OPTO-PHYSIOLOGICAL MODELLING: ON LIGHT INTERACTION IN SKIN TISSUE

NATHAN GOLD

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO

NOVEMBER 2020

Abstract

The future of medical care is digital and mobile based, rooted in smartphone-based imaging and sensor technology. Essential to the development of these next-generation technologies is a rigorous understanding of the mathematical underpinnings of how light and skin tissue interact. In this dissertation we present a mathematical framework and multi-stage methodology to model the interaction of light in skin tissue and recover underlying physiological information from images and videos of skin tissue.

In the first portion of the dissertation, we construct the individual building blocks of our methodology. We review Gaussian Process inference as a tool for regression, study the radiative transport equation as a model of light-interaction in skin tissue, and reconstruct the in-camera imaging processing pipeline from smartphone digital cameras.

The second portion of the dissertation combines these components together into a multi-stage methodology to render skin pixel values in digital images as a function of the underlying pigment generating chromophores, and then recover an inverse chromophore map. We apply this map to both simulated and real skin images to determine the underlying chromophore concentration, and recover vasculature maps depending on the chromophore concentration.

In the final portion, we apply the inverse chromophore map to video frames of skin tissue for imaging diagnostics. We effectively recover the heart rate of a user with excellent accuracy. Finally, we propose a novel methodology to recover a user's electrocardiogram wavefrom from a facial blood flow signal, introducing a method for contactless imaging electrocardiography.

Acknowledgements

"*A PhD can only be done forwards but understood backwards.*"— Andrew Fleck
Paraphrasing Søren Kierkegaard

There is a saying,"*it takes a village...* ", for which a great many things are true; a PhD is no exception. A great number of people have helped me along the way to turn my vision into a reality.

First and foremost, I would like to express my most sincere gratitude to my supervisor, Professor Huaxiong Huang. Without his patience, care, belief, and ceaseless lessons, I would not be the mathematician that I am today. He taught me how to think about applied mathematics, how to solve problems, and to always remember that the journey and what you discover along the way are the reasons for solving a problem - the solution is secondary. From him I observed what true dedication to the craft looks like, and it is something that will be impressed upon me for the rest of my days.

To my committee members, I thank you for all of your support and guidance. Through Professor Steven X. Wang I learned to think about statistics, question everything I am told, and look for the right solution. To Professor Michael Haslam, thank you for believing in me, for being so careful a teacher of numerics, and always ensuring I kept efficiency in mind. To Professor Thomas Salisbury, I thank you for your support, your ability to pierce to the heart of the matter, and for answering any question I had. To my committee chair, Professor Peter Gibson, I thank you for being a superb instructor, for reminding me to always keep a level head, and for your support; our talks in your office and at coffee at Fields have meant a great deal to me. I thank my external examiner Professor Chun Liu for his excellent review and quick reading, and for his boundless energy. To my internal examiner, Professor Sheila Embleton, I thank you for your very detailed review of my work, and for a wonderful set of applications.

The friendships I have formed in graduate school will last a lifetime, and there are many I wish to thank. I want to thank Branislav Nikolič for taking me under his wing, for introducing me to French food, and for always knocking my head back on straight when it got twisted around. And thank you for having Maša and Luka around - they make up for you! To Allysa Lumley for being so great a friend, for being a better mathematician than I could ever hope to be, and for always putting a smile on my face. I want to thank Andrew Fleck for being a

true friend, for laughing ceaselessly, and for reminding me the Leafs still haven't won the cup since '67. To Vishal Siewnarine, I thank you for always making me laugh and for being so good at fluids. To Richard Li, thank you for dealing with me. To Masoud Ataei, I thank you for all the wonderful arguments. To Georgios Katsimpas and Andrea Vaccao, you guys made my life significantly better; a journey to Fields was never complete without you two.

To the members of the Fields CQAM Health Analytics and Multi-disciplinary Modelling Lab, I owe a great deal, both mathematically and personally. To Shixin Xu, for being a supportive friend, and a fabulous colleague; to Xiukun Zhao for working together so well; to Zilong Song for his wonderful joyful energy, and relentless drive to make sure things are correct; to Xiulei Cao and Kian-Chuan Ong for their kindness and mathematical discussions. To Neda Aminnejad, I thank you for your friendship, for working together on great problems, and for sharing lots of stories. To Yi Zhu, thank you for your friendship; I will look back on discussing mathematics with you with great fondness. To Professor Kang Lee, thank you for boundless inspiration, excellent insights, and for collecting such high-quality data. To Professor Arvind Gupta, thank you for being the smartest lab rat ever and for our lively discussions.

During the course of my PhD, I was lucky enough to have the opportunity to travel extensively and work with some of the finest mathematicians I know. I want to thank Professor Marc Thiriet for hosting me in Paris at Laboratoire Jacques-Louis Lions, UPMC, and gratefully acknowledge support from a Mitacs Fellowship. The months I spent in Paris were some of the finest of my life, and changed my view on the world; for this, I am eternally grateful. I learned a great deal about how to think about biology and mathematics from you, as well as about this most wonderful world around us. To Professor Chris Budd, thank you for hosting me at the University of Bath to escape visa issues, and for being a fabulous source of mathematical insight in our discussions. To Professor Shu Takagi, I thank you for hosting me at the University of Tokyo and demonstrating a work ethic that will be my inspiration for years to come. To Professor Samuel Drapeau, we became great friends through our numerous back and forth trips to Shanghai and Toronto, and I thank you for your selfless advice, your immense mathematical ability, and quite literally giving me a bed to sleep on. I would also like to thank Professors Rua Murray and Tim David for hosting me at the University of Canterbury in Christchurch, New Zealand for their hospitality, encouragement, and kindness. Additionally, I would like to thank Professor Dr. Martin G. Frasch for his guidance and advice through our many years of collaboration, for our stimulating scientific and worldly discussions, and for always believing in me. This work was also partially supported by a Natural Science and Engineering Research Council of Canada Postgraduate Doctoral Scholarship.

Finally, thank you to my family for your endless and unwavering support and love. To my parents, Avi and Marlene, there are no words that can describe the gratitude I owe to you for helping me arrive where I am today. You have always encouraged me and supported my interests, and I am proud to be your son. To my brother Alex, thank you for always being willing to drop everything for me - the feeling is mutual and it means more than I express. To Alla and Vladimir, I thank you for your friendship and warmth, and for teaching me

about a world I had never known. To Sasha and Masha, thank you for being such a great brother and sister-in-law duo, your support never failed, and you always cheered me up. To Nataly, you are the light that never goes out. You have stood by me through everything - long nights, early mornings, months away; I could never have accomplished this without you. All of my love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> "*The nurses wear scrubs, but the scrubs are very, very clean. The patients are elsewhere.*"
>
> — Arthur Allen
> *A Hospital Without Patients*

Clinical decision making, and indeed much of modern medicine, is a technology and data-driven endeavour. Diagnostics, disease progression prediction, clinical treatment planning, and medical research itself all rely upon technological advances, data sources, and data analytic methodologies to provide effective care and accurate findings. As technologies have advanced, so too has clinical decision making, leading to improved patient outcomes. Surgical techniques have advanced from large incision surgeries to minimally-invasive laproscopic procedures; precise imaging of the beating heart can be performed with ultrasound echocardiography; and visualisations of cranial blood flow can be captured with magnetic resonance imaging. These areas encompass only a select few avenues for which technology and data analytics have led to evolutions in patient diagnosis, treatments, and quality of care.

Despite these tremendous advances in diagnostic and interventional technologies, the medical field currently finds itself at the onset of a paradigm-shift in how care and monitoring will be provided to both current and future generations of patients, driven by the digital revolution (E. Topol, 2019). The vast majority of adverse medical events happen outside of clinical settings. One particular area of medicine for which this presents a significant challenge is cardiology. In Canada alone, heart disease is the second leading cause of death, after cancer, accounting for approximately 20% of deaths in 2012 (Canada, 2018). Globally, the situation is even more dire, with 8.9 million deaths from heart disease in 2015, representing 45% of all non-communicable disease deaths worldwide (Collaborators and Death, 2016), increasing to 17.9 million deaths in 2017 (Organization, 2017).

Driven by these demands, clinicians have sought to leverage networking and remote sensing

technology to monitor patients outside of clinical care settings, where it matters most (Walsh, E. J. Topol, and Steinhubl, 2014). Technologies pertaining to remote sensing and monitoring offer tremendous value for further insight as they are appealing to the general public and provide information that may otherwise be impossible to obtain in traditional care settings (Sun and Thakor, 2016). Additionally, such collected information may be used for early event detection and prediction (Adamson et al., 2014). The evolution from a single point measurement at a doctor's office or continuous measurement in clinical settings with cumbersome or invasive technology, to continuous non-invasive measurement at home or work, where patient events occur represents a transformative change in the way care may be provided.

The past decade has seen such a transformation take place with the introduction of numerous novel wireless devices for cardiac monitoring (Sana et al., 2020). The most exciting and accessible avenue for the development of patient monitoring outside of clinical settings lies in the smartphone. It is projected that there will be 3.8 billion smartphones in the world by 2021, increasing from the 3.5 billion devices present in 2020, each equipped with a powerful camera sensor for data collection (O'Dea, 2020). As such, it has been predicted that the next generation of accessible medical monitoring and diagnostic devices will be mobile-based, and image-driven (Kranjec et al., 2014; Harford et al., 2019; Hassan et al., 2020). Critical to the development of such next-generation camera-based smartphone diagnostic imaging is a theoretical understanding of the physical processes by which light interacts with participatory media such as skin tissue. Such an understanding is facilitated by mathematical modelling of these processes which will help ensure accurate data collection and robust performance under various operational conditions (Peris and Hu, 2007; Hu et al., 2009).

This dissertation is concerned with the development of a novel modern multi-stage monitoring methodology combining mathematical modelling, machine learning, and signal processing for imaging-based diagnostics and monitoring. The proposed system uses opto-physiological modelling to create a robust model of light-interaction in skin tissue, allowing the recovery of the skin's optical properties and underlying biophysical parameters from digital camera images of skin tissue. We then apply this methodology to video images to recover a subject's vital signs solely from the frames of a video.

The construction of this dissertation consists principally of two parts. The first part consists of Chapters 2 – 5, and provides an overview and implementations of the necessary material to construct the multi-stage monitoring methodology which is the topic of the dissertation. We begin by providing an overview of cardiac physiology and monitoring technology, as well as review the optical and physiological properties of skin tissue in Chapter 2. This introductory chapter provides the necessary background for an understanding of the physiological processes which determine the skin's appearance in a digital image. In Chapter 3 we review Gaussian Processes, a Bayesian inference procedure, and modern computational methods for them. We use Gaussian Processes as the inference engine to recover the optical parameters of skin tissue given an image of the skin. In Chapter 4 we introduce our opto-physiological model of light-interaction in skin tissue through the *radiative transport equation*. We construct a

layered model of skin tissue and compute the reflectance properties of the skin under realistic illumination conditions. In Chapter 5 we review digital camera imaging and present a camera imaging pipeline model for image formulation, as well as present a modification to an existing method in the literature to deduce a smartphone camera's imaging pipeline.

The second part of the dissertation delves into the construction of the novel multi-stage opto-physiological modelling and monitoring methodology for mathematically rigorous camera-based imaging diagnostics. In Chapter 6 we combine the models developed in Chapters 3, 4 and 5 to create a training set of rendered skin images using different biophysical parameters, known as chromophores, which is used to learn an inverse mapping from rendered image pixel values back to the biophysical parameter concentrations. We present results demonstrating the remarkable ability of our novel inverse mapping methodology to accurately recover biophysical parameter concentrations from simulated skin images using both our own rendering methodology and state-of-the-art Monte Carlo simulations from the literature. We then apply the inverse mapping technique to these simulated skin images to form a map of the embedded underlying blood vessels, which we simulate in multiple conditions, including straight, curved, connected, and at different depths. We achieve exceptional accuracy at recovering the positions of these embedded vessels, creating a novel map of the underlying vasculature of a sample of skin tissue. For a first test of the ability of the inverse mapping procedure to be used as an imaging diagnostic technology, we simulate blood flow through the embedded blood vessels of the rendered skin images in a variety of experiments and apply the map to recover the underlying blood flow variations. Finally, we apply the inverse map to images of real skin tissue collected with a smartphone camera to recover the biophysical parameter concentrations, as well as form a vasculature map of the imaged skin.

In Chapter 7, we apply the inverse mapping technique to video frames of subjects and signal processing techniques for robust vital sign extraction through *imaging photoplethysmography*. We provide a rigorous framework for assessing imaging photoplethysmography techniques, and review how existing mehodologies in the literature may be described by this framework, as well as demonstrate how the inverse mapping technqiue we develop solves many of the existing problems in the literature through opto-physiological modelling. We apply our method to an experimentally collected dataset, and demonstrate the state-of-the-art performance of our method for non-contact heart rate estimation in comparison to the existing methods in the literature. As a final contribution, we present, for the first time, imaging electrocardiography, reproducing the electrocardiogram signal of a subject by examining their facial blood flow signal. Chapter 8 concludes the disseration providing suggestions for future work as well as our personal thoughts on the future of imaging photoplethysmography and opto-physiological modelling and their role in the next generation of medicine

We believe that this research will have significant impact on the development of mobile imaging-based diagnostic techniques through the use of principled opto-physiological modelling and machine learning techniques. Further, the introduction of imaging electrocardiography presents a novel diagnostic tool, which has the potential to bring electrocardiographic monitoring to the general population without the requirements for any specialised hardware.

Chapter 2

Physiology, Appearance, and Monitoring

> "*Cardiology is essentially the study of plumbing and electricity. It's amazing how far we can get with that!*"
>
> — Brendan Mullen
> *Personal Communication*

The interaction of propagating light with participatory, or *turbid* media has been a subject of significant research effort for more than 150 years (Stokes, 1862). A turbid medium is a medium for which significant scattering of incident light occurs due to the highly irregular optical non-uniformities of the medium, resulting in considerably diminished light intensity as light passes through (Ryde, 1931). Human skin tissue is a highly scattering, layered, turbid medium, consisting of three or four layers of constituent tissue (Anderson and Parrish, 1981). The appearance of human skin is determined by a variety of factors, which are known as *optical properties*. These optical properties are defined by a large set of biophysical parameters known as *chromophores*, which provide a great deal of information about skin tissue, as well as the state of the underlying system (Steven L. Jacques, 2013).

The cardiovascular system is a dynamic network responsible for meetings the body's various demands for nutrients as delivered by blood (Mohrman and Heller, 2018). Each organ and peripheral tissue of the body is serviced by the cardiovascular system and its extensive vasculature network through hemodynamic interactions of blood flow around the body. Due to the dynamic nature of the cardiovascular system, a patient's *vital signs* are measured through outputs of the cardiovascular system.

In this chapter, we provide a review of the structure and appearance of skin tissue and cardiovascular anatomy and function. We also provide an overview of traditional vital sign

monitoring techniques, especially the *electrocardiogram* and *photoplethysmography* which are the subject of investigation in Chapter 7. In Section 2.1 we review the structure and optical properties of skin, including the absorption and scattering properties of skin tissue with respect to light. In Section 2.2 we give an overview of the anatomy and physiology of the cardiovascular system, as well as discuss the electrical activity of the heart. Section 2.3 discusses clinical monitoring technology.

2.1 Structure and Appearance of Human Skin

2.1.1 Structure

The skin is the largest organ of the body (Anderson and Parrish, 1981). In a 70kg individual, the skin weighs over 5kg and has a surface area of nearly 2m^2. The skin principally consists of three layers: a stratified, cellular *epidermis*; an underlying *dermis* of connective tissues; and a layer of subcutaneous fat, known as the *hypodermis*, in which major blood vessels are located, separated from the rest of the body by a layer of striated muscle. Each macro layer of skin has its own highly differentiated microstructure (McGrath and Uitto, 2016). A labelled cross section of human skin with many constituent parts is shown in Figure 2.1.

The outermost layer of the epidermis is the *stratum corneum*, which acts as a defensive barrier for the skin (McGrath and Uitto, 2016). The thickness of the stratum corneum varies on different parts of the body, being thickest on the palms and soles of the feet, and thinner elsewhere. The typical thickness of the stratum corneum is about 0.001cm (Russell, Wiedersberg, and Delgado-charro, 2009). The typical thickness of the epidermis is 0.021cm, with the typical thickness of the dermis at 0.2cm (Anderson and Parrish, 1981). The thickness of the hypodermis ranges from 0.19 to 0.7cm (Hwang, Hun, and D. J. Kim, 2015).

2.1.2 Absorption

The absorption of light in skin tissue refers to an attenuation of light intensity. Within the visible region of light, absorption is dominated by the chromophores *melanin* and *hemoglobin* (Anderson and Parrish, 1981; Lister, Wright, and Chappell, 2012; Steven L. Jacques, 2013).

Hemoglobin is a protein in red blood cells responsible for oxygen delivery to tissue, and is the principal absorber of light in the dermis. Hemoglobin exists in both the oxygenated state, HbO_2, and deoxygenated state, Hb, depending on its state in the circulation circuit. The strength of absorption at a given wavelength is measured by the molar attenuation coefficient, $\varepsilon(\lambda)$, which has SI units, $\text{L}\cdot\text{mole}^{-1}\cdot\text{cm}^{-1}$, for wavelength λ. In Figure 2.2 we display the molecular extinction coefficient for HbO_2 and Hb with data compiled from Prahl, 1999. Note the strong absorption in the blue region of the spectrum below 450 nm, and then the additional peaks from 500nm to 600nm for both HbO_2 and Hb. This combination of absorption causes hemoglobin to appear red.

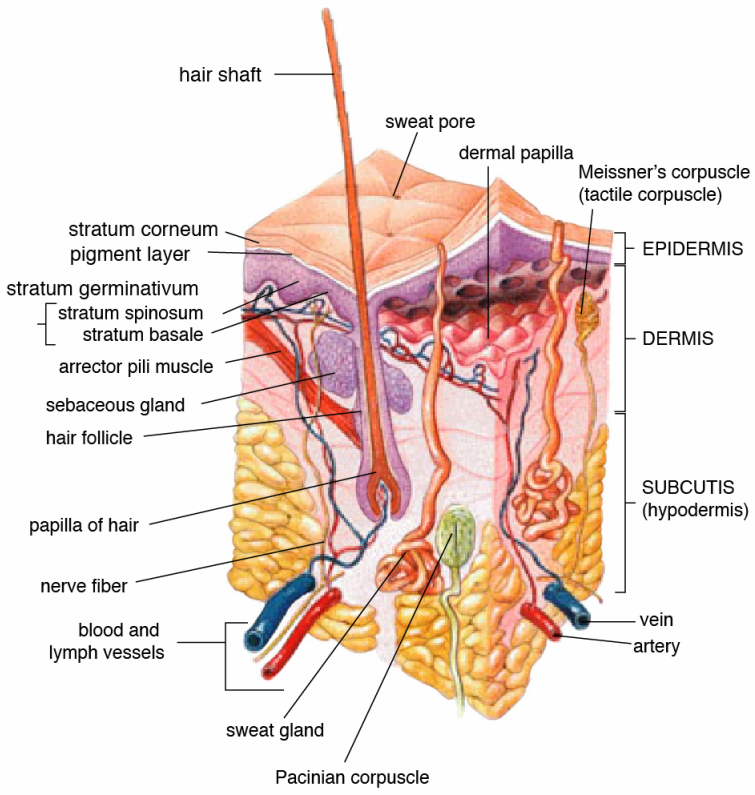

Figure 2.1: Labelled cross section of the layers of the skin. Courtesy of Institute, 2008.

Melanosomes are ordinarly contained in the epidermis. The absorption spectrum of melanin decreases as a function of wavelength, decreasing from the ultraviolet band (< 400nm) to the infrared band (> 700nm). Melanin is the predominant determinant of skin pigmentation, with greater concentrations of epidermal melanin leading to a darker skin appearance than lower concentrations. Melanin concentrations vary from 1.3% in light skin individuals all the way to 43% in dark skinned individuals (Lister, Wright, and Chappell, 2012; Steven L. Jacques, 2013). The two predominant types of melanosomes, are *eumelanin* and *pheomelanin*. Eumelanin is responsible for the dark appearance of skin in higher concentrations, while pheomelanin is responsible for yellow and pink skin pigments. Figure 2.3 displays the molecular extinction coefficients $\varepsilon(\lambda)$ for eumelanin and pheomelanin as a function of wavelength with data compiled from Steven L. Jacques, 1998.

2.1.3 Scattering

Scattering of light also provides a significant contribution to skin's appearance. Scattering describes a change of direction of the photon path of light in the tissue, and occurs due to the interaction of light with molecules in the skin before absorption events. The principal molecules responsible for particulate scattering in the epidermis are β-carotene in the epi-

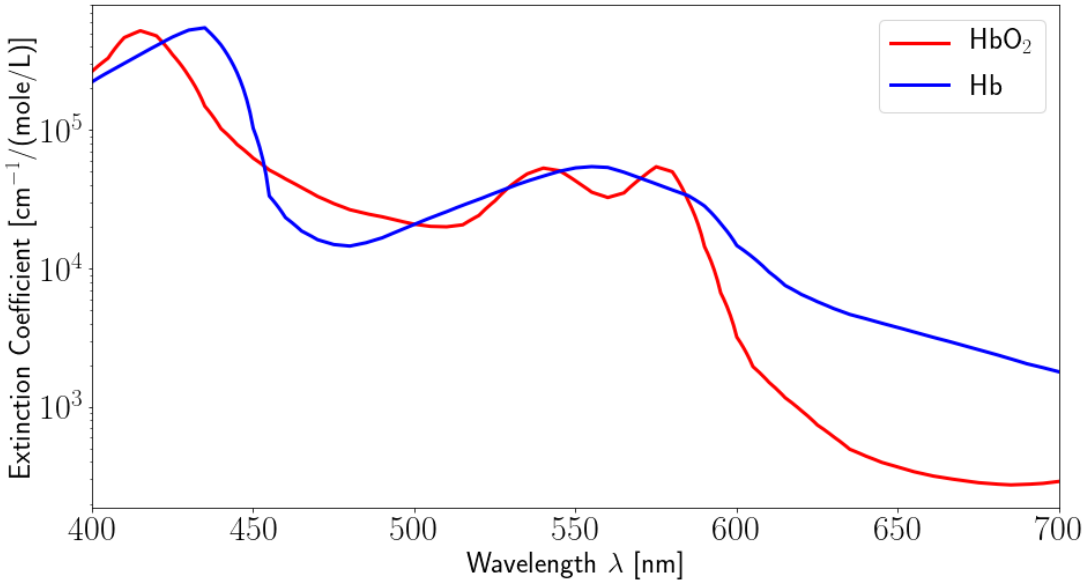

Figure 2.2: Molecular extinction coefficients $\varepsilon(\lambda)$ of oxygenated hemoglobin HbO_2 and deoxygenated hemoglobin Hb in the visible spectrum from 400nm to 700nm. Data compiled from Prahl, 1999.

dermis and collagen in the dermis (Lister, Wright, and Chappell, 2012). Empirical evidence suggests that melanosomes also contribute significantly to the degree of scatter within the epidermis (Bashkatov et al., 2000; Salomatina et al., 2006). In the dermis, blood as well as blood vessels also contribute to scattering events, and play a role in determining the colour of skin. Skin tissue is *strongly* forward scattering (Cheong, Prahl, and Welch, 1990), a fact which we will use in Chapter 4 for modelling light-interaction in skin tissue.

2.2 Cardiovascular System

2.2.1 Anatomy and Physiology

The cardiovascular system responds to the dynamic nutrient demands of the body's organs and periphera by providing variable blood flow. Due to the size of the human body (~ 1 – 2m) and its constituent organs ($\sim 10^{-1}$m), the demand for nutrient transport cannot be met by diffusion. Diffusion operates on a $t \propto x^2$ scale, where t is time, and x is distance. For example, it would take 15.4 hours for nutrients to diffuse over the 1cm thickness of the left ventricle wall (Herring and Paterson, 2018). For this reason, convective transport of nutrients via blood, pumped through the body by the heart, and circulated about the body via the vast vasculature network is required.

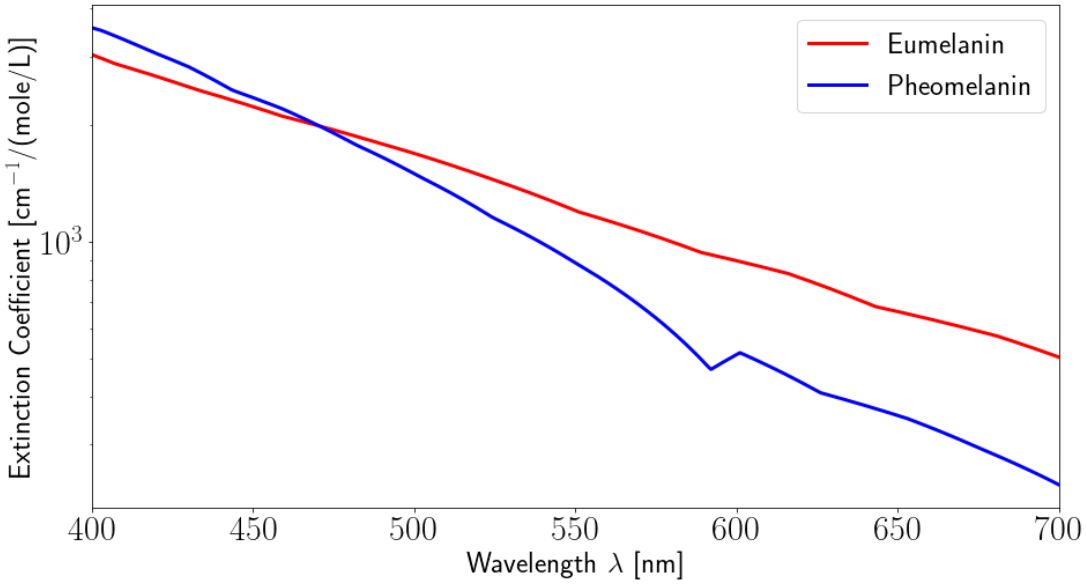

Figure 2.3: Molecular extinction coefficients $\varepsilon(\lambda)$ of eumelanin and pheomelanin in the visible spectrum from 400nm to 700nm. Data compiled from Steven L. Jacques, 1998.

The heart is the cardiac pump component of the cardiovascular system consisting of two synchronized pumps, divided into four chambers, the upper left and right atria, and the lower left and right ventricle, respectively. The left and right components of the heart act as a combined synchronous pump, alternating between cardiac relaxation and contraction, to move blood throughout the body. Cardiac relaxation is known as *diastole*, and cardiac contraction as *systole*. The right side comprised of the right atrium and ventricle pump blood throughout the low-pressure, short pulmonary circulation. The left ventricle propels an equal volume of oxygenated blood to the tissue and organs of the body. Deoxygenated blood is then returned via the *great veins* to the right atrium, which comprises the high-pressure, long systemic circulation (Thiriet, 2008b). The one-way, circular pathway of blood flow in the body was first demonstrated by William Harvey in 1628 (Harvey, 1628). See Figure 2.4 for an anatomical diagram of the heart and great blood vessels.

The volume of blood ejected by one ventricle per unit of time is termed the *cardiac output* (CO). Cardiac output is equal to the *stroke volume* (SV), the blood volume ejected by one ventricle per contraction, multiplied by the *cardiac frequency* or heart rate, (f_c). Stroke volume is defined as the difference between end-diastolic volume, the maximal volume achieved at the end of ventricular filling, and end systolic volume, the residual volume of the ventricle at the end of systolic ejection (Thiriet, 2008b). For a resting adult, the average stroke volume is typically in the range of 70 – 80 mL (Maceira et al., 2006) with a heart beat of 60 – 75 beats/min. This corresponds to a resting cardiac output of $\sim$ 5 – 6 L/min. Output can

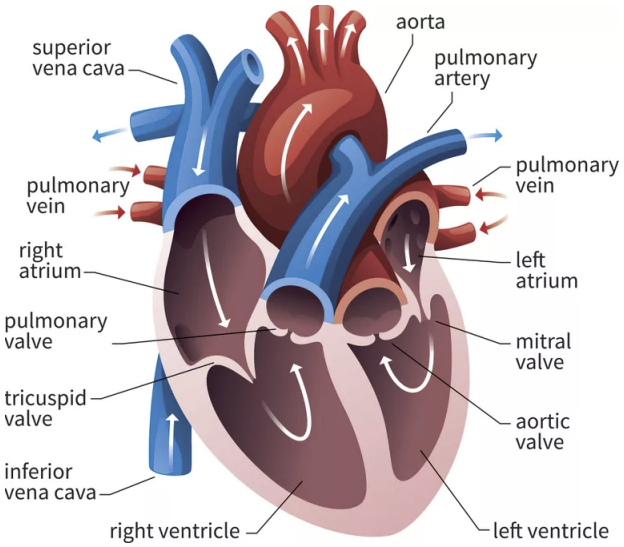

Figure 2.4: Anatomical representation of human heart. Oxygen poor blood flows from the systemic circulation through drainage veins, into the right atrium. Blood is then ejected from the right ventricle to the pulmonary artery, where it is delivered by the pulmonary circulation to the lungs. Oxygen rich blood is then returned to the left atrium via the pulmonary vein. Blood then passes through the mitral valve into the left ventricle, where it is ejected out of the aortic valve, into the aorta. Image courtesy of ThoughtCo.

increase by as much as four to five times during exercise to accomodate increased O_2 demand in the body (Herring and Paterson, 2018). We collect these values Table 2.2.1.

Table 2.1: Physiological quantities of cardiac output at rest in healthy subjects.

End-diastolic Volume	70 – 150 ml
End-systolic Volume	20 – 50 ml
Cardiac Frequency (f_c)	60 – 80 beats/min, 1 – 1.3 Hz
Blood flow rate (q)	4 – 7 L/min (70 - 120 mL / s)

2.2.2 Cardiac Cycle – Electrocardiography

The heart beat is a two stage pumping action consisting of diastole and systole called the *cardiac cycle*. Diastole is the longer phase of pumping, whose length is readily modulated to accomodate increased nutrient demand to various periphera of the body. The cardiac cycle consists of atrial and ventricular contractions. A typical cardiac cycle occurs for ~ 1 s, corresponding to 60 beats/min. We now briefly define phases of the cardiac cycle with respect to the electrocardiogram (ECG) which records the electrical activity of the heart (Cooper, 1986; Fye, 1994).

The ECG was first introduced in 1887 by Augustus Waller (Waller, 1887) and studied

with significant effort by Willem Einthoven (Einthoven, 1903) for which he was awarded the Nobel Prize in 1924. The ECG records the electrical activity of the heart through electrodes known as *leads* placed on specific locations of the body and is a representation of the electrophysiological activity of ion transport during the cardiac cycle (Keener and Sneyd, 2009; Thiriet, 2008a). The electrical activity generates a characteristic periodic waveform with peaks and troughs called the P, Q, R, S, and T waves, each of which corresponds to a particular phase of the cardiac cycle. Morphological characteristics of the ECG waveform including the time between waves and their shapes, are used for diagnostic purposes to detect cardiovascular pathophysiologies such as cardiac arrhythmias like atrial fibrillation, and acute myocardial infarction (De Bacquer et al., 1998; Herring and Paterson, 2018). Typical clinical settings use 12-lead ECG monitors, however, 3-lead ECGs have become popular for relative ease of use and comparable accuracy to the 12-lead standard (Antonicelli et al., 2012).

The cardiac cycle begins with the spontaneous depolarisation of the *sinoatrial* (SA) node found in the wall of the right atrium of the heart. This spreads from cell to cell, creating a sweeping wave of depolarisation across the atria, leading to the ventricles. The depolarisation of the atria is marked by the P-wave of the ECG waveform. Ventricular depolarisation indicates the beginning of ventricular contraction, pumping blood into the systemic circulation, and is represented by the QRS complex of the ECG waveform. The prominant R-wave signals ventricular contraction during the cardiac cycle. Once ventricular contraction is complete, the ventricles are repolarised, which is represented by the T-wave (Herring and Paterson, 2018; Thiriet, 2008b). Figure 2.5 displays the annotated ECG waveform obtained from a 3-lead ECG collected from the Author.

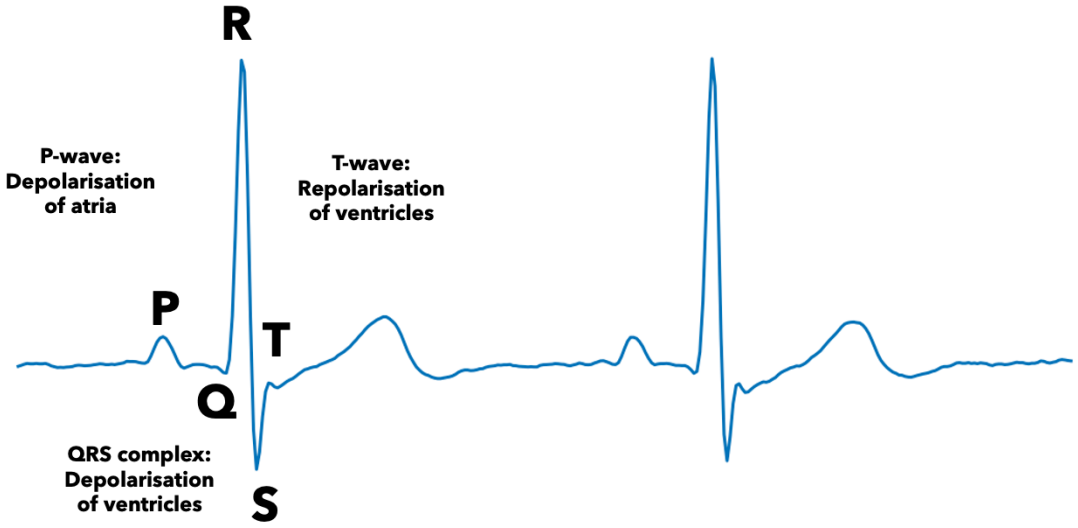

Figure 2.5: ECG waveform obtained from the author with annotated waves obtained from a 3-lead ECG.

Ventricular depolarisation is sudden and powerful, and manifests as the largest and most prominant feature in the ECG waveform. Due to this, the QRS complex, and specifically the R-wave is used as the temporal marker for calculating heart rate.

As mentioned above, abnormalities in the ECG waveform characteristics are represntative of cardiovascular pathophysiologies. Clinicians are trained to recognise such abnormalities for diagnosis. Acute myocardial infarction is typically precipitated by a raised segment between the QRS complex and the T-wave, in a condition known as *ST elevation*. Recently, there has been significant interest in designing automated detection methodologies for ECG abnormalities using deep learning practices (Hannun et al., 2019; Ribeiro et al., 2020) – see Gao, 2019 for a recent review and comments on these methodologies.

2.3 Monitoring – Photoplethysmography

Photoplethysmography (PPG) is a non-invasive circulatory signal related to the pulsatile blood volume in a segment of tissue (Resiner et al., 2008). It is related to *plethysmography*, the measurement of pulsatile tissue volume via measuring the sum total of volume changes in blood vessels. PPG is an optical technique, utilizing the reflective and transmitive properties of human skin and subcutaneous tissue to determine pulsatile blood volume (Moraes et al., 2018). First introduced by Alrick Hertzman in 1937 and 1938 (Hertzman and Spealman, 1937; Hertzman, 1938), photoplethysmography has become a mainstay of clinical monitoring, most commonly used for *pulse oximetry* to measure blood oxygen concentration (Sinex, 1999).

PPG performs optical monitoring by measuring changes to blood volume in subdermal microvascular tissue occuring due to pulsatile blood flow changes. As blood flow in the vascular network is pulsatile in nature due to the cardiac cycle of blood ejection from the heart (and respective blood return), heart rate can be accurately detected from PPG using blood volume variations. A photosensor is coupled with a photo-emitter, typically a light-emitting diode (LED), which pulses to send light through a segment of tissue. The *pulse wave* – pulsatile signal of blood volume – is detected by the photodiode as a voltage perturbation to the original signal. A pulse of blood causes an increase in the optical density and path length through the illuminated tissue. Pulsatile blood flow causes a momentary increase in the density of erythrocytes in the vascular tissue, resulting in a local increase in light absorbing heamoglobin (Hb) (Resiner et al., 2008; E. D. Chan, Michael M. Chan, and Mallory M. Chan, 2013). This perturbation manifests as a decrease in PPG voltage signal, which is inverted in practice so that it correlates to the increased blood volume. Figure 2.6 displays the PPG waveform collected from a finger based PPG device obtained from the Author.

The PPG signal is typically collected in two ways, *transmission* PPG, where the photosensor and photo-emitter are on opposite side of the tissue, and *reflectance* PPG, where the photosensor and photo-emitter are on the same side of the tissue (Allen, 2007; Resiner et al., 2008). Transmission PPG devices provide high quality signals, but are limited in where they

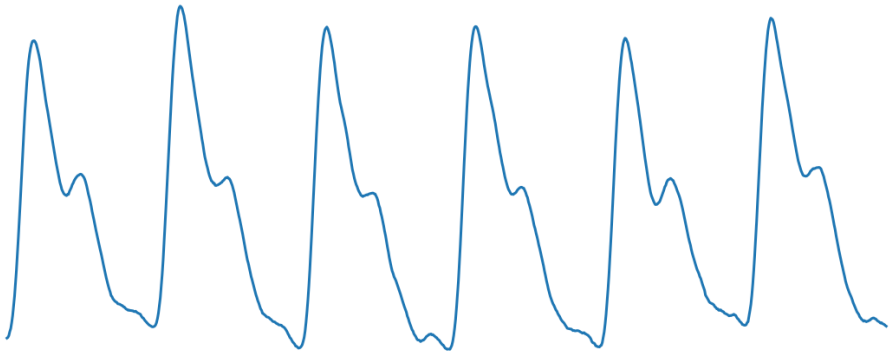

Figure 2.6: Photoplethysmogram (PPG) obtained from a finger based reflectance PPG device. The PPG waveform displays the periodic characteristic of pulsatile blood flow, with the valley of each cycle representing the diastolic phase of the cardic cycle, and the peak the systolic phase.

may be placed on the body (Tamura et al., 2014). Transmission PPG devices are commonly attached to the finger, earlobe, or toe. Reflectance PPG devices offer increased flexibility for placement, but may suffer from motion artifacts (Tamura et al., 2014). Figure 2.7 displays a schematic representation of both transmission and reflectance PPG devices.

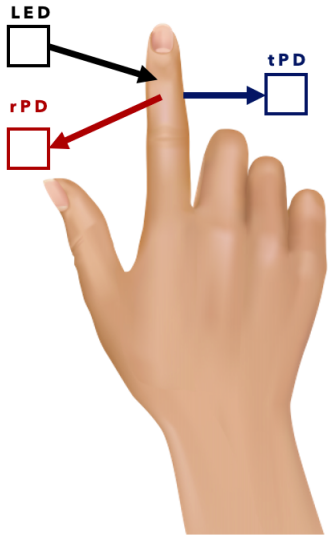

Figure 2.7: Representation of both transmission and reflectance PPG devices. An LED emits light into skin tissue which is then collected by either a reflectance photodiode (rPD) or a transmission photodiode (tPD).

The pulsatile component of the PPG waveform is called the *alternating current* (AC) component, whose fundamental frequency is the heart rate (Allen, 2007). The AC component is superimposed onto a large baseline *direct current* (DC) component that relates to tissues and average blood volume. The DC component itself slowly varies due to respiration, vasomotor activity, thermoregulation, and many other factors (Allen, 2007). Figure 2.8 displays a cross-section of an artery and vein during systole and diastole, demonstrating the pulsatile AC component and non-pulsatile DC components of the arteries, and the absence of such volume changes in veins.

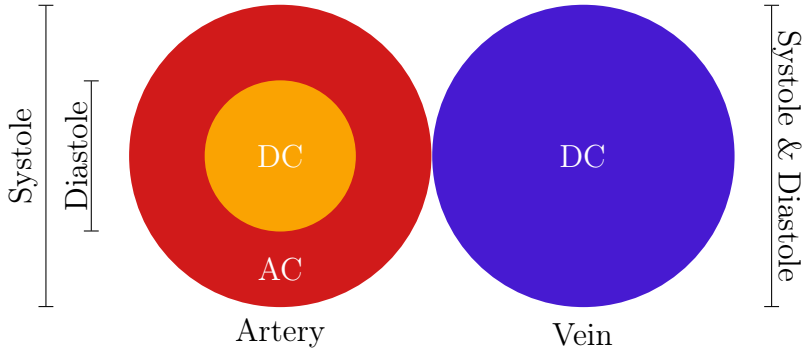

Figure 2.8: Arterial and venous cross sectional diagram displaying pulsatile (AC) and non-pulsatile (DC) compartments of the blood vessel. Only the artery contains a pulsatile component.

2.3.1 Bouguer-Beer-Lambert Law

To describe the principal of photoplethysmography we introduce the first, and simplest, model of light interaction in a turbid medium, the Bouguer-Beer-Lambert Law (Bouguer, 1729; Lambert, 1760; Beer, 1852).

The Bouguer-Beer-Lambert law relates the attenuation of light to the material properties of the material through which it is propagating. The total power extracted by absorption or scattering by a unit volume is described by the *attenuation* coefficient, μ_t, defined as

$$\mu_t = \mu_a + \mu_s \tag{2.3.1}$$

where μ_a and μ_s are the *absorption* and *scattering* coefficients, measured in cm^{-1}, defined as the ratio between the power absorbed or scattered in a unit volume and the power incident upon a unit area. The fraction of total power extracted by scattering is given by the *single-scattering albedo*, Λ,

$$\Lambda = \frac{\mu_s}{\mu_t} = \frac{\mu_s}{\mu_a + \mu_s}. \tag{2.3.2}$$

Consider a beam of light entering a material, and define the z axis parallel to the beam. The change of power passing through a volume element is given by,

$$\frac{dP}{dz} = -\mu_t(z)P \tag{2.3.3}$$

a first order linear equation. Integrating (2.3.3), we obtain the Bouguer-Beer-Lambert law

$$P(z) = P_e \exp\left\{-\int_0^z \mu_t(s)\, ds\right\} \tag{2.3.4}$$

for $P_e = P(0)$, the power emitted at $z = 0$. Thus, the power of a beam of light decays exponentially as it moves through the participatory medium. The exponential term is known as the *optical thickness* $\tau(z)$ of the medium. For a homogeneous medium $\tau(z) = \mu_t(z)$ (Ishimaru, 1978; L. V. Wang and H.-i. Wu, 2007).

In a vessel, the Bouguer-Beer-Lambert law gives the absorbance as $A = \epsilon\ell c$ where ϵ is the extinction coefficient of hemoglobin at a specfied wavelength (a combination of the HbO_2 and Hb coefficients), ℓ is the path length of light through the blood vessel, and c is the hemoglobin concentration (Allen, 2007). The AC component of the PPG waveform is represented as changes in absorbance over time. Let A_t be the total absorbance of light defined as the linear combination of arterial and venous absorbances, A_a, A_v, respectively. That is,

$$A_t = A_a + A_v = \epsilon_a \ell_a c_a + \epsilon_v \ell_v c_v \tag{2.3.5}$$

where ϵ_a, ℓ_a, c_a and ϵ_v, ℓ_v, c_v represent the extinction coefficient, path length, and hemoglobin concentrations in the arteries and veins, respectively.

Differentiating and assuming that extinction and concentrations are held fixed, we find

$$\frac{dA_t}{dt} = \epsilon_a c_a \frac{d\ell_a}{dt} + \epsilon_v c_v \frac{d\ell_v}{dt} \tag{2.3.6}$$

.

As the arteries dilate and constrict significantly more than the veins during the cardiac cycle – see Figure 2.8 – we may assume $\frac{d\ell_a}{dt} >> \frac{d\ell_v}{dt}$, to a reasonable approximation, and thus, (2.3.6) reduces to

$$\frac{dA_t}{dt} = \epsilon_a c_a \frac{d\ell_a}{dt}. \tag{2.3.7}$$

That is, the measured change in total absorbance is dominated by the change in absorbance due to arterial blood. In Chapter 7 we will generalise this principal when we consider imaging based photoplethysmography.

Chapter 3

Gaussian Processes

> "*de Finetti predicted a paradigm shift to Bayesian Methods – in 50 years, post-2020.*"
>
> — Sharon McGrayne
> *The Theory That Would Not Die*

In this chapter, we provide an introduction to Gaussian Processes (GP), a Bayesian machine learning technique for probabilistic inference (Carl E. Rasmussen and Williams, 2006). Gaussian processes are flexible learning models which have achieved state-of-the-art performance for many applications. In comparison to other machine learning techniques, such as deep learning (LeCun, Bengio, and Hinton, 2015), GPs provide several attractive advantages including principled uncertainty representation, the ability to incorporate model priors with little expert advice, and rapid adaptation to any dataset size (Carl Edward Rasmussen and Ghahramani, 2001; Soleimani, Hensman, and Saria, 2018). For these reasons, we chose Gaussian Processes as the machine learning procedure which we use to approximate the inverse map between chromophore concentrations and pixel values in images of skin tissue in Chapter 6.

In Section 3.1 we provide an overview of Bayesian inference, constructing the building blocks for Gaussian Process inference procedures include the concept of prior and posterior probability distributions, and the likelihood. In Section 3.2 we introduce Gaussian Process regression to learn the functional form connecting an input set to observed outputs, and discuss training for parameter tuning. Finally in Section 3.3 we discuss scalable Gaussian Process inference techniques using modern parallel graphical processing unit (GPU) architectures to enable the use of Gaussian Processes for datasets of millions of data points.

3.1 Bayesian Inference

Bayesian inference, unlike traditional frequentist statistical analysis, is centred about the idea of assuming and incorporating prior knowledge and beliefs about the data generation process and functional dependencies that may be observed (Berger, 1985; Jaynes, 2003; Schölkopf and Smola, 2002). Suppose we observe a data generating process which produces a dataset $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ collected in $\mathcal{D} = \{x_1, y_1), \ldots, (x_N, y_N)\}$ consisting of N observations. A Bayesian inference procedure produces an inference about the functional relationship between x and y by combining our prior knowledge with observations about the process.

3.1.1 Prior Distributions

Let f be a hypothesis about the process mapping x to y. We typically have some vague ideas regarding the outcome we expect, be it that the function is reasonably smooth, some correlation must exist between the inputs, or some parametric form exists which we believe relates x and y. We encode this information in a distribution of hypothesis $p(f)$, called the *prior*. Formally, the prior is a probabilistic representation of our hypotheses about the mapping process.

Prior distributions often vary in complexity and form as a function of our knowledge of the hypothesis f mapping inputs x to y. If we know that f has a certain functional form, say that it may be written as a linear combination of a finite number of polynomials, we can prescribe a *parametric* prior on f and restrict our inference to recover a function of this form.

In some cases, we often know little more about f than that values $f(x_i)$ are correlated and distributed according to a multivariate Gaussian distribution with zero mean and a covariance matrix K. While this does provide additional structure for inference, this is a reasonably general assumption and covers a broad class of processes. In such a case, the prior distribution takes the form,

$$p(f) = (2\pi)^{m/2} \det(K)^{-1} \exp\left\{-\frac{1}{2} X^\top K^{-1} X\right\}. \tag{3.1.1}$$

We note here that we only assume knowledge about the function values $f(x_i)$ and have not specified a prior on f itself.

Even more generally, we may only have knowledge that f is more likely to be smooth and take small values than be rough with large variations. This gives rise to a *nonparametric* prior of the form

$$-\log p(f) = ||f||_2^2 + \left|\left|\frac{\partial f}{\partial x}\right|\right|_2^2 + c \tag{3.1.2}$$

for a constant c, where nonsmooth functions with large $||\partial_x f||_2^2$ are less likely to occur. Indeed, the nonparametric prior (3.1.2) gives rise to Gaussian Processes.

The definition of a proper prior distribution is a highly subjective task, greatly influencing the results of the inference procedure, and has been a topic of debate in the community for many years (Jeffreys, 1961; Finetti, 1972; Shafer, 1981; Good, 1983; Berger, 1985; Jaynes, 2003). However, the choice of model typically has a *significantly* greater impact on the answer than the choice of prior, and will typically be just as subjective (Berger, 1985). Our personal view on the matter can be best summarized by Box, 1980,

> "In the past, the need for probabilities expressing prior belief has often been thought of, not as a necessity for all scientific inference, but rather as a feature peculiar to Bayesian inference. This seems to come from the curious idea that an outright assumption does not count as a prior belief... I believe that it is impossible logically to distinguish between model assumptions and the prior distribution of the parameters."

3.1.2 Likelihood

Now that we have encoded our pre-conceived notions and hypotheses f into a prior distribution $p(f)$, we can formalize our knowledge of given information about the process which maps the inputs x to a value to which noise is added to yield the outcomes y. The information is encoded into a conditional probability distribution $p(y|x, f(x))$, which we assume to either be known, or have a known parametric form. The distribution $p(y|x, f(x))$ is called the *likelihood*.

The most common likelihood function is the Gaussian likelihood. In the case of independent and identically distributed (iid) generated data, the likelihood function factorizes as,

$$p(Y|X, f) = \prod_{i=1}^{m} p\left(y_i|x_i, f(x_i)\right). \tag{3.1.3}$$

If $p(y_i|x_i, f(x_i))$ is the true underlying distribution, then $p(Y|X, f)$ describes how likely it is that the sample X and Y was generated by f. Bayesian inference is centred on the reverse of this reasoning, and describes how likely it is that f explains the data X and Y.

3.1.3 The Posterior Distribution

We now perform inference by using *Bayes' Rule* to combine the likelihood and prior into the *posterior* distribution or the *evidence*. By the Laws of Conditional Probability, with X as the included set of parameters for clarity, we have,

$$p(Y, f|X) = p(Y|f, X)p(f). \tag{3.1.4}$$

Conditioning f on Y gives,

$$p(f|Y,X)p(Y) = p(Y,f|X) \tag{3.1.5}$$

where $p(f|Y,X)$ is the posterior distribution, which gives the probability of f being the underlying function mapping X to Y. Combining (3.1.4) and (3.1.5) yields Bayes' rule for the posterior,

$$p(f|X,Y) = \frac{p(Y|f,X)p(f)}{p(Y)}. \tag{3.1.6}$$

As f itself does not enter into $p(Y)$, it is convention to represent the posterior in terms of only the prior and the likelihood,

$$p(f|X,Y) \propto p(Y|f,X)p(f) \tag{3.1.7}$$

which is enough to determine which hypothesis f is more likely.

The denominator of (3.1.6), $p(Y)$ is called the *marginal likelihood*. It is computed by marginalising the numerator with respect to f, as

$$p(Y) = \int p(Y|f,X)p(f)\,df. \tag{3.1.8}$$

Thus, we can write the posterior in (3.1.6) as

$$p(f|X,Y) = \frac{p(Y|f,X)p(f)}{\int p(Y|f,X)p(f)\,df}. \tag{3.1.9}$$

Computing the integral for the marginal likelihood (3.1.8) is often analytically intractable, and must be approximated (MacKay, 2003; Murphy, 2012). The most common techniques to do so are Markov Chain Monte Carlo (MCMC) integration (Robert and Casella, 2004) and variational inference (Blei, Kucukelbir, and McAuliffe, 2017).

Given that we have the posterior, we can compute the *predictive distribution* for y given a new input x, as

$$p(y|X,Y,x) = \int p(y|f,x)p(f|X,Y)\,df \tag{3.1.10}$$

where $p(y|X,y,x)$ tells us what observation y we are most likely to make for new input x, given previous observations X,Y.

3.2 Gaussian Process Regression

We begin our exposition of Gaussian Processes with a formal definition as a stochastic process, and then proceed to study them as a tool to extend the notion of a random variable to that of a random function for regression.

Consider a $\mathbb{R}^d$-valued stochastic process $X_t = \{X_t, 0 \leq t \leq \infty\}$. X_t is a *Gaussian Process* (GP) if for any $k \in \mathbb{N}$, and $0 \leq t_1 \leq t_2 \leq \cdots \leq t_l \leq \infty$, the random vector $(X_{t_1}, \ldots, X_{t_k})$ has a joint Gaussian distribution. If the distribution of $(X_{t+t_1}, \ldots, X_{t+t_k})$ does not depend on t, we say that the process is *stationary* (Karatzas and Shreve, 1988).

The finite dimensional distributions of a Gaussian Process X_t are completely determined by its expectation vector and covariance matrix,

$$m(t) = \mathbb{E}[X_t], t \geq 0 \tag{3.2.1}$$

$$k(s,t) = \mathbb{E}\left[(X_s - m(s))(X_t - m(t))^\top\right], s, t \geq 0. \tag{3.2.2}$$

We then have that $(X_{t_1}, \ldots, X_{t_k}) \sim N(m(t), K)$ for covariance matrix K generated from $k(s,t)$.

It is typical to define a Gaussian Process using the index set as time. The most famous example of such a process is the *zero-mean* GP – defined by $m(t) = 0, t \geq 0$ – one dimensional Brownian motion with covariance function $k(s,t) = s \wedge t$ (Karatzas and Shreve, 1988; Revuz and Yor, 1991).

For our use of Gaussian Processes, we take a general view of the index set of the process as the set of all possible input values - a function space view (Carl E. Rasmussen and Williams, 2006). This allows us to use GPs as a prior distribution over functions $f(\boldsymbol{x}) \sim \mathcal{GP}(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$ where each function is a random variable indexed by $\boldsymbol{x} \in \mathbb{R}^d$, with *mean function* $\mu : \mathbb{R}^d \to \mathbb{R}$ and *covariance function* or *kernel* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

3.2.1 Covariance Functions

The kernel k is used to encode information regarding the covariance between inputs of the GP. It controls the properties of the likely functions under a Gaussian process $f(\boldsymbol{x})$. The specification of the kernel determines the inductive biases of a GP model, that is, whether we expect smooth functions, periodic functions, slowly varying functions, Brownian motion, etc. The kernel itself has parameters that may vary, which are called *hyperparameters*. The values of the hyperparameters are determined through a learning procedure, which we describe below. We begin by showing that the covariance function indeed gives rise to a positve definite covariance matrix.

Theorem 3.1. *The function $k(\boldsymbol{x}, \boldsymbol{x}')$ is well defined and symmetric, and the covariance matrix K is positive definite.*

Proof. By definition,

$$[\text{cov}(f(\boldsymbol{x}_1)\ldots,f(\boldsymbol{x}_d))]_{i,j} = \text{cov}(\boldsymbol{x}_i,\boldsymbol{x}_j)). \tag{3.2.3}$$

As such, K_{ij} is a function of only two arguments $\boldsymbol{x}_i, \boldsymbol{x}_j$, and thus $k(\boldsymbol{x},\boldsymbol{x}')$ is well-defined. Further, we have that k is symmetric by definition of the covariance.

Consider the quadratic form $\boldsymbol{y}^\top K\boldsymbol{y}$ for $\boldsymbol{y}\in\mathbb{R}^d$. For K to be positive definite, this quadratic form need be non-negative.

$$\boldsymbol{y}^\top K\boldsymbol{y} = \boldsymbol{y}^\top\left[\text{cov}(f(\boldsymbol{x}_i),f(\boldsymbol{x}_j)\right]\boldsymbol{y} = \text{Var}\left(\sum_{i=1}^{d} y_i f(x_i)\right) \geq 0. \tag{3.2.4}$$

□

Indeed, the converse to this theorem holds as well; that is, any kernel which satistifies these properties comes from a Gaussian process. It is in fact this property which permits us to generate kernels without having to prove the existence of the associated Gaussian process.

The most common choice of kernel in the machine learning community is the *squared-exponential* (SE) kernel,

$$k_{SE}(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2\exp\left\{-\frac{||\boldsymbol{x}-\boldsymbol{x}'||_2^2}{2\ell^2}\right\} + \sigma_n^2\delta_{\boldsymbol{x},\boldsymbol{x}'} \tag{3.2.5}$$

where ℓ is the *characteristic length scale* over which the function values change, σ_f^2 controls the scale of the functions, and σ_n^2 is the noise variance. The characteristic length scale describes the distance between inputs $\boldsymbol{x}$ in feature space. A larger ℓ describes slowly varying functions, while smaller values of ℓ produce functions with greater variability.

