Supporting Information for

A data-driven sequencer that unveils latent "codons" in synthetic copolymers.

Yusuke Hibi[1*], Shiho Uesaka[1] and Naito Masanobu[1*]

Correspondence to: hibi.yusuke@nims.go.jp and naito.masanobu@nims.go.jp

This PDF file includes:

Materials and Methods

Computational Methods

Supplementary Text

Supplementary Figures S1 to S10

Supplementary Tables S1 to S2

Captions for Data S1 to S5

References cited in this supporting information document.

Other Supplementary Materials for this manuscript include the following:

Data S1 to S5 available in DOI: 10.26434/chemrxiv-2022-mw76d-v2.

Materials and Methods

Materials

Monomers were purchased from Tokyo Chemical Industry and passed through a thin pad of alumina right before polymerization. Poly(styrene-alt-methyl methacrylate) was purchased from Polymer Source (P1604-SMMAalt). For benchmark compositional analysis, living anion polymerized poly(methyl methacrylate) (M) and polystyrene (S) were purchased from Shodex (M_n=178,00) and Tosoh (M_n=102,00), respectively. Free-radical polymerized poly(ethyl methacrylate) (E) was purchased from Polymer Source. Reversible addition and fragmentation chain transfer (RAFT) agent, 2-(dodecylthiocarbonothioylthio)-2-methylpropionic acid (DDMAT), was purchased from Aldrich. Azobis(isobutyronitrile) (AIBN) was purchased from Fujifilm Wako Pure Chemical Corporation, purified by recrystallization from methanol, and stored at 0 °C. Monomers were purchased from Tokyo Chemical Industry and passed through alumina column right before use. All the other high purity chemicals were purchased from Fujifilm Wako Pure Chemical Corporation and used without further purification.

A synthetic procedure of random copolymers via free-radical copolymerization

A typical polymerization procedure is given here. In an open vial, *n*-butyl acrylate (B) (1.15 mL; 8 mmol), styrene (S) (0.23 mL; 2 mmol) and dimethyl 2,2′-azobis(isobutyrate) (23 mg; 0.1 mmol in toluene 0.1 mL) were placed. After sealing the vial with a septum,

nitrogen gas was bubbled just for one minute. The reaction mixture was stirred for an hour at 60 ºC, and then dropped into methanol/water = 9/1 (v/v). The precipitated polymer was dried overnight under vacuum and subjected to pyrolysis-MS measurements for dataset preparation. High-temperature polymerization (150 ºC) was conducted without addition of the initiator/toluene solution. No unexpected or unusually high safety hazards were encountered.

A synthetic procedure of B/S random copolymers via living radical copolymerization [1]

To a round flask reactor, DDMAT 72.9 mg (0.2 mmol) and AIBN 9.9 mg (0.06 mmol) were placed. Styrene 2.29 mL (20 mmol), *n*-butyl acrylate 2.87 mL (20 mmol) and 1,4-dioxane (2 mL) were all combined, bubbled with nitrogen gas for 30 mins and then transfer to the degassed reactor at room temperature. The system was heated up to 70 ºC to start polymerization. Small portions of the polymerization solution were taken out at 5, 10, 25, 58 h. The withdrawn solutions were subjected to ^{1}H NMR for conversion calculation and pyrolysis-MS for direct sequencing. The left solutions were dropped into methanol/water =9/1 (v/v). The precipitated polymers were dried under vacuum at 100 ºC for 24 h and subjected to ^{1}H and ^{13}C NMR for triad sequencing.

Sample preparation procedures for the benchmark test of the E/M/S ternary polymer film

A typical procedure for preparing ternary films is here given. A polymer mixture solution was prepared by dissolving 2.0 mg of each E/M/S in 294 mg of 1,4-dioxane (totally 2 wt% solution). The solution was then dropped in a copper pan. The solution weight was quickly measured, typically around 10 mg. The polymer weight is calculated by the solution

weight and polymer concentration, typically around 0.20 mg. The pan was left in atmosphere overnight, yielding a drop casted film for pyrolysis-MS measurement.

Pyrolysis-MS measurements

Polymer samples were pyrolyzed on a heater (ionRocket; Biochromato) at heating rate of 50 ºC/min from 50 ºC to 500 ºC after two minutes preheating period at 50 ºC. One sample thus took 12 mins measurement time. The pyrolyzed gases were continuously ionized by excited He gas with DART-ion source (DART-OS; IonSense). MS spectra were recorded with LCMS-2020 (Shimadzu) in positive-ion mode at 50 scan/min, yielding 550 spectra per a sample. The mass range was 50-1500 m/z and the interval scale was 0.05 m/z with mass resolution of 2000. The spectra were output in CDF file format, converted to Numpy format with a Python module, netCDF4. The spectra were further formatted for data size reduction as described in "Spectra formatting" section. The formatted spectral datasets used in this study are presented as Data S1-S5. All the data processing were conducted on Python3.7 on a Windows 11 laptop computer with a processing unit of AMD Ryzen9 4900HS without external GPU assistance. The total processing time depended on the dataset size and further specified mass range, which was not beyond 3 hours in this study.

Computational Methods

Mathematical notations

The notation mostly follows the common convention in signal processing. ${\mathbb{R}_+}^{N\times M}$ and $\mathbb{R}^{N\times M}$ represent a non-negative real matrix and real matrix, respectively, with $N \times M$ dimensions. For a matrix $\boldsymbol{X} \in \mathbb{R}^{N\times M}$, $\boldsymbol{X}_{n:} \in \mathbb{R}^{1\times M}, \boldsymbol{X}_{:m} \in \mathbb{R}^{N\times 1}, X_{nm} \in \mathbb{R}$ respectively represent the n^{th} row vector, m^{th} column vector, and (n, m)-element of the matrix $\boldsymbol{X}$. $\boldsymbol{X}^T$ represents a transposed matrix of $\boldsymbol{X}$. $\|\boldsymbol{X}\|_F$ represents Frobenius norm of $\boldsymbol{X}$. $\|\boldsymbol{X}_{n:}\|_1$ and $\|\boldsymbol{X}_{n:}\|_2$ respectively represent ℓ_1- and ℓ_2-norm of n^{th} row vector of $\boldsymbol{X}$. For a square matrix of $\boldsymbol{X}$, $Tr(\boldsymbol{X})$ represents the trance of $\boldsymbol{X}$. $\mathbf{1}_N$, $\mathbf{11}_N$ and $\boldsymbol{I}_N$ respectively represent a N-dimensional all-ones vector, (N, N)-dimensional all-ones matrix and N-dimensional identity matrix.

Spectra formatting

For simplicity, in the main text, we described as if pyrolysis-MS was 1D-spectrum and dataset was $\boldsymbol{X} \in {\mathbb{R}_+}^{N\times D}$ (N: sample number, D: channel number). However, as described in pyrolysis-MS measurement section, a single pyrolysis-MS was 2D-spectrum consisting of 550 1D-spectra recorded at different temperatures. The first 100 spectra corresponded to the preheating period. The mass range was 50-1500 m/z with intervals of 0.05 m/z. The original matrix size for a single sample was thus 550×29001. The absolute peak-intensities of DART pyrolysis-MS are sensitively affected by sample weights and ambient environments, especially humidity. As we collected data over a year in a room without controlling humidity in Japan, there should be a huge fluctuation in the relative

humidity (20-100%), making direct comparison of spectral intensities among different samples invalid. The spectral intensities were thus scaled by an internal standard peak and sample weight. As an internal standard, we used a background peak at 391 m/z contained in the first 100 spectra from preheating period, attributable to a proton adduct of di(2-ethylhexyl)phthalate, which is a plasticizer continuously released from plastics and equilibrated in a room. After scaling the spectral intensities by the internal standard peaks and sample weights, the first 100 spectra were removed from the data matrix. The spectral dimension per a sample became 450×29001. As such high temperature resolution (1 ^{o}C) is unnecessary, the spectra were averaged into *T*-temperature bands (typically $T \in [10, 20]$). Since the mass resolution in our setup was not high enough to precisely measure mass with 0.05 m/z accuracy, the m/z intervals were doubled to 0.1 m/z and then gaussian filter was applied; smoothing sequential 4 channels (0.4 m/z) with standard deviation of 2 channels (0.2 m/z). For further reducing data size, we specified the mass range depending on the dataset (see Table S2). In this study, we used 5 datasets: S/B binary triad sequencing (Fig. 1A), benchmark compositional analysis of E/M/S ternary polymer films (Fig. 2A), M/S binary triad sequencing (Fig. 3), M/S/B ternary triad sequencing (Fig. S8), and S/B binary pentad sequencing (Fig. 4A). For the benchmark dataset of 24 samples, the mass range was limited to 50-1000 m/z, since a region over 1000 m/z showed very subtle peaks only, deriving a 20×9501 matrix for a single sample and 460×9501 matrix for the entire dataset. For triad sequencing of S/B binary, M/S binary and M/S/B ternary, the mass range was limited to 50-500 m/z, as the fragment distribution longer than triads cannot be expressed with 5 bases (see Fig. S1). For pentad sequencing of S/B, mass range was limited to 100-670 m/z, for

covering up to pentads (the largest pentad is BBBBB at 641 m/z for proton adduct and 658 m/z for ammonium ion adduct). For sequencing, temperature range was further specified within 200-450 ºC and averaged into T=10 (for triad sequencing) and T=15 (for pentad sequencing). Each spectrum was ℓ_1-normalized and the resulting spectral dataset, $\widetilde{\boldsymbol{X}} \in \mathbb{R}_+^{NT\times D}$ (N: sample number, T: temperature-band number, D: channel number), was subjected to the first NMF (see Fig. S4). The ℓ_1-norms of all the NT-spectra were preserved in a diagonal matrix of $\boldsymbol{L}_{\widetilde{\boldsymbol{X}}} \in \mathbb{R}_+^{NT\times NT}$ for recovering the spectral intensities after the first NMF: if the normalized $\widetilde{\boldsymbol{X}}$ was factorized into $\widetilde{\boldsymbol{X}} \approx \widetilde{\boldsymbol{A}}\boldsymbol{S}$, then the output $\widetilde{\boldsymbol{A}}$ was replaced with $\boldsymbol{L}_{\widetilde{\boldsymbol{X}}}\widetilde{\boldsymbol{A}}$ for recovering the absolute intensities.

Designing outline of RQMS algorithm

The RQMS algorithm is outlined in Fig. S4 as a flowchart. The RQMS algorithm is composed of three main parts: the two NMFs described in the main text and a filter for automatically identifying and removing contaminants/backgrounds fragments from the system based on canonical correlation analysis[2] (CCA). In the following sections, the mathematical derivations and pseudo-codes of the first NMF, CCA-filter and second NMF are described in this order.

Derivations of the first NMF

For simplicity, we wrote $\boldsymbol{X} \approx \boldsymbol{A}\boldsymbol{S}$ in the main text and above section, where $\boldsymbol{X} \in \mathbb{R}_+^{N\times D}$ and $\boldsymbol{A} \in \mathbb{R}_+^{N\times M}$ are sample-wise spectra and fragment abundances (FAs). However, in real implementation, $\widetilde{\boldsymbol{X}} \in \mathbb{R}_+^{NT\times D}$ should be subjected to the first NMF rather

than $\boldsymbol{X} \in {\mathbb{R}_+}^{N \times D}$ (see Fig. S5). The first NMF is thus $\widetilde{\boldsymbol{X}} \approx \widetilde{\boldsymbol{A}}\boldsymbol{S}$ factorizing a dataset matrix $\widetilde{\boldsymbol{X}}$ into spectrum-wise FA $\widetilde{\boldsymbol{A}} \in {\mathbb{R}_+}^{NT \times M}$ and M-fragments basis spectra $\boldsymbol{S} \in {\mathbb{R}_+}^{M \times D}$. The derivation mostly followed the previously proposed NMF by Shiga *et al.* [3] with ARD [4] and soft orthogonal constraint (ARD-SO-NMF). For a while, identical and independent distributed (i.i.d.) Gaussian noise with σ^2-variance was assumed in a probabilistic model generating $\widetilde{\boldsymbol{X}}$ from $\widetilde{\boldsymbol{A}}\boldsymbol{S}$:

$$p\left(\widetilde{\boldsymbol{X}} \middle| \widetilde{\boldsymbol{A}}, \boldsymbol{S}, \sigma^2\right) = \prod_{i=1}^{NT} \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{\left(\widetilde{\boldsymbol{X}}_{im} - \left(\widetilde{\boldsymbol{A}}\boldsymbol{S}\right)_{im}\right)^2}{2\sigma^2}\right\}.$$

To automatically determine the fragment number M, automatic relevance determination (ARD) based on sparse modeling was introduced, assuming exponential distribution for prior distribution of $\widetilde{\boldsymbol{A}}$ column-wisely parameterized by $\boldsymbol{\lambda} \in \mathbb{R}^M$, i.e.:

$$p\left(\tilde{A}_{im} \middle| \lambda_m\right) = \frac{1}{\lambda_m} \exp\left(-\frac{\tilde{A}_{im}}{\lambda_m}\right) \; s.t. \lambda_m > 0, (i = 1, \dots, NT, m = 1, \dots, M),$$

$$p\left(\widetilde{\boldsymbol{A}} \middle| \boldsymbol{\lambda}\right) = \prod_{i=1}^{NT} \prod_{m=1}^{M} p\left(\tilde{A}_{im} \middle| \lambda_m\right).$$

Note that half-gaussian distribution is also available as the prior distribution of $\widetilde{\boldsymbol{A}}$ instead of exponential distribution, as presented in [3]. Assuming sparseness on $\widetilde{\boldsymbol{A}}$ is reasonable as a certain fragment is only generated in specific samples and temperature bands. As greater λ_m tolerates greater $\left\|\tilde{A}_{:m}\right\|_1 = \sum_{i=1}^{NT} \tilde{A}_{im}$, λ_m represents the importance of the m^{th} fragment throughout the dataset. ARD starts with a large integer M and then deletes components when their λ_m becomes close to zero while updating, outputting a model with suitable component number. The total probabilistic model becomes:

$$p(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{A}}, \boldsymbol{S}) = p(\widetilde{\boldsymbol{X}}|\widetilde{\boldsymbol{A}}, \boldsymbol{S}, \sigma^2)p(\widetilde{\boldsymbol{A}}|\boldsymbol{\lambda})p(\boldsymbol{S})p(\boldsymbol{\lambda}|a, b)$$

where uniform distribution for $p(\boldsymbol{S})$ on the hypersphere of $\|\boldsymbol{S}_{m:}\|_2 = 1 \; for \; m = 1, \dots, M$, and inverse-gamma distribution for $p(\boldsymbol{\lambda}|a, b)$ parameterized by (a, b) are assumed, i.e.,

$$p(\boldsymbol{\lambda}|a, b) = \prod_{m=1}^{M} p(\lambda_m|a, b) = \frac{b^a}{\Gamma(a)} {\lambda_m}^{-(a+1)} \exp\left(-\frac{b}{\lambda_m}\right) for \; m = 1, \dots, M$$

where a is a hyperparameter very close to 1 adjusting sparsity, here fixed $a = 1 + 10^{-16}$ and $\Gamma(\cdot)$ represents the gamma function. Empirically, b was determined from the relationship to expectation of $\tilde{A}_{im}(13)$:

$$E(\tilde{A}_{im}) = \frac{b}{a-1}. \tag{1}$$

This was further correlated to $E(X_{id})$ by

$$E(\tilde{X}_{id}) = \sum_{m=1}^{M} E(\tilde{A}_{im})E(S_{md}) = M \cdot E(\tilde{A}_{im})E(S_{md}).$$

