Bachelor thesis

Philip Mark Suskin

# Development of an Algorithmic Estimation of a Good Signal-to-Noise Ratio Threshold for Reconstruction in Magnetic Particle Imaging

November 9, 2022

supervised by:
Prof. Dr.-Ing. Tobias Knopp
Dr. rer. nat. Martin Hofmann

Hamburg University of Technology
Institute for Biomedical Imaging
Schwarzenbergstraße 95
21073 Hamburg

University Medical Center Hamburg-Eppendorf
Section for Biomedical Imaging
Martinistraße 52
20246 Hamburg

Ich versichere an Eides statt, die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt zu haben.

Hamburg, den 07.11.2022

# Contents

# 1 Introduction

Magnetic particle imaging (MPI) is a tomographic medical imaging method [1] based on the determination of iron oxide nanoparticle concentration. As such, MPI is a tracer-based imaging method. A comparison to other prominent imaging modalities is visualized in table 1.1 [2].

| | CT | MRI | PET | SPECT | MPI |
|---|---|---|---|---|---|
| spatial resolution | 0.5 mm | 1 mm | 4 mm | 10 mm | <1 mm |
| acquisition time | 1 s | 1 s - 1 h | 1 min | 1 min | <0.1 s |
| sensitivity | low | low | high | high | high |
| quantifiability | yes | no | yes | yes | yes |
| harmfulness | X-ray | heating | $\beta/\gamma$ radiation | $\gamma$ radiation | heating |

**Table 1.1:** Quantitative comparison of different imaging modalities

MPI imaging comprises two key components: signal acquisition and signal reconstruction.

Signal acquisition is conducted by exciting iron oxide nanoparticles. This is done using *send coils* that apply a time-dependent external magnetic field. Together with a *selection field*, the signals generated by excitation provide a spatial encoding of the local nanoparticle concentration and are acquired via *receive coils* which measure the magnetic flux density. The selection field is necessary to assign the emitted particle signals to local concentrations and is implemented via a static magnetic field with a single field free point (FFP), for which the field vector is zero.

After signal acquisition, the resulting spatial encoding is reconstructed into a tomogram. This is done by solving an ill-posed inverse problem based on the following linear system of equations:

$$\boldsymbol{S}\boldsymbol{c} = \boldsymbol{u} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{S} \in \mathbb{C}^{K \times M}$ is the system matrix, $\boldsymbol{c} \in \mathbb{R}_+^M$ is the particle concentration vector, $\boldsymbol{u} \in \mathbb{C}^K$ is the measurement vector, and $\boldsymbol{\epsilon} \in \mathbb{C}^K$ is a noise component [3]. A solution is generally found by solving the resulting least squares problem:

$$\underset{\boldsymbol{c} \in \mathbb{R}_+^M}{\operatorname{argmin}} \, \|\boldsymbol{S}\boldsymbol{c} - \boldsymbol{u}_{meas}\|_2^2, \tag{1.2}$$

where $\boldsymbol{u}_{meas}$ denotes the raw measurement vector into which the noise component is integrated. Since the linear system is ill conditioned, however, *regularization* is implemented through three core techniques that aim to penalize non-plausible solutions in addition to minimizing the residual $\boldsymbol{S}\boldsymbol{c} - \boldsymbol{u}_{meas}$.

One regularization technique is Tikhonov regularization, which utilizes a penalization term $\|\boldsymbol{c}\|_2^2$ scaled by a regularization constant $\lambda \in \mathbb{R}_+$. This extends the least squares problem to the following:

$$\underset{\boldsymbol{c}\in\mathbb{R}_+^M}{\operatorname{argmin}} \|\boldsymbol{S}\boldsymbol{c} - \boldsymbol{u}_{meas}\|_2^2 + \lambda\|\boldsymbol{c}\|_2^2. \tag{1.3}$$

Though it reduces noise, Tikhonov regularization can lead to decreases in spatial resolution depending on the value for $\lambda$.

Another technique involves specifying the number of iterations $i$ executed by iterative solvers. While the accuracy of the least squares solution initially improves as the number of iterations increases, excessive iterations cause the solution to be fitted to the noise. Iterative solvers such as the Kaczmarz algorithm are implemented in Julia [4], a programming language designed for numerical and scientific computing, by the MPIReco package [5]. A typical reconstruction that uses three regularization techniques is illustrated with the following high-level reconstruction call:

```
reconstruction(fnSF, fnMeas; SNRThresh = 2.5,
               lambda = 0.001, iterations = 3)
```

The third regularization technique, which is the focus of this thesis, is the implementation of an SNR threshold $\Theta$ in order to remove rows (frequency components) below a certain SNR from the linear system. More formally, this means that original index set $I_{K_0} = \{1, 2, ..., K_0\}$ containing all frequency components is reduced to the index set $I_{K_\Theta} = \{k \in I_{K_0} \mid SNR(k) \geq \Theta\}$. This modifies the problem to the following:

$$\underset{\boldsymbol{c}\in\mathbb{R}_+^M}{\operatorname{argmin}} \|\tilde{\boldsymbol{S}}_\Theta\boldsymbol{c} - \tilde{\boldsymbol{u}}_{meas,\Theta}\|_2^2 + \lambda\|\boldsymbol{c}\|_2^2, \tag{1.4}$$

where $\tilde{\boldsymbol{S}}_\Theta \in \mathbb{C}^{K_\Theta \times M}$ and $\tilde{\boldsymbol{u}}_{meas,\Theta} \in \mathbb{C}^{K_\Theta}$ are the modified system matrix and measurement vector, respectively, and $K_\Theta = |I_{K_\Theta}|$ is the cardinality of the modified index set. While the exclusion of frequency components through an increasing SNR threshold reduces noise in the solution, the spatial resolution is also reduced, which can lead to blurring for large enough thresholds.

The influence of the SNR threshold in the context of respective frequency components [6] is visualized graphically for a minimum frequency of 80 kHz in figure 1.1. Note that a minimum frequency shifts the frequency component index sets, such that an index set $I_K = \{1, 2, ..., K\}$ implicitly starts at the minimum frequency.

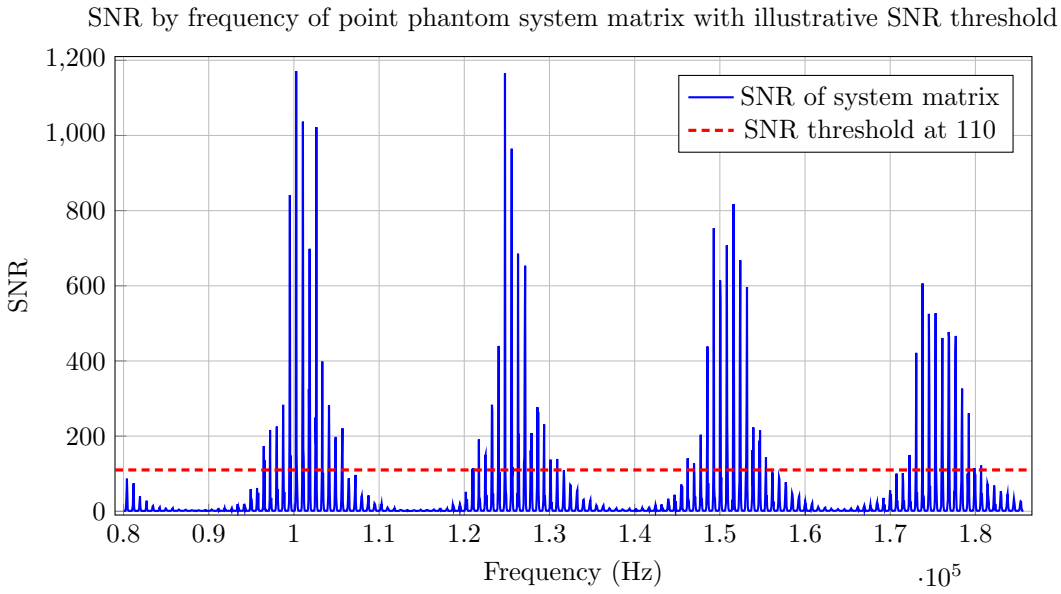



**Figure 1.1:** Visualization of the effect of the SNR threshold on frequency components

The impact of the exclusion of frequency components through the SNR threshold is further illustrated in figure 1.2, in which the number of remaining frequency components is plotted as a function of the SNR threshold for two system matrices. It can be observed that frequency component count has a non-linear relationship to SNR threshold. As such, shifts in the SNR threshold result in non-linear changes in the spatial resolution of reconstructions. Instead, changes in the spatial resolution of reconstructions depend on the SNR threshold in **logarithmic** fashion.

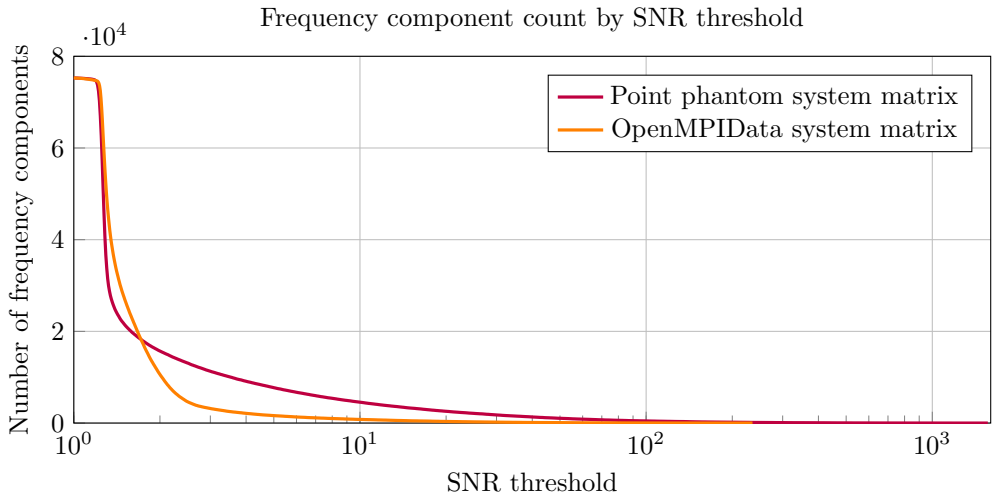



**Figure 1.2:** Relationship between frequency component count and SNR threshold

Since the ideal SNR threshold has been chosen by hand up until this point, an algorithmic selection is a desirable improvement to the reconstruction process.

# 2

# Problem Statement

The innate human ability to assess the quality of an image without the knowledge of its exact contents is one that is difficult to model computationally. Problems of this kind fall under the category of NRIQA (non-reference image quality assessment), for which there exists a wide range of metrics with emphases on varying graphical attributes [7]. A quality assessment metric tailored to MPI reconstructions should, more specifically, emphasize penalizing distortion, particularly noise and blur. Additionally, instead of calculating a standardized score for **overall** image quality, the metric only has to generate scores which have a well-defined ordinal relationship to the scores of other reconstructions **for a specific measurement** at various SNR thresholds.

A set of 10 reconstructions of a point phantom measurement at an iron oxide concentration of 6.25 mmol/L (more formally introduced in section 3.1) generated at logarithmically equidistant SNR thresholds can be found in figure 2.1. The reconstructions are sorted left-to-right then top-to-bottom by the SNR threshold used for reconstruction and are normalized via scaling by the reciprocal of the maximum pixel brightness. They are presented using the viridis colormap.

