Appendix A. Additional experiments and results

We performed several additional experiments to study our algorithms. Results from these experiments are presented here.

Appendix A.1. Sample size experiment

For this experiment, we generated a synthetic dataset using the mean shape and SSM from the pelvis dataset. We deformed the mean pelvis shape by known shape parameters sampled within ±3 SD, sampled oriented points from the deformed shapes, and then applied known transformations within realistic intervals to the sampled points. Experiments were run with 500, 1000, 1500 and 2000 sample points to evaluate the performance of our algorithms with increasing number of samples.

For each algorithm, 3 sets of experiments were run with 0, 5, and 10 modes used to deform the mean shape. In this experiment, the same number of modes are used by our algorithms to estimate the deformed shape as were used to deform the mean shape in order to evaluate our algorithms' performance with different modes without bias. As before, when 0 modes are used, the algorithms used are the corresponding non-deformable algorithms performing rigid registration between the mean shape and points sampled from it. 10 registrations were performed in each set with known transformations sampled from the intervals $[0, 15]$ mm and $[0, 9]^\circ$ for translational and rotational offsets, respectively, and applied to points sampled from the deformed shapes. Noise was added to both the position and orientation of the sampled points, and two experiments were designed based on different noise models.

Appendix A.1.1. Experiment 1: Isotropic position noise

In this experiment, an isotropic noise model with SD of $1 \times 1 \times 1\,\text{mm}^3$ for positional noise was used to generate samples. Additionally, an isotropic noise model with SD of 2° for angular noise was used for D-IMLOP. For GD-IMLOP, the anisotropic angular noise model also had a SD of 2° and e was set to 0.5. The noise model assumed by each of the algorithms was the same as the noise model used to generate samples for each of the algorithms.

All three of our algorithms produced small errors in recovering the shape and registering the sampled points to the recovered shape with the different number of samples. D-IMLOP and GD-IMLOP outperformed D-IMLP due to the added information provided by the normals (Fig. A.15). Over all runs, D-IMLP produced a mean tRE of $0.71\,(\pm 0.80)$ mm, while D-IMLOP and GD-IMLOP produced mean tREs of $0.30\,(\pm 0.32)$ mm ($p < 0.001$ compared to D-IMLP) and $0.29\,(\pm 0.35)$ mm ($p < 0.001$ compared to D-IMLP), respectively. D-IMLOP slightly outperformed GD-IMLOP in recovering rotation. This is because D-IMLOP solves a relatively simpler objective function since it only models isotropic orientation noise. Since the orientation noise in the samples used for D-IMLOP is also isotropic, the algorithm is able to converge toward the correct rotation quickly. For GD-IMLOP, however, the objective function is more complicated since it models anisotropic orientation noise making GD-IMLOP slower to converge to the correct rotation in some cases.

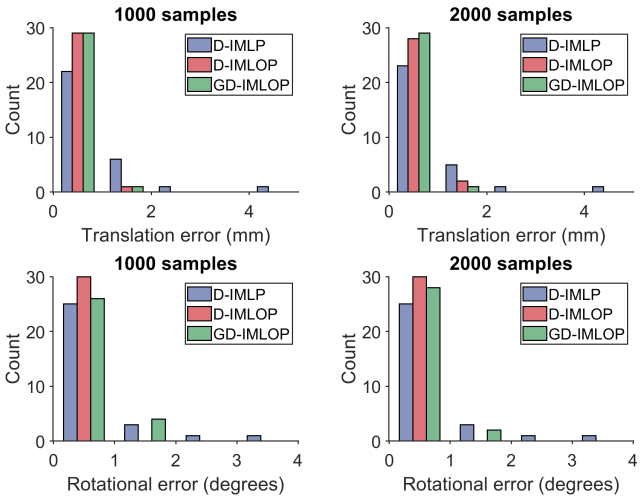

Fig. A.15. Sample size experiment: translation (top) and rotation (bottom) errors produced using (from L to R) 1000 and 2000 sample points from the pelvis model in Exp. 1. Bars in the histogram are transparent to show all three algorithms.

Further, metrics produced by our algorithms, like the objective function (total match error) or the residual error (Mahalanobis distance), show correlation with the tRE and can be used to assign confidence to the computed registration. We used empirically chosen thresholds to determine which trials succeeded and which did not using the residual error. We do this retrospectively in the simulated experiments because we have access to ground truth and can use it to learn how to associate residuals errors with success in clinical or other experiments where ground truth in not available. Using thresholds such that there were no false positives, our algorithms were always able to correctly detect successful registrations and showed an increase in successful registrations with increasing number of sample points (Fig. A.15, Table A.1), where success is defined as registrations producing tRE less than 1.5 mm for the pelvis dataset. In fact, all D-IMLOP and GD-IMLOP trials using greater than 500 sample points produced successful registrations.

Appendix A.1.2. Experiment 2: Anisotropic position noise

Second, an anisotropic noise model with 1 mm SD in each direction in plane and 2 mm out of plane was used for positional noise (or, $1 \times 1 \times 2\,\text{mm}^3$). For angular noise, the parameters were the same as in Exp. 1.

