Spatiotemporal Filtering in Super Resolution Ultrasound

Jun Ikeda

ABSTRACT. Ultrasound imaging is a medical imaging technique leveraging acoustic waves to conduct internal imaging of the body. Furthermore, as it does not require tracers or ionizing radiation to conduct imaging, there is much to be desired of with respect to its application in clinic. Leveraging these benefits while being scalable to 4D imaging (3D in space, 1D in time), there is a lot of potential for ultrasound. Exploration of Super-Resolution Ultrasound (SR-US) has been particularly a hot topic among ultrasound circles, leveraging microbubble contrast agents to image vasculature at high spatiotemporal resolution. In the filtering of SR-US images, the Singular Value Decomposition is used for the formation of a bandpass filter to isolate vasculature signal from tissue signal. However, forming this filter appears to be through trial and error, and computing the SVD is computationally expensive. In response to these questions, more rigorous formation of the SVD bandpass filter is explored, as well as computationally efficient ways to compute the SVD to further unlock ultrasound's potential for clinical application.

INTRODUCTION. In ultrasound imaging, acoustic waves are emitted by an ultrasound transducer's piezoelectric elements in contact with the surface of the skin as the transducer probes the area of interest to be imaged. Through the lifetime of these acoustic waves, they propagate through the body and interact with different tissues in the body. These different tissues exhibit different acoustic impedance characteristics. It is at these boundaries of different tissues where a large difference in acoustic impedances causes for the acoustic wave is largely reflected and delivers information back to the ultrasound transducer. The direction, timing, amplitude and phase are all information interpreted by the ultrasound transducer. Common wave phenomena such as reflection, refraction, diffraction, scattering occurs and provides information of the wave's interactions with structures inside the body.

Ultrasound is unlike other imaging modalities, exhibiting complementing spatial and temporal resolution that is otherwise compromised or traded off in other imaging modalities such as MRI, CT, and PET. Furthermore, it is nonionizing and has no adverse effects on the individuals it images. But through the leveraging of acoustic wave phenomena, the intrinsic properties of sound waves cause there to be a tradeoff between frequency and penetration. While resolution of the ultrasound image is proportional to frequency of the wave leveraged, the penetration of the wave is inversely proportional to the wave frequency. Navigating this tradeoff is difficult as high-resolution ultrasound images cannot be produced as you want to go deeper into the body. Working around this frequency-resolution tradeoff is done through the methodology introduced in Super-Resolution Ultrasound (SRUS).

Super resolution concerns a research field that attempts to overcome the diffraction barrier in imaging. The diffraction barrier is defined as about half-wavelength, where structures closer than the half-wavelength cannot be differentiated by an imaging system. However, physical, and computational techniques can be leveraged to reproduce a super-resolved image.

Fluorescence Photoactivation Localization Microscopy (FPALM) is the predecessor to SRUS and largely influenced the methodology and acquisition protocols used in SRUS. FPALM is a widefield imaging technique that acquires images based on the entire sample being exposed to the light source. This is on the contrary to point-scanning imaging techniques such as confocal microscopy where out of focus light is rejected and the specimen is scanned to reconstruct a super-resolved image.

In FPALM, the sample under scrutiny is stained with fluorophores. Then, a small number of these fluorophores are stochastically activated, and an image is captured. This protocol is then repeated where the fluorophores in the "on" state revert to the "off" state indefinitely while other fluorophores would turn "on." Only a few isolated fluorophores are in the "on" state at every acquisition. Maintaining a low density of activated fluorophores in a high contrast to the backdrop allows for no overlap between fluorophore signals. These fluorophore signals are then localized as point sources [7].

The Point Spread Function (PSF) is a powerful tool used to reproduce super-resolved image. By knowing the response of your imaging system to a localized point source—analogous to an impulse response—deconvolution of your image can be done with the PSF to localize the point sources. In FPALM, the fluorophores were point sources to be deconvolved.

A convolution operation takes two mathematical functions and produces a new operation conserving property of both original functions.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

Convolution in the imaging domain is conducted with a smaller filter or kernel that is scanned across the image.

A deconvolution operation is the inverse of the convolution operation. It disentangles two mathematical functions.

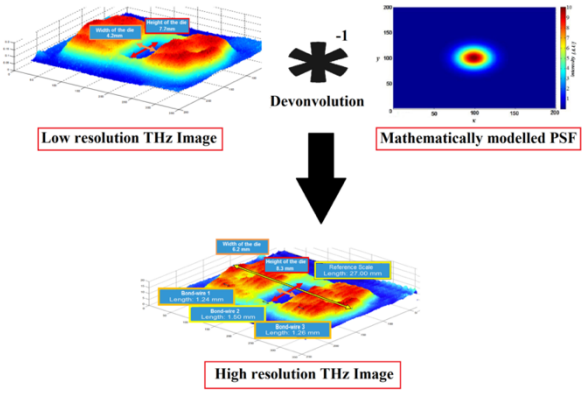

Fig. 1. A mathematically estimated PSF used to deconvolve a low-resolution image to compute a highly resolved image.

