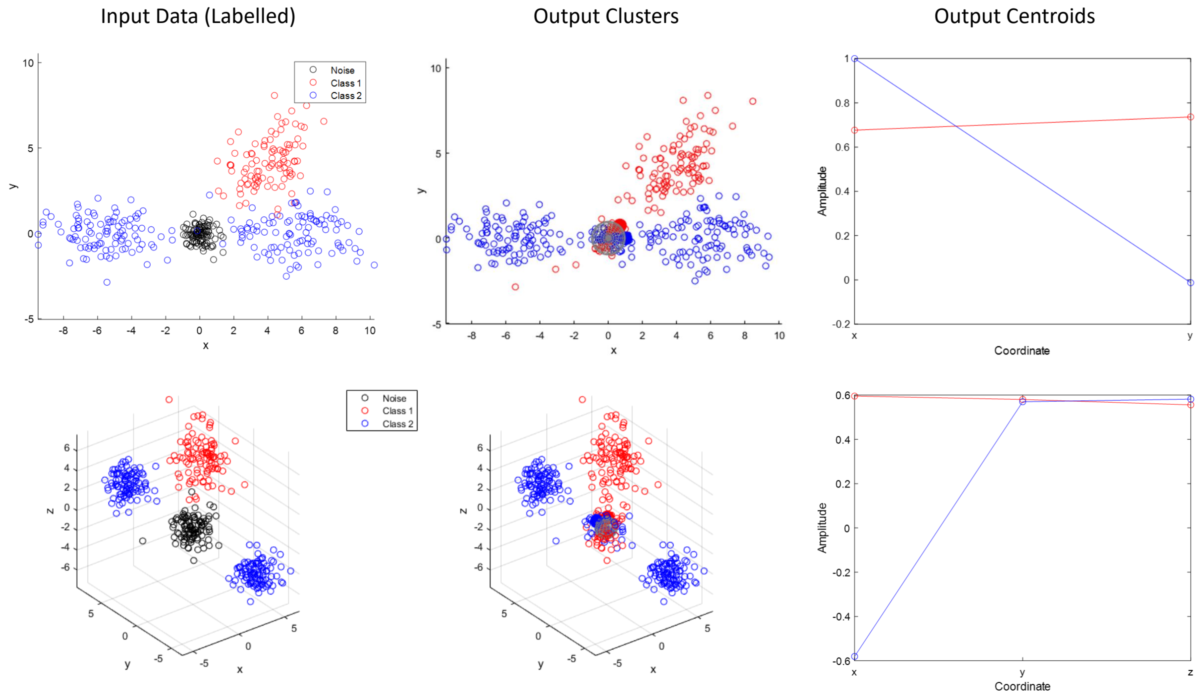
With the performance of the algorithm validated on synthetic data, it is now appropriate to test its performance on real PARS data to perform feature extraction.

Results on Real PARS Data

The use of the clustering algorithm for performing feature extraction is demonstrated here on both an unstained human breast tissue slide and on freshly resected murine brain tissue.

The standard projection PARS image of the unstained, formalin-fixed paraffin-embedded (FFPE) human breast tissue slide is shown in Figure 5.6. This image, again, was captured and provided by Benjamin R. Ecclestone with gratitude from the author. 266 nm excitation and 637 nm detection wavelengths were used to capture this image. A boxed-in region in Figure 5.6 indicates a selection of TD signals used for feature learning. Note that an input of $K = 6$ was used arbitrarily here. The task of selecting the number of clusters, K, is not addressed here in this thesis, but rather is assumed to be known based on external information.

The learned features, or centroids, $\vec{c}_i$, are shown in Figure 5.7, sorted by number of data points within the given cluster. A wide variety of signal shapes are present, each representing a component of the data. Figure 5.8 shows extracted feature amplitudes, α_j, as images for each of the six learned features. The amplitude and polarity (arbitrary,


Input Data (Labelled)

Output Clusters

Output Centroids

Figure 5.4: The clustering algorithm is validated in both 2D (top row) and 3D (bottom row). Synthetic data are generated and shown (left column) with true class labels (red and blue). Additionally, low-intensity zero-mean noise (black) is included to test the clustering algorithm's abilities. The synthetic data are clustered (central column) and shown with cluster labels (red and blue). The unit-circle / -sphere (grey) is shown at the origin. Cluster centroids, $\vec{c}_i$, are shown as large dots on the unit-circle / -sphere, indicating their associated cluster direction. The cluster centroids are shown as a plot (right column), over their two / three dimensions (x, y) / (x, y, z). Although the dimensionality of this data, and therefore the centroids, is very low, one can imagine that for high-dimensional time-domain data, the centroids plotted in this manner would take the form of representative signal shapes.

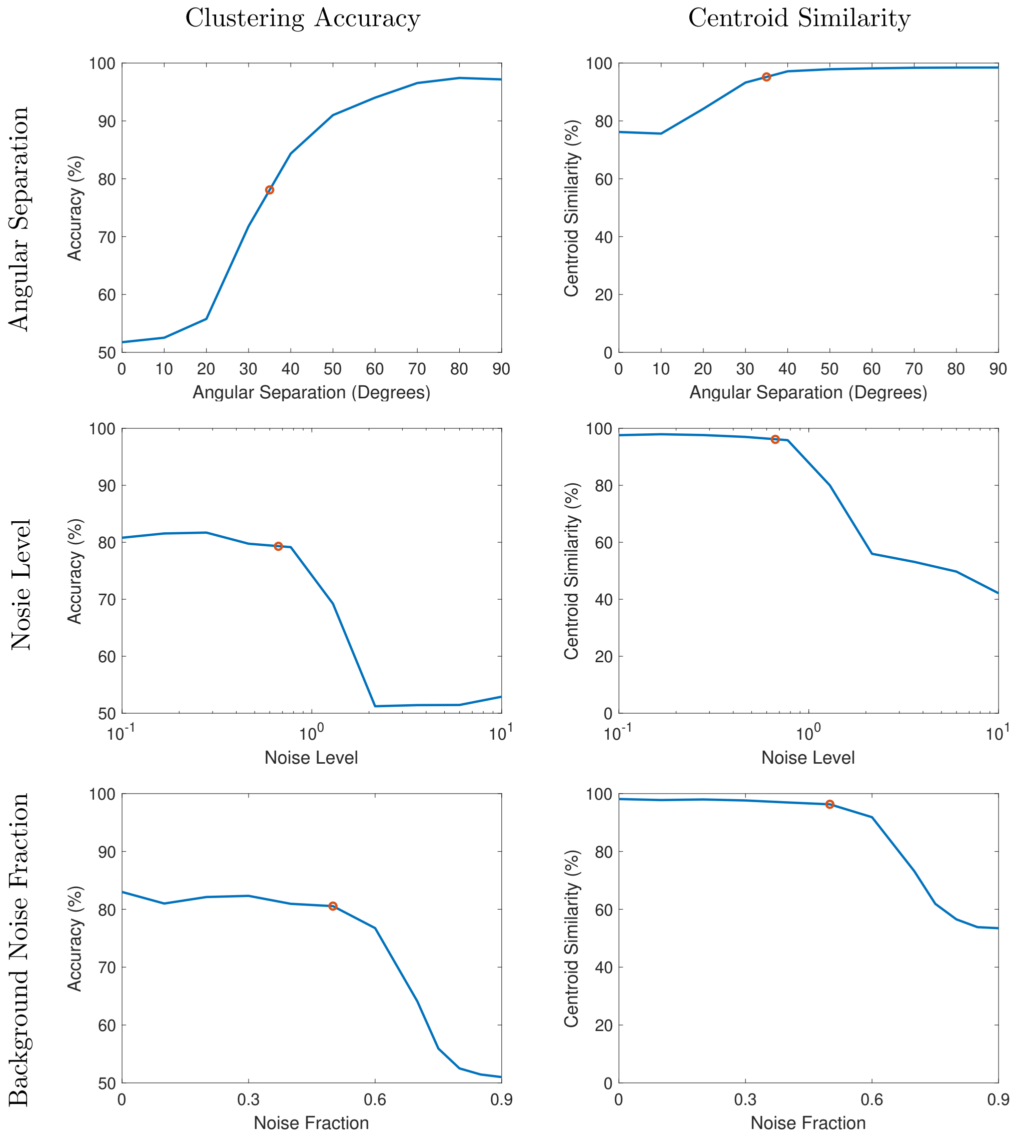


Figure 5.5: The clustering algorithm is evaluated over ranges of class separation, background noise level, and the fraction of background noise present, creating gradients between favourable and adverse conditions. The default parameter settings are marked with orange circles in each plot. Observe that the algorithm performs best at high class separation angles, low noise levels, and for low fractions of background noise. This matches intuition in that class overlap at low separation angles, high noise levels, and high quantities of noise would all negatively influence the algorithm's ability to cluster.

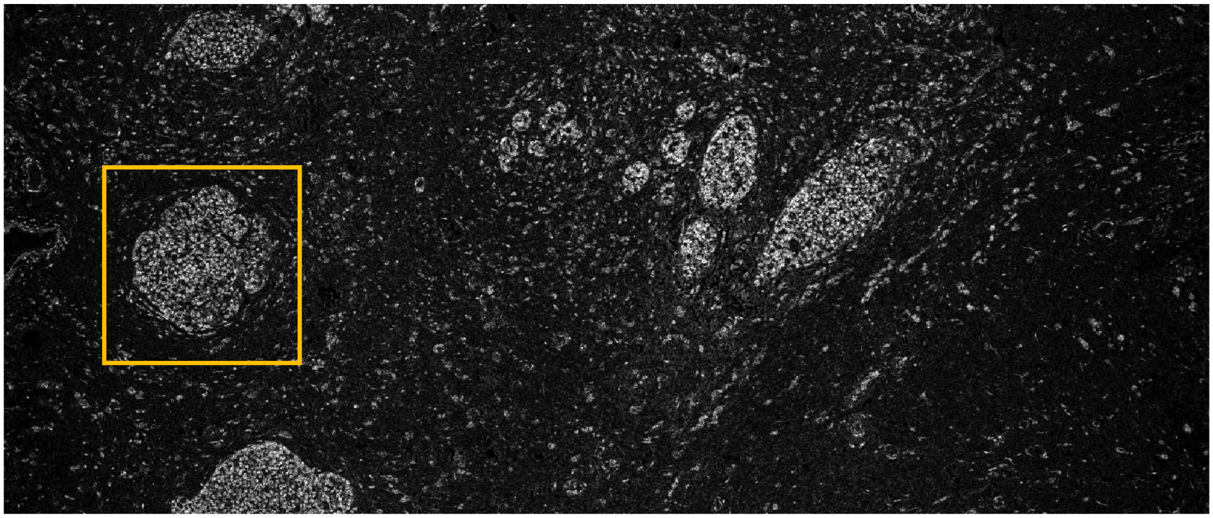


Figure 5.6: A standard TD signal projection PARS image of unstained human breast tissue on a slide, captured with 266 nm excitation and 637 nm detection wavelengths. Signals from the boxed-in region are used for feature learning. Image captured by Benjamin R. Ecclestone. This image was created for this thesis but also submitted as part of article [1].

given the polarity agnosticism of the clustering algorithm, based upon the learned feature polarity) are shown using colour (positive in red vs. negative in green). It is clear that each feature extracts information associated with specific tissue structures. Feature 1 is clearly the most striking of the learned features, strongly highlighting nuclear contrast. Features 3 and 4 also seem to extract some nuclear contrast; however, not universally throughout the image. Instead, the cluster of cell nuclei where the algorithm was trained, shown in Figure 5.6, are not highlighted, whereas the clusters on the right of the image are. The other three features, 2, 5, and 6, seem to mainly be sensitive to the extracellular matrix or cytochromes — largely being insensitive to cell nuclei.

Additionally, the polarity of the feature amplitudes is not at all random, but rather is "patchy" and likely correlates with additional properties of the tissue, perhaps information related to the *depth* of the absorbing chromophore relative to the focal plane of the PARS microscope. Additional notes regarding signal polarity are provided in the Future Work section (Section 6.1).

Finally, by combining extracted feature amplitudes (absolute value, thus ignoring the effects of polarity), aggregate information is made visible. Figure 5.9 shows a combination of features via assigning each to an RGB colour-channel. Feature 1 maps to red, feature

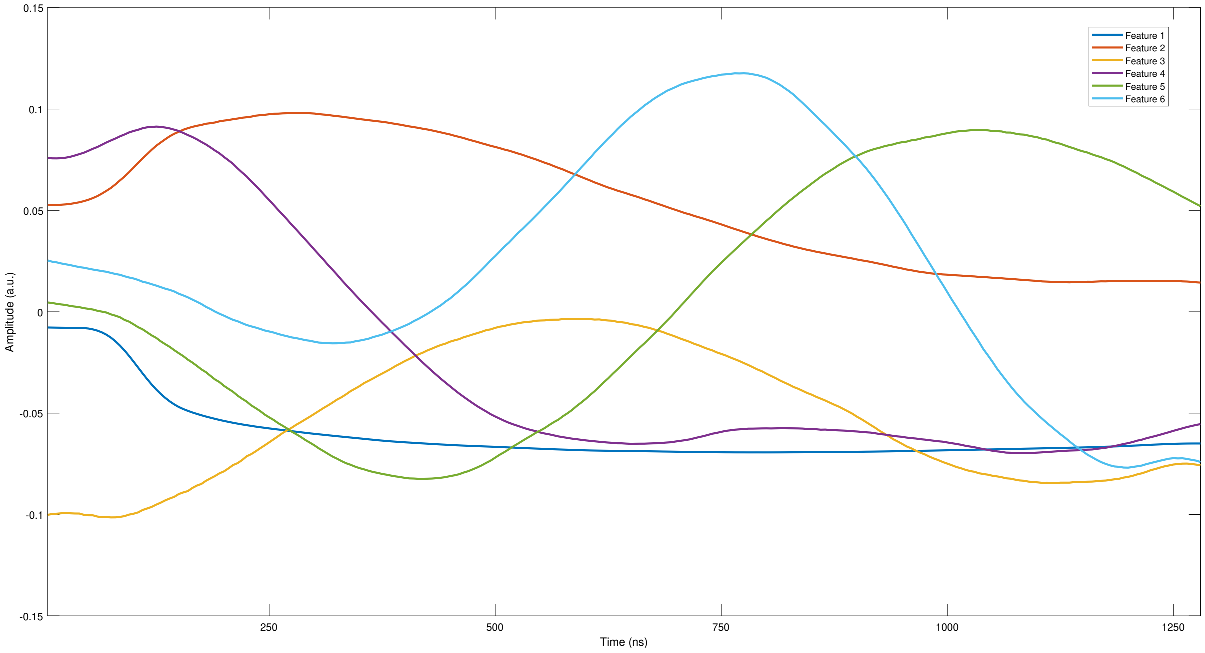


Figure 5.7: The cluster centroids learned, for $K = 6$, from the human breast tissue slide in Figure 5.6. A wide variety of signal shapes are learned and presented here.

