The Gaussian Kernel, Regularization

The Gaussian Kernel

- The Gaussian kernel
- Normalization
- Cascade property, selfsimilarity
- The scale parameter
- Relation to generalized functions
- Separability
- Relation to binomial coefficients
- The Fourier transform of the Gaussian kernel
- Central limit theorem
- Anisotropy
- The diffusion equation

Taken from B. M. ter Haar Romeny,
Front-End Vision and Multi-scale Image Analysis,
Dordrecht, Kluwer Academic Publishers, 2003.
Chapter 3

The Gaussian Kernel

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

- 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*.
- The scale can only take positive values, $\sigma > 0$.
- The scale-dimension is *not* just another spatial dimension.

Normalization

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

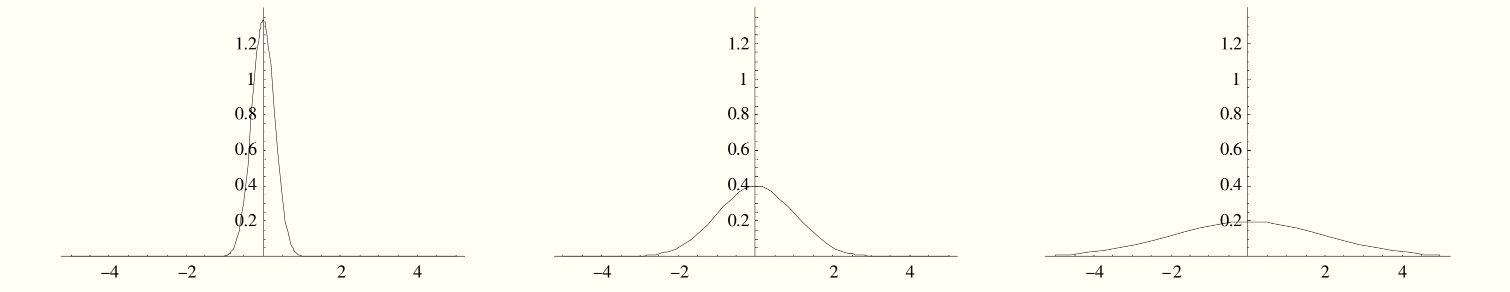
- The term in front of the Gaussian kernel is the normalization constant.

$$\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\, dx = \sqrt{2\pi}\,\sigma$$

- With the normalization constant this Gaussian kernel is a *normalized* kernel, i.e. its integral over its full domain is unity for every σ.

Normalization

- This means that increasing the σ of the kernel reduces the amplitude substantially.



- 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*.

Cascade property, self-similarity

- 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:

$$g_{\text{new}}(\vec{x};\sigma_1^2+\sigma_2^2) = g_1(\vec{x};\sigma_1^2) \otimes g_2(\vec{x};\sigma_2^2)$$

- 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.

The scale parameter

- In order to avoid the summing of squares, one often uses the following parameterization: $2\,\sigma^2 = t$

$$G_{\mathrm{ND}}(\vec{x},\, t) = \frac{1}{(\pi\, t)^{N/2}}\, e^{-\frac{x^2}{t}}$$

- To make the self-similarity of the Gaussian kernel explicit, we can introduce a new *dimensionless* (natural) spatial parameter:

$$\tilde{x} = \frac{x}{\sigma\sqrt{2}}$$

- … and obtain the *natural* Gaussian kernel

$$g_n(\tilde{x};\; t) = \frac{1}{(\pi\, t)^{N/2}}\, e^{-\tilde{x}^2}$$

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.

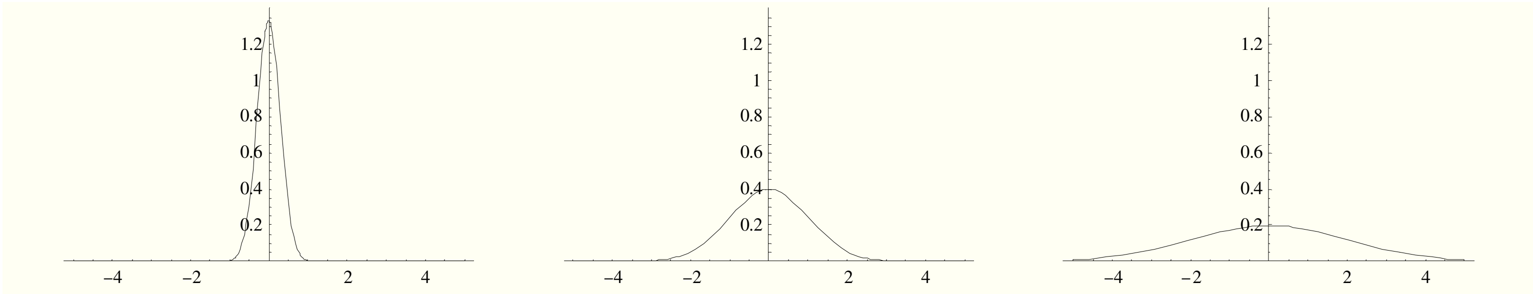
- Focus on some mathematical notions that are directly related to the sampling of values from functions and their derivatives at *selected* points.

