High throughput decomposition of spectra

Dumitru Mirauta[1] **Vladimir Gusev**[1] **Michael Gaultois**[2] **Matthew Rosseinsky**[2]
[1]Department of Computer Science [2]School of Physical Sciences
University of Liverpool
Merseyside, L69 3BX
{d.mirauta, vladimir.gusev, m.gaultois, rossein}@liverpool.ac.uk

Abstract

In order to fully utilise the potential throughput of automated synthesis and characterisation data collection, data analysis capabilities must have matching throughput, which consumes excessive (human) expert time even for small datasets. One such analysis task is unmixing; being able to generally separate, from a sample consisting of multiple components, the individual patterns characteristic of the constituent parts. Such tasks are often complicated by variation of the basis patterns (e.g. peak shifting and broadening in PXRD). Conventional approaches focus on fitting a parameterised subset of transformations or utilising phase space relationships, and so one tuned for PXRD may require extensive modification or retraining before being suitable for another modality. This work aims to build a more robust foundation for unmixing, not specific to a particular spectral modality. A more robust optimisation can be achieved through a more robust cost, and distance/comparison is a vital component of such costs. We construct a non-regressive, distance geometry based framework, in this presentation leveraging Optimal Transport (OT) with a Euclidean ground cost, but lending itself to modification through the use of different distances. This provides a non-parametric approach that allows for arbitrary variation. We show through numerical experiments that our approach can handle fully blind basis discovery despite independent random peak shifting/broadening at various intensities, where matrix factorisation frameworks break down. We also showcase use in smaller data regimes through a laboratory discovery mockup, where our method can flag compositions containing an unknown trace component.

1 Introduction

While characterisation instruments, robotics and expert time can be scarce for all but the largest of labs, compute (in the form of consumer hardware) is universally abundant. Not only that, but full use of the former hinges on sufficiently powerful, robust and algorithmically efficient methods for the latter. There is much to still be discovered in the materials realm, so there is much to be gained from democratisation of the endeavour, towards this goal, it is therefore paramount to get the most out of such scarcity. Regardless, the analysis challenge will grow even further with increased availability of spectral instruments and robotics. In particular, quick and reliable unmixing is important both for fraction estimation/component extraction for further characterisation, and for identification of trace components corresponding to new materials.

A classic approach to unmixing, at least in the setting of non-varying bases, is non-negative matrix factorisation (NMF) (1). This arises from the fact that linear combination $\mathbf{c} = \sum_k a_k \mathbf{b}^k$ can be written as matrix-vector product $\mathbf{c} = \mathbf{B}\mathbf{a}$ and indeed multiple such products as a matrix-matrix product $\mathbf{C} = \mathbf{B}\mathbf{A}$. The NMF method performs a (regression) fit for this matrix relation. Going further, overlapping NMF (oNMF) allows each basis to be varied by convolution, with variations $\widetilde{\mathbf{b}}^k = \boldsymbol{\lambda}^k * \mathbf{b}^k$ composing $\mathbf{c} = \sum_k a_k \widetilde{\mathbf{b}}^k$ (2; 3). Agile Factor Decomposition, a method tuned for

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

PXRD, constrains this to shifts only, e.g. $\boldsymbol{\lambda}^k = (0, 1, 0, \ldots, 0)$, and enforces phase map constraints (4; 5). The topic of unmixing is explored extensively in the signal processing literature (6; 7; 8; 9). Generally such models are fit through regression, often in terms of a pointwise distances. Note, regression in these only corresponds to a maximum likelihood estimate for pointwise independent noise (10). As difference drives the fitting process, an appropriate notion of difference can make sure we are driven the right way.

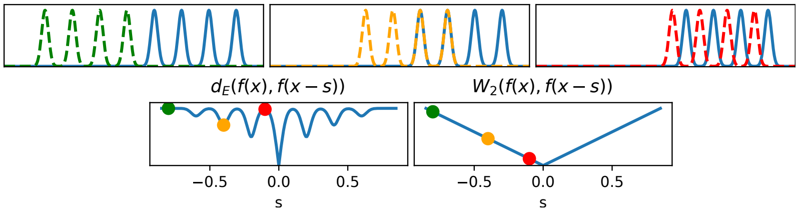


Figure 1: Top: A pattern (blue) and progressive variations of it (green,orange,red). Bottom: False alignments of a pointwise distance vs convexity of a Wasserstein distance (w.r.t. shift).