After localization of these point sources, a superposition of the many frames containing the activated fluorophores can be done to reconstruct a super-resolved image.

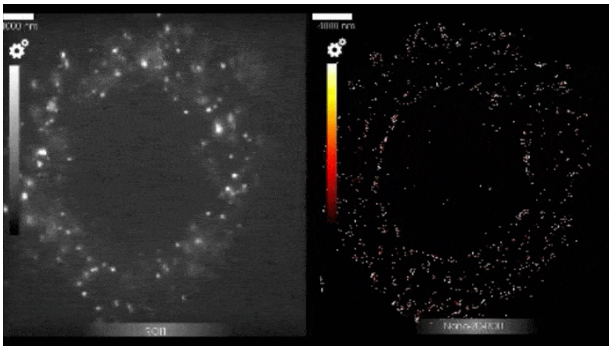

Fig. 2. direct Stochastic Optical Reconstruction Microscopy (dSTORM), where localized fluorophores are superposed to construct a highly resolved image.

SRUS leverages similar acquisition and methodology to conduct super resolution imaging. In super resolution ultrasound, microbubble contrast agents are introduced into the blood stream. These microbubbles are synonymous to the fluorophores in FPALM. These microbubbles oscillate subject to the acoustic pressure wave, backscattering higher-

harmonic signal back to the ultrasound transducer. Due to the nonlinear nature of the backscatter signal, the microbubble can be differentiated from the rest of the ultrasound signal. Superposition of the higher-harmonic signal allows for the imaging of vasculature at high spatial and high temporal resolution without ionizing radiation [6].

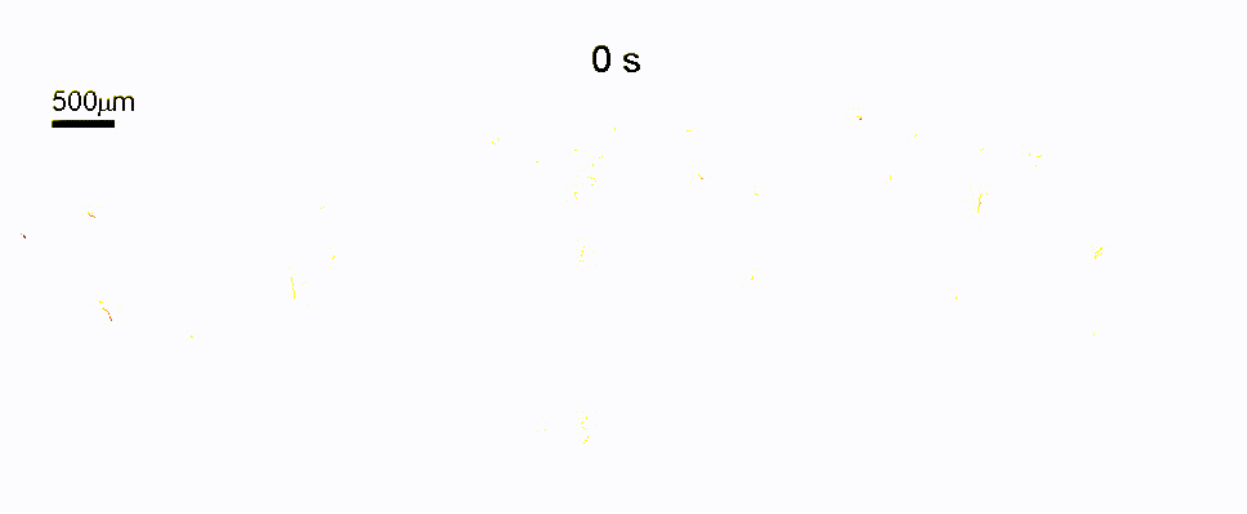

Fig. 3. Superposition of frames containing localized microbubbles allows for a super-resolved reconstruction of rate brain vasculature. However, skull-thinning was employed to conduct imaging of the brain.

In conducting SRUS, the steps the pipeline consist of acquisition, detection, isolation, localization, tracking and mapping [6]. First, SRUS frames are *acquired* where microbubbles are injected into the vasculature. The microbubbles are then *detected* in those frames using filtering methods. Overlapping bubble signals are then discarded, *isolating* localized microbubble signal in each frame that can be later used in the superposition process. *Tracking* of these microbubbles are then done after the localization process to form vectors based off the microbubbles' trajectory. Where these vectors are then *mapped* to image the vasculature. Essentially, the workhorse of SRUS is to reduce information stored in ultrasound videos into linear lines.

Some differences between SRUS and FPALM are that the SRUS frames are sampled at select frequencies to identify the same microbubbles in consecutive frames. These sampling frequencies are estimated based on the microbubble oscillation frequencies when they are subject to the ultrasound wave and can be calculated using microbubble models. This ensures that the sampled SRUS frames are capturing the same bubble and the integrity of the tracking algorithm can be maintained. Furthermore, in contrast to the fluorophores, the microbubbles do not have an "off" state. A low density of microbubbles is maintained so that microbubble signals are not overlapping, and localization can be done. There is an isolation step in SRUS not present in FPALM. Therefore, some bubble signal must be thrown out and implementation of SRUS must be done where the microbubbles are administered in a controlled manner so that the blood stream is not oversaturated.

