TUDelft

Delft University of Technology

Ultrasound imaging through aberrating layers

van der Meulen, P.Q.

DOI

Publication date
2023

Document Version
Final published version

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable).
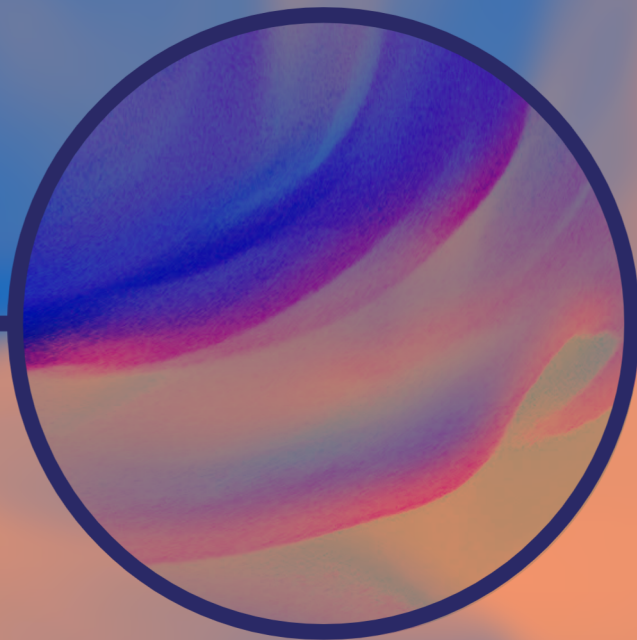Please check the document version above.

Ultrasound imaging through aberrating layers

Pim Quinten van der Meulen

ULTRASOUND IMAGING THROUGH ABERRATING LAYERS

ULTRASOUND IMAGING THROUGH ABERRATING LAYERS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 31 januari 2023 om 15:00 uur

door

Pim Quinten VAN DER MEULEN

Elektrotechnisch Ingenieur, Technische Universiteit Delft, Nederland
geboren te Rotterdam, Nederland

Voor mijn ouders.

CONTENTS

SUMMARY

Whereas aberrating layers are typically viewed as an impediment to medical ultrasound imaging, they can, surprisingly, also be used to our benefit. As long as we can model the effect of an aberrating layer, we can utilize 'model-based imaging', the imaging technique explored throughout this thesis, to reconstruct ultrasound images where traditional beamforming methods would fail, employing the ever increasing computational power available to us nowadays. Not only does this allow us to image through layers, but it also leads to interesting applications, such as 3D ultrasound imaging with spatially undersampled data, using an aberrating 'coding mask'. The formulation of a measurement model, a fundamental part of model-based imaging, also gives insight into the imaging problem mathematically, and allows us to investigate methods for estimating the effect of an aberrating layer 'blindly', i.e., without explicitly measuring it. In this thesis, we thus investigate *(a)*, imaging through a layer when the layer's aberration effect is known, and how it can be applied to imaging with spatially undersampled data, and *(b)*, methods and algorithms for estimating the effect of the aberrating layer without knowing it a priori.

In the first part of this thesis, we illustrate how using model-based imaging can be utilized for 3D ultrasound imaging using a single ultrasound transducer, and equipping it with a plastic coding mask. The plastic mask acts as an analog coder, that scrambles the transmitted and received waves in a manner that is location dependent. As a result, the temporal shape of an ultrasound echo can be used instead of the traditional method of using phase differences between sensors in a sensor array. Imaging is instead accomplished using model-based imaging. By measuring the pulse-echo response of each pixel, we can form an image by solving a regularized linear least squares problem, which takes into account the measured pixel-specific pulse-echo signals. The proposed device and imaging method is then verified experimentally.

In the following chapter, a coding mask design method is proposed for the aforementioned imaging device. A measurement model is formulated where the mask geometry is an explicit parameter to be optimized. After forming this model, a numerical optimization method is proposed and numerically tested. Our numerical experiments show that optimized mask geometries exhibit an energy focusing effect on the region-of-interest, whilst simultaneously decorrelating echo signals between pixels.

In the second part of this thesis, in contrast, we consider methods for calibrating propagation models when the pulse-echo response per pixel is *not* known. The most important calibration challenge we consider is that of imaging through an aberrating layer in front of an ultrasound array. This could be subcutaneous fat or the human skull, for example. In this thesis we formulate a measurement model consisting of a part where wave propagation is *known* (i.e., the assumed homogeneous region behind the aberrating layer, where the contrast image of interest is located), and an *unknown* propagation part, consisting of the Green's functions from an array sensor to any point on the the

interface of the aberrating layer and the imaging medium. We then investigate methods for finding this set of Green's functions without explicitly measuring them (so called 'blind' calibration).

The first proposed method exploits the singular value decomposition of the measurement data in combination with the assumed Toeplitz structure of the matrices representing the aberrating layer's Green's functions. However, the method is lacking in practicality since an additional set of measurements is required with a phase screen mounted on the interface of the aberration layer and the imaging medium. The second method resolves these practical issues by utilizing a covariance matching technique. A sufficiently large set of measurements is obtained where each measurement is different due to e.g. moving particles such as blood flow or micro-bubbles. Using the covariance of the data, algorithms are then defined that can estimate the transfer functions of the aberrating layer from the measurement covariance data.

Finally, we propose a method for estimating the electro-mechanical impulse response of an ultrasound sensor, by simply measuring its pulse-echo response from a flat plate reflector in front of the sensor. Estimating the one-way (electro-mechanical) impulse response then becomes a de-autoconvolution problem, for which we propose a method by solving a semi-definite relaxation of the de-autoconvolution problem.

SAMENVATTING

Hoewel aberrerende lagen doorgaans als een obstakel worden gezien voor medische echografie, kunnen ze verrassend genoeg ook in ons voordeel worden gebruikt. Zolang we het effect van de aberratielaag maar kunnen modelleren, kunnen we gebruik maken van 'model-afgeleide beeldvorming' ('*model-based imaging*'), de beeldvormingstechniek die wordt gebruikt in deze dissertatie, om echografische beelden te reconstrueren wanneer traditionele beeldvormingstechnieken falen, gebruikmakend van de steeds toenemende computationele kracht die ons tegenwoordig ter beschikking staat. Behalve dat dit behulpzaam is voor beeldvorming door aberratielagen heen, leidt dit ook tot interessante toepassingen zoals 3D echografie met spatieel onderbemonsterde data, door gebruik te maken van een zogenaamd coderingsmasker. De formulatie van een meetmodel, een fundamenteel onderdeel van mode-afgeleide beeldvorming, geeft ook wiskundig inzicht in het beeldvormingsprobleem, waardoor het mogelijk wordt om methoden te onderzoeken die het effect van een aberratielaag *blind* kunnen schatten, d.w.z., zonder dit effect expliciet zelf te meten. In deze dissertatie doen we dus onderzoek naar *(a)*, beeldvorming door een aberratielaag heen wanneer het effect van deze laag al bekend is, en hoe dit kan worden gebruikt voor beeldvorming uit spatieel onderbemonstde metingen, en *(b)*, methoden en algoritmen voor het achterhalen van van het effect van de aberratielaag zonder dat die van tevoren bekend is.

In het eerste gedeelte van deze dissertatie illustreren we hoe model-afgeleide beeldvorming kan worden gebruikt voor 3D echografie met slechts een enkele sensor, door die sensor met een plastic coderingsmasker uit te rusten. Het masker doet dienst als een analoog coderingsmechanisme dat de uitgezonden en ontvangen echo's verstoort op een manier die afhankelijk is van de locatie van een reflectie. Hierdoor kan kan de temporele vorm van de echo worden gebruikt in een model-afgeleid beeldvormingsalgoritme, in plaats van door gebruik te maken van faseverschillen tussen een reeks sensoren. Door de puls-echo van elke pixel te meten kan een beeld worden achterhaald door een geregulariseerd kleinste-kwadratenprobleem op te lossen, die rekening houdt met de gemeten pixel-specifieke puls-echo signalen. Het voorgestelde apparaat en bijbehorend reconstructiealgoritme worden daarna experimenteel geverifieerd.

In het daaropvolgende hoofdstuk wordt een methode geïntroduceerd voor het ontwerpen van een coderingsmasker zoals gebruikt in het hiervoor besproken apparaat. Er wordt een meetmodel geformuleerd waar de de vorm van het coderingsmasker een expliciete variabele is waarvoor kan worden geoptimaliseerd. Daarna wordt een numeriek optimilizatiealgoritme geintroduceerd en numeriek getest. Deze experimenten tonen aan dat de geoptimaliseerde maskers energie proberen te focusseren naar het gedefinieerde pixeldomein, en tegelijkertijd proberen echosignalen tussen pixels te decorreleren.

In het tweede gedeelte van deze dissertatie worden methoden voor het kalibreren van propagatiemodellen voorgesteld, waar de pulse-echo respons per pixel *niet* bekend

is. De meest belangrijke uitdaging op dit gebied die wordt bekeken is die van beeldvorming door een aberratielaag heen die zich direct voor een sensorreeks bevindt. Dit kan bijvoorbeeld een laag onderhuids vet zijn, of de menselijke schedel. We formuleren een meetmodel dat bestaat uit een deel waar de propagatie van golven *wel* bekend is (d.w.z., de homogene regio achter de aberratielaag waar de akoestische contrastafbeelding zich bevindt), en een *onbekend* gedeelte, bestaande uit de Greense functies van de sensorreeks naar naar elk punt op het raakvlak tussen de aberrerende laag en het homogene medium. Vervolgens worden methoden onderzocht om deze Greense functies te achterhalen zonder ze expliciet te meten (zogenaamde 'blinde' kalibratie).

De eerst voorgestelde methode exploiteert de singulierenwaardenontbinding van gemeten echodata in combinatie met de veronderstelde Toeplitzstructuur in de matrices die de Greense functies van de aberratielaag representeren. De methode is echter weinig praktisch omdat een tweede set metingen met een fasemasker op het raakvlak tussen de aberratielaag en het beeldvormingsmedium vereist is. De tweede methode lost dergelijke praktische problemen op door gebruik te maken van een *'covariance matching'*-techniek. Een grote verzameling metingen wordt verkregen waar elke meting anders is als gevold van, bijvoorbeeld, bewegende deeltjes in de bloedstroom of *micro-bubbles*. Door te kijken naar de covariantie van de data kunnen vervolgens algoritmen worden ontworpen die de overdrachtsfuncties van aberratielagen kunnen achterhalen uit de covariantie statistieken van de gemeten data.

Als laatste wordt een methode geïntroduceerd voor het schatten van de elektrome-chanische overdrachtsfunctie van een ultrasone sensor, simpelweg door de puls-echo respons van een vlakke plaat gepositioneerd voor de sensor te meten. Het schatten van de enkele-weg (elektromechanische) impulsrespons wordt dan een de-autoconvolutieprobleem, waarvoor een methode wordt voorgesteld door een semi-definiete relaxatie van het originele probleem op te lossen.

1

INTRODUCTION

Soon after launching the Hubble Space Telescope into orbit, it was discovered that one of the lenses in the optical system contained a production error, leading to severe aberrations and loss of resolution in the photos it took of distant galaxies. Service missions were conducted where astronauts tried to repair and update the systems in order to resolve the issue. The problem was partially solved by attaching an optical system that contained the opposite error of the one in place, effectively canceling the aberrating effect of the original system. Since engineers exactly knew the effect of the production error, they were able to come up with a system that took the output of the old system as an input and corrected the aberrations. Even though the original system was flawed, the light waves could still be manipulated to obtain the correct image. In other words, information is not necessarily lost when wavefields are distorted. More importantly, if we know *how* distortion occurs, we can take it into account, and still obtain a good image.

In ultrasound imaging, the central imaging modality in this thesis, a similar situation occurs when ultrasonic waves travel through the human skull, skin, fat, or other aberrating layers. These layers distort ultrasonic waves in an unknown way, reducing image quality. Interestingly, as with the Hubble telescope, this does not necessarily imply that the information about the underlying (medical) ultrasound image is lost. We just need to characterize the distortion, and account for it. In fact, as will be shown, it can be beneficial to intentionally distort the ultrasonic waves in such a way, that more information is extracted for a particular sensing setup.

This has enormous consequences for situations where information is scarce, or where we would like to use fewer sensors without sacraficing imaging quality. Denser information content in the ultrasonic echo wavefields could then help to get better images with the same amount of sensors. Volumetric (3D) imaging, for example, would offer much more complete information to a clinician, as opposed to conventional 2D imaging. The sheer amount of sensors required for 3D imaging, however, results in enormous engineering challenges. After all, if N sensors are needed for a linear array, N^2 sensors are needed for a matrix array with the same resolution. This increasing sensor density makes it challenging to connect all sensors, not to mention the enormous amount of

1

cables needed to connect each sensor to a computer that processes all sensor data to compute an image, and the much increased amount of data that needs to be processed. For such issues (more examples are given in the next section), *compressing* ultrasound waves before they are measured by sensors could offer a solution to the wiring and cabling challenges mentioned above. In this thesis, it will be demonstrated that using a 'coding mask', an aberrating layer that intentionally distorts waves to extract as much information as possible, allows for fewer sensors to be used, which could lead to systems using fewer output channels, less data transfer, and cheaper devices.

This way of imaging can also be considered for biological 'masks', or aberrating layers, such as the human skull, skin, fat, and other layers that distort waves passing through them. Unfortunately, it is rather challenging to construct an 'inverse' skull, skin, or fat layer to cancel the distortion of the skull, as was done with the Hubble telescope. In fact, we would have to find out what that effect is in the first place, and we can't simply remove the layer from a patient to characterize it. Nonetheless, it is of vital importance that we find methods to look through aberrating layers with ultrasound. It would allow researchers to see brain activity non-invasively using the emerging technique of functional ultrasound (fUS) imaging, which could lead to better understanding of the brain, mental diseases, and the workings of the human psyche, but would also help us to obtain more detailed images for patients with higher fat contents or other obstructing tissue layers, leading to improved diagnoses.

1.1. CHALLENGES IN MEDICAL ULTRASOUND IMAGING

Looking through distorting layers with ultrasound is the common theme in this dissertation, and is useful both for looking through aberrating layers, as well as for wavefield compression. In both cases, we are interested in extracting as much information as possible from a (limited) set of measurements. In the case of an aberrating layer such as bone tissue, we would like to find out how this layer distorts, since the layer cannot be removed. In the case of compression, we would like to intentionally introduce an aberrating layer that distorts waves in such a way that as much information as possible about the true image is extracted after measuring the distorted waves, potentially using a spatial sampling below Nyquist rate. In either case, as long as the distortion is characterized, it can be accounted for in the imaging algorithm. The approach of creating undistorted ultrasound images by taking into account the entire wave propagation inside a layer is what binds these two scenarios together, and the much increased computational power available nowadays is what facilitates this technique. Below, these two major challenges in ultrasound imaging that have led to the research presented in this dissertation are discussed in more detail.

1.1.1. SENSOR DENSITY AND TRANSDUCER MINIATURIZATION

One of the first ultrasound arrays was devised in Rotterdam in the Netherlands in 1973 [1, 2], for medical imaging purposes. Each sensor-receiver in the array would transmit a pulse, which reflects on tissue interfaces inside the human body. From the reflected echoes measured on each sensor, a rough map of the patient's tissue directly in front of the corresponding sensor could be inferred, after which these so called 'A-lines' could be

stitched together to form an image. Since then, image reconstruction has advanced by using beamforming techniques [3], where not only the echoes measured on the transmitting transducer, but also on all the other sensors are processed and combined to obtain much better images. Beamforming techniques primarily use the differences in an echo's time-of-arrival between sensors to estimate from which angle an ultrasonic echo originates. Time-of-arrival is used to estimate the depth of an echo, and the signal amplitude is a measure for the scattering intensity. Nowadays, 1D and 2D linear (phased) arrays are available, which use beamforming to compute 2D and 3D images of the human body, respectively.

Ultrasound imaging quality is typically improved by increasing the amount of sensors in an array. As mentioned in the introduction, beamforming with 2D arrays results in a 3D image reconstruction, which provides much more complete information to a clinician than a typical 2D image. However, 2D arrays generate much more measurement data that need to be processed (preferably in real-time, one of ultrasound's major advantages over other imaging modalities such as magnetic resonance imaging (MRI) or computed tomography (CT)), and also cause electronic design problems due to the small amount of sensors and electronics on a small area, thus prohibiting mainstream usage of 3D medical imaging. Such problems include the acoustic transducer design and challenges in electrical interconnections [4].

The same issue plays an important role in spatially limited ultrasound probes, such as ultrasound sensors mounted on intravascular catheters [4]. These catheters are used for minimally invasive ultrasound procedures, and enable the treatment of a range of conditions for which traditionally invasive surgery is required. Since MRI lacks temporal resolution (and requires large, immobile and expensive machines), and CT uses harmful radiation, ultrasound is an excellent candidate to assist a surgeon by visualizing the tissue around the catheter in real time without opening the patient. However, ultrasonic waves lack the ability to penetrate deeply with small wavelengths (for high resolution), or to propagate through bone tissue (e.g., to look through the rib cage during cardiac surgery) at the frequency range which is desired for high spatial resolutions. As a result, engineers started mounting ultrasonic transducers on the catheter tip, so that the imaging takes place as close to the surgical target as possible, without being obstructed, and using small wavelengths.

Real time 3D imaging is highly desirable for catheter-based ultrasound scenarios, and thus suffers from the same challenges in developing 2D arrays as mentioned above. Since space is even more limited than in conventional non-invasive transducers, these challenges become an even greater obstacle. As each sensor in an array would have to be electronically connected to a processor outside the patient for beamforming, many wires would be required which would simply not fit inside the catheter cable, which also contains instruments that are required for surgery. The spatial limitations also mean it is hardly possible to include sophisticated electronics in the catheter that would sufficiently compress ultrasound signals before transmitting them. This leads to suboptimal solutions where only 2D imaging is possible, or where the transducer has to be mechanically moved to scan the region of interest (ROI) to make 3D images. Clearly, there is a need for reducing the amount of measurement data without mechanical scanning, or other measures that reduce frame-rate or are sensitive to motion between frames (e.g.

1

cardiac motion). Addressing this issue would greatly increase the amount of information available to a surgeon, leading to better procedures for a range of minimally-invasive treatments.

Many researchers have approached the compression problem (i.e., reconstructing high quality images from only a few sensors and/or measurements) using the recently developed technique of compressed sensing (CS) [5–8]. Using CS, signals are compressed into a smaller amount of samples by taking different linear combinations of the original signals. The original signal can then be efficiently reconstructed from the (smaller) compressed signal using a convex l_1-regularized optimization problem. Successful reconstruction is only possible if the original signal can be represented using few non-zero coefficients in a suitable signal basis. The theory of CS provides conditions on the linear compression of the original signal to ensure successful signal decompression.

One of the most striking examples of the application of CS is the single-pixel camera [9]. It consists of a single photo diode, onto which different linear combinations of the pixels of the optical input field are projected. Since natural images can easily be sparsely represented using wavelet bases, far fewer of such measurements are needed than there are pixels in the final image. Naturally, this would be an interesting approach to solve the aforementioned challenges in medical ultrasound imaging, and researchers have studied the potential of CS in the field of medical ultrasound imaging ([10–16], to name a few). However, there has been no report of signal bases where actual ultrasound signals or images containing speckle (a noisy component in a US image of human tissue due to the many sub-resolution scatterers throughout the human body) are truly sparse. Moreover, compression mechanisms proposed so far would still require complex electronics that are hard to realize in ultrasound catheters.

In the first part of this thesis, we propose a method where signals are 'naturally' compressed by a an easily manufactured plastic 'coding' mask that distorts the ultrasound fields, and by using model-based imaging, a technique that does not rely on CS or beamforming to reconstruct an image, but instead uses a more complete wave propagation model and a more complicated image estimation technique for imaging. The coding mask is homogeneous, but has a spatially varying height, strongly mixing the 'input' field, resulting in an output field with a strong interference pattern, while ensuring a sufficient signal-to-noise (SNR) ratio by matching the acoustic impedance of water. Instead of using compressive sensing to reconstruct images, we use 'sparse sensing' [17] to optimize the compression mask design. Of course, this imaging approach requires one to image through an aberrating layer (the compression mask), which leads us to the second challenge in the field of ultrasound imaging, discussed next.

1.1.2. IMAGING THROUGH ABERRATING LAYERS

Conventional beamforming techniques for ultrasound imaging implicitly assume that the transmitted ultrasound pulse propagates to each pixel unimpededly, even if there are reflections from scatterers during propagation. The same is assumed for echoes propagating from an acoustic scatterer towards the sensor array. Moreover, it is assumed that a secondary reflection of an echo from another scatterer is negligible. This set of assumptions is known as the Born approximation [18, 19]. It simplifies ultrasound modeling and simulations, and also forms the foundation for ultrasound beamforming to compute ul-

trasound images. The Born approximation is valid when the acoustic tissue inhomogeneities that cause reflections are only weakly scattering (i.e., their acoustic impedance constitutes only a small contrast compared to the overall acoustic impedance of the background medium). The Born approximation is a good approximation for most medical ultrasound scenarios, and is often used as the *de facto* ultrasound model [19–22].

The Born approximation starts to lose its validity for imaging scenarios which cannot be approximated as a homogeneous background medium with small acoustic contrasts. The plastic compression mask mentioned in the previous section, for example, is such a scenario. Although the Born approximation would be valid for waves traveling and scattering inside the tissue of a patient, the propagation through the coding mask is not easily modeled, and needs to be characterized somehow, since waves will bounce around inside the mask before entering the ROI (instead of propagating unimpededly as assumed by the Born approximation).

Another important scenario would be non-invasive ultrasonic imaging of the human brain through the skull. Inside the brain, scattering could be modeled using the Born approximation, but the propagation of echoes or pulses through the skull is unknown, since the skull is highly inhomogeneous, causing many reflections and distorted waves to enter the human brain. Currently, fUS, which enables imaging of brain *activity* using ultrasound arrays (much like in functional magnetic resonance imaging (fMRI), but at higher spatiotemporal resolution, a fraction of the cost of fMRI, and better ease of use), is only possible when the skull is partly removed (in the case of open brain surgery) [23], or thinned (in the case of small adult rodents) [24]. Solving this problem would make it possible to image blood flow to visualize the brain vasculature, and brain activity using fUS [24–26], without removing part of the patient's skull. This would help researchers in understanding the functioning of the human brain, understanding mental disorders, but could also be an invaluable image guidance tool for intra-operative tumor-brain delineation [23].

Researchers have tried to solve this problem for the skull wall particularly using a variety of approaches. Most studies assume the skull can be modeled as a phase screen, i.e., the measured echo signal can be modeled as an unimpeded echo wavefield, with an independent phase change applied to the field on each sensor ([27–29], to name but a few). However, the skull is a highly aberrating layer, and such approaches may only be effective for those parts of the skull which are relatively thin, or when sufficiently large wavelengths are used. A more sophisticated approach is proposed by Fink et. al. [30, 31], which use time-reversal and signal processing techniques to focus energy through the skull, but require one to first experimentally characterize the impulse responses through the skull. A very computationally intensive approach which does not rely on the Born approximation is proposed in [32], where the acoustic wave equations are solved directly. However, this approach requires many sensors around the entire skull, and more importantly, its computational requirements currently make it unsuitable for quickly imaging many ultrasound images, which need to be compounded to obtain a single Doppler or fUS image.

In this thesis, we propose some new steps on the road towards imaging through aberrating layers, whether it is a compression mask or a skull layer. To this end, we propose to model the entire aberrating layer, not as a simple phase mask, but as a full set of Green's

functions (transfer functions) that describe the linear propagation of a wave from one point on one side of the layer, towards any other point on the other side (a similar model as used in [33]). This model would be apt for any aberrating layer besides the skull. We then analyse two methods to estimate this set of transfer functions without the need of first characterizing the layer experimentally, which we call 'blind calibration', since both the skull layer's transfer function as well the the ultrasound contrast image are unknown. We then use *model-based* imaging, the same technique we use to image using a compression mask (see the previous subsection), in order to estimate an image by incorporating these transfer functions into the imaging algorithm.

1.2. MODEL-BASED ULTRASOUND IMAGING

In this dissertation we utilize computational imaging tools that can be used for tackling more complicated ultrasound imaging scenarios, such as the two described above. These computational approaches are able to take into account complex wavefields, such as sophisticated interference patterns created when an ultrasound wave propagates through a (heterogeneous) aberrating layer. We call this computational imaging approach *model-based imaging*. We will demonstrate that, by this approach, and utilizing the large amount of computer power available nowadays, we can still look through some of these layers. In fact, we argue that such layers can be helpful to obtain *more* information about the imaging scene.

Model-based imaging relies on the following **two principles**:

1. An accurate wave propagation model for the imaging scenario. Current imaging methods do not necessarily take into account the full wave propagation model that describes how echo signals look like. For example, the shape of the transmitted pulse, sensor transfer functions, and the spatial interference pattern of transmitted waves are typically not considered. In the first part of this thesis, we experimentally characterize the field transmitted or received through an aberrating layer intentionally used for information compression. Only when using a model-based imaging algorithm can a good image be reconstructed from the ultrasound echoes. In the second part, we look for intelligent ways of characterizing an aberrating layer without a laborious and medically invasive characterization method.

2. The second pillar is that of using an imaging technique that tries to find the image that best explains the measured echo signals given the propagation model from the first principle above. To do this we typically use least squares methods, i.e., mathematical minimization methods, instead of traditional methods that rely on physics-based insights or coherent summation (e.g., delay-and-sum beamforming). This not only allows us to reconstruct images when aberrating layers are present, but also to impose constraints on the reconstructions, based on prior information about, e.g., the image structure.

In Chapter 2, we will describe the physics and mathematics of model-based imaging in more detail.

1

1.3. CONTRIBUTIONS IN THIS DISSERTATION

The contributions of this thesis rely heavily on the two principles of model-based imaging. The main innovations in this thesis consist of the application of model-based imaging to medical ultrasound, showing that it allows us to image with strongly distorted ultrasound fields, and that one can extract more information and create better image reconstructions with less sensors. This way, some first steps are taken towards solving the problem of looking through layers, and the problem of signal compression in ultrasound imaging. This thesis also provides methods for obtaining this required propagation model using a blind technique (both the propagation model and ultrasound image are unknown) as well as a non-blind technique (a controlled calibration environment with a known image is used to find the propagation model). Blind methods are especially useful for scenarios where characterization of the distorting layer is difficult, such as imaging through the human skull.

 More concretely, the contributions of this thesis are as follows:

1. Demonstrate that a 3D ultrasound image can be reconstructed from a single sensor by spatially encoding the transmitted and received ultrasound field using a simple and easily manufactured plastic coding mask. This is done by carefully measuring the pulse-echo signal for each pixel, and by utilizing model-based imaging techniques. The method is successfully tested on lab measurements of two objects, paving the way for 3D imaging without a fully sampled 2D matrix array, as well as simpler and smaller imaging devices. This contribution is presented in Ch. 3, and has been published as:

   > Kruizinga, P., van der Meulen, P., Fedjajevs, A., Mastik, F., Springeling, G., de Jong, N., ... & Leus, G. (2017). Compressive 3D ultrasound imaging using a single sensor. *Science advances, 3*(12), e1701423.

   Related work to this contribution can also be found in:

   > Janjic, J., Kruizinga, P., van der Meulen, P., Springeling, G., Mastik, F., Leus, G., ... & van Soest, G. (2018). Structured ultrasound microscopy. *Applied Physics Letters, 112*(25), 251901.

   However, we will not elaborate on this paper in this dissertation.

2. Propose and compare various numerical optimization methods for finding the best coding mask geometry for a single sensor setup. A convex-relaxation approach, as well as a greedy algorithm are proposed, and we show that the faster greedy algorithm works just as well as the convex approach, while being more computationally efficient. Both methods come close to finding the globally optimal mask for small problem sizes. We demonstrate that these masks find a trade-off between the amounf of information encoded by the coding mask on the one hand, and SNR on the other. The obtained masks outperform randomly designed masks. This contribution is presented in Ch. 4, and has been published as:

   > van der Meulen, P., Kruizinga, P., Bosch, J. G., & Leus, G. (2019). Coding mask design for single sensor ultrasound imaging. *IEEE Transactions on Computational Imaging, 6*, 358-373.

1

Related work to this contribution can also be found in:

> van der Meulen, P., Kruizinga, P., Bosch, J. G., & Leus, G. (2017). Spatial compression in ultrasound imaging. *In 2017 51st Asilomar Conference on Signals, Systems, and Computers* (pp. 1016-1020). IEEE.

> van der Meulen, P., Kruizinga, P., Bosch, J. G., & Leus, G. (2018, October). Joint optimization of coding mask and scan positions for compressive single sensor imaging. In *2018 IEEE International Ultrasonics Symposium (IUS)* (pp. 1-4). IEEE.

However, we will not elaborate on these papers in this dissertation.

3. A fast and simple measurement method for finding the electro-mechanical transfer function of each transducer in an ultrasound array, using simple pulse-echo measurements instead of using hydrophones or other additional transducers. These are needed for accurate ultrasound simulations, and for more accurate model-based imaging methods. To accomplish this, we propose a de-autoconvolution technique based on a convex relaxation, leading to a semi-definite program, to estimate the transfer functions. This contribution is presented in Ch. 5, and has been published as:

> van der Meulen, P., Kruizinga, P., Bosch, J. G., & Leus, G. (2017, September). Impulse response estimation method for ultrasound arrays. In *2017 IEEE International Ultrasonics Symposium (IUS)* (pp. 1-4). IEEE.

4. A method for finding the transfer function of an aberrating layer in front of an ultrasound array, when the ultrasound image is unknown as well. A pulse-echo measurement model is formulated where the layer's transfer function is an explicit variable to be estimated. It is demonstrated that a second set of measurements with a phase mask gives rise to a simple technique for finding the layer's transfer function. It is shown that the method works successfully for laterally-invariant layers, where the layer's transfer function matrix is Toeplitz, and that it leads to better imaging results compared to imaging without calibration. This contribution is presented in Ch. 6, and has been published as:

> van der Meulen, P., Coutiño, M., Kruizinga, P., Bosch, J. G., & Leus, G. (2021, January). Blind calibration for arrays with an aberration layer in ultrasound imaging. In *2020 28th European Signal Processing Conference (EUSIPCO)* (pp. 1269-1273). IEEE.

5. A second, more stable and flexible method for blindly (i.e., without knowledge of the ultrasound contrast image) estimating the transfer function of an aberrating layer in front of an array. Only a single set of measurements without an additional phase mask is needed. The method measures the echoes of many random uncorrelated images, such as from blood flow. Using covariance matching to estimate the layer's transfer function, a manifold optimization problem is defined, and several techniques are proposed for solving it. The technique is tested on a variety of layers simulated in *k-Wave*, including a skull layer, and simulated blood flow data.

Using the obtained transfer function allows one to obtain better images when e.g. a skull-layer is placed in front of the array, by incorporating the layer's transfer function in a model-based imaging algorithm. This contribution is presented in Ch. 7, and at the time of writing has been submitted for publication as:

> van der Meulen, P., Coutiño, M., Kruizinga, P., Bosch, J. G., & Leus, G. (2022). Ultrasonic imaging through aberrating layers using covariance matching. *IEEE Transactions on Computational Imaging*, submitted for publication.

Related work to this contribution can also be found in:

> van der Meulen, P., Kruizinga, P., Bosch, J. G., & Leus, G. (2018, October). Calibration techniques for single-sensor ultrasound imaging with a coding mask. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (pp. 1641-1645). IEEE.

However, we will not elaborate on this paper in this dissertation.

6. Finally, the following contribution has been co-authored by the doctoral candidate, but is not directly related to this thesis, and will not be further discussed.

> Zangabad, R. P., Iskander-Rizk, S., van der Meulen, P., Meijlink, B., Kooiman, K., Wang, T., ... & van Soest, G. (2021). Photoacoustic flow velocity imaging based on complex field decorrelation. *Photoacoustics, 22,* 100256.

The remainder of this work is organized as follows. First, in the next chapter the basics of ultrasound imaging are briefly put in place, both from a physical point of view, as well as a mathematical point of view. This mathematical view of imaging is used throughout this thesis, as opposed to the traditional physics-based approach. It also contains the central equation related to the model-based imaging framework. In the same chapter, we will then briefly explain the main research questions coming in the subsequent chapters from this mathematical point of view, and shortly discuss the main mathematical problems considered in each chapter. Chapters 3 through 7 contain the main research results as described in the list above, and is divided into two parts. The **first part** considers imaging with a single sensor using a coding mask to distort ultrasound fields in an informative way, as well as the task of finding the best coding mask. In these chapters, the propagation model, including the effect of the coding mask, is obtained experimentally prior to imaging, or assumed to be known otherwise. In the **second part**, we consider scenarios where the propagation model is not fully known, and cannot be easily or non-invasively obtained experimentally. We propose two methods for blindly finding the effect of an aberration layer, and a method for finding the electro-mechanical impulse response of an ultrasound transducer in an array. Fig. 1.1 summarizes the two parts of this dissertation visually. Finally, Chapter 8 summarizes and concludes the research in this thesis.

REFERENCES

[1] N. Bom, C. T. Lancée, G. van Zwieten, F. E. Kloster, and J. Roelandt, *Multiscan echocardiography: I. technical description,* Circulation **48**, 1066 (1973).

1
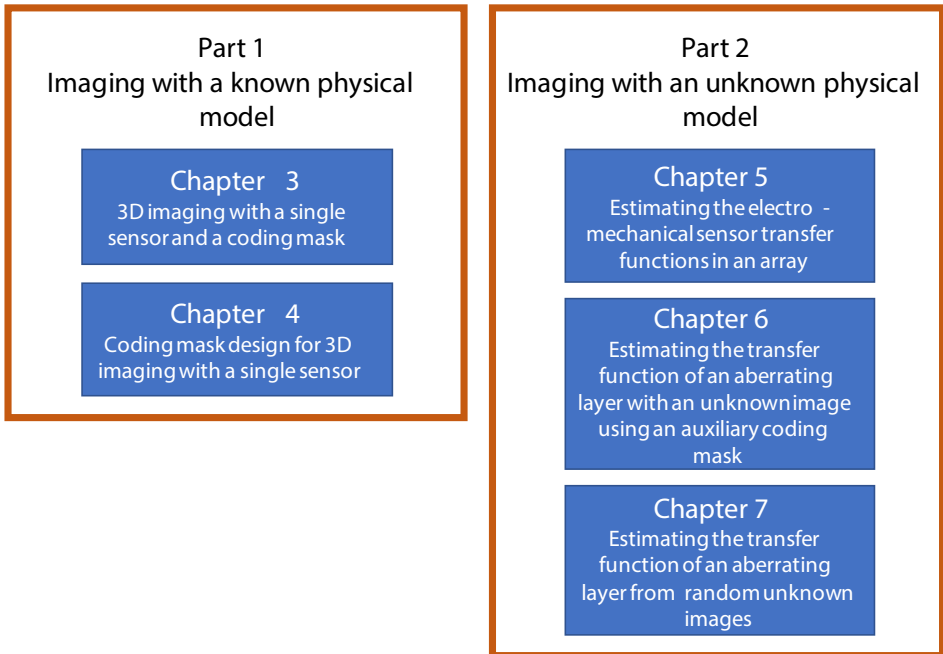
Figure 1.1: Chapter overview of this dissertation.

[2] F. E. Kloster, J. Roelandt, F. J. t. Cate, N. Bom, and P. G. Hugenholtz, *Multiscan echocardiography: II. technique and initial clinical results,* Circulation **48**, 1075 (1973).

[3] K. E. Thomenius, *Evolution of ultrasound beamformers,* in *1996 IEEE Ultrasonics Symposium. Proceedings,* Vol. 2 (IEEE, 1996) pp. 1615–1622.

[4] J. Janjic, *Looking Forward with Minimally Invasive Ultrasound,* Ph.D. thesis (2018).

[5] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,* IEEE Transactions on information theory **52**, 489 (2006).

[6] E. J. Candès and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?* IEEE transactions on information theory **52**, 5406 (2006).

[7] D. L. Donoho, *Compressed sensing,* IEEE Transactions on information theory **52**, 1289 (2006).

[8] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling,* IEEE signal processing magazine **25**, 21 (2008).

[9] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling,* IEEE signal processing magazine **25**, 83 (2008).

[10] H. Liebgott, R. Prost, and D. Friboulet, *Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing,* Ultrasonics **53**, 525 (2013).

[11] C. Quinsac, A. Basarab, and D. Kouamé, *Frequency domain compressive sampling for ultrasound imaging,* Advances in Acoustics and Vibration **2012** (2012).

[12] N. Wagner, Y. C. Eldar, and Z. Friedman, *Compressed beamforming in ultrasound imaging,* IEEE Transactions on Signal Processing **60**, 4643 (2012).

[13] T. Chernyakova and Y. Eldar, *Fourier-domain beamforming: the path to compressed ultrasound imaging,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **61**, 1252 (2014).

[14] M. F. Schiffner and G. Schmitz, *Fast pulse-echo ultrasound imaging employing compressive sensing,* in *Ultrasonics Symposium (IUS), 2011 IEEE International* (IEEE, 2011) pp. 688–691.

[15] G. David, J.-l. Robert, B. Zhang, and A. F. Laine, *Time domain compressive beam forming of ultrasound signals,* The Journal of the Acoustical Society of America **137**, 2773 (2015).

[16] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, *Compressed delay-and-sum beamforming for ultrafast ultrasound imaging,* in *Image Processing (ICIP), 2016 IEEE International Conference on* (Ieee, 2016) pp. 2509–2513.

[17] S. P. Chepuri and G. Leus, *Sparsity-Promoting Sensor Selection for Non-Linear Measurement Models.* IEEE Trans. Signal Processing **63**, 684 (2015).

[18] J. T. Fokkema and P. M. van den Berg, *Seismic applications of acoustic reciprocity* (Elsevier, 1993).

[19] M. Verweij, B. Treeby, K. Van Dongen, and L. Demi, *Simulation of ultrasound fields,* Comprehensive biomedical physics , 465 (2014).

[20] P. R. Stepanishen, *Pulsed transmit/receive response of ultrasonic piezoelectric transducers,* The Journal of the Acoustical Society of America **69**, 1815 (1981).

[21] J. A. Jensen, *A model for the propagation and scattering of ultrasound in tissue,* The Journal of the Acoustical Society of America **89**, 182 (1991).

[22] J. A. Jensen and N. B. Svendsen, *Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **39**, 262 (1992).

1

[23] S. Soloukey, A. J. Vincent, D. D. Satoer, F. Mastik, M. Smits, C. M. Dirven, C. Strydis, J. G. Bosch, A. F. van der Steen, C. I. De Zeeuw, *et al.*, *Functional ultrasound (fus) during awake brain surgery: The clinical potential of intra-operative functional and vascular brain mapping,* Frontiers in neuroscience **13**, 1384 (2020).

[24] R. Rau, P. Kruizinga, F. Mastik, M. Belau, N. de Jong, J. G. Bosch, W. Scheffer, and G. Maret, *3D functional ultrasound imaging of pigeons,* Neuroimage **183**, 469 (2018).

[25] J. Bercoff, G. Montaldo, T. Loupas, D. Savery, F. Mézière, M. Fink, and M. Tanter, *Ultrafast compound doppler imaging: Providing full blood flow characterization,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **58**, 134 (2011).

[26] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, *Functional ultrasound imaging of the brain,* Nature methods **8**, 662 (2011).

[27] L. Nock, G. E. Trahey, and S. W. Smith, *Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,* The Journal of the Acoustical Society of America **85**, 1819 (1989).

[28] M. Karaman, A. Atalar, H. Koymen, and M. O'Donnell, *A phase aberration correction method for ultrasound imaging,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **40**, 275 (1993).

[29] M. O'donnell and S. Flax, *Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **35**, 768 (1988).

[30] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu, *Time-reversed acoustics,* Reports on progress in Physics **63**, 1933 (2000).

[31] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, *Optimal focusing by spatio-temporal inverse filter. II. experiments. application to focusing through absorbing and reverberating media,* The Journal of the Acoustical Society of America **110**, 48 (2001).

[32] L. Guasch, O. C. Agudo, M.-X. Tang, P. Nachev, and M. Warner, *Full-waveform inversion imaging of the human brain,* NPJ digital medicine **3**, 1 (2020).

[33] F. Vignon, J. Aubry, M. Tanter, A. Margoum, and M. Fink, *Adaptive focusing for transcranial ultrasound imaging using dual arrays,* The Journal of the Acoustical Society of America **120**, 2737 (2006).

2

MODEL-BASED ULTRASOUND IMAGING

I N this chapter we introduce the basics of model-based ultrasound imaging as opposed to conventional beamforming. Based on the linear equation central to this imaging method, we then outline each chapter in this thesis, and show how various US imaging problems can be expressed mathematically using this central equation. This then naturally leads to various mathematical (optimization) problems, which we try to solve in the various chapters, leading to solutions for compressive imaging and blind calibration.

NOTATION

In this chapter, a vector $\mathbf{x}$ is represented using lower-case bold symbols, and a matrix $\mathbf{X}$ is represented by upper-case bold symbols. The operator $(.)^{\dagger}$ represents the Moore-Penrose pseudo-inverse operator. The matrix transpose and conjugate transpose are denoted by the operators $(.)^{\mathsf{T}}$ and $(.)^{\mathsf{H}}$, respectively. The statistical expectation operator is denoted as $\mathbb{E}\{.\}$. Finally, the l_2-norm of a vector $\mathbf{x}$ is denoted as $\|\mathbf{x}\|_2$, the l_1-norm of a vector $\mathbf{x}$ as $\|\mathbf{x}\|_1$, and the Frobenius norm of a matrix $\mathbf{X}$ as $\|\mathbf{X}\|_F$.

2.1. TRADITIONAL ULTRASOUND BEAMFORMING

Classical ultrasound imaging is based on the technique of delay-and-sum (DAS) beamforming. Since ultrasound waves travel with a more or less constant speed of sound throughout the body, one can compute at what time an echo from a specific location would be measured on each sensor in an array. One could then simply sum the signal values of the envelope of the echo signals across all sensors at those computed arrival times into a single value, and it should result in a relatively large number, since those signals should sum coherently. Coherent summation is possible since the same pulse is transmitted to (and reflected by) all pixels, and since we can pre-compute the time-of-arrival of the potential echo of each pixel. Conversely, if we use the same operation for a location where no scattering is present, the summation would be a relatively low
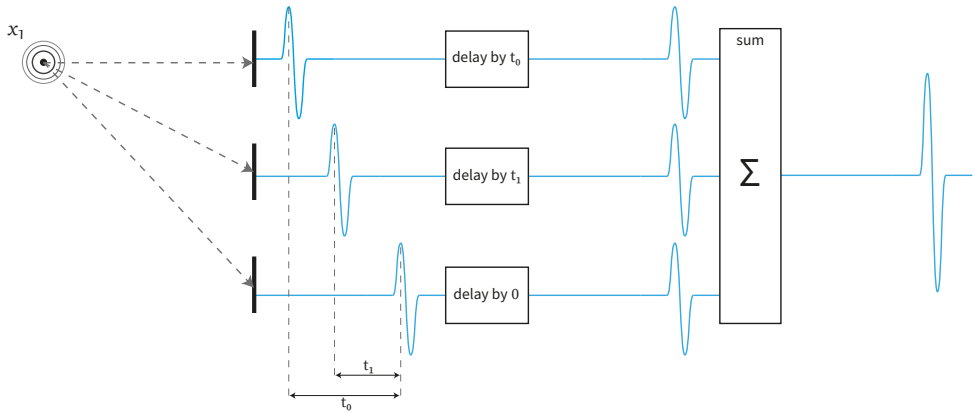
Figure 2.1: Delay-and-sum beamforming works by delaying all channels based on the time-of-arrival of the pixel being beamformed. Signals from that pixel will sum coherently if an echo is present. Applying the same delays to a set of echoes originating from an echo at a different location will not result in coherent summation.

number. DAS beamforming thus consists of repeating this operation for all pixels, after which an image can be formed where high intensities correspond to scattering activity. Fig. 2.1 contains a visual explanation of the DAS beamforming concept.

Various transmit schemes are used in medical ultrasound imaging in combination with DAS beamforming. The most basic scheme focuses energy in one direction, and then beamforms for only that angular sector. This is then repeated until all sectors are covered, and as such requires multiple transmit events in time before a single image can be formed. In order to increase framerate, one could consider to transmit a diverging wave (or other limited diffracted wave) using e.g. a single transmitter, that illuminates the entire ROI instead of a single angular sector. This comes at a loss of image quality since the signal-to-noise ratio (SNR) is lower for all pixels, and thus multiple of such images are often compounded into a single, high SNR image (where each image is obtained by transmitting from a different sensor in the transducer array [1]). A trade-off between SNR and framerate has been found using coherent plane-wave compounding [2], where all sensors transmit in such a way that a plane wave illuminates the entire ROI, at a higher SNR than using a diverging wave. Multiple of these plane wave images with different plane wave angles are coherently compounded to obtain a good image at a very high frame rate. Note that, for all these techniques, the time-of-arrival of an echo is known for each pixel, and so DAS beamforming can be applied.

The DAS approach relies on the fact that in most medical imaging situations waves can travel uninterruptedly with only small inhomogeneities acting as scatterers, so that the background medium speed of sound can be assumed more or less constant (so that travel times can be computed), and it also relies on the fact that more or less the same pulse waveform is received on each sensor in an array from each single reflection (so that coherent summation is possible). In this dissertation we are interested in scenarios where an aberrating layer is placed in front of an ultrasound transducer. In compli-

cated layers like the skull, reflected waves will not directly travel towards the sensors, but instead scatter and reverberate inside the layer, before reaching the sensor array. Moreover, the echo pulse shape of each pixel can be different, since the incident transmit wave varies spatially across pixels. It invalidates the assumptions used by DAS beamforming, which can no longer reconstruct images in such a scenario.

A more thorough approach for these kinds of scenarios would take into account that an echo can have a non-typical time-of-arrival compared to a normal imaging situation, and, more importantly, may have a much different waveform than the pulse originally transmitted into the region of interest (ROI). The received echo waveform could even strongly vary with pixel location, providing additional information not exploited using DAS beamforming. Instead of delaying channels followed by coherent summation, one could then use a more sophisticated imaging technique. If we know what the potential echo signal of each pixel in the ROI looks like, we can use simple least squares approaches to find the image that best explains the measured echo signals, taking into account this prior knowledge. Of course, one has to be able to estimate what the echo of each pixel would look like, which requires extensive and accurate simulations, or could be accomplished by directly measuring these signals experimentally.

We will refer to this approach of ultrasound imaging as *model-based ultrasound imaging*, and it will be shortly explained in this chapter. Using this framework, each chapter in this dissertation will then be explained based on the measurement equation central to this approach.

2.2. BASICS OF MODEL-BASED IMAGING

We consider the scenario where a known pressure pulse is transmitted into the region of interest (ROI), with a homogeneous background medium, and no aberrating layers (Fig. 2.2). Scattering of the pressure pulse occurs when pixels in the ROI have a contrasting speed of sound and/or density compared to its surrounding medium. Since these contrasts are relatively small in biomedical imaging, we can safely ignore multiple reflections between the scatterers. This approximation is known as the *Born approximation* [3, 4], and greatly simplifies the relation between the contrast image and measurement data (in fact, it becomes a *linear* relation). The output signal on a receiving sensor is now a linear superposition of the echo of each individual pixel, since we can assume an echo is not strongly changed by other reflectors, and thus does not generate additional echoes with enough energy to invalidate the Born approximation. Hence, the output signal on sensor m related to pixel n, $a_{m,n}(t)$, can be pre-computed for each of the N pixels. Let x_n be the scattering intensity of the n-th pixel (which is zero in case of no scattering), then the measured output signal $y_m(t)$ on sensor m is equal to

$$y_m(t) = \sum_{n=1}^{N} a_{m,n}(t) x_n + n_m(t), \tag{2.1}$$

where $n_m(t)$ represents additive measurement noise. Since all $a_{m,n}(t)$ are known *beforehand* (we know the pulse transmitted into the ROI, and at what time an echo would be measured for each pixel), the only unknowns are the coefficients x_n. Imaging thus becomes a matter of solving the linear equation (2.1) for x_n. This relation is more compactly denoted using a matrix-vector formalism. After sampling the continuous signals
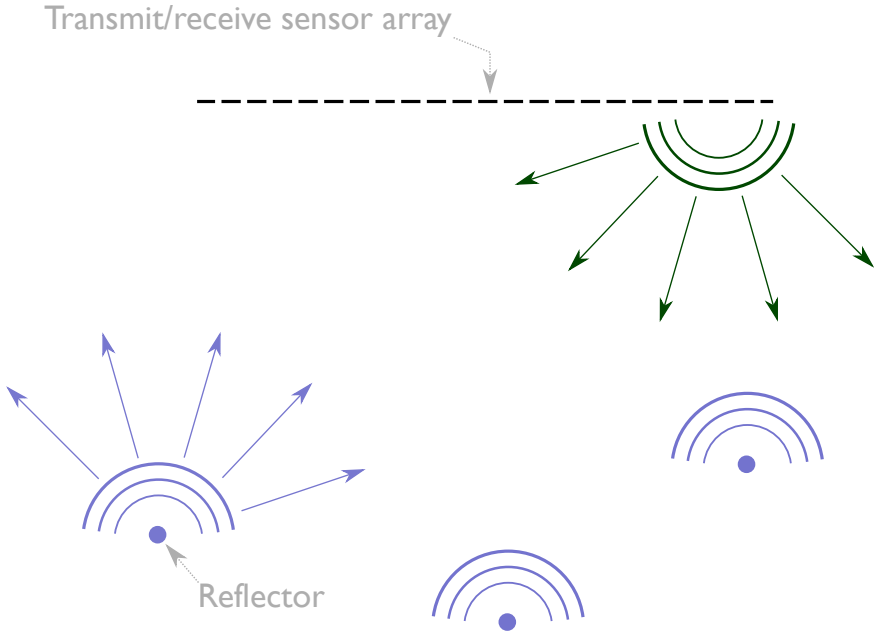
Figure 2.2: A typical ultrasound imaging scenario. One or more elements in the array transmit a pulse into the ROI. Acoustic contrasts will cause echoes of the incident pulse, which are reflected back towards the transducer array. To obtain an image, it is assumed secondary reflections are negligible. Since the speed of sound of the background medium is known, it can be pre-computed when the potential echo of each pixel should arrive, which is exploited by e.g. DAS beamforming.

$y_m(t)$ and $a_{m,n}(t)$, their L samples can be stored in the vectors $\mathbf{y}_m \in \mathbb{C}^L$ and $\mathbf{a}_{m,n} \in \mathbb{C}^L$. For the sake of generality, $\mathbf{y}$ and $\mathbf{a}_{m,n}$ can be complex for various reasons: e.g., one does image reconstruction in the frequency domain, or the measured signals have been shifted to baseband to obtain an IQ-signal. However, the relationship between the measurements, model, and unknown image remains linear, and so everything discussed in this chapter applies to those situations. We can now write the vectorized version of (2.1), and stack them for all M sensors into a single equation:

$$
\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1,1} & \mathbf{a}_{1,2} & \ldots & \mathbf{a}_{1,N} \\ \mathbf{a}_{2,1} & \mathbf{a}_{2,2} & \ldots & \mathbf{a}_{2,N} \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_{M,1} & \ldots & \ldots & \mathbf{a}_{M,N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_M \end{bmatrix},
\tag{2.2}
$$

which we shortly denote as

$$
\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}.
\tag{2.3}
$$

To obtain an estimated image $\mathbf{x}$ from $\mathbf{y}$, the following optimization problem can be

solved, to find the image $\mathbf{x}$ that best explains the measurements $\mathbf{y}$, given the model $\mathbf{A}$:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2. \tag{2.4}$$

Unfortunately, $\mathbf{A}$ is typically an ill-posed matrix, so that a small change in $\mathbf{y}$, e.g., due to measurement noise or model errors, can cause a wildly different solution $\mathbf{x}$. For this reason, one typically adds a penalty term to the cost function above:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda f(\mathbf{x}), \tag{2.5}$$

where the scalar parameter λ determines the trade-off between the penalty function and the model error. Although model-based imaging takes into account more information using the matrix $\mathbf{A}$, one now has to choose an appropriate regularizer function $f(\mathbf{x})$ and tune the regularization weight parameter λ [5]. A typical example is the Tikhonov regularizer $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$, which penalizes large values in $\mathbf{x}$. As a result, there is a unique and stable solution, at the cost of smoothing $\mathbf{x}$, causing a loss in resolution. In the field of compressive sensing [6–9], it was discovered that the the l_1 regularizer $f(\mathbf{x}) = \|\mathbf{x}\|_1$ can exactly reconstruct $\mathbf{x}$ if the true $\mathbf{x}$ is known to be sparse, and under certain conditions on $\mathbf{A}$. The regularizer function can therefore be used to exploit prior information about the solution $\hat{\mathbf{x}}$. Since model-based imaging is computationally intensive compared to DAS, it is often not used for typical imaging scenarios with no aberration layer present. DAS is computationally more efficient, does not require the pre-computation or storage of a large matrix, and typically leads to comparable resolution.