4 maps to green, and feature 5 maps to blue. These features were selected based on the variation in contrast observed between them in Figure 5.8. Generally, it can be seen that a combination of features 1 and 4 (red and green combined to form a yellow / orange colour) yield nuclear contrast. Feature 5 (blue), tends to be most sensitive to extracellular contrast. Based on comments from clinicians, feature 1 on its own (i.e., in places where feature 4 (green) is weak), may be indicative of the presence of cytokines. This is seen in the dark-red region, center-right, just below and surrounding the regions of densely packed nuclei.

Next, an image of freshly resected murine (mouse) brain tissue is examined. This image, again, was captured and provided by Benjamin R. Ecclestone with gratitude from the author. 266 nm excitation and 405 nm detection wavelengths were used. No additional processing was done to this tissue prior to imaging. Figure 5.10 shows the standard projection of this PARS image for reference. In a similar vein as with the brain tissue, the clustering algorithm is run on this image to learn centroids, $\vec{c}_i$, before feature extraction is used to extract associated feature amplitudes, α_j. Finally, the extracted feature amplitudes are combined as an RGB image.

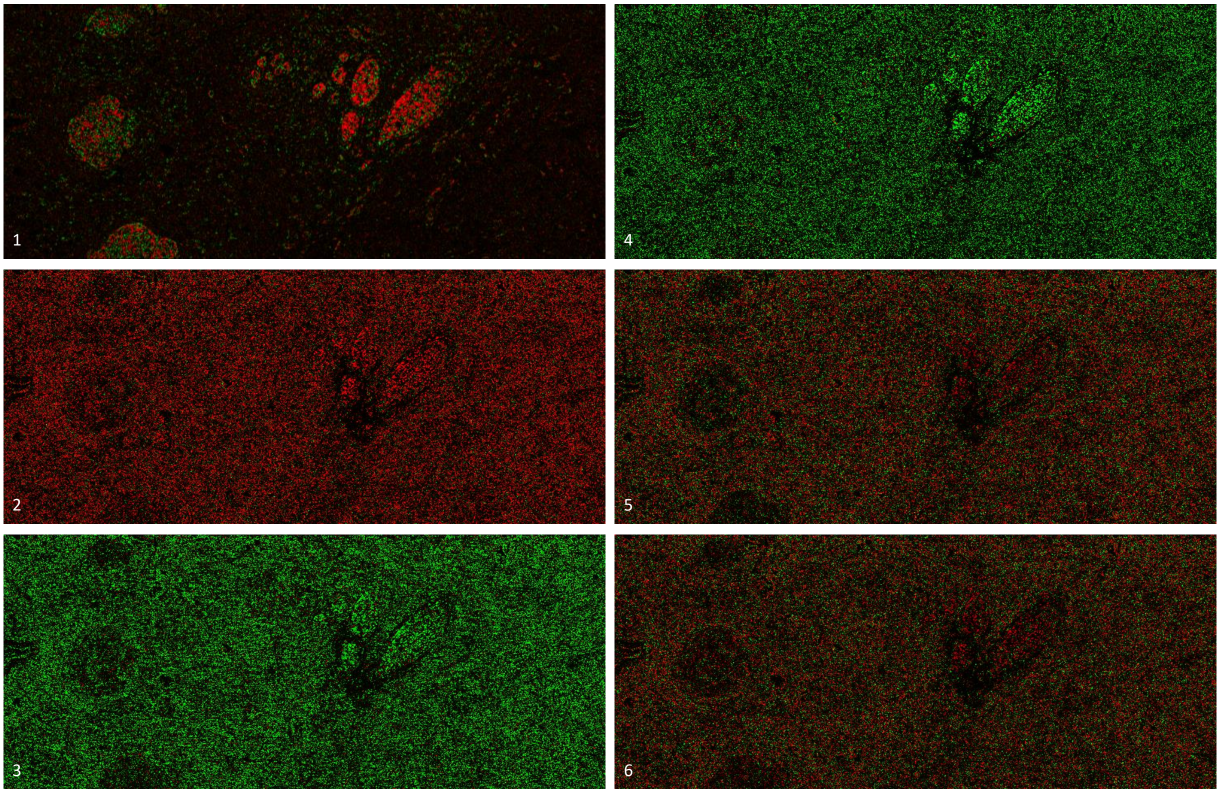


Figure 5.8: Feature images of the human breast tissue slide of Figure 5.6 for each of the associated centroids from Figure 5.7 are shown with signal polarity indicated by colour (positive in red vs. negative in green). It is clear that each feature extracts information related to specific structures in the tissue.

Feature learning was performed over the entire image, with $K = 3$ clusters specified. Figure 5.11 shows the learned centroids, $\vec{c}_i$. The shapes tend to have a sharp initial rise and then a comparatively longer decay at varying rates. In this sample, the learned feature shapes are less diverse than was seen in Figure 5.7; however, this may be partially due to the fact that fewer cluster centroids were learned. Figure 5.12 shows sets of learned centroids for each of $K = 2 \ldots 5$ requested clusters. Indeed, for increasing K, the learned centroids become more diverse and complex, rather than simply "adding" a new centroid to the set learned for lower K. To be clear, in this thesis, no assumptions are made regarding what the *correct* number of clusters, K, for a given PARS image is. The correct number of clusters is unknown and therefore selecting K remains a subject of future work. Prior

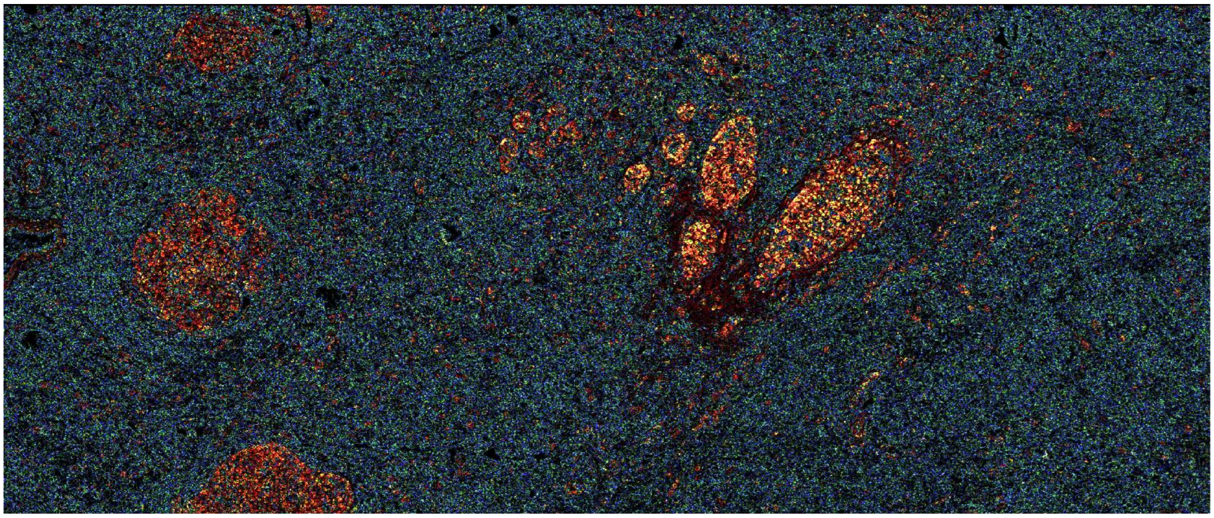


Figure 5.9: Three of the human breast tissue slide feature images from Figure 5.8 are combined (via taking the absolute value and assigning each an RGB colour-channel). Feature 1 maps to red, feature 4 maps to green, and feature 5 maps to blue. Interesting structures are clearly visible in the tissue based on colour. Although feature meanings may be rather uncertain, this result represents an initial step towards virtual staining through PARS. This image was created for this thesis but also submitted as part of article [1].

knowledge of the components within imaged tissues may prove to useful in selecting K; however, detailed studies would be required to test this and remain to be done.

Finally, without first plotting individual extracted feature amplitude images, the extracted amplitudes, α_j, are directly combined in Figure 5.13 by assigning each feature to one of the RGB colour-channels. Here, feature 1 maps to red, feature 2 maps to green, and feature 3 maps to blue. A strong correspondence between colour (i.e., extracted feature) and tissue structure is seen here. Structures visually similar to the soma (cell body structure of neuron containing the nucleus) and axons (fibrous, myelin covered component of neuron responsible for transmitting impulses) [224] are present and coloured in this image.

To observe more detail, a sub-section of Figure 5.13 is shown in Figure 5.14. The structures resembling the soma and axons of neurons are more clearly visible here. The structures and contrast separation seen here resembles that of labeled multi-fluorescence microscopy [225]. If similar levels of specificity are available in PARS via the method presented here, this would represent an incredible advance in the capabilities of label-free PARS microscopy.

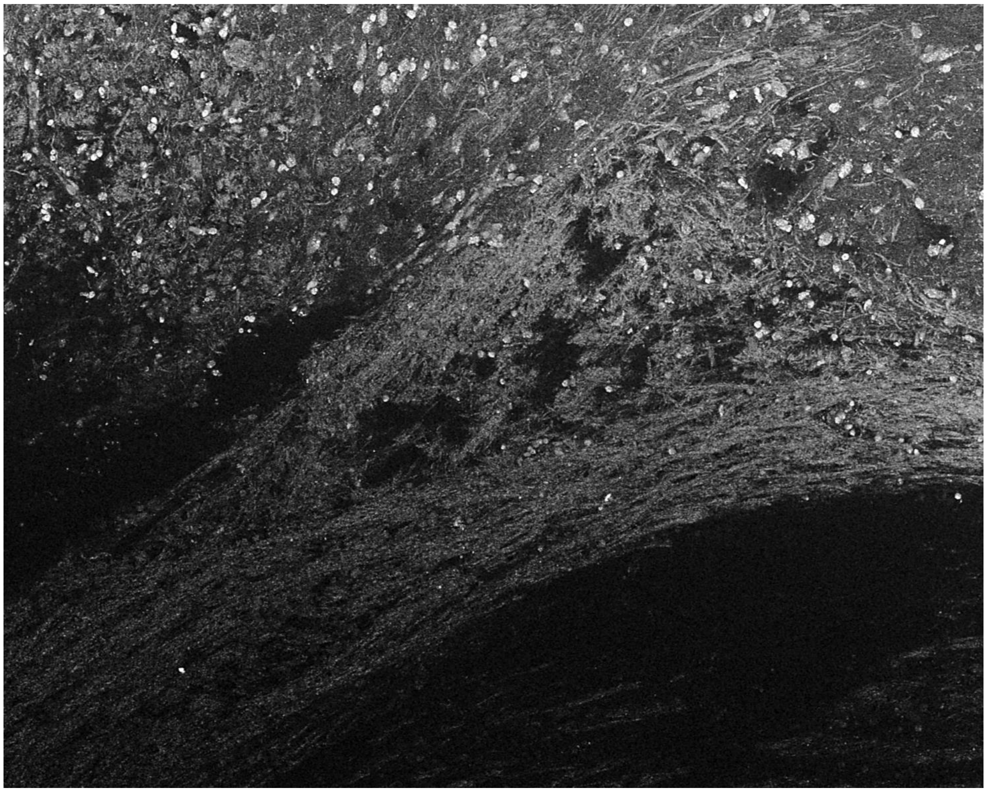


Figure 5.10: A standard TD signal projection PARS image of freshly resected, unstained, murine (mouse) brain tissue. Image captured by Benjamin R. Ecclestone. This image was created for this thesis but also submitted as part of article [2].

5.3 Chapter Conclusion

To conclude this chapter, several methods starting from fairly simple linear filtering via a matched filter and ending with a novel feature learning / extraction scheme based on a modified K-means clustering algorithm were developed and explored. PARS imaging contrast was improved, allowing finer structures to be discerned and enabling the possibility of using lower fluences when imaging, making the modality safer for clinical use. The effectiveness of matched filtering was limited because only a *single* characteristic signal shape was used to extract amplitude information. This limitation was overcome by developing a modified K-means algorithm to learn characteristic signal shapes as cluster centroids. The algorithm was designed such that it was polarity-agnostic and insensitive to low-level noise. This method was first validated on synthetically generated data in low dimensions

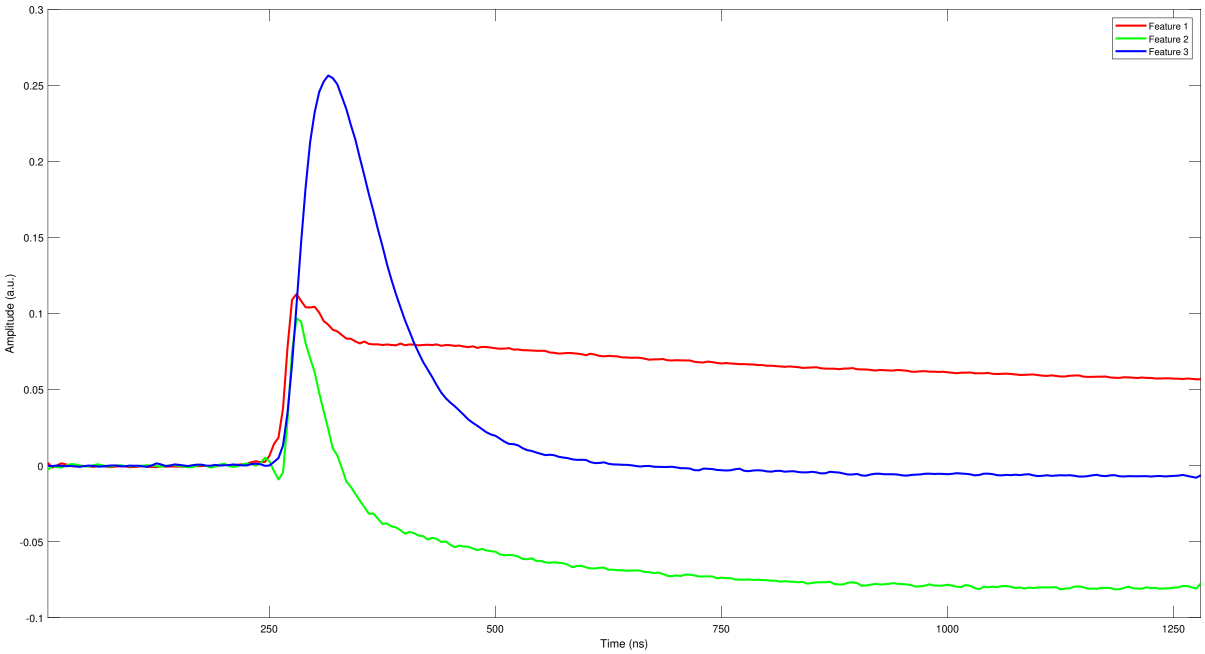


Figure 5.11: The cluster centroids learned, for $K = 3$, from the freshly resected murine brain tissue in Figure 5.10. Three very distinct signal shapes are found. Note that the colouring scheme (red, green, and blue) is carried over to Figure 5.13, where the three extracted feature images are combined into a single RGB image.

