

# Convex Biclustering

Eric C. Chi<sup>1,3,\*</sup>, Genevera I. Allen<sup>2</sup>, Richard G. Baraniuk<sup>1</sup>

1. Department of Electrical Engineering, Rice University

2. Department of Statistics, Rice University

3. Department of Statistics, North Carolina State University

\*Contact author: [echi@rice.edu](mailto:echi@rice.edu)

**Keywords:** Clustering, Convex Optimization, Fused Lasso, Reproducible Research, Structured Sparsity

In the biclustering problem, we seek to simultaneously group observations and features. While biclustering has applications in a wide array of domains, ranging from text mining to collaborative filtering, the problem of identifying structure in high dimensional genomic data motivates this work. In this context, biclustering enables us to identify subsets of genes that are co-expressed only within a subset of experimental conditions. We present a convex formulation of the biclustering problem that possesses a unique global minimizer and an iterative algorithm, COBRA, that is guaranteed to identify it. Our approach generates an entire solution path of possible biclusters as a single tuning parameter is varied. We also show how to reduce the problem of selecting this tuning parameter to solving a trivial modification of the convex biclustering problem. The key contributions of our work are its simplicity, interpretability, and algorithmic guarantees. We demonstrate the advantages of our approach, which includes stably and reproducibly identifying biclusterings, on simulated and real microarray data.