An imaging method that does not fully solve (2.3), but is more computationally efficient, is the estimated solution

$$\hat{\mathbf{x}} = \mathbf{A}^H \mathbf{y}, \tag{2.6}$$

known as the matched filter estimator in communication and radar literature. If $\mathbf{A}^H\mathbf{A}$ is approximately equivalent to the identity matrix, it is straightforward to see that the estimator approximately solves (2.3). This estimator (2.6) acts as a correlator, since it takes the inner product of each column of $\mathbf{A}$ with $\mathbf{y}$: it checks how similar the expected echo signals for each pixel are compared to the measurement $\mathbf{y}$ by computing the unnormalized correlation coefficient through the inner product of the two vectors. As pointed out in [10], for typical ultrasound imaging scenarios without an abberating layer this reduces to the DAS beamformer if one has no knowledge about the ultrasound pulse transmitted into the imaging medium. If the pulse is unknown, one would assume a delta-pulse $\delta(t)$ is transmitted and reflected by scatterers. Thus, each vector $\mathbf{a}_{m,n}$ in (2.1) consists of zeros and has unity magnitude on the sample at which an echo is expected to arrive for pixel n at sensor m. The matched filter would thus simply sum M values from $\mathbf{y}$ (since each column in $\mathbf{A}$ has only M non-zero entries) for each pixel, namely the samples in $\mathbf{y}$ at which an echo is expected at sensor m. Denoting the n-th column vector of $\mathbf{A}$ as $\mathbf{a}_n$, using $\tau_{m,n}$ as the expected time-of-arrival for an echo from pixel n on sensor m, and using the DAS matrix $\mathbf{A}_{\mathrm{DAS}}$ as just described, we have for the corresponding matched filter

estimate:

$$\hat{x}_{n,\text{DAS}} = \mathbf{a}_n^{\mathsf{H}} \mathbf{y} \tag{2.7}$$

$$= \sum_{m=1}^{M} \mathbf{a}_{m,n}^{\mathsf{H}} \mathbf{y}_m \tag{2.8}$$

$$= \sum_{m=1}^{M} y_m[\tau_{m,n}], \tag{2.9}$$

where $y_m[\tau]$ represents the signal y at sensor m at sample τ. This is thus equivalent to delay-and-sum beamforming.

The model-based imaging approach has been applied to (medical) ultrasound imaging without abberrating layers already in the 1990's [11, 12], and in various other works throughout time [13, 14]. Recently, it is of interest to researchers due to the increased amount of computational resources and in the context of compressive sensing [10, 15]. The model-based imaging approach becomes more interesting for scenarios where one can no longer beamform based purely on time-of-arrival information. In contrast, model-based imaging is able to reconstruct images in non-conventional scenarios where waves are strongly distorted, but the Born approximation still holds. At the other end of the spectrum, as opposed to DAS-based beamforming, there is full waveform inversion (FWI), an imaging algorithm that gained immense popularity in seismic imaging [16, 17]. Instead of using the Born approximation, it tries to solve the acoustic wave equations using gradient descent techniques. Although this is not a convex optimization problem, very good results are obtained. This method is in the same spirit of model-based imaging, since it does not make any assumptions on the background medium wave speed, transmitted pulses, and even foregoes the Born approximation. However, this comes at the cost of much larger computational times, which are tolerable in seismic imaging since real-time imaging is not required. With model-based imaging we try to find a balance between the computational imaging burden and imaging quality, where DAS and FWI would form the two outer extremes on this scale. There are many CPU and memory efficient algorithms to solve (2.5) available for various regularizer functions, and there is still an active research community for (efficiently) solving ill-posed linear inverse problems.

In this thesis we discuss a few scenarios where model-based imaging is required, the common theme being that of imaging through aberrating layers located just in front of an (array of) ultrasound sensor(s). These scenarios will be briefly discussed in the remainder of this chapter, where we will give an extended outline of this thesis based on the preceding equations. Each chapter discussed below is based on a corresponding research paper, which is mentioned at the beginning of each chapter.

2.3. DISSERTATION OUTLINE

2.3.1. 3D ULTRASOUND IMAGING WITH A SINGLE SENSOR AND A CODING MASK

In Chapter 3 we show a first example of a scenario where DAS is no longer useful, and the use of model-based imaging is required. It considers a large single transducer, equipped with a coding mask, capable of reconstructing 3D images using only a single sensor. The
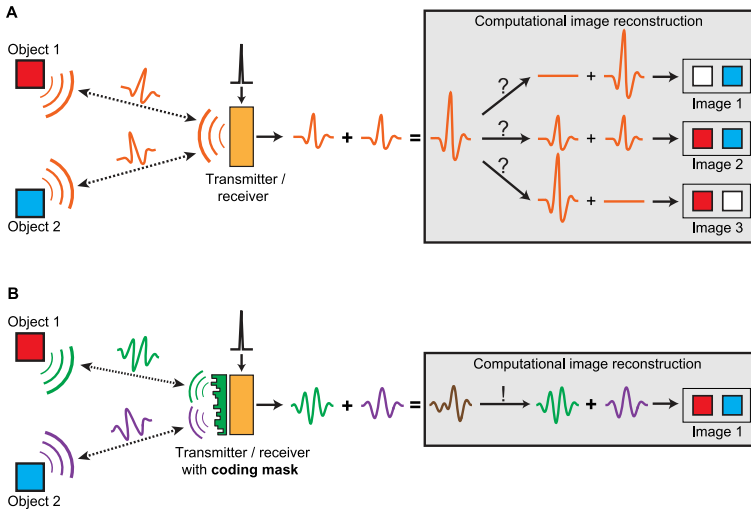
Figure 2.3: Ultrasound imaging with a single sensor and a coding mask (Chapter. 3). In **(A)** no coding mask is used. The echo signal of each scatterer at the same range will look the same, and arrive at the same time, denying the possibility of resolving two or more scatterers. In **(B)** a coding mask distorts waves in such a way that a different echo signal is received from each scatterer. Since we can characterize the echo of each pixel using a calibration procedure, we known them beforehand, and the two scatterers can now be resolved based on their pulse-shape.

transmitted ultrasound pulses travel through the coding mask, which strongly distorts the transmitted ultrasound field, so that a different pulse is scattered toward each pixel (Fig. 2.3). In this case the DAS algorithm would not be able to reconstruct an image, since there is not an array of sensors to exploit time-of-arrival differences between sensors. With model-based imaging, however, we can experimentally measure the response of each pixel, use those measurements to construct **A**, and solve (2.4) to successfully obtain an estimate of a contrast image. Not only does this demonstrate the usefulness of model-based imaging, but it also shows it can play a role in reducing the number of sensors needed for 3D imaging, and furthermore paves the way for miniaturized ultrasound devices for e.g. minimally-invasive intravascular imaging applications. Only a single sensor is needed, strongly reducing data output and the amount of cables needed to transfer measurement data.

2.3.2. Optimizing the geometry of a coding mask for single sensor imaging

Building on the work presented in Chapter. 3, we try to find the aberrating mask that enables the best imaging performance in Chapter. 4. We specifically consider the type of coding masks used in Chapter. 3: a homogeneous plastic disc with varying thickness to create local phase changes, resulting in a coded echo field (Fig. 2.4). In order to design these pressure fields, an approximate model for the output field based on the local mask thicknesses is utilized. Similarly structured phase masks and signal models have also
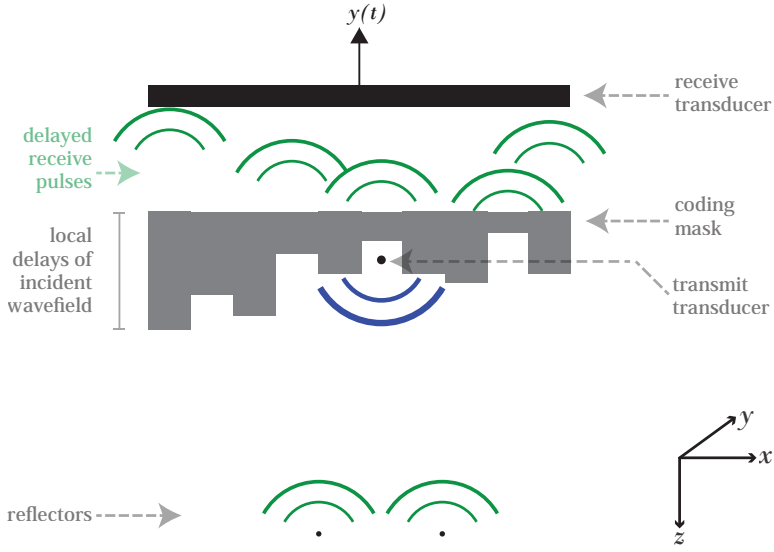
Figure 2.4: The coding mask geometry used in Chapter. 4. It consists of a homogeneous material with a different speed of sound than the ROI medium. Due to the different heights of the pillars, echo wavefields are distorted, enabling distinction between scatterers at the same range. In Chapter. 4 we assume a separate transmitter is used to illuminate the ROI.

been used to design arbitrary pressure fields by other researchers [18, 19]. Each pillar in the mask is seen as an independent channel that delays the incident wavefield based on its length (Fig. 2.5). In our work, we use this approximate model to pre-compute the **A**-matrix. The resulting **A**-matrix can then be used to predict how well images can be reconstructed from measurements obtained with the corresponding coding mask.

To this end, we assume the scattering coefficients in **x** are normally distributed with known covariance matrix $\mathbf{C_x}$. We fill further assume that a Wiener estimator is used to estimate **x**, which takes into account the covariance matrix of the distribution of **x**. Under this assumption, we can compute the error covariance matrix when using the Wiener estimator [20] as

$$\mathbf{C_e} = \mathbb{E}\left((\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathsf{H}}\right) \tag{2.10}$$

$$= \left(\mathbf{C_x}^{-1} + \mathbf{A}^{\mathsf{H}}\mathbf{C_n}^{-1}\mathbf{A}\right)^{-1}, \tag{2.11}$$

where $\mathbf{C_n}$ is the known noise covariance matrix, and $\hat{\mathbf{x}}$ represents the estimated image. After discretizing the mask surface into pillars, we collect the heights of all pillars in the ordered set $\mathcal{W}$. If $\mathcal{W}$ is known, we can also compute **A** based on the aforementioned approximate delay-based model. Thus, **A** becomes parameterized by $\mathcal{W}$, which we will denote as $\mathbf{A}(\mathcal{W})$. We will try to find an optimized mask by minimizing the mean squared error (MSE) under additive zero-mean i.i.d. white Gaussian measurement noise, which is a function of $\mathbf{C_e}$. That is, we will use the sum of the error variances over all pixels.
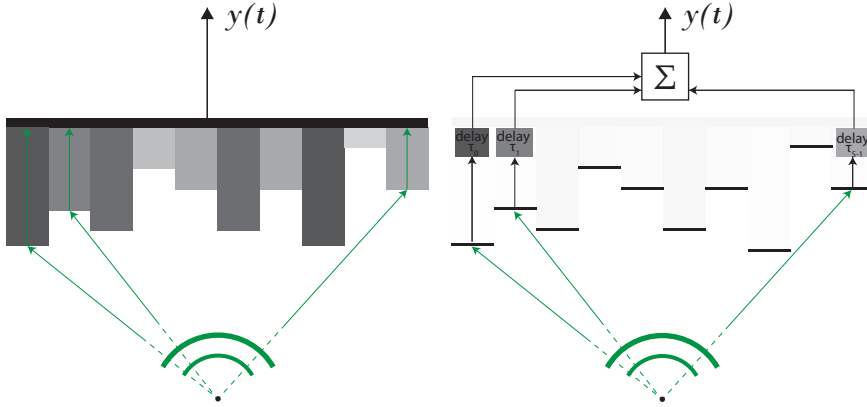
Figure 2.5: The approximate model used to predict the echo signal for each pixel in Chapter 4. Each mask pillar is seen as an independent channel which delays the local part of the incoming wavefront based on its length and mask speed of sound. The final output signal $y(t)$ is the integral of all delayed channel signals. We optimize the mask by finding the best length for each pillar, such that echo signals from different pixels are resolvable.

Substituting $\mathbf{A}(\mathcal{W})$ into (2.11), we thus want to minimize the function

$$f(\mathcal{W}) = \text{trace}(\mathbf{C_e}) \tag{2.12}$$

$$= \text{trace}\left(\left(\mathbf{C_x}^{-1} + \mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{C_n}^{-1}\mathbf{A}(\mathcal{W})\right)^{-1}\right), \tag{2.13}$$

with respect to $\mathcal{W}$. This results in a discrete optimization problem, for which we propose a convex relaxation approach, as well as a greedy discrete optimization approach. We are able to show that the optimized masks come close to the true optimal mask for small problem sizes. For realistic scenarios, we observe that the optimized masks try to focus energy in the ROI, increasing SNR, while at the same time maximizing decorrelation between echo signals (i.e., columns of $\mathbf{A}$) of neighbouring pixels.

2.3.3. IMPROVING THE ACCURACY OF THE SCATTERING MODEL BY ESTIMATING THE TRANSMITTED ULTRASOUND PULSE

In Part 2 of this thesis, we consider how to increase the accuracy of the $\mathbf{A}$-matrix. It goes without saying that errors in $\mathbf{A}$ will lead to imaging artifacts in estimates of $\mathbf{x}$. This ranges from a (slight) loss of resolution for imaging through (subcutaneous) fat layers, to completely incomprehensible images when imaging through bone tissue. In case of the latter, such as the human skull, waves are distorted so strongly that the $\mathbf{A}$ matrix typically used for free propagation of waves in a homogeneous medium no longer corresponds to reality, resulting in catastrophic image reconstructions when using a 'normal' $\mathbf{A}$ matrix.

In Chapter 5, we start with finding the electro-mechanical transfer function of each element in an ultrasound array, so that it can be included in $\mathbf{A}$, potentially leading to improved imaging results. Typically, one controls the pulse that is sent to a transmitter. This is not necessarily the same pulse that will propagate inside the ROI, mostly

due to the electro-mechanical transfer function of the transducer that is applied on both transmit and receive. The combination of the transmit pulse and the application of the electro-mechanical transfer function on both transmit and receive form the actual measurements in the data vector **y**, and thus needs to be included in **A**.

One way to measure the forward field of a transducer would consist of exciting it with a single δ-like pulse, and using a hydrophone to measure the actual transmitted pulse. The pulse-echo impulse response for each combination of transmit and receive elements in an array can then be found by convolving the impulse responses of the corresponding array elements. However, hydrophones are expensive, measurements are sensitive to the position of the hydrophone, and the procedure becomes laborious for large arrays.

For that reason, a method is presented where a flat plate reflector is placed in front of the transducer array, which acts as a mirror. In that case, one would measure a delayed version of the impulse response $h(t)$ convolved with itself, repeated for each array element. The problem considered in this chapter is how to estimate $h(t)$ from the measurement signal $x(t) = h(t) * h(t)$. Repeating this simple procedure for each element in an ultrasound array would then give the one-way electro-mechanical impulse response of each transducer. We present a de-autoconvolution algorithm using semi-definite programming, exploiting the circular convolution property of the DFT transform of $x(t)$.

2.3.4. BLIND ESTIMATION OF THE EFFECT OF AN ABERRATING LAYER
In the remainder of Part 2, we are interested in finding the **A** matrix without measuring the pulse-echo signal of each pixel explicitly. We call this a '*blind*' calibration approach, since neither **x** nor **A** is known. In Chapters 6 and 7, we propose two techniques for blind calibration when there is an aberrating layer placed between the ultrasound array and the ROI (Fig. 2.6). The first one based on obtaining two measurements using an additional auxiliary coding mask is discussed in Chapter. 6, and a second, more flexible and stable method based on covariance matching [21] is presented in Chapter. 7.

The first step towards solving this problem is the insight that it is well known how waves propagate between the aberration layer and pixels in the ROI. Suppose a non-aberrated pulse is transmitted into the ROI with a homgeneous background medium with a known speed of sound, and propagates to each pixel. Acoustically contrasting pixels will scatter the pulse towards the aberrating layer. The component in the **A** matrix covering the propagation path up until this point is known, and is easily computed. The only unknown factor, then, is the propagation from the 'input'-field on the virtual array in Fig. 2.6 towards the true sensor array. Even though there is scattering and reverberation inside the layer, we can still assume the propagation is *linear*, ignoring non-linear effects such as absorption. In other words, there is a Green's function from each point on the virtual array to each sensor in the actual array. Consequently, the 'output'-field of the aberrating layer as measured by the true array, is a linear combination of the entire input field on the virtual array.

This enables us to define a new measurement equation, where the matrix **A** now describes the propagation from a (point-like) transmitter that does not suffer from aberration, towards pixels, and then towards the virtual array. We further define the matrix **H**, which defines the linear operation exercised by the aberrating layer on input fields in
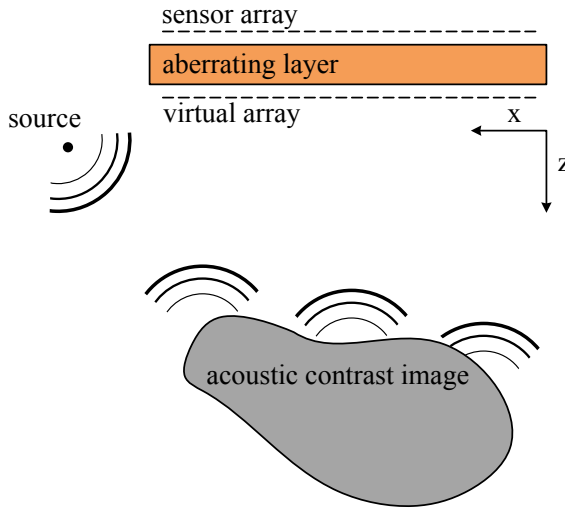
Figure 2.6: The imaging scenario considered in Chapters 6 and 7. An aberrating layer is present in front of the ultrasound array. Its effect on incoming wavefronts is seen as a linear transformation from the virtual array just in front of the layer, toward the true array sensors. In these chapters, we present algorithms for estimating this linear transformation without knowing the acoustic contrast image **x**.

the virtual array. We then obtain the following measurement equation, which is used in both Chs. 6 and 7:

$$\mathbf{y} = \mathbf{HAx} + \mathbf{n}. \tag{2.14}$$

Thus, we have separated the complete **A**-matrix into known and unknown propagation components. This strongly reduces the amount of unknowns, especially since we can typically expect some kind of structure in **H** that can be exploited. Having obtained this equation, the question is: how to estimate **H** and **x**, knowing **A**, and measuring **y**? This is a bi-linear measurement equation, and these are typically hard to solve.

In Chapter 6 we make a first attempt at blind calibration. There, we assume that a set of P images $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_P]$ can be measured, so that we have a first set of measurements

$$\mathbf{Y}_1 = \mathbf{HAX}. \tag{2.15}$$

Next, we use a phase mask (similar to the one used in Chs. 3 and 4), that locally changes the phase of the 'input' field placed directly in front of the aberrating layer. Using the same set of images, we then obtain a second set of measurements:

$$\mathbf{Y}_2 = \mathbf{H\Theta AX}, \tag{2.16}$$

where Θ is a diagonal matrix containing the phase shift per channel of the phase mask. Now, the difference between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ can be exploited by multiplying the pseudo-

inverse of one with the other. Defining $\mathbf{Z} = \mathbf{Y}_2\mathbf{Y}_1^{\dagger}$, we have:

$$\mathbf{Z} = \mathbf{Y}_2\mathbf{Y}_1^{\dagger} \tag{2.17}$$

$$= \mathbf{H}\Theta\mathbf{A}\mathbf{X}(\mathbf{A}\mathbf{X})^{\dagger}\mathbf{H}^{-1} \tag{2.18}$$

$$= \mathbf{H}\Theta\mathbf{H}^{-1}. \tag{2.19}$$

Since Θ is a diagonal matrix, $\mathbf{H}$ can be obtained as the eigenvector matrix of $\mathbf{Z}$. Of course, $\mathbf{H}$ is only known up to a column permutation and a scaling per column, which can be resolved by using prior knowledge on the structure of $\mathbf{H}$, and using the fact that Θ is known. Moreover, it is required that $\mathbf{H}$ is invertible, that the second set of measurements can be obtained using the same $\mathbf{X}$ as in the first set of measurements, and that $\mathbf{A}\mathbf{X}$ is full row-rank. Furthermore, ill-posedness of $\mathbf{Y}_1$ can make the method sensitive to noise in $\mathbf{Y}_2$.

In Chapter 7 we propose a more flexible and stable technique for estimating $\mathbf{H}$, compared to the previous technique. It does not require two measurements, nor the use of an additional phase mask. It exploits the second-order statistics of $\mathbf{y}$ instead of working with the raw data directly. A single set of measurements are obtained, which are viewed as realizations of a random process with covariance matrix $\mathbf{C_x}$. In that case, the covariance matrix $\mathbf{C_y}$ of $\mathbf{y}$ is defined as:

$$\mathbf{C_y} = \mathbb{E}\{\mathbf{yy}^{\mathsf{H}}\} = \mathbb{E}\{\mathbf{H}\mathbf{A}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\} + \mathbb{E}\{\mathbf{nn}^{\mathsf{H}}\} \tag{2.20}$$

$$= \mathbf{H}\mathbf{A}\mathbf{C_x}\mathbf{A}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + \mathbf{C_n}, \tag{2.21}$$

where $\mathbf{C_n}$ is the noise covariance matrix. As a result of working with covariance data, we no longer need to know $\mathbf{x}$, but only the covariance matrix of $\mathbf{x}$, which is a much less demanding requirement. For example, one could obtain many measurements of blood flowing inside a patient, assuming there is enough time between measurements such that images are uncorrelated, and thus $\mathbf{C_x} = \mathbf{I}$. Blood flow is easily separated from static scattering in a set of measurements by applying a high-pass filter on each sample across the frame-dimension.

Using covariance matching [21], we would then like to find the $\mathbf{H}$ that best explains the measured $\mathbf{C_y}$, by minimizing:

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{H}} \|\hat{\mathbf{C}}_{\mathbf{y}} - \mathbf{H}\mathbf{A}\mathbf{C_x}\mathbf{A}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} - \mathbf{C_n}\|_F^2, \tag{2.22}$$

where $\hat{\mathbf{C}}_{\mathbf{y}}$ is the estimated measurement data covariance matrix which is estimated from a large set of measurements. However, this problem is hard to solve since the term within the Frobenius norm is quadratic w.r.t. $\mathbf{H}$. A straightforward solution, in the noiseless case, to estimate $\mathbf{H}$ from $\mathbf{C_y}$ is found as:

$$\hat{\mathbf{H}} = \mathbf{C_y}^{\frac{1}{2}}\mathbf{B}^{-\frac{1}{2}}, \tag{2.23}$$

where $\mathbf{B} = \mathbf{A}\mathbf{C_x}\mathbf{A}^{\mathsf{H}}$. This can be verified by inserting $\hat{\mathbf{H}}$ into (2.21). However, any orthogonal matrix $\mathbf{Q}$, such that

$$\hat{\mathbf{H}} = \mathbf{C_y}^{\frac{1}{2}}\mathbf{Q}\mathbf{B}^{-\frac{1}{2}}, \tag{2.24}$$

2

will also solve (2.21) for **H**. Our solution is to enforce a structure on **H**, which can be obtained by simulating various aberration layers and analyzing their corresponding **H** matrix in accurate wavefield simulators such as *k-Wave* [22] (an accurate ultrasound simulator cabaple of simulating the propagation of waves inside an aberrating layer). We will denote the parameters of this parameterization by $\boldsymbol{\theta}$, and the parameterized **H** matrix as $\mathbf{H}(\boldsymbol{\theta})$. We then try to minimize the following cost function, ensuring a solution is found that lies within the parameterized space for **H**, while also solving (2.21):

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{Q}}\} = \arg\min_{\boldsymbol{\theta}, \mathbf{Q}} \|\mathbf{C}_{\mathbf{y}}^{\frac{1}{2}} \mathbf{Q} \mathbf{B}^{-\frac{1}{2}} - \mathbf{H}(\boldsymbol{\theta})\|_F^2, \text{ s.t. } \mathbf{Q} \in \mathcal{Q}, \tag{2.25}$$

where $\mathcal{Q}$ represents the set of orthogonal matrices (i.e., **Q** lies on the Stiefel manifold). Various ways of solving this optimization problem are proposed in Chapter. 7, and the algorithm is tested on a variety of aberration layers, simulated in the *k-Wave* toolbox. We demonstrate that it allows imaging through a skull layer, and also shows that measurements from a blood flow vasculature can serve as random images with $\mathbf{C_x} = \mathbf{I}$. Hence, one could calibrate through a skull-layer using blood flow to generate random images, followed by B-mode, power Doppler, or fUS imaging of that same data.

REFERENCES

[1] J. A. Jensen, O. Holm, L. Jerisen, H. Bendsen, S. I. Nikolov, B. G. Tomov, P. Munk, M. Hansen, K. Salomonsen, J. Hansen, *et al.*, *Ultrasound research scanner for real-time synthetic aperture data acquisition,* IEEE transactions on ultrasonics, ferro-electrics, and frequency control **52**, 881 (2005).

[2] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, *Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **56**, 489 (2009).

[3] J. T. Fokkema and P. M. van den Berg, *Seismic applications of acoustic reciprocity* (Elsevier, 1993).

[4] M. Verweij, B. Treeby, K. Van Dongen, and L. Demi, *Simulation of ultrasound fields,* Comprehensive biomedical physics , 465 (2014).

[5] P. C. Hansen, *Discrete inverse problems: insight and algorithms* (SIAM, 2010).

[6] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,* IEEE Transactions on information theory **52**, 489 (2006).

[7] E. J. Candès and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?* IEEE transactions on information theory **52**, 5406 (2006).

[8] D. L. Donoho, *Compressed sensing,* IEEE Transactions on information theory **52**, 1289 (2006).

[9] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling,* IEEE signal processing magazine **25**, 21 (2008).

[10] G. David, J.-l. Robert, B. Zhang, and A. F. Laine, *Time domain compressive beam forming of ultrasound signals,* The Journal of the Acoustical Society of America **137**, 2773 (2015).

[11] R. Stoughton and S. Strait, *Source imaging with minimum mean-squared error,* The Journal of the Acoustical Society of America **94**, 827 (1993).

[12] J. Shen and E. S. Ebbini, *A new coded-excitation ultrasound imaging system. i. basic principles,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **43**, 131 (1996).

[13] F. Lingvall, *A method of improving overall resolution in ultrasonic array imaging using spatio-temporal deconvolution,* Ultrasonics **42**, 961 (2004).

[14] R. Lavarello, F. Kamalabadi, and W. D. O'Brien, *A regularized inverse approach to ultrasonic pulse-echo imaging,* IEEE transactions on medical imaging **25**, 712 (2006).

[15] M. F. Schiffner and G. Schmitz, *Fast pulse-echo ultrasound imaging employing compressive sensing,* in *2011 IEEE International Ultrasonics Symposium* (IEEE, 2011) pp. 688–691.

[16] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation,* Geophysics **49**, 1259 (1984).

[17] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics,* Geophysics **74**, WCC1 (2009).

[18] M. Brown, D. Nikitichev, B. Treeby, and B. Cox, *Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles,* Applied Physics Letters **110**, 094102 (2017).

[19] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, *Holograms for acoustics,* Nature **537**, 518 (2016).

[20] S. M. Kay, *Fundamentals of statistical signal processing* (Prentice Hall PTR, 1993).

[21] B. Ottersten, P. Stoica, and R. Roy, *Covariance matching estimation techniques for array signal processing applications,* Digital Signal Processing **8**, 185 (1998).

[22] B. E. Treeby and B. T. Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields,* Journal of biomedical optics **15**, 021314 (2010).

I

PART I

3

COMPRESSIVE 3D ULTRASOUND IMAGING USING A SINGLE SENSOR

Three-dimensional ultrasound is a powerful imaging technique, but it requires thousands of sensors and complex hardware. Very recently, the discovery of compressive sensing has shown that the signal structure can be exploited to reduce the burden posed by traditional sensing requirements. In this spirit, we have designed a simple ultrasound imaging device that can perform three-dimensional imaging using just a single ultrasound sensor. Our device makes a compressed measurement of the spatial ultrasound field using a plastic aperture mask placed in front of the ultrasound sensor. The aperture mask ensures that every pixel in the image is uniquely identifiable in the compressed measurement. We demonstrate that this device can successfully image two structured objects placed in water. The need for just one sensor instead of thousands paves the way for cheaper, faster, simpler, and smaller sensing devices and possible new clinical applications.

3.1. INTRODUCTION

Ultrasound imaging is a widely used technique in medical decision-making, mainly due to the fact that ultrasound devices use nonionizing radiation and are relatively low-cost. By carefully aiming the transducer, the operator can use ultrasound to obtain real-time images of specific cross sections of the body. The contrast in ultrasound images is due to sound speed and density differences between tissues. This is why, for example, a baby's skull in a watery womb forms a clear picture in ultrasound imaging. Almost all ultrasound images are made with transducers composed of many tiny sensors (between 64 and 10,000) that can transmit short ultrasonic bursts (typically 1 to 40 MHz). These bursts are then reflected by the tissue and recorded by the same sensor array. The time delay between transmission of the burst and detection of the echoes defines where the tissue is located. The strength of the reflected echo contains information on the density of the tissue.

The sensors that make up these arrays are made mainly from piezoelectric crystals. When a voltage is applied to the material, the crystal vibrates and produces a short ultrasonic burst of a few cycles [2]. When the sensors are activated with appropriate delays, the emitted ultrasound wave travels along a straight line (in the form of a narrow beam) and can be focused on a point of interest. With these focused beams, a whole volume can be scanned (Fig. 3.1A). In terms of receiving the signal reception, similar delays can be used to selectively enhance the signals from scatterers along the beam of interest. This type of focusing that uses delayed signals in transmitting and/or receiving is named beamforming and has been a source of active academic research for many years [3]. One advantage of applying beamforming in receive mode is that it can be carried out digitally after the received ultrasound field has been recorded (Fig. 3.1B). An alternative technique applied in ultrasound imaging is to use an unfocused ("plane") wave during transmission. The focus is then regained upon reception, allowing for a faster frame rate because the medium is not "scanned" line by line with a focused beam.

The majority of ultrasound imaging is performed using transducer arrays that have their sensors distributed along one dimension (Fig. 3.1, A and B), thereby offering a two-dimensional (2D) cross-sectional view of the inside of the human body. To obtain a full 3D view, one needs to mechanically move or rotate the 1D array or to create a 2D array that allows beam steering in two directions. These 2D arrays have been shown to produce stunning high-resolution 3D images, for example, of faces of fetuses [4, 5]. However, 3D ultrasound imaging is a long way from achieving the popularity of 2D ultrasound imaging apart from in certain medical disciplines, such as obstetrics, cardiology, and image-guided intervention. This is surprising given that a full 3D view should in almost all cases allow for a better assessment of the specific medical question. One of the main reasons for this low acceptance of 3D ultrasound imaging is that 3D imaging requires a high hardware complexity (>1000 sensors on a small footprint, integrated electronics, high data rates, etc.). Most of this is necessary to cope with the constraints imposed by the Nyquist theorem when sampling the received radio frequency ultrasound signals.

Fortunately, as shown by recent discoveries in statistics [6–8], the classical idea that digitization of analog signals (such as ultrasound waves) demands uniform sampling at rates twice as high as the highest frequency present (known as the Nyquist theorem) is
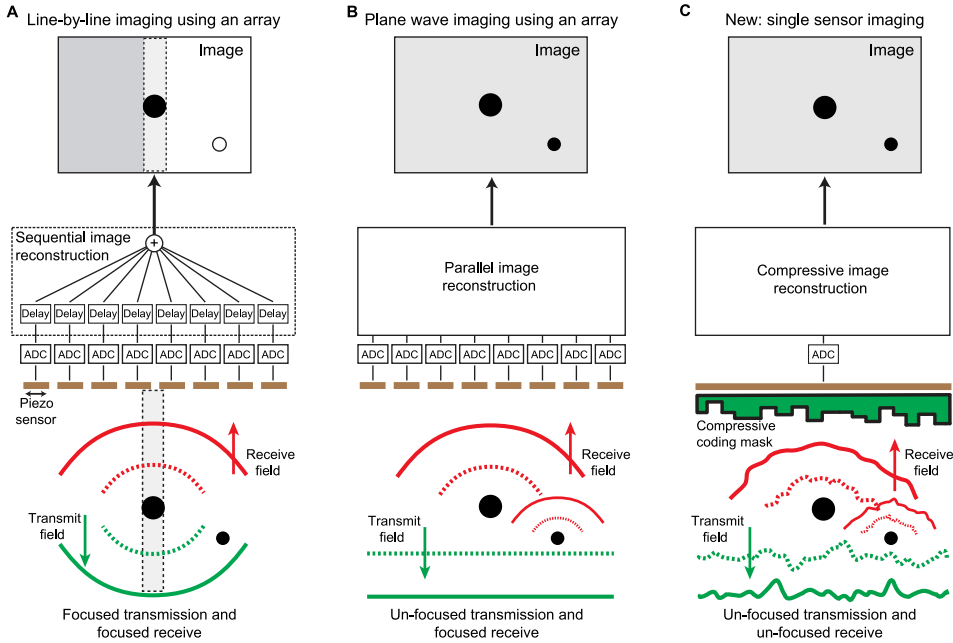
Figure 3.1: Techniques of ultrasound imaging formation. Conventional ultrasound imaging requires an array of sensors to **(A)** focus the transmitted and received wave field or **(B)** only the received wave field by postprocessing the individual signals after analog-to-digital conversion (ADC). We propose using only one sensor without focusing in transmission or receive **(C)**.

no longer a rule carved in stone. This discovery has opened up a very active field of research known as compressive sensing (CS) [9]. One of the leading thoughts in CS is that signals generally contain some natural structure. This structure can be exploited to compress the signal to one that is many times smaller than the original, as is done, for example, in jpeg compression of images. However, until recently, this size reduction was always done after the signal had been acquired at full Nyquist rate. CS now allows the compression to be merged with the sensing. As an immediate consequence, the number of measurements required to recover the signal of interest can be drastically reduced, potentially leading to cheaper, faster, simpler, and smaller sensing devices. In the case of ultrasound imaging, this has already led to the development of successful new strategies for sampling and image reconstruction [10–13].

One of the most striking implementations of CS has been the design of the single-pixel camera by Duarte et al. [14]. Here, the authors managed to reconstruct images using an adjustable mirror mask and a single photodetector. By projecting the image onto the photodetector using random mask patterns, they were able to recover the actual image from a number of measurements much less than the number of image pixels, hence the name compressive imaging. This work has inspired many others to try and use the idea in a similar way for different applications [15–17], including within the ultrasound research community.

Our work follows the concept central to compressive imaging: that of projecting the object/image information through a set of incoherent functions onto a single measurement. A successful strategy for obtaining these compressed measurements is to apply a so-called "coded aperture" [18–20], In this case, the information—whether it entails different light directions, different frequency bands, or any other information that is conventionally needed to build up an image—is coded into the available aperture and associated measurement. Smart algorithms are then needed to decode the retrieved measurements to form an image.

Along these lines, here we introduce for the first time in ultrasound imaging a simple compressive imaging device that can create 3D images. Our device contains one large piezo sensor that transmits an ultrasonic wave through a simple plastic coding mask. Local variations in the mask thickness (Fig. 3.1C) cause local delays, which scrambles the phase of the wave field. Similar techniques were demonstrated by Melde et al. [21] and Brown et al. [22] to produce holograms that generate predefined (photo)acoustic fields. This enables a complex interference pattern to propagate inside the volume, removing ambiguity among echoes from different pixels, as illustrated in Fig. 3.2. The interference pattern propagates through the medium, scatters from objects within the medium, and then propagates back through the coding mask onto the same ultrasound sensor, providing a single compressed ultrasound measurement of the object.
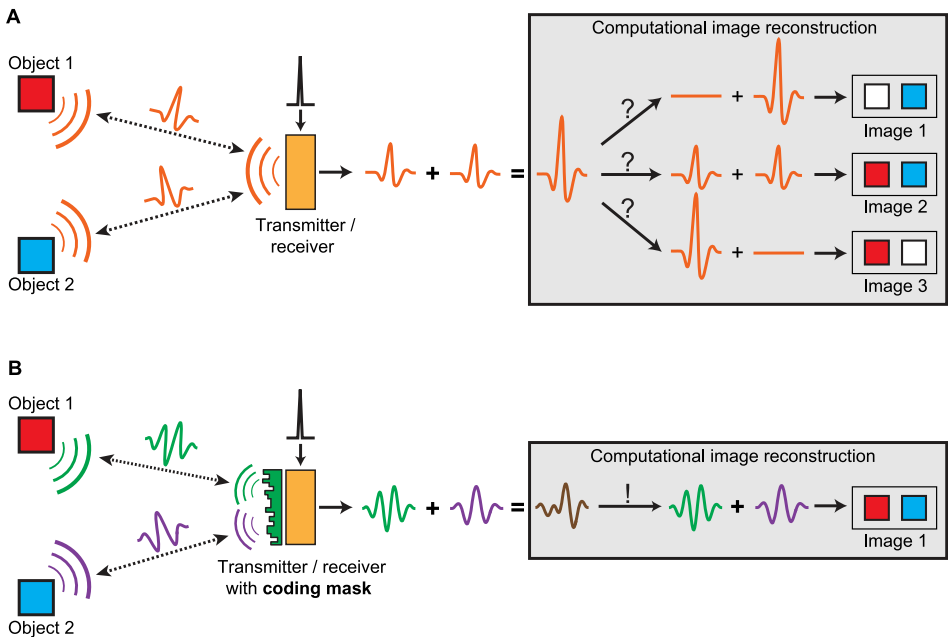
Figure 3.2: A spatiotemporally varying ultrasound field allows for compressive imaging. **(A)** One sensor (transmitter and receiver) cannot distinguish between two objects when the transmitted wave field has no spatiotemporal diversity. **(B)** A simple coding mask can introduce a spatiotemporally varying wave field, which allows unique signal separation between two objects.

Using our device, we aim to mitigate the hardware complexity associated with conventional 3D ultrasound. The manufacturing costs of our device will be much lower. A simple, cheap, single-element transducer is used, and the plastic coding mask can be produced for less than a euro. These lower costs enable broader use of these 3D compressive imaging devices, for example, for long-term patient monitoring. Because the imaging is 3D, finding and maintaining the proper 2D view does not require a trained operator. One could also envision other applications, such as minimally invasive imaging catheters that are too thin to accommodate the hundreds or thousands of electrical wires currently needed for 3D ultrasound imaging. Here, we discuss this new imaging concept and demonstrate experimentally that a simple setup with a coding mask allows for 3D ultrasound imaging—thereby paving the way for an entirely new method of imaging in which the complexity is shifted away from the hardware and toward computing power.

3.2. RESULTS

3.2.1. ULTRASOUND WAVE FIELD DIVERSITY USING A CODED APERTURE

For single-sensor compressive imaging, the 3D object information needs to be projected onto a set of incoherent basis functions via a coded aperture. This is accomplished by breaking the phase uniformity of the ultrasound wave in both transmission and reception. Conventionally, the loss of phase uniformity (for example, through aberration) is unwanted because it leads to blurring and artifacts in the image. In our case, however, the pattern of this interference is known and is used to our benefit.

The coding mask placed in front of the ultrasound sensor consists of a plastic material that supports ultrasound propagation with a speed of around 2750 m/s, in contrast to the speed in water, which is 1480 m/s at room temperature. Varying the mask thickness consequently varies the time the ultrasound wave spends within the plastic before entering the medium, in our case water. For an ultrasound wave with a central frequency of 5 MHz to be delayed by a full wavelength with respect to the rest of the wave requires a local thickness increase of 641 μm. When the thickness of the plastic varies spatially, the phase uniformity of the propagating wave is broken, initiating a random deterministic interference pattern in the medium. The mask covers the complete aperture of the sensor (diameter, 12.7 mm), and the thickness randomly varies between 0.1 and 1 mm. For more details, see the Materials and Methods section. It is important to note that the mask generates an interference pattern that essentially stretches in four dimensions (three spatial and one temporal). So for every location in 3D space (every pixel), we have one unique ultrasound signal that originates from that location. Figure 3.3 and movie S1 show how the natural focusing of the ultrasound sensor is altered after we place our coding mask in front of the sensor.

After excitation with a short high-voltage pulse, the piezo sensor used to insonify the medium produces a short ultrasound burst. Although short, the limited bandwidth ensures the burst has a certain length (time) and contains several (positive-negative) cycles. After propagating through the mask, the pulse length increases further due to the distortion of the wavefront by the coding mask (Fig. 3.4A). These waves will propagate spherically in the medium and will interfere constructively and destructively, a process
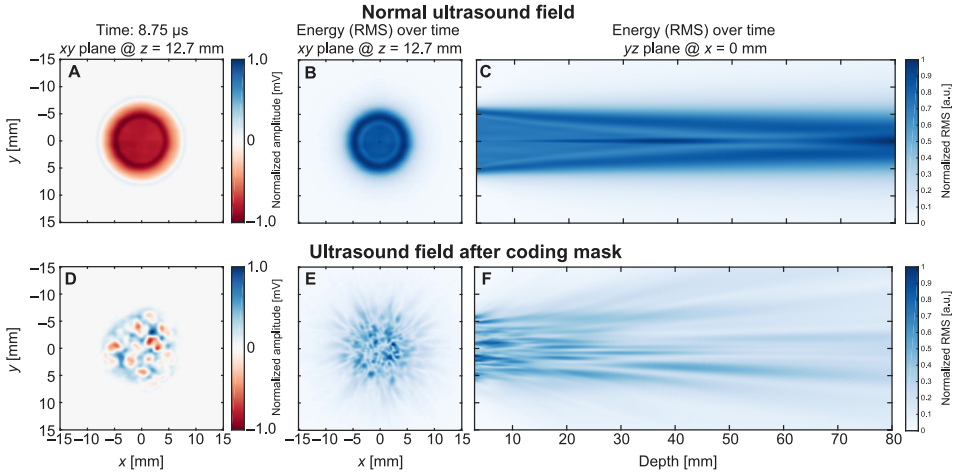
Figure 3.3: A random delay coding mask breaks the phase uniformity of the ultrasound transmission to enable compressive imaging. The top panel shows the ultrasound field transmitted from a normal piezoelectric sensor recorded by a small microphone (hydrophone) in front of the sensor. Note that the transmitted wave focuses naturally into a narrower bundle after propagation. The bottom panel shows the same transmitted wave field after propagation through the plastic coding mask. Here, the phase uniformity of the wave is completely lost. The images in **(A)** and **(D)** show one time sample (at 8.75 μs) of the transmitted wave field in a 2D plane (12.7 mm away from the sensor surface) perpendicular to the propagation axis, before and after the addition of the coding mask. The images in **(B)** and **(E)** show a projection of the total ultrasound energy over time in the same 2D plane that is used for **(A)** and **(D)**. The images **(C)** and **(F)** show the energy projection over time in a 2D plane parallel to the ultrasound propagation axis.

that stretches out longer in time than an undistorted wavefront does. Consequently, the available spatial bandwidth increases compared to an undistorted burst by a mask-less sensor. This effect is highlighted in Fig. 3.4B, which shows two spatial frequency spectra: one at the beginning of the pulse and one spectrum a few microseconds later. Here, we assume that most of the acoustic energy is directly transmitted through the coding mask; yet, in reality, some of the energy will be reflected back from the mask-water interface and possibly redirected in the medium through the mask in a second pass after bouncing back from the mask-transducer interface, thereby increasing the spatial variability even more.

Any ultrasound imaging problem requires information about the spatial ultrasound field, typically acquired by array sampling. By modeling the aperture mask as a collection of point sensors, each having different transmit/receive delays, the mask can still be regarded as a sensor array. However, all these point sensor signals are subsequently summed by the piezo sensor just after they have passed though the mask, resulting in a single compressed measurement, as depicted in the right panel of Fig. 3.1 (for a further comparison between a sensor array and a single sensor with a coding mask, see text S2 and Figs. 3.1 and 3.7). The delays produced by the mask create complex spatiotempo-
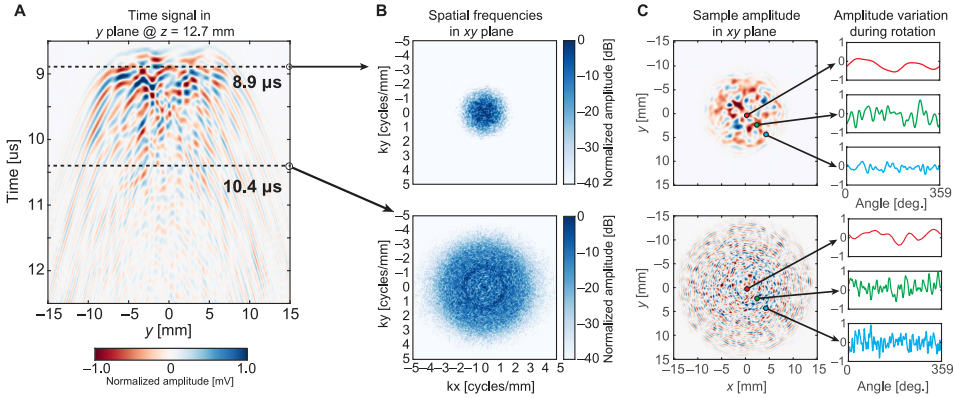
Figure 3.4: Spatiotemporal diversity as well as wave field rotation provides a unique signal for every pixel. (A) A 2D slice (time and y dimension) of the 4D ultrasound field (time and three spatial dimensions), where $x = 0$ and $z = 12.7$ mm. Every time sample of this 4D wave field has unique spatial properties, as is shown in **(B)** and **(C)**. **(B)** The spatial ultrasound field at a later point in time contains higher spatial frequencies, which is desirable for reconstructing objects with high resolution. The possibility of rotating the mask offers another dimension of signal variation to be used for image reconstruction. **(C)** Analysis of two samples based on their spatial properties with respect to coding mask rotation. For pixels further away from the center, the local variance is higher than for pixels closer to the center.

ral interference patterns that ensure that each pixel generates a unique temporal signal in the compressed measurement. This unique pixel signature enables direct imaging without the need for uncompressed spatial measurements.

Because of the limitations in temporal bandwidth, as well as in mask thickness distribution, it is not guaranteed that all pixel echo signals are uncorrelated. To introduce more diversity between the pixels, we chose to rotate the mask in front of the sensor such that the interference pattern is rotated along with it, thereby obtaining additional measurements containing new information. Mask translations (in x, y, or z dimension) would provide similar pixel diversity. Note that for a sufficiently good mask, mask rotation or translation is not necessarily required but significantly decreases interpixel correlations. The result on the ultrasound field as received by every pixel in one xy plane is depicted in Fig. 3.4C, where the effect of rotation in terms of temporal variety is shown at two points in time for three different pixels in the field of view.

3.2.2. Pixel diversity
If we model the ultrasound field of the coded aperture by approximating the mask aperture as a collection of point sources and sensors (see Materials and Methods section), we can analyze the coding performance of the aperture. To this end, we compute the pulse-echo signals for pixels in the central (xy plane) area at a fixed depth (z) relative to the sensor and then compute their cross-correlations. Ideally, all signals are uncorrelated, so that any superposition of pulse-echo signals can be uniquely broken down into its composing individual pulse-echo signals (as in Fig. 3.2), in which case the sig-

nals for each pixel can be unambiguously resolved. Figure 3.5B shows the histogram of interpixel correlations for a plane parallel to the sensor surface at a depth of 12.7 mm (aperture diameter size). We computed these correlations without a mask for several rotations. Because the wave field phase without using a mask is highly uniform, most pulse-echo signals are highly correlated. However, adding a mask to the single sensor causes the distribution to be zero-mean and removes the high correlations around +1 and -1. As expected, these results illustrate that adding more measurements by rotating the mask causes correlations to be distributed more narrowly around zero; pulse-echo signals become more orthogonal. This tells us that the use of rotation removes ambiguities and consequently increases resolvability of scatterers. Here, we do not include the signal correlations over the depth dimension, because generally speaking, the decorrelation in the acoustic propagation direction comes naturally with the pulse-echo delay, or in other words, using only one sensor, we can estimate the depth of an object, but we cannot see whether the object is located left or right from the sensor.

Instead of looking at the pixel-pixel correlation in an entire xy imaging plane, we can also observe the local correlations. Such observations are demonstrated in Fig. 3.5D, which visualizes the correlations between the center pixel and all other local pixels. A higher correlation between pixels results in lower resolution due to their increased ambiguity. These figures show that the resolution in x and y worsens as the depth increases and that the resolution in z is better than that in xy, which is often the case for ultrasound imaging.

For a more detailed analysis, we can also use the Cramer-Rao lower bound (CRLB) and apply it in the aforementioned approximate model [23]. The CRLB is a bound on the best attainable estimation variance for any unbiased estimator (for example, the estimated parameter can be the image itself). That is, the CRLB is related to the average squared estimation error. Hence, a lower CRLB suggests that the best obtainable resolution is lower as well. The actual estimator (beamformer) used in this paper is described in the next section and does not necessarily achieve the CRLB. Consequently, the error covariance of the actual estimator is typically greater than the CRLB, but the CRLB is a good measure to study trends in the resolution. The CRLB is derived for the case of estimating the position of a single scatterer given a pulse-echo measurement with additive Gaussian noise (see the Materials and Methods section for more details). For a 3D problem, this results in estimating three unknowns (x, y, and z coordinates) from N measurements, where N is the total number of samples. Although position estimation and imaging are not the same, one could argue that the best value obtained for positional error spread (for example, 3 SD of error) nevertheless is indicative for the achievable imaging resolution.

