Computational Vision
U. Minn. Psy 5036
Daniel Kersten
Lecture 7: Image Modeling, Linear Systems

■ Initialize standard library files:

In[1]:=
```
Off[General::spell1];
```

Outline

Last time

■ Bayesian decision theory applied to perception

The task is important: Some causes of image features are important to estimate (primary variables, S_{prim}), and some are not S_{sec}.

Use the joint probability to characterize the "universe" of possibilities:

$$p\,(S_{prim}, S_{sec}, I) \tag{1}$$

(Directed) Graphs provide a way of modeling what random variables influences other random variables, and correspondingly how to decompose the joint into a simpler set of factors (conditional probabilities).

Through conditioning on what is known, and integrating out what we don't care to estimate, we arrive at the posterior:

$$p\,(S_{prim} \mid I) \tag{2}$$

And by Bayes' theorem the posterior probability is given by:

$$p[S_{prim} \mid I] = \frac{p[I \mid S_{prim}]\, p[S_{prim}]}{p[I]}$$

$p[I \mid S_{prim}]$ is the likelihood, and is determined by the generative model
$p[S_{prim}]$ is the prior model
$p[I]$ is fixed

The task determines how to use the posterior.

Picking the most probable value of S_{prim} (MAP) results in the fewest errors on average

■ Counter-intuitive consequences:

Inference: Fruit classification example

Pick most probable color--Answer "red"

Pick most probable fruit--Answer "apple"

Pick most probable fruit AND color--Answer "red tomato", ***not*** "red apple"

Moral of the story: Optimal inference depends on the precise definition of the task

■ Generalization to degrees of desired precision leads to Bayesian *decision* theory

Slant estimation example

Today

Generative modeling.

Introduction to modeling image structure:

image-based vs. scene-based

Linear systems (linear image-intensity based modeling)

Optical Resolution: Diffraction limits to spatial resolution

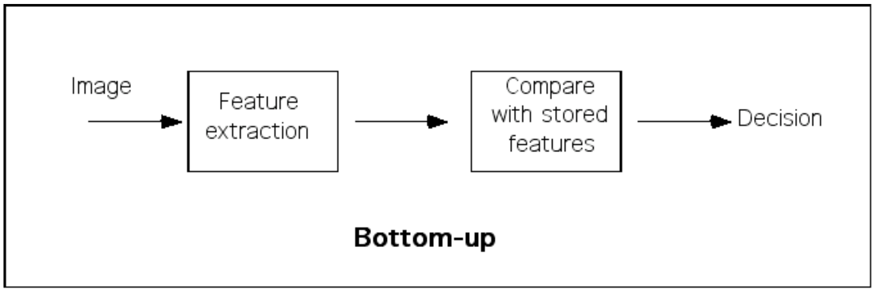
Point spread function, convolutions and blur

Overview of image modeling

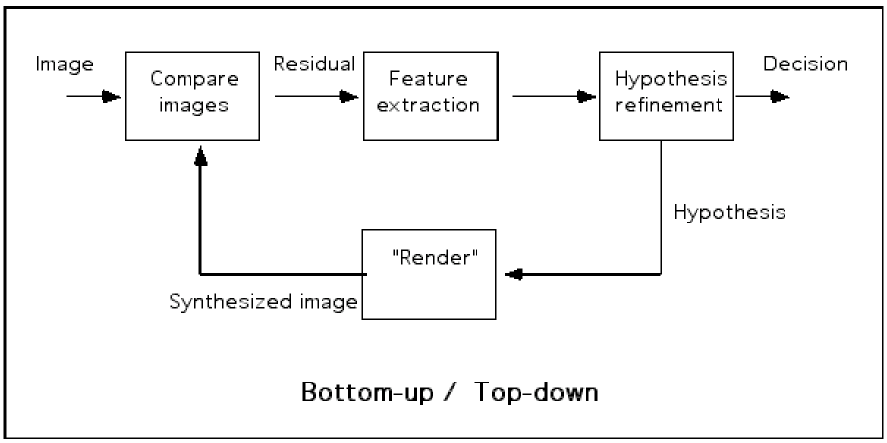
Generative models for images: rationale

■ Characterize the knowledge required for inference

Feedforward procedures:



Pattern theory perspective: "analysis by synthesis" (Grenander, Mumford)



■ Easier to characterize information flow: Mapping is is many-to-one

Can be used to model the likelihood of scene descriptions, p(I|S) (i.e. the probability of an image description given a scene description)

■ Provides tools to specify the independent variables in psychophysics, and vision models

■ Two basic concepts: Photometric (intensity, color) & geometric variation

■ Two more basic concepts: Local and global representations

E.g. an edge can be locally represented in terms of the contrast and orientation at a point of an image. But a long edge (or contour) can also be represented by function with a small number of parameters (e.g. slope and intercept of straight line, or if curved, as a polynomial curve).

An image can be represented in terms of a list of intensities at each location (local), or as we will see shortly, as a linear combination of global patterns (sine-wave gratings or other "basis functions").

- **Two fundamental classes of image modeling:**

  1) image- (or appearance) based

  2) scene based

As an example, ***image-based modeling*** tools are provided in software packages like:

Adobe Photoshop or GIMP, or Adobe Illustrator

For an introduction to image manipulation using *Mathematica*, see:

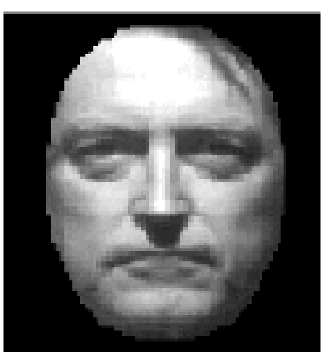
http://library.wolfram.com/infocenter/Articles/1906/

Scene-based modeling tools are provided in 3D graphics packages like: Maya, 3DS Studio Max, or software development packages like OpenGL.

Image-based modeling

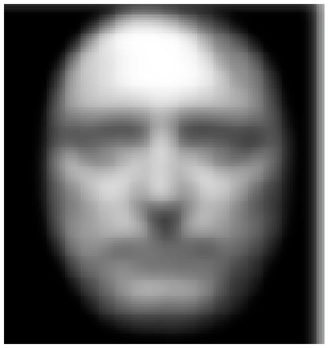
Pattern theory perspective (Grenander, Mumford)