- These mathematical functions are the *generalized functions*, i.e. the Dirac Delta-function, the Heaviside function and the error function.

Dirac delta function

- δ(x) is everywhere zero except in $x = 0$, where it has infinite amplitude and zero width; its area is unity.

$$\lim_{\sigma \downarrow 0} \left(\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} \right) = \delta(x)$$



The error function

- The integral of the Gaussian kernel from $-\infty$ to x is the *error function*, or *cumulative* Gaussian function:

$$\int_0^x \frac{1}{\sigma\sqrt{2\pi}} \operatorname{Exp}\left[-\frac{y^2}{2\sigma^2}\right] dy$$

- The result is $\frac{1}{2}\operatorname{Erf}\left[\frac{x}{\sqrt{2}\,\sigma}\right]$

  so re-parameterzing is needed: $x \to \frac{x}{\sigma\sqrt{2}}$

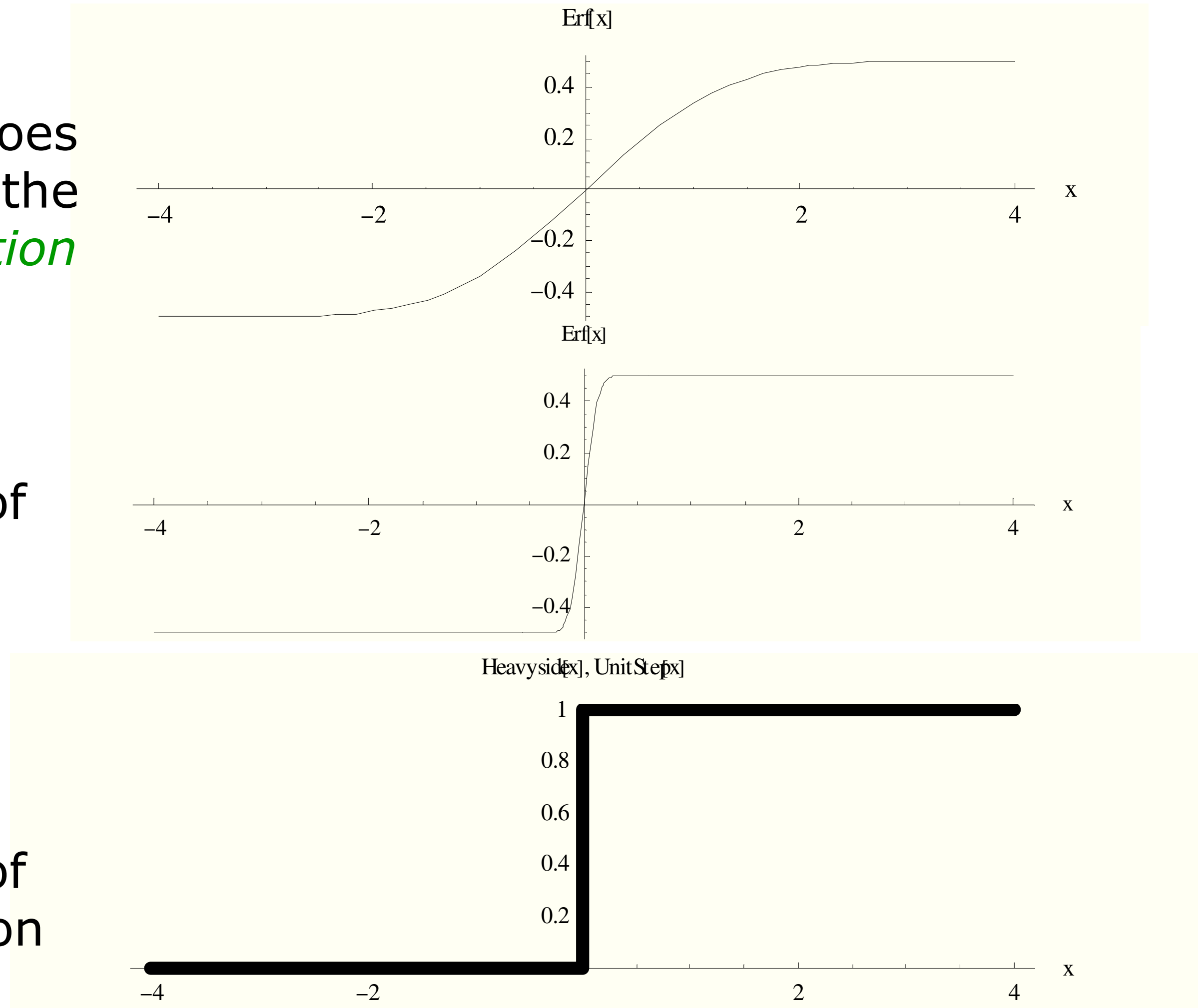
  -> natural coordinates!

