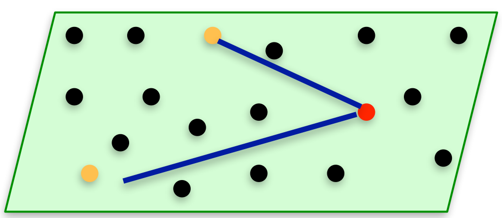
JHU vision lab

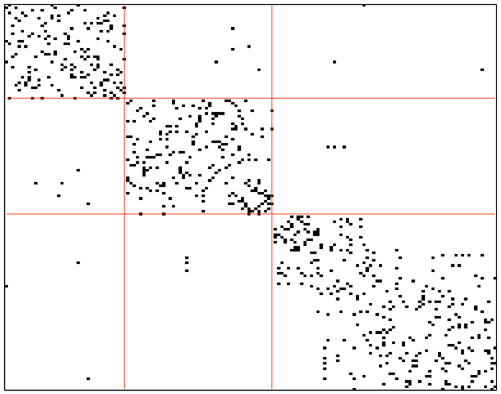
Sparse Subspace Clustering (SSC)

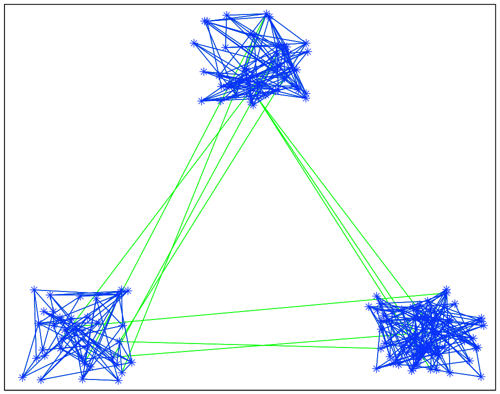
Ehsan Elhamifar and René Vidal

Sparse Subspace Clustering: Spectral Clustering

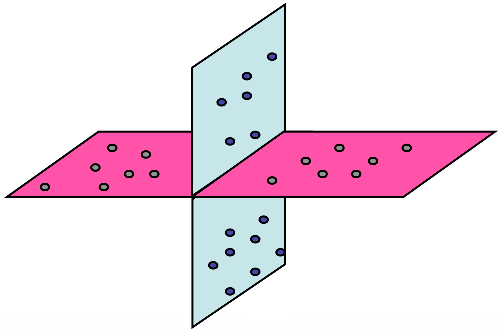
- Spectral clustering
  - Represent data points as nodes in graph G
  - Connect nodes i and j with weight c_{ij}
  - Infer clusters from Laplacian of G





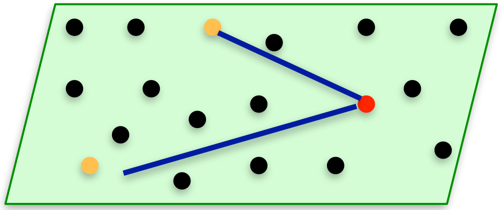


- How to define a good affinity matrix C for subspaces?
  - points in the same subspace: $c_{ij} \neq 0$
  - points in different subspaces: $c_{ij} = 0$

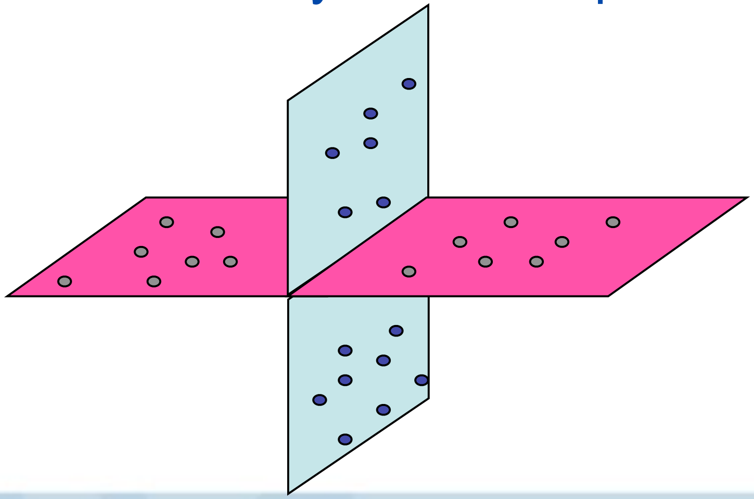


Sparse Subspace Clustering: Spectral Clustering

- Spectral curvature clustering (SCC) (Chen-Lerman '08)
  - Define multiway similarity as normalized volume of *d+1* points



