7

Feature Extraction

What are features?

Features are inherent properties of data, independent of coordinate frames etc.

Dimension of a feature:

- 0: point feature (often defined by n equations for n coordinates)
- 1: line-like feature (n-1 equations)
- 2: surface-like feature
- etc.
- n: region-type feature (typically defined by a single inequality)

Region-type features

A feature is often indicated by high or low values of a derived field.

Example: vortical regions in a flow field have been defined by

- large magnitude of vorticity $\boldsymbol{\omega}(\mathbf{x}) = \nabla \times \mathbf{v}(\mathbf{x})$
- high absolute helicity $\boldsymbol{\omega}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})$ or normalized helicity $\frac{\boldsymbol{\omega}(\mathbf{x})}{\|\boldsymbol{\omega}(\mathbf{x})\|} \cdot \frac{\mathbf{v}(\mathbf{x})}{\|\mathbf{v}(\mathbf{x})\|}$
- positive pressure Laplacian $\nabla \cdot \nabla p(\mathbf{x})$
- positive second invariant of the velocity gradient $\nabla \mathbf{v}(\mathbf{x})$
- two negative eigenvalues of $\frac{\nabla \mathbf{v}(\mathbf{x})^2 + \left(\nabla \mathbf{v}(\mathbf{x})^T\right)^2}{2}$

The latter three definitions are parameter-free (preferred in feature definitions).

Point features in scalar fields

Point features in scalar fields:

- local minima/maxima
- saddle points

occur at zero gradient $\nabla s(\mathbf{x}) = 0$ (n scalar equations),
(places where height field is horizontal).

The above point features are the places where the contour line or isosurface changes its topology when the level is varied from min to max.

The contour tree (or Reeb graph) describes the split and join events.

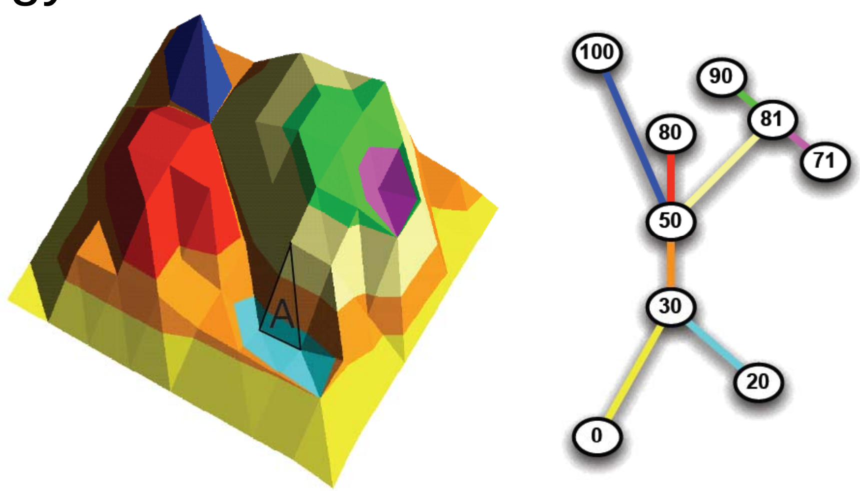


Image credit: S. Dillard

Line-like features in scalar fields

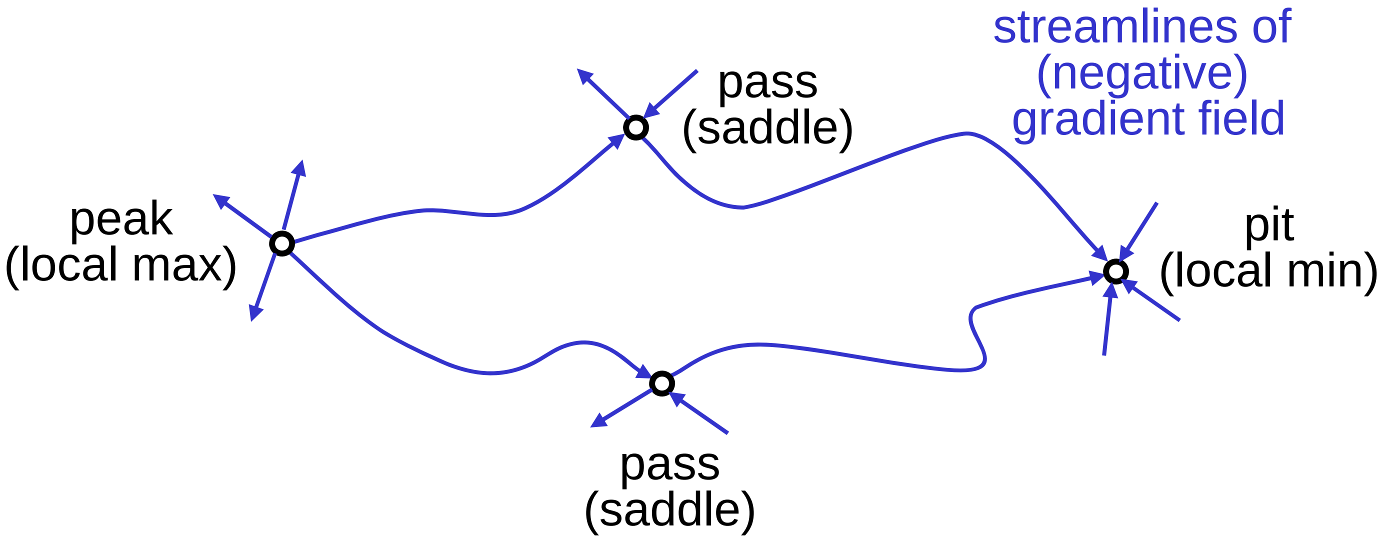
Line-like features in 2D scalar field:

Watersheds describe ridges/valleys of a height field s(**x**):

integrate the gradient field $\nabla s(\mathbf{x})$ (backward/forward),

starting at saddle points.

The watersheds provide a segmentation of the domain into so-called Morse-Smale complexes.



Watersheds require integration, are therefore not locally detectable.

Alternative definition of ridges/valleys (in *n*D scalar fields)?

