VDA all Questions

Visual Data Analytics (Technische Universität München)

a) Explain the visualization pipeline. What are the four stages?

General model of visualisation process user can interact at each of these stages and it consists of 4 stages:

Data acquisition:
Filtering/Enhancing:
Visualization Mapping:
Rendering:

b) Explain the data acquisition stage. What are three general cases?

The stage where we collect data.
Databases
Simulation
Sensors
-> Leads to raw data (global climate sim, projected densities from CT, sonar)

c) Explain the filtering/enhancement stage. Give at least two examples.

The stage where raw data turns into derived data by resampling to grid, interpolation, data reduction, co-registration ,remove non relevant data. etc.

d) Explain the visualization mapping stage. Give at least two examples.

The stage where the derived data becomes graphical primitives which has renderable representation. EX: scalar field -> isosurface, vector field -> vectors, any flow vis, and volume vis

e) In which stage of the visualization pipeline happens resampling to a regular grid?

Filtering/Enhancement Stage

f) In which stage of the visualization pipeline are the viewpoint and lighting parameters specified?

Rendering Stage

g) In which stage of the visualization pipeline happens lighting and shading?

Rendering Stage

h) In which stage of the visualization pipeline are colors assigned to every voxel?

Visualization Mapping Stage (transfer function)

i) In which stage of the visualization pipeline happen smoothing and noise suppression?

Filtering/Enhancement

Data Representation

a) Discuss independent vs. dependent variables in data. Give at least two examples each.

Dimension of the domain of the problem -> Type and dimnsion of data to be visualised

Independent Variables: Spatial Coordinates, Time

Dependent Variables: Temperature, Density, Velocity vector

b) What are the independent and dependent variables in a 3D spatial curve $\phi : \mathbb{R} \rightarrow \mathbb{R}^3$

Independent: $\mathbb{R}$
Dependent: $\mathbb{R}^3$

c) What are the independent and dependent variables in a 3D vector field?

Independent: $\mathbb{R}^3$
Dependent: $\mathbb{R}^3$
Each vector given at every location (3D)

d) What type of attribute are the following: categorical, ordinal, or quantitative?
   a. Type of cheese (e.g., Swiss, Brie) (Ca)
   b. Tire pressure (e.g., 2.3 bar, 2.5 bar) (Qu)
   c. First name (e.g., Alice, Bob) (Ca)
   d. Unemployment rate (e.g., 6%, 10%) (Qu)
   e. T-Shirt sizes (e.g., medium, large) (Or)

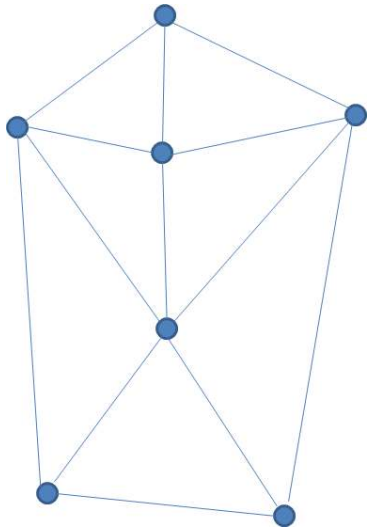
e) Draw an illustration of a Cartesian grid. Describe how such a grid is different from a regular grid? Which information needs to be specified explicitly for such a grid?
Both height and weight are partitioned into equal parts for Cartesian grid whereas in the regular grid only the partitions of the height are equal given in grids and width are equal but width!=height. Position implicit, neighbourhood implicit

f) What is a curvilinear grid? How is it characterized? How is it different from an unstructured grid? Which information needs to be specified explicitly for such a grid?

Curvilinear grid is a non-orthogonal grid where the neighborhood information is specified implicitly whereas it is explicitly specified for unstructured grids.
Unstructured grid: tetrahedra, hexahedra. Both neighborhood and positions are explicit.
Curvilinear grid: Non orthogonal, positions explicitly and neighborhood implicitly

Data Interpolation

a) For the triangulation shown below, proof or disproof that this triangulation is a Delaunay triangulation. Your proof should be geometrically, meaning that you either illustrate the Delaunay property in the figure or illustrate that this property is violated.



Draw a circle for each triangle. Make sure no other vertex is in circle

b) An interpolation function $f(x) = \sum^{N} w_i \varphi(\|p_i - x\|)$ is a weighted sum of *N radial functions*

$\varphi(r) = e^{-r_2}$ where $\|p - x\|$ is the distance between the points p_i and x.