By approximating $E(\tilde{X}_{id})$ as the mean of $\tilde{X}$ (μ_X) and substituting $E(S_{md}) = \sqrt{D}$, this becomes:

$$\mu_X = M \frac{b}{(a-1)\sqrt{D}}$$

Then, b can be determined as:

$$b = \frac{\mu_X(a-1)\sqrt{D}}{M}. \tag{2}$$

A negative-log likelihood function now becomes

$$L(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{A}}, \boldsymbol{S}, \boldsymbol{\lambda}) = -\log\left[p(\widetilde{\boldsymbol{X}}|\widetilde{\boldsymbol{A}}, \boldsymbol{S})p(\widetilde{\boldsymbol{A}}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|a, b)\right]$$

$$= \frac{1}{2\sigma^2}\left\|\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{A}}\boldsymbol{S}\right\|_F^2 + \frac{DNT}{2} log 2\pi\sigma^2 + (NT + a + 1)\sum_{m=1}^{M} log\lambda_m$$

$$+ \sum_{m=1}^{M}\frac{1}{\lambda_m}\left(b + \sum_{i=1}^{NT}\tilde{A}_{im}\right) + M(log\Gamma(a) - alogb).$$

As this function is convex over $\boldsymbol{\lambda}$, the updating rule for $\boldsymbol{\lambda}$ can be determined so that $\frac{\partial L}{\partial \boldsymbol{\lambda}} \equiv \mathbf{0}$, i.e.:

$$\lambda_m = \frac{b + \sum_{i=1}^{NT}\tilde{A}_{im}}{NT + a + 1}, for\ m = 1, \dots, M \tag{3}$$

The derivation so far has completely followed ARD-SO-NMF derivation reported by Shiga *et. al*. [3]. The i.i.d. assumption is, however, significantly violated in MS data as noise variance of a channel correlates with its signal intensity. The noise distribution does not follow Poisson distribution neither [5]. We propose to estimate a noise matrix $\mathbf{E} \in \mathbb{R}^{NT \times D}$ from the residuals of ordinary least square regression based on natural isotope peaks. The estimated noise of d^{th} channel throughout the dataset becomes:

$$\boldsymbol{E}_{:d} = \widetilde{\boldsymbol{X}}_{:d} - \boldsymbol{M}^{(d)}\left(\boldsymbol{M}^{(d)T}\boldsymbol{M}^{(d)}\right)^{-1}\boldsymbol{M}^{(d)T}\widetilde{\boldsymbol{X}}_{:d}, for\ d = 1, \dots, D, \tag{4}$$

where $\boldsymbol{M}^{(d)} = \widetilde{\boldsymbol{X}}_{:[d-30, d-20, d-10, d+10, d+20, d+30]}$. As m/z intervals are 0.1 m/z, the channel set, $[d - 30, d - 20, d - 10, d + 10, d + 20, d + 30]$, should include ± 3 m/z isotope peaks. To intuitively understand this concept, consider a noiseless MS dataset where intensity ratio of isotopic peaks is precisely identical to natural isotope-abundance, and thus constant throughout all the spectra. The regression residuals thus become zeros. Conversely, in a real dataset, the intensity ratio of isotopic peaks is no longer constant among spectra owing to

signal noise, yielding non-zero $\boldsymbol{E}$ reflecting the deviations from natural isotope-abundance ratios. A channel-wise covariance matrix $\boldsymbol{R} \in {\mathbb{R}_+}^{D\times D}$ is thus obtained by $\boldsymbol{R} = \frac{1}{NT}\boldsymbol{E^T E}$. The likelihood function is then rewritten as

$$p(\widetilde{\boldsymbol{X}}|\widetilde{\boldsymbol{A}},\ \boldsymbol{S},\boldsymbol{R}) = \frac{1}{\sqrt{2\pi}^{DNT}\sqrt{|\boldsymbol{R}|}^{NT}}\exp\left\{-Tr\left[(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})\boldsymbol{R}^{-1}(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})^T\right]\right\}.$$

Note that $\boldsymbol{R}$ is a constant matrix, not necessitating updates. The entire negative log-likelihood function now becomes

$$\begin{aligned} L(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{A}},\boldsymbol{S},\boldsymbol{\lambda}) &= -\log\left[p(\widetilde{\boldsymbol{X}}|\widetilde{\boldsymbol{A}},\boldsymbol{S},\boldsymbol{R})p(\widetilde{\boldsymbol{A}}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|a,b)\right] \\ &= Tr\left[(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})\boldsymbol{R}^{-1}(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})^T\right] + \frac{DNT}{2}log2\pi + \frac{NT}{2}log|\boldsymbol{R}| \\ &+ (NT+a+1)\sum_{m=1}^{M} log\lambda_m + \sum_{m=1}^{M}\frac{1}{\lambda_m}\left(b+\sum_{i=1}^{NT}\tilde{A}_{im}\right) \\ &+ M(log\Gamma(a) - alogb). \end{aligned}$$

By simplifying about $\boldsymbol{\lambda}$ using Eq. 3 and dropping constant terms, this can be rewritten as:

$$L(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{A}},\boldsymbol{S},\boldsymbol{\lambda}) = Tr\left[(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})\boldsymbol{R}^{-1}(\widetilde{\boldsymbol{X}}-\widetilde{\boldsymbol{A}}\boldsymbol{S})^T\right] + (NT+a+1)\sum_{m=1}^{M} log\lambda_m.$$

The objective function is minimized in a framework of hierarchical alternating least square (HALS) [6], a vector-wise updating methods known as the fastest NMF compatible with orthogonal constraints [7]. For convenience, the following column-vector notations are used:

$$\boldsymbol{a}_m \equiv \widetilde{\boldsymbol{A}}_{:m} \in {\mathbb{R}_+}^{NT\times 1}, \boldsymbol{s}_m \equiv \boldsymbol{S}_{m:}{}^T \in {\mathbb{R}_+}^{D\times 1}, for\ m = 1,..,M.$$

The key of HALS is representing the residual, $\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{A}}\boldsymbol{S}$, as $\widetilde{\boldsymbol{X}}^{(m)} - \boldsymbol{a}_m{\boldsymbol{s}_m}^T$ $(m = 1,\dots,M)$, where $\widetilde{\boldsymbol{X}}^{(m)} = \widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{A}}\boldsymbol{S} + \boldsymbol{a}_m{\boldsymbol{s}_m}^T$. The negative-log likelihood now can be written

separately by *M*-components as:

$$L\left(\widetilde{\boldsymbol{X}}, \boldsymbol{a}_m, \boldsymbol{s}_m, \lambda_m(\boldsymbol{a}_m)\right)$$

$$= Tr\left[\left(\widetilde{\boldsymbol{X}}^{(m)} - \boldsymbol{a}_m {\boldsymbol{s}_m}^T\right)\boldsymbol{R}^{-1}\left(\widetilde{\boldsymbol{X}}^{(m)} - \boldsymbol{a}_m {\boldsymbol{s}_m}^T\right)^T\right] + (NT + a + 1)log\lambda_m,$$

$$for\ m = 1, \ldots, M.$$

The orthogonal constraints on rows of $\boldsymbol{S}$ can be included in L as a penalty term of $w_O \xi_m \boldsymbol{s}_m^T \boldsymbol{s}^{(m)}$ where $\boldsymbol{s}^{(m)} \equiv \sum_{j \neq m}^{M} \boldsymbol{s}_j$. This represents non-orthogonality between the m^{th} component and all the others; ξ_m is a Lagrange multiplier when the orthogonal constraint is best-satisfied under the strict non-negative constraints, which is further soften by hyperparameter $w_O \in [0, 1]$. Higher w_O imposes the stronger orthogonal constraints on the rows of $\boldsymbol{S}$. Note that even if $w_O = 1$ is given, the rows of $\boldsymbol{S}$ would not be strictly orthogonal, as ξ_m has been estimated without nonnegative-constraints [7]. The full objective function to be minimized was finally derived as:

$$L\left(\widetilde{\boldsymbol{X}}, \boldsymbol{a}_m, \boldsymbol{s}_m, \lambda_m(\boldsymbol{a}_m)\right)$$

$$= Tr\left[\left(\widetilde{\boldsymbol{X}}^{(m)} - \boldsymbol{a}_m {\boldsymbol{s}_m}^T\right)\boldsymbol{R}^{-1}\left(\widetilde{\boldsymbol{X}}^{(m)} - \boldsymbol{a}_m {\boldsymbol{s}_m}^T\right)^T\right] + (NT + a + 1)log\lambda_m$$

$$+ w_O\, \xi_m \boldsymbol{s}_m^T \boldsymbol{s}^{(m)}.$$

The gradients of L over $\boldsymbol{a}_m$ and $\boldsymbol{s}_m$ are, respectively,

$$\frac{\partial L}{\partial \boldsymbol{a}_m} = \left(\boldsymbol{a}_m {\boldsymbol{s}_m}^T - \widetilde{\boldsymbol{X}}^{(m)}\right)\boldsymbol{R}^{-1}\boldsymbol{s}_m + \frac{1}{\lambda_m}\mathbf{1}_{NT},$$

$$\frac{\partial L}{\partial \boldsymbol{s}_m} = \boldsymbol{R}^{-1}\left(\boldsymbol{s}_m \boldsymbol{a}_m^T - {\widetilde{\boldsymbol{X}}^{(m)}}^T\right)\boldsymbol{a}_m + w_O\, \xi_m \boldsymbol{s}^{(m)}.$$

The $\boldsymbol{a}_m$ and $\boldsymbol{s}_m$ minimizing L are obtained by setting the gradients zeros:

$$\boldsymbol{a}_m = \frac{\widetilde{\boldsymbol{X}}^{(m)}\boldsymbol{R}^{-1}\boldsymbol{s}_m - \frac{1}{\lambda_m}\,\mathbf{1}_{NT}}{\boldsymbol{s}_m^T\boldsymbol{R}^{-1}\boldsymbol{s}_m}, \tag{5}$$

$$\boldsymbol{s}_m = \widetilde{\boldsymbol{X}}^{(m)^T}\boldsymbol{a}_m - w_O\,\xi_m\boldsymbol{R}\boldsymbol{s}^{(m)}. \tag{6}$$

Note that the scaling factor for $\boldsymbol{s}_m$ was dropped as $\boldsymbol{s}_m$ will be normalized after each update. The updated $\boldsymbol{a}_m$ and $\boldsymbol{s}_m$ are projected to non-negative orthant after each update by, e.g.,

$$\boldsymbol{a}_m \leftarrow \frac{\boldsymbol{a}_m + |\boldsymbol{a}_m|}{2},$$

where $|\boldsymbol{a}_m|$ represents a vector having absolute values of $\boldsymbol{a}_m$ elements. The estimation of ξ_m is done by multiplying $\boldsymbol{s}^{(m)^T}\boldsymbol{R}$ to Eq. 6 from the left side with using the strict orthogonal constraint of $\boldsymbol{s}^{(m)^T}\boldsymbol{s}_m = 0$ and $w_O = 1$, deriving

$$-\boldsymbol{s}^{(m)^T}\widetilde{\boldsymbol{X}}^{(m)^T}\boldsymbol{a}_m + \xi_m\boldsymbol{s}^{(m)^T}\boldsymbol{R}\boldsymbol{s}^{(m)} = 0,$$

$$\xi_m = \frac{\boldsymbol{s}^{(m)^T}\widetilde{\boldsymbol{X}}^{(m)^T}\boldsymbol{a}_m}{\boldsymbol{s}^{(m)^T}\boldsymbol{R}\boldsymbol{s}^{(m)}} \tag{7}$$

Using based on Eq. 1-7, we propose the following code.

Algorithm 1: Pseudo-code of the first NMF

Input: ℓ_1-normalized data matrix: $\widetilde{\boldsymbol{X}} \in {\mathbb{R}_+}^{NT\times D}$, orthogonal constraint weight: w_O, initial component number: M, iteration number: itr, margining threshold: $t \in [0.9, 0.99]$.

Output: spectra-wise fragment abundance: $\widetilde{\boldsymbol{A}} \in {\mathbb{R}_+}^{NT\times M}$, M-fragment spectra: $\boldsymbol{S} \in {\mathbb{R}_+}^{M\times D}$.

Initialization

initialize $\widetilde{\boldsymbol{A}}$ by Eq. 1

initialize $\boldsymbol{S}$ by random selection of M-row vectors from $\widetilde{\boldsymbol{X}}$ and ℓ_2-normalization.

calculate b by Eq. 2

initialize $\boldsymbol{\lambda}$ by Eq. 3

calculate $\boldsymbol{E}$ by Eq. 4 and $\boldsymbol{R} = \frac{1}{NT}\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{E}$

Repeat until convergence criteria are satisfied:

for $m = 1, \ldots, M$:

calculate $\boldsymbol{s}^{(m)} \leftarrow \sum_{j \neq m}^{M} \boldsymbol{s}_j$ and calculate ξ_m by Eq. 7

update $\boldsymbol{s}_m$ by Eq. 6.

project $\boldsymbol{s}_m$ to the non-negative orthant

ℓ_2-normalize $\boldsymbol{s}_m$

for $m = 1, \ldots, M$:

update $\boldsymbol{a}_m$ by Eq. 5

project $\boldsymbol{a}_m$ to the non-negative orthant

Merging very similar components

for $k = 1, \ldots, M-1$:

for $m = k+1, \ldots, M$:

if $\boldsymbol{s}_k^T \boldsymbol{s}_m > t$:

$\boldsymbol{a}_k \leftarrow \boldsymbol{a}_k + \boldsymbol{a}_m,\ \boldsymbol{a}_m \leftarrow \boldsymbol{0}$

if $\boldsymbol{a}_k^T \boldsymbol{a}_m > t\|\boldsymbol{a}_k\|\|\boldsymbol{a}_m\|$:

$\boldsymbol{s}_k \leftarrow \boldsymbol{s}_k + \frac{\|\boldsymbol{a}_m\|_1}{\|\boldsymbol{a}_k\|_1}\boldsymbol{s}_m,\ \ s = \|\boldsymbol{s}_k\|_2,\ \ \boldsymbol{s}_k \leftarrow \frac{\boldsymbol{s}_k}{s}$

$\boldsymbol{a}_k \leftarrow s\boldsymbol{a}_k, \boldsymbol{a}_m \leftarrow \boldsymbol{0}$

update $\boldsymbol{\lambda}$ by Eq. 3

return $\widetilde{\boldsymbol{A}}$ and $\boldsymbol{S}$

Note that after each update of $\widetilde{\boldsymbol{A}}$ and $\boldsymbol{S}$, the components having similar spectra or abundance distribution over the dataset are merged, reducing the component number down. By merging the components based on not only their spectra but also abundance distribution, a series of fragments peaks belonging to an identical codon would be unified in a single spectrum, deriving more interpretable outputs.

