8. Differentiation and regularization

8.1 Regularization

Regularization is the technique to make data behave well when an operator is applied to them. Such data could e.g. be functions, that are impossible or difficult to differentiate, or discrete data where a derivative seems to be not defined at all. In scale-space theory, we realize that we do physics. This implies that when we consider a system, a small variation of the input data should lead to small change in the output data.
Differentiation is a notorious function with 'bad behaviour'. Here are some examples of non-differentiable functions:

```
<< FrontEndVision`FEV`;
Block[{$DisplayFunction = Identity},
  p1 = Plot[Exp[-Abs[x]], {x, -2, 2}, PlotStyle -> Thickness[.01]];
  p2 = Plot[UnitStep[x - 1], {x, -2, 5}, PlotStyle -> Thickness[.01]];
  p3 = Plot[Floor[4 Sin[x]], {x, 0, 4 π}, PlotStyle -> Thickness[.01]];
  p4 = ListPlot[Table[Sin[4 π / √i], {i, 2, 40}], PlotStyle -> PointSize[.02]]];
Show[GraphicsArray[{{p1, p2}, {p3, p4}}], ImageSize -> 300];
```

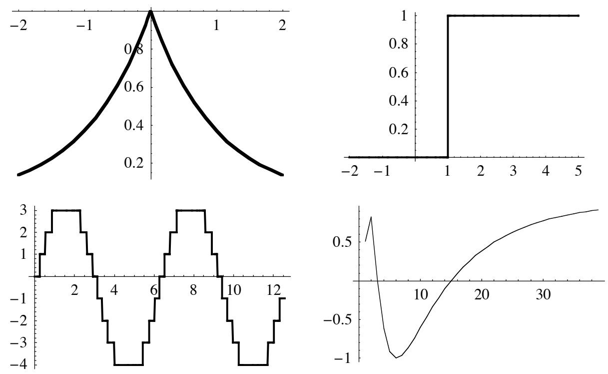

Figure 8.1 Some functions that can not be differentiated.

In mathematical terms it is said that the operation of differentiation is *ill-posed*, the opposite of *well-posed*. Jacques Hadamard (1865–1963) [Hadamard1902] stated the conditions for well-posedness:

- The solution must exist;
- The solution must be uniquely determined;
- The solution must depend continuously on the initial or boundary data.

The first two requirements state that there is one and only one solution. The third requirement assures that if the initial or boundary data change slightly, it should also have a limited impact on the solution. In other words, *the solution should be stable*.

Regularization is a hot topic. Many techniques are developed to regularize the data, each based on a constraint on how one wishes the data to behave without sacrificing too much. Well known and abundantly applied examples are:

- *smoothing* the data, convolution with some extended kernel, like a 'running average filter' (e.g. $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$) or the Gaussian;
- *interpolation*, by a polynomial (multidimensional) function;
- *energy minimization*, of a cost function under constraints [Mumford1985a, Mumford1989a, Mumford1994a];
- *fitting a function* to the data (the best known examples are splines, i.e. polynomials fitting a curve or a surface up to some order [DeBoor1978]. The *cubic splines* are named so because they fit to third order, x^3 ;
- *graduated convexity* [Blake1987];
- *deformable templates* ('snakes') [McInerney1996];
- *thin plates* or *thin plate splines* [Bookstein1989] (see also mathworld.wolfram.com/ThinPlateSpline.html);
- *Tikhonov regularization*, discussed in detail in the next section.

However, smoothing before the differentiation does *not* solve the ill-posedness problem. The crucial difference between the approaches above and scale-space theory is that the first methods *change the data*, your most valuable source of information, before the operation (e.g. differentiation) is applied. The derivative is taken of the regularized data.

When we recall the importance of doing a measurement uncommitted, we surely should *not* modify our data in any way. We need a regularization of the operator, not the operand. Actually, the only control we have when we do a measurement is in our measurement device. There we can change the size, location, orientation, sensitivity profiles etc. of our filtering kernels. That is something completely different from the methods described above. It is one of the cornerstones in scale-space theory that the *only control allowed* is in the filters. As such, scale-space theory can be considered the 'theory of apertures'.