In the context of PXRD, a small change in a crystal or measurement conditions (e.g. temperature) ought to result in a small change in lattice planes, therefore a small change in Bragg angle. This should mean that an interference spike that otherwise occurs at 2θ, occurs at $2\theta + \Delta$; this is more so a domain-wise perturbation (side-to-side), than a pointwise independent change in intensity.

2 Our methodology

Here we will showcase a high level overview of our work, a more detailed presentation and analysis will be provided in an upcoming paper. While in EMD-NMF (11), OT is essentially used to compare already combined patterns, and while in Wasserstein dictionary learning (12) OT averaging is used as a combination mechanism, in our work we will seek to use OT to describe how each basis pattern transforms.

Let variation of a basis pattern merely rearrange its mass/intensity distribution, for instance shifting by Δ merely takes mass from x and deposits it at $x + \Delta$. Then even after superposition, a mixture consists of these rearranged basis parts. These parts can be arbitrarily small and their movements need not be assumed correlated. We can attempt to trace back these parts, by considering all possible origins and a proposed likelihood for each. We require two simple principles; a closer origin is likelier than further, and all available mass must be accounted for.

To begin with, consider a single mixture $\mathbf{c}$ and a fully known set of basis patterns $\mathbf{b}^k$. Then, let p_{ij}^k be the amount of mass in $\mathbf{c}$ at position j that originates from position i of $\mathbf{b}^k$ (see Fig. 2 - middle), then applying a penalty d_{ij} for movement from position i to j, for instance $d_{ij} = (x_i - x_j)^2$ as we use here, we define the total cost of the separation that this p_{ij}^k suggests

$$\sum_{ijk} d_{ij} p_{ij}^k. \tag{1}$$

For mass conservation, we require $\sum_{ik} p_{ij}^k = c_j$ and $\sum_j p_{ij}^k = a_k b_i^k$ where $a_k = \sum_{ij} p_{ij}^k$ and $\sum_i b_i^k = \sum_j c_j = 1$. We call minimisation of cost 1 with these constraints the separation problem. In this, one may observe multiple coupled transport costs (see Supplementary material C), between original bases $\mathbf{b}^k$ and the varied counterparts $\widetilde{\mathbf{b}}^k$ that constitute $\mathbf{c} = \sum_k a_k \widetilde{\mathbf{b}}^k$. This is a (well defined) linear problem in p_{ij}^k, which has a unique solution. Further, in 1D with certain d_{ij}, there is a unique feasible (overall) plan which can be constructed in linear time with no minimisation, at least for given abundances a_k. If we take advantage of this, we need only optimise for a_k.

During (partially or fully) blind unmixing, some or all of the basis patterns may be unknown, and must be inferred from multiple mixed examples. While reasoning how variations due to shifted bases may be distributed in Euclidean space is difficult due to coordinate permutation, i.e. we might have $(b_1 b_2 \ldots b_N) \rightarrow (0 b_1 b_2 \ldots b_{N-1})$ for $\mathbf{b} \rightarrow \widetilde{\mathbf{b}}$, the convexity of the Wasserstein distance to this

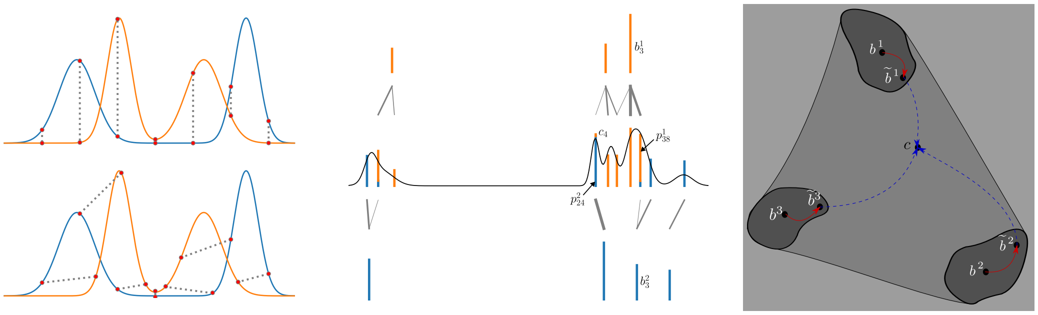


Figure 2: Left: Corresponding parts of the compared patterns according to; a pointwise distance such as $\|\mathbf{b} - \widetilde{\mathbf{b}}\|_2$ (top), and Wasserstein (bottom). Middle: Visualisation of the flexible parts-based representation that the separation program gives. Right: Metric view of basis variation and mixing, with dark grey representing possible basis variations, lighter representing possible combinations and lightest - the whole space of patterns.

