Data Visualization

Fall 2017

The Data as a Quantity

- Quantities can be classified in two categories:
  - **Intrinsically continuous** (scientific visualization, or scivis)

    e.g. pressure, temperature, position, speed , density, force, color, light intensity etc.
  - **Intrinsically discrete** (information visualization, or infovis)

    e.g. text, hypertext, content of web pages, database records etc.
- **Sampled data** – originally continuous data represented in a finite approximative form
- Corollary of the difference between sampled and discrete data:

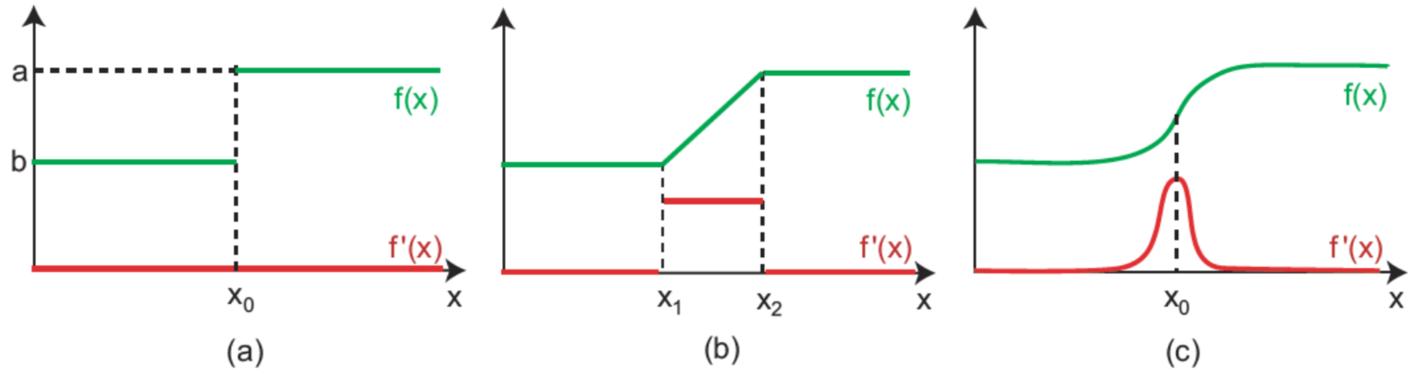
  In the case of sampled data, we can go back to a continuous approximation of the original (intrinsically) continuous data

Continuous Data

- Continuous data as a function:

$$f : \mathrm{D} \to \mathrm{C}$$

- Intuitive interpretation of continuity:



Discontinuous function | First-order C^0 cont. | High-order C^k cont.

Source: Data Visualization: Principles and Practice

Datasets and Dimensions

- Let the triplet *D* = (D, C, f) define a continuous dataset
- We assume that f : D → C
- D refers to a function domain, C is the function codomain
- d is the geometrical dimension of the space R^d into which D is embedded
- s <= d represents the topological dimension of D itself (e.g. plane in the Euclidean space R^3 has s = 2

  s is number of independent variables required to represent the domain D)
- Codimension of an object of some d and s is the difference d – s

Datasets and Dimensions

- Virtually all data-visualization apps fix geometrical dimension d = 3
- Only the topological dimension varies s = {1, 2, 3}
  - s = 1 corresponds to the curve
  - s = 2 corresponds to the surface
  - s = 3 corresponds to the volumetric datasets
- Topological dimension and dataset dimension are used interchangeably
- Topological dimension is important in case of sampled datasets and grid cells

Datasets and Dimensions

- Function values are called dataset attributes
- The dimensionality c of the function codomain C is also called the attribute dimension
- Typically ranges from 1 to 4

Sampled Data

- Typically, continuous functional representation of data is not available
- Moreover, several operations (filtering, simplification, analysis etc.) on such data are not efficient
- Visualization apps work with sampled datasets
- Important operations relating continuous and sampled data:
  - **Sampling** – quite straightforward
  - **Reconstruction** – more complicated, the goal is to recover an approximated version of the original continuous data