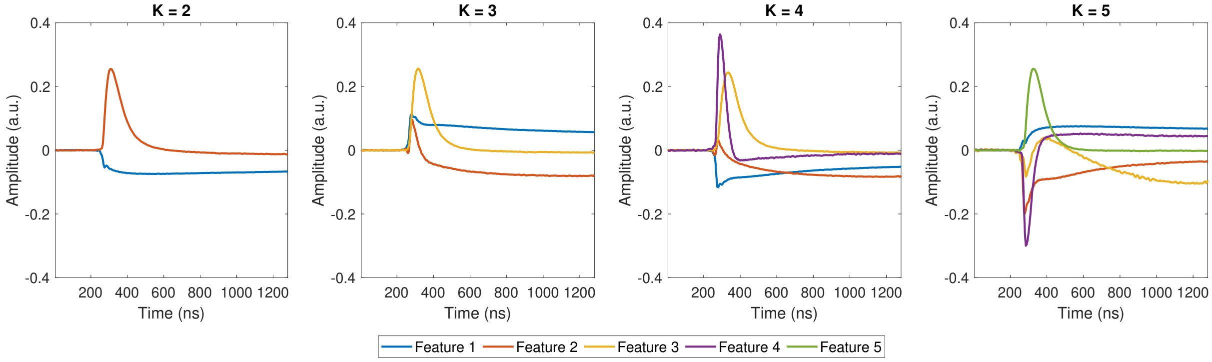


Figure 5.12: With increasing numbers of requested clusters, K, it is seen that the learned centroids become more specific and complex, rather than simply "adding" a new centroid to the set learned for lower K. Shown here are cluster centroids learned, for $K = 2 \ldots 5$, from the freshly resected murine brain tissue in Figure 5.10.

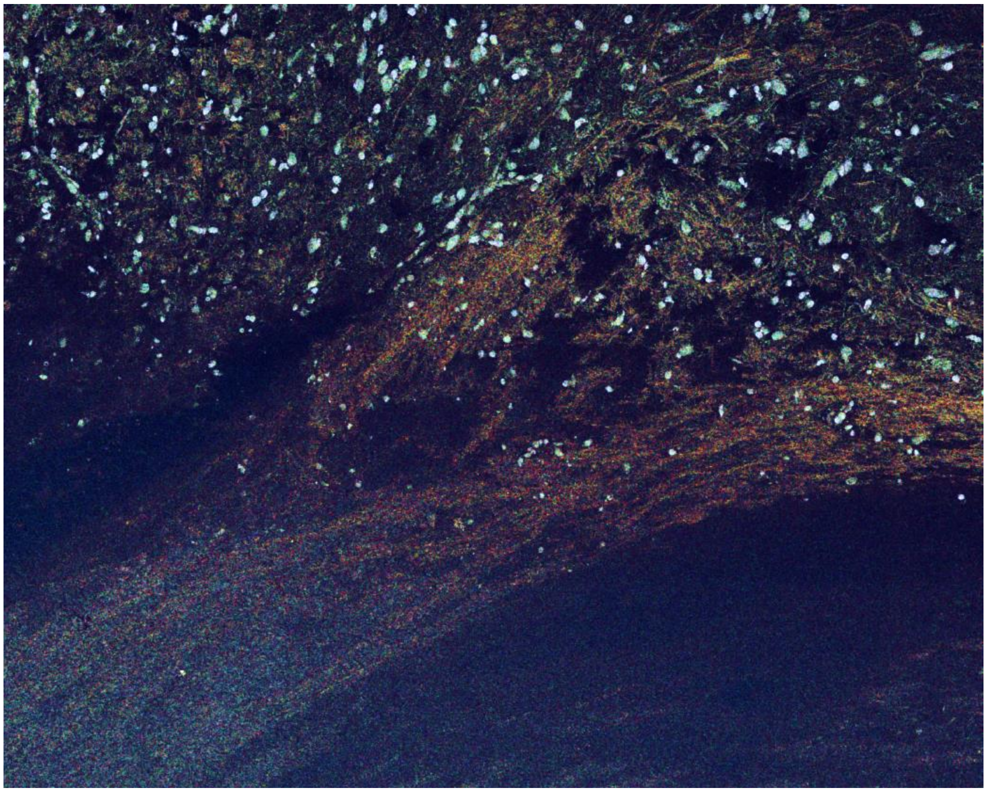


Figure 5.13: Three feature images associated with centroids shown in Figure 5.11, learned from the freshly resected murine brain tissue of Figure 5.10, are combined (via taking the absolute value and assigning each an RGB colour-channel, consistent with the colouring scheme used in Figure 5.11). A strong correspondence between colour (i.e., extracted feature) and tissue structure is seen here. Structures visually similar to the soma (globular components) and axons (fibrous components) are present and coloured in this image. This image was created for this thesis but also submitted as part of article [2].

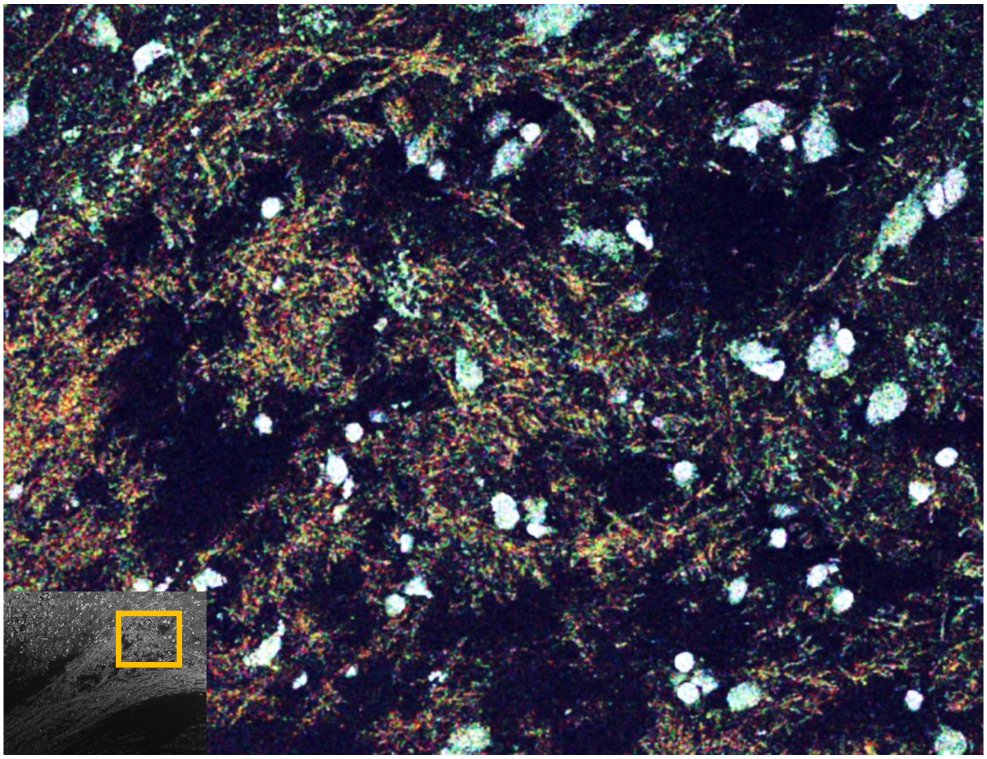


Figure 5.14: A detailed view of the same combined feature image as in Figure 5.13. The zoomed-in region is indicated in the lower left. A clearer view of structures resembling the soma and axons of neurons is presented here.

where its performance was easily understood, before being applied to PARS images of an unstained human breast tissue slide and freshly resected murine brain tissue. In both cases, a set of characteristic signal shapes were learned, yielding intriguing extracted feature amplitudes, visualised as false-coloured images. These images revealed correlations between the extracted features and clinically relevant tissue structures such as cell nuclei, the extracellular matrix, and various components of neurons. The contrast derived here, all without the use of exogenous dyes or labels, and captured with only one excitation wavelength, represents an important step forward in the field of PARS microscopy.

Chapter 6

Conclusion

The work presented here represents developments towards achieving chromophore specificity in Photoacoustic Remote Sensing. The key contributions surrounded formulating the inverse problem of chromophore unmixing (Chapter 3), studying and characterizing the non-linear optical phenomenon of Stimulated Raman Scattering in optical fiber for use in a multi-wavelength light source for PARS microscopy (Chapter 4), and finally developing a clustering algorithm capable of learning characteristic PARS time-domain signal shapes for later use in feature extraction enabling the representation of TD signals by the contribution of individual component shapes / features (Chapter 5).

In Chapter 3, several developments towards chromophore unmixing in PARS via multi-wavelength imaging were made. The question of how best to select excitation wavelengths was addressed, a solution to the unmixing problem was formulated with additional constraints and considerations for PARS were specifically included, and finally *in-vivo* blood oxygen saturation estimation was undertaken in PARS images of a rat retina.

It was concluded that the most suitable method for selecting excitation wavelength for unmixing specific targets was based on minimizing the magnitude of expected estimation error. In principle, this would make the estimates as accurate as possible, assuming no other unknown sources of error are present. The blood oxygen saturation estimates presented in Figure 3.14 constitute the first ever report of non-contact *in-vivo* photoacoustic measurements of oxygen saturation in retina, representing a significant step towards the clinical use of PARS as a diagnostic tool for many ophthalmic diseases.

In Chapter 4, a comprehensive study was performed to understand the impacts that several fiber and laser parameters have on multi-wavelength light generation through SRS in single-mode optical fiber. Additionally, a simulation of the SRS effect in optical fiber was developed and used to aid in understanding SRS from a theoretical standpoint and to create correspondence with the experimentally measured results.

Regarding the effects of the explored parameters, it was found that longer fiber lengths, greater input intensities, and the use of highly polarization-maintaining fibers enable high-order Stokes wave generation. The temporal pulse width does not affect the SRS process throughout the nano-second range; however, in the pico-second range, pulse walk-off and higher-order optical effects severely limit SRS. Lastly, pulse repetition rate was found not to affect the SRS process. Simulated spectra showed excellent correspondence to experimentally measured spectra and also showcased the central erosion effect of the temporal pulse profile, whereby only the tails of pulses remain after SRS conversion and are separated temporally. The effects of using centrally eroded pulses in PARS remain unknown. The study performed here allows better informed decisions to be made regarding parameter selection for the creation of multi-wavelength light sources for applications of PARS unmixing.

Chapter 5 focused on extracting meaningful information from PARS TD signals. Matched filtering was explored initially, before developing a clustering algorithm to learn *multiple* characteristic signal shapes as cluster centroids. The algorithm was designed such that it was scalable and robust to low-level noise. This method was applied to PARS images of an unstained human breast tissue slide and freshly resected murine brain tissue.

The matched filter proved effective in improving PARS image contrast, allowing finer structures to be discerned and enabling the possibility of using lower fluences when imaging, which would make PARS safer for clinical use. The feature learning and extraction method, when applied to tissue samples, resulted in colourized images revealing correlations between the extracted features and clinically relevant tissue structures such as cell nuclei, the extracellular matrix, and various cellular components of neurons. The contrast in these images was derived without the use of exogenous dyes or labels, and was captured with only one excitation wavelength. This development represents an important step forward in label-free PARS microscopy.

In conclusion, the three main contributions surrounding chromophore unmixing, multi-wavelength light generation, and time-domain feature extraction serve to advance PARS as a biomedical imaging modality. The research undertaken in this thesis enabled the first *in-vivo* non-contact photoacoustic measurement of blood oxygen saturation in an eye. Furthermore, the ability to extract additional information from the time-domain content of PARS signals dramatically changes the way PARS signals can be analysed, giving direct access to information related to the components of the underlying target, and offering a complementary alternative to the more conventional multi-wavelength imaging approaches for discerning targets.

The contributions of this thesis lead to substantial future developments to PARS microscopy, as outlined below in the Future Work section.

6.1 Future Work

Although much technical ground is covered in this thesis, there remain many experiments and research areas that should be delved more deeply into in order to gain a better understanding and to properly make technical use of the research presented in this thesis. A brief summary is listed here:

- Section 6.1.1 outlines next steps towards accurate chromophore unmixing in PARS,
- Section 6.1.2 discusses further exploration surrounding the use of SRS, and
- Section 6.1.3 discusses future directions related to PARS TD signal feature extraction.

6.1.1 Accurate Chromophore Unmixing in PARS

Paramount in being able to achieve accuracy in chromophore unmixing in PARS, is having *consistency* in measurements. Although much formulation was done in Section 3.1 of Chapter 3, regarding analysing which combination of excitation wavelengths would optimally minimize the estimation error covariance and deriving an inverse model, any inverse problem is *impossible* to solve (or make any reasonable estimate of a solution) if the forward process is not consistent. Note that the lack of consistency means that the forward process is in fact not well understood! There are effects at play that the forward model does not take into account that result in inconsistent measurements. The forward model (estimating the forward process) does not match the true forward process, causing the accuracy of the solution to subsequently suffer. Thus, it would be incredibly worthwhile to pursue improving the consistency of measurements.

The main suspected sources of inconsistency come as a result of instrument focus issues due to chromatic aberrations at the objective lens, and temporal variations in the excitation and detection lasers; however, there may be other effects at play that are not yet known.

Chromatic Aberration

Refractive optics, e.g., lenses, inherently result in chromatic aberration. This is because refractive index is *dependant* on wavelength, and thus the focus of a lens varies by wavelength as well. Although achromatic lenses (designed for identical focus at two wavelengths) and apochromatic lenses (identical focus at three wavelengths) exist[1], for applications where high magnification is desired — which is often the case in PARS microscopy — high numerical aperture lenses are used, which limit the depth of focus, resulting in significant variation in focus by wavelength. Furthermore, in the application of retinal imaging, where the lens of the eye is used for focusing, there is *no* ability to directly control the chromatic aberration.

Avoiding chromatic aberration, at least in cases where the focusing optic is selected by the operator / creator of the PARS microscope, can be achieved through the use of *reflective* focusing optics, as was done in article [226] in a PAM system. Reflective objective lenses do not cause chromatic aberration since the path of the light does not depend on wavelength. A critical downside of the available reflective objective lenses, such as the LMM series, available from Thorlabs Inc. [227], is a central obstruction inherent to the design, which negatively affects the focal spot. Two reflecting surfaces are used in these lenses: a primary annular concave mirror which faces towards the target, and a secondary convex mirror that is centered and faces the internally towards the microscope. The central secondary mirror is held in place by three curved spider vanes and causes poor (highly non-Gaussian) beam quality and negatively affects the focal spot size and shape. Figure 6.1 shows a diagram of the lens, a view of the secondary mirror and its supporting vanes, and a through-focus spot diagram which illustrates the effects that the obstruction has on the focal spot for minor deviations in focus. Thus, one must balance the benefits of being free of chromatic aberration with the drawbacks of the poor focal spot quality of this type of lens. Single-element reflective focusing mirrors are also available and may offer improved focus spot quality compared to the reflective objective lens design discussed here, while suffering less from undesired diffractive effects.

Temporal Variation

Pulse-to-pulse variation in the excitation laser may take the form of differences in pulse energy or polarization state, both of which *strongly* affect multi-wavelength light generation through SRS. Similarly, the continuous-wave detection laser also exhibits similar temporal

---

[1] There is no consensus on a precise definition of the terms *achromatic* and *apochromatic*. The point made here is that there are lens designs which *mitigate* chromatic aberration.