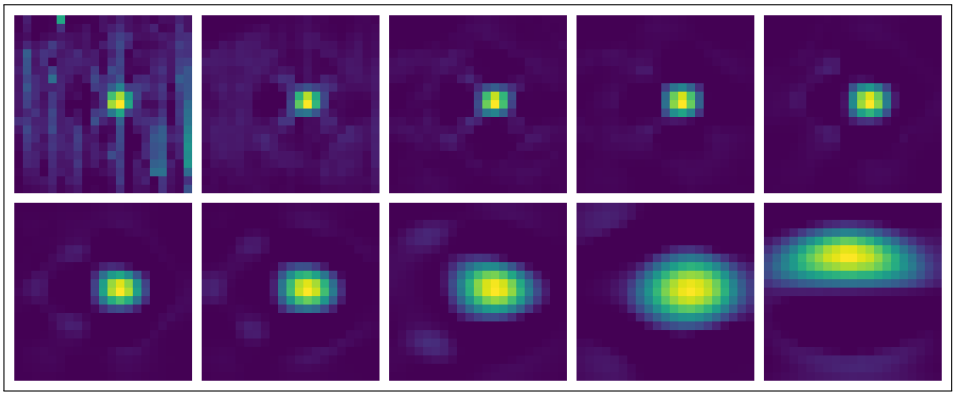


**Figure 2.1:** Grid of 10 normalized reconstructions of point phantom at concentration of 6.25 mmol/L

This set of reconstructions illustrates the transition from noisy to blurry reconstructions as the SNR threshold increases. An additional set of reconstructions of the same point phantom measurement, this time at an iron oxide concentration of 0.2 mmol/L, is displayed in figure 2.2.

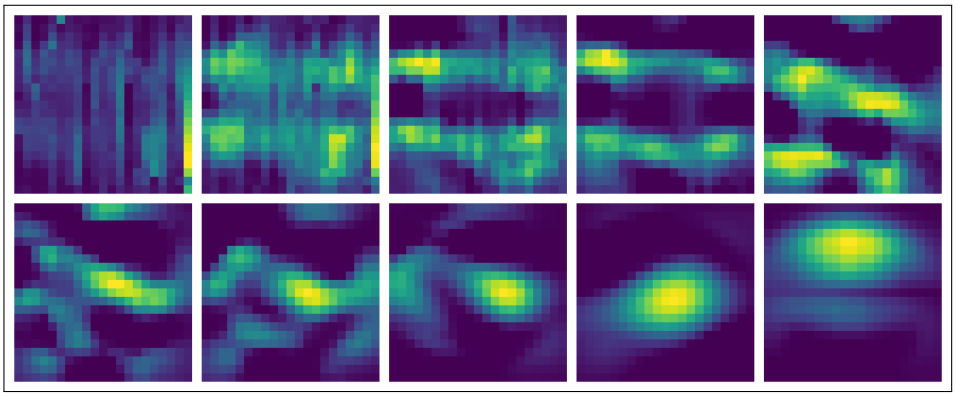


**Figure 2.2:** Grid of 10 normalized reconstructions of point phantom at concentration of 0.2 mmol/L

This set of reconstructions, while also exhibiting a transition from noisy to blurry reconstructions, demonstrates that this transition is not uniform across different MPI measurements. Thus, there exists no single SNR threshold capable of generating the visually optimal reconstruction for every measurement.

Nevertheless, in both sets of reconstructions, and as a common theme throughout various measurements, a respective optimum can be found between noisy reconstructions at low SNR thresholds and blurry reconstructions at high SNR thresholds. Designing a metric that scores reconstructions according to this kind of observation and devising an algorithm that finds the consequent optimal SNR threshold for reconstruction as reliably and as efficiently as possible encapsulates the problem discussed in this thesis.

# 3 Experiments & Methods

Experiments were conducted on various MPI measurements in order to find a suitable algorithm for the estimation of an optimal SNR threshold with respect to reconstruction quality. The algorithm should comprise two modular components, namely a quality assessment metric and an optimization algorithm, both of which are analyzed on their performance and validated by qualified human reference.

## 3.1 Data utilized for analysis

In order to establish a basis for analysis, a composition of various phantom measurements from two datasets was employed.

The first dataset involves a point phantom with various iron oxide concentrations [8]. The point phantom consists of a sphere of 2.4 mm diameter which contains tracer material with variable iron oxide concentration, starting at 400 mmol/L with dilutions by a factor of 2 for subsequent measurements until reaching a concentration of 0.2 mmol/L, totaling 12 measurements. For each measurement, a $21 \times 21$ pixel 2-dimensional phantom resulting from slicing the corresponding $21 \times 21 \times 24$ voxel reconstruction was used for analysis. The slice was taken from the xy plane by taking the brightest pixel along the the z axis[1] for each coordinate in the plane.

The other dataset stems from OpenMPIData [9]. Therein are measurements for three 3-dimensional phantoms, denoted as concentration, resolution, and shape phantoms respectively. The three phantoms are displayed in figure 3.1. Slices were taken from $37 \times 37 \times 37$ voxel reconstructions of these phantoms to form $37 \times 37$ pixel 2-dimensional phantoms to use for analysis.

The concentration phantom consists of eight cubes of 8 µL volume that form the vertices of a rectangular prism. Adjacent cubes are spaced 12 mm apart[2] within the xy plane and 6 mm apart along the z axis. Each cube contains a different iron oxide concentration, starting at 100 mmol/L with dilutions by a factor of 1.5 for subsequent cubes.

[1] In the interval from 7 to 9.
[2] Distance from center to center.

**(a)** Concentration **(b)** Resolution **(c)** Shape

**Figure 3.1:** OpenMPIData phantoms

A numbered CAD sketch for the concentration phantom is displayed to the right, with the corresponding concentrations from 1 to 8 as follows: [44.4, 100, 29.6, 8.77, 19.7, 66.6, 13.1, 5.85]. Six 2-dimensional phantoms resulting from slices of the reconstructed concentration phantom were used for analysis. Two slices were taken from each principal plane, i.e. the xy, xz, and yz planes, with one slice positioned at 1/3 of the depth of the remaining axis and the other positioned at 2/3 of the depth of the remaining axis.

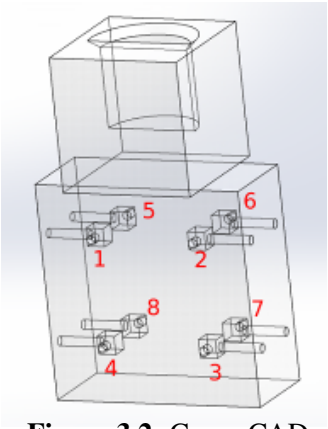



**Figure 3.2:** Conc. CAD

The resolution phantom consists of 5 thin cylindrical tubes, each with iron oxide concentrations of 50 mmol/L. The tubes stem from a common origin and diverge from each other at varying angles within the xy and yz plane (20–30° in the xy plane and 10–15° in the yz plane). Three 2-dimensional phantoms resulting from slices of the reconstructed resolution phantom were used for analysis. One slice was taken from each principal plane by slicing at 1/2 of the depth of the remaining axis.

The shape phantom consists of a cone of 683.9 µL volume with an iron oxide concentration of 50 mmol/L. The cone has a slightly elliptical base and an apex angle of approximately 10° in the xy plane. Three 2-dimensional phantoms resulting from slices of the reconstructed shape phantom were used for analysis. One slice was taken from each principal plane by taking the brightest pixel along the remaining axis for each coordinate in the plane.

## 3.2 Human SNR threshold picks

Since image quality is subjective at its core, human quality assessments were generated by choosing optima out of sets of reconstructions. Specifically, visually optimal reconstructions were chosen from sets of 10 normalized reconstructions per phantom, for which the SNR

thresholds used for reconstruction are logarithmically equidistant across the interval $[1, \Theta_1]$, where $\Theta_1$ is defined as the SNR threshold for which the respective system matrix has exactly one frequency component. Two such sets have previously been presented in figures 2.1 and 2.2.

Picks were conducted independently by 10 volunteers of the IBI (Institute for Biomedical Imaging). They serve as a reference to validate valuations of metric quality and algorithm performance. A complete collection of all reconstruction grids from which picks were conducted is located in the appendix under figures 7.1 and 7.2, with the corresponding table of the picks in table 7.3.

## 3.3 Algorithmic SNR threshold searches

The first step toward formulating a suitable algorithm for finding the optimal SNR threshold in terms of reconstruction quality involves establishing a quality assessment metric that penalizes noise and blur. Functionally, this means that the metric generates monotonic functions when plotted against degree of image distortion while applying increasing levels of noise and blur individually. Both functions have identical monotonicity to each other (either increasing or decreasing) and ideally have slopes far from zero, as this simplifies the resulting optimization problem.

Once a metric has been established, an algorithm to solve the resulting optimization problem must be devised. The optimization problem aims to maximize an objective function defined by the proposed quality assessment metric as a function of the SNR threshold used to reconstruct the images that the metric is applied to. Neither the shape nor the derivative of the objective function is known. As such, a derivative-free optimization algorithm is required to determine the optimal SNR threshold with respect to reconstruction quality. All explicitly defined reconstruction parameters used during analysis (aside from the SNR threshold) are listed by phantom in table 7.1 in the appendix.

### 3.3.1 Quality assessment metric

In order to determine a metric tailored to the quality assessment of reconstructed MPI images, multiple regularization norms were tested on their responses to blur and noise in images of various artificial phantoms. The norms evaluated include L1, L2, and TV (total variation). They each have specific definitions for application to images [10]. For consistency, all images to which the norms are applied are normalized discrete grayscale images, with normalization implemented via scaling by the reciprocal of the maximum pixel brightness.

Applied to a discrete grayscale image $I$ of size $M \times N$ pixels, where $I(m, n)$ denotes the brightness of the pixel in the $m$-th row and $n$-th column, the norms are defined as follows:

$$\boldsymbol{L1}(I) = \sum_{m=1}^{M} \sum_{n=1}^{N} |I(m, n)| \tag{3.1}$$

$$\boldsymbol{L2}(I) = \sum_{m=1}^{M} \sum_{n=1}^{N} I(m, n)^2 \tag{3.2}$$

$$\boldsymbol{TV}(I) = \sum_{m=1}^{M-1} \sum_{n=1}^{N} |I(m, n) - I(m+1, n)| + \sum_{m=1}^{M} \sum_{n=1}^{N-1} |I(m, n) - I(m, n+1)| \tag{3.3}$$

*Distortion curves* were generated by plotting each norm as a function of the SSIM[3] (structural similarty index) between a distorted image and the original. The images used for plotting were generated by hand-drawing phantoms on a $300 \times 300$ pixel canvas[4] and scaling the canvas down to a $30 \times 30$ image using linear interpolation. Blur was applied to the images via a gaussian kernel sized between 0 and 3 at intervals of 0.1. Noise was applied via absolute-valued AWGN (additive white Gaussian nosie) with a standard deviation between 0 and 0.3 at intervals of 0.01. An example of a hand-drawn phantom is displayed to the right, and the corresponding distortion curves, in which each norm is independently normalized between 0 and 1, are shown in figures 3.4 and 3.5. In addition, the application of presented distortion methods to this phantom is visualized in figure 3.6.

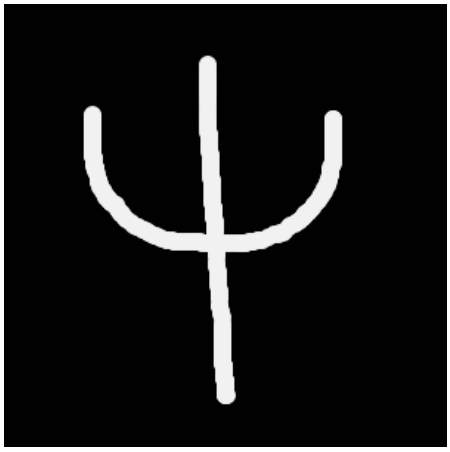


**Figure 3.3:** Trident phantom

Distortion curves were generated for a wide array of hand-drawn phantoms with various graphical features in order to investigate potential trends in shared monotonicity for each norm.