The Heavyside function

- When the inner scale σ of the error function goes to zero, we get the *Heavyside function* or *unitstep function*.

- The derivative of the Heavyside function is the Delta-Dirac function.

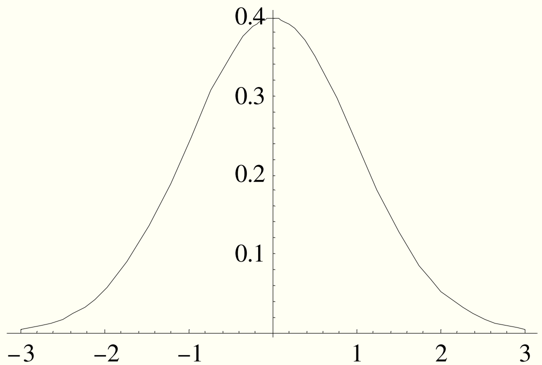
- The derivative of the error function is the Gaussian kernel.

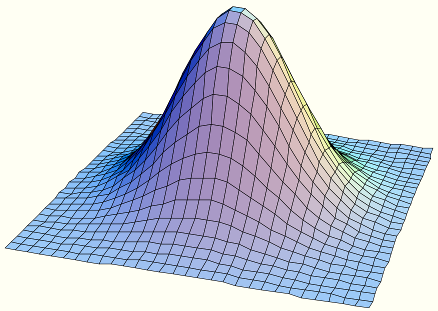


Separability

- The Gaussian kernel for dimensions higher than one, say *N*, can be described as a regular product of *N* one-dimensional kernels.

$$g_{2D}(x, y; \sigma^2) = g_{1D}(x; \sigma^2) + g_{1D}(y; \sigma^2)$$



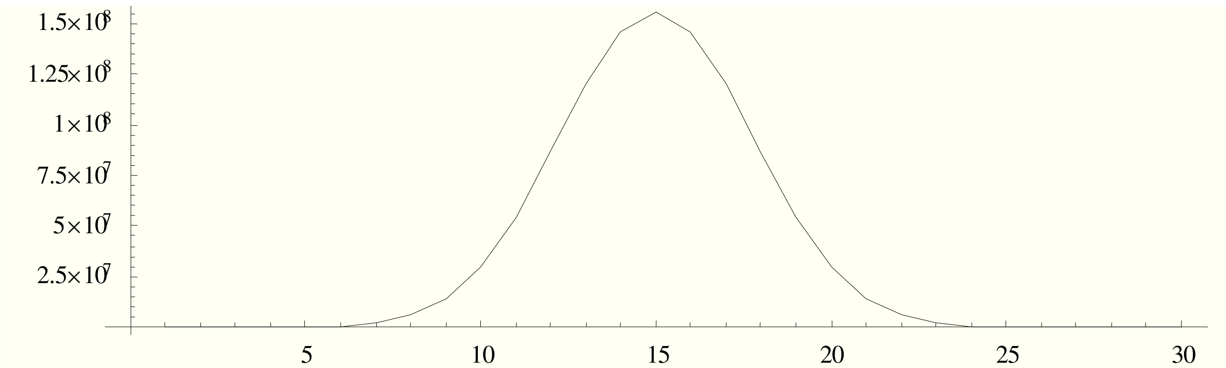


Relation to binomial coefficients

```
Expand[(x + y)^30]
```

$$x^{30} + 30\,x^{29}\,y + 435\,x^{28}\,y^2 + 4060\,x^{27}\,y^3 + 27405\,x^{26}\,y^4 + 142506\,x^{25}\,y^5 + 593775\,x^{24}\,y^6 + 2035800\,x^{23}\,y^7 + 5852925\,x^{22}\,y^8 + 14307150\,x^{21}\,y^9 + 30045015\,x^{20}\,y^{10} + 54627300\,x^{19}\,y^{11} + 86493225\,x^{18}\,y^{12} + 119759850\,x^{17}\,y^{13} + 145422675\,x^{16}\,y^{14} + 155117520\,x^{15}\,y^{15} + 145422675\,x^{14}\,y^{16} + 119759850\,x^{13}\,y^{17} + 86493225\,x^{12}\,y^{18} + 54627300\,x^{11}\,y^{19} + 30045015\,x^{10}\,y^{20} + 14307150\,x^9\,y^{21} + 5852925\,x^8\,y^{22} + 2035800\,x^7\,y^{23} + 593775\,x^6\,y^{24} + 142506\,x^5\,y^{25} + 27405\,x^4\,y^{26} + 4060\,x^3\,y^{27} + 435\,x^2\,y^{28} + 30\,x\,y^{29} + y^{30}$$