The tREs from this experiment were higher than those in Exp. 1 but by less an 0.1 mm for all algorithms. Most D-IMLP trials and all D-IMLOP and GD-IMLOP trials produced successful registrations (Table A.1). D-IMLOP performs slightly better than GD-IMLOP in recovering rotation due to the same reason as in Exp. 1 (Fig. A.16). However, GD-IMLOP is able to achieve smaller tREs than D-IMLOP since it outperforms D-IMLOP in recovering translation and shape parameters (Fig. A.17). Again, over all runs, D-IMLP produced a mean tRE of $0.76\,(\pm 0.75)$ mm, while D-IMLOP and GD-IMLOP produced mean tREs of $0.39\,(\pm 0.33)$ mm ($p < 0.001$ compared

Table A.1. Sample size experiment: percent successful trials, i.e., trials producing registrations with tRE less than 1.5 mm. All successful trials were correctly detected as successful.

		Algorithm (success in %)		
	samples	D-IMLP	D-IMLOP	GD-IMLOP
Exp. 1	500	83.33	96.67	96.67
	1000	86.67	100.00	100.00
	1500	83.33	100.00	100.00
	2000	86.67	100.00	100.00
Exp. 2	500	76.67	100.00	100.00
	1000	76.67	100.00	100.00
	1500	76.67	100.00	100.00
	2000	76.67	100.00	100.00

to D-IMLP) and 0.33 (± 0.28) mm ($p < 0.001$ compared to D-IMLP), respectively. An improvement with increasing number of sample points was observed in this experiment as well (Fig. A.17). Note that errors show some increase with increasing number of modes because, as mentioned in the experiment setup, number of modes used by our algorithms to estimate the deformed target shape was the same as the number of modes used to generate the target shape. The residual errors from all three algorithms were, again, found to be correlated with the tRE (Fig. A.19) and, therefore, can be used to assign confidence to registrations. Using empirically found thresholds, we used the residual errors produced by our algorithms to automatically classify trials as successful or unsuccessful and again found that our algorithms were always able to correctly classify successful registrations (Table A.1).

Since D-IMLOP and GD-IMLOP are able to filter out matches where positions align but orientations do not, we see faster improvement in the quality of matched points with these two algorithms than with D-IMLP. (Fig. A.18). Here, the error shown is the distance at each iteration between the matched points computed by each algorithm and the *true* location of the matched point. This true location is simply the location of the points sampled from the deformed model shape before any noise or transformation was applied. This illustrates the improvement afforded by orientation information.

Appendix A.2. Noise model experiment

This experiment was designed to evaluate the stability of our algorithms with different noise models. A synthetic dataset was generated using the pelvis data in a similar way as described in Sec. Appendix A.1. The differences are that experiments in this section are run with a fixed sample size of 500, and for each algorithm, 11 sets of 25 experiments each are run with increasing number of modes used to deform the mean shape in each set, starting at 0 and going up to 10 modes. Again, the same number of modes are used by our algorithms to recover the deformed shape as were used to generate the deformed shape. Different noise models were used to add noise to both the position and orientation of the sampled points, and the same noise models were assumed by our algorithms. Four experiments were designed based on how the different noise models were varied.

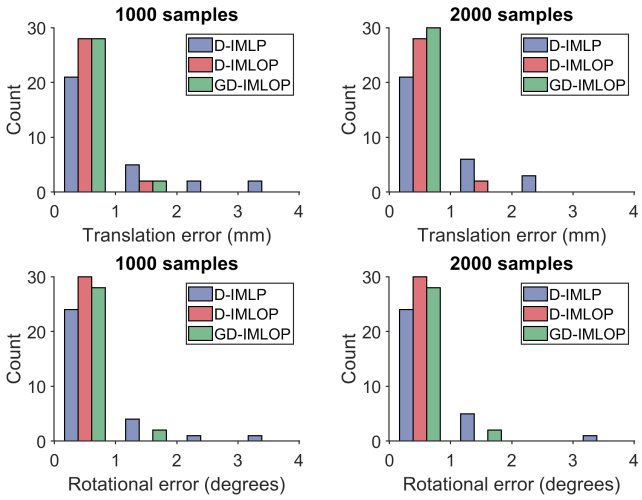

Fig. A.16. Sample size experiment: translation (top) and rotation (bottom) errors produced using (from L to R) 1000 and 2000 sample points from the pelvis model in Exp. 2. Bars in the histogram are transparent to show all three algorithms.

Appendix A.2.1. Experiment 1: Varying isotropic position noise

For the first experiment, we used 5 isotropic noise models with SDs of $1 \times 1 \times 1\,\text{mm}^3$, $2 \times 2 \times 2\,\text{mm}^3$, $3 \times 3 \times 3\,\text{mm}^3$, $4 \times 4 \times 4\,\text{mm}^3$, and $5 \times 5 \times 5\,\text{mm}^3$ for positional noise. For D-IMLOP, an isotropic noise model with SD of 2° for angular noise was used, while an anisotropic noise model with SD of 2° and $e = 0.5$ was used for GD-IMLOP.

