In the Nonnegative Matrix Factorization (NMF) problem, we are given an $m$ by $n$ nonnegative matrix $M$ and our goal is to express this matrix as the sum of at most $r$ rank one, nonnegative matrices. The smallest $r$ for which this is possible is the nonnegative rank of $M$. This problem has a rich history spanning quantum mechanics, probability theory, data analysis, polyhedral combinatorics, communication complexity, demography, chemometrics, etc. In the past decade NMF has become enormously popular in machine learning, where a factorization is used to discover latent relationships in the data and is found using a variety of local search heuristics. Vavasis proved that this problem is NP-complete.

We initiate a study of when this problem is solvable in polynomial time, and find that for most interesting applications the problem is in fact efficiently solvable. We prove that when $r$ is constant, the problem can be solved exactly in polynomial time (and we give approximation algorithms for the approximate version too). We complement this with a hardness result that rules out substantial improvement: If this problem can be solved in time $(nm)^{o(r)}$, then 3-SAT has sub-exponential time algorithms.

Lastly, we give an algorithm that runs in time polynomial in $n, m$ and $r$ under the separability condition identified by Donoho and Stodden. Separability is believed to hold in many practical settings. To the best of our knowledge, this last result is the first example of a polynomial-time algorithm that provably works under a non-trivial condition on the input and we believe that this will be an interesting and important direction for future work.

This is joint work with Sanjeev Arora, Rong Ge and Ravi Kannan.