- **Original**

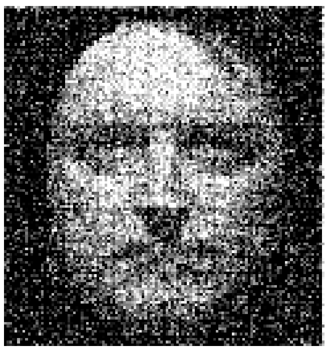


- **Superposition: blur, additive noise, transparency**

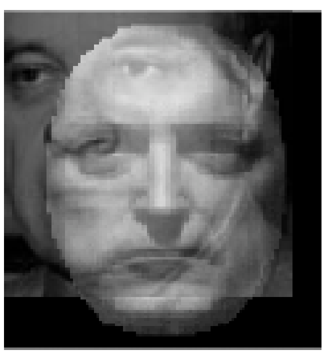
photometric



added noise

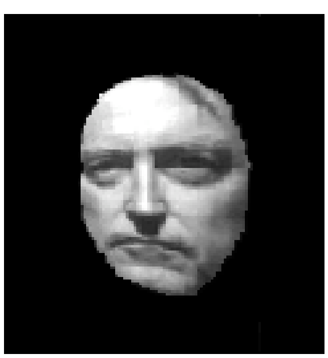


transparency: sum of two images



■ Domain warping (morphs)

geometric



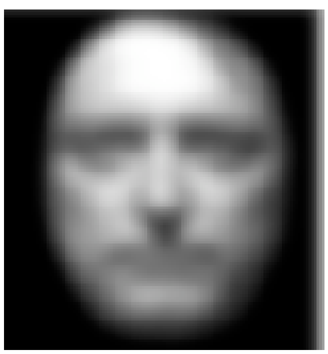
■ Domain interruption (occlusion)

geometric

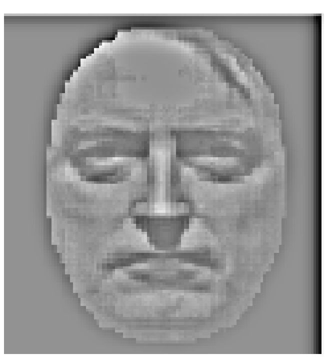
■ Processes occur over multiple scales

photometric and geometric

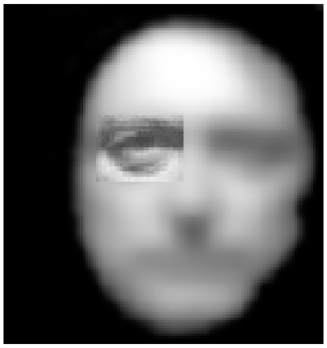
Low-pass



High-pass

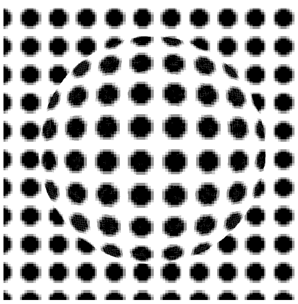


Large scale structures, and small scale structures

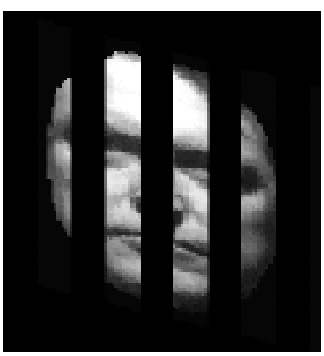


We'll see later how local edges can be seen as a kind of cooperation over multiple scales

Texture (regularities at small scales) can interact with regularities at larger scales (surface shape) to produce patterns like this:



■ Combinations



We need various combinations of transformations to account for the kind of variations that we'd see in other forms of the same face. For example, a 3D rotation produces both domain warping and occlusion. A beard is another form of domain interruption. Expressions are a combination of warping, occlusion, and dis-occlusion.

David Mumford

Ulf Grenander

(From: http://www.dam.brown.edu/ptg/ and http://www.dam.brown.edu/people/mumford/)

Four classes of image-based models

■ Linear intensity-based

(Photometric)

Basic idea is to represent an image $\vec{I}$ in terms of linear combinations or sums or other images $\vec{I_i}$:

$$\vec{I} = m_1 * \vec{I_1} + m_2 * \vec{I_2} + m_3 * \vec{I_3} + \ldots$$

where m_i's are scalars. You can think of each image as a point in an N-dimensional space, where N is the number of pixels, and the dimensions of the point correspond to the pixel intensities.

Linear models get interesting when we begin to talk about *linear transformations*. The powerful idea behind linear systems is that a transformed image can be represented by the sum of its transformed "sub-images" or "basis images". This is the principle of superposition that will crop up repeatedly later.

Basis sets $\{\vec{I_i}\}$:

Complete:

Means you have enough basis images that you can construct any image from a weighted sum of the basis images

$$\vec{I} = m_1 * \vec{I_1} + m_2 * \vec{I_2} + m_3 * \vec{I_3} + \ldots$$

Will lead us to linear shift-invariant systems--fourier synthesis

application: optics of the eye

efficient representation of natural images -- wavelets (sometimes "over-complete")

application: V1 cortical neuron representation, edge detection

V1, MT motion-selective neurons

Incomplete:

lossy compression, principal components analysis (PCA)

characterizing image variation in specialized domains

application: illumination variation for fixed views of an object

■ Non-linear intensity-based

(Photometric)

Point non-linearities (e.g. gamma correction for computer display)

Polynomial expansions

Contrast normalization ()

application: V1 cortical cell representation

■ Linear geometry-based

Affine:

$\{x_i, y_i\} \to \{x'_i, y'_i\}$, $\{x'_i, y'_i\} = \mathbf{M}.\{x_i, y_i\}$, where **M** is a 2x2 matrix

rigid translations, rotations, scale and shear

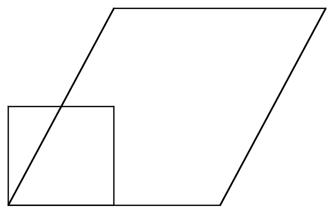
(e.g. in 2D, all the transformations that result from applying the 2x2 M matrix

to each of the coordinates of points $\{x_i, y_i$ in an image)

Application: viewpoint variation

You can play with the values in the matrix M below to transform the rectangle x into new shapes x2:

In[2]:=
```
x = {{0, 0}, {0, 1}, {1, 1}, {1, 0}, {0, 0}};
M = {{1, 2}, {2, 0}};
x2 = Dot[M, #] & /@ {{0, 0}, {0, 1}, {1, 1}, {1, 0}, {0, 0}};
Show[Graphics[Line[x]], Graphics[Line[x2]]];
```



■ Non-linear geometry-based

Warps (Morphs)

$\{x_i, y_i\} \to \{x'_i, y'_i\}$, where the transformation is no longer a single matrix, and each point can get mapped in arbitrary ways. By warp or morph, we usually mean a smooth mapping.

Application:within-category variation for an object,or objects

finding the "average" face

e.g. Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgments of female attractiveness. Nature, 368, 239-242.

Scene-based modeling: 3D Computer graphics models

■ Scenes

Variable classes : extended surfaces, solid objects, diffuse "stuff"/particles, lights, and camera (eye).

