Single Cluster Graph Partitioning

For Robotics Applications

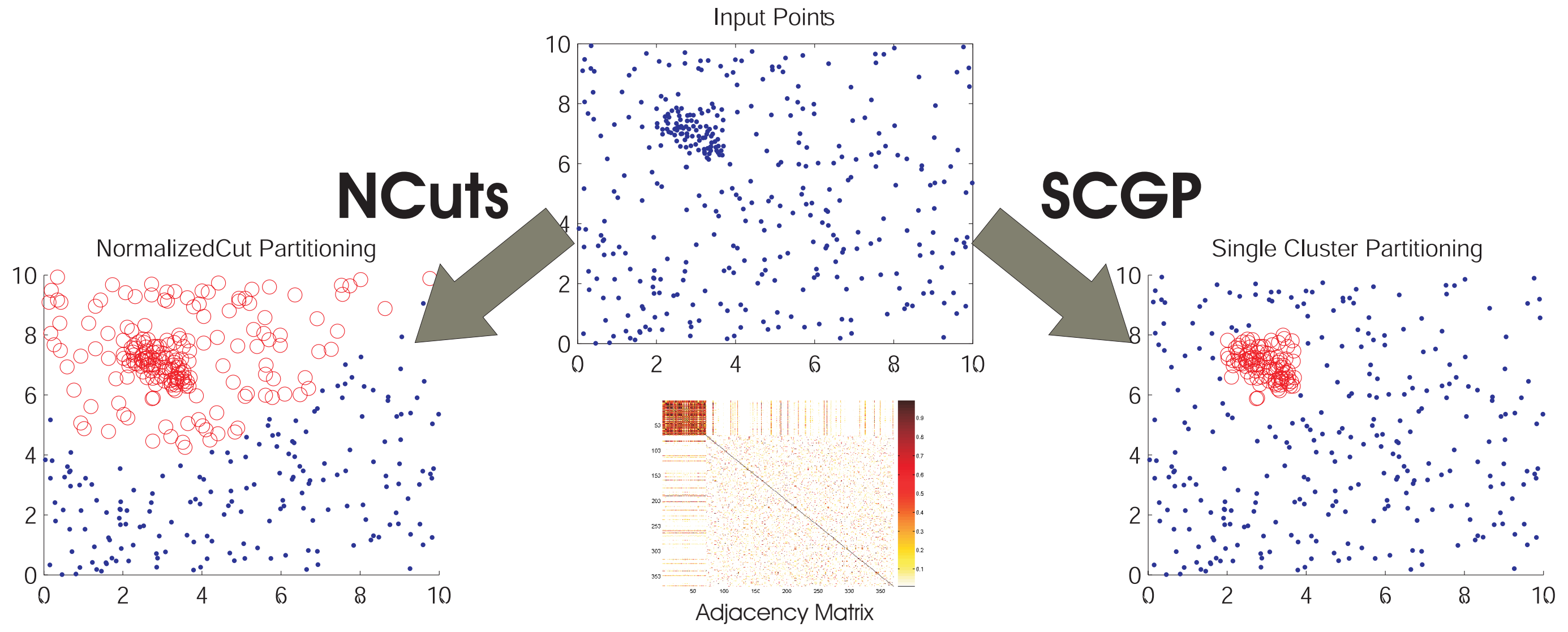
CSAIL

Edwin Olson
Matthew Walter
John Leonard
Seth Teller

Introduction

We have developed a spectral clustering algorithm that identifies a *single* cluster, rejecting points which are not part of a coherent cluster.

This is different from existing clustering algorithms, such as NCuts and MinMaxCuts which assume that all points belong to a logical cluster. When all of the input points *do* belong to clusters (such as a mixture of Gaussians), these algorithms perform very well. However, they produce poor results when there is only a single cluster amidst noise.



Formulation (symmetric case)

Given N points, $p_{1...N}$, define a pairwise consistency function, $f(p_i,p_j)$ that is large for consistent points and small otherwise. Form the NxN adjacency matrix A such that $A_{ij} = f(p_i,p_j)$.