Derivation of the canonical correlation analysis filter (CCA-filter)

Backgrounds and/or contaminants are potentially included in the *M*-components output from the first NMF, which may distort the analytic results. We here describe our development of CCA-filter, automatically identifying and removing the contaminated components from the *M*-components. After removing *M'*-contaminants, $\widetilde{\boldsymbol{A}} \in {\mathbb{R}_+}^{NT\times M}$ and $\boldsymbol{S} \in {\mathbb{R}_+}^{M\times D}$ are respectively reduced to $\widetilde{\boldsymbol{A}} \in {\mathbb{R}_+}^{NT\times(M-M')}$ and $\boldsymbol{S} \in {\mathbb{R}_+}^{(M-M')\times D}$. For simplicity, in the main text and following second NMF derivation, M rather than $M - M'$ is consistently used for representing the component number.

To use CCA-filter, a background spectrum $\boldsymbol{X}_{BG} \in {\mathbb{R}_+}^{T\times D}$ needs to be included in a dataset before the first NMF is applied. If some contaminants are expected, e.g., residual solvents in cast-film samples, they can be mixed and measured in a single spectrum, which can be used as $\boldsymbol{X}_{BG}$. Conceptually, CCA-filter finds out *M'*-spectra from ℓ_2-normalized $\boldsymbol{S} \in {\mathbb{R}_+}^{M\times D}$, of which peak patterns are also included in $\boldsymbol{X}_{BG}$. Assume we would like to examine *m*-th component if it is contaminant or fragment generated from the samples. The first step

is classifying M-spectral set $\boldsymbol{S}$ into two groups: spectral sets $\boldsymbol{Y} \in \mathbb{R}_+^{M_{sim} \times D}$ and $\boldsymbol{Z} \in \mathbb{R}_+^{M_{dis} \times D}$, respectively similar and dissimilar to $\boldsymbol{S}_{m:}$. Here, M_{sim} and M_{dis} are the numbers of similar and dissimilar spectra, respectively. The classification is based on cosine similarity. After the classification, the following inequalities should be hold:

$$\boldsymbol{S}_{m:}\boldsymbol{Y}_{m':}^{T} \geq t_1, for\ m' = 1, \dots, M_{sim},$$

$$\boldsymbol{S}_{m:}\boldsymbol{Z}_{m':}^{T} < t_1, for\ m' = 1, \dots, M_{dis},$$

where $t_1 \in [0, 1]$ is a threshold, in this study fixed at $t_1 = 0.2$. Note that $\boldsymbol{S}_{m:}$ is always classified into $\boldsymbol{Y}$, and put at the top row of $\boldsymbol{Y}$. The dissimilar spectral set $\boldsymbol{Z}$ is further combined with $\boldsymbol{X}_{BG}$:

$$\boldsymbol{Z} \leftarrow \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{X}_{BG} \end{pmatrix}.$$

$\boldsymbol{Y}$ and $\boldsymbol{Z}$ are then mean-subtracted:

$$\overline{\boldsymbol{Y}} = \boldsymbol{Y}\left(\boldsymbol{I}_D - \frac{1}{D}\boldsymbol{1}\boldsymbol{1}_D\right) \in \mathbb{R}^{M_{sim} \times D}, \tag{8}$$

$$\overline{\boldsymbol{Z}} = \boldsymbol{Z}\left(\boldsymbol{I}_D - \frac{1}{D}\boldsymbol{1}\boldsymbol{1}_D\right) \in \mathbb{R}^{(M_{dis}+T) \times D}. \tag{9}$$

CCA is then applied to the two spectral set. Generally, the purpose of CCA is to generate a pair of spectra most similar to each other from two different spectral sets by linear combination, here $\overline{\boldsymbol{Y}}$ and $\overline{\boldsymbol{Z}}$. Let two coefficient vectors be $\boldsymbol{u} \in \mathbb{R}^{M_{sim}}$ and $\boldsymbol{v} \in \mathbb{R}^{M_{dis}+T}$ for $\overline{\boldsymbol{Y}}$ and $\overline{\boldsymbol{Z}}$, respectively. The new mixed spectra are defined by $\boldsymbol{y} \equiv \boldsymbol{u}^T\overline{\boldsymbol{Y}} \in \mathbb{R}^{1 \times D}$ and $\boldsymbol{z} \equiv \boldsymbol{v}^T\overline{\boldsymbol{Z}} \in \mathbb{R}^{1 \times D}$. The similarity of these two spectra is evaluated with correlation coefficient:

$$\rho = \frac{\boldsymbol{u}^T \boldsymbol{V}_{yz} \boldsymbol{v}}{\sqrt{\boldsymbol{u}^T \boldsymbol{V}_{yy} \boldsymbol{u}} \sqrt{\boldsymbol{v}^T \boldsymbol{V}_{zz} \boldsymbol{v}}},$$

where $\boldsymbol{V}_{YY} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}^T/D, \boldsymbol{V}_{ZZ} = \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^T/D, \boldsymbol{V}_{YZ} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Z}}^T/D$. The CCA problem now becomes:

$$(\boldsymbol{u}^*, \boldsymbol{v}^*) = \underset{\boldsymbol{u},\boldsymbol{v}}{arg\,max}\, \rho.$$

The optimum $\binom{\boldsymbol{u}^*}{\boldsymbol{v}^*} \in \mathbb{R}^{M_{sim}+M_{dis}+T}$ are given as the solution of generalized eigenvalue problem:

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{V}_{yz} \\ \boldsymbol{V}_{yz}^T & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_{yy} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{zz} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_{M_{sim}+M_{dis}+T} \end{pmatrix} \tag{10}$$

where $\binom{\boldsymbol{U}}{\boldsymbol{V}}$ consists of all the eigen column-vectors $\binom{\boldsymbol{u}^*}{\boldsymbol{v}^*}$ in a descending order of the corresponding eigenvectors, i.e., $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_{M_{sim}+M_{dis}+T}$. Each eigen value represents the correlation coefficient of the paired spectra of $\boldsymbol{y}$ and $\boldsymbol{z}$, synthesized via corresponding eigen vectors $\binom{\boldsymbol{u}^*}{\boldsymbol{v}^*}$. Again, all the basis spectra similar to the examined $\boldsymbol{S}_{m:}$ are contained in $\boldsymbol{Y}$, while the background *T*-spectra are contained in $\boldsymbol{Z}$. If $\boldsymbol{Y}$ and $\boldsymbol{Z}$ can generate very similar spectra, $\boldsymbol{S}_{m:}$ is suspected of being derived from backgrounds or contaminations. Practically, we collect all the eigenvectors $\binom{\boldsymbol{u}^*}{\boldsymbol{v}^*}$ of which eigenvalues satisfy $\rho > t_2$ for given $t_2 \in [0.9, 0.99]$. If the first element of $\boldsymbol{u}^*$ corresponding to the coefficient of $\boldsymbol{S}_{m:}$ significantly contributes to $\boldsymbol{u}^*$, i.e., if $\frac{|u^*{}_1|}{\|\boldsymbol{u}^*\|_1} \geq t_3$ for $t_3 \in [0, 1]$, then the *m*-th component is judged as a background/contamination component and removed from the system. The whole process is summarized in Algorithm 2.

Algorithm 2: CCA-filter

Input: ℓ_1-normalized basis-spectra of the 1stNMF output; $\boldsymbol{S} \in {\mathbb{R}_+}^{M \times D}$, a background spectrum; $\boldsymbol{X}_{BG} \in {\mathbb{R}_+}^{T \times D}$, thresholds: $t_1 \in [0,1], t_2 \in [0,1], t_3 \in [0,1]$, in this study $(t_1, t_2, t_3) = (0.2, 0.9, 0.5)$ is consistently used.

Output: a list of the components judged as background components.

for $m = 1, \ldots, M$:

  for $m' = 1, \ldots, M$:

    classify $\boldsymbol{S}_{m':}$ into $\boldsymbol{Y}$ if $\boldsymbol{S}_{m:}{\boldsymbol{S}_{m':}}^T \geq t_1$ else into $\boldsymbol{Z}$

$\boldsymbol{Z} \leftarrow \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{X}_{BG} \end{pmatrix}$

calculate $\overline{\boldsymbol{Y}}$ and $\overline{\boldsymbol{Z}}$ by Eq. 8-9

calculate $\boldsymbol{V}_{YY} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}^T/D, \boldsymbol{V}_{ZZ} = \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^T/D, \boldsymbol{V}_{YZ} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Z}}^T/D$

obtain $\boldsymbol{U}^* = ({\boldsymbol{u}^*}_1, \ldots, {\boldsymbol{u}^*}_{M_{sim}+M_{dis}+T})$ and $(\rho_1, .., \rho_{M_{sim}+M_{dis}+T})$ by solving Eq. 10

find Q such that $\rho_Q \geq t_2 \; and \; \rho_{Q+1} < t_2$

for $q = 1, \ldots, Q$:

  if $\frac{|{\boldsymbol{U}^*}_{1q}|}{\|\boldsymbol{U}^{*}_{:q}\|_1} \geq t_3$, add the component m in a background list

return background list

All the components attributed to backgrounds/contaminants are then removed from the columns of $\widetilde{\boldsymbol{A}}$ and rows of $\boldsymbol{S}$.

Non-negative least square fitting

In the following section, we frequently use non-negative least square (NNLS) fitting. This

finds the best non-negative coefficient vector $\boldsymbol{x} \in \mathbb{R}_+{}^n$ for approximating $\boldsymbol{y} \in \mathbb{R}^m$ as a linear combination of *n*-column vectors of a constant matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$, i.e.,

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|^2, s.t. \boldsymbol{x} \geq \boldsymbol{0}.$$

Various algorithms have been proposed to quickly solve this problem, including alternating direction multiplier methods (ADMM) [8]. We here use ADMM-NNLS proposed by Fu. *et. al.* [9], and use a notation of $\boldsymbol{x}^* = NNLS(\boldsymbol{y}, \boldsymbol{\Phi})$. For each of *L*-vectors set, $\boldsymbol{Y} = [\boldsymbol{y}_1, .., \boldsymbol{y}_L]$, the corresponding coefficient vector $\boldsymbol{x}_l{}^* \ (l = 1, .., L)$ is separately calculated by:

$$\boldsymbol{x}_l{}^* = NNLS(\boldsymbol{Y}_{:l}, \boldsymbol{\Phi}) \ for \ l = 1, .., L$$

This is simply written in a matrix form as:

$$\boldsymbol{X}^* = NNLS(\boldsymbol{Y}, \boldsymbol{\Phi})$$

where $\boldsymbol{X}^* = [\boldsymbol{x}_1{}^*, .., \boldsymbol{x}_L{}^*]$ is the set of optimal non-negative coefficient vectors. Also, a similar problem with sum-to-one constraints:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|^2, s.t. \boldsymbol{x} \geq \boldsymbol{0}, \boldsymbol{x}^T \boldsymbol{1_n} = 1.$$

can be solved by ADMM as proposed by Fu. *et. al.* [9]. This is called fully constrained least square (FCLS) [10], whose solution is written as:

$$\boldsymbol{x}_l{}^* = FCLS(\boldsymbol{Y}_{:l}, \boldsymbol{\Phi}) \ for \ l = 1, .., L,$$

or in a matrix form:

$$\boldsymbol{X}^* = FCLS(\boldsymbol{Y}, \boldsymbol{\Phi}).$$

Derivation of the second NMF

The first NMF outputs spectrum-wise fragment abundance (FA) $\widetilde{\boldsymbol{A}} \in \mathbb{R}^{NT \times M}$, which is

converted to $\boldsymbol{A} \in \mathbb{R}_+^{N\times M}$ by sample-wise *T*-spectra integration (Fig. S4), as temperature-independency of FA is unnecessary for compositional analysis. NMF still should be applied to $\widetilde{\boldsymbol{X}} \in \mathbb{R}_+^{NT\times D}$ rather than $\boldsymbol{X} \in \mathbb{R}_+^{N\times D}$, i.e., $\boldsymbol{X} \approx \boldsymbol{AS}$ for straightforwardly deriving $\boldsymbol{A}$, as the temperature-dependent spectral modulation facilitates identifying fragment-spectra $\boldsymbol{S}$ (Fig. S5). The second NMF is applied to $\boldsymbol{A}$, i.e., $\boldsymbol{A} \approx \boldsymbol{CB}$. The factorized $\boldsymbol{C} \in \mathbb{R}_+^{N\times K}$ represents *K*-polymers weight fraction in *N*-samples, and $\boldsymbol{B} \in \mathbb{R}_+^{K\times M}$ represents FAs of pure *K*-polymers. To emphasize the row stochastic condition of $\boldsymbol{C}$, the rows of $\boldsymbol{A}$ is ℓ_1-normalized in advance [11]:

$$\boldsymbol{A} \leftarrow \boldsymbol{L_A}^{-1}\boldsymbol{A},$$

where $\boldsymbol{L_A} = diag(\|\boldsymbol{A}_{1:}\|_1, \dots, \|\boldsymbol{A}_{N:}\|_1)$, which is later used for determining spectral norm of basis FAs, $\boldsymbol{B}$. We evaluate the approximate residuals by Riemann metrics, i.e.,

$$D_{\boldsymbol{G}}(\boldsymbol{A}|\boldsymbol{CB}) = Tr[(\boldsymbol{A}-\boldsymbol{CB})\boldsymbol{G}(\boldsymbol{A}-\boldsymbol{CB})^T],$$

where $\boldsymbol{G} = \boldsymbol{SS}^T \in \mathbb{R}_+^{M\times M}$ [12]. Note that rows of $\boldsymbol{S}$ should be ℓ_2-normalized in advance for calculating $\boldsymbol{G}$. A lower triangular matrix $\boldsymbol{L} \in \mathbb{R}^{M\times M}$ is obtained via Cholesky decomposition, $\boldsymbol{G} = \boldsymbol{LL^T}$ [13]. The metrics can be rewritten as:

$$D_{\boldsymbol{G}}(\boldsymbol{A}|\boldsymbol{CB}) = Tr[(\boldsymbol{A}-\boldsymbol{CB})\boldsymbol{LL^T}(\boldsymbol{A}-\boldsymbol{CB})^T] = Tr\left[\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}\widehat{\boldsymbol{B}}\right)^T\right] = \left\|\widehat{\boldsymbol{A}}-\boldsymbol{C}\widehat{\boldsymbol{B}}\right\|_F^2$$

where $\widehat{\boldsymbol{A}} = \boldsymbol{AL} \in \mathbb{R}^{N\times M}, \widehat{\boldsymbol{B}} = \boldsymbol{BL} \in \mathbb{R}^{K\times M}$. Thus, conducting NMF, $\boldsymbol{A} \approx \boldsymbol{CB}$, with residuals evaluation in Riemann metrics is equivalent to factorizing $\widehat{\boldsymbol{A}} \approx \boldsymbol{C}\widehat{\boldsymbol{B}}$ in Euclidean space. The optimization problem now becomes:

$$\min_{\boldsymbol{C},\widehat{\boldsymbol{B}}} \frac{1}{2}\left\|\widehat{\boldsymbol{A}}-\boldsymbol{C}\widehat{\boldsymbol{B}}\right\|_F^2 + \frac{\alpha}{2}vol\left(\widehat{\boldsymbol{B}}\right) + \frac{\beta}{2}nonorth\left(\widehat{\boldsymbol{B}}\right), \quad (11)$$

$$s.t. \widehat{\boldsymbol{B}} = \boldsymbol{BL}, \boldsymbol{B} \geq 0, \boldsymbol{C} \geq 0, \boldsymbol{C}\boldsymbol{1}_K = \boldsymbol{1}_N.$$

where $vol(\widehat{\boldsymbol{B}})$ is the volume term of the simplex spanned by row vectors of $\widehat{\boldsymbol{B}}$ [14], $nonorth(\widehat{\boldsymbol{B}})$ is a non-orthogonality term of row vectors of $\widehat{\boldsymbol{B}}$ [15], and $\alpha > 0, 1 > \beta > 0$ are balancing parameters. The volume-minimization term, $vol(\widehat{\boldsymbol{B}})$, is introduced by Fu *et al.* [14]:

$$vol(\widehat{\boldsymbol{B}}) = \log\left|\det\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T + \tau \boldsymbol{I}_K\right)\right| = \log|\det(\boldsymbol{H})|,$$

where $\boldsymbol{H}(\widehat{\boldsymbol{B}}) = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T + \tau \boldsymbol{I}_K$ and τ is a small regularization parameter, here fixed at $\tau = 10^{-8}$. This volume term ensures ground-truth identifiability under "sufficiently scattered" condition [14]. Roughly speaking in chemistry words, if every component has several high-purity samples in a dataset, over 50 wt% at least, the unique solution $\boldsymbol{C}^*$ coincides with the true composition. As this criterion is often violated in practical datasets, we found that introducing $nonorth(\widehat{\boldsymbol{B}})$ would be helpful, particularly for biased dataset (Fig. S2, here, $\alpha = \beta = 0.1$). The non-orthogonality term is defined for penalizing non-diagonal non-zero elements [15], i.e.:

$$nonorth(\widehat{\boldsymbol{B}}) = Tr\left(\boldsymbol{\Lambda}\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T - diag(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T)\right)\right),$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{K\times K}$ is a symmetric Lagrange multiplier matrix. Importantly, the orthogonal constraints are imposed on $\widehat{\boldsymbol{B}}$ rather than $\boldsymbol{B}$, restricting heavy peak-overlapping among different polymers in the original spectral space, rational in MS. In contrast, the non-negative constraints are imposed on $\boldsymbol{B}$ rather than $\widehat{\boldsymbol{B}}$, as $\boldsymbol{L}$ and thus $\widehat{\boldsymbol{B}}$ may take negative elements. Fu *et. al.* [14] reported that volume-minimized factorization is very sensitive to outliers, and suggested to discount the importance of outliers by introducing a weight matrix [14]:

$$w_n = \frac{p}{2}\left(\left\|\widehat{\boldsymbol{A}}_{n:} - \boldsymbol{C}_{n:}\widehat{\boldsymbol{B}}\right\|^2 + \varepsilon\right)^{\frac{p-2}{2}} \tag{12}$$

$$\boldsymbol{W} = diag(w_1, \ldots w_N)$$

where $p \in (0,2]$ and ε is a small regularization positive number. Smaller p more strongly discounts the samples with large fitting error. In this work, $p \in \{1, 1.5\}$ and $\varepsilon = 10^{-8}$ were used. With $\boldsymbol{W}$, the optimization problem is rewritten as:

$$\min_{\boldsymbol{C},\widehat{\boldsymbol{B}}} \frac{1}{2} Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right)^T\right] + \frac{\alpha}{2} vol\left(\widehat{\boldsymbol{B}}\right) + \frac{\beta}{2} nonorth\left(\widehat{\boldsymbol{B}}\right),$$

$$s.t. \widehat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{L}, \boldsymbol{B} \geq 0, \boldsymbol{C} \geq 0, \boldsymbol{C}\boldsymbol{1}_K = \boldsymbol{1}_N.$$

We solved this optimization by updating $\boldsymbol{C}$ and $\widehat{\boldsymbol{B}}$ alternatively based on block coordinate descent (BCD) theory [16]. Assume we have $(\boldsymbol{C}^{(t)}, \widehat{\boldsymbol{B}}^{(t)})$ after t^{th} updating. The initial $\widehat{\boldsymbol{B}}^{(0)}$ can be selected out from the N-rows of $\widehat{\boldsymbol{A}}$ via vertex component analysis (VCA) [17]. As the penalty terms are not related to $\boldsymbol{C}$, updating $\boldsymbol{C}$ based on a fixed $\widehat{\boldsymbol{B}}^{(t)}$ is straightforward by using the forementioned ADMM-FCLS, i.e.:

$$\boldsymbol{C}^{(t+1)^T} = FCLS\left(\widehat{\boldsymbol{A}}^T, \widehat{\boldsymbol{B}}^{(t)^T}\right). \tag{13}$$

Note that $\boldsymbol{W}$ is a diagonal matrix and thus negligible for updating $\boldsymbol{C}$. Now we focus on updating $\widehat{\boldsymbol{B}}$ based on the fixed $\boldsymbol{C}^{(t)}$. To efficiently obtain the updating rule from $\widehat{\boldsymbol{B}}^{(t)}$ to $\widehat{\boldsymbol{B}}^{(t+1)}$, majorizer function of $vol\left(\widehat{\boldsymbol{B}}\right)$ is introduced. After t^{th} update, based on fixed $\boldsymbol{H}^{(t)} = \boldsymbol{H}(\widehat{\boldsymbol{B}}^{(t)})$, the following tangent inequality holds:

$$\log|\det(\boldsymbol{H})| \leq \log\left|\det\left(\boldsymbol{H}^{(t)}\right)\right| + Tr\left[\left(\boldsymbol{\nabla}_{\boldsymbol{H}^{(t)}} \log|\det(\boldsymbol{H})|\right)^T\left(\boldsymbol{H} - \boldsymbol{H}^{(t)}\right)\right],$$

here $\boldsymbol{\nabla}_{\boldsymbol{H}^{(t)}} \log|\det(\boldsymbol{H})|$ is the gradient of $\log|\det(\boldsymbol{H})|$ over $\boldsymbol{H}$ at $\boldsymbol{H}^{(t)}$. As

$\nabla_{\boldsymbol{H}^{(t)}} \log|\det(\boldsymbol{H})| = \boldsymbol{H}^{(t)^{-T}}$, $vol(\widehat{\boldsymbol{B}})$ can be majorized as:

$$vol(\widehat{\boldsymbol{B}}) = \log|\det(\boldsymbol{H})| \leq Tr\left[\boldsymbol{H}^{(t)^{-1}}\boldsymbol{H}\right] + const = Tr\left[\boldsymbol{F}^{(t)}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right] + const,$$

where $\boldsymbol{F}^{(t)} = \boldsymbol{H}^{(t)^{-1}}$ and $const$ is a constant term not related to $\widehat{\boldsymbol{B}}$. After replacing $vol(\widehat{\boldsymbol{B}})$ with the majorizer $Tr\left[\boldsymbol{F}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right]$, all the penalty terms are combined as:

$$\frac{\alpha}{2}vol(\widehat{\boldsymbol{B}}) + \frac{\beta}{2}nonorth(\widehat{\boldsymbol{B}}) \leq \frac{\alpha}{2}Tr\left(\boldsymbol{F}^{(t)}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) + \frac{\beta}{2}Tr\left(\boldsymbol{\Lambda}^{(t)}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) + const$$

$$= \frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) + const,$$

where $\boldsymbol{V} = \alpha\boldsymbol{F}^{(t)} + \beta\boldsymbol{\Lambda}^{(t)}$. $\widehat{\boldsymbol{B}}^{(t+1)}$ is then updated by solving the following problem:

$$\widehat{\boldsymbol{B}}^{(t+1)} = \arg\min_{\widehat{\boldsymbol{B}}} \frac{1}{2}Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)^T\right] + \frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) \qquad (14)$$

$$s.t.\, \widehat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{L}, \boldsymbol{B} \geq 0.$$

To simplify the complicated constraints, $\widehat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{L}$ is embedded into the objective function by using Lagrangian multiplier $\boldsymbol{Z} \in \mathbb{R}^{K\times M}$ in a framework of ADMM:

$$\left(\boldsymbol{B}^{(t+1)}, \widehat{\boldsymbol{B}}^{(t+1)}\right) = \arg\min_{\boldsymbol{B},\widehat{\boldsymbol{B}}} \max_{\boldsymbol{Z}} \left\{f(\widehat{\boldsymbol{B}}) + Tr\left[\boldsymbol{Z}^T\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}\boldsymbol{L}\right)\right] - \frac{1}{2\mu}\|\boldsymbol{Z} - \boldsymbol{Z}'\|_F^2\right\},$$

$$s.t.\, \boldsymbol{B} \geq 0,$$

where $f(\widehat{\boldsymbol{B}}) \equiv \frac{1}{2}Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)^T\right] + \frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right)$ and μ is a hyperparameter for ADMM (in this work, $\mu = 1$ is consistently used.). Here, $\boldsymbol{Z}'$ represents $\boldsymbol{Z}$ in the previous cycle. The objective function is maximized over $\boldsymbol{Z}$ when

$$\boldsymbol{Z} = \boldsymbol{Z}' + \mu\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}\boldsymbol{L}\right). \qquad (15)$$

The maximized objective function over $\boldsymbol{Z}$ now becomes

$$g\left(\boldsymbol{B}, \widehat{\boldsymbol{B}}; \boldsymbol{Z}\right) \equiv f\left(\widehat{\boldsymbol{B}}\right) + \frac{\mu}{2}\left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}\boldsymbol{L} + \frac{1}{\mu}\boldsymbol{Z}\right\|_F^2, \boldsymbol{B} \geq 0,$$

which is then minimized over $\left(\boldsymbol{B}, \widehat{\boldsymbol{B}}\right)$. $\left(\boldsymbol{B}, \widehat{\boldsymbol{B}}, \boldsymbol{Z}\right)$ is thus cyclically updated via Algorithm3.

Algorithm 2: ADMM for solving the optimization of Eq. 14

input: $\boldsymbol{B}^{(t)}, \widehat{\boldsymbol{B}}^{(t)}$, hyperparameter μ ($\mu = 1$ consistently used in this study), function $g\left(\boldsymbol{B}, \widehat{\boldsymbol{B}}\right)$

output: $\boldsymbol{B}^{(t+1)}, \widehat{\boldsymbol{B}}^{(t+1)}$

initialize: $q = 0, \boldsymbol{B}_q \leftarrow \boldsymbol{B}^{(t)}, \widehat{\boldsymbol{B}}_q \leftarrow \widehat{\boldsymbol{B}}^{(t)}, \boldsymbol{Z}_q \leftarrow \boldsymbol{0}$

repeat until convergence:

$\widehat{\boldsymbol{B}}_{q+1} = \underset{\widehat{\boldsymbol{B}}}{\operatorname{argmin}}\, g\left(\widehat{\boldsymbol{B}}; \boldsymbol{B}_q, \boldsymbol{Z}_q\right)$, solved by Eq. 16 as described below

$\boldsymbol{B}_{q+1} = \underset{\boldsymbol{B} \geq \boldsymbol{0}}{\operatorname{argmin}}\, g\left(\boldsymbol{B}; \widehat{\boldsymbol{B}}_{q+1}, \boldsymbol{Z}_q\right)$, solved by Eq. 17 as described below

$\boldsymbol{Z}_{q+1} = \boldsymbol{Z}_q + \mu\left(\widehat{\boldsymbol{B}}_{q+1} - \boldsymbol{B}_{q+1}\boldsymbol{L}\right)$ (Eq. 15)

$q \leftarrow q + 1$

$\boldsymbol{B}^{(t+1)} \leftarrow \boldsymbol{B}_q, \widehat{\boldsymbol{B}}^{(t+1)} \leftarrow \widehat{\boldsymbol{B}}_q$.