operation and more (13; 14; 15), keeps variations local, giving us clear geometry in Wasserstein space (see Fig. 2 - right). Just as simplex structure can be exploited in linear mixing (6; 7), the structure present here can be exploited by looking for joint decompositions that have good clustering (in terms of Fréchet variance). See also methods considering basis distributions (16; 17; 18). This optimisation can be broken into alternating between solving for variations given a basis (separation) and solving for the basis given variations (OT Fréchet averages). We might intuit that, just as an ordinary sample mean converges to the distribution mean with increasing samples, that in order for this aggregate to sufficiently reflect a true basis pattern, a sufficient amount of varied examples are required (i.e. mixtures in which the basis appears in non-trivial amount).

By definition, trace components are not significantly represented and cannot be deduced through averages, however, we can still detect their presence. We need only watch out for uncharacteristically high separation costs, indicating that our basis has subpar explanatory power in a given mixture, and that some mass may in fact be transported from a distribution not included in our basis. With a modification to the separation problem, where we allow for a residual, this anomaly can be estimated and illustrated. This may be done on demand, after detection through the aforementioned means.

3 Experimental results

Here we share a preview of our results in applying our approach on both synthetic patterns and laboratory PXRD data. As part of other work, we have successfully applied our methods to Raman, XPS and even hyperspectral remote sensing data.

3.1 Synthetic stress test

In order to illustrate the advantage over classical approaches, we generate many mixed spectra datasets with large and complex variation (of bases) prior to basis combination. With some abuse of notation, our bases can be expressed as $b^i = \sum_{k=1}^{N_G} I_k^i \mathcal{N}(\mu_k^i, \sigma_k^i)$. Basis variations follow a similar format, with the same peak intensities I_k^i, however the bases of each mixture c^j will use independently perturbed parameters $\mu_k^{ij}, \sigma_k^{ij}$ (see Supplementary material A for specifics). Mixtures themselves are then generated by combining these randomly varied bases in random proportions. We generate 3 bases and 800 mixtures per dataset, and 400 datasets per variability preset (see supplementary B for a visualised example). We include the usual root mean square error (RMSE) metrics, as well as the average absolute difference in percentage abundances (AADPA) which may be more interpretable. Between inferred abundances $\mathbf{A}$ and ground truth abundances $\mathbf{A}_t$, this is calculated as $\frac{100}{KM} \sum_{i=1}^{K} \sum_{j=1}^{M} |a_i^j - (a_t)_i^j|$.

For each generated dataset, we attempt to derive both the abundances and the basis itself (blind unmixing), each tested method is run eight times and the best result is recorded. We can see that for the medium and high presets NMF and oNMF begin to break down, while the accuracy of our method is less impacted.

Variability	Approach	Metric		
		AADPA$(\mathbf{A}, \mathbf{A}_t)$	RMSE$(\mathbf{A}, \mathbf{A}_t) \times 100$	RMSE$(\mathbf{B}, \mathbf{B}_t)$
Low	NMF	5.921 ± 2.539	8.040 ± 3.546	1.793 ± 0.586
	oNMF	6.813 ± 2.269	9.112 ± 2.988	2.253 ± 0.896
	Our approach	$\mathbf{4.289} \pm 1.959$	$\mathbf{5.571} \pm 2.720$	$\mathbf{1.663} \pm 0.516$
Medium	NMF	10.624 ± 3.873	14.623 ± 5.255	3.214 ± 0.926
	oNMF	10.735 ± 3.010	14.282 ± 3.881	4.292 ± 1.521
	Our approach	$\mathbf{5.954} \pm 2.008$	$\mathbf{8.049} \pm 2.819$	$\mathbf{2.564} \pm 0.590$
High	NMF	14.346 ± 3.873	19.498 ± 4.979	4.283 ± 0.944
	oNMF	13.759 ± 3.122	18.214 ± 3.894	5.788 ± 1.695
	Our approach	$\mathbf{8.295} \pm 2.393$	$\mathbf{11.396} \pm 3.310$	$\mathbf{3.628} \pm 0.721$

Table 1: Mean and standard deviation of decomposition metrics across the 400 trials.