While it is the most common choice, it has been argued that the SE kernel produces functions which are too smooth to reflect natural phenomena (Stein, 1999). As a result, it has been suggested to use the *Matérn* kernel, which Stein, 1999 named for Matern, 1960,

$$K_\nu(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{||\boldsymbol{x}-\boldsymbol{x}'||_2}{\ell}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{||\boldsymbol{x}-\boldsymbol{x}'||_2}{\ell}\right) \tag{3.2.6}$$

with parameters $\nu,\ell>0$, $\Gamma(\cdot)$ is the Gamma function, and K_ν is a modified Bessel function of the second kind (Abramowitz and Stegun, 1965),

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x)-I_\nu(x)}{\sin(\nu\pi)} \tag{3.2.7}$$

where $I_\nu(x)$ is a modified Bessel function of the first kind,

$$I_\nu(x) = \sum_{m=0}^{\infty}\frac{1}{m!\Gamma(m+\nu+1)}\left(\frac{x}{2}\right)^{2m+\nu}. \tag{3.2.8}$$

The Matérn kernel simplifies dramatically for positive half-integer orders $\nu = p + \frac{1}{2}$, $p \in \mathbb{N}$, in which case the kernel is a product of an exponential function and p-th order polynomial,

$$k_{v=p+1/2}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left\{-\frac{\sqrt{2\nu}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2\right\} \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{n=0}^{\infty} \frac{(p+n)!}{n!(p-i)!} \left(\frac{\sqrt{8\nu}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2\right)^{p-n}. \tag{3.2.9}$$

The most common use cases in machine learning, as well as in this dissertation, are $\nu = 3/2$ and $\nu = 5/2$. In these cases, the Matérn kernels are given as,

$$k_{3/2}(\boldsymbol{x}, \boldsymbol{x}') = \left(1 + \frac{\sqrt{3}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2\right) \exp\left\{-\frac{\sqrt{3}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2\right\} \tag{3.2.10}$$

$$k_{5/2}(\boldsymbol{x}, \boldsymbol{x}') = \left(1 + \frac{\sqrt{5}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2 + \frac{5}{3\ell^2}||\boldsymbol{x} - \boldsymbol{x}'||_2^2\right) \exp\left\{-\frac{\sqrt{5}}{\ell}||\boldsymbol{x} - \boldsymbol{x}'||_2\right\}. \tag{3.2.11}$$

For $\nu = 1/2$ in one-dimension, Matérn kernel will generate an Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930) with covariance function $k(\boldsymbol{x}, \boldsymbol{x}') = \exp\{-||\boldsymbol{x} - \boldsymbol{x}'||_2/\ell\}$. For an extensive discussion of covariance functions, see Carl E. Rasmussen and Williams, 2006.

3.2.2 Predictions with Gaussian Processes

Let us now consider using GPs for regression. Suppose we have a training set of N observations $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$, with $X = \{\boldsymbol{x}_i\}_{i=1}^N$ where we want to determine the functional relationship between y and $\boldsymbol{x}$, modelled as $y = f(\boldsymbol{x})$. In practice, rather than observing $f_i = f(\boldsymbol{x}_i)$ directly, we observe a version of f_i corrupted by noise, which we model as $y = f(\boldsymbol{x}) + \epsilon$ for additive Gaussian noise $\epsilon \sim N(0, \sigma_n^2)$. Given a test set X_* of inputs of size $N_* \times D$, we want to predict the function outputs $\boldsymbol{f}_*$.

Let the prior on the regression function f be a GP,

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{3.2.12}$$

with mean function $m(\boldsymbol{x})$ and kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ defined as in (3.2.1) and (3.2.2). By definition for a finite set of points X, this process defines a joint Gaussian distribution on function values $\boldsymbol{f}$,

$$p(\boldsymbol{f}|X) = N(\boldsymbol{f}|\boldsymbol{\mu}, K) \tag{3.2.13}$$

where $\boldsymbol{\mu} = (m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x}_N))$ and $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

With the iid Gaussian noise ϵ, the prior on the noisy observations has covariance given by $\text{cov}(\boldsymbol{y}) = K_{XX} + \sigma_n^2 I = \tilde{K}_{XX}$, where $K_{X,X} = K(X, X) = k(X, X) \in \mathbb{R}^{N \times N}$. The joint density

of the observed data and the noise-free functions on the test points, $\boldsymbol{f}_*$ is a Gaussian,

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{pmatrix} = N\left(\boldsymbol{0}, \begin{pmatrix} K_{XX} & K_* \\ K_*^\top & K_{**} \end{pmatrix}\right) \tag{3.2.14}$$

where $K_* = k(X, X_*) \in \mathbb{R}^{N \times N_*}$, and $K_{**} = k(X_*, X_*) \in \mathbb{R}^{N_* \times N_*}$. The case with noise also fits the same formalism without noise, but requires a different kernel than we present here. Conditioning upon past observations by application of (B.2.6)–(B.2.10), we find the *predictive distribution* – the posterior – of $\boldsymbol{f}_*$, $p(\boldsymbol{f}_x|X_*, X, \boldsymbol{y})$ for Gaussian Process regression as

$$p(\boldsymbol{f}_*|X_*, X, \boldsymbol{y}) = N(\boldsymbol{f}_*|\boldsymbol{\mu}_*, \Sigma_*) \tag{3.2.15}$$

$$\boldsymbol{\mu}_* = K_*^\top \tilde{K}_{XX}^{-1} \boldsymbol{y} \tag{3.2.16}$$

$$\Sigma_* = K_{**} - K_*^\top \tilde{K}_{XX}^{-1} K_*. \tag{3.2.17}$$

For a single test input $\boldsymbol{x}_*$, (3.2.16) and (3.2.17) reduce to

$$p(f_*|\boldsymbol{x}_*, X, \boldsymbol{y}) = N(f_*|\bar{f}_x, \mathrm{Var}(f_*) \tag{3.2.18}$$

$$\bar{f}_* = \boldsymbol{k}_*^\top \tilde{K}_{XX}^{-1} \boldsymbol{k}_* \tag{3.2.19}$$

$$\mathrm{Var}(f_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^\top \tilde{K}_{XX}^{-1} \boldsymbol{k}_* \tag{3.2.20}$$

where $\boldsymbol{k}_* = k(X, \boldsymbol{x}_*)$ is the vector of covariances between the test input point $\boldsymbol{x}_*$ and the n training points.

In Figures 3.1 – 3.3 we demonstrate Gaussian Process regression with three different covariance functions. Each displays five functions sampled from a GP prior with a different kernel before and after conditioning on observed data (Left and Right, respectively, in each figure.) The mean function for each is denoted with a thick blue curve, while the sampled functions are displayed as well. The shaded region in each figure displays the uncertainty in the predictive distribution, plotting $\mathbb{E}[f(\boldsymbol{x})] \pm 2\mathrm{std}(f(\boldsymbol{x}))$. Note how the uncertainty grows when there is significant distance between observations, and shrinks between subsequent observations that are close together. Figure 3.1 displays GP regression with an SE kernel, with infinitely smooth sampled functions. Figure 3.2 displays GP regression with a Matérn 3/2 kernel (3.2.10). Note how the sampled functions are more rough, allowing for more variability in the expressed functions. Figure 3.3 displays GP regression with a Matérn 5/2 kernel (3.2.11), which is the suggested kernel choice to model natural phenomena (Stein, 1999; Carl E. Rasmussen and Williams, 2006).

3.2.3 Training a Gaussian Process

In the previous section, the Gaussian Proccess posterior predictive distribution was determined with fixed kernel parameters – the *hyperparameters* of the model. The functions that are sampled for GP regression prediction to form the mean (3.2.16) and covariance (3.2.17) of the

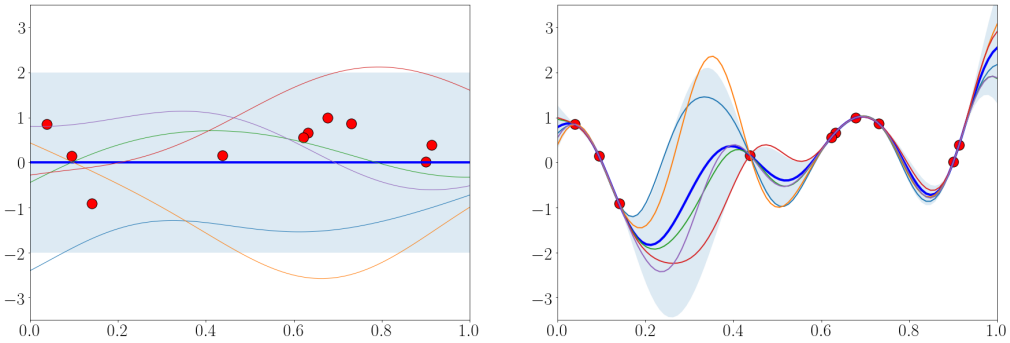

Figure 3.1: GP Regression with squared-exponential kernel. Left: Five functions sampled from a GP prior with an SE Kernel. Right: Five samples from a GP posterior, after conditioning on the 10 observations. The thick blue line represents the mean function value, and the red points represent observations. The shaded region represents $\mathbb{E}[f(\boldsymbol{x})] \pm 2\text{std}(f(\boldsymbol{x}))$

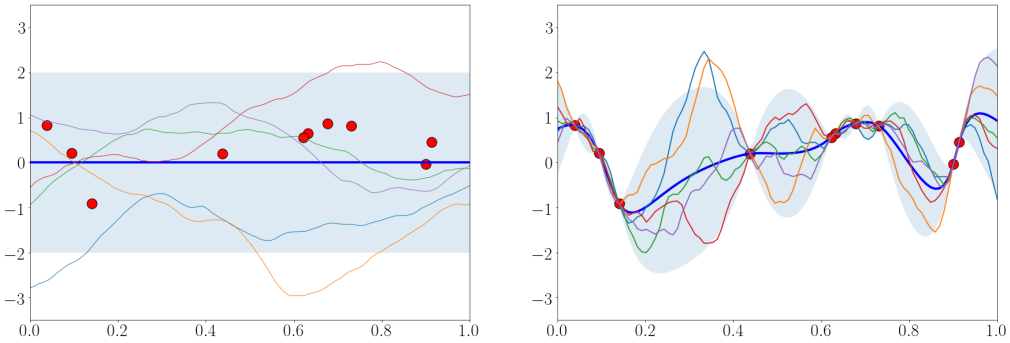

Figure 3.2: GP Regression with Matérn 3/2 kernel. Left: Five functions sampled from a GP prior with a Matérn 3/2 Kernel. Right: Five samples from a GP posterior, after conditioning on the 10 observations. The thick blue line represents the mean function value, and the red points represent observations. The shaded region represents $\mathbb{E}[f(\boldsymbol{x})] \pm 2\text{std}(f(\boldsymbol{x}))$

posterior distribution $p(\boldsymbol{f}_*|X_*, X, \boldsymbol{y})$ (3.2.15) depend on these hyperparameters to control the functions that are generated. By training a GP we mean tuning these model hyperparameters to improve performance.

The procedure by which a Gaussian Process is trained is through *Type-II Maximum Likelihood* (ML-II) estimation (Carl E. Rasmussen and Williams, 2006). ML-II proceeds by maximising the marginal likelihood of a model with respect to its hyperparameters, θ; that is, the marginal likelihood is the utility or loss function of the GP model.

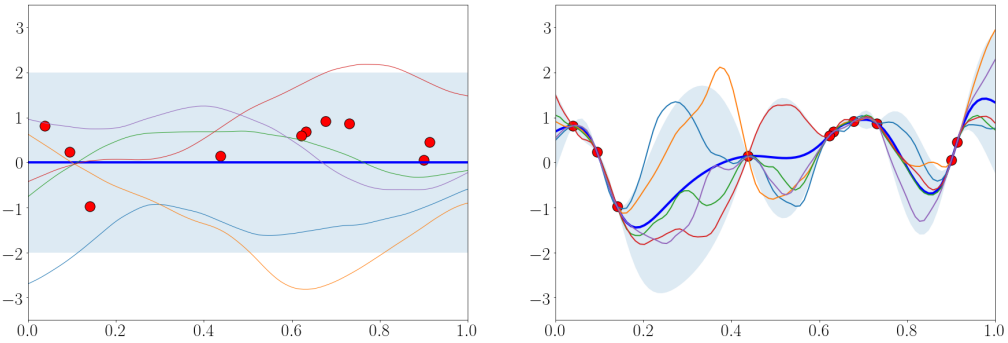

Figure 3.3: GP Regression with Matérn 5/2 kernel. Left: Five functions sampled from a GP prior with a Matérn 5/2 Kernel. Right: Five samples from a GP posterior, after conditioning on the 10 observations. The thick blue line represents the mean function value, and the red points represent observations. The shaded region represents $\mathbb{E}[f(\boldsymbol{x})] \pm 2\text{std}(f(\boldsymbol{x}))$

The *marginal likelihood* $p(\boldsymbol{y}|X)$ for a GP is given as in (3.1.8), where the prior is Gaussian $p(\boldsymbol{f}|X) = N(\boldsymbol{f}|\boldsymbol{0}, K)$,

$$\log p(\boldsymbol{f}|X) = -\frac{1}{2}\boldsymbol{f}^\top K^{-1}\boldsymbol{f} - \frac{1}{2}\log\det K - \frac{N}{2}\log 2\pi. \tag{3.2.21}$$

and the conditional likelihood given f is a factorized Gaussian $p(\boldsymbol{y}|\boldsymbol{f}) = N(\boldsymbol{y}|\boldsymbol{f}, \sigma_n^2 I)$. By (B.2.12) and (B.2.15), the integral can be computed directly to yield the *log marginal likelihood*,

$$\log p(\boldsymbol{y}|X) = \log \int p(\boldsymbol{y}|\boldsymbol{f}, X)p(\boldsymbol{f}|X)\, d\boldsymbol{f} \tag{3.2.22}$$

$$= -\frac{1}{2}\boldsymbol{y}^\top \tilde{K}_{XX}^{-1}\boldsymbol{y} - \frac{1}{2}\log\det \tilde{K}_{XX} - \frac{N}{2}\log 2\pi. \tag{3.2.23}$$

The marginal likelihood of a GP is the probability of data given only the kernel hyperparameters. This brings about an *automatic Occam's razor* (Carl Edward Rasmussen and Ghahramani, 2001) such that the simplest models which explain the data are automatically favoured without the need for cross-validation or regularisation. All of the covariance matrices formed the models depend implicitly upon the kernel hyperparameters θ.

Let us now write the marginal likelihood as a loss function.

$$\mathcal{L}(\theta) = \log p(\boldsymbol{y}|X, \theta) = \underbrace{-\frac{1}{2}\boldsymbol{y}^\top \tilde{K}_{XX}^{-1}\boldsymbol{y}}_{\text{model fit}} \underbrace{-\frac{1}{2}\log\det \tilde{K}_{XX}}_{\text{complexity penalty}} \underbrace{-\frac{N}{2}\log 2\pi}_{\text{normalisation}}. \tag{3.2.24}$$

Model fit decreases monotonically as a function of length scale, ℓ, as the model becomes less flexible and losses ability to adapt to variations in the data. Conversely, the negative

complexity penalty term increases with length scale, as the model becomess less complex as the length scale increases.

To determine the hyperparameters by maximising the log marginal likelihood, we take the gradient with respect to the hyperparameters.

$$\nabla_\theta \mathcal{L} \equiv \frac{\partial}{\partial \theta_j} \log p(\boldsymbol{y}|X,\theta) = \frac{1}{2}\boldsymbol{y}^\top \tilde{K}_{XX}^{-1} \frac{\partial \tilde{K}_{XX}}{\partial \theta_j} \tilde{K}_{XX}^{-1}\boldsymbol{y} - \frac{1}{2}\mathrm{tr}\left(\tilde{K}_{XX}^{-1}\frac{\partial \tilde{K}_{XX}}{\partial \theta_j}\right) \tag{3.2.25}$$

where we have used matrix derivative identities (B.1.1) for the first term, and (B.1.2) for the second, respectively. The hyperparameters are then learned through maximising the log marginal likelihood (3.2.24) through gradient descent. Since a typical GP has fewer hyperparameters to optimise in comparison to most parametric models, fewer iterations are required for training.

3.3 Scalable Gaussian Process Inference

To use a Gaussian Process inference procedure, we must compute the predictive distribution (3.2.15), the loss (3.2.24), and its gradient (3.2.25). Common to all of these equations are three operations which dominate the time complexity: (1) the linear solve $\tilde{K}_{XX}^{-1}\boldsymbol{y}$, (2) calculation of the log determinant $\log\det \tilde{K}_{XX}$, and (3) the computation of the trace term $\mathrm{tr}(\tilde{K}_{XX}^{-1}\nabla_\theta \tilde{K}_{XX})$.

Standard GP implementations perform these steps via a Cholesky decomposition (Golub and Van Loan, 2013) of the covariance matrix $\tilde{K}_{XX} = LL^\top$ for a lower triangular matrix L (Carl E. Rasmussen and Williams, 2006). The columns of $L = [\boldsymbol{l}^{(1)} \cdots \boldsymbol{l}^{(k)}]$ are computed recursively as

$$\boldsymbol{l}^{(i)} = \begin{pmatrix} \mathbf{0} \\ d_i \\ \frac{1}{d_i}\left(\boldsymbol{k}^{(i)}_{(i+1):n} - \sum_{j=i}^{i-1} l_i^{(j)} \boldsymbol{l}^{(j)\top}_{(i+1):n}\right) \end{pmatrix} \tag{3.3.1}$$

where $\boldsymbol{k}^{(i)}$ is the i-th column of $\tilde{K}_{XX}$ and $d_i = \left(k_i^{(i)} - \sum_{j=1}^{i-1} l_i^{(j)}\right)^{1/2}$.

As such, training of a GP takes $\mathcal{O}(n^3)$ operations, $\mathcal{O}(n^2)$ storage requirements, and $\mathcal{O}(n)$ and $\mathcal{O}(n^2)$ operations to compute the predictive mean $\boldsymbol{\mu}_*$ and covariance Σ_*, respectively, for each data point. This computational complexity has made GP methods prohibitively expensive to compute for datasets larger than $n = 10000$ points, and has been the bottleneck to widespread use of GPs in large scale data problems (Wilson, Dann, and Nickisch, 2015). Additionally, the recursive nature of the Cholesky decomposition is not easily parallelised, which prohibits extensive use of modern GPU acceleration.

A variety of approximation methods have been proposed to improve the scalability of Gaussian Process inference to larger datasets. Some of the first approximation methods

were *inducing point* methods (Quinonero-Candela and Carl E. Rasmussen, 2005; Snelson and Ghahramani, 2006; Titsias, 2009), which can be viewed as replacing the exact covariance matrix $\tilde{K}_{XX}$ with an approximation $\tilde{K}^*_{XX}$ evaluated at m inducing points, where $m < n$. Other approximations include random feature expansions (Rahimi and Recht, 2007; Le, Sarlos, and Smola, 2013) where the kernel is expressed in a random Fourier bases, or stochastic variational optimisation (Hensman, Fusi, and Lawrence, 2013; Salimbeni and Deisenroth, 2017).

Recently, there has been a shift in the community towards the exploration of the use of *Krylov subspace methods* using matrix-vector multiplications (MVMs) (Golub and Van Loan, 2013). While Krylov subspace methods have had a long history in the numerical analysis community, the machine leanring community has only much more recently begun to use them, to dramatic effect. Cunningham, Shenoy, and Sahani, 2008 and Murray, 2009 were some of the first authors to propose MVMs in GP inference with significant speedup. Wilson and Nickisch, 2015 used a Kronecker product structured kernel interpolation (SKI) matrix with fast MVMs to train a GP with 60,000 data points at less than 1% of the runtime cost of approximate inducing point methods with less than half the error!

In Chapter 6 we apply GP regression to approximate the inverse chromophore map from rendered skin pixels from a simulated training containing 65,536 skin pigments (256×256), which is a reasonably large data set for traditional GP inference. As such, we use the SKI approximation method of Wilson and Nickisch, 2015, which we will briefly describe here. SKI and other scalable GP inference procedures have been implemented in the GPyTorch Python package (Gardner et al., 2018), building upon the multipurpose automatic differentiation and neural network PyTorch Python package (Paszke et al., 2017).

The method begins with an inducing point method to approximate the exact kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ with an approximate kernel $\tilde{k}(\boldsymbol{x}, \boldsymbol{x}')$ for fast computations. Using the subset of regressors (SoR) (Silverman, 1985), the approximate kernel is

$$\tilde{k}_{SoR}(\boldsymbol{x}, \boldsymbol{x}') = K_{\boldsymbol{x},U} K_{U,U}^{-1} K_{U,\boldsymbol{x}'} \tag{3.3.2}$$

for m inducing points $U = \{\boldsymbol{u}\}_{i=1}^m$, $K_{\boldsymbol{x},U} \in \mathbb{R}^{1\times m}$, $K_{U,U}^{-1} \in \mathbb{R}^{m\times m}$, $K_{U,\boldsymbol{x}'} \in \mathbb{R}^{m\times 1}$ are the covariance matrices generated with the exact kernel $k(\boldsymbol{x}, \boldsymbol{x}')$. K_{SoR} is an $n \times n$ matrix of rank at most m.

Structured Kernel Interpolation (SKI) approximates the matrix of cross covariances for the kernel evaluated at the training points X and inducting points U, $K_{X,U} \in \mathbb{R}^{n\times m}$ by interpolating on the $K_{U,U} \in \mathbb{R}^{m\times m}$ to form,

$$K_{X,U} \approx W K_{U,U} \tag{3.3.3}$$

where $W \in \mathbb{R}^{n\times m}$ is a sparse matrix of interpolation weights. Wilson and Nickisch, 2015 use local cubic spline interpolation on equispaced grids, giving W 4 non-zero entries per row.

The SoR approximation to $K_{X,X}$ is

$$K_{X,X} \approx K_{X,U} K_{U,U}^{-1} K_{U,X} \approx W K_{U,U} K_{U,U}^{-1} K_{U,U} W^\top = W K_{U,U} W^\top = K_{\text{SKI}}. \tag{3.3.4}$$

Fast matrix vector products $K_{\text{SKI}}\boldsymbol{y}$ can be computed for efficient inference, costing $\mathcal{O}(n+m^2)$ computations and $\mathcal{O}(n+m^2)$ storage for W sparse. The matrix vector solves $K_{\text{SKI}}^{-1}\boldsymbol{y}$ can be computed with linear conjugate gradients with $j << n$ iterations for convergence, and $\log\det K_{\text{SKI}}$ can be computed using standard eigenvalue solvers, all of which are implemented in GPyTorch.

Chapter 4

Radiative Transport Equation

"*If taking the Taylor Series doesn't solve your problem, I don't know what will.*"

— Anonymous Graduate Student
A Blackboard Somewhere

The interaction of light with a scattering medium may be described by photon transport phenomena, which is a well-studied problem. In this chapter, we introduce the *radiative transport equation* (RTE), a linearised Boltzmann equation which describes the transport phenomena of photons in the skin. The RTE serves as the model of light-interaction in skin tissue in our multi-stage modelling methodology, and will be used to compute the reflectance of the skin which constitutes the image formulation process we describe in Chapter 6.

We begin in Section 4.1 by deriving the RTE in general terms from conservation laws to give a complete picture of the complexity of the equation. In Section (4.2) we write the RTE as a model for radiative transport in skin tissue and discuss the parameters of the equation. The RTE is a famously difficult partial-integral differential equation to solve, for which analytic solutions exist in only the very simplest geometry (Ishimaru, 1978). To accomodate this, we review the P_N approximations in Section 4.3, providing a complete derivation of the general infinite system of coupled partial differential equations by expanding the RTE in a spherical harmonic basis (Bal, 2009; Egger and Schlottbom, 2019). We then discuss in detail the diffusion approximation in Section 4.4, where we derive the mismatched refractive interface boundary conditions for the diffusion equation (A. D. Kim and Keller, 2003; Martelli et al., 2010). Finally in Section 4.5 we compute the diffusive reflectance for a model of layered skin media which will be used in Chapter 6 for skin image rendering as inspired by the Fourier tranform and Kubelka-Munk technique of (Donner and Jensen, 2005).

4.1 Derivation

The radiative transport equation may be derived in two ways, using either a physically motivated phenomenological approach satisfying conservation properties, or directly from Maxwell's equations of electromagnetisism. We show here, for completeness, a derivation of the RTE from the phenomenological perspective. An exhaustive derivation of the RTE directly from Maxwell's equations may be found in Mishchenko, 2006. We will mainly follow the derivation of the RTE due to Case and Zweifel, 1967 who employ conservation laws.

We begin with a general derivation of the equation describing photon transport phenoma. This analysis first appeared i nthe literature with respect to neutron scattering and transport and has been adapted to photon transport instead. Let $\phi(\boldsymbol{r}, \boldsymbol{v}, t)$ be the *angular particle density* in a space of positions $\boldsymbol{r}$ and velocities $\boldsymbol{v}$ as a function of time. Alternatively, ϕ represents the energy density of high-frequency waves in the space of positions and wavenumbers as a function of time (Bal, 2009). The spatial domain X is a convex, bounded, open subset of $\mathbb{R}^d$, $d \geq 2$ with C^1 boundary ∂X. The space of velocities V is a convex, bounded, open subset of $\mathbb{R}^d$ or the unit sphere $\mathbb{S}^{d-1}$.

Let $\phi(\boldsymbol{r}, \boldsymbol{v}, t)\, d\boldsymbol{r}\, d\boldsymbol{v}$ be the expected number of neutrons or particles in a volume element $d\boldsymbol{r}$ about $\boldsymbol{r}$ whose velocities lie in the element of velocity space $d\boldsymbol{v}$ about $\boldsymbol{v}$ at time t. We take the unit vector $\hat{\boldsymbol{s}} = \boldsymbol{v}/v$ for particle speed v, for which then lies in $\mathbb{S}^{d-1}$. As a result, $d\boldsymbol{v} = v^2\, dv\, d\boldsymbol{s}$. Define the *angular current*,

$$j(\boldsymbol{r}, \boldsymbol{v}, t) = \boldsymbol{v}\phi(\boldsymbol{r}, \boldsymbol{v}, t). \tag{4.1.1}$$

Then $j(\boldsymbol{r}, \boldsymbol{v}, t) \cdot \hat{\boldsymbol{n}}\, dS\, d\boldsymbol{v}\, dt$ is the number of particles in $d\boldsymbol{v}$ about $\boldsymbol{v}$ crossing area element dS with unit normal $\hat{\boldsymbol{n}}$ in time dt.

The *energy dependent density* or *photon fluence* in the case of photons is,

$$\Phi(\boldsymbol{r}, t) = \int_V \phi(\boldsymbol{r}, \boldsymbol{v}, t)\, d\boldsymbol{v}. \tag{4.1.2}$$

The *density* or *photon current/flux* is given by,

$$\boldsymbol{J}(\boldsymbol{r}, t) = \int_V \boldsymbol{v}\phi(\boldsymbol{r}, \boldsymbol{v}, t)\, d\boldsymbol{v}. \tag{4.1.3}$$

Both Φ and $\boldsymbol{J}$ constitute the *measurable* quantities of light propagation in participatory media. Both will be essential in deriving the reflectance used in Chapter 6 to produce rendering of skin.

Define $q(\boldsymbol{r}, \boldsymbol{v}, t)$ as the *angular source density*. Then $q(\boldsymbol{r}, \boldsymbol{v}, t)\, d\boldsymbol{r}\, d\boldsymbol{v}\, dt$ is the number of photons inserted into $d\boldsymbol{r}$ at $\boldsymbol{r}$ and $d\boldsymbol{v}$ at $\boldsymbol{v}$ over time dt.

As photons move through the volume V, they experience collisions with other particles. Define the *mean free path* between collisions for a photon of velocity $\boldsymbol{v}$ as $\ell(\boldsymbol{v})$. On average, a

photon of velocity $\boldsymbol{v}$ will suffer $\frac{v}{\ell}$ collisions per second, from which we can define the collision rate,

$$\frac{v}{\ell(\boldsymbol{r},\boldsymbol{v})}\phi(\boldsymbol{r},\boldsymbol{v},t)\,d\boldsymbol{r}\,d\boldsymbol{v}.$$

The inverse mean free path is the *macroscopic cross section*, $\sigma(\boldsymbol{r},\boldsymbol{v}) = \ell(\boldsymbol{r},\boldsymbol{v})^{-1}$, which is a weighted sum of the cross sections of all the photons and particles present. From this, we have that the scattering reaction rate for photons of velocity $\boldsymbol{v}$ at $\boldsymbol{r}$ is $v\phi(\boldsymbol{r},\boldsymbol{v},t)\sigma(\boldsymbol{r},\boldsymbol{v})\,d\boldsymbol{r}\,d\boldsymbol{v}$. When a collision occurs, $c(\boldsymbol{r},\boldsymbol{v})$ secondary photons will result. For an absorption event $c = 0$, and for scattering $c = 1$.

We define the scattering function $k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')$, where $v'k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\phi(\boldsymbol{r},\boldsymbol{v}',t)\,d\boldsymbol{v}'\,d\boldsymbol{r}\,d\boldsymbol{v}\,dt$ describes the probable number of photons in $d\boldsymbol{r}$ at $\boldsymbol{r}$ emitted/scattered into $d\boldsymbol{v}$ at $\boldsymbol{v}$ in time dt due to collisions taking place induced by photons of velocity in $d\boldsymbol{v}'$ at $\boldsymbol{v}'$. That is, if a collion occurs at time t, there is a probability of a photon of velocity $\boldsymbol{v}'$ being scattered into $\boldsymbol{v}$. We make the simplifying assumption that secondary photons appear instantaneously upon collision. By definition of the cross section,

$$\int_V k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\,d\boldsymbol{v} = c(\boldsymbol{r},\boldsymbol{v}')k(\boldsymbol{r},\boldsymbol{v}'). \tag{4.1.4}$$

We may now derive the radiative transport equation. Consider the change dN in time dt of the number of photons with velocity in $d\boldsymbol{v}$ about $\boldsymbol{v}$ in volume Ω with surface $\partial\Omega$ about $\boldsymbol{r}$. To ensure conservation, we have,

$$dN = d\boldsymbol{v}\,dt\int_\Omega \frac{\partial\phi(\boldsymbol{r},\boldsymbol{v},t)}{\partial t}\,d\boldsymbol{r} \tag{4.1.5}$$

and,

$$\begin{aligned} dN = &-(a)\text{ net neutrons flowing out of }\partial\Omega\text{ in }dt\\ &-(b)\text{ neutrons encountering collisions in }\Omega\text{ in }dt\\ &+(c)\text{ secondary neutrons of velocity }\boldsymbol{v}\text{ produced in }\Omega\text{ in }dt\text{ by collisions}\\ &+(d)\text{ neutrons of velocity }\boldsymbol{v}\text{ produced in }\Omega\text{ in }dt\text{ by sources} \end{aligned} \tag{4.1.6}$$

The net photons flowing out of $\partial\Omega$ in dt are described by,

$$(a) = d\boldsymbol{v}\,dt\int_{\partial\Omega} j(\boldsymbol{r},\boldsymbol{v},t)\cdot\hat{\boldsymbol{n}}\,dS = d\boldsymbol{v}\,dt\int_{\partial\Omega}\boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\cdot\hat{\boldsymbol{n}}\,dS \tag{4.1.7}$$

where $\hat{\boldsymbol{n}}$ is the unit outward normal to $\partial\Omega$.

Photons encountering collisions in Ω in time dt are described by,

$$(b) = d\boldsymbol{v}\,dt\int_\Omega \frac{v\phi(\boldsymbol{r},\boldsymbol{v},t)}{\ell(\boldsymbol{r},\boldsymbol{v})}\,d\boldsymbol{r}. \tag{4.1.8}$$

The secondary photons of velocity $\boldsymbol{v}$ produced in Ω in dt by collisions are,

$$(c) = d\boldsymbol{v}\, dt \int_{\Omega}\int_{V} v'\phi(\boldsymbol{r},\boldsymbol{v}',t)k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\, d\boldsymbol{v}'\, d\boldsymbol{r}. \tag{4.1.9}$$

Finally, the photons of velocity $\boldsymbol{v}$ produced in Ω in dt by sources are,

$$(d) = d\boldsymbol{v}\, dt \int_{\Omega} q(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r}. \tag{4.1.10}$$

Inserting (4.1.7) – (4.1.10) into (4.1.6), we have,

$$\begin{aligned} d\boldsymbol{v}\, dt \int_{\Omega} \frac{\partial\phi(\boldsymbol{r},\boldsymbol{v},t)}{\partial t}\, d\boldsymbol{r} &= -d\boldsymbol{v}\, dt \int_{\partial\Omega} \boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\cdot\hat{\boldsymbol{n}}\, dS - d\boldsymbol{v}\, dt \int_{\Omega} \frac{v\phi(\boldsymbol{r},\boldsymbol{v},t)}{\ell(\boldsymbol{r},\boldsymbol{v})}\, d\boldsymbol{r} \\ &\quad + d\boldsymbol{v}\, dt \int_{\Omega}\int_{V} v'\phi(\boldsymbol{r},\boldsymbol{v}',t)k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\, d\boldsymbol{v}'\, d\boldsymbol{r} \\ &\quad + d\boldsymbol{v}\, dt \int_{\Omega} q(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r}. \end{aligned} \tag{4.1.11}$$

Re-arranging the above we get,

$$\begin{aligned} &d\boldsymbol{v}\, dt \int_{\Omega} \frac{\partial\phi(\boldsymbol{r},\boldsymbol{v},t)}{\partial t}\, d\boldsymbol{r} + d\boldsymbol{v}\, dt \int_{\partial\Omega} \boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\cdot\hat{\boldsymbol{n}}\, dS + d\boldsymbol{v}\, dt \int_{\Omega} \frac{v\phi(\boldsymbol{r},\boldsymbol{v},t)}{\ell(\boldsymbol{r},\boldsymbol{v})}\, d\boldsymbol{r} \\ &= d\boldsymbol{v}\, dt \int_{\Omega}\int_{V} v'\phi(\boldsymbol{r},\boldsymbol{v}',t)k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\, d\boldsymbol{v}'\, d\boldsymbol{r} + d\boldsymbol{v}\, dt \int_{\Omega} q(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r}. \end{aligned} \tag{4.1.12}$$

By the divergence theorem, we have

$$\int_{\partial\Omega} \boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\cdot\hat{\boldsymbol{n}}\, dS = \int_{\Omega} \nabla\cdot\boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r},$$

and further, as $\boldsymbol{v}$ is independent of space,

$$\int_{\Omega} \nabla\cdot\boldsymbol{v}\phi(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r} = \int_{\Omega} \boldsymbol{v}\cdot\nabla\phi(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r}.$$

Combining all of this, we have,

$$\begin{aligned} &\int_{\Omega} \frac{\partial\phi(\boldsymbol{r},\boldsymbol{v},t)}{\partial t}\, d\boldsymbol{r} + \int_{\Omega} \boldsymbol{v}\cdot\nabla\phi(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r} + \int_{\Omega} \frac{v\phi(\boldsymbol{r},\boldsymbol{v},t)}{\ell(\boldsymbol{r},\boldsymbol{v})}\, d\boldsymbol{r} \\ &\qquad = \int_{\Omega}\int_{V} v'\phi(\boldsymbol{r},\boldsymbol{v}',t)k(\boldsymbol{r},\boldsymbol{v},\boldsymbol{v}')\, d\boldsymbol{v}'\, d\boldsymbol{r} + \int_{\Omega} q(\boldsymbol{r},\boldsymbol{v},t)\, d\boldsymbol{r} \end{aligned} \tag{4.1.13}$$

and since Ω was an arbitary volume, we finally arrive at the radiative transport equation in its general form as,

$$\frac{\partial \phi(\boldsymbol{r}, \boldsymbol{v}, t)}{\partial t} + \boldsymbol{v} \cdot \nabla \phi(\boldsymbol{r}, \boldsymbol{v}, t) + v\sigma(\boldsymbol{r}, \boldsymbol{v})\phi(\boldsymbol{r}, \boldsymbol{v}, t) = \int_V v' k(\boldsymbol{r}, \boldsymbol{v}, \boldsymbol{v}')\phi(\boldsymbol{r}, \boldsymbol{v}', t)\, d\boldsymbol{v}' + q(\boldsymbol{r}, \boldsymbol{v}, t) \tag{4.1.14}$$

where $\ell^{-1} = \sigma$ is the mean free path length and cross section.

4.2 Radiative Transport in Skin Tissue

Now that we have defined the radiative transport equation in general terms, we will proceed to write the radiative transport equation in the biologically relevant terms that pertains to light transport in skin tissue. We consider propagation in the unit sphere $\mathbb{S}^{d-1}$ replacing $\boldsymbol{v}$ with $\hat{\boldsymbol{s}}$ such that $||\hat{\boldsymbol{s}}||_2 = 1$. Additionally, the cross section terms are denoted with the transmission or attenuation coefficient $\mu_t(\boldsymbol{r})$ as given in (2.3.1). The scattering function k is assumed to be only a function of the angle between directions $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{s}}'$.

With these modifications, the radiative transport equation for photon propagation in skin tissue becomes

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \hat{\boldsymbol{s}} \cdot \nabla + \mu_t(\boldsymbol{r})\right) \phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \mu_s(\boldsymbol{r}) \int_{\mathbb{S}^{d-1}} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}')\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}', t)\, d\hat{\boldsymbol{s}} + q(\boldsymbol{r}, \hat{\boldsymbol{s}}, t). \tag{4.2.1}$$

where $\mu_s(\boldsymbol{r})$ is the scattering coefficient derived from the attenuation coefficient, and c is the speed of light subject to boundary and initial conditions,

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = 0, \quad \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}} < 0, \quad \boldsymbol{r} \in \partial V \tag{4.2.2}$$

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, 0) = 0. \tag{4.2.3}$$

The degree of anisotropy of scattering is measured by g, the mean value of the cosine of the scattering angle μ where $\cos\mu = \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}'$,

$$g = 2\pi \int_0^\pi (\cos\mu)\, k(\cos\mu) \sin\mu\, d\mu \tag{4.2.4}$$

where $g = 0$ represents isotropic scattering and $g = 1$ is entirely forward directed scattering. As mentioned in Section 2.1.3, biological tissue is highly forward scattering, with $g = 0.99$ for whole human blood, and $g > 0.9$ typically used for skin modelling (Cheong, Prahl, and Welch, 1990; A. D. Kim and Keller, 2003).

The radiative transport equation is notoriously difficult to solve, with analytic solutions existing for only the simplest cases (Fan, An, and Ying, 2019). Techniques for solving the RTE to simulate light propagation in skin tissue either involve Monte Carlo simulations (L.

Wang, S. L. Jacques, and Zheng, 1995; C. J. Zoller et al., 2018), or deterministic schemes based on differential-integral equations depending on different discretisations in the spatial and angular domains including Dedner and Vollmoller, 2002; Klose and Larsen, 2006; Frank et al., 2007; Godoy and Liu, 2012; Asllanaj et al., 2019 to name just a few of the many authors who have contributed to this field.

In the next sections, we will discuss the most commonly used approximation method for the RTE, the P_N approximations, and use it to derive the diffusion approximation, which we will ultimately use to compute the skin's reflectance properties under incident light.

4.3 The P_N Approximations

The domain of definition of the RTE, $\Omega \times V \times [0, T]$ possesses an inherent tensor product structure (Bal, 2009). Due to this, it is natural to expand the photon density $\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)$ into a series expansion (Case and Zweifel, 1967; Marchuck and Lebedev, 1986; S R Arridge, 1999),

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \sum_n \phi_n(\boldsymbol{r}, t) H_n(\hat{\boldsymbol{s}}). \tag{4.3.1}$$

This expansion permits the formal recasting of the RTE as an infinite system of coupled partial differential equations for the moments $\phi_n(\boldsymbol{r}, t)$. The most common, and perhaps best-suited, choice of basis functions $H_n(\hat{\boldsymbol{s}})$ are the spherical harmonics, which form a complete orthonormal system in $L^2(V)$, correspond to the eigenfunctions of the scattering operator. Additionally, the product $\hat{\boldsymbol{s}} H_n(\hat{\boldsymbol{s}})$ leads to a sparse coupling of the moment equations which arise from the spherical harmonics expansion.

Using the spherical harmonic ansatz (4.3.1), we express the density and source terms as,

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \psi_{l,m}(\boldsymbol{r}, t) Y_{l,m}(\hat{\boldsymbol{s}}) \tag{4.3.2}$$

$$q(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} q_{l,m}(\boldsymbol{r}, t) Y_{l,m}(\hat{\boldsymbol{s}}) \tag{4.3.3}$$

where $Y_{l,m}(\hat{\boldsymbol{s}})$ are the spherical harmonics (A.3.1) and normalisation factor $((2l+1)/4\pi)^{1/2}$ is used for convenience of expression.

Additionally, the fluence and flux vectors may be written as

$$\Phi(\boldsymbol{r}, t) = \psi_{0,0}(\boldsymbol{r}, t) \tag{4.3.4}$$

$$\boldsymbol{J}(\boldsymbol{r}, t) = \begin{pmatrix} \frac{1}{\sqrt{2}} (\psi_{1,-1}(\boldsymbol{r}, t) - \psi_{1,1}(\boldsymbol{r}, t)) \\ \frac{1}{i\sqrt{2}} (\psi_{1,-1}(\boldsymbol{r}, t) + \psi_{1,1}(\boldsymbol{r}, t) \\ \psi_{1,0}(\boldsymbol{r}, t) \end{pmatrix}. \tag{4.3.5}$$

By the addition theorem for spherical harmonics (A.3.5), we can express the scattering kernel as,

$$k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') = \sum_{l}^{\infty} \left(\frac{2l+1}{4\pi} \right) k_l P_l(\cos\theta) \tag{4.3.6}$$

$$= \sum_{l}^{\infty} \sum_{m=-l}^{l} k_l Y_{l,m}^*(\hat{\boldsymbol{s}}') Y_{l,m}(\hat{\boldsymbol{s}}) \tag{4.3.7}$$

for Legendre polynomials $P_l(\cos\theta)$ (A.1.1). Substituting these expansions into the RTE (4.2.1) we have,

$$\begin{aligned}
&\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r}) \right) \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \psi_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}) \\
&\qquad + \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \hat{\boldsymbol{s}} \cdot \nabla \psi_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}) \\
&= \mu_s(\boldsymbol{r}) \int_{\mathbb{S}^{d-1}} \left(\sum_{l}^{\infty} \sum_{m=-1}^{l} k_l Y_{l,m}^*(\hat{\boldsymbol{s}}') Y_{l,m}(\hat{\boldsymbol{s}}) \right) \\
&\qquad \times \left(\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \psi_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}) \right) d\hat{\boldsymbol{s}}' \\
&\qquad + \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} q_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}).
\end{aligned} \tag{4.3.8}$$

We now take the inner product with $Y_{l,m}^*(\hat{\boldsymbol{s}})$ term-by-term to exploit the orthogonal structure of the spherical harmonics (A.3.3). Applying the intermediate algebraic manipulations, relegated to Appendix C, we arrive at the infinite set of coupled partial differential equations known as the P_N approximations,

$$\begin{aligned}
&\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r}) \right) \psi_{l,m}(\boldsymbol{r},t) + \frac{1}{2l+1} \left(\frac{\partial}{\partial z} [((l-m+1)(l+m+1))^{1/2} \psi_{l+1,m}(\boldsymbol{r},t) \right. \\
&\qquad + ((l-m)(l+m))^{1/2} \psi_{l-1,m}(\boldsymbol{r},t)] \\
&\qquad - \frac{1}{2} \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y} \right) [((l+m)(l+m-1))^{1/2} \psi_{l-1,m-1}(\boldsymbol{r},t) \\
&\qquad - ((l-m+1)(l-m+2))^{1/2} \psi_{l+1,m-1}(\boldsymbol{r},t)] \\
&\qquad - \frac{1}{2} \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y} \right) [-((l-m)(l-m-1))^{1/2} \psi_{l-1,m+1}(\boldsymbol{r},t) \\
&\qquad \left. + ((l+m+1)(l+m+2))^{1/2} \psi_{l+1,m+1}(\boldsymbol{r},t) \right) \\
&\qquad = \mu_s(\boldsymbol{r}) k_l \psi_{l,m}(\boldsymbol{r},t) + q_{l,m}(\boldsymbol{r},t).
\end{aligned} \tag{4.3.9}$$

The P_N approximations are obtained by assuming $\psi_{l,m} = 0$ for $l > N$. This gives rise to $(N+1)^2$ coupled first-order partial differential equations in the general case. Due to the highly forward scattering nature of human skin, the photon density $\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)$ will be a smooth function of $\hat{\boldsymbol{s}}$. This results in fast decay of the moments $\psi_{l,m}(\boldsymbol{r}, t)$ in the expansions as $n \to \infty$. Thus, the P_N approximations yield excellent results for approximating the photon density in a highly scattering medium with N small (Ishimaru, 1978; A. D. Kim and Keller, 2003; Martelli et al., 2010).

4.3.1 The P_1 Approximation

For $N = 1$, we obtain four equations,

$$\begin{aligned}\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\psi_{0,0}(\boldsymbol{r}, t) + \frac{\partial}{\partial z}\psi_{1,0}(\boldsymbol{r}, t) + \frac{1}{\sqrt{2}}\frac{\partial}{\partial x}(\psi_{1,-1}(\boldsymbol{r}, t) - \psi_{1,1}(\boldsymbol{r}, t)) \\ + \frac{1}{i\sqrt{2}}\frac{\partial}{\partial y}(\psi_{1,-1}(\boldsymbol{r}, t) + \psi_{1,1}(\boldsymbol{r}, t)) = \mu_s(\boldsymbol{r})k_0\psi_{0,0}(\boldsymbol{r}, t) + q_{0,0}(\boldsymbol{r}, t)\end{aligned} \tag{4.3.10}$$

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\psi_{1,0}(\boldsymbol{r}, t) + \frac{1}{3}\frac{\partial}{\partial z}\psi_{0,0}(\boldsymbol{r}, t) = \mu_s(\boldsymbol{r})k_1\psi_{1,0}(\boldsymbol{r}, t) + q_{1,0}(\boldsymbol{r}, t) \tag{4.3.11}$$

$$\begin{aligned}\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\psi_{1,-1}(\boldsymbol{r}, t) + \frac{\sqrt{2}}{6}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_{0,0}(\boldsymbol{r}, t) \\ = \mu_s(\boldsymbol{r})k_1\psi_{1,-1}(\boldsymbol{r}, t) + q_{1,-1}(\boldsymbol{r}, t)\end{aligned} \tag{4.3.12}$$

$$\begin{aligned}\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\psi_{1,1}(\boldsymbol{r}, t) - \frac{\sqrt{2}}{6}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\psi_{0,0}(\boldsymbol{r}, t) \\ = \mu_s(\boldsymbol{r})k_1\psi_{1,1}(\boldsymbol{r}, t) + q_{1,1}(\boldsymbol{r}, t).\end{aligned} \tag{4.3.13}$$

Given the general form of the P_1 approximations above, we consider the more specific case of the RTE in spatial dimension $\mathbb{R}^3$ with a view towards the eventual derivation of the diffusion approximation for the fluence Φ, by writing the P_1 approximations in terms of the fluence Φ and flux $\boldsymbol{J}$. The spherical harmonics for $n = 1$ are

$$\begin{aligned}Y_{0,0}(\theta, \varphi) = \frac{1}{\sqrt{4\pi}}, \quad Y_{1,0}(\theta, \varphi) = \sqrt{\frac{3}{4\pi}}\cos\theta \\ Y_{1,-1}(\theta, \varphi) = \sqrt{\frac{3}{8\pi}}\sin\theta e^{-i\varphi}, \quad Y_{1,1}(\theta, \varphi) = -\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\varphi}.\end{aligned} \tag{4.3.14}$$

The fluence $\Phi(\boldsymbol{r}, t)$ then becomes,

$$\Phi(\boldsymbol{r}, t) = \int_{4\pi} \psi_{0,0}Y_{0,0} + \psi_{1,-1}Y_{1,-1} + \psi_{1,0}Y_{1,0} + \psi_{1,1}Y_{1,1}\, d\Omega \tag{4.3.15}$$

for solid angle $d\Omega$, where from (4.3.14),

$$\int_{4\pi} \psi_{0,0} Y_{0,0}\, d\Omega = 4\pi \psi_{0,0}(\boldsymbol{r}, t) Y_{0,0}(\hat{\boldsymbol{s}}) \tag{4.3.16}$$

$$\int_{4\pi} \psi_{1,-1} Y_{1,-1}\, d\Omega = \sqrt{\frac{3}{8\pi}} \psi_{1,-1} \int_0^{2\pi} \int_0^{\pi} \sin^2\theta e^{-i\varphi}\, d\theta d\varphi = 0 \tag{4.3.17}$$

$$\int_{4\pi} \psi_{1,0} Y_{1,0}\, d\Omega = 2\pi \sqrt{\frac{3}{4\pi}} \psi_{1,0} \int_0^{\pi} \cos\theta \sin\theta\, d\theta d\varphi = 0 \tag{4.3.18}$$

$$\int_{4\pi} \psi_{1,1} Y_{1,1}\, d\Omega = -\sqrt{\frac{3}{8\pi}} \psi_{1,-1} \int_0^{2\pi} \int_0^{\pi} \sin^2\theta e^{-i\varphi}\, d\theta d\varphi = 0 \tag{4.3.19}$$

and thus,

$$\Phi(\boldsymbol{r}, t) = 4\pi \psi_{0,0}(\boldsymbol{r}, t) Y_{0,0}(\hat{\boldsymbol{s}}) = \sqrt{4\pi} \psi_{0,0}(\boldsymbol{r}, t) \tag{4.3.20}$$

which says that the isotropic term of the spherical harmonic expansion (4.3.2) is the photon fluence divided by the solid angle 4π.

Similarly, taking the inner product of 4.3.2 with $\hat{\boldsymbol{s}}$ in the flux (4.1.3) with application of (A.3.2) and (A.3.3) where the flux is as in (4.3.5)

$$\boldsymbol{J}(\boldsymbol{r}, t) = \sqrt{\frac{2\pi}{3}} \begin{pmatrix} \psi_{1,-1}(\boldsymbol{r}, t) - \psi_{1,1}(\boldsymbol{r}, t) \\ -i(\psi_{1,-1}(\boldsymbol{r}, t) + \psi_{1,1}(\boldsymbol{r}, t) \\ \sqrt{2}\psi_{1,0}(\boldsymbol{r}, t) \end{pmatrix} \tag{4.3.21}$$

and $\hat{\boldsymbol{s}}$ is represented as in (A.3.23) gives,

$$\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}} = \frac{\pi}{3} \sum_{m=-1}^{1} \psi_{1,m}(\boldsymbol{r}, t) Y_{1,m}(\hat{\boldsymbol{s}}) \tag{4.3.22}$$

where $\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}} = |\boldsymbol{J}(\boldsymbol{r}, t)| \cos\alpha$ where $\cos\alpha = \boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}$.

Thus, the radiance $\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)$ in the P_1 approximation is

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \frac{1}{4\pi} \Phi(\boldsymbol{r}, t) + \frac{3}{4\pi} \boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}. \tag{4.3.23}$$

The radiance in (4.3.23) is known as the *diffusion expansion*. Substituting the diffusion expansion into the RTE (4.2.1) and integrating gives the first coupled equation of the P_1 approximation in terms of Φ and $\boldsymbol{J}$,

$$\frac{\partial \Phi(\boldsymbol{r}, t)}{\partial t} + \mu_a \Phi(\boldsymbol{r}, t) + \nabla \cdot \boldsymbol{J}(\boldsymbol{r}, t) = \boldsymbol{q}(\boldsymbol{r}, t) \tag{4.3.24}$$

Additionally, substituting the diffusion expansion into the RTE (4.2.1), multiplying by $\hat{\boldsymbol{s}}$ and integrating yields the second equation,

$$\frac{1}{c}\frac{\partial \boldsymbol{J}(\boldsymbol{r},t)}{\partial t} + (\mu_a + \mu_s')\boldsymbol{J}(\boldsymbol{r},t) + \frac{1}{3}\nabla\Phi(\boldsymbol{r},t) = 0 \tag{4.3.25}$$

where μ_s' is the *reduced scattering coefficient*

$$\mu_s' = \mu_s(1-g). \tag{4.3.26}$$

A complete derivation is given in Appendix C in Section C.2.

4.4 Diffusion Approximation

We now introduce the diffusion approximation to the RTE, which takes the P_1 approximation written in terms of fluence $\Phi(\boldsymbol{r},t)$ and the flux $\boldsymbol{J}(\boldsymbol{r},t)$, (4.3.24) and (4.3.25) and reduces them to a single equation only in $\Phi(\boldsymbol{r},t)$.

We assume that the time variation of the diffuse flux $\boldsymbol{J}(\boldsymbol{r},t)$ over time period $1/c(\mu_a+\mu_s')$ is negligible with respect to the vector itself,

$$(\mu_a + \mu_s')|\boldsymbol{J}(\boldsymbol{r},t)| \gg \left|\frac{1}{c}\frac{\partial \boldsymbol{J}(\boldsymbol{r},t)}{\partial t}\right| \tag{4.4.1}$$

and so $\partial_t \boldsymbol{J}(\boldsymbol{r},t)$ in (4.3.25) is negligable, yielding

$$\boldsymbol{J}(\boldsymbol{r},t) = -D\nabla\Phi(\boldsymbol{r},t) \tag{4.4.2}$$

Fick's law, with diffusion constant,

$$D = \frac{1}{3(\mu_a+\mu_s')} \tag{4.4.3}$$

In the diffusion expansion (4.3.23), this allows us to express the radiance as a function of the fluence $\Phi(\boldsymbol{r},t)$ alone as,

$$\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t) = \frac{1}{4\pi}\Phi(\boldsymbol{r},t) - \frac{3}{4\pi}D\nabla\Phi(\boldsymbol{r},t)\cdot\hat{\boldsymbol{s}}. \tag{4.4.4}$$

Substituting (4.4.2) into (4.3.24) gives the diffusion equation approximation to the RTE,

$$\frac{1}{c}\frac{\partial\Phi(\boldsymbol{r},t)}{\partial t} + \mu_a\Phi(\boldsymbol{r},t) - \nabla\cdot(D\nabla\Phi(\boldsymbol{r},t)) = \boldsymbol{q}(\boldsymbol{r},t). \tag{4.4.5}$$

4.4.1 Boundary Conditions

The vacuum boundary conditions for the RTE specify that no photons travel in an inward direction at the boundary except for the source term

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = 0, \quad \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}} < 0 \tag{4.4.6}$$

where $\hat{\boldsymbol{n}}$ is the unit normal to Σ at $\boldsymbol{r} \in \Sigma$, where Σ is the interface between the medium and the external non-scattering region. The diffusion equation cannot satisfy this condition exactly. The diffuse radiance $\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)$, $r \in \Sigma$ entering from any direction $\hat{\boldsymbol{s}}$ pointing toward the medium can only originate from reflection at the boundary. Due to a refractive index mismatch at the interface Σ between the medium and the external non-scattering medium, the radiance $\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}', t)$ along $\hat{\boldsymbol{s}}'$, the mirror direction to $\hat{\boldsymbol{s}}$ with respect to Σ, is

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = R_F(\hat{\boldsymbol{s}}' \cdot \hat{\boldsymbol{n}})\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}', t) \tag{4.4.7}$$

where $R_F(\hat{\boldsymbol{s}}' \cdot \hat{\boldsymbol{n}})$ is the *Fresnel reflection coefficient* for unpolarised light exiting the medium,

$$R_F(\hat{\boldsymbol{s}}' \cdot \hat{\boldsymbol{n}}) = \begin{cases} \frac{1}{2}\left(\frac{n_i \cos\theta - n_e \cos\theta_r}{n_i \cos\theta + n_e \cos\theta_r}\right)^2 + \frac{1}{2}\left(\frac{n_e \cos\theta - n_i \cos\theta_r}{n_e \cos\theta + n_i \cos\theta_r}\right)^2, & 0 < \theta < \theta_c \\ 1, & \theta \geq \theta_c \end{cases} \tag{4.4.8}$$

where n_i is the refractive index of the scattering medium, n_e is the refractive index of the external medium, θ is the angle of incidence, $\cos\theta = \hat{\boldsymbol{s}}' \cdot \hat{\boldsymbol{n}}$, and $\theta_r = \sin^{-1}((n_i/n_e)\sin\theta)$ is the angle of refraction, computed from *Snell's law*. $n_r = n_i/n_e$ is the relative refractive index of the scattering medium.