One interesting challenge that arises in SRUS but not in FPALM, is the isolation of microbubble signal and rejection of tissue signal to image the vasculature at high spatiotemporal resolution. To accomplish this, the detection stage of the SRUS pipeline must be robust, employing many signal processing and filtering

techniques to isolate the microbubble signal. To these ends, spatiotemporal filtering techniques applied in SRUS is explored.

Spatiotemporal Filtering Techniques in SRUS. The most elementary method for removing tissue signal is by leveraging a *high-pass filter*. As the reflected acoustic signal due to microbubble interaction exhibit higher harmonics, a high-pass filter can be used to isolate it. Choosing this frequency threshold can be done by a priori knowledge with regards to the frequency range of the backscattered signal possible due to size and composition of the microbubbles. However, this is not a scalable technique at varying imaging depths and varying wave frequencies.

Singular Value Decomposition (SVD) is a powerful tool leveraged in SRUS filtering to isolate microbubble signal and reject tissue signal. The power of SVD lies in its ability to represent spatiotemporal information in a decomposition of three matrices. This decomposition allows for spatiotemporal information to establish the basis for filtering and motion is the metric by which filtering is conducted.

Machine Learning approaches have also been used in SRUS to filter and detect microbubbles. However, SVD is explored more thoroughly in this paper for the sake of robustness and potential of clinical application.

Eigen Decomposition. The SVD is a generalization of the eigen decomposition, where a square matrix can be factorized and be represented in terms of its eigenvectors and eigenvalues. The eigen decomposition is where the eigenvector v behaves as a scalar and merely scales the matrix when multiplied with matrix A. And the eigenvalue lambda corresponds to the scalar which has equivalent effect on matrix A as the eigenvector. These eigenvectors can be organized into a matrix Q and eigenvalues into a diagonal matrix Lambda.

$$Av = \lambda v$$

$$AQ = Q\Lambda$$

$$A = Q\Lambda Q^{-1}$$

Singular Value Decomposition. The SVD is more powerful than the eigen decomposition as it can be applied to any matrix shape. After decomposition, matrix U's columns contains the orthonormal basis for the column space of M. While V transposed contains the orthonormal basis for the row space of M. Lastly, Sigma's diagonal contains the singular values.

$$M = U\Sigma V^T$$

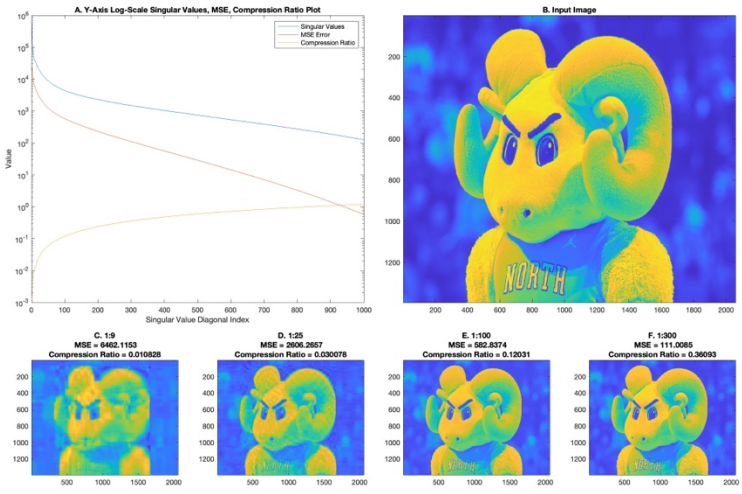

Fig. 4. Application of SVD to a 2D image of Ramses. SVD in this scenario is leveraged to compress an image. Plotted are the singular values that decrease as you go down the diagonal's indices. As more singular values are used to reproduce the original matrix from the decomposition, more information is retained. Earlier singular values contain more information than later singular values.

Spatiotemporal Filtering in SRUS by SVD: To conduct spatiotemporal filtering of a stack of SRUS frames, the 3D tensor must be rearranged into a 2D array. By doing so, the orthonormal matrices of U and V hold physical relevance to the SRUS images as they span the column and row spaces of matrix M respectively. The frames are rearranged such that the Casorati matrix contains the information of a same pixel over time in each row and the information at each time point is in each column [5].

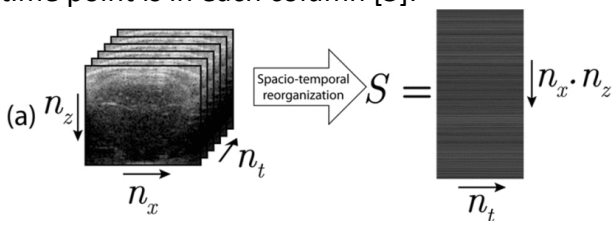

Fig. 5. Visual depiction of how 3D information is reorganized to construct the Casorati matrix. While this is done for the 3D case (2D in space, 1D in time) the method could be expanded upon to 4D data (3D in space, 1D in time) however this means that your matrix size increase by a factor of n and longer acquisition times also cause your matrix to increase as well.