We observed that registrations produced by our algorithms degraded as the noise in the sample points increased, but tREs also showed improvement going from D-IMLP to D-IMLOP to GD-IMLOP (Fig. A.20, top row). Over all registrations, D-IMLP produced a mean tRE of 1.55 (± 0.96) mm, while D-IMLOP and GD-IMLOP produced mean tREs of 0.75 (± 0.60) mm ($p < 0.001$ compared to D-IMLP) and 0.40 (± 0.22) mm ($p < 0.001$ compared to both D-IMLP and D-

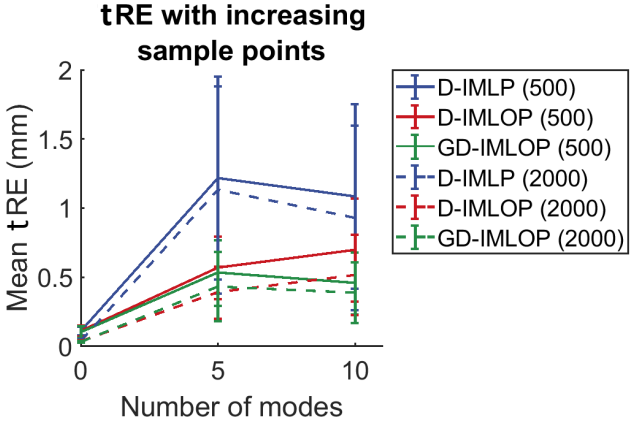

Fig. A.17. Sample size experiment: mean tRE with increasing number of samples from Exp. 2 (shown with 500 and 2000 sample points for simplicity).

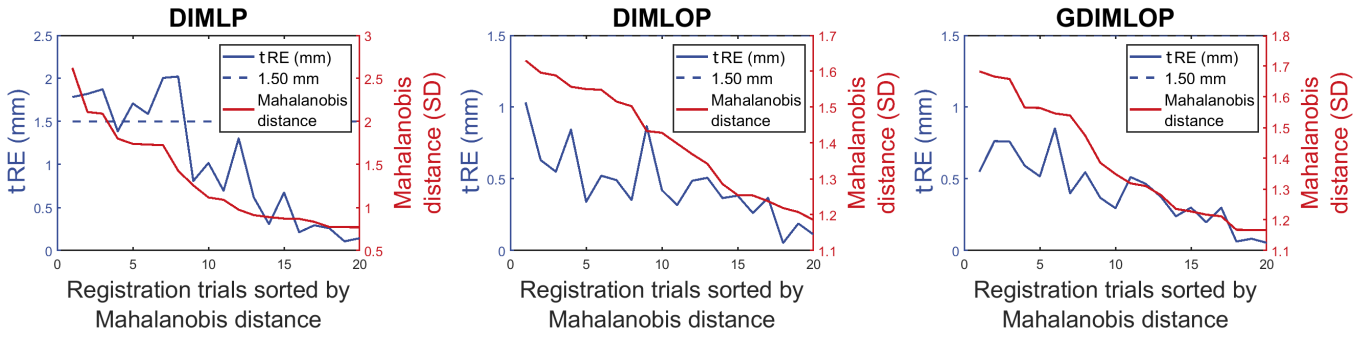

Fig. A.19. Residual errors compared against tRE using 2000 sample points in Exp. 2 of the sample size experiment. The two measures exhibit correlation with correlation coefficients (left to right) of 0.85, 0.72 and 0.86. Results from trials using 0 modes are ignored here to focus on the deformable algorithms.

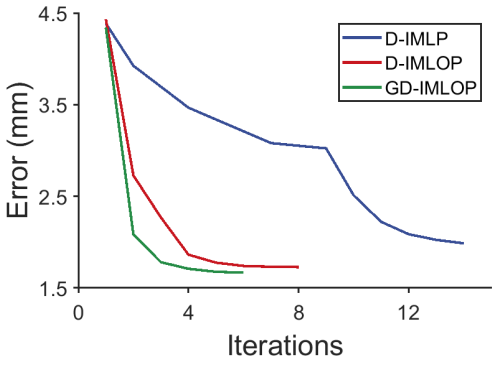

Fig. A.18. Sample size experiment: evolution of the quality of matched points for each algorithm with increasing iterations in Exp. 2 with 2000 sample points. The plot shows distances between matched points at each iteration and the location of points sampled from the deformed model shape. Added orientations drastically improve the quality of matched points.

IMLOP), respectively. Further, while noise increased 125× (5× along each dimension), tREs only degraded 1.78× at 1 mode and 2.18× at 10 modes for D-IMLP, 1.48× at 1 mode and 2.33× at 10 modes for D-IMLOP, and 1.44× at 1 mode and 1.26× at 10 modes for GD-IMLOP. By compensating for the noise in the samples, our algorithms are able to limit the effects of increasing noise on the registration.

The tREs increase slightly as we add more modes because, as described in the experiment setup, the number of modes used by our algorithms to estimate the deformed target shape was the same as the number of modes used to generate the deformed target shape. That is, for 0 modes, points were sampled from the mean shape without any deformation and no shape parameters were used in the optimization, resulting in the rigid versions of our algorithms. Whereas when 10 modes are used, the mean shape is deformed along 10 mode directions and 16 parameters are optimized by our algorithms (10 shape parameters as well as 3 rotation and 3 translation parameters). The objective function and the residual errors are again found to be strongly correlated with the tRE, which again can be used to distinguish between successful and unsuccessful registrations (Fig. A.21).

Appendix A.2.2. Experiment 2: Varying anisotropic position noise

For the second experiment, anisotropic noise models with SDs of $1\times1\times2\,\text{mm}^3$, $2\times2\times3\,\text{mm}^3$, $3\times3\times4\,\text{mm}^3$, $3\times3\times5\,\text{mm}^3$, and $4\times4\times5\,\text{mm}^3$ for positional noise were used. For angular noise, the parameters were the same as in Exp. 1.