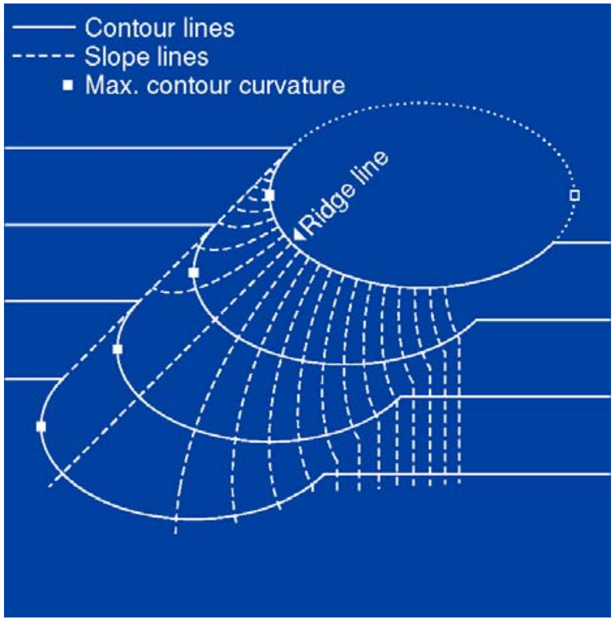
Local minima/maxima:

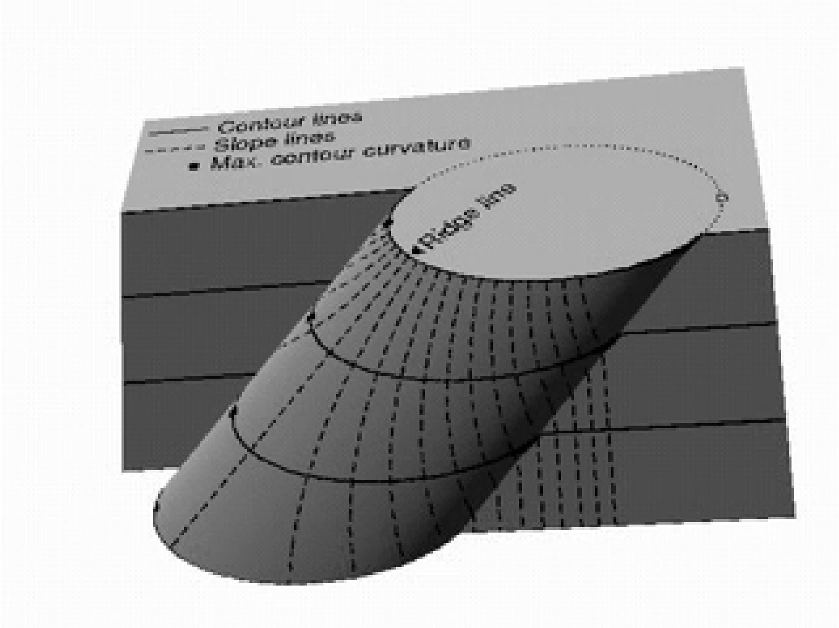
- Known at least since differentiation was invented (17^{th} century)!
- What is the natural extension to 1D?

Often used concepts:

- profile-based ridges
- curvature extrema on height contours

Counter-example for both (Wiener 1887!): "inclined elliptic cylinder"





Question: How are local maxima most naturally extended to 1D features?

Answer: height ridges.

Surprisingly, a formal definition of height ridges was given only in the 1990s (Eberly, Lindeberg), based on Haralick's definition (1983).

In contrast, local minima/maxima are known for centuries.

De Saint-Venant (1852) defined a concept similar to height ridges.

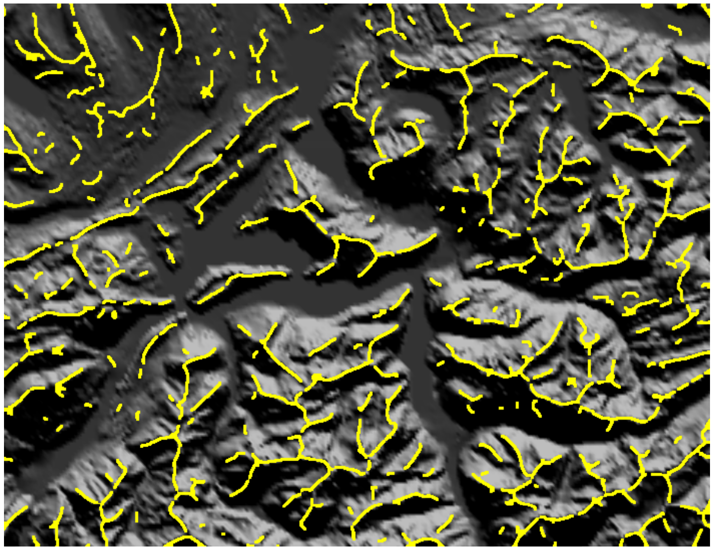


Image credit: P. Majer

At a given point $\mathbf{x}_0$ the scalar field has the Taylor approximation

$$s(\mathbf{x}_0 + \mathbf{x}) = s(\mathbf{x}_0) + \nabla s \cdot \mathbf{x} + \mathbf{x}^T \mathbf{H} \mathbf{x} + O\left(|\mathbf{x}|^3\right)$$

where **H** is the Hessian matrix of second derivatives

$$\mathbf{H} = \left(\frac{\partial^2 s(\mathbf{x})}{\partial x_i \partial x_j} \right)_{ij}$$

H has real eigenvalues and orthogonal eigenvectors.

By taking the eigenvectors as the coordinate frame, **H** becomes the diagonal matrix

$$\mathbf{H} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

A point $\mathbf{x} \in \mathbb{R}^n$ is a local maximum of s(**x**) if for all *n* axes:

- the first derivatives are zero: $s_{x_1} = \cdots = s_{x_n} = 0$
- the second derivatives are negative: $s_{x_1 x_1}, \cdots, s_{x_n x_n} < 0$

In the appropriate coordinate frame, this generalizes to:

A point $\mathbf{x} \in \mathbb{R}^n$ is on a *d*-dimensional height ridge of s(**x**) if for the first *n*-*d* axes:

- first derivatives are zero: $s_{x_1} = \cdots = s_{x_{n-d}} = 0$
- second derivatives are negative: $s_{x_1 x_1}, \cdots, s_{x_{n-d} x_{n-d}} < 0$

Appropriate coordinate frame means: axes are

- aligned with eigenvectors of **H**
- ordered by absolute eigenvalues: $|\lambda_1| \geq \cdots \geq |\lambda_n|$