- The coefficients of this expansion are the *binomial coefficients* ('n over m')



The Fourier transform

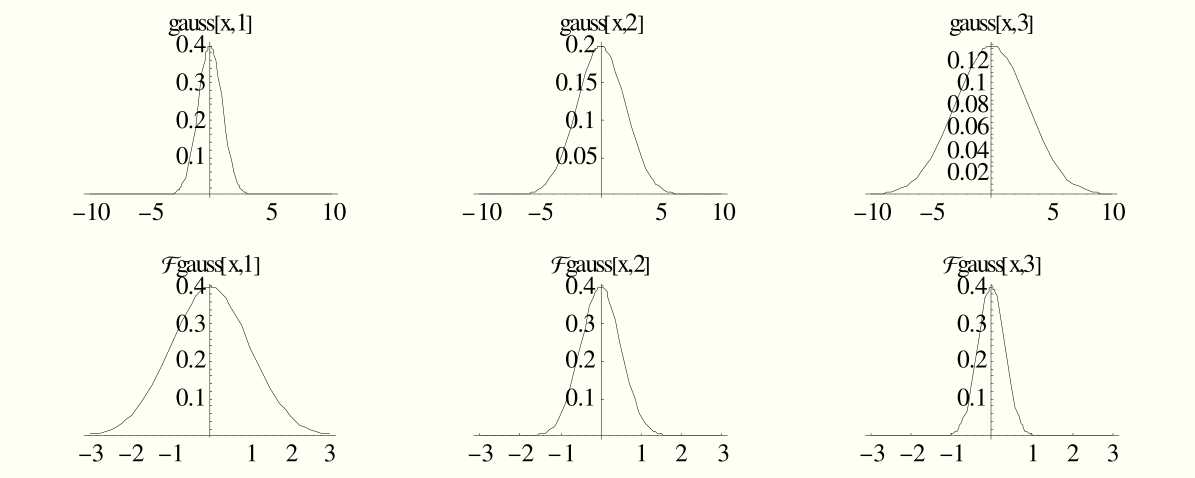
- the Fourier transform:

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

- the inverse Fourier transform:

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

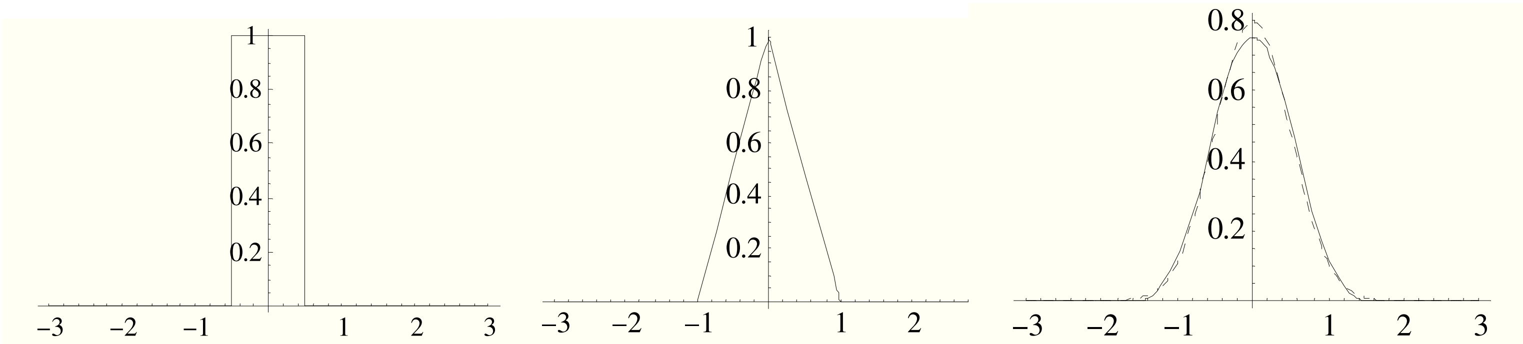
- The Fourier transform of the Gaussian function is again a Gaussian function, but now of the frequency ω.



Central limit theorem

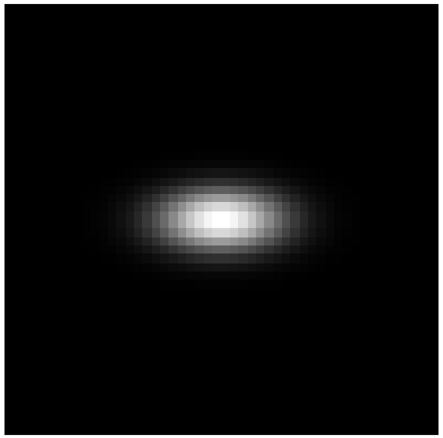
- The *central limit theorem*: any repetitive operator goes in the limit to a Gaussian function.

