3. The Gaussian kernel

"Everybody believes in the exponential law of errors: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation" (Lippman in [Whittaker1967, p. 179]).

3.1 The Gaussian kernel

The Gaussian (better Gaußian) kernel is named after Carl Friedrich Gauß (1777-1855), a brilliant German mathematician. This chapter discusses many of the nice and peculiar properties of the Gaussian kernel.

```
<< FEVinit`; << FEVFunctions`;
Show[Import["Gauss10DM.gif"]];
```


Figure 3.1. The Gaussian kernel is apparent on every German banknote of DM 10,- where it is depicted next to its famous inventor when he was 55 years old. The new Euro replaces these banknotes.

The Gaussian kernel is defined in 1-D, 2D and N-D respectively as

$$G_{1D}(x; \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2}{2\sigma^2}}, \; G_{2D}(x,\, y; \sigma) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}}, \; G_{ND}(\vec{x}; \sigma) = \frac{1}{\left(\sqrt{2\pi}\,\sigma\right)^N}\, e^{-\frac{|\vec{x}|^2}{2\sigma^2}}$$

The σ determines the *width* of the Gaussian kernel. In statistics, when we consider the Gaussian probability density function it is called the *standard deviation*, and the square of it, σ^2, the *variance*. In the rest of this book, when we consider the Gaussian as an aperture function of some observation, we will refer to σ as the *inner scale* or shortly *scale*. In the whole of this book the scale can only take positive values, $\sigma > 0$. In the process of observation σ can never become zero. For, this would imply making an observation through an infinitesimally small aperture, which is impossible. The factor of 2 in the exponent is a matter of convention, because we then have a 'cleaner' formula for the diffusion equation, as we will see later on. The semicolon between the spatial and scale parameters is conventionally put there to make the difference between these parameters explicit. The scale-dimension is *not* just another spatial dimension, as we will thoroughly discuss in the remainder of this book.

3.2 Normalization

The term $\frac{1}{\sqrt{2\pi}\,\sigma}$ in front of the one-dimensional Gaussian kernel is the normalization constant. It comes from the fact that the integral over the exponential function is not unity: $\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} \, dx = \sqrt{2\pi}\,\sigma$. With the normalization constant this Gaussian kernel is a *normalized* kernel, i.e. its integral over its full domain is unity for every σ. This means that increasing the σ of the kernel reduces the amplitude substantially. Let us look at the graphs of the normalized kernels for $\sigma = 0.3$, $\sigma = 1$ and $\sigma = 2$ plotted on the same axes:

```
Unprotect[gauss]; gauss[x_, σ_] := ────────── Exp[- ──── ];
                                    σ √‾2‾π‾        2 σ²

Block[{$DisplayFunction = Identity}, {p1, p2, p3} =
    Plot[gauss[x, σ = #], {x, -5, 5}, PlotRange -> {0, 1.4}] & /@ {.3, 1, 2}];
Show[GraphicsArray[{p1, p2, p3}], ImageSize -> {450, 130}];
```
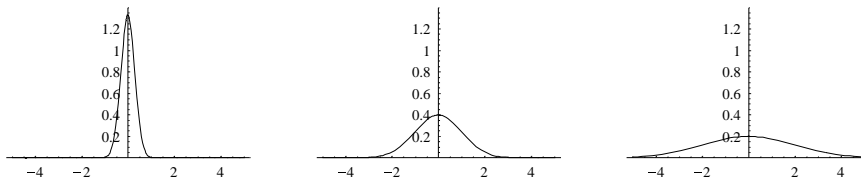

Figure 3.2. The Gaussian function at scales $\sigma = .3$, $\sigma = 1$ and $\sigma = 2$. The kernel is normalized, so the area under the curve is always unity.

The normalization ensures that the average greylevel of the image remains the same when we blur the image with this kernel. This is known as *average grey level invariance*.

3.3 Cascade property

The shape of the kernel remains the same, irrespective of the σ. When we convolve two Gaussian kernels we get a new wider Gaussian with a variance σ^2 which is the sum of the variances of the constituting Gaussians: $g_{\text{new}}(\vec{x}; \sigma_1^2 + \sigma_2^2) = g_1(\vec{x}; \sigma_1^2)\, g_2(\vec{x}; \sigma_2^2)$.

```
σ =.; FullSimplify[ ∫_{-∞}^{∞} gauss[x, σ₁] gauss[α - x, σ₂] dx,
    {σ₁ > 0, Im[σ₁] == 0, σ₂ > 0, Im[σ₂] == 0} ]
```

$$\frac{e^{-\frac{\alpha^2}{2(\sigma_1^2 + \sigma_2^2)}}}{\sqrt{2\pi}\,\sqrt{\sigma_1^2 + \sigma_2^2}}$$

This phenomenon, i.e. that a new function emerges that is similar to the constituting functions, is called *self-similarity*. The Gaussian is a *self-similar function*. Convolution with a Gaussian is a linear operation, so a convolution with a Gaussian kernel followed by a convolution with again a Gaussian kernel is equivalent to convolution with the broader kernel. Note that the *squares* of σ add, not the σ's themselves. Of course we can concatenate as many blurring steps as we want to create a larger blurring step. With analogy to a cascade of waterfalls spanning the same height as the total waterfall, this phenomenon is also known as the *cascade smoothing property*.