Despite having more freedom, oNMF sometimes performs worse than NMF, this illustrates that residual minimisation may not be sufficient as a driving force for good decompositions. A closer fit of the overall sum, does not require good components. After all, one could arbitrarily divide a mixture into irrelevant components and have no residual. In contrast our method can express more subtle changes, and is able to look for the least changes required, yet is capable of describing more if necessary. A smaller standard deviation in our metrics also suggests more consistent decompositions.

3.2 Prototypical discovery problem

As a proof of concept, a laboratory dataset was made to test whether we could blindly identify an unknown component. Four substances were chosen, mixed in different amounts, then the PXRD spectra of the mixtures as well as those of three of the four basis substances were shared, the mixing ratios and the fourth were not initially shared. A total of 16 mixtures were made, 6 of which contained small amounts of the unknown substance (11.5%, 9.9%, 8.3%, 6.5%, 4.7% and 2.9% by mass).

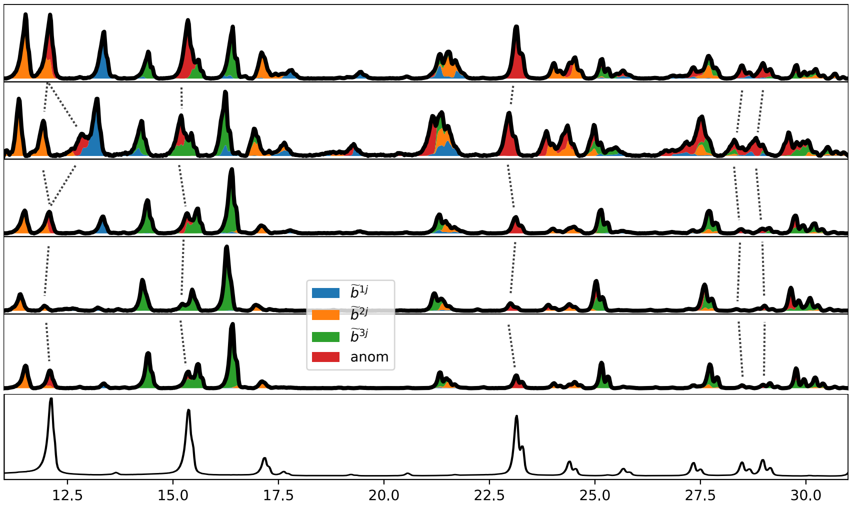


Figure 3: Summary of high level search for unknown, augmented through our tool. Note the variation in peak positions and that not all anomalous area is due to the unknown. Top: Commonality across anomalies, betraying the unknown. Bottom: true initially unknown PXRD pattern.

We applied our modified separation approach to this, and used it to organise the mixtures from most to least anomalous. Looking at the most anomalous, our computed anomalies highlighted peaks which likely belonged to a non-basis pattern, as they were common across the anomalies (which can also contain preprocessing artefacts or occasional misassignments), we later linked this to the true hidden PXRD pattern (see Fig. 3). Note, this process of extracting further commonality from the most anomalous area, though in this instance manual, could be further automated.

References

[1] Daniel Lee and Hyunjune Seung. Algorithms for non-negative matrix factorization. *Adv. Neural Inform. Process. Syst.*, 13, 02 2001.

[2] Paris Smaragdis. Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs. In Carlos G. Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation*, pages 494–499, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-30110-3.

[3] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-Ichi Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, September 2009. doi: 10.1002/9780470747278. URL `https://doi.org/10.1002/9780470747278`.

[4] Yexiang Xue, Junwen Bai, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Johan Bjorck, Liane Longpre, Santosh K. Suram, Robert B. van Dover, John Gregoire, and Carla P. Gomes. Phase-mapper: An ai platform to accelerate high throughput materials discovery, 2016.

[5] Di Chen, Yiwei Bai, Sebastian Ament, Wenting Zhao, Dan Guevarra, Lan Zhou, Bart Selman, R. Bruce van Dover, John M. Gregoire, and Carla P. Gomes. Automating crystal-structure phase mapping by combining deep learning with constraint reasoning. *Nature Machine Intelligence*, 3 (9):812–822, September 2021.

