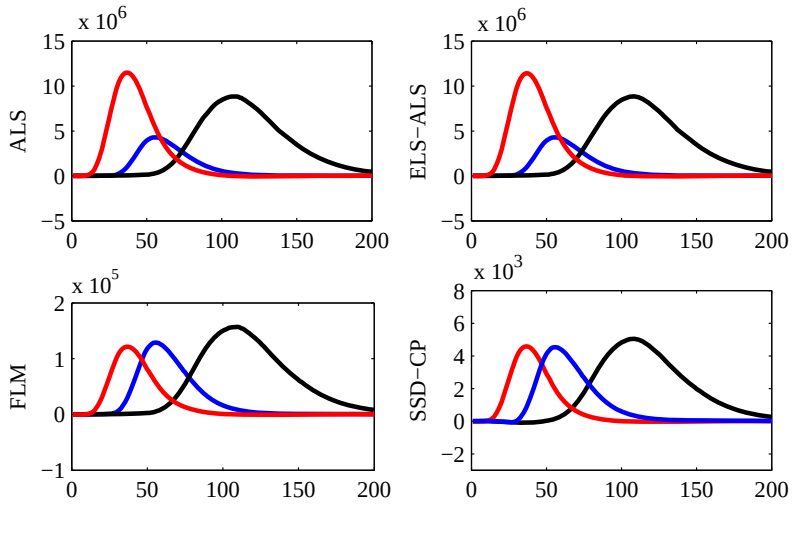
4.1.4 Multiway analysis of amino acids fluorescence data

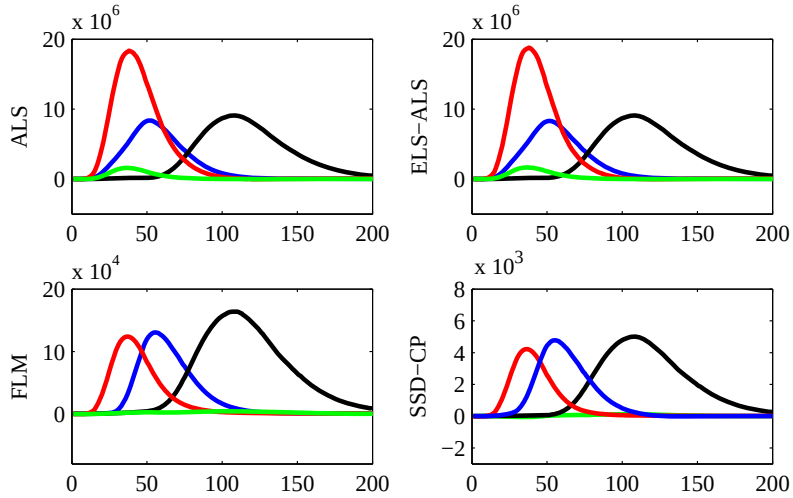
In this section, we examine the effect of overfactoring, i.e., overestimating the array rank, on the performance of different CP methods. To this end, we use the Amino acids fluorescence data downloaded from [Ami, 1998] and described in [Ami, 1998], [Bro, 1998], [Kiers, 1998]. This data set comprises five simple laboratory-made samples. Each sample contains different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence (excitation $250-300$ nm, emission $250-450$ nm, 1 nm intervals) on a PE LS50B spectrofluorometer with excitation slit-width of 2.5 nm, an emission slit-width of 10 nm and a scan-speed of 1500 nm/s [Ami, 1998]. Therefore, the array to be decomposed is of size $(5 \times 61 \times 201)$. Since each individual amino acid gives a rank-1 contribution to the data, we can ideally describe these data with three canonical factors [Ami, 1998]. We applied three classical CP methods, namely ALS [Harshman and Lundy, 1994], ELS-ALS [Comon et al., 2009], FLM [Phan et al., 2013c] and the SSD-CP method proposed in this paper, to canonically decompose the amino acids fluorescence data. Then we compared the emission-mode factors estimated by the four methods by varying the rank P of the CP decomposition from 3 to 5 as shown in figure 4.4.

For each algorithm, we scaled the p-th column of the emission-mode loading matrix by the norm of the rank-1 three-way array produced by the outer product of p-th columns of each of the three loading matrices. Note that, regarding the three non-direct methods, they are run several times with different random initial guess in order to use the best initialization.

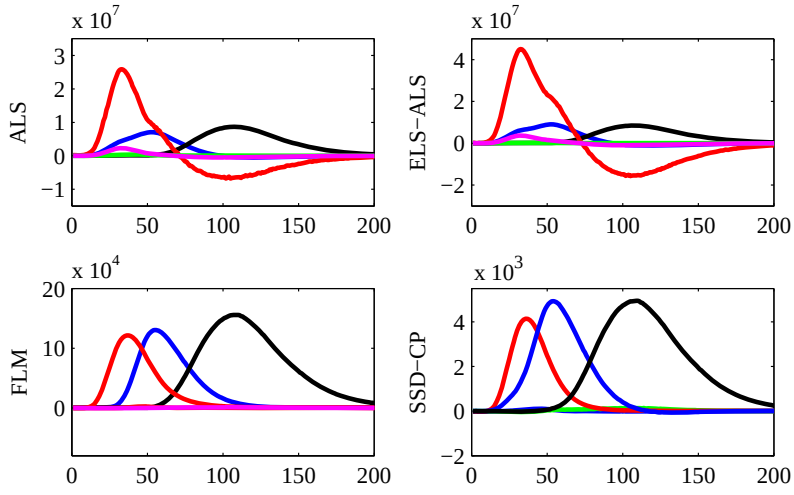
As displayed in figure 4.4(a), for $P = 3$, all CP algorithms generate almost the same emission factors corresponding to three amino acids. As shown in figures 4.4(b-c), in the case of $P = 4$ and $P = 5$, the factors obtained from the ALS and ELS-ALS change, however the proposed SSD-CP technique and the FLM algorithm generate three factors that match with the true factors while the forth and the fifth estimated factors are very small. These results show that SSD-CP and FLM are robust with respect to overfactoring contrary to the other methods.



(a)



(b)



(c)

Figure 4.4: The emission mode factors with the estimated number of components equal to (a) $P = 3$, (b) $P = 4$ and (c) $P = 5$.