3.4 The scale parameter

In order to avoid the summing of squares, one often uses the following parametrization: $2\sigma^2 \rightarrow t$, so the Gaussian kernel get a particular short form. In N dimensions $G_{ND}(\vec{x}, t) = \frac{1}{(\pi\tau)^{N/2}} e^{-\frac{x^2}{\tau}}$. It is this t that emerges in the diffusion equation $\frac{\partial L}{\partial t} = \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} + \frac{\partial^2 L}{\partial z^2}$. It is often referred to as 'scale'(like in: differentiation to scale, $\frac{\partial L}{\partial t}$), but a better name is *variance*.

To make the self-similarity of the Gaussian kernel explicit, we can introduce a new *dimensionless* spatial parameter, $\tilde{x} = \frac{x}{\sigma\sqrt{2}}$. We say that we have *reparametrized* the x-axis. Now the Gaussian kernel becomes: $g_n(\tilde{x}; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\tilde{x}^2}$, or $g_n(\tilde{x}; t) = \frac{1}{(\pi\tau)^{N/2}} e^{-\tilde{x}^2}$. In other words: if we walk along the spatial axis in footsteps expressed in scale-units (σ's), all kernels are of equal size or 'width'(but due to the normalization constraint not necessarily of the same amplitude). We now have a 'natural' size of footstep to walk over the spatial coordinate: a unit step in x is now $\sigma\sqrt{2}$, so in more blurred images we make bigger steps. We call this basic Gaussian kernel the *natural* Gaussian kernel $g_n(\tilde{x}; \sigma)$. The new coordinate $\tilde{x} = \frac{x}{\sigma\sqrt{2}}$ is called the *natural coordinate*. It eliminates the scale factor σ from the spatial coordinates, i.e. it makes the Gaussian kernels similar, despite their different inner scales. We will encounter natural coordinates many times hereafter.

The spatial extent of the Gaussian kernel ranges from $-\infty$ to $+\infty$, but in practice it has negligeable values for x larger then a few (say 5) σ. The numerical value at x=5σ, and the area under the curve from x=5σ to infinity (recall that the total area is 1):

```
gauss[5, 1] // N
Integrate[gauss[x, 1], {x, 5, Infinity}] // N

1.48672 × 10⁻⁶

2.86652 × 10⁻⁷
```

The larger we make the standard deviation σ, the more the image gets blurred. In the limit to infinity, the image becomes homogenous in intensity. The final intensity is the average intensity of the image. This is true for an image with infinite extent, which in practice will never occur, of course. The boundary has to be taken into account. Actually, one can take many choices what to do at the boundary, it is a matter of concensus. Boundaries are discussed in detail in chapter 5, where practical issues of computer implementation are discussed.

3.5 Relation to generalized functions

The Gaussian kernel is the physical equivalent of the *mathematical point*. It is not strictly local, like the mathematical point, but *semi-local*. It has a *Gaussian weighted extent*, indicated by its inner scale σ. Because scale-space theory is revolving around the Gaussian function and its derivatives as a physical differential operator (in more detail explained in the next chapter), we will focus here on some mathematical notions that are directly related, i.e. the mathematical notions underlying sampling of values from functions and their derivatives at *selected* points (i.e. that is why it is referred to as sampling). The mathematical functions involved are the *generalized functions*, i.e. the Delta-Dirac function, the Heavyside function and the error function. We study in the next section these functions in more detail.