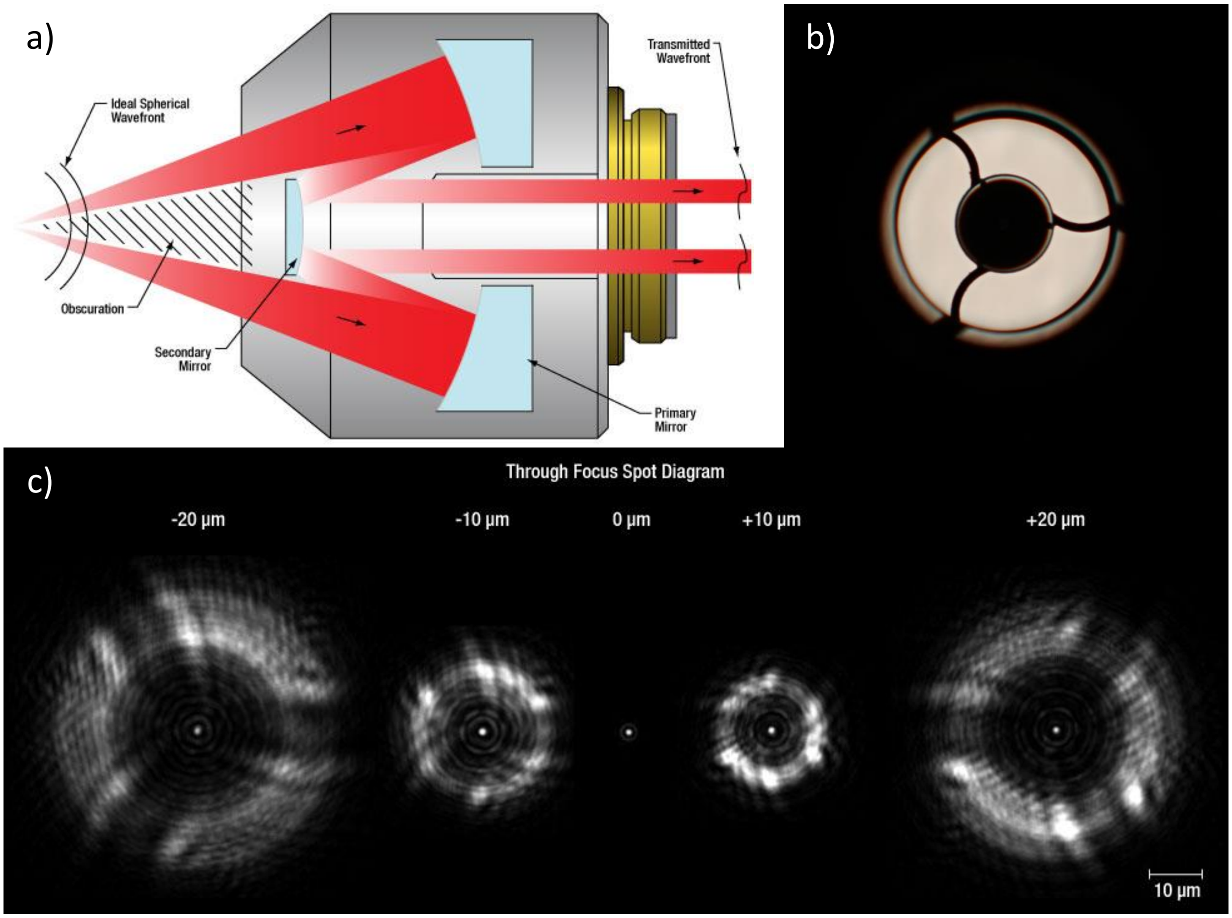


Figure 6.1: A diagram of the reflective objective lens architecture available from Thorlabs Inc. is shown in panel a). An incoming ideal spherical wave-front (from the left) first reflects off of the annular primary concave mirror before reflecting off of the central secondary convex mirror. The central part of the spherical wave-front is obscured by the secondary mirror and the vanes that hold it. Panel b) shows a view of the secondary mirror and the three curved spider vanes holding it. Panel c) is a through-focus spot diagram, illustrating the obscuration effect near the focal plane caused by the secondary mirror and the vanes holding it. It is clear that for even minor deviations from the focus (just micrometers) severe diffractive effects occur, leading to a highly non-Gaussian spot. Images copied from Thorlabs Inc. Reflective Microscope Objectives product web-page [227].

variations resulting in measurement inconsistency. Beyond impacts on SRS, fluctuations in the polarization state of both excitation and detection lasers ultimately result in intensity variations in the light reaching the target. This effect arises as a result of angled reflections along the beam paths, whereby parallelly polarized (referred to as *p* state) vs. perpendicularly polarized (referred to as *s* state) light, with respect to the plane of incidence, has different reflectivity [228]. Indeed, this effect is a part of everyday life as the underlying premise behind using polarized sunglasses to avoid glare when driving. In the angled reflection of sunlight off a road, the *s* (horizontally) polarized light is reflected far more strongly than the *p* (vertically) polarized light; so, by blocking the horizontally polarized light, significant glare (i.e., reflection) is removed. What this means for angled reflections in the beam paths of a PARS microscope is that changes in the polarization states of the beams cause corresponding changes in the intensity of reflected light.

The temporal variation in the excitation and detection lasers may be reduced simply by using more stable lasers. Barring that solution, simply monitoring the outputs of the two lasers, and correcting for deviations may prove effective. Monitoring can be achieved by incorporating partially-reflecting "pick-off" mirrors and photo-diodes into the design of the PARS microscope. This addition to the system would produce accurate readings of excitation and detection laser beam power, which is always desirable when performing quantitative imaging. Changes in polarization state can be more problematic. Polarization affects how light reflects off of angled planar surfaces, with one polarization state reflecting more strongly than its orthogonal counterpart. This means that polarization affects the interaction of light with all angled surfaces in the PARS microscope, including the suggested "pick-off" mirrors to be incorporated. Thus, variation seen by the monitoring photodiodes could be a result of changes in polarization in addition to changes in intensity. To eliminate this effect, polarization filters could be installed directly after the excitation and detection lasers, causing all light entering the PARS system to be of a fixed pre-determined polarization state. All variations seen by the proposed monitoring system would then be representative of the power reaching the target. Note that calibration would be required to accurately infer power at the target.

Paths Forward when Measurement Consistency is Achieved

Measurement consistency means that repeated PARS measurements, using fixed imaging parameters, of a static target have limited variability. In mathematical terms, if the PARS microscope is thought of as an abstract system with forward process, $s(t) = \mathrm{PARS}(\vec{x})$, with inputs (imaging parameters and the target) given by vector, $\vec{x}$, and output measurements, $s(t)$, then having consistency means that for a set of recorded measurements for a given fixed input, $\{s_i(t) = \mathrm{PARS}(\vec{x})\}$, the only variation in $\{s_i(t)\}$ is due to the presence of time-stationary system noise, ν_i, on top of some underlying noise-free forward model, $g(\vec{x})$:

$$s_i(t) = g(\vec{x}) + \nu_i. \tag{6.1}$$

Furthermore, assuming the noise, ν_i, is zero-mean, then for a set of N repeated measurements,

$$\lim_{N\to\infty} \mathrm{mean}\{s_i(t)\}_{i=1}^{N} = g(\vec{x}), \tag{6.2}$$

meaning that averaged measurements become increasingly close to the underlying noise-free forward model as the number of measurements, N is increased. If measurement consistency were achieved, then two possible paths forward exist depending on the situation at hand:

1. The forward process is unknown or not well understood.

   Despite not fully understanding the forward process, supervised learning methods can be applied to estimate the solution to a given unmixing problem. The most simple of these would be a look-up table. In the context of blood sO_2 estimation, if measurements over a range of *known* sO_2 values are made, then a look-up table can be formed, and interpolation can readily be used to construct an approximate solution. Furthermore, ideally many measurements would be made for each known sO_2 value, thus allowing for the construction of prior probability distributions conditioned for each associated sO_2 value, and an improved ability to statistically infer true sO_2 from measurements.

2. The physics of the forward process are understood and can be incorporated into a mathematical forward model.

   If the model allows, an "exact" inverse model may be found, thus allowing for direct estimation of the underlying concentrations, similar to what was pursued in Chapter 3. If an "exact" inverse model does not exist, many techniques exist to address this, not least known of which is artificial neural-networks [229–231], which are capable (with an appropriate choice of network architecture and sufficient training) of approximating any non-linear function.

With these paths forward established, the clear next step in improving unmixing accuracy is to work towards increasing measurement consistency.

Updates to Unmixing Formulation

Updates to the unmixing formulation of Chapter 3 to allow for improved accuracy would include explicitly allowing for over-constrained estimation problems (i.e., more measurements than unknowns), and following the SRS characterization presented in Chapter 4, incorporating and accounting for broadband excitation.

With the formulation presented in Equation (3.18), no changes are required to admit over-constrained estimation. The only difference from the analysis presented in Chapter 3 is that the mixing matrix would simply be non-square; however, the methods and constraints developed to solve for unknown chromophore concentrations would still apply.

Updates to the mixing matrix would be necessary to properly incorporate the use of broadband SRS excitation into the model. Rather than assuming excitation light is concentrated at a *discrete* wavelength, the entire *spectrum* of the light should be considered. For multi-wavelength light generated through SRS, the resulting intensity spectrum will be denoted as $I_{SRS,i}(\lambda)$, for a given parameter setting, indexed by i.

In Equation (3.24), a combined mixing matrix, $H = \Phi E R$ was presented, composed of a diagonal matrix of fluences, Φ, the extinction matrix, E, and another diagonal matrix of diffuse reflectance, R. Recall from the discussion on page 48 preceding Equation (3.24), that fluence, ϕ_{λ_i}, at wavelength λ_i is given by $\phi_{\lambda_i} = I_{\lambda_i}/A_{\lambda_i}$, where A_{λ_i} is the focal spot area. Therefore, the diagonal matrix of fluences, Φ, can be decomposed as $\Phi = \mathbf{I}\mathbf{A}^{-1}$, where $\mathbf{I}$ and $\mathbf{A}$ are diagonal matrices of intensity and focal spot area, respectively. Therefore, $H = \mathbf{I}\mathbf{A}^{-1}ER$. For a diffraction-limited focus, $A_{\lambda_i} \propto \lambda_i^2$; however, for broad-band light, the focal spot area varies with wavelength. To address this complication, consider a system of equations involving only a *single* excitation spectrum, $I_{SRS,i}(\lambda)$, and a *single* absorber with extinction given by $\varepsilon_j(\lambda)$. The term $\mathbf{I}\mathbf{A}^{-1}E$ can then conveniently be evaluated explicitly through integration[2], as

$$\mathbf{I}\mathbf{A}^{-1}E = \int I_{SRS,i}(\lambda) A_\lambda^{-1} \varepsilon_j(\lambda)\,\mathrm{d}\lambda. \tag{6.3}$$

Then, accepting arbitrary proportionality rather than strict equality, A_λ can be replaced by λ^2:

$$\mathbf{I}\mathbf{A}^{-1}E \overset{\propto}{=} \int I_{SRS,i}(\lambda) \lambda^{-2} \varepsilon_j(\lambda)\,\mathrm{d}\lambda. \tag{6.4}$$

---

[2] It may seem odd at first to evaluate a matrix product through integration; however, for vector spaces based on sets of functions, it is commonplace to define inner products through integration. In fact, this type of approach underpins much of Fourier analysis, where periodic functions are thought of as vectors. Without further detail, this type of inner product is the premise for matrix multiplication where the entries are functions.

In a similar manner, the general multiple-excitation multiple-absorber system can be expressed as a matrix. For the example of sO_2 estimation, the absorbers are oxy- and deoxyhemoglobin, HbO_2 and Hb, respectively, resulting in

$$\mathbf{IA}^{-1}E \overset{\propto}{=} \begin{bmatrix} \int I_{SRS,1}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{HbO_2}}(\lambda)\,\mathrm{d}\lambda & \int I_{SRS,1}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{Hb}}(\lambda)\,\mathrm{d}\lambda \\ \int I_{SRS,2}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{HbO_2}}(\lambda)\,\mathrm{d}\lambda & \int I_{SRS,2}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{Hb}}(\lambda)\,\mathrm{d}\lambda \end{bmatrix}, \tag{6.5}$$

for two excitation spectra, $I_{SRS,1}(\lambda)$ and $I_{SRS,2}(\lambda)$. Note that this expression is in fact consistent with the previous discrete wavelength model via the replacement $I_{SRS,i}(\lambda) = \delta(\lambda - \lambda_k)$, which reduces the integrals to single-point samples of the extinction spectra at λ_k. To practically compute $\mathbf{IA}^{-1}E$, based on recorded SRS spectra with discrete λ, integrals would be replaced with summations, resulting in

$$\mathbf{IA}^{-1}E \overset{\propto}{=} \begin{bmatrix} \sum_\lambda I_{SRS,1}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{HbO_2}}(\lambda) & \sum_\lambda I_{SRS,1}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{Hb}}(\lambda) \\ \sum_\lambda I_{SRS,2}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{HbO_2}}(\lambda) & \sum_\lambda I_{SRS,2}(\lambda)\lambda^{-2}\varepsilon_{\mathrm{Hb}}(\lambda) \end{bmatrix}. \tag{6.6}$$

This formulation allows $\mathbf{IA}^{-1}E$, and therefore H, to be evaluated for target absorption spectra as a function of arbitrary SRS spectra, or even more specifically, the key parameters discussed in Chapter 4 (fiber type, fiber length, and laser input intensity) associated with the SRS spectra.

For a simple system implementation, a single fiber for multi-wavelength light generation could be used. This would mean the fiber type and length are fixed. In that case, H is then a function of the laser input intensities used. Such a system would be operated by switching the excitation laser intensity to select SRS spectra. If rapid (pulse-to-pulse) switching is desired, which may allow for single-pass imaging, an electro-optic modulator could be employed to modulate the intensity of the beam at suitably high rates. More complicated system architectures could be implemented however, involving multiple fibers and potentially multiple lasers. The output beams from each fiber would need to be co-aligned prior to reaching the microscope objective lens. These architectures would allow for greater flexibility in selecting SRS spectra. Note that a similar analysis as was done in Section 3.1, where optimal excitation wavelengths were selected, should be done for SRS spectra, informing which fiber types and laser parameters will yield optimal SRS for a given unmixing problem.

As was briefly discussed on page 47 at the end of Section 3.2.1, one last additional step towards improving chromophore unmixing accuracy would be to characterize the system noise. By doing so, an empirical prior model of noise, $p_\nu()$, could be created — and would likely be more accurate than the results of assumed Gaussianity, made in Equation (3.9). Repeated from Equation (3.18), the general unmixing solution formulation of

Equation (3.11) would then be updated to include a likelihood term for the noise residual, $p_\nu(\vec{s} - E\vec{c})$, such that

$$\widehat{\vec{c}} = \arg\max_{\vec{c}} \Big\{ p_\nu(\vec{s} - E\vec{c}) - \sum_i w_i \Psi_i(\vec{c}) \Big\}. \tag{6.7}$$

This formulation has the advantage that the likelihood term associated with observing a given noise residual is more accurate. This empirical model would likely vary from system to system, and even between targets. A few sources of noise to consider would include temporal variation in the excitation and detection lasers (with specific noise characteristics associated with each wavelength used), measurement noise at the photo-diode, and electrical noise associated with cabling and the DAQ system. Thus, the model may require several parameters and be somewhat complex. Alternatively, one could observe the system abstractly as whole, rather than examining its constituent parts, and carefully learn noise statistics from observed data. Regardless of the approach taken, having a good understanding of noise may be critical for accurate unmixing.