[3] PSNR and RMSE were also used as measures of image distortion and yielded practically identical results regarding monotonicity.

[4] Phantoms were drawn on an initially black canvas by filling in white circles with a radius of 6 pixels wherever the mouse is dragged.

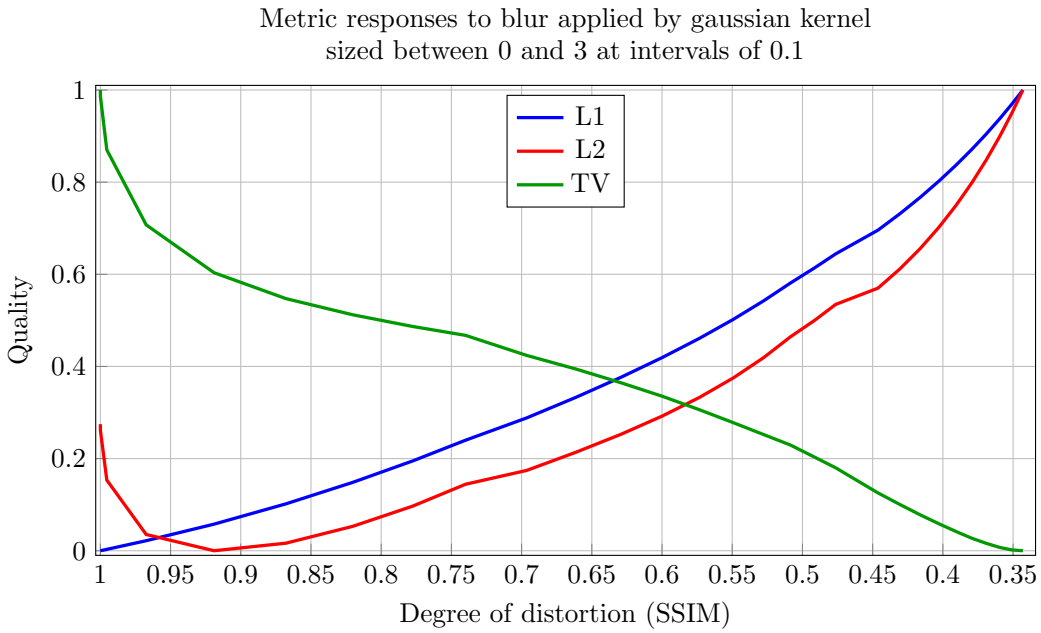



**Figure 3.4:** Blur curve of trident phantom

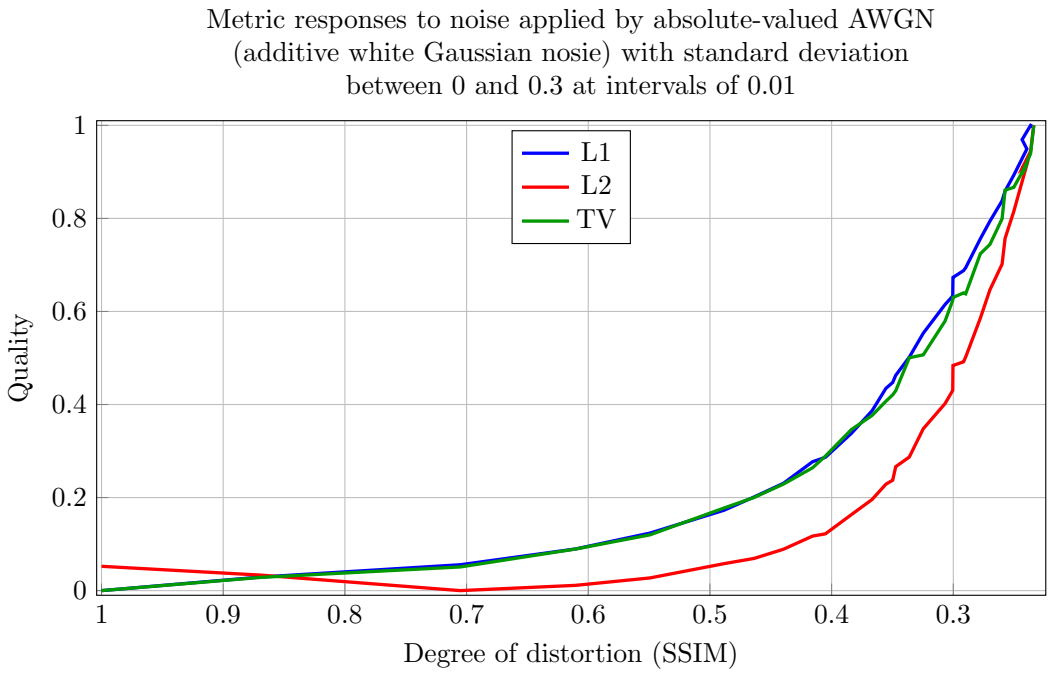



**Figure 3.5:** Noise curve of trident phantom

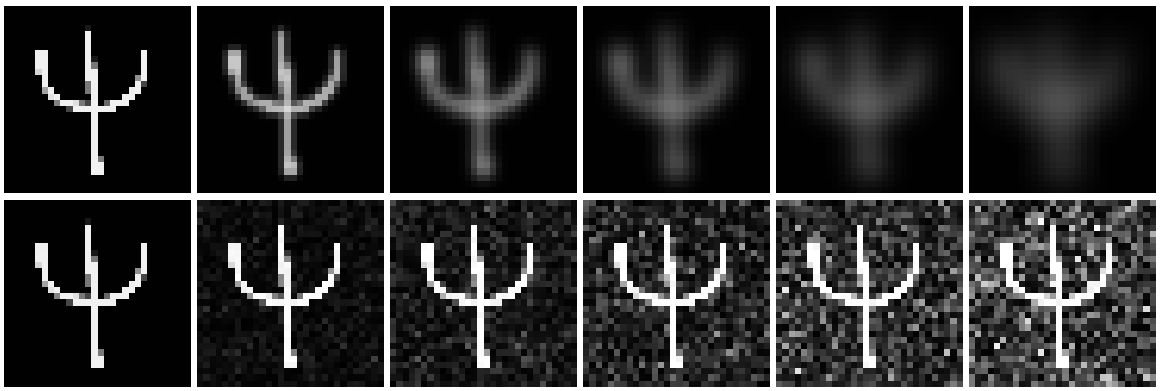


**Figure 3.6:** Blur (top row) applied via gaussian kernel between 0 and 3 at intervals of 0.6. Noise (bottom row) applied via absolute-valued AWGN with standard deviation between 0 and 0.3 at intervals of 0.06

### 3.3.2 Optimization algorithm

To determine a suitable optimization algorithm for finding SNR thresholds that maximize reconstruction quality, LIPO, a derivative-free global optimization algorithm based on the global optimization of Lipschitz functions [11], and grid search were analyzed on their ability to find optima with constraints on the number (and cost) of function evaluations.

Since the underlying reconstruction quality as a function of SNR threshold is not known, the maximum found in a 100 point grid search was used as the ground truth, i.e. the assumed global maximum of the function, for each respective phantom.

LIPO was chosen for analysis due to its particularly good performance for few function evaluations and its focus on finding the tallest peak of the objective function rather than convergence along this peak [12]. These properties align with the optimization of reconstruction quality, since each function evaluation involves a reconstruction and is thereby expensive in terms of time and resources, and since high precision with respect to the exact value of the optimal SNR threshold is not as important as finding an optimum in its vicinity, as slight deviations in the SNR threshold result in negligible differences in image quality.

In order to visualize LIPO's performance for a limited number of function evaluations, its iterative search path was logged for different *frequency component budgets*. Frequency component budgets are implemented into LIPO as stopping conditions, which halt the search if the running sum of frequency components encountered in previously executed reconstructions would exceed the budget if the next reconstruction were to be executed. Because grid search acts as a baseline optimization algorithm, budgets were defined as the sums of frequency components in 5, 10, 15, and 20 log-scaled grid searches. This way, LIPO can be compared to grid search for a variety of budgets. The budgets are listed in table 7.2 in the appendix.

Comparisons were conducted by analyzing the deviation of LIPO's prevailing optimal SNR threshold from the true optimum for each iteration and comparing these deviations to those

of the grid searches, with emphasis on the relevant budgets. Deviations were quantified using a *geometric norm*. For an *optimization outcome*, i.e. the SNR threshold of the result of optimization, $\Theta_{out}$ and a true optimal SNR threshold $\Theta_{true}$, the geometric norm is defined as follows:

$$\delta = \frac{\Theta_{out}}{\Theta_{true}} \tag{3.4}$$

LIPO's optimization outcomes themselves were also analyzed for each budget in order to visualize LIPO's performance in the context of its respective objective function.

# 4 Results

## 4.1 Human SNR threshold picks

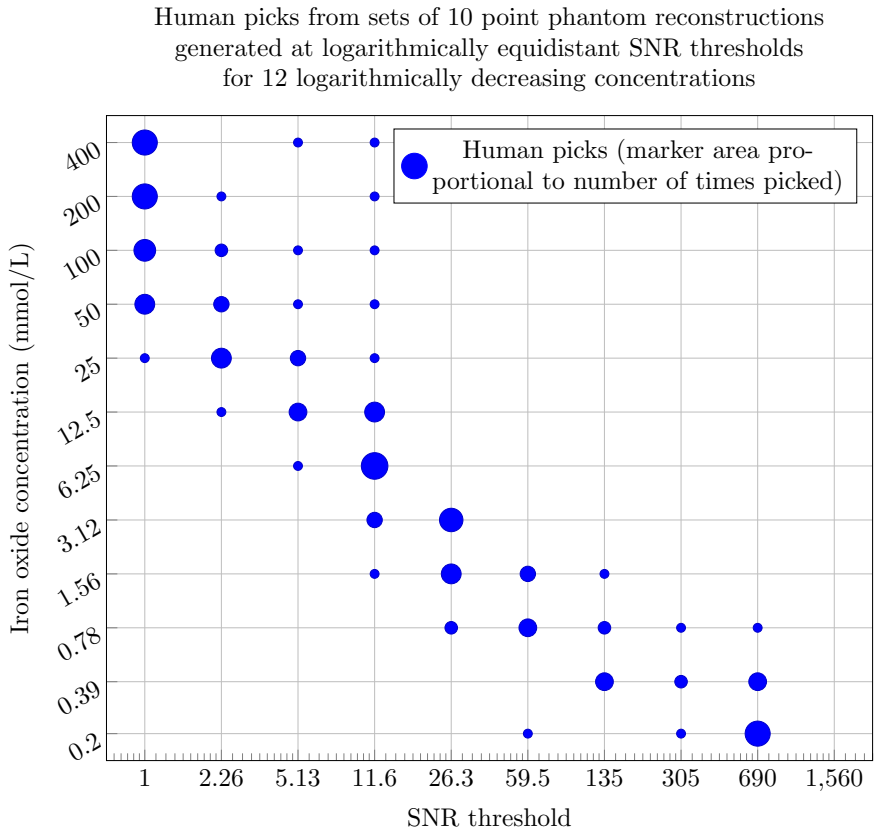
Due to differing priorities regarding graphical attributes of images (i.e., emphasis on high resolution or on minimal noise), a certain degree of variance exists within the picks of the volunteers for each set of reconstructions. The picks recorded for each reconstruction set have been condensed to illustrate general tendencies and variance in figure 4.1. In the case of the point phantom reconstruction sets, it can be observed that picks shift toward higher SNR thresholds as the iron oxide concentration decreases. However, for the OpenMPIData reconstruction sets, picks generally center around similar low SNR thresholds.

