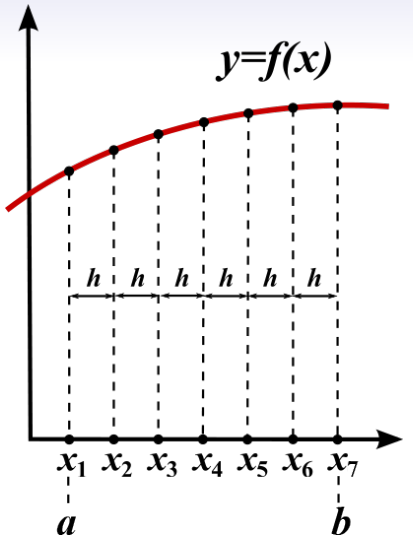
Ritz-Galerkin method

Sergiy Bubin

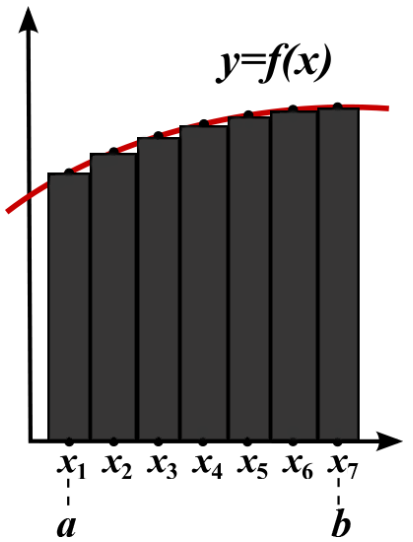
Department of Physics
Nazarbayev University

Analytic approximation of a function

What we have done previously is that we represented a function (a solution of a differential equation of an eigenvalue problem) as an array of points on a grid.



You can think about this from a different pespective. Essentially we represent the function as a combination of bins. If you define a bin as a sharp square function $b_i(x)$ then $f(x)$ can be represented as a linear combination of such functions:



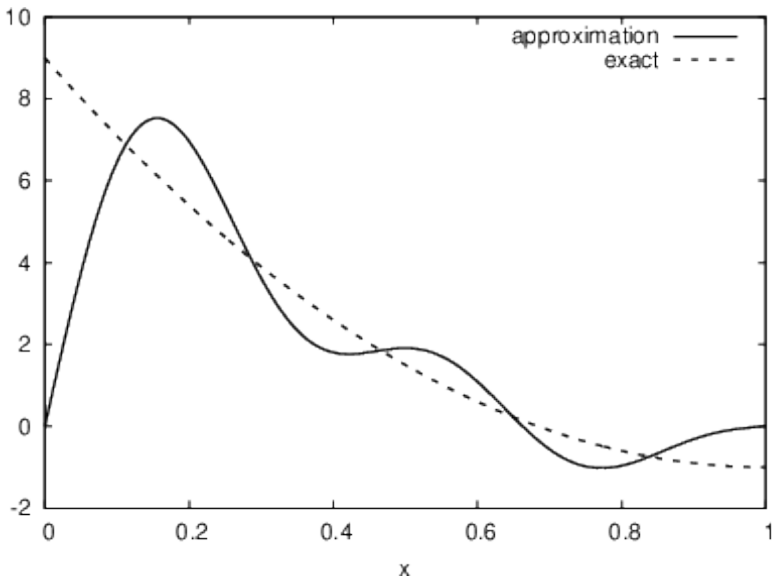
$$f(x) = \sum_{i=1}^{N} c_i b_i(x), \quad b_i(x) = \begin{cases} 1, & x_i - \frac{h}{2} < x < x_i + \frac{h}{2} \\ 0, & \text{otherwise} \end{cases}$$