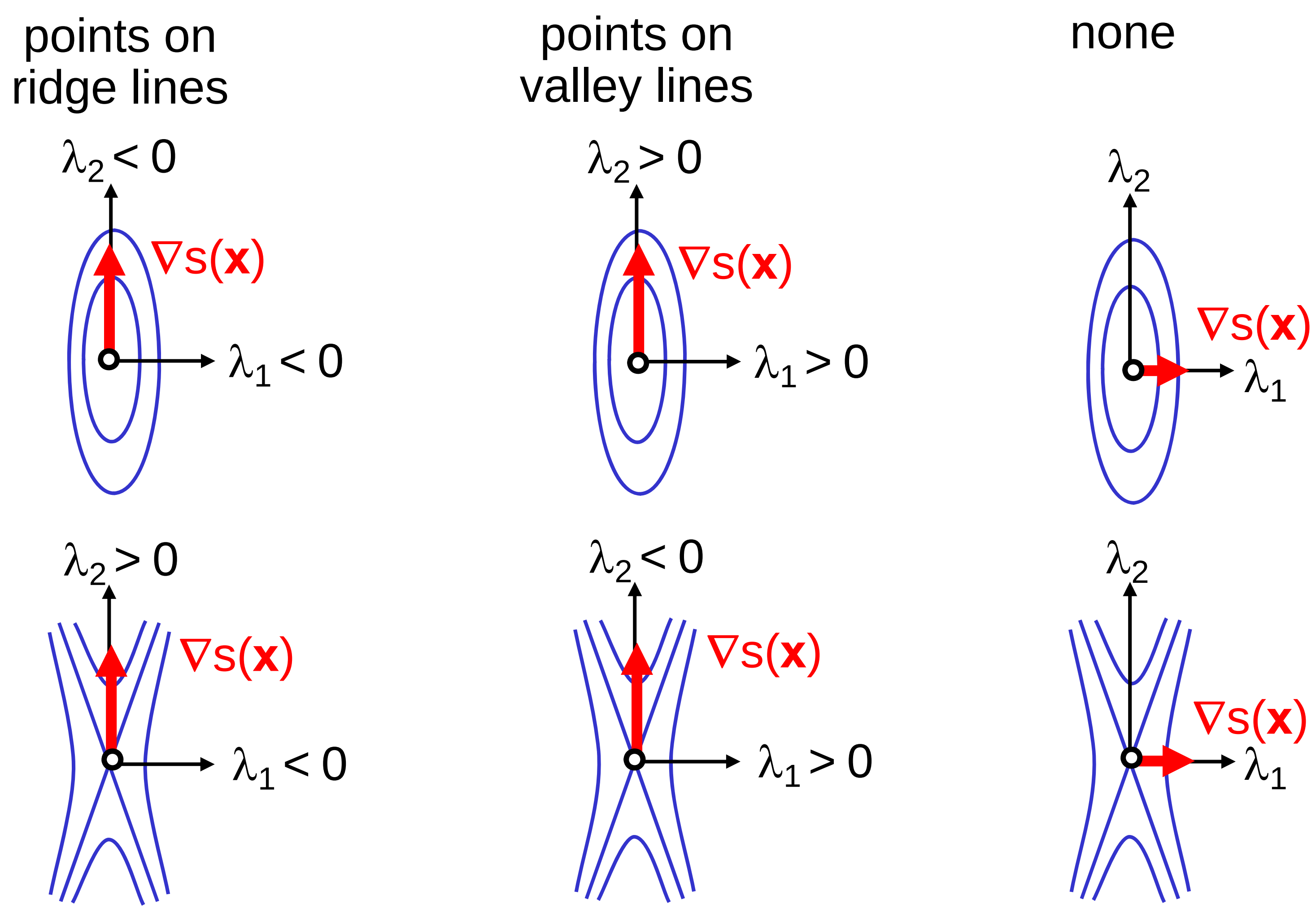
Remark: We used Lindeberg's definition. In Eberly's definition axes are ordered by signed eigenvalues: $\lambda_1 \leq \cdots \leq \lambda_n$

This is slightly weaker (accepting more points).

Example: scalar field, (1D) height ridge according to Eberly and Lindeberg:

image credit: P. Majer

Sketch of cases (with Lindeberg's definition, $|\lambda_1| \geq |\lambda_2|$)



"Circular gutter" example (Koenderink / van Doorn):

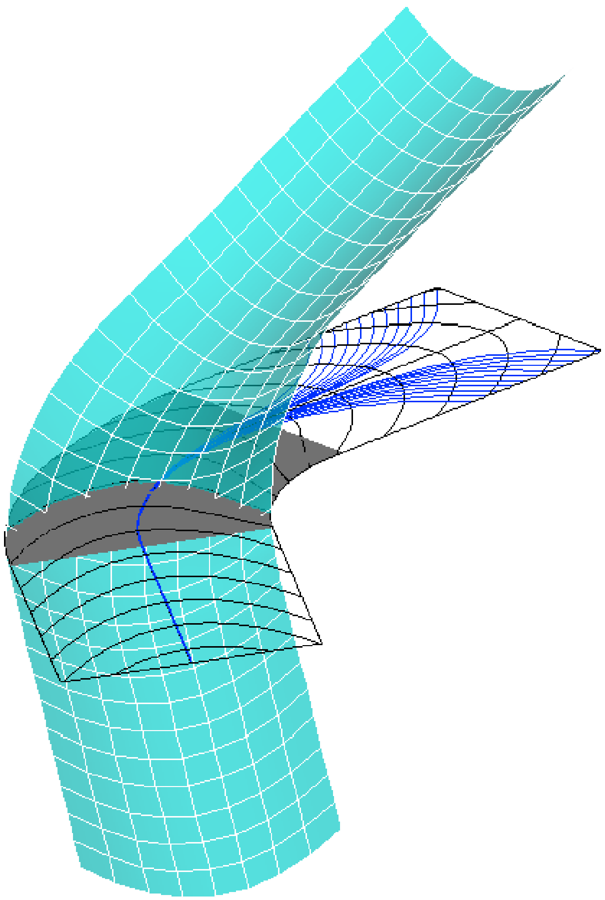
Height field in polar coordinates:

$$z(r,\varphi) = (1-r)^2 + k\varphi$$

- k describes the steepness in the tangential direction.
- Profiles in radial sections are parabolas:

$$z(r,\varphi) = (1-r)^2 + \text{const}$$

- Lowest points in sections $\varphi = \text{const}$ lie on the asymptote circle $r = 1$.