Sampled Data

- Reconstruction employs interpolation of the values of the function between its sample points
- Sampling strategies: uniform, non-uniform
- Sampled dataset should be accurate (up to an user-specified error), minimal (resp. to an error), generic (operations), efficient (algorithmically), and simple (implementation) [Schroeder et al. 2006]

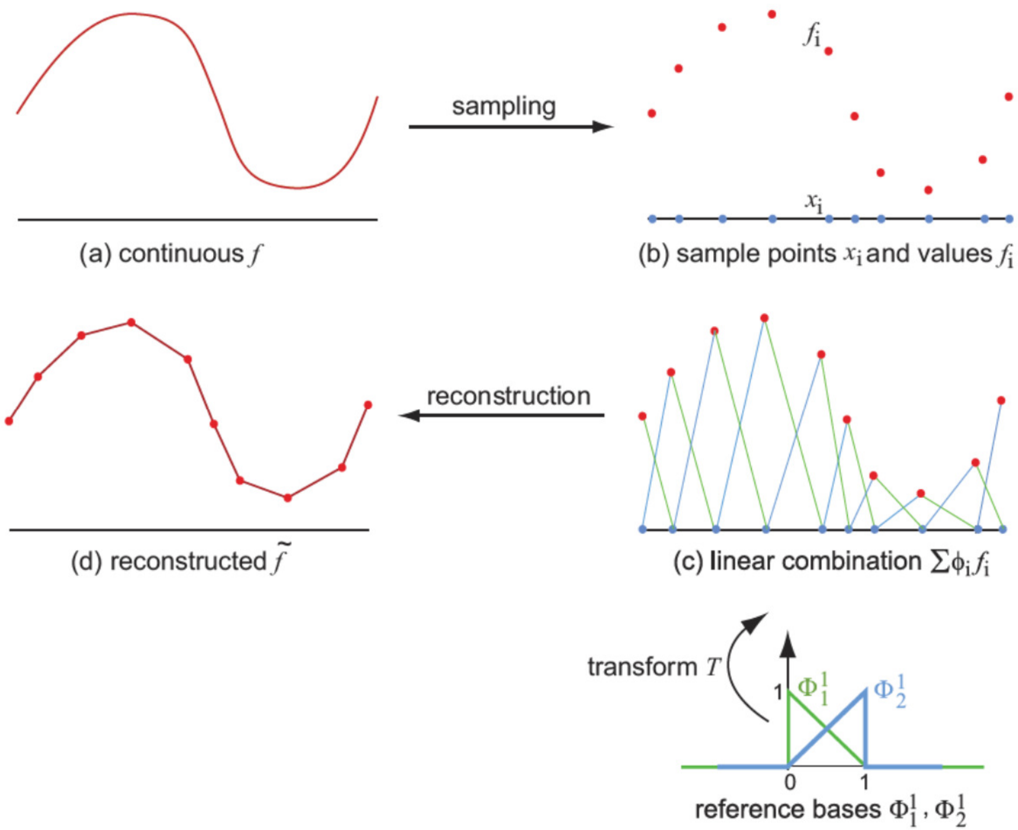
Sampled Data

- Sampled dataset $\{p_i, f_i\}$ consists of a set of N sample points and values
- Interpolation - the reconstructed function should equal the original one at all sample points, i.e. $\tilde{f}(p_i) = f(p_i) = f_i$
- One way to define reconstruction function: $\tilde{f} = \sum_{i=1}^{N} f_i \phi_i$

Basis/interpolation functions

- Subsequently, we get $\sum_{i=1}^{N} f_i \phi_i(p_j) = f_j, \forall j$
- Orthogonality basis function $\phi_i(p_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- Normality of basis function $\sum_{i=1}^{N} \phi_i(x) = 1, \forall x \in \mathrm{D}$