To analyze the behavior and performance of the mask, we compute the CRLB for several points in 3D space, comparing the bounds for measurements obtained with no and four rotations (Fig. 3.5E). We distinguish two main factors that contribute to the CRLB. First, rotation of the mask significantly reduces the CRLB, and a very large performance gain is obtained by using more than one rotation. This effect is apparent when comparing the CRLB for several rotations in the graphs in Fig. 3.5E. Second, when we analyze the CRLB for no rotation, we see that the geometry of the problem determines the performance, which is visible in the second panel in Fig. 3.5E. That is, the performance
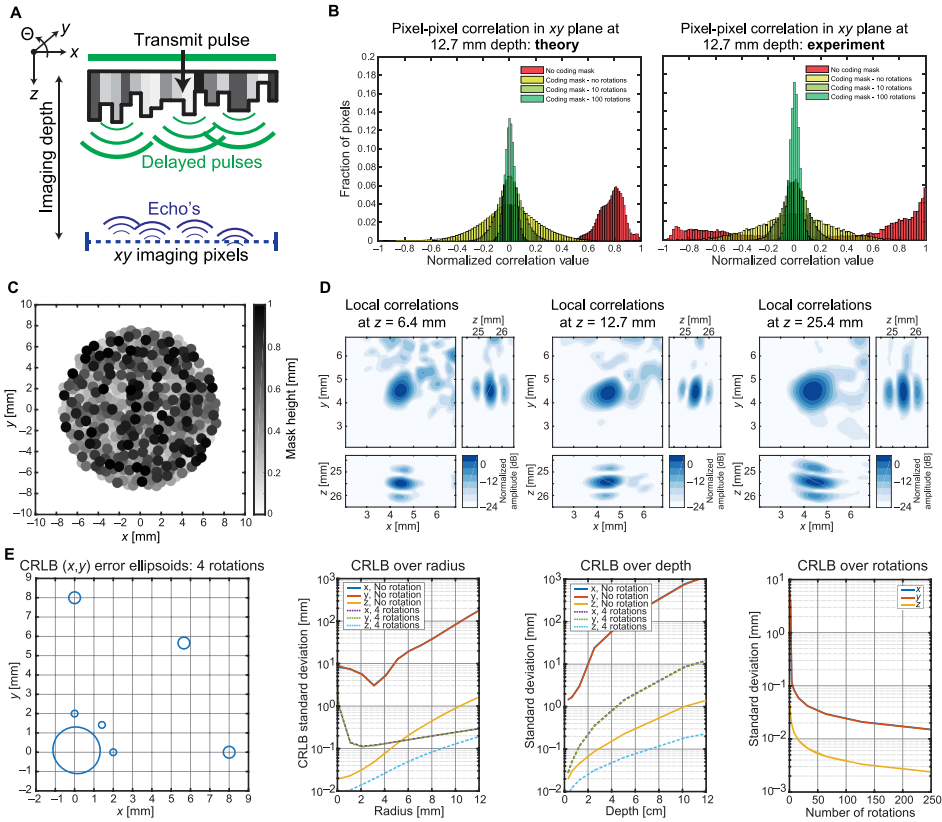
Figure 3.5: Random delays in the coding mask generate signal variability for unique pixel reconstruction. **(A)** Schematic drawing of the approximated ultrasound field in a plane inside the medium parallel to the sensor surface. For every pixel in this plane, we want the echo signals to be as uncorrelated/orthogonal as possible. **(B)** The histograms show the cross-correlation values of a large set of pixels in the plane, obtained using an approximate simulation model. When the plastic delay mask is rotated, the cross-correlation values distribute closer to zero, which suggests better image reconstruction. **(C)** Mask layout where the gray values indicate the thickness variations, causing local delays of the ultrasound field. **(D)** Local correlations for a pixel at a fixed (x, y) position at several depths using 72 mask rotations. **(E)** CRLB analysis. The leftmost panel shows the level curves for one SD of position estimation error for a depth z of 12.7 mm. The center graphs show the SD in x, y, and z over radius (distance to the center of rotation) at depth of 12.7 mm, and over depth for a pixel at a distance of 3 mm from the center. The rightmost graph shows how the CRLB changes as more measurements are added by rotation for a pixel at a distance of 3 mm from the center and at a depth of 12.7 mm.

for a small radial distance is dependent on the specific mask layout and worsens as the received energy decreases inversely proportional to the distance from the sensor. For computing the CRLB graphs, we used a constant noise energy such that the signal-to-noise ratio of a pixel at $(x, y, z) = (0, 0, 12.7$ mm$)$ is equal to 10 dB.

3.2.3. SIGNAL MODEL AND IMAGE RECONSTRUCTION

Because only one sensor is used, there is no possibility of applying standard beam-forming techniques typically used in conjunction with sensor arrays. These techniques mainly consider the geometry of the problem and do not require detailed information on the ultrasound transmission itself. In our case, it is essential that this detailed information is available because we only have one sensor observation. Thus, instead of applying a standard geometric operation to multiple sensor observations, we attempt to explain the received signal as a linear combination of point scatterer echo signals [24, 25]. Without loss of generality, let us represent the unknown true image by a vector **v**, containing N pixels at a desired resolution. In our case, we use a pixel size smaller than two times the smallest wavelength detected (0.120 mm). We found that this discretization size was a good trade-off between model accuracy and memory usage. Furthermore, we assume that the data obtained by the measurement device are linearly related to the image **v** and can be modeled as

$$\mathbf{u} = \mathbf{Hv},$$

where **u** is an M-dimensional vector containing the sampled pulse-echo signals, which are concatenated when using multiple rotations, such that M is equal to the number of time samples times the number of rotations. The M-dimensional vector **n** represents the additive noise. Note that in this model, the system matrix or so-called observation matrix **H** is determined by the measurement device (hardware) and contains all pulse-echo signals of all the pixels in **v**. For more details on the dimensions of **H**, a comparison to multisensor arrays, and the relation to CS, see text and figs. S2 and movie S3.

Having defined **u**, **H**, and **v**, and assuming that a linear propagation and scattering model holds, imaging can be accomplished using a wide variety of linear inversion methods. To demonstrate the system, we imaged two plastic letters that were positioned at a distance of 17 and 23 mm away from the sensor. Here, **H** contained pulse-echo signals for all pixels within a 3D volume containing the scattering objects and was constructed using the calibration method discussed in the next section. As seen in Fig. 3.6 (B, D, and F), both letters were successfully reconstructed, demonstrating the 3D imaging capabilities of the presented methods. For Fig. 3.6 (B and D), we imaged a small volume containing each letter and applied regularized least squares inversion using the LSQR algorithm [26] with limited iterations to regularize the problem, which is known to be equivalent to a filtered singular value decomposition inversion [27].

For the full 3D reconstruction shown in Fig. 3.6F, we made use of the sparsity of these two letters in water by applying the sparsity-promoting basis pursuit denoising (BPDN) algorithm [28]. As can be seen, this prior knowledge about the image could be effectively exploited to improve image quality, significantly improving the dynamic range from 15 to 40 dB. Unlike in CS reconstructions, sparsity is not necessary for linear inversion due to the local correlations of the encoding. However, sparsity can be exploited to improve
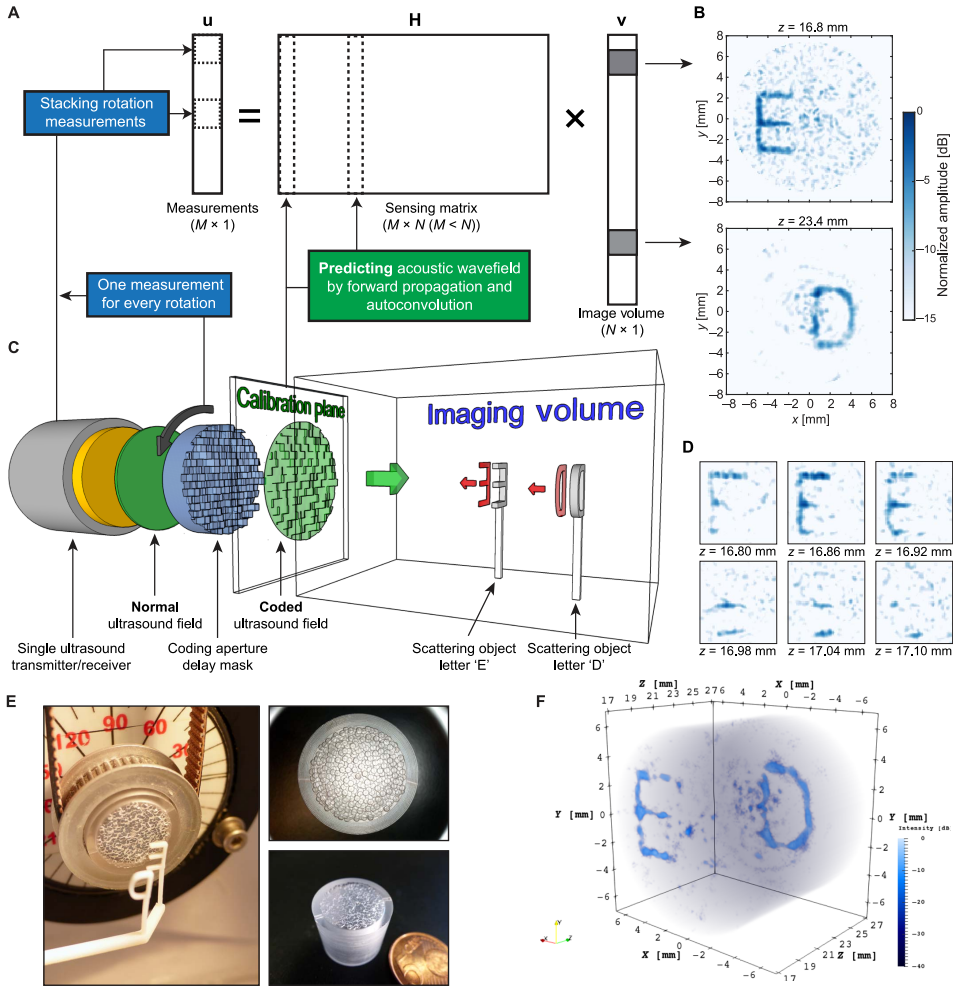
Figure 3.6: Compressive 3D ultrasound imaging using a single sensor. (A) Schematic sketch of the signal model involved in this type of compressive imaging. Each column of the observation matrix **H** contains the ultrasound pulse-echo signal that is associated with a pixel in 3D space, which is contained in the image vector **v**. By rotating the coding mask in front of the sensor, we obtain new measurements that can be stacked as additional entries in the measurement vector **u** and additional rows in **H**. **(B)** Result of solving part of the image vector v using an iterative least squares technique. The two images are mean projections of six pixels along the z dimension [individual z slices shown in **(D)**]. **(C)** Schematic overview of the complete imaging setup. A single sensor transmits a phase uniform ultrasound wave through a coding mask that enables the object information (two plastic letters "E" and "D") to be compressed to a single measurement. Rotation of the mask enables additional measurements of the same object. **(D)** Reconstruction of the letter "E" in six adjacent z slices. A small tilt of the letter (from top left corner to bottom right corner) can be observed, demonstrating the potential 3D imaging capabilities of the proposed device. **(E)** Image showing the two 3D-printed letters and the plastic coding mask with a rubber band for rotating the mask over the sensor. The two right-hand panels show close-ups of the plastic coding mask. **(F)** 3D rendering of the complete reconstructed image vector **v**, obtained by BPDN. The images shown in **(B)** and **(D)** were obtained using 72 evenly spaced mask rotations, and the full 3D image in **(F)** was obtained using only 50 evenly spaced rotations to reduce the total matrix size.

the reconstruction as discussed in 3.C.

3.2.4. CONSTRUCTING THE SYSTEM MATRIX H

For computing the column signals of **H**, it is crucial to know the entire spatiotemporal wave field. To this end, we adopt a simple calibration measurement in which we spatially map the impulse signal (using a small hydrophone and a translation stage) in a plane close to the mask surface and perpendicular to the ultrasound propagation axis (Fig. 3.6C). This recorded wave field can then be propagated to any plane that is parallel to the recorded plane using the angular spectrum approach [29, 30]. Note that this procedure is required only once and is unique for the specific sensor and coding mask. According to the reciprocity theorem, we are able to obtain the pulse-echo signal by performing an autoconvolution of the propagated hydrophone signal with itself [31]. Using these components, we can then compute the pulse-echo ultrasound signal for every point in 3D space and subsequently populate the columns of **H**. The additional rotation of the coding mask can be regarded as a rotation of the ultrasound field inside the medium. Because every rotation provides a new measurement of the same image **v**, we can make use of this information by stacking these new signals as additional rows in the matrix **H**. The more independent rows we add to **H**, the better conditioned our image reconstruction problem becomes.

More formally, let $\mathbf{u}_i$ denote the ith sampled pulse-echo signal of length K, where K is the number of temporal samples obtained from one pulse-echo measurement and i indexes the various measurements obtained from different rotation angles. Let $\mathbf{H}_i$ denote the measurement matrix for rotation i of size $K \times N$, where N is the number of pixels. Then, if R rotations are performed, **u** and **H** are defined as follows: $\mathbf{u} = \left[\mathbf{u}_0^\mathsf{T} \ \mathbf{u}_1^\mathsf{T} \ldots \mathbf{u}_{R-1}^\mathsf{T}\right]^\mathsf{T}$ and $\mathbf{H} = \left[\mathbf{H}_0^\mathsf{T} \ \mathbf{H}_1^\mathsf{T} \ldots \mathbf{H}_{R-1}^\mathsf{T}\right]^\mathsf{T}$, where **u** is length M, with $M = KR$, and **H** has size $M \times N$. Furthermore, because each $\mathbf{H}_i$ is obtained by propagating and rotating the same calibration measurement (see "Acoustic calibration procedure" section in the Materials and Methods), each $\mathbf{H}_i$ can be computed from a reference measurement matrix $\tilde{\mathbf{H}}$: $\mathbf{H}_i = f(\tilde{\mathbf{H}}, \phi_i)$, where ϕ_i is the angle of rotation corresponding to measurement i, and the function $f(\tilde{\mathbf{H}}, \phi_i)$, computes the matrix $\mathbf{H}_i$ from the reference matrix $\tilde{\mathbf{H}}$ and input angle ϕ_i by appropriately rotating the acoustic field implicitly stored in $\tilde{\mathbf{H}}$. Consequently, **H** is expressed as $\mathbf{H} = \left[f(\tilde{\mathbf{H}}, \phi_0)^\mathsf{T} \ f(\tilde{\mathbf{H}}, \phi_1)^\mathsf{T} \ldots f(\tilde{\mathbf{H}}, \phi_{R-1})^\mathsf{T}\right]^\mathsf{T}$.

3.3. DISCUSSION

We have demonstrated that compressive 3D ultrasound imaging can be successfully performed using just a single sensor and a simple aperture coding mask, instead of the thousand or more sensors conventionally required. The plastic mask breaks the phase uniformity of the ultrasound field and causes every pixel to be uniquely identifiable within the received signal. The compressed measurement contains the superimposed signals from all object pixels. The objects can be resolved using a regularized least squares algorithm, as we have proved here by successfully imaging two letters ("E" and "D"). Furthermore, we have provided an in-depth analysis of the qualities of our coding aperture mask for compressive ultrasound imaging. We believe that this technique will pave the way for an entirely new means of imaging in which the complexity is shifted away from the hard-

ware side and toward the power of computing.

The calibration procedure that we use and our imaging device both share common ideas with the time-reversal work conducted by Fink and co-workers [31–35]. Time-reversal ultrasound entails the notion that any ultrasound field can be focused back to its source by re-emitting the recorded signal back into the same medium. Consequently, the ultrasonic waves can be focused onto a particular point in space and time if the impulse signal of that point is known. Using a reverberant cavity to create impulse signal diversity and a limited number of sensors, Fink et al. have shown that a 3D medium can be imaged by applying transmit focusing with respect to every spatial location. Hence, they need as many measurements as there are pixels, resulting in an unrealistic scenario for real-time imaging. Our work differs in the fact that rather than focusing our ultrasound waves to a single point, we encode the whole 3D volume onto one spatiotemporal varying ultrasound field, from which we then reconstruct an image based on solving a linear signal model. This has the advantage of requiring only a few measurements to image an entire 3D volume, in contrast to time-reversal work. This difference in acquisition time is crucial for medical imaging where the object changes over time, for example, in a beating heart. Another important distinction is the small impedance difference between our coding mask and imaging medium (less than a factor of 2). As a consequence, there is very little loss of acoustic energy as compared to reverberant cavities that inherently require high impedance mismatch to ensure signal diversity [34, 35].

Besides the time-reversal work, our technique also shares some common ground with the 2D localization work by Clement et al. [36–38]. They successfully showed that frequency-dependent wave field patterns can be used to localize two point scatterers in a 2D plane using only one sensor in receive. Instead of solving a linear system as we propose here, the authors used a dictionary containing measured impulse responses and a cross-correlation technique to find the two point scatterers. By contrast, we propose to induce signal diversity using local delays in both transmit and receive. This type of "wave field coding" can be applied to many other imaging techniques and seems to offer a much higher dynamic range than the slowly varying frequency-dependent wave fields. As a result, we are able to move beyond the localization of isolated point scatterers but perform actual 3D imaging as we show in this paper.

The findings reported in this paper demonstrate that compressive imaging in ultrasound using a coded aperture is possible and practically feasible. Our device has disadvantages and limitations and does not deliver similar functionality as existing 3D ultrasound arrays. For example, we cannot focus the transmission beam, making techniques such as tissue harmonic imaging that require high acoustic power less accessible [38]. In addition, the mechanical rotation of the mask is expected to introduce a fair amount of error in the prediction of the ultrasound field and limits the usability when it comes down to imaging moving tissue or the application of ultrasound Doppler. Besides, a rotating wave field becomes less effective at pixels closer to the center of rotation—assuming a spatially uniform generation of spatial frequencies. Future systems should consider the incorporation of other techniques to introduce signal variation, such as controlled linearized motion, or a mask that can be controlled electronically, similar to the digital mirror devices used in optical compressive imaging [14]. Furthermore, the angular spectrum approach method that predicts the ultrasound field in-

side the medium is not flawless in cases where the medium contains strong variations in sound speed and density. These issues may possibly be solved, however, by incorporating more advanced ultrasound field prediction tools that can deal with these kinds of variations [39]. This will increase the computational burden of the image reconstruction even further but may potentially lead to better results than what we have shown here.

One of the fundamental ideas of this paper is to prove that spatial wave compression allows for compressive imaging. Hence, this technique is applicable not only to ultrasound but also to many other imaging and localization techniques that normally rely on spatial sampling of coherent wave fields. With respect to clinical ultrasound, we believe that the proposed technique may prove to be useful in cases where signal reduction is beneficial, for example, in minimally invasive imaging catheters. Furthermore, we think that a coded aperture mask could potentially be used in conjunction with conventional 1D ultrasound arrays to better estimate the out-of-plane signals and possibly extend the normal 2D imaging capabilities to 3D.

Alternative coded aperture implementations for ultrasound compressive imaging are certainly possible. However, the range of the physical parameters that we can vary to create a useful coded aperture for ultrasound seems less diverse compared to optics. The properties of light are such that it can easily be blocked very locally (for example, by absorption), which is hardly possible with ultrasound waves because of their mechanical nature and larger wavelengths. The idea of applying local delays in the ultrasound field to uniquely address every pixel therefore seems more practical. Ultimately, it will be these kinds of physical parameters that will dictate whether and to what extent true compressive imaging is possible. Restricting ourselves to local time delays, we have successfully shown in this paper that the physics of ultrasound allow for the ultimate compression of a 3D volume in a few measurements with a single sensor and a simple coding mask.

3.4. Materials and methods

3.4.1. Mask manufacturing

Our aim was to make a spatiotemporally varying ultrasound field by locally introducing phase shifts in the wave transmitted by the sensor. We did so by introducing a mask that delays the wave locally, similar to the function of a phase spatial light modulator in optics [40]. The aberration mask that we put in front of the sensor was made from an 11-mm-thin circular plastic layer (Perspex), in which we had drilled holes 1 mm in diameter (Fig. 3.5C). The depth and positions of these holes were randomly chosen by an algorithm, allowing for a depth ranging between 0.1 and 1 mm and for a maximum of 0.5-mm overlap between the holes. The holes were drilled using a computer numeric control machine (P60 HSC, Fehlmann). The choice for 1-mm holes was made because of convenience of manufacturing. At room temperature, the speed of sound propagation in water is approximately 1480 m/s, whereas the speed of sound in the Perspex material is 2750 m/s. Perspex has good properties in terms of processing, high sound speed, and low acoustic absorption.

3.4.2. ACOUSTIC SETUP

For the transmission and reception of ultrasonic waves, we used one large sensor of piezocomposite material that had an active diameter of 12.7 mm and a center frequency of 5 MHz (C309-SU, Olympus Panametrics NDT Inc.). The sensor was coupled to a pulser/receiver box (5077PR, Olympus Panametrics NDT Inc.) that can generate a short -100 V transmit spike and amplifies signals in receive up to 59dB. No filtering was applied in receive. The sensor with the mask was mounted in a tank filled with water measuring $200 \times 300 \times 200$ mm. In the water, we placed two plastic letters that had been 3D-printed (Objet30, Stratasys Ltd.). To enhance the ultrasound reflection from these letters, we sprayed only the letters with silicon carbide powder (size, k800). The two letters "E" and "D" refer to the names of our academic institutions, Erasmus Medical Center and Delft University of Technology.

The thin plastic coding mask was part of a larger round casing that fitted snugly around the sensor (Fig. 3.6E). To this casing, we fitted a cogwheel that allows the casing to be rotated in front of the sensor using a small rubber band and an electronic rotary stage (T-RS60A, Zaber Technologies). By rotating the aberration mask, we were essentially able to rotate the interference pattern inside the stationary medium, allowing for multiple observations of the same object. We allowed a thin water film between the plastic mask and the sensor surface to ensure optimal coupling between the piezo material and the plastic mask. With this setup, it took about 1 min to acquire the signals needed to reconstruct the 3D volume shown in Fig. 3.6.

3.4.3. ACOUSTIC CALIBRATION PROCEDURE

Before we could use our device for imaging, we needed to know the signature of our interference pattern. That is, we needed to know what the transmission and receive wave fields looked like in both space and time, so that we could populate our system matrix **H**. Note that for every spatial point, we had a unique temporal ultrasound signal. This calibration step is an essential component in this kind of compressive imaging [41]. The better **H** is known, the better the reconstruction of the image. For our calibration procedure, we recorded the transmission wave field in a plane perpendicular to the ultrasound propagation axis using a 3D translation stage (BiSlide MN10-0100-M02-21, Velmex) and a broadband needle hydrophone (0.5 to 20 MHz, 0.2-mm diameter; Precision Acoustics) coupled to a programmable analog-to-digital converter (DP310, Acqiris USA) using 12 bits per sample and a 200-MHz sampling rate.

For the results presented in this paper, we recorded the wave field in a 30×30 mm plane located 2.5 mm in front of the aberration mask using a step size of 0.12 mm (less than two times the highest spatial frequency detected) in both the x and y directions (Figs. 3.3 and 3.4). The signal of every pixel was obtained using the angular spectrum approach to address pixels at greater depth, using autoconvolution to account for the pulse-echo acquisition, and using linear interpolation to account for additional rotation measurements; all of which were implemented in MATLAB software, release 2015b (The MathWorks). We noticed that our sensor was not completely axisymmetric, meaning that the mask wave interaction changes slightly upon mask rotation. This issue was solved by measuring four rotated planes (0°, 90°, 180°, and 270°). We generated all intermediate planes by linear interpolation between each pair of neighboring planes. The

calibration procedure took around 1.5 hours, after which the complete spatiotemporal impulse response of the sensor was known. Note that this calibration is sensor- and mask-specific and, in principle, only needs to be done once.

3.4.4. Approximate model and CRLB

To analyze the CRLB for the estimation of a single scatterer from one or more pulse-echo measurements, we derived the expressions for computing the Fisher information matrix, the inverse of which was computed numerically to obtain the CRLB covariance matrix. The following assumptions and conventions were used:

- Any noise or other errors can be modeled as zero mean white Gaussian additive noise, with variance σ^2. Hence, this analysis is only valid if noise sources, such as electronic, quantization, or jitter noise, can be modeled as Gaussian random variables, and their total effect is compounded into the signal-to-noise figure.

- The ultrasound field due to the coding mask can be approximated as a collection of point sources, where the point sources are located at the surface of the mask. The point in time at which each virtual point source is excited is determined by the time of arrival of the ultrasound field from the true sensor location through the mask to the end of the mask.

- The impulse response for the main sensor is taken as being equal to a Gaussian pulse: $h(t) = \exp\left(-\frac{t^2}{2l^2}\right)$, where l determines the pulse width or frequency bandwidth.

- The origin of the spatial axes corresponds to the center of the main sensor surface, with the sensor lying in the xy plane and the z axis perpendicular to the sensor surface.

- The sensor is excited by a single Dirac delta pulse.

- The position of the virtual source j is denoted as $\mathbf{s}_j \in \mathbf{R}^3$, and the set of all $\mathbf{s}_j$ is denoted as $\mathcal{S}$. Vector variables are indicated by bold typesetting.

- The ith element of a vector $\mathbf{k}$ is denoted as $[\mathbf{k}]_i$.

Under the assumptions above, the forward pressure wave field measured at position $\boldsymbol{\theta} = [\mathbf{x}, \mathbf{y}, \mathbf{z}]^\mathsf{T}$ can be written as

$$p(t, \boldsymbol{\theta}) = \sum_{j=0}^{|\mathcal{S}|-1} h(t - \tau(\boldsymbol{\theta}, \mathbf{s}_j)) \tag{3.1}$$

where $\tau(\boldsymbol{\theta}, \mathbf{s}_j) = \frac{1}{c_1}[\mathbf{s}_j]_z \frac{1}{c_0}\|\mathbf{s}_j - \boldsymbol{\theta}\|_2$. This can be regarded as the convolution of the sensor impulse response with the mask impulse response: $p(t, \boldsymbol{\theta}) = h(t) * m(t, \boldsymbol{\theta}, \mathcal{S})$. The echo signal can then be expressed as $a(t, \boldsymbol{\theta}) = h(t) * m(t, \boldsymbol{\theta}, \mathcal{S}) * m(t, \boldsymbol{\theta}, \mathcal{S}) * h(t)$, or simply $a(t, \boldsymbol{\theta}) = p(t, \boldsymbol{\theta}) * p(t, \boldsymbol{\theta})$, which is equal to the autoconvolution of the forward pressure

wave field at $\boldsymbol{\theta}$. Under the white Gaussian noise assumption, it can be shown that the (i,j)-th entry of the Fisher information matrix $I(\boldsymbol{\theta}) \in \mathbf{R}^{3\times3}$ is equal to

$$I(\boldsymbol{\theta})_{i,j} = \frac{1}{\sigma^2} \frac{\partial \mathbf{a}(\boldsymbol{\theta})^\top}{\partial [\boldsymbol{\theta}]_i} \frac{\partial \mathbf{a}(\boldsymbol{\theta})}{\partial [\boldsymbol{\theta}]_j} \tag{3.2}$$

where $\mathbf{a}(\boldsymbol{\theta})$ is a vector representing the discretized signal $a(t,\boldsymbol{\theta})$. Next, we derived the derivative of $\mathbf{a}(\boldsymbol{\theta})$. We started by computing the derivative of $a(t,\boldsymbol{\theta})$:

$$\frac{d}{d[\boldsymbol{\theta}]_i} a(t,\boldsymbol{\theta}) = \frac{d}{[\boldsymbol{\theta}]_i} p(t,\boldsymbol{\theta}) * p(t,\boldsymbol{\theta}) \tag{3.3}$$

$$= \frac{d}{[\boldsymbol{\theta}]_i} \int_{-\infty}^{\infty} p(t-T,\boldsymbol{\theta}) p(T,\boldsymbol{\theta}) dT \tag{3.4}$$

$$= \int_{-\infty}^{\infty} \frac{d}{[\boldsymbol{\theta}]_i} p(t-T,\boldsymbol{\theta}) p(T,\boldsymbol{\theta}) dT \tag{3.5}$$

$$= 2p(t,\boldsymbol{\theta}) * \frac{d}{[\boldsymbol{\theta}]_i} p(t,\boldsymbol{\theta}). \tag{3.6}$$

Using the definition of $p(t,\boldsymbol{\theta})$ and $\tau(\boldsymbol{\theta},\mathbf{s}_j)$, we find that

$$\frac{d}{[\boldsymbol{\theta}]_i} p(t,\boldsymbol{\theta} = \sum_{j=0}^{|\mathscr{S}|} \frac{[\boldsymbol{\theta}]_i - [\mathbf{s}]_j}{\|\boldsymbol{\theta} - \mathbf{s}_j\|_2^2} h(t-\tau(\boldsymbol{\theta},\mathbf{s}_j)) \left(\frac{1}{c_0 l^2}(t-\tau(\boldsymbol{\theta},\mathbf{s}_j)) - \frac{1}{\|\boldsymbol{\theta}-\mathbf{s}_j\|_2} \right) \tag{3.7}$$

which can be discretized to obtain $d\mathbf{a}(\boldsymbol{\theta})/d[\boldsymbol{\theta}]_i$ for a single measurement. From $I(\boldsymbol{\theta})$, the CRLB is found as $I(\boldsymbol{\theta})^{-1}$.

3.4.5. IMAGE RECONSTRUCTION

Here, we formalized our image reconstruction problem in a set of linear equations $\mathbf{u} = \mathbf{Hv} + \mathbf{n}$. This formalization allowed for a multitude of solvers to be applied to this problem. We applied two inversion methods to find $\mathbf{v}$ from $\mathbf{u}$. The first was the least squares estimate, which finds the image $\mathbf{v}$ that minimizes the squared error between the model $\mathbf{Hv}$ and observation $\mathbf{u}$ as follows: $\hat{\mathbf{v}} = \text{argmin}_{\mathbf{v}'} \|\mathbf{u} - \mathbf{Hv}'\|_2^2$. This estimate is known to be the minimum variance unbiased estimator if the linear model is valid, and the noise $\mathbf{n}$ is the white Gaussian. To implement this estimator, we used the LSQR algorithm [26], which is especially appropriate for large-scale sparse systems. By limiting the number of iterations, one can regularize the reconstruction problem [see, for example, the study by Hansen [27]]; this also allows for balancing of the resolution, the noise robustness, or the contrast.

For the reconstructed letters shown in Fig. 3.6 (B and D), we used about 15 iterations, which took several seconds to compute. The results did not seem to vary much between 7 and 40 iterations. Above 50 iterations, we observed a serious loss in contrast as the algorithm starts to fit the solution to the noise and errors present in our $\mathbf{H}$ matrix. For the complete reconstruction of the 3D volume shown in Fig. 3.6F, we applied a BPDN algorithm [28], exploiting the sparsity of the image because we were only interested in two reflecting letters in an otherwise empty medium (water). For this purpose, we chose to use the iterative two-step shrinkage/thresholding algorithm, named TwIST, using l_1-norm regularization [42] for implementation, although other algorithms that are capable

of l_1-based optimization are also viable. By this approach, we minimized the BPDN cost-function $\hat{\mathbf{v}} = \arg\min_{\mathbf{v}'} \|\mathbf{u} - \mathbf{H}\mathbf{v}'\|_2^2 + \lambda \|\mathbf{v}'\|_1$, where λ is a dimensionless scalar that weights the amount of regularization. The BPDN algorithm regularization parameter λ was chosen by defining a region of interest (ROI) containing the reconstructed letters and an ROI outside the letters and computed the contrast ratio as $20\log_{10}(I_{\text{letter}}/I_{\text{background}})$ for a range of λ values, where I_{letter} or $I_{\text{background}}$ are the average values of $|\mathbf{v}|$ over the pixels in the corresponding ROI. We then chose the λ that offered the highest contrast ratio. Using this procedure, we observed a contrast of 9 dB for regularized least squares and 29 dB for the l_1 regularization. For this proof-of-principle study, it shows that selecting a suitable λ is possible in a similar way as in real-life clinical ultrasound imaging, where the operator tunes some critical image reconstruction parameters (such as speed of sound) by visual feedback on image quality. In further research, it should be investigated if λ should be tuned for image content or a globally optimal λ can be established. If the latter is the case, an optimal λ should be established over large sets of training images, and its generality tested on independent test images. The 3D rendering of the letters shown in Fig. 3.6F were performed using the open-source visualization tool ParaView (www.paraview.org) [43].

REFERENCES

[1] P. Kruizinga, P. van der Meulen, A. Fedjajevs, F. Mastik, G. Springeling, N. de Jong, J. G. Bosch, and G. Leus, *Compressive 3d ultrasound imaging using a single sensor,* Science advances **3**, e1701423 (2017).

[2] K. K. Shung and M. Zippuro, *Ultrasonic transducers and arrays,* IEEE Engineering in Medicine and Biology Magazine **15**, 20 (1996).

[3] K. E. Thomenius, *Evolution of ultrasound beamformers,* in *1996 IEEE Ultrasonics Symposium. Proceedings,* Vol. 2 (IEEE, 1996) pp. 1615–1622.

[4] A. Fenster, D. B. Downey, and H. N. Cardinal, *Three-dimensional ultrasound imaging,* Physics in medicine & biology **46**, R67 (2001).

[5] E. Merz, *25 years of 3d ultrasound in prenatal diagnosis (1989–2014),* Ultraschall in der Medizin-European Journal of Ultrasound **36**, 3 (2015).

[6] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,* IEEE Transactions on information theory **52**, 489 (2006).

[7] D. L. Donoho, *Compressed sensing,* IEEE Transactions on information theory **52**, 1289 (2006).

[8] C. E. Shannon, *Communication in the presence of noise,* Proceedings of the IRE **37**, 10 (1949).

[9] R. G. Baraniuk, *Compressive sensing [lecture notes],* IEEE signal processing magazine **24**, 118 (2007).

[10] N. Wagner, Y. C. Eldar, and Z. Friedman, *Compressed beamforming in ultrasound imaging,* IEEE Transactions on Signal Processing **60**, 4643 (2012).

[11] H. Liebgott, A. Basarab, D. Kouame, O. Bernard, and D. Friboulet, *Compressive sensing in medical ultrasound,* in *2012 IEEE International Ultrasonics Symposium* (IEEE, 2012) pp. 1–6.

[12] C. Quinsac, A. Basarab, and D. Kouamé, *Frequency domain compressive sampling for ultrasound imaging,* Advances in Acoustics and Vibration **2012** (2012).

[13] M. F. Schiffner and G. Schmitz, *Fast pulse-echo ultrasound imaging employing compressive sensing,* in *2011 IEEE International Ultrasonics Symposium* (IEEE, 2011) pp. 688–691.

[14] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling,* IEEE signal processing magazine **25**, 83 (2008).

[15] L. Wang, L. Li, Y. Li, H. C. Zhang, and T. J. Cui, *Single-shot and single-sensor high/super-resolution microwave imaging based on metasurface,* Scientific reports **6**, 1 (2016).

[16] D. Shrekenhamer, C. M. Watts, and W. J. Padilla, *Terahertz single pixel imaging with an optically controlled dynamic spatial light modulator,* Optics express **21**, 12507 (2013).

[17] N. Huynh, E. Zhang, M. Betcke, S. R. Arridge, P. Beard, and B. Cox, *A real-time ultrasonic field mapping system using a fabry pérot single pixel camera for 3d photoacoustic imaging,* in *Photons Plus Ultrasound: Imaging and Sensing 2015,* Vol. 9323 (International Society for Optics and Photonics, 2015) p. 93231O.

[18] E. Caroli, J. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino, *Coded aperture imaging in x-and gamma-ray astronomy,* Space Science Reviews **45**, 349 (1987).

[19] R. F. Marcia, Z. T. Harmany, and R. M. Willett, *Compressive coded aperture imaging,* in *Computational Imaging VII,* Vol. 7246 (International Society for Optics and Photonics, 2009) p. 72460G.

[20] X. Peng, G. J. Ruane, and G. A. Swartzlander Jr, *Randomized aperture imaging,* arXiv preprint arXiv:1601.00033 (2016).

[21] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, *Holograms for acoustics,* Nature **537**, 518 (2016).

[22] M. Brown, D. Nikitichev, B. Treeby, and B. Cox, *Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles,* Applied Physics Letters **110**, 094102 (2017).

[23] S. M. Kay, *Fundamentals of statistical signal processing* (Prentice Hall PTR, 1993).

3

[24] R. Lavarello, F. Kamalabadi, and W. D. O'Brien, *A regularized inverse approach to ul-
      trasonic pulse-echo imaging,* IEEE transactions on medical imaging **25**, 712 (2006).

[25] R. Stoughton and S. Strait, *Source imaging with minimum mean-squared error,* The
      Journal of the Acoustical Society of America **94**, 827 (1993).

[26] C. C. Paige and M. A. Saunders, *Lsqr: An algorithm for sparse linear equations and
      sparse least squares,* ACM Transactions on Mathematical Software (TOMS) **8**, 43
      (1982).

[27] P. C. Hansen, *Discrete inverse problems: insight and algorithms* (SIAM, 2010).

[28] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pur-
      suit,* SIAM review **43**, 129 (2001).

[29] Y. Du, H. Jensen, and J. A. Jensen, *Investigation of an angular spectrum approach
      for pulsed ultrasound fields,* Ultrasonics **53**, 1185 (2013).

[30] P. R. Stepanishen and K. C. Benjamin, *Forward and backward projection of acoustic
      fields using fft methods,* The Journal of the Acoustical Society of America **71**, 803
      (1982).

[31] M. Fink, *Time reversal of ultrasonic fields. i. basic principles,* IEEE transactions on
      ultrasonics, ferroelectrics, and frequency control **39**, 555 (1992).

[32] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, *Building three-dimensional images
      using a time-reversal chaotic cavity,* IEEE transactions on ultrasonics, ferroelectrics,
      and frequency control **52**, 1489 (2005).

[33] C. Draeger, J.-C. Aime, and M. Fink, *One-channel time-reversal in chaotic cavi-
      ties: Experimental results,* The Journal of the Acoustical Society of America **105**, 618
      (1999).

[34] N. Etaix, M. Fink, and R. K. Ing, *Acoustic imaging device with one transducer,* The
      Journal of the Acoustical Society of America **131**, EL395 (2012).

[35] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, *Time reversal kaleidoscope: A
      smart transducer for three-dimensional ultrasonic imaging,* Applied physics letters
      84, 3879 (2004).

[36] P. J. White and G. T. Clement, *Two-dimensional localization with a single diffuse
      ultrasound field excitation,* ieee transactions on ultrasonics, ferroelectrics, and fre-
      quency control **54**, 2309 (2007).

[37] F. C. Meral, M. A. Jafferji, P. J. White, and G. T. Clement, *Two-dimensional image
      reconstruction with spectrally-randomized ultrasound signals,* IEEE transactions on
      ultrasonics, ferroelectrics, and frequency control **60**, 2501 (2013).

[38] G. T. Clement, *Two-dimensional ultrasound detection with unfocused frequency-
      randomized signals,* The Journal of the Acoustical Society of America **121**, 636
      (2007).

[39] B. E. Treeby and B. T. Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields,* Journal of biomedical optics **15**, 021314 (2010).

[40] N. Savage, *Digital spatial light modulators,* Nature Photonics **3**, 170 (2009).

[41] A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, *Imaging with nature: Compressive imaging using a multiply scattering medium,* Scientific reports **4**, 1 (2014).

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

[43] J. Ahrens, B. Geveci, and C. Law, *Paraview: An end-user tool for large data visualization,* The visualization handbook **717** (2005).

[44] E. J. Candes and T. Tao, *Decoding by linear programming,* IEEE transactions on information theory **51**, 4203 (2005).

3

3.A. ON THE SIZE OF THE SYSTEM MATRIX

Standard measurement systems are designed to provide a large number of measurements, generally much larger than the number of unknowns ($M \gg N$), while guaranteeing that the condition number of the system matrix **H** is small enough, such that the problem of estimating **v** from **u** (**u** = **Hv** + **n**) is a well-posed problem. In current ultrasound devices for instance, this can be done by transmitting a short pulse through the tissue as a spatial beam or plane wave (repeated for several spatial beams or plane waves), followed by high-rate sampling (above the Nyquist rate) at every sensor (see also Fig. 3.1a and b). In compressive sensing imaging devices, however, far fewer measurements are taken, thereby reducing the imaging time considerably. In such scenarios, the number of measurements is often smaller than the number of unknowns ($M \leq N$); even when M is larger than N, the system matrix H is typically ill-conditioned in compressive ultrasound imaging. In our case M is typically much smaller (between 10 and 100 times) than N. These values are highly dependent on pulse bandwidth, pulse length after the coding mask, number of rotations, size of imaging volume, and pixel size. In our case M is much smaller than N (10: 1 for Fig. 3.6 (b and d) and 100:1 for Fig. 3.6 (f), thus we need to constrain the solution space using some prior information. We describe these in the Materials and Methods section.We attribute the successful reconstructions with fewer measurements than pixels to the fact that the dependence between columns in H is typically very local in space. Keeping in mind that the pixel size can be chosen arbitrarily, consider a large pixel size such that H contains more measurements than unknowns. As a consequence, the system becomes over-determined, but the modelling error due to reflectors that lay in between pixel positions will be rather large . To reduce this type of error, one would like to increase the pixel density by adding columns to H, resulting in fewer measurements than unknowns. It is obvious that the columns of the pixels that were added will be highly correlated to neighbouring pixels. Consequently , the solution ambiguity related to an echo signal coming from that region will be locally restricted to that image region. Combined with techniques like Tikhonov regularization, the least squares estimator will still be able to allocate energy to the corresponding region. The reconstruction will not be exact, but the allocated energy will be in approximately the correct part of the image.

3.B. COMPARISON WITH MULTISENSOR ARRAY IMAGING

The imaging technique we propose relies on the inversion of complicated wavefields, making it difficult to compare our technique one-to-one with conventional geometrical beamforming techniques. For instance, our transmit beam is deliberately chaotic, and we have no ability of focusing this beam , unlike with normal sensor arrays. However, in the context of imaging by inversion we are able to compare the proposed single sensor with coding mask with a multi-sensor array. To this end we have extended our simulation analysis as described in the Materials and Methods section to further understand the influence of adding more sensors and the use of rotations on the imaging performance. In short, H was computed as described in "Approximate model and Cramer -Rao lower bound"of the Materials and Methods section. We pre-defined an image **v** (schematic drawing of a Toucan , in a xy-plane at a depth of 12.7 mm) and computed
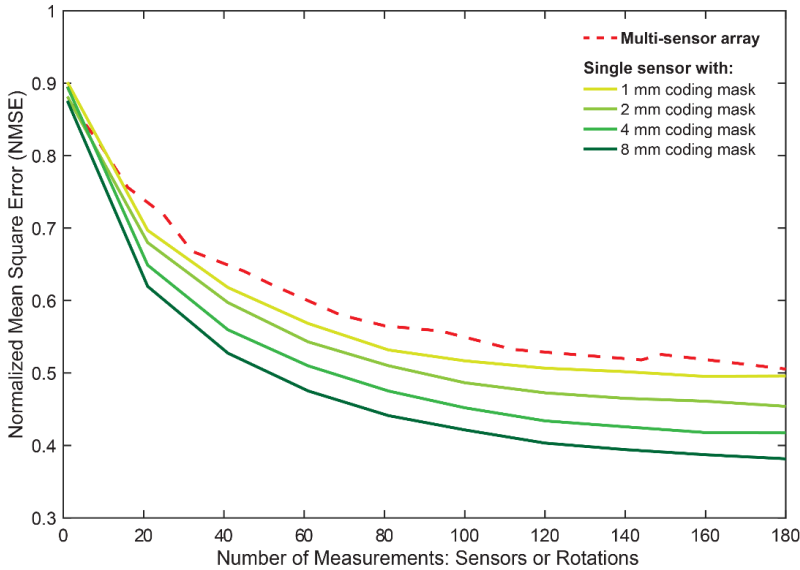
Figure 3.7: Imaging performance for a single sensor with coding mask and normal sensor arrays without coding mask. We computed a system matrix **H** for a normal array with varying sensor densities and for the proposed single sensor with coding mask with several thicknesses and varying rotations. A pre-defined image **v** (schematic drawing of a Toucan, in a xy-plane at a depth of 12.7 mm) was used to compute **u** via **u** = **Hv**. We then compared the estimated image $\hat{\mathbf{v}}$ with the true image **v** by computing the normalized mean square error. The results are shown above. A tradeoff in nmse between number of sensors in an array and number of rotations with a coding mask can be observed. By increasing the mask thickness more information is compressed in the measurement resulting in a better nmse for the same amount of measurements.

u via **u** = **Hv** + **n**. We then compared the estimated image $\hat{\mathbf{v}}$ with the true image **v** by computing the normalized mean square error (nsme) ($\|\mathbf{v} - \hat{\mathbf{v}}\|_2^2 / \|\mathbf{v}\|_2^2$). For this analysis we used similar settings as were used for Fig. 3.5 . We considered the same aperture and coding mask. For the case of the multi-sensor array we removed the coding mask and subdivided the aperture in multiple sensors and excited all sensor subdivisions simultaneously. We compared the nmse as a function of mask rotations and number of sensors. The idea is that every mask rotation provides one unique measurement that should be comparable with the number of spatial measurements that can be obtained with multiple sensors in an array. Since the degree of wavefield complexity largely depends on the available delays in the coding mask, we also vary the local mask thickness from 1 to 8 mm. The overall results of these simulations can be observed in Fig. 3.7. Reconstruction examples can be seen in Fig. 3.8 and movie S3.

3.C. ON THE RELATION TO COMPRESSIVE SENSING

For the results in Fig. 3.6f we used a technique more related to common CS methods. Traditional CS techniques require that the image **v** is sparse in some domain which can
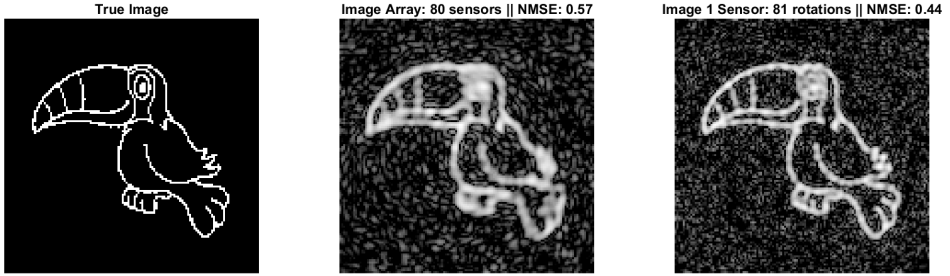
Figure 3.8: Image reconstruction example for a sensor array and a single sensor with coding mask for a comparable amount of measurements. Left panel shows the true image in a xy-plane at a depth of 12.7mm that was used to simulate the measurements for a normal array and single sensor with mask. Middle panel shows the reconstruction for an array with 80 sensors. The right panel shows the reconstruction for a single sensor with a 8 mm mask and 81 mask rotations.

be represented by a basis matrix **G**. The representation **w** of the image in this domain is computed as $\mathbf{w} = \mathbf{Gv}$, and assuming **G** is invertible, the measurement equation can instead be written as

$$\mathbf{u} = \mathbf{Hv} = \mathbf{HG}^{-1}\mathbf{w}$$

and solved for **w**. If **w** is sufficiently sparse in the domain represented by **G**, and the matrix $\mathbf{HG}^{-1}$ satisfies the restricted isometry property (RIP) [44], CS techniques using l_1-based regularization can be used to estimate **w** from **u**, and then compute $\mathbf{v} = \mathbf{G}^{-1}\mathbf{w}$ to obtain an image. In our experiments, the image is already sparse in the spatial domain itself, so that $\mathbf{G} = \mathbf{I}$ (where **I** is the identity matrix) is a sparse basis for our problem. For images that are not spatially sparse, bases such as the wavelet basis can be used to exploit sparsity, although we have not further explored this option. For the full 3D reconstruction shown in Fig. 3.6F, we made use of the sparsity of these two letters in water by applying the sparsity promoting basis pursuit de-noising (BPDN) algorithm. As can be seen, this prior knowledge about the image could be effectively exploited to improve image quality, significantly improving the dynamic range from 9 to 29dB.

4

CODING MASK DESIGN FOR SINGLE SENSOR ULTRASOUND IMAGING

We study the design of a coding mask for pulse-echo ultrasound imaging. We are interested in the scenario of a single receiving transducer with an aberrating layer, or 'mask', in front of the transducer's receive surface, with a separate co-located transmit transducer. The mask encodes spatial measurements into a single output signal, containing more information about a reflector's position than a transducer without a mask. The amount of information in such measurements is dependent on the mask geometry, which we propose to optimize using an image reconstruction mean square error (MSE) criterion. We approximate the physics involved to define a linear measurement model, which we use to find an expression for the image error covariance matrix. By discretizing the mask surface and defining a discrete number of mask thickness levels per point on its surface, we show how finding the best mask can be posed as a variation of a sensor selection problem. We propose a convex relaxation in combination with randomized rounding, as well as a greedy optimization algorithm to solve this problem. We show empirically that both algorithms come close to the global optimum. Our simulations further show that the optimized masks have better a MSE than nearly all randomly shaped masks. We observe that an optimized mask amplifies echoes coming from within the region of interest (ROI), and strongly reduces the correlation between echoes of pixels within the ROI.

4.1. Introduction

I N recent years there has been a considerable interest in reducing the amount of measured sensor data in the ultrasound imaging field. Data compression has great potential, for example, in minimally invasive surgery, where the number of data cables is limited by the available space inside the catheter. This prohibits transferring all data from a dense imaging array through the catheter, making it very challenging to design high frame-rate, high quality imaging devices for this scenario. One possible approach to avoid using a Nyquist-sampled sensor array (either in time or space), is to have the data somehow compressed during sensing. That is, to reduce the amount of sampled data without any complicated electronic hardware actively compressing the measured array signals.

Recently, the field of compressed sensing (CS) showed that such compression is possible if the signal to be estimated has a sparse representation in a known domain [2–5]. This led to a large number of studies in ultrasound using CS in an attempt to get better image reconstructions using less data (see, for example [6–12], amongst others). In compressive sensing, it is assumed that the compressed data is a linear combination of the original, uncompressed, signals. Moreover, there are probabilistic performance guarantees if the compression is implemented by a *random* linear combination of the original signals. However, implementing such a random linear combination in hardware is not necessarily easy, since these compressors might have to combine samples across both space and time, resulting in large sample buffers. Another popular compression technique is randomly subsampling in time and/or space. However, this means that a large amount of information is not used at all (the ignored samples are not in any way present in the compressed measurement), and additionally results in a degradation of SNR. More importantly, we are currently not aware of a domain where ultrasound images are very sparse.

Some of the most striking examples of CS for imaging have been demonstrated in the field of optical imaging, the most well-known being the single-pixel camera [13]. There, random linear combinations of the image are obtained by spatial light modulators (SLM), which basically integrate a different subselection of the image pixels per measurement. This has led to a great number of works using SLM setups for compressed imaging for a single sensor ([14–16], to name a few). Instead of subsampling or implementing a (random) linear compressor in hardware, one could also place a contrasting medium in front of the sensor array. In optics, for example, researchers have placed a heterogeneous medium between the sensor array and the imaging scene, causing multiple scatterings inside this heterogeneous medium, after which the scrambled light field is measured [17–20]. This scrambled light field can be seen as a random linear combination of all image pixels, and CS can be used for image reconstruction. The random compression thus occurs naturally, and is achieved without any electronic hardware. Similar techniques (not employing CS) have also been used in the medical ultrasound community, by transmitting randomized waves and measuring on a single sensor [21], or by placing a heterogeneous cavity in front of a single sensor [22].

We recently demonstrated a similar approach for 3D ultrasound imaging [23, 24], where we have shown how a single sensor with a 'coding mask' can successfully reconstruct 3D objects from multiple pulse-echo measurements, facilitating the design of new

imaging devices with reduced output data. Instead of placing a heterogeneous scattering medium in front of the single sensor, we used a homogenous, but irregularly shaped piece of plastic, to which we refer as a 'coding mask' (Fig. 4.1). In contrast to CS, we have shown that reconstruction is possible without sparse regularization techniques, removing the necessity of an a priori known sparse domain for the ultrasound images.

The main idea of this imaging approach is the following. Consider the experimental setup shown in Fig. 4.1, but without a coding mask. Due to the lack of spatial sampling, one has to rely on the information in the temporal dimension of the measured pulse-echo signal. Only the distance to a reflector can be determined (based on an echo's time-of-arrival (TOA)), but nothing can be said about the direction-of-arrival (DOA), since the pulse-echo signals will look roughly the same for all directions[1]. However, once we place a small layer of irregularly shaped plastic in front of the receiving sensor, in which ultrasound waves have a different speed of sound than the imaging medium, the received wavefields are significantly distorted. If the mask shape is designed correctly, this distortion is direction-dependent, and we are now able to infer the DOA in addition to the distance to the object by exploiting the signal structure in the time-domain waveforms. Through collecting additional measurements by rotating or translating the sensor with the coding mask, the image object is probed from multiple points of view.

In our experiments so far, we used randomly shaped masks to remove ambiguities between pixel signals. Naturally, we would like to know if random masks are a good design choice, and if not, how the coding mask should be shaped to get the best possible imaging performance. In this paper, we investigate coding mask design algorithms using sensor selection techniques, and propose a convex and a greedy optimization program. As such, this study falls within the larger framework of (sparse) ultrasound array design [25, 26]. Using various simulations, we demonstrate that the masks obtained using our techniques exhibit better imaging mean square error (MSE) compared to a random mask design.

To design a mask, we will first introduce an approximate model, where the mask is discretized in all dimensions (x, y, z). Based on this model, we want to minimize the imaging MSE, which is a function of the error covariance matrix, and consequently, of the measurement matrix mapping the discretized image domain to the sampled pulse-echo measurement domain. We will show how optimization problems of this form relate to sensor selection problems. These problems are typically solved in literature using convex relaxations employing the ℓ_1-norm [27, 28], or using more efficient greedy methods by formulating a sub-modular cost-function [29–33]. These are relatively well-understood techniques, and various functions of the reconstruction error covariance matrix can be optimized near-optimally using such convex and greedy optimization algorithms. There are some fundamental differences however, between our problem and the typical sensor selection problem.