Results from this experiment show the same trends as those from Exp. Appendix A.2.1 (Fig. A.20, middle row). Over all runs, D-IMLP produced a mean tRE of 1.97 (± 1.06) mm, while D-IMLOP and GD-IMLOP produced mean tREs of 0.80 (± 0.56) mm ($p < 0.001$ compared to D-IMLP) and 0.38 (± 0.20) mm ($p < 0.001$ compared to both D-IMLP and D-IMLOP), respectively. Additionally, while noise increased 40× (4× along each dimension in plane and 2.5× out of plane), the tREs only degraded 1.08× at 1 mode and 1.29× at 10 modes for D-IMLP, 1.54× at 1 mode and 1.78× at 10 modes for D-IMLOP, and 1.12× at 1 mode and 1.21× at 10 modes for GD-IMLOP.

Appendix A.2.3. Experiment 3: Varying orientation noise

In the third experiment, we used one isotropic noise model and one anisotropic noise model with SDs of $1\times1\times1\,\text{mm}^3$ and $1\times1\times2\,\text{mm}^3$, respectively, for positional noise, and used orientation noise models with SDs of 2°, 4°, 6°, 8° and 10° for each positional noise model. $e = 0.5$ was used for GD-IMLOP.

We observe that all D-IMLP runs with isotropic positional noise produce similar results and all runs with anisotropic positional noise also produce similar results. This is expected since D-IMLP does not take any orientation information into consideration, so varying the orientation noise model has no statistically significant effect on results from D-IMLP. Further, as seen in the previous two experiments, mean tREs from trials with isotropic positional noise (1.04 ± 0.85 mm) are statistically significantly lower ($p < 0.001$) than those from trials with anisotropic noise (1.74 ± 1.14 mm) (Fig. A.20, bottom-left). The difference in tREs between runs with isotropic and anisotropic positional noise was smaller for D-IMLOP (0.53 ± 0.42 mm and 0.61 ± 0.42 mm, respectively) and GD-IMLOP (0.51 ± 0.42 mm and 0.58 ± 0.46 mm, respectively), although still statistically significant ($p < 0.001$). As with D-IMLP, changing angular noise did not affect registration results from D-IMLOP and GD-IMLOP statistically significantly since the large number of samples overwhelmed the relatively small changes in the noise model (Fig. A.20, bottom-middle and bottom-right).