The total diffusive radiance directed inward is then equal to the internally reflected outward directed radiance,

$$-\int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}<0} \phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\, d\Omega = \int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}>0} R_F(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\, d\Omega. \tag{4.4.9}$$

We now substitute the diffusive radiance (4.3.23) into (4.4.9) to derive the boundary condition. Taking $\boldsymbol{J}(\boldsymbol{r}, t) = J_u\hat{\boldsymbol{u}} + J_n\hat{\boldsymbol{n}}$ where $\hat{\boldsymbol{u}}$ is tangential to Σ, $\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}} = \cos\theta$, and $\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{s}} = \cos\varphi\sin\theta$, the left hand side becomes,

$$\begin{aligned}\int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}<0} \phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\, d\Omega = \frac{1}{4\pi}\Phi(\boldsymbol{r}, t)\left(2\pi\int_{\pi/2}^{\pi} \cos\theta\sin\theta\, d\theta\right) + \frac{3}{4\pi}J_u\int_{2\pi}(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{s}})(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\, d\Omega \\ + \frac{3}{4\pi}J_n\int_{2\pi}(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})^2\, d\Omega\end{aligned} \tag{4.4.10}$$

which becomes,

$$
\begin{aligned}
&-\frac{1}{4}\Phi(\boldsymbol{r},t)+\frac{3}{4\pi}\left(J_u\int_0^{2\pi}\int_{\pi/2}^{\pi}\cos\varphi\cos\theta\sin^2\theta\,d\theta\,d\varphi+J_n\int_0^{2\pi}\int_{\pi/2}^{\pi}\cos^2\theta\sin\theta\,d\theta\,d\varphi\right)\\
&=-\frac{1}{4}\Phi(\boldsymbol{r},t)+\frac{3}{4\pi}\left(\frac{2\pi}{3}J_n\right).
\end{aligned}
\tag{4.4.11}
$$

Thus the left hand side of (4.4.9),

$$
\int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}<0}\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\,d\Omega=-\frac{1}{4}\Phi(\boldsymbol{r},t)+\frac{1}{2}\boldsymbol{J}(\boldsymbol{r},t)\cdot\hat{\boldsymbol{n}}. \tag{4.4.12}
$$

The right hand side is

$$
\begin{aligned}
\int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}>0}R_F(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\,d\Omega&=\frac{1}{2}\Phi(\boldsymbol{r},t)\int_0^{\pi/2}R_F(\cos\theta)\cos\theta\sin\theta\,d\theta\\
&+\frac{3}{4\pi}\left(J_u\int_{2\pi}R_F(\cos\theta)(\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{s}})(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\,d\Omega+J_n\int_{2\pi}R_F(\cos\theta)(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})^2\,d\Omega\right)
\end{aligned}
\tag{4.4.13}
$$

where

$$
\int_{2\pi}R_F(\cos\theta)(\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{s}})(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\,d\Omega=\int_0^{2\pi}\int_0^{\pi/2}R_F(\cos\theta)\cos\varphi\sin^2\theta\cos\theta\,d\theta=0. \tag{4.4.14}
$$

Thus the right hand side is

$$
\begin{aligned}
\int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}>0}R_F(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}})\,d\Omega&=\frac{1}{2}\Phi(\boldsymbol{r},t)\int_0^{\pi/2}R_F(\cos\theta)\cos\theta\sin\theta\,d\theta\\
&+\frac{3}{2}J_n\int_0^{2\pi}R_F(\cos\theta)\cos^2\theta\sin\theta\,d\theta.
\end{aligned}
\tag{4.4.15}
$$

Equating (4.4.12) and (4.4.15) and defining the coefficient A as,

$$
A=\frac{1+3\int_0^{\pi/2}R_F(\cos\theta)\cos^2\theta\sin\theta\,d\theta}{1-2\int_0^{\pi/2}R_F(\cos\theta)\cos\theta\sin\theta\,d\theta}\equiv\frac{1+R_s(n_r)}{1-R_s(n_r)} \tag{4.4.16}
$$

where $R_s(n_r)$ represents the diffuse Fresnel reflectance, we arrive at the Robin boundary condition for the diffusion approximation,

$$
\Phi(\boldsymbol{r},t)-2A\boldsymbol{J}(\boldsymbol{r},t)\cdot\hat{\boldsymbol{n}}\Big|_{\boldsymbol{r}\in\Sigma}=0. \tag{4.4.17}
$$

From Fick's law, this becomes

$$\Phi(\boldsymbol{r},t) - 2AD\frac{\partial}{\partial n}\Phi(\boldsymbol{r},t)\bigg|_{\boldsymbol{r}\in\Sigma} = 0. \tag{4.4.18}$$

Egan, Hilgeman, and Reichman, 1973 demonstrated that the diffuse Frensel reflectance $R_s(n_r)$ is well approximated by

$$R_s(n) = \begin{cases} -0.4399 + \frac{0.7099}{n_R} - \frac{0.3319}{n_r^2} + \frac{0.0636}{n_r^3}, & n_r < 1 \\ 0.0635 n_r + 0.6681 + \frac{0.7099}{n_r} - \frac{1.4399}{n_r^2}, & n_r > 1. \end{cases} \tag{4.4.19}$$

The Robin BC (4.4.18) states that the derivative of the fluence along the normal to the boundary, $\partial_n\Phi(\boldsymbol{r},t)$ is proportional to the fluence itself $\Phi(\boldsymbol{r},t)$. A decrement of $\Phi(\boldsymbol{r},t)$ may be extrapolated into the non-scattering region by assuming $\partial_n\Phi(\boldsymbol{r},t)$ remains constant in the non-scattering region to the value of the boundary. The distance from the boundary Σ at which the fluence is extrapolated to zero is the *extrapolated distance* z_b

$$z_b = 2AD. \tag{4.4.20}$$

The *extrapolated boundary condition* (ECB) assumes $\Phi(\boldsymbol{r},t) = 0$ on the surface at z_b.

4.5 Diffusive Reflection in Layered Media

The skin is a layered, highly forward scattering, participatory medium, as mentioned in Section 2.1, for which the diffusion equation serves as an excellent approximation to RTE. To determine the reflectance from light-interaction in the skin, we solve the diffusion approximation to the RTE in layered skin tissue building upon the Kubelka-Munk Fourier technique of Donner and Jensen, 2005 who applied the method to translucent scattering materials for computer graphic simulations.

4.5.1 Green's Function in the Infinite Medium

We first consider the essential building block of the diffusive solution in layered skin, the infinite medium.

Consider the diffusion equation (4.4.5) with a spatially indepedent diffusion constant D and spatially and temporally isotropic source of unitary strength,

$$q_0(\boldsymbol{r},t) = \delta^3(\boldsymbol{r})\delta(t) \tag{4.5.1}$$

where $\delta(\cdot)$ is the Dirac delta function as

$$\left(\frac{1}{c}\frac{\partial}{\partial t} - D\nabla^2 + \mu_a\right)\Phi(\boldsymbol{r},t) = q_0(\boldsymbol{r},t). \tag{4.5.2}$$

The solution will have the form $\Phi(\boldsymbol{r},t) = \Phi_0(\boldsymbol{r},t)\exp\{-\mu_a ct\}$ where $\Phi_0(\boldsymbol{r},t)$ is the solution in the non-absorbing external medium. For $\Phi_0(\boldsymbol{r},t)$, we set $\mu_a = 0$ and due to spherical symmetry use separation of variables (Evans, 2010) to find a solution of the form $\Phi_0(\boldsymbol{r},t) = \Phi_{0x}(x,t)\Phi_{0y}(y,t)\Phi_{0z}(z,t)$, for which (4.5.2) becomes

$$\left(\frac{1}{c}\frac{\partial}{\partial t} - D\nabla^2 + \mu_a\right)\Phi_0(\boldsymbol{r},t) = \delta(x)\delta(y)\delta(z)\delta(t). \tag{4.5.3}$$

subject to

$$\Phi_0(\boldsymbol{r},t) = 0,\ t < 0 \tag{4.5.4}$$
$$\Phi_0(\boldsymbol{r},t) \to 0,\ |\boldsymbol{r}| \to \infty. \tag{4.5.5}$$

Taking the spatial Fourier transform, we have,

$$\frac{1}{c}\frac{\partial}{\partial t}\tilde{\Phi}_0(\boldsymbol{\xi},t) + (\xi_1+\xi_2+\xi_3)D\tilde{\Phi}_0(\boldsymbol{\xi},t) = \delta(t) \tag{4.5.6}$$

where

$$\tilde{\Phi}_{0x}(\xi_1,t) = \mathcal{F}(\Phi_{0x})(x,t) = \int_{\mathbb{R}} \Phi_{0x}(x,t)e^{-i\xi_1 x}\,dx \tag{4.5.7}$$

and simiarly for $\tilde{\Phi}_{0y}(y,t)$, $\tilde{\Phi}_{0z}(z,t)$.

Integrating yields solution,

$$\tilde{\Phi}_0(\boldsymbol{\xi},t) = \begin{cases} A\exp\left\{-(\xi_1+\xi_2+\xi_3)Dct\right\}, & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{4.5.8}$$

To find the constant of integration A, integrate over $(0^-, 0^+)$, yielding

$$\frac{1}{c}\left(\tilde{\Phi}_0(\boldsymbol{\xi},0^+) - \tilde{\Phi}_0(\boldsymbol{\xi},0^-)\right) = 1 \tag{4.5.9}$$

whereby $\tilde{\Phi}_0(\boldsymbol{\xi},0^-) = 0$, we have that $A = c$.

Taking the inverse Fourier transform yields,

$$\Phi_0(\boldsymbol{r},t) = \begin{cases} \frac{c}{(2\pi)^3}\left(\frac{\pi}{Dct}\right)^{3/2}\exp\left\{-\frac{x^2+y^2+z^2}{4Dct}\right\}, & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{4.5.10}$$

which thus gives the Green's function,

$$\Phi(\boldsymbol{r},t) = \frac{c}{(4\pi Dct)^{3/2}}\exp\left\{-\frac{r^2}{4Dct} - \mu_a ct\right\},\ t > 0 \tag{4.5.11}$$

where $r^2 = x^2 + y^2 + z^2$. From Fick's Law, the Green's function for the flux for $t > 0$ is,

$$\boldsymbol{J}(\boldsymbol{r}, t) = \frac{r}{16(\pi Dc)^{3/2} t^{5/2}} \exp\left\{-\frac{r^2}{4Dct} - \mu_a ct\right\} \hat{\boldsymbol{r}} \tag{4.5.12}$$

and the Green's function for the radiance is obtained from the diffusion expansion (4.4.4)

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \frac{1}{(4\pi)^{5/2}(Dct)^{3/2}} \left(c + \frac{3|r|}{2t}(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{s}})\right) \exp\left\{-\frac{r^2}{4Dct} - \mu_a ct\right\} \tag{4.5.13}$$

Integrating over time gives the Green's functions for stationary or steady-state sources as,

$$\int_0^\infty \Phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)\, dt = \frac{c}{(4\pi DC)^{3/2}} \int_0^\infty t^{-3/2} \exp\left\{-\frac{r^2}{4DC}\frac{1}{t} - \mu_a ct\right\} dt \tag{4.5.14}$$

which is an integral of the form (Gradshteyn and Ryzhik, 1980)

$$\int_0^\infty t^{\nu-1} \exp\left\{-\frac{B}{t} - \gamma t\right\} dt = 2\left(\frac{B}{\gamma}\right)^{\nu/2} k_v(2\sqrt{\gamma B}) \tag{4.5.15}$$

where k_v is a modified spherical Bessel function of the second kind (Abramowitz and Stegun, 1965), with

$$k_{v+1/2}(x) = k_{-v-1/2}(x) \tag{4.5.16}$$

$$\sqrt{\frac{\pi}{2x}} k_{1/2}(x) = \frac{\pi}{2x} e^{-x}, \quad \sqrt{\frac{\pi}{2x}} k_{3/2}(x) = \left(1 + \frac{1}{x}\right) e^{-x}. \tag{4.5.17}$$

As such, the stationary fluence is given as,

$$\Phi(\boldsymbol{r}) = \frac{1}{4\pi Dr} \exp\left\{-\mu_{eff}|r|\right\} \tag{4.5.18}$$

where μ_{eff} is the effective attenuation coefficient,

$$\mu_{eff} = \sqrt{\frac{\mu_a}{D}} = \sqrt{3\mu_a \mu_s'} \tag{4.5.19}$$

Similarly, the stationary flux and radiance

$$\boldsymbol{J}(\boldsymbol{r}) = \frac{1}{4\pi r}\left(\frac{1}{4} + \mu_{eff}\right) \exp\left\{-\mu_{eff}|r|\right\} \hat{\boldsymbol{r}} \tag{4.5.20}$$

$$\phi(\boldsymbol{r}, \boldsymbol{s}) = \frac{1}{16\pi^2 rD}\left(1 + 3\left(\frac{D}{|r|} + \mu_{eff} D\right)(\hat{r} \cdot \hat{\boldsymbol{s}})\right) \exp\left\{-\mu_{eff}|r|\right\} \tag{4.5.21}$$

4.5.2 Slab Geometry

We now consider the problem of the RTE in a medium bound by two parallel planes, known as the *slab geometry*. The slab geometry has been used as a model of human skin tissue for research for many years (Ishimaru, 1978; Groenhuis, Ferwerda, and Ten Bosch, 1983; Keijzer, Star, and Storchi, 1988; Hielscher et al., 1995; Tuchin, 2000).

The solution to the RTE in the slab geometry is obtained using the *method of images* (Ockendon et al., 2003) with the extrapolated boundary condition (ECB), whereby we assume the fluence is equal to zero at the extrapolated surface $z_b = 2AD$. The method of images constructs the solution of the fluence inside the slab as a superposition of solutions for the infinite medium, derived in Subsection 4.5.1.

For an isotropic source $\boldsymbol{q}(\boldsymbol{r}, t) = \delta^3(\boldsymbol{r} - \boldsymbol{r}_s)\delta(t)$, $\boldsymbol{r}_s = (0, 0, z_s)$ and $0 \leq z_z \leq d$ for a slab of thickness d. From the ECB, the fluence is zero at the extrapolated surfaces outside the slab at $z_b = 2AD$. By the method of images, the boundary conditions are obtained from augmenting the real source in $\boldsymbol{r}_S$ an infinite number of pairs of positive and negative sources in an infinite medium with the same optical properties of the slab. In Figure 4.1 we show a schematic of the slab geometry, along with the locations $\boldsymbol{r}_i^+$ of the positive sources and $\boldsymbol{r}_i^-$ of the negative sources.

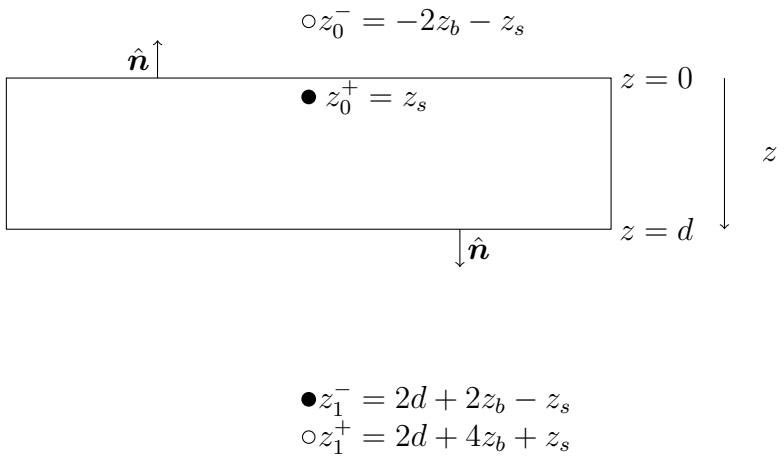

Figure 4.1: Model of slab geometry and method of images sources.

Thus, the fluence inside the slab is comprised of terms of fluence from an infinite medium, where the postive terms from sources at $\boldsymbol{r}_i^+$ and the negative terms from the sources at $\boldsymbol{r}_i^-$ are placed along the z-axis at,

$$z_i^+ = 2i(d + 2z_b) + z_s \tag{4.5.22}$$

$$z_i^- = 2i(d + 2z_b) - 2z_b - z_s \tag{4.5.23}$$

for $i = 0, \pm 1, \pm 2, \ldots$. The Green's function for the fluence at $\boldsymbol{r}$ is the sum of the contributions of the individual contributions from the dipole array,

$$\begin{aligned}\Phi(\boldsymbol{r}, t) =& \frac{c}{4\pi Dct)^{3/2}} \exp\left\{-\frac{\rho^2}{4Dct} - \mu_a ct\right\} \\ &\times \sum_{i=-\infty}^{\infty} \left(\exp\left\{-\frac{(z - z_i^+)^2}{4Dct}\right\} - \exp\left\{-\frac{(z - z_i^-)^2}{4Dct}\right\}\right).\end{aligned} \tag{4.5.24}$$

where $\rho^2 = x^2 + y^2$. From Fick's law, and the diffusion expansion, the flux and radiance are,

$$\boldsymbol{J}(\boldsymbol{r}, t) = -D\nabla\Phi(\boldsymbol{r}, t) \tag{4.5.25}$$

$$\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t) = \frac{1}{4\pi}\Phi(\boldsymbol{r}, t) - \frac{3}{4\pi} D\nabla\Phi(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}. \tag{4.5.26}$$

We can now compute the *reflectance* and *transmittance*, which constitute the power cross the surface at $z = 0$ and $z = d$, respectively. The reflectance constitutes the observed or measured light exiting from the skin tissue and is what we use in Chapter 6 for skin rendering. The Green's function for the reflectance $R(\rho, t)$ is

$$R(\rho, t) = \int_{\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}>0} (1 - R_F(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}}))\, \phi(\rho, z = 0, \hat{\boldsymbol{s}}, t)(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}})\, d\Omega \tag{4.5.27}$$

$$= \boldsymbol{J}(\rho, z = 0, t) \cdot \hat{\boldsymbol{n}} \tag{4.5.28}$$

From Fick's law,

$$R(\rho, t) = D\frac{\partial}{\partial z}\Phi(\rho, t)\bigg|_{z=0}. \tag{4.5.29}$$

Similarly, the transmittance $T(\rho, t)$ is

$$T(\rho, t) = -D\frac{\partial}{\partial z}\Phi(\rho, t)\bigg|_{z=d} \tag{4.5.30}$$

Using these expressions, we arrive at the time-resolved reflectance and transmittance

$$\begin{aligned}R(\rho, t) =& -\frac{1}{2(4\pi Dc)^{3/2} t^{5/2}} \exp\left\{-\frac{\rho^2}{4Dct} - \mu_a ct\right\} \\ &\times \sum_{i=-\infty}^{\infty} \left(z_{3i} \exp\left\{-\frac{z_{3i}^2}{4Dct}\right\} - z_{4i} \exp\left\{-\frac{z_{4i}^2}{4DCt}\right\}\right)\end{aligned} \tag{4.5.31}$$

$$T(\rho, t) = \frac{1}{2(4\pi Dc)^{3/2} t^{5/2}} \exp\left\{-\frac{\rho^2}{4Dct} - \mu_a ct\right\} \tag{4.5.32}$$

$$\times \sum_{i=-\infty}^{\infty} \left(z_{1i} \exp\left\{-\frac{z_{1i}^2}{4Dct}\right\} - z_{2i} \exp\left\{-\frac{z_{2i}^2}{4DCt}\right\}\right) \tag{4.5.33}$$

where the dipole arrays are,

$$z_{1i} = (1-2i)d - 4iz_b - z_s \tag{4.5.34}$$
$$z_{2i} = (1-2i)d - (4i-2)z_b + z_s \tag{4.5.35}$$
$$z_{3i} = -2id - 4iz_b - z_s \tag{4.5.36}$$
$$z_{4i} = -2id - (4i-2)z_b + z_s. \tag{4.5.37}$$

for $i = 0, \pm 1, \pm 2, \ldots$. The steady-state reflectance and transmittance are derived by integrating (4.5.31) and (4.5.32), respectively.

$$\begin{aligned} R(\rho) &= \int_0^\infty R(\rho, t)\, dt \\ &= -\frac{1}{4\pi} \sum_{i=-\infty}^{\infty} \left(z_{3i}(\rho^2 + z_{3i}^2)^{-3/2} \left(1 + \left(\frac{\mu_a(\rho^2 + z_{3i}^2)}{D} \right)^{1/2} \right) \exp\left\{ - \left(\frac{\mu_a(\rho^2 + z_{3i}^2)}{D} \right)^{1/2} \right\} \right. \\ &\quad \left. - z_{4i}(\rho^2 + z_{4i}^2)^{-3/2} \left(1 + \left(\frac{\mu_a(\rho^2 + z_{4i}^2)}{D} \right)^{1/2} \right) \exp\left\{ - \left(\frac{\mu_a(\rho^2 + z_{4i}^2)}{D} \right)^{1/2} \right\} \right) \end{aligned} \tag{4.5.38}$$

$$\begin{aligned} T(\rho) &= \int_0^\infty T(\rho, t)\, dt \\ &= \frac{1}{4\pi} \sum_{i=-\infty}^{\infty} \left(z_{1i}(\rho^2 + z_{1i}^2)^{-3/2} \left(1 + \left(\frac{\mu_a(\rho^2 + z_{1i}^2)}{D} \right)^{1/2} \right) \exp\left\{ - \left(\frac{\mu_a(\rho^2 + z_{1i}^2)}{D} \right)^{1/2} \right\} \right. \\ &\quad \left. - z_{2i}(\rho^2 + z_{2i}^2)^{-3/2} \left(1 + \left(\frac{\mu_a(\rho^2 + z_{2i}^2)}{D} \right)^{1/2} \right) \exp\left\{ - \left(\frac{\mu_a(\rho^2 + z_{2i}^2)}{D} \right)^{1/2} \right\} \right). \end{aligned} \tag{4.5.39}$$

4.5.3 Multiple Layer Interlayer Scattering

The above derived reflectance (4.5.38) and (4.5.39) hold for a single layered slab, with the implicit assumption that the media above and below the slab have the same index of refraction. Human skin, however, is a multi-layered material, where there is a difference in the refractive index of different layers. The refractive index varies from 1.47 to 1.55 in the stratum corneum, 1.41 to 1.49 in epidermal tissue, and 1.36 to 1.41 in dermal tissues over the visible wavelength (Lister, Wright, and Chappell, 2012). Researchers have sought to resolve this through introducing complex boundary conditions and approximations at the various interfaces, however, these result in complex and expensive to evaluate expressions for the fluence (Dayan, Havlin, and Weiss, 1992; Ripoll and Nieto-Vesperinas, 1999; Faris, 2002; Liemert and Kienle, 2010).

The effect of Fresnel reflectance between mismatched refractive index layers can be modelled with extended boundary conditions from the single slab case. Let n_{top} and n_{bottom} be the

refractive indices of two interfaces of the slab. The different Fresnel reflectances yield different EBCs at both the top and bottom of the slab,

$$\Phi(\boldsymbol{r},t) - 2A_{top}D\frac{\partial}{\partial n}\Phi(\boldsymbol{r},t)\bigg|_{z=0} = 0 \tag{4.5.40}$$

$$\Phi(\boldsymbol{r},t) - 2A_{bottom}D\frac{\partial}{\partial n}\Phi(\boldsymbol{r},t)\bigg|_{z=d} = 0 \tag{4.5.41}$$

for A_{top} and A_{bottom} computed as

$$A_{top} = \frac{1 + R_s(n_{top})}{1 - R_s(n_{top})} \tag{4.5.42}$$

$$A_{bottom} = \frac{1 + R_s(n_{bottom})}{1 - R_s(n_{bottom})}. \tag{4.5.43}$$

These will yield different extrapolated boundaries where the fluence is zero, however, we can satisfy both by mirroring about the distance from the top extrapolated boundary to the bottom extrapolated boundary, giving,

$$z_i^+ = 2i(d + z_{b,top} + z_{b,bottom}) + z_s \tag{4.5.44}$$

$$z_i^- = 2i(d + z_{b,top} + z_{b,bottom}) - z_s + 2z_{b,top} \tag{4.5.45}$$

where $z_{b,top}$ is computed with A_{top} and $z_{b,bottom}$ is computed with A_{bottom}, respectively. If $n_{top} = n_{bottom}$, then (4.5.44) and (4.5.45) reduces to (4.5.22) and (4.5.23), respectively.

To account for interlayer scattering between the layers of the slabs, we are inspired by Donner and Jensen, 2005 who convolve the reflectance and transmittance profiles of adjacent layers using *Kubelka-Munk* theory.

Kubelka-Munk (KM) theory describes light propagation through turbid media using the 1D radiative transport equation (Kubelka, 1948; Nobbs, 1985). The KM equations are a two-dimensional system of linear equations in the forward and backward flux, which can be derived empirically, or from a perturbation analysis of the 1D RTE (Ilan and A. D. Kim, 2019). The KM equations are extended to multiple layers through multiplying the reflectances and transmittances of adjacent layers together (Kubelka, 1954; Baranowski et al., 2011).

As we are applying diffusion theory, it is reasonable to assume that all inter-layer interactions are due to multiple scattering events (Cheong, Prahl, and Welch, 1990; Donner and Jensen, 2005; Steven L. Jacques, 2013). This permits modelling scattering one layer at a time. Consider a plane-parallel layered medium with subsequent slab layers i and j. Let R_i, R_j and T_i, T_j be the reflectance and transmittance for layers i and j, respectively. A base transmittance profile of both layers including the effect of internal reflection is determined by convolving the profiles of the individual layers,

$$T_{ij} = T_i * T_j. \tag{4.5.46}$$

This implicitly assumes that none of the light transmitted to layer j is reflected to layer i. Since light T_i reaches the top of layer j, light $T_i * R_j$ will be scattered back, of which $T_i * R_j * R_1$ will be scattered back into layer j, and $T_1 * R_2 * R_1 * T_2$ will be transmitted out of the bottom of layer j. To account for these contributions (4.5.46) is amended as

$$T_{ij} = T_i * T_j + T_i * R_j * R_i * T_j + T_i * R_j * R_i * R_j * R_1 * T_j + \cdots \tag{4.5.47}$$

Figure 4.2 describes the contributions from each multiple scattering event. This series of convolutions can be evaluated as pointwise products in Fourier space,

$$\begin{aligned} \tilde{T}_{ij} &= \tilde{T}_i\tilde{T}_j + \tilde{T}_i\tilde{R}_j\tilde{R}_j\tilde{R}_i\tilde{T}_j + \tilde{T}_i\tilde{R}_j\tilde{R}_i\tilde{R}_j\tilde{R}_i\tilde{T}_j + \cdots \\ &= \tilde{T}_i\tilde{T}_j(1 + \tilde{R}_j\tilde{R}_i + (\tilde{R}_j\tilde{R}_i)^2 + (\tilde{R}_j\tilde{R}_i)^3 + \cdots) \end{aligned} \tag{4.5.48}$$

where $\tilde{T}$ and $\tilde{R}$ are the Fourier transformed diffuse transmittance and reflectance, respectively. The sum in the brackets is a geometric series, and as $\tilde{R}_j\tilde{R}_i < 1$,

$$\tilde{T}_{ij} = \frac{\tilde{T}_i\tilde{T}_j}{1 - \tilde{R}_j\tilde{R}_i} \tag{4.5.49}$$

Repeating this analysis for the reflectance, as in Figure 4.2, we have

$$\tilde{R}_{ij} = \tilde{R}_i + \frac{\tilde{R}_i\tilde{R}_j\tilde{T}_i}{1 - \tilde{R}_j\tilde{R}_i}. \tag{4.5.50}$$

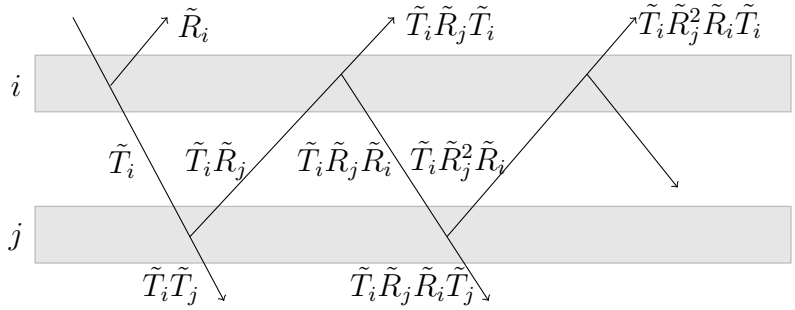

Figure 4.2: Reflectance and transmittance for two layer plane parallel slab geometry medium. Layers i and j with reflectance and transmittance given for each.

Recursive application of (4.5.49) and (4.5.50) yields the total transmittance $\tilde{T}_{12\ldots n}$ and reflectance $\tilde{R}_{12\ldots n}$ for the nth layer,

$$\tilde{T}_{12\ldots n} = \frac{\tilde{T}_{12\ldots n-1}\tilde{T}_n}{1 - \tilde{R}_{12\ldots n-1}\tilde{R}_n} \tag{4.5.51}$$

$$\tilde{R}_{12\ldots n} = \tilde{R}_{12\ldots n-1} + \frac{\tilde{T}^2_{12\ldots n-1}\tilde{R}_n}{1 - \tilde{R}_{12\ldots n-1}\tilde{R}_n} \tag{4.5.52}$$

where $\tilde{T}_{12\ldots n-1}$ represents the total transmittance of the previous $n-1$ layers, and $\tilde{R}_{12\ldots n-1}$ represents the total reflectance of the previous $n-1$ layers.

4.5.4 Reflectance Simulation

For the purpose of recovering chromophore concentrations from digital photographs and video images, we must consider the appropriate length scales of the problem. In the setting of optical tomography, the typical length scale of an object of interest is 10 to 100 mm (S. R. Arridge, Cope, and Delpy, 1992; S R Arridge, 1999). In the setting of video monitoring of vital signs, the exposed skin of a subject is many orders of magnitude larger, with the additional complication of the camera placed up to several meters away from the subject. The speed of light is approximately 0.3 mm/ps, and so the average residence time of a photon in an object is on the order of nanoseconds. As such any source terms as well as inflowing light will be essentially constant in time, for which a stationary solution $\phi(\boldsymbol{r}, \boldsymbol{v})$ will be reached immediately. As such, we compute the steady-state properties of light transport for use in Chapter 6 for skin rendering.

Given that we have defined the steady-state reflectance and transmittance for each layer of skin tissue, as well as the geometric series formulae for the total reflectance (4.5.52) and total transmisttance (4.5.51) we must compute them across the visible wavelengths to produce rendered skin images. To evaluate the accuracy of the diffusion approximation and Kubelka-Munk inter-layer scattering, we compare the simulation results with Monte Carlo simulations of photon propagation using the same physical properties using the MCML code package (L. Wang, S. L. Jacques, and Zheng, 1995). The MCML simulations used 10^6 photon packets, and each took several hours to compute, while the diffusion approximation can be computed rapidly in less than a second. This extreme computational expense for Monte Carlo simulations preclude their use Chapter 6, where we simulate more than 60,000 skin tones across the visible light spectra.

We simulate the reflectance across the optical light spectrum from 400nm to 700nm using a three layer model of skin tissue, consisting of the stratum corneum, epidermis, and dermis. We obtain the absorption coefficients, μ_a, reduced scattering coefficients, μ_s', and indices of refraction n from the literature (Meglinksi and Matcher, 2002; Salomatina et al., 2006). Note that absorption and reduced scattering coefficients are dependent upon the wavelength of light, and thus we compute the total reflectance at different wavelengths. The refractive indices and layer depths remain fixed across wavelengths and are displayed in Table (4.1). The refractive indices are taken from Ding et al., 2006. The absorption and reduced scattering coefficient values are found in Table 4.2.

Table 4.1: Refractive indices n and tissue depths d for skin layers.

Layer	Refractive Index n	Depth d [mm]
Stratum Corneum	1.52	0.01
Epidermis	1.44	0.21
Dermis	1.39	2.0

In Figures 4.3 through 4.6, we display the results of the diffuse reflectance calculations using

Table 4.2: Absorption coefficients and reduced scattering coefficients used for multilayer diffuse reflectance calculations. SC denotes the stratum corneum.

	$\mu_a(\lambda)$		[mm]	$\mu_s'(\lambda)$		[mm]
	SC	Epidermis	Dermis	SC	Epidermis	Dermis
400	0.27	0.42	1.78	12	10	7.5
500	0.13	0.20	0.150	8.4	7	4.5
600	0.07	0.11	0.056	5.5	5	3
700	0.04	0.07	0.007	2	1.2	2.5

the absorption and reduced scattering coefficients at 400, 500, 600, and 700 nm of light as given in Table 4.2.

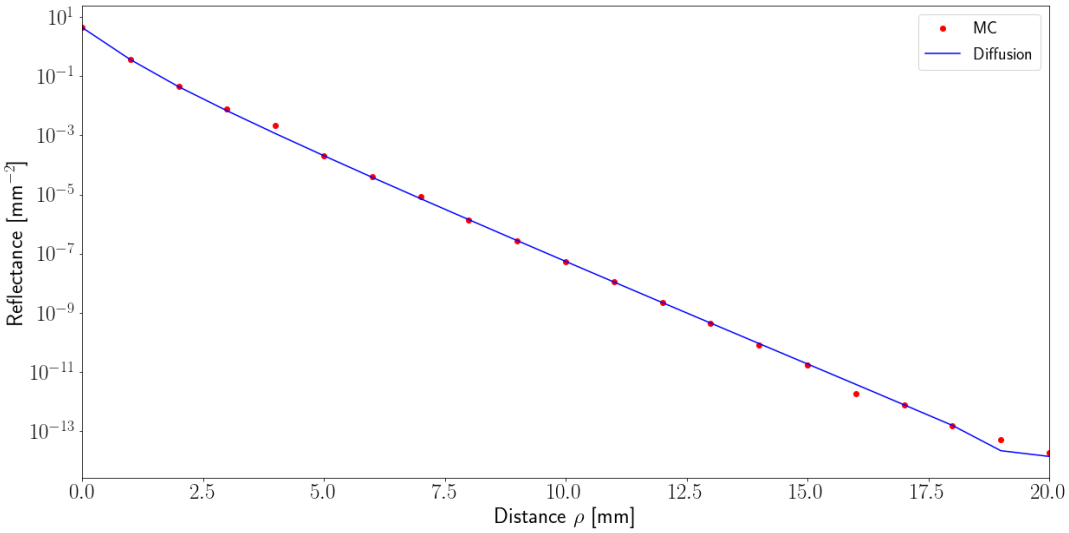

Figure 4.3: Skin reflectance computed using diffusion approximation and Monte Carlo for a three layer skin model using absorption and reduced scattering coefficients at light wavelength 400 nm. The distance ρ is measured as the penetration depth into the tissue.

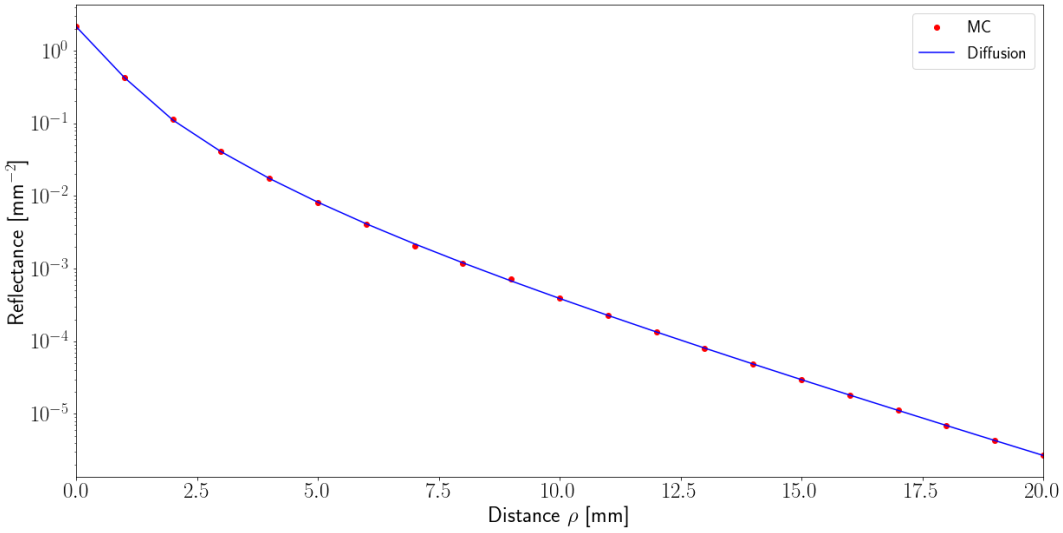

Figure 4.4: Skin reflectance computed using diffusion approximation and Monte Carlo for a three layer skin model using absorption and reduced scattering coefficients at light wavelength 500 nm. The distance ρ is measured as the penetration depth into the tissue.

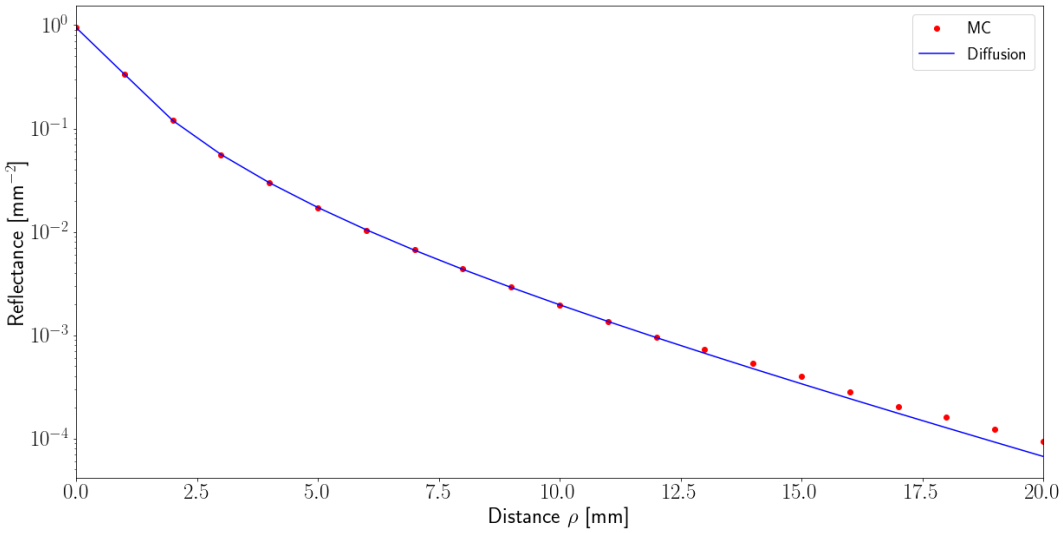

Figure 4.5: Skin reflectance computed using diffusion approximation and Monte Carlo for a three layer skin model using absorption and reduced scattering coefficients at light wavelength 600 nm. The distance ρ is measured as the penetration depth into the tissue.

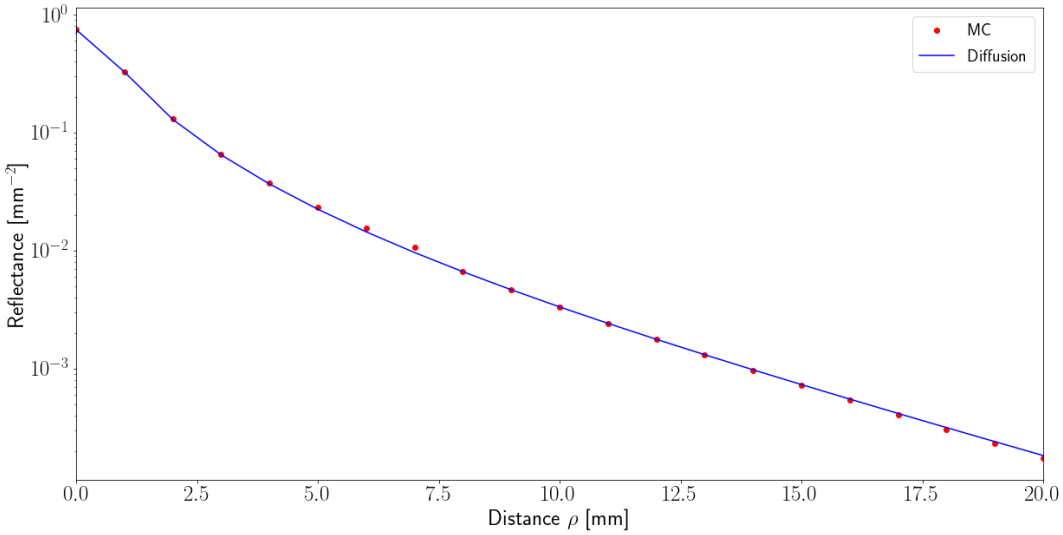

Figure 4.6: Skin reflectance computed using diffusion approximation and Monte Carlo for a three layer skin model using absorption and reduced scattering coefficients at light wavelength 700 nm. The distance ρ is measured as the penetration depth into the tissue.

Chapter 5

Camera Imaging Processing Pipeline

> "*Digital cameras are not designed to be light measuring devices; they are designed to produced visually pleasing photographs.*"
>
> — Michael S. Brown
> *ICCV Tutorial*

The final component of the proposed multi-stage monitoring methodology is a model of digital image formation. The vast majority of computer vision algorithms assume that cameras act as accurate light measuring devices which display images that are directly related to the radiance of the captured scene. In practice, however, this is not the case (Brown, 2019). Digital cameras apply a host of algorithmic processing steps to transform the captured scene radiance to a visually pleasing photograph to be displayed to the end-user. For this reason, a detailed model of the in-camera imaging pipeline is required to correctly render skin images.

In Section 5.1 we present a self-contained introduction to digital photography and the digital camera imaging processing pipeline. We discuss the constituent components of the imaging pipeline, as well as the features which may be estimated through statistical procedures. In Section 5.2 we recover the spectral sensitivity functions of an Apple iPhone 6S (Cupertino, CA) using statistical inference principal component analysis due to Jiang et al., 2013. In section 5.3 we present a modification of the procedure of Chakrabarti, Xiong, et al., 2013 to model the in-camera colour processing pipeline, using Gaussian processes, which we apply to approximate the proprietary image processing algorithms of the iPhone 6S.

5.1 Image Formation Model

A digital camera is a device for capturing photographs in a digital memory system. Over the past 30 years, there have been unprecedented advances in digital camera technology ranging from sensor miniaturization, image reproduction ability, and rapidly falling costs due to economies of scale. Modern digital cameras found in smartphones have followed an evolution from extremely low-quality image reproducers to devices which rival professional grade dedicated cameras at a fraction of the cost. There are two reasons for this evolution: *(i)* advances in *Complementary metal-oxide-semiconductor* (CMOS) sensors, and *(ii)* back-side illumination (Klug and Shimpi, 2010). A CMOS or *active pixel sensor* is an imaging sensor with each pixel consisting of a photodetector and one or more active transistors (Fossum and Hondongwa, 2014). In comparison to traditional *charge-coupled device* (CCD) camera sensors, CMOS sensors can be produced at very low costs, consume very little power, and can be made in compact form factors, all of which make them ideal for smartphones (Fossum, 1993). Back-side illumination refers to the manufacturing of the camera's sensor array with its wiring below the photodiode, as opposed to the wiring above the photodiode, as is found in *front-side illuminated sensors* - see Figure 5.1. This configuration allows more light to reach the photosensitive photodiode of the sensor, thus producing a higher quality image. Within the past 10 years, almost all smartphone cameras have transitioned to back-side illuminated CMOS camera sensors, tremendously improving imaging quality.

An idealised digital camera may be viewed as an abstracted light-measuring device (Sharma, 2003). In reality, however, the image the camera captures and the image displayed to the user are quite different. The captured sensor information goes through an image-processing pipeline in the camera's processing unit which returns a visually appealing image to the user (Rang, Prasad, and Brown, 2014; Brown, 2019). In this section, we review a camera's imaging processing pipeline and describe each feature which must be applied to produce the finished image or video returned to the user[1].

The in-camera image processing pipeline proceedes as follows; see the flowchart in Figure 5.2 for reference.

1. *Camera Lens Array*: Light is collected through the camera's lens array and focused into the photosensor.
2. *Sensor with colour filter array*: Focused light passes through a colour filter array, splitting the light into Red, Green, and Blue (RGB) wavelengths, and then striking the charged photodiode photosensor.
3. *Raw-image processing*: The imaging sensor signal is amplified and digitized. Black light subtraction is applied to cancel sensor noise. Defective pixels in the CMOS sensors due to manufacturing are interpolated.

[1]What we describe here is a general camera image processing pipeline. Individual camera and smartphone manufacturers have their own proprietary procedures which are applied to each image, however, they always typically follow the general processing pipeline as outlined here.

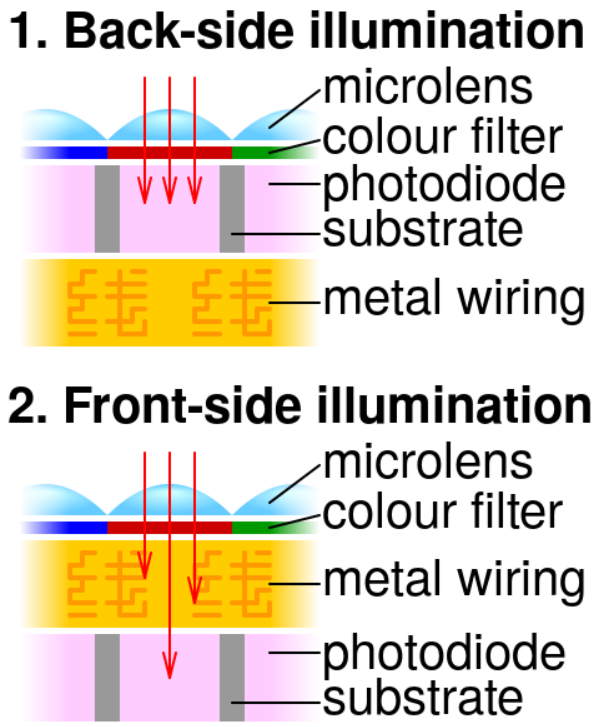

Figure 5.1: Comparison between 1. Back-side illumination and 2. Front-side illumination CMOS digital camera sensors. The metal conducting wiring of the camera array is moved below the photosensitive photodiode in the back-side illuminated sensor, allowing more light to reach the photodiode than in the Front-side illuminated sensor. Diagram courtesy of Cmglee, 2019.

4. *RGB Demosaicing*: The colour filter array placed over the pixel sensors allows only certain wavelengths of light through. To collect RGB values, interpolation is performed for each pixel.
5. *Noise reduction*: Images are slighly blurred to account for sensor noise. Smartphones apply aggresive noise reduction due to the small sensor size.
6. *White Balance and Colour Space Transform*: White balancing applies a diagonal matrix to the RAW RGB pixel values to account for the sensor's response to the image illuminant. The white balanced image is then transformed to a device independent colour space for rendering.
7. *Colour Manipulation*: The image is made to look pleasing here through the application of propreitary algorithms, known as non-linear tone-curve mapping.
8. *Mapping to sRGB Output*: The tone-curved images are mapped from the modified device independent colour space to the sRGB colour space (Incorporated, 2005).

9. *Encoding and Compression*: The image is encoded and compressed using a compression algorithm to prepare for storage.
10. *Save to file*: The image is stored in digital memory for viewing.

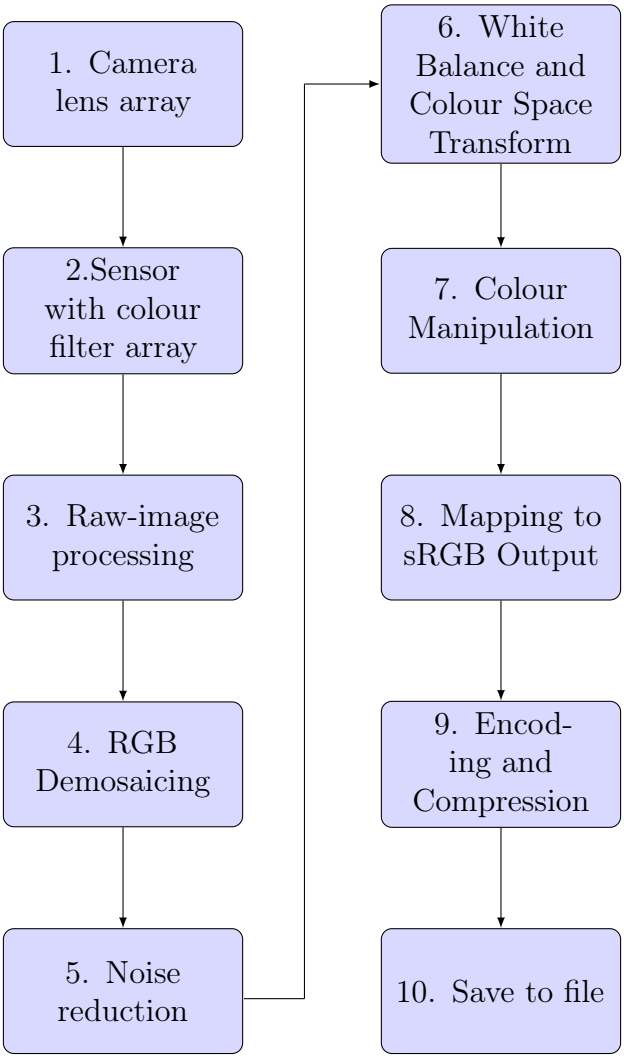

Figure 5.2: In-camera image processing pipeline

The RAW RGB pixels formed in Steps 1 through Steps 5 of the imaging pipeline (see Figure 5.2) may be modelled as a Lambertian colour model (Wyszecki and Stiles, 1982) where the pixel values i_C, $C = \{R, G, B\}$ are generated as,

$$i_C = \int_0^\infty E(\lambda) S_C(\lambda) R(\lambda)\, d\lambda \tag{5.1.1}$$

where λ is the wavelength, $E(\lambda)$ is an illuminant model of light on the imaged subject, $S_c(\lambda)$, is the camera's spectral response function for $C = \{R, G, B\}$ colour bands, and $R(\lambda)$ is the

imaged subject's reflectance. As cameras are manufactured with spectral filters to only permit light in the visible spectral range [400nm, 700nm], Equation (5.1.1) need only be integrated over this range. Further, we discretise the spectral range into m evenly spaced intervals of length $\Delta\lambda$ (typically 10nm), resulting in,

$$i_C = \int_{400}^{700} E(\lambda) S_C(\lambda) R(\lambda)\, d\lambda \approx \boldsymbol{i}_c = \sum_{i=1}^{n} E(\lambda_i) S_C(\lambda_i) R_C(\lambda_i) \Delta\lambda \tag{5.1.2}$$

In matrix, form we have,

$$\boldsymbol{i}_C = \boldsymbol{s}_C \boldsymbol{E} \boldsymbol{R} \tag{5.1.3}$$

where $\boldsymbol{i}_C$ is a $1 \times n$ vector for n pixels, $\boldsymbol{s}_C$ is a $1 \times m$ vector,$\boldsymbol{E} = \text{diag}(E(400\text{nm}), \ldots, E(700\text{nm}))$, and $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m)$, where $\boldsymbol{r}_m = (R_m(400\text{nm}), \ldots, 700\text{nm}))^\top$.

The remaining portions of the image processing pipeline then convert the recorded RAW pixel values to the rendered coloured image displayed to the observer. We model this following similar transformations to S. J. Kim et al., 2012 and Chakrabarti, Xiong, et al., 2013. The imaging pipeline forms a deterministic map $\mathbb{J} : \mathbb{R}^3 \to \mathbb{R}^3$ from the RAW tricolor sensor measurements at pixel $\boldsymbol{x} \in [0, 1]^3$ to the corresponding rendered colour space values $\boldsymbol{y} \in \{[0, 255] \cap \mathbb{Z}\}^3$. The mapping $\mathbb{J} : \boldsymbol{x} \to \boldsymbol{y}$ is given by,

$$\tilde{\boldsymbol{y}} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \end{pmatrix} = \begin{pmatrix} f(\boldsymbol{v}_1^\top \boldsymbol{x}) \\ f(\boldsymbol{v}_2^\top \boldsymbol{x}) \\ f(\boldsymbol{v}_3^\top \boldsymbol{x}) \end{pmatrix} \tag{5.1.4}$$

$$\boldsymbol{y} = Q\left(B(\tilde{\boldsymbol{y}}) + \begin{pmatrix} g_1(\tilde{\boldsymbol{y}}) \\ g_2(\tilde{\boldsymbol{y}}) \\ g_3(\tilde{\boldsymbol{y}}) \end{pmatrix} \right) \tag{5.1.5}$$

where $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3 \in \mathbb{R}^3$ defines a linear colour space transform, $B(\cdot)$ is a bounding function, restricting the values to $[0, 255]$, g_1, g_2, g_3 are colour gamut mapping functions, and $Q(\cdot)$ is a quantization function mapping the arguments to 8-bit integers. The linear colour transformation vectors $\boldsymbol{v}_i$ may be grouped in a matrix $T_v \in \mathbb{R}^{3\times3}$.

The model in (5.1.4) applies a per-channel polynomial model as given in Grossberg and Nayar, 2004 and Chakrabarti, Scharstein, and Zickler, 2009. We take f as $d = 7$ degree polynomial,

$$f(x) = \sum_{i=0}^{d} \alpha_i x^i \tag{5.1.6}$$

where α_i are model parameters. The application of T_v and the polynomial f are denoted in Step 6 of the camera imaging pipeline in Figure 5.2.

As in S. J. Kim et al., 2012, the per-channel polynomial map is augmented by a colour correction gamut-mapping function $g(\cdot)$ which is used to account for the differences in

the colour processing due to the camera manufacturer's proprietary colour manipulation algorithms. This is in reference to Step 7 of the camera imaging pipeline in Figure 5.2.

In the following sections we describe methods for recovering both the camera's spectral sensitivity functions from captured RAW images, and colour mapping functions from pairs of RAW and colour mapped rendered images.

5.2 Spectral Sensitivity Function Estimation

The spectral sensitivity functions of each camera are proprietary trade-secrets of the camera manufacturer or may be unkown. As such, they must either be measured in a laboratory using a monochromator and spectrometer Nakamura, 2006; Mauer and Wueller, 2009; Burggraaff et al., 2019 which is time consuming and costly, or estimated from a recorded RAW image. In this section we will describe the implemented methodology used to estimate the camera's spectral sensitivity functions, $S_C(\lambda)$, $C = \{R, G, B\}$ from (5.1.1).

Some of the first work to attempt to estimate spectral sensitivities was done by Hardeberg, Brettel, and Schmitt, 1998 who used known illumination conditions and images with known scene reflectances. Their results, however, were unreliable due to the low intrisic dimensionality of a scene's spectral reflectance (Hardeberg, Brettel, and Schmitt, 1998; Hubel, Sherman, and Farrell, 1994). Other work has since followed using different illumination conditions, including LED-based emissive targets (Dicarlo, Montgomery, and Trovinger, 2004) and colour targets under different LED illuminants (Urban et al., 2010).

From a statistical estimation perspective, work on estimating the spectral response functions was done by Slater and Healey, 1998 and Grossberg and Nayar, 2004, inspired by analysis performed for the spectrum of daylight performed almost 60 years ago by Judd, 1964, who proposed that the spectrum of daylight can be well represented by a small number of parameters. Extensive work was performed by Zhao et al., 2009 who collected a data from 12 cameras and compared three different types of basis functions - Fourier bases, polynomial bases, and radial bases - to model the spectral response functions. They found that RBF performed the best for recovery. Inspired by these works, Jiang et al., 2013 collected calibrated reflecton images captured from 28 different camera models under standard illuminant conditions and used principal component analysis (PCA) (Bishop, 2006) to recover a two-dimensional space of spectral sensitivities with remarkable accuracy. They found that just two eigenvalues were able to explain more than 97% of variance in their collected dataset.

Recent work concerning spectral sensitivity has tried to better approximate the more specific components of the camera's image formulation process (Qiu and Xu, 2016), constraint based linear estimation techniques (Finlayson, Darrodi, and Mackiewicz, 2016) and non-linear constraint optimisation (Prasad, Nguyen, and Brown, 2013). From a different perspective, other researchers have begun to use neural network and deep learning based techniques (Chaji et al., 2018; Kaya, Can, and Timofte, 2018) with respectable results, although these methods have yet to become widespread in the literature.

To recover the spectral sensitivity function of the Apple iPhone 6S used for experimental video recordings, we apply the PCA-based method of Jiang et al., 2013 due to its ease of implementation and use, as well as very high accuracy.

5.2.1 Experimental Setup and Data Collection

The PCA based method of Jiang et al., 2013 requires the capture of a RAW pixel image with known spectral reflectance content. To do so, we used the Halide Camera App (Lux Optics LLC., San Mateo, CA.) which permits the capture of RAW pixel image data, installed on an iPhone 6S. Using the Halide App, we photographed a Macbeth Colour checker chart, the X-Rite ColorChecker Passport Photo 2 (X-Rite, Inc., Grand Rapids, MI), which contains 24 Pantone colour swatches with known spectral reflectances. The ColorChecker chart is displayed in Figure 5.3. The spectral reflectances across the visible wavelength for each colour swatch were recorded at 5 nm increments by McCamy, Marcus, and Davidson, 1976 for reference. In Figure 5.4 we display the scaled spectral reflectance factors of each of the 24 colour swatches of the X-Rite ColorChecker Passport Photo 2. The images are captured under a known illuminant LED system, replicating the D65 CIE illuminant to ensure a known spectral illuminantion in the photo (*Standard Method of Assessing the Spectral Quality of Daylight Simulators for Visual Appraisal and Measurement of Colour* 2004). We capture both the RAW pixel and colour-rendered image pair. We additionally capture an entirely black image with the Halide App to determine the iPhone camera's noise response for black-light subtraction.

The captured RAW images are stored as large files (11 Mb) in the .DNG image format, which cannot be displayed on conventional image displays. To convert the image to a readable format for display purposes, we use the *dcraw* software package (Decoding raw digital photos in Linux, Dave Coffin).[2] Black light subtraction is applied to the RAW image by subtracting from the image of the ColorChecker chart the black image response.