8.2 Regular tempered distributions and test functions

The formal mathematical method to solve the problems of ill-posed differentiation was given by Laurent Schwartz [Schwartz1951] (see figure 8.2) as was noted by Florack [Florack1994a]. The following is adapted from [Niessen1997a]. A *regular tempered distribution* associated with an image is defined by the action of a *smooth test function* on the image. Smooth is here used in the mathematical definition, i.e. infinitely differentiable, or C^∞.

The class of smooth test functions ϕ (also called the *Schwartz space* $\mathcal{S}(\mathbb{R}^D)$) is large. It comprises all smooth functions that decrease sufficiently fast to zero at the boundaries. They are mathematically defined as the functions ϕ that are C^∞ and whose derivative to whatever order goes faster to zero to any polynomial. Mathematically stated:

$$\phi \in \mathcal{S}(\mathbb{R}^D) \iff \phi \in C^\infty(\mathbb{R}^D) \wedge \sup \| x^m \partial_{i_1 \ldots i_n} \phi(x) \| < \infty$$

for all m and n. As we consider any dimension here, m and n are multi-indices.

Let us give an example of such a function. The Gaussian kernel has the required properties, and belongs to the class of smooth test functions. E.g. its goes faster to zero then e.g. a 13$^{\text{th}}$ order polynomial:

```
Limit[x^13 Exp[-x^2], x → ∞]
```

```
0
```

Gaussian derivatives are also smooth test functions. Here is an example for the third order:

```
Limit[x^13 ∂_{x,x,x} Exp[-x^2], x → ∞]
```

```
0
```

The reason that the function e^{-x^2} suppresses any polynomial function, is that a series expansion leads to polynomial terms of any desired order:

```
Series[Exp[-x^2], {x, 0, 15}]
```

$$1 - x^2 + \frac{x^4}{2} - \frac{x^6}{6} + \frac{x^8}{24} - \frac{x^{10}}{120} + \frac{x^{12}}{720} - \frac{x^{14}}{5040} + O[x]^{16}$$

▲ Task 8.1 Find a number of functions that fulfill the criteria for being a member of the class of smooth test functions, i.e. a member of the Schwartz space.

A *regular tempered distribution* T_L associated with image $L(x)$ is defined as:

$$T_L = \int_{-\infty}^{\infty} L(x)\, \phi(x)\, dx$$

The testfunction 'samples' the image, and returns a scalar value. The derivative of a regular tempered distribution is defined as:

$$\partial_{i_1 \ldots i_n} T_L = (-1)^n \int_{-\infty}^{\infty} L(x)\, \partial_{i_1 \ldots i_n} \phi(x)\, dx$$

Thus the image is now 'sampled' with the derivative of the test function. This is the key result of Schartz' work. It is now possible to take a derivative of all the nasty functions we gave as examples above. We can now also take derivatives of our discrete images. But we still need to find the testfunction ϕ. Florack [Florack1994b] found the solution in demanding that the derivative should be a new observable, i.e. that the particular test function can be interpreted as a linear filter.
The choice for the filter is then determined by physical considerations, and we did so in chapter 2 where we derived the Gaussian kernel and all its partial derivatives as the causal non-committed kernels for an observation.

We saw before that the Gaussian kernel and its derivatives are part of the Schwartz space.

```
Show[Import["Laurent Schwartz.jpg"], ImageSize -> 150];
```