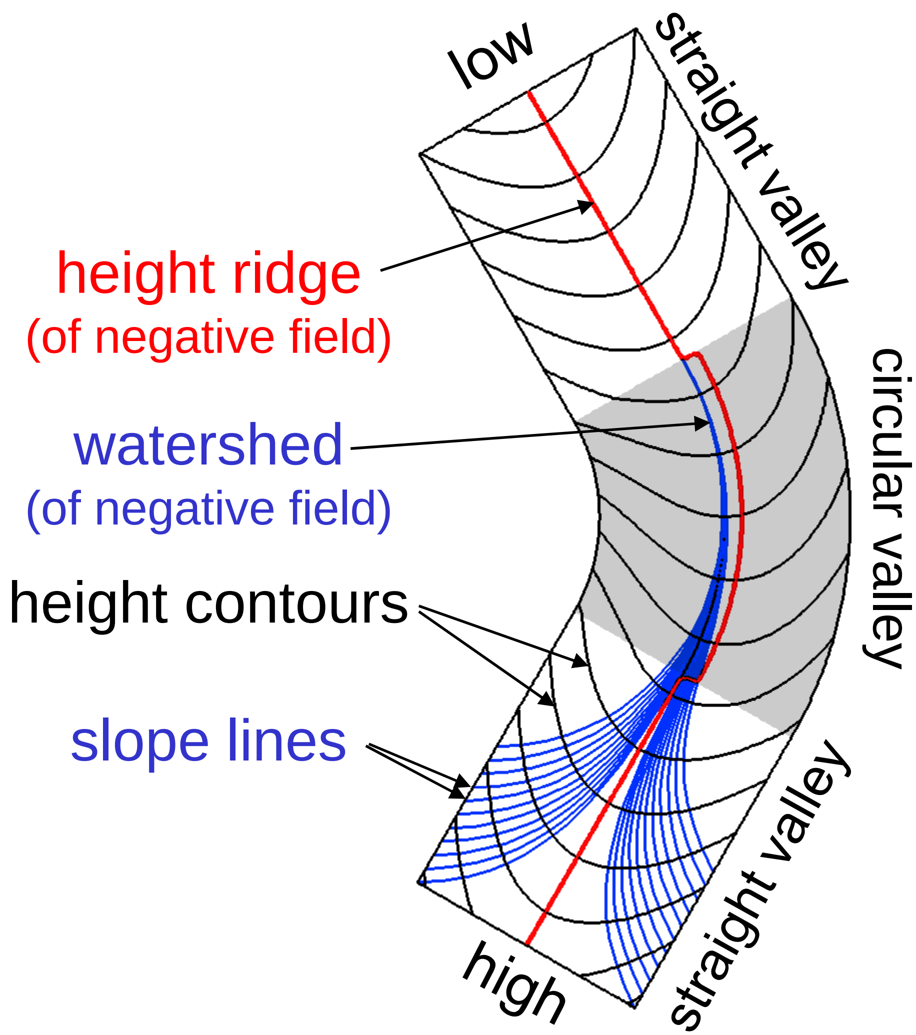


Circular gutter (with two straight segments added)

Circular gutter:

Height ridge deviates
(in the circular part)
from the solution
given by radial
profiles.

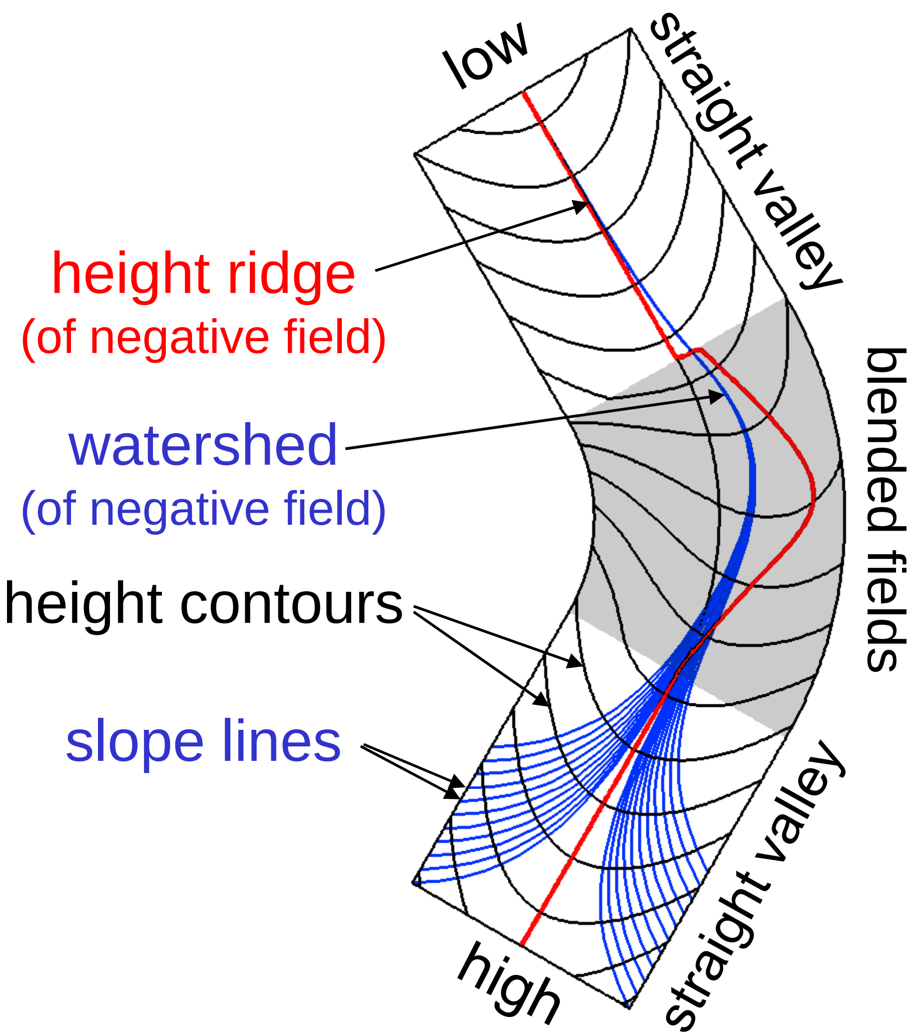
"Counter-example"
for height ridges.



Blended height fields
- (replacing the circular part by a blend of the two height fields):

Watershed deviates
- (in the lower part) from obvious symmetric valley line.

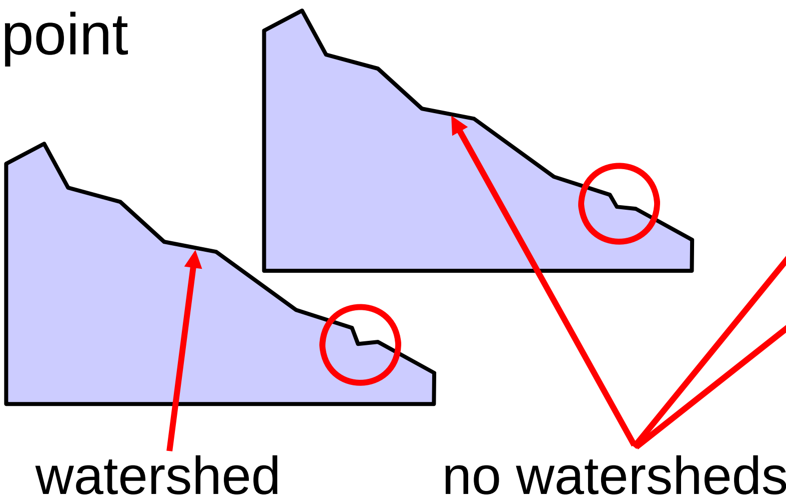
"Counter-example" for
- watersheds.



The use of watersheds vs. height ridges is still heavily discussed in computer vision (Koenderink/van Doorn '93).