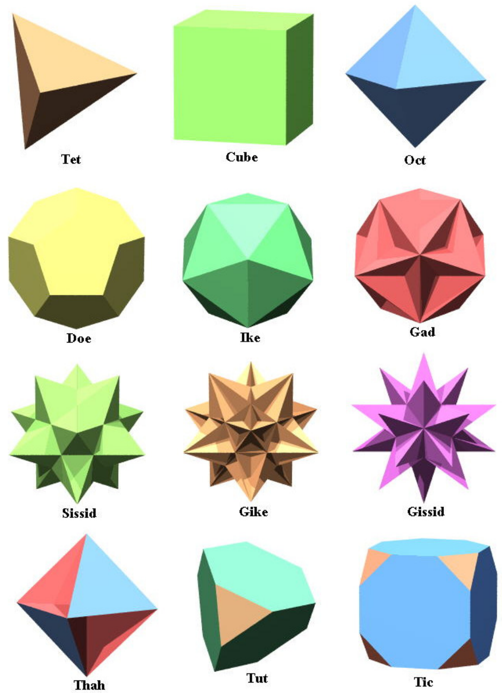
Sampled Data



Source: Data Visualization: Principles and Practice

Grid

- A grid (aka mesh) is a subdivision of a domain D into a collection of cells (aka elements)
- Most commonly used cells:
    - Polylines in R
    - Polygons in R^2
    - Polyhedra in R^3
- Union of cells cover entire domain D and cells are non-overlapping, and vertices are sample points



Grid

constant, zero-order continuity global basis function

- Simplest set of basis function: $\phi_i^0(x) = \begin{cases} 1, & x \in c_i \\ 0, & x \notin c_i \end{cases}$
- Sample points are inside the grid cells
- Nearest-neighbor interpolation
- Virtually no computation cost
- Work with any cell shape and in any dimension
- Provide a poor, staircase-like approximation
- We can provide a better (i.e. more continuous) reconstruction of the original function

Grid

- Linear basis function – next simplest basis functions
- Need to make some assumption about the cell types used in the grid
- Assume quadrilateral cells having 4 vertices
- Reference quad cell in R^2: $v_1=(0,0)$, $v_2=(1,0)$, $v_3=(1,1)$, $v_4=(0,1)$

Set of local basis functions

$$
\begin{aligned}
\Phi_1^1(r,s) &= (1-r)(1-s),\\
\Phi_2^1(r,s) &= r(1-s),\\
\Phi_3^1(r,s) &= rs,\\
\Phi_4^1(r,s) &= (1-r)s;
\end{aligned}
$$

reference coordinates

Grid

- Given any cell type having n vertices p_i in R^3, we define transformation T that maps from a point r, s in reference cell coordinate system to a point x, y, z in the actual cell as follows

$$(x, y, z) = T(r, s) = \sum_{i=1}^{n} p_i \Phi_i^1(r, s)$$

- T maps the reference cell to the world cell
- T^{-1} maps points x, y, z in the world cell to points r, s in the reference cell

Grid

- Having T^{-1}, we can rewrite the reconstruction function $\tilde{f} = \sum_{i=1}^{N} f_i \phi_i$ for quad cell as