## 4.2 Algorithmic SNR threshold searches

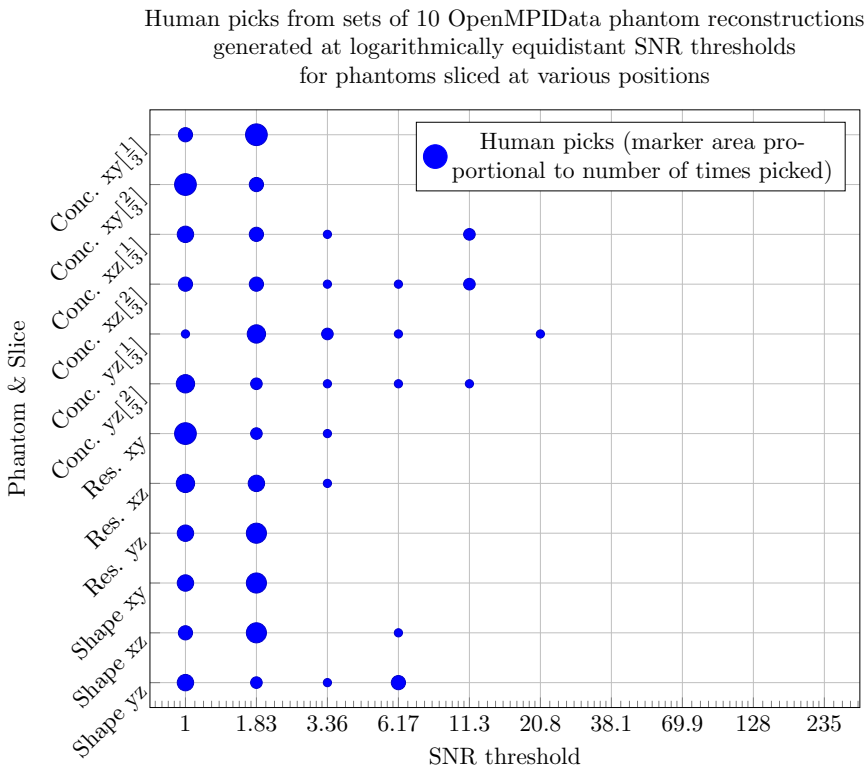
### 4.2.1 Quality assessment metric

Analyzing large numbers of hand-drawn phantoms for trends regarding shared monotonicity in distortion curves for each metric showed that, for the majority of phantoms, L1 is monotonically increasing for both noise and blur. For a subset of phantoms, L2 is also monotonically increasing for both forms of distortion, but it is clear from the results that L2 is, at best, equally as good as L1. In many cases, L2 will "curl up" for low levels of distortion, thereby disrupting monotonicity. TV generally generates the opposite monotonicity for noise as it does for blur, with the blur curve being monotonically decreasing and the noise curve being monotonically increasing. An instance of each of these general observations can be seen in the distortion curves of the previously introduced trident phantom in figures 3.4 and 3.5.

Exceptions to the usual shared monotonicity generated by L1 were found in two classes of phantoms, namely **scattered non-sparse phantoms** and **spacious dense phantoms**. For phantoms of either of these classes, the L1 norm fails to generate satisfactory shared monotonicity for the blur and noise curves.




**(a)** Point phantom




**(b)** OpenMPIData

**Figure 4.1:** Overview of human picks from sets of reconstructions

A model phantom from the scattered non-sparse class is displayed to the right and the corresponding distortion curves are displayed in figures 4.3 and 4.4. It is important to note that, in this case, the noise curves of each norm only exhibit SSIM values in the interval [0.9, 1], and that, as a result, the dip in the noise curve for each metric at an SSIM of approximately 0.999, while disruptive of monotonicity, is not particularly problematic, since an image with such high SSIM is optically nearly identical to the original image. The flatness of the L1 blur curve for SSIM higher than 0.5 indicates that the L1 norm is non-responsive to low/mild blurring. In this case, the relative difference in the L1 norm between the model phantom blurred to a 0.38 SSIM[1] and the original phantom only amounts to approximately 3.36%. Contrarily to most phantom classes, TV can be observed to generate substantially shared monotonicity. In addition, the steepness of both TV distortion curves indicate that it is, in contrast to L1, highly responsive to both noise and blur. These observations demonstrate that the use of TV instead of L1 is a viable option for scattered non-sparse phantoms.

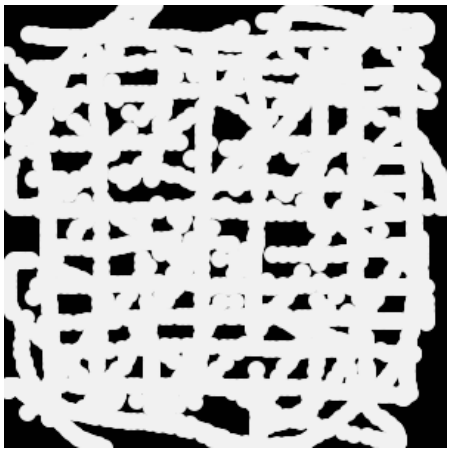


**Figure 4.2:** Model scattered non-sparse phantom

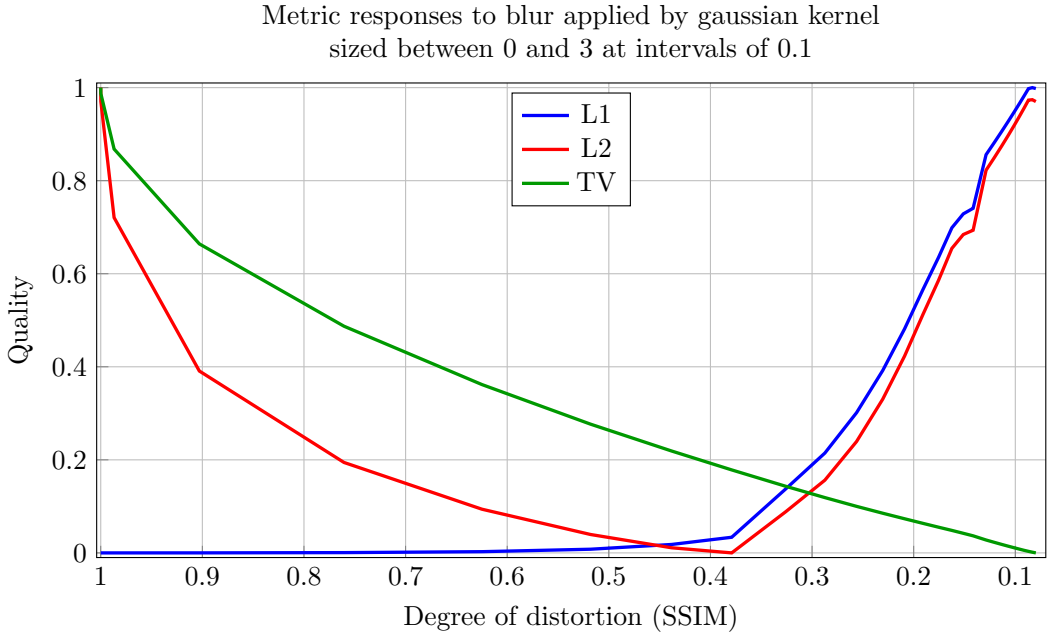



**Figure 4.3:** Blur curve of scattered non-sparse phantom

[1]This level of blurring is achieved with a Gaussian kernel size of 1.

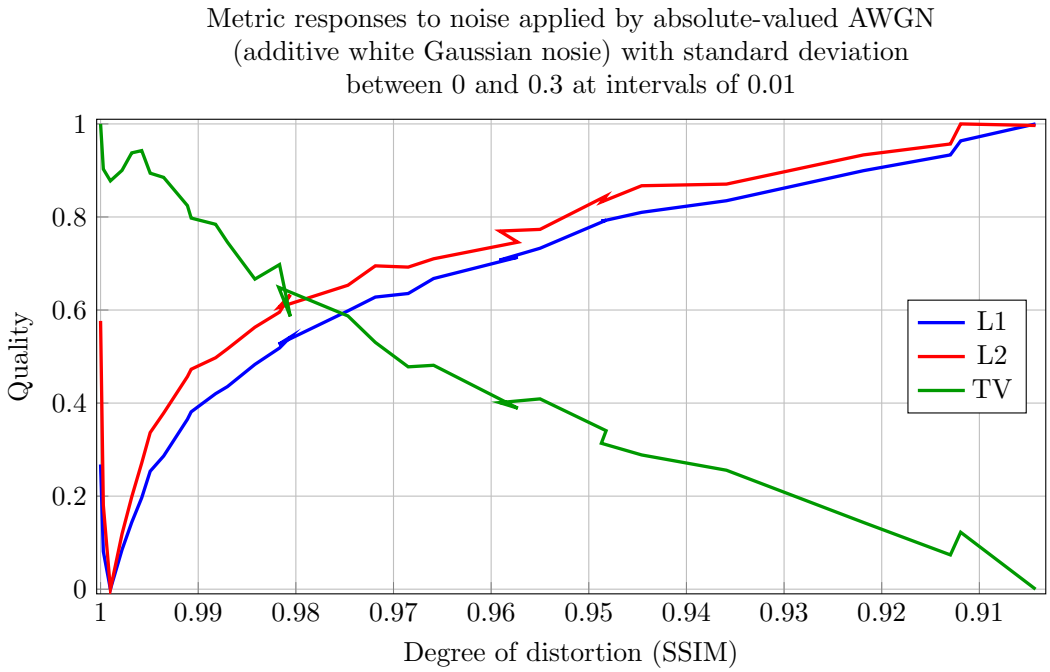



**Figure 4.4:** Noise curve of scattered non-sparse phantom

A model phantom from the spacious dense class is displayed to the right and the corresponding distortion curves are displayed in figures 4.6 and 4.7. For this phantom, the L1 norm generates curves with opposing monotonicity, as the noise curve retains essentially increasing monotonicity and the blur curve generates decreasing monotonicity. Similar behavior is shared by L2 and TV, indicating that the qualities of spacious dense phantoms at various levels of distortion are not easily assessed by any of the three proposed metrics.

**Figure 4.5:** Model spacious dense phantom

The significance of the results presented in this section is contingent on two key assumptions. First, that the transition from noisy reconstructions at low SNR thresholds to blurry reconstructions at high SNR thresholds is relatively "distinct", such that noise and blur are largely mutually exclusive in reconstructions and such that, for any measurement, there exists a reconstruction with low levels of both forms of distortion at some SNR threshold that is in between the SNR thresholds of reconstructions that are distorted predominantly by differing forms of distortion. Second, that the hand-drawn linearly downscaled phantoms are representative of real MPI phantoms and that the noise and blur applied to these phantoms is more or less the same as that found in MPI reconstructions. As a way to validate the practical reliability of the proposed metrics for real MPI measurements, the metrics were further used to plot human picks in the context

of reconstruction quality as a function of SNR threshold. The plots for each MPI phantom are displayed in figures 7.3 and 7.4 in the appendix. They demonstrate that the metrics shown in this section to adequately assess the quality of distorted images in theory indeed lead to optimal SNR thresholds that lie closely to the best picks of human volunteers.