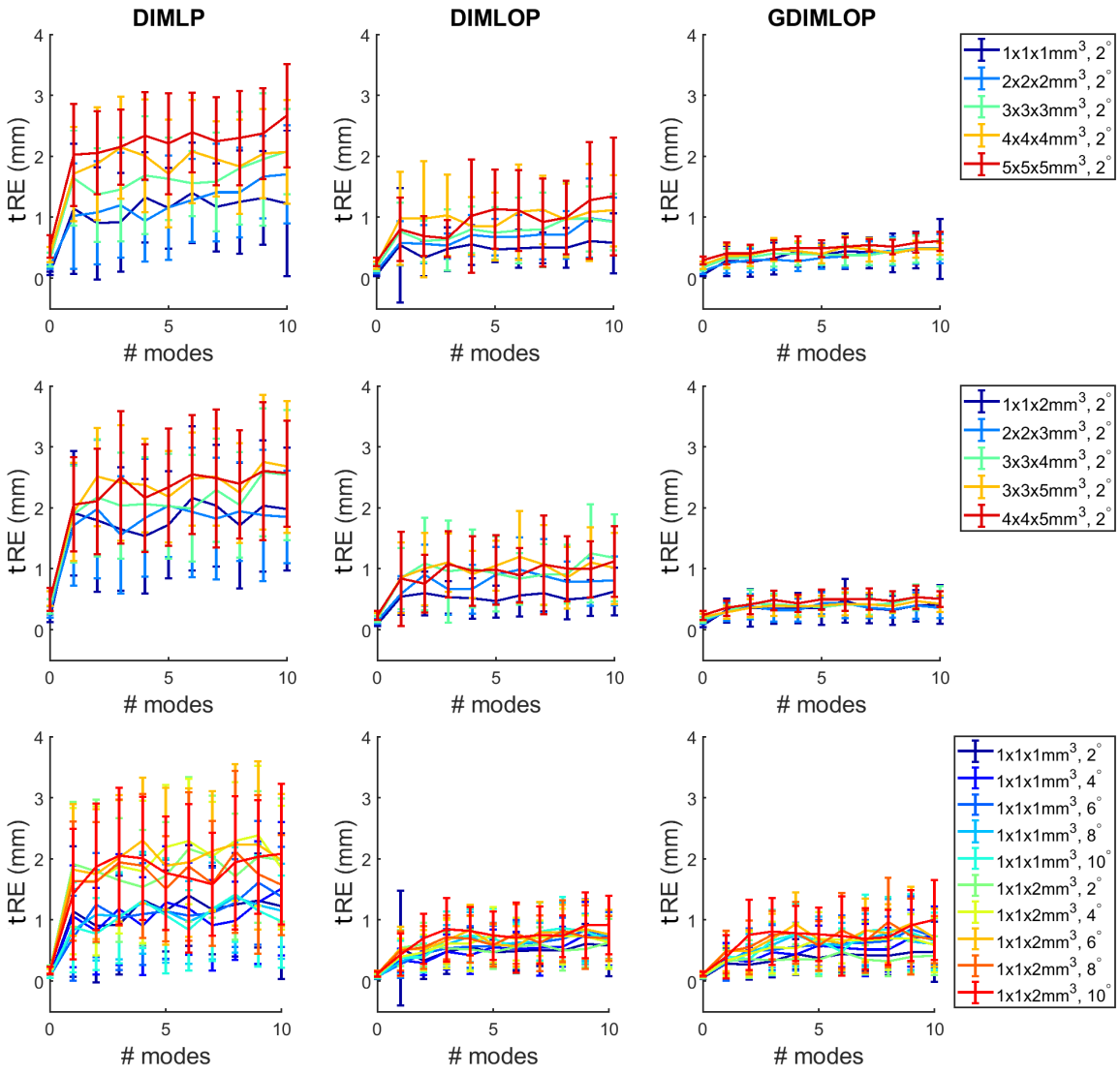

Fig. A.20. Noise model experiment: mean tREs produced by our algorithms (L-R) D-IMLP, D-IMLOP and GD-IMLOP in Exps. 1 (top), 2 (middle) and 3 (bottom). Note that the errors are increasing with increasing modes only because for this experiment the number of modes used to estimate the shapes equals the number of modes used to simulate a new shape from which points were sampled.

Appendix A.2.4. Experiment 4: Noise parameter sweep

In the final experiment, the sample points are generated with a particular noise model for both position and orientation data. However, we assume that this noise model is unknown to our algorithms. Sample points are generated with anisotropic positional and angular noise with SD $2{\times}2{\times}4\,\text{mm}^3$ and 10° ($e = 0.5$), respectively. We perform a hyper-parameter sweep and run our algorithms with different isotropic and anisotropic positional and angular noise assumptions to see how well our algorithms perform with inaccurate noise model assumptions.

This experiment was performed to observe the behavior of our algorithms when the noise model assumed is different from the noise in the sampled points. Since the noise in the generated sampled points, anisotropic in both position and orientation ($2 \times 2 \times 4\,\text{mm}^3$ and 10° ($e = 0.5$), respectively), can be best explained by GD-IMLOP, we expect it to outperform the other two algorithms. Interestingly, D-IMLOP (red) outperforms GD-IMLOP (green) with less conservative noise estimates (Fig. A.23). This is expected since D-IMLOP optimizes a simpler cost function. Therefore, when the noise assumption is optimistic, D-IMLOP converges faster than GD-IMLOP. However, as the noise assumption becomes more conservative, GD-IMLOP's performance improves or stabilizes while D-IMLOP's deteriorates since GD-IMLOP models the noise in the sample points more accurately. Here, the tRE values are less important than the trends shown by D-IMLOP and GD-IMLOP. D-IMLOP performs well with less conservative noise models since the noise in the samples is close enough to the least conservative noise model assumed by D-IMLOP. However, with larger sample noise, less conservative noise estimates will introduce errors in matches. The improving trend with increasingly conservative noise estimates exhibited by GD-IMLOP allows us to make very conservative noise estimates when sample noise is unknown and still expect reasonable registration results.

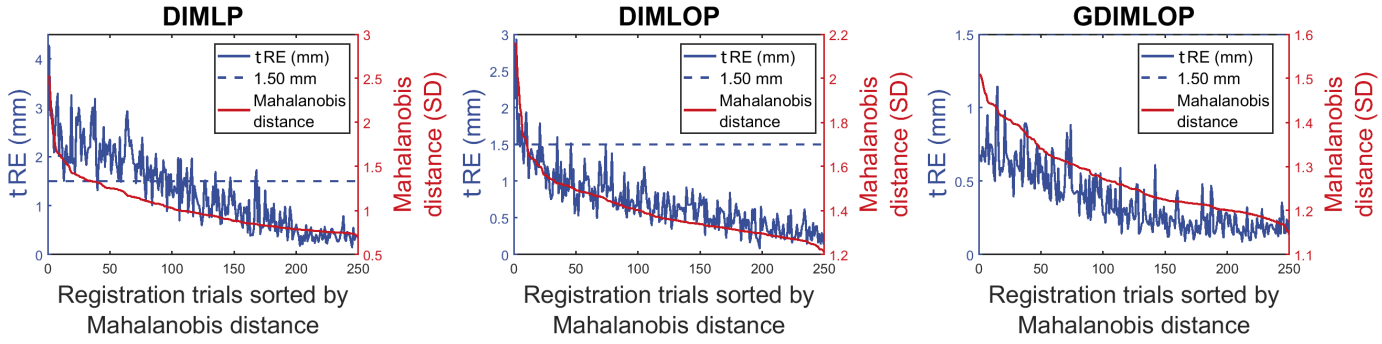

Fig. A.21. Residual errors compared against tRE using 500 sample points with $2 \times 2 \times 2\,\text{mm}^3$ SD positional noise and 2° SD angular noise in Exp. 1 of the noise model experiment. The two measures exhibit correlation with correlation coefficients (L-R) of 0.86, 0.88 and 0.83.