The iPhone camera sensor uses a Blue, Green, Green, Red (BGGR) Bayer colour filter array to filter red, green, and blue wavelengths of light, as displayed in Figure 5.5. The Bayer filter is arranged in a mosaic pattern, which must be demosaiced to form the image. We apply a simple demosaicing procedure of selecting each filtered pixel value and assigning it to the proper RGB container.

Once the image has been demosaiced, we use an image registration algorithm to find the centre location of each of the 24 colour swatches. The algorithm requires the user to select the four corners of the ColorChecker chart, and then imposes a 4×6 grid structure onto the image, extracting the RAW RGB pixel values from the centre of each grid location. These pixel values are then used for analysis. See Figure (5.6) for a demonstation of the image registration algorithm.

[2]We run the command *./dcraw -4 -D -j -v -t 0* on the .DNG raw image to decode the raw image format.

Figure 5.3: Macbeth 24 swatch ColorChecker chart. Each colour swatch is manufactured to have precise known specral reflectance for colour constancy and accuracy in digital photography.

5.2.2 Algorithmic Assumptions

We make the following assumptions for the PCA-based spectral sensitivity estimation algorithm.

(A1) Spatial invariance–$S_C(\lambda)$, $C = R, G, B$.

(A2) Non-negativity–$S_C(\lambda) \geq 0$, $C = R, G, B$.

(A3) Luther condition (Nakamura, 2006)

The Luther condition states that the camera spectral sensitivity functions need to be a linear transformation of the CIE-1931 2-degree colour matching functions (CMF), a set of functions designed to ensure colour constancy between observers (*Standard Method of Assessing the*

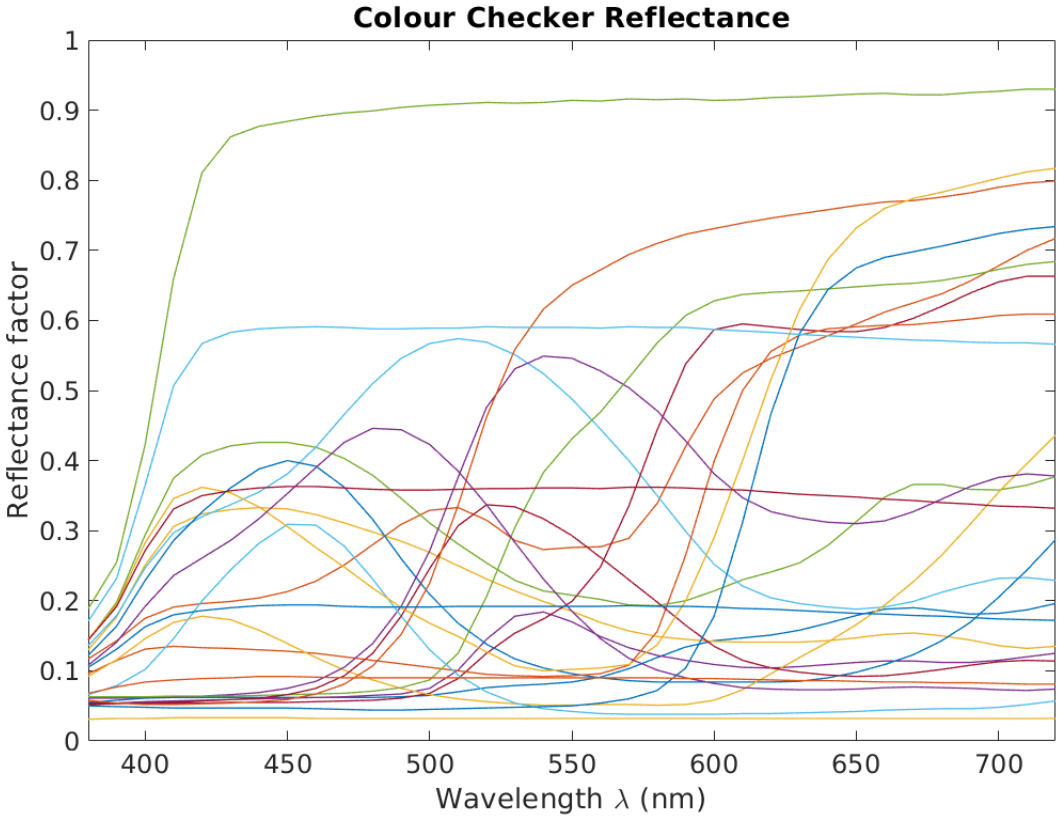

Figure 5.4: Scaled reflectance factors for the Macbeth X-Rite ColorChecker Passport Photo 2 as a function of wavelength from 380nm to 720nm. Each line corresponds to the scaled reflectance for each colour swatch. We have omitted the legend to avoid display clutter.

Spectral Quality of Daylight Simulators for Visual Appraisal and Measurement of Colour 2004).

$$\begin{pmatrix} \bar{\boldsymbol{x}} \\ \bar{\boldsymbol{y}} \\ \bar{\boldsymbol{z}} \end{pmatrix} = T \begin{pmatrix} \boldsymbol{s}_R \\ \boldsymbol{s}_G \\ \boldsymbol{s}_B \end{pmatrix} \tag{5.2.1}$$

where $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{z}})^\top$ is the CIE-1931 2-degree CMF, $(\boldsymbol{s}_R, \boldsymbol{s}_G, \boldsymbol{s}_B)^\top$ are the spectral sensitivities of a digital camera, and $T \in \mathbb{R}^{3\times 3}$ is a full-rank matrix to be determined. Due to the Bayer colour filters, however, the Luther condition is typically satisified for consumer-grade cameras.

Assumptions (A1)–(A3) define the digital camera's space of spectral sensitivity functions, which is convex by construction.

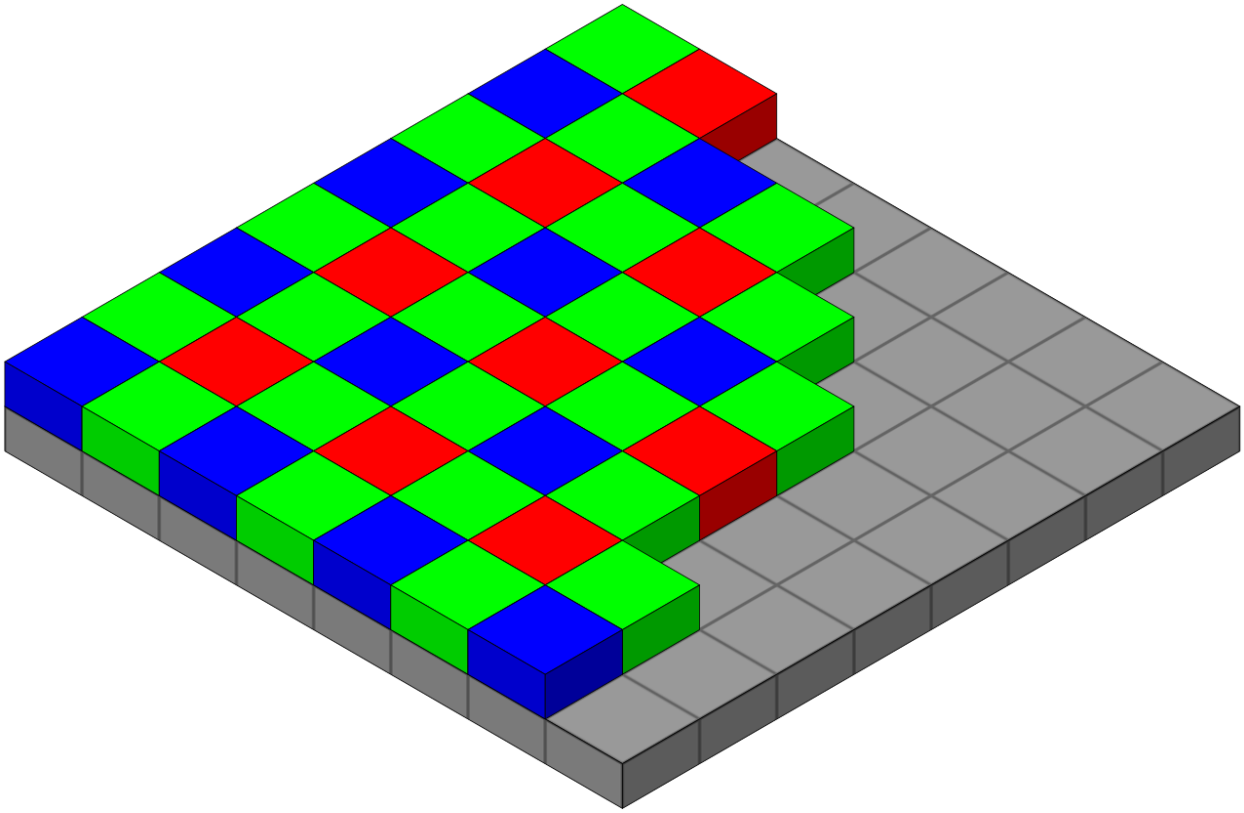

Figure 5.5: Blue Green Green Red (BGGR) Bayer colour filter placed above a digital camera's photosensor.

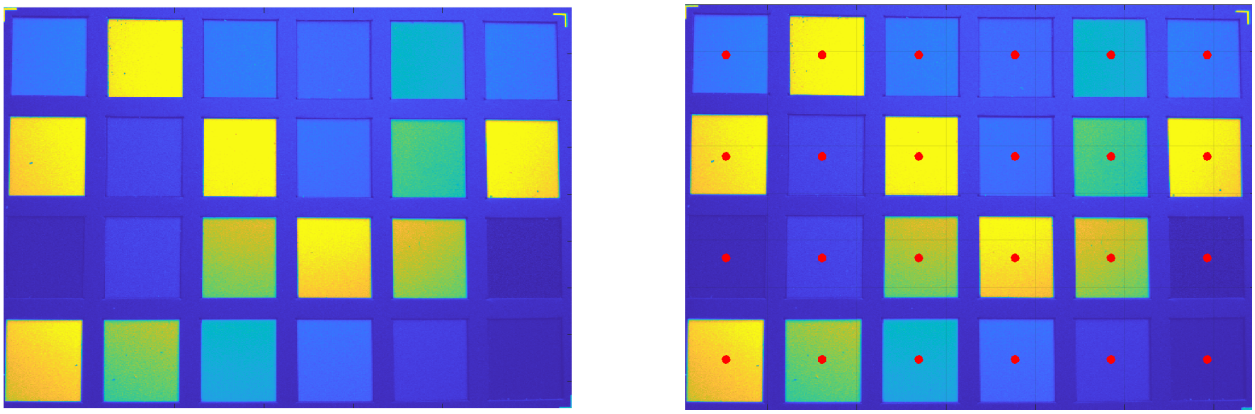

Figure 5.6: X-Rite ColorChecker Passport Photo 2 image registration demonstation. Left: The registered ColorChecker chart. Right: The registered ColorChecker chart with imposed grid structure and colour swatch centre locations marked with red circles.

The spectral sensitivity functions are also normalised to be between zero and one,

$$\max_{\lambda} \boldsymbol{s}_{C,n}(\lambda) = 1,\ C = R, G, B \tag{5.2.2}$$

where $\boldsymbol{s}_{C,n}$ is the normalised spectral sensitivity, and $\boldsymbol{s}_C = g_C\, \boldsymbol{s}_{C,n}$, $g_C \geq 0$, $C = R, G, B$. This normalisation is imposed to improve the performance of the statistical estimation procedure to capture the spread of the sensitivity functions.

5.2.3 PCA-Based Sensitivity Estimation

We then perform principal component analysis (PCA) on the demosaiced and normalised RAW pixel values for each colour channel separately. For the iPhone 6S image, we found that the first two eigenvalues or principal components captured 97.28% of the variance of the data, which is in agreement with the findings of Jiang et al., 2013.

That is, we decompose the sensitivity functions as $\boldsymbol{s}_{C,n} = \sigma_C \boldsymbol{A}_C$ for $C = R$, G, G where $\sigma = [\sigma_1, \sigma_2] \in \mathbb{R}^{1\times 2}$ are the principal components and $\boldsymbol{A}_C = [\boldsymbol{a}_{C,1}, \boldsymbol{a}_{C,2}]^\top \in \mathbb{R}^{2\times 33}$ is the eigenvector matrix. Equation (5.1.3) is then,

$$\boldsymbol{i}_C = g_C \sigma_C \boldsymbol{A}_C \boldsymbol{E} \boldsymbol{R}. \tag{5.2.3}$$

Given that we use a known illuminant $\boldsymbol{E}$, we recover the sensitivity functions from the demosaiced and normalised RAW pixel values by inversion, as

$$g_C \sigma_C = \boldsymbol{i}_C \left(\boldsymbol{A}_C \boldsymbol{E} \boldsymbol{R}\right)^{+} \tag{5.2.4}$$

for $C = R, G, B$, where $+$ is the Moore-Penrose pseudoinverse (B.1.13) The sensitivity functions are then given as,

$$\boldsymbol{s}_C = g_C \boldsymbol{s}_{C,n} = g_C \sigma_C \boldsymbol{A}_C = \boldsymbol{i}_C \left(\boldsymbol{A}_C \boldsymbol{E} \boldsymbol{R}\right)^{+} \boldsymbol{A}_C \tag{5.2.5}$$

for $C = R, G, B$. The estimated normalized RGB sensitivity functions for the iPhone 6S are displayed in Figure 5.7.

5.3 Colour Mapping Estimation

There have been several proposed methods to model the colour gamut functions. S. J. Kim et al., 2012 proposed a linear radial basis function (RBF) (Buhmann, 2003) approach to model the inverse gamut mapping g^{-1} to recover the colour mapping function from the paired RAW and colour rendered images. Chakrabarti, Xiong, et al., 2013 propose a similar approach and use a support-vector regression model with Gaussian RBFs. A different approach was taken by Li, Jia, and Zhang, 2018 who used a logarithmic tone-curve for the colour gamut map.

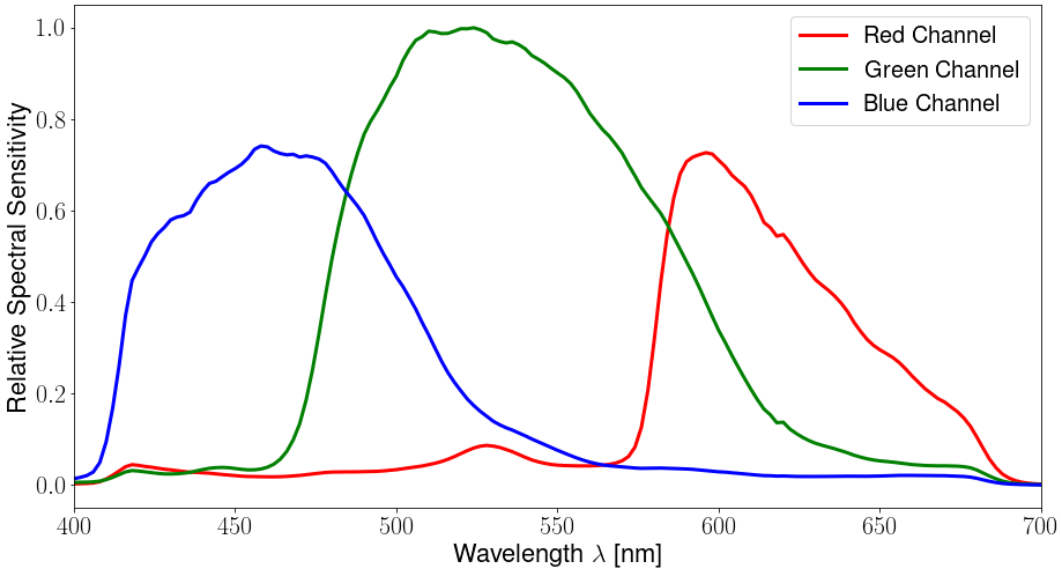

Figure 5.7: Estimated RGB spectral sensitivity functions of iPhone 6S.

To estimate the colour mapping functions of the iPhone 6S, we apply the methodology of Chakrabarti, Xiong, et al., 2013, however, we modify their method to better deal with our limited data size. While Chakrabarti, Xiong, et al., 2013 analysed a data set of more than 8000 colour swatches under different illuminantion conditions, our data set is far more limited to only 24 colour swatches from the X-Rite Colour ColorChecker Passport 2. Given the excellent performance of Gaussian process regression models with small data setes, we replace the support-vector regression model used for colour gamut mapping 5.1.5 with a Gaussian process regression model with a squared-exponential kernel, as described in Section 3.2.

Under a GP model, the deviations resulting from camera processing as given by g_c, $c \in \{1, 2, 3\}$ in 5.1.5 take the form,

$$g_c(\tilde{\boldsymbol{y}}) \sim \mathcal{GP}(\boldsymbol{\mu}_y, k_{SE}(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{y}}')). \tag{5.3.1}$$

with mean function $\boldsymbol{\mu}_y$, and kernel $k_{SE}(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{y}}')$.

The model parameters are determined using the paired RAW and JPEG images collected using the Halide and conventional camera application for the iPhone 6S as described in Section 5.2.1. The paired pixel values $\{(\boldsymbol{x}_t, \boldsymbol{y}_t)\}_{t=1}^{T}$, $T = 24$ form the calibration set. Parameters are estimated by solving the optimisation problem,

$$\{\hat{\boldsymbol{v}}_c\}, \{\hat{\alpha}_i\} = \underset{\{\boldsymbol{v}_c\},\{\alpha_i\}}{\arg\min} \sum_t \sum_c ||f\left(\boldsymbol{v}_c^\top \boldsymbol{x}_t\right) - y_{t:c}||_2^2 \tag{5.3.2}$$

where $\{\alpha_i\}$ are restricted such that f is monotonically increasing.

The cost (5.3.2) is minimised by alternating optimisation over either the linear $\{\boldsymbol{v}_c\}$ or polynomial $\{\alpha_i\}$ parameters, respectively, while holding the other parameters fixed. As the cost is quadratic in $\{\alpha_i\}$, fixing $\{\boldsymbol{v}_c\}$ yields a standard quadratic program problem, where the monotonicity restriction is a linear inequality constraint (Boyd and Vandenberghe, 2010). We use gradient descent to find the linear parameters $\{\boldsymbol{v}_c\}$ with fixed $\{\alpha_i\}$.

The alternating direction optimisation scheme is begun assuming $f(\boldsymbol{x}) = \boldsymbol{x}$ and setting $\{\boldsymbol{v}_c\}$ using a least squares solution, and proceeding until convergence. As the cost (5.3.2) is not convex, however, we have no guarantees of convergence to a global minimum. Chakrabarti, Xiong, et al., 2013 suggest a procedure whereby restarting the optimisation multiple times with different estimates of $\{\boldsymbol{v}_c\}$ acts as random deviations abou the current optimum, and will improve the convergence results.

GP regression is used to model the colour gamut mapping functions $\{g_c\}$ to fit the deviation from the rendered colour image $\tilde{\boldsymbol{y}}$ from the camera model and the JPEG image $\boldsymbol{y}$, where $\tilde{\boldsymbol{y}}$ is computed using 5.1.4 with the parameters learned above.

To evaluate the performance of the fitted colour mapping pipeline for the iPhone 6S, we compute the root mean-squared error (RMSE),

$$RMSE = \left(\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \right)^{1/2} \tag{5.3.3}$$

between the predicted JPEG values, $\hat{\boldsymbol{y}}$ and the actual JPEG values $\boldsymbol{y}$ in terms of the gray levels for an 8-bit image. We find an RMSE of 13.45, which is higher than the RMSE values mentioned in Chakrabarti, Xiong, et al., 2013. This discrepancy between errors is due to sample size. Chakrabarti, Xiong, et al., 2013 trained their model on more than 8000 samples of RAW and JPEG pixel values, compared to the 24 pixel samples we were able to obtain. As such, we believe the estimated colour mapping from RAW to rendered JPEG for the iPhone 6S is a reasonable representation of the true on-board processing algorithms.

Chapter 6

Skin Rendering and Inverse Mapping

> "*Light in tissue is not only heavily scattered but also attenuated. Therefore the degree of signal realistically obtainable varies by orders of magnitude if the operating point is changed.*"
>
> — Simon Arridge
> *Inverse Problems*

In the previous chapters we have constructed the constituent parts of the multi-stage opto-physiological modelling methodology for skin rendering and inverse chromophore recovery including diffuse reflectance modelling in Chapter 4, a camera imaging pipeline in Chapter 5, and the proposed inverse modelling framework of Gaussian Processes introduced in Chapter 3. In this chapter, we will combine these components to form rendered skin images in the RGB colour palette, and then learn the inverse chromophore mapping to recover the chromophore concentrations given an image of skin tissue.

Section 6.1 begins by forming a multiple layered model of human skin and computing the reflectance of incident light in skin tissue by explicitly modelling the absorption and scattering coefficients of each layer of skin, including inter-layer reflectance and transmittance. In Section 6.2 we combine the computed reflectance of incident light in skin tissue with the recovered camera imaging processing pipeline to form a comprehensive training set of skin colour swatches as rendered by an Apple iPhone 6S. We follow this in Section 6.3 by using a Gaussian Process based methodology to learn the inverse mapping from RGB pixel values of skin images to the chromophore concentrations using the rendered training set from the previous section. In Section 6.4 we demonstrate the results of the inverse chromophore map by applying it to both real and simulated skin images to uncover a map of the underlying subdermal vasculature across a variety of conditions. In addition, we implement a simplistic simulation of blood flow through the embedded blood vessels of a skin image by varying

the chromophore concentration in the vessels, and apply the inverse chromophore map to accurately recover the temporal chromophore dynamics. This serves as the first example of the ultimate application of the inverse chromophore mapping methodology for robust imaging diagnostics in Chapter 7.

6.1 Skin Rendering

We model the skin as a three layer plane-parallel slab medium consisting of *(i). stratum corneum, (ii). epidermis*, and *(iii). dermis* as in Section 4.5. For each layer, we compute the total reflectance and transmittance as,

$$R = \int_0^\infty R(\rho) 2\pi\rho \, d\rho \tag{6.1.1}$$

$$T = \int_0^\infty T(\rho) 2\pi\rho \, d\rho \tag{6.1.2}$$

where $R(\rho)$ and $T(\rho)$ are computed from (4.5.38) and (4.5.39), respectively. Applying the recursive layered transmittance and reflectance formulae (4.5.51) and (4.5.52), we compute the reflectance and transmittance for each layer individually, requiring us to compute six quantities, $R_1(\lambda), R_2(\lambda), R_3(\lambda)$, and $T_1(\lambda), T_2(\lambda), T_3(\lambda)$, where we have included the dependence on the wavelength of light λ as the absorption coefficients $\mu_a(\lambda)$ and $\mu_a'(\lambda)$ are dependent on the wavelength due to the constituent chromophores of the skin. In the following subsections, we explicitly model the absorption and scattering coefficients for each layer of the skin.

6.1.1 Layer 1: Stratum Corneum

The first layer of skin tissue is the very thin stratum corneum. The absorption and scattering coefficients are (Krishnaswamy and Baranoski, 2004)

$$\mu_{1,a}(\lambda) = \mu_a^{base}(\lambda) + \mu_a^{cs}(\lambda) \tag{6.1.3}$$

$$\mu_{1,s}(\lambda) = 2 \cdot 10^5 \cdot \lambda^{-1.5} + 2 \cdot 10^{12} \cdot \lambda^{-4} \tag{6.1.4}$$

where $\mu_a^{base}(\lambda)$ is the absorption coefficient of baseline skin, and μ_a^{cs} is the absorption coefficient of β-carotene in the stratum corneum,

$$\mu_a^{base}(\lambda) = 0.244 + 85.3 \exp\left\{\frac{-(\lambda - 154)}{66.2}\right\} \tag{6.1.5}$$

$$\mu_a^{cs} = \frac{\epsilon^{car}(\lambda) c^{cs}}{m^{car}} \tag{6.1.6}$$

where $\epsilon^{car}(\lambda)$ is the extinction coefficient of β-carotene, c^{cs} is the concentration of β-carotene in the stratum corneum, and m^{car} is the molecular weight of β-carotene. Each symbol is listed in Table 6.1. The values for these parameters are noted in Table 6.2. The extinction coefficient curves are found in the literature from Steven L. Jacques, Prahl, and Lindsey, 2001.

6.1.2 Layer 2: Epidermis

The second layer of skin below the stratum corneum is the epidermis. We make the simplifying assumption that the epidermis only contains melanin and does not contain hemoglobin. The absorption and scattering coefficients are (Krishnaswamy and Baranoski, 2004; Donner, Weyrich, et al., 2008),

$$\mu_{2,a}(\lambda) = (\gamma_e \mu_a^{em}(\lambda) + (1-\gamma_e)\mu_a^{pm}(\lambda))\theta_m + (\mu_a^{base}(\lambda) + \mu_a^{ce}(\lambda))(1-\theta_m) \tag{6.1.7}$$

$$\mu_{2,s}(\lambda) = \mu_{1,s}(\lambda) \tag{6.1.8}$$

where,

$$\mu_a^{em}(\lambda) = \frac{\epsilon^{em}(\lambda)c^{cm}}{m^{em}} \tag{6.1.9}$$

$$\mu_a^{pm}(\lambda) = \frac{\epsilon^{pm}(\lambda)c^{pm}}{m^{pm}} \tag{6.1.10}$$

$$\mu_a^{ce}(\lambda) = \frac{\epsilon^{car}(\lambda)c^{ce}}{m^{car}} \tag{6.1.11}$$

where μ_a^{em}, μ_a^{pm}, μ_a^{ce} are the absorption coefficients of eumelanin, pheomlenanin, and β-carotene in the epidermis, respectively; ϵ^{em}, ϵ^{pm}, are the extinction coefficients of eumelanin, and pheomlenanin, respectively; c^{em}, c^{pm}, c^{ce} are the concentrations of eumelanin, pheomelanin, and β-carotene in the epidermis, respectively. Each symbol is listed in Table 6.1. The values for these parameters are noted in Table 6.2.

6.1.3 Layer 3: Dermis

The third layer of skin below the epidermis is the dermis. This is where blood vessels are located above the layer of subcutaneous fat making up the last layer of tissue, the *hypodermis*. We make the simplifying assumption that hemoglobin is only contained in the dermis, and melanin is absent from the dermis. The absorption and scattering coefficients are (Krishnaswamy and Baranoski, 2004; Donner, Weyrich, et al., 2008),

$$\mu_{3,a}(\lambda) = \left(\gamma_d \mu^{ohb}(\lambda) + (1-\gamma_d)\mu_a^{dhb}(\lambda) + \mu_a^{bil}(\lambda) + \mu_a^{cd}(\lambda)\right)\theta_h + \mu_a^{base}(\lambda)(1-\theta_h) \tag{6.1.12}$$

$$\mu_{3,s} = 0.5\mu_{2,s}, \tag{6.1.13}$$

where

$$\mu_a^{ohb}(\lambda) = \frac{\epsilon^{ohb}(\lambda)c^{hb}}{m^{hb}} \tag{6.1.14}$$

$$\mu_a^{dhb}(\lambda) = \frac{\epsilon^{dhb}(\lambda)c^{hb}}{m^{hb}} \tag{6.1.15}$$

$$\mu_a^{bil}(\lambda) = \frac{\epsilon^{bil}(\lambda)c^{bil}}{m^{bil}} \tag{6.1.16}$$

$$\mu_a^{cd}(\lambda) = \frac{\epsilon^{car}(\lambda)c^{cd}}{m^{car}} \tag{6.1.17}$$

$$\tag{6.1.18}$$

where ϵ^{ohb}, ϵ^{dhb}, and ϵ^{bil} are the extinction coefficients of oxygenated hemoglobin, deoxygenated hemoglobin, and bilirubin, a chromophore which gives skin a yellow appearance, respectively; c^{hb}, c^{bil}, and c^{cd} are the concentrations of hemoglobin, bilirubin, and β-carotene in the dermis, respectively; m^{hb} and m^{bil} are the molecular weights of hemoglobin and bilirubin, respectively. All symbols can be found in Table 6.1. Typical parameter values and concentrations can be found in Table 6.2 and ranges for the variables can be found in Table 6.3.

Using the absorption and scattering coefficients for each layer of skin, we simulate the total reflectance and transmittance for a three layer skin model across the visible wavelength. In Figure 6.1 we display the total reflectance across the visible wavelength for several different chromophore concentrations. The left side depicts the reflectance computed by holding the volume fraction of melanin θ_m fixed and varying the volume fraction of hemoglobin θ_h. On the right, we show the converse situation of holding the volume fraction of hemoglobin θ_h fixed and varying the volume fraction of melanin θ_m.

In Figure 6.2 we display the total transmittance across the visible wavelength for several different chromophore concentrations. Like in Figure 6.1 the left side depicts the transmittance computed by holding the volume fraction of melanin θ_m fixed and varying the volume fraction of hemoglobin θ_h. On the right, we show the converse situation of holding the volume fraction of hemoglobin θ_h fixed and varying the volume fraction of melanin θ_m.

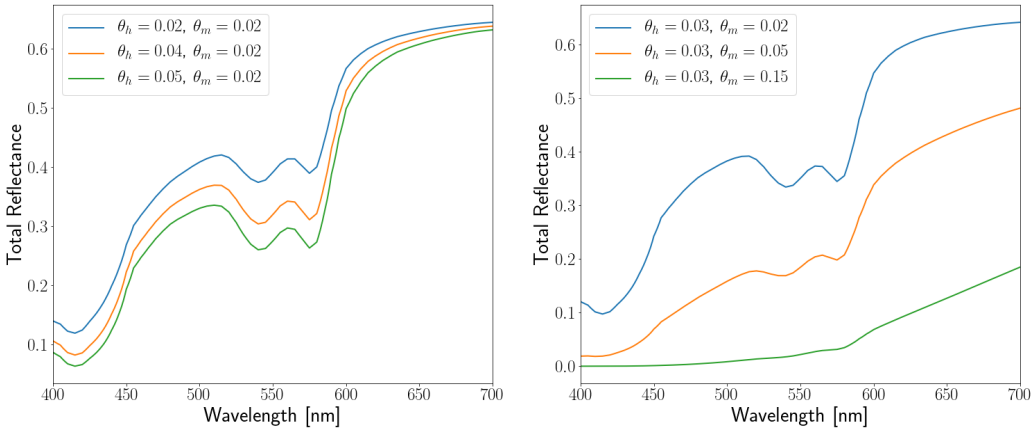

Figure 6.1: Total reflectance computed by varying the chromophore concentrations. Left: Reflectances observed by varying the volume fraction of hemoglobin θ_h and holding the volume fraction of melanin θ_m fixed. Right: Reflectances observed by varying the volume fraction of melanin θ_m and holding the volume fraction of hemoglobin θ_h fixed.

Table 6.1: List of symbols in skin layers.

Symbol	Definition
c^{cs}	concentration of β-carotene in stratum corneum
c^{em}	concentration of eumelanin in epidermis
c^{pm}	concentration of phaeomelanin in epidermis
c^{ce}	concentration of β-carotene in epidermis
c^{hb}	concentration of hemoglobin
c^{bil}	concentration of bilurin
c^{cd}	concentration of β-carotene in the blood
d^{sc}	depth of stratum corneum
d^{e}	depth of epidermis
d^{der}	depth of dermis
$\epsilon^{car}(\lambda)$	extinction coefficient of β-carotene
$\epsilon_a^{em}(\lambda)$	extinction coefficient of eumelanin
$\epsilon^{pm}(\lambda)$	extinction coefficient of phaeomelanin
$\epsilon_a^{ohb}(\lambda)$	extinction coefficient of oxygenated hemoglobin
$\epsilon^{dhb}(\lambda)$	extinction coefficient of de-oxygenated hemoglobin
$\epsilon^{bil}(\lambda)$	extinction coefficient of bilurin
γ_e	ratio control between eumelanin and phaeomelanin
γ_d	ratio control between oxy and de-oxygenated hemoglobin
$\mu_a^{base}(\lambda)$	absorption coefficient of baseline skin
$\mu_a^{cs}(\lambda)$	absorption coefficient of β-carotene in stratum corneum
$\mu_a^{em}(\lambda)$	absorption coefficient of eumelanin
$\mu_a^{pm}(\lambda)$	absorption coefficient of pheomelanin
$\mu_a^{ce}(\lambda)$	absorption coefficient of β-carotene in epidermis
$\mu_a^{ohb}(\lambda)$	absorption coefficient of oxygenated hemoglobin
$\mu_a^{dbb}(\lambda)$	absorption coefficient of de-oxygenated hemoglobin
$\mu_a^{bil}(\lambda)$	absorption coefficient of bilurin
$\mu_a^{cd}(\lambda)$	absorption coefficient of β-carotene in dermis
m^{car}	molecular weight of β-carotene
m^{em}	molecular weight of eumelanin
m^{pm}	molecular weight of phaeomelanin
m^{hb}	molecular weight of hemoglobin
m^{bil}	molecular weight of bilrubin
θ_h	volume fraction (%) of dermis occupied by hemoglobin
θ_m	volume faction (%) of epidermis occupied by melanosomes

Table 6.2: Parameter values.

Symbol	Value
c^{cs}	$2.1 \cdot 10^{-4}$ g/L
c^{ce}	$2.1 \cdot 10^{-4}$ g/L
c^{cd}	$7.0 \cdot 10^{-5}$ g/L
c^{em}	80 g/L
c^{pm}	12 g/L
c^{hb}	150 g/L
c^{bil}	0.05 g/L
γ_e	0.61
γ_d	0.75
m^{car}	537 g/mole
m^{em}	545 g/mole
m^{pm}	318 g/mole
m^{hb}	66500 g/mole
m^{bil}	585 g/mole

Table 6.3: Variable typical values

Variable	Ranges
θ_m	1.3 – 43%
θ_h	2 – 8%
d^{sc}	0.001 cm
d^e	0.021 cm
d^d	0.2 cm
g	0.9

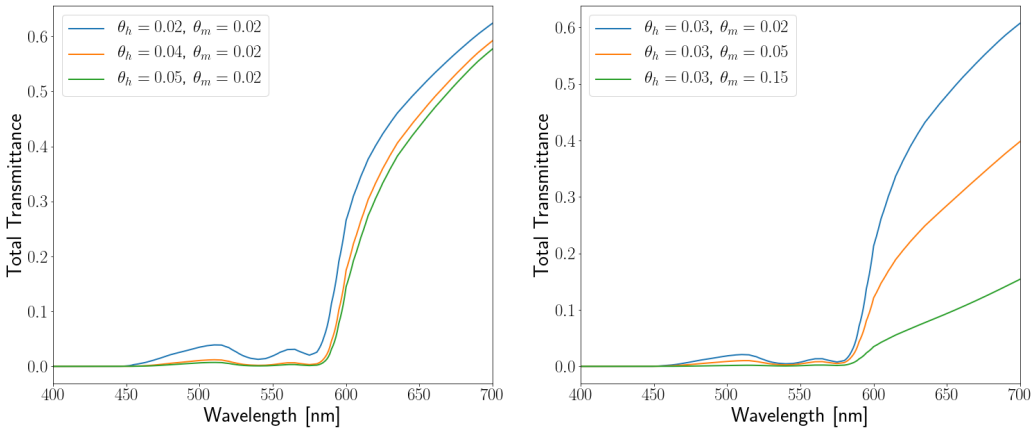

Figure 6.2: Total transmittance computed by varying the chromophore concentrations. Left: Transmittances observed by varying the volume fraction of hemoglobin θ_h and holding the volume fraction of melanin θ_m fixed. Right: Transmittances observed by varying the volume fraction of melanin θ_m and holding the volume fraction of hemoglobin θ_h fixed.

6.2 Camera Model of Skin Colour

To form a rendering of skin in a digital image we now apply the digital camera imaging pipeline as descibed in Chapter 5. We compute the layered skin reflectance across a range of chromophores and convolve with a model of the illuminant and the spectral response function of the camera as in the Lambertian camera model (5.1.1). Thus, for each pixel response, let $i_C, C = \{R, G, B\}$ be the pixel response, $E(\lambda)$ the model of the illuminant, and $S_C(\lambda)$ be the camera's spectral response function for pixel C. Thus,

$$i_C = \int_{400}^{700} E(\lambda) S_C(\lambda) R(\lambda)\, d\lambda \approx \boldsymbol{i}_c = \sum_{i=1}^{n} E(\lambda_i) S_C(\lambda_i) R_C(\lambda_i) \Delta\lambda. \tag{6.2.1}$$

We use the spectral response functions for the iPhone 6S we recovered in Section 5.2 for S_C and the CIE F2 standard illuminant (*Standard Method of Assessing the Spectral Quality of Daylight Simulators for Visual Appraisal and Measurement of Colour* 2004). We use the F2 standard illuminant for flourescent interior lighting, as we believe this will be the principal environment for which skin images will be captured for analysis. There exist other choices of illuminants such as the D50, D55, and D65 illuminants which represent the spectral illuminant for exterior settings under natural sunlight. Due to the modular construction of our methodology, the rendered skin pixel values can be computed across a number of environments by changing the illuminant.

We compute the RAW skin pixel values by exhaustively varying the volume fraction of hemoglobin θ_h and the volume fraction of melanin θ_m over physiologically relevant ranges.

The volume fraction of hemoglobin varies from 2% to 8%, while the volume fraction of melanin varies from 1.3% to 43% to simulate a variety of skin tones from very light skin tones through very dark skin tones. Once the Lambertian camera model is computed (6.2.1), we then apply the recovered in-camera imaging processing pipeline from Section 5.3 to render the skin pixel values as they would appear if they were photographed using the onboard camera of an Apple iPhone 6S. In Figure 6.3 we display the results of the imaging processing pipeline using the computed skin reflectance from exhaustive variation of the chromophore values.

The mains steps in the computation of the rendered skin image area

1. Reflectance and transmittance calculation for each skin layer
2. Total reflectance calculation using recursive formula
3. Lambertian colour model RAW image formulation
4. Digital camera image processing for rendered RAW image

The significant computational challenges are encountered in the reflectance and transmittance calculation for each layer for each chromophore concentration and the total reflectance calculation with the recursive formula for every concentration. The rendered skin image as displayed in Figure 6.3 consisting of 256×256 skin tones is formed in approximately 4.3 hours, for which the RAW image formulation and digital camera imaging pipeline are computed in less than one second.

6.3 Gaussian Process Chromophore Mapping

To recover the chromophore concentrations from an image of skin tissue using the RGB pixel values, we must compute the inverse mapping between the rendered or captured pixel values and the chromophore concentrations which cause their appearance. We begin with an abstraction of the skin image pixel value formulation process.

Let $\Theta = (\theta_h, \theta_m)$ be the vector of chromophores used to generate the skin pixels (R, G, B). The skin pixels are formed from the above procedure as

$$(R, G, B) = f(\Theta, E, S_C) \tag{6.3.1}$$

for $C = \{R, G, B\}$ for S_C. Since both the illuminant E and the spectral sensitivity functions S_C are known, the pixel values reduce to being a function of only the unknown chromophore values Θ. To recover Θ for the RGB pixel values, we must find f^{-1}.

There have been several studies that have attempted to find the inverse mapping from skin RGB pixel values to chromophore concentrations. Cotton and Claridge, 1996 performed one of the first studies to propose a biophysical inspired model of skin appearance from light-interaction in skin tissue, using a simpified Kubelka-Munk model of light propagation. They investigated the appearance of melanomas through biophysical rendering. They later

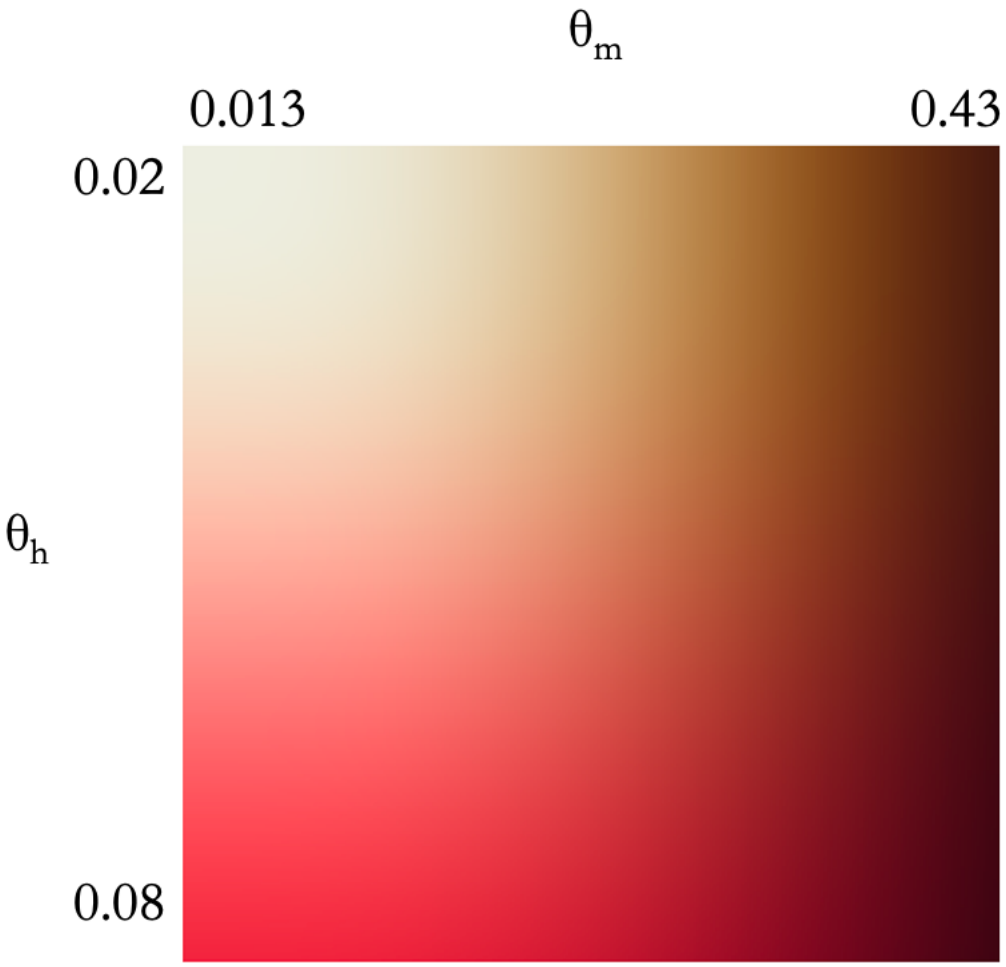

Figure 6.3: Rendered skin images produced by exhaustively varying the volume fractions of hemoglobin and melanin and applying an in-camera imaging processing pipeline. The volume fraction of hemoglobin θ_h varies from 0.02 to 0.08 down the vertical axis, while the volume fraction of melanin θ_m varies from 0.013 to 0.43 across the vertical axis.

extended this work in Claridge and Steve J. Preece, 2003; Claridge, Cotton, et al., 2003; Stephen J. Preece and Claridge, 2004 where they proposed an inverse mapping between their rendered skin images and the chromophores. The essential assumption of this inverse model is that f is a one-to-one mapping between the RGB pixel values and the chromophores Θ. Unfortunately, due to the image transformation processing pipeline and camera colour constancy, this assumption only holds for the true radiometric RAW pixel values, and not for the rendered image displayed by a digital camera. As such, these models will severely underperform in many conditions. Doi and Tominaga, 2003 also proposed the use of KM theory and used a least-squares solution to the inverse problem.

An alternative approach was provided by Tsumura, Haneishi, and Miyake, 1999 who used independent component analysis (ICA) to recover a map of the chromophores from digital images. While effective in many condiitons, the ICA model assumes linear dependence between chromophore quantities, which, due to the complex structure of skin tissue and light-interaction, is very likely to not generally hold (Tsumura, Haneishi, and Miyake, 2000). More recently, Alotaibi and Smith, 2018 has suggested the use of principal component analysis (PCA) to recover the approximate inverse mapping, with a simplified model of skin reflectance. There have also been attempts to use deep learning to both generate skin images and inverse maps, however, this work is still in its infancy (Saito et al., 2017; Huynh et al., 2018).

To approximate f^{-1} we use a Gaussian Process (GP) regression procedure as described in Chapter 3. That is, we construct $f_{GP} \approx f^{-1}$,

$$f_{GP}(R, G, B) = \Theta \tag{6.3.2}$$

where $f_{GP} \sim \mathcal{GP}(\mu, k)$ for mean μ and kernel k.

We implementd the structured kernel interpolation (SKI) GP approximation given in Section 3.3 in the GPyTorch Python package. We experimented with several different covariance functions, however found the best performance with using a Matern $\nu = 5/2$ kernel, and used a 25×25 inducing point approximation to the full covariance matrix.We performed Type II maximum likelihood estimation to tune the GP kernel hyperparameters, using the Adam stochastic optimization procedure (Kingma and Ba, 2015) with a learning rate of 0.1. Training was performed for 400 iterations, which took approximately 11 minutes on a workstation equipped with a Nvidia 1080 Ti GPU. We achieved a negative log-likelihood of -1.3218. Further iterations result in marginal improvements to the negative log-likelihood, however, this results in poorer fit to the data, due to overfitting. In Figure 6.4 we display the loss history of the negative log-likelihood as a function of training iterations for the Adam optimizer training the GP inverse map. In Figure 6.5 we compare the recovered GP inverse model to the actual hemoglobin values θ_h and the absolute error between the set values and the recovered values. We find relatively excellent agreement between the set θ_h and the values recovered from the GP model.

It is important to note that for the simulation of the reflected light from skin tissue and the rendered skin images that we have used relatively general coefficients and models of the

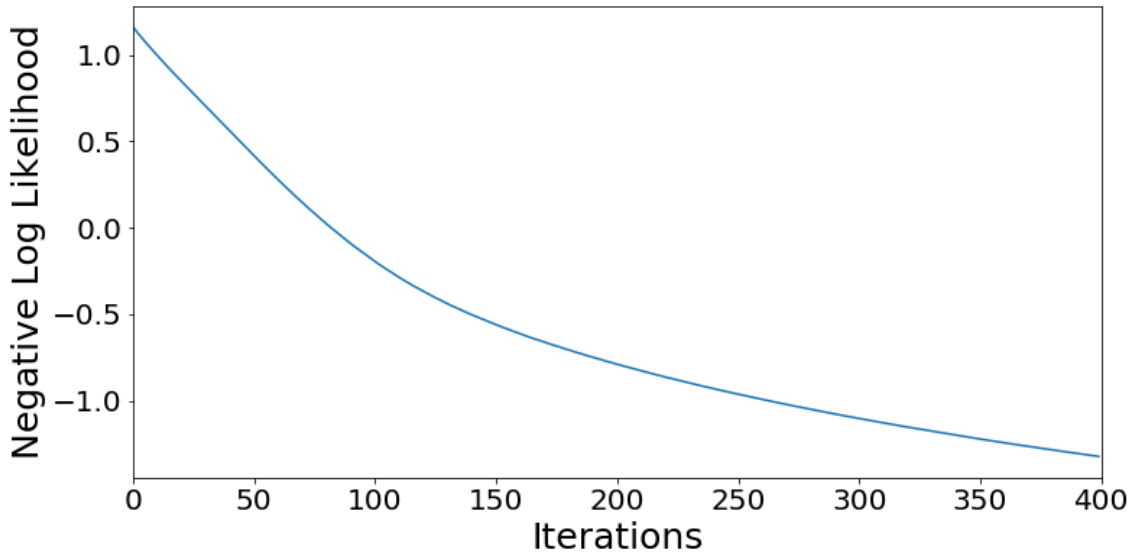

Figure 6.4: Loss history of the negative log-likelihood as a function of training iteration of the Adam optimizer for the Gaussian Process regression function.

absorption and scattering coefficients in each layer tissue. As such, our results are reasonably general across the majority of adult skin, however, there will be age dependent differences which arise due to the changing properties of skin tissue due to age. As skin ages, it loses its elasticity and a percentage of its moisture content, which will in some way affect the absorption coefficients in the layered skin model. While the scattering coefficients that we use in the model are general and consist of only Rayleigh and Mie scattering, the absorption coefficients will change due to these different properties. Since the purpose of this dissertation is concerned with the development of a general and modular methodology for the recovery of chromophore concentrations from digital images, the main source of error that we control arises from other factors such as differences in illumination and the digital camera imaging pipeline, while skin specific differences related to age and different mechanical properties will arise as a secondary source of error. Modelling the functional forms of the absorption coefficients under different skin properties is beyond the scope of this dissertation, but will prove fruitful as a direction of future research so as to generalise the method across a diverse population.

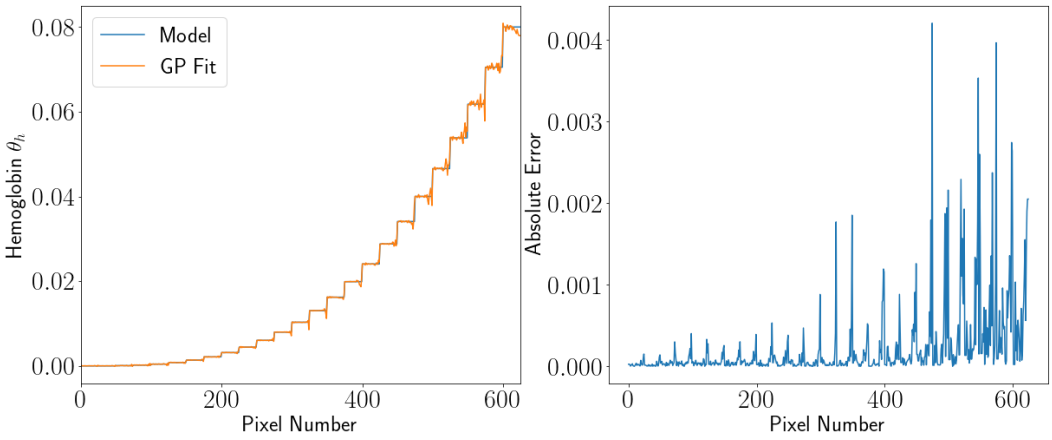

Figure 6.5: Model result and absolute error comparing the model hemoglobin values θ_h and the Gaussian Process approximated inverse map recovered from the rendered skin pixel values. Left: Gaussian Process approximated inverse map and true hemoglobin values comparison. Right: Absolute error for each rendered pixel value.

6.4 Vasculature Recovery

6.4.1 Monte-Carlo Simulations

Given the approximate Gaussian Process inverse chromophore map f_{GP}, we now apply the map to both simulated and real images to identify the hemoglobin distribution in the skin, and to recover a map of the underlying vasculature. As the blood vessels contain a greater concentration of hemoglobin than the rest of the surrounding tissue, and have a regular geometric structure, the application of f_{GP} to skin images should permit recovery of the vasculature by identifying these regions.

We first apply the inverse chromophore map to simulated skin images with blood vessels embedded in them. C. Zoller and Kienle, 2019 recently presented state-of-the-art software for Monte Carlo light propagation simulations in skin tissue with embedded blood vessels, as well as an RGB image rendering scheme. Their simulations use 10^7 photon packets for a two-layered skin model consisting of the epidermis and dermis. They produce a single precision lookup table (LUT) of stored results of roughly 10 Gb in size, as well as present a MATLAB (Mathworks In., Natick, Massachusetts) interface to control the parameters and render skin images.

Unlike our imaging processing pipeline of Chapter 5 which is designed to replicate the Apple iPhone 6S, C. Zoller and Kienle, 2019 apply an idealised camera model to produce their rendered images. Their camera model proceeds as,

$$\boldsymbol{I}_{\text{raw}} = C^{\top}\text{diag}(E)R \tag{6.4.1}$$

where they us the CIE-D65 illuminant (*Standard Method of Assessing the Spectral Quality of Daylight Simulators for Visual Appraisal and Measurement of Colour* 2004). The images are then transformed to linear RGB space,

$$\boldsymbol{I}_{\text{lRGB}} = T_{\text{xyz2rgb}}(SC^{+})\text{diag}(C^{\top}I)^{-1}\boldsymbol{I}_{\text{raw}} \tag{6.4.2}$$

where S is the CIE-1931 2-degree colour matching function, C^{+} pseudo-inverse, and

$$T_{\text{xyz2rgb}} = \begin{pmatrix} 3.2406 & -1.537 & -0.498 \\ -0.968 & 1.8758 & 0.0415 \\ 0.0557 & -0.204 & 1.0570 \end{pmatrix} \tag{6.4.3}$$

is a map from the CIE-XYZ colour space – the standardised colour constancy colour space – to the linear RGB colour space. Gamma-correction is then applied $\boldsymbol{I}_{\text{sRGB}} = \boldsymbol{I}_{\text{lRGB}}^{1/\gamma}$. To ensure accurate results, we apply the same imaging pipeline to our computed reflectance values and retrain the inverse chromophore map. Our training results are very similar to the results in Section 6.3.

The Monte Carlo simulations allow for the generation of different hemoglobin and melanin concentrations as well as different vessel depths for skin appearance with embedded blood vessels. The simulated images are formed with a blood vessel embedded along the centre line of the image, with symmetric skin tissue about the vessel on either side. Due to the variations of absorption and scattering properties caused by the blood vessel, we evaluate the accuracy of the method by properly identifying the location of the blood vessels, and identifying the hemoglobin concentration in the simulated skin tissue far away from the vessel. One significant challenge we pose to the vessel identification method is the ability to identify both deep vessels, as well as identify vessels in darkly pigmented skin tissue with high melanin concentrations due to the highly scattering properties of melanin.

The computational framework of the Monte Carlo simulations forms a 1.6 cm $\times$ 1.6 cm computational grid $x \in [-0.8, 0.8]$ cm, $y \in [-0.8, 0.8]$ cm, or 604 pixels $\times$ 604 pixels in the exported images from the simulations. This corresponds to a 0.00265 cm/pixel discretisation for image rendering and inverse chromophore recovery.

We evaluate the accuracy of the recovered vasculature map by comparing the computed vessel radius with the simulated vessel radius with respect to the 0.0265 mm/pixels discretisation accuracy. The radius of the blood vessel from the recovered vasculature map is computed by extracting the pixels with the greatest concentration of hemoglobin. We perform three condition sets of Monte Carlo simulations using the simulation package of C. Zoller and Kienle, 2019. Each simulation set consisted of fixed melanin and hemoglobin volume fractions, θ_m and θ_h, respectively, with varying vessel depth and radius.The embedded vessel depths

are $d = 0.05$, $d = 0.10$, $d = 0.15$ cm, with radii $r = 0.04$, $r = 0.06$, $r = 0.08$ cm, leading to nine simulations for each condition. The melanin and hemoglobin volume fractions are set, respectively, as, *(i).* $\theta_m = 0.025$, $\theta_h = 0.025$, *(ii).* $\theta_m = 0.22$, $\theta_h = 0.025$, and *(iii).* $\theta_m = 0.41$, $\theta_h = 0.025$. In Tables 6.4 – 6.6 we display the results of the recovered vasculature map for simulation conditions *(i)* – *(iii)*, respectively. In each of the simulation conditions, the inverse chromophore map achieves excellent vessel recovery performance as measured by absolute error. For simulation *(i)*, we achieve absolute error 0.0017 ± 0.0010 (mean $\pm$ standard devation); for simulation *(ii)*, we achieve absolute error 0.0057 ± 0.0036; for simulation *(iii)*, we achieve absolute error 0.0142 ± 0.0054.

As expected, the error increases as a function of increasing melanin volume fraction, due to the highly light scattering properties of melanin. Once reflected light interacts with the embedded blood vessel, it must travel back through the skin tissue to be captured by the camera. In rendered skin tissue with large concentrations of melanin, the increase of light scattering events due to this larger melanin concentration causes a blending of the light about the vessel, not constrained to the exact vessel position. We see this effect in the recovered vasculature map, as the highest error in the estimated radius is in simulation condition *(iii)* where the melanin concentration is greatest. Despite this, the precise position of the vessel is still recovered with excellent accuracy in every simulation condition.

A possible solution for the error scaling as a function of melanin concentration is to implement subject specific models for each group of skin pigments. The proposed general model we present here may be used at first application to identify the melanin concentration in the skin tissue, and then apply a secondary model afterward, following the same principals of construction of the proposed inverse chromophore map. The secondary model would be trained on much more specific data pertaining to hemoglobin changes in only a select range of melanin volume fractions, dramatically increasing the local sample size of the training set for the model, allowing for specific and more accurate results for each group of skin pigments. While this would increase the computational burden of the modelling in the training portion of model development, the application of the secondary skin pigment specific map would yield the same trivial computational expense the first inverse chromophore map takes.

Table 6.4: Predicted vessel radius from recovered vasculature map and absolute error for inverse chromophore map f_{GP} applied to Monte Carlo simulated skin images with an embedded vessel at varying depths d and radii r with *(i).* $\theta_m = 0.025$, and $\theta_h = 0.025$.