■ Objects & surfaces

Shape

Articulations

Material & texture

■ Illumination

Points and extended

Ray-tracing

Radiosity

■ Camera

Projection geometry:

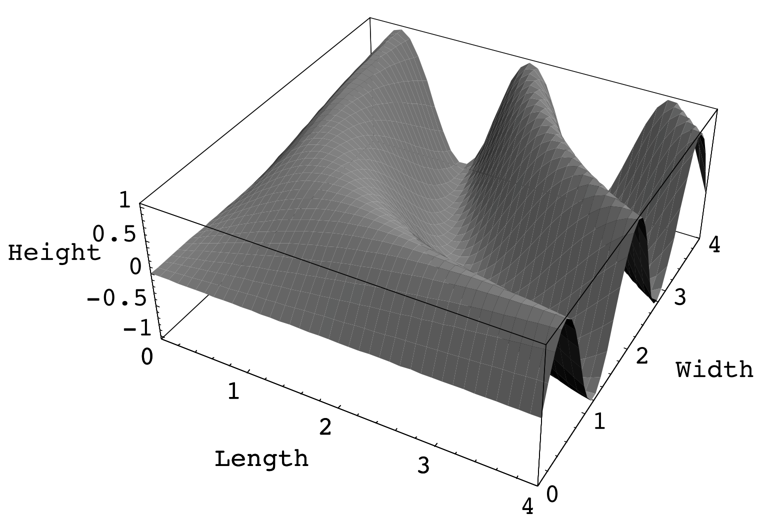
perspective, orthographic

focus

■ *Mathematica* provides some primitive 3D modeling tools

Mathematica provides tools to model surfaces and objects in 3D. Here is an example of a surface plot.

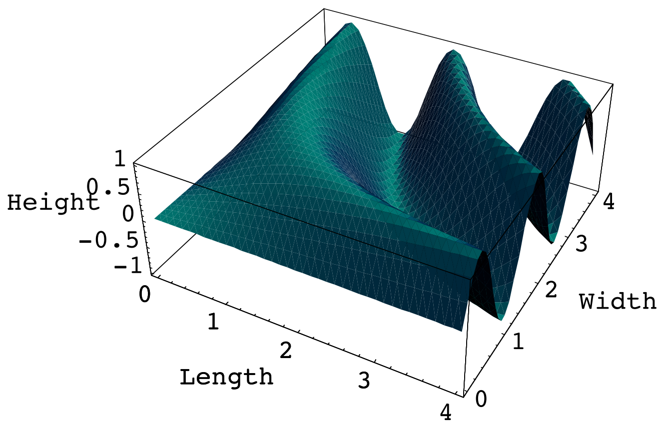
In[6]:=
```
gsin = Plot3D[Sin[x y], {x, 0, 4}, {y, 0, 4}, PlotPoints → 40, Mesh → False,
   AxesLabel → {"Length", "Width", "Height"}, AmbientLight → GrayLevel[0],
   LightSources → {{{1., 1., 1.}, RGBColor[1 / 4, 1 / 4, 1 / 4]},
     {{0., 1., 1.}, RGBColor[1 / 4, 1 / 4, 1 / 4]}}, ViewPoint → {1.3, -2.4, 2.}];
```



Try playing with the camera and lighting parameters.

Using the SurfaceColor directive, you can also specify some simple surface properties, such as the amount and degree of specularities:

In[7]:=
```
gsinShiny = Insert[Graphics3D[gsin],
   SurfaceColor[RGBColor[0, 0.3, 0.6], RGBColor[0, 0.7, 0.3], 3], {1, 1}];
Show[gsinShiny];
```



■ Preview of general inference issues for vision:

Edge ambiguity

Projection invariants--for inference

Pros and cons

Pros of image-based modeling:

closer in some sense to the image data representations that visual perceptual inference deals with.

The so-called "proximal" stimulus corresponds to image information or features directly tied to the image.

Cons of image-based modeling:

often far from the kinds of representations useful for representing scenes, and planning actions

scene-based modeling specifies the properties of the "distal stimulus", like depth, shape, material.

Moral: Good to be familiar with both kinds of generative image modeling

In vision, by the most common type of modeling begins with linear-intensity based models, which we treat next.

Linear systems & images: Motivation

Often when you look through a reflecting glass, you can see one image superimposed transparently on another. Superposition is one of the image transformations of pattern theory that crops up in many different circumstances other than transparency, from early vision to high-level vision. The idea is that a particular image intensity pattern can be expressed as a weighted sum of other images. Being able to combine weighted sums of images can be useful in domains quite different from reflections in a window.

Here is a simple thing you could do by combining images. You could calculate the "average" face. This has been done, and the average face is beautiful, if tending to be a bit fuzzy. This raises the question: when is it appropriate to model image variation in terms of sums? It turns out that taking the average of a set of images, such as faces, is tricky because there is not only photometric variation, but also geometric variation. I.e. the faces need to be put in register, and geometric registration is a non-trivial problem, and one that has been studied in many domains such as biometrics.

Rather than adding images up, suppose we have an image, and want to decompose it as a sum of other sub-images. A natural question to ask is: what should the other "sub" images be? The answer, will depend on the task. For example, below we will learn to model how the optics of the eye transform an input image to retinal image. A natural "sub-image" is the image of a point of light. These sub-images are spatially quite localized.

But we will see there are advantages to using sub-images that are not spatially localized, and are global. These sub-images are sinusoidal gratings.

Later, when we study early neural mechanisms in vision, we will see that models of image coding can be understood as a decomposition of the visual image into image-like sub-components that make a local/global trade-off. We will see evidence for neural basis images that look like two-dimensional "wavelets" that if summed up (with appropriate weights) would give us back the orginal image.

As another example, it is possible to model the space of human faces in terms of a linear sum of "basis faces" whose number is carefully chosen to be as small as possible.

Much later in this course, we will see a less obvious consequence of superposition of light in a method to deal with illumination variation in object recognition. A good approximation to the space of all images of an object can be obtained by summing a small set of appropriate sub-images.

In order to understand how to use superposition, we need to understand the principles of linear systems analysis. And in order to introduce linear systems, we return to the question we asked in lecture 2:

What are the physical limits to vision?

And in particular:

What are the physical/optical limits to spatial resolution?

Optical Resolution

Factors limiting spatial resolution

As we have noted earlier, the physical limits to visual light discrimination and resolution are a consquence of the quantum and wave nature of light. We have seen how the particle aspect of light limits resolution of intensity differences. The wave nature of light limits resolution of spatial differences.

As two identical small points of light get closer and closer, we eventually reach a point at which we can no longer tell whether we are looking at one or two spots of light. What are the limits to our ability to resolve spatial separation?

Physical/optical

Diffraction

(Recall Paul Dirac's famous statement about photons:

"*Each photon interferes only with itself. Interference between photons never occurs*.")