Compute the weights w_i such that the function $f(p_i)$ interpolates the data points $p_1 = 1$, $p_2 = 3, p_3 = 3.5$ with corresponding scalar values f_1 f1= 1, f_2= 0, $f_3 = {}^{1/4}$

The table below shows approximate values for $\varphi(r)$ with respect to different distances r.

r	0	0.5	1	1.5	2	2.5	3
$\varphi(r)$	1	$4/5$	$2/5$	$1/10$	0	0	0

F(x)= w1 φ (||p1-x||) +w2 φ(||p2-x||)+.... For all points.
Then get eqns f(p1)= 1 =w1.φ(dist)+ .... get an eqn for each point
finally solve linear eqn and find w1,w2,w3

Data Interpolation

c) Given the interpolation function $f(x)$ from the previous assignment, compute the interpolated value at point $x = 2$.

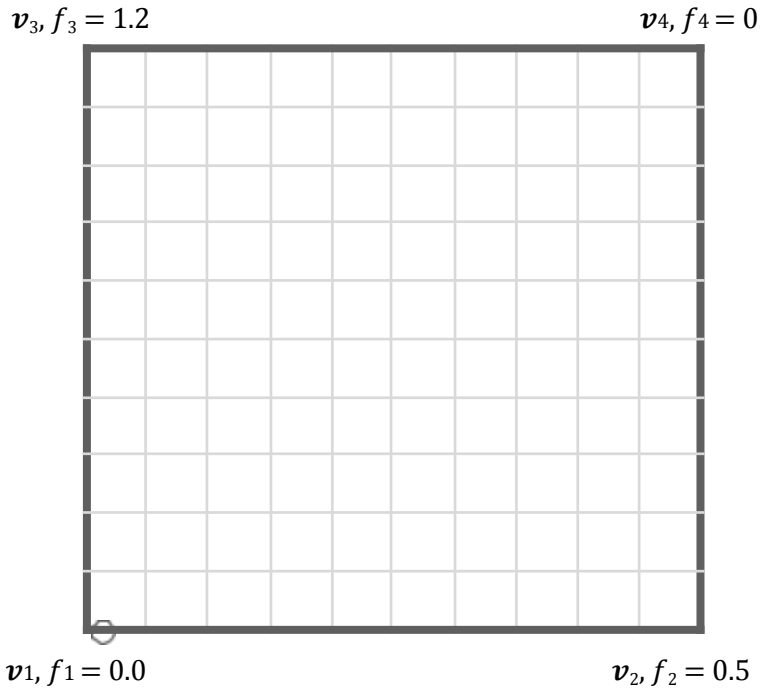
Put x into general eqn. f(2) = w1* $\varphi(||1\text{-}x||)+$ w2* $\varphi(||3\text{-}x||)+$....

Data
Interpolation

a) For a tetrahedron with vertices A = (0,0,0), B = (1,0,0), C = (0,1,0), D = (0,0,1) and the corresponding scalar values f_A, f_B, f_C, f_D, the linear interpolation function $f(x, y, z) = 1 - x + 2 \cdot y - 2 \cdot z$ is given.
Compute the concrete scalar values at the four vertices.
Just put given values into function Fa= 1, Fb =0, Fc= 3, Fd= -1

b) For the tetrahedron given in the previous assignment, assume that the scalar values inside the tetrahedron are interpolated via barycentric interpolation. Compute the gradient of the interpolated scalar field at the points P = (0.5, 0.25, 0.5) and Q = (0, 0, 0).
derive formula above with respect to x,y,z. Get the gerived vector and put P and Q in

c) Given is the following quadrilateral cell with its four vertices $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_4$ and the corresponding scalar values f_1, f_2, f_3, f_4. A grid is shown for orientation purpose, i.e., it does not affect the interpolation.



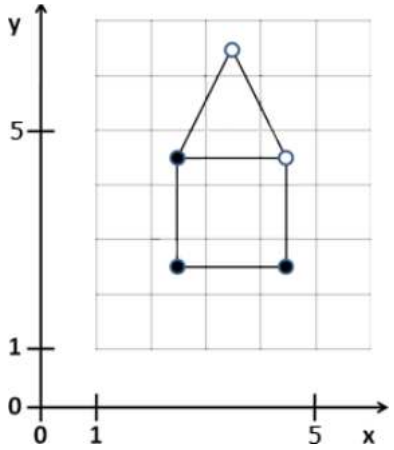
Draw the iso-contour (bold) for the iso-value 0.4 into the same illustration, i.e., all points in the interior of the cell which have a value equal to 0.4. Bi-linear interpolation is assumed for interpolation.

Find each line with the given value and mark the point. Then connect them along either x axis or y axis. (with curve)

d) In the figure below, a quadrilateral cell is shown. Scalar values are given at the cell corners and bi-linear interpolation of these values is used to reconstruct scalar values across the cell. Is it possible that at the four edge points marked by dots the same scalar value is reconstructed? Explain your answer.