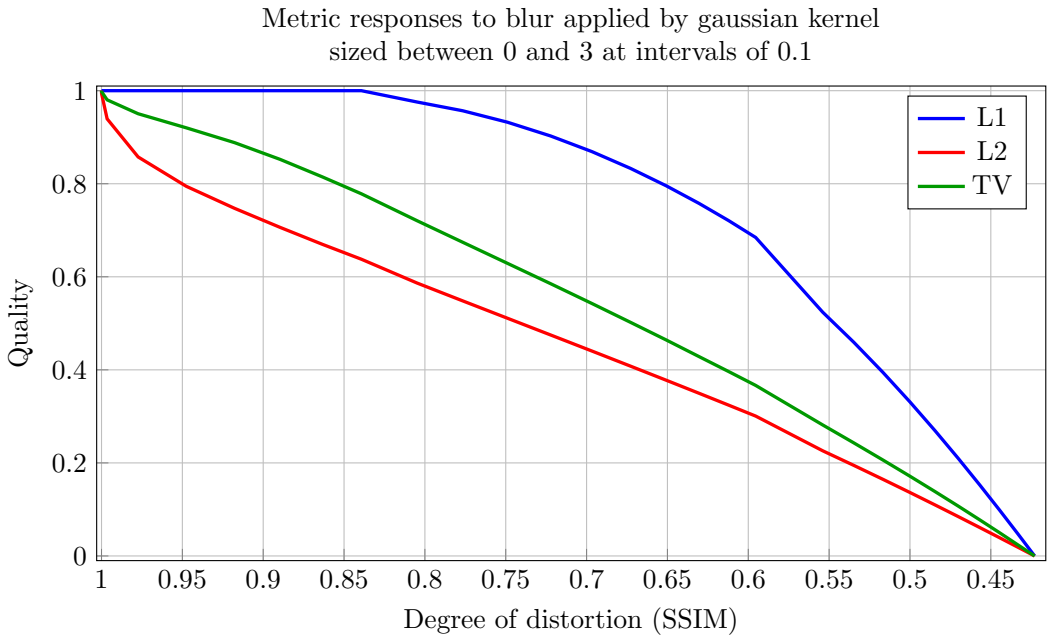



**Figure 4.6:** Blur curve of spacious dense phantom

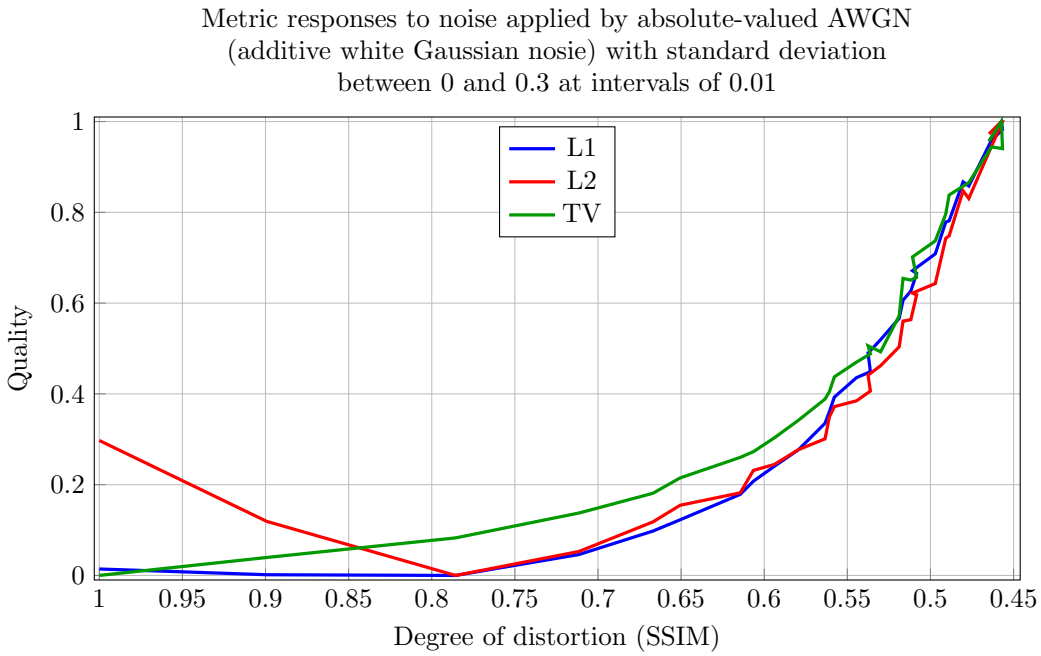



**Figure 4.7:** Noise curve of spacious dense phantom

### 4.2.2 Optimization algorithm

Comparison of LIPO and grid search yielded differing results for each phantom. LIPO's comparative performance was observed to be heavily tied to the position of the optimal SNR threshold for reconstruction quality, with higher optimal SNR thresholds leading to lower optimization outcome deviations for all budgets. This is demonstrated using the point phantom at an iron oxide concentration of 0.2 mmol/L, which, when using L1 as a quality assessment metric[2], has an optimal SNR threshold of approximately 381.1. Resulting deviations for LIPO and grid search are depicted in figure 4.8. It can be observed that LIPO's deviation already falls below that of the 20 point grid search before it has fully expended the 5 point grid search budget. A further observation is that deviations don't worsen for larger budgets. Since LIPO is deterministic and the order of its search path is not influenced by stopping conditions, an iterative search path with an increased budget is always a continuation of the original iterative search path. This, by extension, means that the optimization outcomes of executions with higher budgets deviate from the true optimum strictly equally to or less than the outcomes of executions with lower budgets.

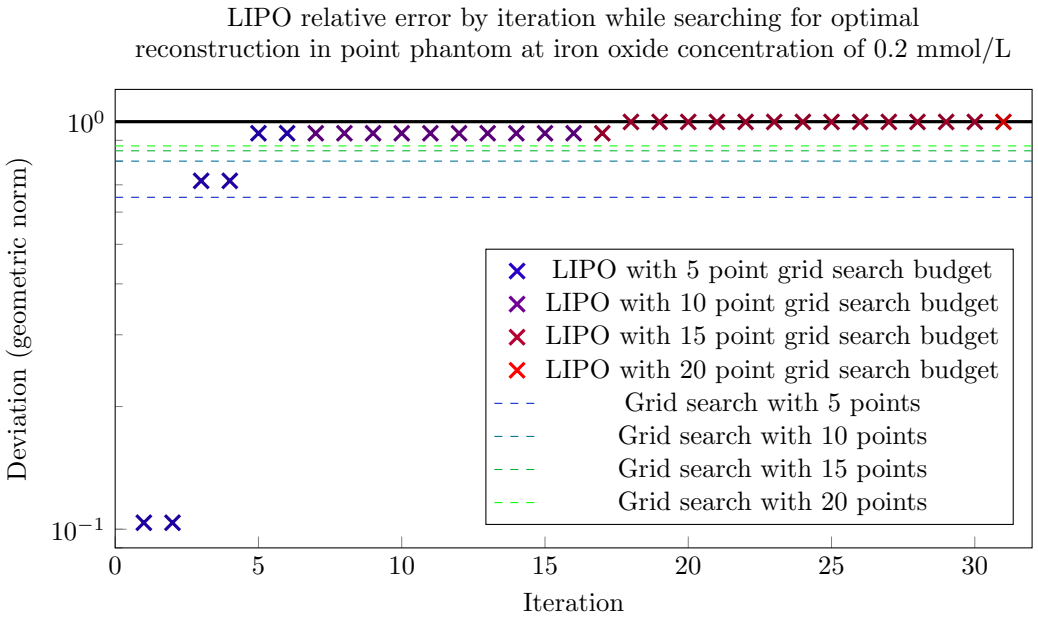



**Figure 4.8:** LIPO and grid search optimization outcome deviations for point phantom at an iron oxide concentration of 0.2 mmol/L

In addition to the deviations, LIPO's optimization outcomes for each budget are illustrated in figure 4.9. Here it can be observed that LIPO's optimization outcomes for the 15 and 20 point grid search budget are in fact of higher quality than the ground truth optimum (defined by the optimum of a 100 point log-scaled grid search, as mentioned in section 3.3.2), with

[2]More precisely, −L1 is used, so that the objective function is to be maximized.

the optimization outcomes achieving a quality of approximately -61.59 and the ground truth optimum amounting only to approximately -62.41.

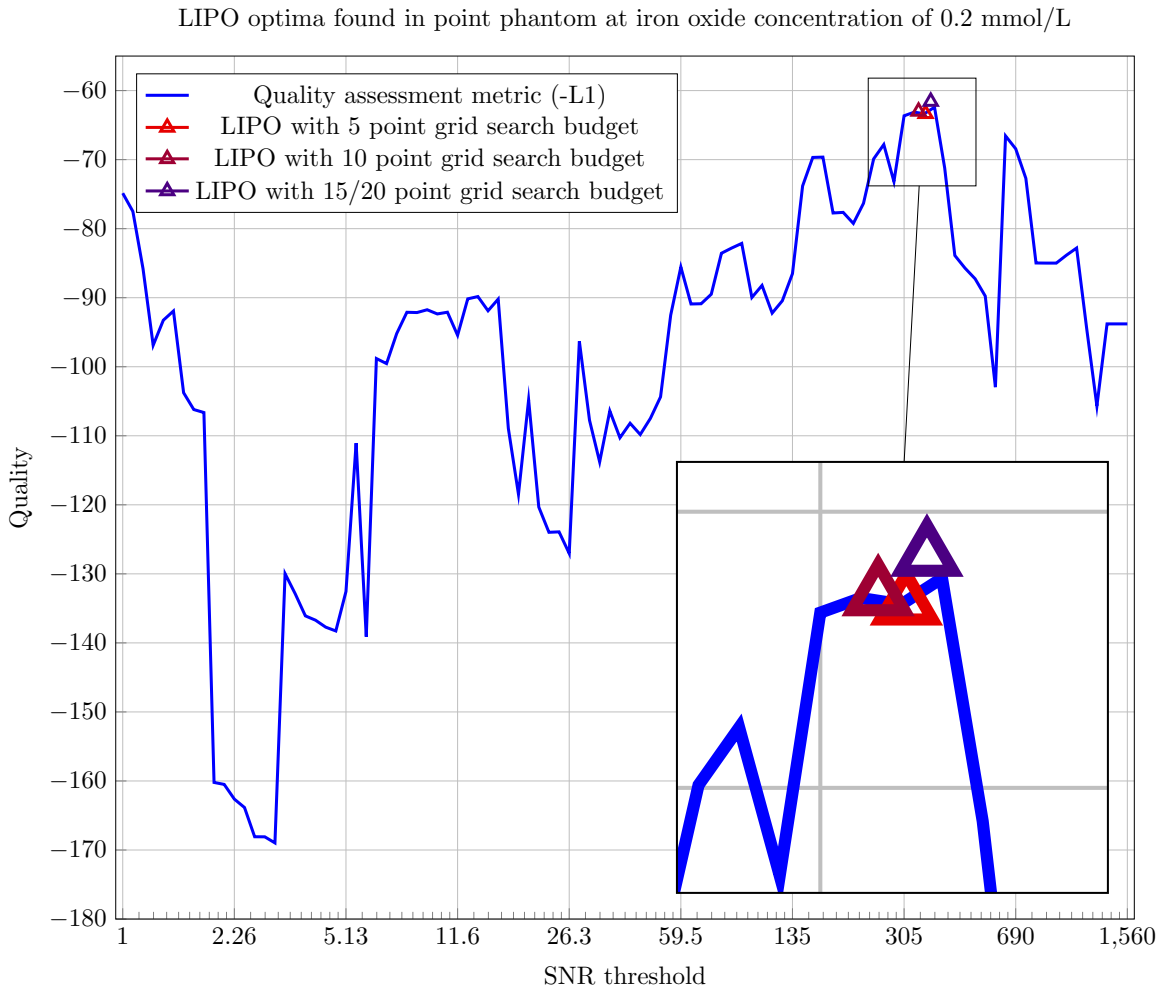



**Figure 4.9:** LIPO optimization outcomes by budget for point phantom at an iron oxide concentration of 0.2 mmol/L

By contrast, LIPO's comparative performance drops for optimal SNR thresholds near 1. The OpenMPIData shape phantom sliced on the xy plane, for example, has a maximum at an SNR threshold of approximately 1.247. The deviations for LIPO and grid search are depicted in figure 4.10, and positions of LIPO's optimization outcomes are visualized in figure 4.11. It is important to note that, for this phantom, TV is used as the quality assessment metric. This is because, due to the large quantity of noise artifacts integrated into the shape phantom measurement, the phantom can be categorized as scattered and non-sparse, which has previously been discussed to warrant the use of TV instead of L1 in section 4.2.1. Contrarily to the previous phantom, LIPO's deviation is at least as high as that of a grid search of identical budget for all proposed budgets.