Optical aberrations

spherical, chromatic, focus, astigmatism

Biological

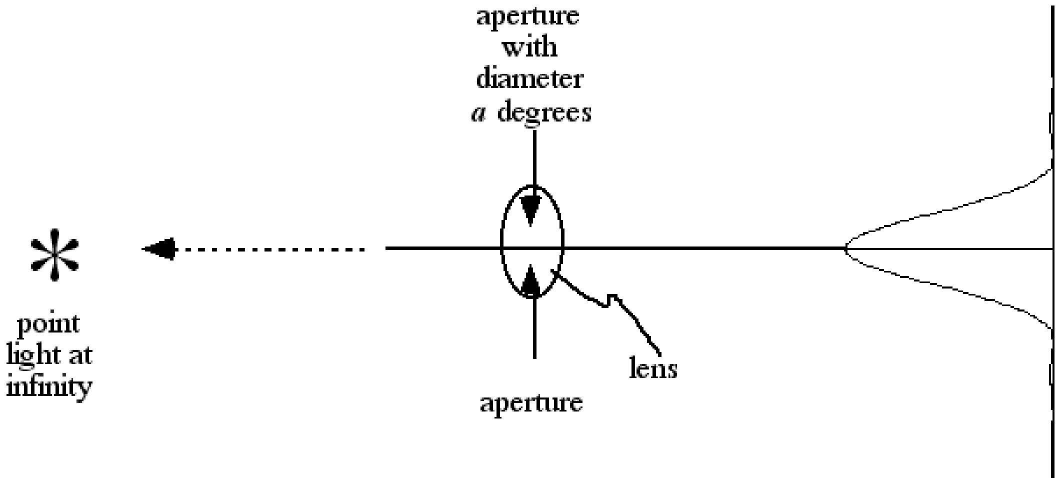
Sampling--retinal receptor array is discrete

Neural convergence (may lead to loss of resolution information)

Diffraction

As with limits to intensity discrimination, our understanding of the limits to spatial resolution begins with the physics of light. Diffraction effects are caused by the constructive and destructive interference of many wavefronts. (The word interference is a result of the same underlying causes, but typically is used when talking about wave phenomena involving only a small number of wavefronts).

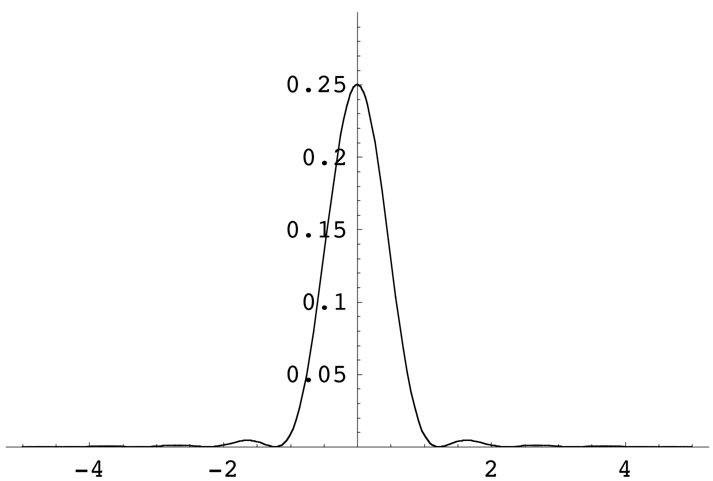
It is possible to mathematically describe what the image of point of light looks like under ideal optical conditions. The image is not a point, because of diffraction. The many wavefront sources around the aperture produce a characteristic pattern for a circular pupil called an Airy disk (1834):



This distribution of light has a very precise form depending on the pupil shape. For a circular pupil, the light energy distribution as a function of distance x away from the center axis, is given by the Airy disk function:

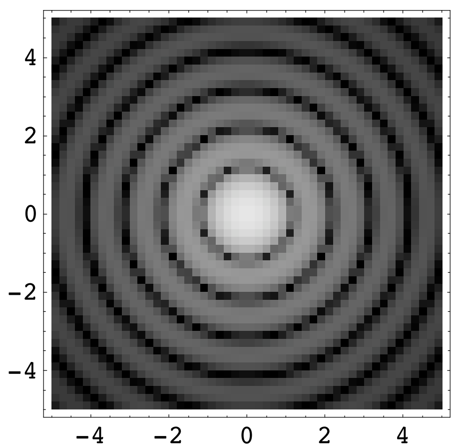
$$\left[\frac{J_1(\pi x)}{\pi x}\right]^2$$

In[9]:=
```
Airy[x_, y_] := If[(x == 0) && (y == 0) , 1 / 4,
   (BesselJ[1, Pi Sqrt[x^2 + y^2]] / (Pi (Sqrt[x^2 + y^2]))) ^2 ];
Plot[Airy[x, 0], {x, -5, 5}, PlotRange → {0, .3}];
```



where J_1 is a first order Bessel function--which if you haven't run into them before, is another one of a large class of "special functions" mathematicians, engineers, and physicists keep in their bag of tricks because they come in handy (See Gaskill, 1978 for details about imaging optics; Also look at section 6.9 in Horn, 1986 for a discusion of the zero and first order Bessel functions in the context of Hankel transforms, and see the *Mathematica* Help Browser for the differential equation for which Bessel functions are the solution). We won't go into the details of the derivation. Here is a two dimensional density plot, where the intensity is squashed by a log function so you can see the ripples (which actually fade away rather quickly as you can check by playing with the PlotRange in the linear cross-section plot above):

In[11]:=
```
LogAiry[x_, y_] := Log[0.00001 + Airy[x, y]];
DensityPlot[LogAiry[x, y], {x, -5, 5}, {y, -5, 5}, Mesh → False,
  PlotPoints → 50];
```



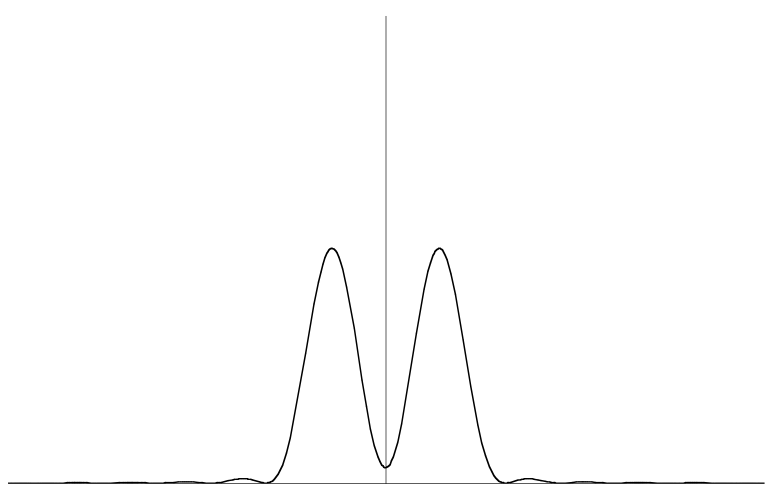
Some of the earliest psychophysics on the limits to spatial resolution was done by astronomers. In 1705, Hooke determined that resolution by eye of two stars was pretty good around 3 min, but became difficult around 1 min of arc. Can we account for human spatial resolution in terms of diffraction limits?

With 2 points of light, the diffraction patterns overlap. Imagine two of the above Airy functions superimposed exactly on top of each other

In[13]:=
```
twopoints[x_,d_] := Airy[x+d,0] + Airy[x-d,0];
```

Try plotting the intensity distribution of two points with several separations: d= 1, 0.5, and 0:

In[14]:=
```
Plot[twopoints[x, 1], {x, -7, 7}, Ticks → None,
  PlotRange → {{-7, 7}, {0, .5}}];
```



What if the light level was very low--is the above two-point plot representative of the signal?