- Local subspace affinity (LSA) (Yan-Pollefeys '06)
  - Use the angles between locally fitted subspaces as similarity



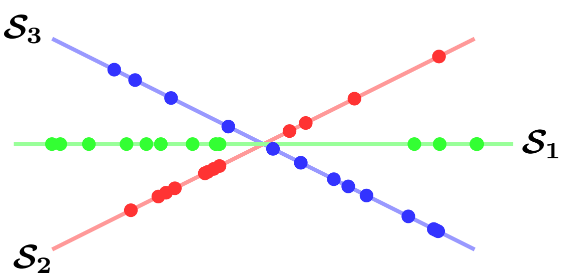
Sparse Subspace Clustering: Intuition

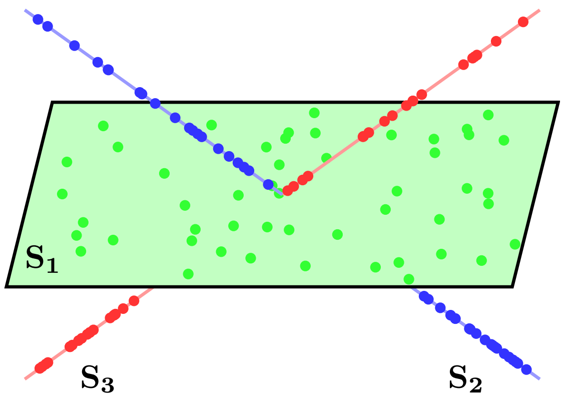
- Data in a union of subspaces are self-expressive

$$\boldsymbol{x}_j = \sum_{i=1}^{N} c_{ij} \boldsymbol{x}_i \implies \boldsymbol{x}_j = X \boldsymbol{c}_j \implies X = XC$$

- Union of subspaces admits subspace-sparse representation

$c_{ij} \neq 0 \implies \boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belong to the same subspace





- Under what conditions is the solution to P_0 subspace-sparse?

$$P_0: \quad \min_{\boldsymbol{c}_j} \|\boldsymbol{c}_j\|_0 \quad \text{s.t.} \quad \boldsymbol{x}_j = X\boldsymbol{c}_j, \quad c_{jj} = 0$$

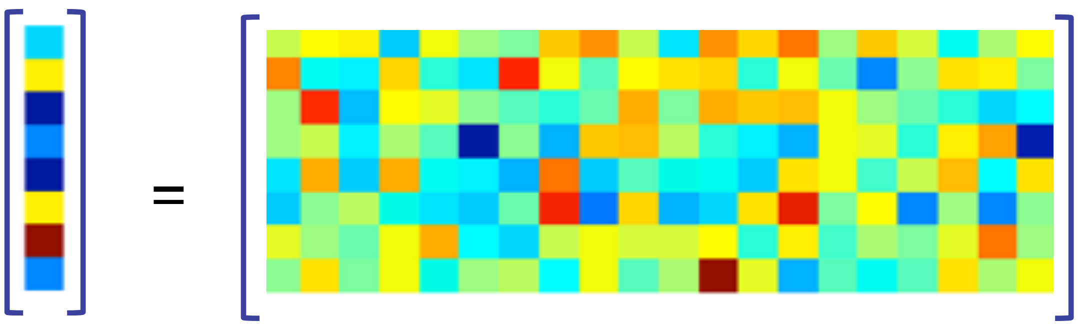
E. Elhamifar and R. Vidal. Sparse Subspace Clustering. CVPR 2009.
E. Elhamifar and R. Vidal. Clustering Disjoint Subspaces via Sparse Representation. ICASSP 2010.
E. Elhamifar and R. Vidal. Sparse Subspace Clustering: Algorithm, Theory and Applications. TPAMI 2013.