Return $\boldsymbol{B}^{(t+1)}, \widehat{\boldsymbol{B}}^{(t+1)}$

As the objective function $g\left(\widehat{\boldsymbol{B}}; \boldsymbol{B}_q, \boldsymbol{Z}_q\right)$ has a simple quadratic form without constraints about $\widehat{\boldsymbol{B}}$, $\widehat{\boldsymbol{B}}_{q+1}$ can be determined in a closed form by setting the derivative of g zero, i.e.:

$$\frac{\partial g}{\partial \widehat{\boldsymbol{B}}} = \left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\boldsymbol{C}^{(t+1)} + \boldsymbol{V} + \mu\boldsymbol{I}_K\right)\widehat{\boldsymbol{B}} - \boldsymbol{C}^{(t)^T}\boldsymbol{W}\widehat{\boldsymbol{A}} - \boldsymbol{B}_q\boldsymbol{L} + \frac{1}{\mu}\boldsymbol{Z} \equiv \boldsymbol{0}$$

$$\widehat{\boldsymbol{B}}_{q+1} = \left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\boldsymbol{C}^{(t)} + \boldsymbol{V} + \mu\boldsymbol{I}_K\right)^{-1}\left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\widehat{\boldsymbol{A}} + \boldsymbol{B}_q\boldsymbol{L} - \frac{1}{\mu}\boldsymbol{Z}\right). \quad (16)$$

As $g\left(\boldsymbol{B}; \widehat{\boldsymbol{B}}_{q+1}, \boldsymbol{Z}_q\right)$ has a simple quadratic form about $\boldsymbol{B}$ with non-negative constrains, the NNLS algorithm can be directly applied, i.e.:

$$\boldsymbol{B}_{q+1}{}^T = NNLS\left(\left(\widehat{\boldsymbol{B}}_{q+1} + \frac{1}{\mu}\boldsymbol{Z}\right)^T, \boldsymbol{L}^T\right). \quad (17)$$

Updating rules for $\left(\widehat{\boldsymbol{B}}, \boldsymbol{B}, \boldsymbol{C}\right)$ to solve the original optimization problem (Eq. 11) have been obtained so far, but updating rule for the multiplier matrix $\boldsymbol{V} = \alpha\boldsymbol{F}^{(t)} + \beta\boldsymbol{\Lambda}^{(t)}$ remains unknown. $\boldsymbol{F}^{(t)}$ is directly calculated from $\widehat{\boldsymbol{B}}^{(t)}$:

$$\boldsymbol{F}^{(t)} = \left(\widehat{\boldsymbol{B}}^{(t)}\widehat{\boldsymbol{B}}^{(t)^T} + \tau\boldsymbol{I}_K\right)^{-1}. \quad (18)$$

$\boldsymbol{\Lambda}^{(t)}$ is estimated from strict orthogonal conditions, i.e., $\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T = diag\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) \equiv \boldsymbol{D}$, $\beta = 1$ and

$$\frac{\partial f\left(\widehat{\boldsymbol{B}}; \boldsymbol{C}^{(t)}\right)}{\partial \widehat{\boldsymbol{B}}} = \boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}} - \widehat{\boldsymbol{A}}\right) + \left(\alpha\boldsymbol{F}^{(t)} + \boldsymbol{\Lambda}^{(t)}\right)\widehat{\boldsymbol{B}} \equiv 0. \quad (19)$$

The multiplier matrix $\boldsymbol{\Lambda}^{(t)}$ at the strcit condition is obtained by multiplying $\widehat{\boldsymbol{B}}^T$ from the right side of Eq. 19:

$$\boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\boldsymbol{C}^{(t)}\boldsymbol{D} - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\right) + \left(\alpha\boldsymbol{F}^{(t)} + \boldsymbol{\Lambda}^{(t)}\right)\boldsymbol{D} \equiv 0,$$

$$\boldsymbol{\Lambda}^{(t)} = \boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\boldsymbol{D}^{-1} - \boldsymbol{C}^{(t)}\right) - \alpha\boldsymbol{F}^{(t)}. \quad (20)$$

By combining Eq. 18 and Eq. 20, updating rule of $\boldsymbol{V}$ is obtained as:

$$\boldsymbol{V} = \alpha(1-\beta)\left(\widehat{\boldsymbol{B}}^{(t)}\widehat{\boldsymbol{B}}^{(t)^T} + \tau\boldsymbol{I}_K\right)^{-1} + \beta\boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\boldsymbol{D}^{-1} - \boldsymbol{C}^{(t)}\right). \quad (21)$$

The volume penalty term is interpretable as a shrinking force imposed on the simplex

spanned by the rows of $\widehat{\boldsymbol{B}}$ [18]. Conversely, the orthogonal term is an expanding force, as evident from Eq. 21; the weight $\beta \in [0, 1]$ buffers the first term of volume-minimization. As the residual-minimization term, $\left\|\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right\|_F^2$, is also interpretable as expanding force [18], solving the original problem of Eq. 11 is conceptually identical to finding the equilibrium point of the three expanding and shrinking forces imposed on the simplex. Based on the updating rules, we propose the following algorithm to solve the second NMF.

Algorithm 3: Pseudo-code for the second NMF for solving Eq. 11

Input: output of the first NMF (sample-wise FA: $\boldsymbol{A} \in {\mathbb{R}_+}^{N \times M}$, fragment spectra: $\boldsymbol{S} \in {\mathbb{R}_+}^{M \times D}$), basis polymer number: *K*, weights for penalty terms: (α, β), weight for outliers: *p*

Output: polymer weight fraction: $\boldsymbol{C} \in {\mathbb{R}_+}^{N \times K}$, FAs of basis polymers per unit weight: $\boldsymbol{B} \in {\mathbb{R}_+}^{K \times M}$

Initialization

calculate $\boldsymbol{L}$ via Cholesky decomposition of $\boldsymbol{S}\boldsymbol{S}^T$

set $\widehat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{L}$

initialize $\widehat{\boldsymbol{B}}^{(0)}$ by selecting out *K*-rows from $\widehat{\boldsymbol{A}}$ via VCA algorithm

set $\boldsymbol{B}^{(0)} = \widehat{\boldsymbol{B}}^{(0)}\boldsymbol{L}^{-1}$

initialize $\boldsymbol{C}^{(0)}$ by ${\boldsymbol{C}^{(0)}}^T = FCLS({\widehat{\boldsymbol{A}}^T}, {\widehat{\boldsymbol{B}}^{(0)}}^T)$

initialize $\boldsymbol{W}$ by Eq. 12

initialize $\boldsymbol{V}$ based on $\boldsymbol{B}^{(0)}$ and $\boldsymbol{C}^{(0)}$ by Eq. 21

Repeat until convergence criteria is satisfied:

update $\widehat{\boldsymbol{B}}, \boldsymbol{B}$ by algorithm 3

update $\boldsymbol{C}$ by Eq. 13

update $\boldsymbol{W}, \boldsymbol{V}$ by Eq. 12 and S18, respectively

return $\boldsymbol{C} \in {\mathbb{R}_+}^{N \times K}$ and $\boldsymbol{B}$

As row vectors of $\boldsymbol{A}$ have been ℓ_1-normalized in advance, the factorized $\boldsymbol{B}$ are also row stochastic in noiseless cases [11]. We still need to recover the norms of row vectors of $\boldsymbol{B}$ based on $\boldsymbol{L_A}$, which are the preserved norms of $\boldsymbol{A}$ rows before being normalized. For the norm-recovered $\boldsymbol{L_A A}$, the following equation holds:

$$\boldsymbol{L_A A} \approx \boldsymbol{L_A C B} = \boldsymbol{L_A C}\left({\boldsymbol{L_{B'}}}^{-1}\boldsymbol{B'}\right) = \left(\boldsymbol{L_{\widetilde{A}} C}{\boldsymbol{L_{B'}}}^{-1}\right)\boldsymbol{B'} = \boldsymbol{C'B'},$$

where $\boldsymbol{B'} = \boldsymbol{L_{B'}B}$ is the norm-recovered variant of $\boldsymbol{B}$, $\boldsymbol{L_{B'}} = diag(\|{\boldsymbol{B'}}_{1:}\|_1, \dots, \|{\boldsymbol{B'}}_{N:}\|_1)$ is the spectral norm of the basis spectral, and $\boldsymbol{C'} = \boldsymbol{L_{\widetilde{A}} C}{\boldsymbol{L_{B'}}}^{-1}$ is the fraction of the norm-recovered basis $\boldsymbol{B'}$. Because the row-stochastic conditions of $\boldsymbol{C'}$ still should be hold if the spectral intensities of raw spectra $\widetilde{\boldsymbol{X}}$ have been properly corrected (see the section of "Spectra Formatting"), $\boldsymbol{L_{B'}}$ can be determined so that $\boldsymbol{C'}\mathbf{1}_K = \mathbf{1}_N$ would be best satisfied. For sequencing, monomer composition ratio would be also useful to determine $\boldsymbol{L_{B'}}$ particularly when the intensity correction of $\widetilde{\boldsymbol{X}}$ is difficult and row-stochastic conditions of $\boldsymbol{C'}$ are unreliable. The reconstructed basis spectra $\boldsymbol{P}$ with absolute intensities in original *D*-dimensional spectral space thus can be obtained as $\boldsymbol{P} = \boldsymbol{B'S}$ (simply written as $\boldsymbol{P} = \boldsymbol{BS}$ in the main text), which showed good consistency both in spectral shapes and absolute intensities with the observed basis spectra in benchmark test (Fig 2A).

Sequential projection of a target pyrolysis-MS spectrum

Here, we describe how to project a targeted 2D MS spectrum, $\widetilde{\boldsymbol{X}}_t \in {\mathbb{R}_+}^{T\times D}$ of a target copolymer onto the subspaces spanned by the rows of $\boldsymbol{S} \in {\mathbb{R}_+}^{M\times D}$ and subsequently $\boldsymbol{B} \in {\mathbb{R}_+}^{K\times M}$. The first projection onto $\boldsymbol{S}$-subspace can be simply done via NNLS fitting:

$$\widetilde{\boldsymbol{A}}_t{}^T = NNLS\left(\boldsymbol{R}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}}_t{}^T, \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{S}^T\right),$$

where noise covariance $\boldsymbol{R} = \frac{1}{NT}\boldsymbol{E^T E} \in \mathbb{R}^{D\times D}$ adjusts channel-wise importance based on a noise-matrix $\mathbf{E} \in \mathbb{R}^{NT\times D}$ estimated from isotope peaks. The spectrum-wise FA $\widetilde{\boldsymbol{A}}_t$ is integrated along the temperature axis to be converted into $\boldsymbol{A}_t \in {\mathbb{R}_+}^{1\times M}$, which is subsequently projected onto $\boldsymbol{B}$-subspace:

$$\boldsymbol{C}_t{}^T = NNLS\left(\boldsymbol{L}^T\boldsymbol{A}_t{}^T, \boldsymbol{L}^T\boldsymbol{B}^T\right),$$

$$\boldsymbol{C}_t \leftarrow \frac{\boldsymbol{C}_t}{\|\boldsymbol{C}_t\|_1},$$

where $\boldsymbol{C}_t \in {\mathbb{R}_+}^{1\times K}$ is the fraction of K-basis polymers and $\boldsymbol{L} \in \mathbb{R}^{M\times M}$ is obtained by Cholesky decomposition of $\boldsymbol{SS}^T$. Importantly, $\boldsymbol{C}_t$ is scale-invariant owing to ℓ_1-normalization. This means the weight of the target polymer subjected to pyrolysis-MS is unnecessary, allowing direct sequencing without chemical purification. In contrast, sample weights included in dataset and used for learning $\boldsymbol{S}$ and $\boldsymbol{B}$ should be accurately measured, as they influence the absolute intensities of the estimated basis spectra, i.e., $\boldsymbol{P} = \boldsymbol{BS} \in {\mathbb{R}_+}^{K\times D}$.

Prediction of sequence distribution from the monomer reactivity ratio

We here use the following notations; monomer reactivity ratios for monomer 1 and monomer 2: r_1 and r_2; molar fraction of monomer 1 in polymerization solution: *f*; feed molar fraction: f^0; molar fraction of monomer 1 in instantaneously generated polymers: *F*; unpolymerized total monomer concentration: $[M]$; initial monomer concentration: $[M]_0$; monomer conversion: $conv = 1 - \frac{[M]}{[M]_0}$. As well known, Lewis-Mayo equation [19] predicts monomer fraction in instantaneously generated polymers:

$$\frac{F}{1-F} = \frac{f}{1-f}\frac{r_1 f + (1-f)}{f + r_2(1-f)}.$$

By solving the equation, F can be written as a function of f under given r_1, r_2:

$$F = \frac{f + (r_1 - 1)f^2}{(r_1 + r_2 - 2)f^2 + 2(1-r_2)f + r_2} \equiv F(f; r_1, r_2).$$

Skeist correlated monomer conversion and *f* via the following equation [20]:

$$log\frac{[M]}{[M]_0} = \int_{f^0}^{f} \frac{df}{F(f; r_1, r_2) - f}.$$

As the right hand is all about *f*, the integral can be numerically calculated from a given f^0 to f, allowing us to calculate *conv* given *f*. Next, we calculate the triad fraction based on r_1, r_2 and *f*. According to the first order Markov terminal model [21], triad fraction in instantaneously generated (or elongated in living polymerization) polymers, $\{T_{111}, T_{112}, T_{212}, T_{222}, T_{221}, T_{121}\}$, is determined by the following equations:

$$T_{111} = (1 - P_{1\to 2})^2,$$

$$T_{112} = 2P_{1\to2}(1 - P_{1\to2}),$$

$$T_{212} = (P_{1\to2})^2,$$

$$T_{222} = (1 - P_{2\to1})^2,$$

$$T_{221} = 2P_{2\to1}(1 - P_{2\to1}),$$

$$T_{121} = (P_{2\to1})^2,$$

where $P_{1\to2} = \frac{1}{1+r_1 f/(1-f)}$ is the probability of monomer2 adduction to monomer 1 radical, and $P_{2\to1} = \frac{1}{1+r_2(1-f)/f}$ is the probability of monomer 1 adduction to monomer 2 radical. After separately calculating *conv* and $\{T_{111}, T_{112}, T_{212}, T_{222}, T_{221}, T_{121}\}$ based on given (f, r_1, r_2), we plotted $\{T_{111}, T_{112}, T_{212}, T_{222}, T_{221}, T_{121}\}$ as a function of *conv*, as shown in Fig. S9. Note that what we can observe is accumulated triad fraction, not instantaneously generated triad fraction. We thus further integrated the instantaneously generated triad fraction along conversion, so that we can compare the predicted and observed accumulated triad fraction in Fig. 4C. To predict the triad fraction of S/B copolymers synthesized at 70 °C (Fig. 4C), we used the reported value $(r_S, r_B) = (0.70,\ 0.17)$ at 60 °C[22].

Determination of the S/B triad fraction via ^{1}H NMR

As shown in Fig. S10C, -O-CH_2- protons of B-units around 4.1-3.5 ppm are sensitive to the adjacent monomer species, splitting into three major peaks, {*X*, *Y*, *Z*} according to the literature [21]. The triad fraction and peak fraction are correlated by the following equations:

$$\begin{pmatrix} T_{BBB} \\ T_{BBS} \\ T_{SBS} \end{pmatrix} = \begin{pmatrix} 1 & 1-\sigma & (1-\sigma)^2 \\ 0 & \sigma & 2\sigma(1-\sigma) \\ 0 & 0 & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

where (x, y, z) is the peak-area fraction of {X, Y, Z} and $\sigma = 0.85$ is a coisotactic parameter[21]. For determining S-centered triad fraction, ^{13}C NMR is sometimes available. The quaternary carbon of the benzene ring is known for being sequence-sensitive. However, as shown in Fig. S10B, SSS, SSB, BSB peaks significantly overlap with some hardly attributable peaks, hindering quantitative analysis. We thus used only B-centered triad fraction determined from ^{1}H NMR, for verifying our RQMS pentad sequencing (Fig. 4B-C). Also see the following section for converting RQMS pentad-fractions to B-centered triad fraction so that NMR and RQMS are comparable.

Downgrading the S/B pentad-fraction.