When we take the limit as the inner scale goes down to zero, we get the mathematical delta function, or Delta-Dirac function, δ(x). This function, named after Dirac (1862-1923) is everywhere zero except in $x = 0$, where it has infinite amplitude and zero width, its area is unity.

$$\lim_{\sigma\downarrow0}\ \left(\frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{x^2}{2\sigma^2}}\right)\ =\ \delta(x).$$

δ(x) is called the *sampling function* in mathematics, because the Dirac delta function adequately samples just one point out of a function when integrated. It is assumed that $f(x)$ is continuous at $x = a$:

```
<< "DiracDelta'"
```

$$\int_{-\infty}^{\infty}\text{DiracDelta[x - a] f[x] dx}$$

```
f[a]
```

The *sampling property of derivatives* of the Dirac delta function is illustrated below:

$$\int_{-\infty}^{\infty}\text{D[DiracDelta[x], \{x, 2\}] f[x] dx}$$

```
f''[0]
```

The integral of the Gaussian kernel from -∞ to x is a famous function as well. It is the *error function*, or *cumulative* Gaussian function, and is defined as:

```
σ =.;
```
$$\text{err[x\_, }\sigma\text{\_] }=\int_{0}^{x}\frac{1}{\sigma\sqrt{2\pi}}\ \text{Exp}\left[-\frac{y^2}{2\sigma^2}\right]\text{dy}$$

$$\frac{1}{2}\ \text{Erf}\left[\frac{x}{\sqrt{2}\ \sigma}\right]$$

The *y* in the integral above is just a dummy integration variable, and is integrated out. The *Mathematica* error function is `Erf[x]`. In our integral of the Gaussian function we need to do the reparametrization $x \to \frac{x}{\sigma\sqrt{2}}$. Again we recognize the natural coordinates. The factor $\frac{1}{2}$ is due to the fact that integration starts halfway, in x = 0.

$$\sigma = 1.;\ \text{Plot}\left[\frac{1}{2}\ \text{Erf}\left[\frac{x}{\sigma\sqrt{2}}\right],\ \{x,\ -4,\ 4\},\right.$$
$$\left.\text{AspectRatio} \to .3,\ \text{AxesLabel} \to \{\text{"x"},\ \text{"Erf[x]"}\}\right];$$
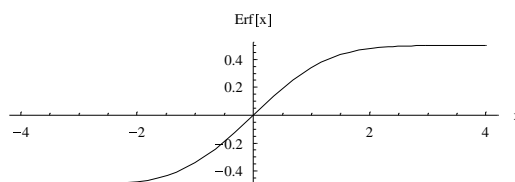

Figure 3.3. The error function **Erf[x]** is the cumulative Gaussian function.

When the inner scale σ of the error function goes to zero, we get in the limiting case the so-called *Heavyside function* or *unitstep function*. The derivative of the Heavyside function is the Delta-Dirac function, just as the derivative of the error function of the Gaussian kernel.

```mathematica
σ = .1; Plot[ 1/2 Erf[ x/(σ √2) ], {x, -4, 4},
  AspectRatio -> .3, AxesLabel -> {"x", "Erf[x]"}];
```
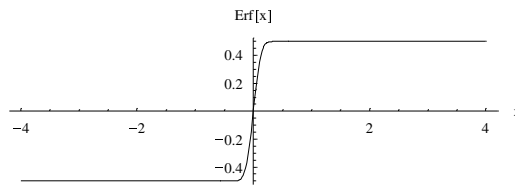

Figure 3.4. For decreasing σ the Errorfunction begins to look like a stepfunction. The Errorfunction is the Gaussian blurred step-edge.

```mathematica
Plot[UnitStep[x], {x, -4, 4},
  DisplayFunction -> $DisplayFunction, AspectRatio -> .3,
  AxesLabel -> {"x", "Heavyside[x], UnitStep[x]"}, PlotStyle -> Thickness[.015]];
```
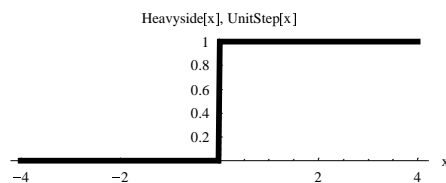

Figure 3.5. The Heavysidefunction is the generalized unit stepfunction. It is the limiting case of the Errorfunction for lim $\sigma \to 0$.

The derivative of the Heavyside stepfunction is the Deltafunction again:

```mathematica
D[UnitStep[x], x]
```

```mathematica
DiracDelta[x]
```

3.6 Separability

The Gaussian kernel for dimensions higher then one, say *N*, can be described as a regular product of *N* one-dimensional kernels. Example: $g_{2D}(x, y; \sigma_1^2 + \sigma_2^2) = g_{1D}(x; \sigma_1^2) \, g_{1D}(y; \sigma_2^2)$ where the space in between is the product operator. The regular product also explains the exponent *N* in the normalization constant for *N*-dimensional Gaussian kernels in (0). Because higher dimensional Gaussian kernels are regular products of one-dimensional Gaussians, they are called *separable*. We will use quite often this property of *separability*.