[6] W. Ma, J. M. Bioucas-Dias, T. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1):67–81, 2014.

[7] R. Heylen, M. Parente, and P. Gader. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 1844–1868, 2014.

[8] José M. Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[9] Ricardo Borsoi, Tales Imbiriba, José Bermudez, Cédric Richard, Jocelyn Chanussot, Lucas Drumetz, Jean-Yves Tourneret, Alina Zare, and Christian Jutten. Spectral variability in hyperspectral data unmixing: A comprehensive review, 01 2020.

[10] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005.

[11] R. Sandler and M. Lindenbaum. Nonnegative matrix factorization with earth mover's distance metric for image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (8):1590–1602, 2011.

[12] Morgan A. Schmitz, Matthieu Heitz, Nicolas Bonneel, Fred Ngolè, David Coeurjolly, Marco Cuturi, Gabriel Peyré, and Jean-Luc Starck. Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM Journal on Imaging Sciences*, 11(1): 643–678, jan 2018.

[13] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2020.

[14] Rémi Flamary. Optimal transport for machine learning, 2019. `https://remi.flamary.com/biblio/hdr.pdf`.

[15] Bjorn Engquist, Brittany Hamfeldt, and Yunan Yang. Optimal transport for seismic full waveform inversion. *Communications in Mathematical Sciences*, 14, 02 2016. doi: 10.4310/CMS.2016.v14.n8.a9.

[16] D. Stein. Application of the normal compositional model to the analysis of hyperspectral imagery. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*, pages 44–51, 2003. doi: 10.1109/WARSD.2003.1295171.

[17] Xiaoxiao Du, Alina Zare, Paul Gader, and Dmitri Dranishnikov. Spatial and spectral unmixing using the beta compositional model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):1994–2003, 2014. doi: 10.1109/JSTARS.2014.2330347.

[18] Yuan Zhou, Anand Rangarajan, and Paul D. Gader. A gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing*, 27 (5):2242–2256, 2018. doi: 10.1109/TIP.2018.2795744.

Supplementary material A

For bases, peak parameters are drawn

$$I_k^i \sim |\mathcal{N}(1, 0.25)|, \quad \mu_k^i \sim U(0.2, 0.8) \quad \text{and} \quad \sigma_k^i \sim 0.02 \cdot 2^{\mathcal{N}(0,0.5)},$$

while for their variations in mixture j

$$\mu_k^{ij} = \mu_k^i + \Delta\mu_k^{ij}, \quad \sigma_k^{ij} = \sigma_k^i \cdot \gamma_k^{ij}.$$

with

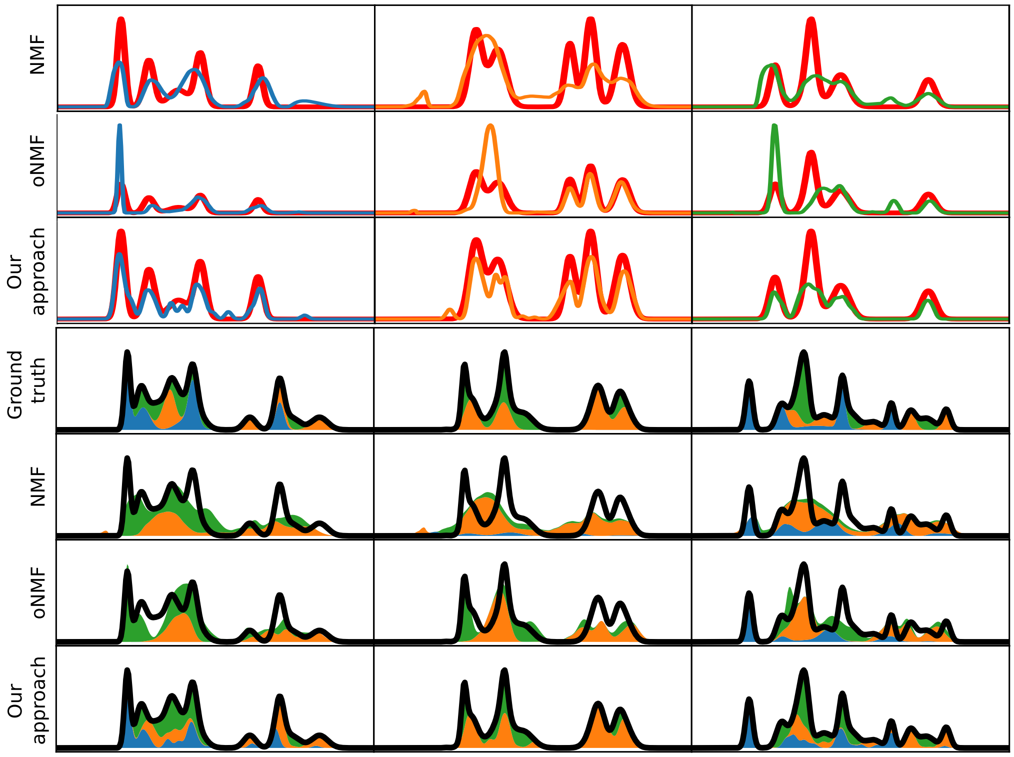
$$\Delta\mu_k^{ij} \sim \mathcal{N}(0, \sigma_{\Delta\mu}) \quad \text{and} \quad \gamma_k^{ij} \sim 2^{\mathcal{N}(0,\sigma_p)},$$