When we draw isolines we cannot draw the asymptote so they are not same values with same isolines

e) In the figure below, a 2D Cartesian grid is shown. It has a constant spacing of 1 between adjacent grid points along either axis. The origin of this grid is at (1, 1). A second grid is shown, which consists of one triangle with vertices at (2.5, 4.5), (3.5, 6.5), (4.5, 4.5) and one quadrilateral with vertices (2.5, 2.5), (2.5, 4.5), (4.5, 4.5), (4.5, 2.5). At the vertices of the second grid, scalar values are given. These values are equal to 1 at vertices marked with a filled circle and 0 at vertices marked with a non-filled circle.



Compute the barycentric coordinates of the points (3.5, 4.5) and (3.5, 2.5) with respect to the triangle.

For the quadrilateral, determine the coefficients *a*, *b*, *c*, *d* of the bi-linear interpolation function $f(x, y) = a + b \cdot x + c \cdot y + d \cdot x \cdot y$ which interpolates the scalar values at the vertices of the quadrilateral.

Barycentrix coordinates = one with alphas. (split into areas) and area values are alphas.

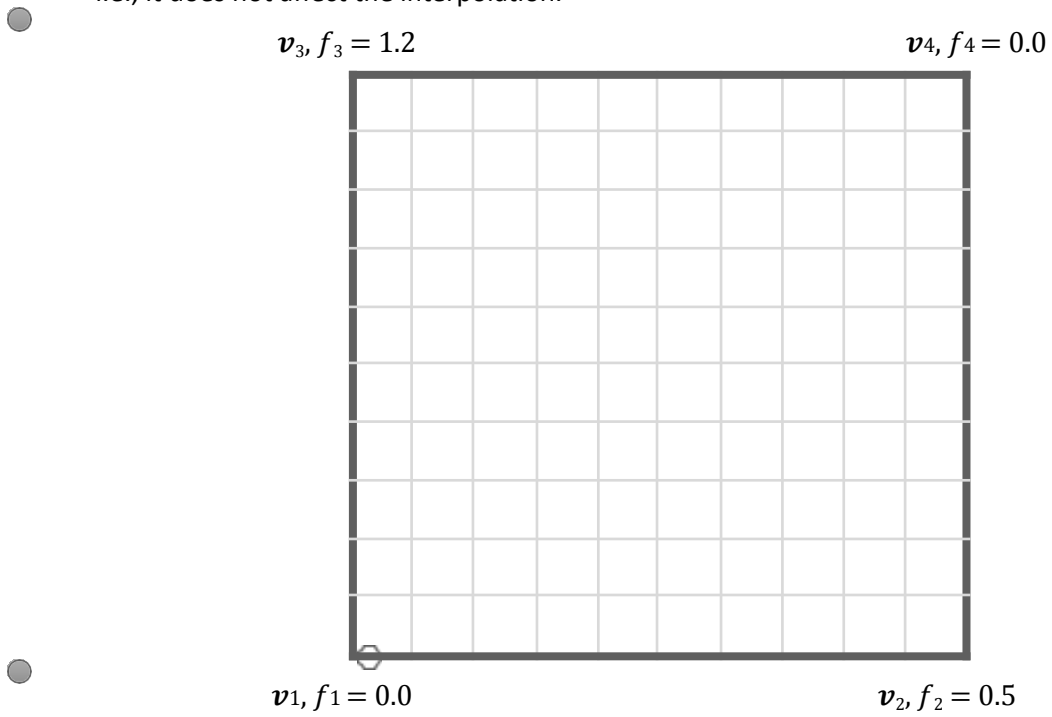
If the point is out of triangle= 2.5*α1+3.5*α2+4.5α3 = 3.5
4.5*α1+6.5*α2+4.5*α3 = 2.5
α1+α2+α3= 1

Bi-linear interpolation part =
1 = a +2.5b +2.5c +2.5*2.5 d
1 = a +4.5b +2.5c +2.5*4.5 d
1 = a +2.5b +4.5c +2.5*4.5 d
0 = a +4.5b +4.5c +4.5*4.5 d

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Isolines & Isosurfaces

a) Given is the following quadrilateral cell with its four vertices $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_4$ and the corresponding scalar values f_1, f_2, f_3, f_4. A grid is shown for orientation purpose, i.e., it does not affect the interpolation.



Draw the approximated iso-lines (dashed) for the iso-value 0.4 into the illustration using the *marching squares algorithm*. Use the *mid-point decider* for ambiguous cases.

Also, draw the true iso-contour (bold) for the iso-value 0.4 into the same illustration, i.e., all points in the interior of the cell which have a value equal to 0.4. Bi-linear interpolation is assumed for interpolation.

We first assign + and - based on isovalue to the points. Then midpoint decider (1.2+0.5+0.0+0.0/4) it says center is +. Result is different then original isolines. For the solving issue use asymptotic decider.

b) Name the three components of the Phong illumination model.
Ambient reflection, diffuse reflection, specular reflection

c) How can a perfect mirror be simulated via the Phong illumination model?
If n goes infinity (from specular formula) the perfect mirror happens

d) Let L be an incoming light ray and M a diffusely reflecting material. Complement the illustration on the right in Figure 7 according to the illustration on the left. Hint: Take into account the direction and strength of the reflection.