	$r_1 = 0.04$		$r_2 = 0.06$		$r_3 = 0.08$	
d	$\hat{r}$	$\lvert r_1 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_2 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_3 - \hat{r}\rvert$
0.05	0.0398	0.0002	0.0609	0.0009	0.0795	0.0005
0.10	0.0424	0.0024	0.0609	0.0009	0.0822	0.0022
0.15	0.0424	0.0024	0.0636	0.0036	0.0822	0.0022

Table 6.5: Predicted vessel radius from recovered vasculature map and absolute error for inverse chromophore map f_{GP} applied to Monte Carlo simulated skin images with an embedded vessel at varying depths d and radii r with *(ii)*. $\theta_m = 0.22$, and $\theta_h = 0.025$.

	$r_1 = 0.04$		$r_2 = 0.06$		$r_3 = 0.08$	
d	$\hat{r}$	$\lvert r_1 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_2 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_3 - \hat{r}\rvert$
0.05	0.0451	0.0051	0.0609	0.0009	0.0822	0.0022
0.10	0.0477	0.0077	0.0663	0.0063	0.0822	0.0022
0.15	0.053	0.013	0.0689	0.0089	0.0842	0.0048

Table 6.6: Predicted vessel radius from recovered vasculature map and absolute error for inverse chromophore map f_{GP} applied to Monte Carlo simulated skin images with an embedded vessel at varying depths d and radii r with *(iii)*. $\theta_m = 0.41$, and $\theta_h = 0.025$.

	$r_1 = 0.04$		$r_2 = 0.06$		$r_3 = 0.08$	
d	$\hat{r}$	$\lvert r_1 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_2 - \hat{r}\rvert$	$\hat{r}$	$\lvert r_3 - \hat{r}\rvert$
0.05	0.0504	0.0104	0.0663	0.0063	0.0875	0.0075
0.10	0.0583	0.0183	0.0716	0.0116	0.0954	0.0154
0.15	0.0636	0.0236	0.0795	0.0195	0.0954	0.0154

To assess the hemoglobin recovery accuracy of the inverse chromophore map, we simulate 20 skin samples with embedded vessels in the dermis at a fixed depth 0.13 cm using the Monte Carlo simulation package of C. Zoller and Kienle, 2019. The results of the hemoglobin estimation for different melanin concentrations are presented in Table 6.7. We performed 20 simulations with fixed hemoglobin volume fractions $\theta_h = 0.025$ and $\theta_h = 0.030$, (10 simulations each) respectively, varying the melanin concentration over the biophysically relevant ranges to observe the effect of increased scattering due to increasing melanin concentration. For $\theta_h = 0.025$ we achieve a relative error of 0.07 ± 0.050 (mean $\pm$ standard deviation). For $\theta_h = 0.030$ we achieve a relative error of 0.11 ± 0.099. We can see that the error in the hemoglobin volume fraction recovery increases with increasing melanin content, which demonstrates the difficulty caused by increased scattering.

In Figures 6.6 – 6.9 we present the results of the recovered hemoglobin distribution and embedded vasculature for four prototypical cases. Portion a) of each figure displays the rendered skin image with an embedded vessel produced using Monte Carlo simulation software from C. Zoller and Kienle, 2019; portion b) depicts the recovered hemoglobin distribution and vasculature map. In Figure 6.6 we present the results for a embedded vessel of radius $r = 0.04$ cm, at a depth of $d = 0.05$ cm, with $\theta_m = 0.025$ and $\theta_h = 0.025$ in the skin tissue. We observe that the vessel is clearly distinguished as a region of greater hemoglobin concentration, and

Table 6.7: Predicted hemoglobin volume fractions $\tilde{\theta}_h$ and relative errors for inverse chromophore map f_{GP} applied to Monte Carlo simulated skin images with embedded vessels at different melanin and hemoglobin volume fractions θ_m and θ_h, respectively.

θ_m	θ_h	$\tilde{\theta}_h$	$\lvert\theta_h - \tilde{\theta}_h\rvert/\theta_h$	θ_h	$\tilde{\theta}_h$	$\lvert\theta_h - \tilde{\theta}_h\rvert/\theta_h$
0.02	0.025	0.025	0	0.030	0.030	0
0.05	0.025	0.024	0.04	0.030	0.030	0
0.10	0.025	0.025	0	0.030	0.031	0.03
0.15	0.025	0.023	0.08	0.030	0.029	0.03
0.20	0.025	0.026	0.04	0.030	0.027	0.10
0.25	0.025	0.027	0.08	0.030	0.027	0.1
0.30	0.025	0.022	0.12	0.030	0.026	0.13
0.35	0.025	0.023	0.08	0.030	0.026	0.13
0.40	0.025	0.022	0.12	0.030	0.022	0.27
0.43	0.025	0.021	0.16	0.030	0.021	0.30

scattering is observed about the blood vessel. Figure 6.7 displays the results of applying f_{GP} to simulated skin tissue with an embedded vessel of radius $r = 0.04$ cm, at a depth of $d = 0.25$ cm, with $\theta_m = 0.025$ and $\theta_h = 0.025$. Even at a a significantly greater depth, the inverse chromophore map accurately recovers the position of the blood vessel in the centre line of the image with a narrow vessel diameter. Figure 6.8 displays the result of applying f_{GP} to simulated skin tissue with an embedded vessel of radius of $r = 0.075$ cm, at a depth of $d = 0.13$ cm with $\theta_m = 0.22$ and $\theta_h = 0.025$. We observe the excellent recovery of the embedded vessel, even with a high concentration of scattering melanin, suggesting that the depth of the vessel and the increased presence of melanin provide non-trivial contributions to the ability of the inverse chromophore map to recover the vessel position. Finally in Figure 6.9 we present the most challenging case of applying f_{GP} to simulated skin tissue with an embedded vessel of radius $r = 0.08$ cm, at a depth of $d = 0.15$ cm, with $\theta_m = 0.41$ and $\theta_h = 0.03$. Due to the very high concentration of the highly light scattering melanin, the inverse chromophore map gives a greater concentration of hemolgobin in tissue around the vessel, however, the exact position of the vessel along the image's centre line is still remarkably well recovered. These results suggest that the inverse chromophore map will provide highly accurate vessel recovery results across all ranges of skin tones, making the method widely applicable.

6.4.2 Curved and Connected Embedded Vessels

The state-of-the-art embedded vessel Monte Carlo simulations of C. Zoller and Kienle, 2019 only permit the simulation of a straight embedded vessel. As blood vessels in real skin tissue follow tortuous curved routes, with vessels connected together and embedded at different depths, we also tested the ability of the inverse chromophore map to recover curved vessels.

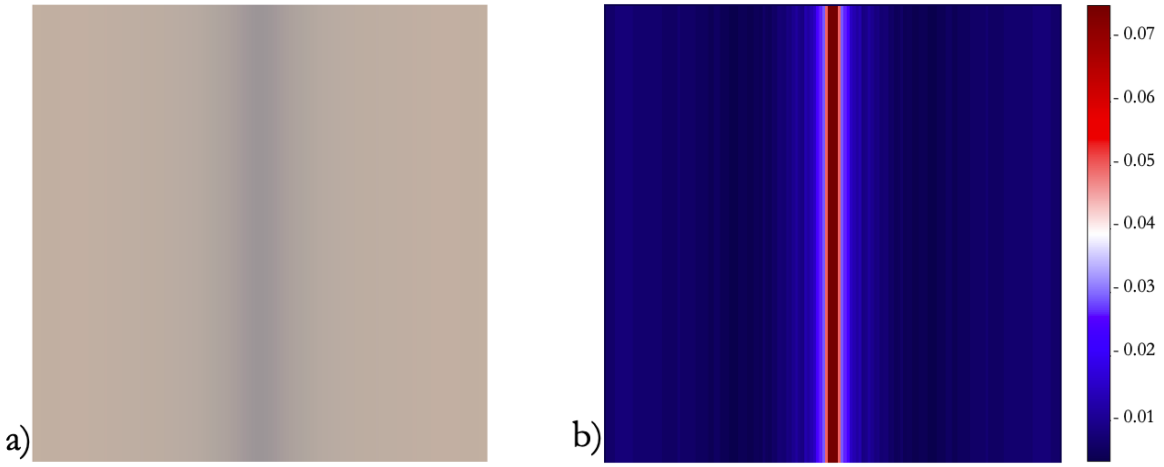

Figure 6.6: Simulated skin tissue with embedded vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin with embedded vessel of radius $r = 0.04$ cm, at depth $d = 0.05$ cm, $\theta_m = 0.025$, $\theta_h = 0.025$. b) Recovered hemoglobin distribution. Note the very narrow concentration of hemoglobin about the blood vessel, indicating excellent vessel recovery ability.

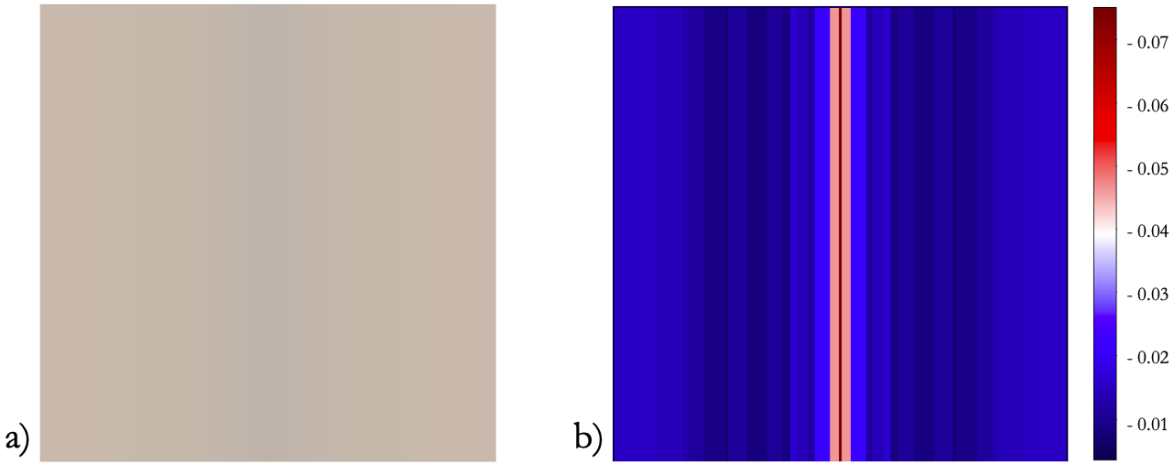

Figure 6.7: Simulated skin tissue with embedded vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin with embedded vessel of radius $r = 0.04$ cm, at depth $d = 0.15$ cm, $\theta_m = 0.25$, $\theta_h = 0.025$. b) Recovered hemoglobin distribution. Note the very narrow concentration of hemoglobin about the blood vessel, even at significant depth.

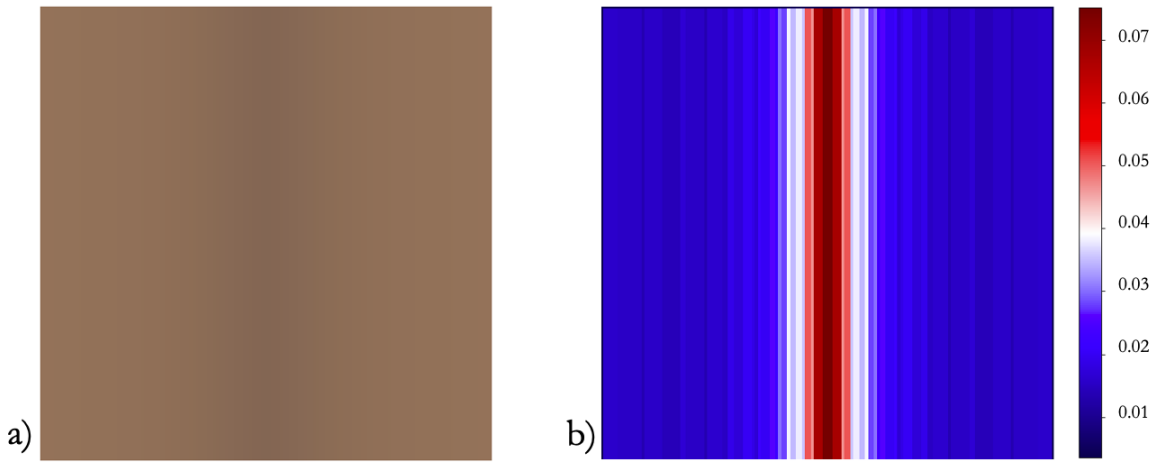

Figure 6.8: Simulated skin tissue with embedded vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin with embedded vessel of radius $r = 0.06$ cm, at depth $d = 0.13$ cm, $\theta_m = 0.22$, $\theta_h = 0.025$. b) Recovered hemoglobin distribution. Note the increased concentration of hemoglobin in the blood vessel area and surrounding tissue, as well as the ability of the inverse chromophore map to recover the greater concentration of hemoglobin in the vessel centre.

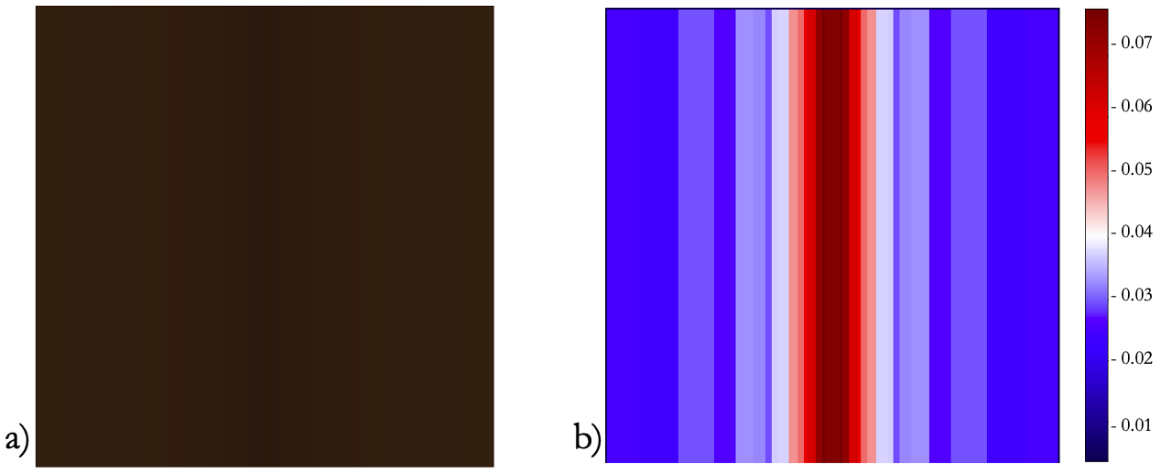

Figure 6.9: Simulated skin tissue with embedded vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin with two embedded vessels of radius $r = 0.08$ cm, at depth $d = 0.15$ cm, $\theta_m = 0.41$, $\theta_m = 0.025$. b) Recovered hemoglobin distribution. The inverse chromophore map is able to accurately recover the vessel position and higher concentration of hemoglobin in the vessel even with a signficant concentration of light scattering melanin.

We used the skin rendering model from Equation 6.2.1 to simulate different curved vessel structures, by creating a layered image consisting of a base overlayer of skin with set θ_m and θ_h, and one or two sublayers consisting only of the blood vessels with greater θ_h, at different depths.

In Figure 6.10 we display the results of the inverse chromophore map applied to a simulated skin image with a curved and connected embedded blood vessel. The skin tissue is simulated with $\theta_m = 0.04$ and $\theta_h = 0.025$, and the vessel is embedded at a depth of $d = 0.2$ cm with a greater hemoglobin concentration of $\theta_h = 0.0275$. The inverse chromophore map exactly recovers the position of the curved vessel, as well as correctly recovers the simulated hemoglobin volume fraction in the vessel and skin tissue.

In Figure 6.11 we display the results of the inverse chromophore map applied to a simulated skin image with a curved and connected embedded blood vessel. The skin tissue is simulated with $\theta_m = 0.30$ and $\theta_h = 0.025$, and the vessel is embedded at a depth of $d = 0.2$ cm with a greater hemoglobin concentration of $\theta_h = 0.0275$. The inverse chromophore map exactly recovers the position of the curved vessel, as well as correctly recovers the simulated hemoglobin volume fraction in the vessel and skin tissue.

In Figure 6.12 we display the results of the inverse chromophore map applied to a simulated skin image with two curved embedded blood vessels at different depths. The skin tissue is simulated with $\theta_m = 0.04$ and $\theta_h = 0.025$. One vessel is embedded at a depth of $d = 0.2$ cm and the second vessel is embedded at a depth of $d = 0.25$ cm each with a greater hemoglobin concentration of $\theta_h = 0.0275$. The inverse chromophore map exactly recovers the position of the two curved vessels, as well as correctly recovers the simulated hemoglobin volume fraction in the vessels and skin tissue.

In Figure 6.13 we display the results of the inverse chromophore map applied to a simulated skin image with two curved embedded blood vessels at different depths. The skin tissue is simulated with $\theta_m = 0.20$ and $\theta_h = 0.025$. One vessel is embedded at a depth of $d = 0.2$ cm and the second vessel is embedded at a depth of $d = 0.25$ cm each with a greater hemoglobin concentration of $\theta_h = 0.0275$. The inverse chromophore map exactly recovers the position of the two curved vessels, as well as correctly recovers the simulated hemoglobin volume fraction in the vessels and skin tissue.

In Figure 6.14 we display the results of the inverse chromophore map applied to a simulated skin image with two curved embedded blood vessels at different depths with different hemoglobin concentrations in each vessel. The skin tissue is simulated with $\theta_m = 0.03$ and $\theta_h = 0.0215$. One vessel is embedded at a depth of $d = 0.2$ cm with a hemoglobin volume fraction of $\theta_h = 0.0245$, and the second vessel is embedded at a depth of $d = 0.25$ cm with a greater hemoglobin concentration of $\theta_h = 0.0275$. The inverse chromophore map exactly recovers the position of the two curved vessels, as well as correctly recovers the simulated hemoglobin volume fraction in the vessels and the skin tissue.

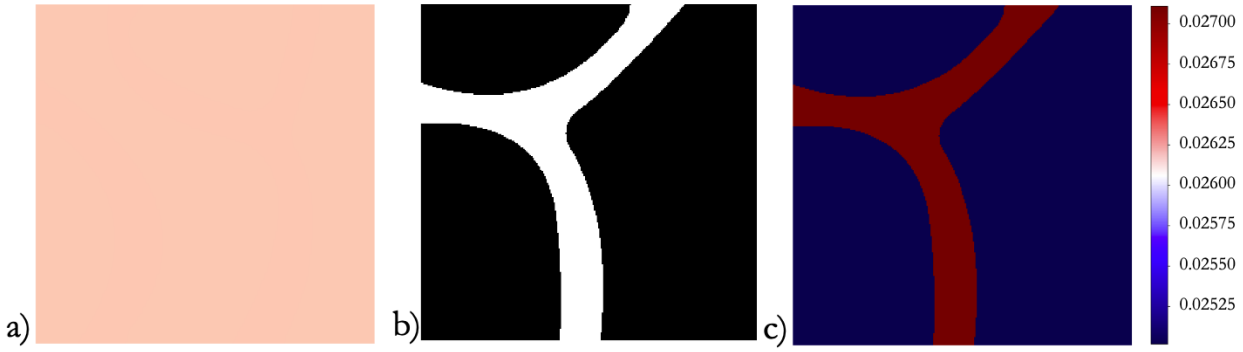

Figure 6.10: Simulated skin tissue with an embedded curved vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin generated from $\theta_m = 0.04$, $\theta_h = 0.025$ with an embedded curved vessel at depth $d = 0.2$ cm, with vessel $\theta_h = 0.0275$. b) The embedded vessel schematic. c) Recovered hemoglobin distribution and distinguished blood vessel.

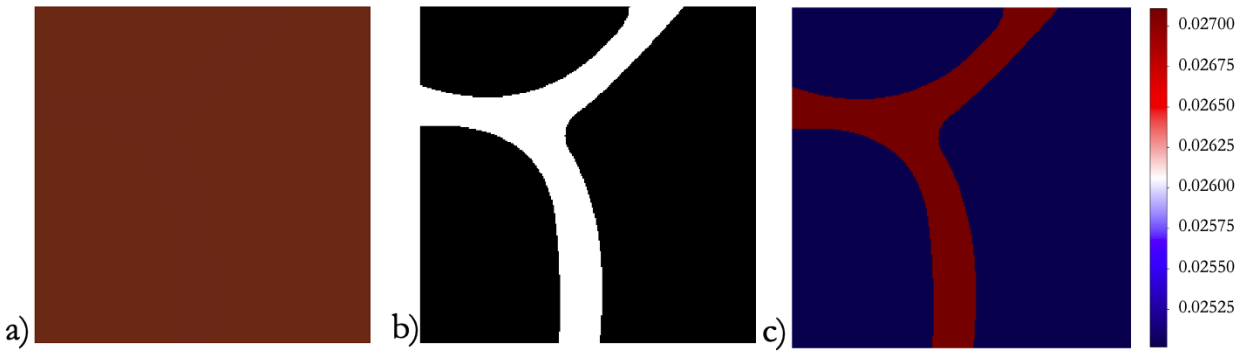

Figure 6.11: Simulated skin tissue with an embedded curved vessel and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin generated from $\theta_m = 0.30$, $\theta_h = 0.025$ with an embedded curved vessel at depth $d = 0.2$ cm, with vessel $\theta_h = 0.0275$. b) The embedded vessel schematic. c) Recovered hemoglobin distribution and distinguished blood vessel.

Figure 6.12: Simulated skin tissue with two embedded curved vessels at different depths with the same hemoglobin volume fraction and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin generated from $\theta_m = 0.04$, $\theta_h = 0.025$ with two embedded curved vessels at depth $d = 0.2$ cm, and $d = 0.25$ cm, respectively, with vessel $\theta_h = 0.0275$ in each. b) The embedded vessel schematic. c) Recovered hemoglobin distribution and distinguished blood vessels.

Figure 6.13: Simulated skin tissue with two embedded curved vessels at different depths with the same hemoglobin volume fraction and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin generated from $\theta_m = 0.20$, $\theta_h = 0.025$ with two embedded curved vessels at depth $d = 0.2$ cm, and $d = 0.25$ cm, respecitvely, with vessel $\theta_h = 0.025$ in each. b) The embedded vessel schematic. c) Recovered hemoglobin distribution and distinguished blood vessels.

Figure 6.14: Simulated skin tissue with two embedded curved vessels at different depths with different hemoglobin volume fractions and recovered hemoglobin distribution using inverse chromophore map f_{GP}. a) The simulated skin generated from $\theta_m = 0.03$, $\theta_h = 0.0215$ with two embedded curved vessels at depth $d = 0.2$ cm with vessel $\theta_h = 0.0245$, and $d = 0.25$ cm with vessel $\theta_h = 0.0275$, respecitvely. b) The embedded vessel schematic. c) Recovered hemoglobin distribution and distinguished blood vessels.

6.4.3 Temporal Hemoglobin Variation

In the following chapter, the inverse chromophore map will be applied to video frames of people to extract their vital sign information. As a first test of our proposed methodology, we performed a simplified simulation of blood flow through the embedded blood vessels in images by applying sinusoidal variation to the hemoglobin content in each of the embedded and curved blood vessels from Subsection 6.4.2. As discussed in Chapter 2, the pumping action of the heart creates a pulse wave causing oscillatory behaviour of the hemoglobin content in the blood vessels. To each embedded vessel we applied,

$$f(t) = A\cos(2\pi\omega t) + B \tag{6.4.4}$$

where A is the amplitude of hemoglobin variation in the blood vessels, ω is the frequency, and B is the baseline hemoglobin content in each vessel.

We perform several simulations to study the effects of hemoglobin variation on the results of the inverse chromophore map, as well as extract a pseudo-pulse rate via Fourier analysis. Each simulation uses the curved vessel schematic with two curved vessels at different depths with different baseline hemoglobin volume fractions, as described in the simulations in Subsection 6.4.2 and which can be seen in part b) of Figures 6.14. This configuration represents a somewhat realistic representation of physical blood flow along the blood vessels as would be seen in an imaging system. Each simulation is set at a sample rate of 60 Hz – matching the sample rate of a smartphone camera – and is two minutes in length. The pixel values in each RGB colour channel of the rendered images are spatially averaged to form one set of

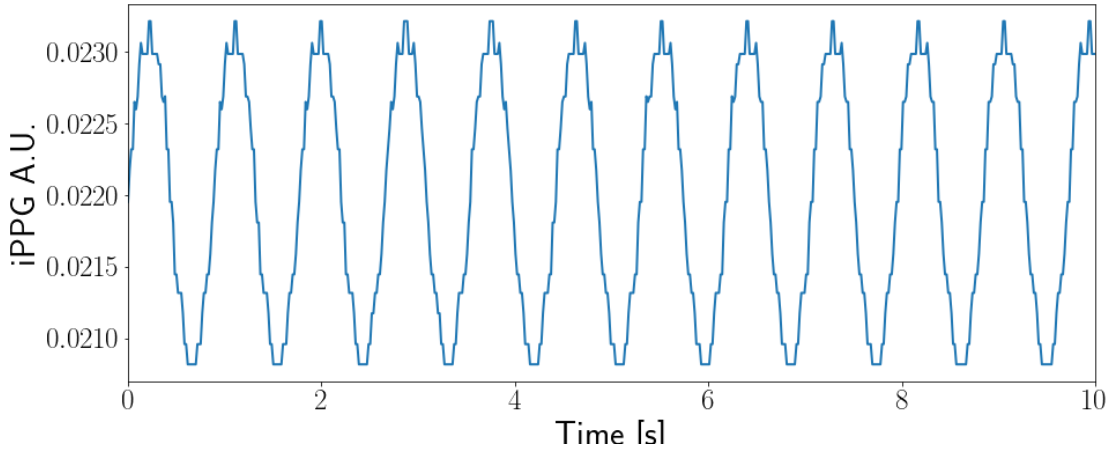

Figure 6.15: Ten second sample of inverse chromophore map applied to averaged pixel content from rendered skin images with two embedded curved vessels at depths $d = 0.2$ cm and $d = 0.25$ cm, respectively. The hemoglobin volume fraction in each vessel varies in phase with each other.

RGB pixel values per frame. As we describe in detail in Chapter 7, spatial averaging reduces camera quantisation noise.

The first simulation consists of each blood vessel's hemoglobin content varying in phase, with the amplitudes in each vessel determined as the difference between the baseline hemoglobin volume fraction in the vessel and the surrouding skin's hemoglobin volume fraction. The frequency is set at $\omega = 1.083$, which corresponds to 65 beats-per-minute (BPM) for the length of the simulation. A 10 second sample of recovered hemoglobin variation from the inverse chromophore map f_{GP} from frame-to-frame is displayed in Figure 6.15. The inverse chromophore map is able to accurately recover the sinusoidal varation in the hemoglobin volume fraction, which suggests that the inverse chromophore methodology will scale well to real skin colour variation.

The second simulation consists of the hemoglobin volume fraction in each vessel varying out of phase with each other. The variation is given as,

$$f_1(t) = A\cos(2\pi\omega_1 t) + B \tag{6.4.5}$$
$$f_2(t) = A_2\cos(2\pi\omega_1 t + \phi) + B_2 \tag{6.4.6}$$

where A_2 and B_2 are the amplitude and vertical shift for the hemoglobin variation in the second vessel signal, and ϕ is the phase shift between the signals in each vessel. Additionally, at one minute, we increase the frequency to represent a change in heart rate. This is represented

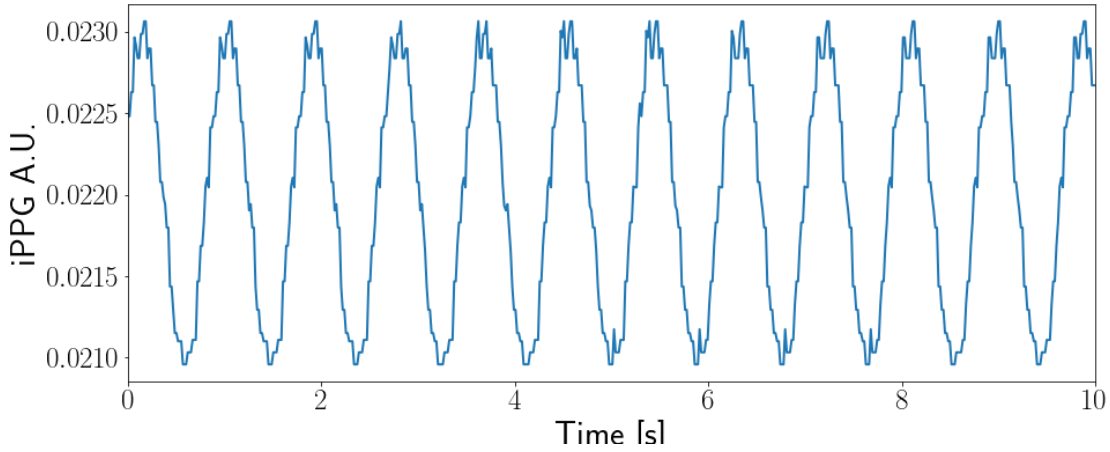

Figure 6.16: Ten second sample of inverse chromophore map applied to averaged pixel content from rendered skin images with two embedded curved vessels at depths $d = 0.2$ cm and $d = 0.25$ cm, respectively. The hemoglobin volume fraction in each vessel varies out of phase with each other, with phase difference displayed here of $\frac{\pi}{3}$.

as,

$$f(t) = \begin{cases} A\cos(2\pi\omega_1 t) + B, & t < 60 \\ A\cos(2\pi\omega_2 t) + B, & t \geq 60 \end{cases} \tag{6.4.7}$$

where $\omega_1 = 1.083$ and $\omega_2 = 1.25$, corresponding to pulse rates of 65 BPM and 75 BPM, respectively. In Figure 6.16 we display a 10 second sample of the extracted hemoglobin content from the inverse chromophore map, with phase difference $\phi = \frac{\pi}{3}$. We note that the phase difference causes a change in the peak behaviour of the signal, as well as adding a bump to the signal between the peaks and valleys of each period.

To evaluate if the signal recovered from the inverse chromophore map contains the same frequency content as the original sinusoidal hemoglobin content variation from the rendered skin frames, we take the short time Fourier transform (STFT) of the signal and inspect its frequency information via the spectrogram. In Figure 6.17 we display the spectrogram of the recovered hemoglobin signal from the inverse chromophore map. We see clearly that the dominant frequency is identified at 65 BPM, and that at one minute of simulation time, the dominant frequency changes to 75 BPM, indicating the definitive ability of the inverse chromophore map to respond to changes in the temporal dynamics of the hemoglobin content in the vessels.

As a final investigation, we added a second frequency component to the hemoglobin volume fraction variation of one of the vessels. Together, the hemoglobin variation in the vessels is

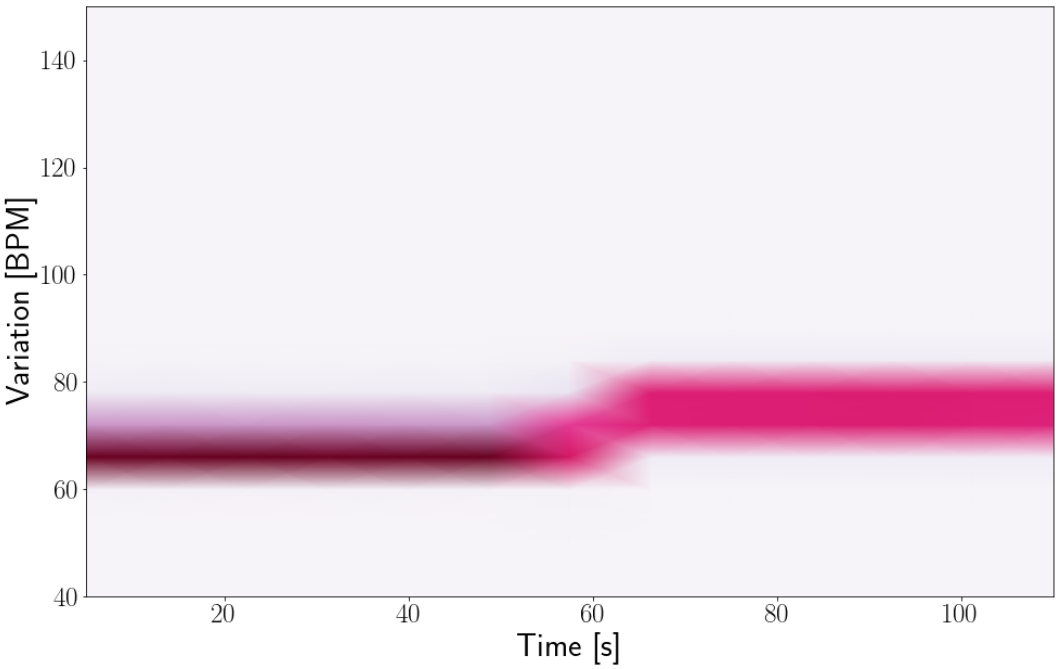

Figure 6.17: Spectrogram of inverse chromophore map applied to averaged pixel content from rendered skin images with two embedded curved vessels at depths $d = 0.2$ cm and $d = 0.25$ cm, respectively. The hemoglobin volume fraction in each vessel varies out of phase with each other, with phase difference of $\frac{\pi}{3}$. The inverse chromophore map recovers the same dominant frequency content as the original signal, with the largest component at 65 BPM for the first minute of the simulation, and 75 BPM after 60 seconds of simulation time.

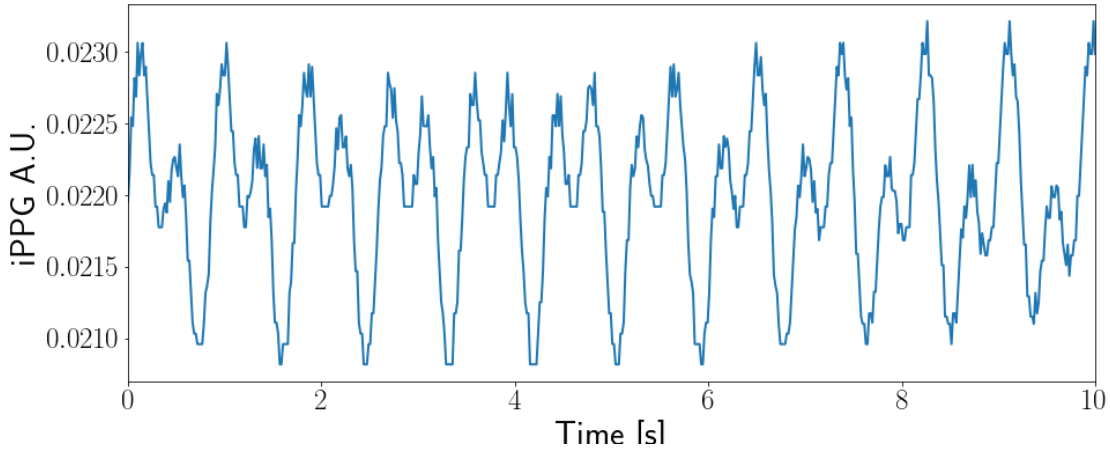

Figure 6.18: Ten second sample of inverse chromophore map applied to averaged pixel content from rendered skin images with two embedded curved vessels at depths $d = 0.2$ cm and $d = 0.25$ cm, respectively. The hemoglobin volume fraction in each vessel varies out of phase with each other, with phase difference of $\frac{\pi}{3}$. In one of the vessels, there is an added frequency component.

given as,

$$f_1(t) = A\cos(2\pi\omega_1 t) + B \tag{6.4.8}$$
$$f_2(t) = A_2\cos(2\pi\omega_1 t + 2pi\omega_2 t + \phi) + B_2 \tag{6.4.9}$$

where we set $\omega_1 = 1.083$ and $\omega_2 = 1.116$ corresponding to 65 BPM and 67 BPM respectively, and $\phi = \frac{\pi}{3}$. In Figure 6.18 we display a 10 second sample of recovered hemoglobin volume fraction from the inverse chromophore map. We observe the contribution of the different frequency components in one of the vessels on the recovered signal, as well as the phase difference appearing over time in the signal.

6.4.4 Real Skin Image Vasculature Recovery

The chromophore map was also tested on a collected data set of skin images from eight participants (five males, three females). The images were captured in an interior setting under ambient light conditions using an Apple iPhone 6S. The patch of skin selected was the base of wrist. We applied the inverse chromophore map f_{GP} learned on rendered skin images processed with the Apple iPhone 6S imaging processing pipeline as described in Chapter 5.

To produce a robust vasculature map, and to improve visual appearance, we apply several post-processing steps to uncover a skeletonized vasculature from the hemoglobin mapped image. We first apply a Contrast Limited Adaptive Histogram Equalisation (CLAHE)

algorithm (Zuiderveld, 1994) which performs local contrast enhancement. CLAHE uses histograms computed over different regions of the image to enhance local details. We apply CLAHE as a first step to enhance the distinguishment of regions with higher concentrations of hemoglobin in the skin image.

After the application of CLAHE, we then invert the colours of the image and apply the Sato tubeness filter (Sato et al., 1998) to detect discontinuous and continuous regions in the image. Originally designed to enhance curvilinear structures such as vessels in 3D medical images, the Sato filter computes the eigenvalues of the Hessian of the image to calculate the similarity of an image region to a defined tube. These post-processing steps improve the quality of the captured vasculature, and allow explicit extraction of the blood vessels from an image of skin tissue.

In Figures 6.19 – 6.22 we display the results of the inverse chromophore mapping and the recovered vasculature for four participants. a) of each figure displays the original colour image of the skin tissue; b) of each displays the hemoglobin concentration in each part of the skin tissue as a result of applying the inverse chromophore map f_{GP}; c) displays the CLAHE processed image to show the enhanced contrast of the hemoglobin map; d) displays the extracted vasculature map as a result of the Sato Filter. We draw particular attention to Figure 6.20 where the inverse chromophore map f_{GP} is able to accurately identify marks on the skin due to an irritation as areas of increased hemoglobin concentration. This suggests the possibility for the use of this methodology as a means of dermatological investigation and diagnosis of skin diseases and irrations related to increased blood flow. In figure 6.22 we apply the methodology to an image containing background values as well as skin tissue, for which the hemoglobin distribution and vasculture map are still reasonably recovered, but with edge effects on the boundary of the skin and the background. Such boundary effects can be eliminated by applying an image registration technique to remove non-skin pixels in the image (Dahmani, Cheref, and Larabi, 2020).

In the final vessel images, the Sato tubeness filter appears to be identifying features related to body hair. These identified features are likely due to the filter design to extract tube or cylinder like shapes, and may be confused with smaller vessels and the microvasculature. While this potentially poses a problem for the ability of the inverse map to accurately extract smaller regions of the vasculature, this issue may be alleviated through the use of a video, rather than a still image. As explained in Chapter 2, the movement of hemoglobin is most concentrated through the blood vessels. Thus, the application of the inverse chromophore map to sequential video frames of the imaged region will allow for accurate formation of the vessel map without additional features by extracting only those features which significantly vary in hemoglobin concentration.

Aside from the above mentioned potential use case as a dermatological tool for investigating skin conditions, the recovered vasculature map has several other possible applications including biometric identification, clinical assistance, and precision medicine. A person's vasculature is a feature truly unique to each individual; thus a vasculature map could serve as a biometric

tool for personal identification. A user would simply point their camera at a registered region of their skin, perhaps their hand or wrist, and image registration algorithms can be used to match the recovered vasculature map back to the user's saved information. Additionally, the vasculature map may serve as phlebotomy assistant, accurately identifying the position of blood vessels for needles and intravenous insertions. One could also envision the usage of a vasculature map as the geometric framework for realistic blood flow simulations, unique to each individual. These simulations may have potential impact for improving treatment courses and decision making in cardiology and endocrinology (Rowell, 1990; Charkoudian, 2003; Colberg et al., 2003).

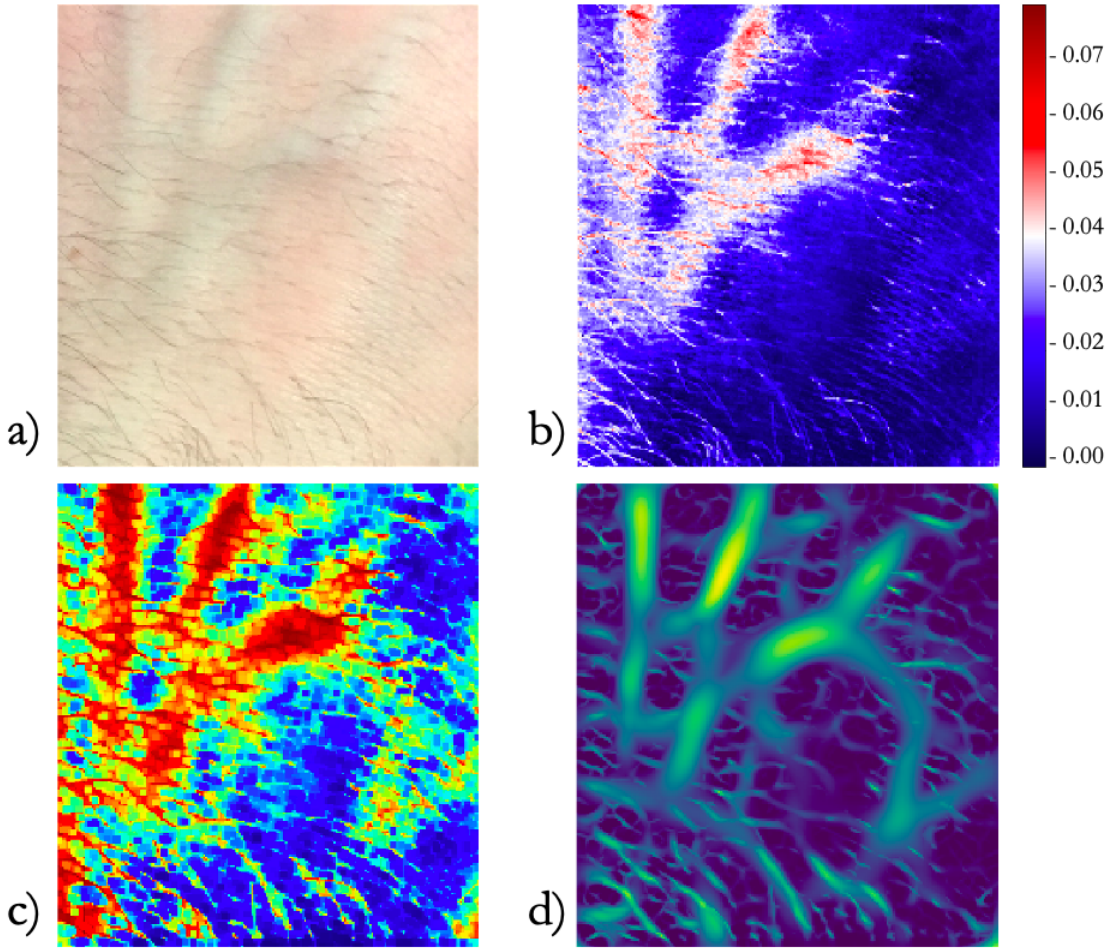

Figure 6.19: Hemoglobin map and vasculature of base of the wrist. a) Original colour image. b) Hemoglobin distribution. c) CLAHE rendered image. d) Sato filtered vasculature map

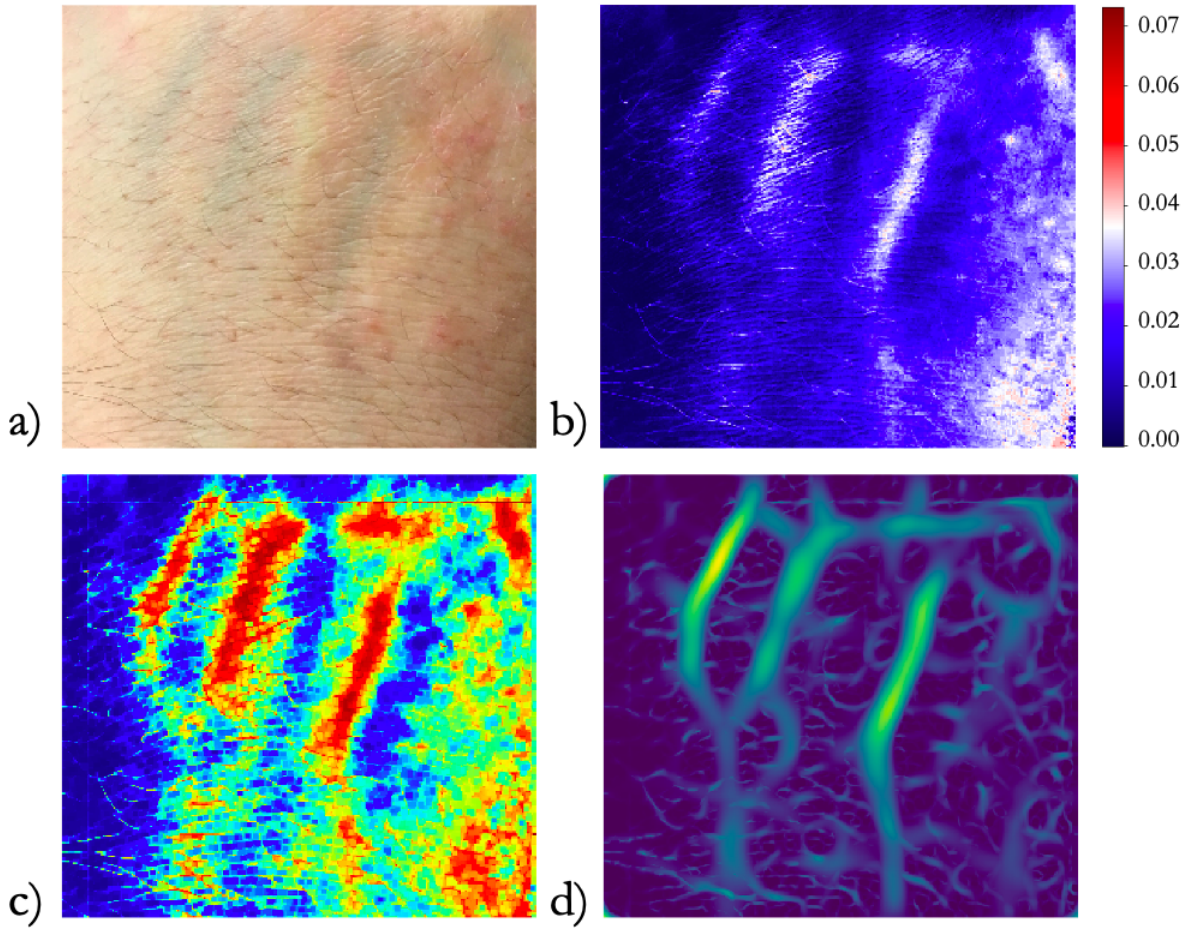

Figure 6.20: Hemoglobin map and vasculature of base of the wrist. a) Original colour image. b) Hemoglobin distribution. c) CLAHE rendered image. d) Sato filtered vasculature map

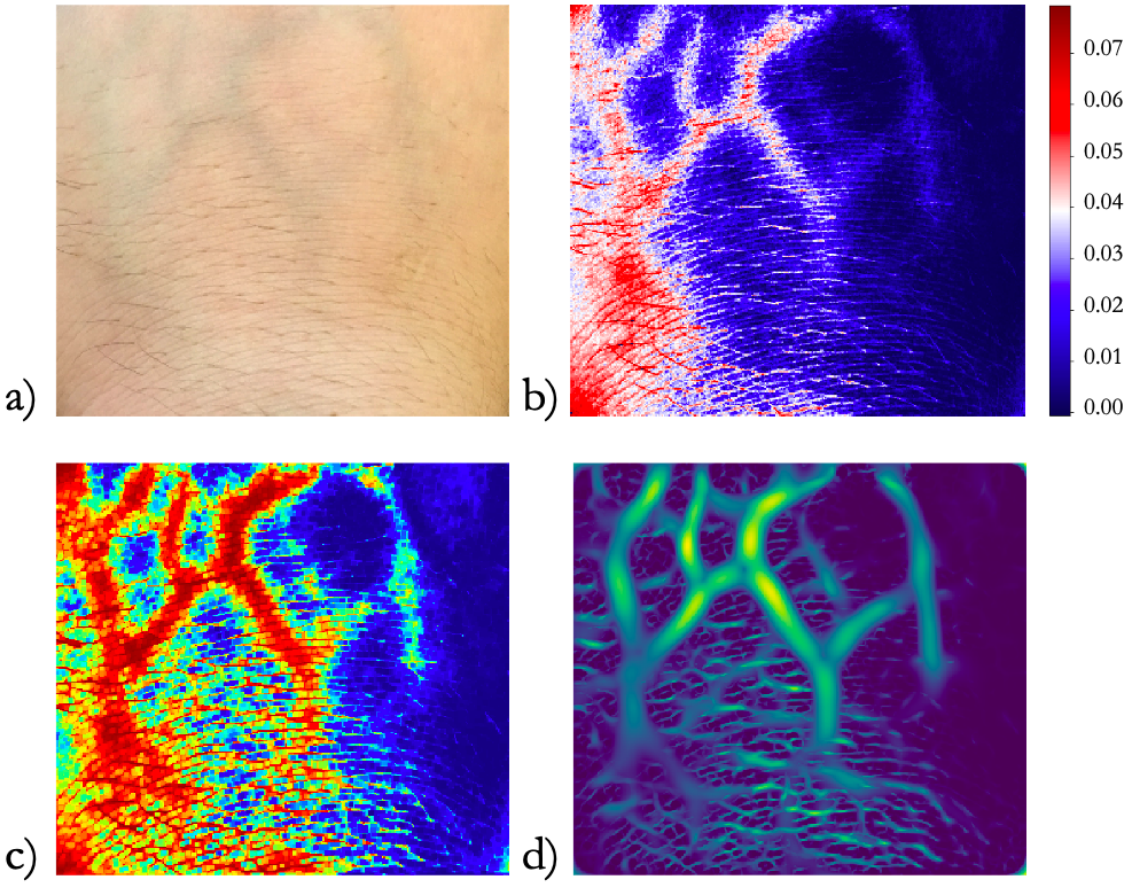

Figure 6.21: Hemoglobin map and vasculature of base of the wrist. a) Original colour image. b) Hemoglobin distribution. c) CLAHE rendered image. d) Sato filtered vasculature map

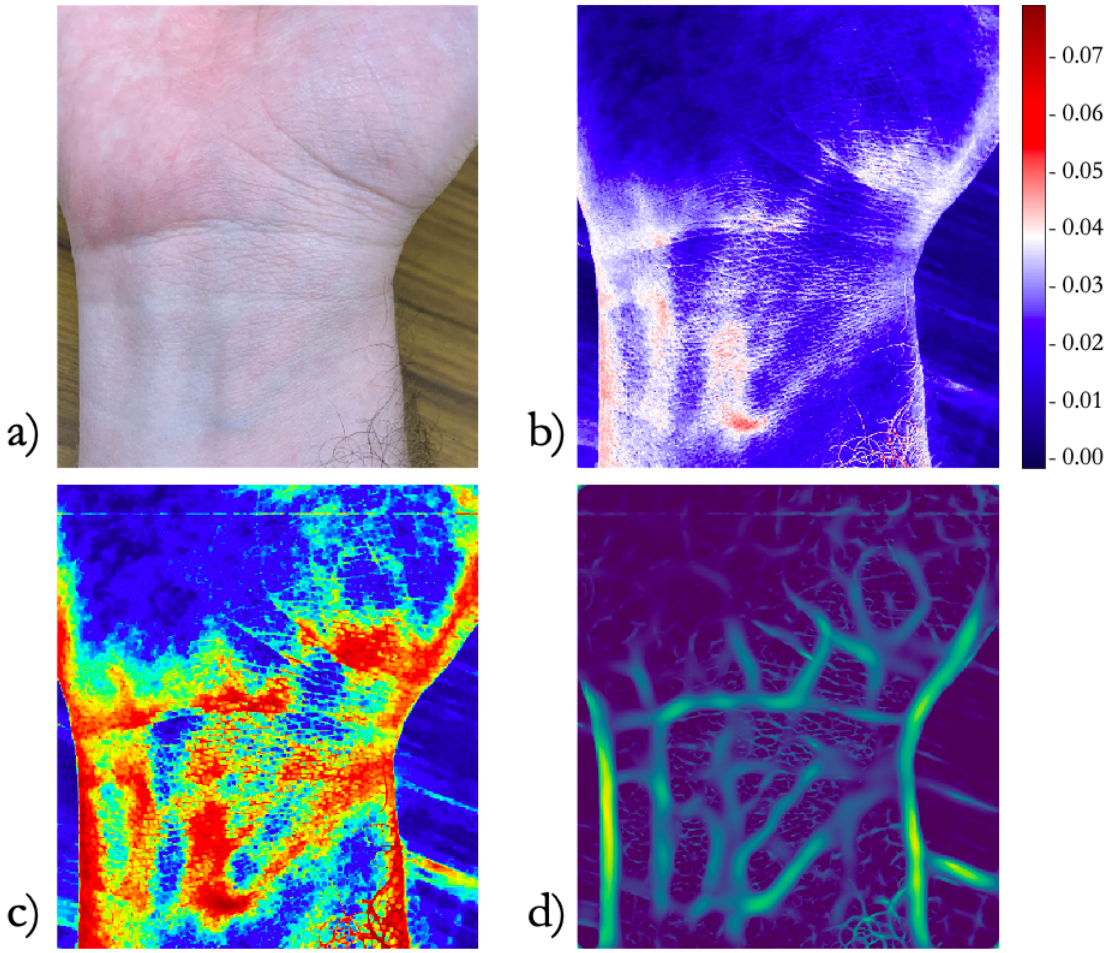

Figure 6.22: Hemoglobin map and vasculature of base of the wrist. a) Original colour image. b) Hemoglobin distribution. c) CLAHE rendered image. d) Sato filtered vasculature map

Chapter 7

Imaging Diagnostics

> "*With advancements in hardware... a multifunctionality and compact ubiquitous personal monitoring system would be feasible.*"
>
> — Yu Sun and Nitish Thakor
> *IEEE Review*

The past chapters have been concerned with the development of a multi-stage methodology to construct an inverse chromophore map for the recovery of chromophore concentrations from digital images of skin tissue. As we demonstrated in Subsection 6.4.3 through our simplified simulation of blood flow in the rendered skin images, the application of the inverse chromophore map to sequential skin images permits the extraction of underlying temporal dynamics. In this chapter we apply the inverse chromophore map to video images of skin tissue, demonstrating its capacity to recover the underlying cardiovascular dynamics which modulate chromophore concentrations and ultimately skin pixel values. Further, we demonstrate the novel ability of the inverse chromophore map to extract the same temporal information contained in the electrocardiogram signal of a user from only the facial blood flow signal.

With the goal of recovering these dynamics we begin with a review of *imaging Photoplethysmography* (iPPG) (sometimes refered to as remote PPG in the literature) and present a mathematical framework for which all existing iPPG techniques can be understood in Section 7.1. We describe how the existing methods in the literature may be explained with this framework, and discuss how our proposed methodology overcomes the technical and mathematical challenges faced by the existing methods in the literature for robust iPPG monitoring. This is followed in Section 7.2 by applying the inverse chromophore map developed in Chapter 6 to facial videos of human subjects to recover facial blood flow signals. This novel methodology represents the formulation of a rigorous opto-physiological model of iPPG

for which the temporal blood flow signals may be accurately recovered, filling an existing gap in the literature as noted by Sun and Thakor, 2016. We apply signal processing techniques to the facial blood signals to recover the subjects' heart rate, and demonstrate excellent agreement with traditional clinical gold-standard monitoring technology, as well as improved performance over existing methodologies in the literature which we implement in a Python software package. In Section 7.3 we present an original application of this technique for the contactless recovery of a representation of the electrocardiogram (ECG) waveform of a subject from their facial blood flow signal in a set of physiological experimental conditions.

7.1 Imaging Photoplethysmography

Imaging or remote Photoplethysmography (iPPG) permits the contactless monitoring of cardiovascular activity of a subject through the detection of subtle colour changes of human skin tissue modulated by cardiac activity using a digital video camera (Sun and Thakor, 2016). The first suggestion of an iPPG system using videos was made in a 1996 German patent application (Blazek, Rutten, and Such, 1996), although the field sat largely dormant for more than a decade after this initial proposition. In 2008, Verkruysse, Svaasand, and Nelson, 2008 published their landmark paper demonstrating the implementation of a iPPG system using a digital RGB camera and ambient light to measure heart rate at a distance. This paper spurred a renewed interest in the field, and the last decade has seen an explosion of research effort in the field of iPPG technology (Zaunseder et al., 2018).

iPPG techniques pose several practical advantages for personal use. While multiple lead ECGs provide gold-standard clinical measurements to diagnose and monitor cardiopathophysiologies, there are certain downside to ECG systems which make alternative imaging techniques practical. The attachment of ECG leads to the skin may cause discomfort for many and raise the risk of skin infections (Lestari, Ryll, and Kramer, 2013). In Neonatal Intensive Care Units (NICU) cardiac monitoring is essential, however infant skin is very prone to irritation from lead attachment, which makes iPPG solutions very valuable in such a setting (Klaessens et al., 2014; Davis et al., 2014). Further, the potential to remotely monitor patients outside of a clinical setting is extremely attractive to medical practitioners.