Finally, we see that D-IMLP (blue) is unaffected by changing orientation noise, which is expected since D-IMLP does not take orientation into account. tREs using D-IMLP are either stable or show a gradual trend downward as position noise becomes more conservative. The other noticeable trend shows that D-IMLP performs slightly worse as the anisotropy in the noise estimates increases. This trend is slightly visible in the curves for D-IMLOP and GD-IMLOP as well, although they are not as noticeable since the orientation information available to these algorithms is able to drive the errors down considerably.

Appendix A.3. Outlier experiment

Using a synthetic dataset generated using the right nasal cavity model from the sinus dataset, we study how robust our algorithms are to outliers. This dataset was again generated similarly as described before (Sec. Appendix A.1); the difference being that in this setup, there are 6 sets of 10 experiments each for all three algorithms. The number of modes used to deform the mean shape increases by 2 in each set, starting at 0 and going up to 10 modes. Again, the same number of modes are used to estimate the deformed target shape using our algorithms as were used to generate the deformed target shape.

All sample points were generated with isotropic noise in position data with $1 \times 1 \times 1\,\text{mm}^3$ SD, and anisotropic noise in orientation data with 2° SD and $e = 0.5$. Experiments were conducted with 0%, 10%, and 20% outliers in the generated point samples. Outliers are generated by perturbing the position and orientation of a particular number of samples randomly in the range $[2, 5]$ mm and $[2, 5]°$, respectively.

Outliers are identified and rejected using the chi-square test, in the same way as in the corresponding rigid algorithms described earlier (Billings et al., 2015; Billings and Taylor, 2015). Under the assumption of correspondences and generalized Gaussian noise, the square Mahalanobis distance between the matched points in 3D space can be assumed to distributed as the sum of squares of three independent Gaussian distributions, each representing one dimension of position data (Danilchenko and Fitzpatrick, 2011). Therefore, a match is rejected if the square Mahalanobis distance is greater than the chi-square inverse cumulative density function with 3 degrees of freedom at $p = 0.95$. For orientation data, a match that passes the outlier test based on the position component can still be rejected if $\hat{\mathbf{y}}_{\mathbf{n}_i}^{\mathbf{T}} \hat{\mathbf{x}}_{\mathbf{n}_i} < \cos(\theta_{\text{thresh}})$, where θ_{thresh} is set according to the circular SD. We set our threshold to 3 times the circular SD.

Although the performance of our algorithms is worse in the presence of outliers, we are able to detect them, as explained above (Sec. Appendix A.3), and limit their effect on errors. The degradation in performance as outliers increase 2× from 10% to 20% is, at worst, 1.45× (using 4 modes), 1.72× (using 2 modes) and 1.42× (using 4 modes) for D-IMLP, D-IMLOP and GD-IMLOP, respectively. At best, the performance is almost identical for all algorithms: 1.12×, 0.97× and 1.10× for D-IMLP, D-IMLOP and GD-IMLOP, respectively, all using 10 modes (Fig. A.22). As outliers increase from 0% to 20%, the degradation in performance is, at worst, 2.81×, 2.18× and 2.53× for D-IMLP, D-IMLOP and GD-IMLOP, respectively, all using 4 modes. At best, the performance degrades 1.79× (using

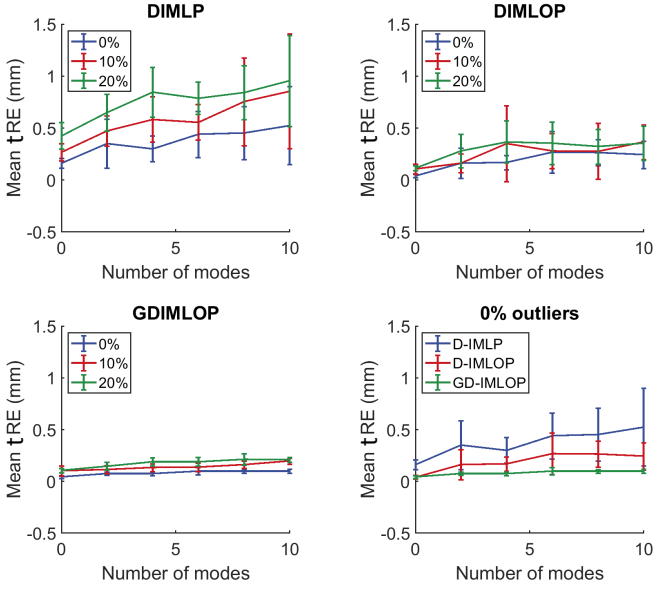

Fig. A.22. Outlier experiment: mean tRE with different number of outliers for D-IMLP (top-left), D-IMLOP (top-right), and GD-IMLOP (bottom-left), and for all three algorithms using sample points with 0% outliers (bottom-right). Note that the errors are increasing with increasing modes only because for this experiment the number of modes used to estimate the shapes equals the number of modes used to simulate the deformed shape from which points were sampled.