After reorganizing the stack of SRUS images in this way, SVD decomposition can be then done on the Casorati matrix to conduct spatiotemporal filtering across all frames. The U matrix spanning column space of the Casorati Matrix has spatial relevance, while the V matrix spanning row space has temporal relevance. Furthermore, the singular values correspond to spatiotemporal coherence which can be used to construct a bandpass filter [5].

Tissue signal should have high spatiotemporal coherence as it is relatively unchanging throughout the video, described by the singular values towards the beginning. On the other hand, blood signal should have low spatiotemporal coherence as it is varying over time, relating to the singular values towards the end.

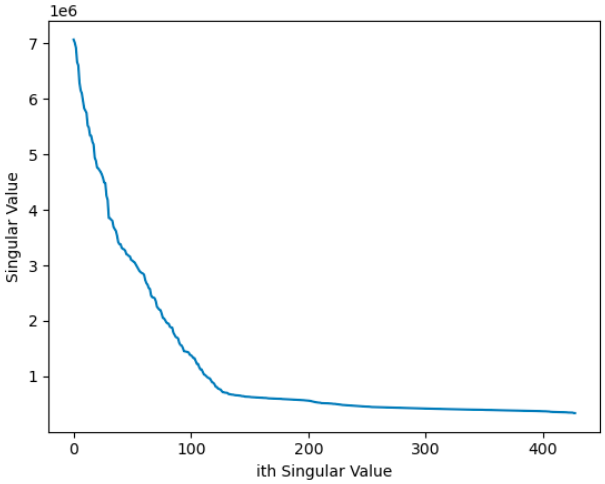

Fig. 6. Plot describing the singular values with respect its index for a 3D ultrasound video.

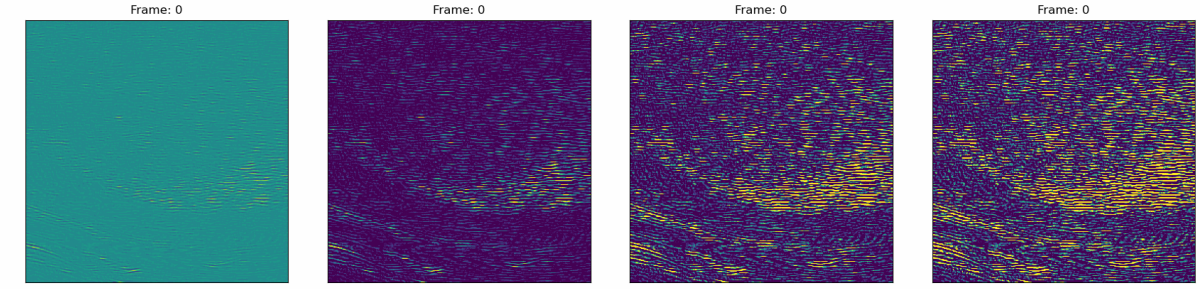

Fig. 7. An example SVD in ultrasound. From left to right are depicted the original video, reconstruction done by first 5 singular values, by 100th to 150th singular values, and the 200th to 428th singular values. The dimensions of the frames were (257, 257) with 428 samples in time. Therefore, the dimensions of the Cosorati Matrix were (66049,428). SVD reconstruction was done by economic SVD, therefore dimensions of U, S's diagonal, V were (66048, 428), 428, and (428, 428) respectively.

The Hilbert Transform is used in to further process the image. The Hilbert Transform computes an analytic signal from a real signal—extending the real signal to the complex plane. These companion imaginary components can be used to better filter the data.

After computing these complex values of the signal, an envelope can be formed by finding the magnitude of the complex signal. This smoothes the data. The application of the Hilbert Transform is intuitive on 1D data. While in the 2D case, the transform is applied column wise to the matrix of interest. This is a powerful tool to filter ultrasound data but requires a good amount of computational power to support it.

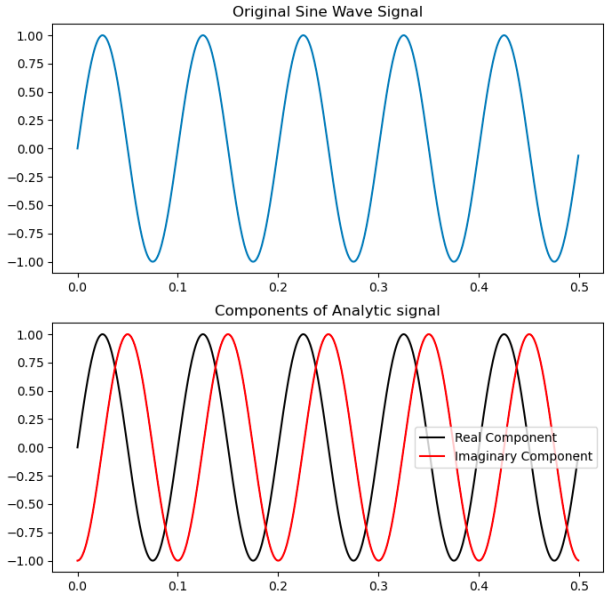

Fig. 8. An example of applying the Hilbert Transform on a sine wave signal. It converts the original real-valued signal into the complex plane. Essentially shifting the original signal by a phase of 90°. By leveraging these two components, an envelope can be formed by computing the magnitude of the signal using the real and imaginary components. It accentuates signal magnitude, eradicating any deconstructive interference of the signal due to phase differences.