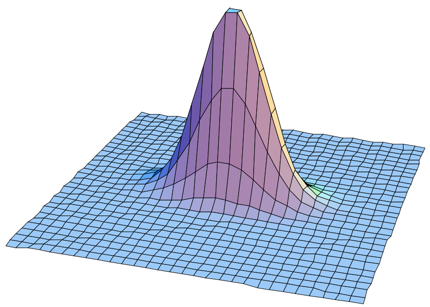
- Example: a repeated convolution of two blockfunctions with each other.

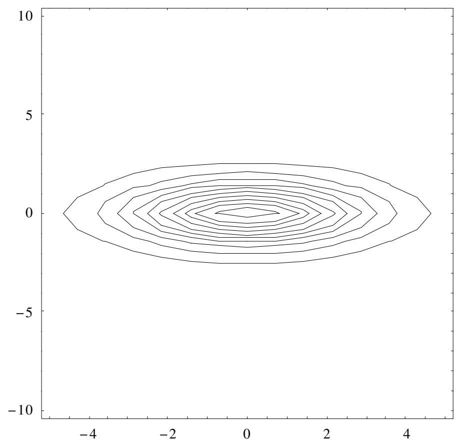


Anisotropy

- The Gaussian kernel as specified above is *isotropic*, which means that the behavior of the function is in any *direction* the same.
- When the standard deviations in the different dimensions are not equal, we call the Gaussian function *anisotropic*.







The diffusion equation

- The Gaussian function is the solution of the linear diffusion equation:

$$\frac{\partial L}{\partial t} = \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} = \Delta L$$

- 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.
- $\Delta L = \nabla \cdot (D \nabla L)$

Summary

- The normalized Gaussian kernel has an area under the curve of unity.
- Two Gaussian functions can be cascaded, 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 Dirac Delta function. The cumulative Gaussian function is the Error function, which is the 'blurred version' of the Heavyside stepfunction.
- 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.
- The Fourier transform of a Gaussian kernel acts as a low-pass filter for frequencies. 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, temperature, …) over space under the force of a gradient.

The Gaussian Kernel, Regularization

Differentiation and regularization

- Regularization
- Regular tempered distributions and testfunctions
- An example of regularization
- Relation regularization and Gaussian scale space

Taken from B. M. ter Haar Romeny,
Front-End Vision and Multi-scale Image Analysis,
Dordrecht, Kluwer Academic Publishers, 2003.
Chapter 8

Regularization

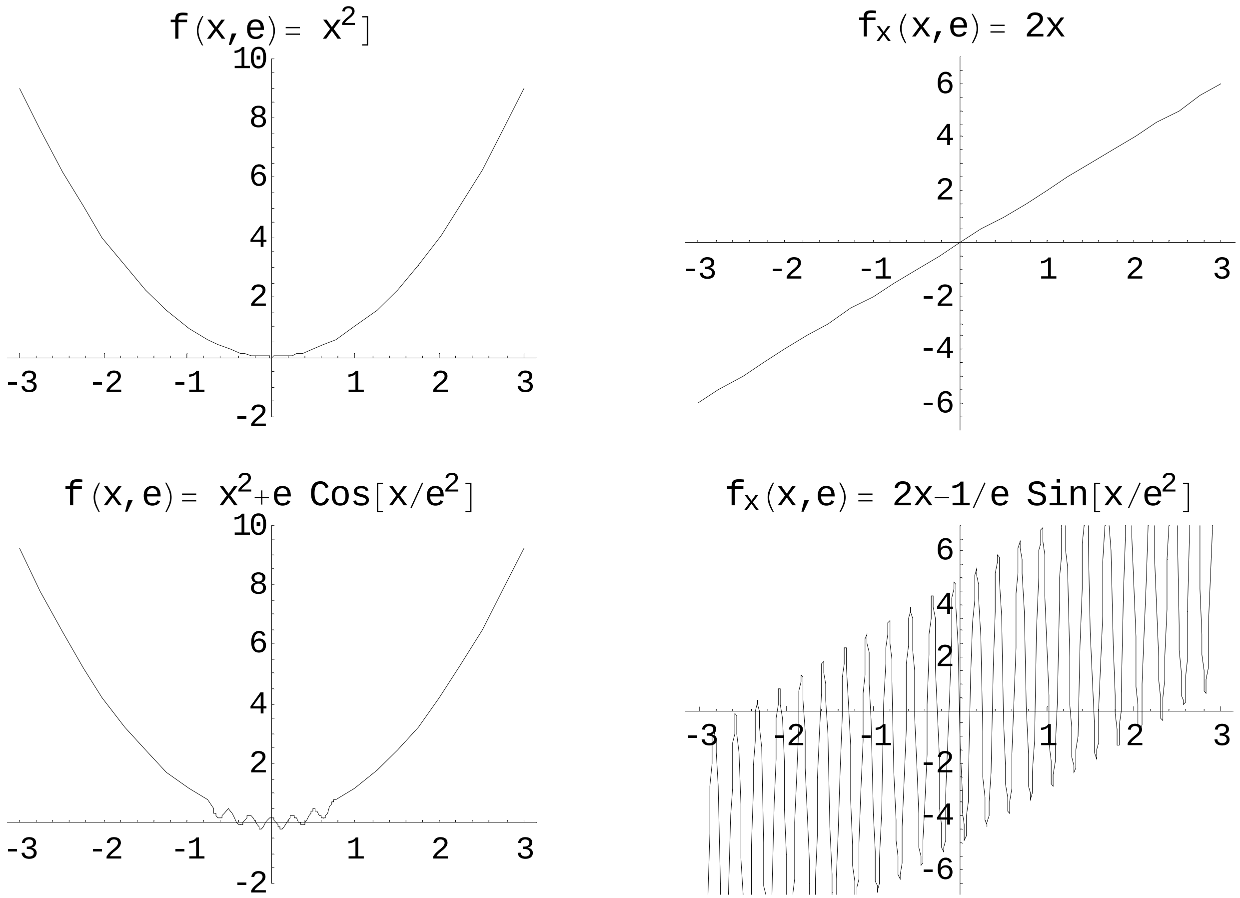
- Regularization is the technique to make data behave well when an operator is applied to them.