As NMR is not sensitive enough to determine S/B pentad distribution, we downgraded the RQMS pentad-sequence results to B-centered triad fractions so that RQMS and NMR results are comparable. We first prepared a matrix $\boldsymbol{T}_{\mathrm{B}} \in {\mathbb{R}_{+}}^{9\times 3}$ connecting nine basis sequence-defined copolymers to B-centered 3 triads abundances, as shown in Table S2. We then obtained pentad sequencing results, $\boldsymbol{C}_t \in {\mathbb{R}_{+}}^{1\times K}$ (here $K = 9$) via procedure described in the section of "*Sequential projection of a target pyrolysis-MS spectrum onto the learned subspaces"*. The B-centered triad fractions can be obtained as $\boldsymbol{C}_t\boldsymbol{T}_{\mathrm{B}}$ after normalization so that $\boldsymbol{C}_t\boldsymbol{T}_{\mathrm{B}}$ satisfies row-stochastic conditions. S-centered triads fraction can be similarly calculated as well with $\boldsymbol{T}_{\mathrm{S}}$ (table S2), which is not used in this study since S-centered triads fraction is not accessible via NMR analysis as mentioned above.

Supplementary Text

Definition of sequencing

In the main text, “sequencing” is defined sequence distribution analysis. This definition could be controversial, since Lutz *el. al.* suggested to distinguish “sequencing” and “sequence analysis” [23]. According to their definition, “sequencing” refers to determining uniformly-defined sequence of a monodispersed polymer ensemble, while “sequence analysis” refers to analyzing sequence tendencies of a polydisperse ensemble. The classification is thus based on the nature of the objective copolymer ensembles. As our proposed method is applicable to either mono/poly-dispersed ensembles, we do not distinguish sequencing and sequence distribution analysis in our manuscript. The sequence-defined copolymers would span the probability-simplex and thus be located at the vertices, while polydisperse copolymers would be located at interior points of the simplex of which coordinates represent the codon composition, as described in the main text. Note that we here assume periodically sequenced copolymers as sequence-defined copolymers. We do not assume arbitrarily yet uniformly sequenced copolymers as an target copolymer ensemble, since such ultimate sequence-controlled copolymerization has not yet been developed except for unpractical single-monomer-addition “oligomerization” strategy, e.g., see [24].

Supplementary Figures

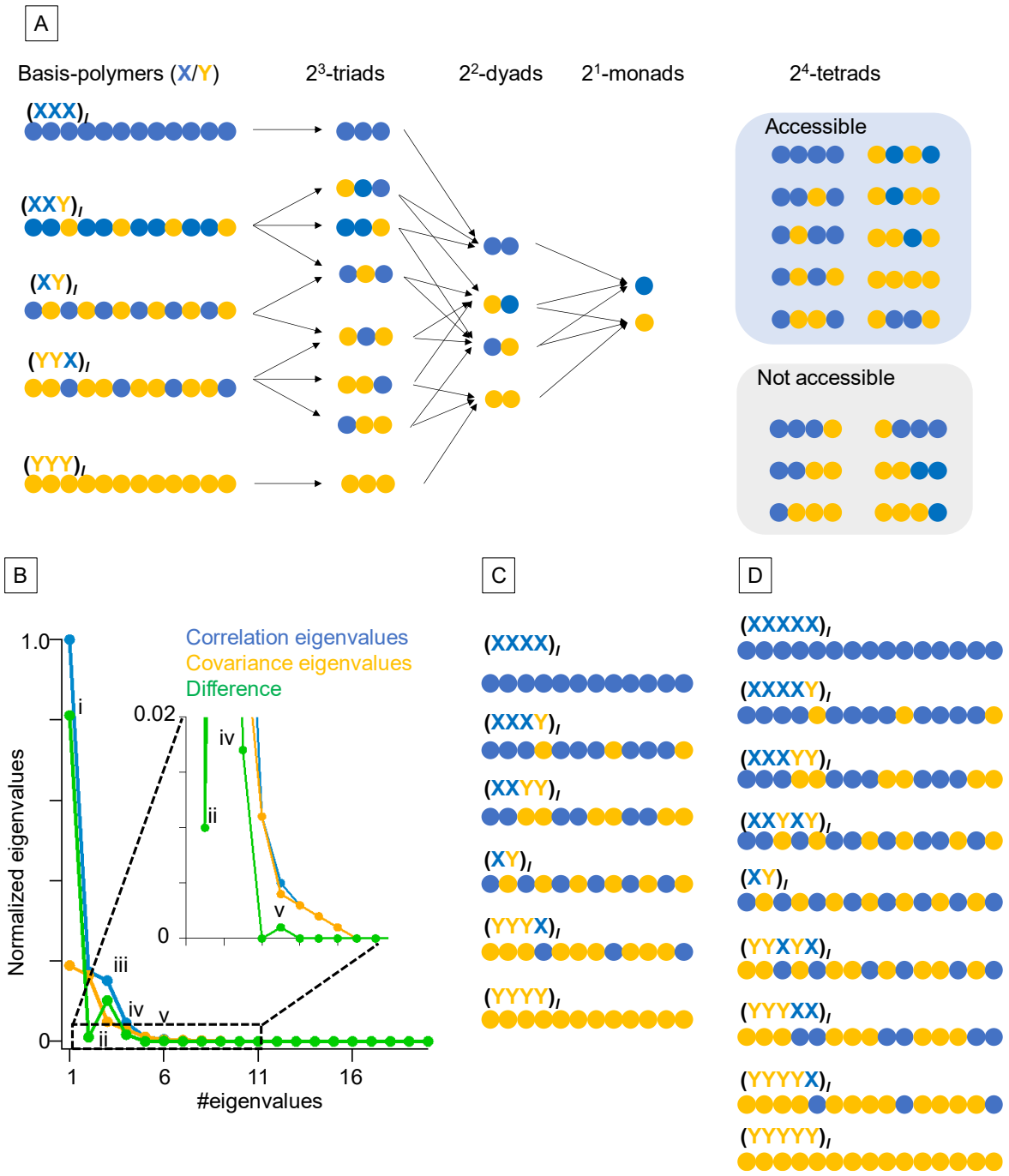

Fig. S1. Assumed basis copolymers for RQMS sequencing. (**A**) Five–basis copolymers for binary triad sequencing. All the 2^3–triads are included in the five basis polymers. Pyrolysis

randomly cleaves the polymer chains generating various length of codons not limited to triad. Nevertheless, as the generation probability of a shorter codon mainly depends on the structure of the one-unit longer codon, the abundances of shorter codons than triad can be written in hierarchical chain structure. This means there are only five-basis patterns of triads-dyads-monads peak series with fixed intensity ratios, whose linear combination can express any triads-dyads-monads distributions generated from arbitrarily sequenced copolymers. To intuitively understand this, it would be helpful to consider the tetrads included in the five–basis copolymers. There are 2^4–tetrads, six of which are not included in the five basis polymers. Therefore, when the basis number *K* is set five, accessible codon length is limited to triads. To express all the tetrad possibilities, we need a different basis set of six copolymers as shown in (C). In this way, we can suppress rapid increase in *K* for longer codon analysis though the possible number of codon sequences exponentially increases. Note that XXY and YXX are different triads which potentially generate different fragments patterns, but can be summarized in a single basis-copolymer $(XXY)_l$, because the abundance ratio of XXY and YXX is identical in any copolymers. Importantly, setting $K = 5$ is not necessarily appropriate for triad sequencing. For example, consider a monomer combination with monomer reactivity ratios of $r_X = 0$ and $r_Y > 0$. In this case, the possible triads are limited to $(XY)_l$, $(YYX)_l$, and $(YYY)_l$, as XX dyad is forbidden. *K* thus should be three, not five. Significantly, our developed sequence necessitates no prior knowledge nor assumption about the chemical structures of the basis copolymers. Our sequencer only necessitates the number of basis copolymers; therefore, statistical determination of the basis number *K* from the spectral dataset could be more effective as shown in (**B**). As well known, *K* can be determined based

on the differences between the eigenvalues of correlation and covariance matrices; the number of the non-zero difference components, marked as i~v, corresponds to K [25]. Here, the dataset for S/B triad sequencing (Fig. 1A) was used as an example, clearly indicating that K should be set as five. (**C**) The six–basis copolymers for tetrad sequencing. (**D**) The nine–basis copolymers for tetrad sequencing.

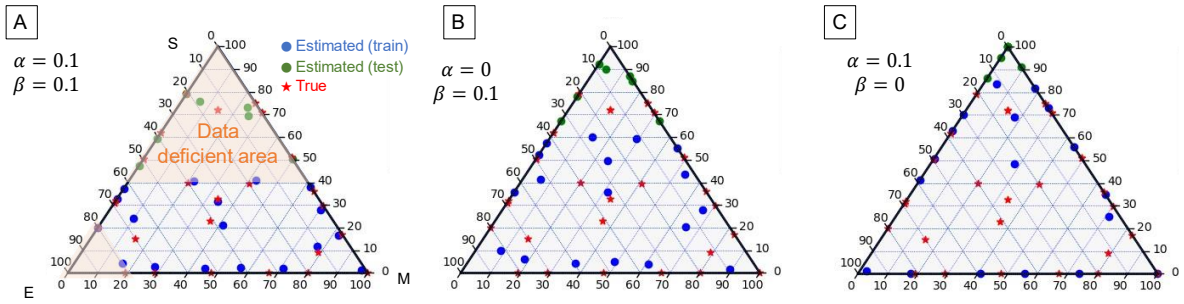

Fig. S2. Benchmark compositional analysis of the M/S/E ternary films based on a biased dataset lacking S-rich samples. The S-rich datapoints beyond 40 wt% S-fraction (green points) were not used for leaning $\boldsymbol{S}$ and $\boldsymbol{B}$, but were sequentially projected onto the subspaces spanned by $\boldsymbol{S}$ and $\boldsymbol{B}$ pre-learned from blue datapoints. The data deficient areas are represented by orange triangles. The weighting parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, for balancing the penalty terms of volume-minimization and orthogonal constraints in the second NMF (see Eq. 11 in "Derivation of the second NMF" and the related sections) are (**A**) $\boldsymbol{\alpha} = \beta = 0.1$, (**B**) $\alpha = 0, \beta = 0.1$ and (**C**) $\alpha = 0.1, \beta = 0$. Only when both constraints are active (A), RQMS algorithm output accurate result.

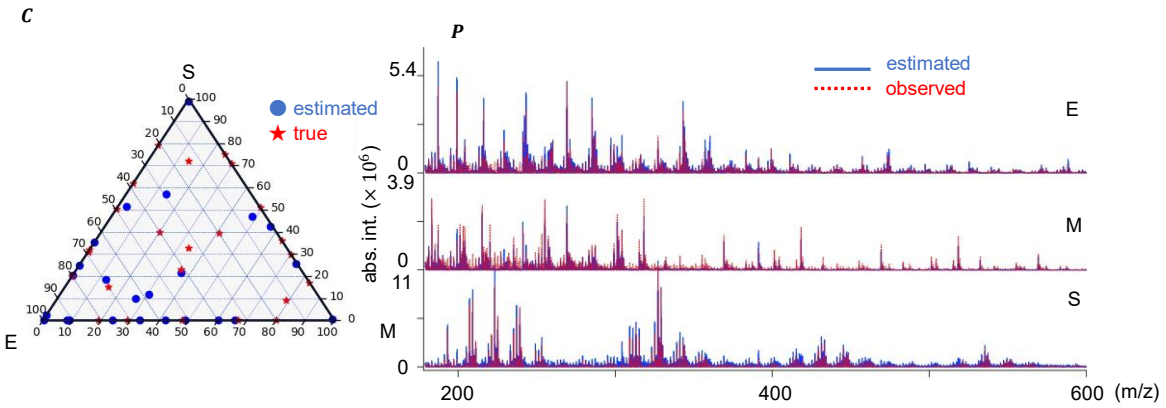

Fig. S3. Benchmark compositional analyses of E/M/S ternary films via single-step NMF, $\boldsymbol{X} \approx \boldsymbol{CP}$, with $K = 3, \alpha = \beta = 0.1$ (see details of these hyper parameters in the derivation of the second NMF section). The estimation errors of both $\boldsymbol{C}$ and $\boldsymbol{P}$ became much greater as compared to Fig. 2A, suggesting the effectiveness of the two-step NMF for RQMS.

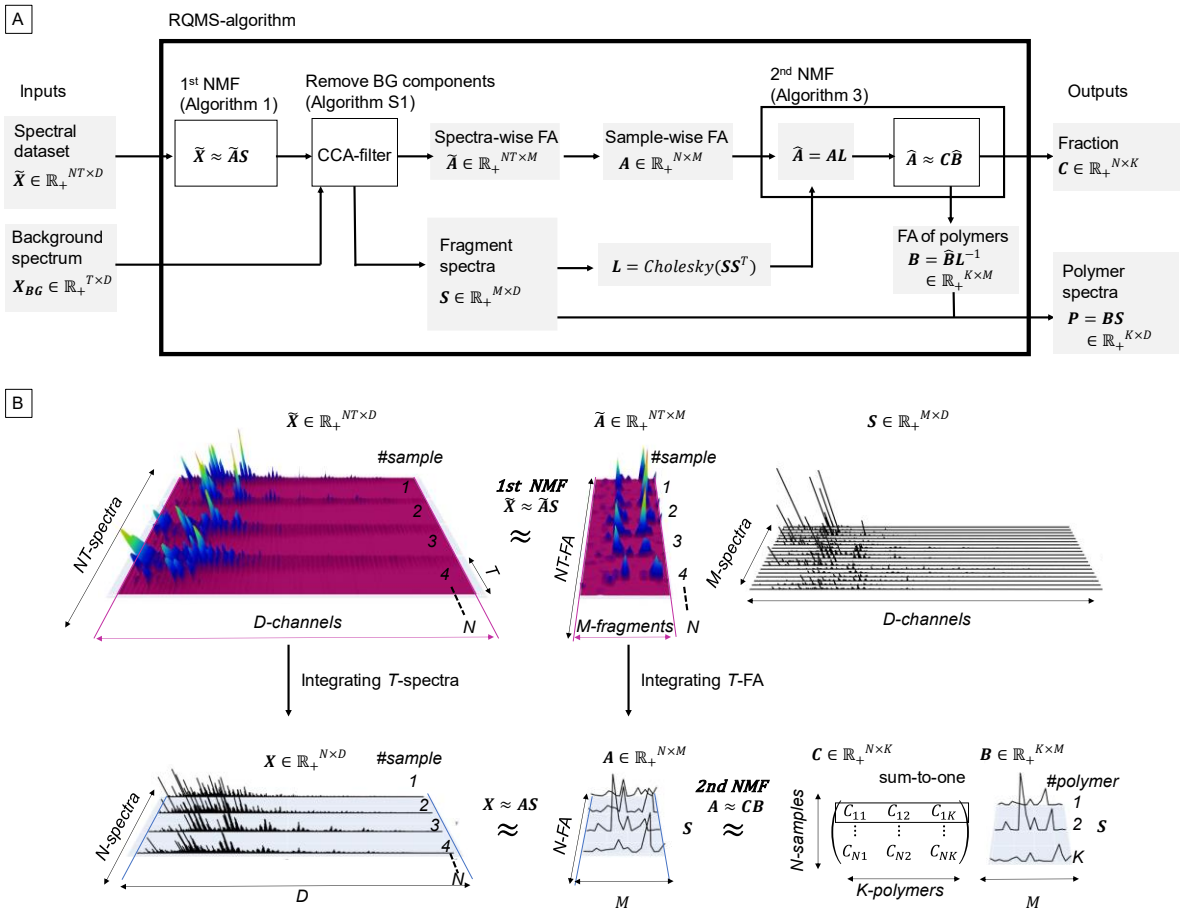

Fig. S4.