$$c_i = f(x_i)$$

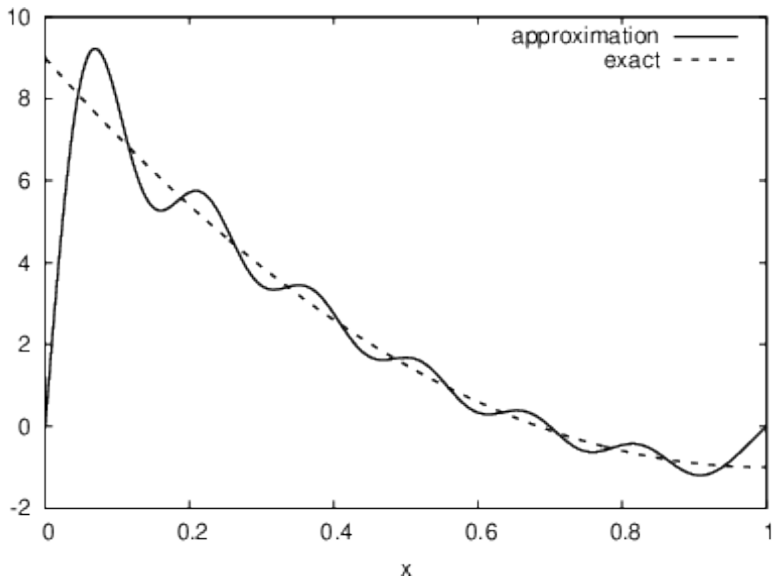
Analytic approximation of a function

What if instead of square bins we use some other (more sophisticated and, hopefully, better) basis functions? Indeed, this is what we do when we represent a function as a truncated Fourier series. Obviously the more the terms we use in the Fourier series the better it approximates our function on a given interval.

$$f(x) = \sum_{i=0}^{N} c_i \sin((i+1)\pi x)$$



$N = 3$



$N = 11$

Analytic approximation of a function

The big question is: how do we choose coefficients c_i? What are our criteria? There may be different choices that give the result that in some way can be considered a close approximation to $f(x)$.
When we solve a linear differential equation with an unknown function $f(x)$, e.g.

$$Lf(x) = g(x),$$

we have the left and right hand sides of it. If we approximate $f(x)$ as

$$f_{\text{appr}}(x) = \sum_{i=1}^{N} c_i \varphi_i(x),$$

then we may naturally require that the left hand side deviates from the right hand side as little as possible. That is, the residual

$$r(x) \equiv Lf_{\text{appr}} - g$$

is minimized. What does it mean "minimized"? There may be different measures of it.

Inner product of functions

For further discussion we need to define the inner product of two functions $p(x)$ and $q(x)$ on a given interval/domain

$$\langle p|q\rangle = \int_a^b p(x)^* q(x) dx,$$

where the asterix $(*)$ stands for complex conjugation. In many situations we deal with real functions so complex conjugation operation is not needed.

Another important concept is the orthogonality of two functions. Functions $p(x)$ and $q(x)$ are orthogonal if

$$\langle p|q\rangle = \langle q|p\rangle = 0.$$

Galerkin orthogonality

Since we have a set of N basis functions φ_i at our disposal we may require that the residual $r(x) = Lf_{\text{appr}} - g$ is orthogonal to the subspace spanned by $\{\varphi_i\}$. In other words, $r(x)$ should be orthogonal to each φ_i. This property is the Galerkin orthogonality.
This yields a system of equations:

$$\langle \varphi_j | r \rangle = 0, \qquad j = 1, \ldots, N$$

or

$$\langle \varphi_j | L \sum_{i=1}^{N} c_i \varphi_i - g \rangle = 0, \qquad j = 1, \ldots, N$$

which is equivalent to a system of N linear algebraic equations:

$$\sum_{i=1}^{N} c_i \underbrace{\langle \varphi_j | L \varphi_i \rangle}_{A_{ji}} = \underbrace{\langle \varphi_j | g \rangle}_{b_j}, \qquad j = 1, \ldots, N$$

$$Ac = b$$

Choice of basis

The most important part of the Galerkin type methods is the choice of an appropriate basis $\{\varphi_i\}$, $i = 1 \dots N$

- The basis should be good enough to represent the solution accurately (at least in the domain where we are interested)
- It should provide systematic convergence if we increase N (i.e. the basis should be complete)
- It should satisfy the boundary conditions of the problem
- All "inner products" (i.e. integrals A_{ji} and b_j) with basis functions $\{\varphi_i\}$ should be computable and these computations should be numerically efficient.

Galerkin method example 1

Let us solve the differential equation

$$f''(x) + f(x) = 2x(x-1),$$

with the boundary conditions $y(0) = 0$ and $y(1) = 0$. In this case

$$L = \frac{d^2}{dx^2} + 1 \qquad g(x) = 2x(x-1).$$

The exact analytic solution to this equation is known and is given by

$$f(x) = 2(x-2)(x+1) + 4\cos x + 4\tan(1/2)\sin x$$

Let us choose some appropriate basis, $N = 3$:

$$\varphi_1(x) = x(1-x), \qquad \varphi_2(x) = x^2(1-x)^2, \qquad \varphi_3(x) = x^3(1-x)^3.$$

Each basis function satisfies the above boundary conditions, so any linear combination of them will obviously satisfy them too. Our trial (i.e. approximate) solution is then

$$f_{\text{appr}}(x) = \sum_{i=1}^{3} c_i \varphi_i(x).$$

Galerkin method example 1

When we substitute the trial function into the expressions for projections, namely

$$\langle \varphi_j | r \rangle = \int_0^1 \varphi_j(x) \left[L f_{\text{appr}}(x) - g(x) \right] dx = 0,$$

or, equivalently,

$$\int_0^1 \varphi_j(x) \left[f''_{\text{appr}}(x) + f_{\text{appr}}(x) - 2x(x-1) \right] dx = 0.$$

we obtain 3 equations for j=1, 2, and 3.