Consider the toy problem in Figure 1. It has eight points; pairwise consistency is boolean-valued, as indicated by edges in the graph. The adjacency matrix is shown in Figure 2. Our question is: "what set of points is maximally self-consistent?"

Our goal is to find a binary-valued, Nx1 indicator vector v such that we accept p_i if u_i=1. The indicator vector is how we represent a set of points.

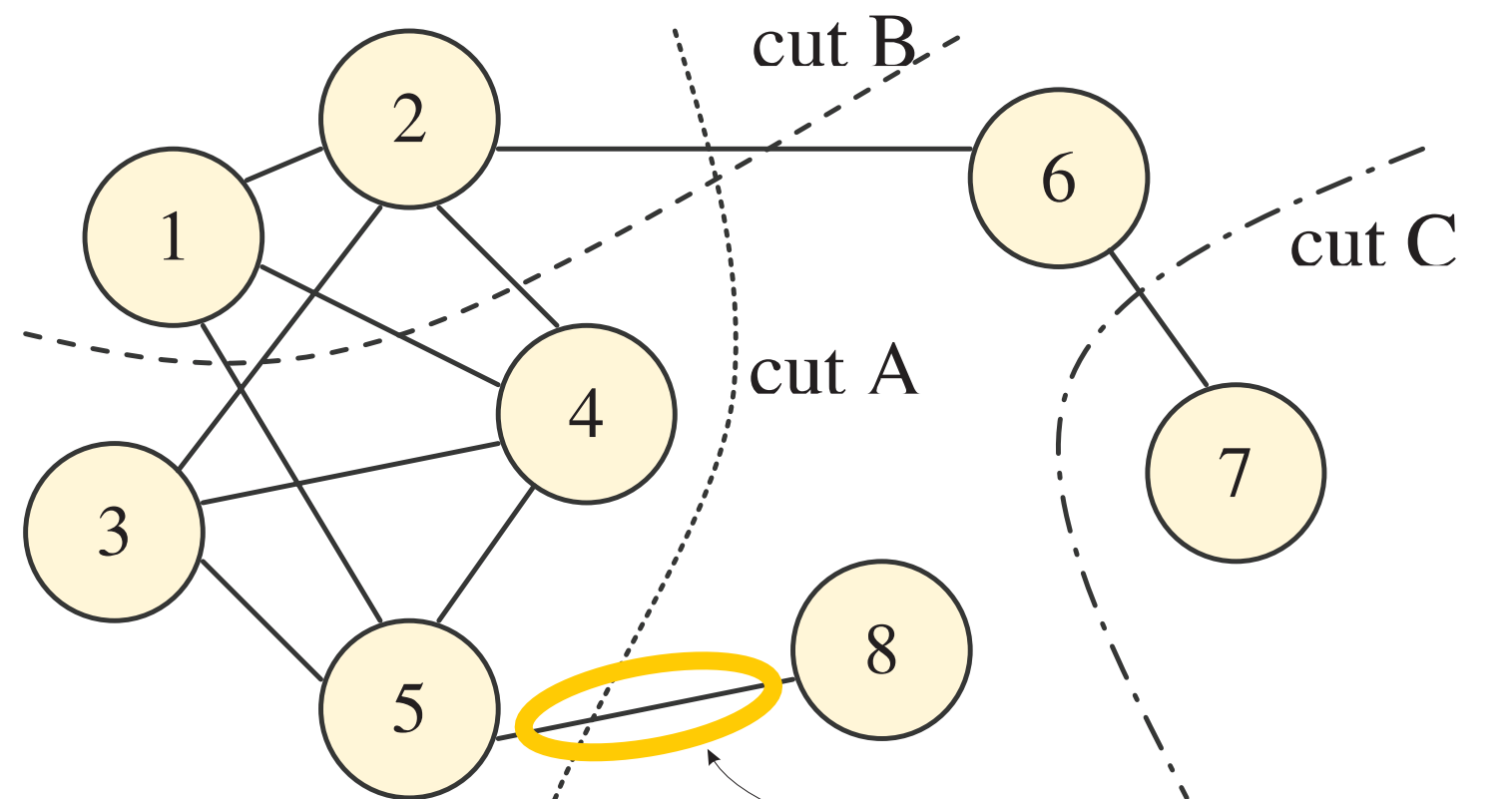
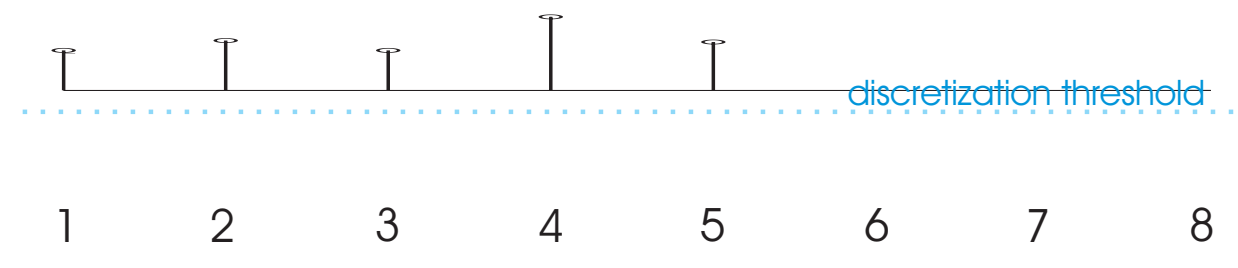