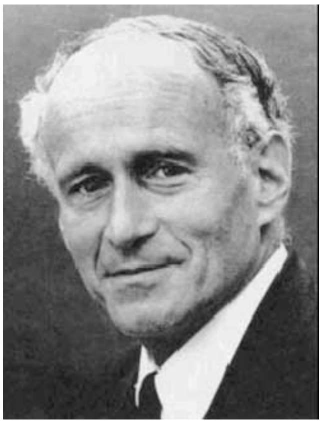

Figure 8.2 Laurent Schwartz (1915 -). Schwartz spent the year 1944-45 lecturing at the Faculty of Science at Grenoble before moving to Nancy where he became a professor at the Faculty of Science. It was during this period of his career that he produced his famous work on the theory of distributions. Harald Bohr presented a Fields Medal to Schwartz at the International Congress in Harvard on 30 August 1950 for his work on the theory of distributions. Schwartz has received a long list of prizes, medals and honours in addition to the Fields Medal. He received prizes from the Paris Academy of Sciences in 1955, 1964 and 1972. In 1972 he was elected a member of the Academy. He has been awarded honorary doctorates from many universities including Humboldt (1960), Brussels (1962), Lund (1981), Tel-Aviv (1981), Montreal (1985) and Athens (1993).

So we can now define a well-posed derivative of an image $L(x)$ in the proper 'Schwartz way':

$$\partial_{i_1 \ldots i_n} L(x) = (-1)^n \int_{-\infty}^{\infty} L(y)\, \partial_{i_1 \ldots i_n} \phi(y, x)\, dy$$

We have no preference for a particular point where we want this 'sampling' to be done, so we have linear shift invariance: $\phi(y; x) = \phi(y - x)$. We now get the result that taking the derivative of an image is equivalent with the *convolution* of the image with the *derivative of the test function*:

$$\partial_{i_1 \ldots i_n} L(x) = \int_{-\infty}^{\infty} L(y)\, \partial_{i_1 \ldots i_n} \phi(y - x)\, dy$$

The set of test functions is here the Gaussian kernel and all its partial derivatives. We also see now the relation with *receptive fields*: they are the Schwartz test functions for the visual system. They take care of making the differentiation regularized, well posed. Here is a comparison list:

Mathematics	⇔	Smooth test function
Computer vision	⇔	Kernel, filter
Biological vision	⇔	Receptive field

Of course, if we relax or modify the constraints of chapter 2, we might get other kernels (such as the Gabor kernels if we are confine our measurement to just a single spatial frequency). As long as they are part of the Schwartz space we get well posed derivatives.

The key point in the reasoning here is that there is no attempt to smooth the data and to take the derivative of the smoothed result, but that the differentiation is done *prior to* the smoothing. Differentiation is transferred to the filter. See for a full formal treatment on Schwartz theory for images the papers by Florack [Florack1992a, Florack1994a, Florack1996b, Florack1997a].

The theory of distribution is a considerable broadening of the differential and integral calculus. Heaviside and Dirac had generalized the calculus with specific applications in mind. These, and other similar methods of formal calculation, were not, however, built on an abstract and rigorous mathematical foundation. Schwartz's development of the theory of distributions put methods of this type onto a sound basis, and greatly extended their range of application, providing powerful tools for applications in numerous areas.

8.3 An example of regularization

The classical example of the regularization of differentiation by the Gaussian derivative is the signal with a high-frequency disturbance $\epsilon \cos(\omega x)$. Here ϵ is a small number, and ω a very high frequency.

We compare the mathematical derivative with convolution with the Gaussian derivative. First we calculate the mathematical derivative:

```
∂x (L[x] + ϵ Cos[ω x])
```

```
-ϵ ω Sin[x ω] + L′[x]
```

For large ω the disturbance becomes very large. The disturbance can be made arbitrarily small, provided that the derivative of the signal is computed at a sufficiently coarse scale σ in scale-space:

$$\texttt{gx[x\_, σ\_] := } \partial_x \left(\frac{1}{\sqrt{2\pi}\,\sigma} \mathrm{E}^{-\frac{x^2}{2\sigma^2}} \right);$$

$$\texttt{Simplify}\left[\int_{-\infty}^{\infty} \epsilon\, \texttt{Cos}[\omega\,(x-\alpha)]\, \texttt{gx}[\alpha, \sigma]\, d\alpha,\ \{\omega > 0, \sigma > 0\}\right]$$

$$-e^{-\frac{1}{2}\sigma^2\omega^2}\,\epsilon\,\omega\,\texttt{Sin}[x\,\omega]$$

8.4 Relation regularization ⟺ Gaussian scale-space

When data are regularized by one of the methods above that 'smooth' the data, choices have to be made as how to fill in the 'space' in between the data that are not given by the original data. In particular, one has to make a choice for the order of the spline, the order of fitting polynomial function, the 'stiffness' of the physical model etc. This is in essence the same choice as the scale to apply in scale-space theory. In fact, it is becoming clear that there are striking analogies between scale-space regularization and other means of regularization.