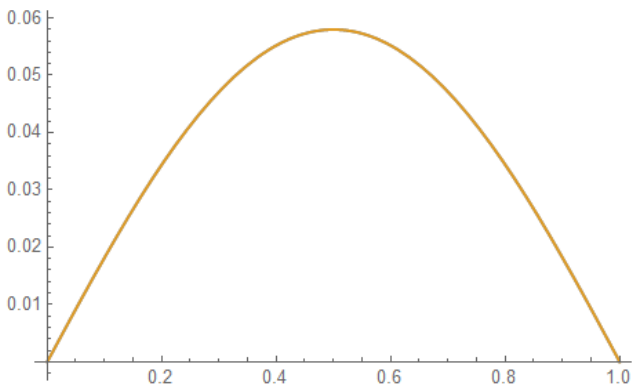
Galerkin method example 1

If we work out the integrals and do some basic algebra, we will get the following system of 3 linear equations:

$$\begin{aligned} -\frac{3}{10}c_1 + \frac{5}{84}c_2 - \frac{4}{315}c_3 &= \frac{1}{15}, \\ \frac{5}{84}c_1 + \frac{11}{630}c_2 + \frac{61}{13860}c_3 &= -\frac{1}{70}, \\ -\frac{4}{315}c_1 + \frac{61}{13860}c_2 - \frac{73}{60060}c_3 &= \frac{1}{315}. \end{aligned} \tag{1}$$

which has the following solution:

$$c_1 = -\frac{1370}{7397} \approx -0.18521,$$
$$c_2 = \frac{50688}{273689} \approx 0.185203,$$
$$c_3 = -\frac{132}{21053} \approx -0.00626989.$$



Galerkin method example 2

Galerkin method can also be used to solve eigenvalue problems. Let us consider the Schrödinger equation for 1D harmonic oscillator:

$$H\Phi_i(x) = E_i\Phi_i(x) \tag{2}$$

where the Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x). \tag{3}$$

and

$$V(x) = \frac{m\omega x^2}{2}.$$

The Ritz theorem says

$$\varepsilon = \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \geq E_1, \tag{4}$$

Here E_1 is the exact energy of the ground state. The equality holds if and only if $\psi \equiv \Phi_1$.

Galerkin method example 2

Use

$$\psi(x) = \sum_{i=1}^{N} c_i \varphi_i(x). \tag{5}$$

ϕ_i's are basis functions, c_i are the expansion coefficients (linear variational parameters). Minimization with respect to c_i's yields the generalized eigenvalue problem:

$$\mathcal{H}c = \varepsilon \mathcal{S}c, \tag{6}$$

$$\mathcal{H}_{ij} = \langle \varphi_i | H | \varphi_j \rangle, \quad \mathcal{S}_{ij} = \langle \varphi_i | \varphi_j \rangle, \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}.$$

There are actually N solutions (roots) to equation (6). The mini-max theorem guarantees that

$$\varepsilon_1 \geq E_1, \quad \varepsilon_2 \geq E_2, \quad \ldots, \quad \varepsilon_N \geq E_N. \tag{7}$$

Galerkin method example 2

Gaussian basis functions

$$\varphi_i(x) = \left(\frac{\alpha_i}{\pi}\right)^{1/2} e^{-\alpha_i (x-s_i)^2}. \tag{8}$$