Figure 1. A simple toy problem

$$
\begin{array}{cc}
 & \text{1 .... points ... N} \\
A = & \begin{pmatrix}
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{pmatrix} \text{1 .... points ... N}
\end{array}
$$

Figure 2. Adjacency matrix for the toy problem



= cut A

Figure 3. Solution for the toy problem

The merit of a cut should increase with the number of edges in the inlier set, but should be penalized by the number of nodes. This leads us to the following heuristic:

$$r(u) = \frac{u^T A u}{u^T u}.$$

This metric computes the number of edges connecting the inliers, and divides by the total number of inliers. On the toy problem in figure 1, the marked cuts have the following scores:

Cut A	Cut B	Cut C
1.6	0.5	1.4

As intuitively desired, cut A has the highest merit. Now, we must determine how to solve for the best cut.

Solution

It is not known how to directly maximize the metric function r(u), when u is constrained to be discrete-valued. We follow the strategy of other spectral clustering algorithms by relaxing u to be continuous-valued. We can differentiate r(u) with respect to u:

$$\nabla r(u) = \frac{Auu^T u - u^T Auu}{(u^T u)^2} = \frac{Au - ru}{u^T u}$$

$$Au = ru.$$

This is an eigenvalue problem; the value r is maximized by setting u to the dominant eigenvector of A.

We now have a continuous-valued indicator vector, but in most applications, we need to know the discrete set membership. Several approaches are possible:

- Find the discrete vector which maximizes the metric by thresholding the continuous indicator vector.
- Find the (normalized) discrete vector that has the greatest dot product with the continuous indicator vector.

Both methods produce good (and similar) results. See Figure 3.