First, the inverse error covariance matrix of the typical sensor selection problem is *linear* with respect to the selection variables, whereas for our selection problem it is

---

[1] To be more accurate, the pulse-echo signal for a flat circular sensor will look exactly the same for all reflectors on a circle with the same elevation angle, for all azimuth angles. The azimuth and elevation are with respect to the plane in which the sensor is positioned. However, pulse-echo signals will only change slowly with respect to the elevation angle, so there will be high ambiguity for reflectors with similar DOA.
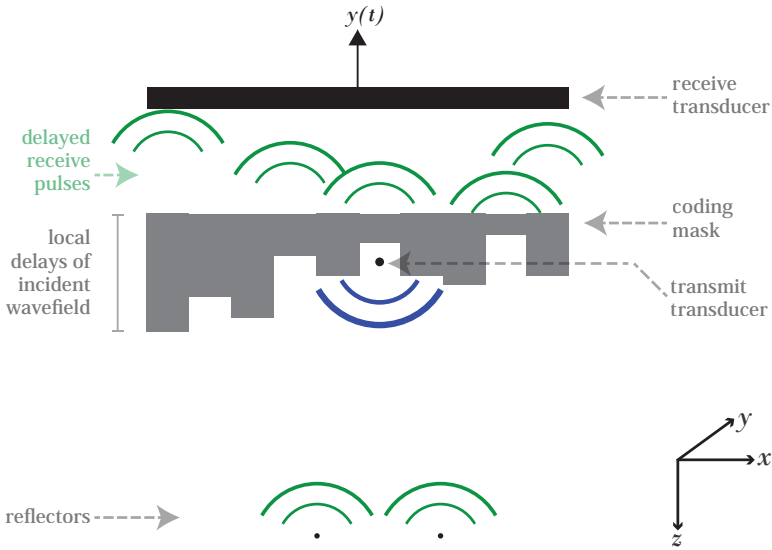
Figure 4.1: Experimental setup considered in this paper. A point source emits a spherical undistorted wave into the ROI. Reflected echoes propagate through the coding mask before being measured by a single transducer. The coding mask has a different speed of sound than its surrounding medium, and distorts the received wavefield in a DOA-dependent manner. As a result, scatterers can be resolved based on the information residing in the temporal waveforms of the pulse-echo signal.

quadratic, making the cost-function non-convex. We are currently not aware of any sensor selection literature studying the same problem structure. Hence, we propose to linearize the problem by lifting the optimization variables to a higher dimensional space and impose a block-matrix sparsity constraint on the matrix formed by these lifted coefficients.

Second, in our optimization problem we want to select one sensor from a specific set, select another one from another set, etc., for a given number of such sets. This second obstacle is also considered in [34], but we are not aware of any similar studies. In [34], each sensor position is associated with a small pool of candidate sensors, corresponding to different sensor types. However, since the inverse covariance matrix in [34] is still a linear function of the selection variables, we cannot employ similar techniques to solve our problem.

In [35], we made a first attempt at mask optimization. There, we proposed a model similar to the one in this work, but using a different convex relaxation, utilizing ℓ_1-reweighting [36]. In this paper, we use a different convex relaxation with a smaller feasible set, without ℓ_1-reweighting, and we also propose a more efficient greedy coordinate-descent optimization scheme. Furthermore, we provide an analysis of the optimized masks, lacking in our previous work.

The remainder of this paper is organized as follows. In Sec. II, we describe an approximate linear measurement model for a masked transducer. In Sec. III, we show how

this model can be parameterized using a discrete selection variable. Using this parameterization, we then formulate the mask optimization problem and propose a convex relaxation, and greedy selection algorithm to find the best mask parameters. Simulation results are discussed in Sec. IV, where we show various examples to evaluate the performance of the optimized masks as compared to the true optimum and a large set of randomly generated masks. In Sec. V the resulting masks are discussed, and the paper is concluded in Sec. VI.

4.1.1. NOTATION

Throughout this paper we use the following notational conventions. Bold lower-case variables (e.g., $\mathbf{x}$) represent vectors, and bold upper-case variables (e.g., $\mathbf{A}$) represent matrices. The transpose and Hermitian transpose are denoted by $^\mathsf{T}$ and $^\mathsf{H}$, respectively. We use $\mathbf{W} \succeq \mathbf{0}$ to indicate that $\mathbf{W}$ is positive semi-definite. Alternatively, the set $\mathbb{S}_+$ is the set of positive semi-definite matrices. By $\mathrm{diag}(\mathbf{W})$ we represent the vector formed by taking the coefficients on the diagonal of $\mathbf{W}$, whereas $\mathrm{diag}(\mathbf{w})$ is the diagonal matrix with the vector $\mathbf{w}$ on its diagonal. Next, $\|\mathbf{w}\|_0$, $\|\mathbf{w}\|_1$ and $\|\mathbf{w}\|_2$ denote the cardinality, ℓ_1-norm, and Euclidean norm of $\mathbf{w}$, respectively. To index matrices or vectors, we use $[\mathbf{W}]_{i,j}$ to represent the (i, j)-th entry of $\mathbf{W}$. In the case of a 3D vector representing a (x, y, z)-coordinate in Cartesian space, we use $[\mathbf{r}]_z$ to index the z-component in $\mathbf{r}$. Finally, $\mathbf{1}$ is the all-one vector, $\mathbf{I}$ denotes the identity matrix, and $\mathbb{E}(\cdot)$ represents the expectation operator. The Gaussian and complex Gaussian distribution with mean μ and covariance matrix $\mathbf{C}$ are written as $\mathcal{N}(\mu, \mathbf{C})$ and $\mathcal{CN}(\mu, \mathbf{C})$, respectively.

4.2. SIGNAL MODEL

We will assume a pulse-echo setup with separate co-located transmit and receive transducers (Fig. 4.1), and the coordinate system shown in the same figure. Throughout this paper, we assume that a point source is located at $(x, y, z) = (0, 0, 0)$, and the masked receive transducer in a plane at $z = 0$. As will be explained at the end of this section, we cannot easily take into account the effect of the mask on the transmitted wave, so we will assume that the transmit transducer emits a spherical wave that is unaffected by the mask.

We first discretize the mask surface in the width and length (x and y) dimensions into many small patches, and regard each of these discrete points as an independent sensor *channel*. Each channel delays the reflected echo field according to the thickness of the mask at the channel and the speed of sound inside the mask. Put differently, we assume that every channel takes the received pulse-echo ultrasound field at the mask-medium interface as channel 'input'. The ultrasound field propagates through each channel independently to arrive at the transducer. This approximate model is visualized in Fig. 4.2.

The basic idea behind the approximate physical model just described is based on ray-tracing physics, and relies on the assumption that the mask is *thin enough* to assume that if a part of the echo wavefront is incident on the mask at position (x, y), it will reach the transducer surface in a sufficiently close neighborhood of (x, y), even with strong wave refraction. Consequently, only a phase change occurs proportional to the mask thickness around (x, y), leading to the independent channels as shown in Fig. 4.2.
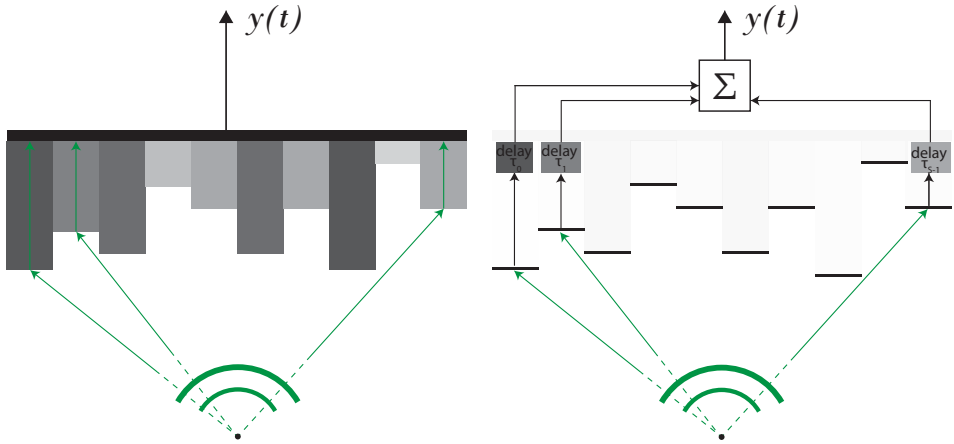
Figure 4.2: Left: we approximate the physics of a transducer with a mask by discretizing the transducer surface into separate propagation channels. Right: approximate signal model. Each channel delays the incident wavefield, and the spatial integration by the transducer surface is approximated by summing all delayed channel signals.

Further assuming that scattering sources are not extremely close to the transducer, the incident wave angle will be limited, reducing the effect of refraction. This requires a proper discretization of the mask surface: if the mask patches (and consequently, the patch size on the transducer surface) are taken too large , they do not sufficiently sample the incident wavefield, and will neither accurately approximate the incident wave on the transducer surface, nor the output signal of the transducer. On the other hand, as patches get smaller, it becomes possible that most of the wave starts interfering into other channels, causing the model to lose its validity. To avoid both cases, we will use a patch size of around 0.3 to 0.5 times the wavelength inside the mask material throughout this work, i.e., wavefields are sampled close to the required spatial sampling (Nyquist) rate of 0.5λ, but not strongly exceeding it.

This approximate model has been succesfully applied in the design of holographic masks, where it was experimentally verified that these approaches can be used to design phase masks that generate a desired pressure pattern [37, 38], and we partly base the validity of our model on these studies. We have also compared this model to a true simulated wavefield obtained using the k-wave simulator [39] in the Supplementary Material accompanying this paper.

Next, we will formalize the model laid out in the previous paragraphs into a linear signal model. Assume that the output signal of the transducer, ignoring the electro-mechanical transfer function for now, is the integral of the pressure field over the transducer surface (after propagation through the mask). If we denote the spatial 3D coordinate-vectors that include all points in space on the transducer surface by X, then the transducer output signal $y(t)$, given the spatio-temporal pressure field $y(t, \mathbf{r})$ on the receiving

transducer surface (after propagating through the mask), is:

$$y(t) = \int_{\mathbf{r} \in X} y(t, \mathbf{r}) d\mathbf{r}, \tag{4.1}$$

where $\mathbf{r}$ is a 3D spatial vector. For our model, if the sensor surface X is discretized into S channels, this can be approximated by

$$y(t) = \sum_{s=0}^{S-1} y(t, \mathbf{r}_s), \tag{4.2}$$

where $\{\mathbf{r}_s\}_{s=0}^{S-1}$ are the sampled positions on the transducer surface.

In the remainder of this work we will work in the frequency domain instead of the time domain. Needless to say, equations (4.1) and (4.2) are equally valid in the frequency domain if $y(t)$ and $y(t, \mathbf{r}_s)$ are replaced by their frequency domain equivalents. We will denote the sampled continuous time Fourier transform of $y(t)$ and $y(t, \mathbf{r}_s)$ by $y[\omega]$ and $y_s[\omega]$, respectively, with $\omega \in \Omega$, sampled at frequencies in the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{N-1}\}$. We will use equidistantly taken samples in the pulse bandwidth of the positive side of the frequency spectrum.

Next, we want to express the pressure fields $y_s[\omega]$ on the transducer surface in terms of the image we try to recover. The pixelized image consists of M pixel reflection coefficients, denoted by the vector $\mathbf{x} \in \mathbb{R}^M$. We assume that the transmitted wavefield to each pixel is known. By using the Born approximation (see e.g. [40, 41]), we can define a measurement model that is linear with respect to the reflection coefficients; any measurement $y_s[\omega]$ is a superposition of the individual pulse-echo signals $\{a_{s,m}[\omega]\}_{m=0}^{M-1}$ of all scatterers, ignoring multiple reflections between scatterers:

$$y_s[\omega] = \sum_{m=0}^{M-1} x[m] a_{s,m}[\omega]. \tag{4.3}$$

Here, $x[m]$ is the m-th component of $\mathbf{x}$, corresponding to the scattering intensity of pixel m, and $a_{s,m}[\omega]$ is the pulse-echo signal of pixel m at frequency ω measured at position $\mathbf{r}_s$.

Using the approximate model described earlier, and illustrated in Fig. 4.2, we have the following expression for the term $a_{s,m}[\omega]$ in (4.3):

$$a_{s,m}[\omega] = \frac{p[\omega]}{4\pi(\|\mathbf{r}_{px,m}\|_2 + \|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2 + \|[\mathbf{r}_{ch,s}]_z\|_2)} \times$$
$$\exp\left(-j\frac{\omega}{c_0}(\|\mathbf{r}_{px,m}\|_2 + \|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2) - j\frac{\omega}{c_1}\|[\mathbf{r}_{ch,s}]_z\|_2\right). \tag{4.4}$$

Here, $\mathbf{r}_{px,m}$ is the 3D position vector of pixel m, and $\mathbf{r}_{ch,s}$ is the 3D position vector of the mask-medium interface of channel s. The variables c_0 and c_1 are the medium and mask speed of sound, respectively. The signal $p[\omega]$ is the frequency domain representation of the excitation pulse $p(t)$. The first term accounts for the geometric spreading of the pressure field, and the exponential term is the phase shift due to the travelling distance from the transmitting transducer to the pixel and back to the masked receive sensor. In

other words, it is the total delay of a single path from transmit to the 'output' of a single mask channel, as visualized in Fig. 4.2. This total path and its corresponding delays can be broken down into several parts, following next. The distance $\|\mathbf{r}_{px,m}\|_2$ is the transmit path length from the transmitter to pixel m, and the sum $\|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2 + \|[\mathbf{r}_{ch,s}]_z\|_2$ is the total receive path from pixel m to the mask-medium interface of channel s, and the echo propagation through the channel. Similarly, the first terms in the exponent yield a phase shift due to the transmit/receive paths in the medium, whereas the last term represents the phase shift due to the propagation through mask channel s.

Storing all samples $\{y_s[\omega]\}_{\omega \in \Omega}$ in the vector $\mathbf{y}_s \in \mathbb{C}^N$, and all $\{a_{s,m}[\omega]\}_{\omega \in \Omega}$ in the vectors $\mathbf{a}_{s,m} \in \mathbb{C}^N$, we can use matrix-vector notation to write (4.3) as:

$$\begin{aligned} \mathbf{y}_s &= \begin{bmatrix} \mathbf{a}_{s,0} \ \mathbf{a}_{s,1} \ ... \ \mathbf{a}_{s,M-1} \end{bmatrix} \mathbf{x} \\ &= \mathbf{A}_s \mathbf{x}. \end{aligned} \tag{4.5}$$

Similarly, we can rewrite (4.2) into its discretized, frequency domain counterpart as:

$$\mathbf{y} = \sum_{s=0}^{S-1} \mathbf{y}_s = \sum_{s=0}^{S-1} \mathbf{A}_s \mathbf{x} = \mathbf{A}\mathbf{x}, \tag{4.6}$$

with $\mathbf{A} = \sum_{s=0}^{S-1} \mathbf{A}_s$. Assuming that measurements are corrupted by complex Gaussian noise in the frequency domain, our final measurement model becomes:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{4.7}$$

where $\mathbf{n} \in \mathbb{C}^N$ represents complex Gaussian noise with covariance matrix $\mathbf{C_n}$: $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_n})$. Thus, the imaging problem consists of estimating $\mathbf{x}$ from $\mathbf{y}$. If additional measurements are acquired by, for example, translation of the masked transducer, they can be easily incorporated by adding additional rows to this system of equations (see (4.10) in the next section). In the next section, we parameterize $\mathbf{A}$ as a function of the mask shape in order to optimize for the mask geometry.

In (4.6), the total measurement matrix $\mathbf{A}$ is separated into independent channels, resulting in a summation of matrices. Our convex mask optimization algorithm in the next section relies on the fact that we can pre-compute each $\mathbf{A}_s$ for a pre-determined number of mask thickness levels. However, it is impossible to pre-compute $\mathbf{A}_s$ if it is not known what pressure field is emitted by each scatterer, which in turn can only be known if the transmit mask is known. Although in [23, 24] we used the same mask in transmit and receive, in this work we assume that a separate (co-located) transducer is used for transmit, and the masked transducer for receive. This makes it always possible to pre-compute any $\mathbf{A}_s$ using (4.4), circumventing this problem. Specifically, throughout this paper, we will assume that a point source (without a mask), located in the center of the receiver, transmits a spherically diverging wave.

Finally, we point out that the greedy algorithm proposed in this paper does not suffer from this limitation. However, for the sake of consistency and comparison to the convex algorithm, we will use the non-masked co-located transmitter setup throughout this paper for all simulations.
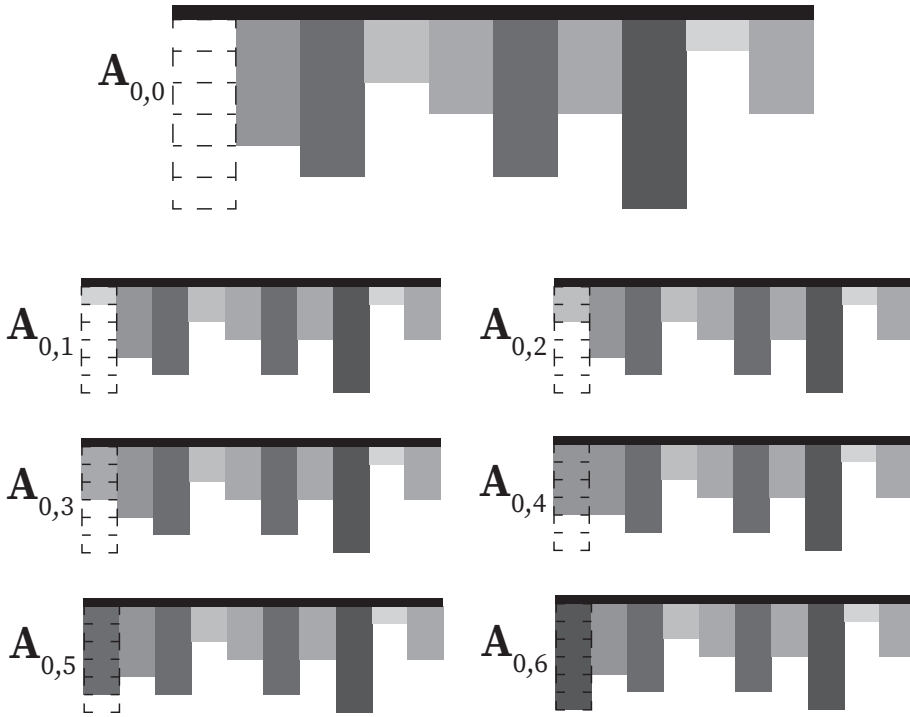
Figure 4.3: Each mask channel is discretized in the mask thickness dimension, and each mask thickness level is associated with a specific measurement matrix for that channel. The goal of the mask optimization problem is to select one thickness level per channel such that the imaging MSE is minimized.

4.3. MASK OPTIMIZATION BY SENSOR SELECTION

Having formulated the linear measurement model in the previous subsection, we now want to use it to find how to optimally choose the mask thickness levels, i.e., we want to find the best channel lengths or delays. We will again use discretization (of the channel lengths), and show that channel length design can be posed as a sensor selection problem. Such problems are hard to optimize exactly, but can be either relaxed into a convex problem formulation, or approximately solved using a greedy optimization algorithm. Using a specific thickness level for a channel is equivalent to using a particular matrix $\mathbf{A}_s$ for that channel (Fig. 4.3). Consequently, our goal is to choose one particular $\mathbf{A}_s$ per mask channel, such that the trace of the error covariance matrix of the mask-encoded measurement system is optimal.

First, we rewrite (4.5) in terms of the measurement matrices for all thickness levels. Suppose there are R potential thickness levels for each channel s. Hence, there are R candidate matrices $\{\mathbf{A}_{s,0},\mathbf{A}_{s,1},\ldots,\mathbf{A}_{s,R-1}\}$ that can be used for $\mathbf{A}_s$ at channel s. Let $w[r + Rs] \in \{0,1\}$ indicate whether sensor candidate $r \in \{0,1,\ldots,R-1\}$ is used for channel s or

not. Then (4.7) can be rewritten as

$$\mathbf{y}(\mathbf{w}) = \sum_{s=0}^{S-1} \sum_{r=0}^{R-1} w[r + Rs]\mathbf{A}_{s,r}\mathbf{x} \tag{4.8}$$

where $\mathbf{w} = [w[0], w[1], \ldots, w[RS-1]]^\top$. To make our notation simpler, we define $\mathbf{A}[r + Rs] = \mathbf{A}_{s,r}$, and linearize s and r into a single index $t = r + Rs$. Consequently we can write (4.8) as:

$$\mathbf{y}(\mathbf{w}) = \sum_{t=0}^{RS-1} w[t]\mathbf{A}[t]\mathbf{x} \tag{4.9}$$
$$= \mathbf{A}(\mathbf{w})\mathbf{x},$$

where $\mathbf{A}(\mathbf{w}) = \sum_{t=0}^{RS-1} w[t]\mathbf{A}[t] = \sum_{s=0}^{S-1} \sum_{r=0}^{R-1} w[r + Rs]\mathbf{A}_{s,r}$. Note that in this formulation only one sensor should be selected per channel, resulting in only S non-zero values for $\mathbf{w}$, and each non-zero value has to be within a specific support range of $\mathbf{w}$. The vector $\mathbf{w}$ is the mask optimization variable, and has to lie within the constraints just described.

In case K different pulse-echo events are taken using the same coding mask, the corresponding measurement equations can be stacked into a single system of equations, assuming $\mathbf{x}$ does not change in between measurements:

$$\mathbf{y}(\mathbf{w}) = \begin{bmatrix} \mathbf{y}_0(\mathbf{w}) \\ \mathbf{y}_1(\mathbf{w}) \\ \vdots \\ \mathbf{y}_{K-1}(\mathbf{w}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0(\mathbf{w}) \\ \mathbf{A}_1(\mathbf{w}) \\ \vdots \\ \mathbf{A}_{K-1}(\mathbf{w}) \end{bmatrix} \mathbf{x} = \mathbf{A}(\mathbf{w})\mathbf{x}. \tag{4.10}$$

For example, the sensor and the mask can be rotated (as in [23]), or translated (as in [24]) to obtain additional information of $\mathbf{x}$. Note however, that the same mask is used in each case, and that $\mathbf{A}(\mathbf{w})$ is still linear in $\mathbf{w}$.

This model can also be generalized to include multiple measurements involving differently shaped masks. In that case, a number of Q measurements with Q distinct masks are stacked vertically into a larger system of equations:

$$\begin{bmatrix} \mathbf{y}(\mathbf{w}_0) \\ \mathbf{y}(\mathbf{w}_1) \\ \vdots \\ \mathbf{y}(\mathbf{w}_{Q-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\mathbf{w}_0) \\ \mathbf{A}(\mathbf{w}_1) \\ \vdots \\ \mathbf{A}(\mathbf{w}_{Q-1}) \end{bmatrix} \mathbf{x}, \tag{4.11}$$

which we denote as

$$\mathbf{y}(\mathcal{W}) = \mathbf{A}(\mathcal{W})\mathbf{x},$$

where $\mathcal{W}$ represents the set $\{\mathbf{w}_q\}_{q \in \mathcal{Q}}$, $\mathcal{Q} = \{0, 1, \ldots, Q-1\}$. Again, $\mathbf{A}(\mathcal{W})$ is still linear in $\mathcal{W}$. The aforementioned constraint on the number of selected channels means that if $\mathbf{w}_{q,s}$ only contains the entries of $\mathbf{w}_q$ that correspond to the R selection coefficients for

channel s, i.e., $\mathbf{w}_{q,s} = \left[\, w_{q,s}[0],\ w_{q,s}[1],\ \ldots,\ w_{q,s}[R-1] \,\right]^{\mathsf{T}}$, then we want $\|\mathbf{w}_{q,s}\|_0 = 1$ for all channels s.

To optimally select mask thickness levels, we use $\hat{\mathbf{x}}$ to denote an estimate of $\mathbf{x}$, and define the image estimation error as $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. We will utilize the covariance matrix of this image estimation error throughout this work, defined as

$$\mathbf{C_e} = \mathbb{E}\left((\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathsf{H}}\right). \tag{4.12}$$

We will try to find an optimized mask by minimizing the mean squared error (MSE) under additive zero-mean i.i.d. white Gaussian measurement noise, which is a function of $\mathbf{C_e}$. Since $\mathbf{A}(\mathscr{W})$ is typically ill-conditioned, some prior information about the solution is required. We assume that the covariance matrix $\mathbf{C_x}$ of $\mathbf{x}$ is known, and that a Wiener estimate of the image is used, so that we can minimize the Bayesian MSE (the error covariance associated with the Wiener estimator). More specifically, we minimize

$$f(\mathscr{W}) = \frac{1}{M}\text{trace}(\mathbf{C_e}) \tag{4.13}$$

$$= \frac{1}{M}\text{trace}\left(\left(\mathbf{C_x}^{-1} + \mathbf{A}(\mathscr{W})^{\mathsf{H}}\mathbf{C_n}^{-1}\mathbf{A}(\mathscr{W})\right)^{-1}\right) \tag{4.14}$$

where $\mathbf{C_x}$ and $\mathbf{C_n}$ are the image and noise covariance matrices, respectively. Throughout this paper we assume that $\mathbf{C_x} = \sigma_x^2 \mathbf{I}$ and $\mathbf{C_n} = \sigma_n^2 \mathbf{I}$.

With these choices for the covariance matrices, the Wiener estimator can be rewritten as a regularized least squares estimator, i.e., $\hat{\mathbf{x}} = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y}(\mathscr{W}) - \mathbf{A}(\mathscr{W})\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$, where λ depends on the ratio between σ_x^2 and σ_n^2. In this case, $\lambda\|\mathbf{x}\|_2^2$ only serves to make sure that a unique solution is found without large outliers in $\hat{\mathbf{x}}$, due to the ill-posedness of the inverse imaging problem, as well as to increase robustness to noise. From this perspective, minimizing $f(\mathscr{W})$ can also be seen as a way to decrease the ill-posedness of the inverse imaging problem, by finding a mask that increases the orthogonality between the columns of $\mathbf{A}(\mathscr{W})$. As a result, we expect that the resulting mask will also be beneficial for other (non-linear) imaging algorithms besides the Wiener estimator used here, since it is expected to decrease the ill-posedness of the problem.

Furthermore, it is expected that the optimal solution of $f(\mathscr{W})$ is not overly sensitive to incorrect values for σ_x^2 and σ_n^2, as long as the following two extreme cases are avoided. The first one where the ratio of these variances causes the inverse term in $f(\mathscr{W})$ to be dominated by $\mathbf{C_x}^{-1}$, and the second case where $\mathbf{C_x}^{-1}$ is negligible, so that no regularizing term is present. By choosing these values sufficiently in between these two extreme cases, we hope to find a mask that is relatively robust to the variances of $\mathbf{x}$ and $\mathbf{n}$.

Ideally, the optimization problem is then posed as

$$\min_{\mathscr{W}} \quad f(\mathscr{W}) \tag{4.15}$$

$$\text{s.t.} \quad \|\mathbf{w}_{q,s}\|_0 = 1$$

$$\mathbf{w}_q \in \{0,1\}^{RS}$$

for all $q \in \mathscr{Q}$ and $s \in \mathscr{S}$. However, neither the objective function nor the constraints are convex with respect to $\mathscr{W}$, and it is hard to solve this problem exactly. Hence, we will relax (4.15) into a problem that is convex, and also propose a more efficient coordinate descent type algorithm.

4.3.1. MASK OPTIMIZATION BY CONVEX RELAXATION

To find a convex relaxation of (4.15), both the cost function and the two constraints in (4.15) have to be relaxed. We will discuss each of these separately.

1. In the typical sensor selection problem, $\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W})$ is an affine function of the selection variables $\mathcal{W}$, and the resulting cost function is convex since the composition of a convex function with an affine function is convex. In our case, however, the input matrix is not affine with respect to the selection variables, and hence our objective function is not necessarily convex. To see that it is indeed not, note that, using (4.9) and (4.11), $\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W})$ is equal to:

$$\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W}) = \sum_{q=0}^{Q-1}\sum_{t=0}^{RS-1}\sum_{u=0}^{RS-1} w_q[t]\,w_q[u]\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u], \tag{4.16}$$

which is quadratic w.r.t. $\mathcal{W}$. As a simple counter-example of the convexity of (4.13), consider the simple case where $Q = 1$, so that $\mathbf{A}(\mathcal{W})$ only depends on a single selection vector $\mathbf{w}$, and consider that each $\mathbf{A}[t] \in \mathbb{C}^{1\times1}$. Then

$$\mathbf{C_e}^{-1} = \sigma_x^{-2} + \sigma_n^{-2}\mathbf{A}(\mathbf{w})^{\mathsf{H}}\mathbf{A}(\mathbf{w}) \tag{4.17}$$

$$= \sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{B}\mathbf{w}, \tag{4.18}$$

where the coefficients of $\mathbf{B} \in \mathbb{S}_+^{RS\times RS}$ consist of the 1×1 matrices $\{\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u]\}_{t,u=0}^{RS-1}$ (using (4.16)). Consequently, $\mathrm{trace}(\mathbf{C_e}) = 1/(\sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{B}\mathbf{w})$, and it is straightforward to see that this function is not convex with respect to $\mathbf{w}$.

To circumvent this problem, we absorb the quadratic terms $w_q[t]\,w_q[u]$ into a single variable $W_q[t, u]$:

$$\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W}) = \sum_{q=0}^{Q-1}\sum_{t=0}^{RS-1}\sum_{u=0}^{RS-1} W_q[t, u]\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u]. \tag{4.19}$$

Whereas the cost-function (4.13) was quadratic in the elements of $\mathbf{w}_q$, it is linear with respect to the variables $W_q[t, u]$. Expression (4.16) is only equivalent to (4.19) if the variables $W_q[t, u]$ have a certain structure. Specifically, if we store the variables $W_q[t, u]$ in a matrix $\mathbf{W}_q$ such that $\left[\mathbf{W}_q\right]_{t,u} = W_q[t, u]$, then equivalence between (4.16) and (4.19) only holds if $\mathbf{W}_q = \mathbf{w}_q\mathbf{w}_q^{\mathsf{T}}$. In other words, $\mathbf{W}_q$ has to be a matrix with rank$(\mathbf{W}_q) = 1$. Although the cost-function is now convex with respect to $\mathbf{W}_q$, the additionally required rank constraint is not. Keeping in mind that if $\mathbf{w}_q \in \{0, 1\}^{RS}$, then diag$(\mathbf{w}_q\mathbf{w}_q^{\mathsf{T}}) = \mathbf{w}_q$, a common technique [42, 43] to relax the constraint $\mathbf{W}_q - \mathbf{w}_q\mathbf{w}_q^{\mathsf{T}} = \mathbf{0}$ is:

$$\mathbf{W}_q - \mathbf{w}_q\mathbf{w}_q^{\mathsf{T}} \succeq \mathbf{0}, \tag{4.20}$$

$$\mathrm{diag}(\mathbf{W}_q) = \mathbf{w}_q \tag{4.21}$$

for all $q \in \mathcal{Q}$.

2. For the second relaxation, we relax the l_0 quasi-norm to the ℓ_1 norm:

$$\|\mathbf{w}_{q,s}\|_1 = 1. \tag{4.22}$$

Since each $\mathbf{W}_q \succeq \mathbf{0}$, the diagonals of each $\mathbf{W}_q$ are non-negative, and thus we can write this constraint as $\mathbf{w}_{q,s}^\top \mathbf{1} = 1$.

3. Finally, we have to relax the binary constraint $\mathbf{w}_q \in \{0,1\}^{RS}$. We relax it to the box constraint

$$\mathbf{w}_q \in [0,1]^{RS}. \tag{4.23}$$

This constraint is already implied by the previous two relaxations. Since each $\mathbf{W}_q \succeq 0$, the diagonals of each $\mathbf{W}_q$ are non-negative. Hence the entries of $\mathbf{w}_q$ are all greater or equal to zero. Together with this observation, due to the constraint $\mathbf{w}_{q,s}^\top \mathbf{1} = 1$, the diagonal elements of each $\mathbf{W}_q$ cannot be larger than one. Consequently, the constraint $\mathbf{w}_q \in [0,1]^{RS}$ does not have to be explicitly enforced in our optimization routine.

So far we have not taken into account the full structure that a rank-1 matrix with entries in $\{0,1\}$ and $\|\mathbf{w}_{q,s}\|_0 = 1$ has. We will thus further exploit the following two properties that $\mathbf{W}_q$ should exhibit.

1. First, if $\mathbf{w}_q$ is a solution to the original problem (4.15), the matrix

$$\mathbf{W}_q = \mathbf{w}_q \mathbf{w}_q^\top \tag{4.24}$$
$$= \begin{bmatrix} \mathbf{W}_{q|0,0} & \mathbf{W}_{q|0,1} & \cdots & \mathbf{W}_{q|0,S-1} \\ \mathbf{W}_{q|1,0} & \mathbf{W}_{q|1,1} & \cdots & \mathbf{W}_{q|1,S-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{q|S-1,0} & \mathbf{W}_{q|S-1,1} & \cdots & \mathbf{W}_{q|S-1,S-1} \end{bmatrix}$$

will have S block-matrices of size $R \times R$ on its diagonal that are themselves diagonal. In other words, each matrix $\mathbf{W}_{q|s,s}$ is in the set $\mathscr{D}$ of $R \times R$ diagonal matrices.

2. If $\|\mathbf{w}_{q,s}\|_0 = 1$ for all s and q, then there is only one non-zero entry in each submatrix of (4.24), and that entry should be equal to one. This implies

$$\mathbf{1}^\top \mathbf{W}_{q|s,s'} \mathbf{1} = 1 \tag{4.25}$$

for all $s, s' \in \{0, 1, \ldots, S-1\}$. Seen from a different point of view, we impose an ℓ_1 sparsity constraint on all the $R \times R$ block-matrices $\mathbf{W}_{q|s,s'}$ instead of just on all vectors $\mathbf{w}_{q,s}$. In fact, the sparsity constraint $\mathbf{1}^\top \mathbf{w}_{q,s} = 1$ is implied by the two constraints we just described, and hence can be omitted now.

3. It is straightforward to see that each entry of $\mathbf{W}$ should be non-negative, so we add the constraint

$$\mathbf{W} \geq 0. \tag{4.26}$$

We will use the set $\mathcal{W}' = \{\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_{Q-1}\}$ to summarize the relaxed problem as follows:

$$\min_{\mathcal{W}'} \quad f(\mathcal{W}') \tag{4.27}$$

$$\text{s.t.} \quad \mathbf{W}_q - \text{diag}(\mathbf{W}_q)\text{diag}(\mathbf{W}_q)^{\mathsf{T}} \succeq 0$$

$$\mathbf{W}_{q|s,s} \in \mathscr{D}$$

$$\mathbf{1}^{\mathsf{T}}\mathbf{W}_{q|s,s'}\mathbf{1} = 1$$

$$\mathbf{W} \geq 0$$

for all $s, s' \in \mathscr{S}$ and $q \in \mathscr{Q}$. Note that, as discussed before, the box and sparsity constraints are implicitly included. Since all the constraints on $\mathbf{w}_q$ are now enforced through constraints on the matrices $\mathbf{W}_q$, we replaced the constraints (4.20) and (4.21) by the first constraint in (4.27).

4.3.2. RANDOM ROUNDING SCHEME

The solution to the convex program above is not guaranteed to have rank 1 with entries according to the original discrete problem (4.15). One could take the leading eigenvector of $\hat{\mathbf{W}}_q$ as the closest solution rank-1 estimate, however it will not necessarily obey the constraints in the original problem formulation (4.15), and one would still have to round the vector to the discrete solution set in (4.15). Instead, we will interpret $\hat{\mathbf{W}}_q$ as a covariance matrix to generate multiple candidate solutions, and project those solutions to the original constraint set. This approach is favored in literature since it has proven itself in many studies, and even has theoretical performance guarantees for a number of discrete optimization problems [44–46]. Although we are currently unable to provide a lower bound on the optimality of this approach, we have observed that this approach typically outperforms solutions obtained from the leading eigenvector. Note that, if $\hat{\mathbf{W}}_q$ turns out to be rank-1, the random vectors generated using $\hat{\mathbf{W}}_q$ as covariance matrix will anyway be equal to the leading eigenvector of $\hat{\mathbf{W}}_q$, up to a scalar value.

To describe this in more detail, we use $\hat{\mathbf{W}}_q$ as a covariance matrix, and generate many random $\tilde{\mathbf{w}}_q \in \mathbb{R}^{RS}$ from a Gaussian distribution:

$$\tilde{\mathbf{w}}_q \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{W}}_q), \tag{4.28}$$

and round each $\tilde{\mathbf{w}}_q$ by selecting the mask thickness level per channel with the highest absolute value:

$$[\hat{\mathbf{w}}_{q,s}]_r = \begin{cases} 1, & \text{if} \quad r = \arg\max_v |[\tilde{\mathbf{w}}_{q,s}]_v| \\ 0, & \text{otherwise} \end{cases}, \tag{4.29}$$

assuming each $\hat{\mathbf{w}}_{q,s}$ has a unique maximum. If there are multiple maxima, we randomly pick one of them.

The intuition behind the random rounding method comes from the following observation (adapting the interpretation in [44] for the problem discussed here). Consider again the example in (4.17), where $Q = 1$, and each $\mathbf{A}[t] \in \mathbb{C}^{1 \times 1}$. The original optimization problem cost function can then be rewritten as

$$\min_{\mathbf{w}} (\sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{B}\mathbf{w})^{-1}.$$

One could instead replace $\mathbf{w}$ by a random variable $\boldsymbol{v}$ which has covariance $\mathbf{W}$, $\boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{W})$, and try to minimize:

$$\min_{\mathbf{W} \in \mathbf{S}_+} \mathbb{E}_{\boldsymbol{v}} \left((\sigma_x^{-2} + \sigma_n^{-2} \boldsymbol{v}^\mathsf{T} \mathbf{B} \boldsymbol{v})^{-1} \right). \tag{4.30}$$

That is, find a covariance matrix $\mathbf{W}$ for the random variable $\boldsymbol{v}$, such that the expected value of the cost function above is minimized. Since all terms are now scalar values, this can be rewritten as (maximizing the inverse of the above equation, using the properties of the trace, and using $\mathbb{E}\boldsymbol{v}\boldsymbol{v}^\mathsf{T} = \mathbf{W}$):

$$\max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{v}} \left(\boldsymbol{v}^\mathsf{T} \mathbf{B} \boldsymbol{v} \right)) \tag{4.31}$$

$$= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{v}} \left(\mathrm{trace}(\boldsymbol{v}^\mathsf{T} \mathbf{B} \boldsymbol{v}) \right))$$

$$= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{v}} \left(\mathrm{trace}(\boldsymbol{v}\boldsymbol{v}^\mathsf{T} \mathbf{B}) \right))$$

$$= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathrm{trace}(\mathbf{W}\mathbf{B})).$$

This is similar in form to a lifted problem, since we have replaced $\boldsymbol{v}\boldsymbol{v}^\mathsf{T}$ by a matrix $\mathbf{W}$. This result suggests that random solutions generated using $\mathbf{W}$ will give good solutions on average, when $\mathbf{W}$ is a higher-dimensional, lifted, version of $\mathbf{w}$.

4.3.3. GREEDY MASK OPTIMIZATION

In this subsection we further describe a much more efficient greedy selection algorithm, since the convex program (4.27) scales polynomially with increasing problem size (more channels, thickness levels, etc.). It is based on the observation that minimizing over only a single channel has a low computational cost, since we would only have to solve the problem

$$\min_{\mathbf{w}_{q,s}} f(\mathcal{W}), \text{ s.t. } \|\mathbf{w}_{q,s}\|_0 = 1, \tag{4.32}$$

which is easily solved by trying all R possible solutions for the current channel s and measurement q. The algorithm operates by iterating through all channels, selecting the thickness level that most decreases the adopted cost function per channel. This is repeated, until either a local minimum or a maximum number of iterations K_{max} is reached. The algorithm is described in Alg. 1. In Alg. 1, we use $f(\mathbf{w}_{q,s})$ to indicate that f is evaluated with $\mathbf{w}_{q,s}$ as a variable, but keeping all other $\mathbf{w}_{q',s'}$ fixed.

Referring back to the description of the approximate mask model, we point out that one could also directly define each $\{\mathbf{A}_s\}_{s\in\mathcal{S}}$ as a function of the mask thickness parameter for each channel. We were able to cast it as a discrete selection problem by sampling the continuous variable for the mask thickness, say $\mathbf{z} \in \mathbb{R}^S$, inside the valid mask thickness range. Keeping this in mind, Alg. 1 acts as a coordinate descent algorithm on $\mathbf{z}$, by optimizing only one element (channel) in $\mathbf{z}$ at a time. Instead of doing a line search over the allowed mask thickness range, it uses a one-dimensional grid search over R predefined values. This line search is relatively cheap, since the maximum mask thickness

---

Algorithm 1 Greedy optimization scheme

1: Input: $\mathcal{W} = \{\mathbf{w}_q\}_{q=0}^{Q-1}$, I_{max}
2: Output: $\hat{\mathcal{W}} = \{\hat{\mathbf{w}}_q\}_{q=0}^{Q-1}$
3: $\hat{\mathbf{w}}_q = \mathbf{w}_q$, $q = 0, 1, \dots, Q-1$
4: $i = 0$
5: **do**
6:     $f_{min} = f(\hat{\mathcal{W}})$
7:     **for** $q = 0, 1, \dots, Q-1$ **do**
8:         **for** $s = 0, 1, \dots, S-1$ **do**
9:             $\hat{\mathbf{w}}_{q,s} = \arg\min_{\mathbf{w}_{q,s}} f(\mathbf{w}_{q,s})$, s.t. $\|\mathbf{w}_{q,s}\|_0 = 1$
10:            $i = i + 1$
11:        **end for**
12:    **end for**
13: **while** $f(\hat{\mathcal{W}}) < f_{min}$, and $i \le I_{max}$

---

is typically confined to a small number. Using a grid search has major advantages, since the proposed algorithm now easily avoids local minima in the current $[\mathbf{z}]_s$ it is optimizing (if the grid on $[\mathbf{z}]_s$ is sufficiently dense). In this way, many local minima are easily avoided, nor does the algorithm need to compute the cost-function gradient or higher order derivatives.

The greedy routine can either be used to refine the rounded solution (4.29) or it can be initiated using an appropriate starting mask. In the first case, the convex solution forms a good starting point, and Alg. 1 will improve it slightly using few iterations. In the latter case, running the convex program (4.27) is avoided at the risk of arriving at a local minimum. In the following numerical experiments, we will use a flat starting mask as an initial guess for the greedy algorithm.

Finally, we point out that this algorithm is not dependent on pre-computing all $\mathbf{A}[t]$, and can be used for configurations where the transmitted pulse passes through the mask both on transmit and receive, using the same transducer for both events. This is due to the fact that $\mathbf{A}$ can be re-computed for the current selection of mask-channels per inner iteration of Alg. 1. For the sake of comparison to the convex program, we have not further investigated this situation.

4.4. Simulation results

To evaluate the proposed mask design algorithms, we consider various imaging scenarios. We first consider two scenarios with a small problem size, where we show that our methods perform well, compared to the true optimum, which we find using an exhaustive search. After these example problems, we consider some larger problems where we design masks for more realistic imaging scenarios. In all simulations, except the first one, we use a Gaussian-modulated excitation pulse $e(t) = w(t)\cos(2\pi f_0 t)$, with $f_0 = 4$ MHz. The function $w(t) = \exp(-(t/f_s)^2/(2\sigma_w^2))$ is a smooth Gaussian window, with f_s the sampling frequency, and $\sigma_w = 3.74$. This pulse is shown in Fig. 4.4. For the greedy algorithm, we set $I_{max} = 5QS$ in all simulations. For the medium we use a speed of sound
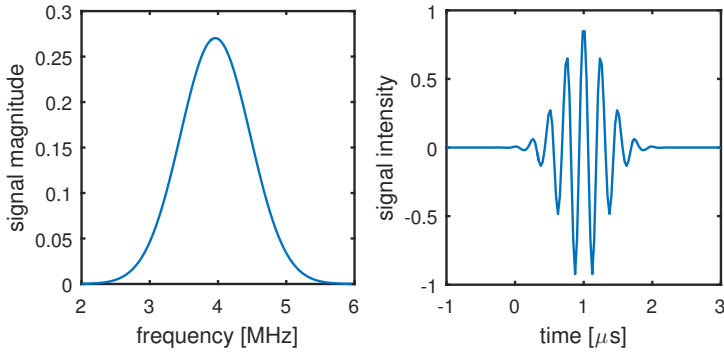
Figure 4.4: Frequency spectrum and time-domain signal of the transmit pulse used in our simulated experiments.

of $c_0 = 1491$ m/s, and we assume that the plastic mask has a speed of sound $c_1 = 2730$ m/s. We consider the following three simulation scenarios throughout this section:

Scenario 1 Rotation of a masked circular transducer in the (x, y) plane.

Scenario 2 Linear translation of a masked linear transducer in the x direction, moving in the x dimension.

Scenario 3 Using a different mask per measurement, with fixed transducer position.

4.4.1. SMALL EXAMPLE 1
In this example (using *scenario 2*), we choose all parameters such that we can find the true optimal mask of (4.15) by an exhaustive search within reasonable time. Although the resulting problem parameters do not correspond to a physically meaningful scenario (due to a very small ROI, and using only a few temporal frequencies), it is the only way to compare how our algorithms perform compared to the true optimum (and hence mostly tests the mathematical validity of our methods). To do this, we investigate a 2D problem, meaning that all pixel and channel positions are within the (x, z) plane with $y = 0$. We use a Cartesian pixel grid of 8x2 ($M = 16$), with pixel coordinates equidistantly placed between -2 and 2 mm in the x-dimension, and at 0.5 and 1.0 cm in the z-dimension, with $y = 0$ for all pixel positions. The non-zero measurement frequencies are equidistantly spaced between 5 and 6 MHz ($N = 11$), which corresponds to taking 120 samples in the time-domain for a sampling frequency of 12 MHz. The mask channels ($S = 8, R = 7$) are equidistantly spaced between -.45 and .45 mm in the x-dimension, at $y = 0$ and $z = 0$, with a maximum mask thickness of 0.39 mm. This corresponds to discretization steps 2.8 times smaller than the centre wavelength in water (.37 mm) in the x-dimension (0.13 mm channel width), and 5.7 times smaller in the mask thickness dimension (0.065 mm per channel height segment). The noise variance was chosen such that the output SNR

$$\text{SNR} = \text{trace}\{\mathbb{E}\{\mathbf{yy}^H\}\}/\text{trace}\{\mathbb{E}\{\mathbf{nn}^H\}\} \tag{4.33}$$

$$= \text{trace}\{\sigma_x^2 \mathbf{AA}^H\}/\text{trace}\{\sigma_n^2 \mathbf{I}\} \tag{4.34}$$
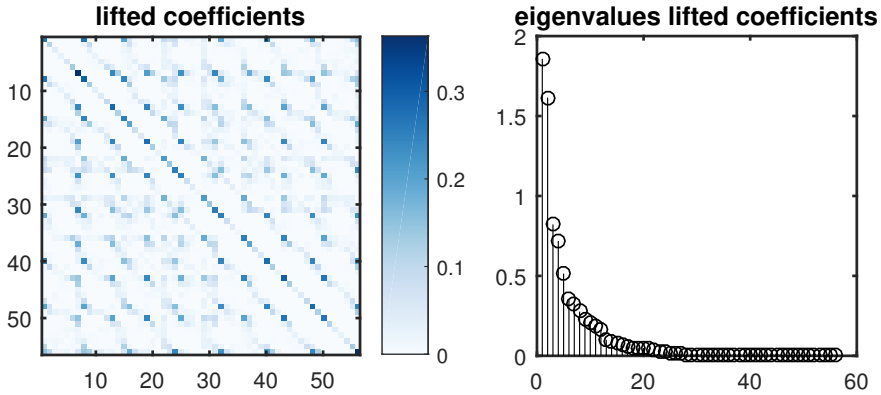
Figure 4.5: Left: optimal solution $\hat{\mathbf{W}}$ to problem (4.27) for the simulation described in Sec. 4.4.1 . Right: sorted eigenvalue spectrum of the matrix to the left.

of a random mask is approximately[2] 0 dB. To add more structure to the problem, we add more measurements using the same mask ($Q = 1, K = M = 16$), by spatially shifting the masked transducer to the same x-coordinates as the pixels in the x-dimension (*scenario 2*).

In Fig. 4.5, the solution $\hat{\mathbf{W}}$ to the convex problem (4.27) is visualized. The constraints are able to obtain a matrix that is already quite sparse, and of low rank. Fig. 4.6 shows the MSE (4.13) distribution for a set of 10,000 randomly generated sets of masks, 10,000 masks generated using random rounding based on the solution of the convex problem (4.27), the MSEs of the proposed methods, the MSE of the true optimum, as well as a the MSE of a completely flat mask. The latter can still reconstruct images since multiple spatial measurements are obtained by translation of the transducer. The random rounding procedure generates solutions with a higher MSE concentration close to the true optimum than the distribution of completely random masks, and the majority of these masks also outperform the flat mask scenario. The greedy algorithm is able to find the true optimum, and provides a mask that is slightly better than the mask obtained using the convex program (4.27), while being much more computationally efficient.

For a complete performance description, we take the optimized and random masks from Fig. 4.6, and compute their MSE if the SNR changes, but the same mask is used. The resulting MSEs are shown in Fig. 4.7 (Fig. 4.6 shows the MSE distribution for one particular SNR in Fig. 4.7). The vertical bars in Fig. 4.7 indicate the mean MSE plus or minus the MSE standard deviation for 1,000 randomly generated masks. Even though all methods are optimized for a particular SNR, Fig. 4.7 shows that all methods are able to generate a mask that performs well over the entire SNR range, although most of the MSE can be gained in the -15 to 10 dB range. As a benchmark, we also show the MSE of the optimal mask per SNR, as well as the MSE of the mask obtained using the greedy algorithm by

---

[2]Note that, even if σ_x^2 is known, we can only determine $\|\mathbf{y}\|_2^2$ once the mask is known, since each mask can have a different measurement energy for the same image $\mathbf{x}$. Hence, we cannot choose σ_n^2 to give an exact output SNR. In this example, the exact output SNR of the optimized mask is 2.8 dB.
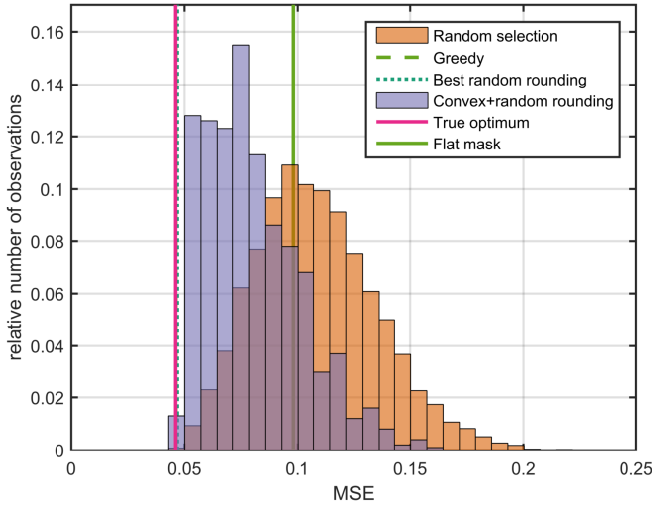
Figure 4.6: MSE distribution for the problem described in Sec. 4.4.1. We show the MSE distribution for a set of 10,000 randomly generated masks, and 10,000 masks generated using the proposed random rounding procedure, the MSEs of the proposed methods, the MSE of the best mask, as well as a flat mask.

optimizing per SNR (the best mask can be different for each SNR). In this example, we can see that all proposed masks, even if they are optimized for a single SNR, are able to perform close to optimal over the entire SNR range, demonstrating the robustness of the optimization method to incorrect values of the noise and image covariance matrices.