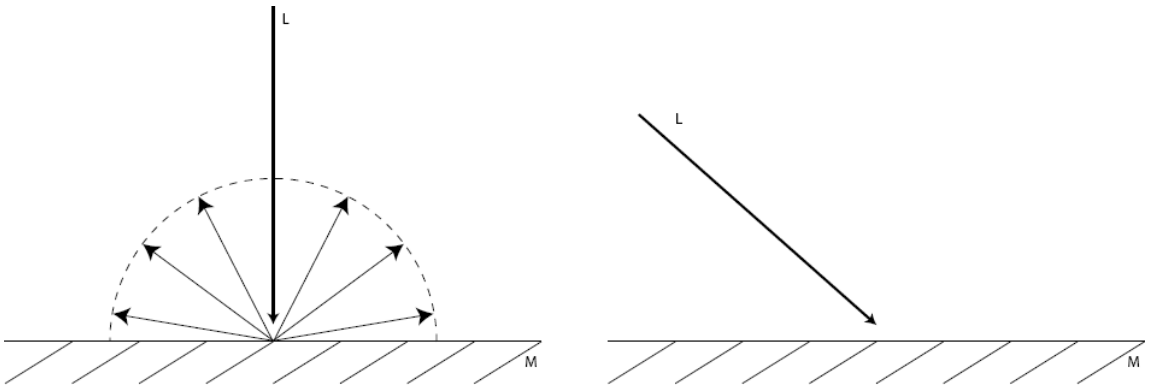
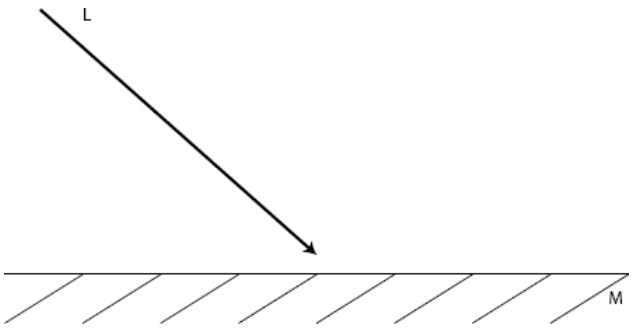


Figure 7

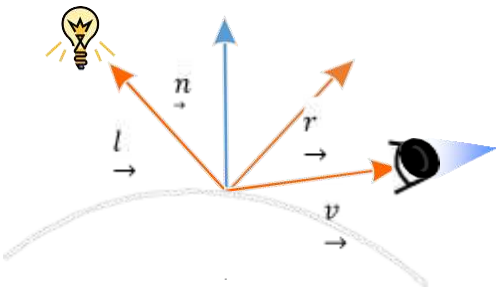
Since L = N we would have the largest diffuse reflection for the figure on the left. For the right we have an angle between L and N cos is smaller. We still have reflection but smaller.

e) Let L be an incoming light ray and M a specular reflecting material. The viewer is positioned at the light source and is looking along L. Complement Figure 8 according to Figure 7.



Reflection doesnt reach the viewer since r does not hit viewer

f) Below is an illustration of the Phong illumination model. All indicated vectors are normalized (i.e., their length is one). Answer whether the following statements are true or false:



$\vec{n}$ … Normal vector of the surface

$\vec{l}$ … Vector pointing to the light source

$\vec{r}$ … Reflected light ray

$\vec{v}$ … Vector pointing to the viewer (view vector)

The specular reflection is based on the scalar/dot product of $\vec{\boldsymbol{n}}$ and $\vec{\boldsymbol{v}}$	FALSE
The diffuse reflection is based on the scalar/dot product of $\vec{\boldsymbol{n}}$ and $\vec{\boldsymbol{l}}$	TRUE
The ambient light is based on the scalar/dot product of $\vec{\boldsymbol{l}}$ and $\vec{\boldsymbol{r}}$	FALSE
The specular reflection is independent of the view vector $\vec{\boldsymbol{v}}$	FALSE

g) Compute the specular reflection at a surface point $P = (3, 1, -2)^T$ using the Phong lighting model. The normal at the point is $\vec{n} = (0,1,0)^T$. Moreover, the specular reflection coefficient of the surface is $k_s = 0.5$ and the specular exponent is $n = 2$. The position of a point light source is $L_{pos} = (1, 3, -1)^T$ and the camera position is at $E_{pos} = (5, 2, -2)^T$. Both the point light source and the surface color are white with RGB-values $(1, 1,1)$.

Hint: You can compute the reflected light ray as $\vec{r} = 2(\vec{n} \cdot \vec{l})\vec{n} - \vec{l}$, where $\vec{n}$ and $\vec{l}$ need to be normalized.

