Combinatorial Vector Fields and the Structure of Fitness Landscapes

Bärbel M. R. Stadler

Max Planck Institute for Mathematics in the Sciences, Leipzig

Vysoka Lipa, Oct 2009
Fitness Landscapes

\[ L = (V, E, f) \]
Graph \( G(V, E) \), cost function \( f : V \rightarrow \mathbb{R} \)

- play a role in combinatorial optimization, evolutionary processes, folding of RNA and protein, ... 
- we assume that we want to minimize \( f \) (e.g. energy)
Walks on Landscapes

- **Gradient walks**: steepest descent, i.e., always take the neighbor with the lowest values of $f$
- **Adaptive walks**: accept any “downhill step”

Both gradient walks and adaptive walks have been used to characterize the *ruggedness* of the landscape (Stu Kauffmann, 1980s):

Idea: short walks $\Rightarrow$ many local minima $\Rightarrow$ rugged walks are longer on smoother landscapes
2nd application: partitioning of a landscape into valleys/basins:

**Idea:** assign to each vertex of $G$ a local minimum.

**Problem:** not unique

Even for gradient walks, unique only under certain circumstances (e.g. if $f$ is invertible on the neighborhoods $N[x]$ of all $x \in V$.)

Combinatorial Vector Fields

Consider walks as **dynamical systems** on the landscape. Formal framework: Robin Forman’s concept of *combinatorial vector fields* Defined on *simplicial complexes* $K$ map $\eta : K \rightarrow K \cup \emptyset$

with

1. if $\eta(\sigma) = \tau \neq \emptyset$, then $\dim(\tau) = \dim(\sigma) + 1$ and $\sigma < \tau$
2. if $\eta(\sigma) = \tau$ then $\eta(\tau) = \emptyset$
3. for every $\tau \in K$, $|\eta^{-1}(\tau)| \leq 1$

Graphs are also simplicial complexes
Formally much simpler:
map $\eta: (V \cup E) \to E \cup \{\emptyset\}$
for every edge $e$, $\eta(e) = \emptyset$ and $|\eta^{-1}(e)| \leq 1$.
for every vertex $v$, $\eta(v) = \emptyset$ ($v$ rest point) or $\eta(v) = e$
... sufficient to consider $\eta: V \to E \cup \{\emptyset\}$
Equivalently: interpret a c.v.f. on $G$ as orientation on a subset of edges such that each vertex has at most one out-edge.
Connection between landscape and combinatorial vector field? 

Lyapunov function associated with a combinatorial vector field: $f$ (weakly) decreases along each trajectory. 

(... skipping a lot formal stuff ...)

$f$ decreases along all oriented edges of the c.v.f.

Idea: consider the set of combinatorial vector fields for which the given landscape $f$ “is” a Lyapunov function

These c.v.f. represent the adaptive walks.
Neutral neighbors, $f(x) = f(y)$ is a problem because it destroys uniqueness of walks.

**Idea:** decompose the landscape/graph into patches where $f$ is constant.
\[N^> [x] = \{x\} \cup \{y \in V \mid \{x, y\} \in E \land f(y) < f(x)\}\]

For any subgraph \(H\) of \(G\) we define \(\vec{H}\) by the following sets of vertices and edges:

\[V(\vec{H}) = \bigcup_{x \in V(H)} N^> [x]\]

\[E(\vec{H}) = E(H) \cup \{\{x, y\} \in E \mid x \in V(H), y \in N^> (x)\}\]

We will call the subgraphs \(\Pi = \{\vec{G^f(x)} \mid x \in V\}\) the shelves of the landscape.
The sets $V(A)$, $A \in \Pi$, form a partition of $V$. The sets $E(A)$, $A \in \overrightarrow{\Pi}$, form a partition of $E$. 
**Generalization of Basins**

**Idea:** assign each vertex to more than one local minimum.

Define weights how the vertex is to be split up between the local minima.
Weights and Partition Functions

Begin with (Boltzmann) weights of edges. More weight for steeper edges.

Natural weight for a given c.v.f:
product of edge weights

\[ \omega(\eta) = \prod_{(x,y) \in \eta} \omega(\{x, y\}) \]

Technical devices: Partition functions over all combinatorial vector fields with all sorts of restrictions.

\[ Z = \sum_{\eta} \omega(\eta) \]

\[ Z_{(u,w)} = \sum_{\eta : (u,w) \in \eta} \omega(\eta). \]

... restricted to individual shelves, etc, etc, etc, ...
Reachability

Compute the relative weight of all c.v.f. that contain a path from $x$ to $y$.

Of particular interest: $y$ is local minimum

$$P(x \rightsquigarrow y) = \frac{1}{Z} \sum_{\eta: x \rightsquigarrow y} \omega(\eta)$$

Idea: Decompose paths to pieces within shelves. Recursively compute $P(x \rightsquigarrow y)$ from contributions on the shelves.

$\implies$ can be done (in terms of partition functions on individuals shelves).
Reachability ($\mathbb{P}(x \leadsto y) > 0$) defines the valley structure.

Partition of $V$ w.r.t. to sets of reachable minima/basins.
Conclusion

- Representation of walk in terms of combinatorial vector fields
- Landscape $\sim$ Lyapunov function
- Decomposition into “shelves” deals with neutrality of $f$
- Probabilities to reach local minimum $y$ from any starting point $x$ is computable by recursion analogous to the barriers program.
- Generalized “valley” structure where vertices are “divided up” between all local minima that can be reached from them.
- Potential applications for RNA folding kinetics as extension of barriers: e.g. provided more accurate estimate of transition rates between valleys.
Acknowledgements

- Peter
- Jürgen Jost
- Danijela Horak