- Such data could e.g. be functions, that are impossible or difficult to differentiate, or discrete data where a derivate seems to be not defined at all.

- From physical principles – images are physical entities - 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 behavior'.

Regularization



- Differentiation is not well-defined.

Regularization

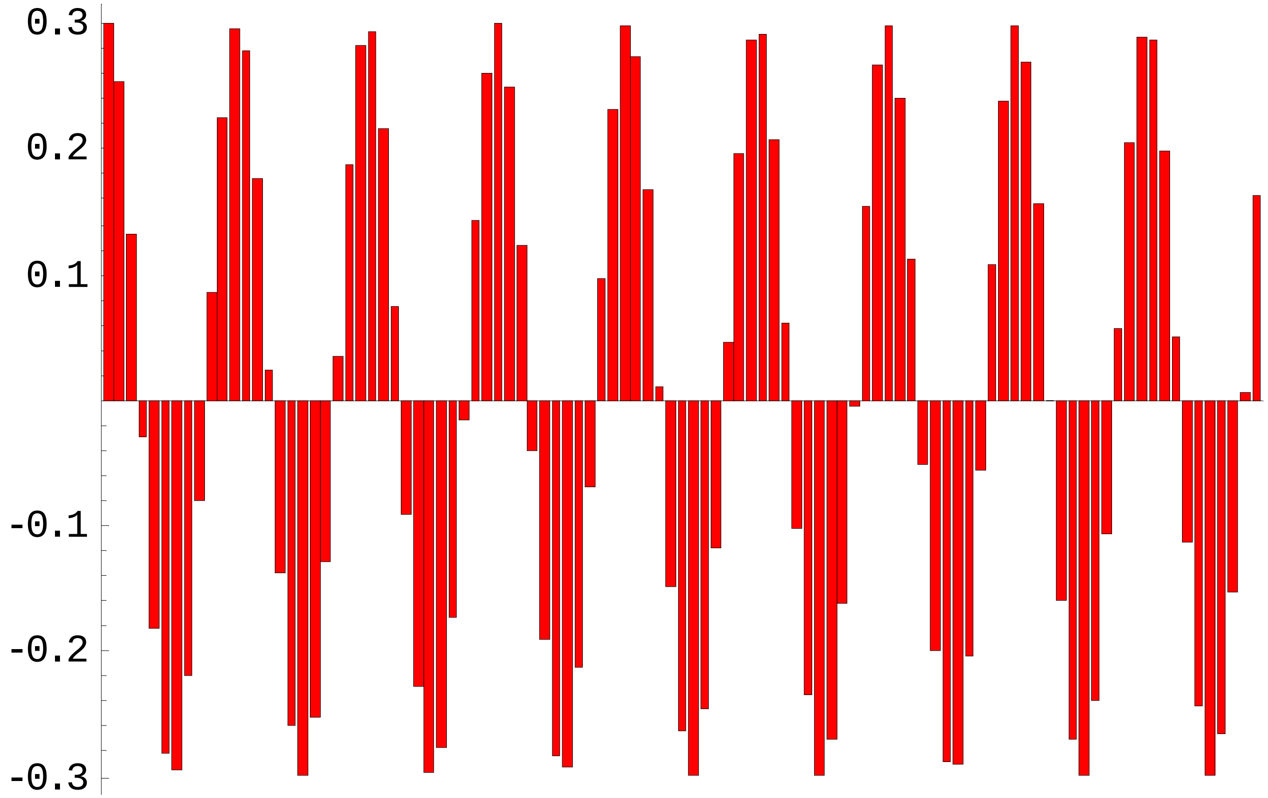
- A solution is well-defined in the sense of Hadamard if the solution
  - Exists
  - Is uniquely defined
  - Depends continuously on the initial or boundary data