$$\tilde{f}(x, y, z) = \sum_{i=1}^{4} f_i \Phi_i^1 (T^{-1}(x, y, z))$$

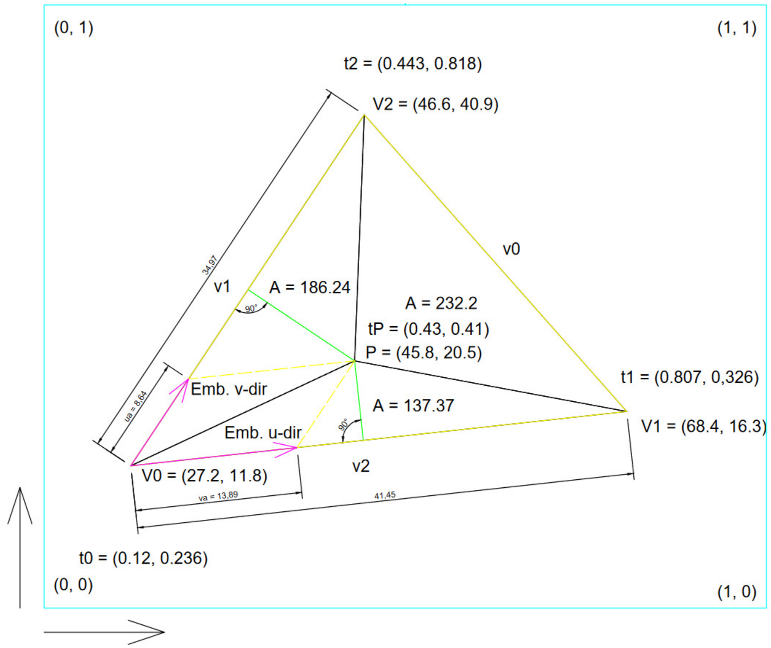
- To compute T^{-1}, we have to invert the expression $T(r, s) = \sum_{i=1}^{n} p_i \Phi_i^1 (r, s)$
- Given a rectangular cell, this yields

$$T_{\text{rect}}^{-1}(x, y, z) = (r, s) = \left(\frac{(p - p_1) \cdot (p_2 - p_1)}{\|p_2 - p_1\|^2}, \frac{(p - p_1) \cdot (p_4 - p_1)}{\|p_4 - p_1\|^2} \right)$$

- Now, we have a way to reconstruct a piecewise C^1 function from samples on any rectangular grid. Arbitrary quad cells require some numerical solution for r, s

Barycentric Coordinates

- A simplex is a convex hull of k+1 points in a k-dimensional space
- Barycentric coordinates provide a simple way to interpolate over simplices



Barycentric Coordinates for Triangles

- Describe location of point in a triangle in relation to the vertices
- $\mathbf{p} = (u, v, w) = u\,\mathbf{v}_0 + u\,\mathbf{v}_1 + w\,\mathbf{v}_2$ and $u + v + w = 1$ where $u, v, w \geq 0$
- In the same way, we can interpolate any quantity (or function) along the triangle
- Coordinates are the signed area of the opposite subtriangle divided by area of the triangle