Center for IMAGING SCIENCE

Sparse Subspace Clustering: Basics of Sparsity

- Underdetermined linear system:

$$\boldsymbol{x} = X\boldsymbol{c}$$



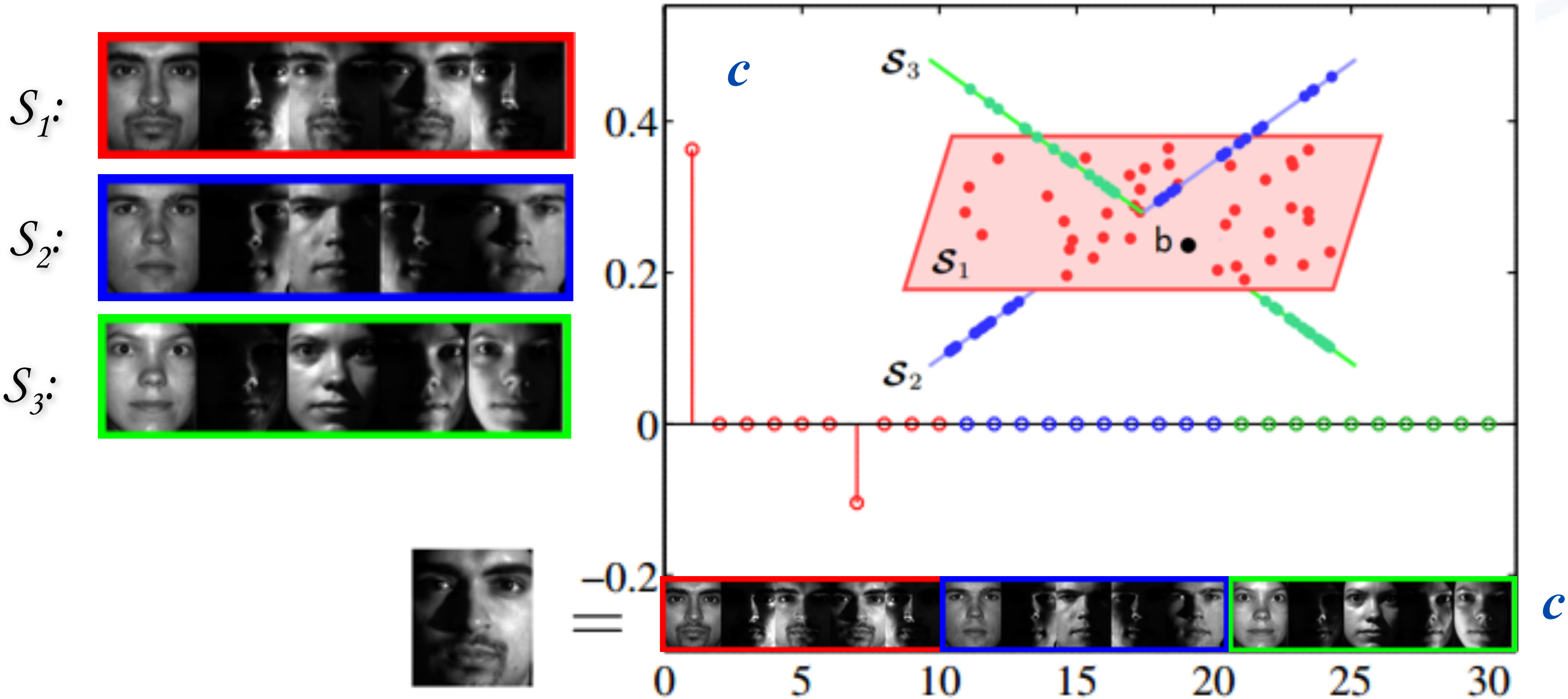
$$\boldsymbol{x} \in \mathbb{R}^D \qquad X \in \mathbb{R}^{D \times N}; \quad D \ll N \qquad \begin{bmatrix} ? \\ ? \\ \vdots \\ ? \end{bmatrix} \quad \boldsymbol{c} \in \mathbb{R}^N$$

- Solution is not unique:
  - Sparsest solution $P_0 : \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_0 \quad \text{s.t.} \quad \boldsymbol{x} = X\boldsymbol{c}$ Intractable!
  - Convex relaxation $P_1 : \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \quad \text{s.t.} \quad \boldsymbol{x} = X\boldsymbol{c}$ Efficient!