6.1.2 Further SRS Study and Exploration

In Chapter 4, many fibers were tested and a simulation of the SRS process in single-mode fiber was developed; however, still, there remain additional questions that arose in hindsight, knowing what was only learned through completing the research project. Briefly, a few of these questions are:

1. What would an optimal temporal pulse width be?
2. Can multiple fibers be cascaded (connected sequentially) to extend the range of wavelengths that can be generated?
3. What are the impacts on PARS imaging when centrally eroded pulses (see Figure 4.11 in Section 4.3.2) are used?

Answering these questions would serve to maximize the utility of SRS in PARS microscopy and perhaps beyond.

Optimal Temporal Pulse Width

Regarding question #1, in Section 4.2.5 it was found that 2 ps pulses were too short, where additional wavelength generation through SRS was severely hampered by higher-order optical phenomena. Meanwhile, for the same pulse energy, 1.5 ns pulses showed improvement over 3 ns pulses, by virtue of having roughly twice the pulse intensity. What must be found then is the threshold between using longer pulses with sub-optimal intensity and using more intense pulses at sub-optimal (shorter) widths, for a given pulse energy. One might assume that for a given pulse width that is known to perform well, such as 1.5 ns, the intensity could be increased arbitrarily to achieve higher-order Stokes wave generation; however, doing so increases the pulse energy and thus the average power, which ultimately causes the fiber tip to burn if it is too high. It is desirable to image at high pulse repetition rates, but simultaneously, the average power must remain low enough not to cause burning. Thus, it would be meaningful to find the optimal pulse width for a given pulse energy (as determined by the desired pulse repetition rate for the imaging application and the maximal average power that can be used).

Cascading Fiber to Extend SRS Spectrum

Next, regarding question #2, one must consider the reason a single fiber is inadequate for generating light of arbitrarily long wavelength. In general, it was found that the limiting factor on the extent of the generated SRS spectra was attenuation in the fiber at higher wavelengths. This attenuation occurs due to the wave-guiding behaviour of optical fiber. A specific parameter of interest is the normalized frequency [232], V, often used to determine the cut-off wavelength for single-mode operation of a given fiber. For stepped-index fiber, with core radius, a, core refractive index, n_{core}, and cladding refractive index, n_{cladding}, the normalized frequency for wavelength, λ, is given [232] by

$$V = \frac{2\pi a}{\lambda}\sqrt{n_{\text{core}}^2 - n_{\text{cladding}}^2}. \tag{6.8}$$

If $V \leq 2.4048$ (the eigen-value of the fundamental eigen-mode), only the fundamental mode propagates. This is the condition for single-mode operation. Related to the normalized frequency is the normalized guide index, b, which is shown in Figure 6.2 as a function of V for various modes. A given mode only propagates if $b \geq 0$. For decreasing V, b also decreases until it finally reaches 0. At wavelengths longer than the cut-off wavelength, guidance becomes progressively weaker until finally the fiber ceases to guide ($b = 0$). This is referred to as the fundamental mode cut-off wavelength. Therefore, it is important to select a fiber with core radius and refractive indices such that V remains greater than the fundamental mode cut-off for the longest wavelength desired. Perhaps by transitioning into another fiber just prior to when the next Stokes wavelength would reach the fundamental cut-off, a broader range of wavelengths may be generated.

Splicing fibers together is possible; however, care must be taken to mitigate other unwanted effects. Generally, fibers meant for longer wavelengths have larger cores. This

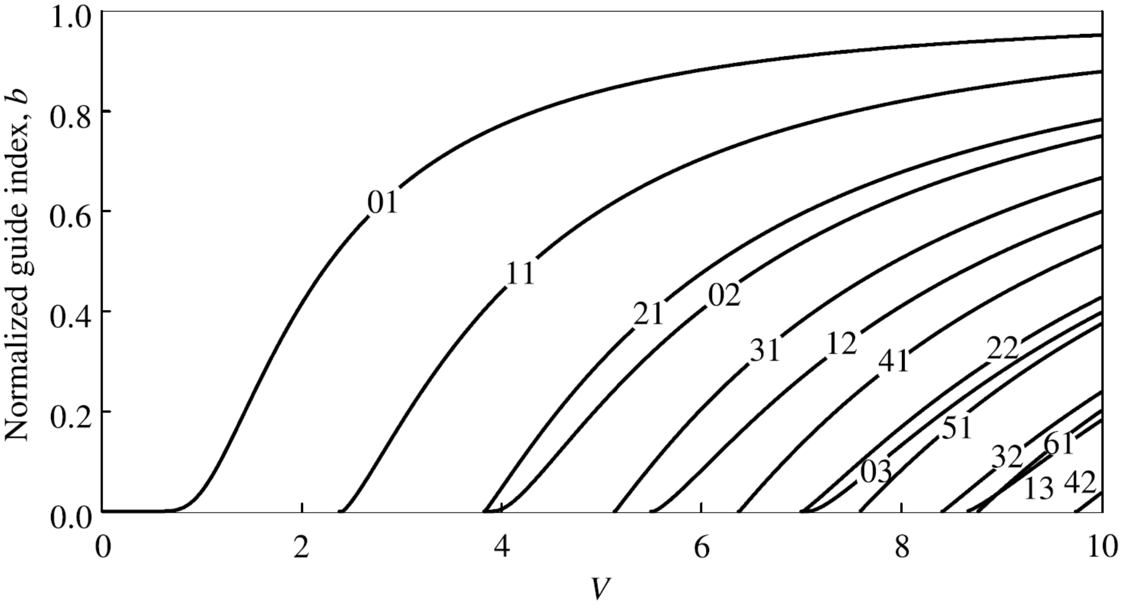


Figure 6.2: The normalized guide index, b, is shown as a function of normalized frequency, V, for several LP modes (labeled on each curve) of step-index fiber. For decreasing V, fewer and fewer modes are able to propagate. Notice that for $V \leq 2.4048$, only the fundamental (01) mode remains. Figure copied from textbook "Photonic Devices" by Jia-Ming Liu [232].

means there will be a transition between fiber of a relatively small core to fiber of a relatively larger one (as fiber meant for longer wavelengths is appended). The change in core size means a change in effective refractive index, and thus an impedance mismatch appears between fibers. Obvious, unwanted reflections occur here due to the mismatch. Multi-mode operation will likely occur as well, for any of the shorter wavelengths. This presents additional challenges as modal dispersion will cause the pulse to separate by mode, with higher-order modes propagating more slowly, lagging behind. In addition, imperfections in the splice may cause further losses. Care must also be taken to ensure the axes of polarization are aligned between the fibers; however, there are existing techniques and technologies available to address this (such as the Fujikura ARCMASTER FSM-100P, which is used at Photomedicine Labs).

Impacts of Central Erosion Effect on PARS Imaging

Lastly, to address question #3, one must consider the time-scale on which the photoacoustic effect takes place. Based on [5], there are two quantities of interest: the thermal relaxation time, τ_{th}, and the stress relaxation time, τ_s. The thermal relaxation time characterizes thermal diffusion out of the heated region. For a given thermal diffusivity (material-dependent), α_{th}, and characteristic dimension, d_c, of the heated region, the thermal relaxation time is given approximately by:

$$\tau_{th} \approx \frac{d_c^2}{\alpha_{th}}. \tag{6.9}$$

The stress relaxation time characterizes pressure propagation from the heated region. For a given speed of sound (again, material-dependent), ν_a, and characteristic dimension, the stress relaxation time is given approximately by

$$\tau_s \approx \frac{d_c}{\nu_a}. \tag{6.10}$$

As a rough approximation, these quantities can be evaluated based on the material properties of water (as an approximation of animal tissue). For $\alpha_{th} = 1.43 \times 10^{-3}$ cm^2/s, $\nu_a = 0.1480$ cm/μs [5], and characteristic dimension based on the focal spot diameter of 532 nm light through the 10X objective regularly used at Photomedicine Labs (University of Waterloo), $d_c = 1.25782 \times 10^{-4}$ cm,

$$\tau_{th} \approx \frac{d_c^2}{\alpha_{th}} = \frac{(1.25782 \times 10^{-4}\ \text{cm})^2}{(1.43 \times 10^{-3}\ \text{cm}^2/\text{s})} \approx 10\ \mu\text{s}, \tag{6.11}$$

and

$$\tau_s \approx \frac{d_c}{\nu_a} = \frac{(1.25782 \times 10^{-4}\ \text{cm})}{(0.1480\ \text{cm}/\mu\text{s})} \approx 1\ \text{ns}. \tag{6.12}$$

Clearly, the thermal relaxation time is greater than the temporal spacing between the tails of centrally eroded pulses shown in Figure 4.11 (approximately 2 ns); however, the stress relaxation time is of a similar scale. Therefore, experimental testing is certainly warranted in order to better understand the impacts on PARS microscopy.

6.1.3 Feature Extraction and Target Labelling

Chapter 5 focused on extracting additional information from PARS TD signals compared to the conventional methods; however, the specific application of emulating the contrast of H&E staining was not addressed. Additionally, during the study, kernel methods [233, 234], in particular Kernel K-Means, were informally explored with little success. Perhaps a second look at these methods, with the background of an improved understanding of the problem, would result in more compelling results.

H&E Stain Emulation via Optimal Linear Projection

H&E staining is an example of highly specific tissue labelling, with hematoxylin staining nuclear structures a blue colour, and eosin staining cytoplasm pink. The specificity of this staining lends itself well to supervised learning approaches for virtual staining. One could imagine a projection of the PARS TD signals that specifies the contrast associated with hematoxylin and eosin stain.

If one-to-one H&E-to-PARS images are collected, perhaps by imaging the tissue with PARS prior to staining and imaging via bright-field microscopy, then it would be possible to establish a metric of error between the projected PARS image and the ground truth H&E. Through discriminant analysis, such as by using Fisher's Linear Discriminant [235, 236], a discriminant could be learned to optimally distinguish between components that should be associated with hematoxylin vs. with eosin. Because not all structures would be stained, a third class may be necessary, requiring slightly more elaborate methods (multiple discriminants or ensemble methods). Nonetheless, optimization could be used to find a suitable discriminant, ideally yielding a linear transformation of the PARS TD signals that can be used to project signals directly to intensities of virtual H&E labels.

Kernel Methods

In the early stages of approaching the problem of how best to extract additional information from PARS TD signals, kernel methods were explored with little success and thus were not presented in Chapter 5, nor explained in the background (Section 2.4). Nonetheless, with the prevalence of kernel methods broadly throughout the field of machine learning, it perhaps deserves a second look. In this section, the basic theory of kernel methods is explained before discussing how these methods might be applied.

For the tasks of classification and clustering data, it is advantageous to be able to linearly separate inherent / underlying classes of the data. If this is not possible, methods such as PCA and K-means, using distance or similarity metrics, including Euclidean Distance and Cosine Similarity, will fail. A canonical example of this is given in Figure 6.3. The data set is comprised of two classes in two dimensions, where the first class is distributed

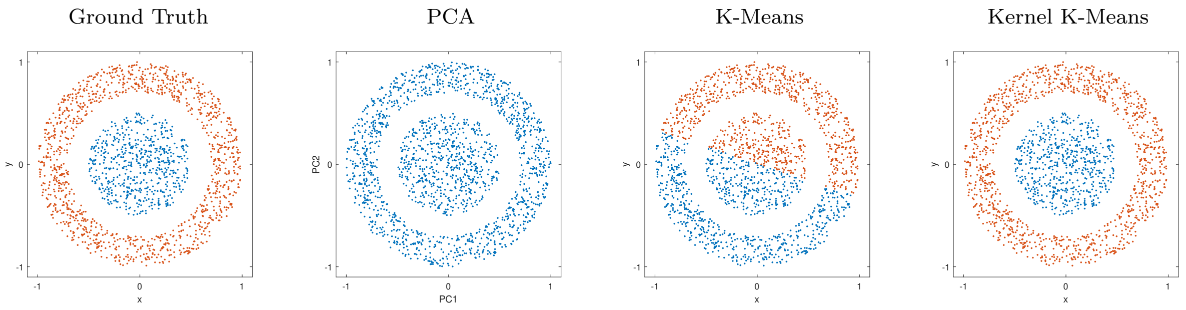


Figure 6.3: A concentric circle data-set, with ground truth class labels shown through colour, is presented in the left panel, and processed using PCA, K-Means, and Kernel K-Means. Using PCA does not yield linearly separable classes, and K-Means is not able to separate the underlying classes. In contrast to the results of these methods, Kernel K-Means, with the radial kernel described in Equation (6.13), is able to accurately separate the classes.

such that all points lie within the unit circle centered at the origin, and the second class is distributed annularly around the first class, with a radial gap separating the classes. PCA is unable to produce principal components where the classes are linearly separable. This is because the principal component space is a *continuous, linear* transformation (rotation, stretching, and skewing) of the original space, which cannot move points of the inner class such that they lie outside of the surrounding annular class, which would be required for linear separation. K-means also fails to cluster the data according to its underlying structure despite the fact that visually, the two clusters are clearly distinct. Again, this is because K-means can only produce linearly separated clusters, given that cluster membership is assigned based on distance to the mean (for any pair of cluster means, a linear cluster boundary, exactly halfway between the means, is implied). The same principle applies to many other data sets, such as "c-shaped" classes, and "spiral-shaped" data, as presented in Chapter 10 of [237].

To overcome this major issue, the data may be transformed to a domain / space where the classes are linearly separable. Consider the following transformation on data points $p = (x, y) \in \mathbb{R}^2$:

$$\begin{aligned} \phi \colon \mathbb{R}^2 &\to \mathbb{R}^3 \\ p &\mapsto \phi(p) = (x^2, \sqrt{2}xy, y^2). \end{aligned} \tag{6.13}$$

In the transformed space, the data from the canonical example, $\phi(p)$, are linearly separable by a plane. Now, performing tasks such as PCA and K-Means requires computing distances in the transformed space. Indeed, in the transformed space, K-Means can successfully cluster the data, as shown in the right panel of Figure 6.3. It can be shown algebraically [233, 234] that the distance calculations amount to computing inner products (generalized dot-products). For a well-chosen and simple transformation, like the one in Equation (6.13), computing the transformation, distances between points, and even inner products / dot-products of the transformed points is simple; however, for more complicated transformations such as the Gaussian Radial Basis Function, the transformed space may be of very high or even infinite dimension, making the computation of the transform as well as working in the transformed space extremely prohibitive. Fortunately, it is possible to compute inner products in the transformed space without actually performing the transformation! This is done through a *kernel* function.