Finally, we show the PSF in dB (computed as $10\log_{10}$) for the second pixel (Fig. 4.8). We define the PSF for pixel n as the image when the measurement is equal to: $\mathbf{y} = \mathbf{a}_n(\mathcal{W})$, the n-th column of $\mathbf{A}(\mathcal{W})$. For comparison, we took the PSF of a random mask with an expected MSE of 0.106, close to the average performance of random masks according to the distribution in Fig. 4.6. The PSFs of the different masks re-affirm the performance as predicted by Fig. 4.6.

4.4.2. SMALL EXAMPLE 2

In the previous example, we constrained the problem parameters to be able to use an exhaustive search to find the global optimum. In this second small example, we make the problem slightly larger, by using more channels and multiple masks, and using *scenario 3* as described in Sec. 7.2. As a result, the global optimum cannot be found within reasonable time, and is not shown in the results of this experiment. Furthermore, the distribution of random masks will be further away from the true optimum, unlike in Fig. 4.6. Using random rounding, however, we can still do well.

Instead of shifting the transducer, it remains in the same position, but uses $Q = 12$ different masks to obtain different measurements. We optimize the mask for 5 pixels on a line at $z = 5$ mm, located between -1 and 1 mm. The mask channels are located between -.45 and .45 mm ($S = 8$), with a maximum mask thickness of .26 mm ($R = 5$).
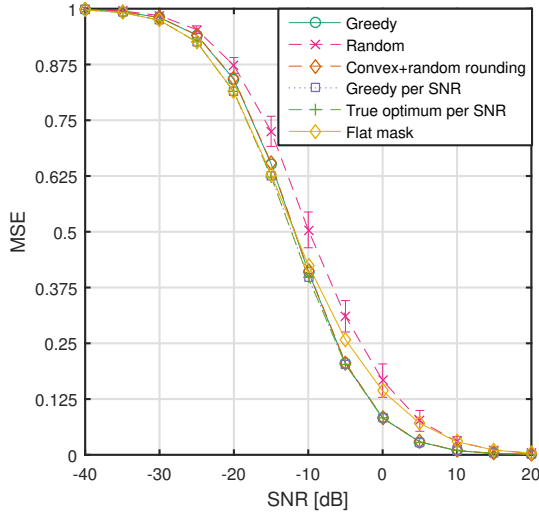
Figure 4.7: Performance of various mask design strategies in terms of MSE for various output SNR scenarios, for the problem described in Sec. 4.4.1. Each curve shows the performance of a single mask that was optimized for an SNR of 0 dB. The crosses represent the mean MSE of 1,000 randomly designed masks, with the vertical bars showing the standard deviation. The MSE of the true optimal mask per SNR is shown, as well as the MSE obtained by using a greedy algorithm for each SNR.

The measurement frequencies are equidistantly spaced between 5 and 6 MHz ($N = 101$).

We generate 1,000 random masks and 1,000 randomly rounded masks, as well as a greedily generated mask. The MSE distributions are shown in Fig. 4.9. As mentioned before, the random mask design strategy does not work well, and covers a higher range of MSEs. In Fig. 4.10 the PSF is shown for the optimized image domain. As predicted by Fig. 4.9, some performance gain is possible, although not as great as in the previous simulation in Sec. 4.4.1.

4.4.3. MASK DESIGN FOR A SINGLE ROTATING MASK

In this subsection we optimize a single rotating mask (*scenario 1*). We assume that all the channels are located within a 5 mm disc. The only difference between the measurements is a rotation of the sensor and the mask relative to the pixels. We only consider the greedy approach so that we can use realistic problem sizes with acceptable computation times. We discretize the mask into 1245 square channels of equal size (0.13 mm), equivalent to approximately 2.9 channels per wavelength. Each channel was divided into 11 uniformly distributed mask thickness levels (0.13 mm per mask thickness segment), with a maximum mask thickness of 0.65 mm, equivalent to approximately 5.7 channels per wavelength. The 2D ROI of 5 by 5 mm is discretized into 20 by 20 pixels at a depth of 6 mm. The transducer is rotated 40 times, evenly divided over 360 degrees ($K = 40$). Note that, although there are multiple measurements, we optimize for a single mask ($Q = 1$). During optimization, σ_n^2 was chosen such that the expected SNR of a random mask is
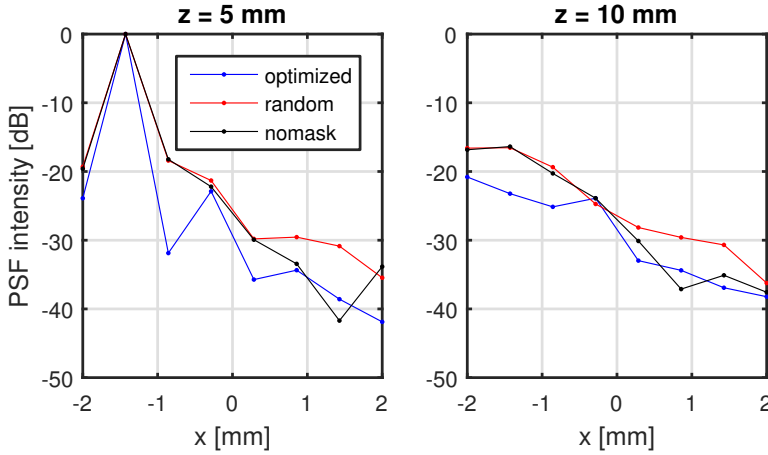
Figure 4.8: PSF for a pixel at $x = -1.4$mm, $z = 5$mm for the scenario of Sec. 4.4.1. The subfigures show slices of the PSF at different depths.

approximately 10 dB. We use the transmit pulse shown in Fig. 4.4, sampling it in the frequency domain using $\Delta f = 40$ kHz, or $N = 101$.

The imaging performance for various SNRs is shown in Fig. 4.11, although the mask was optimized for a single SNR. Again, it is clear that the optimized mask is able to consistently outperform all random masks. Moreover, the MSE distribution of the random masks is relatively narrow, meaning that it would take an enormous amount of tries to randomly generate a mask that performs just as well as our proposed techniques. Furthermore, the expected SNR (4.33) for the optimal mask is equal to 15.5 dB, whereas the SNR averaged over 100 random masks, is equal to 9.0 dB, a gain of 6.5 dB for the optimized setup.

To illustrate the imaging performance, we take a letter-shaped test image (zero-mean and unit-energy, averaged over all pixels) to generate a noisy test measurement $\mathbf{y}$. Fig. 4.12 shows the greedily optimized mask and a particular randomly generated mask. Fig. 4.13 shows some example estimated images using the masks in Fig. 4.12, choosing σ_n^2 such that the optimized mask has an SNR of 20 dB for the test phantom. The random mask, in contrast, causes measurements to have an SNR of 13.1 dB using the same noise variance and test phantom. We see from Fig. 4.13 that the optimized mask is better able to correctly estimate the phantom.

The aforesaid SNR gain is caused by a focusing effect that becomes apparent if we visualize the pulse-echo amplitude per pixel for a large region and a single frequency, shown in Fig. 4.14. The mask acts as a spatial filter that boosts energy of echoes originating from the ROI, whereas a random mask does not have this capability, resulting in a lower SNR.

In addition to an increased SNR, the optimized mask also causes lower similarity between pulse-echo signals of different pixels. We visualize this in Fig. 4.15 by plotting the normalized PSF (computed as $\mathbf{A}^H \mathbf{a}_i$, with each column of $\mathbf{A}$ normalized so that the Euclidean norm is one) for various points in space. The optimized mask strongly decorre-
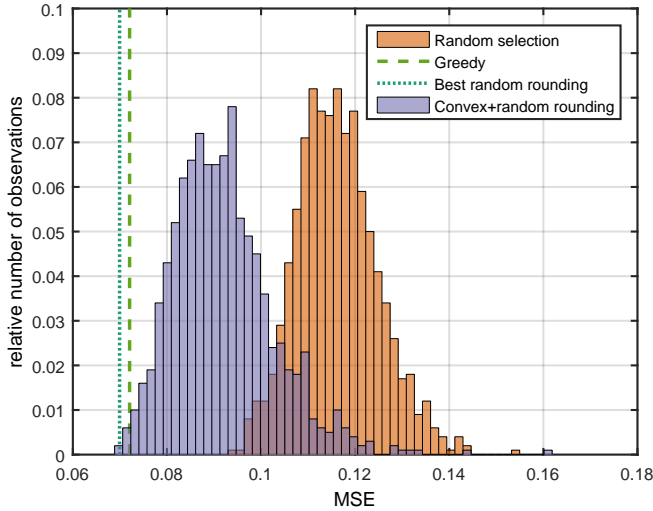
Figure 4.9: MSE distribution for the problem of Sec. 4.4.2, for a set of 1,000 randomly generated masks, as well as 1,000 masks generated using the proposed random rounding procedure, and the MSEs of the proposed methods.

lates between pulse-echo signals from different pixels, whereas the random mask always has relatively high correlation across almost the entire ROI. The PSF is typically best at the edges, where the difference between different measurements obtained by rotation is the largest.

These results show that the proposed design algorithm makes columns of $\mathbf{A}(\mathcal{W})$ more orthogonal. This implies that other, possibly non-linear imaging methods will also benefit from the designed mask, since the increased orthogonality generally reduces ambiguity about $\mathbf{x}$, making the estimation of $\mathbf{x}$ easier for *any* imaging method.

4.4.4. MASK DESIGN FOR MULTIPLE MASKS

For our final simulation example, we consider an experiment where the transducer is not rotated or translated in any way, but instead the mask is changed between pulse-echo measurements (*scenario 3*). We use the greedy method to find $Q = 40$ masks, and compare the performance to 40 randomly generated masks. All other parameters are the same as in the previous simulation for a single rotating mask (Sec. 4.4.3). We choose σ_n such that the SNR of a random mask is approximately 10 dB.

Using the greedy approach with a flat starting mask, we obtain the masks shown in Fig. 4.16. In Fig. 4.18 we show the corresponding SNR curves. The expected SNR of the optimized mask is 13.7 dB, whereas the random mask SNR is 9.8 dB, averaged over 100 random masks, an SNR gain of nearly 4 dB.

Fig. 4.17 shows an example reconstruction for an SNR of 20 dB. We can see that using the optimized masks, the true image is better reconstructed compared to a set of random masks. From Fig. 4.18 we observe an MSE improvement for all SNRs. Since the random
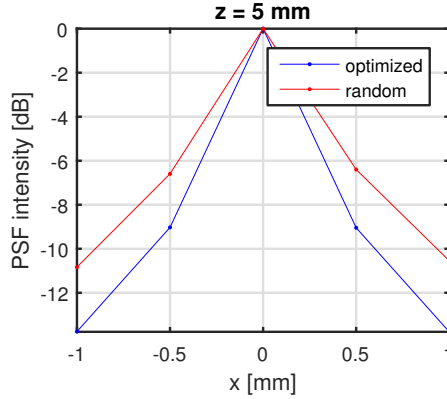
Figure 4.10: PSF for the simulations of Sec. 4.4.2 for a pixel at $x = 0$mm, $z = 5$mm.

masks show a very narrow MSE distribution, the optimized mask outperforms nearly all random masks in terms of the MSE (4.13). Although the MSE gain may not seem like much, the example reconstruction in Fig. 4.17 visualizes that a small MSE difference can have a non-negligible impact on the visual reconstruction quality. We observe that the multiple mask setup exhibits better imaging performance than a rotating mask with the same amount of measurement data. Finally, we remark that the imaging performance for this setup is better than the rotating mask configuration, since it does not restrict all masks to be related by rotation (although the algorithm is free to output such a set of masks).

### 4.4.5. MASK DESIGN FOR A ROTATING MASK WITH PIXELS ON A CIRCLE
In this subsection we analyse a particular example that admits an intuitive physical interpretation of the results. We use a single mask ($Q = 1$), and obtain new measurements by rotating the mask ($K = 30$). To make the results easily interpretable, we only consider pixels uniformly spaced on a circle with a radius of 1.8 mm ($M = K = 30$), where the center of the circle coincides with the center of rotation of the masked transducer.

Consider the measurement matrix $\mathbf{A}_{\omega_k}$, containing only the measurements for frequency ω_k. Assuming the angle of rotation is equivalent to the angle between pixels, each row or column is a circular shift of the previous row. As a result, each $\mathbf{A}_{\omega_k}$ is a circulant matrix, and the total measurement matrix is a concatenation of circulant matrices:

$$\mathbf{A} = \left[\mathbf{A}_{\omega_0}^{\mathsf{T}} \ \mathbf{A}_{\omega_1}^{\mathsf{T}} \ \dots \ \mathbf{A}_{\omega_{K-1}}^{\mathsf{T}} \right]^{\mathsf{T}}. \tag{4.35}$$

Consequently, the singular values $\mathbf{s} \in \mathbb{R}_+^M$ of the SVD of $\mathbf{A}$ can be expressed as the square root of the element-wise sum of the squared magnitude of the eigenvalues over all circulant submatrices:

$$\mathbf{s} = \sqrt{\sum_{k=0}^{K-1} |\mathbf{s}_{\omega_k}|^2}, \tag{4.36}$$
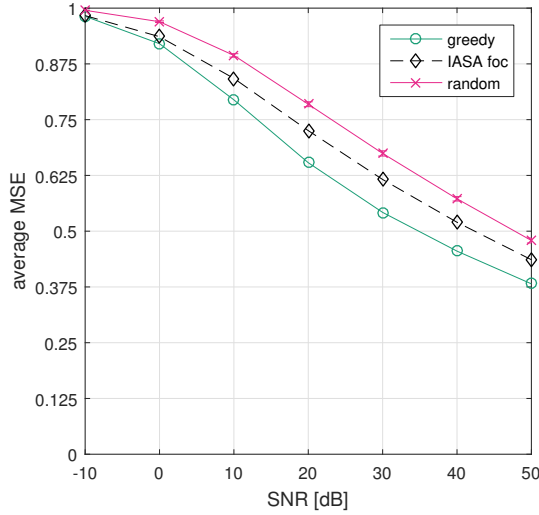
Figure 4.11: MSE performance for the scenario of Sec. 4.4.3, for various SNR scenarios, optimizing for a single rotating mask. The crosses represent the mean MSE of 100 randomly designed masks, with the vertical bars showing the standard deviation. The IASA method is discussed in the Discussion section of this paper.

where $\mathbf{s}_{\omega_k}$ is the vector of eigenvalues of submatrix $\mathbf{A}_{\omega_k}$. Since each submatrix is circulant, its eigenvalues are equivalent to the DFT spectrum of its first column. Therefore, we can replace each $\mathbf{s}_k$ by the DFT of the first column of $\mathbf{A}_{\omega_k}$:

$$\mathbf{s} = \sqrt{\sum_{k=0}^{K-1} |\hat{\mathbf{a}}_{\omega_k}|^2}, \tag{4.37}$$

where $\hat{\mathbf{a}}_{\omega_k}$ is such a DFT. Note that in (4.36) and (4.37), the square root and magnitude operators are assumed to be taken element wise.

This yields an intuitive way of comparing two masks. Due to the reciprocity principle, the columns/rows of $\mathbf{A}_{\omega_k}$ can be interpreted in two ways. First, they can be viewed as the transducer output signal for each pixel, if only that single pixel is transmitting a wavefield $p(\omega)$. Alternatively, they can be seen as the complex pressure field measured at the pixel locations if the transducer is *transmitting* the pulse $p(\omega)$ through the mask towards the pixel locations. Since the singular values of $\mathbf{A}$ can be expressed as a function of the Fourier transformed columns of the sub-matrices $\mathbf{A}_{\omega_k}$, the mask with the best (i.e., slowest decrease of the) singular value spectrum is the mask that generates a pressure field that has high spatial bandwidth, i.e., has a balanced spectrum with its energy spread across all the spatial frequencies as evenly as possible and has a high total energy.

For this particular case, the mask can also be analyzed from another point of view: the measured signal (per temporal frequency) equals a circular convolution of the image map, and a good mask results in a filter that preserves as much frequency content of the image as possible.
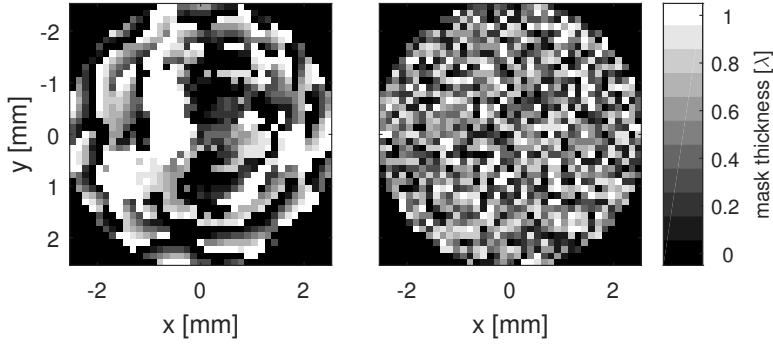
Figure 4.12: Optimized and random mask used for the results in Fig. 4.13, using a single sensor with a rotating mask, for the simulation scenario of Sec. 4.4.3. The mask thickness values are displayed as a fraction of the wavelength inside the mask.

We illustrate the preceding with a simulated example. In addition to the parameters mentioned, we use the same transmit pulse as in Fig. 4.4 ($N = 101$ or $\Delta f = 40$ kHz). All channels ($S = 1245$) are located on a circle of 5 mm radius. We define $R = 6$ mask thickness levels up to a thickness of 0.65 mm, and $M = 30$ pixels on a ring of 5 mm diameter.

In Fig. 4.19, we show the spatial frequency spectra of the columns of all $\mathbf{A}_{\omega_k}, \omega_k \in \Omega$, in addition to $\mathbf{s}$ as computed in (4.37), and the singular values $\mathbf{s}$ sorted according to their magnitude. We do this for the greedily optimized mask as well as for a particular random mask. From Fig. 4.19 we observe that the optimization algorithm tries to find a mask such that the total spatial frequency bandwidth over the ring of pixels is increased. As a result, the singular values of the optimized matrix decrease more slowly, and are better conditioned. This is achieved due to each temporal frequency generating energy in a specific spatial frequency. For example, in the left panel of Fig. 4.19, the blue temporal frequencies generate high spatial frequencies, whereas temporal frequencies in the red spectrum focus energy in the lower spatial frequencies. The random mask is unable to accomplish the same result, and depressions in the blue curve are not compensated by any other temporal frequencies. We further observe that the total energy is much higher for the optimized mask, resulting in an increased SNR.

4.5. DISCUSSION

In Figs. 4.14 and 4.19 we showed that an optimized mask (for a rotating measurement configuration) acts as a spatial filter, enhancing echoes coming from within the ROI. In the angular frequency domain this means that the mask should act as an angular filter with a passband for directions within the ROI. Hence, as a rule of thumb, the structure within the mask should contain spatial features within the same angular frequency band. For example, if pixels are located in the far-field at an incident angle of 0 degrees, the mask can coherently sum signals from this ROI using a flat mask. On the other hand, if pixels are located at an incident angle of 90 degrees, a sinusoidal profile with the same period as the pulse wavelength will coherently sum echoes from that direction, and average out echoes originating from 0 degrees. This is more complicated in the near field,
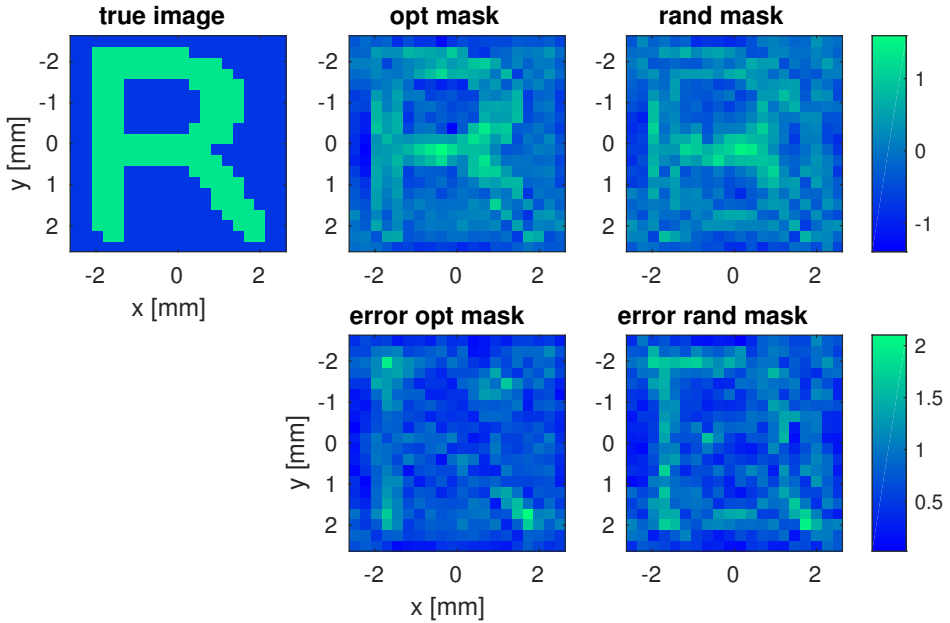
Figure 4.13: Reconstruction of a letter R image for a single sensor with a rotating mask, using 40 rotations, for the simulation of scenario Sec. 4.4.3. The top figures show the true image and its reconstructions using both the optimized mask and a random mask. The bottom figures show the error maps for each reconstruction result.

but the same principle holds. To illustrate this point, we show several masks optimized for ROIs parallel to the transducer (for simplicity, we only define pixels on a circle with the same (x, y)-coordinates as the edge of the mask) at different depths (Fig. 4.20). As the ROI is located further away, its angular range decreases, and consequently, spatial features in the mask profile contain a lower range of spatial frequencies.

Beside angular filtering, the mask should decorrelate echoes from different pixels. Optimized masks thus don't consist of pure harmonics, but of more complex patterns, in order to bring about this decorrelation. As discussed in Sec. 4.4.5, for a ring of pixels this means that the optimized mask should cause high variation of the wavefield w.r.t. this ring. Generalizing this to a fully sampled ROI by adding more rings is more complicated, but one could imagine that different wavelength intervals in the transmitted pulse can focus on different rings. One should, roughly speaking, try to design a mask that if used for transmit, causes a high spatial variation in the transmitted wavefield. Hence, an alternative design methodology would consist of first defining a wavefield at the ROI with high spatial bandwidth, and propagate it back in time towards the transducer to obtain a phase mask (similar to [37, 38]). Needless to say, this should be co-designed with the desired angular filtering discussed before, and algorithms presented in this paper try to find a balance between these two criteria.

Still, we have explored this alternative design strategy as suggested in the above study. First, we designed a pressure field for a single frequency ($f = 4$ MHz at $z = 6$ mm), with
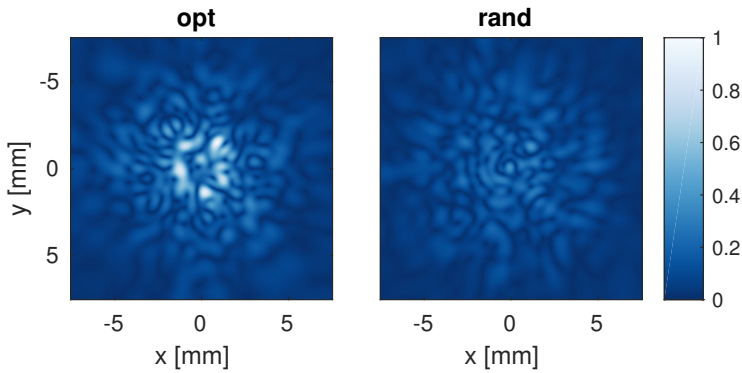
Figure 4.14: Normalized amplitude per pixel at $f = 4$ MHz for an optimized rotating mask (Sec. 4.4.3) in Fig. 4.12, and a random mask. Note that the visualized energy extends beyond the imaging ROI, which is only 5 by 5 mm.
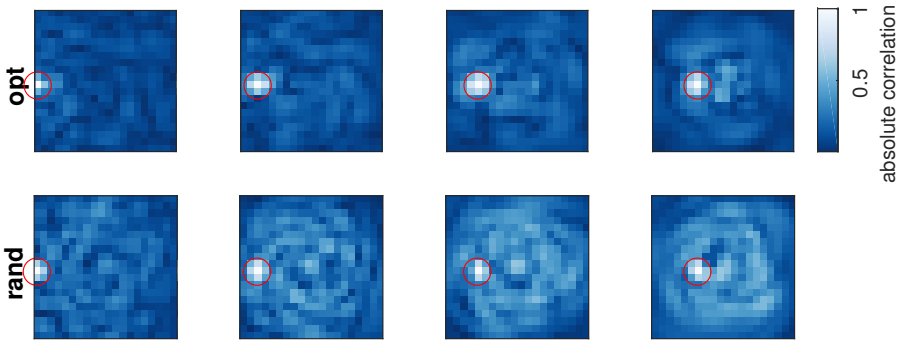
Figure 4.15: Absolute PSFs for various points in space, indicated by the red circle. Top: using the greedily optimized mask. Bottom: using a randomly selected mask. The columns in **A** are energy-normalized before computing the PSFs. Simulation scenario of Sec. 4.4.3.

high spatial bandwidth, and which has most energy focused inside the pixel ROI within this z-plane. We then use the Iterative Angular Spectrum Approach (IASA) to approximate a phase mask that generates this pressure plane, following the method proposed in [37]. This method starts with a flat mask, finds the associated pressure field in the image plane, and compares it to the desired pressure field. It then changes the complex amplitudes in the obtained forward field to the desired amplitudes, but keeping the phases unchanged. The field is then propagated back in time to the transducer plane, where the mask profile is adapted to cause the required phase changes to generate the back-propagated pressure field. This is repeated until the algorithm converges to a phase mask. For further details on this method, we refer to the original paper [37]. As a desired pressure field we generated a white complex Gaussian field, and applied a Gaussian window in both the space and k-space domain to obtain a focused bandlimited target field

for the image plane. The exact details for the target field are shown in Appendix 4.A.

The general performance of the resulting mask for Gaussian images is included in Fig. 4.11. We also visualize the reconstructed image for the resulting mask in Fig. 4.21, where it is compared to the mask obtained with the proposed greedy method. The IASA-based design is superior to random mask design, but is not able to surpass the method proposed in this work, although the difference in reconstruction quality for the non-Gaussian (sparse) phantom in Fig. 4.21 is relatively small (the reconstruction MSEs for this particular phantom are 0.65 and 0.70 for the greedy and IASA-based method, respectively). Of course, the crucial design step in this strategy amounts to choosing a proper target field for the phase mask to produce, and in this experiment we have tried to find a good target field by trial-and-error. However, it is not clear what the 'best' forward field would be. Although initial investigation agrees with our suggestions in this paragraph, (e.g., a focused forward field seemed to give better results than an unfocused beam), broad-bandedness of the spatial target distribution does not seem to be sufficient to acquire a good mask with this method. Hence, this design strategy can probably be significantly improved by choosing a more intelligent pressure distribution than the one used in this section.

4.5.1. Mask design for multiple sensors

The optimization framework discussed in this work could also be extended to include multiple sensors, each with possibly their own mask. In that case, one would have the following measurement equation for L sensors, combining equations (4.10) and (4.11):

$$\mathbf{y}(\mathscr{W}) = \begin{bmatrix} \mathbf{y}_0(\mathbf{w}_0) \\ \mathbf{y}_1(\mathbf{w}_1) \\ \vdots \\ \mathbf{y}_{L-1}(\mathbf{w}_{L-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0(\mathbf{w}_0) \\ \mathbf{A}_1(\mathbf{w}_1) \\ \vdots \\ \mathbf{A}_{L-1}(\mathbf{w}_{L-1}) \end{bmatrix} \mathbf{x} = \mathbf{A}(\mathscr{W})\mathbf{x}, \tag{4.38}$$

where each $\mathbf{A}_l, (l = 0, 1, \dots, L)$ represents the measurement equations for sensor l. The equations are still linear with respect to $\mathscr{W}$, and the same proposed techniques can be applied.

4.6. Conclusions

Using an approximate linear signal model, we formulated mask optimization as an *additive* sensor selection problem by discretizing each mask channel in the thickness dimension, and optimizing for the imaging MSE, which is a function of the measurement matrix $\mathbf{A}$. Each candidate sensor corresponds to a full measurement matrix, and the final measurement model is the *summation* of the individual sensors' measurement matrices. We have shown that this is not a conventional sensor selection problem, and proposed to solve this problem by transforming the original variables into a larger amount of 'lifted' variables. By further relaxing the problem, and by exploiting the structure among the lifted optimization variables, a convex problem is formulated. To get a solution that is feasible in the original problem formulation, a random rounding technique was proposed. In addition to the convex approach, we also proposed a greedy algorithm to solve the same problem.

Although the proposed algorithms are not guaranteed to provide the globally optimal mask, a small numerical experiment showed that our mask design algorithms come close to the true optimum, and all experiments showed that they outperform randomly generated masks in terms of imaging MSE, even for images with non-white and non-Gaussian statistics. The distribution of masks using the combination of convex optimization and random rounding has a lower MSE range than the distribution of completely random masks. Furthermore, the proposed greedy coordinate-descent algorithm obtained good results close to the global optimum. Our experiments demonstrate that a good mask is able to filter in the angular domain, enhancing echoes coming from within the ROI, while strongly decreasing correlation between the pulse-echo signals of pixels within the ROI.

REFERENCES

[1] P. van der Meulen, P. Kruizinga, J. G. Bosch, and G. Leus, *Coding mask design for single sensor ultrasound imaging,* IEEE Transactions on Computational Imaging **6**, 358 (2019).

[2] E. J. Candes and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?* IEEE transactions on information theory **52**, 5406 (2006).

[3] D. L. Donoho, *Compressed sensing,* IEEE Transactions on information theory **52**, 1289 (2006).

[4] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,* IEEE Transactions on information theory **52**, 489 (2006).

[5] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling,* IEEE signal processing magazine **25**, 21 (2008).

[6] H. Liebgott, R. Prost, and D. Friboulet, *Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing,* Ultrasonics **53**, 525 (2013).

[7] C. Quinsac, A. Basarab, and D. Kouamé, *Frequency domain compressive sampling for ultrasound imaging,* Advances in Acoustics and Vibration **2012** (2012).

[8] N. Wagner, Y. C. Eldar, and Z. Friedman, *Compressed beamforming in ultrasound imaging,* IEEE Transactions on Signal Processing **60**, 4643 (2012).

[9] T. Chernyakova and Y. Eldar, *Fourier-domain beamforming: the path to compressed ultrasound imaging,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **61**, 1252 (2014).

[10] M. F. Schiffner and G. Schmitz, *Fast pulse-echo ultrasound imaging employing compressive sensing,* in *Ultrasonics Symposium (IUS), 2011 IEEE International* (IEEE, 2011) pp. 688–691.

4

[11] G. David, J.-l. Robert, B. Zhang, and A. F. Laine, *Time domain compressive beam forming of ultrasound signals,* The Journal of the Acoustical Society of America **137**, 2773 (2015).

[12] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, *Compressed delay-and-sum beamforming for ultrafast ultrasound imaging,* in *Image Processing (ICIP), 2016 IEEE International Conference on* (Ieee, 2016) pp. 2509–2513.

[13] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling,* IEEE signal processing magazine **25**, 83 (2008).

[14] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, *A single-pixel terahertz imaging system based on compressed sensing,* Applied Physics Letters **93**, 121105 (2008).

[15] L. Gao, J. Liang, C. Li, and L. V. Wang, *Single-shot compressed ultrafast photography at one hundred billion frames per second,* Nature **516**, 74 (2014).

[16] A. C. Fannjiang, *Exact localization and superresolution with noisy data and random illumination,* Inverse Problems **27**, 065012 (2011).

[17] R. Fergus, A. Torralba, and W. T. Freeman, *Random lens imaging,* (2006).

[18] A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, *Imaging with nature: Compressive imaging using a multiply scattering medium,* Scientific reports **4**, 5552 (2014).

[19] V. Boominathan, J. K. Adams, M. S. Asif, B. W. Avants, J. T. Robinson, R. G. Baraniuk, A. C. Sankaranarayanan, and A. Veeraraghavan, *Lensless imaging: A computational renaissance,* IEEE Signal Processing Magazine **33**, 23 (2016).

[20] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, *Diffuser-Cam: lensless single-exposure 3D imaging,* Optica **5**, 1 (2018).

[21] F. C. Meral, M. A. Jafferji, P. J. White, and G. T. Clement, *Two-dimensional image reconstruction with spectrally-randomized ultrasound signals,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **60**, 2501 (2013).

[22] N. Etaix, M. Fink, and R. K. Ing, *Acoustic imaging device with one transducer,* The Journal of the Acoustical Society of America **131**, EL395 (2012).

[23] P. Kruizinga, P. van der Meulen, A. Fedjajevs, F. Mastik, G. Springeling, N. de Jong, J. G. Bosch, and G. Leus, *Compressive 3D ultrasound imaging using a single sensor,* Science advances **3** (2017).

[24] J. Janjic, P. Kruizinga, P. van der Meulen, G. Springeling, F. Mastik, G. Leus, J. G. Bosch, A. F. van der Steen, and G. van Soest, *Structured ultrasound microscopy,* Applied Physics Letters **112**, 251901 (2018).

[25] A. Austeng and S. Holm, *Sparse 2-D arrays for 3-D phased array imaging-design methods,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **49**, 1073 (2002).

[26] B. Diarra, M. Robini, P. Tortoli, C. Cachard, and H. Liebgott, *Design of optimal 2-D nongrid sparse arrays for medical ultrasound,* IEEE Transactions on Biomedical Engineering **60**, 3093 (2013).

[27] S. Joshi and S. Boyd, *Sensor selection via convex optimization,* IEEE Transactions on Signal Processing **57**, 451 (2009).

[28] S. P. Chepuri and G. Leus, *Sparsity-Promoting Sensor Selection for Non-Linear Measurement Models.* IEEE Trans. Signal Processing **63**, 684 (2015).

[29] F. Bian, D. Kempe, and R. Govindan, *Utility based sensor selection,* in *Proceedings of the 5th international conference on Information processing in sensor networks* (ACM, 2006) pp. 11–18.

[30] A. Krause and C. Guestrin, *Near-optimal observation selection using submodular functions,* in *AAAI*, Vol. 7 (2007) pp. 1650–1654.

[31] M. Shamaiah, S. Banerjee, and H. Vikalo, *Greedy sensor selection: Leveraging submodularity,* in *Decision and Control (CDC), 2010 49th IEEE Conference on* (IEEE, 2010) pp. 2572–2577.

[32] M. Coutino, S. Chepuri, and G. Leus, *Near-optimal greedy sensor selection for MVDR beamforming with modular budget constraint,* in *Signal Processing Conference (EUSIPCO), 2017 25th European* (IEEE, 2017) pp. 1981–1985.

[33] M. Coutino, S. P. Chepuri, and G. Leus, *Submodular sparse sensing for gaussian detection with correlated observations,* IEEE Transactions on Signal Processing **66**, 4025 (2018).

[34] O. M. Bushnaq, T. Y. Al-Naffouri, S. P. Chepuri, and G. Leus, *Joint sensor placement and power rating selection in energy harvesting wireless sensor networks,* in *Signal Processing Conference (EUSIPCO), 2017 25th European* (IEEE, 2017) pp. 2423–2427.

[35] P. van der Meulen, P. Kruizinga, J. G. Bosch, and G. Leus, *Spatial compression in ultrasound imaging,* in *Signals, Systems, and Computers, 2017 51st Asilomar Conference on* (IEEE, 2017) pp. 1016–1020.

[36] E. J. Candes, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization,* Journal of Fourier analysis and applications **14**, 877 (2008).

[37] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, *Holograms for acoustics,* Nature **537**, 518 (2016).

[38] M. Brown, D. Nikitichev, B. Treeby, and B. Cox, *Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles,* Applied Physics Letters **110**, 094102 (2017).

4

[39] B. E. Treeby and B. T. Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields,* Journal of biomedical optics **15**, 021314 (2010).

[40] M. Verweij, B. Treeby, K. Van Dongen, L. Demi, and A. Brahme, *Simulation of ultrasound fields,* Comprehensive biomedical physics , 465 (2014).

[41] J. T. Fokkema and P. M. van den Berg, *Seismic applications of acoustic reciprocity* (Elsevier, 1993).

[42] C. Helmberg, *Semidefinite programming for combinatorial optimization,* (2000).

[43] A. d'Aspremont and S. Boyd, *Relaxations and randomized methods for nonconvex QCQPs,* EE392o Class Notes, Stanford University , 1 (2003).

[44] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, *Semidefinite relaxation of quadratic optimization problems,* IEEE Signal Processing Magazine **27**, 20 (2010).

[45] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,* Journal of the ACM (JACM) **42**, 1115 (1995).

[46] Y. Nesterov, *Semidefinite relaxation and nonconvex quadratic optimization,* Optimization methods and software **9**, 141 (1998).

4

APPENDICES

4.A. IASA-BASED MASK DESIGN

To obtain a mask using the IASA method, one has to first choose a target pressure field in the image plane that the phase mask is supposed to produce. We designed a focused wavefield with maximum spatial bandwidth, by first generating a completely complex Gaussian random pressure field. Next, we filtered this field by applying a circular, symmetric Gaussian window in the 2D DFT of this field. The radius of this window was chosen approximately equal to the wavenumber in water, choosing $f = 4$ MHz, the central frequency of the transmit pulse. After transforming back to the spatial domain, we again applied a circular Gaussian window, so that most energy is focused in the pixel ROI. We show the resulting pressure field in Fig. 4.22. The pressure field produced by the resulting mask is limited in spatial bandwidth due to the limited size of the aperture.

4
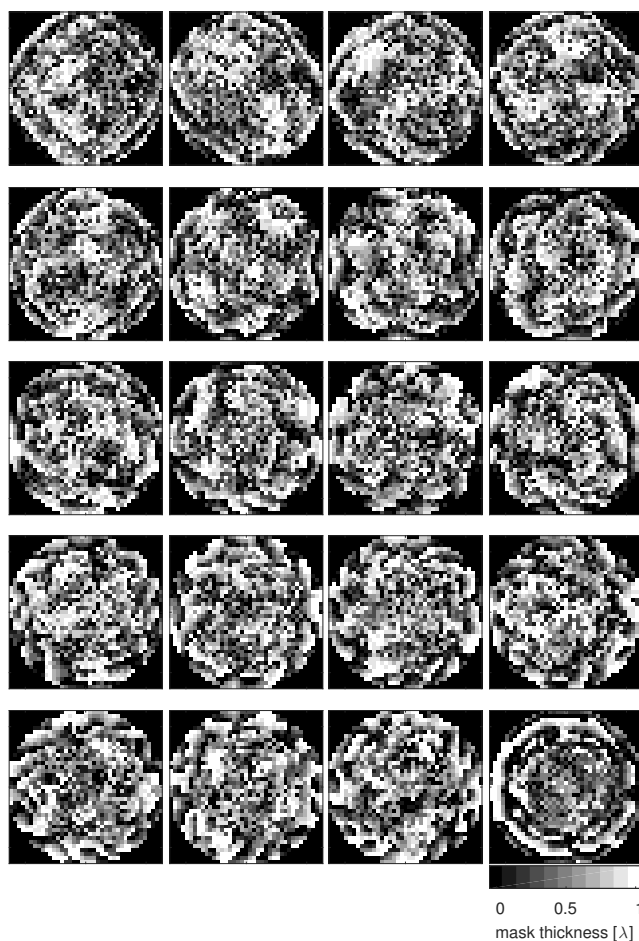
Figure 4.16: A subset of the optimized masks for the problem of Sec. 4.4.4 with $Q = 40$, using Alg. 1. The mask thickness values are displayed as a fraction of the wavelength inside the mask. The algorithm optimizes one mask at a time. The first mask is shown in the top left corner, while the third mask is shown to the right of the first mask, etc. (every second mask is plotted).
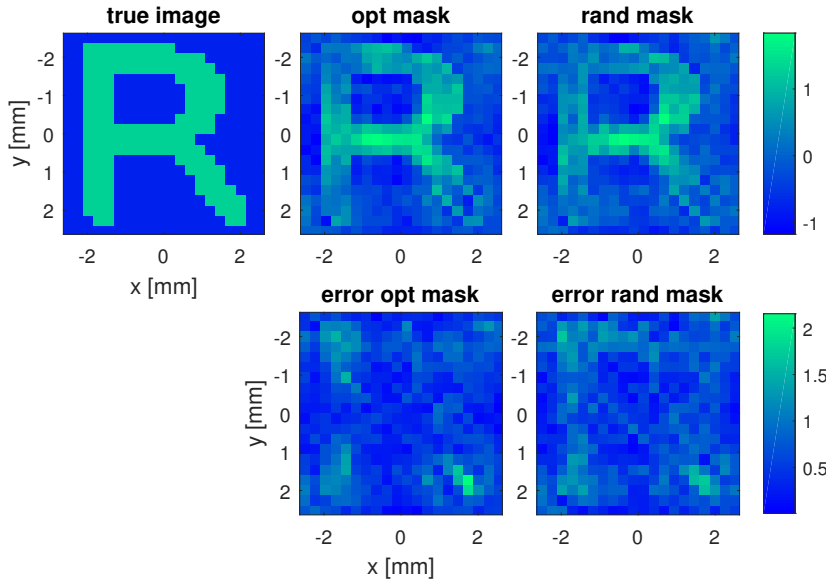
Figure 4.17: Reconstruction of a letter R image combining the measured data of 40 masks obtained using the greedy optimization scheme, for an output SNR of 20 dB, for the simulation scenario of Sec. 4.4.4. The top figures show the true image and its reconstructions using both the optimized mask and a random mask. The bottom figures show the error maps for each reconstruction result.
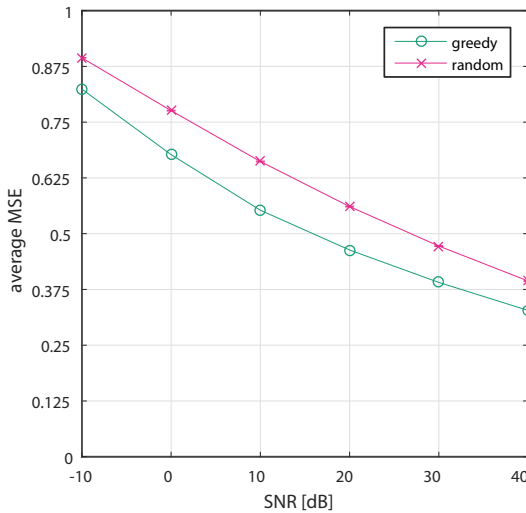
Figure 4.18: MSE performance distribution for various SNR scenarios, optimizing for $Q = 40$ masks. The optimized mask was designed assuming an SNR of 10 dB, but performs relatively well across the entire SNR range. The crosses represent the mean MSE of 100 randomly designed masks, with the vertical bars showing the standard deviation. Simulation scenario of Sec. 4.4.4.
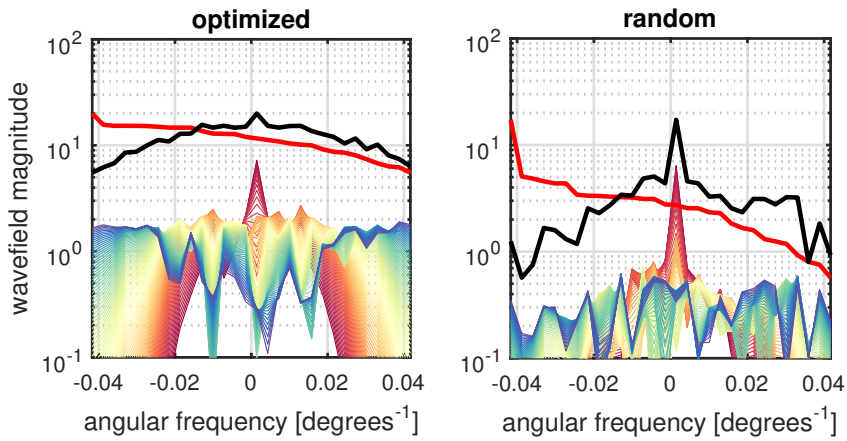
Figure 4.19: Spatial frequency and singular value spectra for the optimized mask and a particular random mask with pixels on a circle (Sec. 4.4.5). The color-coded curves show the spatial frequency spectrum of the complex pressure field of each temporal pulse frequency ω_k (starting with low temporal frequencies in red, ending with high frequencies in blue). The black curve is the compounded spatial frequency spectrum of all temporal frequencies, as computed per (4.37), which is equivalent to the singular values of **A**. The red curve shows the singular values of **A**, sorted by magnitude.
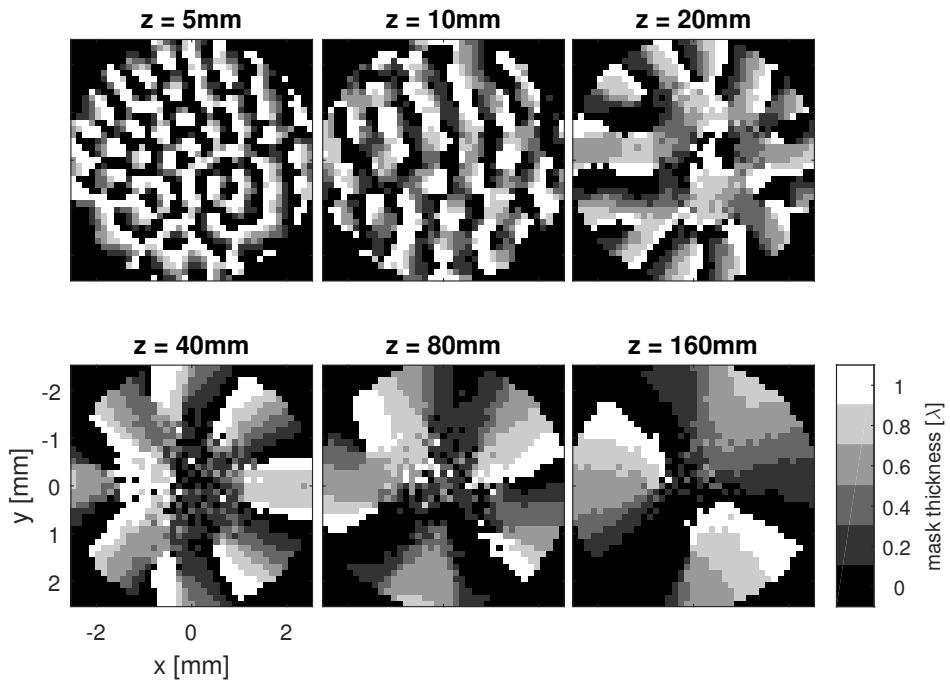
Figure 4.20: Optimized masks for various depths. For simplicity, pixels are positioned on a ring, and the excitation signal is a single frequency.
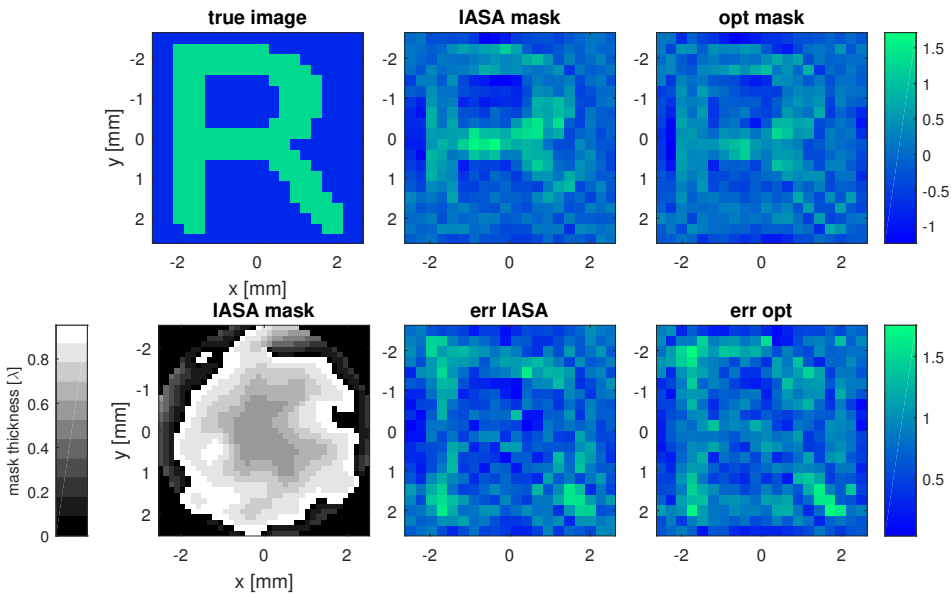
Figure 4.21: Bottom left: coding mask obtained using the IASA-method [37], by finding a phase mask that gives a focused pressure field in the image ROI with high spatial bandwidth. Top/bottom: image reconstruction comparison between the IASA-based mask design method and the proposed method in this work.

4
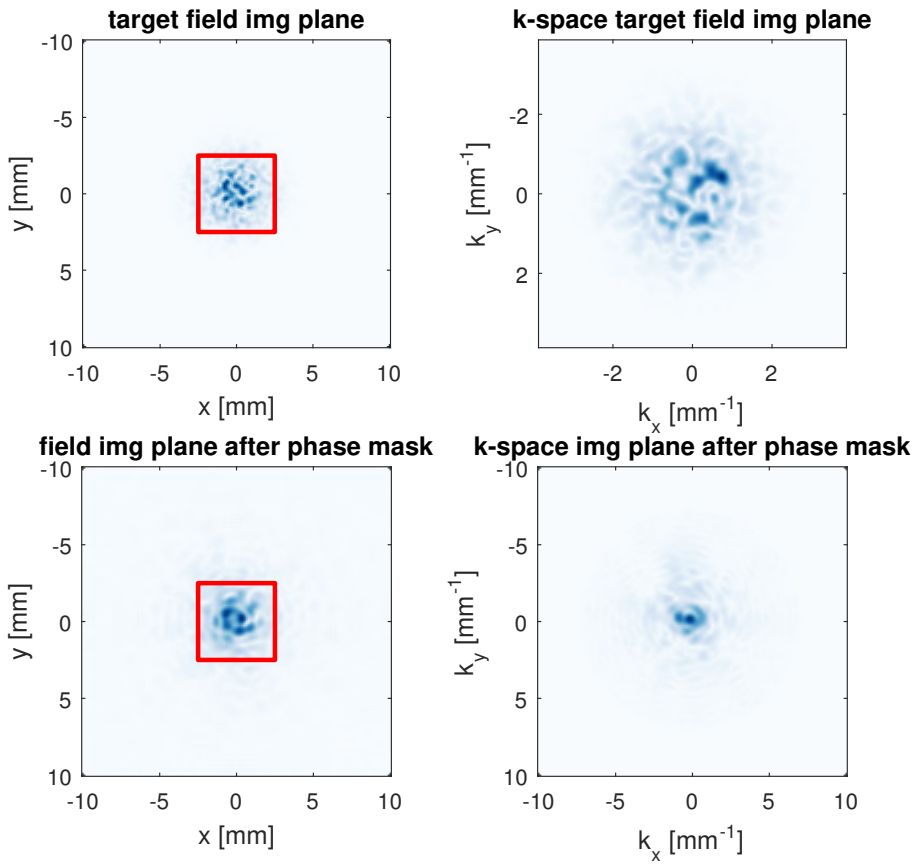
Figure 4.22: Top: target pressure field for $f = 4$ MHz in the image plane. Bottom: the field in the image plane produced with the mask obtained through the IASA method. All figures show the magnitude of the complex pressure fields. The red box indicates the pixel ROI used for optimization in Sec. 4.4.3.

II

PART II

5

IMPULSE RESPONSE ESTIMATION METHOD FOR ULTRASOUND ARRAYS

We present a method for estimating the one-way electro-mechanical impulse response or transfer function of self-reciprocal ultrasound transducers. The one-way impulse response is needed for forward field simulations, or for pulse-echo simulations and excitation code design when the one-way impulse response per array element is different. Using a flat plate reflector that is positioned parallel to the transducer surface, the resulting pulse-echo signal is measured. Since the transducer is self-reciprocal, the transmit and receive impulse responses are equivalent. Consequently, the measured signal is the autoconvolution of the one-way impulse response. We propose a new de-autoconvolution algorithm to obtain the one-way impulse response from such a signal. The proposed measurement procedure is especially time-efficient for large arrays, and does not rely on hydrophones or additional transducers. Experimental results are shown to demonstrate the effectiveness of the proposed method.