3.7 Relation to binomial coefficients

Another place where the Gaussian function emerges is in expansions of powers of polynomials. Here is an example:

```
Expand[(x + y)^20]
```

$x^{20} + 20 x^{19} y + 190 x^{18} y^2 + 1140 x^{17} y^3 + 4845 x^{16} y^4 + 15504 x^{15} y^5 + 38760 x^{14} y^6 + 77520 x^{13} y^7 +$
$125970 x^{12} y^8 + 167960 x^{11} y^9 + 184756 x^{10} y^{10} + 167960 x^9 y^{11} + 125970 x^8 y^{12} + 77520 x^7 y^{13} +$
$38760 x^6 y^{14} + 15504 x^5 y^{15} + 4845 x^4 y^{16} + 1140 x^3 y^{17} + 190 x^2 y^{18} + 20 x y^{19} + y^{20}$

The coefficients of this expansion are the *binomial coefficients* ($\binom{n}{m}$) ('n over m'):

```
ListPlot[Table[Binomial[20, n], {n, 1, 20}],
  PlotStyle -> {PointSize[.015]}, AspectRatio -> .3];
```
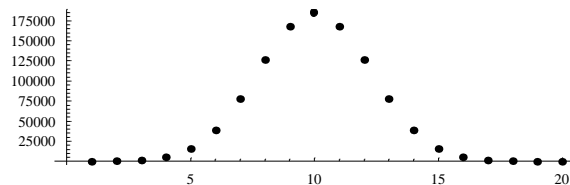

Figure 3.6. Binomial coefficients approximate a Gaussian distribution for increasing order.

And here in two dimensions:

```
ListPlot3D[Table[Binomial[20, n] Binomial[20, m],
{n, 1, 20}, {m, 1, 20}], PlotRange -> All];
```
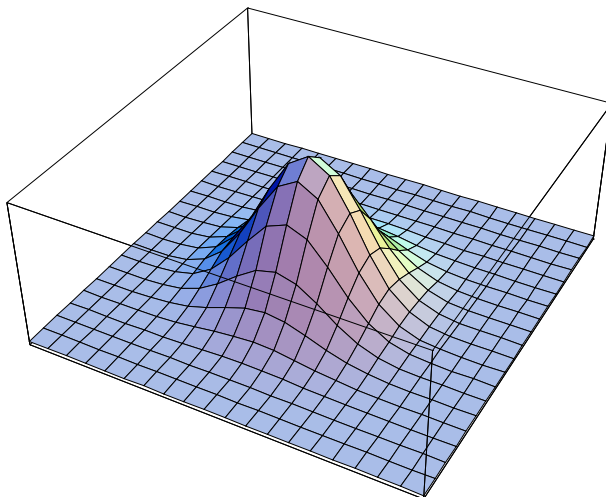

Figure 3.7. Binomial coefficients approximate a Gaussian distribution for increasing order. Here in 2 dimensions we see separability again.

3.8 Central limit theorem

We see in the paragraph above the relation with the *central limit theorem*: any repetitive operator goes in the limit to a Gaussian function. Later, when we study the discrete implementation of the Gaussian kernel and discrete sampled data, we will see the relation between interpolation schemes and the binomial coefficients.

We study a repeated convolution of two blockfunctions with each other:

```
<< Calculus`DiracDelta`
f[x_] := UnitStep[1 / 2 + x] + UnitStep[1 / 2 - x] - 1;
g[x_] := UnitStep[1 / 2 + x] + UnitStep[1 / 2 - x] - 1;
```

```
Plot[f[x], {x, -3, 3}, ImageSize -> 200];
```
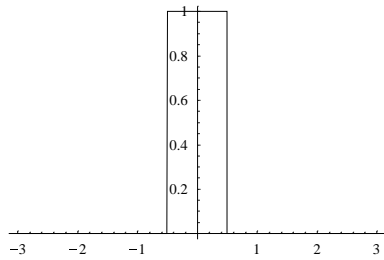

Figure 3.8. The analytical blockfunction is a combination of two Heavyside unitstep functions.

We calculate analytically the convolution integral

```
h1 = Integrate[f[x] g[x - x1], {x, -∞, ∞}]
```

$-1 - (-1 + x1)\, \text{UnitStep}[1 - x1] +$
$x1\, \text{UnitStep}[-x1] - x1\, \text{UnitStep}[x1] + (1 + x1)\, \text{UnitStep}[1 + x1]$

```
Plot[h1, {x1, -3, 3}, PlotRange -> All, ImageSize -> 200];
```
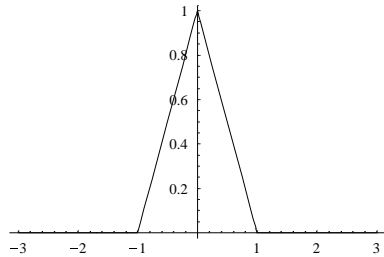

Figure 3.9. One times a convolution of a blockfunction with the same blockfunction gives a triangle function.