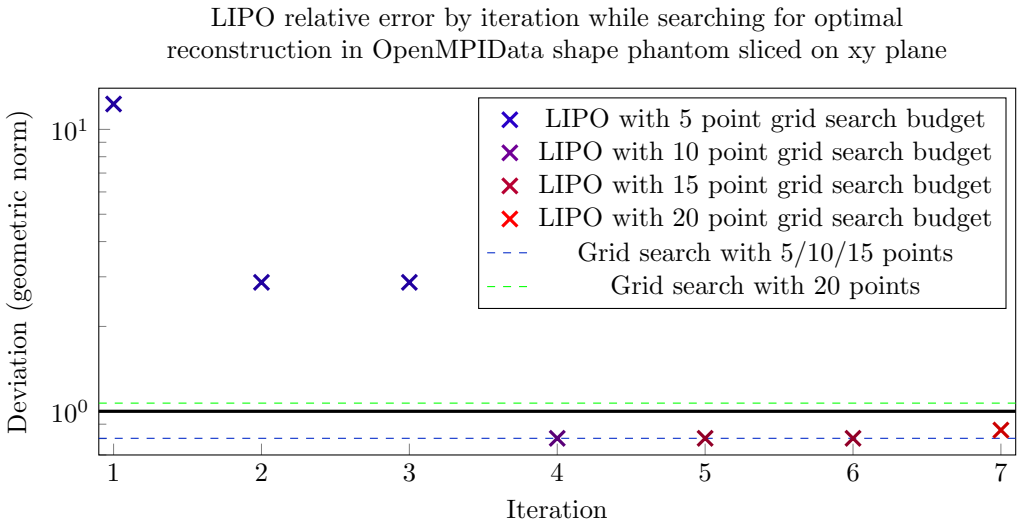



**Figure 4.10:** LIPO and grid search optimization outcome deviations for OpenMPIData shape phantom sliced on xy plane

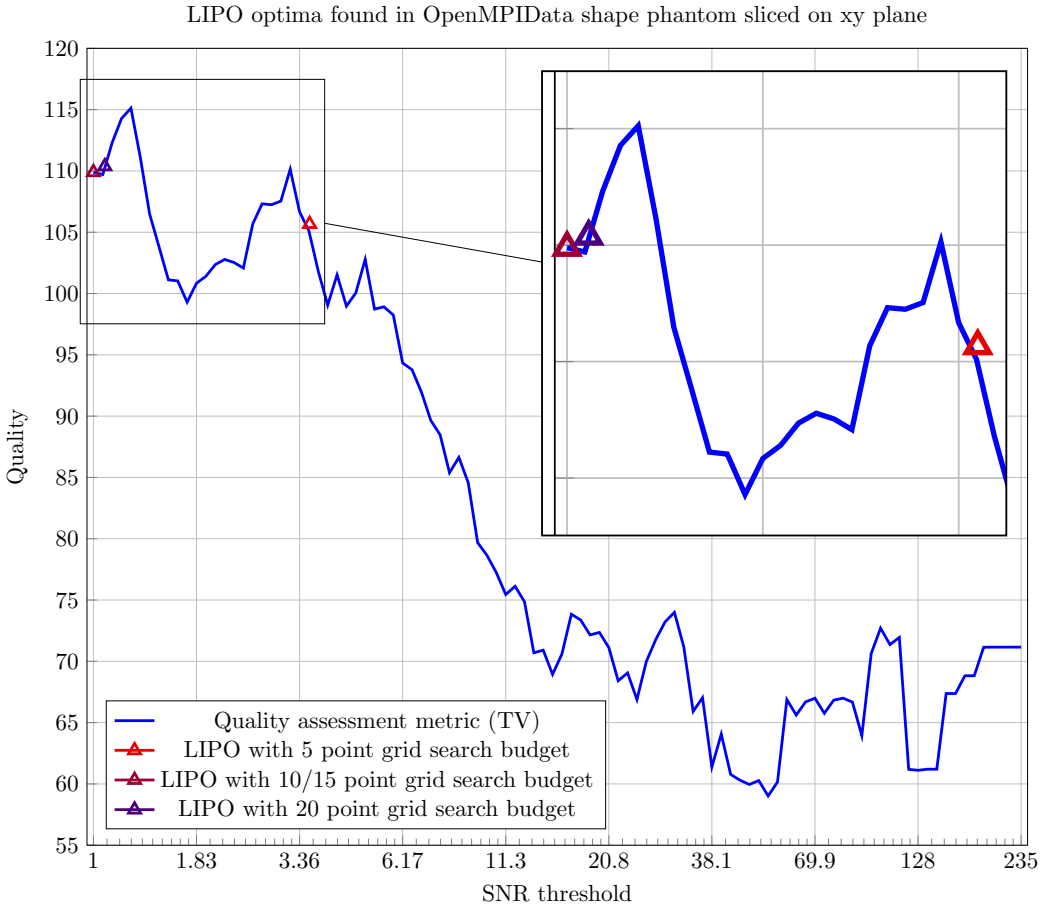



**Figure 4.11:** LIPO optimization outcomes by budget for OpenMPIData shape phantom sliced on xy plane

# 5 Discussion

The results attained through analysis provide insight on the viability of an algorithmic estimation of the optimal SNR threshold regarding reconstruction quality. However, there are caveats which concern parameters of this algorithm.

Since L1 is the most consistently viable metric across all types of phantoms, it represents a good candidate for a default quality assessment metric. However, classes of phantoms for which L1, or even all three presented norms, fail to properly assess quality, have been shown to exist (see section 4.2.1). For this reason, it is important that the quality assessment metric remain interchangeable and open to improvement. The interchangeability of the quality assessment metric reveals a distinct advantage that grid search has over iterative optimization algorithms, namely that, if its reconstructions are cached, it allows for the retroactive substitution of the quality assessment metric without the need for additional reconstructions. By contrast, iterative optimization algorithms must be run practically[1] from scratch if the quality assessment metric is substituted.

A further point regarding the choice of quality assessment metric pertains to the identification and distinction of the scattered non-sparse and spacious dense phantom classes introduced in section 4.2.1. Intuitive approaches based on image processing methods, such as quantification of gradient and integral values of surface plots [13], are possible, with basic methods yielding satisfactory assessments of which class, if either, a phantom belongs into. It is important to note, however, that without prior knowledge of the appearance of the phantom of interest, the reliability of these methods could be heavily reduced, since there is no guarantee that an initial reconstruction at a certain SNR threshold is representative of the underlying phantom and thereby a useful indicator of the phantom class.

Establishing a default frequency component budget is not straight forward, since it depends heavily on the priorities and resources of the user. The results of the analysis of the optimization algorithm indicate that a budget of $1.25 \cdot K_0$, where $K_0$ represents the total number of frequency components in the system matrix used in reconstruction, constitutes a reasonable recommendation for most purposes. This recommendation is subject to change, however, if constraints on the SNR threshold search interval are implemented. Constraints on the search interval, particularly those that exclude low SNR thresholds, generally lower the budget necessary to acquire equally good optimization outcomes. If *a priori* information is known

[1] Some iterative optimization algorithms, including LIPO, can make use of *a priori* information about the objective function such as points on the function to improve the optimization process.

or assumptions are made regarding the position of the optimal SNR threshold, constraining the search interval accordingly can be beneficial in terms of resources and/or reconstruction quality. However, constraints must be used cautiously, as they inherently risk inadvertently excluding the optimal SNR threshold.

As the results stand, the use of LIPO instead of grid search does not generally appear advantageous. Even disregarding the aforementioned benefit of grid search regarding the retroactive substitution of the quality assessment metric, LIPO is not a compelling choice, as it fails to keep up with grid search for optima at low SNR thresholds. For the proposed budgets, LIPO was observed to only consistently outperform a grid search of corresponding budget for optimal SNR thresholds roughly above 5. LIPO's poor performance for optima at low SNR thresholds is predominantly attributable to its search routine, in which the objective function is initially sampled before homing in on the maximum. For the proposed budgets, this initial sampling makes up a significant portion of the budget, leaving LIPO no room to close in on expensive optima. However, this indicates that for sufficiently large budgets, LIPO's comparative performance improves, even for optimal SNR thresholds near 1, since the cost of initial sampling is not as substantial relative to the entire budget. This claim is supported by empirical trials using a budget of approximately $5 \cdot K_0$.

Additional improvements to LIPO's performance could conceivably be made by incorporating frequency component count as a function of SNR threshold for the system matrix used in reconstruction (see figure 1.2) into LIPO's search algorithm, such that LIPO has *a priori* information about the cost of objective function evaluations. This could allow LIPO's search path to be biased toward high SNR thresholds and initially focus less on precision for low SNR thresholds, resulting in better optimization outcomes without a change to the budget. This suggestion notably is theoretically applicable to any iterative optimization algorithm. The choice of optimization algorithm is not restricted to LIPO or grid search and remains open to improvement.

# 6
# Conclusion

Based on the findings in this thesis, an SNR threshold optimization algorithm comprising parameters for quality assessment metric, optimization algorithm, and frequency component budget is proposed. The algorithm is designed to estimate an SNR threshold that optimizes reconstruction quality, thereby alleviating the need for manual selection.

# Acknowledgements

I would like to express great appreciation to my advisers, Dr. Martin Hofmann and Konrad Scheffler, for their constant guidance and support. I'd also like to extend my gratitude to all 10 volunteers from the IBI for their help in experimentation. Finally, I'm grateful to Prof. Dr.-Ing. Tobias Knopp for giving me this opportunity.