Computational Complexity

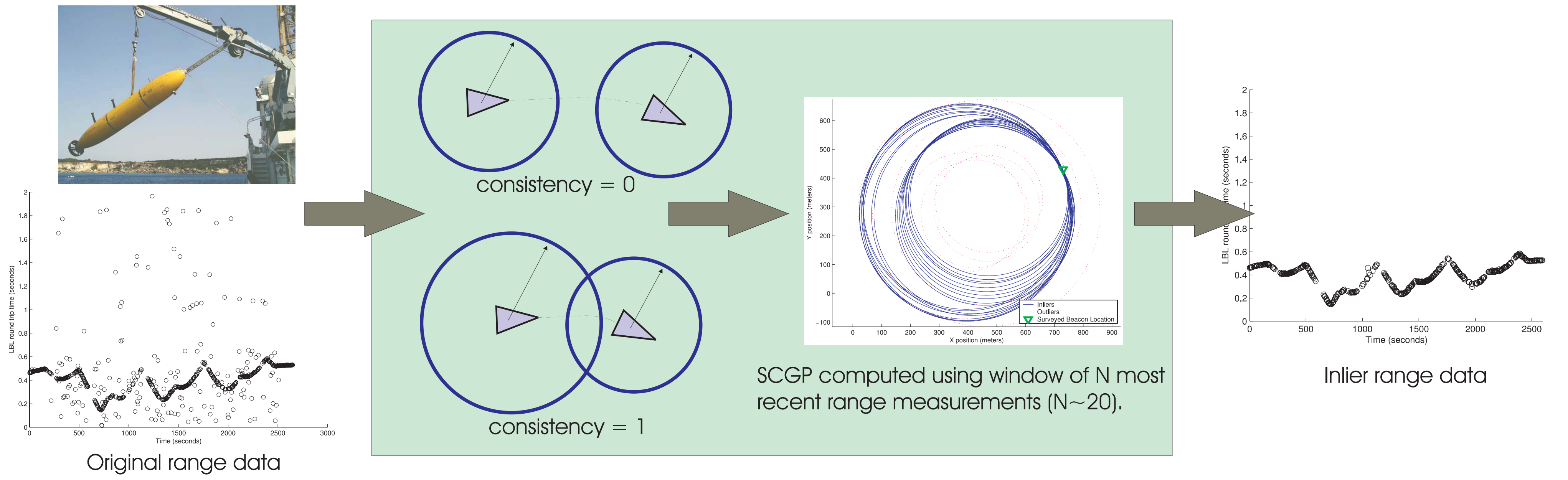
SCGP is $O(N^2)$ in both time and space. The slowest operation is computing the dominant eigenvector, but the Power Method can be used to find a good approximation.

http://cgr.csail.mit.edu

Outlier Rejection on Range-Only SLAM data

Sonar range measurements to stationary beacons were corrupted by extensive noise. We constructed an adjacency matrix by determining whether two range measurements had at least one intersection. A window of a few dozen measurements was used to causally filter the data.

SCGP was able to filter outliers without requiring a prior on the beacon positions.



Original range data

SCGP computed using window of N most recent range measurements (N~20).