RQMS algorithm. (**A**) Flow chart, outputting $\boldsymbol{C}$ and $\boldsymbol{P}$ based on the input $\widetilde{\boldsymbol{X}} \in \mathbb{R}_+^{NT\times D}$. In the main text, for simplicity, we consistently assume 1D-spectral dataset $\boldsymbol{X} \in \mathbb{R}_+^{N\times D}$ as the input. However, a single spectrum of pyrolysis-MS consists of *T*-spectra recorded at different *T*-temperature bands (see "Spectra formatting" section), and thus, the dataset is $\widetilde{\boldsymbol{X}} \in \mathbb{R}_+^{NT\times D}$ rather than $\boldsymbol{X} \in \mathbb{R}_+^{N\times D}$. After the first NMF, spectrum-wise FA $\widetilde{\boldsymbol{A}} \in \mathbb{R}_+^{NT\times M}$ is sample-wisely integrated to sample-wise FA $\boldsymbol{A} \in \mathbb{R}_+^{N\times M}$, as the temperature dependency of FA is not important for compositional analysis. Background/contaminated components are

removed by our developed CCA-filter (see "Derivation of the CCA-filter" section). To conduct the second NMF in Riemann metrics considering the non-orthogonality of $\boldsymbol{S}$ (also see Fig. S6), the factorization is conducted for $\widehat{\boldsymbol{A}} \equiv \boldsymbol{AL} \in \mathbb{R}^{N\times M}$, where $\boldsymbol{L} \in \mathbb{R}^{M\times M}$ is the lower matrix obtained by Cholesky decomposition of $\boldsymbol{G} \equiv \boldsymbol{SS}^T$ (see "Derivation of the second NMF" section). (**B**) The relationship between $\widetilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ are graphically presented.

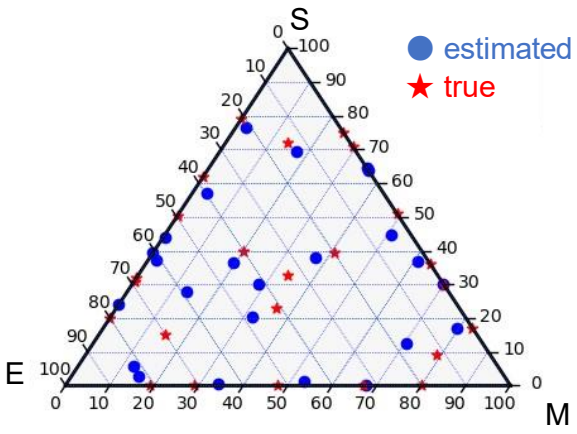

Fig. S5.

Inputting $\boldsymbol{X} \in {\mathbb{R}_+}^{N \times D}$ to RQMS algorithm derived a poor result in benchmark compositional analysis of E/M/S ternary films, suggesting the first NMF should be applied to $\widetilde{\boldsymbol{X}} \in {\mathbb{R}_+}^{NT \times D}$, rather than $\boldsymbol{X} \in {\mathbb{R}_+}^{N \times D}$. All the parameters were set identical to those for Fig. 2A, except for setting temperature-band number $T = 1$ ($T = 20$ for Fig. 2A).

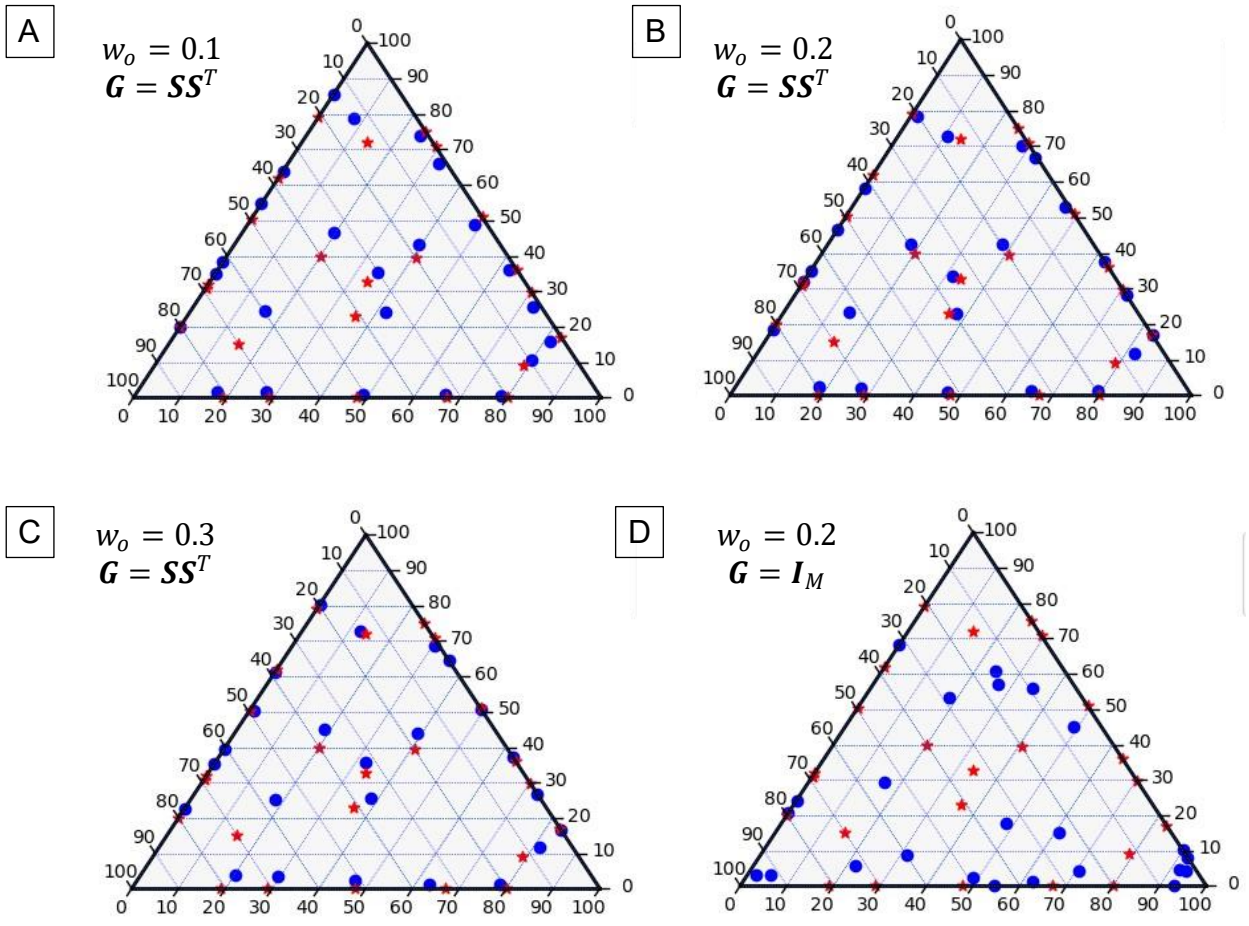

Fig. S6.

Benchmark compositional analyses of E/M/S ternary films. The residuals of the second NMF in RQMS algorithm were evaluated with Riemann metrics, i.e., $D_{\boldsymbol{G}}(\boldsymbol{A}|\boldsymbol{C}\boldsymbol{B}) = Tr[(\boldsymbol{A} - \boldsymbol{C}\boldsymbol{B})\boldsymbol{G}(\boldsymbol{A} - \boldsymbol{C}\boldsymbol{B})^T]$, where (**A**, **B**, **C**) $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{S}^T$ and (**D**) $\boldsymbol{G} = \boldsymbol{I}_M$ corresponding to Euclidean distance. (**A**, **B**, **C**) To show the robustness of the estimated composition $\boldsymbol{C}$, the weight of orthogonal constraints in the first NMF was varied, i.e., $w_o \in \{0.1, 0.2, 0.3\}$. All the outputs were similar and less sensitive to the first NMF model selection. As the best result was obtained when $w_o = 0.2$, we consistently use this condition in this study. (**D**) When $\boldsymbol{G} =$

$\boldsymbol{I}_M$, the estimation was not reliable. Considering non-orthogonality of $\boldsymbol{S}$ by introducing $\boldsymbol{G}$ is thus critically important.

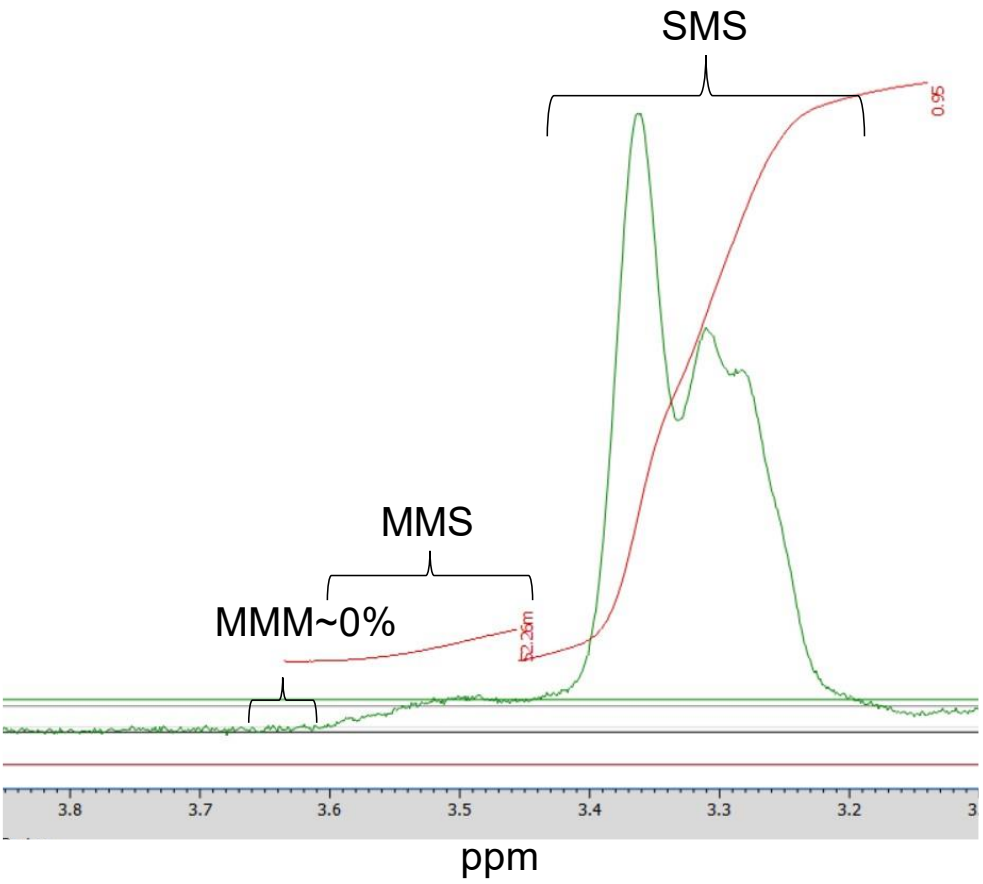

Fig. S7.

^{1}H NMR spectrum of commercially available $(MS)_{l}$ copolymer measured in $CDCl_3$. The copolymer was synthesized in the presence of ethyl aluminum sesquichloride [26]. The methoxy protons sensitive to the sequence are presented. According to the literature[26], MMM, MMS, SMS peaks roughly split, allowing *qualitative* analysis. As here shown, there was no MMM and very little MMS, suggesting highly alternating sequence. We did not conduct further quantitative analysis via NMR by decoupling its tacticity.

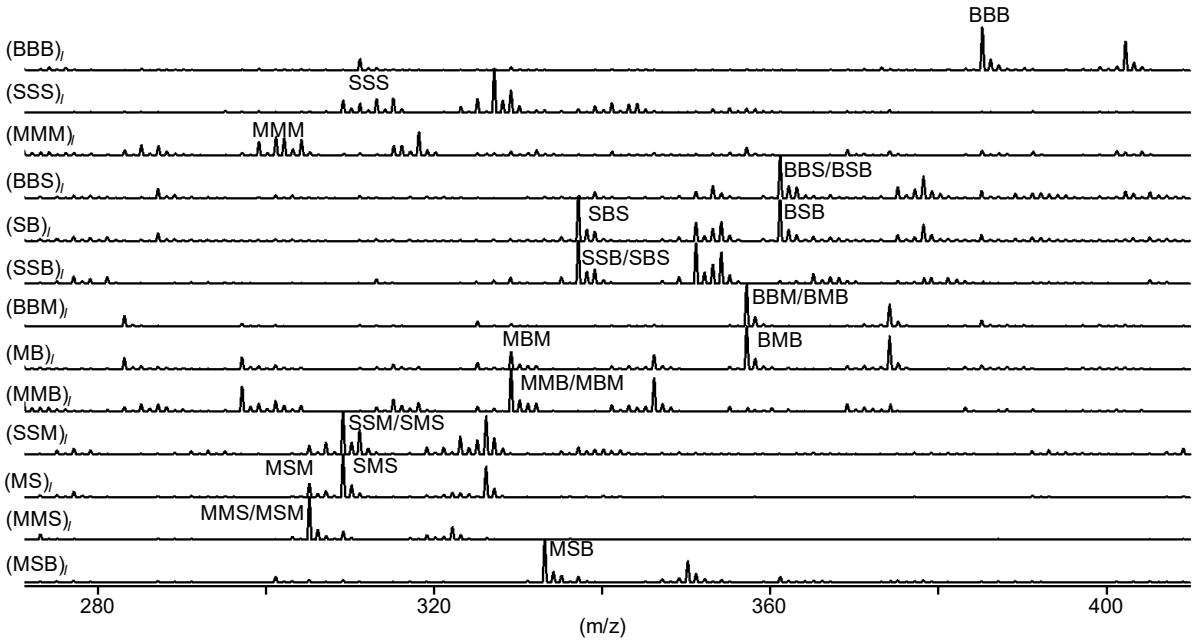

Fig. S8.

The estimated 13 basis spectra of $(BBB)_l$, $(SSS)_l$, $(MMM)_l$, $(BBS)_l$, $(SB)_l$, $(SSB)_l$, $(BBM)_l$, $(MB)_l$, $(MMB)_l$, $(SSM)_l$, $(MS)_l$, $(MMS)_l$, and $(MSB)_l$ for M/S/B ternary triad sequencing. The coupled peaks with 17 m/z intervals are attributable to proton- and ammonium-adducted fragments.