■ A back-of-the-envelope calculation

When are the points distinguishable?

To answer these questions rigorously would require a signal detection theory model based on the noisiness of the light samples obtained to make a decision (see Geisler, 1984; 1989). This would require bringing what we learned about Poisson statistics into the picture. But let's ignore the photon noise for the moment, and assume that the temporal integration time is sufficient to absorb lots of photons. Then we can make some "back of the envelope calculations just based on diffraction. Let's proceed with a little less rigor, and ask: when are two points separated far enough so that the first zeros of the 2 Airy disks coincide?

The distance from the peak to the first zero is given by: $x = 1.22\lambda/a$, where a is the pupil diameter, so the distance between the two peaks when separated by twice this distance is:

$$d = \frac{2 \times 1.22\lambda}{a} \; radians$$

$$a = 2 \times 10^{-3} meters, \; \lambda = 555 \times 10^{-9} meters$$

picking 2mm pupil, and the wavelength corresponding to peak photopic sensitivity. We can then calculate the separation between the two points to be: **d = 0.04 degrees** of visual angle, or 2.3' of arc.(By the way, it is useful to bear in mind that, in the human visual system:

receptor spacing ~ 0.008 degrees in the fovea, or, about 0.48'.

■ ..but wait a minute...let's check our assumptions

Now let's how what we've just done is related to Hooke's observation. At first you might think that 2' of arc is pretty close to what Robert Hooke reported in 1705. But there were a couple of things wrong with our choices for both pupil size and wavelength. These figures aren't really appropropriate for scotopic vision where the pupil would be larger, and we should use the peak sensitivity for dark adapted rod vision.

If the pupil is 8 mm in diameter, and the wavelength is 505 nm, the predicted d would be smaller--about 0.008 degrees, which corresponds to 0.5' of arc. This is too small to account for the observed limit to human resolution of between 1 and 3' of arc for spatial resolution of two points in the dark.

Further, we were actually quite liberal in allowing this big of a separation between the two points. In fact, if there were no intensity fluctuations to worry about (imagine averaging the image of the two points over a long period of time), then there would be no theoretical limit to the separation one could discriminate, if the observer had an exact model of the two images (two vs one point) to look for. Factors other than diffraction enter to limit spatial resolution of two points in the dark. The main limitation is the separation of the rods, and the degree of neural convergence.

Although our ability to resolve two points for scotopic vision is worse than we would calculate based on diffraction-blur alone, there are conditions under bright light were our spatial resolution comes very close to the diffraction limit. But we need some more powerful mathematical tools to make this point solid. Later we will introduce a technique--spatial frequency analysis--to study spatial resolution in general, that will include diffraction as a special case. One of the consequences of spatial frequency analysis of imaging optics is that we will see that there is information which is lost when going through the optics, and noise averaging will not help to get it back.

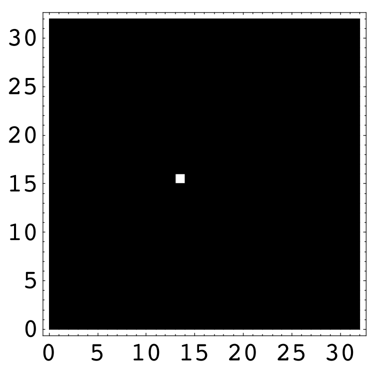
But first let's generalize the idea of the diffraction function to a point source, and see what it could buy us. This will give us a preview of the kind of predictions we would like to be able to make about image quality, and introduce a simple mathematical notion that has broad uses.

Point spread functions, blur, and convolution

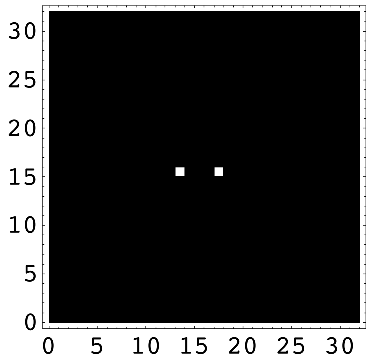
Going from two points to lots of points

Incoherent light adds. So we can use the principle of superposition to predict what the image of two points of light will look by summing up the images corresponding to what each would look like alone. This is the key idea behind linear sysems analysis pursued in greater detail below.

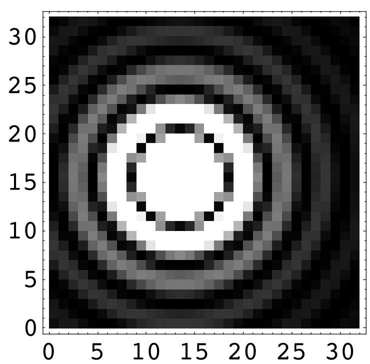
In[15]:=
```
offset = 2;
onepoint = Table[0.0, {32}, {32}];
onepoint[[16, 16 - offset]] = 1.0;
ListDensityPlot[onepoint];
```



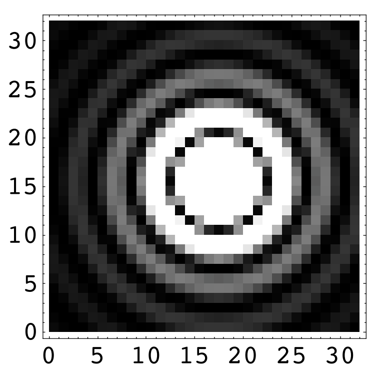
In[19]:=
```
anotheronepoint = Table[0.0, {32}, {32}];
anotheronepoint[[16, 16 + offset]] = 1.0;
twopoints = (onepoint + anotheronepoint ) / 2;
ListDensityPlot[twopoints];
```



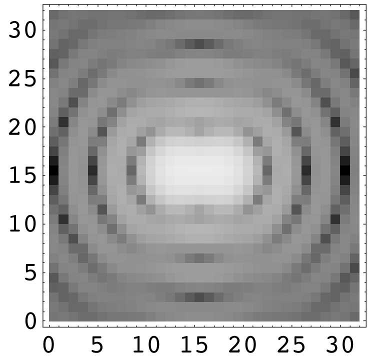
In[23]:=
```
oneblurredpoint = Table[Airy[(x - 16) / 4, (y - (16 - offset)) / 4],
   {x, 1, 32}, {y, 1, 32}];
ListDensityPlot[oneblurredpoint, Mesh -> False];
```



In[25]:=
```
anotheroneblurredpoint =
  Table[Airy[(x - 16) / 4, (y - (16 + offset)) / 4], {x, 1, 32}, {y, 1, 32}];
ListDensityPlot[anotheroneblurredpoint, Mesh -> False];
```



In[27]:=
```
ListDensityPlot[Log[oneblurredpoint + anotheroneblurredpoint],
  Mesh -> False];
```



If we have N points, then we just add up the images of each of them. This kind of operation occurs so frequently in many domains, that we will now formalize and generalize what we've just illustrated.