Watersheds:

+ are slope lines of height field (=streamlines of gradient field)
- depend on boundaries
- require existence of a saddle point



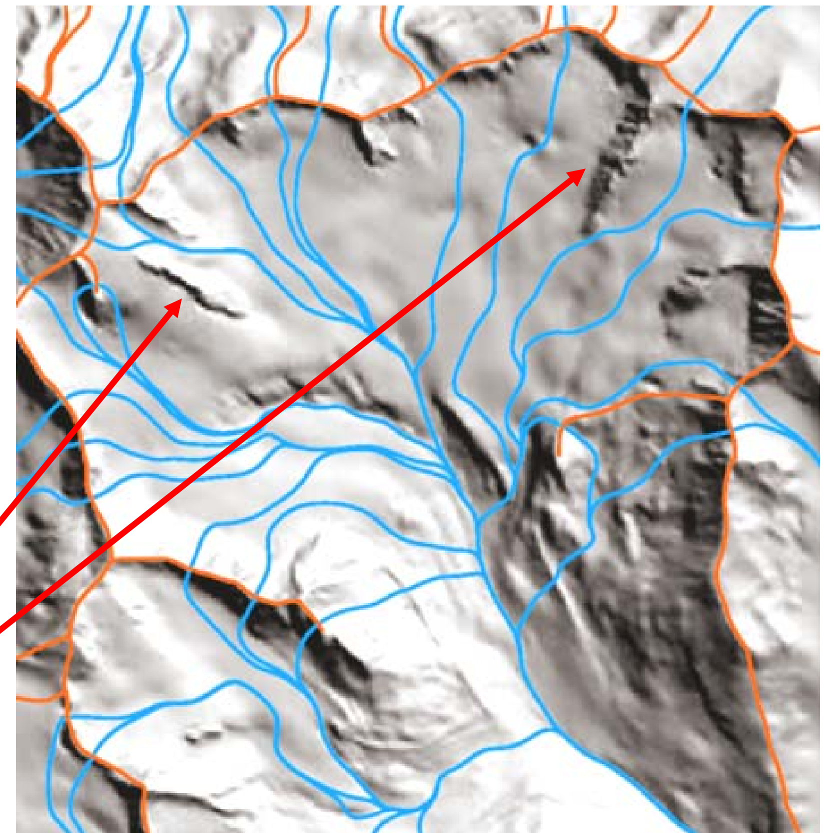
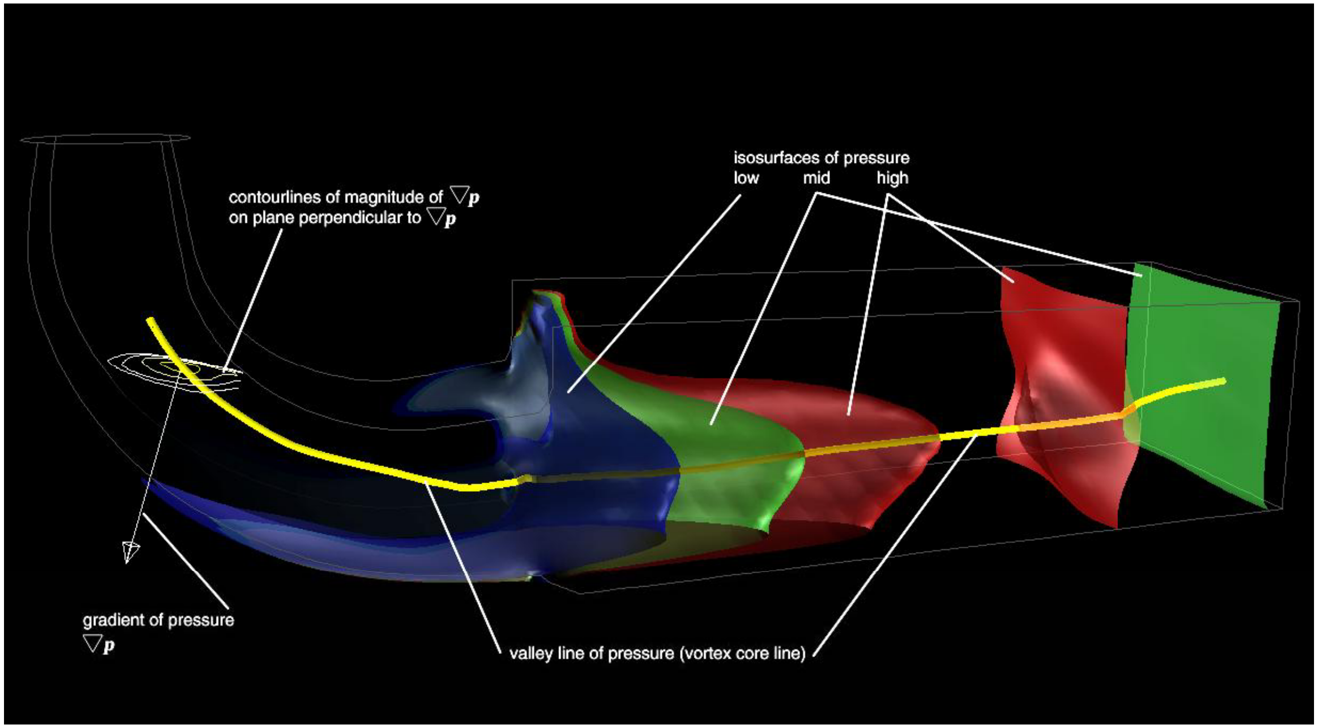


Image credit: C. Steger

Height ridges in 3D scalar fields can be used for defining/detecting vortex core lines. These are

- by Kida and Miura: height ridges (valley lines) of pressure



- by Ahmad/Kenwright/Strawn: height ridges of vorticity magnitude

Geometric features of surfaces

On surfaces in 3-space, 0- and 1-dimensional features can be defined by the (differential) geometry alone.

Geometric features vs. features of a field.

Examples of geometric features (not a core subject of SciVis), based on principal curvatures $\kappa_1, \kappa_2, \ |\kappa_1| \geq |\kappa_2|$

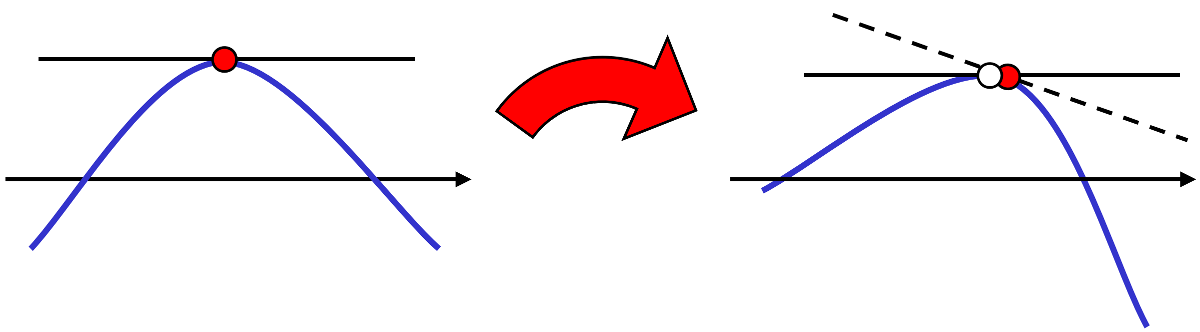
- umbilic points: $\kappa_1 = \kappa_2$
- curvature ridges: Loci of points where $|\kappa_1|$ is a maximum along the associated curvature line

Image credit: Y. Ohtake

The term "ridge" can refer to either height ridges or curvature ridges.