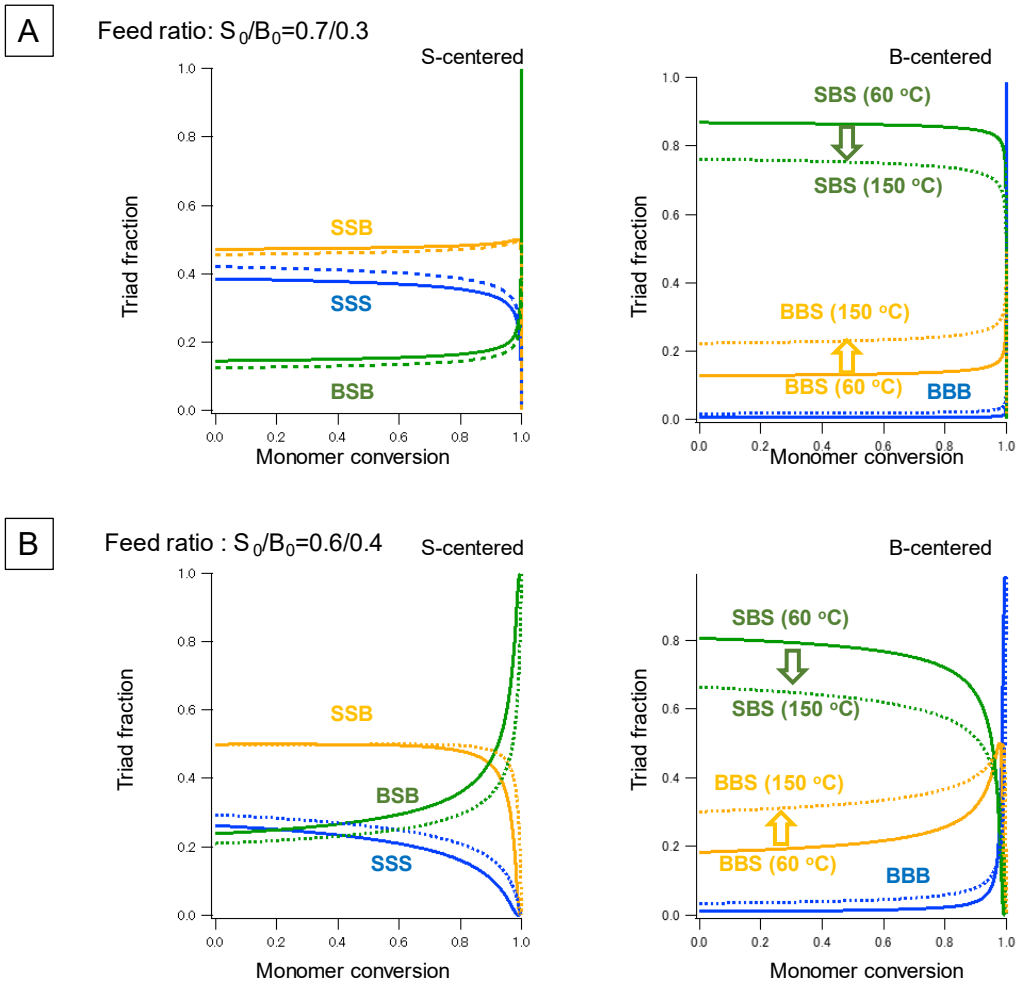

Fig. S9.

S/B instantaneously generated triad-fractions as functions of monomer conversion with the initial monomer molar fraction of (**A**) S_0/B_0 = 0.7/0.3 and (**B**) S_0/B_0 = 0.6/0.4 at polymerization temperature of 60 °C (solid lines) and 150 °C (dotted lines). S-centered and B-centered triad fractions are separately depicted. Monomer reactivity ratios are $(r_S, r_B) = (0.70, 0.17)$ and $(r_S, r_B) = (0.78, 0.34)$ at 60 °C and 150 °C polymerization temperatures[22]. Assume we need a copolymer having high SSSBB pentad fraction. As S/B is

inherently alternating monomer combination ($r_S < 1, r_B < 1$), achieving high SSSBB fraction is challenging. This can be intuitively understood as follows; for yielding high SSS fraction, S_0 should be sufficiently high, however, which would derive low BBS fraction owing to highly selective SBS generation. By increasing temperature from 60 ºC to 150 ºC, this SBS selectivity among B-centered triads is reduced, affording higher BBS generation even under high S_0 conditions as indicated by allows. This strategy was helpful for BBBSS pentad as well. In the dataset for pentad sequencing, we thus included copolymers synthesized at 150 ºC with the monomer feed ratios around S_0/B_0=0.6/0.4 (for SSSBB) and S_0/B_0=0.2/0.8 (for BBBSS). For further detailed mathematical procedures, see "Prediction of sequence distribution from the monomer reactivity ratio" section. What if your monomer combination has no reported values of monomer reactivity ratio? In such case, we still can conduct triad sequencing based on a roughly prepared small dataset. Since triad fraction and monomer reactivity ratio are correlated via Alfrey-Mayo equation, the monomer reactivity ratio can be reversely calculated from the triad fraction[27]. Based on the estimated monomer reactivity ratio, the dataset can be further designed and expanded for more advanced sequencing.

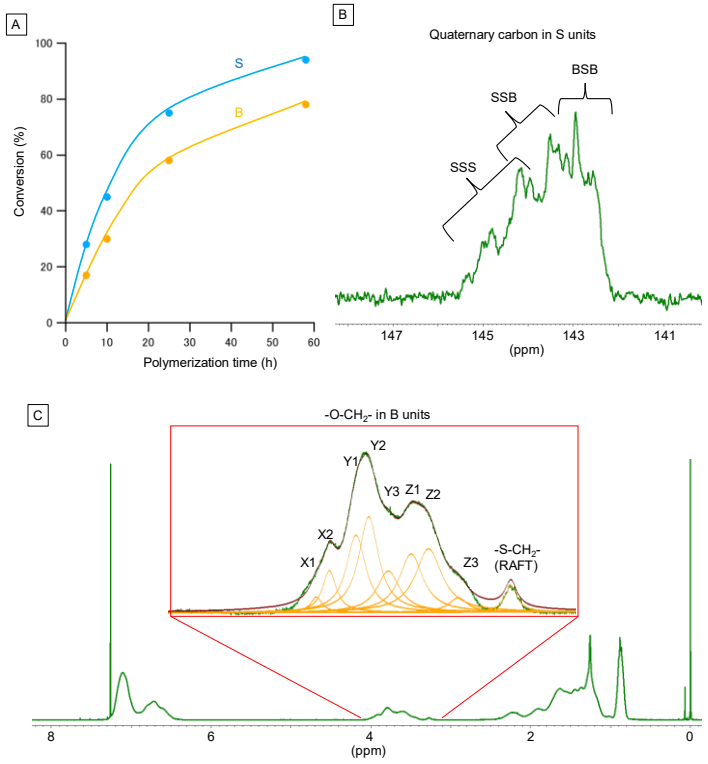

Fig. S10.

S/B random copolymer synthesis via RAFT polymerization. Polymerization conditions: $[S]_0/[B]_0/[DDMAT]_0/[AIBN]_0$=20/20/0.2/0.06 mmol in 1,4-dioxane 2 mL at 70 °C. (**A**) Monomer-polymer conversion curves of S and B. Small portions of the polymerization solution were taken out at 5 h, 10 h, 25 h, 58 h, for investigating sequence modulation along main-chains, as shown in Fig. 4B. (**B**) ^{13}C NMR spectrum of S/B copolymers synthesized via RAFT copolymerization from 1/1 monomer feed ratio. Peaks of the sequence-sensitive quaternary carbon involved in S units are shown. NMR conditions: 10 wt% copolymer solution in $CDCl_3$ at 25 °C; 5 s relaxation time; 9000 scans in 15 h measurement time; referencing $CDCl_3$ central peak at 77.0 ppm. The rough peak attribution of SSS, SSB, and BSB followed the reported paper[21]. (**C**) ^{1}H NMR spectrum of S/B copolymers synthesized

via RAFT copolymerization. NMR conditions: 30 s of relaxation delays in $CDCl_3$ at 25 ^oC. The inset shows magnified peaks of sequence-sensitive -O-CH_2- protons involved in B monomers. The small peak on the right is attributable to -S-CH_2- protons involved in DDMAT, which is not related to sequence analysis. The overlapped peaks X, Y, Z were deconvoluted via Lorentzian peak fitting, yielding peak-area fraction of (*x*, *y*, *z*), which was further converted to B-centered triad fractions as described in "Determination of the S/B triad fraction via ^1H NMR" section. As simple 3-peaks fitting gave huge fitting errors, we here used 8 peaks, {X1, X2, Y1, Y2, Y3, Z1, Z2, Z3}, without a good scientific excuse. The green and brown are observed and fitted curves, respectively.

Table S1.

Hyperparameters used in this study.

Dataset	Sample number *N*	Temp. range (°C) Mass range(m/z)	First NMF				Second NMF		
			w_o	Merging threshold	Initial *M*	iteration	K	$\alpha = \beta$	p
Data S1 (Fig. 1A, S/B triad)	21	[200, 450] [50, 500]	0.2	0.99	30	5000	5	0.05	1
Data S2 (Fig. 2A, benchmark)	24	[50, 450] [50, 1000]	0.1/0.2/0.3	0.99	30	3000	3	0.1	1/1.5
Data S3 (Fig. 3, M/S triad)	30	[200, 450] [50, 500]	0.2	0.99	30	3000	5	0.05	1
Data S4 (Fig. S8, M/S/E triad)	84	[200, 450] [50, 500]	0.2	0.99	60	10000	13	0.05	1
Data S5 (Fig. 4, S/B pentad)	81	[200, 450] [100, 670]	0.2	0.99	60	10000	9	0.01	1

Table S2.

Pentad-to-triad transforming matrices, $\boldsymbol{T}_{\mathrm{B}}$ and $\boldsymbol{T}_{\mathrm{S}}$. We used only $\boldsymbol{T}_{\mathrm{B}}$ for downgrading RQMS pentad sequencing results to B-centered triad fraction to be comparable to NMR sequencing.

Sequence-defined copolymers	B-centered triad matrix, $\boldsymbol{T}_{\mathrm{B}}$			S-centered triad matrix, $\boldsymbol{T}_{\mathrm{S}}$		
	BBB	BBS	SBS	SSS	SSB	BSB
$(BBBBB)_I$	5	0	0	0	0	0
$(BBBBS)_I$	2	2	0	0	0	1
$(BBBSS)_I$	1	2	0	0	2	0
$(BBSBS)_I$	0	2	1	0	0	2
$(BS)_I$	0	0	2.5	0	0	2.5
$(SSBSB)_I$	0	0	2	0	2	1
$(SSSBB)_I$	0	2	0	1	2	0
$(SSSSB)_I$	0	0	1	2	2	0
$(SSSSS)_I$	0	0	0	5	0	0

Data S1-S5.

The full spectral datasets and sample information for Fig. 1A, Fig. 2A, Fig. 3, Fig. S8, and Fig. 4C. All the spectra recorded as CDF files have been converted into CSV files, and formatted as described in "Spectra Formatting" section, to be ready-to-use. All the spectral datasets are available in DOI: 10.26434/chemrxiv-2022-mw76d-v2.

References

1 G. Moad, E. Rizzardo and S. H. Thang, *Aust. J. Chem.* 2012, **65**, 985–1076.
2 X. Zhuang, Z. Yang and D. Cordes, *Hum. Brain. Mapp*. 2020, **41**, 3807–3833.
3 M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori and T. Tanji, *Ultramicroscopy*, 2016, **170**, 43–59.
4 V. Y. F. Tan and C. Févotte, *IEEE Trans. Pattern. Anal. Mach. Intell.* 2013, **35**, 1592–1605.
5 M. Anderle, S. Roy, H. Lin, C. Becker and K. Joho, *Bioinformatics*, 2004, **20**, 3575–3582.
6 A. Cichocki and A.-H. Phan, *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* 2009, **E92-A**, 708–721.
7 K. Kimura, Y. Tanaka, M. Kudo, D. Phung and H. Li, *Mach. Learn. Res.* 2014, **39**, 129-141.
8 S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Found. Trends Mach. Learn*, 2010, 3, 1–122.
9 X. Fu, W. K. Ma, K. Huang and N. D. Sidiropoulos, in *ICASSP,* 2016, 2534–2538.
10 D. C. Heinz and C.-I. Chang, *IEEE Trans. Geosci. Remote. Sens*. 2001, **39**, 529–545.
11 X. Fu, K. Huang, N. D. Sidiropoulos and W. K. Ma, *IEEE Signal Process Mag*, 2019, **36**, 59–80.
12 W. Liu and N. Zheng, *Pattern Recognit. Lett*, 2004, **25**, 893–897.
13 T. Yoshida, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2011, **6804 LNAI**, 214–219.
14 X. Fu, K. Huang, B. Yang, W. K. Ma and N. D. Sidiropoulos, *IEEE Trans. Signal Process.* 2016, **64**, 6254–6268.
15 C. Ding, T. Li, W. Peng and H. Park, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2006, 126–135.
16 J. Kim, Y. He and H. Park, *J. Glob. Optim.* 2014, **58**, 285–319.
17 J. M. P. Nascimento and J. M. B. Dias, *IEEE Trans. Geosci. Remote Sens.* 2005, **43**, 898–910.

18 L. Miao and H. Qi, *IEEE Trans. Geosci. Remote Sens.* 2007, **45**, 765–777.
19 F. R. Mayo and F. M. Lewis, *J. Am. Chem. Soc.* 1944, **66**, 1594–1601.
20 I. Skeist, *J. Am. Chem. Soc.* 1946, **68**, 1781–1784.
21 A. S. Brar and C. V. V. Satyanarayana, *Polym. J.* **24**, 879-887 (1992).
22 L. K. Kostanski and A. E. Hamielec, *Polymer (Guildf).* 1992, **33**, 3706–3710.
23 H. Mutlu and J. F. Lutz, *Angew. Chem. Int. Ed.* 2014, **53**, 13010–13019.
24 Y. Hibi, M. Ouchi and M. Sawamoto, *Nat Commun*, 2016, **7**, 1–9.
25 C. I. Chang and Q. Du, *IEEE Trans. Geosci. Remote Sens.* 2004, **42**, 608–619.
26 H. Hirai, T. Tanabe and H. Koinuma, *J. Polym. Sci., Polym. Chem. Ed*. 1979, **17**, 843–857.
27 D. J. T. Hill, A. P. Lang, J. H. O'Donnell and P. W. O'Sullivan, *Eur. Polym. J.* 1989, **25**, 911–915.