An essential result in scale-space theory was shown by Mads Nielsen. He proved that the well known and much applied method of regularization as proposed by Tikhonov and Arsenin [Tikhonov and Arsenin 1977] (often called 'Tikhonov regularization') is essentially equivalent to convolution with a Gaussian kernel [Nielsen1996b, Nielsen1997a, Nielsen1997b]. Tikhonov and Arsenin tried to regularize functions with a formal mathematical approach from variational calculus called the method of Euler-Lagrange equations.

This method studies a function of functions, and tries to find the minimum of that function given a set of constraints. Their proposed formulation was the following: Make a function $E(g) = \int_{-\infty}^{\infty} (f-g)^2 \, dx$ and minimize this function for g. f and g are both functions of x. The function g must become the regularized version of f, and the problem is to find a function g such that it deviates as little as possible from f. The difference with f is taken with the so-called 2-norm, $(f-g)^2$, and we like to find that g for which this squared difference is minimal, *given a constraint*.

This constraint is the following: we also like the first derivative of g to x (g_x) to behave well, i.e. we require that when we integrate the square of g_x over its total domain we get a finite result. Mathematically you see such a requirement sometimes announced as that the functions are 'mapped into a Sobolev space', which is the space of square integrable functions.

The method of the Euler-Lagrange equations specifies the construction of an equation for the function to be minimized where the constraints are added with a set of constant factors λ_i, one for each constraint, the so-called Lagrange multipliers. In our case: $E(g) = \int_{-\infty}^{\infty} (f-g)^2 + \lambda_1 \, {g_x}^2 \, dx$. The *functional* $E(g)$ is called the Lagrangian and is to be minimized with respect to g, i.e. we require $\frac{dE}{dg} = 0$.

In the Fourier domain the calculations become a lot more compact. $\tilde{f}$ and $\tilde{g}$ are denoted as the Fourier transforms of f and g respectively. The famous theorem by Parceval states that the Fourier transform of the square of a function is equal to the square of the function itself. Secondly, we need the result that the Fourier transform of a derivative of a function is the Fourier transform of that function multiplied with the factor $-i\omega$. So $\mathcal{F}\left(\frac{\partial g(x)}{\partial x}\right) = -i\omega\,\mathcal{F}(g(x))$ where $\mathcal{F}$ denotes the Fourier transform.

For the square of such a derivative we get the factor ω^2, because the square of a complex function z is the function z multiplied with its *conjugate* ($i \to -i$), denoted as z^*, so

$z^2 = z\,z^*$ which gives the factor $(-i\,\omega)\,(-i\,\omega)^* = \omega^2$. So finally, because the Fourier transform is a linear operation, we get for the Lagrangian $\tilde{E}$:

$$\tilde{E}(\tilde{g}) = \mathcal{F}\,\{\int_{-\infty}^{\infty}(f-g)^2 + \lambda_1\,g_x{}^2\,d\omega\} = \int_{-\infty}^{\infty}(\tilde{f}-\tilde{g})^2 + \lambda_1\,\tilde{g}_x{}^2\,d\omega = \int_{-\infty}^{\infty}(\tilde{f}-\tilde{g})^2 + \lambda_1\,\omega^2\,\tilde{g}^2\,d\omega$$

$$\frac{d\tilde{E}(\tilde{g})}{d\tilde{g}} = 2\,(\tilde{f}-\tilde{g})^2 + 2\,\lambda_1\,\omega^2\,\tilde{g} = 0, \text{ so } (\tilde{f}-\tilde{g}) + \lambda_1\,\omega^2\,\tilde{g} = 0 \leftrightarrow \tilde{g} = \frac{1}{1+\lambda_1\,\omega^2}\,\tilde{f}.$$

The regularized function $\tilde{g}$ is (in the Fourier domain, only taking into account a constraint on the first derivative) seen to be the product of two functions, $\frac{1}{1+\lambda_1\,\omega^2}$ and $\tilde{f}$, which product is a convolution in the spatial domain.