Curvature ridges are not appropriate as features of a scalar field (height field).

Reason: Invariance under rotation (tilting).



Line-like features in vector fields

Height ridges of a scalar field s(**x**) are definable by the gradient field $\mathbf{v}(\mathbf{x}) = \nabla s(\mathbf{x})$ alone:

- **H** is its Jacobian $\nabla\mathbf{v}(\mathbf{x})$, and
- s(**x**) itself is not needed.

A necessary condition for a height ridge is:

$\mathbf{v}(\mathbf{x})$ is an eigenvector of $\nabla\mathbf{v}(\mathbf{x})$

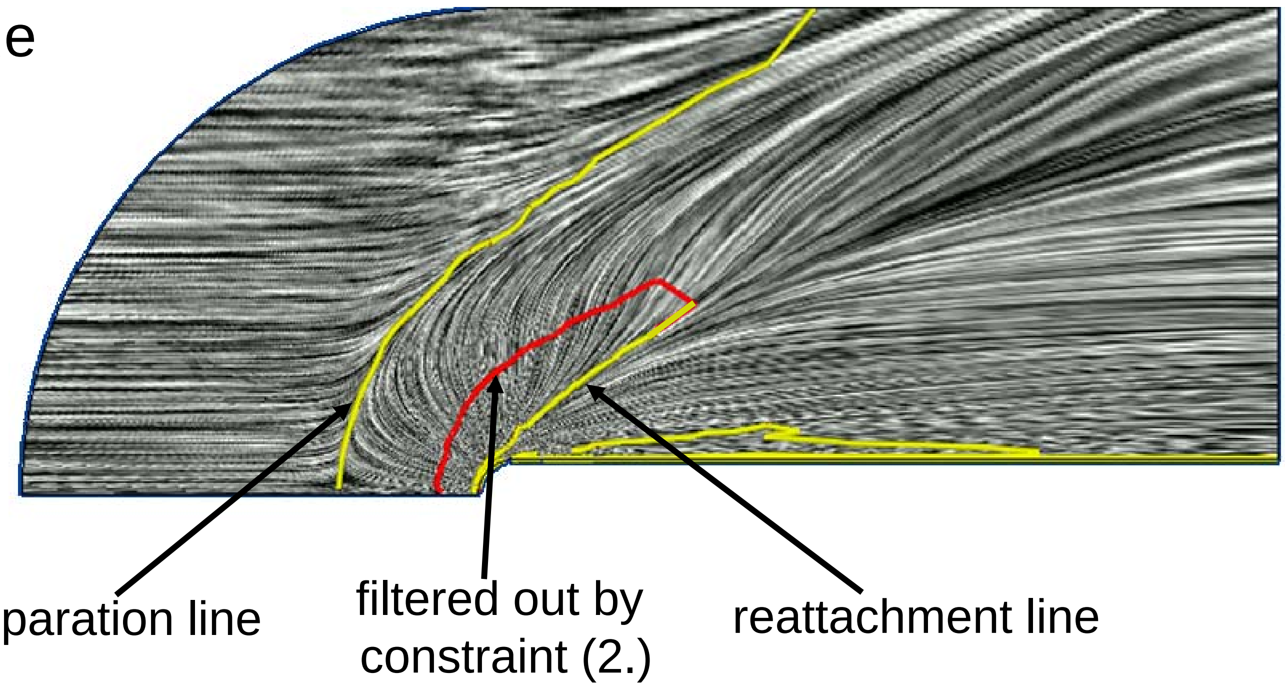
The gradient is a conservative (irrotational) vector field.

Let's now extend this to general vector fields.

Separation lines in boundary shear flow (Kenwright):

A point **x** lies on a separation or reattachment line if

1. at **x** the field vector $\mathbf{v}(\mathbf{x})$ is an eigenvector of $\nabla\mathbf{v}(\mathbf{x})$
2. the corresponding eigenvalue is the one with smaller absolute value



Vortex core lines (Sujudi / Haimes):

A point **x** lies on a vortex core line if

1. at **x** the field vector $\mathbf{v}(\mathbf{x})$ is an eigenvector of $\nabla\mathbf{v}(\mathbf{x})$
2. the two other eigenvalues of $\nabla\mathbf{v}(\mathbf{x})$ are complex

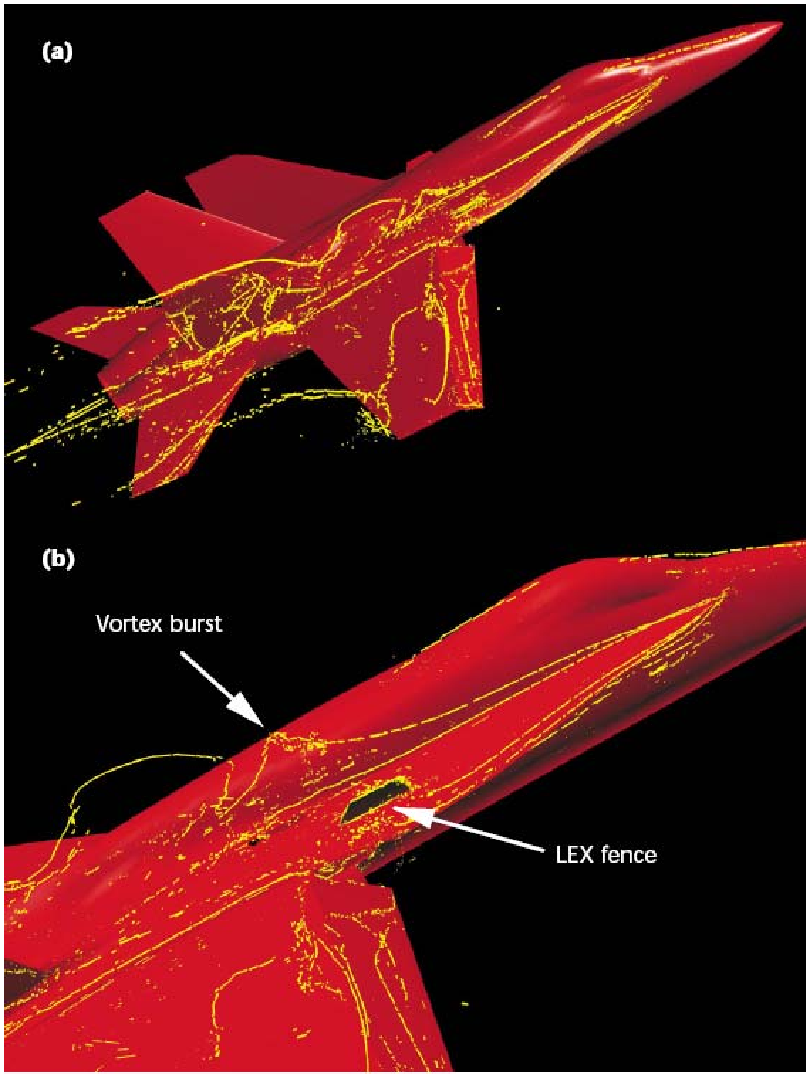


Image credit: D. Kenwright

Alternative definitions of vortex core lines:

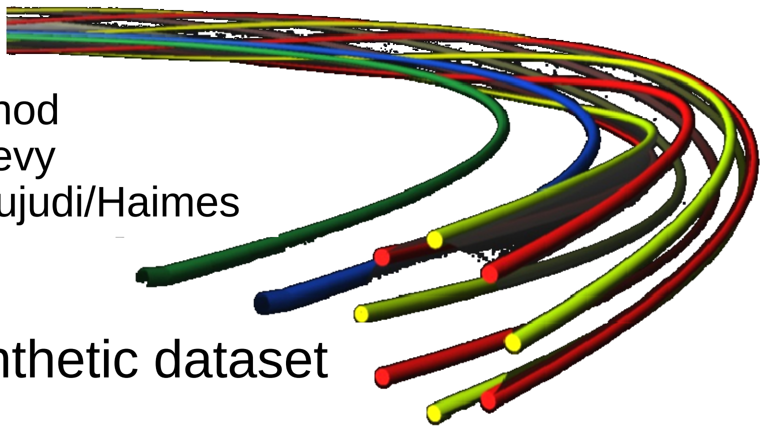
- According to Levy et al., longitudinal vortices have high normalized helicity (or small angles between velocity and vorticity).

  → vortex core line criterion: $\mathbf{v}(\mathbf{x})$ is (anti-) parallel to $\omega(\mathbf{x})$.

- Singer and Banks' method:
  - find a first point on the core line
  - repeat
    - predict next point along $\omega(\mathbf{x})$
    - correct to pressure minimum in normal plane of $\omega(\mathbf{x})$
  - compute vortex hull

Image credit: D. Banks

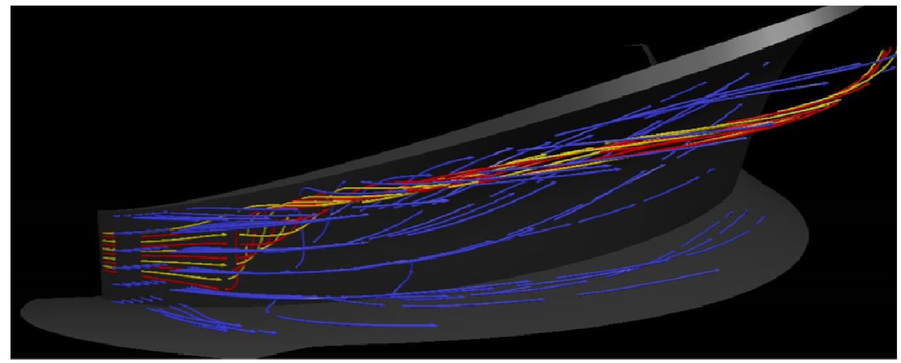
Deviation of locally computed vortex core lines:

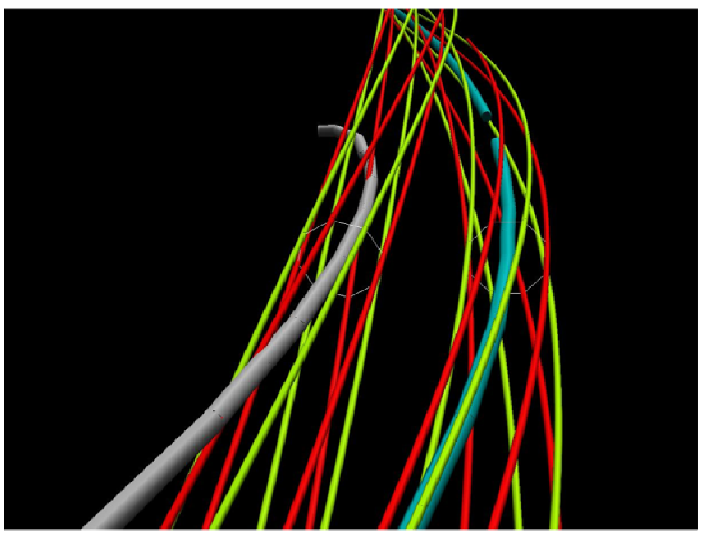


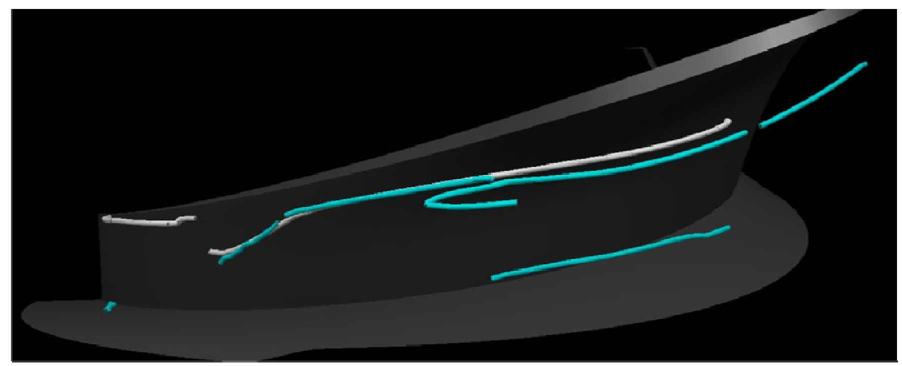
Method
- Levy
- Sujudi/Haimes

Synthetic dataset

Francis turbine runner







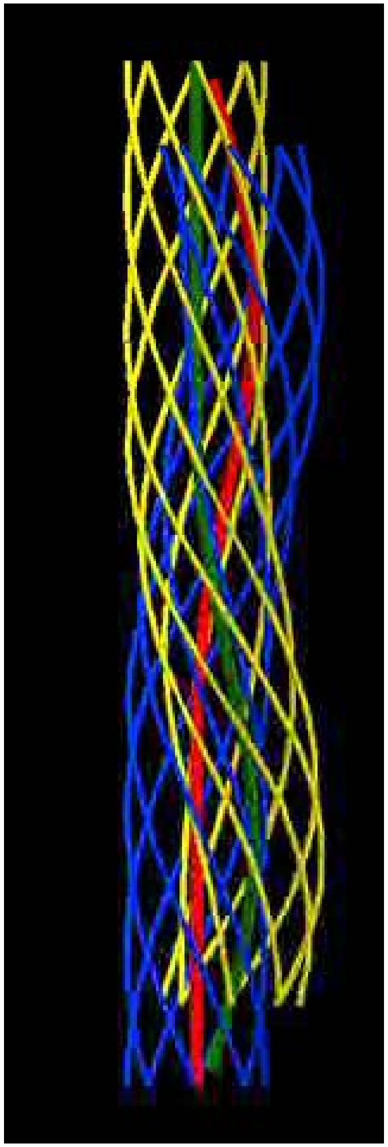
Method
- Sujudi/Haimes
- 2nd order

Discussion: local vs. global features

Global features: e.g. streamlines.