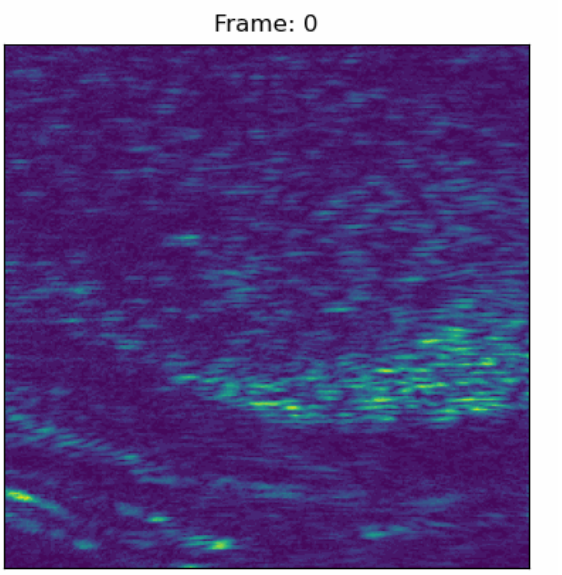

Fig. 9. Application of the Hilbert Transform to unfiltered US data. It allows for greater between structures of varying acoustic impedances in the image.

Exploring spatiotemporal filtering in SRUS, earlier literature lacked robustness on the construction of their spatiotemporal filter to allow for good signal separation. Furthermore, computational costs of SRUS are a bottleneck for its potential in clinic [5,6].

THE QUESTIONS TO ANSWER. Is there a more rigorous way to differentiate between tissue and microbubble singular values? Computing SVD is expensive, how can it be sped up for clinical purposes?

HYPOTHESIS. By leveraging spatial similarity and accelerated SVD algorithms, SR-US pipelines can be formed to produce high resolution ultrasound images efficiently

METHODS. *Algorithms* can be leveraged to conduct adaptive filtering and increase computational efficiency of the SVD. To accomplish the former goal of more robust spatiotemporal filtering in SRUS, spatial similarity can be used to inform threshold selection.

A spatial similarity matrix can be formed by taking matrix U which contains the spatial eigenvectors. The magnitude of every entry in matrix U is computed and normalized with respect to the column means and column standard deviations (columns being synonymous to spatial eigenvectors). Then the correlation of this new matrix can be taken by multiplying itself by its transpose and scaling it by dividing by the product of the spatial dimensions [3].

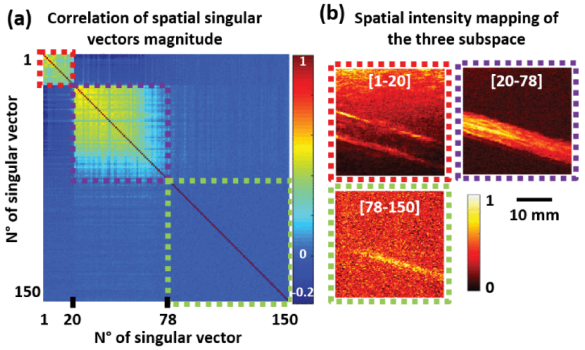

Fig. 11. An ideal spatial similarity matrix and the square partitioning that can be leveraged to robustly segment tissue, microbubble, and noise signal.

For the second challenge of efficient computation, the SVD is computationally expensive as the original matrix M must be used so that the eigenvectors of MM^T and eigenvectors of M^TM are computed. The singular values correspond to the square root of the eigenvalues.

To improve the computational efficiency of the SVD, the dimensionality of the Casorati matrix can be reduced so that fewer targets require computing. The dimension reduction could be done so that dimensions of the matrix concerning noise and/or tissue are eliminated. This methodology is called power SVD (pSVD) [2].

The easier index to determine would be n_{noise} which we could use to shrink our matrix before calculating the SVD.

First, we can assume that the singular values of noise will have similar amplitudes. The noise subspace can be identified by using a squared ratio of singular values [2].

$$n_{noise} = i\ s.t. \frac{s_i^2}{s_{i-1}^2} > 0.99$$

The temporal eigenvectors can be found by computing the covariance matrix of C, the Cosorati Matrix and finding its eigen decomposition. The square root of the eigenvalues are the singular values. Calculating the squared ratio of consecutive singular values can be done to determine the singular value index of the noise subspace. Lastly, a spatial similarity matrix can be computed using the matrices produced by the eigen decomposition to determine n_{tissue}, the singular value index corresponding to the tissue subspace.