The first result is that this first order regularization can be implemented with a spatial filtering operation. The filter $\frac{1}{1+\lambda_1\,\omega^2}$ in the spatial domain looks like this:

```
g1[x_] = InverseFourierTransform[1/(1 + λ1 ω^2), ω, x] // Simplify
```

$$e^{-\text{Abs}[x]}\sqrt{\frac{\pi}{2}}$$

```
λ1 = 1; Plot[g1[x], {x, -3, 3}, ImageSize -> 150];
```

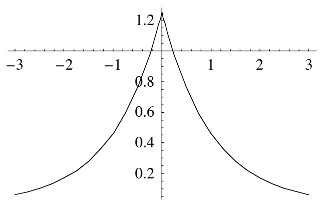

Figure 8.3 Filter function proposed by Castan et al. [Castan1990].

It is precisely the function proposed by Castan et al. [Castan1990]. The derivative of this filter is not well defined, as we can clearly see.

This is a first result for the inclusion of the constraint for the first order derivative. However, we like our function $\tilde{g}$ to be regularized with *all* derivatives behaving nicely, i.e. square integrable. When we add the constraint of the second derivative, we get two Lagrangian multipliers, λ_1 and λ_2:

$\tilde{E}(\tilde{g}) = \int_{-\infty}^{\infty}(\tilde{f}-\tilde{g})^2 + \lambda_1\,\tilde{g}_x{}^2 + \lambda_2\,\tilde{g}_{xx}\,d\omega = \int_{-\infty}^{\infty}(\tilde{f}-\tilde{g})^2 + \lambda_1\,\omega^2\,\tilde{g}^2 + \lambda_2\,\omega^4\,\tilde{g}^2\,d\omega$ and we find in a similar way for $\tilde{g}$:

$$\frac{dE(\tilde{g})}{d\tilde{g}} = 2\,(\tilde{f}-\tilde{g})^2 + 2\,\lambda_1\,\omega^2\,\tilde{g} + \lambda_2\,\omega^4\,\tilde{g} = 0 \leftrightarrow \tilde{g} = \frac{1}{1+\lambda_1\,\omega^2+\lambda_2\,\omega^4}\,\tilde{f}.$$

This is a regularization involving well behaved derivatives of the filtered $\tilde{f}$ to second order. This filter was proposed by Deriche [Deriche1987], who made a one-parameter family of this filter by setting a relation between the $\lambda' s$: $\lambda_1 = 2\sqrt{\lambda_2}$. The dimensions of λ_1 and λ_2 are correctly treated by this choice. When we look at the Taylor series expansion of the

Gaussian kernel in the Fourier domain, we see that his choice is just the truncated Gaussian to second order:

```
Simplify[FourierTransform[1/(Sqrt[2 π] σ) E^(-x^2/(2 σ^2)), x, ω], σ > 0]
```

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

```
Series[E^(1/2 σ^2 ω^2), {ω, 0, 10}]
```

$$1+\frac{\sigma^2\omega^2}{2}+\frac{\sigma^4\omega^4}{8}+\frac{\sigma^6\omega^6}{48}+\frac{\sigma^8\omega^8}{384}+\frac{\sigma^{10}\omega^{10}}{3840}+O[\omega]^{11}$$