5.1. INTRODUCTION

Accurate knowledge of the one-way electro-mechanical impulse response or transfer function is important for e.g. model-driven beamforming techniques and accurate ultrasound simulations. Typical transducer calibration procedures (e.g. [2, 3]) result in an estimate of the *pulse-echo* impulse response, which are only useful for pulse-echo simulations: they estimate the combined effect of the transmit and receive impulse responses. For computing the forward field, however, the *one-way* impulse response is of interest. Another example is that of an array transducer where each element may have a somewhat different impulse response, e.g. due to imprecisions in the transducer production process. As a result, the pulse-echo impulse response from one array element to another may be different for each combination of elements. In this case, it is convenient to know the pulse-echo response for each combination of transmit and receive elements for either simulations, model-based imaging, or excitation code design.

Even if one is only interested in the pulse-echo transfer function, measuring the pulse-echo impulse response for each combination of transmit and receive elements is a very laborious process for large arrays. In that case, it would be more convenient to use a hydrophone to measure the transmitted pulse of each transducer excited by a delta-pulse. The pulse-echo impulse response between transmit/receive pairs is then found by convolving the corresponding impulse responses. However, hydrophone setups are expensive, the procedure is time-consuming, and the measurement is sensitive to the hydrophone position relative to the transducer. Our proposed method does not rely on hydrophone measurements, but uses the pulse-echo measurement of the reflection from a flat reflector. An additional benefit of this method over the hydrophone-based characterization is that the signal to noise ratio (SNR) of the pulse-echo measurements may be much better if the transducer under test has a large measurement surface compared to the hydrophone surface. Moreover, the transducer is more likely to have a high sensitivity in its own bandwidth compared to hydrophones, resulting in increased SNR.

Many other characterization procedures exist to determine the one-way impulse response, but require experimental setups that directly measure the transducer input and output voltage and currents (see e.g. [4–8]). This type of method typically requires some knowledge of the incident or transmitted pressure field, hence involving additional transducers, which need to be carefully aligned for each measurement in an array. Alternatively, they assume a specific type of wave is transmitted to avoid using additional transducers.

Here we present a simple procedure to estimate the one-way impulse response using only one digitized pulse-echo measurement per array element by means of a new de-autoconvolution algorithm, not involving hydrophones. If the transducer under test is self-reciprocal, the transmit and receive impulse responses are equivalent [4, 9]. Consequently, the measurement resulting from pulse-echo characterizations such as in [2, 3] is the autoconvolution of the one-way impulse response. Hence we propose to retrieve the one-way impulse response from such a measurement in conjunction with a newly developed de-autoconvolution algorithm, although other algorithms are available as well [10–13]. More specifically, we measure the pulse-echo response of a flat plate reflector, which is especially time-efficient for large linear arrays. The flat reflector only needs to be aligned once after which all the measurements can be obtained for all array elements

without any adjustments or alignments between measurements.

In Section II, we will first describe the measurement setup and the de-autoconvolution algorithm. In Section III we show the experimental results of the proposed technique, after which we provide a conclusion and discussion in Sections IV and V.

5.2. METHODS

5.2.1. CHARACTERIZATION PROCEDURE

Here we propose a simple calibration procedure to obtain autoconvolved measurements of the one-way impulse response. However, we would like to point out that the de-autoconvolution algorithm described in the next section should work with any measurement procedure that obtains an autoconvolved impulse response.

We suggest to place a flat plate reflector in front of the ultrasound transducer, parallel to the transducer surface, and measure the pulse-echo response after exciting the transducer by a delta-pulse. Since flat reflectors act as an acoustic mirror, the reflection does not change the temporal ultrasound waveforms before reflecting. Assuming the experiment takes place in a homogeneous medium, and assuming delta-pulse excitation, the sampled pulse-echo signal $x[n]$ can be modelled as: $x[n] = h_t[n] * g_t[n] * g_r[n] * h_r[n]$, where $h_t[n]$ and $h_r[n]$ are the electro-mechanical transmit and receive impulse responses of the transducer, respectively, and '*' denotes temporal convolution. The signals $g_t[n]$ and $g_r[n]$ represent the Green's functions for the wavefield propagation to and from the flat reflector (see e.g. [14–16] for convolutive ultrasound models). In the described setup, it is straightforward to see that $g_t[n] = g_r[n]$. Assuming self-reciprocal transducers, $h_t[n] = h_r[n]$, and consequently $x[n]$ can be expressed as $x[n] = h[n] * g[n] * g[n] * h[n]$. This measurement can be de-autoconvolved to obtain the forward field $h[n] * g[n]$. If the flat reflector is placed appropriately, $g[n] = \delta[n - \tau]$, where τ depends on the distance to the flat reflector, such that $x[n] = h[n - \tau] * h[n - \tau]$. The time-shifted one-way impulse response $h[n - \tau]$ can then be obtained using de-autoconvolution.

For this setup it is important to align the flat reflector parallel to the transducer surface and place it at an appropriate distance to avoid sensor diffraction effects, such that $g[n] = \delta[n - \tau]$. If one is interested in the pulse-echo response of every sensor in a linear array, the flat reflector can simply be properly aligned with the array surface, after which the pulse-echo signal of each array sensor is measured one by one. This way, proper reflector alignment only needs to be done once for a large collection of sensors.

5.2.2. DE-AUTOCONVOLUTION ALGORITHM

Suppose we have measured the pulse-echo signal $x[n] = h[n] * h[n]$, and that the time-of-arrival delay τ is included in $h[n]$ from now on. Since linear convolution in the time-domain equals multiplication in the frequency domain, the magnitude spectrum of $h[n]$ can be obtained by taking any square root of each frequency of the Fourier transform of $x[n]$. Unfortunately, this results in a large set of solutions since all combinations of both the positive and negative square root are valid.

If the Discrete Fourier Transform (DFT) is used instead of a Discrete Time Fourier Transform (DTFT), multiplication in the DFT domain equals *circular* convolution in the

time domain. Fortunately, linear autoconvolution can be expressed as a circular auto-convolution by zero-padding $h[n]$ until it has length $2N-1$ samples, where N is the original length of $h[n]$, followed by a circular autoconvolution. Denote the set of solutions formed by taking any combination of positive and negative roots of the DFT of $x[n]$ by C. Any of these solutions solve the circular autoconvolution problem, whereas we are only interested in the linear version of the problem. It is easy to see that any solution to the linear problem is located in the set C as well. Moreover, they can be identified since such solutions must be a zero-padded signal. That is, a solution $\hat{h}[n]$ such that $\hat{h}[n] = 0$ for $n = N+1, N+2, ..., 2N-1$ is a solution to the linear autoconvolution problem. Hence, to find a solution to the linear problem we look for solutions with this structure in the set C, which is equivalent to finding a solution for which the energy of the 'tail' of the estimated signal is zero. Next, we show how finding such a solution can be cast as a semi-definite programming problem.

Let the matrix $\bar{\mathbf{F}} \in \mathbb{C}^{(N-1)\times(2N-1)}$ denote the lower half of the inverse DFT matrix, corresponding to the tail of the signal in the time domain if right-multiplied with a DFT signal. Represent $x[n]$ by the vector $\mathbf{x} \in \mathbb{R}^{2N-1}$, and define the vector $\tilde{\mathbf{x}}$ as the point-wise positive square root of the DFT of $\mathbf{x}$. Finally, let $\mathbf{p}$ be a $(2N-1)$-dimensional vector with components equal to ± 1, corresponding to selecting either the positive or negative root in $\tilde{\mathbf{x}}$. To minimize the tail energy, we solve the following problem:

$$\min_{\mathbf{p}} \quad \mathbf{p}^{\mathsf{H}}\mathrm{diag}(\tilde{\mathbf{x}})^{\mathsf{H}}\bar{\mathbf{F}}^{\mathsf{H}}\bar{\mathbf{F}}\mathrm{diag}(\tilde{\mathbf{x}})\mathbf{p} \quad \text{s.t.} \quad p_i \in \{-1, 1\}, \tag{5.1}$$

where $^{\mathsf{H}}$ represents the conjugate transpose, and $\mathrm{diag}(\tilde{\mathbf{x}})$ represents the square matrix where the diagonal is equal to $\tilde{\mathbf{x}}$. Once $\mathbf{p}$ is obtained, it can be pointwise multiplied with $\tilde{\mathbf{x}}$ to obtain the estimate for $h[n]$ in the DFT domain. Problem (5.1) is a binary quadratic optimization problem which is known to be NP-complete. One way to still obtain a solution that (hopefully) comes close to the optimal solution is by means of semidefinite relaxation. After defining $\mathbf{P} = \mathbf{p}\mathbf{p}^{\mathsf{H}}$ and $\mathbf{Q} = \mathrm{diag}(\tilde{\mathbf{x}})^{\mathsf{H}}\bar{\mathbf{F}}^{\mathsf{H}}\bar{\mathbf{F}}\mathrm{diag}(\tilde{\mathbf{x}})$, we have that $\mathbf{p}^{\mathsf{H}}\mathbf{Q}\mathbf{p} = \mathrm{Tr}(\mathbf{p}^{\mathsf{H}}\mathbf{Q}\mathbf{p}) = \mathrm{Tr}(\mathbf{Q}\mathbf{p}\mathbf{p}^{\mathsf{H}}) = \mathrm{Tr}(\mathbf{Q}\mathbf{P})$. Taking into account that $\mathbf{p}$ in (5.1) is real, problem (5.1) can then be equivalently written as

$$\min_{\mathbf{P}\in\mathbb{R}^{(2N-1)\times(2N-1)}} \quad \mathrm{Tr}(\mathbf{Q}\mathbf{P}) \tag{5.2}$$

$$\text{s.t.} \quad [\mathbf{P}]_{ii} = 1,$$
$$\mathbf{P} \succeq 0,$$
$$\mathrm{rank}(\mathbf{P}) = 1$$

If this problem is relaxed by removing the rank constraint, the resulting convex problem can be solved by the following semi-definite program:

$$\min_{\mathbf{P}\in\mathbb{R}^{(2N-1)\times(2N-1)}} \quad \mathrm{Tr}(\mathbf{Q}\mathbf{P}) \tag{5.3}$$

$$\text{s.t.} \quad [\mathbf{P}]_{ii} = 1,$$
$$\mathbf{P} \succeq 0$$

To obtain a solution that still comes close to a rank one solution, we use random hyper-plane rounding as in [17].
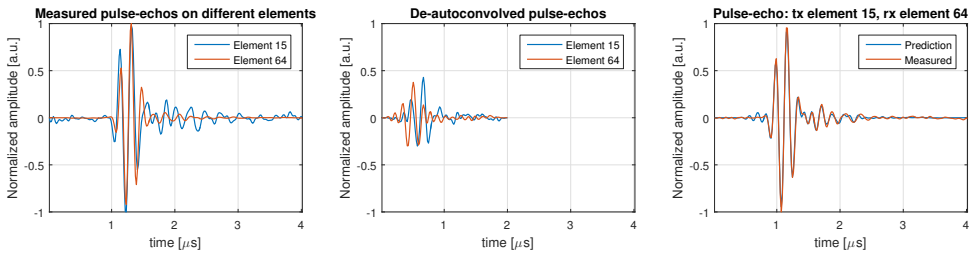
Figure 5.1: Example to illustrate estimation and validation of the one-way impulse responses, including propagation effects, for two different elements, based on pulse-echo measurements of a thin wire in water. Left: the measured *pulse-echo* responses for two different elements. Each signal is obtained by transmitting and receiving on the same element. The signals are significantly different. Centre: the estimated *one-way* impulse responses obtained using de-autoconvolution. Right: comparison of the estimated and measured pulse-echo when transmitting on element 15 and receiving on element 64. The predicted echo is obtained by convolving the estimated one-way impulse responses. Note: we aligned pulses with different time-of-arrivals so they are more easily compared, and only plot from the moment echoes arrive (time-scales are not absolute).

5.3. RESULTS

5.3.1. NUMERICAL SIMULATIONS

We first demonstrate the effectiveness of the de-autoconvolution on arbitrary signals. We generated 1000 random zero-mean unit-variance white Gaussian signals $h[n]$, and autoconvolved them to obtain $x[n]$. In all cases we were able to exactly reconstruct the original signal $h[n]$. Next, we simulated noisy measurements with additive zero-mean white Gaussian noise $v[n]$, $x[n] = h[n] * h[n] + v[n]$, and using an SNR of 20 dB. In this case, the average correlation between the estimated and true impulse response is 0.85 on average, with standard deviation 0.23. Of these simulations, 77% of the estimates has a correlation higher than 0.9. Exact conditions for recovery and noise robustness for this particular algorithm are a topic of future research.

5.3.2. EXPERIMENT: FLAT REFLECTOR IN WATER

We placed a flat perspex reflector parallel to a 128 element linear array (Philips/ATL L7-4) at a distance of 25 mm, coupled to an open ultrasound system (Vantage 256 channels, Verasonics Inc. Kirkland WA USA). To avoid diffraction effects, the reflector was placed as parallel as possible to the array. The elements were excited one by one by a delta-like pulse to obtain the autoconvolved impulse-responses for each element, and were then de-autoconvolved with the proposed algorithm to obtain $\hat{h}[n]$. To validate the results we subsequently measured pulse-echo responses from a perspex flat reflector using different transmit and receive elements. We then compared those measurements to the convolution of the estimated one-way impulse responses of those elements. The comparison is quantified by computing the normalized correlation between the measurement and the estimate. Figure 5.2 shows the average normalized correlation value between all elements in the array with the same inter-distance between transmit and receive el-
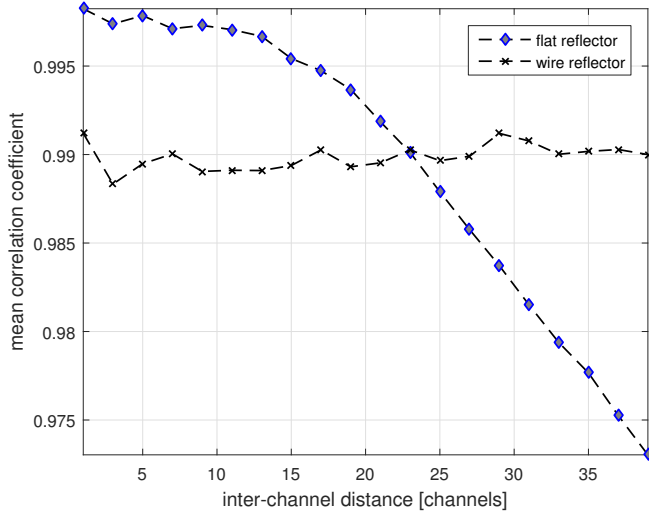
Figure 5.2: Mean correlation between measured pulse-echo impulse response and the convolution of the estimated one-way impulse responses of those elements. The horizontal axis indicates the distance between transmit and receive elements.

ements. For larger inter-channel distance, the angle of arrival and departure becomes larger, and the role of diffractive effects become non-negligible. The Green's function $g[n]$ can no longer be considered to be a shifted delta-pulse, and we are comparing the estimate $\hat{h}_t[n] * \hat{h}_r[n]$ to the measured signal $h_t[n] * g_t[n] * g_r[n] * h_r[n]$. This is apparent from the downward trend in Fig. 5.2.

5.3.3. EXPERIMENT: THIN WIRE IN WATER

Instead of estimating the impulse response $h[n]$, we now consider estimating $h[n] * g[n]$ for a single point in space. To this end, we placed a thin wire in water, and kept its position fixed relative to the centre of the array, at a depth of 25 mm. By transmitting and receiving on the same channel, we basically measure the autoconvolution of $h[n] * g[n]$, which describes both the transducer transfer function as well as the propagation effects. Since the wire is in a fixed position, $g[n]$ is different for each element due to a different incident angle. After estimating $h[n] * g[n]$, we can predict the pulse-echo signal when transmitting and receiving on different elements. Figure 5.1 shows some example measurements and estimates. Figure 5.3 shows the radio frequency signals (RF) before and after de-autoconvolution. After de-autoconvolution, the RF resembles an RF where the wire itself is the ultrasound source. We further compare the normalized correlation coefficient similarly to the experiments with the flat reflector. The results are shown in Fig. 5.2. In this case, diffractive effects are included in the term $g[n]$, so that the estimation performance does not decrease as the inter-channel distance increases.
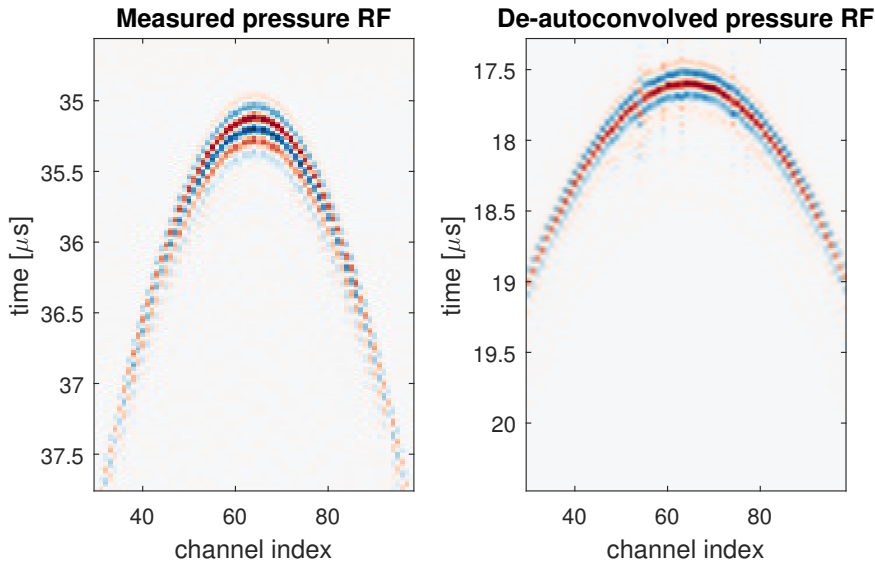
Figure 5.3: Left: measured RF consisting of doing a pulse-echo measurement with the same transmit and receive channels. Right: de-autoconvolved RF of the left panel. The time-of-arrivals are two times shorter, and the pulse-lengths have decreased.

5.4. DISCUSSION

An inherent property of the proposed method is that it estimates the transfer function of the entire system involved in the measurement, including the effect of the cable transfer functions, for example. As a result, simulations are more accurate by including all these effects and system electronics that are complicated to model [18]. An essential condition for successful estimation using our method is the self-reciprocality of the ultrasound transducer. This condition can be violated if different equipment or signal paths are used on transmit or receive, prohibiting successful calibration.

Some non-causal oscillation is visible at the beginning of the predicted signal in Fig. 5.3 in the right panel. This can be caused by an incorrect selection of the sign of the root in the vector **p**, resulting in an incorrect phase for the corresponding frequency component.

Topics for future research are noise robustness, conditions for convergence of the semi-definite relaxation, and the use of additional constraints such as the limited bandwidth of a transducer, which is currently not exploited and could improve noise-robustness. The proposed de-autoconvolution algorithm should be compared to other de-autoconvolution methods in an ultrasound characterization context. The de-autoconvolution technique could potentially be used for estimating transmission signals for time-reversal focusing [19]. Finally, the fact that de-autoconvolution shortens the RF pulse-lengths could be interesting in the context of image reconstruction.

5.5. CONCLUSION

We outlined a de-autoconvolution algorithm and proposed a calibration procedure to use this algorithm for estimating the one-way impulse response of self-reciprocal ultrasound transducers. After careful alignment of a flat reflector, all characterization measurements for the entire array are rapidly obtained. We provide simulations to demonstrate the performance of the algorithm for random Gaussian signals with a white power spectral density. We show experimental results to illustrate its effectiveness with an array where the one-way impulse response is different for each element. The predicted pulse-echo signals based on the estimated one-way impulse responses correlate well with experimental measurements. We discussed the validity of the self-reciprocality condition and provided an outlook on feature research topics.

REFERENCES

[1] P. van der Meulen, P. Kruizinga, J. G. Bosch, and G. Leus, *Impulse response estimation method for ultrasound arrays,* in *2017 IEEE International Ultrasonics Symposium (IUS)* (IEEE, 2017) pp. 1–4.

[2] T. L. Szabo, B. Ü. Karbeyaz, R. O. Cleveland, and E. L. Miller, *Determining the pulse-echo electromechanical characteristic of a transducer using flat plates and point targets,* the Journal of the Acoustical Society of America **116**, 90 (2004).

[3] T. Gehrke, F. Cheikhrouhou, and H. M. Overhoff, *Derivation of the impulse response of ultrasonic transducers by experimental system identification,* in *Ultrasonics Symposium, 2005 IEEE*, Vol. 1 (IEEE, 2005) pp. 349–352.

[4] K. Beissner, *Free-field reciprocity calibration in the transition range between near field and far field,* Acta Acustica united with Acustica **46**, 162 (1980).

[5] W. Shou, S. Duan, P. He, R. Xia, and D. Qiari, *Calibration of a focusing transducer and miniature hydrophone as well as acoustic power measurement based on free-field reciprocity in a spherically focused wave field,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **53**, 564 (2006).

[6] A. L. Lopez-Sanchez and L. W. Schmerr, *Determination of an ultrasonic transducer's sensitivity and impedance in a pulse-echo setup,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **53** (2006).

[7] S. Zhang, C. M. Kube, Y. Song, and X. Li, *A self-reciprocity calibration method for broadband focused transducers,* The Journal of the Acoustical Society of America **140**, EL236 (2016).

[8] K. Brendel and G. Ludwig, *Calibration of ultrasonic standard probe transducers,* Acta Acustica united with Acustica **36**, 203 (1976).

[9] L. L. Foldy and H. Primakoff, *A general theory of passive linear electroacoustic transducers and the electroacoustic reciprocity theorem. i,* The journal of the acoustical society of America **17**, 109 (1945).

[10] R. Gorenflo and B. Hofmann, *On autoconvolution and regularization,* Inverse Problems **10**, 353 (1994).

[11] K. Choi and A. D. Lanterman, *An iterative deautoconvolution algorithm for nonnegative functions,* Inverse problems **21**, 981 (2005).

[12] Z. Dai and P. K. Lamm, *Local regularization for the nonlinear inverse autoconvolution problem,* SIAM Journal on Numerical Analysis **46**, 832 (2008).

[13] S. Bürger and J. Flemming, *Deautoconvolution: A new decomposition approach versus tigra and local regularization,* Journal of Inverse and Ill-posed Problems **23**, 231 (2015).

[14] P. R. Stepanishen, *Pulsed transmit/receive response of ultrasonic piezoelectric transducers,* The Journal of the Acoustical Society of America **69**, 1815 (1981).

[15] J. A. Jensen and N. B. Svendsen, *Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **39**, 262 (1992).

[16] A. Lhémery, *Impulse-response method to predict echo-responses from targets of complex geometry. part i: Theory,* The Journal of the Acoustical Society of America **90**, 2799 (1991).

[17] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,* Journal of the ACM (JACM) **42**, 1115 (1995).

[18] T. Kim, S. Shin, H. Lee, H. Lee, H. Kim, E. Shin, and S. Kim, *Matlab/simulink pulse-echo ultrasound system simulator based on experimentally validated models,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **63**, 290 (2016).

[19] M. Fink, *Time reversal of ultrasonic fields. i. basic principles,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **39**, 555 (1992).

5

6

BLIND CALIBRATION FOR ARRAYS WITH AN ABERRATION LAYER IN ULTRASOUND IMAGING

We consider the scenario of finding the transfer function of an aberrating layer in front of an ultrasound array. We are interested in blindly estimating this transfer function without prior knowledge of the unknown ultrasound sources or ultrasound contrast image. The algorithm gives an exact solution if the matrix representing the aberration layer's transfer function is full rank, up to a scaling and reordering of its columns, which has to be resolved using some prior knowledge of the matrix structure. We provide conditions for the robustness of blind calibration in noise. Numerical simulations show that the method becomes more robust for shorter wavelengths, as the transfer function matrices then tend to be less ill-conditioned. Image reconstruction from simulated data using the k-Wave toolbox show that a well calibrated model removes some of the distortions introduced by an uncalibrated model, and improves the resolution for some of the sources.

6.1. INTRODUCTION

IN this contribution we consider the scenario of an aberrating layer in front of an ultrasound (US) array, as depicted in Fig. 7.1. In medical ultrasound imaging, such layers could represent the skin and fat layers to which the ultrasound probe is pressed to image internal organs. Furthermore, there is typically a small layer to match the impedance of the US probe and human skin, which causes some phase changes across the array. Another example is the human skull, which distorts ultrasound waves too strongly to use the resulting pulse-echo data to reconstruct an image of the brain with current imaging techniques. Recently, we have also investigated the use of so called 'coding masks', thin random phase masks, to reconstruct ultrasound images with a single US element. To successfully image using a coding mask, precise calibration is required but the current method measures the response of each pixel individually, which is time consuming. In all these cases, we are interested in estimating the effect of the aforementioned layers to incorporate it into the image reconstruction process.

A straightforward way of solving this problem would be to place sources or reflectors in known positions of the probe's field of view (FOV), and then adjust our model so as to match the expected response for that particular source to the measured response. Needless to say, this approach becomes problematic when we desire to adjust our ultrasound model for the human skull or the skin, since this would require us to place a multitude of sources inside the patient's skull or body. Moreover, such methods may be prone to positioning errors of such sources as well. Ideally, then, we would use a 'blind' calibration procedure that relies on some other principle to correct layer aberrations.

In this paper, we generalize all problems above using the model as visualized in Fig. 7.1. We define a 'virtual' array, just in front of the aberration layer, and assume there is a linear transfer function from each point on the virtual array to each other point on the real array. The US field in the real array is then a linear combination of the field incident on the virtual array. We furthermore assume that the remainder of the FOV is a homogeneous medium, with only small inhomogeneities in the medium formed by the US contrast image that we want to reconstruct.

The problem of correcting for phase changes due to a phase screen in front of an US array has been studied before, using both blind and non-blind techniques. However, most of the studies focus on thin phase layers which cause a time delay for each point on the array (e.g., [2–4]). These methods only consider a time shift caused per element of the array, but do not consider situations where each sensor in the array may be influenced by multiple points on the virtual array.

Focusing ultrasound waves through the skull has been achieved using time-reversal techniques [5], and spatiotemporal filter design [6]. Unfortunately, these methods typically require one to first measure waves *inside* the brain. An interesting alternative approach is one where two arrays are placed opposite to one another on each side of the skull [7], avoiding the need for measurements inside the skull, but requiring two arrays. Furthermore, in [7] it is assumed that one of the skull walls can be approximated as a thin phase layer, which is not necessarily the case.
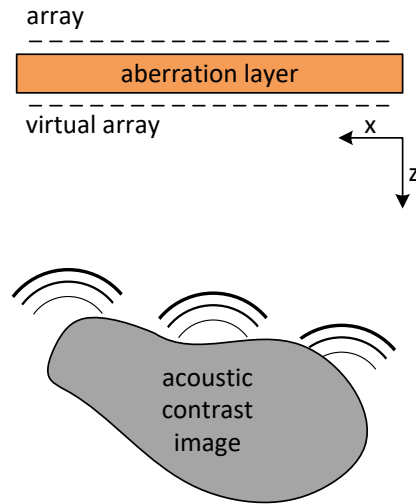
Figure 6.1: An aberration layer distorts incident wavefields, leading to incorrect estimation of the acoustic contrast image. We are interested in estimating the transfer function from the virtual array just before the aberration layer to the real array.

6

NOTATION

We will represent a vector $\mathbf{x}$ in lower-case bold, and a matrix $\mathbf{X}$ in upper-case bold. The operator $\|.\|_F$ represents the Frobenius norm, $\star$ represents the Khatri-Rao product, and $*$ denotes a temporal convolution. The notation $\mathbf{X}^\dagger$ represents the Moore-Penrose pseudo-inverse of $\mathbf{X}$. Finally, diag$\mathbf{c}$ stands for a diagonal matrix with the vector $\mathbf{c}$ on its diagonal.

6.2. SIGNAL MODEL AND PROBLEM FORMULATION

We assume that a linear array is placed in a homogeneous medium with linear propagation of US waves. We can either consider the case of a set of sources transmitting towards the array, or that a separate transmitter illuminates the FOV, where inhomogeneities (acoustic contrast) in the acoustic properties of the propagation medium scatter echoes towards the array. In the latter case, each contrast then acts as an independent source. If the contrasts are sufficiently weak, higher order reflections between reflectors can be ignored according to the Born approximation (see e.g. [8, 9]), which is typically used in ultrasound imaging. In that case, all measurements are linear w.r.t. sources and/or ultrasound contrast images, and we can use the following convolutional formulation to model measurements [10–12]. Denote the number of array elements by M, and the measurement of element m by $y_m(t)$. We divide the FOV in N small pixels, and assume that

each from K sources transmits a signal $s_k(t)$. Our model then becomes:

$$y_m(t) = \sum_{n=1}^{N} x_n \, g_{m,n}^{(r)}(t) * \left(\sum_{k=1}^{K} g_{n,k}^{(t)}(t) * s_k(t) \right), \tag{6.1}$$

where $g_{m,n}^{(r)}(t)$ is the Green's function for a wave travelling from pixel n towards array element m, and x_n is the scatter intensity coefficient of pixel n. The signal $g_{n,k}^{(t)}(t)$ is the Green's function from source k to pixel n. Here, the superscript $^{(r)}$ refers to the receive Green's function (from a pixel to an array element), and the superscript $^{(t)}$ refers to the transmit Green's function.

Typically, the transfer functions and excitation signals are known beforehand, and the goal then is to estimate x_n from y_m. Since the y_m's are completely linear w.r.t. the x_n's, we can stack the measurements of all sensors for a single temporal frequency bin l into a single vector $\mathbf{y}_l \in \mathbb{C}^M$, and write:

$$\mathbf{y}_l = \mathbf{A}_l \mathbf{x}, \tag{6.2}$$

where $\mathbf{x} \in \mathbb{R}^N$ contains the coefficients x_n, and $\mathbf{A}_l \in \mathbb{C}^{M \times N}$ can be constructed using (6.1). This equation is typically used for solving the inverse imaging problem, i.e., finding the contrast image $\mathbf{x}$ from measurements $\mathbf{y}_l$, and is the equation we will try to solve for obtaining an image in the results section.

For the results in this paper we will mostly focus on calibration from many different transmit signals $s_k(t)$, using the equation below. From (6.1), we can express measurements as a linear function of the sources:

$$\begin{aligned} \mathbf{y}_l &= \mathbf{G}_l^{(r)} \mathrm{diag}\mathbf{x}\mathbf{G}_l^{(t)} \mathbf{s}_l \\ &= \mathbf{G}_l \mathbf{s}_l, \end{aligned} \tag{6.3}$$

where $\mathbf{G}_l \triangleq \mathbf{G}_l^{(r)} \mathrm{diag}\mathbf{x}\mathbf{G}_l^{(t)}$ has dimensions $M \times K$, and $\mathbf{s}_l \in \mathbb{C}^K$. If, instead of contrast sources scattering towards the array, there are K sources directly transmitting towards the array themselves, we set $\mathbf{G}_l^{(t)} = \mathbf{I}$, and $\mathbf{x}$ will be an all-zero vector with ones for the pixels where a transmitting source is active. *The latter will be the case for the simulations in this paper.* For the remainder of this paper, we drop the frequency subscript l, and we assume calibration is done on each frequency bin separately.

In the case of the presence of an aberration layer, there is an additional transfer function in the frequency domain from the ultrasound field just before the aberration layer to the ultrasound field in the positions of the array. That is, we introduce a virtual array just before the aberration layer, and assume there is a linear transfer function from each virtual array element to each element on the real array (Fig. 7.1). Hence, we can express the measurements $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{HGs}, \tag{6.4}$$

where $\mathbf{H} \in \mathbb{C}^{M \times M}$ describes the linear mapping from US fields in the virtual array to the real array, for a single temporal frequency. Different from the original inverse imaging problem, the problem addressed in this paper is that of estimating $\mathbf{H}$ using equation (6.4) without knowledge of either $\mathbf{s}$ or $\mathbf{x}$.

6.3. CALIBRATION ALGORITHM

We start by trying to estimate $\mathbf{H}$ based on (6.4). Assume that P transmit events are used, using a K-element transmit array. We can stack the transmitted signals $\{\mathbf{s}(p) \in \mathbb{C}^K\}$, $p = 1,\ldots,P$ in a transmit matrix $\mathbf{S} = [\mathbf{s}(1)\ \mathbf{s}(2)\ \ldots\ \mathbf{s}(P)]$, and denote our first set of measurements as

$$\mathbf{Y}_1 = \mathbf{HGS}, \tag{6.5}$$

where $\mathbf{Y}_1 \in \mathbb{C}^{M \times P}$.

Next, we assume that a second set of measurements $\mathbf{Y}_2$ is obtained in a similar way, using the same image $\mathbf{x}$, and the same transmit sequences $\mathbf{S}$, but with a thin phase layer just in front of the virtual array. The thin phase layer causes a phase shift in the frequency domain on each point of the virtual array, so its effect can be modeled as:

$$\mathbf{Y}_2 = \mathbf{H\Theta GS}, \tag{6.6}$$

where $\Theta = \mathrm{diag}\,\boldsymbol{\theta}$, with $\boldsymbol{\theta} \in \mathbb{C}^M$ being the phase shifts of the thin phase layer. Such layers can be implemented using 'kinoforms' [13, 14], i.e., thin layers with a different speed of sound which cause a different time delay for each point on its surface by spatially varying the layer's thickness.

The basic idea of the calibration algorithm is to combine $\mathbf{Y}_1$ and $\mathbf{Y}_2$ as follows. First, $\mathbf{Y}_2$ is projected onto $\mathbf{Y}_1$:

$$\begin{aligned}
\mathbf{Z} &= \mathbf{Y}_2 \mathbf{Y}_1^\dagger \\
&= \mathbf{H\Theta GS(GS)}^\dagger \mathbf{H}^{-1} \\
&= \mathbf{H\Theta H}^{-1}.
\end{aligned} \tag{6.7}$$

for which we require that $\mathbf{H}$ is invertible, and that $\mathbf{GS}$ is full row-rank, so that $\mathbf{GS(GS)}^\dagger = \mathbf{I}$.

This way, we obtain the matrix $\mathbf{Z}$ *which no longer depends on $\mathbf{x}$ or $\mathbf{s}$*. Moreover, the matrix $\mathbf{Z}$ has an eigendecomposition with eigenvectors $\mathbf{H}$ and eigenvalues $\boldsymbol{\theta}$. The aberration matrix can now be found using an eigendecomposition, without any prior knowledge of $\boldsymbol{\theta}$, provided that the eigendecomposition is unique. This can be ensured through the design of the thin phase layer, by making each delay in $\boldsymbol{\theta}$ unique. Through these steps, we have obtained the matrix $\mathbf{H}$ without *any prior knowledge* of the transmit codes $\mathbf{s}$ or the matrix $\mathbf{G}$. Note that this technique would also have worked with many contrast images, using $\mathbf{y} = \mathbf{HAx}$ from (6.2). In that case, $\mathbf{GS}$ in (6.7) would be replaced by $\mathbf{AX}$ and we would obtain $\mathbf{Z}$ as long as $\mathbf{AX}$ is full row-rank.

For the algorithm to be robust to noise, it is required that the matrix $\mathbf{Y}_1$ is well-conditioned, which in turn depends on $\mathbf{H}$, $\mathbf{G}$ and $\mathbf{S}$ to be well-conditioned (and full-rank). If $\mathbf{Y}_1$ is ill-conditioned, even small amounts of noise in $\mathbf{Y}_2$ can be amplified by the small singular values of $\mathbf{Y}_1$, resulting in a bad estimate of $\mathbf{Z}$.

This technique exploits a similar data structure as the one used in the well known ESPRIT algorithm [15]. Both ESPRIT and our method assume that two measurements are available, which are related by a diagonal matrix. In contrast to the ESPRIT algorithm, however, we are not interested in estimating the values on the diagonal of Θ, but

6

in the eigenvector matrix $\mathbf{H}$. From a different point of view, (6.7) is the solution for $\mathbf{Z}$ for solving the equation $\mathbf{Z}\mathbf{Y}_1 = \mathbf{Y}_2$. Thus, by computing the eigendecomposition of $\mathbf{Z}$, we are performing a dynamic mode decomposition [16] on the data $\mathbf{Y}_1$ and $\mathbf{Y}_2$ to obtain $\mathbf{H}$ and $\boldsymbol{\theta}$.

6.3.1. UNIQUENESS

Since eigendecompositions are unique up to a reordering and scaling of the columns of $\mathbf{H}$, the order and scaling of the columns of the eigenvectors obtained from (6.7) are also ambiguous. To find the correct ordering, we assume that $\boldsymbol{\theta}$ is ordered in a known manner. For example, if $\theta_m = \exp(i2\pi m/M)$, $m = 1, \ldots, M$, we may have the prior knowledge that $\{\theta_m\}$ are ordered by a linearly increasing phase component.

The scaling of each eigenvector of $\mathbf{Z}$ can be resolved according to various principles, but is very dependent on the aberration layer being studied. For example, a class of similar layers could be studied by simulations to find a parameterized common basis for the corresponding class of matrices. In the results section of this contribution, we focus on random aberration layers. We empirically found that the transfer function matrices of such layers have a strong Toeplitz structure per frequency (see Fig. 6.2 for a typical matrix of this class). Based on this prior knowledge, we will resolve the scaling issue by first defining a basis for the set of matrices a constant diagonal and constant subdiagonals, $\mathbf{B}_l \in \mathbb{C}^{M \times M}$, $l = 1, \ldots, 2M - 1$. We then try to minimize the Frobenius norm between this set of matrices and the (column-wise) scaled eigenvector matrix $\mathbf{H}_0$ of $\mathbf{Z}$, where $\mathbf{H}_0$ already has its columns re-ordered:

$$\{\hat{\mathbf{b}}, \hat{\mathbf{c}}\} = \arg\min_{\mathbf{b}, \mathbf{c} \neq 0} \left\| \sum_{q=1}^{2M-1} b_q \mathbf{B}_q - \mathbf{H}_0 \mathrm{diag}\mathbf{c} \right\|_F^2, \tag{6.8}$$

where $\mathbf{b} \in \mathbb{C}^{2M-1}$ contains the coefficients b_q, and $\mathbf{c} \in \mathbb{C}^M$ contains the scaling coefficients for each column of $\mathbf{H}_0$.

The solution of this problem, up to a complex scaling, is given by the smallest singular vector of the composite matrix $[\mathbf{B} \ (\mathbf{I} \star \mathbf{H})]$, with $\mathbf{B} = [\mathrm{vec}(\mathbf{B}_1), \ldots \mathrm{vec}(\mathbf{B}_{2M-1})]$. The resulting matrix $\hat{\mathbf{H}} = \mathbf{H}_0 \mathrm{diag}\hat{\mathbf{c}}$ gives us an estimate of $\mathbf{H}$ up to a complex scaling of the entire matrix. Consequently, there is one final ambiguity, which is especially important when one wants to estimate an ultrasound image using all temporal frequencies jointly. The study of this problem is beyond the scope of this contribution, but solutions are available in literature. See e.g. [17] for tackling this problem structure, or [18–20] for solving this problem from covariance data.

6.4. SIMULATION RESULTS

To avoid the robustness issues described in the previous section, the calibration experiments have to be arranged and designed in such a way that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are well-conditioned and full rank matrices. To that end, we will assume the following experimental setup.

First, for simplicity, we will assume that a set of point sources are excited by some signals $\mathbf{S}$, which the sources transmit to the array. We will further place these sources in the far field, evenly distributed across directions of arrival from 0 to 180 degrees w.r.t. the
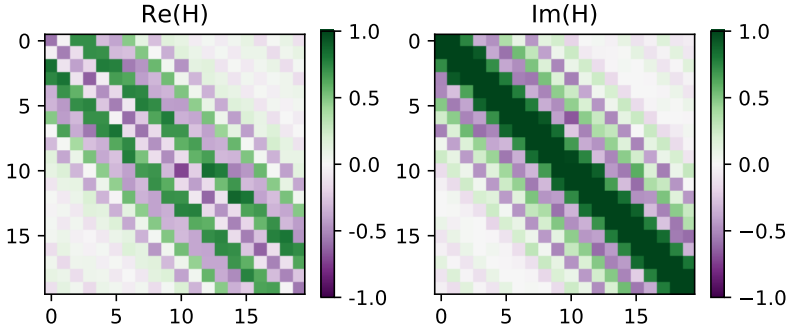
Figure 6.2: An example of a typical **H** matrix of the random aberration layer in Fig. 6.3 for a temporal frequency of 5 MHz.
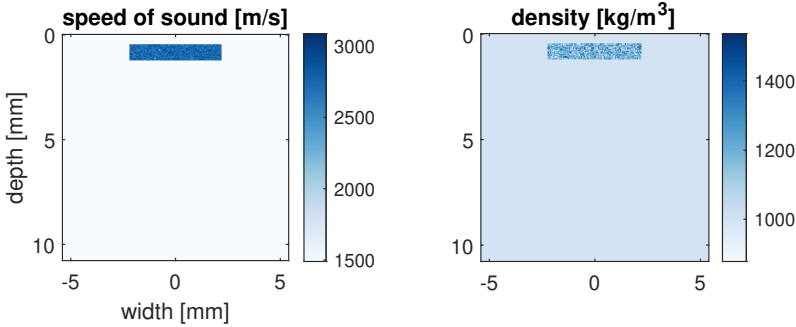
Figure 6.3: Speed of sound and density distribution of the aberration layer used for the simulations in this contribution.

array, causing the incident waves to be nearly plane when arriving at the aberration layer, making **G** well-conditioned if all directions-of-arrival (DOAs) are covered. Note that **G** can also be well conditioned by taking a sufficiently large FOV in the imaging domain. We adopt the far-field in our simulation for simplicity for now. Second, we will use a random aberration layer as visualized in Fig. 6.3, which we empirically found to preserve information (i.e., **H** is a well-conditioned full rank matrix) relatively well. Finally, we use a white Gaussian matrix **S**, so that **S** is full rank with high probability, and fully samples the column space of both **HG** and **HΘG**. As a thin phase layer we choose one that causes a different phase shift for each point on the virtual array:

$$\theta_m = \exp(2\pi i (m - M/2)/M), \; m = 1, \dots, M. \tag{6.9}$$

We obtain the true transfer function of the random layer by sampling **H** column-wise using the k-Wave ultrasound simulation toolbox [21]. The background medium in the simulations has the acoustic properties of water. We then generate $\mathbf{Y}_1$ and $\mathbf{Y}_2$ according to (6.5), (6.6), and (6.9), having obtained the **H** matrix as just described. In k-Wave, the medium is defined on a grid with a spacing of 50 μm, and we used an array of point

sensors with a spacing of 200 μm, with $M = 20$. Signals are sampled in the time-domain using a sampling frequency of 206 MHz. For actual imaging simulations, in the next subsection, we focus on the frequency range of approximately 3-8 MHz.

6.4.1. GENERAL PERFORMANCE

First, we assess the performance of the algorithm for a range of noise levels. We define SNR with respect to $\mathbf{Y}_1$:

$$\text{SNR} = \frac{\|\mathbf{Y}_1\|_F^2}{\sigma_n^2 KM}, \tag{6.10}$$

where σ_n^2 is the noise variance, which we assume to be spatially and temporally white Gaussian i.i.d. noise. For the entries of $\mathbf{S}$ we use a realization of an i.i.d. Gaussian distribution, with $P = 1,000$. To find the correct order of the columns of $\mathbf{H}$, we sort the eigenvalues of $\mathbf{Z}$ according to their phase, and apply the same permutation to the columns of $\mathbf{H}$. To resolve the per-column scaling ambiguity, we solve (6.8) as described earlier. For each frequency and SNR, we compute a noisy $\mathbf{Y}_1$ and $\mathbf{Y}_2$ 100 times, and each time we compute the normalized correlation coefficient between $\mathbf{H}$ and $\hat{\mathbf{H}}$, defined as $\rho = \text{vec}(\mathbf{H})^{\mathsf{H}}\text{vec}(\hat{\mathbf{H}})/(\|\mathbf{H}\|_F\|\hat{\mathbf{H}}\|_F)$. As a final measure of performance, we show the mean absolute correlation over all 100 simulations.

The resulting performance is visualized in Fig. 6.4 for a range of SNRs and various frequencies. Performance increases with an increasing SNR, but also with increasing frequency. For higher frequencies, the size of inhomogeneities becomes closer to the wavelength, causing more scattering within the aberration layer, hence causing a more diverse transfer function matrix $\mathbf{H}$. For lower frequencies, the opposite happens: inhomogeneities are relatively small compared to the wavelength, and the layer mostly acts as a homogeneous layer. Such layers typically have a low-pass behaviour, i.e., the wavefield is smoothed out during propagation through the layer, reducing higher frequency content in the original wavefield. Consequently, the corresponding $\mathbf{H}$ matrix tends to be rank deficient or ill-conditioned, making it more difficult to estimate it from noisy measurements using the proposed method.

6.4.2. IMAGING EXPERIMENT

Next, we demonstrate the difference in terms of image reconstruction of a model calibrated using the proposed method, and an uncalibrated model. The parameter and simulation settings remain unchanged. For calibration, we use $P = 1,000$, using an SNR of 30 dB, but for the image reconstruction itself we only use one transmit event from the sources to the array. Furthermore, no noise is added to the array measurements used for image reconstruction since errors in the image are typically dominated by side and grating lobes. Note that in the calibration experiment we used far-field sources as described above, whereas in this simulation we consider sources positioned between the near and far field. All the sources inject a Gaussian pulse centered at 5 MHz, with a -6 dB bandwidth of approximately 4.5 MHz into the medium.

To obtain an image, we use the LSQR algorithm [22] to solve (6.2), using the result after 6 iterations to obtain a regularized estimate. For the uncalibrated model, we use $\mathbf{H} = \mathbf{I}$ to solve (6.2). The resulting image reconstructions are shown in Fig. 6.5 and 6.6. As
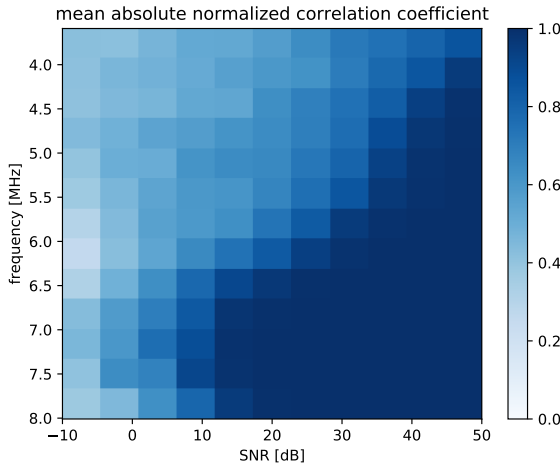
Figure 6.4: Absolute correlation between the estimated transfer function matrix $\hat{\mathbf{H}}$ and the true $\mathbf{H}$ for various temporal frequencies and SNR values.

can be seen, the layer mostly distorts by spatially shifting and stretching the true image, which is resolved by using the calibrated model. Moreover, without calibration some of the sources are not visible, such as the fourth source in the axial slice of the image in Fig. 6.6, and the rightmost source in the lateral slice.

6.5. CONCLUSION

In this contribution we considered the problem of estimating the transfer function matrix of an aberrating layer in front of an (ultrasound) array. The proposed method does not require knowledge of the ultrasound image/sources, but is dependent on the proper conditioning of the blind calibration problem. Consequently, blind calibration becomes more robust to additive noise for higher temporal frequencies, where the transfer function matrices become more well-conditioned. In noiseless scenarios, the algorithm gives an exact solution of the transfer function matrix, up to a scaling and permutation of its columns. This ambiguity is resolved in this paper by assuming that the transfer function matrix has a strong Toeplitz structure, which we empirically found to be the case for random aberration layers. Finally, we demonstrated that using a calibrated model for imaging reduces the geometric distortion introduced by a random aberration layer, and additionally is able to reveal sources that are hard to resolve for the uncalibrated model.

REFERENCES

[1] P. van der Meulen, M. Coutino, P. Kruizinga, J. G. Bosch, and G. Leus, *Blind calibration for arrays with an aberration layer in ultrasound imaging*, in *2020 28th European Signal Processing Conference (EUSIPCO)* (IEEE, 2021) pp. 1269–1273.
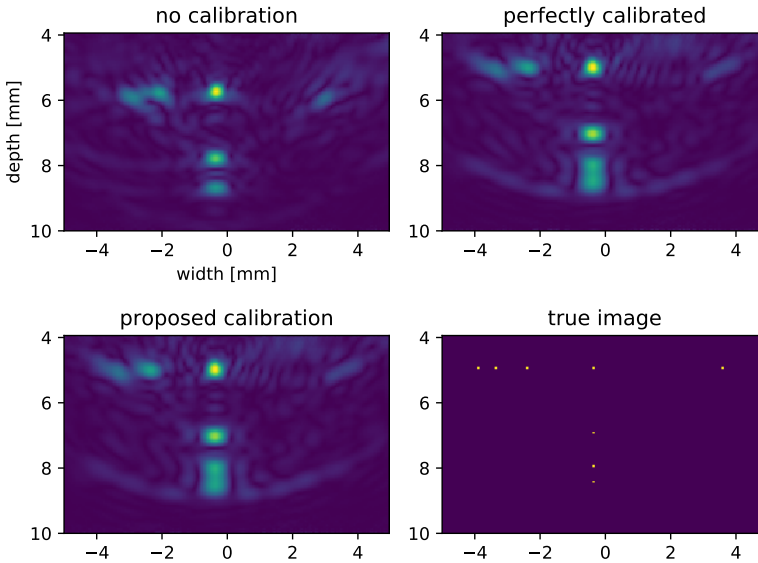
Figure 6.5: Image reconstructions of a resolution phantom in k-Wave. We compare reconstruction with no calibration, a perfectly calibrated model, and a calibrated model using the proposed method.

6

[2] L. Nock, G. E. Trahey, and S. W. Smith, *Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,* The Journal of the Acoustical Society of America **85**, 1819 (1989).

[3] M. Karaman, A. Atalar, H. Koymen, and M. O'Donnell, *A phase aberration correction method for ultrasound imaging,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **40**, 275 (1993).

[4] M. O'donnell and S. Flax, *Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **35**, 768 (1988).

[5] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu, *Time-reversed acoustics,* Reports on progress in Physics **63**, 1933 (2000).

[6] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, *Optimal focusing by spatio-temporal inverse filter. ii. experiments. application to focusing through absorbing and reverberating media,* The Journal of the Acoustical Society of America **110**, 48 (2001).

[7] F. Vignon, J. Aubry, M. Tanter, A. Margoum, and M. Fink, *Adaptive focusing for transcranial ultrasound imaging using dual arrays,* The Journal of the Acoustical Society of America **120**, 2737 (2006).
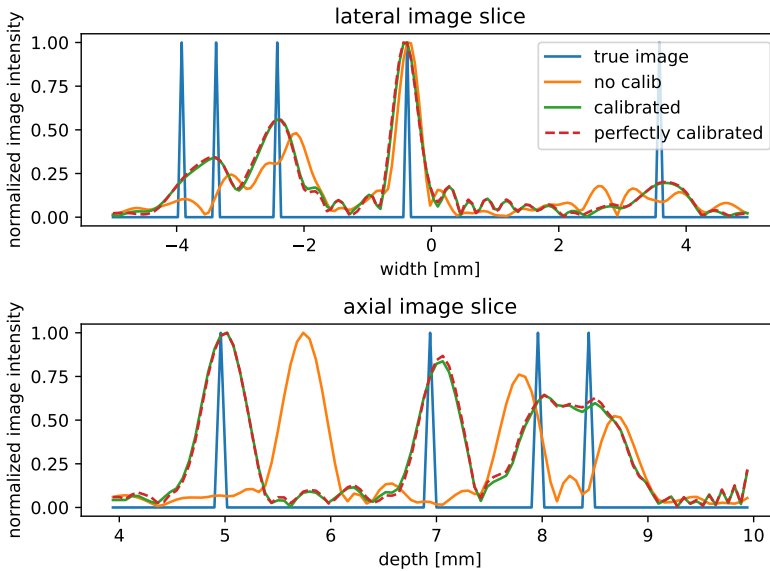
Figure 6.6: Lateral and axial profiles of the reconstructions of Fig. 6.5.

[8] J. T. Fokkema and P. M. van den Berg, *Seismic applications of acoustic reciprocity* (Elsevier, 1993).

[9] M. Verweij, B. Treeby, K. Van Dongen, and L. Demi, *Simulation of ultrasound fields,* Comprehensive biomedical physics , 465 (2014).