- **Theorem**: if *X* is incoherent, solutions *c** are the same

Tropp,"Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Information Theory, 2004.
Candes and Tao, Decoding by linear programming. IEEE Trans. on Information Theory, 2005.
Donoho, For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," Communications on Pure and Applied Mathematics, 2006.

Sparse Subspace Clustering: Beyond Sparsity



$$P_0: \quad \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_0 \quad \text{s.t.} \quad \boldsymbol{x} = X\boldsymbol{c}$$

Is this a sparse recovery problem?

Center for IMAGING SCIENCE

Sparsity versus Subspace Sparsity

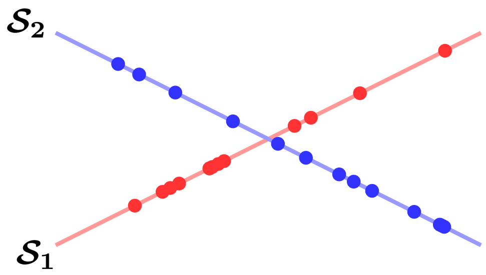
- Classical sparse recovery results do not apply to data in a union of subspaces

Sparse Recovery	**Subspace Sparse Recovery**
Incoherent dictionary	Data in subspaces is not coherent
Sparsest solutions are unique	Sparsest solutions are not unique

Sparse Subspace Clustering: Noiseless Data

- Under what conditions on the subspaces and the data is the solution to P_1 subspace-sparse?
  - Point by point: $$\min_{\boldsymbol{c}_j} \|\boldsymbol{c}_j\|_1 \quad \text{s.t.} \quad \boldsymbol{x}_j = X\boldsymbol{c}_j, \quad c_{jj} = 0$$
  - All points: $$\min_{C} \|C\|_1 \quad \text{s.t.} \quad X = XC, \quad \operatorname{diag}(C) = 0$$
- **Theorem 1**: If the subspaces are independent, then C is subspace-sparse

$$\dim\Big(\bigoplus_{i=1}^{n} S_i\Big) = \sum_{i=1}^{n} \dim(S_i)$$



E. Elhamifar and R. Vidal. Sparse Subspace Clustering. CVPR 2009.
E. Elhamifar and R. Vidal. Sparse Subspace Clustering: Algorithm, Theory and Applications. TPAMI 2013.

Sparse Subspace Clustering: Noiseless Data

- Independence may be too restrictive: e.g., articulated motions

- **Theorem 2**: If the subspaces are sufficiently separated and the data are well distributed inside the subspaces, then *C* is subspace-sparse if for all *i=1,…, n,*

$$\text{(incoherence)} \quad \mu_i = \max_{j \neq i} \cos(\theta_{ij}) < r_i \quad \text{(inradius)}$$

- **Theorem 3**:
  - *n d*-dimensional subspaces drawn independently, uniformly at random
  - $\rho d + 1$ points per subspace drawn independently, uniformly at random
  - P_1 recovers a subspace-sparse representation with high probability if

$$d < \frac{c^2(\rho)\log(\rho)}{12\log(N)} D$$

E. Elhamifar and R. Vidal. Clustering Disjoint Subspaces via Sparse Representation. ICASSP 2010.
E. Elhamifar and R. Vidal. Sparse Subspace Clustering: Algorithm, Theory and Applications. TPAMI 2013.
M. Soltanolkotabi, E. Candes. A geometric analysis of subspace clustering with outliers. Annals of Statistics, 40(4):2195– 2238, 2013.