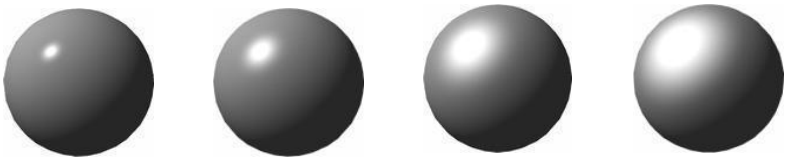
C = ks*Oc*(r.v)^n we need r and n

l_vector = L_pos - P => for r use the formula

v_vector = E_pos - P then put all into the formula.

NOTE : ALL VECTORS MUST BE UNIT LENGTH, NORMALIZEE!!!!

h) In the four images below, a sphere is rendered using the Phong illumination model. Four different specular exponents (2, 4, 16, and 64) were used to create the specular reflection. Write the specular exponent that was used to create the rendering below each image.



64 16 4 2

Larger values cause to smaller specular area

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Volume Rendering

a) Name an algorithm commonly used in *indirect volume visualization*.

Marching cubes

Why is it considered to be "indirect" compared to "direct" volume rendering?

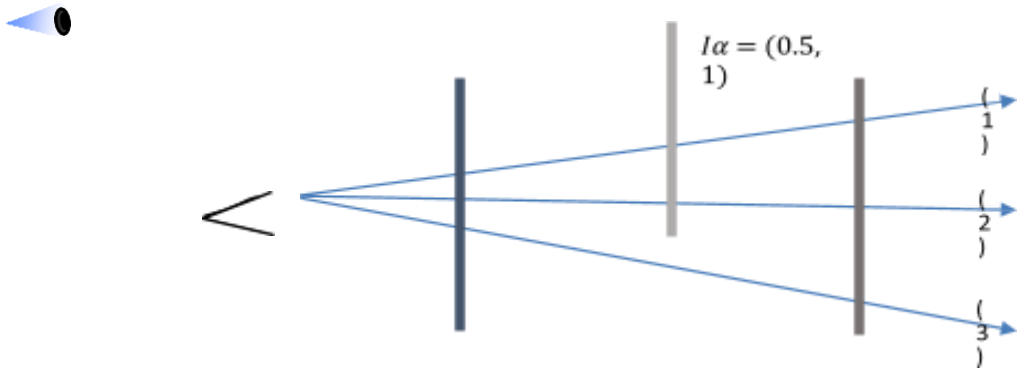
For the indirect volume rendering we first have to construct intermediate representation of data in this example its surface representation from original data.

For the direct example: Ray casting

b) Which optical/visual properties are assigned by a transfer function in volume rendering?

Color and opacity assigned by T func.

c) A scene consisting of 3 objects (vertical lines) with different intensities (I) and opacities (α) is shown. The α-value (second component) represents the object's opacity, where 0 = ´completely transparent´ and 1 = ´completely opaque´. The 3 objects are ordered as shown. For the three rays starting at the viewpoint, determine the intensity that is seen along these rays using 1) front-to-back **α-compositing** for the upper ray, 2) the compositing scheme **Average** for the middle ray, and 3) the compositing scheme **Maximum** for the lower ray. Specify exactly how the intensities are combined in either case.



Front to back alpha =

$Cout = Cin + (1\text{-}alpha\_in)\ alpha * C$

$Alpha\_out = alpha\_in + (1\text{-}\ alpha\_in)alpha$ DO this for each surface

$Average = (1+0.5+0.75)/3$

$Max = \text{argmax}(1,0.75) = 1$

d) Light/color has been emitted at a point along the viewing ray. How is it diminished (due to absorption) in a homogeneous, semi-transparent medium? Draw a typical curve.

It starts to emitted for a point at start. then diminishes in time until the 0

e) How do you get values along the viewing ray (from volume data)?

Data in the volumetric data set is giving in grid then we ray cast then interpolate for the points that are not the data points directly for the interpolation scheme example trilinear , or some higher order interpolations

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

f) Which compositing schemes do you know (for combining values along the viewing ray)?
bonus alpha compositing,
first hit= stop as soon as a value is larger than threshold. similar to marching cubes but better because it operates per pixel
Maximum intensity projection = The maximum value for each viewing ray is shown
Averaging = Values along the viewing ray is averaged