Grid

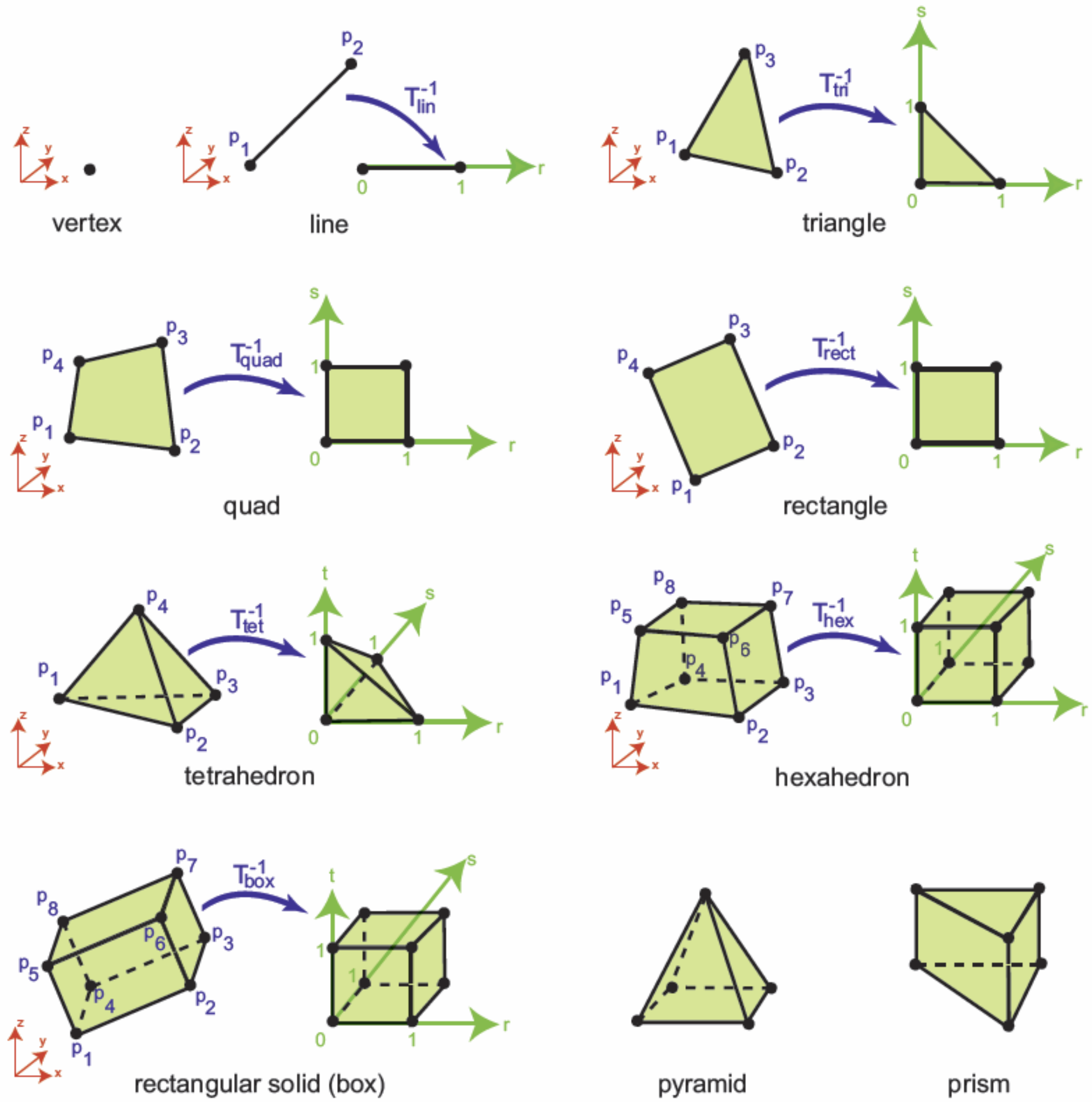
- Same situation but with triangle cell (simplest cell in 2D)
- Three linear basis functions

$$\Phi_1^1(r,s) = 1 - r - s,$$
$$\Phi_2^1(r,s) = r,$$
$$\Phi_3^1(r,s) = s.$$

- The transformation T^{-1} for triangular cells

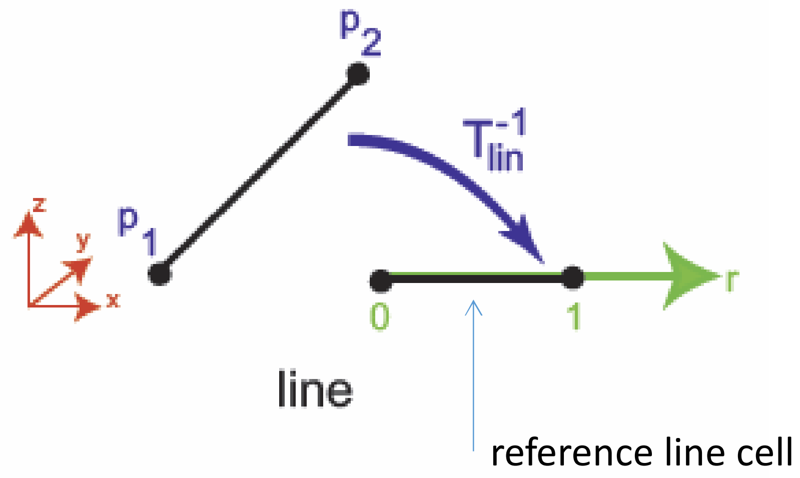
$$T_{\text{tri}}^{-1}(x,y,z) = (r,s) = \left(\frac{\|(p-p_1)\times(p_3-p_1)\|}{\|(p_2-p_1)\times(p_3-p_1)\|}, \frac{\|(p-p_1)\times(p_2-p_1)\|}{\|(p_3-p_1)\times(p_2-p_1)\|} \right)$$