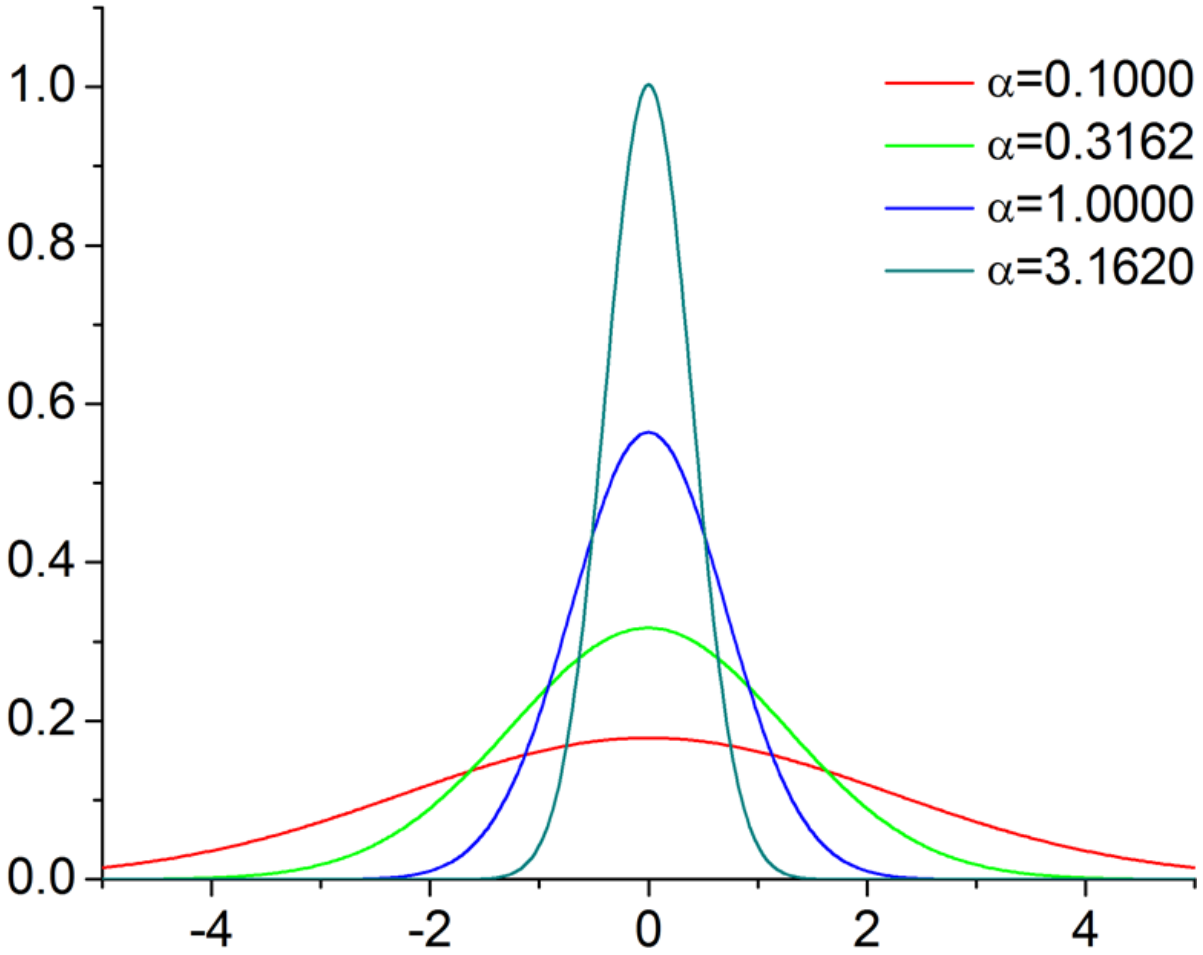
Here α_i and s_i are nonlinear variational parameters. Matrix elements are easily calculable:

$$\langle \varphi_i | \varphi_j \rangle = \left(\frac{2\sqrt{\alpha_i \alpha_j}}{\alpha_i + \alpha_j}\right)^{1/2} \exp\left(-\frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}(s_i - s_j)^2\right), \tag{9}$$

$$\langle \varphi_i | -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} | \varphi_j \rangle = \frac{\hbar^2}{m}\frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}\left(1 - \frac{2\alpha_i \alpha_j}{\alpha_i + \alpha_j}(s_i - s_j)^2\right)\langle \varphi_i | \varphi_j \rangle, \tag{10}$$