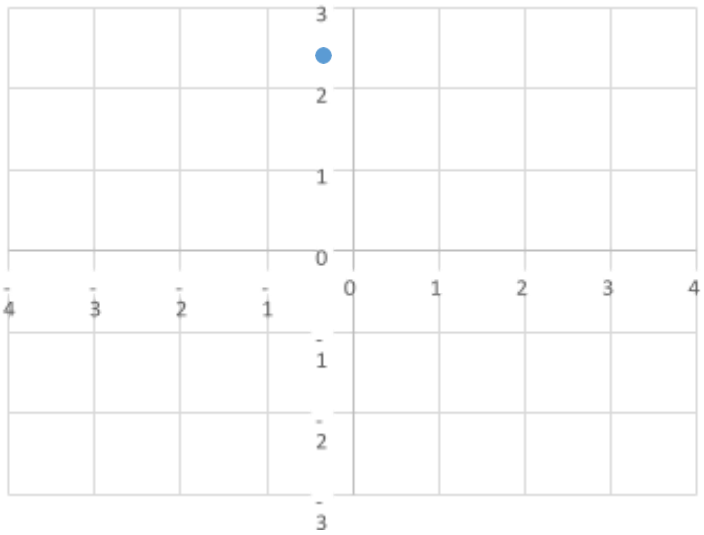
Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Vector Field Visualization (part I)

a) Calculate the Jacobian matrix **J** of the 3D vector field $\boldsymbol{v}(x, y, z) = (y^2 - 1, x - y, xz)^T$.

b) Calculate the Jacobian matrix, divergence and curl of the 3D vector field $v(x, y, z) = (\cos(xy)\ , \sin(x)\ , z)^T$**.**

c) Given a 2D vector field $\boldsymbol{v}(x, y) = (y, -x)^T$. Compute the next point $\boldsymbol{x}_1$ of a stream line with seed point $\boldsymbol{x}_0 = (0, 2)^T$ using the *midpoint integration* method (also known as Runge-Kutta of 2nd order) with a step size $\Delta t = 1$. Draw the resulting point and the used vector(s) in the illustration below (don't forget to annotate them). Specify exactly how the point and vectors are calculated.



d) In Figure 6, a vector field at four different times is shown. The grid vertex marked by a dot is selected as initial position for particles that are seeded into the flow. Assume that a particle can only move diagonally, vertically, or horizontally (depending on the vector at the particles current position), and that it always moves from the current grid point to the next grid point in the respective direction.

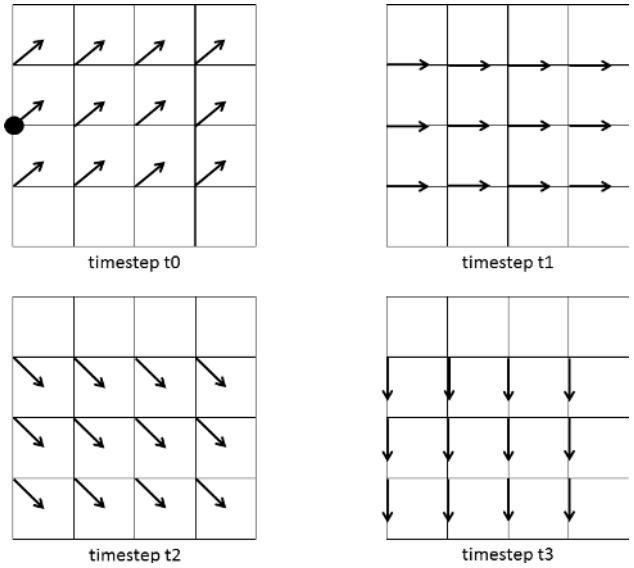
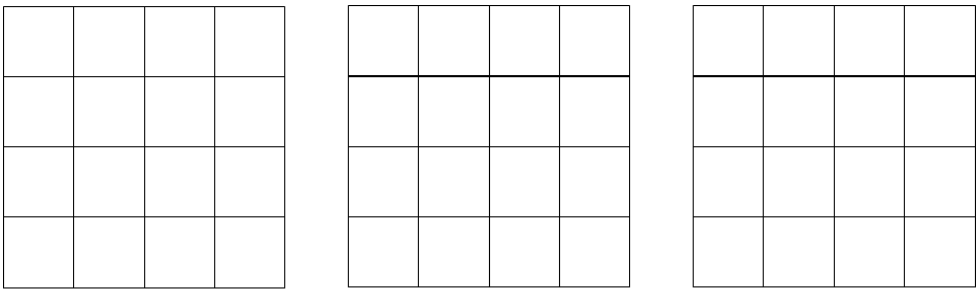


Figure 6

Illustrate in the grids given below a) the pathline of a particle released into the flow at the marked position at time t0, b) the streakline of particles released into the flow at the marked position at time t0, and c) the streamline of a particle released into the flow at the marked position in the fixed vector field at time t1.



e) Which characteristic line approaches do you know for *unsteady/time-varying* flow? What happens if you apply them to steady flows?

Pathline and Streakline. They become the same with streamlines.

f) Answer whether the following statements are true or false.

The Jacobian matrix at a point in a constant 3D vector field has non-zero elements on the main diagonal.	F
If the Jacobian matrix at every point in a 3D vector field is the identity matrix, then the vector field is divergence free.	F
The divergence at every point in a 3D vector field is a scalar value.	T
Streamlines in a steady 3D vector field never cross.	T
Path lines in a time-varying 2D vector field never cross.	F

g) Give two examples for direct flow visualization.

