Non-convex quasi-norm-based normalization of MALDI-MS Imaging Data

Luis Mancera[1], Lyna Sellami[1], Philippa Hart[1], Fiona Henderson[2], Hervé Boutin[2], Adam McMahon[2], Omar Belgacem[1]

[1]Shimadzu, Manchester, UK; [2]Wolfson Molecular Imaging Centre, University of Manchester, UK

Overview

- We explore the use of non-convex quasi-norms as normalization factor for MALDI-MS Imaging datasets
- As a lower norm parameter is used, results show less artefacts but less uniform normalization range. A trade-off value should be chosen
- Results shown for decreasing concentrations of angiotensin peptide by itself and mixed with E.Coli (DH5α) suspension

Introduction

Normalization is used as a pre-step before applying multivariate techniques or to improve visibility in MALDI imaging datasets. Total ion count (TIC) normalization calculates the normalization factor for each pixel as the l1-norm of the signal intensities. It has been found to be less prone to artefacts compared to variations based on higher norms (i.e., l2 or RMS, l∞ or maxima-based). We investigate the use of generalized lp-quasi-norms, with $0 < p < 1$. As p gets smaller, the increasing influence of smaller signal intensities in the normalization factor tend to be even less prone to artefacts, but normalization ranges become less uniform across pixels. A reasonable trade-off is proposed within $0.75 \leq p \leq 1$.

Theoretical background

Normalization is the application of a scaling factor to all signal intensities in a mass spectrum.

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

Factor calculated as a particular case of the generalized lp-norm. When:

- **p = 1**, the result is the sum of all intensities in the spectrum (TIC)
- **p = 2**, the factor becomes the vector norm
- **p →∞**, the norm is approximated by the maximum intensity
- **0 < p < 1**, the norm becomes a quasi-norm, holding all properties of a norm except for triangle inequality

Experimental Methods

Sample preparation:

- Decreasing concentrations of the angiotensin II peptide (m/z 1046.54) corresponding to 76.48, 38.24, 15.29, 7.64, 3.82, 1.92, 0.95, 0.764, 0.47, 0.38 pmol/µl were prepared. 0.5 µl of each solution were spotted on the ITO glass slide (left side)
- Same solutions were mixed with E.coli (DH5α) suspension (Takara Bio, Japan) v:v. 1 µl of the latter was spotted on the ITO glass slide (right side)
- The plate was sprayed with CHCA matrix (5mg/ml) using the SunCollect sprayer (SunChrom, Friedrichsdorf), 20 layers at 10 µl/min, except from an area that was covered (stripe in the middle of the plate) in order to get no signal (real zero)
- Data were acquired using the MALDI-7090 (Shimadzu, Manchester, UK) in reflectron positive mode, 100 shots/raster position, 150 µm spatial resolution, 100 µm laser diameter

Normalization:

- Raw data were exported to imzML using MALDI Solutions (Shimadzu, Manchester, UK). After normalization (C++), data exported back to modified imzML files. Figures generated using R, and images using Biomap (http://www.maldi-msi.org)

Results & discussion

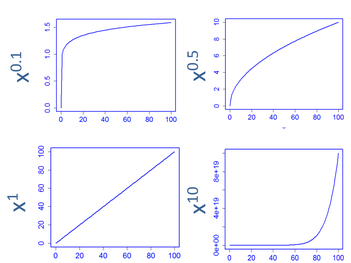


Figure 1: Contribution of a single value x to the normalization factor for different p-values.

- **p < 1** produces logarithmic functions that increases the relative weight of small intensities in the calculation of the factor
- **p ≈ 0**, at the limit, all nonzero intensities contribute a value of 1 to the factor
- **p > 1** increasingly inverse effect. At the limit, only the highest intensities contribute

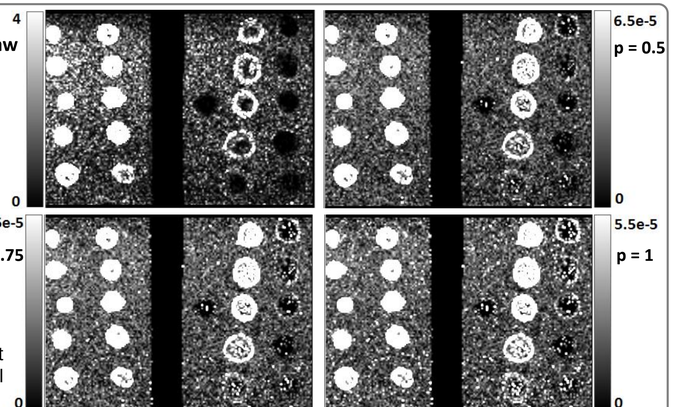
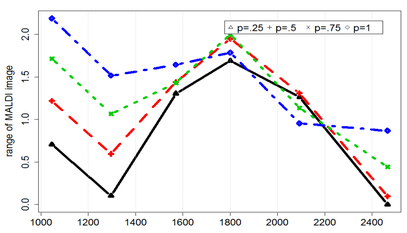


Figure 2: Effect of the normalization process for several p values. Images of angiotensin peptide (m/z 1046.54, lateral res. 150 µm). Spots correspond to the quantities of angiotensin loaded, from top to bottom and left to right, 38 pmol, 19 pmol, 7.6 pmol, 3.8 pmol, 1.9 pmol, 0.95 pmol, 0.47 pmol, 0.38 pmol, 0.23 pmol and 0.19 pmol. With (right side of the plate) or without (left side of the plate) mixing with E.coli suspension.

- Figure 2, top left image, shows a weak signal on the mixtures of angiotensin and E. Coli suspension, mostly at lower concentrations
- p increases ➔ signal enhanced on those spots, visibility and sensitivity improved on smaller intensities but equalization effect over the amplitudes across the image also make some false positives. Note that false positives do not appear in the absolute zero stripe in the middle of the image, but they get more significant in those areas of original lower intensities
- On our experiments we found that values within $0.75 \leq p \leq 1$ give the best results for most cases of visualization and data pre-processing
- Figure 3 shows the range between maximum and minimum values for MALDI images generated at several peaks of interest of angiotensin peptide, normalized using various lp-norms
- As p is increased, the range of values across pixels in the dataset gets more uniform, explaining the increased importance of signals with smaller relative amplitude but also the increased difficulty to separate signal from noise



Conclusions

- Generalized lp-norm-based normalization can be extended beyond traditional $p \geq 1$ values to non-convex values in the range $0 < p < 1$.
- The resulting normalization factor is more heavily affected by smaller amplitudes as the value of p approaches zero.
- This parameter can be tuned to achieve a trade-off between the equalization of amplitudes across different pixels and the appearance of visual artefacts and false positives.
- The best value will depend on the experiment, but we have found that a value between 0.75 and 1 often provides good results.

References

[1] Rasmussen, G.T.; Isenhour, T.L.. J. Chem. Inf. Comput. Sci. (1979), 19: 179-186

[2] Deininger, S.O.; Cornett, D.S.; Paape, R., Becker, M; Pineau, C.; Rauser, S.; Walch, A.; Wolsky, E. Anal Bioana. Chem (2011) 401:167-181.