- The operation is the problem, not the function.

- And what about discrete data?

Regularization

- How should the derivative of this thing look like?



- Regularize the data or the operation?

Regular tempered distributions and testfunctions

- Laurent Schwartz: use the Schwartz space with smooth test functions
  - Infinitely differentiable
  - decrease fast to zero at the boundaries

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

- Construct a regular tempered distribution
  - i.e. the integral of a test function and "something"

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

- The regular tempered distribution now has the nice properties of the test function.
- It can be regarded as a probing of "something" with a mathematically nice filter.

Regular tempered distributions and testfunctions

- Smooth test functions
  - Infinitely differentiable
  - decrease fast to zero at the boundaries
- For example a Gaussian.

The regular tempered distribution

=

The filtered image

- Now everything is well-defined, since integrating is well-defined.
- Do everything "under the integral"

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

- No data smoothing needed.

An example of regularization

$$\partial_x (L[x] + \epsilon \, \mathrm{Cos}[\omega x])$$

$$-\epsilon \, \omega \, \mathrm{Sin}[x \, \omega] + L'[x]$$

$$\int_{-\infty}^{\infty} \epsilon \, \mathrm{Cos}[\omega \, (x - \alpha)] \, \partial_x \left(\frac{1}{\sqrt{2 \pi} \, \sigma} \, E^{-\frac{x^2}{2 \sigma^2}} \right) d\alpha$$

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

Regularization and 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.

- The well known and much applied method of regularization as proposed by Tikhonov and Arsenin (often called 'Tikhonov regularization') is essentially equivalent to convolution with a Gaussian kernel.

Regularization and Gaussian scale space

- Try to find the minimum of a functional E(g), where g is the regularized version of f, given a set of constraints.

$$E(g) = \int_{-\infty}^{\infty} (f - g)^2 \, dx$$

- This constraint is the following: we also like the first derivative of g to x (g_x) to behave well: we require that when we integrate the square of over its total domain we get a finite result.

- 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 , one for each constraint, the so-called Lagrange multipliers.

Regularization and Gaussian scale space

- The functional becomes

$$E(g) = \int_{-\infty}^{\infty} (f - g)^2 + \lambda_1 \, {g_x}^2 \, dx$$

- The minimum is obtained at

$$\frac{d\,E}{d\,g} = 0$$

- Simplify things: go to Fourier space:
  - 1) Parceval theorem: the Fourier transform of the square of a function is equal to the square of the function itself.
  - 2) $\mathcal{F}\left(\frac{\partial g(x)}{\partial x}\right) = -i\,\omega\,\mathcal{F}(g(x))$

$$(-i\,\omega)\,(-i\,\omega)^* = \omega^2$$

Regularization and Gaussian scale space

- Therefore:

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

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

- So

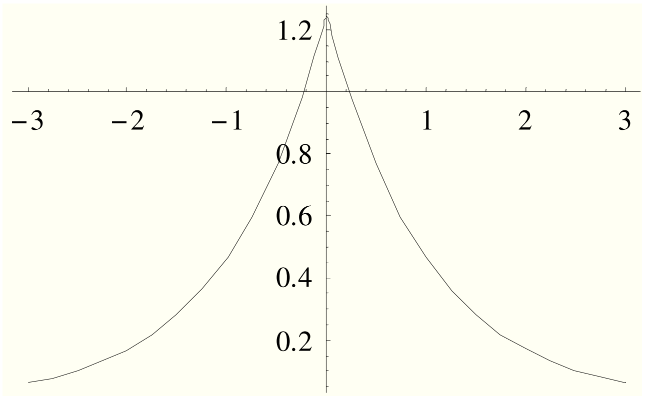
$$\frac{d\,\tilde{E}(\tilde{g})}{d\,\tilde{g}} = -2\,(\tilde{f}-\tilde{g}) \;+ 2\,\lambda_1\,\omega^2\,\tilde{g} = 0$$

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

Regularization and Gaussian scale space

- Back in the spatial domain:

$$\frac{1}{1+\lambda 1\,\omega^2} \to e^{-\mathrm{Abs}[x]}\sqrt{\frac{\pi}{2}}$$



- This is a first result for the inclusion of the constraint for the first order derivative.
- However, we like our function to be regularized with *all* derivatives behaving nicely, i.e. square integrable.