While the explosion of research effort has been focussed on algorithmic development and the consistency of the application of new signal and image processing techniques to the iPPG field, a comprehensive mathematical framework for opto-physiological modelling has been lagging behind (Sun and Thakor, 2016). In this section we describe a framework to model iPPG and assess how the exisiting literature fits into the framework. Afterwards, we show how our inverse chromophore mapping methodology permits the explicit extraction of cardiac activity.

Consider a constant spectral intensity light source illuminanting a section of skin tissue and a remote RGB camera recording the skin. Due to the pulsatile nature of blood flow, as described in Section 2.2, subtle colour changes will occur due to cardiac activity, due to

the movement of chromophores – here, hemoglobin – during the cardiac cycle. While the reflectance of each skin pixel can be modelled using the radiative transport theory of Chapter 4, the simpler Shafer dichromatic model is sufficient to describe the time varying reflection in the RGB channels (W. Wang et al., 2017). That is, the RGB values of the k-th pixel in an image sequence is described by

$$\boldsymbol{C}_k(t) = E(t)(\boldsymbol{v}_s(t) + \boldsymbol{v}_d(t)) + \boldsymbol{v}_n(t) \tag{7.1.1}$$

where $\boldsymbol{C}_k(t)$ is the RGB channels of the k-th skin pixel (in columns), $E(t)$ is the time-varying illuminant intensity level which is affected by specular reflection $\boldsymbol{v}_s(t)$ and diffuse reflection $\boldsymbol{v}_d(t)$; $\boldsymbol{v}_n(t)$ represents quantisation noise of the camera sensor.

Specular reflectance $\boldsymbol{v}_s(t)$ describes the mirror like reflection from an object, and thus does not contain any pulsatile information from cardiac variation; its spectral composition is equivalent to the illuminant. The time dependence of $\boldsymbol{v}_s(t)$ arises from bodily-motion, changing the distance between the skin, camera, and illuminant.

$$\boldsymbol{v}_s(t) = \boldsymbol{w}_s(s_0 + s(t)) \tag{7.1.2}$$

where $\boldsymbol{w}_s$ is the unit colour vector of the light spectum, s_0 is the stationary component of the specular reflection, and $s(t)$ is the time-varying component, induced by motion.

The diffuse reflectance, as described in Section 4.4, is controlled by chromophores. For the time-varying pixel components, we have

$$\boldsymbol{v}_d(t) = \boldsymbol{w}_d d_0 + \boldsymbol{w}_p p(t) \tag{7.1.3}$$

where $\boldsymbol{w}_d$ is the unit colour vector of the skin, d_0 is the stationary reflectance strength, $\boldsymbol{w}_p$ describes the relative pulsatile strength in the RGB channels, and $p(t)$ represents the pulse-signal; that is, $p(t)$ represents the cardiac acitivity as observed through the chromophore variation.

Combing the stationary components of the specular and diffuse reflectance

$$\boldsymbol{w}_c c_0 = \boldsymbol{w}_s s_0 + \boldsymbol{w}_d d_0 \tag{7.1.4}$$

and writing the illuminant as $E(t) = E_0(1 + e(t))$ for stationary component and E_0 and motion induced time varying component $E_0 e(t)$ yields the pixel values (7.1.1) as

$$\boldsymbol{C}_k(t) = E_0(1 + e(t))\left(\boldsymbol{w}_c c_0 + \boldsymbol{w}_s s(t) + \boldsymbol{w}_p p(t)\right) + \boldsymbol{v}_n(t) \tag{7.1.5}$$

Given (7.1.5), the purpose of an iPPG algorithm is to extract $p(t)$ from $\boldsymbol{C}_k(t)$.

Rather than considering the pixel values $\boldsymbol{C}_k(t)$ directly, the vast majority of existing iPPG algorithms use spatially averaged RGB pixel values $\boldsymbol{C}(t)$ (Verkruysse, Svaasand, and Nelson, 2008; Poh, D. J. McDuff, and Picard, 2010; Lewandowska et al., 2011; Gerard de Haan and Jeanne, 2013; G. de Haan and Van Leest, 2014; Blackford, Estepp, et al., 2016; E. J. Wang

et al., 2017; Blackford and Estepp, 2017). Spatial averaging serves two purposes, (*i*). it reduces camera quantisation error by taking a coalition of pixel values in a region, and (*ii*). it reduces computational complexity for image processing algorithms.

Expanding (7.1.5)

$$\begin{aligned}\boldsymbol{C}(t) &= E_0\, c_0 + \boldsymbol{w}_s\, E_0\, s(t) + \boldsymbol{w}_p\, E_0\, p(t) \\ &\quad + \boldsymbol{w}_c\, E_0\, c_0\, e(t) + \boldsymbol{w}_s\, E_0\, s(t)\, e(t) + \boldsymbol{w}_p\, E_0\, p(t)\, e(t) \qquad (7.1.6)\\ &\approx \boldsymbol{w}_c\, E_0\, c_0 + \boldsymbol{w}_c\, E_0\, c_0\, e(t) + \boldsymbol{w}_s\, E_0\, s(t) + \boldsymbol{w}_p\, E_0\, p(t) \qquad (7.1.7)\end{aligned}$$

where (7.1.7) results from AC modulations in the pulsatile signal being orders of magnitude smaller than the quasi-stationary DC term, as described in Section 2.3. This results in the contribution of products of time-varying terms being negligible. (7.1.7) suggests that the spatially averaged observed signal $\boldsymbol{C}(t)$ is a linear mixture of the time varying source signals, $s(t)$, $p(t)$, $e(t)$. Thus, iPPG algorithms of the literature extract the pulse signal $p(t)$ from $\boldsymbol{C}(t)$ by defining a projection system so as to decompose $\boldsymbol{C}(t)$.

The most straightforward procedure for decomposition is the method of Verkruysse, Svaasand, and Nelson, 2008, called *GREEN*, who simply take the spatial average of the Green pixels over a large region of the face and then apply filtering algorithms to the resulting signal to smooth its appearance. This model assumes the contributions to the signal from movement $s(t)$ and the illuminant $e(t)$ are trivial. Despite its relatively simple formulation, the *GREEN* method performs remarkably well in many situations, but suffers from performance decay under challenging conditions such as motion and varying lighting conditions.

Given the linear mixing of (7.1.7), other authors have used Blind Source Separation (BSS) techniques for demixing $\boldsymbol{C}(t)$ into its constituent sources. The BSS techniques for iPPG write the signal as

$$\boldsymbol{Y}(t) = W\boldsymbol{C}(t) \qquad (7.1.8)$$

where $\boldsymbol{Y}(t)$ are the factorised source signals made up of the pulse signal and noise, and W is the demixing matrix. Poh, D. J. McDuff, and Picard, 2010 apply independent component analysis (ICA) for estimating the demixing matrix W, while Lewandowska et al., 2011 use principal component analysis (PCA). ICA and PCA methods differ regarding their signal relationship assumptions. ICA assumes the components of $\boldsymbol{Y}(t)$ are independent and non-Gaussian, while PCA estimates the eigenvalues of W from the covariance of the RGB signals, requiring the variations of the pulse signal and noise to be sufficiently uncorrelated.

While BSS techniques offer an improvement in performance, both ICA and PCA are general statistical methods, and do not take into account the specifics of skin reflectance. Motivated by this Gerard de Haan and Jeanne, 2013; G. de Haan and Van Leest, 2014 extended the iPGG methodology to use specific colour vector information to constrain demixing. These methods eliminate the dependence of $\boldsymbol{C}(t)$ on DC-level average skin reflection, which includes the illuminant colour and inherent skin colour through temporal normalisation. The RGB

signals are temporally normalised by division with the temporal mean, either calculated over the length of the signal, or over a moving window of fixed length. The temporal mean is

$$\overline{\boldsymbol{C}}(t) = \boldsymbol{w}_c \, E_0 \, c_0 \tag{7.1.9}$$

which defines a unique diagonal normalisation matrix N,

$$N \, \overline{\boldsymbol{C}}(t) = N \, \boldsymbol{w}_c \, E_0 \, c_0 = \mathbf{1}. \tag{7.1.10}$$

N temporally normalises $\boldsymbol{C}(t)$ as

$$\begin{aligned} \boldsymbol{C}_n(t) &= N \, ()\boldsymbol{w}_c \, E_0 \, c_0 + \boldsymbol{w}_c \, E_0 \, c_0 \, e(t) + \boldsymbol{w}_s \, E_0 \, s(t) + \boldsymbol{w}_p \, E_0 \, p(t)) \\ &= \underbrace{\mathbf{1} \, (1 + e(t))}_{\text{intensity}} + \underbrace{N \, \boldsymbol{w}_s \, E_0 \, s(t)}_{\text{specular}} + \underbrace{N \, \boldsymbol{w}_p \, E_0 \, p(t)}_{\text{pulse}}. \end{aligned} \tag{7.1.11}$$

The intensity term describes illuminant intensity variation along $\mathbf{1}$, the direction of temporal normalisation, and is present simultaneously and equally in each RGB channel. The specular term gives the temporal variation along the scaled specular reflectance direction, and depends on the illuminant spectrum, as well as the camera spectral sensitivity functions. Finally, the pulse component represents the pulse-induced temporal variations, which iPPG seeks to extract. It has been observed that the Green channel has the largest amplitude, followed by the Blue channel, and subsequently the Red channel. As discussed in Section 2.1, this is due to the optical absorption spectrum of hemoglobin, being largest in the Greeen and Blue light optical wavelengths and smaller in the Red region.

To proceed, the quasi-stationary DC component is removed, giving

$$\tilde{\boldsymbol{C}}_n(t) = \boldsymbol{C}_n(t) - \mathbf{1} = \mathbf{1} \, e(t) + N \, \boldsymbol{w}_s \, E_0 \, s(t) + N \, \boldsymbol{w}_p \, E_0 \, p(t) \tag{7.1.12}$$

where $\tilde{\boldsymbol{C}}_n(t)$ is the zero-mean RGB colour signal.

The *CHROM* method of Gerard de Haan and Jeanne, 2013 and subsequent methods (G. de Haan and Van Leest, 2014; W. Wang et al., 2017) restrict the colour variation to only the pulsatile directions by projecting $\tilde{\boldsymbol{C}}_n(t)$ onto a direction $\boldsymbol{z} \in \mathbb{R}^3$ such that an estimate $\hat{p}(t)$ is recovered proportional to $p(t)$,

$$\hat{p}(t) = \tilde{\boldsymbol{C}}_n(t)^\top \, \boldsymbol{z} = kp(t) \tag{7.1.13}$$

for proportionality constant $k \neq 0$. *CHROM* uses chrominance-signals by projecting $\tilde{\boldsymbol{C}}_n(t)$ onto the plane orthogonal to specular variation. A standardised skin-tone vector is assumed, $\boldsymbol{w}_{skin} = (0.77, 0.51, 0.38)^\top$ (determined from experiments) which is constructed to appear the same under different illuminants. The mapping matrix is then defined as

$$M^{-1} \boldsymbol{w}_{skin} = \mathbf{1} \tag{7.1.14}$$

M is then applied to the zero mean colour signals

$$M\tilde{\boldsymbol{C}}_n(t) = M\left(\mathbf{1}\, e(t) + N\, \boldsymbol{w}_s\, E_0\, s(t) + N\, \boldsymbol{w}_p\, E_0\, p(t)\right) \tag{7.1.15}$$

where the specular reflection vector $N\, \boldsymbol{w}s\, E_0$ is mapped into the white light direction,

$$M\, N\, \boldsymbol{w}_s\, E_0 = k'\mathbf{1} \tag{7.1.16}$$

for proportionality constant k'. $M\, \tilde{\boldsymbol{C}}_n(t)$ is then projected onto an orthogonal plane to $\mathbf{1}$, so as to be independent from the specular reflection

$$\begin{aligned} \boldsymbol{S}(t) &= P_c\, M\, \tilde{\boldsymbol{C}}_n(t) \\ &\approx P_c\, M\, \mathbf{1}\, e(t) + P_c\, M\, N\, \boldsymbol{w}_p\, E_0\, p(t) \end{aligned} \tag{7.1.17}$$

such that

$$P_c\, M\, N\, \boldsymbol{w}_s\, E_0 \approx k'\, P_c\, \mathbf{1} = \mathbf{0} \tag{7.1.18}$$

for $P_c \in \mathbb{R}^{2\times 3}$ initial projection matrix defining a plane in the temporally normalised RGB space,

$$P_c = \begin{pmatrix} 1 & -1 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix} \tag{7.1.19}$$

The projection matrix $P_c\, M$ creates a pair of in-phase/anti-phase signals in $\boldsymbol{S}(t)$ representing the motion induced and pulse induced variations of the colour signal and is given as

$$P_c\, M = \begin{pmatrix} 3 & -2 & 0 \\ \frac{3}{2} & 1 & -\frac{3}{2} \end{pmatrix} \tag{7.1.20}$$

The pair of signals are then combined by *alpha tuning* to form $\hat{p}(t)$,

$$\hat{p}(t) = S_1(t) - \alpha\, S_2(t),\ \alpha = \frac{\sigma(S_1)}{\sigma(S_2)} \tag{7.1.21}$$

for standard deviation $\sigma(\cdot)$.

Given the general mathematical framework for iPPG measurement, we can position our proposed methodology with respect to the current literature. The proposed multi-stage methodology to recover the inverse chromophore map serves as an ideal candidate for an iPPG technique, by allowing explicit extraction of the $p(t)$ signal from a video of skin tissue. As the cardiac signal $p(t)$ is determined by cardiac activity pumping blood, and therefore hemoglobin about the body, the application of the inverse chromophore map f_{GP} to each frame of a video image of skin tissue returns the temporal variation of hemoglobin in the region of imaged skin. The temporal variation of hemoglobin is a direct representation of the variation of skin tone caused by cardiac activity, and is thus an accurate representation of

$p(t)$. The inverse chromophore map thus bypasses directly the technical and mathematical challenges faced by existing iPPG techniques in the literature by extracting the hemoglobin variation signal directly. In the following sections we will discuss the use of f_{GP} directly on skin images collected from video frames to form a time series of temporal hemoglobin variation for analysis.

It is worth mentioning that there has been an emergence in the literature of deep learning techniques for iPPG monitoring (Chen and D. McDuff, 2018; Luo et al., 2019; Bousefsaf, Pruski, and Maaoui, 2019; Yu et al., 2019). Chen and D. McDuff, 2018 propose an attention-based deep learning system to analyse frame-to-frame variation of facial videos to capture the relevant $p(t)$ signal. Their methodology is directly rooted in the theory of light-interaction in skin tissue, and achieves impressive results in numerous use cases. Finally, Luo et al., 2019 propose the use a machine learning based system to estimate one's blood pressure from a facial video. It is our belief that such systems may benefit tremendously from including a rigorous model of light-interaction in skin as we propose here.

7.2 Heart Rate Recovery

The first application of the inverse chromophore map f_{GP} for vital sign recovery will be to extract the heart rate from a video of a person's face. We begin by giving an overview of the experimental data collected for our analysis, then describe the processing procedures applied to the video frames to extract the hemoglobin variation signal, and then compare our results with clinical ground-truth data collected from the experiments. We also compare the performance of our method against methods in the literature which we implemented in a Python software package.

7.2.1 Data

To evaluate the use of our inverse chromophore methodology to recover the facial blood flow signal and ultimately the temporal cardiac signal $p(t)$, we collected a data set of eight subjects (fives males, three females) to particpate in a controlled experiment involving a video recording and simulatenous vital sign monitoring with gold-standard clinical measurement devices. Informed consent was obtained for all participants. Participants sat in a dark room with a single LED light source illuminating the participant. An Apple iPhone 6S (Apple, Cupertino, CA) was placed approximately 1m from each participant on a stand to avoid unnecessary movement. Participant's vital signs were simultaneously monitored using a Biopac MP160 monitoring system (BIOPAC Systems, Goleta, CA). A 3-lead ECG was used to record ECG, a reflectance based finger PPG system recorded the finger photoplethysmogram signal, and non-invasive blood pressure measurements were taken with a calibrated blood pressure cuff using a FinaPres monitor.

Participants were presented with a grip strength testing device which they were instructed to squeeze as hard as possible for a set period of time upon a signal. The experiment proceeded

according to the following time course: participants first sat for two minutes to establish a baseline signal, after which a spoken communication was given to squeeze the grip strength testing device for 90 seconds. Once 90 seconds of gripping elapsed, the participants were instructed to release their grip on the device and relax for the remaining six minutes and 30 seconds. In total, the experiment took approximately 10 minutes per participant.

Participants were recorded at 1920×1080 resolution at 60 frames-per-second, which was the maximal framerate the camera could record at. The simultaneously recorded vital signs were all measured at 200 Hz. Upon the completion of the experiment, the videos and vital signs files were exported for analysis. Each video is about 1.7 Gb in size, which is fairly large, given the high resolution and frame rate.

7.2.2 Video Processing

The collected videos were processed using custom written Python code. The videos are read from disk to memory using the OpenCV software library (Bradski, 2000). The very high resolution of the videos presents a computational bottle neck for real-time computing, as it is expensive to work with video frames of such large size. As we are focussed on only analysing skin in the face of the participant, we first implement a procedure to identify only the skin regions in the video frame. In Figure 7.1 we display our processing steps to produce a facial skin mask. We first locate the face in the video frame using a Viola-Jones face detector (Viola and Jones, 2004), crop the video frame about the face, and then construct an image pyramid of the reduced image (Burt, 1981). While standard image pyramid techniques apply a Gaussian filter to the image to smooth its appearance as the size is reduced, this would compromise the integrity of the skin pixel values. As such, we do not apply a Gaussian pyramid, but instead scale down the size of the frame by sampling each second pixel in the pyramid structure, this dramatically reduces the size of the image to less than 10,000 pixels. Figure 7.1 a) displays the cropped and scaled facial image.

Given that the participant is unconstrained, there is a possibility for movement from frame-to-frame. This movement would create difficulty for our analysis, since the movement of the skin pixels may create spurious changes in the pixel content, which would cause the inverse chromophore map to perform poorly. To alleviate this problem, we implement a facial landmark tracker placing 86 landmarks on the face to accurately track the face from frame-to-frame (Kazemi and Sullivan, 2014). Figure 7.1 b) displays the landmarks.

The final step of pre-processing is to isolate the facial skin region by applying a mask to the face. We form the mask by constructing a mesh over the face, using the facial landmarks as the vertices of the polygon shaped mask components. By constructing this mask, we effectively isolate the skin from the surrounding image, and also exclude the eyes and lips. The facial skin mask is displayed in Figure 7.1 c).

Once the facial mask is formed, we then take the spatial average of all the pixels in the mask in each colour channel to form $\boldsymbol{C}(t)$, as above, and write to memory. The entire

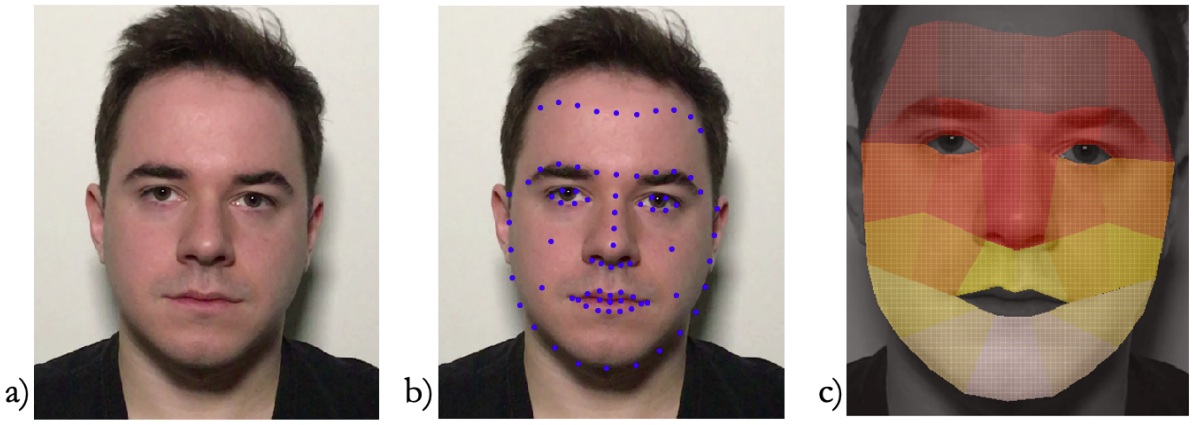

Figure 7.1: Facial cropping and masking. a) The original video frame cropped to the face as a result of Viola-Jones face detection. b) Landmarks placed on the face to identify the facial regions. c) Masking procedure applied to each facial region to form the facial mask.

video processing procedure is relatively fast, and operates typically around 63 frames per second, allowing for real-time video processing at the speed at which the camera records the video.

7.2.3 Heart Rate Calculation

Given the spatially averaged pixel vector $\boldsymbol{C}(t)$, we apply the inverse chromophore map f_{GP} to recover the hemoglobin concentration over the facial region, $h(t)$. Even though quantisation noise has been alleviated due to spatial averaging, and the effects of facial movement are largely managed by the facial tracking algorithm, the collected signal is still noisy, and thus must be processed using signal processing techniques. For our analysis, we first standardise the signal over a moving window of length 6 seconds with a 5 second overlap,

$$h_n(t) = \frac{h(t) - \overline{h}_w(t)}{\sigma(h_w(t))} \tag{7.2.1}$$

where $\overline{h}_w(t)$ is the mean of $h(t)$ in the window, and $\sigma(h_w(t))$ is the standard deviation. We then apply a 6-th order Butterworth bandpass filter (Butterworth, 1930) with cutoff frequencies 0.4 Hz and 3 Hz (24 BPM and 180 BPM), respectively. In Figure 7.2 we display a 10 second sample of the iPPG signal acquired from the facial hemoglobin map and the finger PPG signal acquired using a reflectance hotoplethysmogram attached to the finger. We point out excellent agreement between the waveform shapes, and also point out the pronounced presence of the dichrotic notch in the iPPG signal which is representative of the reflected wave of the blood pulse (Shi et al., 2009). We also note the minute discrepancy in timing of the signals that appears towards the end of the time window. This difference is due to the

difference in time it takes for the blood pulse from the heart to reach the face (10 ms) and the finger (80 – 120 ms) (Moraes et al., 2018).

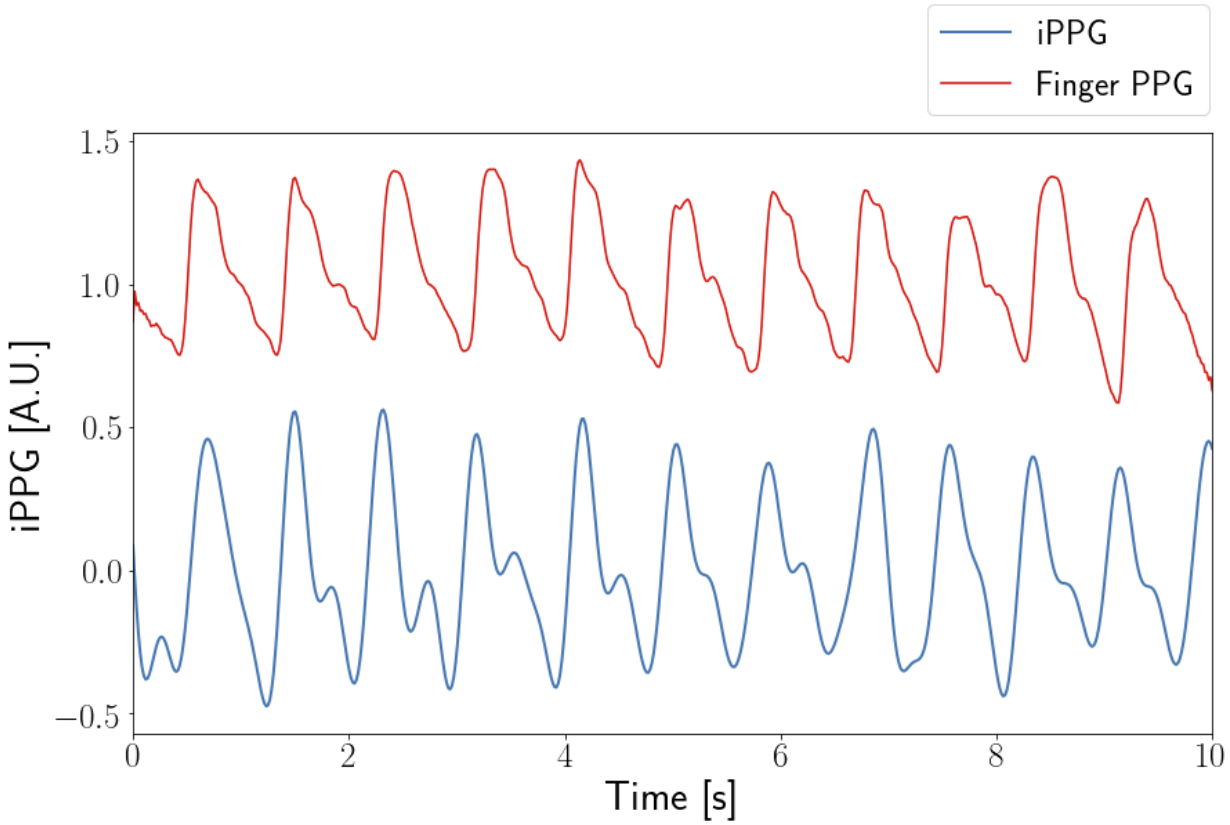

Figure 7.2: Comparison between the iPPG signal acquired from the facial hemoglobin map and the finger PPG signal acquired using a reflectance photoplethysmogram attached to the finger over the course of ten seconds. The minute discrepancy in timing that begins to appear towards the end of the time window is due to the difference in time it takes for the blood pulse from the heart to reach the face (10 ms) in comparison to the finger (80 – 120 ms).

We now apply Fourier analysis to the extracted iPPG signal to determine the heart rate. We apply a short-time Fourier transform (STFT) to the signal using a 6s moving window to match the standardisation. In Figure 7.3 we display the spectrogram of the iPPG signal over the time course of the experiment. We note the increase in heart rate activity expected due to the grip task, and the subsequent lowering of the heart rate as the participant relaxes.

To compute the instantaneous heart rate, we apply a curve extraction algorithm due to Cicone and H. T. Wu, 2017,

$$c^* = \arg\max_{c \in \mathbb{N}^{n_t}} \sum_{t=1}^{n_t} \log |S_{\text{iPPG}}(t, c(t))| - \lambda \sum_{t=2}^{n_t} |c(t) - c(t-1)| \in \mathbb{N}^{n_t} \tag{7.2.2}$$

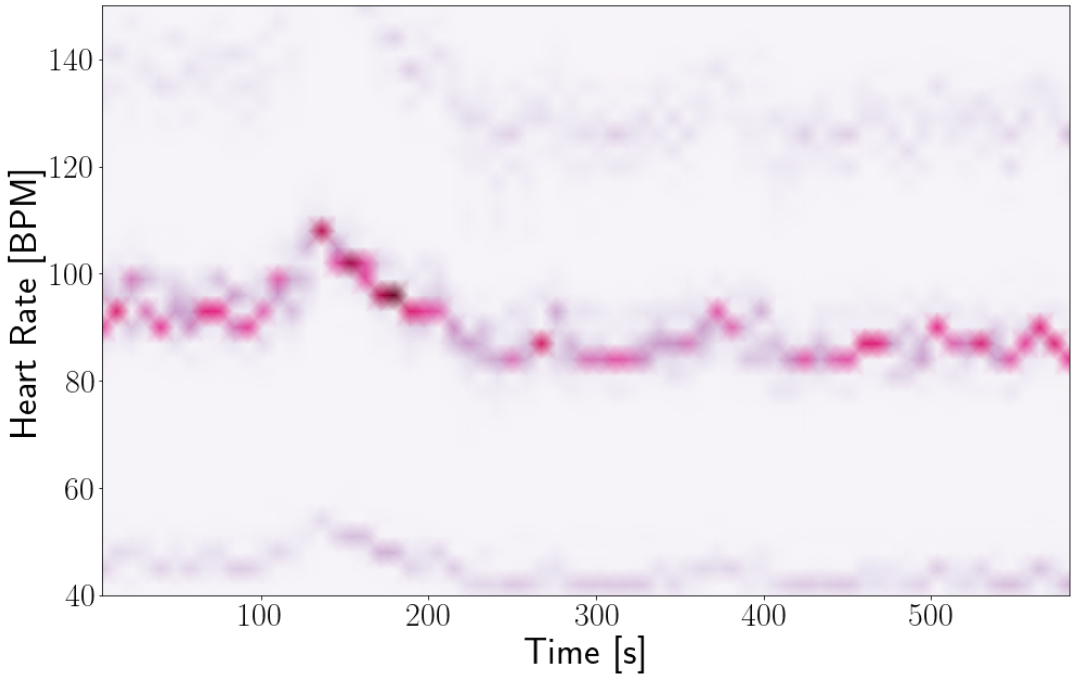

Figure 7.3: Spectrogram of the hemoglobin iPPG signal over the course of the experiment. Note the highest concentration of signal power is within range of the heart rate, and the rise in heart rate that is observed during the grip task.

where $S_{\text{iPPG}} \in \mathbb{C}^{n_t \times (n_t/2)}$ is the STFT of the iPPG signal, with $S_{\text{iPPG}}(t, f)$ the STFT coefficient at time t/ω, and frequency f/T, where n_t is the number of frames, T is the length of the video, ω is the sample rate, and $\lambda > 0$ is a regularisation parameter. The instantaneous heart rate from the iPPG signal is

$$HR = \frac{c^*}{T} \in \mathbb{R}^{n_t}. \tag{7.2.3}$$

We compare this to the instantaneous heart rate estimated from the ECG signal computed using the BioSPPy Python Package (Carreiras et al., 2015). In Figure 7.4 we display a comparison between the instantaneously computed heart rate from the iPPG signal from (7.2.2) and the instantaneous heart rate computed with the ECG.

To compare the performance of our method against the existing methods in the literature, we wrote a Python software package of implementations of the methods mentioned in Section 7.1. The methods are *GREEN* (Verkruysse, Svaasand, and Nelson, 2008), *ICA* (Poh, D. J. McDuff, and Picard, 2010), *CHROM* (Gerard de Haan and Jeanne, 2013) and its variant *POS* (W. Wang et al., 2017). We measure the root mean square error between the computed

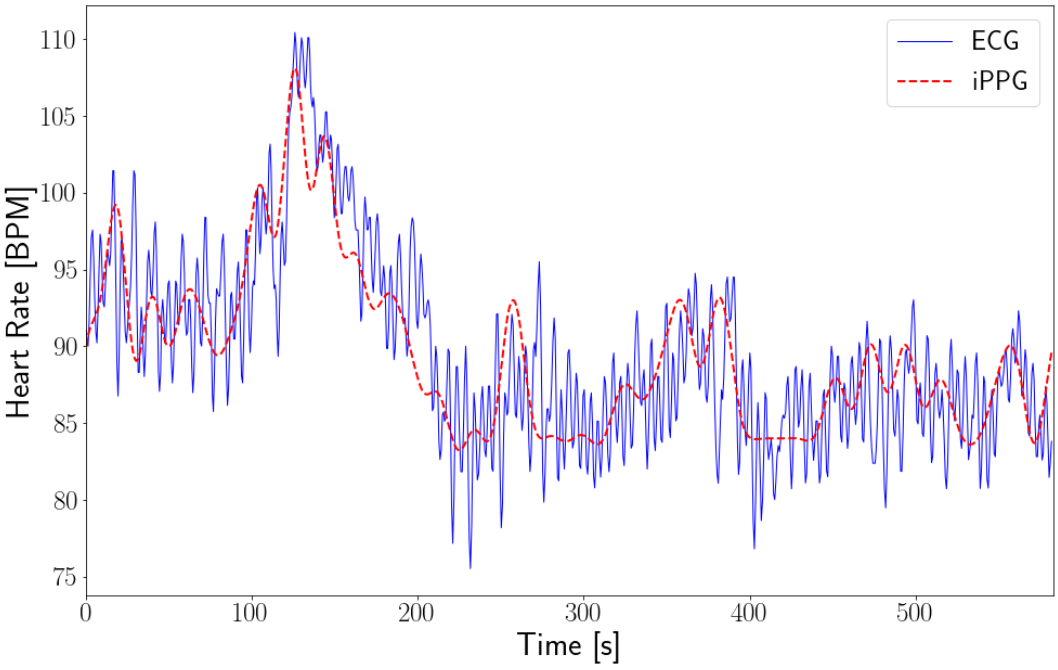

Figure 7.4: Comparison of instantaneous heart rate calculated from the ECG and from the extracted curve of the iPPG spectrum.

instantaneous heart rate for each method and the instantaneous heart rate as calculated from the ECG. The results are displayed in Talble 7.1. Our method achieves the best performance as measured by RMSE. The POS and CHROM methods perform adequately well, while the GREEN method achieves reasonable performance. The ICA method applied to the same skin mask as our method as the region of choice displays similar performance characteristics to our method.

In addition to evaluating the inverse chromophore map for heart rate extraction using the entire face, we also evaluated the performance of the methodology by only applying the inverse chromophore map to the forehead region. In some situations, a view of the entire face may be difficult to obtain, such as clinical settings where face mask wearing is required or for individuals with facial hair, obstructing visible regions of the face. Accordingly, we select the forehead region for evaluation, since it is typically exposed in most settings, even when mask wearing is required. The results of the application of our method as well as the existing methods in the literature are displayed in Table 7.2. With the exception of the GREEN method (Verkruysse, Svaasand, and Nelson, 2008) which uses the forehead region in its standard implementation, all of the methods sustain a slight decrease in performance due to the smaller pixel region to analyse. Our method still achieves the best performance, with

Table 7.1: Comparison of instantaneous heart rate accuracy as measured from iPPG techniques.

Method	RMSE (mean ± std)
GREEN	10.03 ± 3.85
ICA	4.18 ± 0.70
CHROM	5.37 ± 1.12
POS	4.82 ± 0.65
HEM iPGG (Ours)	3.64 ± 0.41

very acceptable error over all experiments. These results suggest that iPPG techniques and especially the inverse chromophore map are well suited for use in settings where face masks are required to be worn or an individual has dense facial hair, leaving only the forehead region of the face exposed.

Table 7.2: Comparison of instantaneous heart rate accuracy as measured from iPPG techniques using only forehead region of the face.

Method	RMSE (mean ± std)
GREEN	10.03 ± 3.85
ICA	5.75 ± 1.17
CHROM	6.83 ± 1.31
POS	7.02 ± 1.46
HEM iPGG (Ours)	5.12 ± 0.85

7.3 Imaging Electrocardiography

For the final application of the inverse chromophore map we present in this dissertation, we endeavour to recover the ECG signal of a subject using the recovered iPPG blood flow signal from Section 7.2. This represents an original *imaging electrocardiography* technique which may be used for cardiac patient monitoring outside of a clinical setting without the requirement for specialised equipment such as wearable ECG devices. Abnormalities in the ECG signal represent the clinical gold standard for diagnosing cardiopathophysiologies such as acute myocardial infarction, and thus a low-cost measurement of the ECG signal without the need for specialised equipment has significant potential for monitoring (De Luca et al., 2004; Meischke et al., 1995; Chew, Moliterno, and Herrmann, 2002).

More specifically, shifts in the ST-segment of the ECG signal are a highly sensitive and specific biomarker of ischemia and are known as a harbinger of acute myocardial infarction (Herring and Paterson, 2018). In the U.S., the average time from symptom onset to hospital arrival for acute myocardial infarction is 2.5 to 3 hours (Walsh, E. J. Topol, and Steinhubl, 2014). This time delay for treatment may result in lethal arrhythmias and irreversible myocardial

infarction. As such, the ability of patients to identify such changes outside of clinical settings represents a significant step forward in patient-specific medical care, and should lead to improved clinical outcomes.

To estimate the ECG signal from the facial blood flow iPPG signal, we form a linear mapping between the real Fourier transform coefficients of each signal over the same time course. Once this linear mapping is learned, we then apply the inverse Fourier transform to recover the ECG signal. This work is partially inspired by Tian et al., 2020 who apply a machine learning technique known as dictionary learning to the finger based PPG signal to recover the ECG waveform. Our method differs from their method in that we use a simple linear regression framework to learn the ECG waveform, incurring minimal computational expense. Additionally, we use only the recoverd facial hemoglobin variations measured from a smartphone, rather than the PPG signal acquired from the finger, which requires a specialised photoplethysmogram device.

Let $\boldsymbol{\upsilon} \in \mathbb{R}^T$ and $\boldsymbol{\gamma} \in \mathbb{R}^T$ be the iPPG and ECG signals, respectively, of length T with n cardiac cycles. Note that we downsample the ECG signal from 200 Hz in the experiments to the 60 Hz frame rate of the iPPG signal. The first step of the process is to temporally synchronise the ECG waveform the facial blood flow iPPG signal by the relative timings of the cardiac cycle. As discussed in Section 2.3 the diastolic trough of the PPG cycle corresponds to the R wave of the ECG waveform in a cardiac cycle; hence, we synchronise the two signal by matching the diastolic trough of the facial blood flow iPPG signal to the R wave of the ECG. Since the time delay between the pulse waveform in the heart and the facial arteries is small – 10 ms – there is no need to match different cardiac cycles together. We identify the R waves of the ECG, and the diastolic troughs and systolic peaks of the iPPG signal using a peak finding algorithm (Scholkmann, Boss, and Wolf, 2012). In Figure 7.5 a) we display the unsynchronised ECG and iPPG waveforms along with the R waves of the ECG signal and the systolic peaks and diastolic troughs of the iPPG signal. In Figure 7.5 b) we display the synchronised ECG and iPPG signals.

The signals are then segmented into individual cardic cycles according to the cardiac cycle length. The cardiac cycle length is computed as the temporal difference between subsequent diastolic troughs of the synchronised iPPG signal. While the cardiac cycle length may be more accurately computed as the temporal difference between the R peaks, this data will not be available in practice, and as such we rely on the iPPG signal alone. Each waveform is segmented using 30% of the data to the left of each diastolic trough, and the remaining 70% of the data to the right of the diastolic trough, so as to capture the P-wave of the ECG cycle, signifiying SA node depolarisation. Each segment is then standardised by subtracting the mean of each segment and dividing by the standard deviation of each segment. In Figure 7.6 we display a single cardiac cycle of the temporally synchronised, detrended, and standardised ECG and iPPG facial blood flow signals. We then interpolate the segments so that each segment is of length m, and collect the iPPG and ECG signals into matrices $\boldsymbol{\Upsilon}, \boldsymbol{\Gamma} \in \mathbb{R}^{n \times m}$.

We take the Fourier transform of $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Gamma}$ to get $\tilde{\boldsymbol{\Upsilon}}$, $\tilde{\boldsymbol{\Gamma}}$ and retain only the real coefficients for learning the map between the two. We then split the data into a training and test set $\tilde{\boldsymbol{\Upsilon}}_{train} \in \mathbb{R}^{n_{train} \times m}$, $\tilde{\boldsymbol{\Upsilon}}_{test} \in \mathbb{R}^{n_{test} \times m}$, $\tilde{\boldsymbol{\Gamma}}_{train} \in \mathbb{R}^{n_{train} \times m}$, $\tilde{\boldsymbol{\Gamma}}_{test} \in \mathbb{R}^{n_{test} \times m}$, where $n_{train} + n_{test} = n$, with 80% of the data used for the training set n_{train}, and the remaining 20% used for the test set n_{test}.

Training consists of solving a linear regression problem to learn $\tilde{\boldsymbol{\Gamma}}$ from $\tilde{\boldsymbol{\Upsilon}}$. The analytic solution to the linear regression problem is the map G

$$G = (\tilde{\boldsymbol{\Upsilon}}_{train}^{\top} \tilde{\boldsymbol{\Upsilon}}_{train})^{-1} \tilde{\boldsymbol{\Upsilon}}_{train}^{\top} \tilde{\boldsymbol{\Gamma}}_{train}. \tag{7.3.1}$$

We then evaluate $G\tilde{\boldsymbol{\Upsilon}}_{train} = \tilde{\boldsymbol{\Gamma}}^*$ where $\tilde{\boldsymbol{\Gamma}}^*$ is our estimate of the $\tilde{\boldsymbol{\Gamma}}_{train}$. We then apply the inverse Fourier transform to $\tilde{\boldsymbol{\Gamma}}^*$ to recover $\boldsymbol{\Gamma}^*_{test}$.

To assess the performance of our imaging electrocardiography methodology, we perform two sets of experiments using our experimentally collected data sets. As the experimental data consists of two different conditions over the time course of recording, once while the participant is at rest, and one while the participant is performing a physiologically challenging strength exercise, this serves as an appropriate setting for the evaluation of the imaging electrocardiography method to distinguish physiolgical changes outside of clinical settings.

In each experiment, we divide the iPPG signal generated from the inverse chromophore map and the ECG signal into *no-stress* and *stress* segments depending on the time frame of the experiment. We then train the linear map on the no-stress segment and test on another no-stress segment, or test on the stress segment. Conversely, we also train on the stress segment and then test on the stress segment and no-stress segments, respectively. The more difficult, and practically relevant, case is the application of a learned map to recover the ECG signal for a subject without calibration data. We performed a leave-one-out training procedure, where the coefficient mapping was learned on seven of the eight participants and then tested on the final participant without any exposure to the ECG of the participant, again repeating the no-stress and stress training and testing procedure.

The second experiment followed the same no-stress and stress training and testing procedure, both on the same subject and for applying a learned map to different subjects, but used only the forehead region of each participants faced.

The performance of the method for each experiment is displayed in Table 7.3, where Face-Same refers to the case where the method is trained using the entire face and tested on the same participant; Face-Transfer refers to the case where the method is trained using the entire face and test on the different participant using leave-one-out cross validation; Forehead-Same refers to the case where the method is trained using only the forehead and tested on the same participant; Forehead-Transfer refers to the case where the method is trained using only the forehead and tested on a different participant using leave-one-out cross validation. As expected, the performance for the imaging electrocardiography method is

greatest using the whole face, and applying the learned map to the same participant within the same segment. The error naturally increases when the learned map is applied to different participants, however, interestingly, the performance when the training and testing segment coincide is excellent and comparable to the performance when the participant is the same. This suggests that a careful data-collection procedure is required when training the algorithm for large scale use, as the training set must have a sufficiently broad range of ECG features and physiological conditions to achieve desirable performance.

When the mapping is learned only using the hemoglobin variation signal from the forehead region, the performance of the imaging electrocardiography method further declines with respect to using the entire face, however still performs reasonably well. A possible explanation for this difference in performance when using the hemoglobin signal from the entire face and using the hemoglobin signal from only the forehead lies in the arterial blood supply to the face and the facial vasculature structure. The face is supplied with blood from the *external* and *internal carotid arteries* which service different regions of the face with blood (Arx et al., 2018). The external carotid arteries supplies the cheeks, lateral temples, tip of the nose, mouth, and chin with blood, while the internal carotid arteries supply the forehead, brows, eyes, and upper nose. See Figure 7.7 for reference. When only the forehead region is used to extract the hemoglobin variation signal rather than the entire face, important information concerning cardiovascular dynamics from the external carotid arteries may be lost, leading to a decrease in performance. Additionally, there is a greater amount of camera quantisation noise by averaging over a smaller spatial region, potentially leading to a loss of performance.

Table 7.3: Imaging electrocardiography performance measured by relative root mean square error (mean ± standard deviation) for no-stress and stress training and testing procedure for same subject and novel subject examples, as well as forehead hemoglobin based example. The training → testing examples are: NS → NS (no-stress → no-stress), NS → S (no-stress → stres), S → S (stress → stress), and S → NS (stress to no-stress).

	Relative RMSE			
	NS → NS	NS → S	S → S	S → NS
Face-Same	0.505 ± 0.082	0.748 ± 0.146	0.547 ± 0.087	0.686 ± 0.073
Face-Transfer	0.779 ± 0.096	1.324 ± 0.261	0.702 ± 0.094	1.289 ± 0.195
Forehead-Same	1.270 ± 0.128	1.393 ± 0.315	1.341 ± 0.106	1.385 ± 0.184
Forehead-Transfer	1.258 ± 0.229	1.415 ± 0.443	1.445 ± 0.281	1.458 ± 0.371

In Figure 7.8 we display a comparison between the learned ECG signal from the imaging electrocardiography method applied to the whole face and the actual ECG signal for the same participant. The method is trained and tested on the no-stress segment. We observe the excellent agreement between the two signals, with all of the constituent waveforms of the ECG morphology accurately reconstructed.

Figure 7.9 displays a comparison between the learned ECG signal from the imaging electrocardiography method applied to the whole face and the actual ECG signal for the same participant with training and testing both applied to the stress region. The method is able to accurately reconstruct the signal with minimal error and accurately captures qualitatively different features of the ECG signal during the stress time segment such as the supressed P-wave.

Figure 7.10 displays a comparison between the learned ECG signal from the imaging electrocardiography method applied to the whole face and the actual ECG signal for the leave-one-out participant transfer procedure. The method is trained on the no-stress segment for all but one of the participants and tested on the no-stress segment of the remaining participant. We observe again excellent agreement between the two signals.

In Figure 7.11 we display a comparison between the learned ECG signal from the imaging electrocardiography method applied to the whole face and the actual ECG signal for the leave-one-out participant transfer procedure. The method is trained on the stress segment for all but one of the participants and tested on the no-stress segment of the remaining participant. The representative waveforms display excellent agreement, however there is a small time-shift between the two signals which occurs at different times. This is caused by slight differences between the cycle matching where by the trough of the facial hemoglobin cycle and the R-peak of the ECG signal do not exactly align for every cycle.

Figure 7.12 a demonstration of the imaging electrocardiography method applied to the hemoglobin variation signal from the forehead compared to the actual ECG signal. The method is trained on the stress segment of the experiment and tested on the stress segment of the experiment for the same participant. The essential morphology of the waveform is captured, however there are variations in the values of some of the waveform features in the learned ECG signal that do not appear when the entire face is used for the inverse chromophore map.

Finally, Figure 7.13 displays a comparison between the learned ECG signal from the imaging electrocardiography method applied to forehead and the actual ECG signal for the leave-one-out participant transfer procedure. The method is trained on the no-stress segment for all but one of the participants and tested on the stress segment of the remaining participant. The method encounters distinct difficulties in this case where by features of the ECG waveform are absent, exaggerated, or incorrectly displayed.

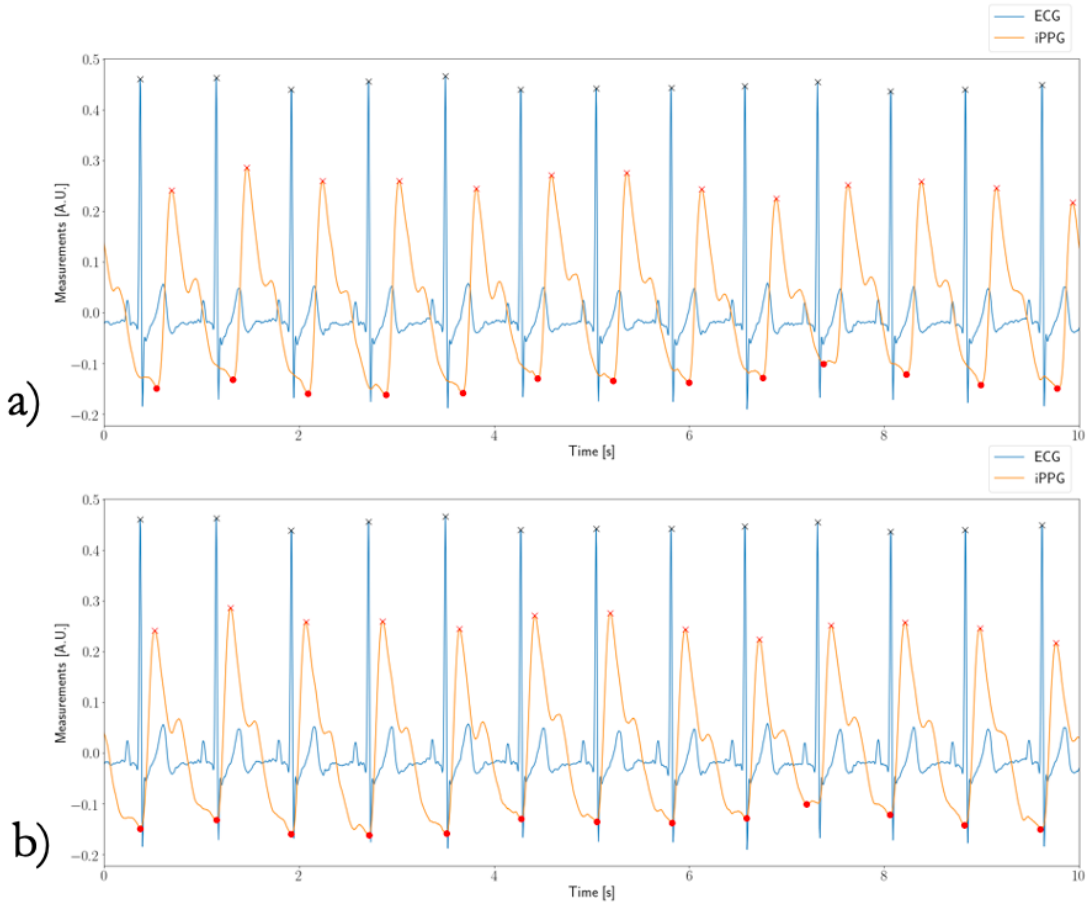

Figure 7.5: Comparison beween ECG and facial blood flow iPPG signal over a 10s time window. a) The unsynchronised ECG and iPPG waveforms with R waves denoted by black crosses, the systolic peaks of the iPPG waveform denoted with red crosses, and the diastolic troughs of the iPPG waveform denoted with red circles. b) The temporally synchronised ECG and iPPG waveforms formed by matching the R waves to diastolic troughs. The R waves are denoted by black crosses, the systolic peaks of the iPPG waveform denoted with red crosses, and the diastolic troughs of the iPPG waveform denoted with red circles.

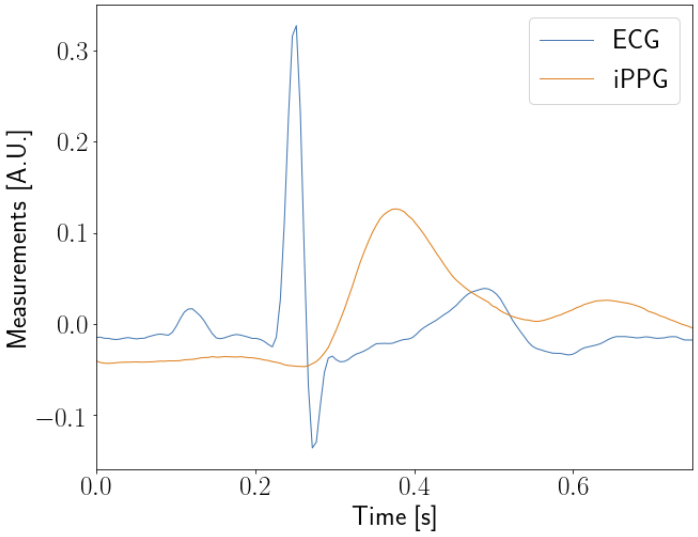

Figure 7.6: Single cardiac cycle of the temporally synchronised, detrended, and standardised ECG and iPPG facial blood flow signals.

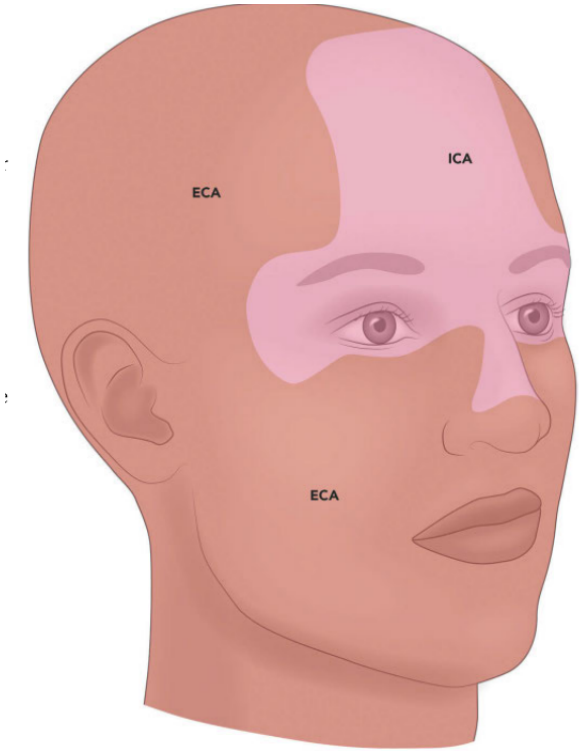

Figure 7.7: Schematic of regions of the face as supplied blood by the external carotid arteries (ECA) and internal carotid arteries (ICA). Figure originally appeared in Arx et al., 2018 and is reproduced with permission.

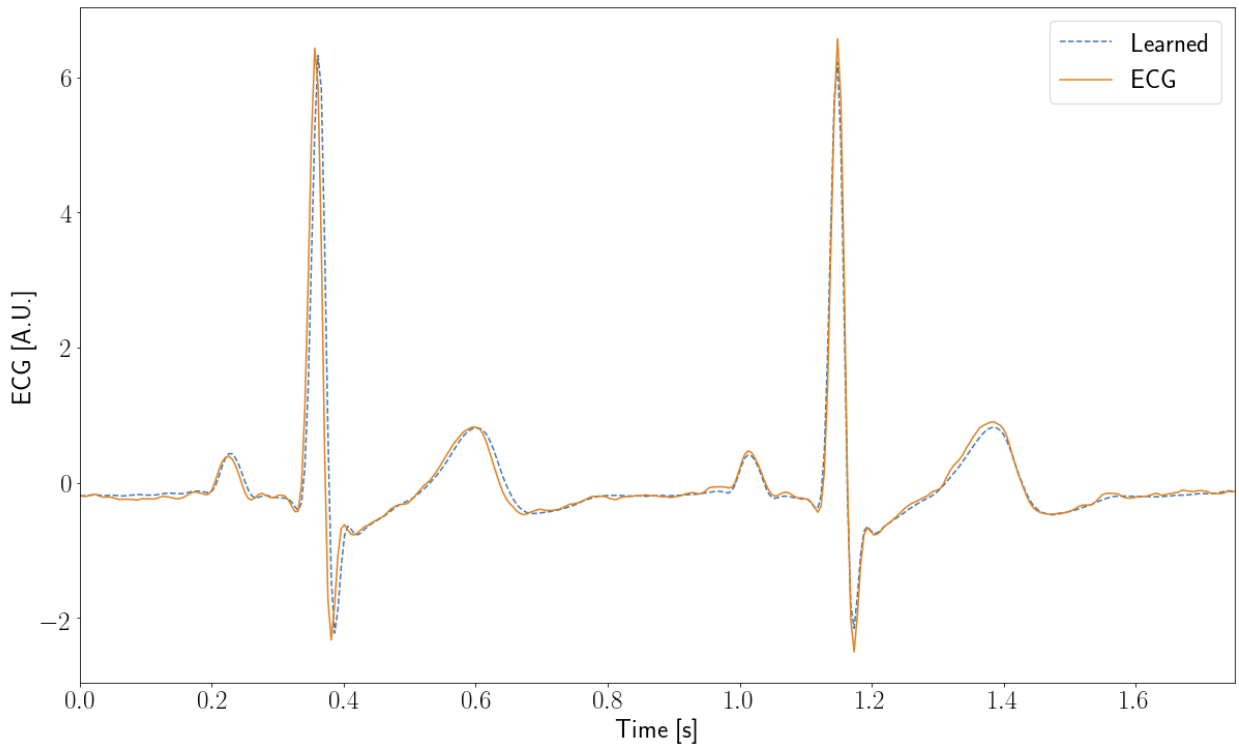

Figure 7.8: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method for the same participant trained and tested on the no-stress segment.