The next convolution is this function convolved with the block function again:

```
h2 = Integrate[(h1 /. x1 -> x) g[x - x1], {x, -∞, ∞}]
```

$-1 + \frac{1}{8}\, (1 - 2\, x1)^2 + \frac{1}{8}\, (1 + 2\, x1)^2 + \frac{1}{8}\, (3 - 4\, x1 - 4\, x1^2) + \frac{1}{8}\, (3 + 4\, x1 - 4\, x1^2) +$

$\frac{1}{8}\, \left(-4 + (3 - 2\, x1)^2\, \text{UnitStep}\left[\frac{3}{2} - x1\right] + (1 - 2\, x1)^2\, \text{UnitStep}\left[-\frac{1}{2} + x1\right]\right) +$

$\frac{1}{4}\, \left(-(1 - 2\, x1)^2\, \text{UnitStep}\left[\frac{1}{2} - x1\right] - (1 + 2\, x1)^2\, \text{UnitStep}\left[\frac{1}{2} + x1\right]\right) +$

$\frac{1}{8}\, \left(-4 + (1 + 2\, x1)^2\, \text{UnitStep}\left[-\frac{1}{2} - x1\right] + (3 + 2\, x1)^2\, \text{UnitStep}\left[\frac{3}{2} + x1\right]\right)$

We see that we get a result that begins to look more towards a Gaussian:

```
Plot[{h2, gauss[x1, .5]}, {x1, -3, 3}, PlotRange -> All,
  PlotStyle -> {Dashing[{}], Dashing[{0.02, 0.02}]}, ImageSize -> 300];
```
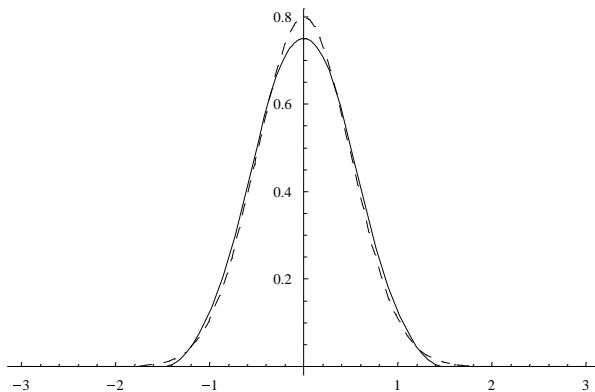

Figure 3.10. Two times a convolution of a blockfunction with the same blockfunction gives a function that rapidly begins to look like a Gaussian function. A Gaussian kernel with $\sigma = 0.5$ is drawn in a dotted mode for comparison.

The real Gaussian is reached when we apply an infinite number of these convolutions with the same function. It is remarkable that this result applies for the infinite repetition of *any* convolution kernel. This is the *central limit theorem*.

▲   **Task 3.1 Show the central limit theorem in practice for a number of other arbitrary kernels.**

3.9 Anisotropy

The Gaussian kernel as specified above is *isotropic*, which means that the behaviour of the function is in any *direction* the same. For 2D this means the Gaussian function is circular, for 3D it looks like a fuzzy sphere.

```
PlotGradientField[-gauss[x, 1] gauss[y, 1],
  {x, -3, 3}, {y, -3, 3}, PlotPoints -> 20];
```
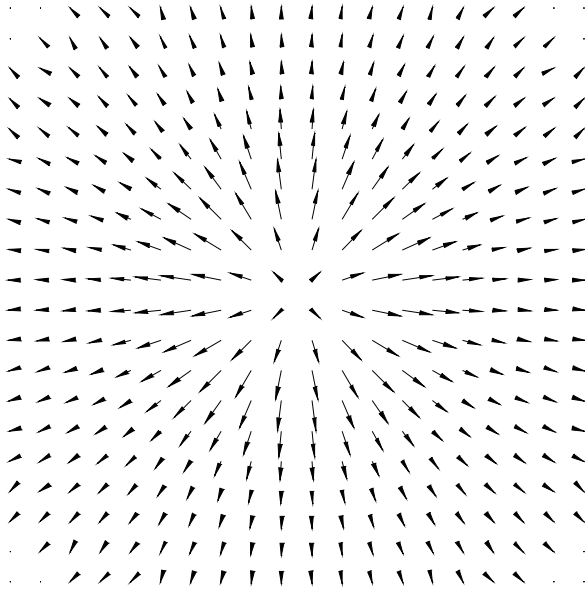

Figure 3.11. The slope of an isotropic Gaussian function is indicated by arrows here. There are circularly symmetric, i.e. in all directions the same, from which the name *isotropic* derives. The arrows are in the direction of the *normal* of the intensity landscape, and are called *gradient vectors*. We will study such gradient vectors in great detail in chapter 10.