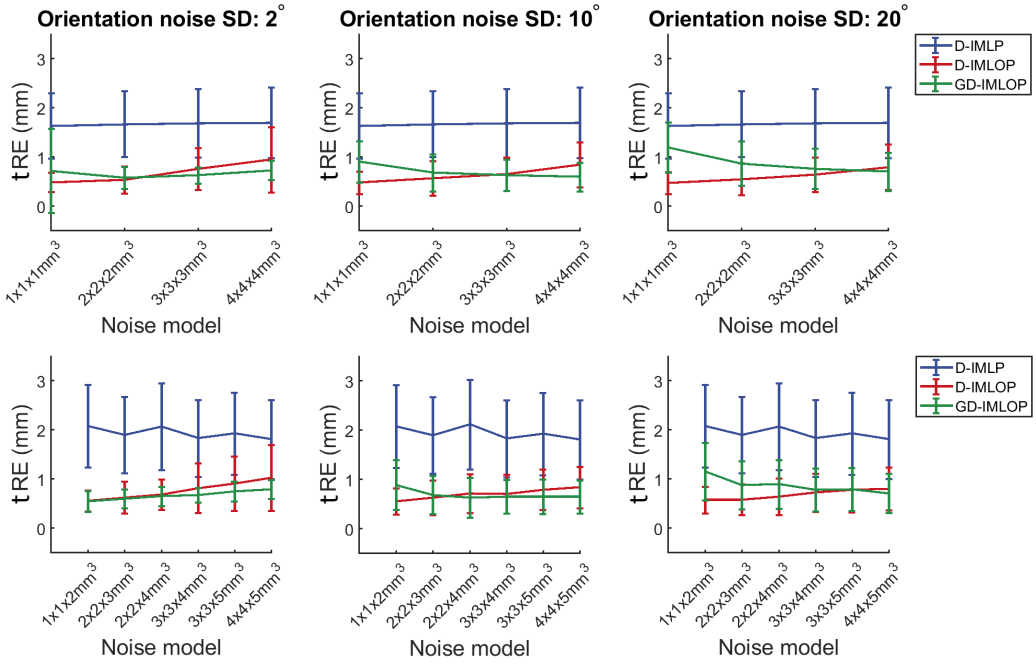

Fig. A.23. Noise model experiment: mean tREs produced by our algorithms with different isotropic (top) and anisotropic (bottom) position noise assumptions, labeled on the x-axis, and different orientation noise assumptions, with standard deviations of 2° (left), 10° (middle) and 20° (right) in Exp. 4.

6 modes), 1.21× (using 8 modes) and 1.93× (using 6 modes) for D-IMLP, D-IMLOP and GD-IMLOP, respectively. However, we cannot compare this to the degradation in the quality of samples points since the initial set had 0% outliers (making any increase an ∞ increase).

Although the addition of outliers produced statistically significant increases in tREs for D-IMLP and GD-IMLOP, the increase in tREs from 0% to 10% outliers and from 10% to 20% outliers was not statistically significant for D-IMLOP. However, the increase in tREs from 0% to 20% outliers was statistically significant for D-IMLOP as well (Table A.2). Further, Fig. A.22 (bottom-right) shows that without any outliers in the sampled point set, GD-IMLOP performs best, followed by D-IMLOP, and then D-IMLP (as seen in previous experiments). Fig. A.22 (bottom-right) also shows a very strong result in that although with 10 modes, GD-IMLOP must optimize over 10 extra parameters, the degradation in the tRE is negligible (~ 0.05 mm) with 0% outliers (green).

Table A.2. Outlier experiment: mean tREs produced by each of our algorithms with 0%, 10%, and 20% outliers in the samples points. * indicates statistically significant increase in tREs between 0% and 10% outliers; † indicates statistically significant increase in tREs between both 10% and 20% outliers and 0% and 20% outliers; ‡ indicates statistically significant increase in tREs between 0% and 20% outliers ($p < 0.001$).

	Algorithm (tREs in mm)		
outliers	D-IMLP	D-IMLOP	GD-IMLOP
0%	0.37 ± 0.25	0.19 ± 0.15	0.08 ± 0.03
10%	0.58 ± 0.36*	0.26 ± 0.23	0.14 ± 0.05*
20%	0.75 ± 0.30†	0.30 ± 0.18‡	0.18 ± 0.05†

Appendix A.4. Scale experiment

In this experiment, we evaluate how well our algorithm is able to recover scale in addition to rotation, translation, and shape parameters. We use the same dataset that was generated in Sec. Appendix A.3 with 0% outliers. However, the sample points are scaled by some known amount in the range [0.7, 1.3].

Results show that our methods can successfully estimate scale in addition to rotation, translation, and shape parameters. Although our methods perform better when there is one fewer parameter to optimize over (Table A.3), with an additional scale parameter, tSEs using all three algorithms and tREs using D-IMLOP and GD-IMLOP still remain consistently below 1 mm (Fig. A.24, left and middle). Errors in recovering scale also reflect D-IMLOP and GD-IMLOP's performance, with mean errors ~ 0.01 and SD < 0.01 (Fig. A.24, right).

Appendix A.5. Non-medical data experiment

Our previous experiments test the generalizability of our algorithms within the medical field. With the following experiment, we test our algorithms on non-medical data to test their

Table A.3. Scale experiment: mean tREs produced by each of our algorithms with and without scale. * indicates statistically significant ($p < 0.001$) increase in tREs when optimizing over an additional scale parameter.