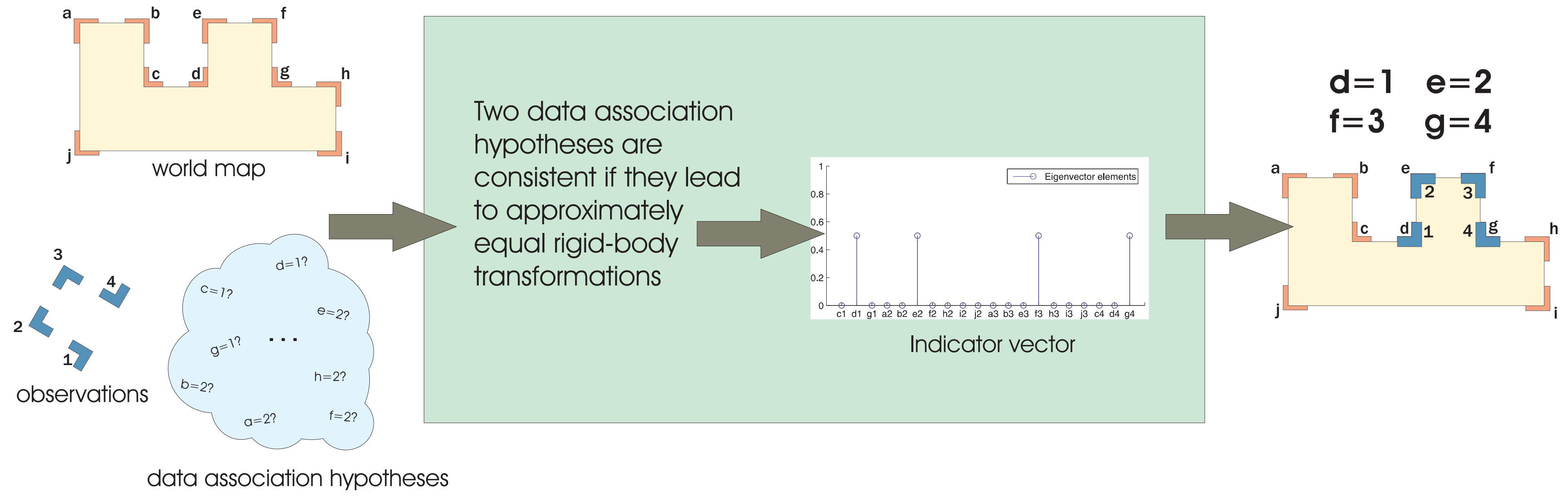
Inlier range data

Data Association on Corner Features

SCGP can perform data association in polynomial time, providing a fast (but approximate) alternative to Joint Compatibility Branch and Bound (JCBB).

In this toy problem, corner features are extracted from the environment and matched against a world map. A large number of possible corner associations are generated, and the pairwise consistency of these hypotheses are tested.

SCGP produces the correct data associations.



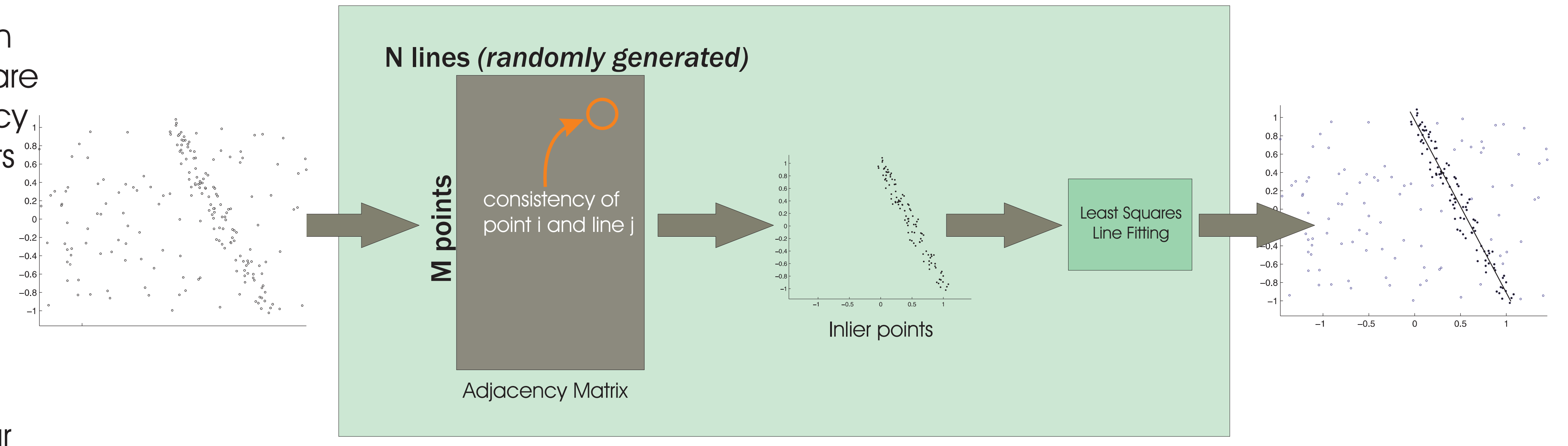
Two data association hypotheses are consistent if they lead to approximately equal rigid-body transformations

Line Fitting / Parameter Estimation

Line fitting is an application of non-symmetric/non-square SCGP. As with RANSAC, a number of hypotheses are randomly generated. The adjacency matrix tests the consistency of points versus lines.

SCGP extracts a set of inlier points, then fits a line to those points.

Performance is competitive, and often better, than RANSAC for similar computational complexity.


N lines *(randomly generated)*
M points
consistency of point i and line j
Adjacency Matrix
Inlier points
Least Squares Line Fitting