By recycling computed matrices and binning the noise lower bound and tissue upper bounds for the singular value indices early, SVD computation can be done efficiently. And by using the spatial similarity matrix, robust SRUS filtering can also be accomplished in parallel.

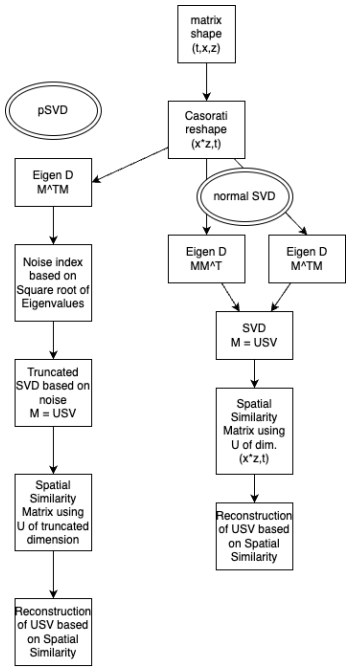

Fig. 12. Pipeline of pSVD compared with normal SVD. pSVD would likely work most efficiently when there are more spatial samples than temporal samples.

To experimentally test this pipeline for SRUS imaging, the algorithm should first be applied to a flow

phantom experiment. This is because variable flow rate and variable motor rate can be explored, observing the pipeline's performance for different blood flow velocities and scanning speeds. Another benefit of these phantom experiments is the presence of a ground truth that can be leveraged to assess the performance of pSVD.

The next step after these experiments would be in-vivo application of this pipeline. The two experiments that would be conducted would be imaging of the rat and neonate brain as SRUS has largely explored the use of ultrasound on the brain. This is because the skull exhibits huge impedance with the skin and attenuates a lot of the ultrasound signal before it can reach image the brain.

RESULTS. While I was unable to make a robust algorithm to partition squares of the spatial similarity matrix to identify tissue, microbubble, and noise signal, I was able to reproduce the spatial similarity matrix for my 3D ultrasound data. The boundaries between tissue, microbubble and noise signal are somewhat apparent.

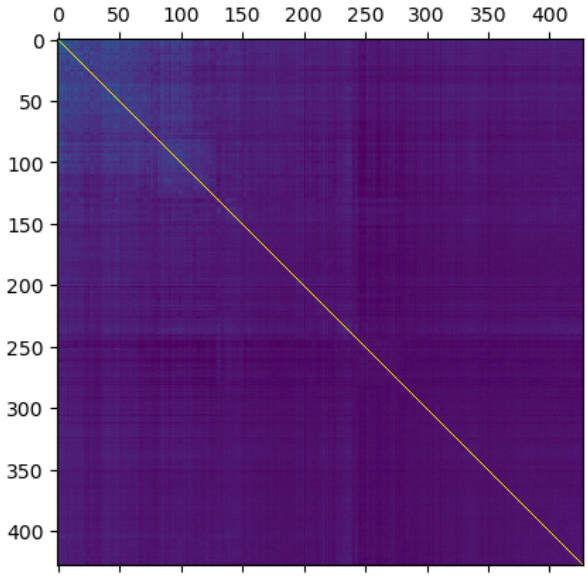

Fig. 10. Spatial similarity matrix of the 3D ultrasound data. The 3D data was likely not to be as highly resolved so the square partitioning was not as definitive.

Without a square fitting algorithm for the spatial similarity matrix, I approximated about 0-125 to be the index for tissue noise and 250 to 428 for the index of noise. Therefore, the microbubble signal should be described by the singular values between 125 and 250.

After successful computation of the spatial similarity matrix, the pSVD was applied and the squared singular value ratios were computed. These ratios were then smoothed using a 3-kernel moving average across the singular values and were normalized. From this, a noise index was approximated.

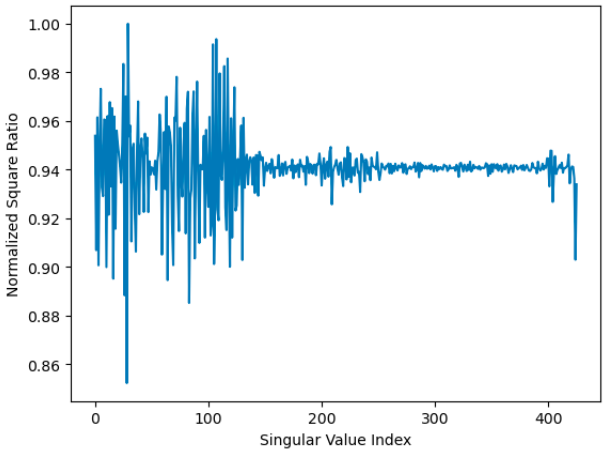

Fig. 11. The squared ratio of the singular values after a 3-kernel smoothing step and normalization. There were some unexpected oscillations in the data. Through this, it was estimated that the noise index was index at either 29, 104 and 107.