# 7 Appendix

| Point phantom iron oxide concentration (mmol/L) | lambda | iterations | minFreq | frames | recChannels | numAverages |
|---|---|---|---|---|---|---|
| 400 | 0.0005 | 3 | 80,000 | 301:1300 | 1:3 | 1000 |
| 200 | 0.003755 | 3 | 80,000 | 3087:4086 | 1:3 | 1000 |
| 100 | 0.01786 | 3 | 80,000 | 5873:6872 | 1:3 | 1000 |
| 50 | 0.055835 | 3 | 80,000 | 8658:9657 | 1:3 | 1000 |
| 25 | 0.1359 | 3 | 80,000 | 11444:12443 | 1:3 | 1000 |
| 12.5 | 0.2815 | 3 | 80,000 | 14229:15228 | 1:3 | 1000 |
| 6.25 | 0.5213 | 3 | 80,000 | 17015:18014 | 1:3 | 1000 |
| 3.12 | 0.8891 | 3 | 80,000 | 19800:20799 | 1:3 | 1000 |
| 1.56 | 1.424 | 3 | 80,000 | 22586:23585 | 1:3 | 1000 |
| 0.78 | 2.17 | 3 | 80,000 | 25371:26370 | 1:3 | 1000 |
| 0.39 | 3.1774 | 3 | 80,000 | 28157:29156 | 1:3 | 1000 |
| 0.2 | 4.5 | 3 | 80,000 | 30942:31941 | 1:3 | 1000 |
| **OpenMPIData** | 0.001 | 3 | 80,000 | 1:1000 | 1:3 | 1000 |

**Table 7.1:** Reconstruction parameters by phantom

| **Grid search point count** | **System matrix** | | | |
|---|---|---|---|---|
| | OpenMPIData | | Point phantom | |
| | Relative to $K_0$ | Absolute | Relative to $K_0$ | Absolute |
| 5 | 1.03 | 78003 | 1.1 | 83258 |
| 10 | 1.26 | 94989 | 1.38 | 104350 |
| 15 | 1.57 | 118972 | 1.68 | 127369 |
| 20 | 1.98 | 149616 | 2 | 151581 |

**Table 7.2:** Frequency component budgets by system matrix and points in grid search

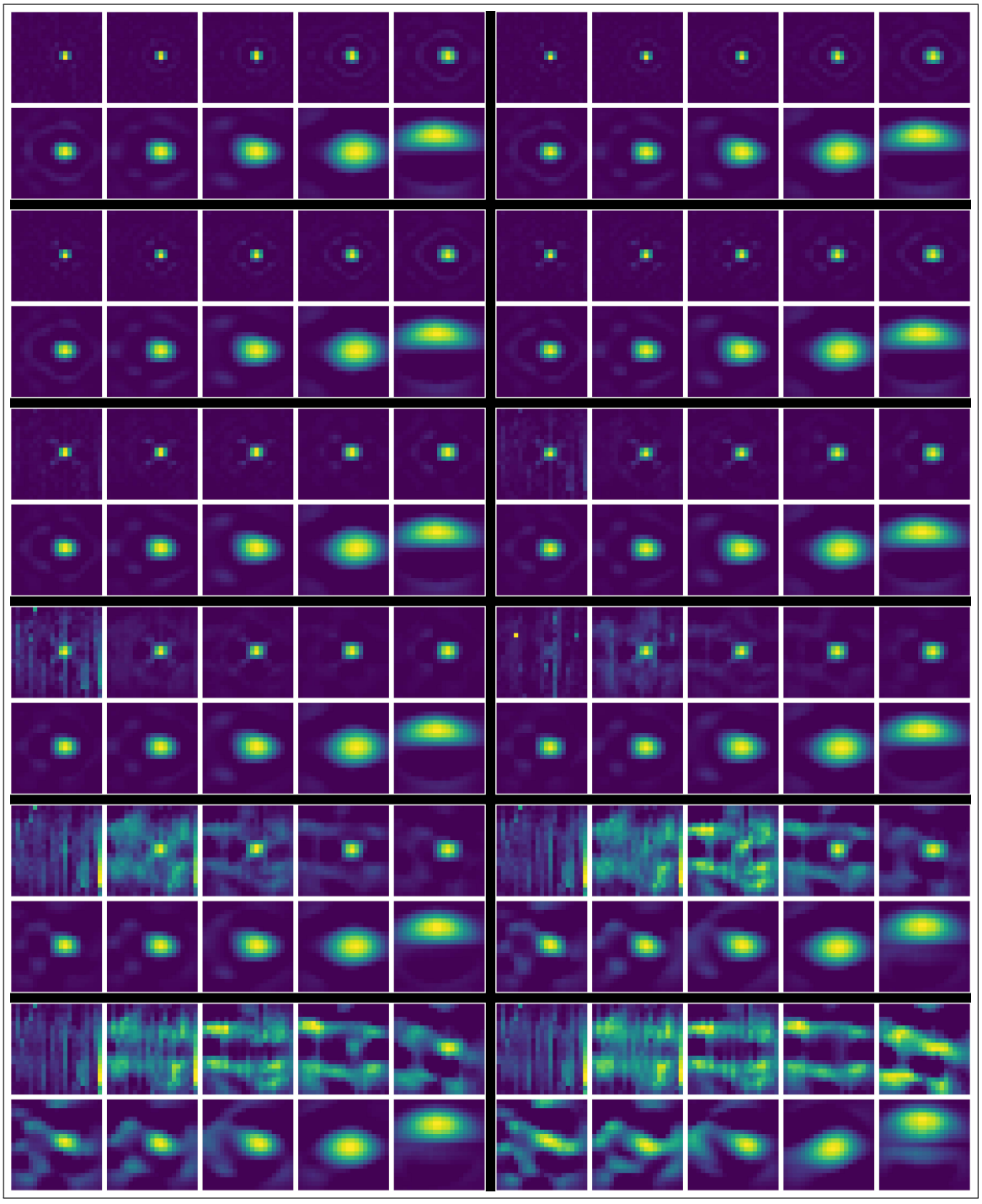


**Figure 7.1:** Point phantom grids sorted left-to-right then top-to-bottom by iron oxide concentration from [400, 200, 100, 50, 25, 12.5, 6.25, 3.12, 1.56, 0.78, 0.39, 0.2] mmol/L

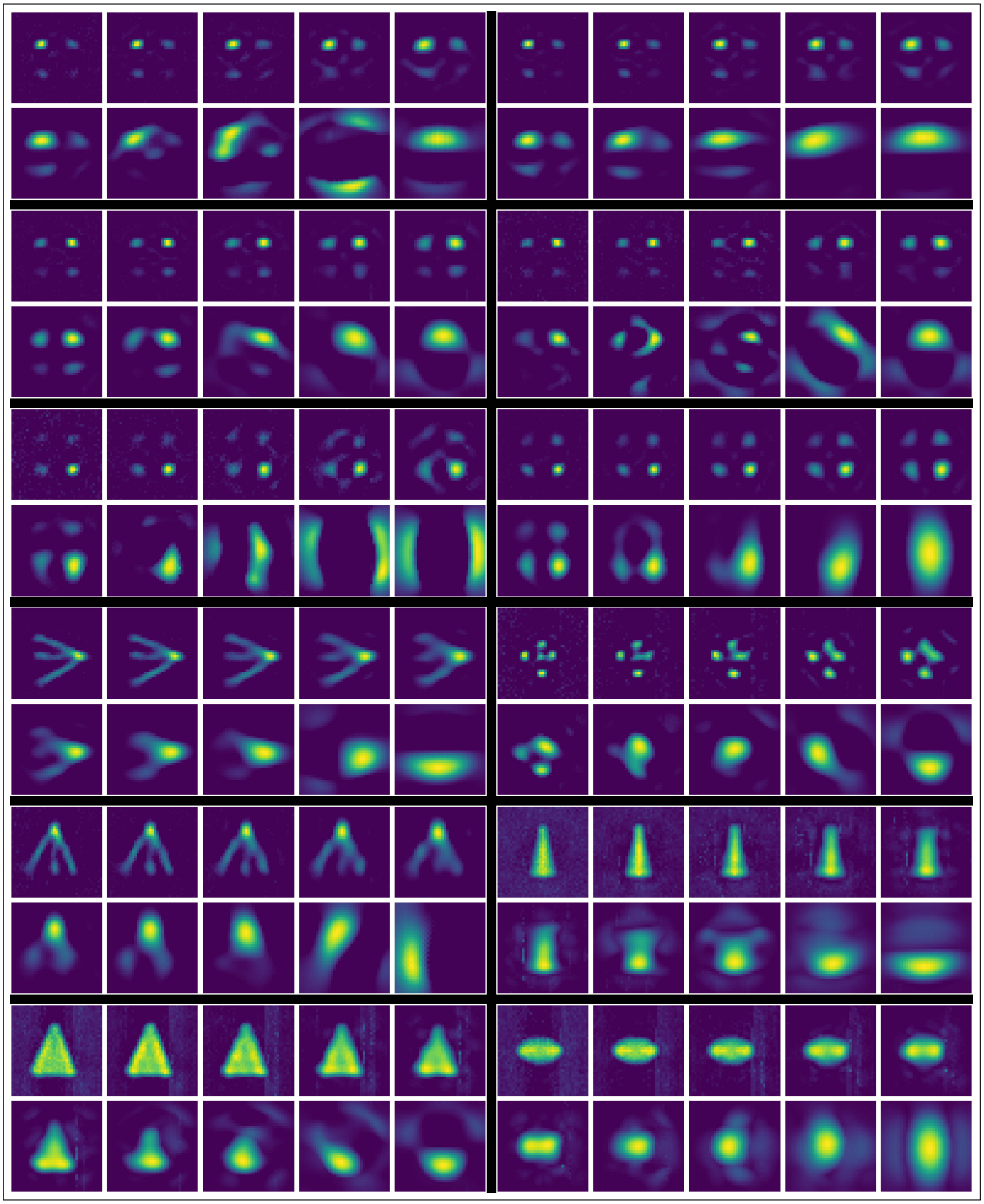


**Figure 7.2:** OpenMPIData phantom grids sorted left-to-right then top-to-bottom by phantom and slice from [Conc. xy[$\frac{1}{3}$], Conc. xy[$\frac{2}{3}$], Conc. xz[$\frac{1}{3}$], Conc. xz[$\frac{2}{3}$], Conc. yz[$\frac{1}{3}$], Conc. yz[$\frac{2}{3}$], Res. xy, Res. xz, Res. yz, Shape xy, Shape xz, Shape yz]

| **Iron oxide concentration (mmol/L)** | **Volunteer picks** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 400 | 1 | 1 | 1 | 1 | 1 | 4 | 3 | 1 | 1 | 1 |
| 200 | 1 | 1 | 1 | 1 | 1 | 4 | 2 | 1 | 1 | 1 |
| 100 | 1 | 1 | 1 | 1 | 2 | 4 | 3 | 1 | 2 | 1 |
| 50 | 1 | 1 | 1 | 1 | 2 | 4 | 3 | 2 | 1 | 2 |
| 25 | 2 | 2 | 1 | 2 | 3 | 3 | 3 | 2 | 2 | 4 |
| 12.5 | 3 | 4 | 2 | 4 | 3 | 3 | 4 | 3 | 4 | 4 |
| 6.25 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 3.12 | 5 | 4 | 4 | 5 | 4 | 5 | 5 | 5 | 5 | 5 |
| 1.56 | 5 | 5 | 4 | 7 | 5 | 6 | 5 | 5 | 6 | 6 |
| 0.78 | 6 | 8 | 5 | 9 | 5 | 7 | 6 | 7 | 6 | 6 |
| 0.39 | 8 | 9 | 7 | 9 | 7 | 7 | 8 | 9 | 7 | 9 |
| 0.2 | 9 | 9 | 9 | 9 | 9 | 9 | 8 | 9 | 6 | 9 |

| **Phantom & Slice** | **Volunteer picks** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Conc. xy[$\frac{1}{3}$] | 1 | 1 | 1 | 1 | 1 | 4 | 3 | 1 | 1 | 1 |
| Conc. xy[$\frac{2}{3}$] | 1 | 1 | 1 | 1 | 1 | 4 | 2 | 1 | 1 | 1 |
| Conc. xz[$\frac{1}{3}$] | 1 | 1 | 1 | 1 | 2 | 4 | 3 | 1 | 2 | 1 |
| Conc. xz[$\frac{2}{3}$] | 1 | 1 | 1 | 1 | 2 | 4 | 3 | 2 | 1 | 2 |
| Conc. yz[$\frac{1}{3}$] | 2 | 2 | 1 | 2 | 3 | 3 | 3 | 2 | 2 | 4 |
| Conc. yz[$\frac{2}{3}$] | 3 | 4 | 2 | 4 | 3 | 3 | 4 | 3 | 4 | 4 |
| Res. xy | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Res. xz | 5 | 4 | 4 | 5 | 4 | 5 | 5 | 5 | 5 | 5 |
| Res. yz | 5 | 5 | 4 | 7 | 5 | 6 | 5 | 5 | 6 | 6 |
| Shape xy | 6 | 8 | 5 | 9 | 5 | 7 | 6 | 7 | 6 | 6 |
| Shape xz | 8 | 9 | 7 | 9 | 7 | 7 | 8 | 9 | 7 | 9 |
| Shape yz | 9 | 9 | 9 | 9 | 9 | 9 | 8 | 9 | 6 | 9 |