Kernel functions are defined such that for two points, p and q, in the input domain (continuing with the example of Figure 6.3, $\mathbb{R}^2$), the kernel yields the inner product of these points in the transformed domain:

$$\begin{aligned} \kappa\colon \mathbb{R}^2 &\to \mathbb{R} \\ (p,q) &\mapsto \kappa(p,q) = \langle \phi(p), \phi(q) \rangle. \end{aligned} \tag{6.14}$$

For the transformation from Equation (6.13), the associated kernel function is the following, which with simple algebraic manipulations, can be simplified:

$$\begin{aligned} \kappa(p,q) &= \langle \phi(p), \phi(q) \rangle \\ &= \langle (x_p^2, \sqrt{2}x_p y_p, y_p^2), (x_q^2, \sqrt{2}x_q y_q, y_q^2) \rangle \\ &= x_p^2 x_q^2 + 2x_p y_p x_q y_q + y_p^2 y_q^2 \\ &= (x_p x_q + y_p y_q)^2 \\ &= \langle (x_p, y_p), (x_q, y_q) \rangle^2 \\ &= \langle p, q \rangle^2 \end{aligned} \tag{6.15}$$

Thus, the inner product in the transformed domain ($\mathbb{R}^3$) can be computed via κ without ever performing the transformation ϕ!

As explained earlier, classification problems are not always linearly separable. To address this, complex feature transformations, possibly to infinite-dimensional spaces, may be employed, making the underlying classes linearly separable in the transformed domain. Because of the complexity of such transformations, actually computing the transformations and computing distances in the transformed spaces are prohibitive. This is where kernel functions are extremely useful, allowing distances in the transformed spaces to be computed directly.

While PARS TD signals are already in fairly high-dimensional space, often more than a thousand dimensions, there may still be advantages to performing a feature transformation, such that linear separation of classes or *types* of signals may be possible. Methods such as Kernel PCA and Kernel K-Means [233, 234] may provide highly effective methods for analysing and separating TD signals based on the underlying target. Based on the current understanding of PARS signals, explained in Section 5.2, where signals of a given type can vary in amplitude and polarity, it may be possible to design either a feature transformation or a kernel catered specifically towards the problem of separating PARS TD signals. This remains an area of continued interest and research.

References

[1] N. Pellegrino, P. Fieguth, and P. H. Reza, "K-means for noise-insensitive multi-dimensional feature learning," *arXiv preprint arXiv:2202.07754*, 2022, Submitted to the 26th International Conference on Pattern Recognition (ICPR 2022).

[2] N. Pellegrino, B. R. Ecclestone, P. Fieguth, and P. H. Reza, "Time-domain feature extraction for target-specificity in photoacoustic remote sensing microscopy," *arXiv preprint arXiv: arXiv:2203.04316*, 2022, Submitted to Optics Letters.

[3] Z. Hosseinaee *et al.*, "In-vivo functional and structural retinal imaging using multiwavelength photoacoustic remote sensing microscopy," *Scientific Reports*, vol. 12, no. 1, p. 4562, Mar. 2022, ISSN: 2045-2322. DOI: `10.1038/s41598-022-08508-2`. [Online]. Available: `https://doi.org/10.1038/s41598-022-08508-2`.

[4] B. Burbridge and E. Mah, *Undergraduate diagnostic imaging fundamentals*. University of Saskatchewan, 2017.

[5] L. V. Wang and H.-i. Wu, *Biomedical optics: principles and imaging*. John Wiley & Sons, 2012.

[6] P. L. Allan, G. M. Baxter, and M. J. Weston, *Clinical Ultrasound, 2-Volume Set E-Book: Expert Consult: Online and Print*. Elsevier Health Sciences, 2011.

[7] J. C. Hobbins, *Obstetric ultrasound: artistry in practice*. John Wiley & Sons, 2008.

[8] J. Cohen-Adad and C. Wheeler-Kingshott, *Quantitative MRI of the spinal cord*. Academic Press, 2014.

[9] S. A. Smith, J. J. Pekar, and P. C. Van Zijl, "Advanced mri strategies for assessing spinal cord injury," *Handbook of clinical neurology*, vol. 109, pp. 85–101, 2012.

[10] B. M. Ellingson, N. Salamon, and L. T. Holly, "Imaging techniques in spinal cord injury," *World neurosurgery*, vol. 82, no. 6, pp. 1351–1358, 2014.

[11] Y. Kumar and D. Hayashi, “Role of magnetic resonance imaging in acute spinal trauma: A pictorial review,” *BMC musculoskeletal disorders*, vol. 17, no. 1, pp. 1–11, 2016.

[12] S. Sippel, K. Muruganandan, A. Levine, and S. Shah, “Use of ultrasound in the developing world,” *International journal of emergency medicine*, vol. 4, no. 1, pp. 1–11, 2011.

[13] R. Guo, G. Lu, B. Qin, and B. Fei, “Ultrasound imaging technologies for breast cancer detection and management: A review,” *Ultrasound in medicine & biology*, vol. 44, no. 1, pp. 37–70, 2018.

[14] H. E. Martz, C. M. Logan, D. J. Schneberk, and P. J. Shull, *X-ray Imaging: fundamentals, industrial techniques and applications*. CRC Press, 2016.

[15] M. Hoheisel, “Review of medical imaging with emphasis on x-ray detectors,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 563, no. 1, pp. 215–224, 2006.

[16] W. A. Kalender, “X-ray computed tomography,” *Physics in Medicine & Biology*, vol. 51, no. 13, R29, 2006.

[17] P. J. Withers *et al.*, “X-ray computed tomography,” *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 1–21, 2021.

[18] G. Katti, S. A. Ara, and A. Shireen, “Magnetic resonance imaging (mri)–a review,” *International journal of dental clinics*, vol. 3, no. 1, pp. 65–70, 2011.

[19] S. Geethanath and J. T. Vaughan Jr, “Accessible magnetic resonance imaging: A review,” *Journal of Magnetic Resonance Imaging*, vol. 49, no. 7, e65–e77, 2019.

[20] *Optical imaging*, Dec. 2020. [Online]. Available: `https://www.nibib.nih.gov/science-education/science-topics/optical-imaging`.

[21] J. M. Schmitt, “Optical coherence tomography (oct): A review,” *IEEE Journal of selected topics in quantum electronics*, vol. 5, no. 4, pp. 1205–1215, 1999.

[22] M. Ang *et al.*, “Optical coherence tomography angiography: A review of current and future clinical applications,” *Graefe’s Archive for Clinical and Experimental Ophthalmology*, vol. 256, no. 2, pp. 237–245, 2018.

[23] A. Yasin Alibhai, C. Or, and A. J. Witkin, “Swept source optical coherence tomography: A review,” *Current Ophthalmology Reports*, vol. 6, no. 1, pp. 7–16, 2018.

[24] Z. Sun, D. Yang, Z. Tang, D. S. Ng, and C. Y. Cheung, “Optical coherence tomography angiography in diabetic retinopathy: An updated review,” *Eye*, vol. 35, no. 1, pp. 149–161, 2021.

[25] M. Pluta, *Advanced Light Microscopy vol. 1 Principles and Basic Properties*. Elsevier, 1988.

[26] S. H. Gage, "Modern dark-field microscopy and the history of its development," *Transactions of the American Microscopical Society*, vol. 39, no. 2, pp. 95–141, 1920.

[27] W. Chambers, T. J. Fellers, and M. W. Davidson, *Darkfield illumination*. [Online]. Available: https://www.microscopyu.com/techniques/stereomicroscopy/darkfield-illumination.

[28] M. Pluta, "Phase contrast microscopy," *Advanced light microscopy*, vol. 2, 1993.

[29] C. Burch and J. Stock, "Phase-contrast microscopy," *Journal of Scientific Instruments*, vol. 19, no. 5, p. 71, 1942.

[30] N. Panwar *et al.*, "Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare," *Telemedicine and e-Health*, vol. 22, no. 3, pp. 198–208, 2016.

[31] R. Bernardes, P. Serranho, and C. Lobo, "Digital ocular fundus imaging: A review," *Ophthalmologica*, vol. 226, no. 4, pp. 161–181, 2011.

[32] E. Pead *et al.*, "Automated detection of age-related macular degeneration in color fundus photography: A systematic review," *survey of ophthalmology*, vol. 64, no. 4, pp. 498–511, 2019.

[33] U. Iqbal, "Smartphone fundus photography: A narrative review," *International Journal of Retina and Vitreous*, vol. 7, no. 1, pp. 1–12, 2021.

[34] R. F. Guthoff, A. Zhivov, and O. Stachs, "In vivo confocal microscopy, an inner vision of the cornea–a major review," *Clinical & experimental ophthalmology*, vol. 37, no. 1, pp. 100–117, 2009.

[35] D. V. Patel and C. N. McGhee, "Quantitative analysis of in vivo confocal microscopy images: A review," *Survey of ophthalmology*, vol. 58, no. 5, pp. 466–475, 2013.

[36] O. E. Olarte, J. Andilla, E. J. Gualda, and P. Loza-Alvarez, "Light-sheet microscopy: A tutorial," *Advances in Optics and Photonics*, vol. 10, no. 1, pp. 111–179, 2018.

[37] P. K. Poola, M. I. Afzal, Y. Yoo, K. H. Kim, and E. Chung, "Light sheet microscopy for histopathology applications," *Biomedical engineering letters*, vol. 9, no. 3, pp. 279–291, 2019.

[38] Y. Wan, K. McDole, and P. J. Keller, "Light-sheet microscopy and its potential for understanding developmental processes," *Annual review of cell and developmental biology*, vol. 35, pp. 655–681, 2019.

[39] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature methods*, vol. 2, no. 12, pp. 910–919, 2005.

[40] B. O. Leung and K. C. Chou, "Review of super-resolution fluorescence microscopy for biology," *Applied spectroscopy*, vol. 65, no. 9, pp. 967–980, 2011.

[41] P. A. Santi, "Light sheet fluorescence microscopy: A review," *Journal of Histochemistry & Cytochemistry*, vol. 59, no. 2, pp. 129–138, 2011.

[42] K. A. Antonio and Z. D. Schultz, "Advances in biomedical raman microscopy," *Analytical chemistry*, vol. 86, no. 1, pp. 30–46, 2014.

[43] S. G. da Costa, A. Richter, U. Schmidt, S. Breuninger, and O. Hollricher, "Confocal raman microscopy in life sciences," *Morphologie*, vol. 103, no. 341, pp. 11–16, 2019.

[44] P. Beard, "Biomedical photoacoustic imaging," *Interface focus*, vol. 1, no. 4, pp. 602–631, 2011.

[45] J. Yao and L. V. Wang, "Photoacoustic microscopy," *Laser & photonics reviews*, vol. 7, no. 5, pp. 758–778, 2013.

[46] W. Liu and H. F. Zhang, "Photoacoustic imaging of the eye: A mini review," *Photoacoustics*, vol. 4, no. 3, pp. 112–123, 2016.

[47] Z. Hu, X. Wang, Q. Liu, and Y. Paulus, "Photoacoustic imaging in ophthalmology," *Int J Ophthalmol Eye Res*, vol. 3, no. 8, pp. 126–32, 2015.

[48] V. P. Nguyen and Y. M. Paulus, "Photoacoustic ophthalmoscopy: Principle, application, and future directions," *Journal of imaging*, vol. 4, no. 12, p. 149, 2018.

[49] R. D. Glickman, "Photoacoustic imaging and sensing: A new way to see the eye," *Journal of Ocular Pharmacology and Therapeutics*, vol. 37, no. 3, pp. 162–171, 2021.

[50] P. Hajireza, W. Shi, K. Bell, R. J. Paproski, and R. J. Zemp, "Non-interferometric photoacoustic remote sensing microscopy," *Light: Science & Applications*, vol. 6, no. 6, e16278–e16278, 2017.

[51] P. H. Reza, K. Bell, W. Shi, J. Shapiro, and R. J. Zemp, "Deep non-contact photoacoustic initial pressure imaging," *Optica*, vol. 5, no. 7, pp. 814–820, 2018.

[52] S. Abbasi, D. Dinakaran, G. Bigras, J. R. Mackey, and P. H. Reza, "All-optical label-free human breast tissue block histology using photoacoustic remote sensing," *Optics Letters*, vol. 45, no. 17, pp. 4770–4773, 2020.

[53] B. R. Ecclestone *et al.*, "Three-dimensional virtual histology in unprocessed resected tissues with photoacoustic remote sensing (pars) microscopy and optical coherence tomography (oct)," *Scientific Reports*, vol. 11, no. 1, 2021. DOI: `10.1038/s41598-021-93222-8`.

[54] B. R. Ecclestone *et al.*, “Improving maximal safe brain tumor resection with photoacoustic remote sensing microscopy,” *Scientific Reports*, vol. 10, no. 1, pp. 1–7, 2020.

[55] B. R. Ecclestone *et al.*, “Towards virtual biopsies of gastrointestinal tissues using photoacoustic remote sensing microscopy,” *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 3, 2021.

[56] B. Ecclestone, D. Dinakaran, and P. H. Reza, “Single acquisition label-free histology-like imaging with dual-contrast photoacoustic remote sensing microscopy,” *Journal of Biomedical Optics*, vol. 26, no. 5, 2021.

[57] K. Bell *et al.*, “Reflection-mode virtual histology using photoacoustic remote sensing microscopy,” *Scientific reports*, vol. 10, no. 1, pp. 1–13, 2020.

[58] S. Abbasi *et al.*, “Chromophore selective multi-wavelength photoacoustic remote sensing of unstained human tissues,” *Biomedical optics express*, vol. 10, no. 11, pp. 5461–5469, 2019.

[59] N. J. Haven, K. L. Bell, P. Kedarisetti, J. D. Lewis, and R. J. Zemp, “Ultraviolet photoacoustic remote sensing microscopy,” *Optics letters*, vol. 44, no. 14, pp. 3586–3589, 2019.

[60] N. J. Haven, P. Kedarisetti, B. S. Restall, and R. J. Zemp, “Reflective objective-based ultraviolet photoacoustic remote sensing virtual histopathology,” *Optics Letters*, vol. 45, no. 2, pp. 535–538, 2020.

[61] J. Zhou *et al.*, “Miniature non-contact photoacoustic probe based on fiber-optic photoacoustic remote sensing microscopy,” *Optics Letters*, vol. 46, no. 22, pp. 5767–5770, 2021.

[62] Z. Hosseinaee *et al.*, “Functional photoacoustic remote sensing microscopy using a stabilized temperature-regulated stimulated raman scattering light source,” *Optics Express*, vol. 29, no. 19, pp. 29 745–29 754, 2021.

[63] K. L. Bell, P. H. Reza, and R. J. Zemp, “Real-time functional photoacoustic remote sensing microscopy,” *Optics letters*, vol. 44, no. 14, pp. 3466–3469, 2019.

[64] N. D. Wangsa-Wirawan and R. A. Linsenmeier, “Retinal oxygen: Fundamental and clinical aspects,” *Archives of ophthalmology*, vol. 121, no. 4, pp. 547–557, 2003.

[65] F. Colliez, B. Gallez, and B. F. Jordan, “Assessing tumor oxygenation for predicting outcome in radiation oncology: A review of studies correlating tumor hypoxic status and outcome in the preclinical and clinical settings,” *Frontiers in oncology*, vol. 7, p. 10, 2017.