■ Point spread function (impulse response function)

We see that a theory of spatial resolution based on diffraction alone is not enough to account for spatial resolution. One of the reasons is that the actual pattern at the back of the eye due to a point source of light, in general, is not an Airy function. This is because of other optical aberrations (e.g. spherical, chromatic, astigmatism) in addition to diffraction, that blur the image.

An optical system's response to a delta function input (point of light) is called a point spread function (PSF):

$$\delta(x,y) \to PSF(x,y)$$

The Dirac delta function is an infinite spike at 0, but its area is 1:

```
Clear[x];
Integrate[DiracDelta[x], {x, -Infinity, Infinity}]
```

Out[65]= 1

(Recall that we encountered the Dirac delta function earlier in the formalism to represent discrete probability distributions as densities.)

Below we will study convolutions. One example of a convolution is the following integral ("convolution of Dirac delta function with a function f1()). It may look sophisticated, but the net result is that it is equivalent to evaluating f1[x] at t:

In[66]:=
```
Integrate[f1[t - x] DiracDelta[x], {x, -Infinity, Infinity}]
```

Out[66]= f1(t)

A delta function, for us, is a unit point source of light--it has no size, but the light intensity adds up to 1--this is a tricky definition, because it means that the intensity at (0,0 has to be infinite (the delta function is also called the unit impulse, e.g.

see Horn, chapter 6). It is a useful idealization that can make calculations easier. The convolution property of Dirac delta functions is important and crops up frequently.

For the special case of diffraction limited imaging with a circular aperture, the PSF is the Airy disk function.

Suppose for the moment that we have measured the point spread function (or "impulse response") of an optical imaging system--e.g. your eye. With enough information we could calculate it too--this is a branch of Fourier Optics. In any case, if we know the point spread function of the eye's optics, could we predict the form of the image on the retina, given an arbitrary input image? Yes. It is fairly straightforward if we can assume that the optics are linear and spatially homogeneous (a fairly good approximation for the optics of the eye over small patches called isoplanatic patches). Given an input image l(x,y), the output r(x,y) is given by the convolution of l(x,y) with the point spread function, PSF(x,y):

$$r(x,y) = \int_{-\infty}^{+\infty} l(x - x', y - y')PSF(x', y')dx'\,dy'$$

The idea is to treat each point in the input image as generating a smeared-out pattern (i.e.the PSF) on the output image whose intensity value is scaled by the input intensity l(x,y).Then add up all the contributions over the whole image.The short hand for convolution is

$$r(x,y) = l(x,y) * PSF(x,y)$$

Among other properties, you can show that the convolution operation commutes: f*g = g*f.

Understanding convolutions is important, not only for understanding optical image formation, but also because the formalism is applied in a number of models of the neural processing of images. Convolution is often used as a step to describe how image information is transformed by neural networks. So, r(x,y) could also be "neural image" response due to some neural analog of the point spread function. We will see too that convolution or pre-blurring is an important step in edge detection algorithms in computer vision. But more on this later.

If we know the point spread function of the eye's optics, we could predict the image intensity pattern for an arbitrary input. Further, given a linear approximation to neural image transmission, we might be able to find a "neural point spread function", and thereby predict the form that neural images might take at various stages of early vision. In some cases, the neurophysiologist's analog for the neural point spread function is receptive field. One has to be careful, because this analogy is only good to the extent that neural networks are spatially homogeneous (the receptive field of a single neuron can be modeled by the same set of weights from each input; the "spread" function corresponds to the weights assigned multiple outputs given a single input unit), and behave as if they are linear--e.g. calculating the response as the area under the input pattern times the receptive field. Under a number of instances, these are reasonable approximations.

■ Convolutions with Mathematica

Suppose we have measured the PSF, let's see how various PSF patterns affect the input image. First, we will get a sample image file, Fourier128x128.jpeg.

Under the **Input** menu, you can use **Get File Path** to find the directory of a file that you'd like to read in, and then set the current working directory to that location.

You can use this combination of function calls to bring up a window to hunt for a file whose directory you want to be the default working directory:

In[30]:=
```
SetDirectory[DirectoryName[Experimental`FileBrowse[False]]]
```

Out[30]= /Users/kersten/Desktop

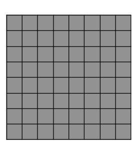
Once you've set your working directory, you can read in a jpeg file. You have to select the [[1,1]] component of the list to pick off the data for the array:

In[31]:=
```
fourier = Import["Fourier128x128.jpeg"][[1, 1]];
ListDensityPlot[fourier, Frame → False, Mesh → False,
  AspectRatio → Automatic];
```

You can get the width and height using: **Dimensions[fourier]**. Try it.

We set up a simple kernel in which each element is 1/64 in an 8x8 array. Convolution replaces each pixel intensity with the average value of the 64 values in the square image region surrounding it:

In[33]:=
```
kernel = Table[1 / 64, {i, 1, 8}, {j, 1, 8}];
ListDensityPlot[kernel, Frame → False, AspectRatio → Automatic];
```



In[35]:=
```
blurFourier = ListConvolve[kernel, fourier];
ListDensityPlot[blurFourier, Frame → False, Mesh → False,
  AspectRatio → Automatic];
```

In[37]:=
```
kernel = Table[Airy[x, y], {x, -2.3, 2.3, .3}, {y, -2.3, 2.3, .3}];
```

In[38]:=
```
blurFourier = ListConvolve[kernel, N[fourier]];
ListDensityPlot[blurFourier, Frame → False, Mesh → False,
  AspectRatio → Automatic];
```

It is technically hard to measure the response to a delta function for most real (physical or biological) devices--infinite or near infinite values in the input fall outside any approximately linear range the device might have. An alternative is to measure the responses to other basic image patterns that are formally related to point spread functions. This problem leads us to a more general consideration of linear systems theory, and in particular to the spatial frequency or Fourier analysis of images. Spatial frequency has a number of other nice properties over point spread function characterizations. One, alluded to earlier, is that it will make explicit the kind of information that is lost, irrespective of the noise considerations.

Exercise: Astigmatism

Modify the above blurring kernel so that it blurs more in the horizontal than the vertical direction. This would simulate the effect of astigmatism.