Are vortex core lines streamlines?

Here is a "counter-example":



How to compute line-like features?

Instead of explicitly computing eigenvectors for height ridges, Sujudi-Haimes core lines, etc.:

Make use of observation:

v is eigenvector of **A** if and only if **Av** is parallel to **v** (because **Av** = λ**v**)

Recipe:

- compute **w** = **Av** as a derived field
- find places where **v** and **w** are parallel (or one of them is 0).
- apply constraints
- apply post-filtering (vortex strength, etc.)

Parallel vectors operator:

given **v**, **w**: returns points where **v** and **w** are parallel

Implementation:

- in 2D: $\mathbf{v} \times \mathbf{w} = 0$ is just a contour line problem
- In 3D: $\mathbf{v} \times \mathbf{w} = 0$ is 3 equations for 3 unknowns:
  - equations are linearly dependent
  - can be solved with Marching Cubes like method

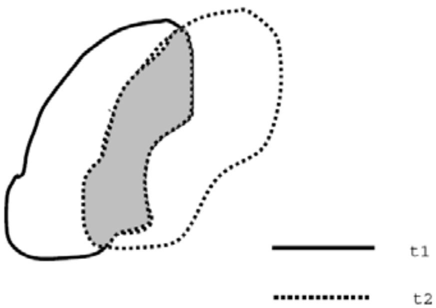
Tracking of features

In time-dependent data, features are usually extracted for single time steps.

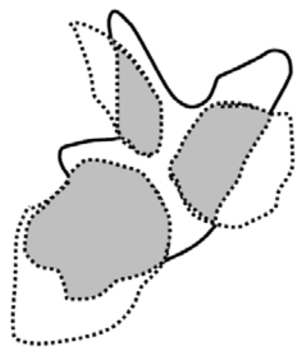
How to recognize a feature in a different time step?

Some methods are:

- Decide on spatial overlap (Silver et al.)
  - appropriate for region-type features
  - detects motion and events (split, merge, birth, death)



motion



split event

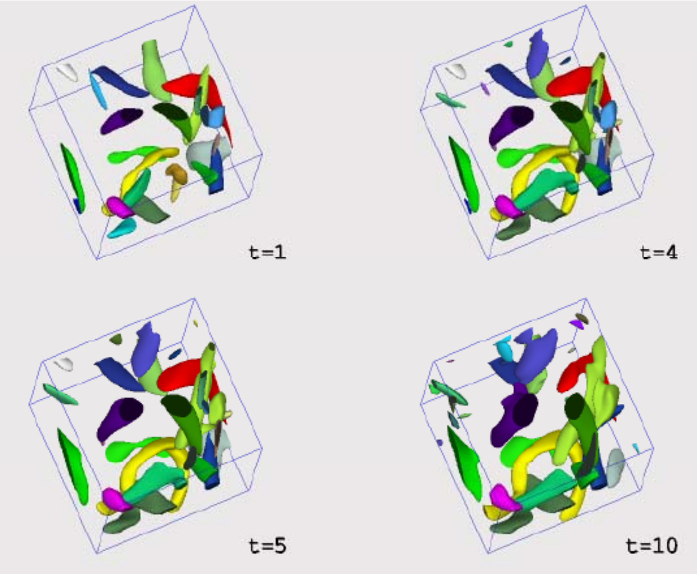
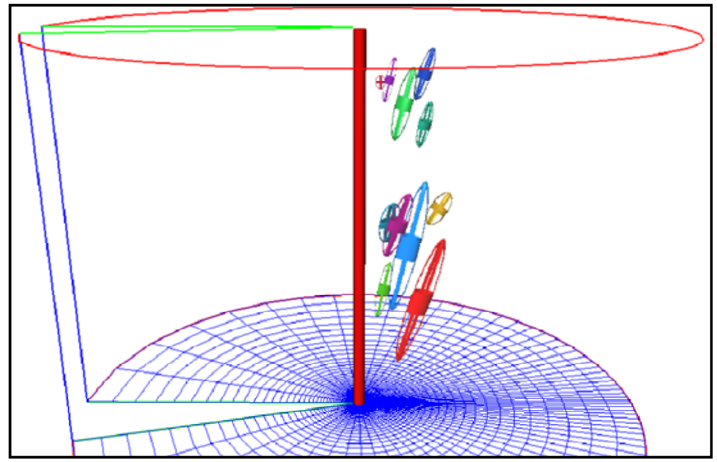


Image credit: D. Silver

- Decide on feature attributes (Reinders)
  - use attributes such as position, shape (fitted ellipsoid), orientation, spin, data values, etc.
  - combine with motion prediction



Flow past tapered cylinder.
Vortices represented by ellipsoids

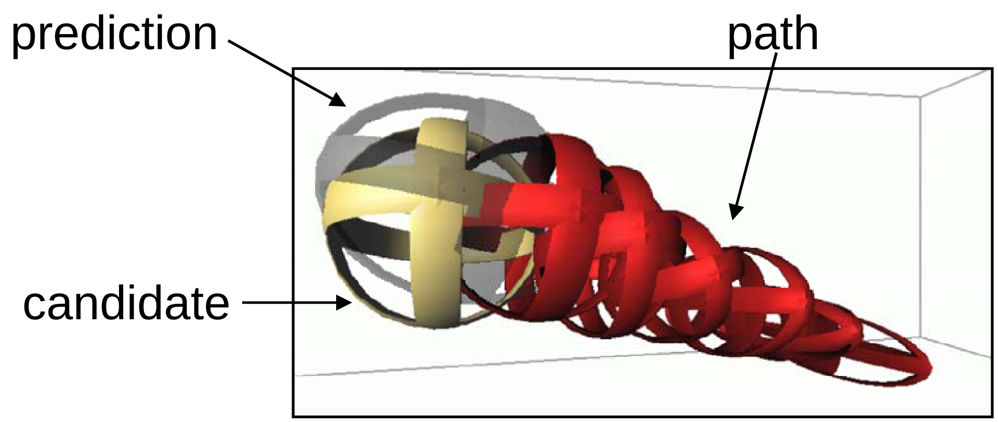


Image credit: F. Reinders

- Lift the feature extraction method to space-time domain.
  Examples:
  - Critical points in 2-space + time (Tricoche):
    Equations $u(x,y,t)=0, v(x,y,t)=0$ yield lines when solved in an "extruded" (x,y,t) grid.
    Features move along these lines, no explicit tracking needed.
  - Vortex core lines in 3-space + time (Bauer):
    Feature extraction yields a 2D mesh in 4-space.
    Time-slice is a line-like feature.