- Continue the procedure:

$$\tilde{E}(\tilde{g}) = \int_{-\infty}^{\infty} (\tilde{f} - \tilde{g})^2 + \lambda_1 \, \tilde{g}_x{}^2 + \lambda_2 \, \tilde{g}_{xx}^2 \, d\omega$$
$$= \int_{-\infty}^{\infty} (\tilde{f} - \tilde{g})^2 + \lambda_1 \, \omega^2 \, \tilde{g}^2 + \lambda_2 \, \omega^4 \, \tilde{g}^2 \, d\omega$$

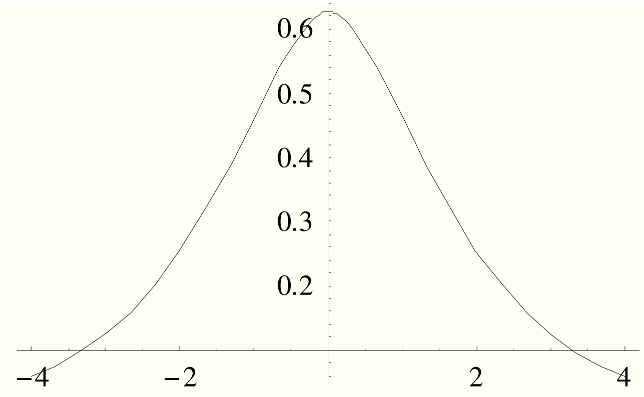
$$\frac{d\,E(\tilde{g})}{d\,\tilde{g}} = 2\,(\tilde{f} - \tilde{g})^2 + 2\,\lambda_1 \, \omega^2 \, \tilde{g} + \lambda_2 \, \omega^4 \, \tilde{g} = 0$$

$$\tilde{g} = \frac{1}{1+\lambda_1 \, \omega^2 + \lambda_2 \, \omega^4} \, \tilde{f}$$

Regularization and Gaussian scale space

- Back in the spatial domain:

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



$$\lambda_1 = 2\sqrt{\lambda_2}$$

- This is a result for the inclusion of the constraint for the first and second order derivative.
- However, we like our function to be regularized with *all* derivatives behaving nicely, i.e. square integrable.

Regularization and Gaussian scale space

- General form:

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

- How to choose the Lagrange multipliers?
  - Cascade property / *scale invariance* for the filters:

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

  - Computing per power one obtains

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

Regularization and Gaussian scale space

- This results in ($s=\frac{1}{2}\sigma^2$, $\alpha=1$)

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

- The Taylor series of the Gaussian in Fourier space:

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

Summary

- Many functions can not be differentiated.
  - 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.
- A well know variational form of regularization is given by the so-called Tikhonov regularization.
  - A functional is minimized in sense with the constraint of well behaving derivatives.
  - Tikhonov regularization with inclusion of the proper behavior of *all* derivatives is essentially equivalent to Gaussian blurring.

Next Time

- **Gaussian derivatives**
  - Shape and algebraic structure
  - Gaussian derivatives in the Fourier domain
  - Zero crossings of Gaussian derivative functions
  - The correlation between Gaussian derivatives
  - Discrete Gaussian kernels
  - Other families of kernels
- **Natural limits on observations**
  - Limits on differentiation: scale, accuracy and order
- **Deblurring Gaussian blur**
  - Deblurring
  - Deblurring with a scale-space approach
  - Less accurate representation, noise and holes
- **Multiscale derivatives: implementations**
  - Implementation in the spatial domain
  - Separable implementation
  - Some examples
  - N-dim Gaussian derivative operator implementation
  - Implementation in the Fourier domain
  - Boundaries

In some weeks

- **The differential structure of images**
  - Differential image structure
  - Isophotes and flowlines
  - Coordinate systems and transformations
  - Directional derivatives
  - First order gauge coordinates
  - Gauge coordinate invariants: examples
    - Ridge detection
    - Isophote and flowline curvature in gauge coordinates
    - Affine invariant corner detection
  - A curvature illusion
  - Second order structure
    - The Hessian matrix and principal curvatures
    - The shape index
    - Principal directions
    - Gaussian and mean curvature
    - Minimal surfaces, zero Gaussian curvature surfaces
  - Third order image structure: T-junction detection
  - Fourth order image structure: junction detection
  - Scale invariance and natural coordinates
  - Irreducible invariants
    - Tensor notation