Exercise:

Now try a kernel = {{0,1,0},{1,-4,1},{0,1,0}};

In[40]:=
```
kernel = {{0, 1, 0}, {1, -4, 1}, {0, 1, 0}} // MatrixForm
```

Out[40]//MatrixForm=

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

■ **Convolution is a special case of a a linear system. Let's look at the general discrete case.**

Linear systems analysis

■ Introduction

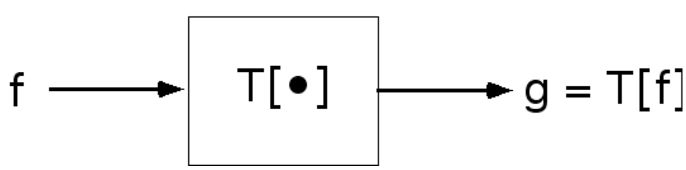
The world of input/output systems can be divided up into linear and non-linear systems. Linear systems are nice because the mathematics that describes them is not only well-known, but also has a mature elegance. On the other hand, it is a fair statement to say that most real-world systems are not linear, and thus hard to analyze...but fascinating if for that reason alone. That nature is usually non-linear doesn't mean one shouldn't familiarize oneself with the basics of linear system theory. Many times a non-linear system has a sufficiently smooth mapping that it can be approximated by a linear one over restricted ranges of parameter values. The assumption of linearity is an excellent starting point--but must be tested.

So exactly what is a "linear system"?

The notion of a "linear system" is a generalization of the input/output properties of a straight line passing through zero. The matrix equation **W.x == y** is a linear system. This means that if **W** is a matrix, **x1** and **x2** are vectors, and a and b are scalars:

W.(a **x**1 + b **x2**) = a **W.x1** + b **W.x2**

This is a consequence of the laws of matrix algebra.The idea of a linear system has been generalized beyond matrix algebra. Imagine we have a box that takes inputs such as f, and outputs g = T[f].



The abstract definition of a linear system is that it satsifies:

T[a f + b g] = a T[f] + b T[g]

where T is the transformation that takes the sum of scaled inputs f, g (which can be functions or vectors) to the sum of the scaled transformation of f and g. The property, that the output of a sum is the sum of the outputs, is sometimes known as the *superposition principle* for linear systems. The fact that linear systems show superposition is good for doing theory, but as we will see later, it limits the kind of computations that can be done with linear systems, and thus with linear neural network models.

Characterizing a linear system by its response to an orthonormal basis set

Suppose we have an unknown physical system, which we model as a linear system with matrix **T**:

In[41]:=
```
T = Table[Random[],{i,1,8},{j,1,8}];
```

We would like to make a simple set of measurements that could characterize **T** in such a way that we could predict the output of **T** to any input. This is the sort of task that engineers face when wanting to characterize, say a stereo amplifier (as a model linear system), so that the output sound can be predicted for any input sound.What kind of measurements would tell us what **T** is? Well, we could just "stimulate" the system with cartesian vectors {1,0,0,0,0,0,0,0},{0,1,0,0,0,0,0,0}, and so forth and collect the responses which would be the columns of **T**. This has two practical problems: 1) for a real physical system, such as your stereo, or a neuron in the eye, this would require stimulating it with a high-intensity audio or light intensity spike, which could damage what you are trying to study; 2) Characterizing the linear system by a matrix **T**, requires n^2 numbers, where **n** is the input signal vector length--and **n** can be pretty big for both audio and visual systems. Problem 2) has a nice solution when **T** is symmetric, and even nicer solution if the rows are shifted versions of each other (this is addressed later). Problem 1) can be addressed by showing that we can characterize **T** with any basis set--so we can pick one that won't blow out the physical system being tested.

The set of Walsh functions is one possible set that has the advantage that the elements that contribute to the "energy", i.e. (the square of the length) are distributed across the vector.

In[42]:=
```
Vectorlength[x_] := N[Sqrt[x.x]]
```

In[43]:=
```
v1 = {1, 1, 1, 1, 1, 1, 1, 1}; w1 = v1/Vectorlength[v1];
v2 = {1,-1,-1, 1, 1,-1,-1, 1}; w2 = v2/Vectorlength[v2];
v3 = {1, 1,-1,-1,-1,-1, 1, 1}; w3 = v3/Vectorlength[v3];
v4 = {1,-1, 1,-1,-1, 1,-1, 1}; w4 = v4/Vectorlength[v4];
v5 = {1, 1, 1, 1,-1,-1,-1,-1}; w5 = v5/Vectorlength[v5];
v6 = {1,-1,-1, 1,-1, 1, 1,-1}; w6 = v6/Vectorlength[v6];
v7 = {1, 1,-1,-1, 1, 1,-1,-1}; w7 = v7/Vectorlength[v7];
v8 = {1,-1, 1,-1, 1,-1, 1,-1}; w8 = v8/Vectorlength[v8];
```

The set $\{w_i\}$ is *complete* (i.e. it spans 8-space in such a way that we can easily express any vector as a linear sum of these basis vectors):

$$g = \sum (g.w_i)\, w_i \tag{3}$$

The set has other nice properties like: *orthogonality* (check out the dot products $w_i.w_j$), and *normality* (what is the vector length of w_i?). When a basis set has both properties (orthogonality and normality), it is called an orthonormal basis set. We will see shortly that orthonormal basis sets make it easy to do calculations.

An arbitrary vector, **g**

In[51]:=
```
g = {2,6,1,7,11,4,13, 29};
```

is the sum of its own projections onto the basis set:

In[52]:=
```
(g.w1) w1  + (g.w2) w2  +(g.w3) w3  +(g.w4) w4  +
(g.w5) w5  +(g.w6) w6  +(g.w7) w7  +(g.w8) w8
```

Out[52]=
```
{2., 6., 1., 7., 11., 4., 13., 29.}
```

Suppose we now do an "experiment" to find out how **T** transforms the vectors of our basis set:, and we put all of these transformed basis elements into a new set of vectors **newW[[i]]. newW** is a matrix for which each row is the response of **T** to a basis vector.

In[53]:=
```
newW = {T.w1,T.w2,T.w3,T.w4,T.w5,T.w6,T.w7,T.w8};
```

Note that **newW** is an 8x8 matrix. So how can we calculate the output of **T**, given **g** without actually running the input through **T**? If we do run the input through **T** we get:

In[54]:=
```
T.g
```

Out[54]=
```
{31.6263, 45.2444, 27.4108, 46.3986, 52.5677, 44.4026, 25.979, 52.5991}
```

But by the principle of linearity, we can also calculate the output by finding the "spectrum" of **g**, and then scaling each of the transformed basis elements by the spectrum and adding them up:

$$T.g = T.\left\{\sum (g.w_i)\, w_i\right\} = \sum (g.w_i)\, T.w_i \tag{4}$$

In[55]:=
```
(g.w1) T.w1 + (g.w2) T.w2 + (g.w3) T.w3 + (g.w4) T.w4 +
(g.w5) T.w5 + (g.w6) T.w6 + (g.w7) T.w7 + (g.w8) T.w8
```

Out[55]=
```
{31.6263, 45.2444, 27.4108, 46.3986, 52.5677, 44.4026, 25.979, 52.5991}
```

Of course, we have already done our "experiment", so we know what the transformed basis vectors are, we stored them as rows of the matrix **newW**. We can calculate what the specturm (**g.w**i) is, so the output of **T** is:

In[56]:=
```
(g.w1) newW[[1]] + (g.w2) newW[[2]] + (g.w3) newW[[3]] +
(g.w4) newW[[4]] + (g.w5) newW[[5]] + (g.w6) newW[[6]] +
(g.w7) newW[[7]] + (g.w8) newW[[8]]
```