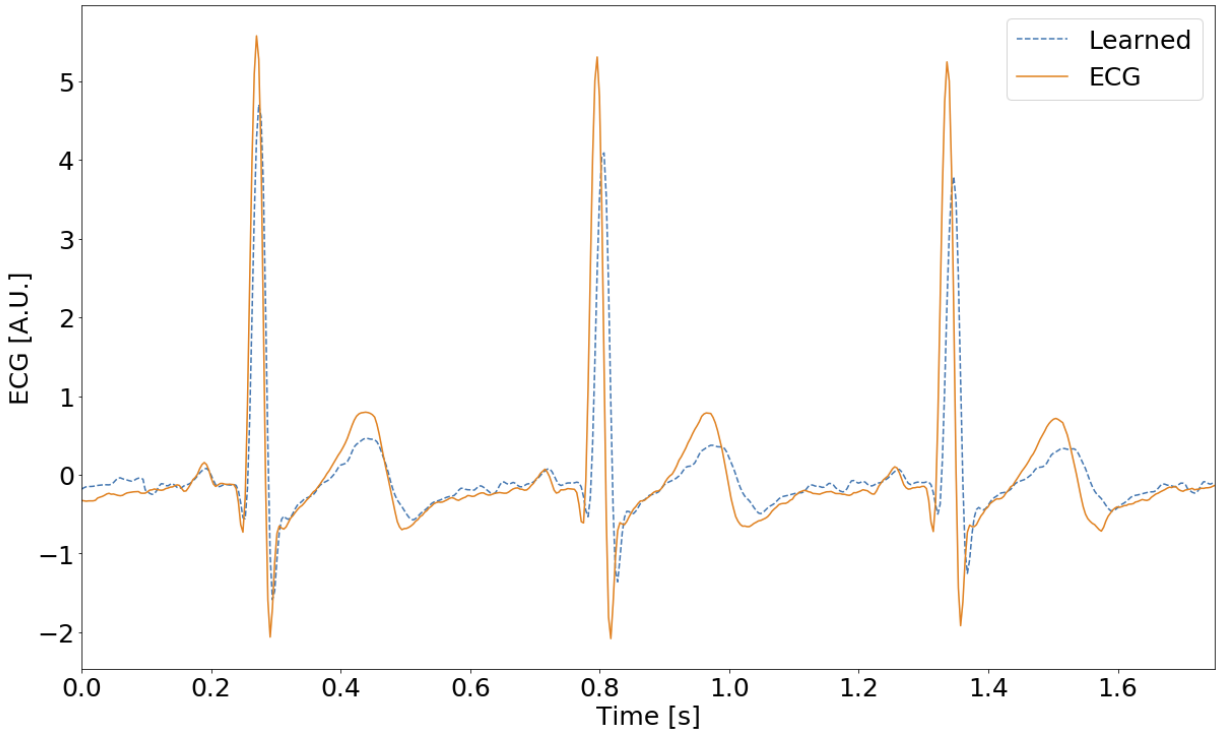

Figure 7.9: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method for the same participant trained and tested on the stress segment.

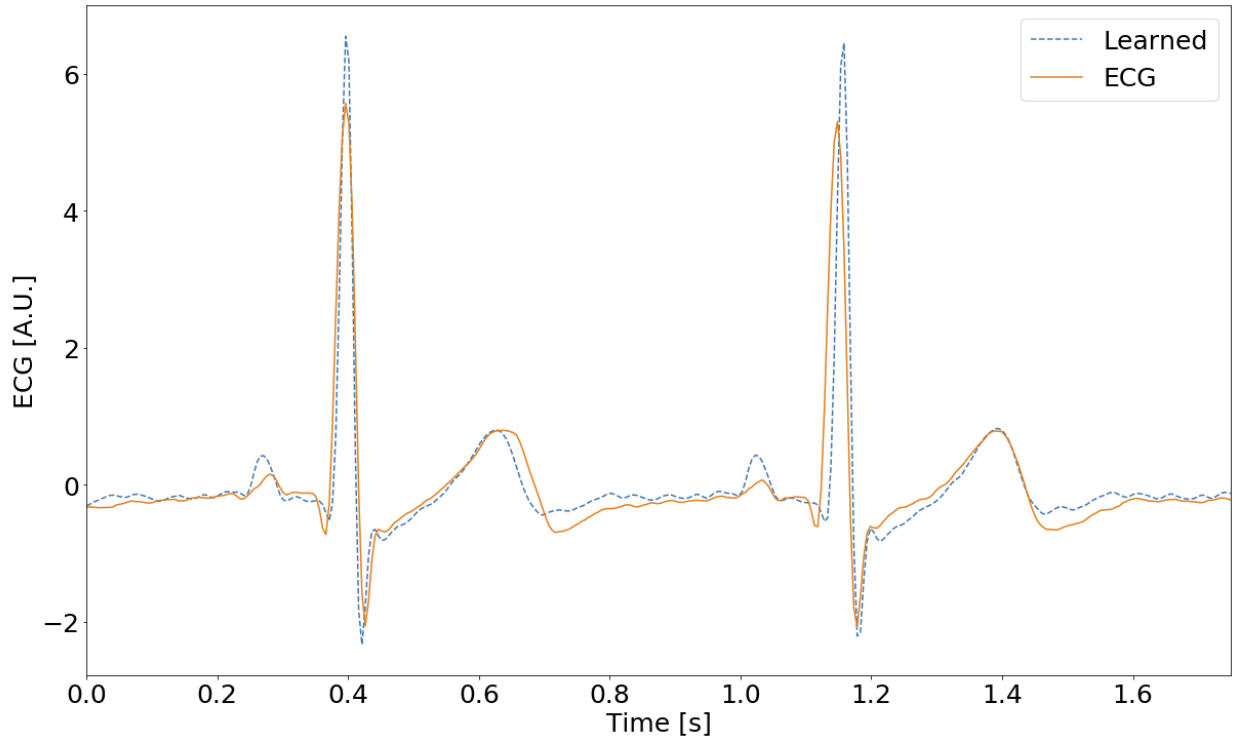

Figure 7.10: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method from the leave-one-out transfer procedure trained and tested on the no-stress segment.

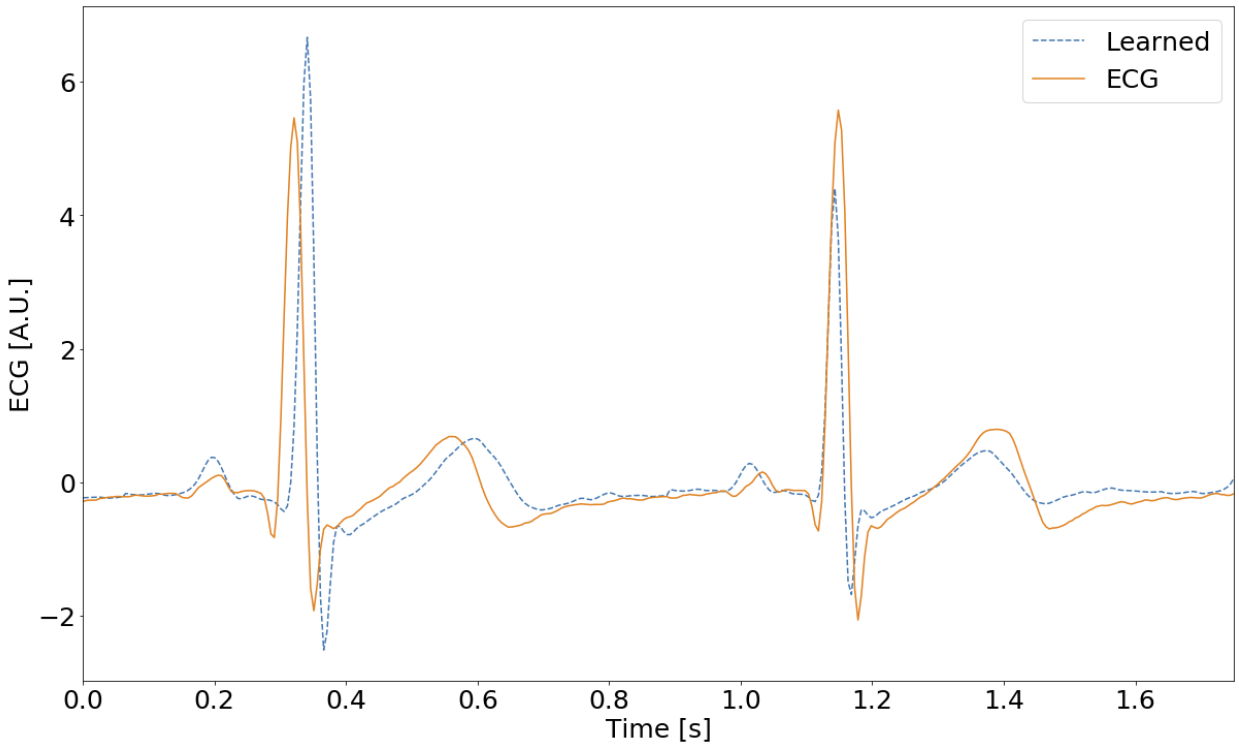

Figure 7.11: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method from the leave-one-out transfer procedure trained on the stress segment and tested on the no-stress segment.

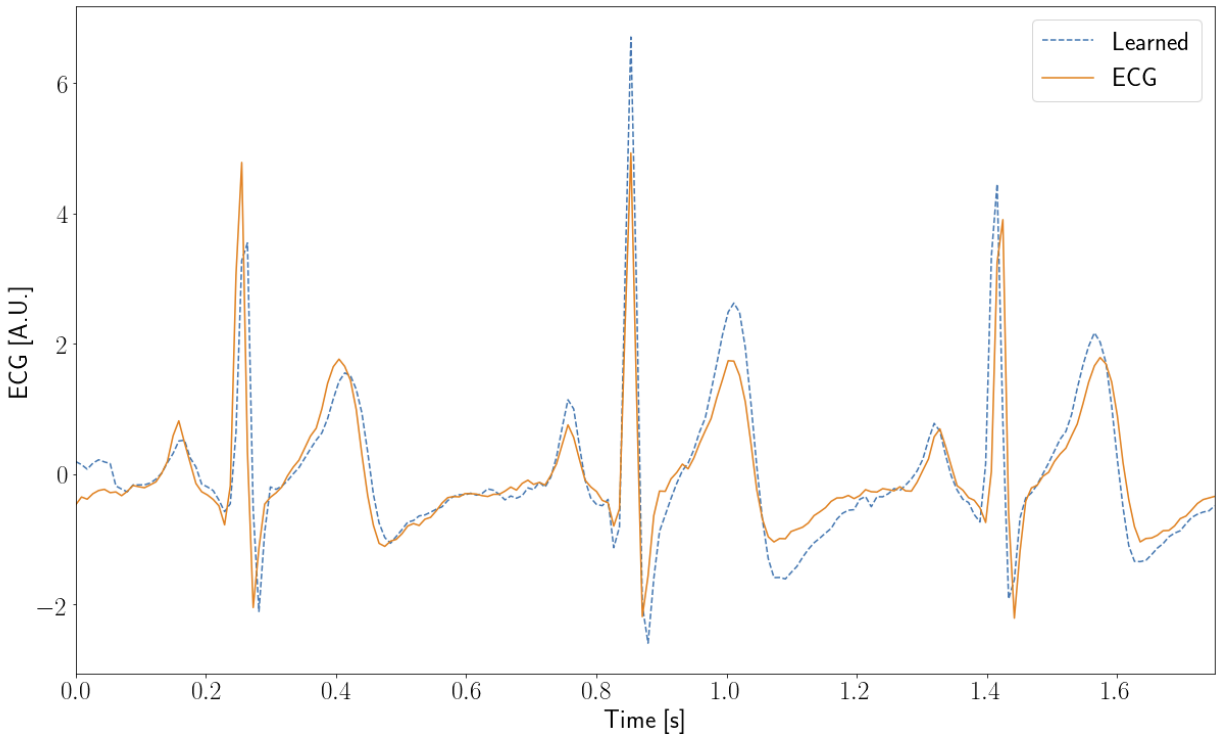

Figure 7.12: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method applied to the forehead hemoglobin signal for the same participant trained and tested on the stress segment.

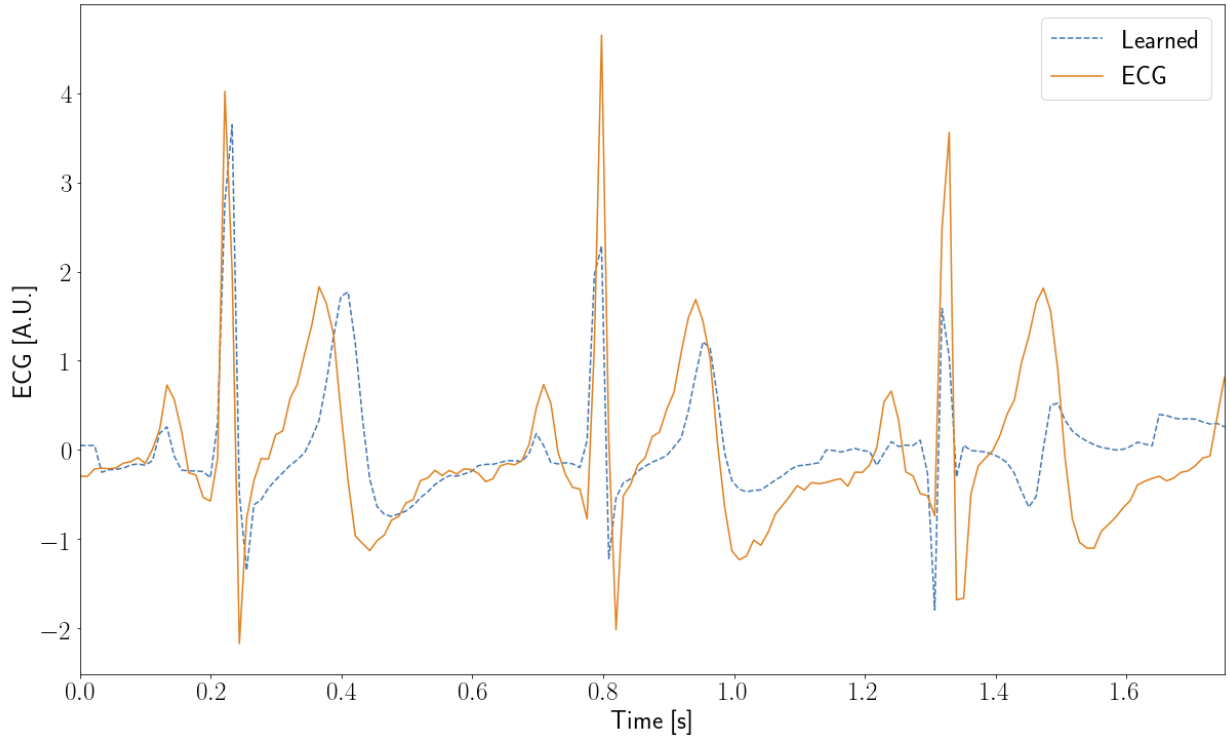

Figure 7.13: Comparison between ECG waveform and learned ECG waveform from the imaging electrocardiography method applied to the forehead hemoglobin signal for the leave-one-out transfer procedure trained on the no-stress segment and tested on the stress segment.

Chapter 8

Conclusion

> "*A lot of the diagnosis and monitoring functions will be done through little devices – smartphones – by the patient with computer assistance. So it's a real big change in the model of how we render healthare.*"
>
> — Eric Topol
> *New technologies put the patient first*

This dissertation has been concerned with the development of a multi-stage opto-physiological modelling methodology for the recovery of chromophore concentrations from digital camera images. We have demonstrated how the recovered chromophore concentrations may be used to render a map of the underlying vasculature of the imaged skin, and we have shown how such a methodology may provide contact free measurements of a person's vital signs using only a smartphone camera. In this chapter, we conclude the dissertation by reviewing our work, discussing limitations to our present work, and finally providing suggestions for future research directions. We conclude with an offering of our personal view of digital medicine and its future.

8.1 Dissertation Review

The dissertation was the sum of two parts, with the first consisting of relevant background material and slight modifications or novel applications of methods in the literature, while the second consisted of entirely novel material. Briefly, the first portion of the dissertation collects all of the necessary components to construct the multistage inverse chromophore mapping methodology, including a review of anatomy, radiative transport for light interaction in skin tissue, and digital camera imaging. In this portion of the dissertation, we present a

novel application of interlayer light scattering to skin tissue, as well as a modification of a camera imaging pipeline transformation to deal with small data sets. In the second part of the disseration, we present the novel inverse chromophore mapping methodology, construct a map of the underlying vasculature from a digital skin image, and apply the methodology for vital sign recovery from sequential video frames of people. In the following we provide a detailed description of each chapter.

We began the dissertation in Chapter 2 with a self-contained review of skin tissue and cardiovascular anatomy. We discussed the skin as a light scattering and absorbing turbid medium, and provided a brief overview of the structure of the skin and its contents. The cardiovascular system and the cardiac cycle was reviewed in detail, including a review of the electrical activity of the heart. We then reviewed photoplethysmography and introduced the Bouguer-Beer-Lambert law as a simple explanation for light interaction in skin tissue.

In Chapter 3 we reviewed Bayesian inference and Gaussian Process (GPs) regression to set the stage for the use of GPs as the inverse chromophore map developed in Chapter 6. We covered the concept of prior probability and the likelihood function, as well as the posterior distribution, and then demonstrated how GP regression performs Bayesian inference to learn a functional relationship between input data and observations. Several covariance functions or kernels were introduced and the differences in functions learned using different kernels was discussed. We also discussed the computational challenges associated with GP inference and reviewed recent developments in the literature using Krylov subspace methods to speed up computation. Details of the structured kernel interpolation method (Wilson and Nickisch, 2015) used in Chapter 6 were given for reference.

Chapter 4 introduced the partial-integral radiative transport equation (RTE) as a model of light-interaction in skin tissue. We derived the RTE using conservation arguments, and then additionally derived the P_N infinite system of coupled partial differential equations approximation by expanding the photon density in a spherical harmonic basis. We then showed how the P_1 approximation gives rise to the famous diffusion approximation to light interaction in skin tissue. We discussed the mismatched index boundary condition that arises from the diffusion approximation and demonstrated how it could be solved using Fresnel reflection coefficients. We derived the Green's function for the case of light propagation in an infinite medium and considered the diffusion approximation in the slab geometry using the Method of Images. Finally, we applied the diffusion approximation and Kubelka-Munk approximation for inter-layer light scattering introduced for general light scattering material in the computer graphics community for the novel application of computeing the reflectance and transmittance for layered skin tissue.

In Chapter 5 we constructed the final piece of the multi-stage modelling methodology by analysing the digital camera imaging processing pipeline. An overview of the digital camera in-camera imaging procesisng pipeline was introduced and we discussed the nonlinear colour mapping techniques digital cameras apply to images to make them appear visually pleasing to the end user. We recovered the spectral sensitivity functions of an Apple iPhone 6S using

PCA analysis from a RAW image of a calibrated colour checker chart. Then we estimated the colour mapping functions of the Apple iPhone 6S camera by analysing a pair of RAW and rendered colour image captures of a calibrated colour checker chart using a modification of an existing method in the literature with Gaussian Process regression rather than the support vector regression technique of the literature to overcome a limited size data set. The estimated colour map was found to have comparable error to results in the literature, however, was limited by the availability of data.

Chapter 6 presented the complete novel multi-stage methodology for rendering skin images. We began with a construction of a three layered model of skin tissue consisting of the stratum corneum, epidermis, and dermis, and computed the total reflectance across the visible wavelength using models of the absorption and scattering coefficients for each layer of skin tissue taken from the literature. We then applied the camera imaging pipeline we estimated in Chapter 5 to form a rendered skin sample in the RGB pixel space by exhaustively varying the chromophore concentrations of hemoglobin and melanin over physiologically relevant ranges. The rendered skin samples were used as training data for a Gaussian process model to recover the chromophore concentrations, given each pixel value. This inverse chromophore map was then applied to both simulated and real skin image to recover the hemoglobin concentration across a variety of simulated skin tones. The accuracy of the recovered hemoglobin volume fraction decreased as the volume fraction of melanin in the simulated skin tissue increased, representative of the increased scattering caused by melanin. The inverse chromophore maps applied to the simulated skin images with embedded vessels were able to identify the positions of the embedded blood vessels with excellent accuracy, forming a map of the underlying vasculature. We also applied the inverse chromophore map to rendered skin images with curved and connected embedded vessels, as well as curved and connected vessels at different depths, for which the method was able to extract the precise position of each vessel across a range of chromophore concentrations. We implemented a simplified simulation of blood flow in the vessels due to the pumping action of the heart by temporally varying the hemoglobin concentration in each embedded blood vessel and applied the inverse chromophore map to the sequentially rendered skin images. The inverse chromophore map recovered the dynamics of the frame-to-frame hemoglobin variation in each case, and was able to identify frequency changes in the simulated signal. We finally applied the inverse chromophore map to experimentally collected images of real skin tisue extracting an underlying vasculature map by identifying vessels with increased hemoglobin concentration and applying image processing techniques.

Finally, in Chapter 7 we applied the inverse vasculature map to video frames to recover a person's vital signs using a smartphone camera. We began the chapter with the presentation of a mathematical framework for image based photoplethysmography systems, and discussed how the methods in the literature fit into the framework. We also discussed how the results of the inverse chromophore map bypassed many of the difficulties encountered by imaging photoplethysmography techniques in the literature by returning the time varying hemoglobin signal when applied to a sequence of video frames. We then applied our method

to an experimentally collected data set of participants performing a cardiovascular challenge task, and were able to accurately identify the heart rate of the participants in different circumstances. We implemented various methods in the literature in a Python package and compared their performance to our methodology, for which we achieved state-of-the art performance. We also noted the ability of the inverse chromophore map to accurately recover a participants heart rate when only applied to the forehead region of their face, forming a clinically relevant case study in situations where a mask, face-covering, or facial hair may obfuscate other regions of the face. In the final section, we introduced our novel imaging electrocardiography methodology by learning a mapping from the facial blood flow signal to the ECG waveform of participants. We trained and tested the imaging electrocardiography method under different experimental physiological conditions where a participant was forced to perform a cardiovascularly challenging task. The imaging electrocardiography method learned the ECG waveform from the facial hemoglobin variation signal with remarkable accuracy across both tasks, as well as in the practically relevant and difficult case of recovering a participant's ECG waveform without any calibration data. We also implemented the imaging electrocardiography method using only the hemoglobin variation recovered from the inverse chromophore map applied to the forehead with reasonable results.

8.2 Limitations

While this disseration provides a rigorous model description of imaging vital sign and camera based diagnostic systems, our work is not without its limitations.

The most significant limitation to our proposed methodology is rooted in data. The construction of our methodology relies upon accurate data in several places, including the recovery of the camera spectral sensitivity functions, the nonlinear mapping function of the in-camera imaging processing pipeline, and estimation of the illuminant. All of these factors may combine to significantly reduce the performance of our methodology. We were limited in the amount of data that we could collect for the camera spectral sensitivity function and colour mapping pipeline estimation due to practical reasons related to image capturing and file storage. We believe that further analysis and data collection will invariably improve the camera imaging pipeline and dramatically reduce the error associated with the colour mapping function to provide a more realistic representation of the camera imaging pipeline.

Further, the experimentally collected data for vital sign estimation is limited in its applications for medical diagnostics. All of the participants were healthy adults without any underlying health conditions, and the videos were captured in a controlled environment, with a fixed and constant source of illuminantion without external stimuli. Accordingly, we can draw limited conclusions as to how our methodology will scale across different health populations including those with cardiovascular conditions. Also, it is likely that the performance of our vital sign methodology will degrade in a less controlled environment due to additional noise sources.

The performance of our imaging electrocardiography technique is another limitation of the study, but one for which the inclusion of more data will lead to performance improvements. Our sample size of eight participants is comparatively small in relation to data sets that should be collected for much larger trials.

Finally, our work is also limited by our choice of diffuse light for the model of light interaction in skin tissue. In the setting of vital sign monitoring, there is likely to be non-trivial specular reflectance from the skin tissue, for which our model neglects. The performance of our inverse chromophore mapping procedure also degrades in the presence of higher concentrations of melanin.

8.3 Future Directions

Given the limitations of our methodology, there are plenty of opportunities for future research related to opto-physiological modelling of light-interaction in skin tissue. That said, we also believe that the proposed methodology provides a mathematically and biophysically rigorous foundation for which future work can be built upon. Our suggestions for future work range in complexity from further experimental studies to much larger scope simulation studies and engineering infrastructure development. Additionally, the following are all required for the development of a productionised version of the methodology for use with participants and potential patients in a variety of settings.

The most immediate direction for further research is in the conduction of two larger scale experiments across a diverse population. The first experiment would be to collect a large sampling of skin tissue photographs to assess the performance of the inverse chromophore map across different skin types. The second experiment would be collect additional videos and simultaneously recorded vital signs from a larger population, including those with pre-existing medical conditions. We plan to collect a data set of participants with cardiovascular conditions to view the performance of our methodology in such a population, as well as to develop classification techniques to identify those with pathologies and those without.

As stated in the limitations, a future research direction should be to include a specific model of specular reflectance to correct for over-saturated pixel values and to more accurately describe the reflected light collected at the camera distance. Additionally, the inclusion of multiple illuminant conditions would improve the range of operating conditions of the method, however, this comes at the cost of increased computational complexity.

The methodology may also be improved by directly simulating a much larger range of skin tones using Monte Carlo simulations of the RTE. Given the rise of GPU based computational practices, one could scale up the number of skin tones that are computed using Monte Carlo, and thus have greater accuracy for skin tone modelling. These simulations would also improve the performance of the inverse chromophore mapping procedure by providing greater fidelity simulation results for skin tones with higher concentrations of melanin. Further,

alternative inverse mapping procedures such as deep learning based approaches should be investigated.

To improve the performance of the methodology for different skin tones, especially more darkly pigmented tones with a larger concentration of melanin, we suggest the development of multiple inverse chromophore maps that are specially tuned to operate for darker pigmented skin. We envision this as a two-stage approach, where the initial inverse chromophore map may be used to identify a cluster of skin tones, for which then a more closely calibrated and tuned model can be applied for each cluster. This would enable increased accuracy for hemoglobin identification, as well as increased operability across diverse lighting conditions with more specific models for each skin type.

Additional avenues for future research involve directly addressing the limitation of camera spectral senstivity estimation and the colour mapping estimation problem. By working directly with a smartphone manufacturer, much of the error related to the camera component of the model can be reduced if not eliminated, as domain specific knowledge of the camera imaging pipeline for each smartphone can be explicitly included in the model. Finally, models for other smartphones should be developed and data should be collected with a variety of smartphones and cameras to increase the range of accessibility of the method.

Cardiology, and indeed much of medicine currently sits at the precipice of a paridgm-shift in the provision of care. We believe that smartphone-based camera imaging diagnostics represents the future of personalised medical care and long distance monitoring. To achieve this goal, a coalition of researchers, clinicians, and engineering capabilities are required to deliver usable and practical solutions to patients. As discussed by Sun and Thakor, 2016, rigorous opto-physiological modelling of light interaction in skin tissue is essential for the futher understanding and development of the photoplethysmogram waveform and remote imaging techniques. We believe that this work serves as a step forward in this direction, and hope for it to serve as a foundation, along with other literature in the area of imaging diagnostics, to build improved and accurate solutions upon. The doctor's office of the future may very well already exist in the smartphones we carry inside our pockets.

Bibliography

Abramowitz, Milton and Irene A. Stegun (1965). *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables (National Bureau of Standards Applied Mathematics Series No. 55)*. DOI: 10.1115/1.3625776. arXiv: 1701.01870. URL: http://appliedmechanics.asmedigitalcollection.asme.org/article.aspx?articleid=1396937.

Adamson, Philip B. et al. (2014). "Wireless pulmonary artery pressure monitoring guides management to reduce decompensation in heart failure with preserved ejection fraction". In: *Circulation: Heart Failure* 7.6, pp. 935–944. ISSN: 19413297. DOI: 10.1161/CIRCHEARTFAILURE.113.001229.

Allen, John (2007). "Photoplethysmography and its application in clinical physiological measurement." In: *Physiological measurement* 28.3, R1–39. ISSN: 0967-3334. DOI: 10.1088/0967-3334/28/3/R01. URL: http://www.ncbi.nlm.nih.gov/pubmed/17322588.

Alotaibi, Sarah and William A.P. Smith (2018). "A Biophysical 3D Morphable Model of Face Appearance". In: *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017* 2018-Janua, pp. 824–832. DOI: 10.1109/ICCVW.2017.102.

Anderson, R. R. and J. A. Parrish (1981). "The Optics of Human Skin". In: *Journal of Investigative Dermatology* 77.1, pp. 13–19. ISSN: 0022202X. DOI: 10.1111/1523-1747.ep12479191. URL: http://dx.doi.org/10.1111/1523-1747.ep12479191.

Antonicelli, Roberto et al. (2012). "Validation of the 3-lead tele-ECG versus the 12-lead tele-ECG and the conventional 12-lead ECG method in older people". In: *Journal of Telemedicine and Telecare* 18.2, pp. 104–108.

Arridge, S R (1999). "Topical Review: Optical Tomography in medical imaging". In: *Inverse Problems* 15, pp. 41–93.

Arridge, S. R., M. Cope, and D. T. Delpy (1992). "The theoretical basis for the determination of optical pathlengths in tissue: Temporal and frequency analysis". In: *Physics in Medicine and Biology* 37.7, pp. 1531–1560. ISSN: 00319155. DOI: 10.1088/0031-9155/37/7/005.

Arx, Thomas von et al. (2018). "The Face – A Vascular Perspective. A literature review". In: *Swiss dental journal* 128.5, pp. 382–392. ISSN: 2296-6498. URL: http://www.ncbi.nlm.nih.gov/pubmed/29734800.

Asllanaj, F. et al. (2019). "Light propagation in biological tissue". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 224, pp. 78–90. ISSN: 00224073. DOI: 10.1016/j.jqsrt.2018.11.001. URL: https://doi.org/10.1016/j.jqsrt.2018.11.001.

Bal, Guillaume (2009). "Inverse transport theory and applications". In: *Inverse Problems* 25.5. ISSN: 02665611. DOI: 10.1088/0266-5611/25/5/053001.

Baranowski, Megan et al. (2011). "The Kubelka-Munk Diffuse Reflectance Formula Revisited". In: *Applied Spectroscopy Reviews* 46.2, pp. 140–165. ISSN: 0570-4928. DOI: 10.1080/05704928.2010.537004.

Bashkatov, Alexey N. et al. (2000). "Optical properties of melanin in the skin and skin-like phantoms". In: *Proceedings of the SPIE Volume 4162, Controlling Tisue Optical Properties: Applications in Clinical Study*.

Beer, August (1852). "Bestimmung der Absorption des rothen Lichts in farbigem Flussigkeiten". In: *Annalen der Physik und Chemie* 85.5, pp. 78–88.

Berger, James O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second. New York, NY: Springer-Verlag New York, Inc.

Bishop, Christopher M (2006). *Pattern Recognition and Machine Learning*. Vol. 4. 4, p. 738. ISBN: 9780387310732. DOI: 10.1117/1.2819119. arXiv: 0-387-31073-8. URL: http://www.library.wisc.edu/selectedtocs/bg0137.pdf.

Bjerhammar, Arne (1951). "Application of calculus of matrices to method of least squares; with special references to geodetic calculations". In: *Transactions of the Royal Institute of Technology, Stockholm* 49.

Blackford, Ethan B. and Justin R. Estepp (2017). "Measurements of pulse rate using long-range imaging photoplethysmography and sunlight illumination outdoors". In: *Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics* 10072.February 2017, 100720S. ISSN: 16057422. DOI: 10.1117/12.2253460.

Blackford, Ethan B., Justin R. Estepp, et al. (2016). "Long-range non-contact imaging photoplethysmography: cardiac pulse wave sensing at a distance". In: *Optical Diagnostics and Sensing XVI: Toward Point-of-Care Diagnostics* 9715.March 2016, p. 971512. ISSN: 16057422. DOI: 10.1117/12.2208130.

Blazek, V., W. Rutten, and O. Such (1996). *A method for space-resolved, non-contacting and functional visualization of dermal perfusion (German)*.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. ISSN: 1537274X. DOI: 10.1080/01621459.2017.1285773. arXiv: 1601.00670. URL: https://doi.org/10.1080/01621459.2017.1285773.

Bouguer, Pierre (1729). *Essai d'optique sur la gradation de la lumiere*. Paris, France: Claude Jombert.

Bousefsaf, Frédéric, Alain Pruski, and Choubeila Maaoui (2019). "3D convolutional neural networks for remote pulse rate measurement and mapping from facial video". In: *Applied Sciences (Switzerland)* 9.20, pp. 1–21. ISSN: 20763417. DOI: 10.3390/app9204364.

Box, George E. P. (1980). "Sampling and Bayes inference in scientific modelling and robustness (with discussion)". In: *Journal of the Royal Statistical Society. Series A* 383-430.

Boyd, Stephen and Lieven Vandenberghe (2010). *Convex Optimization*. Vol. 25. 3, pp. 487–487. ISBN: 9780521833783. DOI: 10.1080/10556781003625177. arXiv: 1111.6189v1. URL: https://web.stanford.edu/%7B~%7Dboyd/cvxbook/bv%7B%5C_%7Dcvxbook.pdf.

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools*.
Brown, Michael S. (2019). "Understanding the in-camera image processing pipeline for compter vision". In: *IEEE International Conference on Computer Vision (ICCV)* Tutorial.
Buhmann, M.D. (2003). *Radial Basis Functions: Theory and Implementations*. Cambridge, UK: Cambridge University Press.
Burggraaff, Olivier et al. (2019). "Standardized spectral and radiometric calibration of consumer cameras". In: *Optics Express* 27.14, p. 19075. ISSN: 1539-4794. DOI: 10.1364/oe.27.019075. arXiv: 1906.04155.
Burt, P. J. (1981). "Fast filter transform for image processing". In: *Computer Graphics and Image Processing* 16, pp. 20–51.
Butterworth, S. (1930). "On the Theory of Filter Amplifiers". In: *Experimental Wireless and the Wireless Engineer* 7, pp. 536–541.
Byron, Frederick W. and Robert W. Fuller (1992). *Mathematics of Classical and Quantum Physics*. Mineola, New York: Dover Publications.
Canada, Public Health Agency of (2018). *Report from the Canadian Chronic Disease Surviellance System: Heart Disease in Canada, 2018*. Tech. rep.
Carreiras, Carlos et al. (2015). *BioSPPy: Biosignal Processing in Python*. URL: https://github.com/PIA-Group/BioSPPy/.
Case, Kenneth M. and Paul F. Zweifel (1967). *Linear Transport Theory*. Reading: Addison-Wesley Publishing Company. DOI: 10.13182/nse68-a17608.
Chaji, Sedigheh et al. (2018). "Estimation of the camera spectral sensitivity function using neural learning and architecture". In: *Journal of the Optical Society of America A* 35.6, p. 850. ISSN: 1084-7529. DOI: 10.1364/josaa.35.000850.
Chakrabarti, Ayan, Daniel Scharstein, and Todd Zickler (2009). "An empirical camera model for internet color vision". In: *British Machine Vision Conference, BMVC 2009 - Proceedings*. DOI: 10.5244/C.23.51.
Chakrabarti, Ayan, Ying Xiong, et al. (2013). "Modeling Radiometric Uncertainty for Vision with Tone-mapped Color Images". In: pp. 1–14. DOI: 10.1109/TPAMI.2014.2318713. arXiv: 1311.6887. URL: http://arxiv.org/abs/1311.6887%7B%5C%%7D0Ahttp://dx.doi.org/10.1109/TPAMI.2014.2318713.
Chan, Edward D., Michael M. Chan, and Mallory M. Chan (2013). "Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations". In: *Respiratory Medicine* 107.6, pp. 789–799. ISSN: 09546111. DOI: 10.1016/j.rmed.2013.02.004. URL: http://dx.doi.org/10.1016/j.rmed.2013.02.004.
Charkoudian, Nisha (2003). "Skin Blood Flow in Adult Human Thermoregulation: How It Works, When It Does Not, and Why". In: *Mayo Clinic Proceedings* 78.5, pp. 603–612.
Chen, Weixuan and Daniel McDuff (2018). "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks BT - Computer Vision – ECCV 2018". In: *15th European Conference on Computer Vision*, pp. 356–373.
Cheong, Wai-Fung, Scott A. Prahl, and Ashley J. Welch (1990). "A Review of the Optical Properties of Biological Tissues". In: *IEEE Journal of Quantum Electornics* 26.12, pp. 2166–2185. ISSN: 00189197. DOI: 10.1109/3.64354.

Chew, Derek P., David J. Moliterno, and Howard C. Herrmann (2002). "Present and potential future paradigms for the treatment of ST-segment elevation acute myocardial infarction". In: *Journal of Invasive Cardiology* 14.Supplement A, 3A–20A.

Cicone, Antonio and Hau Tieng Wu (2017). "How nonlinear-type time-frequency analysis can help in sensing instantaneous heart rate and instantaneous respiratory rate from photoplethysmography in a reliable way". In: *Frontiers in Physiology* 8.SEP, pp. 1–17. ISSN: 1664042X. DOI: 10.3389/fphys.2017.00701.

Claridge, Ela, Symon Cotton, et al. (2003). "From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions." In: *Medical image analysis* 7.4, pp. 489–502. ISSN: 1361-8415. URL: http://www.ncbi.nlm.nih.gov/pubmed/14561553.

Claridge, Ela and Steve J. Preece (2003). "An Inverse method for the recovery of tissue parameters from colour images". In: *Information Processing in Medical Imaging* 2732, pp. 306–317.

Cmglee (2019). *Back-side illumination*.

Colberg, Sheri R. et al. (2003). "Cutaneous blood flow in type 2 diabetic individuals after an acute bout of maximal exercise". In: *Diabetes Care* 26.6, pp. 1883–1888. ISSN: 01495992. DOI: 10.2337/diacare.26.6.1883.

Collaborators, GBD 2015 Mortality and Causes of Death (2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015". In: *The Lancet* 388.10053, pp. 1459–1544.

Condon, E. U. and G. H. Shortley (1935). *The Theory of Atomic Spectra*. Cambridge, UK: Cambridge University Press.

Cooper, James K. (1986). "Electrocardiography 100 years ago. Origins, pioneers, and contributers". In: *New England Journal of Medicine* 315.7, pp. 461–464.

Cotton, Symon and Ela Claridge (1996). "Developing a predictive model of human skin colouring". In: *Medical Imaging 1996: Physics of Medical Imaging* 2708, pp. 814–825. DOI: 10.1117/12.237846.

Cunningham, J. P., K. V. Shenoy, and M. Sahani (2008). "Fast Gaussian process methods for point process intensity estimation". In: *International Conference on Machine Learning*.

Dahmani, Djamila, Mehdi Cheref, and Slimane Larabi (2020). "Zero-sum game theory model for segmenting skin regions". In: *Image and Vision Computing* 99, p. 103925. ISSN: 02628856. DOI: 10.1016/j.imavis.2020.103925. URL: https://doi.org/10.1016/j.imavis.2020.103925.

Davis, Sara et al. (2014). "Continuous non-contact vital sign monitoring in neonatal intensive care unit". In: *Healthcare Technology Letters* 1.3, pp. 87–91. ISSN: 2053-3713. DOI: 10.1049/htl.2014.0077.

Dayan, I., S. Havlin, and G. H. Weiss (1992). "Photon migration in in a two-layer turbid medium: a diffusion analysis". In: *Journal of Modern Optics* 39, pp. 1567–1582.

De Bacquer, D. et al. (1998). "Prognostic value of ECG findings for total, cardiovascular disease, and coronary heart disease death in men and women". In: *Heart* 80.6, pp. 570–577.

De Luca, Giuseppe et al. (2004). "Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts". In: *Circulation* 109.10, pp. 1223–1225.

Dedner, Andreas and Peter Vollmoller (2002). "An adaptive higher order method for solving the radiative transport equation on unstructured grids". In: *Journal of Computational Physics* 178.2, pp. 263–289.

Dicarlo, J. M., G. E. Montgomery, and S. W. Trovinger (2004). "Emissive chart for imager calibration". In: *IS\&T/SID 12th Color imaging Conference*, pp. 295–301.

Ding, Huafeng et al. (2006). "Refractive indices of human skin tissues at eight wavelengths and estimated dispersion relations between 300 and 1600 nm". In: *Physics in Medicine and Biology* 51.6, pp. 1479–1489.

Doi, M. and S. Tominaga (2003). "Spectral Estimaton of Human Skin Colour using the Kubelka-Munk Theory". In: *SPIE/IS&T Electronic Imaging, SPIE* 5008, pp. 221–228.

Donner, Craig and Henrik Wann Jensen (2005). "Light diffusion in multi-layered translucent materials". In: *ACM Transactions on Graphics* 24.3, p. 1032. ISSN: 07300301. DOI: 10.1145/1073204.1073308.

Donner, Craig, Tim Weyrich, et al. (2008). "A layered, heterogeneous reflectance model for acquiring and rendering human skin". In: *ACM Transactions on Graphics* 27.5, p. 1. ISSN: 07300301. DOI: 10.1145/1409060.1409093.

Egan, W. G., T. W. Hilgeman, and J. Reichman (1973). "Determinination of of absorption and scattering coefficients for nonhomogeneous media. 2: Experiment". In: *Applied Optics* 12, pp. 1816–1823.

Egger, Herbert and Matthias Schlottbom (2019). "A Perfectly Matched Layer Approach for PN-Approximations in Radiative Transfer". In: *SIAM Journal of Numerical Analysis* 57.5, pp. 2166–2188.

Einthoven, Willem (1903). "Die galvanometrische Registrirung des menschlichen Elektrokardiogramms, zugleich eine Beurtheilung der Anwendung des Capillar-Elektrometers in der Physiologie". In: *Pflugers Arch* 99, pp. 472–480.

Evans, Lawrence C. (2010). *Partial Differential Equations*. 2nd. Providence, Rhode Island: American Mathematical Society.

Fan, Yuwei, Jing An, and Lexing Ying (2019). "Fast algorithms for integral formulations of steady-state radiative transfer equation". In: *Journal of Computational Physics* 380, pp. 191–211. ISSN: 10902716. DOI: 10.1016/j.jcp.2018.12.014. URL: https://doi.org/10.1016/j.jcp.2018.12.014.

Faris, G. W. (2002). "Diffusion equation boundary conditions for the interface between turbid media: a comment". In: *Journal of the Optical Society of America A* 19.519-520.

Finetti, Bruno de (1972). *Probability, Induction, and Statstics*. New York: Wiley.

Finlayson, Graham, Maryam Mohammadzadeh Darrodi, and Michal Mackiewicz (2016). "Rank-based camera spectral sensitivity estimation". In: *Journal of the Optical Society of America A* 33.4, p. 589. ISSN: 1084-7529. DOI: 10.1364/josaa.33.000589.

Fossum, Eric R. (1993). "Active pixel sensors: are CCDs dinosaurs?" In: *SPIE Proceedings Vol. 1900: Charge-Coupled Devices and Solid State Optical Sensors III*. Ed. by Morley M. Blouke. San Jose, CA, United States: International Society for Optics and Photonics, pp. 2–14.

Fossum, Eric R. and Donald B. Hondongwa (2014). "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors". In: *IEEE Journal of Electron Devices Society* 2.3, pp. 33–43.

Frank, Martin et al. (2007). "Time-dependent simplified PN approximation to the equations of radiative transfer". In: *Journal of Computational Physics* 226.2, pp. 2289–2305.

Fye, W. B. (1994). "A history of the origin, evolution, and impact of electrocardiography". In: *American Journal of Cardiology* 73.13, pp. 937–949.

Gao, Xin (2019). "Diagnosing Abnormal Electrocardiogram (ECG) via Deep Learning". In: *Practical Applications of Electrocardiogram*. Ed. by Umashankar Lakshmanadoss. IntechOpen.

Gardner, Jacob R. et al. (2018). "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration". In: NeurIPS. arXiv: 1809.11165. URL: http://arxiv.org/abs/1809.11165.

Godoy, William F and Xu Liu (2012). "Parallel jacobina-free newton krylov solution of the discrete ordinates method with flux limiters for 3d radiative transfer". In: *Journal of Computational Physics* 231.11, pp. 4257–4278.

Golub, Gene G. and Charles F. Van Loan (2013). *Matrix Computations*. 4th Editio. Baltimore: Johns Hopkins University Press, p. 784.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis: Univeristy of Minnesota Press.

Gradshteyn, I. S. and I. M. Ryzhik (1980). *Table of Integrals, Series, and Products*. New York: Academic Press.

Groenhuis, R. A. J., H. A. Ferwerda, and J. J. Ten Bosch (1983). "Scattering and absorption of turbid materials determined from reflection measurements. 1. Theory". In: *Applied Optics* 22, pp. 2456–2462.

Grossberg, M. D. and S.K. Nayar (2004). "Modeling the space of camera response functions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.10, pp. 1272–1282.

Haan, G. de and A. Van Leest (2014). "Improved motion robustness of remote-PPG by using the blood volume pulse signature". In: *Physiological Measurement* 35.9, pp. 1913–1926. ISSN: 13616579. DOI: 10.1088/0967-3334/35/9/1913.

Haan, Gerard de and Vincent Jeanne (2013). "Robust pulse-rate from chrominance-based rPPG". In: *IEEE Transactions on Biomedical Engineering* 60.10, pp. 2878–2886.

Hannun, Awni Y. et al. (2019). "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". In: *Nature Medicine* 25, pp. 65–69.

Hardeberg, J. Y., H. Brettel, and F. Schmitt (1998). "Spectral characterization of electronic Cameras". In: *Electronic Imaging: Processing, Printing, and Publishing in Color* 3409, pp. 100–109.

Harford, M et al. (2019). "Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: a systematic review". In: *Physiological Measurement* 40.6, 06TR01. ISSN: 1361-6579. DOI: 10.1088/1361-6579/ab1f1d.

Harvey, William (1628). *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus*. Frankfurt: William Fitzer, p. 72.

Hassan, M. A. et al. (2020). "Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study". In: *Measurement: Journal of the International Measurement Confederation* 149, p. 106804. ISSN: 02632241. DOI: 10.1016/j.measurement.2019.07.032. URL: https://doi.org/10.1016/j.measurement.2019.07.032.

Hensman, James, N. Fusi, and Neil D. Lawrence (2013). "Gaussian processes for big data". In: *Uncertainty in Artifical Intelligence*.

Herring, Neil and David J. Paterson (2018). *Levick's Introduction to Cardiovascular Physiology*. Boca Raton, FL: Taylor & Francis. URL: http://www.cvphysiology.com/%7B%5C%%7D5Cnhttp://cvphysiology.com/Blood%20Pressure/BP008.htm.

Hertzman, Alrick B. (1938). "The blood supply of various skin areas as estimated by the photoelectric plethysmograph". In: *Proceedings of the Society of Experimental Biology and Medicine* 38.2, pp. 328–340.

Hertzman, Alrick B. and Clair R. Spealman (1937). "Observations on the finger volume pulse recorded photo-electrically". In: *American Journal of Physiology* 119.3, pp. 334–335.

Hielscher, Andreas H. et al. (1995). "The influence of boundary conditions on the accuracy of diffusion theory in time-resolved reflectance spectroscopy of biological tissues". In: *Physics in Medicine and Biology* 40, pp. 1957–1975.

Hu, Sijung et al. (2009). "Development of effective photoplethysmographic measurement techniques: From contact to non-contact and from point to imaging". In: *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pp. 6550–6553. DOI: 10.1109/IEMBS.2009.5334505.

Hubel, P. M., D. Sherman, and J. E. Farrell (1994). "A comparison of methods of sensor spectral sensitivity estimation". In: *IS\&T/SID 2nd Color imaging Conference*, pp. 45–48.

Huynh, Loc et al. (2018). "Mesoscopic Facial Geometry Inference Using Deep Neural Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City.

Hwang, Kun, Kum Hun, and Dae Joong Kim (2015). "Thickness of skin and subcutaneous tissue of the free flap donor sites: A histologic study". In: *Microsurgery* 36, pp. 54–58.

Ilan, Boaz and Arnold D. Kim (2019). "Radiative transfer of light in strongly-scattering media". In: *Light Scattering Reviews, vol 3: Radiative Transfer and Light Scattering*. Ed. by Alexander A. Kokhanovsky. Springer.

Incorporated, Adobe Systems (2005). *Adobe RGB (1998) Color Image Encoding*. Tech. rep.

Institute, National Cancer (2008). *Anatomy of the Skin*.

Ishimaru, Akira (1978). *Wave Propagation and Scattering in Random Media*. London: Academic Press. ISBN: 0123747023.

Jacques, Steven L. (2013). "Optical properties of biological tissues: A review". In: *Physics in Medicine and Biology* 58.11. ISSN: 00319155. DOI: 10.1088/0031-9155/58/11/R37.

– (1998). *Skin optics summary*. Tech. rep. Oregon Medical Laser Center News.

Jacques, Steven L., Scott A. Prahl, and J. Lindsey (2001). *Optical properties spectra*. URL: http://omlc.ogi.edu/spectra.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge, United Kingdom: Cambridge University Press.

Jeffreys, Harold (1961). *Theory of Probability*. 3rd. London: Oxford University Press.

Jiang, Jun et al. (2013). "What is the space of spectral sensitivity functions for digital color cameras?" In: *Proceedings of IEEE Workshop on Applications of Computer Vision*, pp. 168–179. ISSN: 21583978. DOI: 10.1109/WACV.2013.6475015.

Judd, D. B. (1964). "Spectral distribution of typical daylight as a function of correlated color temperature". In: *Journal of the Optical Society of America A* 54, pp. 1031–1040.

Karatzas, Ioannis and Steven E. Shreve (1988). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag New York, Inc. ISBN: 9781468403046.

Kaya, Berk, Yigit Baran Can, and Radu Timofte (2018). "Towards Spectral Estimation from a Single RGB Image in the Wild". In: arXiv: 1812.00805. URL: http://arxiv.org/abs/1812.00805.

Kazemi, Vahid and Josephine Sullivan (2014). "One millisecond face alignment with an ensemble of regression trees". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Keener, James and James Sneyd (2009). *Mathematical Physiology II: Systems Physiology*. Second. New York, New York: Springer Science+Business Media. ISBN: 9780387793870.

Keijzer, Marleen, Willem M. Star, and Pascal R. M. Storchi (1988). "Optical diffusion in layered media". In: *Applied Optics* 27.9, p. 1820. ISSN: 0003-6935. DOI: 10.1364/ao.27.001820.

Kim, Arnold D. and Joseph B. Keller (2003). "Light propagation in biological tissue". In: *Journal of the Optical Society of America A* 20.1, pp. 92–98. ISSN: 00224073. DOI: 10.1016/j.jqsrt.2018.11.001.

Kim, Seon Joo et al. (2012). "A new in-camera imaging model for color computer vision and its application". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.12, pp. 2289–2302. ISSN: 01628828. DOI: 10.1109/TPAMI.2012.58.

Kingma, Diederik P. and Jimmy Lei Ba (2015). "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15. arXiv: 1412.6980.

Klaessens, John H. et al. (2014). "Development of a baby friendly non-contact method for measuring vital signs: First results of clinical measurements in an open incubator at a neonatal intensive care unit". In: *Advanced Biomedical and Clinical Diagnostic Systems XII* 8935.February 2014, 89351P. ISSN: 16057422. DOI: 10.1117/12.2038353.

Klose, Alexander D. and Edward W. Larsen (2006). "Light transport in biological tissue based on the simplified spherical harmonics equations". In: *Journal of Computational Physics* 220.1, pp. 441–470. ISSN: 10902716. DOI: 10.1016/j.jcp.2006.07.007.

Klug, Brian and Anand Lal Shimpi (2010). "Apple's iPhone 4: Thoroughly Reviewed". In: *AnandTech*.

Kranjec, Jure et al. (2014). "Novel methods for noncontact heart rate measurement: A feasibility study". In: *IEEE Transactions on Instrumentation and Measurement* 63.4, pp. 838–847. ISSN: 00189456. DOI: 10.1109/TIM.2013.2287118.

Krishnaswamy, Aravind and Gladimir V G Baranoski (2004). "A Study on Skin Optics". In: *Technical Report CS200401*, pp. 1–17. ISSN: 0167-7055. DOI: 10.1111/j.1467-8659.2004.00764.x. URL: http://www.imt.liu.se/edu/courses/TBMT36/pdf/Text4.pdf.

Kubelka, Paul (1948). "New contributions to the optics of intensely light-scattering materials." In: *Journal of the Optical Society of America* 38.5, pp. 448–457. ISSN: 00303941. DOI: 10.1364/JOSA.38.000448.

– (1954). "New Contributions to the Optics of Intensely Light-Scattering Materials. Part II: Non-homogeneous layers". In: *Journal of the Optical Society of America* 44, pp. 330–335.

Lambert, Johann Heinrich (1760). *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Augsburg, Germany: Eberhardt Klett.

Le, Q., T. Sarlos, and Alex J Smola (2013). "Fastfood-computing Hilbert space expansions in loglinear time". In: *Proceedings of the 30th International Conference on Machine Learning*, pp. 244–252.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning". In: *Nature* 521, pp. 436–444.

Lestari, Trisasi, Sylvia Ryll, and Axel Kramer (2013). "Microbial contamination of manually reprocessed, ready to use ECG lead wire in intensive care units." In: *GMS hygiene and infection control* 8.1, Doc07. ISSN: 2196-5226. DOI: 10.3205/dgkh000207.

Lewandowska, Magdalena et al. (2011). "Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity". In: *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 405–410.

Li, Hui, Xixi Jia, and Lei Zhang (2018). "Clustering based content and color adaptive tone mapping". In: *Computer Vision and Image Understanding* 168, pp. 37–49.

Liemert, André and Alwin Kienle (2010). "Light diffusion in N-layered turbid media: steady-state domain". In: *Journal of Biomedical Optics* 15.2, p. 025003. ISSN: 1083-3668. DOI: 10.1117/1.3368685.

Lister, Tom, Philip A. Wright, and Paul H. Chappell (2012). "Optical properties of human skin". In: *Journal of Biomedical Optics* 17.9, p. 0909011. ISSN: 1083-3668. DOI: 10.1117/1.jbo.17.9.090901.

Luo, Hong et al. (2019). "Smartphone-Based Blood Pressure Measurement Using Transdermal Optical Imaging Technology". In: *Circulation: Cardiovascular Imaging* 12.8, pp. 1–10. ISSN: 1941-9651. DOI: 10.1161/CIRCIMAGING.119.008857. URL: https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.119.008857.

Maceira, Alicia M. et al. (2006). "Reference right ventricular systolic and diastolic function normalized to age, gender and body surface area from steady-state free precession cardiovascular magnetic resonance". In: *European Heart Journal* 27, pp. 2879–2888.

MacKay, David J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press. ISBN: 9780521642989. DOI: 10.1198/jasa.2005.s54. arXiv: arXiv:1011.1669v3. URL: http://www.tandfonline.com/doi/abs/10.1198/jasa.2005.s54.

Marchuck, G. I. and V. I. Lebedev (1986). *Numerical Methods in the Theory of Neutron Transport*. Chur, Switzerland: Harwood Academic Publishers.

Martelli, Fabrizio et al. (2010). *Light Propagation through Biological Tissue*. Bellingham, Washington: SPIE. ISBN: 9780819476586.

Matern, Bertil (1960). *Spatial Variation*. New York: Spring-Verlag New York.

Mauer, C. and D. Wueller (2009). "Measuring the spectral response with a set of interference filters". In: *IS\&T/SPIE Electronic Imaging*. International Society for Optics and Photonics.

McCamy, C.S., H. Marcus, and J.G. Davidson (1976). "A Color Rendition Chart". In: *Journal of Applied Photographic Engineering* 11.3, pp. 95–99.

McGrath, John A. and Jouni Uitto (2016). "Structure and Function of the Skin". In: *Rook's Textbook of Dermatology*. Ed. by Christopher E. M. Griffiths et al. Ninth Edit. Oxford, UK: Wiley Blackwell. Chap. 2, pp. 2.1–2.48.

Meglinksi, I. V. and S. J. Matcher (2002). "Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions". In: *Physiological Measurement* 23.4, pp. 741–753.

Meischke, Hendrika et al. (1995). "Reasons Patients With Chest Pain Delay or Do Not Call 911". In: *Annals of Emergency Medicine* 25.2, pp. 193–197.

Mishchenko, Michael I. (2006). "Maxwell's equations, radiative transfer, and coherent backscattering: A general perspective". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 101.3, pp. 540–555. ISSN: 00224073. DOI: 10.1016/j.jqsrt.2006.02.065.

Mohrman, David E. and Lois Jane Heller (2018). *Cardiovascular Physiology*. London: McGraw Hill Education.

Moore, E. H. (1920). "On the reciprocal of the general algebraic matrix". In: *Bulletin of the American Mathematical Society* 26.9, pp. 394–395.

Moraes, Jermana L. et al. (2018). "Advances in photopletysmography signal analysis for biomedical applications". In: *Sensors (Switzerland)* 18.6, pp. 1–26. ISSN: 14248220. DOI: 10.3390/s18061894.

Murphy, Kevin P. (2012). *Machine Learning A Probabilistic Perspective*. 320321, pp. 1–100. ISBN: 0070428077. DOI: 10.1007/SpringerReference_302149. arXiv: 0-387-31073-8.