Arrows, color coding

h) What challenges does arrow-based direct flow visualization have?

Ambiguity of arrow stance, occlusion.

i) Give two examples for geometric (integration-based) flow visualization. How do these techniques relate to direct flow visualization?

Streamlines, surfaces. In geometric flow visualization we use streamlines, pathlines where we consider movements of particles whereas in direct flow visualization we directly show the flow.

j) What is Runge-Kutta integration (2nd/4th order)? What's the advantage over Euler integration?

It is a 2nd order ODE integration. More accurate than Euler integration.

k) Characteristic lines are tangential to the flow. What does that mean?

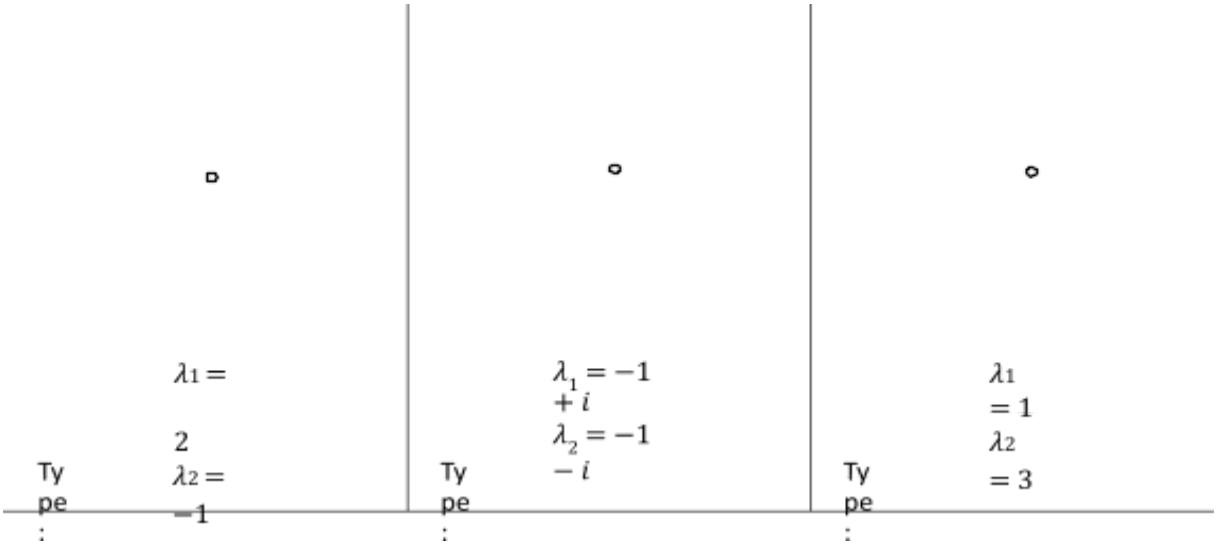
The char lines, such as streamlines and pathlines, are parallel to the vector field of the flow.

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Vector Field Visualization (part II)

a) A critical point in a 2D vector field $\boldsymbol{v}(x, y)$: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a point (x, y) where $\boldsymbol{v}(x, y) = (0,0)^T$. How many critical points does the vector field $\boldsymbol{v}(x, y) = (y^2 - 1, x - y)^T$ have, and where are these points located?

b) Below are three examples of critical points together with the eigenvalues λ_1, λ_2 of the respective Jacobi matrices. Classify each critical point according to the eigenvalues and sketch the typical flow around it.



c) Answer whether the following statements are true or false.

LIC is a local method for visualizing a vector field.	False
The larger the extent of the convolution kernel used in LIC, the lower is the correlation between adjacent intensity values along a streamline.	False
LIC images show high correlation between the intensity values at adjacent streamlines.	False
LIC is restricted to 2D vector fields.	False
The convolution kernel used in LIC must be symmetric.	False

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Visualization Mapping

a) What is the pop-out effect / pre-attentive processing? How can it be used?

Automatic and parallel detection of basic features in visual information. It allows to preattentively (instantly) find the outlier (distinct data) and the speed is independent of the distractor count. Example: one may easily detect a red square within multiple blue squares.

b) Sort the following visual channels according to how accurately humans can compare them starting with the highest accuracy: *2D area – length – curvature – angle/slope*

length - angle/slope - 2D area - curvature

c) What is the difference between separable and integral visual channels?

Separable: Distinguishable channels on the visualization.

Integral: Undistinguishable channels on the visualization.

d) Name an example for fully separable / integral visual channels.

Fully Separable: Color + Position

Integral: Red and green dots (channels) among many colored other dots.

e) Which visual channel(s) can be used in a bar chart? For what types of data?

Length, Color, Size

f) From a perceptual point of view, what works better: Bar charts or pie charts? Why?