Here is how Deriche's filters look like (see figure 8.4):

```
λ1 =.; λ2 =.; g2[x_] =
 InverseFourierTransform[1/(1 + 2 Sqrt[λ2] ω^2 + λ2 ω^4), ω, x] // FullSimplify
```

$$\frac{1}{2\sqrt{\lambda 2}}\left(\sqrt{\frac{\pi}{2}}\left(\text{Cosh}\left[\frac{x}{\lambda 2^{1/4}}\right](\lambda 2^{1/4}+x\,\text{Sign}[x])-(x+\lambda 2^{1/4}\,\text{Sign}[x])\,\text{Sinh}\left[\frac{x}{\lambda 2^{1/4}}\right]\right)\right)$$

and the graph, for $\lambda_2 = 1$:

```
λ2 = 1; Plot[g2[x], {x, -4, 4}, ImageSize -> 150];
```

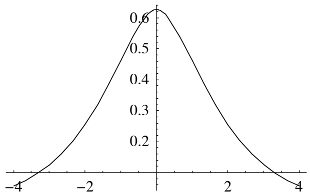

Figure 8.4 Filter function proposed by Deriche [Deriche1987].

From the series expansion of the Gaussian, and the induction from the lower order regularization, we may develop the suspicion that by adding the constraints for *all* derivatives to behave well, we get the infinite series

$\tilde{g} = \frac{1}{1+\lambda_1\,\omega^2+\lambda_2\,\omega^4+\cdots+\lambda_n\,\omega^{2n}}\,\tilde{f} = \tilde{h}\,\tilde{f}$ for $n \to \infty$.

Nielsen showed that the filter $\tilde{h}$ is the Gaussian kernel indeed. The reasoning goes as follows. We have an infinite number of unknowns here, the λ_n's, so we need to come up with an additional constraint that gives us just as many equations, so we can solve uniquely for this system of equations. We have just as many terms ω^{2n}, so we look for a constraint on them. It is found in the requirement that we want *scale invariance* for the filters, i.e. we want two filters $h(s)$ and $h(t)$ to *cascade*, i.e. $h(s \oplus t) = h(s) \otimes h(t)$ where $\otimes$ is the convolution operator. The parameters s and t are the scales of the filters. The operator $\oplus$ stands for the

summation at any norm, which is a compact writing for the definition $a \oplus b = (a+b)^p = a^p + b^p$. It turns out that for $p = 2$ we have the regular addition of the variances of the scales, as we have seen now several times before, due to the requirement of separability and Euclidean metric.
Implementing the cascading requirement for the first order:

$\frac{1}{1+\lambda_1(s\oplus t)\,\omega^2} = \frac{1}{1+\lambda_1(s)\,\omega^2} \cdot \frac{1}{1+\lambda_1(t)\,\omega^2}$ giving
$1 + \lambda_1(s \oplus t)\,\omega^2 = 1 + \lambda_1(s)\,\omega^2 + \lambda_1(t)\,\omega^2 + \lambda_1(t)\,\lambda_1(s)\,\omega^4$ and
$\lambda_1(s \oplus t) = \lambda_1(s) + \lambda_1(t) + \lambda_1(t)\,\lambda_1(s)\,\omega^2$.

We equal the coefficients of powers of ω both sides, so for ω^0 we find $\lambda_1(s \oplus t) = \lambda_1(s) + \lambda_1(t)$ which means that λ_1 must be a linear function of scale, $\lambda_1 = \alpha\, s$. Now for the second order:

$$\frac{1}{1+\lambda_1(s\oplus t)\,\omega^2+\lambda_2(s\oplus t)\,\omega^4} = \frac{1}{1+\lambda_1(s)\,\omega^2+\lambda_2(s)\,\omega^4} \cdot \frac{1}{1+\lambda_1(t)\,\omega^2+\lambda_2(t)\,\omega^4}$$

giving
$$\lambda_1(s \oplus t)\,\omega^2 + \lambda_2(s \oplus t)\,\omega^4 = \lambda_1(s)\,\omega^2 + \lambda_2(s)\,\omega^4 + \lambda_1(t)\,\omega^2 + \lambda_1(t)\,\lambda_1(s)\,\omega^4 + \lambda_1(t)\,\lambda_2(s)\,\omega^6 + \lambda_2(t)\,\omega^4 + \lambda_2(t)\,\lambda_1(s)\,\omega^6 + \lambda_2(t)\,\lambda_2(s)\,\omega^8$$

and equating the coefficients for ω^4 on both sides:

$\lambda_2(s \oplus t) = \lambda_1(t)\,\lambda_1(s) + \lambda_2(t) + \lambda_2(t)$ from which dimension we see that λ_2 must be quadratic in scale, and $\lambda_2 = \frac{\alpha^2\, s^2}{2} = \frac{1}{2}\,{\lambda_1}^2$.