according to variation presets (Table 2).

Preset	$\sigma_{\Delta\mu}$	σ_p	95% of $\Delta\mu_k^{ij}$ are in	95% of σ_k^{ij} are in
Low	0.015	0.15	$[-0.03, +0.03]$	$[0.81\sigma_k^i, 1.23\sigma_k^i]$
Medium	0.025	0.3	$[-0.05, +0.05]$	$[0.66\sigma_k^i, 1.52\sigma_k^i]$
High	0.035	0.5	$[-0.07, +0.07]$	$[0.50\sigma_k^i, 2.00\sigma_k^i]$

Table 2: Basis variation strength presets used, where two variations of the same peak can be found as far as 14% of the domain width apart.

Supplementary material B



Visualised decompositions of a random dataset generated with the high preset. Each row corresponds to a different general unmixing method. Top: Bases (one per column, true in red). Bottom: Basis separation different mixtures (columns).

Supplementary material C

Taking normalised transport plans $a_k \bar{p}_{ij}^k = p_{ij}^k$ we can express Eq. (1) as

$$\sum_k a_k \langle \mathbf{D}, \bar{\mathbf{P}}^k \rangle. \tag{2}$$

For mass conservation, we require $\sum_k a_k \widetilde{\mathbf{b}}^k = \mathbf{c}$, $\widetilde{\mathbf{b}}^k = \mathbf{1}^T \bar{\mathbf{P}}^k$ and $\bar{\mathbf{P}}^k \mathbf{1} = \mathbf{b}^k$ (with $\mathbf{b}^k$ and $\mathbf{c}$ normalised as before). Where

$$\min_{\substack{\bar{\mathbf{P}}^k \\ \text{s.t. } \widetilde{\mathbf{b}}^k = \mathbf{1}^T \bar{\mathbf{P}}^k, \bar{\mathbf{P}}^k \mathbf{1} = \mathbf{b}^k}} \langle \mathbf{D}, \bar{\mathbf{P}}^k \rangle = W_2(\widetilde{\mathbf{b}}^k, \mathbf{b}^k)^2, \tag{3}$$

hence the minimisation of Eq. (1) is equivalent to

$$\min_{\substack{a_k, \widetilde{\mathbf{b}}^k \ \forall k \\ \text{s.t. } \sum_k a_k \widetilde{\mathbf{b}}^k = \mathbf{c}}} \sum_k a_k W_2(\widetilde{\mathbf{b}}^k, \mathbf{b}^k)^2. \tag{4}$$

Note, though written as histograms here, discrete distributions with non-matching supports can be used through separate (squared) distance matrices $\mathbf{D}^k$. Alternatively, with concatenations of $\mathbf{D}^k, \mathbf{P}^k$ and supports it can also be shown that

$$\min_{a_k} W_2 \left(\mathbf{c} = \sum_k a_k \widetilde{\mathbf{b}}^k, \sum_k a_k \mathbf{b}^k \right)^2 \tag{5}$$

is equivalent. It would be tempting to think of this as a fit of the linear superposition of untransformed bases, but note that computing/keeping the transport map so that $\widetilde{\mathbf{b}}^k$ may be recovered is important here. Non-matching supports are also important here, as it helps us separate $\widetilde{\mathbf{b}}^k$.