Cells



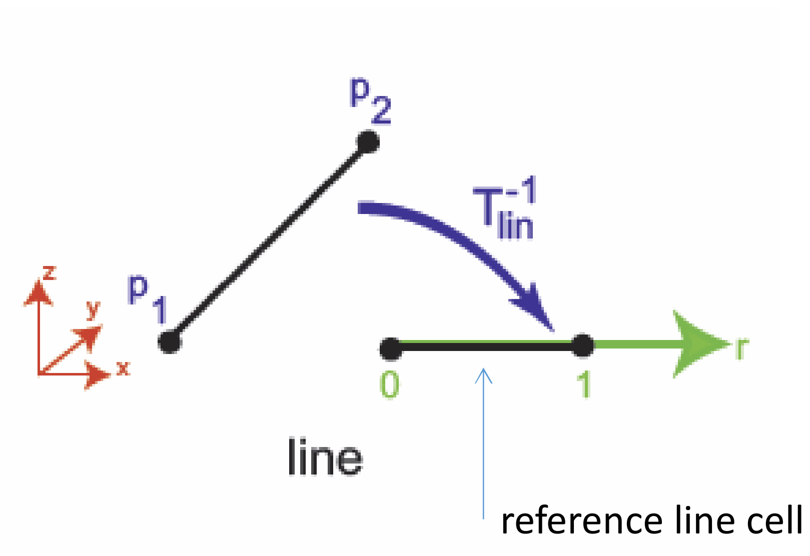
Source: Data Visualization: Principles and Practice

Line Cell



Source: Data Visualization: Principles and Practice

Line Cell

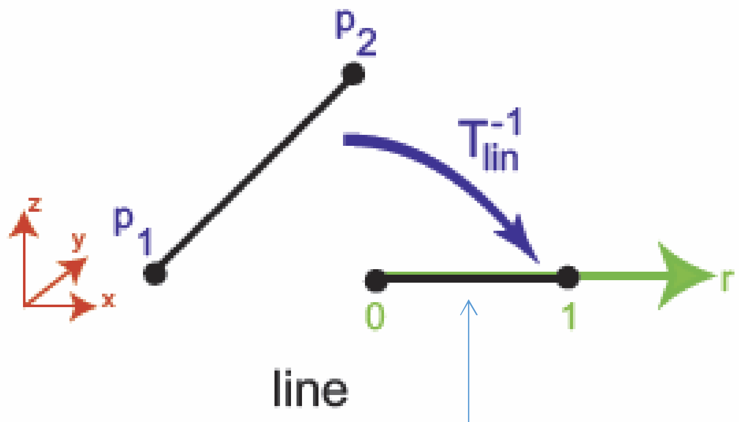


$$\Phi_1^1(r)$$
$$\Phi_2^1(r)$$

$$T_{\text{lin}}^{-1}(x, y, z)$$

Source: Data Visualization: Principles and Practice

Line Cell



$$\Phi^1_1(r) = 1 - r$$
$$\Phi^1_2(r) = r.$$

$$T^{-1}_{\text{lin}}(x, y, z) = \frac{\|p - p_1\|}{\|p_2 - p_1\|}$$

two linear basis functions

Source: Data Visualization: Principles and Practice