Sparse Subspace Clustering: Noisy Data

- Under what conditions on the subspaces and the data is the solution to *LASSO* subspace-sparse?
  - Noiseless (*P₁*): $\min_C \|C\|_1 \quad \text{s.t.} \quad X = XC, \quad \operatorname{diag}(C) = 0$
  - Noise (LASSO): $\min_C \|C\|_1 + \frac{\lambda}{2}\|X - XC\|_F^2 \ \text{s.t.} \ \operatorname{diag}(C) = 0$
- **Theorem 4**: If the subspaces are sufficiently separated, the data are well distributed, the noise is small enough, and the LASSO parameter is well chosen, then *C* is subspace-sparse

$$\mu_i < r_i, \quad \delta < \frac{r_i(r_i - \mu_i)}{3r_i^2 + 8r_i + 2}, \quad \underline{\lambda} < \lambda < \bar{\lambda}$$

Wang, Y.-X. and Xu, H. Noisy sparse subspace clustering. ICML 2013.

Sparse Subspace Clustering: Data with Outliers

- **Assumptions**
  - n d-dimensional subspaces chosen independently, uniformly at random
  - $\rho d + 1$ inliers per subspace chosen independently, uniformly at random
  - $N_{outliers}$ outliers chosen independently and uniformly at random
  - Declare point i as an outlier if the solution to P_1 satisfies

$$\|\boldsymbol{c}_i\|_1 > \lambda(\gamma)\sqrt{D}$$

- **Theorem 5**:
  - P_1 correctly detects all outliers with high probability if

$$N_{outliers} < \frac{1}{D} e^{c\sqrt{D}} - N_{inliers}$$

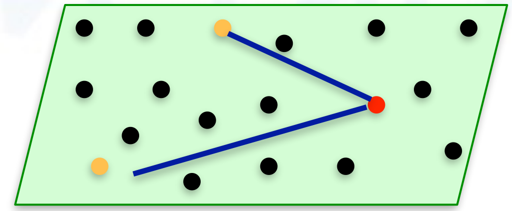
  - P_1 does not detect any inlier as an outlier with high probability if

$$N_{outliers} < D\rho^{c_2 D/d} - N_{inliers}$$

M. Soltanolkotabi, E. Candes. A geometric analysis of subspace clustering with outliers. Annals of Statistics, 40(4):2195– 2238, 2013.

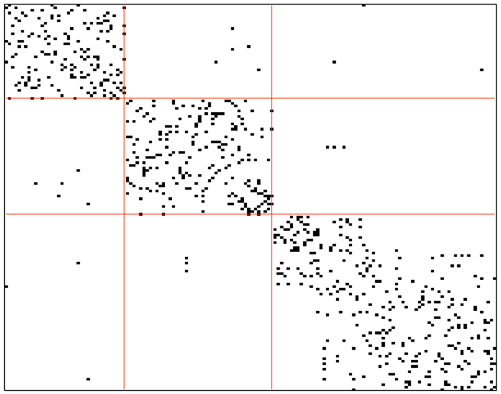
Sparse Subspace Clustering: Algorithm

- Represent data points as nodes in graph G



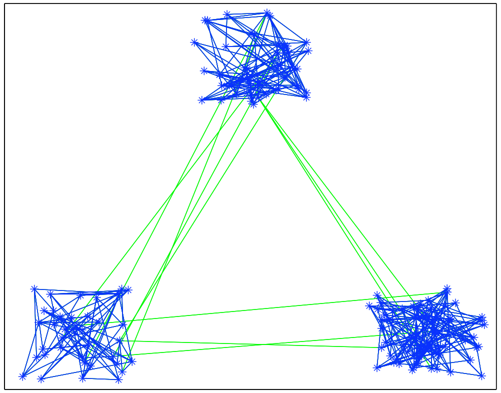
- Find the matrix of sparse coefficients C

$$\min_C \|C\|_1 + \frac{\lambda}{2}\|X - XC\|_F^2 \ \text{ s.t. } \ \operatorname{diag}(C) = 0$$



- Connect nodes i and j by an edge with weight

$$|c_{ij}| + |c_{ji}|$$



- Spectral clustering: apply K-means to the smallest eigenvectors of the Laplacian of G

E. Elhamifar and R. Vidal. Sparse Subspace Clustering. CVPR 2009.
E. Elhamifar and R. Vidal. Clustering Disjoint Subspaces via Sparse Representation. ICASSP 2010.
E. Elhamifar and R. Vidal. Sparse Subspace Clustering: Algorithm, Theory and Applications. TPAMI 2013.

Center for IMAGING SCIENCE