**Table 7.3:** Human picks from sets of reconstructions

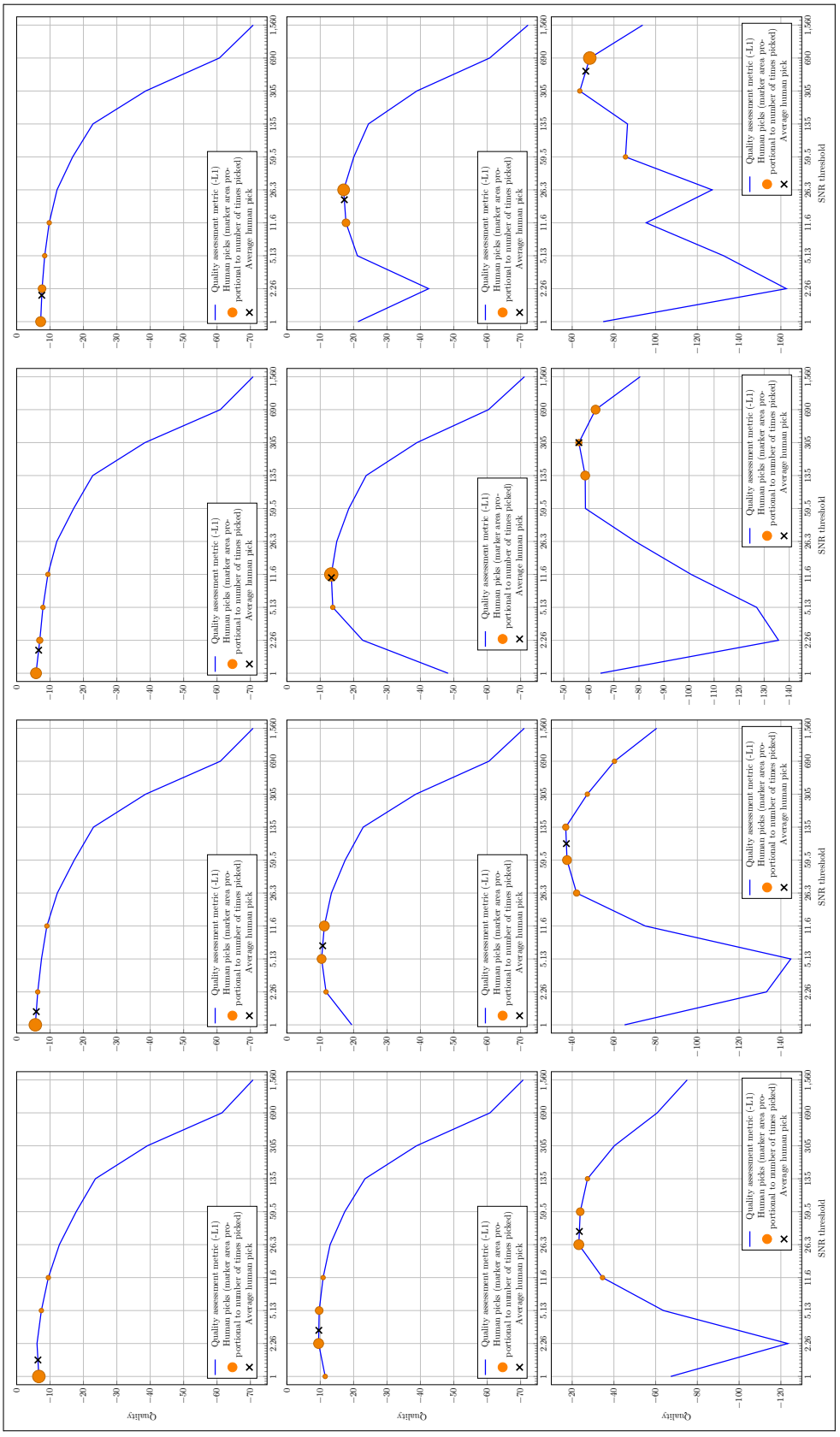



**Figure 7.3:** Human picks in the context of metric quality for point phantoms sorted left-to-right then top-to-bottom by iron oxide concentration from [400, 200, 100, 50, 25, 12.5, 6.25, 3.12, 1.56, 0.78, 0.39, 0.2] mmol/L

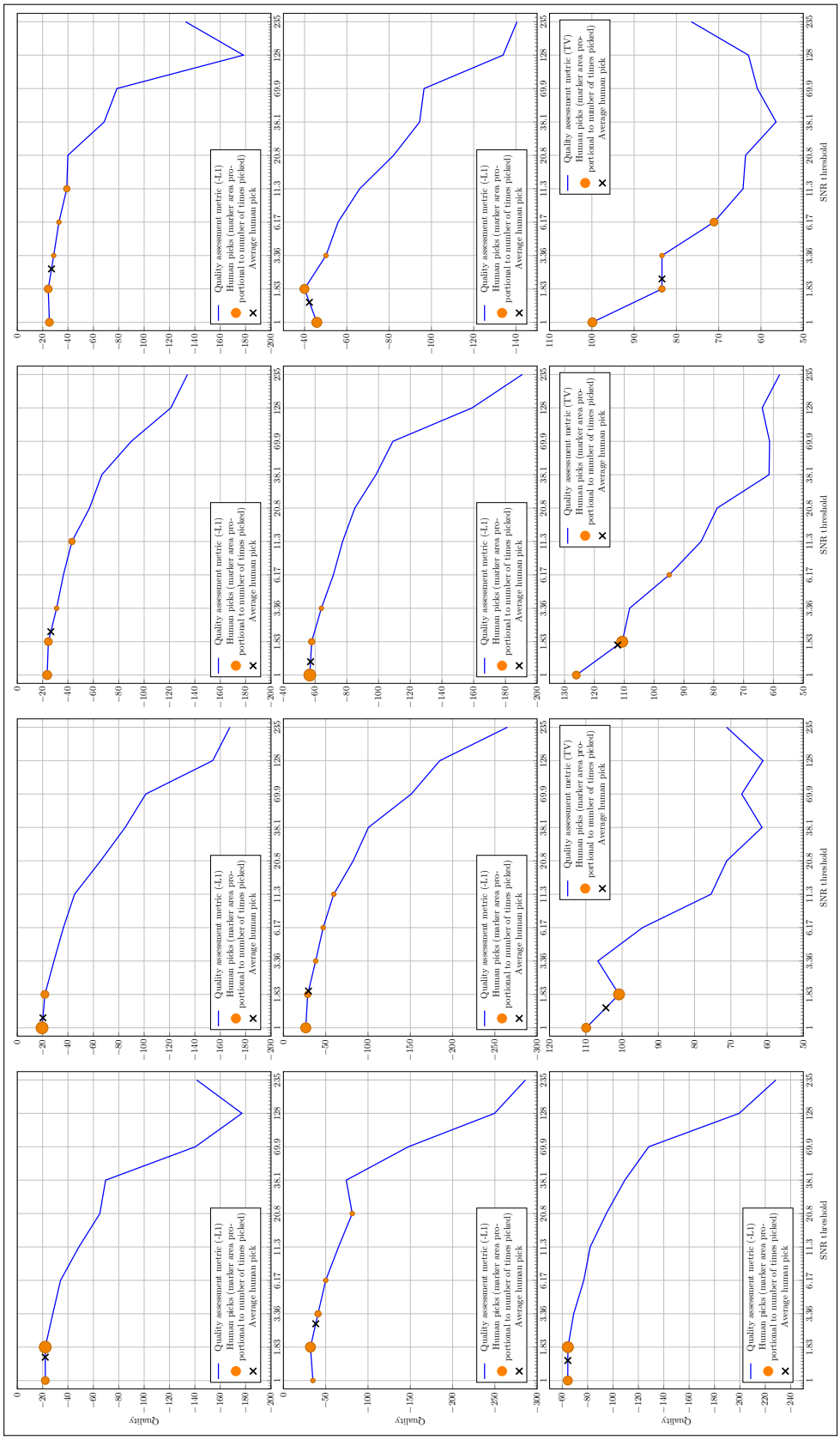



**Figure 7.4:** Human picks in the context of metric quality for OpenMPIData phantoms sorted left-to-right then top-to-bottom by phantom and slice from [Conc. xy[$\frac{1}{3}$], Conc. xy[$\frac{2}{3}$], Conc. xz[$\frac{1}{3}$], Conc. xz[$\frac{2}{3}$], Conc. yz[$\frac{1}{3}$], Conc. yz[$\frac{2}{3}$], Res. xy, Res. xz, Res. yz, Shape xy, Shape xz, Shape yz]

# Bibliography


[1] Tobias Knopp, Nadine Gdaniec, and Martin Möddel. "Magnetic particle imaging: from proof of principle to preclinical applications". In: *Physics in Medicine & Biology* 62.14 (2017), R124.

[2] Tobias Knopp and Thorsten M Buzug. *Magnetic particle imaging: an introduction to imaging principles and scanner instrumentation*. Springer Science & Business Media, 2012.

[3] Tobias Knopp et al. "Weighted iterative reconstruction for magnetic particle imaging". In: *Physics in medicine & biology* 55.6 (2010), p. 1577.

[4] Jeff Bezanson et al. "Julia: A fresh approach to numerical computing". In: *SIAM review* 59.1 (2017), pp. 65–98.

[5] Tobias Knopp et al. "MPIReco. jl: Julia package for image reconstruction in MPI". In: *International Journal on Magnetic Particle Imaging* 5.1-2 (2019).

[6] Jürgen Rahmer et al. "Analysis of a 3-D system function measured for magnetic particle imaging". In: *IEEE transactions on medical imaging* 31.6 (2012), pp. 1289–1299.

[7] Vipin Kamble and KM Bhurchandi. "No-reference image quality assessment algorithms: A survey". In: *Optik* 126.11-12 (2015), pp. 1090–1097.

[8] Marija Boberg et al. "Simultaneous imaging of widely differing particle concentrations in MPI: problem statement and algorithmic proposal for improvement". In: *Physics in Medicine & Biology* 66.9 (2021), p. 095004.

[9] Tobias Knopp et al. "OpenMPIData: An initiative for freely accessible magnetic particle imaging data". In: *Data in brief* 28 (2020), p. 104971.

[10] Tobias Knopp. *RegularizedLeastSquares.jl*. 2017. URL: https://github.com/tknopp/RegularizedLeastSquares.jl (visited on 09/06/2022).

[11] Cédric Malherbe and Nicolas Vayatis. "Global optimization of Lipschitz functions". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 2314–2323.

[12] Davis King. *A Global Optimization Algorithm Worth Using*. 2017. URL: http://blog.dlib.net/2017/12/a-global-optimization-algorithm-worth.html (visited on 09/06/2022).

[13] James F Peters. *Foundations of computer vision: computational geometry, visual image structures and object shape detection*. Vol. 124. Springer, 2017.