[66] F. Z. Yetkin and D. Mendelsohn, “Hypoxia imaging in brain tumors,” *Neuroimaging Clinics*, vol. 12, no. 4, pp. 537–552, 2002.

[67] D. Meadows, “Recent developments with biosensing technology and applications in the pharmaceutical industry,” *Advanced drug delivery reviews*, vol. 21, no. 3, pp. 179–189, 1996.

[68] J. K. Chan, “The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology,” *International journal of surgical pathology*, vol. 22, no. 1, pp. 12–32, 2014.

[69] B. R. Ecclestone *et al.*, “Histopathology for mohs micrographic surgery with photoacoustic remote sensing microscopy,” *Biomedical Optics Express*, vol. 12, no. 1, pp. 654–665, 2021.

[70] S. Abbasi *et al.*, “All-optical reflection-mode microscopic histology of unstained human tissues,” *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[71] A. G. Bell, “On the production and reproduction of sound by light,” in *Proc. Am. Assoc. Adv. Sci.*, vol. 29, 1881, pp. 115–136.

[72] S. Jeon, J. Kim, D. Lee, J. W. Baik, and C. Kim, “Review on practical photoacoustic microscopy,” *Photoacoustics*, vol. 15, p. 100 141, 2019.

[73] G. Li, K. I. Maslov, and L. V. Wang, “Reflection-mode multifocal optical-resolution photoacoustic microscopy,” *Journal of biomedical optics*, vol. 18, no. 3, p. 030 501, 2013.

[74] K. Maslov, H. F. Zhang, S. Hu, and L. V. Wang, “Optical-resolution photoacoustic microscopy for in vivo imaging of single capillaries,” *Optics letters*, vol. 33, no. 9, pp. 929–931, 2008.

[75] C. Liu, J. Chen, Y. Zhang, J. Zhu, and L. Wang, “Five-wavelength optical-resolution photoacoustic microscopy of blood and lymphatic vessels,” *Advanced Photonics*, vol. 3, no. 1, p. 016 002, 2021.

[76] W. Qin, T. Jin, H. Guo, and L. Xi, “Large-field-of-view optical resolution photoacoustic microscopy,” *Optics express*, vol. 26, no. 4, pp. 4271–4278, 2018.

[77] S. Park, C. Lee, J. Kim, and C. Kim, “Acoustic resolution photoacoustic microscopy,” *Biomedical Engineering Letters*, vol. 4, no. 3, pp. 213–222, 2014.

[78] R. Gao, Q. Xue, Y. Ren, H. Zhang, L. Song, and C. Liu, “Achieving depth-independent lateral resolution in ar-pam using the synthetic-aperture focusing technique,” *Photoacoustics*, p. 100 328, 2021.

[79] M. Li, Y. Tang, and J. Yao, "Photoacoustic tomography of blood oxygenation: A mini review," *Photoacoustics*, vol. 10, pp. 65–73, 2018.

[80] W. Choi, D. Oh, and C. Kim, "Practical photoacoustic tomography: Realistic limitations and technical solutions," *Journal of Applied Physics*, vol. 127, no. 23, p. 230 903, 2020.

[81] J. Jo *et al.*, "Photoacoustic tomography for human musculoskeletal imaging and inflammatory arthritis detection," *Photoacoustics*, vol. 12, pp. 82–89, 2018.

[82] J. Yang, S. Choi, and C. Kim, "Practical review on photoacoustic computed tomography using curved ultrasound array transducer," *Biomedical Engineering Letters*, pp. 1–17, 2021.

[83] P. R. Torke, R. Nuster, and G. Paltauf, "Conical ring array detector for large depth of field photoacoustic macroscopy," *Biomedical Optics Express*, vol. 11, no. 5, pp. 2461–2475, 2020.

[84] H. Zhang *et al.*, "A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography," *Computerized Medical Imaging and Graphics*, vol. 84, p. 101 720, 2020.

[85] Y. Wang, C. Li, and R. K. Wang, "Noncontact photoacoustic imaging achieved by using a low-coherence interferometer as the acoustic detector," *Optics letters*, vol. 36, no. 20, pp. 3975–3977, 2011.

[86] Z. Chen, S. Yang, Y. Wang, and D. Xing, "Noncontact broadband all-optical photoacoustic microscopy based on a low-coherence interferometer," *Applied Physics Letters*, vol. 106, no. 4, p. 043 701, 2015.

[87] A. Hochreiner, J. Bauer-Marschallinger, P. Burgholzer, B. Jakoby, and T. Berer, "Non-contact photoacoustic imaging using a fiber based interferometer with optical amplification," *Biomedical optics express*, vol. 4, no. 11, pp. 2322–2331, 2013.

[88] G. Rousseau, B. Gauthier, A. Blouin, and J.-P. Monchalin, "Non-contact biomedical photoacoustic and ultrasound imaging," *Journal of biomedical optics*, vol. 17, no. 6, p. 061 217, 2012.

[89] G. Xu, C. Wang, T. Feng, D. E. Oliver, and X. Wang, "Non-contact photoacoustic tomography with a laser doppler vibrometer," in *Photons Plus Ultrasound: Imaging and Sensing 2014*, International Society for Optics and Photonics, vol. 8943, 2014, p. 894 332.

[90] F. Yang, Z. Chen, and D. Xing, "All-optical noncontact phase-domain photoacoustic elastography," *Optics Letters*, vol. 46, no. 19, pp. 5063–5066, 2021.

[91] N. Ichikawa and Y. Monnai, “Detection of sub-terahertz photoacoustic effect for non-contact material sensing,” *Measurement: Sensors*, vol. 18, p. 100 239, 2021.

[92] D. George, H. Lloyd, R. H. Silverman, and P. V. Chitnis, “A frequency-domain non-contact photoacoustic microscope based on an adaptive interferometer,” *Journal of biophotonics*, vol. 11, no. 6, e201700278, 2018.

[93] C. V. Raman, “A new radiation,” *Indian Journal of physics*, vol. 2, pp. 387–398, 1928.

[94] G. P. Agrawal, “Nonlinear fiber optics,” in *Nonlinear Science at the Dawn of the 21st Century*, Springer, 2000.

[95] S. H. Jack, D. B. Hann, and C. A. Greated, “Influence of the acousto-optic effect on laser doppler anemometry signals,” *Review of scientific instruments*, vol. 69, no. 12, pp. 4074–4081, 1998.

[96] C. Liu, Y. Liang, and L. Wang, “Optical-resolution photoacoustic microscopy of oxygen saturation with nonlinear compensation,” *Biomedical optics express*, vol. 10, no. 6, pp. 3061–3069, 2019.

[97] P. Fieguth, *Statistical image processing and multidimensional modeling.* Springer Science & Business Media, 2010.

[98] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems.* Elsevier, 2018.

[99] A. Tarantola, B. Valette, *et al.*, “Inverse problems= quest for information,” *Journal of geophysics*, vol. 50, no. 1, pp. 159–170, 1982.

[100] M. Benning and M. Burger, “Modern regularization methods for inverse problems,” *Acta Numerica*, vol. 27, pp. 1–111, 2018.

[101] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, “Solving inverse problems using data-driven models,” *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[102] D. Calvetti and E. Somersalo, “Inverse problems: From regularization to bayesian inference,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 10, no. 3, e1427, 2018.

[103] E. C. Cherry, “Some experiments on the recognition of speech, with one and with two ears,” *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[104] M. A. Bee and C. Micheyl, “The cocktail party problem: What is it? how can it be solved? and why should animal behaviorists study it?” *Journal of comparative psychology*, vol. 122, no. 3, p. 235, 2008.

[105] S. Haykin and Z. Chen, “The cocktail party problem,” *Neural computation*, vol. 17, no. 9, pp. 1875–1902, 2005.

[106] A. W. Bronkhorst, “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.

[107] S. Prahl, *Optical absorption of hemoglobin*, `https://omlc.org/spectra/hemoglobin/index.html`, 1999.

[108] G. B. Folland and A. Sitaram, “The uncertainty principle: A mathematical survey,” *Journal of Fourier analysis and applications*, vol. 3, no. 3, pp. 207–238, 1997.

[109] E. H. Moore, “On the reciprocal of the general algebraic matrix,” *Bull. Am. Math. Soc.*, vol. 26, pp. 394–395, 1920.

[110] A. Bjerhammar, *Application of calculus of matrices to method of least squares: with special reference to geodetic calculations*. Elander, 1951.

[111] R. Penrose, “A generalized inverse for matrices,” *Cambridge Philosophical Society Proceedings*, vol. 51, pp. 406–403, 1955.

[112] A. N. Tikhonov, “On the stability of inverse problems,” in *Dokl. Akad. Nauk SSSR*, vol. 39, 1943, pp. 195–198.

[113] A. N. Tihonov, “Solution of incorrectly formulated problems and the regularization method,” *Soviet Math.*, vol. 4, pp. 1035–1038, 1963.

[114] J. Hafner *et al.*, “Regional patterns of retinal oxygen saturation and microvascular hemodynamic parameters preceding retinopathy in patients with type ii diabetes,” *Investigative ophthalmology & visual science*, vol. 58, no. 12, pp. 5541–5547, 2017.

[115] C. Türksever and M. G. Todorova, “Oxygenation in glaucoma patients: Influence of age and oxidative stress,” *Acta Ophthalmologica*, vol. 100, 2022.

[116] M. Sehi *et al.*, “Relationship between retinal blood flow, ganglion cell function and retinal oxygenation in glaucomatous and normal eyes,” *Investigative Ophthalmology & Visual Science*, vol. 55, no. 13, pp. 4731–4731, 2014.

[117] E. Donicova, L. Ramm, R. Augsten, and M. Hammer, “The flicker response of venous oxygen saturation is significantly reduced in the early and late stages of age-related macular degeneration,” *Graefe’s Archive for Clinical and Experimental Ophthalmology*, vol. 258, no. 1, pp. 31–37, 2020.

[118] Z. Hosseinaee, L. Khalili, J. A. T. Simmons, K. Bell, and P. H. Reza, “Label-free, non-contact, in vivo ophthalmic imaging using photoacoustic remote sensing microscopy,” *Optics letters*, vol. 45, no. 22, pp. 6254–6257, 2020.

[119] Z. Hosseinaee *et al.*, "Non-contact, in-vivo, functional, and structural ophthalmic imaging using dual-modal photoacoustic remote sensing microscopy and optical coherence tomography," *Investigative Ophthalmology & Visual Science*, vol. 62, no. 8, pp. 2316–2316, 2021.

[120] K. Bell, L. Mukhangaliyeva, L. Khalili, and P. H. Reza, "Hyperspectral absorption microscopy using photoacoustic remote sensing," *Optics Express*, vol. 29, no. 15, pp. 24 338–24 348, 2021.

[121] M. J. Moore, P. M. Schygulla, E. M. Strohm, and M. C. Kolios, "Single red blood cell oxygenation saturation imaging with multispectral photoacoustic microscopy," in *2016 IEEE International Ultrasonics Symposium (IUS)*, IEEE, 2016, pp. 1–4.

[122] P. Hajireza, A. Forbrich, and R. Zemp, "In-vivo functional optical-resolution photoacoustic microscopy with stimulated raman scattering fiber-laser source," *Biomedical optics express*, vol. 5, no. 2, pp. 539–546, 2014.

[123] S. Tzoumas and V. Ntziachristos, "Spectral unmixing techniques for optoacoustic imaging of tissue pathophysiology," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, no. 2107, 2017.

[124] X. Wang, X. Xie, G. Ku, L. V. Wang, and G. Stoica, "Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography," *Journal of biomedical optics*, vol. 11, no. 2, p. 024 015, 2006.

[125] M. Sivaramakrishnan, K. Maslov, H. F. Zhang, G. Stoica, and L. V. Wang, "Limitations of quantitative photoacoustic measurements of blood oxygenation in small vessels," *Physics in Medicine & Biology*, vol. 52, no. 5, p. 1349, 2007.

[126] A. Needles *et al.*, "Development of a combined photoacoustic micro-ultrasound system for estimating blood oxygenation," in *2010 IEEE International Ultrasonics Symposium*, IEEE, 2010, pp. 390–393.

[127] A. Hussain, W. Petersen, J. Staley, E. Hondebrink, and W. Steenbergen, "Quantitative blood oxygen saturation imaging using combined photoacoustics and acousto-optics," *Optics letters*, vol. 41, no. 8, pp. 1720–1723, 2016.

[128] L. M. Yamaleyeva, Y. Sun, T. Bledsoe, A. Hoke, S. B. Gurley, and K. B. Brosnihan, "Photoacoustic imaging for in vivo quantification of placental oxygenation in mice," *The FASEB Journal*, vol. 31, no. 12, pp. 5520–5529, 2017.

[129] F. Cao, Z. Qiu, H. Li, and P. Lai, "Photoacoustic imaging in oxygen detection," *Applied Sciences*, vol. 7, no. 12, p. 1262, 2017.

[130] J. Chen, Y. Zhang, L. He, Y. Liang, and L. Wang, “Wide-field polygon-scanning photoacoustic microscopy of oxygen saturation at 1-mhz a-line rate,” *Photoacoustics*, vol. 20, p. 100 195, 2020.

[131] F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to optics.* Cambridge University Press, 2017.

[132] A. Ghatak, *Optics.* McGraw-Hill, 2005.

[133] A. Cox, A. J. DeWeerd, and J. Linden, “An experiment to measure mie and rayleigh total scattering cross sections,” *American Journal of Physics*, vol. 70, no. 6, pp. 620–625, 2002.

[134] Y. Shen, *The Principles of Nonlinear Optics.* Wiley, 1984.

[135] R. Stolen and E. Ippen, “Raman gain in glass optical waveguides,” *Applied Physics Letters*, vol. 22, no. 6, pp. 276–278, 1973.

[136] D. K. Pandey, H. L. Kagdada, P. Sanchora, and D. K. Singh, “Overview of raman spectroscopy: Fundamental to applications,” *Modern Techniques of Spectroscopy: Basics, Instrumentation, and Applications*, pp. 145–184, 2021.

[137] C. S. Kumar, *Raman spectroscopy for nanomaterials characterization.* Springer Science & Business Media, 2012.

[138] B. D Patel and P. J Mehta, “An overview: Application of raman spectroscopy in pharmaceutical field,” *Current Pharmaceutical Analysis*, vol. 6, no. 2, pp. 131–141, 2010.

[139] R. Shuker and R. W. Gammon, “Raman-scattering selection-rule breaking and the density of states in amorphous materials,” *Physical Review Letters*, vol. 25, no. 4, p. 222, 1970.

[140] E. Ippen and R. Stolen, “Stimulated brillouin scattering in optical fibers,” *Applied Physics Letters*, vol. 21, no. 11, pp. 539–541, 1972.

[141] R. H. Stolen and C. Lin, “Self-phase-modulation in silica optical fibers,” *Physical Review A*, vol. 17, no. 4, p. 1448, 1978.

[142] M. N. Islam, L. F. Mollenauer, R. H. Stolen, J. R. Simpson, and H.-T. Shang, “Cross-phase modulation in optical fibers,” *Optics letters*, vol. 12, no. 8, pp. 625–627, 1987.