It is of no use to speak of isotropy in 1-D. When the standarddeviations in the different dimensions are not equal, we call the Gaussian function *anisotropic*. An example is the pointspreadfunction of an astigmatic eye, where differences in curvature of the cornea/lens in different directions occur. This show an anisotropic Gaussian with anisotropy ratio of 2 ($\sigma_x / \sigma_y = 2$):

```
gauss[x_, y_, σx_, σy_] := ────────── Exp[-( ──── + ──── )];
                            2 π σx σy        2 σx²   2 σy²

σx = 2; σy = 1; Block[{$DisplayFunction = Identity},
 p1 =
  DensityPlot[gauss[x, y, σx, σy], {x, -10, 10}, {y, -10, 10}, PlotPoints -> 50];
 p2 = Plot3D[gauss[x, y, σx, σy], {x, -10, 10}, {y, -10, 10}, Shading -> True];
 p3 = ContourPlot[gauss[x, y, σx, σy], {x, -5, 5}, {y, -10, 10}]];
Show[GraphicsArray[{p1, p2, p3}], ImageSize -> 500];
```
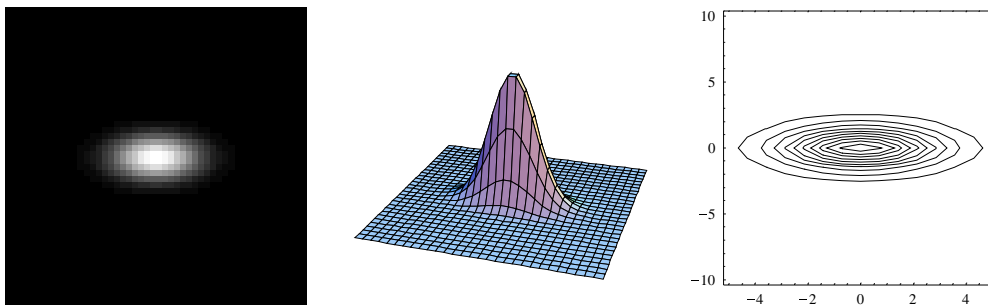

Figure 3.12. An anisotropic Gaussian kernel with anisotropy ratio $\sigma_x / \sigma_y = 2$ in three appearances. Left: `DensityPlot`, middle: `Plot3D`, right: `ContourPlot`.

3.10 The Fourier transform of the Gaussian kernel

We will regularly do our calculations in the Fourier domain, as this often turns out to be analytically convenient or computationally efficient. The basis functions of the Fourier transform $\mathcal{F}$ are the sinusoidal functions $e^{i\omega x}$. The definitions for the Fourier transform and its inverse are:

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

the inverse Fourier transform: $\mathcal{F}^{-1}\{F(\omega)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) \, e^{-i\omega x} \, d\omega$

```mathematica
σ =.; ℱgauss[ω_, σ_] = Simplify[
    1/√(2 π) Integrate[ 1/(σ √(2 π)) Exp[- x^2/(2 σ^2)] Exp[I ω x], {x, -∞, ∞}], {σ > 0, Im[σ] == 0}]
```

$$\frac{e^{-\frac{1}{2}\sigma^2 \omega^2}}{\sqrt{2\pi}}$$

The Fourier transform is a standard *Mathematica* command:

```mathematica
Simplify[FourierTransform[gauss[x, σ], x, ω], σ > 0]
```

$$\frac{e^{-\frac{1}{2}\sigma^2 \omega^2}}{\sqrt{2\pi}}$$

So the Fourier transform of the Gaussian function is again a Gaussian function, but now of the frequency ω. The Gaussian function is the *only* function with this property. Note that the scale σ now appears as a multiplication with the frequency. We recognize a well-known fact: a smaller kernel in the spatial domain gives a wider kernel in the Fourier domain, and vice versa. Here we plot 3 Gaussian kernels with their Fourier transform beneath each plot:

```mathematica
Block[{$DisplayFunction = Identity},
  p1 = Table[Plot[gauss[x, σ], {x, -10, 10}, PlotRange -> All,
      PlotLabel -> "gauss[x," <> ToString[σ] <> "]"], {σ, 1, 3}];
  p2 = Table[Plot[ℱgauss[ω, σ], {ω, -3, 3}, PlotRange -> All,
      PlotLabel -> "ℱgauss[x," <> ToString[σ] <> "]"], {σ, 1, 3}]];
Show[GraphicsArray[{p1, p2}], ImageSize -> {450, 200}];
```
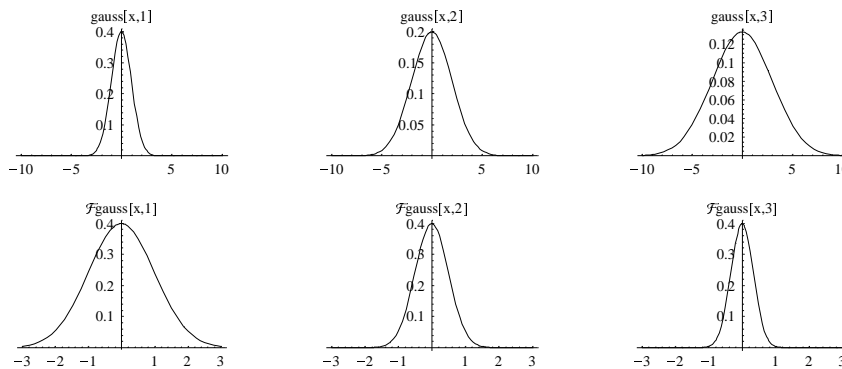
Figure 3.13. Top row: Gaussian function at scales $\sigma=1$, $\sigma=2$ and $\sigma=3$. Bottom row: Fourier transform of the Gaussian function above it. Note that for wider Gaussian its Fourier transform gets narrower and vice versa, a well known phenomenon with the Fourier transform.