Out[56]=
```
{31.6263, 45.2444, 27.4108, 46.3986, 52.5677, 44.4026, 25.979, 52.5991}
```

■ Same thing in more concise notation

Let the basis vectors be the rows of a matrix **W**:

In[57]:=
```
W = {w1, w2, w3, w4, w5, w6, w7, w8};
```

So again, we can project **g** onto the rows of **W**, and then reconstitute it in terms of **W** to get **g** back again:

In[58]:=
```
(W.g).W
```

Out[58]= {2., 6., 1., 7., 11., 4., 13., 29.}

In[59]:=
```
g.Transpose[W].newW
```

Out[59]= {31.6263, 45.2444, 27.4108, 46.3986, 52.5677, 44.4026, 25.979, 52.5991}

The main idea is: characterize and unknown system T by its response to orthonormal vectors. As an exercise, you can:

Show that T==Transpose[newW].W, and that the system output is thus: Transpose[newW].W.g

---

These new basis vectors do span 8-space, but they are not necessarily orthonormal. Under what conditions would they be orthogonal? What if the matrix T was symmetric, and we deliberately chose the basis set to describe our input to be the eigenvectors of T?

Preview: What if the choice of basis set is the set of eigenvectors of T?

We've seen how linearity provides us with a method for characterizing a linear system in terms of the responses of the system to the basis vectors. The problem is that if the input signals are long vectors, say with dimension 40,000, then this set of basis vector responses is really big--$1.6 \mathrm{x} 10^9$.

Construct a symmetric matrix transformation, T. Show that if the elements of the basis set are the eigenvectors of T, then the transformation of any arbitrary input vector x is given by:

$$T[\mathbf{x}] = \sum \alpha_i \lambda_i \mathbf{e}_i$$

Where the α_i are the projections of x onto each eigenvector. Having the eigenvectors of **T** enables us to express the input and output of **T** in terms of the same basis set--the eigenvectors. All **T** does to the input is to scale its projection onto each eigenvector by the eigenvalue for that eigenvector. The set of these eigenvalues is sometimes called the *modulation transfer function* because it describes how the amplitude of the eigenvectors change as they pass through **T**.

Linear systems analysis is the foundation of Fourier analysis, and is why it makes sense to characterize your stereo amplifier in terms of frequency response. But your stereo isn't just any linear system--it has the special property that if you input a sound at time t and measure the response, and then you input the same sound again at a later time, you get the same response, except of course that is is shifted in time. It is said to be a shift-invariant system. The eigenvectors of a shift-invariant system are sinusoids. (The eigenvectors of the symmetric matrix are sinusoids, not just because the matrix was symmetric, but also because each row of the matrix was a shifted version of the previous row--the elements along any given diagonal are identical. This is called a symmetric Toeplitz matrix.)

Sinewave inputs are the eigenvectors of your stereo system. The dimensionality is much higher--if you are interested in frequencies up to 20,000 Hz, your eigenvector for this highest frequency would have least 40,000 elements--not just 8!

This kind of analysis has been applied not only to physical systems, but to a wide range of neural sensory systems. For the visual system alone, linear systems analysis has been used to study the cat retina (Enroth-Cugell and Robson, 1964), the monkey visual cortex, and the human contrast sensivity system as a whole (Campbell and Robson, 1968).

Much empirical analysis has been done using linear systems theory to characterize neural sensory systems, and other neural systems such as those for eye movements. It works wonderfully as long as the linear system approximation holds. And it does do quite well for the lateral eye of the limulus, X-cells and P-cells of the mammalian visual system, over restricted ranges for so-called "simple" cells in the visual cortex, among others. The optics of the simple eye is another example of an approximately linear system. Many non-linear systems can be approximated as linear systems over smooth subdomains.

■ Summary

In summary, if T has n distinct orthogonal eigenvectors, $\mathbf{e}_i$, and known eigenvalues, λ_i, then we have a particularly easy way to calculate the response to an input **x**:

Step 1: Project **x** onto eigenvectors of T using the dot product: $\mathbf{x}.\mathbf{e}_i$

Step 2: Scale each $\mathbf{x}.\mathbf{e}_i$ by the eigenvalue of $\mathbf{e}_i$: $\lambda_i \mathbf{x}.\mathbf{e}_i$

Step 3: Scale each $\mathbf{e}_i$ by $\lambda_i \mathbf{x}.\mathbf{e}_i$

Step 4: Sum these up. That's the response of T to **x**! : $\sum_i \lambda_i \boldsymbol{x}.\boldsymbol{e}_i$

Next time

Linear shift-invariant systems: Fourier analysis of images & optics

- **Sinewave basis elements**

- **The Fast Fourier Transform (FFT)**

Appendices

Fourier integral

Fourier transform : $F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{i \omega t}\, dt$,

Inverse fourier transform : $f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega)\, e^{-i \omega t}\, d\omega$

- **What is the Fourier transform of Cos[x]?**

In[67]:=
```
Clear[x];
FourierTransform[Cos[x], x, f]
```

Out[68]=
$$\sqrt{\frac{\pi}{2}}\, \delta(f-1) + \sqrt{\frac{\pi}{2}}\, \delta(f+1)$$

■ **Some functions have the same shape in Fourier domain as in the spatial domain**

In[69]:=
```
FourierTransform[Exp[-(x / σ)^2], x, f]
```

Out[69]=

$$\frac{e^{-\frac{1}{4} f^2 \sigma^2}}{\sqrt{2} \sqrt{\frac{1}{\sigma^2}}}$$

Note that as the gaussian gets narrower in space, it gets broader in frequency.

References

Campbell, F. W., & Green, D. (1966). Optical and retinal factors affecting visual resolution. *Journal of Physiology (Lond.), 181*, 576-593.

Gaskill, J. D. (1978). Linear Systems, Fourier Transforms, and Optics. New York: John Wiley & Sons.

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *J. Opt. Soc. Am. A, 1,*, 775-782.

Geisler, W. (1989). Sequential Ideal-Observer analysis of visual discriminations. *Psychological Review, 96*(2), 267-314.

Grenander, U. (1996). *Elements of pattern theory*. Baltimore: Johns Hopkins University Press.

Horn, B. K. P. (1986). *Robot Vision*. Cambridge MA: MIT Press.

Kersten, D. (1997). Inverse 3D Graphics: A Metaphor for Visual Perception. *Behavior Research Methods, Instruments, & Computers, 29*(1), 37-46.

Liang, J., & Williams, D. R. (1997). Aberrations and retinal image quality of the normal human eye. *J Opt Soc Am A, 14*(11), 2873-2883.

Mumford, D. (1996). Pattern theory: A unifying perspective. In D. C. Knill & R. W. (Eds.), *Perception as Bayesian Inference* (pp. Chapter 2). Cambridge: Cambridge University Press.

Mumford, David (2002) *Pattern Theory: the Mathematics of Perception.* APPTS Report #02-10, August 2002. [pdf]