After finding the singular value index of noise, the truncated spatial similarity matrix was computed. From the truncated spatial similarity matrix without the square fitting algorithm, the tissue signal was approximated to be at about index 50. Therefore, the microbubble signals were estimated to be contained in the singular values from 50 to 104.

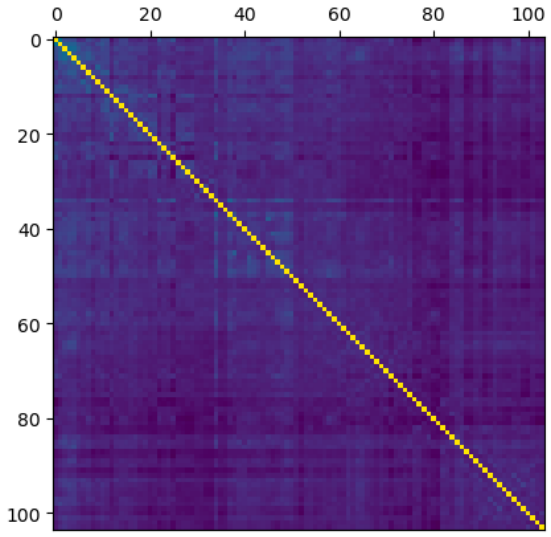

Fig. 12. The spatial similarity matrix after reducing dimensionality of the Cosorati matrix by truncating the matrix to the calculated noise index.

By only using the eigen decomposition which is half of the computation required to calculate the SVD and by incorporating spatial similarity, faster, robust methodology for filtering was achieved. pSVD allowed for a more computationally effective way to reproduce SRUS images.

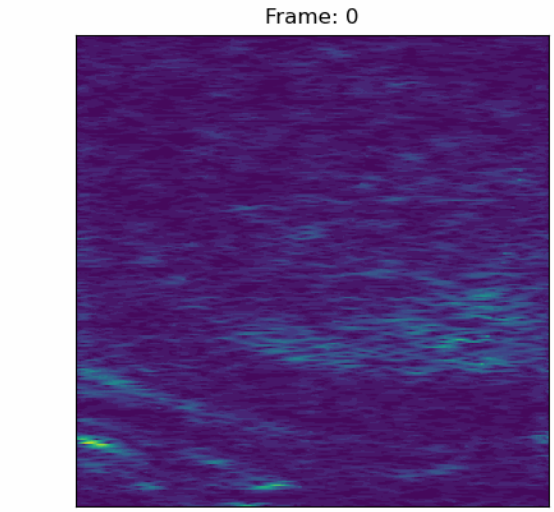

Fig. 13. Reconstruction of the SRUS frames using the 50th to 104th singular values.

For the experimental set-ups, the phantom experiment should be able to reproduce a velocity map of the flow of fluid inside the tubes. It would be beneficial to see the performance of pSVD at varying flow rates and probing speeds.

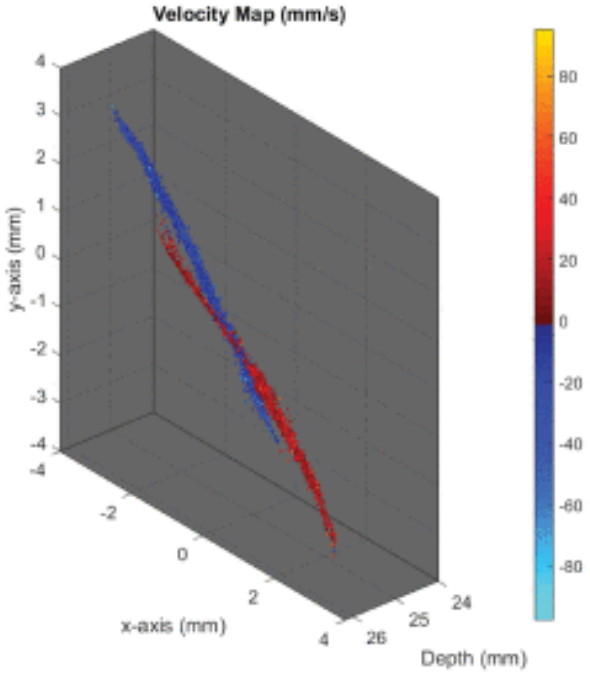

Fig. 14. The velocity map that should be produced from a phantom experiment suing pSVD

For the in vivo studies, spatial similarity matrices should again be reproduced by leveraging pSVD. The computational efficiency of comparing pSVD to normal SVD can be done by calculating the number of operations required for each method. Measuring of time can also be done, however is not as generalizable of a metric as computation times largely depends on the hardware used.

By replicating the same SRUS images by using SVD and pSVD, while incorporating spatial similarity matrices to conduct robust, adaptive filtering, it should confirm our hypothesis that high fidelity SRUS images can be filtered at higher computational efficiency.