	Algorithm (tREs in mm)		
	D-IMLP	D-IMLOP	GD-IMLOP
w/o scale	0.37 ± 0.25	0.19 ± 0.15	0.08 ± 0.03
w scale	1.11 ± 0.45*	0.68 ± 0.34*	0.10 ± 0.03*

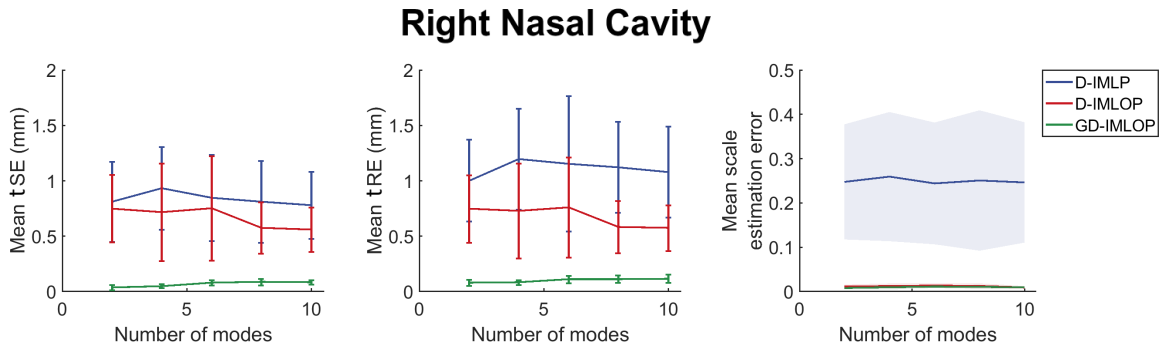

Fig. A.24. Scale experiment: Mean tSE (left) and tRE (middle) using sample points with 0% outliers and scale applied to the sampled points, and mean errors in recovering scale with increasing number of modes (right). Again, the number of modes used to estimate the shapes equals the number of modes used to simulate the deformed shape from which points were sampled.

generalizability outside the medical field. We use a human expression dataset in a leave-n-out experiment by dividing the dataset into a training set and a test set. We use the training set to build a shape model, and estimate the meshes in the test set using the two methods described in Sec. 4.1.

We used 300 meshes in the training set to build an SSM for expressions from a single individual, and tested with the remaining 86 meshes in the test set, also from the same individual. 1000 points were sampled from meshes in the test set with anisotropic position and orientation noise with SDs $1\times1\times2$ mm^3 and 10° ($e = 0.5$), respectively. This simulates a realistic situation in which a scan of a head is obtained using a depth camera, where error is large in the depth direction. Our algorithms make slightly more relaxed noise assumptions, assuming that the position and orientation noise models to have SDs of $2\times2\times4$ mm^3 and 20° ($e = 0.5$), respectively.

Results from this experiment show that our algorithms perform relatively well even with challenging datasets. The assumption that facial expressions are Gaussian distributed is likely an incorrect assumption depending on the dataset (Buciu et al., 2008). Further, the limited number of data points in our dataset was not enough to explain well the complex variations that can exist in human expression. However, our algorithms produced promising results with tSEs and tREs consistently below 1 mm and 2 mm, respectively (Fig. A.25). Over all runs, D-IMLP produced a mean tRE of 1.78 (± 0.63) mm, D-IMLOP produced mean tRE of 1.66 (± 0.60) mm ($p < 0.001$ compared to D-IMLP) and and GD-IMLOP produced mean tRE of 1.38 (± 0.48) mm ($p < 0.001$ compared to both D-IMLP and D-IMLOP), respectively. With more data and more sample points, we can likely perform well enough with our PCA model to attempt to classify the facial expressions (Fig. A.26). The residual errors produced by our algorithms also correlate with the tRE, indicating that our algorithms have the ability to handle such data (Fig. A.27).

Fig. A.26. Non-medical data experiment: this particular target shape (right) has a lot of detail which is necessary to convey the emotion in this face. 1000 sample points are too few to capture this detail resulting in an inaccurate reconstruction (left). However, with 2000 sample points, we are able to estimate this expression better (middle) since more sample points are better able to capture the detail in the target.

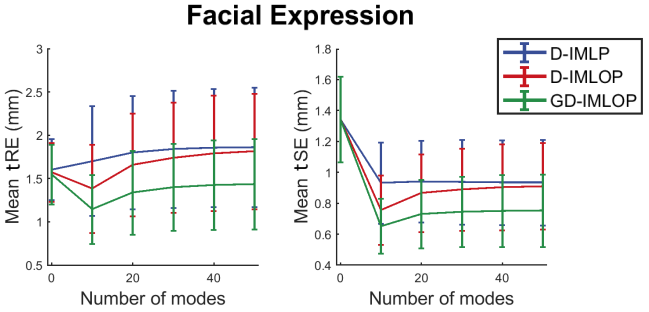

Fig. A.25. Non-medical data experiment: mean tRE (left) and tSE (right) obtained using different number of modes to estimate the test shape using facial expression data.

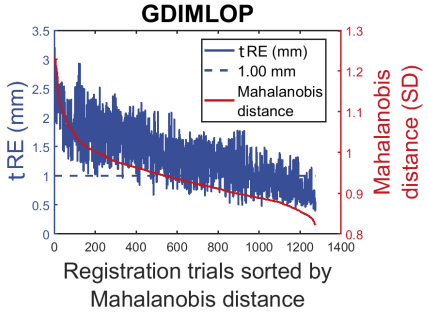

Fig. A.27. Non-medical data exp.: residual errors compared against tRE for GD-IMLOP using facial expression data. The two measures show correlation (correlation coefficient = 0.77).