This reasoning can be extended to higher scale, and the result is that we get the following series:

g$\lambda_1 = \alpha\, s$, $\lambda_2 = \frac{1}{2}\,\alpha^2\, s^2$, $\lambda_3 = \frac{1}{2.4}\,\alpha^4\,\sigma^4$, $\lambda_4 = \frac{1}{2.4.6}\,\alpha^6\,\sigma^6$ etc.

We recall that the series expansion of the Gaussian function $E^{-\frac{1}{2}\sigma^2\,\omega^2}$ is, using

```
Series[E^(1/2 σ^2 ω^2), {ω, 0, 10}]
```

$$1 + \frac{\sigma^2\,\omega^2}{2} + \frac{\sigma^4\,\omega^4}{8} + \frac{\sigma^6\,\omega^6}{48} + \frac{\sigma^8\,\omega^8}{384} + \frac{\sigma^{10}\,\omega^{10}}{3840} + O[\omega]^{11}$$

$$\frac{1}{1+\frac{\sigma^2\,\omega^2}{2}+\frac{\sigma^4\,\omega^4}{2.4}+\frac{\sigma^6\,\omega^6}{2.4.6}+\frac{\sigma^8\,\omega^8}{2.4.6.8}+\frac{\sigma^{10}\,\omega^{10}}{2.4.6.8.10}+O[\omega]^{11}}$$

```
σ = 1.; Plot[Evaluate[InverseFourierTransform[E^(-1/2 σ^2 ω^2), ω, x]],
 {x, -4, 4}, ImageSize -> 300];
```

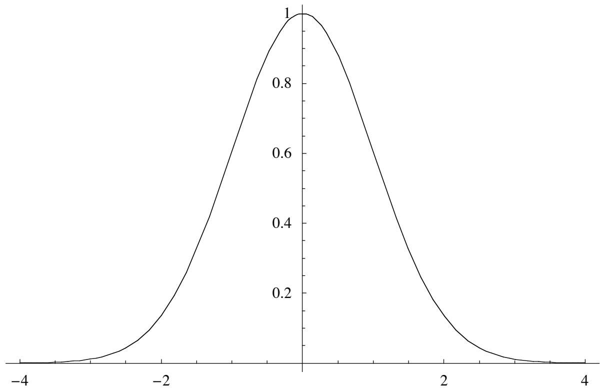

Figure 8.5 The Gaussian kernel ($\sigma = 1$).

When we take the arbitrary constant $\alpha = 1$, we get as the optimal regularization filter, where all derivatives are required to behave normal, precisely the Gaussian kernel! This important result is due to Nielsen [Nielsen1996b, Nielsen1997a, Nielsen1997b]. It has recently been proved by Radmoser, Scherzer and Weickert for a number of other regularization methods that they can be expressed as a Gaussian scale-space regularization [Radmoser1999a, Radmoser2000a].

▲ Task 8.2 Prove the equations for the coefficients λ_n in the section above by induction.

8.5 Summary of this chapter

Many functions can not be differentiated. A sampled image is such a function. The solution, due to Schwartz, is to *regularize* the data by convolving them with a smooth *test function*. Taking the derivative of this 'observed' function is then equivalent to convolving with the derivative of the test function. This is just what the receptive fields of the front-end visual system do: regularization and differentiation. It is one of the key results of scale-space theory.

A well know variational form of regularization is given by the so-called Tikhonov regularization: a functional is minimized in L^2 sense with the constraint of well behaving derivatives. It is shown in this chapter, with a reasoning due to Nielsen, that Tikhonov regularization with inclusion of the proper behaviour of *all* derivatives is essentially equivalent to Gaussian blurring.