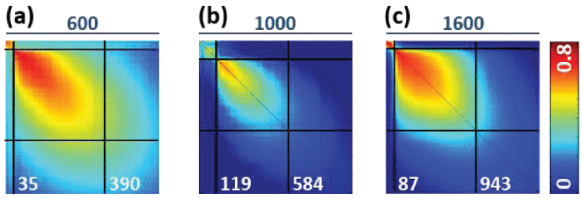

Fig. 15. The spatial similarity matries when SRUS was applied on SRUS images of vasculature. It was found that the signal of microvasculature in the spatial similarity domain appears to have more of an ellipsoidal shape than rectangular.

DISCUSSION The spatial similarity matrix and square ratio of singular values were not as definitive and robust as expected, perhaps because my 3D ultrasound data has low SNR. I also think movement during acquisition greatly affected the reconstructed SRUS video as I'm not sure if the final video really captured any microbubble signal effectively. SVD + spatial similarity also seemed to produce a different SRUS video in the end compared to the pSVD method. Perhaps square segmentation would help with this as I did estimate where the cutoffs were in the former method.

The 3D SRUS data also reflected one of the limitations of filtering with SVD where motion overturns the assumption that tissue, microbubble, and noise subspaces can be separable. Perhaps implementing motion correction can help with this.

Furthermore, while the spatial similarity matrix can be leveraged to construct an adaptive filter for SRUS images, I am not sure what the spatial similarity matrix would look like for different images. In-vivo results reveal that the partitioning of signal is no longer a square but an ellipsoid shape. This causes partitioning to be not as definitive and requires further exploration before its application in clinic.

pSVD also does not make the SVD computation that much efficient. It merely scales one of the dimensions of the Casorati matrix so changes in computational efficiency are only by an order of magnitude of a scalar. If this filtering method was to be expanded upon for 4D acquisitions or longer acquisition times, benefits by truncation of the Casorati matrix may not be sufficient for clinical application. One other approach that has been explored and can be further leveraged to speed up the computation is to leverage stochastic down sampling of the matrix [4]. However, further investigation is required for the spatiotemporal resolution tradeoffs and whether SVD is effective to be applied to such down sampled image.

However, spatiotemporal filtering by SVD is more clinically desirable in some ways than compared to other filtering approaches. High-pass filtering is too elementary relying on trial and error. However, advanced approaches such as machine learning are difficult for the sake of clinical transparency. SVD is merely a re-representation of the same information contained in the SRUS frames so spatiotemporal filtering done this way is more communicable to both patients and doctors.

Conclusion Through this exploration, I saw both the pros and cons of spatiotemporal filtering of SRUS through SVD. While I am skeptical of the practicality of the application of SVD in the clinical setting as a controlled environment is not always guaranteed, its simplicity is elegant and can be effectively communicated to healthcare professionals. It is powerful tool for its differentiation of signal based on spatiotemporal coherence. While in-vivo application of the spatial similarity matrix and pSVD requires further refinement, there are so many other potential changes that can be implemented in other steps of SRUS pipeline to bring it into the clinic. Through this, more exploration should be done to produce high resolution ultrasound at efficient speeds.

REFERENCES

[1] S. Harput *et al.*, “3-D Super-Resolution Ultrasound Imaging With a 2-D Sparse Array,” *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 67, no. 2, pp. 269–277, Feb. 2020, doi: 10.1109/TUFFC.2019.2943646.

[2] B. Pialot, L. Augeul, L. Petrusca, and F. Varray, “A simplified and accelerated implementation of SVD for filtering ultrafast power Doppler images,” *Ultrasonics*, vol. 134, p. 107099, Sep. 2023, doi: 10.1016/j.ultras.2023.107099.

[3] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demene, “Adaptive Spatiotemporal SVD Clutter Filtering for Ultrafast Doppler Imaging Using Similarity of Spatial Singular Vectors,” *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1574–1586, Jul. 2018, doi: 10.1109/TMI.2018.2789499.

[4] U.-W. Lok *et al.*, “Real time SVD-based clutter filtering using randomized singular value decomposition and spatial downsampling for micro-vessel imaging on a Verasonics ultrasound system,” *Ultrasonics*, vol. 107, p. 106163, Sep. 2020, doi: 10.1016/j.ultras.2020.106163.

[5] C. Demene *et al.*, “Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity,” *IEEE Trans. Med. Imaging*, vol. 34, no. 11, pp. 2271–2285, Nov. 2015, doi: 10.1109/TMI.2015.2428634.

[6] K. Christensen-Jeffries *et al.*, “Super-resolution Ultrasound Imaging,” *Ultrasound in Medicine & Biology*, vol. 46, no. 4, pp. 865–891, Apr. 2020, doi: 10.1016/j.ultrasmedbio.2019.11.013.

[7] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, “Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy,” *Biophysical Journal*, vol. 91, no. 11, pp. 4258–4272, Dec. 2006, doi: 10.1529/biophysj.106.091116.

Figures:

- Fig. 1: https://commons.wikimedia.org/wiki/File:PSF_Deconvolution_V.png
- Fig. 2: https://www.abbelight.com/wp-content/uploads/2021/02/TEEE.gif
- Fig. 3: https://www.nature.com/articles/nature16066#Sec1
- Fig. 5: [5]
- Fig. 11: [3]
- Fig. 14: [2]