[143] J. R. Thompson and R. Roy, “Nonlinear dynamics of multiple four-wave mixing processes in a single-mode fiber,” *Physical Review A*, vol. 43, no. 9, p. 4987, 1991.

[144] D. Dougherty, F. Kärtner, H. Haus, and E. Ippen, "Measurement of the raman gain spectrum of optical fibers," *Optics letters*, vol. 20, no. 1, pp. 31–33, 1995.

[145] R. G. Smith, "Optical power handling capacity of low loss optical fibers as determined by stimulated raman and brillouin scattering," *Applied optics*, vol. 11, no. 11, pp. 2489–2494, 1972.

[146] R. Stolen, "Polarization effects in fiber raman and brillouin lasers," *IEEE Journal of Quantum Electronics*, vol. 15, no. 10, pp. 1157–1160, 1979.

[147] R. H. Stolen, C. Lee, and R. Jain, "Development of the stimulated raman spectrum in single-mode silica fibers," *JOSA B*, vol. 1, no. 4, pp. 652–657, 1984.

[148] K. X. Liu and E. Garmire, "Understanding the formation of the srs stokes spectrum in fused silica fibers," *IEEE journal of quantum electronics*, vol. 27, no. 4, pp. 1022–1030, 1991.

[149] K. J. Blow and D. Wood, "Theoretical description of transient stimulated raman scattering in optical fibers," *IEEE Journal of Quantum Electronics*, vol. 25, no. 12, pp. 2665–2673, 1989.

[150] R. H. Stolen, J. P. Gordon, W. Tomlinson, and H. A. Haus, "Raman response function of silica-core fibers," *JOSA B*, vol. 6, no. 6, pp. 1159–1166, 1989.

[151] A. Sokolovskaya, A. Kudryavtseva, G. Brekhovskikh, and M. Sushchinskii, "Effect of temperature on stimulated raman scattering light in substances with various kerr constants," *Sov. Phys.-JETP.*, vol. 30, pp. 633–636, 1969.

[152] A. Sokolovskaya, A. Kudryatseva, T. Zhbanova, and M. SUSHCffiNSKil, "Investigation of stimulated raman scattering and self focusing of light in carbon disulfide," *SOVIET PHYSICS JETP*, vol. 26, no. 2, 1968.

[153] G. Rosman, "High-order comb spectrum from stimulated raman scattering in a silica-core fibre," *Optical and Quantum Electronics*, vol. 14, no. 1, pp. 92–93, 1982.

[154] A. Gomes, V. Da Silva, J. Taylor, B. Ainslie, and S. Craig, "Picosecond stimulated raman scattering in p2o5-sio2 based single mode optical fibre," *Optics communications*, vol. 64, no. 4, pp. 373–378, 1987.

[155] I. Alexandrov and Z. Nesterova, "Cascade and parametric stimulated raman scattering of picosecond light pulse in single-mode optical fiber," in *Conference Proceedings LEOS'96 9th Annual Meeting IEEE Lasers and Electro-Optics Society*, IEEE, vol. 1, 1996, pp. 225–226.

[156] P. French, A. Gomes, A. Gouveia-Neto, and J. Taylor, "Picosecond stimulated raman generation, pump pulse fragmentation, and fragment compression in single-mode optical fibers," *IEEE journal of quantum electronics*, vol. 22, no. 12, pp. 2230–2235, 1986.

[157] D. Koeplinger, M. Liu, and T. Buma, "Photoacoustic microscopy with a pulsed multi-color source based on stimulated raman scattering," in *2011 IEEE International Ultrasonics Symposium*, IEEE, 2011, pp. 296–299.

[158] A. K. Loya, J. Dumas, and T. Buma, "Photoacoustic microscopy with a tunable source based on cascaded stimulated raman scattering in a large-mode area photonic crystal fiber," in *2012 IEEE International Ultrasonics Symposium*, IEEE, 2012, pp. 1208–1211.

[159] P. Hajireza, A. Forbrich, and R. J. Zemp, "Multifocus optical-resolution photoacoustic microscopy using stimulated raman scattering and chromatic aberration," *Optics letters*, vol. 38, no. 15, pp. 2711–2713, 2013.

[160] S.-W. Cho *et al.*, "Optimal generation of ten individual green-to-red raman source for wavelength-dependent real-time or-pam images," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, pp. 1–9, 2018.

[161] F. Zhong, Y. Bao, R. Chen, Q. Zhou, and S. Hu, "High-speed wide-field multiparametric photoacoustic microscopy," *Optics letters*, vol. 45, no. 10, pp. 2756–2759, 2020.

[162] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature extraction*, Springer, 2006, pp. 1–25.

[163] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*. Springer, 2008, vol. 207.

[164] B. Ghojogh *et al.*, "Feature selection and feature extraction in pattern analysis: A literature review," *arXiv preprint arXiv:1905.02845*, 2019.

[165] X. Huang, L. Wu, and Y. Ye, "A review on dimensionality reduction techniques," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 10, p. 1 950 017, 2019.

[166] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

[167] S. Abbasi, K. Bell, and P. Haji Reza, “Rapid high-resolution mosaic acquisition for photoacoustic remote sensing,” *Sensors*, vol. 20, no. 4, p. 1027, 2020.

[168] V. Cizek, “Discrete hilbert transform,” *IEEE Transactions on Audio and Electroacoustics*, vol. 18, no. 4, pp. 340–343, 1970.

[169] R. El Guerjouma, A. Mouchtachi, Y. Jayet, and J. Baboux, “Nondestructive evaluation of graphite by ultrasonic velocity measurement using cross-correlation and hilbert transform methods,” in *IEEE 1992 Ultrasonics Symposium Proceedings*, IEEE, 1992, pp. 829–832.

[170] Y.-F. Chang and S.-C. Chen, “Imaging hilbert-transformed ultrasonic data,” *Research in nondestructive evaluation*, vol. 13, no. 2, pp. 97–104, 2001.

[171] Y. Kim, J. Park, and H. Kim, “Signal-processing framework for ultrasound compressed sensing data: Envelope detection and spectral analysis,” *Applied Sciences*, vol. 10, no. 19, p. 6956, 2020.

[172] Y. Watanabe, F. Sajima, T. Itagaki, K. Watanabe, and Y. Shuto, “High-speed linear detection time domain optical coherence tomography with reflective grating-generated spatial reference delay,” *Applied optics*, vol. 48, no. 18, pp. 3401–3406, 2009.

[173] M. Wojtkowski, V. J. Srinivasan, T. H. Ko, J. G. Fujimoto, A. Kowalczyk, and J. S. Duker, “Ultrahigh-resolution, high-speed, fourier domain optical coherence tomography and methods for dispersion compensation,” *Optics express*, vol. 12, no. 11, pp. 2404–2422, 2004.

[174] U. Haberland, P. Jansen, V. Blazek, and H. J. Schmitt, “Optical coherence tomography of scattering media using frequency-modulated continuous-wave techniques with tunable near-infrared laser,” in *Coherence Domain Optical Methods in Biomedical Science and Clinical Applications*, International Society for Optics and Photonics, vol. 2981, 1997, pp. 20–28.

[175] Y. Zhao, Z. Chen, Z. Ding, H. Ren, and J. S. Nelson, “Real-time phase-resolved functional optical coherence tomography by use of optical hilbert transformation,” *Optics letters*, vol. 27, no. 2, pp. 98–100, 2002.

[176] B. Rao, L. Li, K. Maslov, and L. Wang, “Hybrid-scanning optical-resolution photoacoustic microscopy for in vivo vasculature imaging,” *Optics letters*, vol. 35, no. 10, pp. 1521–1523, 2010.

[177] G. Ku, K. I. Maslov, L. Li, and L. V. Wang, “Photoacoustic microscopy with 2-μm transverse resolution,” *Journal of biomedical optics*, vol. 15, no. 2, p. 021 302, 2010.

[178] W. Bost, F. Stracke, M. Fournelle, and R. Lemor, "Developing a high-resolution photoacoustic microscopy platform," in *4th European Conference of the International Federation for Medical and Biological Engineering*, Springer, 2009, pp. 448–451.

[179] M. J. Moore *et al.*, "Photoacoustic f-mode imaging for scale specific contrast in biological systems," *Communications Physics*, vol. 2, no. 1, pp. 1–10, 2019.

[180] P. Kedarisetti *et al.*, "F-mode ultraviolet photoacoustic remote sensing for label-free virtual h&e histopathology using a single excitation wavelength," *Optics Letters*, vol. 46, no. 15, pp. 3500–3503, 2021.

[181] P. Kedarisetti, N. J. Haven, B. S. Restall, M. T. Martell, and R. J. Zemp, "Label-free virtual h&e histopathology using f-mode ultraviolet photoacoustic remote sensing," in *Photons Plus Ultrasound: Imaging and Sensing 2021*, International Society for Optics and Photonics, vol. 11642, 2021, 116422S.

[182] G. A. Seber and A. J. Lee, *Linear regression analysis.* John Wiley & Sons, 2012.

[183] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis.* John Wiley & Sons, 2021.

[184] Å. Björck, "Least squares methods," *Handbook of numerical analysis*, vol. 1, pp. 465–652, 1990.

[185] C. L. Lawson and R. J. Hanson, *Solving least squares problems.* SIAM, 1995.

[186] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.

[187] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.

[188] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical methods*, vol. 6, no. 9, pp. 2812–2831, 2014.

[189] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[190] B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021.

[191] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[192] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[193] D. L. Donoho and M. Gavish, *The optimal hard threshold for singular values is* $4/\sqrt{3}$, 2013.

[194] Z. Bai, K. P. Choi, and Y. Fujikoshi, "Consistency of aic and bic in estimating the number of significant components in high-dimensional principal component analysis," *The Annals of Statistics*, vol. 46, no. 3, pp. 1050–1076, 2018.

[195] M. Friendly, G. Monette, and J. Fox, "Elliptical insights: Understanding statistical methods through elliptical geometry," *Statistical Science*, vol. 28, no. 1, pp. 1–39, 2013.

[196] D. S. Bernstein, *Matrix mathematics*. Princeton university press, 2009.

[197] C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis*. Addison-Wesley, 1998.

[198] D. W. Harder, *Numerical analysis for engineering - ill-conditioned matrices*, Archived from the original., Jul. 2013. [Online]. Available: `https://web.archive.org/web/20130720042248/https://ece.uwaterloo.ca/~dwharder/NumericalAnalysis/04LinearAlgebra/illconditioned`.

[199] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016.

[200] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[201] J. A. Hartigan, *Clustering algorithms*. John Wiley & Sons, Inc., 1975.

[202] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[203] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on neural networks*, vol. 16, no. 3, pp. 645–678, 2005.

[204] R. Xu and D. C. Wunsch, "Clustering algorithms in biomedical research: A review," *IEEE reviews in biomedical engineering*, vol. 3, pp. 120–154, 2010.

[205] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, vol. 1, 1967, pp. 281–297.

[206] H. Steinhaus *et al.*, "Sur la division des corps matériels en parties," *Bull. Acad. Polon. Sci*, vol. 1, no. 804, p. 801, 1956.

[207] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[208] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[209] R. A. Horn and C. R. Johnson, "Norms for vectors and matrices," *Matrix analysis*, pp. 313–386, 1990.

[210] C. Ramirez, V. Kreinovich, and M. Argaez, "Why $l1$ is a good approximation to $l0$: A geometric explanation," *Journal of Uncertain Systems*, vol. 7, no. 3, pp. 203–207, 2013.

[211] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[212] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of Lp minimization," *Mathematical programming*, vol. 129, no. 2, pp. 285–299, 2011.

[213] N. N. Boustany, S. A. Boppart, and V. Backman, "Microscopic imaging and spectroscopy with scattered light," *Annual review of biomedical engineering*, vol. 12, pp. 285–314, 2010.

[214] O. M. Can and Y. Ülgen, "Estimation of free hemoglobin concentrations in blood bags by diffuse reflectance spectroscopy," *Journal of biomedical optics*, vol. 23, no. 12, p. 127 001, 2018.

[215] G. Zonios and A. Dimou, "Modeling diffuse reflectance from semi-infinite turbid media: Application to the study of skin optical properties," *Optics express*, vol. 14, no. 19, pp. 8661–8674, 2006.

[216] ——, "Modeling diffuse reflectance from homogeneous semi-infinite turbid media for biological tissue applications: A monte carlo study," *Biomedical optics express*, vol. 2, no. 12, pp. 3284–3294, 2011.

[217] C. Lin and R. Stolen, "New nanosecond continuum for excited-state spectroscopy," *Applied Physics Letters*, vol. 28, no. 4, pp. 216–218, 1976.

[218] G. Turin, "An introduction to matched filters," *IRE transactions on Information theory*, vol. 6, no. 3, pp. 311–329, 1960.

[219] D. Brennan, "On the maximum signal-to-noise ratio realizable from several noisy signals," *Proceedings of the Institute of Radio Engineers*, vol. 43, no. 10, pp. 1530–1530, 1955.

[220] R. Burdett, "Amplitude modulated signals: The lock-in amplifier," *Handbook of Measuring System Design*, 2005.

[221] N. Pellegrino *et al.*, "Towards improved imaging contrast with photoacoustic remote sensing through intelligent signal extraction," in *Photons Plus Ultrasound: Imaging and Sensing 2021*, International Society for Optics and Photonics, vol. 11642, 2021, 116421B.

[222] M. O'Searcoid, *Metric spaces.* Springer Science & Business Media, 2006.

[223] D. R. Kurepa, "Tableaux ramifiés d'ensembles. espaces pseudo-distanciés," *CR Acad. Sci. Paris*, vol. 198, pp. 1563–1565, 1934.

[224] A. Woodruff, *What is a neuron?* Aug. 2019. [Online]. Available: `https://qbi.uq.edu.au/brain/brain-anatomy/what-neuron`.

[225] L. L. Acosta, *Utep researchers attempt to map brain function*, 2014. [Online]. Available: `http://news.utep.edu/utep-researchers-attempt-to-map-brain-function/`.

[226] K. Tachi, T. Hirasawa, S. Okawa, A. Horiguchi, K. Ito, and M. Ishihara, "Chromatic-aberration-free multispectral optical-resolution photoacoustic microscopy using reflective optics and a supercontinuum light source," *Applied Optics*, vol. 60, no. 31, pp. 9651–9658, 2021.

[227] *Reflective microscope objectives.* [Online]. Available: `https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6933`.

[228] E. Hecht, *Optics.* Pearson Education, Incorporated, 2017, ISBN: 9780133977226.

[229] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[230] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[231] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning.* Cambridge University Press, 2020.

[232] J.-M. Liu, *Photonic devices.* Cambridge University Press, 2009.

[233] J. Shawe-Taylor, N. Cristianini, *et al.*, *Kernel methods for pattern analysis.* Cambridge university press, 2004.

[234] M. Welling, *A first encounter with machine learning*, 2011.

[235] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[236] G. McLachlan, “Discriminant analysis and statistical pattern recognition. wiley-interscience: Hoboken,” 2004.

[237] S. Marsland, *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC, 2011.