[10] P. R. Stepanishen, *Pulsed transmit/receive response of ultrasonic piezoelectric transducers,* The Journal of the Acoustical Society of America **69**, 1815 (1981).

[11] J. A. Jensen, *A model for the propagation and scattering of ultrasound in tissue,* The Journal of the Acoustical Society of America **89**, 182 (1991).

[12] J. A. Jensen and N. B. Svendsen, *Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **39**, 262 (1992).

[13] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, *Holograms for acoustics,* Nature **537**, 518 (2016).

[14] M. Brown, D. Nikitichev, B. Treeby, and B. Cox, *Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles,* Applied Physics Letters **110**, 094102 (2017).

[15] R. Roy and T. Kailath, *Esprit-estimation of signal parameters via rotational invariance techniques,* IEEE Transactions on acoustics, speech, and signal 995 (1989).

[16] P. J. Schmid, *Dynamic mode decomposition of numerical and experimental data,* Journal of fluid mechanics **656**, 5 (2010).

[17] S. Ling and T. Strohmer, *Self-calibration and bilinear inverse problems via linear least squares,* SIAM Journal on Imaging Sciences **11**, 252 (2018).

[18] S. J. Wijnholds and A.-J. Van Der Veen, *Multisource self-calibration for sensor arrays,* IEEE Transactions on Signal Processing **57**, 3512 (2009).

[19] P. van der Meulen, P. Kruizinga, J. G. Bosch, and G. Leus, *Calibration techniques for single-sensor ultrasound imaging with a coding mask,* in *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (IEEE, 2018) pp. 1641–1645.

[20] K. N. Ramamohan, S. P. Chepuri, D. F. Comesana, and G. Leus, *Blind calibration of sparse arrays for doa estimation with analog and one-bit measurements,* in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2019) pp. 4185–4189.

[21] B. E. Treeby and B. T. Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields,* Journal of biomedical optics **15**, 021314 (2010).

[22] C. C. Paige and M. A. Saunders, *Lsqr: An algorithm for sparse linear equations and sparse least squares,* ACM Transactions on Mathematical Software (TOMS) **8**, 43 (1982).

6

7

ULTRASONIC IMAGING THROUGH ABERRATING LAYERS USING COVARIANCE MATCHING

We consider the scenario of finding the transfer function of an aberrating layer in front of a receiving ultrasound (US) array. We propose a method for blindly estimating this transfer function without exact knowledge of the ultrasound sources or acoustic contrast image, and without directly measuring the transfer function using a separate controlled calibration experiment. Instead, the measurement data of many unknown random images is collected, such as from blood flow, and its second-order statistics are exploited. A measurement model is formulated that explicitly defines the layer's transfer function. A covariance domain problem is then defined to eliminate the image variable, and it is solved for the layer's transfer function using manifold-based optimization. The proposed approach and calibration algorithm are evaluated on a range of challenging and realistic simulations using the k-Wave *toolbox. Our results show that, given a sufficiently efficient parameterization of the layer's transfer function, and by jointly estimating the transfer function at multiple frequencies, the proposed algorithm is able to obtain a successful estimate. Subsequent simulated imaging experiments using the obtained transfer function also show an increased imaging performance in various aberrating layers, including a skull layer.*

7.1. Introduction

Ultrasound imaging is one of the most widely used medical imaging techniques due to its affordability, portability, its capability of imaging at very high frame rates (more than 1,000 Hz), and its real-time feedback for a clinician. Additionally, it does not utilize ionizing radiation or strong magnetic fields, and as such does not create potential adverse health effects for patients. However, it suffers from an inherent trade-off between resolution and penetration depth. Moreover, imaging through certain layers in the human body, such as human skull bone tissue or fat and muscle layers, cause aberrations in the ultrasonic pulse-echo wavefields, resulting in imaging artifacts and loss of resolution [2, 3]. These wavefield aberrations are a result of both non-linear distortions (due to absorption) as well as linear distortions (due to echo time-of-arrival shifts, scattering within the layer, etc.). In this paper we intend to address the linear distortions by estimating the Green's functions through an aberrating layer in front of an ultrasound array.

Most imaging techniques rely on the Born approximation [4, 5], which approximates a pulse-echo scenario by modeling transmitted waves as travelling throughout the region of interest (ROI), such as human tissue, uninterruptedly (as if there was a homogeneous background medium). Small inhomogeneities in human tissue cause small reflections of this forward field, which are also assumed to propagate to a receiving sensor without being altered. Hence, multiple scattering is ignored, and the model consists of an independent wave model for the propagation path from a transmitter to a scatterer, and from a scatterer to each receiver. More importantly, the unknown ultrasound image becomes linear w.r.t. the ultrasonic measurements, simplifying the imaging problem greatly. One of the conditions underlying the Born approximation is that scatterers do not scatter very strongly, so that the transmit field is indeed nearly unaltered by scatterers, and so that second-order (and higher) scattering is negligible. The Born approximation is implicitly assumed by most ultrasound beamforming algorithms, and the resulting beamforming algorithms are successful in scenarios with no aberrating layers and where the image mostly consists of weakly scattering contrasts.

Traditional approaches, however, are not effective when there are strongly aberrating layers placed between the sensor array and the ROI, even if scattering outside these layers is still linear, since it is not known how echoes propagate through this layer to the array of receiving sensors. To image correctly, we would have to know how the acoustic field propagates through the aberrating layer, which would require one to e.g. remove part of the skull, and thus violates the purpose of non-invasive imaging. Hence, we propose a method for *blindly* estimating an aberrating layer's transfer function. That is, to characterize it without an *a priori* known calibration image. Instead, we propose to measure many images of which only the second-order statistics need to be known, a condition which is much less demanding than prior knowledge of the true contrast image. We then use the second-order statistics of the set of measured echo signals in combination with the assumed known covariance of the images, to estimate the layer's transfer function.

An interesting layer in the context of this paper is the human skull. As mentioned earlier, it is known to cause strong aberrations, and prohibits imaging of the brain, as well as non-invasive ablation of (malignant) tissue using high intensity focused ultrasound

(HIFU) [2]. This is highly relevant for recent advances in ultrasound such as functional ultrasound imaging (fUS) [6–8]. Functional US imaging allows for inferring brain activity based on changes in the local haemodynamics that cause measureable changes in the ultrasound Doppler signal. Compared to fMRI, fUS allows for a more affordable, portable, and precise way of imaging brain activity [9]. It helps us to better understand the human brain, and it is a promising tool for surgeons to more accurately distinguish between healthy and malignant tissue when removing a brain tumor, preventing a surgeon from accidentally damaging healthy tissue, which could have disastrous results for the patient [10]. All of these examples urge us to investigate more advanced imaging techniques to deal with the distortions caused by the aforementioned layers.

Most studies address this problem by applying a time shift to each measured signal in the array (i.e., it is assumed aberrations are modeled using a phase screen directly in front of the array), based on a variety of goodness of fit criteria. Some examples from many are the studies of [11–16]. However, these approaches often need to estimate a different phase screen for each imaging direction (direction into the ROI w.r.t. to ultrasound array), since a single phase screen would not be able to model more complex layers for all directions simultaneously. Moreover, one could argue such models are not accurate for more complicated layers where e.g. internal multiple scattering takes place. In such cases, the echo from a reflector as measured on different array elements could be different not only by a time shift, but also by the shape of pulse-echo signal. In addition, this makes it hard to use the methods above, as they rely on delay-and-sum beamforming, which requires signals to be in-phase after aberration correction and delaying before summing. A more holistic approach would view the US field in a single sensor on the array-side of the layer to be a *linear combination* of the *entire* 'input' field on the other side (Fig. 7.1).

Another large body of research is specifically focused on imaging through the human skull. In [17], time-reversal techniques are used to focus waves on transmit inside the brain. This is useful for acoustic therapy, but less so for directly imaging the brain. In [18], a similar goal is achieved by the design of a spatio-temporal filter. However, both of these techniques require one to first measure the transmitted ultrasound waves inside the brain to characterize the joint effect of the imaging setup and the skull, prohibiting non-invasive procedures. Recently, trans-cranial brain vasculature imaging was demonstrated using localization microscopy utilizing micro-bubbles inserted into the brain vasculature [19]. Their work exploits the use of micro-bubbles, which can be isolated in the back scattered echo signals (reflectors are spaced sufficiently apart in the brain vessels to do so), to analyze aberration profiles and model them using a phase-screen for regions in the ROI where bubbles were estimated to have roughly similar aberration profiles.

Another approach was proposed in [20], where two opposing ultrasound arrays are placed on the skull. The effect of the two skull layers on each side has to be estimated. To simplify matters, the authors assume that the skull wall on one side can be approximated as a thin phase layer, allowing for a more complex formulation and estimation of the effect of the other skull wall. However, a thin phase screen is arguably not an accurate approximation of the skull wall and consequently, it is possible that the estimated skull wall's transfer functions are sub-optimal. The technique proposed in our paper only

involves a single skull wall's transfer function, and it does not rely on approximations using thin phase screens.

A more recent approach ignores the Born approximation entirely, and tries to solve the acoustic wave equations using gradient descent optimization techniques [21], known as full waveform inversion (FWI). Their method relies on a sufficient initial estimate of the skull wall, and requires a measurement setup where sensors are distributed across the entire skull, instead of using only an array. Moreover, the technique is currently very computationally intensive, and requires many hours to obtain a single (3D) frame. This currently makes it entirely unsuitable for Doppler or fUS imaging, where hundreds to thousands of frames need to be imaged to obtain a single Doppler or fUS image.

Finally, we have already proposed an SVD-based technique to blindly estimate the effect of the aberrating layer from random images or source signals in the ROI. However, it relies on structural requirements of the aberration layer's transfer functions, and on the setup of the imaging scenario [22]. In the techniques we will discuss in this paper, in contrast to existing studies, we will propose a more flexible and general method, where many kinds of layers can be characterized blindly, and without the need for two sensor arrays.

The method we propose in this new contribution heavily relies on *model-based imaging* techniques. Such techniques take into account the entire wave propagation model when forming an image, and try to find an image that, based on a model that relates the image to measured echo signals, best explains the observed echoes. Such techniques have been used before (see e.g., [23, 24], to just name a few), but are especially relevant to scenarios where traditional imaging techniques like delay-and-sum cannot be utilized. We have particularly shown this in [25, 26], where we intentionally distort transmitted fields using an aberration layer, and show that this allows 3D imaging using just a single sensor. In [25, 26], we carefully characterize the aberrating layer in order to incorporate it into the wave propagation model, after which an image is formed from echo signals based on this model. Hence, even heavily distorted signals where the echo of a single reflector is different on each sensor, and strongly smeared across time could be resolved, as long as it is known *how* the distortion takes place.

In this contribution we will show how to explicitly incorporate an extra variable representing the layer's transfer function into our echo signal model, an approach which has partly been introduced before in [20], but has not seen much use in literature. Specifically, we use the concept of a transfer function from an imaginary (virtual) array just in front of the aberrating layer, to the true sensor after the layer (Fig. 7.1). Next, we will show how to get rid of the contrast image variable, which we assume is unknown. This is done by collecting many measurements, and looking at their second-order statistics, instead of the raw measurements directly. As a result, the contrast image variable is integrated out of the signal model equations, after which the layer's transfer function is the only unknown. Finally, we propose a method based on covariance matching [27], using a manifold optimization technique, to obtain an estimate of the layer's transfer function. Other work in ultrasound imaging exploiting second-order statistics can be found in [28, 29], where the cross-correlations of the measured signals are used. However, these papers are still focused on time delays or assume that the aberrating layer acts identically on each point of the layer. More general layer transfer functions with

interaction between channels are not included.

Using covariance measurements to find the array gain and phase has been studied in signal processing literature before, most notably by Friedlander and Weiss [30]. Covariance matching for array calibration has also been applied before to antenna arrays for radio astronomy imaging [31], and to acoustic vector sensor arrays [32]. However, only phase and gain changes per antenna are considered (instead of linear combinations of inputs), similar to [12, 13, 16] in ultrasonic imaging, instead of more complicated transfer functions. They thus assume a specific structure for the layer's transfer function, and can thus be seen as a special case of our approach. The authors use alternating descent and semidefinite programming methods, which are not directly applicable to the mathematical problem we discuss in this paper. As far as we know, there has been no prior work where the estimation of general transfer functions using covariance matching is considered.

NOTATION
We will represent a vector $\mathbf{x}$ in lower-case bold, and a matrix $\mathbf{X}$ in upper-case bold. The operator $\mathbb{E}\{.\}$ represents the statistical expectation operator, $\|\mathbf{X}\|_F$ the Frobenius norm of $\mathbf{X}$, and $\|\mathbf{x}\|_2$ the l_2-norm of $\mathbf{x}$. The notation $\mathbf{X}^\dagger$ represents the Moore-Penrose pseudo-inverse of $\mathbf{X}$. The matrix transpose and conjugate transpose are denoted using $(.)^\mathsf{T}$ and $(.)^\mathsf{H}$, respectively. Finally, the $\otimes$ symbol represents the Kronecker product.

7.2. SIGNAL MODEL AND PROBLEM FORMULATION
In the case of a homogeneous medium with an acoustic contrast image, and using the Born approximation, we can assume the following linear measurement model for an M element array, and a single temporal frequency l:

$$\mathbf{y}_l = \mathbf{A}_l\mathbf{x} + \mathbf{n}_l, \qquad (7.1)$$

where $\mathbf{y}_l \in \mathbb{C}^M$ groups the array pressure measurements for frequency l, $\mathbf{x} \in \mathbb{R}^N$ represents the scattering coefficients for N pixels, and $\mathbf{A}_l \in \mathbb{C}^{M \times N}$ is a known model that relates the measurements in $\mathbf{y}_l$ to the image $\mathbf{x}$, and thus contains the expected reflected signals at frequency l received by the array for all pixels of interest. Finally, $\mathbf{n}_l$ represents additive measurement noise. The matrix $\mathbf{A}_l$ is well known when no aberration layer is present, and the Born approximation holds. It is easily generated since the Green's functions for the pulse propagation paths (from transmitter, to each pixel, to each receiver) and the speed of sound are known. In that case, the received echo signal for each such path is straightforwardly found by delaying the transmitted pulse based on the total path length. Under the Born approximation, the scattering intensity of each pixel depends only on the difference in speed of sound and density between the pixel and the background speed of sound (contrast image), and is thus real-valued and frequency-independent. With no aberration layer present, image reconstruction boils down to estimating $\mathbf{x}$ from $\mathbf{y}_l$, knowing $\mathbf{A}_l$, using the entire available bandwidth instead of a single frequency.

In the case of an abberation layer, we assume there is a linear transfer function from each point on the virtual array in Fig. 7.1 to each element in the real array. That is,
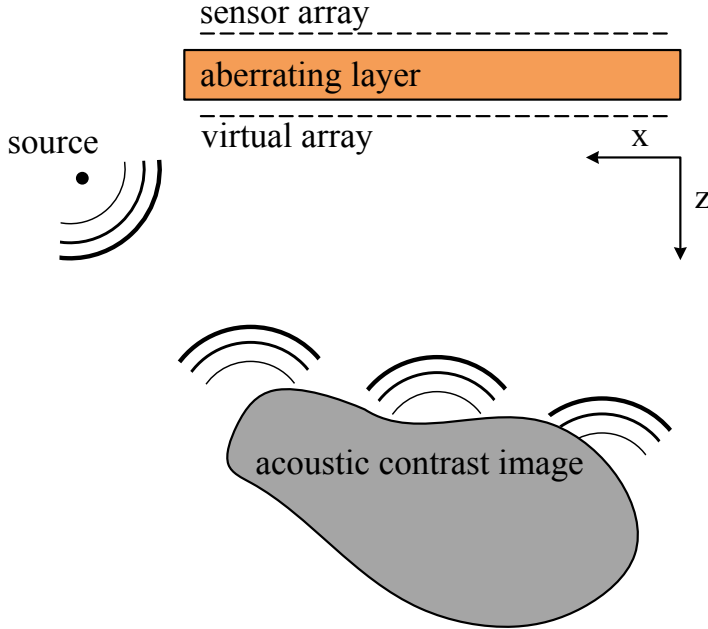
Figure 7.1: An aberration layer distorts incident wavefields, leading to incorrect estimation of the acoustic contrast image. We are interested in estimating the transfer function from each point on the virtual array just before the aberration layer, to each point on the real array.

7

there exists a Green's function $h_{i,j}(t)$ that represents the impulse response for the linear time-invariant channel from point i on the virtual array to sensor j on the real array. The measurement on any given sensor j is then the superposition of all the convolved channel responses from each virtual sensor i to j. In the frequency domain, that means the problem can be approached on a per-frequency basis, and we have the following new measurement model:

$$\mathbf{y}_l = \mathbf{H}_l \mathbf{A}_l \mathbf{x} + \mathbf{n}_l, \tag{7.2}$$

where $\mathbf{A}_l \in \mathbb{C}^{M \times N}$ is now related to the virtual array, and $\mathbf{H}_l \in \mathbb{C}^{M \times M}$ contains the frequency domain transfer function coefficients for frequency l, from each point on the virtual array to each point on the true array (note that in (7.2) it is assumed that waves travel through the aberrating layer only *once*[1], forcing us to use a transmitter that is not impeded by the aberrating layer). Since $\mathbf{H}_l$ represents the transfer function of an aberrating layer, it is unknown, and it is not possible to estimate $\mathbf{x}$.

When considering multiple frequencies, the $\mathbf{A}_l$ matrices of the corresponding L fre-

quencies can be stacked vertically into a bigger matrix $\mathbf{A} \in \mathbb{C}^{ML \times N}$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_L \end{bmatrix}. \tag{7.3}$$

Defining $\mathbf{y} \in \mathbb{C}^{ML}$ and $\mathbf{n} \in \mathbb{C}^{ML}$ in a similar way, we have a new multi-frequency measurement equation

$$\mathbf{y} = \mathbf{HAx} + \mathbf{n}. \tag{7.4}$$

Since convolution in the time domain becomes a product in the frequency domain, $\mathbf{H} \in \mathbb{C}^{ML \times ML}$ will have a block diagonal structure and mostly contains zeros,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & \dots & & \mathbf{H}_L \end{bmatrix}. \tag{7.5}$$

Note that the measurement equation (7.2) is simply a special case of (7.4) with $L = 1$. Furthermore, we point out that the usage of $\mathbf{H}$ generalizes all linear aberration layers. For example, if the aberrating layer is modelled as a phase screen, $\mathbf{H}$ will be a diagonal matrix, and is thus a special case of the formulation (7.4). Another example would be laterally invariant layers (i.e., layers that vary in the z-dimension, but not in the x-dimension), which would have a strictly Toeplitz structure.

In a *blind* calibration scenario, we wish to estimate $\mathbf{H}$ from (7.4), without knowing $\mathbf{x}$. Estimating both $\mathbf{H}$ and $\mathbf{x}$ jointly would result in a bi-linear measurement equation, which is typically challenging to solve. Instead, we consider removing $\mathbf{x}$ from our measurement equation by using covariance measurements, where $\mathbf{x}$ would be integrated out of the measurement equation, and only the covariance matrix of $\mathbf{x}$ would have to be known.

---

[1] In (7.2) and (7.4), $\mathbf{A}$ contains the pulse-echo signal received across the virtual array from each pixel. This implicitly relies on the known Green's functions in a homogeneous medium from transmitter to pixel to each point on the virtual array. Note that it is thus assumed that there is a separate transmitter from which a pulse propagates into the imaging medium, which reflects of contrast inhomogeneities in the ROI towards the virtual and true array. After all, if the array is used for transmission, we would not know the field that is transmitted into the ROI. In that case, (7.4) would contain the variable $\mathbf{H}$ *twice* (as pulses would propagate through the aberration layer both on transmit and receive), and (7.4) would become $\mathbf{y} = \mathbf{HA}\mathrm{diag}(\mathbf{x})\mathbf{A}^\top \mathbf{H}^\top \mathbf{s}$ (where $\mathbf{s} \in \mathbb{C}^{ML}$ would contain the transmitted pulses on the array, and $\mathbf{A}$ now represents the one-way propagation in a linear medium from a pixel to a point on the virtual array). This would make it much harder to estimate $\mathbf{H}$ from $\mathbf{y}$, since $\mathbf{y}$ would be quadratic w.r.t. $\mathbf{H}$ (and (7.9) would be bi-quartic w.r.t. $\mathbf{H}$), and the techniques described in the next section are not trivially applicable to this equation. Hence, it is assumed that a small transmitter is inserted into the medium. When the aberration layer consists of human tissue, this would mean that the calibration procedure would be minimally invasive, since only a small device has to be inserted. As a result, only *one* layer's transfer function has to be estimated, instead of two (like in [20]), and we do not have to estimate one of the layers using e.g. a thin phase mask (sometimes an insufficient approximation), thus retaining the validity of our model equation (7.4). Photoacoustic scenarios would also be covered under the formulation in (7.4), since in that case all pixels in the optic transmit view would be excited simultaneously, and start transmitting a pulse towards the array at the same time.

Therefore, we will interpret **x** as a realization of a random process for which multiple realizations are obtained. Consider for example a liquid medium with moving micro-bubbles or contrast agents, so that each pulse-echo measurement has a different contrast image **x** due to the displacement of the bubbles between measurements. Instead of bubbles, one could consider the flow of blood inside a patient. If stationary reflections are removed prior to calibration (using e.g. a high-pass filter in each pixel across the frame-dimension), one is left with only the signals originating from moving particles inside the blood stream. Assuming the measurement noise **n** is uncorrelated with the contrast image **x**, and defining the covariance matrices, $\mathbf{C_n} = \mathbb{E}\{\mathbf{nn}^H\}$, $\mathbf{C_y} = \mathbb{E}\{\mathbf{yy}^H\}$, and $\mathbf{C_x} = \mathbb{E}\{\mathbf{xx}^H\}$, we have the following measurement equation in the covariance domain:

$$\mathbf{C_y} = \mathbb{E}\{\mathbf{HAxx}^T\mathbf{A}^H\mathbf{H}^H\} + \mathbb{E}\{\mathbf{nn}^H\} \tag{7.6}$$

$$= \mathbf{HAC_xA}^H\mathbf{H}^H + \mathbf{C_n}. \tag{7.7}$$

Interestingly, the image variable **x** has been integrated out of the original measurement equation (7.4), and we now only need to know the second-order statistics of **x**. This is a much less demanding requirement. For example, if the ROI primarily consists of blood flow, and many pulse-echo measurements of that ROI are taken with sufficient time in between, it can be safely assumed that the corresponding **x**-vectors are uncorrelated, and thus $\mathbf{C_x} = \mathbf{I}$. Correlated measurements due to non-moving 'background' tissue can be easily removed from measurements using clutter filtering, as is common in US Doppler imaging. The covariance data of the measurement noise $\mathbf{C_n}$ can be obtained by taking measurements with no ultrasound transmission, and $\mathbf{C_y}$ can be estimated by taking P measurements, and then using the estimate

$$\hat{\mathbf{C}}_\mathbf{y} = \frac{1}{P}\sum_{p=1}^{P}\mathbf{y}^{(p)}\mathbf{y}^{(p)\,H} - \mathbf{C_n}, \tag{7.8}$$

where the known noise covariance matrix has already been subtracted for convenience in our derivations. We are interested, then, in estimating **H** from $\hat{\mathbf{C}}_\mathbf{y}$, knowing $\mathbf{C_x}$ and **A**. This is a covariance matching problem [27], where we want to minimize the following cost function:

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{H}} \|\hat{\mathbf{C}}_\mathbf{y} - \mathbf{HAC_xA}^H\mathbf{H}^H\|_F^2. \tag{7.9}$$

This is not a trivial problem to solve since **H** is *quartic* w.r.t. $\hat{\mathbf{C}}_\mathbf{y}$. In the next section we will provide methods for solving this optimization problem.

7.3. CALIBRATION ALGORITHMS

7.3.1. ALTERNATING MINIMIZATION

Defining $\mathbf{B} \triangleq \mathbf{AC_xA}^H$, one solution to estimate **H** from $\mathbf{C_y}$, assuming it's perfectly known for now, is given by

$$\hat{\mathbf{H}} = \mathbf{C}_\mathbf{y}^{\frac{1}{2}}\mathbf{B}^{-\frac{1}{2}}. \tag{7.10}$$

Here, we define the square root of a matrix $\mathbf{C}$ as any matrix $\mathbf{C}^{\frac{1}{2}}$ such that $\mathbf{C}^{\frac{1}{2}}\mathbf{C}^{\frac{H}{2}} = \mathbf{C}$. However, there is a larger set of solutions, which can be described by

$$\hat{\mathbf{H}} = \mathbf{C}_{\mathbf{y}}^{\frac{1}{2}}\mathbf{Q}\mathbf{B}^{-\frac{1}{2}}. \tag{7.11}$$

where $\mathbf{Q}$ is any orthogonal matrix, resulting in a non-unique solution.

To further narrow down the solution space, we assume that $\mathbf{H}$ can be parameterized by fewer parameters than the total number of entries in $\mathbf{H}$. These parameters will be represented by $\boldsymbol{\theta}$, and the related matrix is then denoted as $\mathbf{H}(\boldsymbol{\theta})$. Hence, to estimate $\mathbf{H}$, we would like to find the correct $\mathbf{Q}$ and $\boldsymbol{\theta}$ by solving the following problem instead of (7.9):

$$\{\hat{\boldsymbol{\theta}},\hat{\mathbf{Q}}\} = \arg\min_{\boldsymbol{\theta},\mathbf{Q}} \|\mathbf{C}_{\mathbf{y}}^{\frac{1}{2}}\mathbf{Q}\mathbf{B}^{-\frac{1}{2}} - \mathbf{H}(\boldsymbol{\theta})\|_F^2, \text{ s.t. } \mathbf{Q} \in \mathcal{Q}, \tag{7.12}$$

where $\mathcal{Q}$ represents the set of orthogonal $ML \times ML$ matrices. In other words, we try to find a solution that will give a zero cost function for (7.9) (any orthogonal $\mathbf{Q}$), while also penalizing the distance to the set of parameterized matrices $\mathbf{H}(\boldsymbol{\theta})$. Using this formulation the cost function is no longer quartic w.r.t. $\mathbf{H}$. However, even if $\mathbf{H}(\boldsymbol{\theta})$ is linearly parameterized, the problem above is not a convex problem, due to the orthogonality constraint.

In order to arrive at a solution for this problem, we propose to alternatingly optimize for $\boldsymbol{\theta}$ and $\mathbf{Q}$, as outlined in Algorithm 2. When $\mathbf{Q}$ is fixed, $\boldsymbol{\theta}$ can be found by solving a linear least squares problem. We assume that $\mathbf{H}$ can be linearly parameterized by $\boldsymbol{\theta}$ with system matrix[2] $\mathbf{G}$, such that $\text{vec}(\mathbf{H}(\boldsymbol{\theta})) = \mathbf{G}\boldsymbol{\theta}$. This way, solving for this step admits a unique global minimum which can be found either iteratively or analytically. When $\boldsymbol{\theta}$ is fixed, (7.12) becomes a weighted orthogonal Procrustes problem (WOPP). The latter is hard to solve, since it does not allow an analytical solution, the cost function contains multiple local minima, and iterative methods are not guaranteed to find the global minimizer (see [33] and the references therein). These two problems are solved alternatingly until the algorithm converges, or a maximum amount of iterations has been completed. In the next subsection we propose an improved version of this algorithm that addresses the problems mentioned above.

When jointly calibrating for multiple frequencies, the block-diagonal structure of $\mathbf{H}$ in (7.5) is enforced by $\mathbf{G}$, while additionally imposing a parameterization for the blocks on the diagonal representing each $\mathbf{H}_l$. Due to the block-diagonal structure of $\mathbf{H}$, *the number of unknowns increases linearly with each added frequency, whereas the number of new covariance measurements increases quadratically.* This is of course limited by the echo signal bandwidth and the Nyquist rate in both space and time. Moreover, space-frequency samples in $\mathbf{C}_{\mathbf{y}}$ tend to decorrelate for large frequency and space differences between entries. The additional information in $\mathbf{C}_{\mathbf{y}}$ consists of the cross-correlations between temporal frequencies, which now also have to be matched by the covariance-matching algorithm. Hence, we expect the probability of finding a unique solution for $\mathbf{H}$ to increase as the bandwidth L increases.

---

[2]A basis $\mathbf{G}$ can be found by simulating the layer of interest in an accurate simulator such was *k-Wave*, and studying the structure in $\mathbf{H}$. For the sake of simplicity, we used a 2D Fourier basis in our experiments, removing the high frequencies from the basis. Finding proper bases is a topic of future research.

The above procedure, summarized in Alg. 2, is an alternating projection algorithm, where $\boldsymbol{\theta}$ and $\mathbf{Q}$ are alternatingly projected onto a linear subspace, and the Stiefel manifold, respectively. We have not been able to find a similar problem or algorithm in literature, and are not aware of any analyses concerning convergence and global and local minima. However, given the non-convex constraint imposed by the Stiefel manifold, it is likely that there are multiple local minima. Moreover, there are at least two global minima, because the solutions $\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{Q}}\}$ and $\{-\hat{\boldsymbol{\theta}}, -\hat{\mathbf{Q}}\}$ should have the same cost function value in (7.12).

In [34], cyclical algorithms are proposed, which use a similar alternating minimization problem, although the authors in [34] are not necessarily looking for a unique solution, since any solution in the feasible set should suffice. In our work, however, we are looking for a specific solution: the one that is as close as possible to the true $\mathbf{H}$. The work does not analyse local and global minima. However, similar to [34], we can point out that (7.9) is zero if and only if (7.12) is also zero. Moreover, if the global minimum of (7.9) is small, then (7.12) should be small as well. It is then argued that the two cost functions are 'nearly' equivalent in the sense that if the global minimizer of (7.9) is sufficiently small, that the sequences minimizing both cost functions can be expected to lead to similar results. However, the cost functions are not equivalent, and will have different local minima and minimization sequences in general.

Algorithm 2 Alternating optimization algorithm

1: Input: $N_{it}, \mathbf{C_y}^{\frac{1}{2}}, \mathbf{B}^{-\frac{1}{2}}, \mathbf{G}$.
2: Initialize: random orthogonal matrix $\hat{\mathbf{Q}}$, $n = 0$.
3: **repeat**
4:     $\hat{\boldsymbol{\theta}} = \mathbf{G}^{\dagger} \text{vec}(\mathbf{C_y}^{\frac{1}{2}} \hat{\mathbf{Q}} \mathbf{B}^{-\frac{1}{2}})$
5:     $\hat{\mathbf{H}} = \text{unvec}\left(\mathbf{G}\hat{\boldsymbol{\theta}}\right)$
6:     Solve the WOPP $\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \|\mathbf{C_y}^{\frac{1}{2}} \mathbf{Q} \mathbf{B}^{-\frac{1}{2}} - \hat{\mathbf{H}}\|_F^2$
7:     $n = n + 1$
8: **until** $n = N_{it}$

7.3.2. Alternating optimization for ill-conditioned $\mathbf{B}^{\frac{1}{2}}$ and low-rank $\mathbf{C_y}^{\frac{1}{2}}$

Algorithm 2 has several drawbacks that we intend to address in this subsection. Most importantly, a WOPP has to be solved in Alg. 2, which requires iterative methods with no guarantee of obtaining a global minimizer. The altered cost function proposed in this subsection will lead to a formulation where only an *unweighted* orthogonal Procrustes problem needs to be solved, for which an analytical solution is available. Another issue is that when $\mathbf{H}$ is low-rank, $\mathbf{C_y}^{\frac{1}{2}}$ is low rank as well. Consequently, there will be no unique solution to the weighted orthogonal Procrustes problem in Alg. 2. Moreover, the matrix $\mathbf{A}$ is typically ill-conditioned in the case of (multi-frequency) imaging problems, making the inversion of $\mathbf{B}^{\frac{1}{2}}$ problematic, and also causing $\mathbf{C_y}^{\frac{1}{2}}$ to be ill-conditioned. The

issue of ill-conditioning is the one we encountered in our experiments, and thus of most importance. For the sake of completeness, we also show how to deal with a low-rank $\mathbf{C_y}$.

To address these issues, we start by observing that instead of solving (7.9), we can also take the matrix square root of the terms in (7.9), similar to the work in [34], to obtain a solution using the following minimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{C_y}^{\frac{1}{2}} - \mathbf{H}(\boldsymbol{\theta})\mathbf{A}\mathbf{C_x}^{\frac{1}{2}}\|_F^2 \tag{7.13}$$

$$= \arg\min_{\boldsymbol{\theta}} \|\mathbf{C_y}^{\frac{1}{2}} - \mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}}\|_F^2. \tag{7.14}$$

Of course, this is only one of many solutions, and similar to the previous subsection, multiplication by an orthogonal matrix $\mathbf{Q}$ of the left term will also provide a solution to the original cost function (7.9). Incorporating this term into the new cost function leads to:

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{Q}}\} = \arg\min_{\boldsymbol{\theta}, \mathbf{Q}} \|\mathbf{C_y}^{\frac{1}{2}}\mathbf{Q} - \mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}}\|_F^2, \text{ s.t. } \mathbf{Q} \in \mathscr{Q}. \tag{7.15}$$

When $\boldsymbol{\theta}$ is fixed, solving for $\mathbf{Q}$ now has become an OPP instead of a WOPP, and the global optimum for the OPP in the current iteration can be found analytically. The solution for $\boldsymbol{\theta}$ when $\mathbf{Q}$ is fixed is found by solving a (linear) least squares problem. In case of an ill-conditioned $\mathbf{B}^{\frac{1}{2}}$, one can use any regularization technique when solving for $\boldsymbol{\theta}$, such as Tikhonov regularization. As discussed in the previous subsection, the local minima and minimization sequences of (7.9) and (7.15) are not equivalent in general, although they should have similar behaviour for small enough values of the cost functions, and (7.9) can only be zero if (7.15) is also zero.

Next, we rewrite the problem to only solve for the column space of $\mathbf{C_y}^{\frac{1}{2}}$ as follows. First, we define the *economy size* SVD of $\mathbf{C_y}$: $\mathbf{C_y} = \mathbf{U}\Sigma\mathbf{U}^H$, and use $\mathbf{C_y}^{\frac{1}{2}} = \mathbf{U}\Sigma^{\frac{1}{2}}\mathbf{U}^H$. Here $\mathbf{U} \in \mathbb{C}^{ML \times k}$, where $k = \text{rank}(\mathbf{C_y})$, and $\Sigma \in \mathbb{C}^{k \times k}$. Next, we rewrite the cost function in (7.15) by left-multiplying the terms in (7.15) by $\mathbf{U}^H$, leading to the problem

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{Q}}\} = \arg\min_{\boldsymbol{\theta}, \mathbf{Q}} \|\Sigma^{\frac{1}{2}}\mathbf{U}^H\mathbf{Q} - \mathbf{U}^H\mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}}\|_F^2, \text{ s.t. } \mathbf{Q} \in \mathscr{Q}, \tag{7.16}$$

or equivalently,

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{P}}\} = \arg\min_{\boldsymbol{\theta}, \mathbf{P}} \|\Sigma^{\frac{1}{2}}\mathbf{P} - \mathbf{U}^H\mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}}\|_F^2, \text{ s.t. } \mathbf{P} \in \mathscr{Q}_{k \times ML}, \tag{7.17}$$

where $\mathbf{P} = \mathbf{U}^H\mathbf{Q}$, and $\mathscr{Q}_{k \times ML}$ represents the set of $k \times ML$ matrices with unit-norm orthogonal rows. The solution to (7.17) for $\mathbf{Q}$ (an OPP) is given by [33]:

$$\hat{\mathbf{P}} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^H, \tag{7.18}$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ collect the k most significant left and right singular vectors of $\mathbf{Z} = \Sigma^{\frac{H}{2}}\mathbf{U}^H\mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}} \in \mathbb{C}^{k \times ML}$, respectively.

Although we have changed the cost function, it is now more easily solved for $\mathbf{Q}$ if $\boldsymbol{\theta}$ is fixed: instead of a WOPP, we only have to solve an OPP, for which a global minimum can be found analytically. Moreover, it avoids computing the inverse of $\mathbf{B}^{\frac{1}{2}}$, and ensures there is a unique solution to the OPP when $\mathbf{C_y}$ is rank-deficient, by only solving for the column space of $\mathbf{C_y}$. We will use cost function (7.17) throughout this paper to avoid the aforementioned issues. The updated algorithm is shown in Alg. 3.

Algorithm 3 Alternating optimization for ill-conditioned $\mathbf{B}^{\frac{1}{2}}$ and low-rank $\mathbf{C_y}^{\frac{1}{2}}$

1: Input: N_{it}, $\mathbf{B}^{\frac{1}{2}}$, $\mathbf{G}$, $\mathbf{U}$, Σ.
2: Initialize: random matrix $\mathbf{P} \in \mathscr{Q}_{k \times ML}$, $n = 0$.
3: **repeat**
4: $\quad \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{C_y}^{\frac{1}{2}} - \mathbf{H}(\boldsymbol{\theta})\mathbf{B}^{\frac{1}{2}}\|_F^2 + \lambda\|\boldsymbol{\theta}\|_2^2$
5: $\quad \hat{\mathbf{H}} = \text{unvec}(\mathbf{G}\hat{\boldsymbol{\theta}})$
6: $\quad$ Decompose: $\Sigma^{\frac{H}{2}}\mathbf{U}^H\hat{\mathbf{H}}\mathbf{B}^{\frac{1}{2}} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^H$
7: $\quad \hat{\mathbf{P}} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^H$
8: $\quad n = n + 1$
9: **until** $n = N_{it}$

7.3.3. MANIFOLD GRADIENT DESCENT

From Alg. 2, line 4, we can directly substitute the solution for $\boldsymbol{\theta}$ into (7.12), resulting in the following simpler cost function where $\mathbf{Q}$ is the only optimization variable:

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}\in\mathscr{Q}} \|\text{vec}(\mathbf{C_y}^{\frac{1}{2}}\mathbf{Q}\mathbf{B}^{-\frac{1}{2}}) - \mathbf{G}\mathbf{G}^\dagger\text{vec}(\mathbf{C_y}^{\frac{1}{2}}\mathbf{Q}\mathbf{B}^{-\frac{1}{2}})\|_2^2,$$

$$= \arg\min_{\mathbf{Q}\in\mathscr{Q}} \|\mathbf{F}\text{vec}(\mathbf{Q})\|_2^2, \tag{7.19}$$

where $\mathbf{F} = (\mathbf{I} - \mathbf{G}\mathbf{G}^\dagger)(\mathbf{B}^{-\frac{T}{2}} \otimes \mathbf{C_y}^{\frac{1}{2}})$. Given a $\mathbf{Q}$, the operator $\mathbf{F}$ first computes the corresponding estimate of $\mathbf{H}$, $(\mathbf{B}^{-\frac{T}{2}} \otimes \mathbf{C_y}^{\frac{1}{2}})\text{vec}(\mathbf{Q})$, and then projects it onto the nullspace of $\mathbf{G}$. Toolboxes such as ManOpt [35] can be used to find a solution to (7.19), which uses a Riemannian based gradient descent method.

In a similar fashion, (7.17) can be rewritten as

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}\in\mathscr{Q}_{k\times ML}} \|\text{vec}(\Sigma^{\frac{1}{2}}\mathbf{P}) - (\mathbf{B}^{\frac{T}{2}} \otimes \mathbf{U}^H)\mathbf{G}\boldsymbol{\theta}\|_2^2 \tag{7.20}$$

$$= \arg\min_{\mathbf{P}\in\mathscr{Q}_{k\times ML}} \|\text{vec}(\Sigma^{\frac{1}{2}}\mathbf{P}) - \tilde{\mathbf{G}}\tilde{\mathbf{G}}^\dagger\text{vec}(\Sigma^{\frac{1}{2}}\mathbf{P})\|_2^2, \tag{7.21}$$

where $\tilde{\mathbf{G}} = (\mathbf{B}^{\frac{T}{2}} \otimes \mathbf{U}^H)\mathbf{G}$. Defining $\tilde{\mathbf{F}} = (\mathbf{I} - \mathbf{G}\mathbf{G}^\dagger)(\mathbf{I} \otimes \Sigma^{\frac{1}{2}})$, we end up with

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}\in\mathscr{Q}_{k\times ML}} \|\tilde{\mathbf{F}}\text{vec}(\mathbf{P})\|_2^2. \tag{7.22}$$

Our initial investigations showed that Alg. 3 converges faster with faster iterations, so we do not consider these gradient descent methods in the Results section. On the other hand, they show what kind of problem is being solved in a more compact manner.

7.4. NUMERICAL AND SIMULATION RESULTS

We will evaluate the proposed algorithm using various simulated aberration layers, in increasing complexity and difficulty. Our main approach will be as follows. First, we will design the layer of interest in the *k-Wave* toolbox [36]. To avoid lengthy simulation times, and to have a reference ground truth, we directly measure the $\mathbf{H}_l$ matrices in *k-Wave*, from which we can form $\mathbf{H}$. To do so, we transmit the excitation pulse (the same one that we would transmit from our transmit transducer) from one point on the virtual array, and measure the entire response on the receiving array. This single simulation is then repeated for each virtual array point. This $\mathbf{H}$ will act as our ground truth throughout most of the experiments, and will also be used to efficiently generate many measurements. The matrix $\mathbf{A}$ is generated assuming wave propagation from a known source location, to each pixel, towards the virtual array. For this, we assume a homogeneous background medium with a speed of sound of 1490 m/s, and a density of 1000 kg/m^3 (similar to human tissue).

Since $\mathbf{H}$ is now known, we can generate synthetic calibration measurements using (7.4) for many random $\mathbf{x}$. To this end, we generate multiple realizations of a spatially white zero mean i.i.d. Gaussian image, with $\mathbf{C_x} = \mathbf{I}$. The contrast images are zero mean, because we assume that the speed of sound of the inhomogeneties (consider e.g. various moving particles in a blood stream) can be either higher or lower than the background medium speed of sound. To every measurement we add a different realization of spatially and temporally white zero mean i.i.d. Gaussian measurement noise, using $\mathbf{C_n} = \sigma_n^2 \mathbf{I}$. From these synthesized measurements, we use the estimated covariance matrix as defined in (7.8), using $\mathbf{C_n} = \sigma_n^2 \mathbf{I}$, to estimate $\mathbf{H}$. We assume the noise covariance is known, since it could be measured, e.g., by measuring on the used ultrasound transducer without any echo signals.

Since the measured $\mathbf{H}$ acts as ground truth, we can compare our estimated $\mathbf{H}$ with the true one. We will use the normalized correlation coefficient

$$\rho = \text{vec}(\mathbf{H})^{\mathsf{H}}\text{vec}(\hat{\mathbf{H}})/(\|\mathbf{H}\|_F \|\hat{\mathbf{H}}\|_F) \tag{7.23}$$

to do so. However, how well $\mathbf{H}$ describes the true propagation throughout the aberrating layer depends on *(a)* whether the assumption of linear propagation holds, and *(b)* whether the virtual array samples the incident ultrasound field sufficiently well. Since we will generate measurement data for estimating $\mathbf{C_y}$ using an assumed true $\mathbf{H}$ as described above, these assumptions are automatically satisfied. However, we will verify if we can truly obtain better imaging performance, by performing an extra imaging experiment (i.e., estimating $\mathbf{x}$ from a single measurement $\mathbf{y}$, using a $\mathbf{y}$ obtained directly from a *k-Wave* simulation, instead of using $\mathbf{H}$ to generate $\mathbf{y}$).

We use Alg. 3 based on the alternative cost function described in Sec. 7.3.2 for all the results, using $k = LM$ (i.e., we don't use a low-rank approximation). We found it unnecessary to use regularization as described by Alg. 3.

For the sake of simplicity, all pixels act as independent sources, omni-directionally transmitting a Gaussian pulse centered around 5 MHz, unless otherwise specified. Note that this does not invalidate our model assumptions, since the Born approximation ignores multiple scattering, implying each acoustic contrast acts as an independent source.

In the final experiments, we will forego these assumptions, and use contrast-based scattering to validate our methods.

7.4.1. TRIPLE-LAYERED ABERRATION LAYER

First, we investigate the aberration layer in Fig. 7.2. It consists of a water-layer 'sandwiched' in between two homogeneous bone layers. It is important to realize that this layer is laterally *space-invariant*, so that the spatial transfer function also becomes space-invariant (i.e., it is only dependent on the lateral distance between the virtual and true array element). Consequently, each $\mathbf{H}_l$ will have a Toeplitz structure, which was also confirmed by inspecting the $\mathbf{H}_l$ matrices obtained from *k-Wave* simulations. Using this Toeplitz structure, $\mathbf{G}$ only requires $2M - 1$ basis vectors per frequency, resulting in less unknowns. We use random scatterers in a ROI of 10 mm width by 8 mm depth for calibration. The calibration is performed for each temporal frequency independently ($L = 1$), and using $P = 500$ frames. Both the true and virtual array have $M = 20$ point-like sensors, placed at 175 μm intervals.

In Fig. 7.2, we show the calibration performance for various SNRs and temporal frequencies. Since the signal power is frequency dependent, we adjust the noise power accordingly to obtain the same SNR for each frequency. In the case of joint multi-frequency calibration in subsequent subsections we don't do this, so that the SNR will vary per frequency. The transfer functions in $\mathbf{H}$ are Toeplitz for all frequencies, and thus the calibration performance depends only on SNR.

Finally, the bottom half of Fig. 7.2 shows the imaging results using our calibration method for a single point and a small collection of points. After estimating $\mathbf{H}$ for all temporal frequencies, all frequencies are combined when estimating the contrast image $\mathbf{x}$. All measurement data $\mathbf{y}$ is obtained from *k-Wave* simulation experiments. All images are reconstructed using the LSQR algorithm [37] to solve (7.2) for $\mathbf{x}$, which is regularized by limiting the amount of LSQR iterations. As can be seen from the single point, the uncalibrated PSF deteriorates in the axial direction due to the multiple reflections in the aberration layer, which are mistaken as echoes from separate sources when no calibration is utilized. This becomes even more clear in Fig. 7.2, bottom right, where the calibrated model is able to resolve most scatterers, in contrast to the uncalibrated model. The calibrated model gives much better reconstruction results, on par with the perfectly calibrated model.

We also attempted to obtain a good image by only tweaking the speed of sound in $\mathbf{A}$ (and $\mathbf{H} = \mathbf{I}$). However, we were not able to obtain a reasonable image reconstruction using this approach, since the multitude of layers reverberate, which is probably the main limiting factor for imaging through this layer.

7.4.2. JOINT MULTI-FREQUENCY CALIBRATION: SMOOTHLY VARYING LAYER

In the previous subsection we showed that calibration is possible using a single frequency if $\mathbf{H}_l$ has a Toeplitz structure, i.e., when the aberration layer is space-invariant in the x-dimension. However, we have encountered significant difficulty calibrating for a single frequency for non-Toeplitz matrices. In such cases, we typically are able to find many solutions with a near-zero cost function.

In this subsection, we resolve this issue by using the multi-frequency approach to
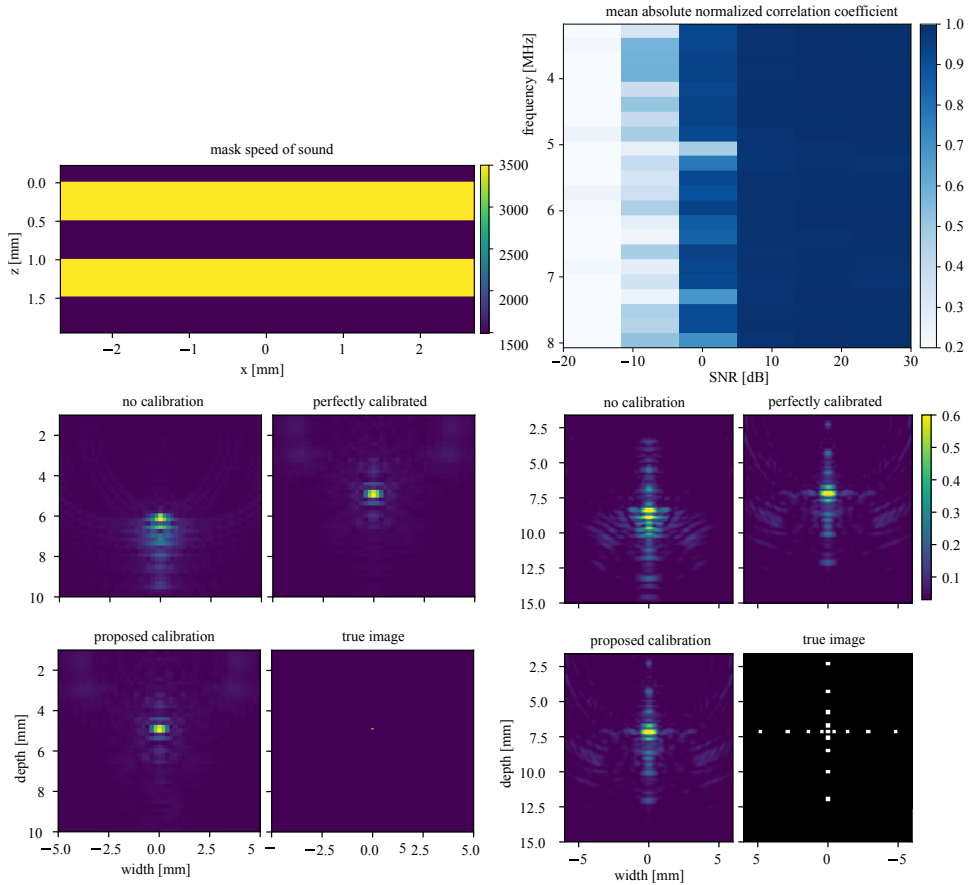
Figure 7.2: Result for the simulations in Sec. 7.4.1. *Top left:* Laterally space-invariant aberration layer. *Top right:* Calibration performance per frequency and SNR. *Bottom left:* Imaging PSF performance comparison with/without calibration, for $P = 500$, and an SNR of 25 dB. *Bottom right:* Test image reconstruction performance comparison with/without calibration for reflectors at various inter-distances, for $P = 500$, and an SNR of 25 dB.
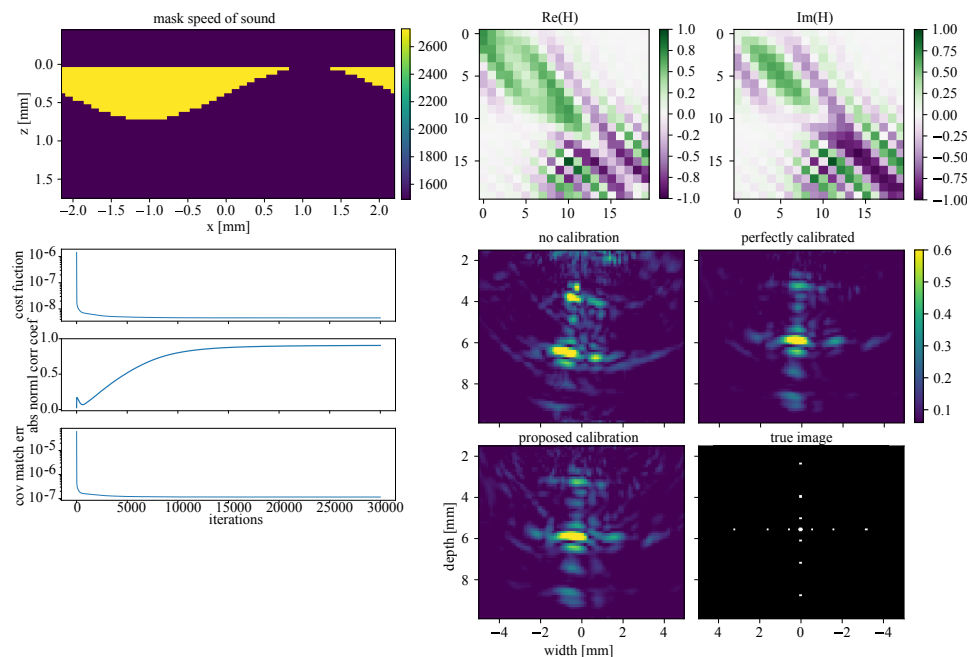
Figure 7.3: *Top left:* Simulated smooth aberration layer used for the simulations in Sec. 7.4.2. *Top right:* **H**-matrix for the aberration layer, for a frequency of 4.4 MHz. *Bottom left:* Various cost functions per iteration for the results of Sec. 7.4.2: proposed cost function (7.17), absolute normalized correlation coefficient (7.23), covariance matching error (7.9). *Bottom right:* Test image reconstruction performance comparison with/without calibration for reflectors at various interdistances, for $P = 500$, and an SNR of 25 dB.

still obtain a good estimate of **H**. We simulate a homogeneous, smoothly shaped layer as shown in Fig. 7.3. For the parameterization θ, we use the fact that the measured $\mathbf{H}_l$-matrices in *k-Wave* have a clear and smooth structure (Fig. 7.3, top right). Although probably not optimal, for now we will use a 2D FFT parameterization containing only the lower frequencies. The random images are located in a ROI of 10 mm in width by 5 mm in depth. Both the virtual and true array contain $M = 20$ elements, spaced at 100 μm intervals.