Bar charts can read length better and humans understand length better than angle as in pie charts.

g) How do Parallel sets work? What kind of data can be shown?

Quantitative data wrt. multiple categorical attributes.

h) How does the ThemeRiver work? Which visual channels are used for which type(s) of data?

Occurrence per topic/category mapped to width of river band. Thematic changes in documents, rearranging bands.

i) How do Horizon Graphs work? How can you read out values at a position?

Split vertically into layered bands and collapse color bands to show values in less vertical space. One may read the values by determining the band number first and then getting the value from the last band and combine.

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Visualization Mapping

a) Which visual channels can be used in a scatterplot besides position?

Size, color etc.

b) How does a scatterplot matrix work? How can you see correlations?

It's a half matrix where each cell represents a relation (shows correlation) between two distinct features.

c) How does linking and brushing work?

Linking and brushing allows interaction between different visualizations. By interacting with one visualization one may affect another one to get a better understanding.

d) Draw a visualization of the following data using *parallel coordinates*.

	Age	Height	Shoe size	Number of children
User 1	36	195	46	0
User 2	42	187	44	3
User 3	25	172	40	1
User 4	32	168	39	0
User 5	34	160	37	2
User 6	55	164	38	2
User 7	40	176	41	1

e) What does it mean when the lines between two axes in a parallel coordinates visualization meet in a point?

If the lines between two axes in a parallel coordinates visualization meet in a point, then the data points in the scatterplot (where the axes are the axes from the parallel coordinates) are located on a line and negatively correlated.

f) What are glyphs? For which type of data are they typically used?

Glyphs are data objects that represent particular data points with multiple features (arrows). They are typically usef for multivariate data.

g) How do star glyphs/stick figures work? How do they encode the data?

Star Glyph: A star composed of equally spaced spikes originating from center where length of the spikes represents value of respective attribute. End of rays connected by line.

Stick Figure: 2D figure with limbs where data is encoded by length, thickness and angle between lines. Allows better recognition of patterns (texture patterns).

h) What are the advantages/disadvantages of a rainbow color map?

Adv: Good for categorical data

Disadv: Perceptually nonlinear, perceptually not ordered

i) What does it mean, when a visual channel (e.g., color) is *perceptually linear/ordered*?

Perceptually Linear: The rate of change in a constant interval around the color mapping is the same (number of different perceivable colors is different on different locations)

Perceptually Ordered: Intuitive order in the color mapping (like luminance mapping)

j) What are the characteristics of a sequential / diverging color map?

Sequential color map is a perceptually ordered map where one end is a color and the other is another color and the middle is interpolating in between those colors. Diverging color map uses an extra color in between the end ones (usually white) and interpolates one end through white and other through white.

Disclaimer: *Below you find some example questions, which should help you prepare for the exam. However, note that the actual questions at the exam can be very different and can cover all material presented in the lecture!*

Visual Analysis of Scientific Data

a) What is the main goal of visual exploration?

Visualization exploration helps one find the unknown/unexpected about the data and generate new hypotheses over it.

b) Which three major areas/concepts are combined in visual analysis/analytics?

Visualization-Interaction-Data Analytics

c) Give examples how multivariate data can be encoded in a spatial context?

Glyphs, scatterplots, parallel coordinates

d) What are challenges when fusing multi-modal data stemming from different data sources?

Coregistration, segmentation etc.

e) Visual data fusion intermixes data in a single visualization using a common frame of reference. Give at least two general approaches.

Layering Techniques (glyphs, color etc.), Multi-volume rendering (coregistration, segmentation)

f) What are three general approaches for comparative visualization (according to the taxonomy of Gleicher et al. 2011)?

Side-by-side comparison(Juxtaposition)

Overlay in same coordinate system(superposition)

Explicit encoding of differences/correlations

g) What is focus+context visualization? Explain the general approach. How is it different from an overview+detail visualization?

The idea is to combine both the important area -the focus- and the context in a single visualization. Whereas the overview+detail visualization uses spatially separate overview / detail and the user has to switch attention between representations.

h) Give at least three examples of visual channels (graphical resources) that can be used for focus-context discrimination.

Color, Style, Blurring, Opacity

i) Give two examples for focus+context visualization techniques which use spatial distortion.

Perspective Wall, Table lens, Fisheye views

j) What is the main idea in clustering? Is clustering a supervised or unsupervised method?

To group close (distance-wise) data into bundles and treat them as similar. Unsupervised

k) What is the main idea in dimensionality reduction? Name one example method? How does it work?

To reduce the number of features to obtain a visualization that makes sense. The motivation is to remove redundant features. Principal Component Analysis (PCA) finds the directions with high variance and use those only.

l) Principal component analysis transforms data from a cartesian coordinate system into another coordinate system. Why is it then still considered a dimensionality reduction method?

Because after changing into another coordinate system, one of the axis would be removed since it has significantly lower variance then the other axis