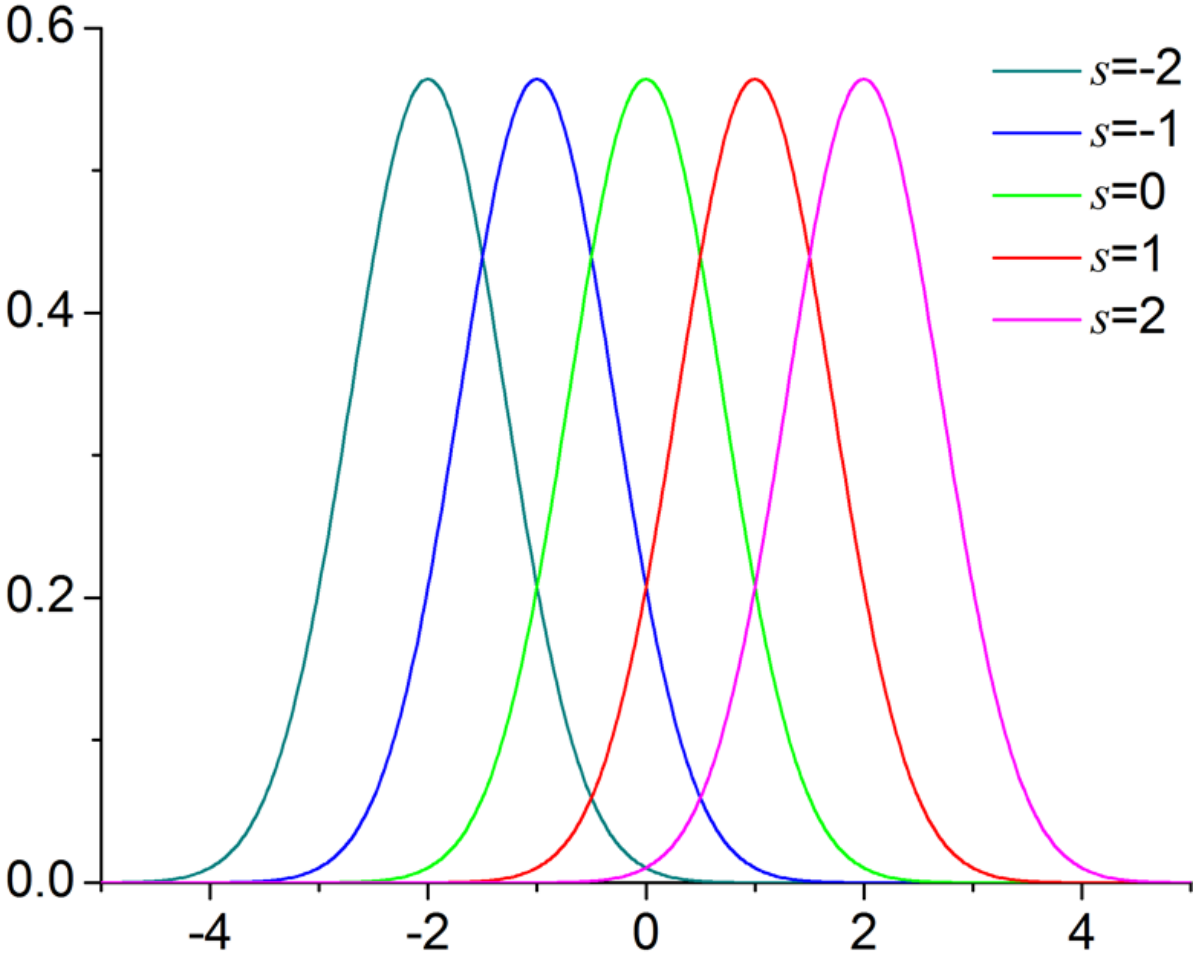
$$V(x) = \frac{m\omega^2 x^2}{2} \; : \; \langle \varphi_i | V(x) | \varphi_j \rangle = \frac{m\omega}{2}\left(\frac{1}{\alpha_i + \alpha_j} + \left(\frac{\alpha_i s_i + \alpha_j s_j}{\alpha_i + \alpha_j}\right)^2\right)\langle \varphi_i | \varphi_j \rangle. \tag{11}$$

Galerkin method example 2



Normalized Gaussian functions with different α_i.

Galerkin method example 2



Normalized Gaussian functions with different s_i and the same α_i.

Galerkin method example 2

Comparison of the eigenenergies of the quantum harmonic oscillator obtained by using shifted and variable-width Gaussian bases. 101 basis functions were used with some appropriately chosen parameters.

Exact	Shifted	Variable-width
0.5	0.500000000000	0.500000000000
1.5	1.500000000000	
2.5	2.500000000000	2.499999999954
3.5	3.500000000009	
4.5	4.500000000059	4.499999998287
5.5	5.500000001890	
6.5	6.500000005356	6.499999967019
7.5	7.500000161985	
8.5	8.500000213739	8.499997966523
9.5	9.500007221469	

Galerkin method summary

- Could be used to solve linear problems (differential equations, eigenvalue problems)
- Easily generalized to problems in multiple dimensions, even more than in 3D
- Requires an appropriate basis (ideally such that all matrix elments can be computed analytically)
- Basis functions may depend on a set of parameters that can be tuned to make convergence even better