There are many names for the Fourier transform $\mathcal{F}\,g(\omega;\sigma)$ of $g(x;\sigma)$: when the kernel $g(x;\sigma)$ is considered to be the point spread function, $\mathcal{F}\,g(\omega;\sigma)$ is referred to as the *modulation transfer function*. When the kernel $g(x;\sigma)$ is considered to be a signal, $\mathcal{F}\,g(\omega;\sigma)$ is referred to as the *spectrum*. When applied to a signal, it operates as a bandpass *filter*. Let us plot the spectra of a series of such filters (with a logarithmic increase in scale) on double logarithmic paper:

```
scales = N[Table[ Exp[t / 3], {t, 0, 8}]]

{1., 1.39561, 1.94773, 2.71828, 3.79367, 5.29449, 7.38906, 10.3123, 14.3919}

spectra =
  LogLinearPlot[𝓕gauss[ω, #], {ω, .01, 10}, DisplayFunction -> Identity] & /@
    scales;
Show[spectra, DisplayFunction -> $DisplayFunction, AspectRatio -> .4,
  PlotRange -> All, AxesLabel -> {"ω", "Amplitude"}, ImageSize -> 500];
```
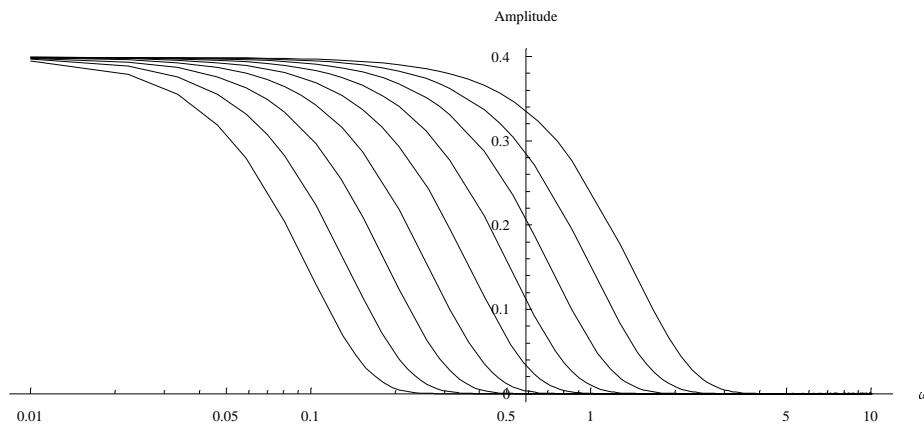

Figure 3.14. Fourier spectra of the Gaussian kernel for an exponential range of scales $\sigma = 1$ (most right graph) to $\sigma = 14.39$ (most left graph). The frequency ω is on a logarithmic scale. The Gaussian kernels are seen to act as *low-pass* filters.

Due to this behaviour the role of receptive fields as bandpass filters has long persisted. But the retina does *not* measure a Fourier transform of the incoming image, as we will discuss in the chapters on the visual system (chapters 6-9).

3.11 The Diffusion Equation

The Gaussian function is the solution of several differential equations.

It is the solution of $\frac{d\,y}{d\,x} = \frac{y(\mu-x)}{\sigma^2}$, because $\frac{d\,y}{y} = \frac{(\mu-x)}{\sigma^2}\,d\,x$, from which we find by integration $\ln\left(\frac{y}{y_0}\right) = -\frac{(\mu-x)^2}{2\,\sigma^2}$ and thus $y = y_0\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. QED.