Murray, I. (2009). "Gaussian processes and fast matrix-vector multiplies". In: *ICML Workshop on Numerical Mathematics in Machine Learning*.

Nakamura, J. (2006). *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL: CRC Press.

Nobbs, James H (1985). "Kubelka- Munk Theory and the Prediction of Reflectance". In: *Review of Progress in Coloration and Related Topics* 15, pp. 66–75.

O'Dea, S. (2020). *Number of smartphone users worldwide from 2016 to 2021*. Tech. rep. Statista.

Ockendon, John et al. (2003). *Applied Partial Differential Equations*. Revised Ed. Oxford, UK: Oxford University Press.

Organization, World Health (2017). *Cardiovascular diseases (CVDs)*. Tech. rep. Geneva.

Paszke, A. et al. (2017). "Automatic differentiation in PyTorch".

Penrose, Roger (1955). "A generalized inverse for matrices". In: *Proceedings of the Cambridge Philosophical Society* 51.3, pp. 406–413.

Peris, Vicente Azorin and Sijung Hu (2007). "Validation of a Monte Carlo platform for the optical modelling of pulse oximetry". In: *Journal of Physics: Conference Series* 85.1. ISSN: 17426596. DOI: 10.1088/1742-6596/85/1/012027.

Petersen, Kaare Breandt and Michael Syskind Pedersen (2007). "The Matrix Cookbook". In: *Citeseer* 16.4, pp. 1–66. ISSN: 09621083. DOI: 10.1111/j.1365-294X.2006.03161.x. arXiv: 0608522 [math].

Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard (2010). "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation". In: *Optics Express* 18.10, p. 10762. DOI: 10.1364/oe.18.010762.

Prahl, Scott A. (1999). *Optical absorption of hemoglobin*. Tech. rep. Oregon Medical Laser Center News.

Prasad, Dilip K., Rang Nguyen, and Michael S. Brown (2013). "Quick approximation of camera's spectral response from casual lighting". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 844–851. DOI: 10.1109/ICCVW.2013.116.

Preece, Stephen J. and Ela Claridge (2004). "Spectral filter optimization for the recovery of parameters which describe human skin". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.7, pp. 913–922. ISSN: 01628828. DOI: 10.1109/TPAMI.2004.36.

Qiu, Jueqin and Haisong Xu (2016). "Camera response prediction for various capture settings using the spectral sensitivity and crosstalk model". In: *Applied Optics* 55.25, p. 6989. ISSN: 0003-6935. DOI: 10.1364/ao.55.006989.

Quinonero-Candela, J. and Carl E. Rasmussen (2005). "A unifying view of sparse approximate gaussian process regression". In: *Journal of Machine Learning Research* 6, pp. 1939–1955.

Rahimi, A. and B. Recht (2007). "Random features for large-scale kernel machines". In: *Neural Information Processing Systems (NIPS)2*.

Rang, Nguyen Ho Man, Dilip K. Prasad, and Michael S. Brown (2014). "Raw-to-raw: Mapping between image sensor color responses". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3398–3405. ISSN: 10636919. DOI: 10.1109/CVPR.2014.434.

Rasmussen, Carl E. and Christopher K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts, p. 266. ISBN: 026218253X. DOI: 10.1142/S0129065704001899. arXiv: 026218253X. URL: http://www.gaussianprocess.org/gpml/chapters/RW.pdf.

Rasmussen, Carl Edward and Zoubin Ghahramani (2001). "Occam's Razor". In: *Neural Information Processing Systems (NIPS)*. DOI: 10.1016/S0140-6736(60)91387-8.

Resiner, Andrew et al. (2008). "Utility of the Photoplethysmogram in Circulatory Monitoring". In: *Anesthesiology* 108.5, pp. 950–958. URL: http://www.ncbi.nlm.nih.gov/pubmed/18431132.

Revuz, Daniel and Marc Yor (1991). *Continuous Martingales and Brownian Motion*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. ISBN: 9781461385868.

Ribeiro, Antonio H. et al. (2020). "Automatic diagnosis of the 12-lead ECG using a deep neural network". In: *Nature Communications* 11, p. 1760.

Ripoll, J. and M. Nieto-Vesperinas (1999). "Index mismatch for diffuse photon density waves at both flat and rough diffuse-diffuse itnerfaces". In: *Journal of the Optical Society of America A* 16.1947-1957.

Robert, C. and G. Casella (2004). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Rowell, Loring B (1990). "Editorial Comment Hyperthermia: A Hyperadrenergic State". In: *Hypertension* 15, pp. 505–507.

Russell, Lisa M, Sandra Wiedersberg, and M Begoña Delgado-charro (2009). "The determination of stratum corneum thickness - an alternative approach". In: *European Journal of Pharmaceutics and Biopharmaceutics* 69.3, pp. 861–870.

Ryde, John Walter (1931). "The scattering of light by turbid media.-Part I". In: *Proceedings of the Royal Society A* 131.817.

Saito, Shunsuke et al. (2017). "Photorealistic Facial Texture Inference Using Deep Neural Networks". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Salimbeni, Hugh and Marc Peter Deisenroth (2017). "Doubly stochastic variational inference for deep Gaussian processes". In: *Advances in Neural Information Processing Systems* 2017-Decem.Nips, pp. 4589–4600. ISSN: 10495258. arXiv: 1705.08933.

Salomatina, Elena Vladimirovna et al. (2006). "Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range". In: *Journal of Biomedical Optics* 11.6, p. 064026.

Sana, Furrukh et al. (2020). "Wearable Devices for Ambulatory Cardiac Monitoring: JACC State-of-the-Art Review". In: *Journal of the American College of Cardiology* 75.13, pp. 1582–1592. ISSN: 15583597. DOI: 10.1016/j.jacc.2020.01.046.

Sato, Y. et al. (1998). "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images". In: *Medical Image Analysis1* 2.2, pp. 143–168.

Scholkmann, Felix, Jens Boss, and Martin Wolf (2012). "An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals". In: *Algorithms* 5.4, pp. 588–603.

Schölkopf, Bernhard and Alex J Smola (2002). *Learning with kernels*. ISBN: 0262194759.

Shafer, G. (1981). "Constructive Probability". In: *Synthese* 48, pp. 1–60.

Sharma, G. (2003). *Digital Color Imaging Handbook*. Boca Raton, FL: CRC Press.

Shi, P. et al. (2009). "Insight into the dichrotic notch in photoplethysmographic pulses from the finger tip of young adults". In: *Journal of Medical Engineering and Technology* 33.8, pp. 628–633.

Silverman, Bernhard W. (1985). "Some aspects of the spline smoothing approach to non-parametric regrssion curve fitting". In: *Journal of the Royal Statistical Society B* 47.1, pp. 1–52.

Sinex, J. E. (1999). "Pulse oximetry: principles and limitations". In: *American Journal of Emergency Medicine* 17, pp. 59–67.

Slater, D. and G. Healey (1998). "What is the spectral dimensionality of illumination function in ourdoor scenes?" In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Snelson, Edward and Zoubin Ghahramani (2006). "Sparse Gaussian processes using pseudo-inputs". In: *Advances in Neural Information Processing Systems*.

Soleimani, Hossein, James Hensman, and Suchi Saria (2018). "Scalable Joint Models for Reliable Uncertainty-Aware Event Prediction". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.8, pp. 1948–1963. ISSN: 01628828. DOI: 10.1109/TPAMI.2017.2742504. arXiv: 1708.04757.

Standard Method of Assessing the Spectral Quality of Daylight Simulators for Visual Appraisal and Measurement of Colour (2004). Tech. rep. CIE Standard, S012/E:2004.

Stein, Michael L. (1999). *Interpolation of Spatial Data: some theory for kriging*. New York: Springer-Verlag.

Stokes, George Gabriel (1862). "IV. On the intensity of light reflected from or transmitted through a pile of plates". In: *Proceedings of the Royal Society of London* 11.

Sun, Yu and Nitish Thakor (2016). "Photoplethysmography Revisited: From Contact to Noncontact, from Point to Imaging". In: *IEEE Transactions on Biomedical Engineering* 63.3, pp. 463–477. ISSN: 15582531. DOI: 10.1109/TBME.2015.2476337.

Tamura, Toshiyo et al. (2014). "Wearable photoplethysmographic sensors—past and present". In: *Electronics* 3.2, pp. 282–302. ISSN: 20799292. DOI: 10.3390/electronics3020282.

Thiriet, Marc (2008a). *Biology and Mechanics of Blood Flows Part I: Biology*. New York, New York: Spring-Verlag New York.

– (2008b). *Biology and Mechanics of Blood Flows Part II: Mechanics and Medical Aspects*. New York, NY: Springer Science+Business Media. ISBN: 9780387748481.

Tian, Xin et al. (2020). "Cross-Domain Joint Dictionary Learning for ECG Reconstruction from PPG". In: pp. 936–940.

Titsias, Michael K. (2009). "Variational learning of inducing variables in sparse gaussian processes". In: *AISTATS*, pp. 567–574.

Topol, Eric (2019). *Deep Medicine*. New York, NY: Basic Books, Hachette Book Group.

Tsumura, Norimichi, Hideaki Haneishi, and Y. Miyake (1999). "Independent Component Analysis of Skin Color Image". In: *Journal of the Optical Society of America A* 16.9, pp. 2169–2176.

– (2000). "Independent Component Analysis of Spectral Absorbance Images in Human Skin". In: *Optical Reviews* 7.6, pp. 479–482.

Tuchin, Valery (2000). *Tissue Optics*. Bellingham, Washington: SPIE Press.

Uhlenbeck, G. E. and L. S. Ornstein (1930). "On the Theory of Brownian Motion". In: *Physical Review* 36, pp. 823–841.

Urban, P. et al. (2010). "Recovering camera sensitivities using target-based reflectances captured under multiple LED-illuminations". In: *16th Workshop on Color Image Processing*, pp. 9–16.

Verkruysse, Wim, Lars O Svaasand, and J Stuart Nelson (2008). "Remote plethysmographic imaging using ambient light". In: *Optics Express* 16.26, pp. 21434–21445. DOI: 10.1038/mp.2011.182.doi.

Viola, Paul and Michael J. Jones (2004). "Robust Real-Time Face Detection". In: *International Journal of Computer Vision* 57.2, pp. 137–154.

Waller, Augustus Waller (1887). "A demonstration on man of electromotive charges accompanying the heart's beat". In: *Journal of Physiology* 8, pp. 229–234.

Walsh, Joseph A., Eric J. Topol, and Steven R. Steinhubl (2014). "Novel wireless devices for cardiac monitoring". In: *Circulation* 130.7, pp. 573–581. ISSN: 15244539. DOI: 10.1161/CIRCULATIONAHA.114.009024.

Wang, Edward Jay et al. (2017). "HemaApp". In: *GetMobile: Mobile Computing and Communications* 21.2, pp. 26–30. ISSN: 2375-0529. DOI: 10.1145/3131214.3131223.

Wang, L., S. L. Jacques, and L. Zheng (1995). "MCML – Monte Carlo modeling of light transport in multi-layered tissues". In: *Computer Methods and Programs in Biomedicine* 47.2, pp. 131–146.

Wang, Lihong V. and Hsin-i Wu (2007). *Bio-Medical Optics*. Hoboken, New Jersey: John Wiley & Sons. ISBN: 9780471743040.

Wang, Wenjin et al. (2017). "Algorithmic Principles of Remote PPG". In: *IEEE Transactions on Biomedical Engineering* 64.7, pp. 1479–1491. ISSN: 15582531. DOI: 10.1109/TBME.2016.2609282.

Wang, Z. X. and D. R. Guo (1989). *Special Functions*. Singapore: World Scientific Publishing Company.

Wilson, Andrew Gordon, Christoph Dann, and Hannes Nickisch (2015). "Thoughts on Massively Scalable Gaussian Processes". In: pp. 1–25. arXiv: 1511.01870. URL: http://arxiv.org/abs/1511.01870.

Wilson, Andrew Gordon and Hannes Nickisch (2015). "Kernel interpolation for scalable structured Gaussian processes (KISS-GP)". In: *32nd International Conference on Machine Learning, ICML 2015* 3, pp. 1775–1784. arXiv: 1503.01057.

Wyszecki, Gunter and W. S. Stiles (1982). *Color Science: Concepts and Methods Quantitative Data and Formulae*. New York, NY: Wiley.

Yu, Zitong et al. (2019). "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement". In: *Proceedings of the IEEE International Conference on Computer Vision* 2019-Octob, pp. 151–160. ISSN: 15505499. DOI: 10.1109/ICCV.2019.00024.

Zaunseder, Sebastian et al. (2018). "Cardiovascular assessment by imaging photoplethysmography – a review". In: *Biomed. Eng.-Biomed. Tech.* 63.5, pp. 617–634.

Zhao, H. et al. (2009). "Estimating basis functions for spectral sensitivity of digital cameras". In: *Meeting on Image Recognition an dUnderstanding (MIRU)*.

Zoller, Christian and Alwin Kienle (2019). "Fast and precise image generation of blood vessels embedded in skin". In: *Journal of Biomedical Optics* 24.01, p. 1. DOI: 10.1117/1.jbo.24.1.015002.

Zoller, Christian Johannes et al. (2018). "Parallelized Monte Carlo software to efficiently simulate the light propagation in arbitrarily shaped objects and aligned scattering media". In: *Journal of Biomedical Optics* 23.06, p. 1. DOI: 10.1117/1.jbo.23.6.065004.

Zuiderveld, K. (1994). "Contrast Limited Adaptive Histogram Equalization". In: *Graphics Gems IV*. Ed. by P. Heckbert. Academic Press.

Appendix A

Spherical Harmonics

This appendix provides an overview of the mathematical properties of Legendre polynomials, associated Legendre polynomials, and spherical harmonics relevant to derivations in the main text of the dissertation.

Traditionally, spherical harmonics arise as solutions of the angular portion of Laplace's equation $\nabla^2 f = 0$ for scalar-values $f : \mathbb{R}^n \to \mathbb{R}$, $n \geq 1$ in spherical coordinates (r, θ, ϕ),

$$\nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r^2 \frac{\partial f}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial f}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 f}{\partial \phi^2} = 0 \tag{A.0.1}$$

We will focus on spherical harmonics as both the solutions of the angular portion of Laplace's equation, as well as a complete set of orthogonal functions on the unit sphere, forming an orthonormal basis of functions on the sphere.

A.1 Legendre Polynomials

The Legendre polynomials may be defined in several ways: *(i)* as series solutions of the Legendre differential equation, or *(ii)* more generally as a complete orthonormal set of polynomials on $[-1, 1]$. The Legendre polynoimals as a complete orthonormal set on $[-1, 1]$ are determined by applying Weierstrass' theorem for an arbitrary continuous function on $[-1, 1]$ with the polynomial basis $\{x^n\}_{n=0}^{\infty}$ and subsequently the Gram-Schmidt orthogonalization process. A complete derivation can be found in Byron and Fuller, 1992.

To explicitly form the Legendre polynomials, we solve the Legendre equation

$$\frac{d}{dx}\left((1 - x^2)\frac{dy}{dx}\right) + l(l+1)y = 0 \tag{A.1.1}$$

which has regular singular points at $x = \pm 1$. Taking a series expansion about the ordinary point $x = 0$, the solution takes the form,

$$y = \sum_{k=0}^{\infty} a_k x^k \tag{A.1.2}$$

for which we determine recurrence relation,

$$a_{k+2} = -\frac{(l-k)(l+k+1)}{(l+2)(l+1)} a_k. \tag{A.1.3}$$

We thus have two linearly indepenent solutions of (A.1.1),

$$y_1 = a_0 \, {}_2F_1\left(-\frac{l}{2}, \frac{l+1}{2}; \frac{1}{2}; x^2\right) \tag{A.1.4}$$

$$y_2 = a_1 \, {}_2F_1\left(-\frac{1-l}{2}, \frac{2+l}{2}; \frac{3}{2}; x^2\right) \tag{A.1.5}$$

for hypergeometric function ${}_2F_1(a, b; c; x)$,

$${}_2F_1(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} x^n \tag{A.1.6}$$

where $(a)_n$ is the rising Pochhammer symbol,

$$(a)_n = \begin{cases} 1, & n = 0 \\ \frac{\Gamma(a+n)}{\Gamma(a)}, & n \geq 1. \end{cases} \tag{A.1.7}$$

The Legendre polynomials are denoted as $P_l(x)$ for $l \geq 0$. To more easily determine the Legendre polynomials for use in spherical harmonics, we use the general formula for the Legendre polynomials given by *Rodrigues' formula* Z. X. Wang and Guo, 1989,

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l. \tag{A.1.8}$$

The Legendre polynomials satisfy a number of useful recurrence relations. We begin with *Bonnet's recursion formula*, with $P_0(x) = 1$, $P_1(x) = x$,

$$(2l+1) x P_l(x) = (l+1) P_{l+1}(x) + l P_{l-1}(x). \tag{A.1.9}$$

Additionally, we have the following recurrence relations involving derivative terms,

$$x P_l'(x) = P_{l+1}'(x) - (l+1) P_l(x) \tag{A.1.10}$$

$$l P_l(x) = x P_l'(x) - P_{l-1}'(x) \tag{A.1.11}$$

$$(2l+1) P_l(x) = P_{l+1}'(x) - P_{l-1}'(x) \tag{A.1.12}$$

$$(x^2 - 1) P_l'(x) = l x P_l(x) - l P_{l-1}(x). \tag{A.1.13}$$

Finally, we state the Legendre polynomial orthogonality result,

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \frac{2}{2n+1}\delta_{mn} \tag{A.1.14}$$

for Kronecker delta δ_{mn}.

A.2 Associated Legendre Polynomials

The associated Legendre polynomials are the solutions of the associated Legendre equation,

$$\frac{d}{dx}\left((1-x^2)\frac{dy}{dx}\right) + \left(l(l+1) - \frac{m^2}{1-x^2}\right)y = 0 \tag{A.2.1}$$

for $l = 0, 1, 2, \ldots$, and $m \in \mathbb{Z}$. The associated Legendre polynomials are given by,

$$P_l^m(x) = (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x). \tag{A.2.2}$$

By Rodrigues' formula (A.1.8) have,

$$P_l^m(x) = (-1)^m\frac{(1-x^2)^{m/2}}{2^l l!}\frac{d^{l+m}}{dx^{l+m}}\left(x^2-1\right)^l. \tag{A.2.3}$$

Further, the associated Legendre differential equation depends on m^2 and there is thus no sign dependence on m. As such, $P_l^m(x)$ and $P_l^{-m}(x)$ are equivalent solutions and will be proportional to each other. It follows immediately from (A.2.3) and the Liebniz differentiation formula,

$$\frac{d^n}{dx^n}(fg) = \sum_{k=0}^{n}\binom{n}{k}f^{(n-k)}g^{(k)} \tag{A.2.4}$$

that,

$$P_l^{-m}(x) = (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x). \tag{A.2.5}$$

The associated Legendre polynomials have two orthogonality results,

$$\int_{-1}^{1} P_l^m(x)P_{l'}^m(x)\,dx = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ll'} \tag{A.2.6}$$

$$\int_{-1}^{1}\frac{P_l^m(x)P_l^{m'}(x)}{1-x^2}\,dx = \frac{1}{m}\frac{(l+m)!}{(l-m)!}\delta_{mm'}. \tag{A.2.7}$$

In addition, the associated Legendre polynomials $\{P_l^m(x)\}$ for a definite m with $l \geq m$ for a complete orthonormal set in $[-1,1]$. Any continuous function f on $[-1,1]$ such that $f(-1) = f(1) = 0$ has associated Legendre polynomial expansion, converging in the mean, of arbitrary order m,

$$f(x) = \sum_{l \geq m} a_l P_l^m(x) \tag{A.2.8}$$

where

$$a_l = \frac{2l+1}{2} \frac{(l-m)!}{(l+m)!} \int_{-1}^{1} f(x) P_l^m(x)\, dx. \tag{A.2.9}$$

The associated Legendre polynomials also satisfy a number of recurrence relations.

$$(2l+1)xP_l^m = (l+m)P_{l-1}^m + (l-m+1)P_{l+1}^m \tag{A.2.10}$$

$$(2l+1)(1-x^2)^{1/2}P_l^m = P_{l-1}^{m+1} + P_{l+1}^{m+1} \tag{A.2.11}$$

$$(2l+1)(1-x^2)^{1/2}P_l^m = (l-m+2)(l-m+1)P_{l+1}^{m-1} - (l+m)(l+m-1)P_{l-1}^{m-1} \tag{A.2.12}$$

Proof. We prove (A.2.10). From (A.1.10), we have,

$$(2l+1)xP_l(x) = (l+1)P_{l+1}(x) + lP_{l-1}(x)$$

Differentiating m times by the Liebniz differentiation formula, we have,

$$(2l+1)xP_l^{(m)}(x) + m(2l+1)P_{l-1}^{(m-1)}(x) = (l+1)P_{l-1}^{(m)}(x) + lP_{l-1}^{(m)}(x)$$

By (A.1.12), we have $(2l+1)mP_{l-1}^{(m-1)}(x) = m\left(P_{l+1}^{(m)}(x) - P_{l-1}^{(m)}(x)\right)$, which gives,

$$(2l+1)xP_l^{(m)}(x) + m\left(P_{l+1}^{(m)}(x) - P_{l-1}^{(m)}(x)\right) = (l+1)P_{l+1}^{(m)}(x) + lP_{l-1}^{(m)}$$

$$(2l+1)xP_l^{(m)}(x) = (l-m+1)P_{l+1}^{(m)}(x) + (l+m)P_{l-1}^{(m)}(x)$$

By definition of the associated Legendre polynomials (A.2.2), we recover (A.1.10).

The remainder of the recurrence relations are proved in a similar manner using the recurrence relations for the Legendre polynomials and repeated differentiation where appropriate. □

A.3 Spherical Harmonics

Spherical harmonics are defined as, for unit vector $\hat{\boldsymbol{s}} = (\theta, \phi)$,

$$Y_{l,m}(\hat{\boldsymbol{s}}) = Y_{l,m}(\theta,\phi) = (-1)^m \left(\left(\frac{2l+1}{4\pi}\right)\frac{(l-m)!}{(l+m)!}\right)^{1/2} P_l^m(\cos\theta)e^{im\phi} \tag{A.3.1}$$

for $l \in \mathbb{N} \cup \{0\}$, $m = -l, -l+1, \ldots, l-1, l$. We include the $(-1)^m$ phase factor of Condon and Shortley, 1935 for convention.

By (A.2.3) we have,

$$Y_{l,-m}(\theta, \phi) = (-1)^m Y^*_{l,m}(\theta, \phi). \tag{A.3.2}$$

The spherical harmonics satisfy the orthogonality property,

$$\int_{\mathcal{S}^{n-1}} Y^*_{l,m}(\hat{\boldsymbol{s}}) Y_{l',m'}(\hat{\boldsymbol{s}})\, d\hat{\boldsymbol{s}} = \delta_{ll'}\delta_{mm'}, \tag{A.3.3}$$

where integration occurs over the solid angle $d\hat{\boldsymbol{s}} = d\Omega = \sin\theta d\theta\, d\phi$.

By completeness, the spherical harmonics are linearly independent and provides that any well-behaved function f of $\hat{\boldsymbol{s}} = (\theta, \phi)$ may be written as a *Laplace series*,

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{l,m}(\theta, \phi). \tag{A.3.4}$$

The spherical harmonics satisfy the famous addition theorem, for any two unit vectors $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{b}}$

$$P_l(\hat{\boldsymbol{a}} \cdot \hat{\boldsymbol{b}}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y^*_{l,m}(\hat{\boldsymbol{a}}) Y_{l,m}(\hat{\boldsymbol{b}}). \tag{A.3.5}$$

More explicitly, for angle γ between $\hat{\boldsymbol{a}} = (\theta, \phi)$ and $\hat{\boldsymbol{b}} = (\theta', \phi')$, we have,

$$\cos\gamma = \hat{\boldsymbol{a}} \cdot \hat{\boldsymbol{b}} = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi') \tag{A.3.6}$$

which allows us to write the addition theorem as,

$$P_l(\cos\gamma) = \sum_{m=-l}^{l} (-1)^m P_l^m(\cos\theta) P_l^{-m}(\cos\theta') e^{im(\phi-\phi')} \tag{A.3.7}$$

$$= P_l(\cos\theta) P_l(\cos\theta') + 2\sum_{m=-l}^{l} \frac{(l-m)!}{(l+m)!} P_l^m(\cos\theta) P_l^m(\cos\theta') \cos m(\phi - \phi'). \tag{A.3.8}$$

Finally we list several important recurrence relations for spherical harmonics,

$$\cos\theta Y_{l,m} = \left(\frac{(l+m)(l-m)}{(2l-1)(2l+1)}\right)^{1/2} Y_{l-1,m} + \left(\frac{(l+m+1)(l-m+1)}{(2l+1)(2l+3)}\right)^{1/2} Y_{l+1,m} \tag{A.3.9}$$

$$\sin\theta e^{i\theta} Y_{l,m} = \left(\frac{(l-m)(l-m-1)}{(2l-1)(2l+1)}\right)^{1/2} Y_{l-1,m+1} + \left(\frac{(l+m+1)(l+m+2)}{(2l+1)(2l+3)}\right)^{1/2} Y_{l+1,m+1} \tag{A.3.10}$$

$$\sin\theta e^{-i\theta} Y_{l,m} = -\left(\frac{(l+m)(l+m-1)}{(2l-1)(2l+1)}\right)^{1/2} Y_{l-1,m-1} + \left(\frac{(l-m+1)(l-m+2)}{(2l+1)(2l+3)}\right)^{1/2} Y_{l+1,m-1} \tag{A.3.11}$$

Proof. We prove (A.3.9) and (A.3.10); (A.3.11) is similar to (A.3.10). By definition, we have,

$$Y_{l-1,m} = \left(\frac{(2l-1)(l-m-1)!}{4\pi(l+m-1)!}\right)^{1/2} P_{l-1}^m(\cos\theta)e^{im\phi} \tag{A.3.12}$$

$$Y_{l+1,m} = \left(\frac{(2l+3)(l-m+1)!}{4\pi(l+m+1)!}\right)^{1/2} P_{l+1}^m(\cos\theta)e^{im\phi}. \tag{A.3.13}$$

By (A.2.10), $(2l+1)\cos\theta P_l^m(\cos\theta) = (l+m)P_{l-1}^m(\cos\theta) + (l-m+1)P_{l+1}^m(\cos\theta)$. Thus,

$$\cos\theta Y_{l,m} = \left(\frac{2l+1)(l-m)!}{4\pi(l+m)!}\right)^{1/2} \left(\left(\frac{l+m}{2l+1}\right) P_{l-1}^m(\cos\theta) + \left(\frac{l-m+1}{2l+1}\right) P_{l+1}^m(\cos\theta)\right) e^{im\phi} \tag{A.3.14}$$

Also,

$$P_{l-1}^m(\cos\theta)e^{im\phi} = \left(\frac{4\pi(l+m-1)!}{(2l-1)(l-m-1)!}\right)^{1/2} Y_{l-1,m} \tag{A.3.15}$$

$$P_{l+1}^m(\cos\theta)e^{im\phi} = \left(\frac{4\pi(l+m+1)!}{(2l+3)(l-m+1)!}\right)^{1/2} Y_{l+1,m} \tag{A.3.16}$$

Inserting (A.3.15) and (A.3.16) the first and second term of the right hand side of (A.3.14), and doing some algebra we arrive at (A.3.9).

To prove (A.3.10) we first have,

$$\sin\theta e^{i\phi} Y_{l,m} = \left(\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right)^{1/2} (1-\cos^2\theta)^{1/2} P_l^m(\cos\theta)e^{i(m+1)\phi}. \tag{A.3.17}$$

By (A.2.11), we have,

$$(1-\cos^2\theta)^{1/2} P_l^m(\cos\theta) = \frac{P_{l-1}^{m+1}(\cos\theta)}{2l+1} - \frac{P_{l-1}^{m+1}(\cos\theta)}{2l+1}. \tag{A.3.18}$$

Then,

$$Y_{l-1,m+1} = \left(\frac{(2l-1)(l-m-2)!)}{4\pi(l+m)!}\right)^{1/2} P_{l-1}^{m+1}(\cos\theta)e^{i(m+1)\phi} \tag{A.3.19}$$

$$Y_{l+1,m+1} = \left(\frac{(2l+3)(l-m)!)}{4\pi(l+m+2)!}\right)^{1/2} P_{l+1}^{m+1}(\cos\theta)e^{i(m+1)\phi}. \tag{A.3.20}$$

Re-arranging, we have,

$$P_{l-1}^{m+1}(\cos\theta)e^{i(m+1)\phi} = \left(\frac{4\pi(l+m)!}{(2l-1)(l-m-2)!}\right)^{1/2} Y_{l-1,m+1} \tag{A.3.21}$$

$$P_{l+1}^{m+1}(\cos\theta)e^{i(m+1)\phi} = \left(\frac{4\pi(l+m+2)!}{(2l+3)(l-m)!}\right)^{1/2} Y_{l+1,m+1} \tag{A.3.22}$$

Inserting (A.3.21) and (A.3.22) into the first and second term of the right hand side of (A.3.18) and subsequently replacing this result on the right hand side of (A.3.17) gives the result after some algebra. □

Finally, we have the unit vector expansion,

$$\hat{\boldsymbol{s}} = \begin{pmatrix} s_x \\ s_y \\ s_x \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\varphi \end{pmatrix} = \sqrt{\frac{4\pi}{3}} \begin{pmatrix} \frac{1}{\sqrt{2}}\left(Y_{1,-1}(\hat{\boldsymbol{s}}) - Y_{1,1}(\hat{\boldsymbol{s}})\right) \\ \frac{i}{\sqrt{2}}\left(Y_{1,-1}(\hat{\boldsymbol{s}}) + Y_{1,1}(\hat{\boldsymbol{s}})\right) \\ Y_{1,0}(\hat{\boldsymbol{s}}) \end{pmatrix} \tag{A.3.23}$$

Appendix B

Mathematical Preliminaries

This appendix contains a collection of useful identities and formulae used throughout the dissertation.

B.1 Matrix and Vector Identities

Here we present many of the matrix identities that appear in the text. These and many more identities can be found in The Matrix Cookbook of Petersen and Pedersen, 2007.

The derivatives of the elements of an inverse matrix are

$$\frac{\partial K^{-1}}{\partial \theta} = -K^{-1}\frac{\partial K}{\partial \theta}K^{-1} \tag{B.1.1}$$

where $\partial_\theta K$ are the elementwise derivatives. Further, for a symmetric positive definite matrix K, the elementwise derivative of its log determinant is,

$$\frac{\partial}{\partial \theta}\log\det K = \operatorname{tr}\left(K^{-1}\frac{\partial K}{\partial \theta}\right) \tag{B.1.2}$$

which follows from Jacobi's formula,

$$\frac{\partial}{\partial \theta}\det K = (\det K)\operatorname{tr}\left(\frac{\partial K}{\partial \theta}\right). \tag{B.1.3}$$

Consider a partioned matrix A,

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix} \tag{B.1.4}$$

where E and H are invertible. The inverse of A is given by the *partioned inverse formula* (Murphy, 2012),

$$A^{-1} = \begin{pmatrix} (A/H)^{-1} & -(A/H)^{-1}FH^{-1} \\ -H^{-1}G(A/H)^{-1} & H^{-1} + H^{-1}G(A/H)^{-1}FH^{-1} \end{pmatrix} \tag{B.1.5}$$

$$= \begin{pmatrix} E^{-1} + E^{-1}F(A/E)^{-1}GE^{-1} & -E^{-1}F(A/E)^{-1} \\ -(A/E)^{-1}GE^{-1} & (A/E)^{-1} \end{pmatrix} \tag{B.1.6}$$

where,

$$A/H = E - FH^{-1}G \tag{B.1.7}$$

$$A/E = H - GE^{-1}F \tag{B.1.8}$$

are the *Schur complements* of A with respect to H and E, respectively. Using (B.1.5), we arrive at the *matrix inversion lemma* or the *Sherman-Morrison-Woodbury formula* for the general partioned matrix A given above,

$$(E - FH^{-1}G)^{-1} = E^{-1} + E^{-1}F(H - GE^{-1}F)^{-1}GE^{-1} \tag{B.1.9}$$

$$(E - FH^{-1}G)^{-1}FH^{-1} = E^{-1}F(H - GE^{-1}F)^{-1} \tag{B.1.10}$$

for E and H invertible. Additionally we have the *matrix determinant lemma*,

$$\det(E - FH^{-1}G) = \det(H - GE^{-1}F)\det(H^{-1})\det(E). \tag{B.1.11}$$

We also have the useful identify for rearranging into quadratic forms. Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$. Then,

$$-\frac{1}{2}\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{x} = -\frac{1}{2}(\boldsymbol{x} - A^{-1}\boldsymbol{b})^\top A(\boldsymbol{x} - A^{-1}\boldsymbol{b}) + \frac{1}{2}\boldsymbol{b}^\top A^{-1}\boldsymbol{b}. \tag{B.1.12}$$

The *Moore-Penrose* pseudoinverse is a generalisation of the inverse matrix to non-square matrices (Moore, 1920; Bjerhammar, 1951; Penrose, 1955). They are most often used to compute a least-squares solution to linear systems without a unique solution. For a matrix $A \in \mathbb{C}^{m\times n}$, the Moore-Penrose pseudoinverse $A^+ \in \mathbb{C}^{n\times m}$ satisfies the four Moore-Penrose condtions (Penrose, 1955; Golub and Van Loan, 2013),

$$AA^+A = A \tag{B.1.13}$$

$$A^+AA^+ = A^+ \tag{B.1.14}$$

$$(AA^+)^* = AA^+ \tag{B.1.15}$$

$$(A^+A)^* = A^+A. \tag{B.1.16}$$

For a vector $\boldsymbol{a}$ independent of a unit vector $\hat{\boldsymbol{s}}$, we have

$$\int_{4\pi} (\hat{\boldsymbol{s}} \cdot \boldsymbol{a})\hat{\boldsymbol{s}}\, d\Omega = \frac{4\pi}{3}\boldsymbol{a} \tag{B.1.17}$$

$$\int_{4\pi} \hat{\boldsymbol{s}}\left(\hat{\boldsymbol{s}} \cdot \nabla\left(\boldsymbol{a} \cdot \hat{\boldsymbol{s}}\right)\right)\, d\Omega = 0 \tag{B.1.18}$$

B.2 Gaussian Identities

The multivariate Gaussian distribution has joint probability density

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-d/2} \det \Sigma^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{B.2.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the *mean* vector and $\Sigma \in \mathbb{R}^{d\times d}$ is the symmetric positive definite *covariance matrix*. This is expressed as $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$.

Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be jointly Gaussian distributed random vectors,

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2) = N(\boldsymbol{\mu}, \Sigma) \tag{B.2.2}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2\end{pmatrix} \quad \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}, \quad \Lambda = \Sigma^{-1} = \begin{pmatrix}\Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22}\end{pmatrix} \tag{B.2.3}$$

where Λ is the *precision matrix*.

The marginals of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by

$$p(\boldsymbol{x}_1) = \int p(\boldsymbol{x}_1, \boldsymbol{x}_2)\, d\boldsymbol{x}_2 = N(\boldsymbol{\mu}_1, \Sigma_{11}) \tag{B.2.4}$$

$$p(\boldsymbol{x}_2) = \int p(\boldsymbol{x}_1, \boldsymbol{x}_2)\, d\boldsymbol{x}_1 = N(\boldsymbol{\mu}_2, \Sigma_{21}). \tag{B.2.5}$$

The conditional distributions are given by,

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{p(\boldsymbol{x}_2)} = N(\boldsymbol{x}_1|\boldsymbol{\mu}_{1|2}, \Sigma_{1|2}) \tag{B.2.6}$$

where the conditional mean $\boldsymbol{\mu}_{1|2}$ and covariance $\Sigma_{1|2}$ are given by,

$$\begin{aligned}
\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) && \text{(B.2.7)}\\
&= \boldsymbol{\mu}_1 - \Lambda_{11}^{-1}\Lambda_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) && \text{(B.2.8)}\\
&= \Sigma_{1|2}\left(\Lambda_{11}\boldsymbol{\mu}_1 - \Lambda_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)\right) && \text{(B.2.9)}\\
\Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Lambda_{11}^{-1}. && \text{(B.2.10)}
\end{aligned}$$

Linear combinations of Gaussians are also Gaussian. Take $\boldsymbol{x} \sim N(\boldsymbol{\mu}_x, \Sigma_x)$ and $\boldsymbol{y} \sim N(\boldsymbol{\mu}_y, \Sigma_y)$. Then,

$$A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{c} \sim N(A\boldsymbol{\mu}_x + B\boldsymbol{\mu}_y + \boldsymbol{c}, A\Sigma_X A^\top + B\Sigma_y B^\top) \tag{B.2.11}$$

The product of two Gaussians is itself Gaussian. Consider two Gaussians $N(\boldsymbol{x}|\boldsymbol{a}, A)$ and $N(\boldsymbol{x}|\boldsymbol{b}, B)$. Their product is

$$N(\boldsymbol{x}|\boldsymbol{a}, A)N(\boldsymbol{x}|\boldsymbol{b}, B) = Z^{-1}N(\boldsymbol{x}|\boldsymbol{c}, C) \tag{B.2.12}$$

$$\boldsymbol{c} = C(A^{-1}\boldsymbol{a} + B^{-1}\boldsymbol{b}) \tag{B.2.13}$$

$$C = (A^{-1} + B^{-1})^{-1} \tag{B.2.14}$$

with normalisation constant Z^{-1},

$$Z^{-1} = (2\pi)^{-d/2} \det(A + B)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{a} - \boldsymbol{b})^\top (A + B)^{-1}(\boldsymbol{a} - \boldsymbol{b})\right\}. \tag{B.2.15}$$

Appendix C

The P_N Approximations Derivation

C.1 Spherical Hamonic Expansion

We here provide detailed derivations of the P_N approximations to the radiative transport equation (4.2.1) as given in Section 4.3. Given the RTE,

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \hat{\boldsymbol{s}}\cdot\nabla + \mu_t(\boldsymbol{r})\right)\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t) = \mu_s(\boldsymbol{r})\int_{\mathbb{S}^{d-1}} k(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{s}}')\phi(\boldsymbol{r},\hat{\boldsymbol{s}}',t)\,d\hat{\boldsymbol{s}} + q(\boldsymbol{r},\hat{\boldsymbol{s}},t), \tag{C.1.1}$$

we take the inner product term-by-term of the spherical harmonic basis expansion 4.3.8 with $Y^*_{l,m}(\hat{\boldsymbol{s}})$ to exploit the orthogonal structure of the spherical harmonics. With the exception of the transport term $\hat{\boldsymbol{s}}\cdot\nabla\phi$, the terms in (4.3.8) decouple. Beginning with the first term we have,

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\int_{\mathbb{S}^{n-1}}\left(\sum_l^\infty\sum_{m=-l}^{l}\left(\frac{2l+1}{4\pi}\right)^{1/2}\psi_{l,m}(\boldsymbol{r},t)Y_{l,m}(\hat{\boldsymbol{s}})\right)Y^*_{l,m}(\hat{\boldsymbol{s}})\,d\hat{\boldsymbol{s}} \tag{C.1.2}$$

$$= \left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\int_0^{2\pi}\int_0^{\pi}\left(\sum_l^\infty\sum_{m=-l}^{l}\left(\frac{2l+1}{4\pi}\right)^{1/2}\psi_{l,m}(\boldsymbol{r},t)Y_{l,m}(\theta,\varphi)\right)Y^*_{l,m}(\theta,\varphi)\sin\theta\,d\theta\,d\varphi \tag{C.1.3}$$

$$= \left(\frac{1}{c}\frac{\partial}{\partial t} + \mu_t(\boldsymbol{r})\right)\psi_{l,m}(\boldsymbol{r},t). \tag{C.1.4}$$

The scattering operator becomes,

$$\begin{aligned}\int_{\mathbb{S}^{n-1}}\mu_s(\boldsymbol{r})\int_{\mathbb{S}^{n-1}}&\left(\sum_l^\infty\sum_{m=-1}^{l}k_l Y^*_{l,m}(\hat{\boldsymbol{s}}')Y_{l,m}(\hat{\boldsymbol{s}})\right)\\ &\times\left(\sum_{l=0}^\infty\sum_{m=-l}^{l}\left(\frac{2l+1}{4\pi}\right)^{1/2}\psi_{l,m}(\boldsymbol{r},t)Y_{l,m}(\hat{\boldsymbol{s}})\right)d\hat{\boldsymbol{s}}'\,d\hat{\boldsymbol{s}}\end{aligned} \tag{C.1.5}$$

and the source term is,

$$\int_{\mathbb{S}^{n-1}} \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{2\pi} \right)^{1/2} q_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}) Y_{l,m}^{*}(\hat{\boldsymbol{s}})\, d\hat{\boldsymbol{s}} = q_{l,m}(\boldsymbol{r},t). \tag{C.1.6}$$

The transport term $\hat{\boldsymbol{s}} \cdot \nabla \phi$ presents difficulty for straightforward evaluation and requires the use of the spherical harmonic recurrence relations (A.3.9)–(A.3.11). We first transform the transport operator using (A.3.23), as

$$\hat{\boldsymbol{s}} \cdot \nabla = \sin\theta \cos\varphi \frac{\partial}{\partial x} + \sin\theta \sin\varphi \frac{\partial}{\partial y} + \cos\theta \frac{\partial}{\partial z}, \tag{C.1.7}$$

then,

$$\begin{aligned}
&\int_{\mathbb{S}^{n-1}} \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \hat{\boldsymbol{s}} \cdot \nabla \psi_{l,m}(\boldsymbol{r},t) Y_{l,m}(\hat{\boldsymbol{s}}) Y_{l,m}^{*}(\hat{\boldsymbol{s}})\, d\hat{\boldsymbol{s}} \\
&= \int_0^{2\pi} \int_0^{\pi} \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \left[\sin\theta \cos\varphi \frac{\partial \psi_{l,m}}{\partial x} + \sin\theta \sin\varphi \frac{\partial \psi_{l,m}}{\partial y} + \cos\theta \frac{\partial \psi_{l,m}}{\partial z} \right] \\
&\qquad Y_{l,m}(\theta,\varphi) Y_{l,m}^{*}(\theta,\varphi) \sin\theta\, d\theta\, d\varphi.
\end{aligned} \tag{C.1.8}$$

For the ∂_z term, we have by the recurrence relation (A.3.9),

$$\begin{aligned}
&\int_0^{2\pi} \int_0^{\pi} \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \frac{\partial \psi_{l,m}}{\partial z} \cos\theta Y_{l,m}(\theta,\varphi) Y_{l,m}^{*}(\theta,\varphi) \sin\theta\, d\theta\, d\varphi \\
&= \int_0^{2\pi} \int_0^{\pi} \sum_{l}^{\infty} \sum_{m=-l}^{l} \left(\frac{2l+1}{4\pi} \right)^{1/2} \frac{\partial \psi_{l,m}}{\partial z} \left[\left(\frac{(l+m)(l-m)}{(2l-1)(2l+1)} \right)^{1/2} Y_{l-1,m} \right. \\
&\left. + \left(\frac{(l+m+1)(l-m+1)}{(2l+1)(2l+3)} \right)^{1/2} Y_{l+1,m} \right] Y_{l,m}^{*}(\theta,\varphi) \sin\theta\, d\theta\, d\varphi.
\end{aligned} \tag{C.1.9}$$

By orthogonality (A.3.3) the term (C.1.9) reduces to

$$\frac{1}{2l+1} \left[\frac{\partial}{\partial z} \left((l+m)(l-m) \right)^{1/2} \psi_{l-1,m} + \left((l+m+1)(l-m+1) \right)^{1/2} \psi_{l+1,m} \right) \right]. \tag{C.1.10}$$

For the ∂_x and ∂_y terms we make use of (A.3.10)–(A.3.11). Consdering the integrands, from Euler's identity, we obtain

$$\begin{aligned}
&\left(\frac{\partial \psi_{l,m}}{\partial x} \sin\theta \cos\varphi + \frac{\partial \psi_{l,m}}{\partial y} \sin\theta \sin\varphi \right) Y_{l,m} Y_{l,m}^{*} \\
&= \left(\frac{\partial \psi_{l,m}}{\partial x} \sin\theta e^{-i\varphi} - i \frac{\partial \psi_{l,m}}{\partial x} \sin\theta \sin\varphi - i \frac{\partial \psi_{l,m}}{\partial y} \sin\theta \cos\varphi + i \frac{\partial \psi_{l,m}}{\partial y} \sin\theta e^{-i\varphi} \right) Y_{l,m} Y_{l,m}^{*}.
\end{aligned} \tag{C.1.11}$$

Combining terms,

$$\sin\theta e^{-i\varphi}\left(\frac{\partial\psi_{l,m}}{\partial x}-i\frac{\partial\psi_{l,m}}{\partial y}\right)Y_{l,m}Y^*_{l,m}+i\sin\theta(\cos\varphi-\sin\varphi)\left(\frac{\partial\psi_{l,m}}{\partial x}+\frac{\partial\psi_{l,m}}{\partial y}\right)Y_{l,m}Y^*_{l,m} \tag{C.1.12}$$

yields,

$$\begin{aligned}\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)\psi_{l,m}\Bigg[&\left(\frac{(l+m)(l+m-1)}{(2l-1)(2l+1)}\right)^{1/2}Y_{l-1,m-1}\\&+\left(\frac{(l-m+1)(l-m+2)}{(2l+1)(2l+3)}\right)^{1/2}Y_{l+1,m-1}\Bigg]Y^*_{l,m}\end{aligned} \tag{C.1.13}$$

Integrating with the orthogonality condition (A.3.3) results in,

$$\begin{aligned}=-\frac{1}{2(2l+1)}\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)[&-\left((l-m)(l+m-1)\right)^{1/2}\psi_{l-1,m+1}\\&+\left((l+m+1)(l+m+2)\right)^{1/2}\psi_{l+1,m+1}].\end{aligned} \tag{C.1.14}$$

Following the same procedure for the other term making use of recurrence relation (A.3.10), we obtain

$$\begin{aligned}\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)\psi_{l,m}\Bigg[&\left(\frac{(l-m)(l-m-1)}{(2l-1)(2l+1)}\right)^{1/2}Y_{l-1,m+1}\\&+\left(\frac{(l+m+1)(l+m+2)}{(2l+1)(2l+3)}\right)^{1/2}Y_{l+1,m+1}\Bigg]Y^*_{l,m}\end{aligned} \tag{C.1.15}$$

Integrating again with the orthogonality condition yields,

$$\begin{aligned}=-\frac{1}{2(2l+1)}\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)[&-\left((l-m)(l+m-1)\right)^{1/2}\psi_{l-1,m+1}\\&+\left((l+m+1)(l+m+2)\right)^{1/2}\psi_{l+1,m+1}].\end{aligned} \tag{C.1.16}$$

Combining (C.1.4) – (C.1.16), yields the infinite set of coupled partial differential equations,

the P_N approximations,

$$\begin{aligned}\left(\frac{1}{c}\frac{\partial}{\partial t}+\mu_t(\boldsymbol{r})\right)\psi_{l,m}(\boldsymbol{r},t)+\frac{1}{2l+1}\Bigg(&\frac{\partial}{\partial z}[((l-m+1)(l+m+1))^{1/2}\psi_{l+1,m}(\boldsymbol{r},t)\\&+((l-m)(l+m))^{1/2}\psi_{l-1,m}(\boldsymbol{r},t)]\\&-\frac{1}{2}\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)[((l+m)(l+m-1))^{1/2}\psi_{l-1,m-1}(\boldsymbol{r},t)\\&-((l-m+1)(l-m+2))^{1/2}\psi_{l+1,m-1}(\boldsymbol{r},t)]\\&-\frac{1}{2}\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)[-((l-m)(l-m-1))^{1/2}\psi_{l-1,m+1}(\boldsymbol{r},t)\\&+((l+m+1)(l+m+2))^{1/2}\psi_{l+1,m+1}(\boldsymbol{r},t)\Bigg)\\&=\mu_s(\boldsymbol{r})k_l\psi_{l,m}(\boldsymbol{r},t)+q_{l,m}(\boldsymbol{r},t).\end{aligned}\tag{C.1.17}$$

C.2 Diffusion Approximation Derivation

In this section we show the intermediate calculations resulting from substituting the diffusion expansion into the RTE to arrive at the coupled P_1 approximation in terms of the fluence Φ and flux $\boldsymbol{J}$.

Substituting the diffusion expansion,

$$\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)=\frac{1}{4\pi}\Phi(\boldsymbol{r},t)+\frac{3}{4\pi}\boldsymbol{J}(\boldsymbol{r},t)\cdot\hat{\boldsymbol{s}}.\tag{C.2.1}$$

into the RTE (4.2.1) and integrating gives a series of terms

$$\int_{4\pi}\frac{1}{c}\frac{\partial\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)}{\partial t}\,d\Omega=\frac{1}{c}\frac{\partial\Phi(\boldsymbol{r},t)}{\partial t}\tag{C.2.2}$$

$$\int_{4\pi}\hat{\boldsymbol{s}}\cdot\nabla\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)\,d\Omega=\int_{4\pi}\nabla\cdot\left(\hat{\boldsymbol{s}}\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)\right)\,d\Omega=\nabla\cdot\int_{4\pi}\hat{\boldsymbol{s}}\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)\,d\Omega=\nabla\cdot\boldsymbol{J}(\boldsymbol{r},t)\tag{C.2.3}$$

where we have used the vector identites $\hat{\boldsymbol{s}}\cdot\nabla\phi=\nabla\cdot(\hat{\boldsymbol{s}}\phi)-\phi\nabla\cdot\hat{\boldsymbol{s}}$, and $\nabla\cdot\hat{\boldsymbol{s}}=0$. Next,

$$\int_{4\pi}\mu_t\phi(\boldsymbol{r},\hat{\boldsymbol{s}},t)\,d\Omega=\mu_t\Phi(\boldsymbol{r},t).\tag{C.2.4}$$

Assuming an isotropic source – $\boldsymbol{q}(\boldsymbol{r},\hat{\boldsymbol{s}},t)$ is independent of $\hat{\boldsymbol{s}}$ gives $\boldsymbol{q}(\boldsymbol{r},\hat{\boldsymbol{s}},t)=\boldsymbol{q}(\boldsymbol{r},t)/4\pi$, which becomes

$$\int_{4\pi}\boldsymbol{q}(\boldsymbol{r},\hat{\boldsymbol{s}},t)\,d\Omega=\boldsymbol{q}(\boldsymbol{r},t).\tag{C.2.5}$$

The scattering term again poses the most difficulty, yielding,

$$\mu_s \int_{4\pi} \left(\int_{4\pi} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \phi(\boldsymbol{r}, \hat{\boldsymbol{s}}', t)\, d\hat{\boldsymbol{s}}' \right) d\hat{\boldsymbol{s}} \tag{C.2.6}$$

$$= \mu_s \int_{4\pi} \int_{4\pi} \left(\frac{1}{4\pi} \Phi(\boldsymbol{r}, t) + \frac{3}{4\pi} \boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}} \right) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') d\hat{\boldsymbol{s}}' d\hat{\boldsymbol{s}} \tag{C.2.7}$$

The first term is

$$\int_{4\pi} \left(\int_{4\pi} \Phi(\boldsymbol{r}, t) k(\hat{\boldsymbol{s}} \cdot \Omega')\, d\hat{\boldsymbol{s}} \right) d\Omega = \Phi(\boldsymbol{r}, t) \int_{4\pi} \left(\int_{4\pi} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}')\, d\Omega' \right) d\Omega \tag{C.2.8}$$

$$= 4\pi \Phi(\boldsymbol{r}, t) \tag{C.2.9}$$

and the second is

$$\int_{4\pi} \int_{4\pi} (\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}')\, k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}')\, d\Omega' d\Omega \tag{C.2.10}$$

$$= |J(\boldsymbol{r}, t)| \int_{4\pi} \left(\int_{4\pi} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}')\, d\Omega \right) \cos\theta'\, d\Omega' \tag{C.2.11}$$

$$= |J(\boldsymbol{r}, t)| \int_{4\pi} \cos\theta'\, d\Omega' \tag{C.2.12}$$

$$= 0. \tag{C.2.13}$$

Combing these terms all together yields

$$\frac{\partial \Phi(\boldsymbol{r}, t)}{\partial t} + \mu_a \Phi(\boldsymbol{r}, t) + \nabla \cdot \boldsymbol{J}(\boldsymbol{r}, t) = \boldsymbol{q}(\boldsymbol{r}, t) \tag{C.2.14}$$

Additionally, substituting the diffusion expansion into the RTE (4.2.1), multiplying by $\hat{\boldsymbol{s}}$ and integrating yields the second equation. Term-by-term, we have,

$$\int_{4\pi} \hat{\boldsymbol{s}} \frac{1}{c} \frac{\partial \phi(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)}{\partial t}\, d\hat{\boldsymbol{s}} = \frac{1}{c} \frac{\partial \boldsymbol{J}(\boldsymbol{r}, t)}{\partial t}. \tag{C.2.15}$$

Using (B.1.17) and (B.1.18), the transport term is

$$\int_{4\pi} \hat{\boldsymbol{s}} \cdot \nabla \left(\frac{1}{4\pi} \Phi(\boldsymbol{r}, t) + \frac{3}{4\pi} \boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}} \right) \hat{\boldsymbol{s}}\, d\Omega = \frac{\nabla \Phi(\boldsymbol{r}, t)}{3}. \tag{C.2.16}$$

The extinction and source terms become,

$$\int_{4\pi} \hat{\boldsymbol{s}} \phi(\boldsymbol{r}, t)\, d\Omega = \mu_t \boldsymbol{J}(\boldsymbol{r}, t) \tag{C.2.17}$$

$$\int_{4\pi} \hat{\boldsymbol{s}} \boldsymbol{q}(\boldsymbol{r}, \hat{\boldsymbol{s}}, t)\, d\Omega = \frac{\boldsymbol{q}(\boldsymbol{r}, t)}{4\pi} \int_{4\pi} \hat{\boldsymbol{s}}\, d\Omega = 0. \tag{C.2.18}$$

The scattering term is,

$$\mu_s \int_{4\pi} \int_{4\pi} \hat{\boldsymbol{s}} \left(\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}', t) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega'\right) d\Omega \tag{C.2.19}$$

$$= \frac{1}{4\pi} \int_{4\pi} \int_{4\pi} \hat{\boldsymbol{s}} \left(\Phi(\boldsymbol{r}, t) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega'\right) \, d\Omega \tag{C.2.20}$$

$$+ \frac{3}{4\pi} \int_{4\pi} \hat{\boldsymbol{s}} \left(\int_{4\pi} \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}'\right) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega' \right) \, d\Omega. \tag{C.2.21}$$

The first integral vanishes by definition of the scattering function, while the second intergral is split via the vector identity $\hat{\boldsymbol{s}} = \hat{\boldsymbol{s}}'(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') + \hat{\boldsymbol{s}}' \times (\hat{\boldsymbol{s}} \times \hat{\boldsymbol{s}}')$,

$$\int_{4\pi} \hat{\boldsymbol{s}} \left(\int_{4\pi} \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}'\right) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega' \right) \, d\Omega = \int_{4\pi} \left(\int_{4\pi} \hat{\boldsymbol{s}} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega \right) \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}\right) d\Omega \tag{C.2.22}$$

$$= \int_{4\pi} \left(\int_{4\pi} \hat{\boldsymbol{s}}'(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega + \hat{\boldsymbol{s}}' \times \int_{4\pi} \hat{\boldsymbol{s}} k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}) \, d\Omega \times \hat{\boldsymbol{s}}' \right) \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}'\right) d\Omega' \tag{C.2.23}$$

By the symmetry of the scattering function $k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}')$ about $\hat{\boldsymbol{s}}$, the second integral term is parallel to $\hat{\boldsymbol{s}}'$, and thus vanishes. The scattering term is therefore,

$$\int_{4\pi} \hat{\boldsymbol{s}} \left(\int_{4\pi} \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}'\right) k(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}') \, d\Omega' \right) \, d\Omega = g \int_{4\pi} \hat{\boldsymbol{s}}' \left(\boldsymbol{J}(\boldsymbol{r}, t) \cdot \hat{\boldsymbol{s}}'\right) \, d\Omega' \tag{C.2.24}$$

$$= \frac{4\pi}{3} g \boldsymbol{J}(\boldsymbol{r}, t). \tag{C.2.25}$$

combing these terms we have

$$\frac{1}{c} \frac{\partial \boldsymbol{J}(\boldsymbol{r}, t)}{\partial t} + (\mu_a + \mu_s') \boldsymbol{J}(\boldsymbol{r}, t) + \frac{1}{3} \nabla \Phi(\boldsymbol{r}, t) = 0 \tag{C.2.26}$$