We calibrate jointly for equidistantly spaced temporal frequencies between 1 and 7.5 MHz using $L = 36$. We form **G** as mentioned before, using the 355 largest 2D FFT coefficients (out of 400). Using $P = 10,000$ and an SNR of 30 dB, we obtain an absolute normalized correlation coefficient of 0.94 after 30,000 iterations. When using an SNR of 20 dB, we obtain an absolute normalized correlation coefficient of 0.91. The used parameterization is very general and could possibly be used for a variety of aberration layers. Tighter parameterizations could lead to improved estimation results and are a topic of future research, although it is encouraging that the proposed method is able to correctly calibrate using many unknowns.

In the bottom left of Fig. 7.3 we show the corresponding cost function. Interestingly, although we used the modified cost function of Sec. 7.3.2, both the alternative cost function and the original covariance matching error decrease in a similar manner.

In the bottom right of Fig. 7.3, we show an image reconstruction for a phantom generated using k-wave. All image reconstructions are obtained using the LSQR algorithm with limited iterations. A number of observations can be made. Firstly, none of the reconstructions are of particularly high quality. This can be attributed to the fact that the $\mathbf{H}_l$ matrices for this aberration layer are *low rank*. In other words, one would not be able to reconstruct (using e.g. a least-squares approach) the 'input' field to the layer from the 'output' field, since **H** is low rank or ill-posed. Thus, given the general model equation $\mathbf{y} = \mathbf{HAx}$, information about **x** is lost due to the conditioning and/or rank of **H**. Secondly, we observe that knowledge of **H** strongly improves the image reconstruction, making it apparent that a cross-like image is present. Finally, the image result using the estimated **H** comes close to the perfect calibration case, with slightly worse background artefacts due to sidelobes etc. Note that sidelobes and other imaging artifacts do not have the same constitution as in imaging scenarios without an aberrating layer, where traditional imaging techniques, and their corresponding analysis, are applicable.

7.4.3. Joint multi-frequency calibration: skull layer

As a more clinically relevant example, we calibrate a skull-like layer, consisting of a porous bone layer sandwiched between two homogeneous bone layers (Fig. 7.4). We exploit the fact that each $\mathbf{H}_l$ matrix can be well represented using the 2D FFT basis, omitting the highest frequencies, so that we have to estimate 342 variables (out of 400) per frequency. The random images are located in a ROI of 15 mm in width and 20 mm in depth. The true and virtual arrays contain $M = 20$ point-like sensors, spaced at 350 μm intervals.

We calibrate for $L = 46$ equidistantly spaced temporal frequencies between 1.8 and 6.0 MHz. Using $P = 10,000$ and an SNR of 30 dB, we obtain an estimate of **H** with an absolute normalized correlation coefficient of 0.75. Fig. 7.5 shows the various cost functions for 30,000 iterations.
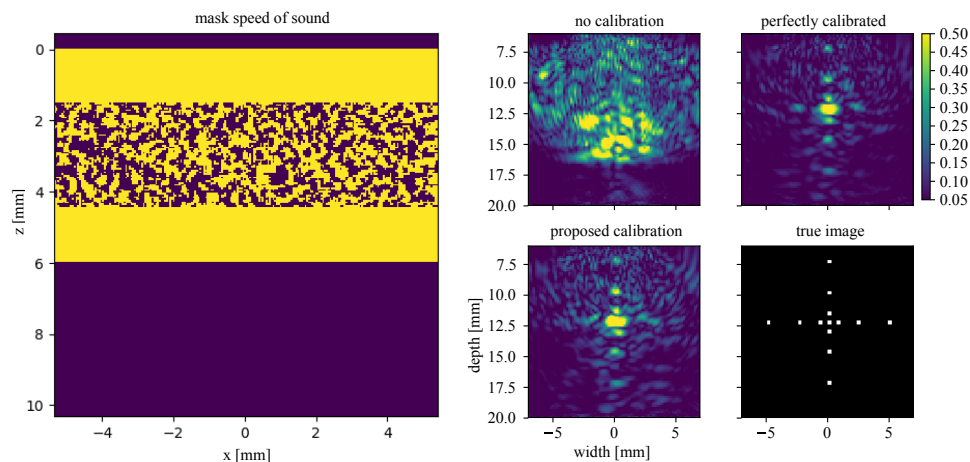
Figure 7.4: *Left:* Results for the simulations in Sec. 7.4.3. Simulated skull-like aberration layer in k-wave, used for multi-frequency calibration experiments. *Right:* Image reconstruction example of simulated k-wave data of a set of point sources with varying inter-distances.
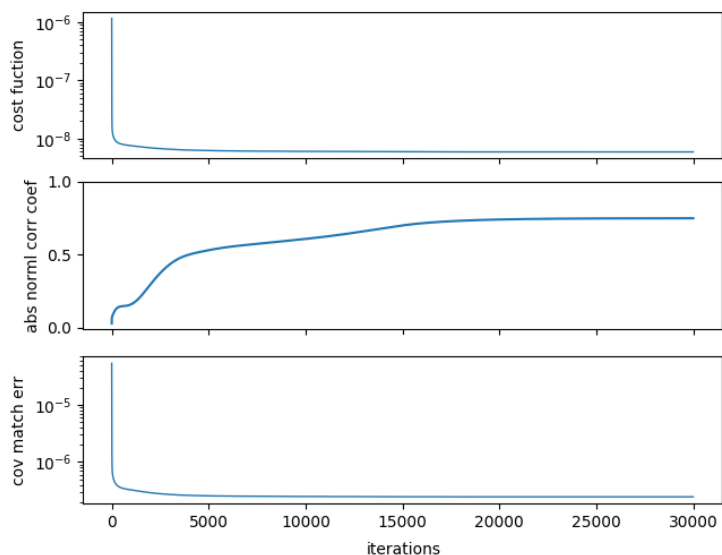
Figure 7.5: Various cost functions per iteration for the results of Sec. 7.4.3: proposed cost function (7.17), absolute normalized correlation coefficient (7.23), covariance matching error (7.9).

In Fig. 7.4 we show the image reconstruction from RF data obtained from a *k-Wave* simulation. Image reconstructions are obtained using the LSQR algorithm to solve (7.2) with limited iterations. In this case, an enormous reconstruction improvement is obtained, as compared to imaging without any model calibration. As in the previous subsections, we tried to find the best image reconstruction for the calibration-less case by simply tweaking the speed of sound, but were not able to find a satisfactory result. The one in Fig. 7.4 was the best one obtained using a lower speed of sound than the background medium.

7.4.4. BLIND CALIBRATION FROM BLOOD FLOW IN A VASCULATURE NETWORK

Next, we consider a more realistic scenario where images are not purely random. Instead, a vasculature tree with flowing particles is used, to emulate a more realistic and practical imaging scenario. Additionally, we will use true contrast reflections by simulating contrast-based reflections in *k-wave* instead of making each reflector a source as in the earlier results. This scenario is thus more realistic, and we can expect more model errors since the Born approximation may not strictly hold.

We use random i.i.d. normally distributed scatterers, where the mean speed of sound and density is that of the background medium (water). This field of random reflectors is shifted towards the array with a fixed distance every measurement (500 μm steps). The random reflectors only occur inside the vasculature tree, and are set to the background medium properties when outside the vasculature. When computing the measurement covariance matrix $\hat{\mathbf{C}}_{\mathbf{y}}$ from such a set of images, we found it to be nearly identical to the measurement covariance matrix from completely random images (where $\mathbf{C_x} = \mathbf{I}$), and hence we will assume $\mathbf{I}$ as the *a priori* known image covariance matrix. *Note that we simulate each of the $P = 1,000$ frames in* k-Wave, *instead of using* $\mathbf{H}$ *to generate many measurements.*

As an aberration layer, we use the same triple layer and Toeplitz basis as used in Sec. 7.4.1. The vasculature is positioned in a ROI from 6 to 22 mm depth, and 20 mm in width. The measurement array consists of $M = 30$ sensors, each one 250 μm wide, with 50 μm between them. The virtual array consists of $M = 30$ point-like sensors, at 300 μm intervals. We calibrate for equidistantly spaced temporal frequencies between 2.6 and 5.6 MHz using $L = 96$. We use only $P = 1,000$ measurements, and an SNR of 30 dB. Since our proposed algorithm relies on the forward field being known (see the footnote in Sec. 7.2), we assume a point-like source transmits a spherical wave into the imaging medium without being aberrated.

In order to remove stationary signals, such as the transmit wave that propagates directly towards the sensor array without scattering of pixels, we form a matrix of all the measured frames. Similar to SVD-based power Doppler [38], the 10 highest singular values are then set to zero. After this procedure, the filtered measurements $\mathbf{y}^{(p)}$ can be retrieved from the SVD-filtered matrix.

To evaluate the performance, we consider the estimated $\mathbf{H}$ matrix to beamform each frame used in the calibration based on the LSQR algorithm with a large damping coefficient as regularization. We then apply a high-pass filter per pixel across all frames, followed by summing the absolute value of all frames to obtain a single power Doppler
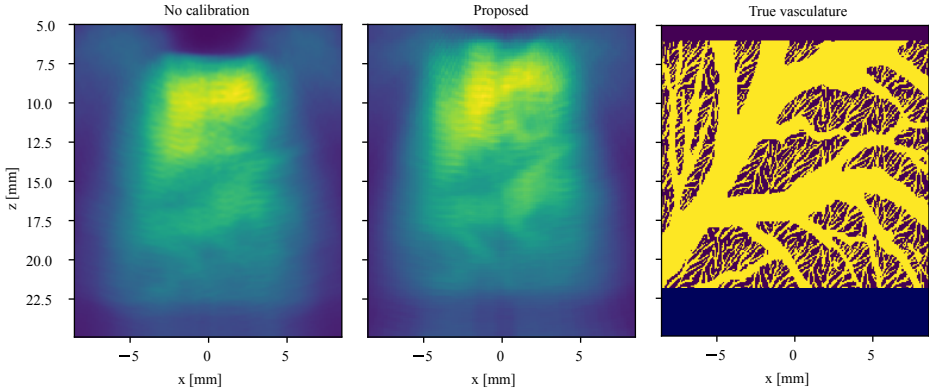
Figure 7.6: Power Doppler images for $P = 1,000$ frames, for the results of Sec. 7.4.4. Each frame is simulated in *k-Wave*, and calibration is done using flow data from the same vasculature as visualized in the rightmost panel. The power Doppler images are normalized to have maximum of 1, and the images have been scaled equally in depth to be able to show image details at larger depths.

image. The result is shown in Fig. 7.6. Since the reconstructed images were suffering from decreasing image intensity for increasing depth and increasing lateral distance from the center (as was the case for the preceding image reconstructions in this section), the compounded power Doppler image has been scaled in those directions for visualization purposes. The power Doppler image using the **H** matrix as estimated by the proposed algorithm shows a better resolved image compared to the case of no calibration. The improvement seems to be especially good in the ROI closer to the array, which can be explained due to the fact that columns of **A** are more uncorrelated in that region, and thus more resolution is lost there if **H** is unknown. As a simple calibration procedure, the speed of sound was somewhat tweaked to compensate for the delays of the aberration layer, but we found that using a speed of sound close to the true speed of sound gives the best result when not using our proposed calibration method. Hence, we expect that the multiple reflections of the sub-layers cause the loss in (axial) resolution. Using our calibration algorithm, however, these are taken into account, leading to the improved image in Fig. 7.6. Importantly, this result suggests that power Doppler imaging (and thus, fUS imaging) through a layer is possible using the vasculature blood flow itself as random images. This is an especially encouraging result for fUS imaging of the brain from cerebral blood flow.

7.5. CONCLUSION & DISCUSSION

We proposed a method for estimating the transfer function of an aberrating layer in ultrasound imaging scenarios by formulating a measurement model in the covariance domain. The ultrasound images do not need to be known, and instead only the second-order statistics of subsequent images are required, a condition that is severely less demanding. The proposed measurement model explicitly exhibits the aberration transfer function, by observering that propagation through an aberrating layer is still a linear

propagation process. By collecting a large number of measurements, the covariance matrix of the raw RF measurements can be estimated, from which the aberration transfer function is estimated. The differences between measurements is caused by the movement of e.g. blood flow or injected micro-bubbles. If enough time is taken between each frame/measurement, the positions of blood particles or micro-bubbles can be modelled as a random process, with no correlation between measurement frames and pixels. We can thus analyse data in the covariance domain where the ultrasound contrast image variable is replaced by the contrast image covariance matrix.

The approach results in a non-convex minimization problem with orthogonality constraints, for which we proposed algorithms to find an estimated transfer function. The proposed method is especially applicable for medical imaging scenarios with sufficient blood flow across the image, such as the human brain. It is minimally invasive since only a small transmitter would need to be inserted, whereas the much larger sensor array can be positioned non-invasively outside the ROI and the aberration layer. Making the method suitable for pulse-echo imaging by transmitting and receiving using the same sensor array seems challenging as one of the main challenges observed in this paper was the fact that the transfer function variables (in the case of non-abberated transmit sensors) are quartic w.r.t. the observed RF data, and pulse-echo imaging would make it bi-quartic (octic), adding another layer of mathematical complexity to the problem.

A second critique on our method is that it does not model non-linear processes such as absorption, which may be more significant in certain aberrating layers than in soft tissue. Consequently, since our model is a purely linear one, it may not be able to capture these non-linear propagation effects sufficiently, and instead tries to capture it in our linear model, leading to incorrect solutions.

Although prior knowledge of the image covariance matrix instead of the actual images themselves severely relaxes the calibration problem, there can still be a model mismatch if there are relatively large areas in the ROI with no varying pixels (e.g. due to a lack of blood flow), or pixels with a relatively different variance compared to other pixels. In that case, using the identity matrix as the image covariance matrix might lead to faulty transfer function estimates. If an acceptable initial image can be obtained precalibration, or using an initial estimate of $\mathbf{H}$, such empty regions can be identified, and omitted from $\mathbf{A}$ to obtain a more accurate model for $\mathbf{C_y}$.

We tested our proposed technique on a range of increasingly challenging calibration scenarios. We used *k-Wave* simulations to generate raw RF measurements to test our algorithm for a varying range of aberration layers. In each case, our technique was able to find a sufficiently good estimate of the transfer function of the aberration layer. Consequently, for each tested layer, image reconstructions were much better using our calibrated model than an uncalibrated model, even enabling imaging through a simulated skull layer.

There are currently no proofs or conditions for the existence of a unique solution to the mathematical calibration problem as formulated in this work. Consequently, we had to find whether this is the case for each scenario empirically. Furthermore, a parameterization of the transfer function matrix is required, which is dependent on the aberration layer characteristics. Nevertheless, we were able to successfully calibrate various layers using general FFT-based parameterizations, for various layers, including a skull layer

with porous bone causing strong wave distortions. For tightly parameterized transfer functions like those for laterally invariant layers, there seems to be a unique solution when calibrating for each single frequency independently, which our algorithm is able to find. For more challenging layers, a joint multi-frequency approach is required. More efficient representations of the layer transfer functions could reduce the number of unknowns and lead to better and/or faster calibration performance, and is a topic of future study.

We finally tested our techniques on a simulated blood flow phantom where the displacement of blood acts as random images. Assuming all images are uncorrelated, we were able to estimate the layer transfer functions, and we subsequently successfully computed a Doppler image of the same measurement data used for calibration. This is an encouraging result for non-invasive functional ultrasound and Doppler imaging through the human skull, where the blood flow itself would act as random contrast images.

REFERENCES

[1] P. van der Meulen, M. Coutiño, J. G. Bosch, P. Kruizinga, and G. Leus, *Ultrasonic imaging through aberrating layers using covariance matching,* Submitted for publication to IEEE Transactions on Computational Imaging (2022).

[2] J. G. Lynn, R. L. Zwemer, A. J. Chick, and A. E. Miller, *A new method for the generation and use of focused ultrasound in experimental biology,* The Journal of general physiology **26**, 179 (1942).

[3] L. M. Hinkelman, T. D. Mast, L. A. Metlay, and R. C. Waag, *The effect of abdominal wall morphology on ultrasonic pulse distortion. part I. measurements,* The Journal of the Acoustical Society of America **104**, 3635 (1998).

[4] J. T. Fokkema and P. M. van den Berg, *Seismic applications of acoustic reciprocity* (Elsevier, 1993).

[5] M. Verweij, B. Treeby, K. Van Dongen, and L. Demi, *Simulation of ultrasound fields,* Comprehensive biomedical physics , 465 (2014).

[6] J. Bercoff, G. Montaldo, T. Loupas, D. Savery, F. Mézière, M. Fink, and M. Tanter, *Ultrafast compound doppler imaging: Providing full blood flow characterization,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **58**, 134 (2011).

[7] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, *Functional ultrasound imaging of the brain,* Nature methods **8**, 662 (2011).

[8] R. Rau, P. Kruizinga, F. Mastik, M. Belau, N. de Jong, J. G. Bosch, W. Scheffer, and G. Maret, *3D functional ultrasound imaging of pigeons,* Neuroimage **183**, 469 (2018).

[9] B.-F. Osmanski, S. Pezet, A. Ricobaraza, Z. Lenkei, and M. Tanter, *Functional ultrasound imaging of intrinsic connectivity in the living rat brain with high spatiotemporal resolution,* Nature communications **5**, 1 (2014).

7

[10] S. Soloukey, A. J. Vincent, D. D. Satoer, F. Mastik, M. Smits, C. M. Dirven, C. Strydis, J. G. Bosch, A. F. van der Steen, C. I. De Zeeuw, *et al.*, *Functional ultrasound (fus) during awake brain surgery: The clinical potential of intra-operative functional and vascular brain mapping,* Frontiers in neuroscience **13**, 1384 (2020).

[11] M. Hirama and T. Sato, *Imaging through an inhomogeneous layer by least-mean-square error fitting,* The Journal of the Acoustical Society of America **75**, 1142 (1984).

[12] M. O'donnell and S. Flax, *Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **35**, 768 (1988).

[13] L. Nock, G. E. Trahey, and S. W. Smith, *Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,* The Journal of the Acoustical Society of America **85**, 1819 (1989).

[14] D. Rachlin, *Direct estimation of aberrating delays in pulse-echo imaging systems,* The Journal of the Acoustical Society of America **88**, 191 (1990).

[15] M. Karaman, A. Atalar, H. Koymen, and M. O'Donnell, *Experimental analysis of a computationally efficient phase aberration correction technique,* in *IEEE 1992 Ultrasonics Symposium Proceedings* (IEEE, 1992) pp. 619–622.

[16] M. Karaman, A. Atalar, H. Koymen, and M. O'Donnell, *A phase aberration correction method for ultrasound imaging,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **40**, 275 (1993).

[17] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu, *Time-reversed acoustics,* Reports on progress in Physics **63**, 1933 (2000).

[18] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, *Optimal focusing by spatio-temporal inverse filter. II. experiments. application to focusing through absorbing and reverberating media,* The Journal of the Acoustical Society of America **110**, 48 (2001).

[19] C. Demené, J. Robin, A. Dizeux, B. Heiles, M. Pernot, M. Tanter, and F. Perren, *Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients,* Nature biomedical engineering **5**, 219 (2021).

[20] F. Vignon, J. Aubry, M. Tanter, A. Margoum, and M. Fink, *Adaptive focusing for transcranial ultrasound imaging using dual arrays,* The Journal of the Acoustical Society of America **120**, 2737 (2006).

[21] L. Guasch, O. C. Agudo, M.-X. Tang, P. Nachev, and M. Warner, *Full-waveform inversion imaging of the human brain,* NPJ digital medicine **3**, 1 (2020).

[22] P. van der Meulen, M. Coutino, P. Kruizinga, J. G. Bosch, and G. Leus, *Blind calibration for arrays with an aberration layer in ultrasound imaging,* in *2020 28th European Signal Processing Conference (EUSIPCO)* (IEEE, 2021) pp. 1269–1273.

7

[23] R. Stoughton and S. Strait, *Source imaging with minimum mean-squared error,* The Journal of the Acoustical Society of America **94**, 827 (1993).

[24] R. Lavarello, F. Kamalabadi, and W. D. O'Brien, *A regularized inverse approach to ultrasonic pulse-echo imaging,* IEEE transactions on medical imaging **25**, 712 (2006).

[25] P. Kruizinga, P. van der Meulen, A. Fedjajevs, F. Mastik, G. Springeling, N. de Jong, J. G. Bosch, and G. Leus, *Compressive 3D ultrasound imaging using a single sensor,* Science advances **3**, e1701423 (2017).

[26] J. Janjic, P. Kruizinga, P. van der Meulen, G. Springeling, F. Mastik, G. Leus, J. G. Bosch, A. F. van der Steen, and G. van Soest, *Structured ultrasound microscopy,* Applied Physics Letters **112**, 251901 (2018).

[27] B. Ottersten, P. Stoica, and R. Roy, *Covariance matching estimation techniques for array signal processing applications,* Digital Signal Processing **8**, 185 (1998).

[28] S.-E. Måsøy, B. Angelsen, and T. Varslot, *Estimation of ultrasound wave aberration with signals from random scatterers,* The Journal of the Acoustical Society of America **115**, 2998 (2004).

[29] T. Varslot, H. Krogstad, E. Mo, and B. A. Angelsen, *Eigenfunction analysis of stochastic backscatter for characterization of acoustic aberration in medical ultrasound imaging,* The Journal of the Acoustical Society of America **115**, 3068 (2004).

[30] A. J. Weiss and B. Friedlander, *Eigenstructure methods for direction finding with sensor gain and phase uncertainties,* Circuits, Systems and Signal Processing **9**, 271 (1990).

[31] S. J. Wijnholds and A.-J. Van Der Veen, *Multisource self-calibration for sensor arrays,* IEEE Transactions on Signal Processing **57**, 3512 (2009).

[32] K. N. Ramamohan, S. P. Chepuri, D. F. Comesana, and G. Leus, *Blind calibration of sparse arrays for doa estimation with analog and one-bit measurements,* in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2019) pp. 4185–4189.

[33] T. Viklands, *Algorithms for the weighted orthogonal procrustes problem and other least squares problems*, Ph.D. thesis, Datavetenskap (2006).

[34] P. Stoica, H. He, and J. Li, *New algorithms for designing unimodular sequences with good correlation properties,* IEEE Transactions on Signal Processing **57**, 1415 (2009).

[35] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, *Manopt, a matlab toolbox for optimization on manifolds,* The Journal of Machine Learning Research **15**, 1455 (2014).

[36] B. E. Treeby and B. T. Cox, *k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields,* Journal of biomedical optics **15**, 021314 (2010).

7

[37] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares,* ACM Transactions on Mathematical Software (TOMS) **8**, 43 (1982).

[38] C. Demené, T. Deffieux, M. Pernot, B.-F. Osmanski, V. Biran, J.-L. Gennisson, L.-A. Sieu, A. Bergel, S. Franqui, J.-M. Correas, *et al.*, *Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases doppler and fultrasound sensitivity,* IEEE transactions on medical imaging **34**, 2271 (2015).

7

8

CONCLUSION AND DISCUSSION

I N this dissertation, we introduced techniques and algorithms for imaging through aberrating layers in the context of medical ultrasound imaging. In the first part, we focused on compressive imaging with a single sensor and an aberrating coding mask. The effect of the coding mask is known *a priori*, and was measured in an experimental calibration setup. We experimentally demonstrated that 3D imaging is possible, and we proposed a method for the design of an optimal coding mask. In the second part, we considered imaging scenarios with an array and an abberating layer where the effect of the layer is *not* known, and propose methods to blindly estimate its transfer function in a minimally-invasive manner. Two techniques and their accompanying algorithms were proposed for finding the transfer function of an aberrating layer using minimally-invasive measurement setups, as well as a simple technique for estimating the electro-mechanical impulse response of each sensor in an array without using hydropohones.

Besides the contributions discussed in the introduction, and each technical chapter, this thesis demonstrated the potential of model-based imaging, the tool underlying each contribution in this thesis. By explicitly defining a linear wave propagation and scattering model, and using simple linear algebraic formulations, we arrived at a model-based imaging framework. It was shown that this framework is easily adapted to explicitly define and include variables for aberrating layers. These formulations then naturally lead to algorithms for imaging through layers, optimizing a compressive coding mask, and for blindly estimating the transfer functions of such layers. As a result, this dissertation offers models and algorithms for complex imaging scenarios, utilizing the large amount of computational power available nowadays, and opens the possibility of compressive ultrasound imaging, and imaging through aberrating layers. More detailed conclusions per chapter are summarized below.

In Chapter 3, it is demonstrated experimentally that 3D imaging is possible using only a single ultrasound transducer. Spatial compression of the ultrasonic wavefields by using a plastic coding mask ensures that a unique temporal pulse-echo signal per pixel in the ROI allows for identifying the direction of an echo signal. The pulse-echo response

of each pixel needs to be measured first, after which a linear scattering measurement model can be defined, which can be inverted to obtain an estimated image using various regularization methods. This chapter shows that imaging using complicated wave interference patterns due to aberrating layers on transmit and/or receive is possible, as long as they can be characterized. In fact, it can be used to our advantage to encode directional information in the temporal dimension instead of inferring it from spatial samples. The work in this chapter provides an avenue for compressive ultrasound imaging, which is relevant for imaging devices where the device is limited by factors as space and money (or other practical limitations such as electronic interconnections etc.).

Chapter 4 considers the imaging device introduced in Chapter 3, and offers optimization algorithms for finding an optimal coding mask shape. To that end, the pillar on each point of the mask is discretized in height, and subsequently incorporated into the linear measurement equation. It is then assumed a Wiener estimator is used for imaging, and the error variance of the image is minimized with respect to the mask shape parameters. This results in a discrete optimization problem. To solve it, a semidefinite relaxation, as well as a greedy discrete optimization algorithm are proposed. Empirical results show that the proposed algorithms come close to the true global optimum. The algorithms try to find masks that keep pulse-echo signals of pixels decorrelated, while focusing energy within the ROI. This manifests itself in the mask geometry since the approximate repetition of features of the mask seems relative to the spatial (angular) bandwidth required to focus on the defined ROI. In addition, the measurement matrix of the optimized masks exhibit a better shaped singular value spectrum than those of random masks, improving the conditioning of these matrices. Numerical experiments show that the masks obtained this way outperform masks with a random profile.

In Part II of the thesis, it is assumed that the measurement matrix is only partially known, and we investigate methods for estimating the unknown components of it. We start with a normal imaging scenario using a linear array, where only the electro-mechanical impulse response of each sensor in the array is unknown in Chapter 5. Knowledge of this impulse response is helpful for more accurate wave simulations and models, especially if the impulse response is different for each array element. We propose a simple calibration procedure, where a flat mirror-like reflector is placed in front of the array, after which the pulse-echo signal of each element is measured. Assuming each element is excited with a delta-like impulse, this means that each element measures the auto-convolved (self-convolved) impulse response of itself. To estimate the one-way impulse response, we then propose a de-autoconvolution algorithm. The proposed procedure is fast, simple, and does not rely on expensive hydrophones. We conclude that the estimated impulse response signals are accurate by comparing convolved combinations of them to measurements with different transmit and receive pairs.

Next, in Chapter 6, we propose a more explicit linear measurement equation for ultrasound imaging scenarios with an aberrating layer, and we introduce a matrix variable representing the transfer function of an aberrating layer. We then propose a method for estimating this layer's transfer function by measuring the echo signals of many images, without explicit knowledge of the ultrasound contrast images themselves. Instead, it is only required that the matrix formed by these images is full-rank. However, the method

has only been demonstrated to work for laterally-invariant layers, where the layer's transfer function matrix is Toeplitz, a structure that is capitalized on by the calibration algorithm. In addition, a second set of measurements is necessary using the same set of images as the first set, but using a phase screen as a second layer. The difference between these two measurements is then exploited by the calibration algorithm. However, obtaining the second set of measurements and using the phase screen may make the method impractical and medically invasive, since the screen needs to be placed between the aberrating layer and the ROI (e.g., against the inside of the skull). Additionally, the method is dependent on good conditioning of several of the matrices involved.

To address the aforementioned issues of the corresponding calibration method, we propose a much more robust and flexible method in Chapter 7. It relies on a single set of many contrast images, where it is assumed that only the second-order statistics of these images is known, and utilizes covariance matching to find the layer's transfer function matrix that best matches the observed measurement covariance data. One could think of a set of random images obtained by taking many measurements of blood or micro-bubble flow, taken at subsequent points in time. The proposed technique is capable of dealing with a variety of layers, even if their transfer function matrix is not Toeplitz. The formulated covariance matching problem leads to a least squares cost-function with manifold constraints, for which we propose to use an alternating projection algorithm. Unfortunately, there is no proof or analysis about the existence of a global optimum. The proposed technique was successfully tested on a variety of layers in the *k-Wave* ultrasound simulator, including a skull-like layer, and a blood flow phantom. After calibration, imaging performance was evaluated on simulated resolution phantoms, and showed that the calibrated model shows a much improved image reconstruction performance compared to a non-calibrated imaging approach. For a successful calibration, a parameterization of the transfer function matrix is required, for which we simply used 2D FFT coefficients. Hence, the method already works without a strong reduction of parameters, and using a very general parameterization. Tighter parameterizations could lead to increasing calibration performance, and is a topic of future research.

8

8.1. FUTURE RESEARCH DIRECTIONS

COMPRESSED ULTRASOUND IMAGING

The discovery of Compressed Sensing has facilitated (medical) imaging using measurements sampled below Nyquist rate for medical imaging modalities, such as MRI [1] and CT imaging [2]. In the field of medical US imaging, however, CS has not (yet) lead to large breakthroughs that address one of the major challenges in medical US imaging: electronic design and wiring challenges for 2D ultrasound arrays.

Although there exists a large body of studies proposing a variety of CS methods for medical US imaging, it appears as though most of these techniques only work for images that are already sparse in the spatial image domain. For example, excellent results are achieved on point scatter phantoms, or on *in-vivo* images that have large empty regions, such as blood vessels in images of carotids. As pointed out by A. Besson *et. al* [3], for dense medical images, results are often subpar, as speckle regions tend to become darker

than in the reference DAS images. Similar results are observed in other works, both for compression in the array domain, as well as the time domain. A comprehensive study was done in [4], where the compressibility of many popular bases on DAS beamformed US images and RF signals are evaluated. It was unable to find a basis in which non-sparse US images can be sparsely represented.

An important open question in the field of compressive sensing for medical US imaging, therefore, is whether there exists any representation basis in which medical US images are truly sparse. A strong case can be made that such a basis does not actually exist, since speckle in medical US images is due to the randomness of sub-wavelength particles in biological tissue. Of course, speckle is not random in a true sense, but for all intents and purposes, they can be modelled as being *i.i.d. normally distributed in terms of position and scattering strength*, with many such sub-wavelength scatterers in a resolution cell. This Gaussian model of speckle explains the universally observed Rayleigh distribution of speckle in medical US images [5], since one often visualizes the absolute value squared of the beamformed RF data. If this is truly the case, then speckle regions in US images cannot be sparsely represented, since random uncorrelated signals cannot be compressed, which would explain the results in CS US studies mentioned above. Note that the same problem would hold for sparse reconstruction techniques for measured RF data (instead of reconstructing the contrast image directly), since the RF data itself is a convolution of these random scatterers and the temporal-spatial impulse response of the imaging system.

Consequently, the randomness of US images should be taken into account in signal models and image reconstruction algorithms. For example, consider the following signal model. We still use the equation $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$, but we add the information that $\mathbf{x}$ is a realization of a white Gaussian i.i.d. process. The *structure* that we see in US images comes from the fact that distinct regions of the same tissue type will have the same randomness (speckle), but with a *different amplitude or variance* compared to other tissue regions. In other words, what distinguishes tissue regions in US images are their spatial second order statistics. Consequently, since the second order statistics of speckle regions should be similar (i.e., piece-wise constant), one could consider *that medical US images are sparse in the spatial (co)variance domain, not directly in the spatial domain itself.* Future research would have to investigate how such models would be best expressed mathematically, and what kind of (Bayesian) estimators would be needed for imaging using such models.

NEW CODING MASK CLASSES AND CODING TECHNIQUES

The coding masks used in Chapter 3 and 4 is a homogeneous mask where the only variability is due to the height of the pillars that constitute the mask. This is only one particular class of coding masks, and one could imagine that more complex mask configurations, e.g. by using a heterogeneous mix of materials, or more complex geometries, could lead to better compression capabilities, SNR, and resolution. Such a new class of coding masks would have to consider *(a)* the capability to decorrelate pulse-echo signals, *(b)* maximize the amount of energy transmitted through the mask, *(c)* minimize the use of mechanical translation of the coding mask for obtaining a diverse set of measurements.

One potentially interesting set of candidates to address the second point are acous-

tic chaotic cavities, which have already been studied in the context of ultrasound imaging [6–9], and more recently for ultrasound imaging with limited sensor arrays as well [10, 11], mostly based on time-reversal techniques [12]. Since these methods use time-reversal, they typically form an image by focusing transmit waves on each individual pixel consecutively, leading to very low frame rates. As such, they are more suited for therapeutic techniques like High Intensity Focused Ultrasound (HIFU), where high resolution and intense focal points are required on transmit in order to ablate malignant tissue. Model-based imaging, in contrast, can create an image from much fewer frames, since it does not have to focus on each pixel during transmit. A concern with acoustic chaotic cavities, however, is that waves are not necessarily fully focused into the region of interest, but also along the sensor plane or even back into sensors prior to illuminating the ROI.

A second approach would not entirely remove sensors in a fully sampled ultrasound array, but would keep them and apply the effect of a coding mask electronically. As a result, the total amount of output channels would still be less than the total amount of sensors in the array, potentially reducing the complexity of wiring and electronics of dense (2D) ultrasound arrays. After all, an array with spatial sampling at Nyquist rate would be able to replicate any field produced by any linear coding mask on transmit, thus allowing us to create the best spatial transmit diversity without the inevitable losses of a coding mask. Furthermore, the effect of the coding mask would be easy to model, since we would have precise control over the electronic mask. Consequently, a calibration procedure to measure the $\mathbf{A}$ matrix would no longer be needed. Finally, this approach entirely removes all mechanical translation, and attains measurement diversity by obtaining multiple measurements with multiple electronic coding masks.

Whether this approach would be feasible is ultimately dependent on the limits of electronic design, and will require an investigation that considers an electronic coding mask class with a small electronic footprint, yet being potent enough for creating spatial wave diversity in the ROI. For example, one could use electronic shifts per channels, followed by spatial summation, emulating the coding principle of the masks used in this thesis. However, there might be more interesting signal operations that can be performed that lead to better results whilst still keeping the complexity of the electronics sufficiently low. Interesting studies exploring spatially diverse wavefields transmitted by phased arrays have been done by Clement *et al.* [13–15], although these studies were not performed in the context of compressive imaging. Random excitation has been investigated for compressive sensing in [16–19], using fewer transmit events, but not addressing the electronic design challenges for 2D arrays. Preliminary research exploring such methods and their electronic design for compressive US imaging as proposed in this section can be found in [20, 21].

IMPROVED CODING MASK DESIGN MODELS AND ALGORITHMS
The coding mask design algorithms proposed in Chapter 4 possess a number of drawbacks that could be addressed by improved mask design algorithms: *(a)* they are computationally intensive, since the $\mathbf{A}$ matrix needs to be calculated for each candidate mask in this iteration; *(b)* since $\mathbf{A}$ is ill-conditioned, a regularization method is needed, resulting in the Wiener estimator used in this thesis. Consequently, a suitable regularization

8

parameter value needs to be chosen prior to mask optimization, even though the regularization parameter might be different when doing the actual imaging with the actually used mask. It has not been checked whether the resulting coding masks perform just as well in imaging scenarios with different regularization parameters, or using common beamformers such as the DAS beamformer. *(c)* The algorithm and the resulting masks are lacking in physical and mathematical insight as to why the resulting mask is a good candidate, besides having a lower cost function relative to random masks.

A first recommendation for addressing most of these points is using the signal models used in Part II of this thesis, $\mathbf{y} = \mathbf{HAx}$, where $\mathbf{H}$ is the transfer function of the coding mask. This gives more insight into the mathematics of the problem, and allows one to optimize directly for $\mathbf{H}$ instead of $\mathbf{A}$. For example, we could consider what the best $\mathbf{H}$ matrix is in relation to $\mathbf{A}$: we know that the row space of $\mathbf{H}$ should lie as much as possible in the column space of $\mathbf{A}$ to preserve information (since $\mathbf{A}$ is often ill-posed, they should be the K largest singular vectors if $\mathbf{H}$ has K rows, instead of the entire column space). After all, if the virtual array of $\mathbf{H}$ is Nyquist sampled, we know that $\mathbf{HA}$ cannot sense more information than the uncompressed $\mathbf{A}$. In other words, a coding mask cannot infer more information about a contrast image than a fully sampled array can.

Consequently, one could aim at finding a good $\mathbf{H}$ matrix by maximizing its singular values (or any other cost function promoting well-posed singular spectra, such as $\mathrm{tr}((\mathbf{HH}^{\mathsf{H}})^{-1})$), while at the same time constraining or promoting the solution to lie in the subspace spanned by the singular vectors of the K largest singular values of $\mathbf{A}$. The subspace constraint ensures that the relevant pixel region is optimized for, and dictates which information in $\mathbf{A}$ to encode. At the same time, the cost function ensures that the coding mask focuses/maximizes energy into the ROI, thereby maximing SNR, while also ensuring that the algorithm spreads information across the entire subspace instead of e.g. just a singular vector, ensuring pulse-echo signal diversity. This gives us more tools for evaluating candidate masks, and would reduce the size of the matrices involved in optimization algorithms since the pixel dimension is no longer directly involved, reducing the computational complexity, and possibly eliminating the need for choosing a regularization parameter.

Additionally, using this signal model formulation gives us tools to analyze US imaging with a coding mask in the context of compressive sensing. The original uncompressed RF field, say $\mathbf{s} := \mathbf{Ax}$, before entering the coding mask, then relates to the coded measurement $\mathbf{y}$ as $\mathbf{y} = \mathbf{Hs}$. One could attempt to reconstruct $\mathbf{s}$ from $\mathbf{y}$ using compressive sensing theory, and design $\mathbf{H}$ in such a way as to minimize coherence with a suitable representation basis for $\mathbf{s}$.

BLIND CALIBRATION OF ABERRATING LAYERS

Blind calibration is an inherently difficult problem due to the bilinear nature of the measurement equation when an unknown aberrating layer is present. Such equations generally do not have unique solutions, and tend to be very challenging to solve. Instead of trying to constrain the solution space of the contrast image $\mathbf{x}$ and the layer transfer function $\mathbf{H}$ (hoping that the constraints would constrict the solution space enough to obtain good enough solutions from local minima), this thesis presented two approaches where $\mathbf{x}$ is eliminated, so that we are only left with the unknown $\mathbf{H}$. The first one consists of

averaging the image out of the measurement equations by analyzing the problem in the covariance domain, and the second of taking different measurements of the *same* image and using the difference between such measurements to estimate **H**.

Both of these approaches require further investigation. For the covariance matching approach, open questions about the mathematical problem that is being solved remain: analysis of the solution space, uniqueness of solutions and existence of global minima all remain unanswered. Since the original covariance matching cost function is hard to solve, future research could also search for new cost functions that should lead to similar solutions as the covariance matching solution, but may be more easy to solve. Large improvements in the performance of this method could be obtained by strongly constraining the structure of the transfer function matrices. In this thesis the solution space of **H** was only constrained by using a Fourier basis where the higher frequencies were removed. This way, the total number of unknowns was reduced from 400 to approximately 350, a very small reduction of unknowns. It is therefore to be expected that much better performance can be obtained by finding a basis that greatly reduces the number of unknowns. One could find such bases by taking the time to study the structures present in **H**-matrices obtained from simulations. A simple approach would consist of measuring the transfer function matrix **H** of a large number of different skull layers, and using PCA or Machine Learning methods to find a suitable basis for **H**.

The work presented using differential measurements in Chapter 6 suffers from the introduction of an invasive phase-screen. One could consider using sub-arrays instead, as is typically done for blind beamforming techniques in communications, such as MUSIC and ESPRIT, so that the phase shifts of the phase screen are instead caused by the lateral displacement of the sub-array (typically achieved by taking the first and last $M-1$ sensors). This approach would have to account for the fact that one cannot consider all reflections to originate in the far-field (as is typically assumed in the field of telecommunications), and would also have to consider what sort of images (that can realistically be obtained in a medical imaging context) would be required to obtain good estimates.

Of course, alternative blind calibration methods could emerge as well. One could, for example, search for a solution of the transfer function such that the reconstructed inverse image optimizes image quality. In that case, one would have to rely on images that contain e.g. isolated scatterers, such as contrast agents or micro-bubbles, and optimize for e.g. the sharpness of those points in the image, or contrast ratios of pixels belonging to different types of tissue.

REFERENCES

[1] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging,* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58**, 1182 (2007).

[2] G.-H. Chen, J. Tang, and S. Leng, *Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets,* Medical physics **35**, 660 (2008).

[3] A. Besson, D. Perdios, F. Martinez, Z. Chen, R. E. Carrillo, M. Arditi, Y. Wiaux, and J.-P. Thiran, *Ultrafast ultrasound imaging as an inverse problem: Matrix-free*

sparse image reconstruction, IEEE transactions on ultrasonics, ferroelectrics, and frequency control **65**, 339 (2017).

[4] Z. Liu, *An Investigation of the Medical Ultrasound Image Sparse Spaces used for the Model-Based Imaging,* Master's thesis (2019).

[5] J. Janjic, *Looking Forward with Minimally Invasive Ultrasound,* Ph.D. thesis (2018).

[6] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, *Building three-dimensional images using a time-reversal chaotic cavity,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **52**, 1489 (2005).

[7] C. Draeger, J.-C. Aime, and M. Fink, *One-channel time-reversal in chaotic cavities: Experimental results,* The Journal of the Acoustical Society of America **105**, 618 (1999).

[8] N. Etaix, M. Fink, and R. K. Ing, *Acoustic imaging device with one transducer,* The Journal of the Acoustical Society of America **131**, EL395 (2012).

[9] T. Sillanpää, K. Longi, J. Mäkinen, T. Rauhala, A. Klami, A. Salmi, and E. Hæggström, *Localizing a target inside an enclosed cylinder with a single chaotic cavity transducer augmented with supervised machine learning,* AIP Advances **11**, 115104 (2021).

[10] B. Cox and P. Beard, *Photoacoustic tomography with a single detector in a reverberant cavity,* The Journal of the Acoustical Society of America **125**, 1426 (2009).

[11] Y. Choi, H. Lee, H. Hong, and W.-S. Ohm, *Two-dimensional virtual array for ultrasonic nondestructive evaluation using a time-reversal chaotic cavity,* The Journal of the Acoustical Society of America **130**, 2720 (2011).

[12] M. Fink, *Time reversal of ultrasonic fields. I. basic principles,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **39**, 555 (1992).

[13] P. J. White and G. T. Clement, *Two-dimensional localization with a single diffuse ultrasound field excitation,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **54**, 2309 (2007).

[14] G. T. Clement, *Two-dimensional ultrasound detection with unfocused frequency-randomized signals,* The Journal of the Acoustical Society of America **121**, 636 (2007).

[15] F. C. Meral, M. A. Jafferji, P. J. White, and G. T. Clement, *Two-dimensional image reconstruction with spectrally-randomized ultrasound signals,* IEEE transactions on ultrasonics, ferroelectrics, and frequency control **60**, 2501 (2013).

[16] M. F. Schiffner and G. Schmitz, *Fast pulse-echo ultrasound imaging employing compressive sensing,* in *2011 IEEE International Ultrasonics Symposium* (IEEE, 2011) pp. 688–691.

[17] M. F. Schiffner, *Random incident waves for fast compressed pulse-echo ultrasound imaging,* arXiv preprint arXiv:1801.00205 (2017).

[18] R. van Sloun, A. Pandharipande, M. Mischi, and L. Demi, *Compressed sensing for ultrasound computed tomography,* IEEE Transactions on Biomedical Engineering **62**, 1660 (2015).

[19] J. Liu, Q. He, and J. Luo, *A compressed sensing strategy for synthetic transmit aperture ultrasound imaging,* IEEE transactions on medical imaging **36**, 878 (2016).

[20] X. Li, *Semi-Controllable Compression Schemes for Ultrasound Imaging,* Master's thesis (2017).

[21] F. Mirzaei, *An Ultrasound Receiver ASIC Employing Compressive Sensing,* Master's thesis (2018).

8

ACKNOWLEDGEMENTS

Here follows a non-exhaustive list of people I would like to sincerely thank for their direct or indirect contribution to this dissertation.

Geert en Pieter, alles begint en eindigt natuurlijk met jullie. Bedankt voor jullie begeleiding, ideeën, geduld, en de vrijheid die jullie mij hebben gegeven tijdens mijn onderzoek. Dankzij jullie heb ik geleerd dat het beste ultrageluid onderzoek ontstaat wanneer je zowel de hardcore signaalverwerking als de akoestiek (misschien iets minder hardcore) beheerst, en ik hoop dat ik mijzelf nu echt een multidisciplinair ingenieur/onderzoeker mag noemen (ik vind in ieder geval van wel). Ik heb er tot de dag van vandaag nog steeds profijt van, en daarvoor dank ik jullie.

Thanks to all my co-authors and collaborators. Andrejs, you are the true OG of the single-sensor paper, to which much of this dissertation is owed. Thanks for letting me take some of the credit, and for being a great office mate during our MSc theses. Jovana and Gijs, it was a pleasure working with you on the ultrasound microscopy paper. Michael, without you there would never have been the Crazy Summer or watching the DotA championships in the lab. Thanks for the great hot sauces (not). Hans, heel erg bedankt voor al je betrokkenheid en manuscript reviews (die ik je vaak last-minute toestuurde) gedurende mijn promotie.

To all my colleagues at CAS, thank you for making me feel at home in Delft, and sharing the pain of doing a PhD. Extra special shout-out to my bois Mario, Tarik, Jamal, and Krishna a.k.a. KPN: you guys made the journey worth-wile, and I'm truly happy to have met you during these times. Thanks for all your support, even during the parts when I was down, as well as for all the laughs, gossip, and for curing my imposter-syndrome.

To all my BME colleagues, thank you for making the weird TUD guy feel welcome in your group, and allowing me to learn the secrets of ultrasound imaging from outside the Delft bubble. Frits en Robert, mijn de-facto kamergenoot op de 19e, en mijn mede-Rotterdam-Zuiderling, respectievelijk: jullie waren een frisse wind temidden van alle academici. Frits, bedankt voor alles wat ik van je heb geleerd over de wetenschap en techniek buiten mijn eigen vakgebied, van fractals tot de details van printers en hoe een EPS is gerelateerd aan een PDF. Dankzij jou besteed ik veel te veel aandacht aan figuren voor mijn publicaties, en ik heb nog regelmatig kritiek op papers waar de verkeerde color maps bij de verkeerde data wordt gebruikt (hoewel ik eerlijk moet toegeven dat ik toch stiekem de jet color map gebruikt als resultaten daardoor overtuigender zijn). Reza, you deserve a special mention. Our breaks outside allowed me to catch my breath and gain a new perspective on my (imagined) problems. Thanks for being a good friend. It was an honor being your paranimph.

Aan alle Alten-boys: ik heb een geweldige tijd met jullie bij ASML gehad. Ik heb nog nooit zo goed kunnen epibreren als met jullie, en heb heel veel gelachen en geleerd. Thx boyz!

Aan mijn nieuwe collega's, (in het bijzonder Arnoud, Vsevolod, Lotte en Aladdin Kaan): heel erg bedankt voor het warme welkom bij Fugro. Jullie zijn allemaal geweldige collega's en heb het enorm naar mijn zin met jullie!

Een speciaal bedankje aan al mijn speciale vrienden die nog niet genoemd zijn. André, Dennis, Emily, Gabriël, Kier, Killian, Leon, Michelle, Rinske. Bedankt dat jullie er zijn geweest door alle ups en downs over de laatste jaren. Ik zeg het misschien niet vaak genoeg, maar jullie zijn echt allemaal tantoe lauw.

Als laatste en belangrijkste: mijn lieve familie. Jullie waren mijn steun en toeverlaat tijdens deze reis. Ik kan altijd op jullie steun rekenen, en daarvoor ben ik dankbaar.

8

Curriculum Vitæ

Pim Quinten van der Meulen was born in the Southern part of Rotterdam in The Netherlands, on January 24, 1991. There, he attended the Erasmiaans Gymnasium as his pre-university education, which he finished in 2009.

Since he never did any homework in high school, he waived his Bachelor's education in Cognitive Artificial Intelligence in 2010 at Utrecht University, after which he started a Bachelor of Engineering in Electrical Engineering at Rotterdam University of Applied Sciences, which he finished in 2014. Inspired to learn more due to his internship at Thales Naval on radar signal processing, he completed a Master's in Signals and Systems in 2016 at Delft University of Technology (cum laude), where he did his thesis research with Geert Leus and Pieter Kruizinga, a collaboration with the Biomedical Engineering group at Erasmus Medical Center. With those same supervisors, he started a PhD in 2016 on biomedical ultrasound imaging, combining signal processing, numerical optimization, and physics, applied in the ultrasound imaging domain. Since then, he has worked for ASML, and since October 2022, he is an imaging and signal processing engineer at Fugro Innovation in The Netherlands.

In this spare time he enjoys making ceramics, playing Dungeons and Dragons, and has varying interests he reads about, ranging from crypto currencies to quantum mechanics, and the occasional novel so as to appear more sophisticated.

LIST OF PUBLICATIONS

[1] Pieter Kruizinga, Pim van der Meulen, Andrejs Fedjajevs, Frits Mastik, Geert Springeling, Nico de Jong, Johannes G Bosch, and Geert Leus. Compressive 3D ultrasound imaging using a single sensor. *Science advances*, 3(12):e1701423, 2017.

[2] Pim van der Meulen, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Impulse response estimation method for ultrasound arrays. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2017.

[3] Pim van der Meulen, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Spatial compression in ultrasound imaging. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 1016–1020. IEEE, 2017.

[4] Jovana Janjic, Pieter Kruizinga, Pim van der Meulen, Geert Springeling, Frits Mastik, Geert Leus, Johan G Bosch, Antonius FW van der Steen, and Gijs van Soest. Structured ultrasound microscopy. *Applied Physics Letters*, 112(25):251901, 2018.

[5] Pim van der Meulen, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Calibration techniques for single-sensor ultrasound imaging with a coding mask. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 1641–1645. IEEE, 2018.

[6] Pim van der Meulen, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Joint optimization of coding mask and scan positions for compressive single sensor imaging. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2018.

[7] Pim van der Meulen, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Coding mask design for single sensor ultrasound imaging. *IEEE Transactions on Computational Imaging*, 6:358–373, 2019.

[8] Pim van der Meulen, Mario Coutiño, Pieter Kruizinga, Johannes G Bosch, and Geert Leus. Blind calibration for arrays with an aberration layer in ultrasound imaging. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1269–1273. IEEE, 2021.

[9] Reza Pakdaman Zangabad, Sophinese Iskander-Rizk, Pim van der Meulen, Bram Meijlink, Klazina Kooiman, Tianshi Wang, Antonius FW van der Steen, and Gijs van Soest. Photoacoustic flow velocity imaging based on complex field decorrelation. *Photoacoustics*, 22:100256, 2021.

[10] Pim van der Meulen, Mario Coutiño, Johannes G. Bosch, Pieter Kruizinga, and Geert Leus. Ultrasonic imaging through aberrating layers using covariance matching. *Submitted for publication to IEEE Transactions on Computational Imaging*, 2022.