It is the solution of the linear diffusion equation, $\frac{\partial L}{\partial t} = \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} = \Delta\,L$. This is a partial differential equation, stating that the first derivative of the (luminance) function $L(x,y)$ to the parameter t (time, or variance) is equal to the sum of the second order spatial derivatives. The right hand side is also known as the

Laplacian (indicated by Δ for any dimension, we call Δ the *Laplacian operator*), or the trace of the Hessian matrix of second order derivatives:

$$\mathtt{hessian2D} = \begin{pmatrix} \mathtt{L_{xx}} & \mathtt{L_{xy}} \\ \mathtt{L_{xy}} & \mathtt{L_{yy}} \end{pmatrix}; \ \mathtt{Tr[hessian2D]}$$

$\mathtt{L_{xx} + L_{yy}}$

$$\mathtt{hessian3D} = \begin{pmatrix} \mathtt{L_{xx}} & \mathtt{L_{xy}} & \mathtt{L_{xz}} \\ \mathtt{L_{yx}} & \mathtt{L_{yy}} & \mathtt{L_{yz}} \\ \mathtt{L_{zx}} & \mathtt{L_{yz}} & \mathtt{L_{zz}} \end{pmatrix}; \ \mathtt{Tr[hessian3D]}$$

$\mathtt{L_{xx} + L_{yy} + L_{zz}}$

The diffusion equation $\frac{\partial u}{\partial t} = \Delta u$ is one of some of the most famous differential equations in physics. It is often referred to as the *heat equation*. It belongs in the row of other famous equations like the Laplace equation $\Delta u = 0$, the wave equation $\frac{\partial^2 u}{\partial t^2} = \Delta u$ and the Schrödinger equation $\frac{\partial u}{\partial t} = i \, \Delta u$.

The diffusion equation $\frac{\partial u}{\partial t} = \Delta u$ is a *linear* equation. It consists of just linearly combined derivative terms, no nonlinear exponents or functions of derivatives.

The diffused entity is the intensity in the images. The role of time is taken by the variance $t = 2\,\sigma^2$. The intensity is diffused over time (over scale) in all directions in the same way (this is called *isotropic*). E.g. in 3D one can think of the example of the intensity of an inkdrop in water, diffusing in all directions.

The diffusion equation can be derived from physical principles: the luminance can be considered a *flow*, that is pushed away from a certain location by a force equal to the gradient. The divergence of this gradient gives how much the total entity (luminance in our case) diminishes with time.

```
<< Calculus`VectorAnalysis`
SetCoordinates[Cartesian[x, y, z]];

Div[ Grad[L[x, y, z]]]
```

$\mathtt{L^{(0,0,2)}[x, y, z] + L^{(0,2,0)}[x, y, z] + L^{(2,0,0)}[x, y, z]}$

A very important feature of the diffusion process is that it satisfies a *maximum principle* [Hummel1987b]: the amplitude of local maxima are always decreasing when we go to coarser scale, and vice versa, the amplitude of local minima always increase for coarser scale. This argument was the principal reasoning in the derivation of the diffusion equation as the generating equation for scale-space by Koenderink [Koenderink1984a].

3.12 Summary of this chapter

The normalized Gaussian kernel has an area under the curve of unity, i.e. as a filter it does not multiply the operand with an accidental multiplication factor. Two Gaussian functions can be cascaded, i.e. applied consecutively, to give a Gaussian convolution result which is equivalent to a kernel with the variance equal to the sum of the variances of the constituting Gaussian kernels. The spatial parameter normalized over scale is called the dimensionless 'natural coordinate'.

The Gaussian kernel is the 'blurred version' of the Delta Dirac function, the cumulative Gaussian function is the Error function, which is the 'blurred version' of the Heavyside stepfunction. The Dirac and Heavyside functions are examples of *generalized functions*.

The Gaussian kernel appears as the limiting case of the Pascal Triangle of binomial coefficients in an expanded polynomial of high order. This is a special case of the central limit theorem. The central limit theorem states that any finite kernel, when repeatedly convolved with itself, leads to the Gaussian kernel.

Anisotropy of a Gaussian kernel means that the scales, or standard deviations, are different for the different dimensions. When they are the same in all directions, the kernel is called *isotropic*.

The Fourier transform of a Gaussian kernel acts as a low-pass filter for frquencies. The cut-off frequency depends on the scale of the Gaussian kernel. The Fourier transform has the same Gaussian shape. The Gaussian kernel is the *only* kernel for which the Fourier transform has the same shape.

The diffusion equation describes the expel of the flow of some quantity (intensity, tempreature) over space under the force of a gradient. It is a second order parabolic differential equation. The linear, isotropic diffusion equation is the generating equation for a scale-space. In chapter 19 we will encounter a wealth on nonlinear diffusion equations.