Least squares approximations with the TI-92

Guido Herweyers

KHBO-Department Industrial Sciences and Technology
Zeedijk 101, 8400 Ostend, Belgium
Email: guido.herweyers@kh.khbo.be

1 Abstract

Let V be a real vector space equipped with an inner product, and W a finite-dimensional subspace of V. If $\mathbf{v}$ is a vector in V, then the best approximation to $\mathbf{v}$ by a vector in W is the orthogonal projection of $\mathbf{v}$ on W. We will use the TI-92 to find this least squares approximation to $\mathbf{v}$ in several applications. In the "continuous case", this will lead to approximating functions defined on an interval by polynomials, and 2π-periodic functions will be approximated by trigonometric polynomials, leading to the Fourier series. We will illustrate graphically the meaning of "best approximation". In the "discrete case", we will compute a least squares fitting to data and the discrete Fourier transform of a sampled 2π-periodic function.

2 Theoretical background

Consider a real vector space V. With each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, we associate a real number $<\mathbf{u},\mathbf{v}>$, the *inner product* of $\mathbf{u}$ and $\mathbf{v}$, satisfying the following axioms for all vectors $\mathbf{u},\mathbf{v}$ and $\mathbf{w}$ in V and all real scalars k:

1. $<\mathbf{u},\mathbf{v}>=<\mathbf{v},\mathbf{u}>$
2. $<\mathbf{u}+\mathbf{v},\mathbf{w}>=<\mathbf{u},\mathbf{w}>+<\mathbf{v},\mathbf{w}>$
3. $<k\mathbf{u},\mathbf{v}>=k<\mathbf{u},\mathbf{v}>$
4. $<\mathbf{v},\mathbf{v}>\geq 0$ with $<\mathbf{v},\mathbf{v}>=0$ if and only if $\mathbf{v}=\mathbf{0}$

A *Euclidean space* is a real vector space equipped with an inner product. Two vectors $\mathbf{u},\mathbf{v}$ in a Euclidean space are called *orthogonal* if and only if $<\mathbf{u},\mathbf{v}>=0$. The *norm* or *length* of a vector $\mathbf{v}$ in a Euclidean space V is denoted by $\|\mathbf{v}\|$ and defined by $\|\mathbf{v}\|=\sqrt{<\mathbf{v},\mathbf{v}>}$. The norm satisfies the following properties [2] for any $\mathbf{u},\mathbf{v}\in V$ and $k\in\mathbb{R}$:

1. $\|k\mathbf{v}\|=|k|.\|\mathbf{v}\|$
2. $\|\mathbf{v}\|\geq 0$ and $\|\mathbf{v}\|=0$ if and only if $\mathbf{v}=\mathbf{0}$
3. $\|\mathbf{u}+\mathbf{v}\|\leq\|\mathbf{u}\|+\|\mathbf{v}\|$ (triangle inequality)
4. $|<\mathbf{u},\mathbf{v}>|\leq\|\mathbf{u}\|.\|\mathbf{v}\|$ (Cauchy-Schwarz inequality)

Note that if $\mathbf{v}\in V$, $\mathbf{v}\neq\mathbf{0}$, then $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is a vector with norm 1 or a *unit vector*; we say that $\mathbf{v}$ has been *normalized*.
The *distance* between two vectors $\mathbf{u}$ and $\mathbf{v}$ is denoted by $d(\mathbf{u},\mathbf{v})$ and defined by
$d(\mathbf{u},\mathbf{v})=\|\mathbf{u}-\mathbf{v}\|$.

For the remainder of this paper, we will work with the following Euclidean spaces:

1. *Discrete case*
   $V = \mathbb{R}^n$ $(n \in \mathbb{N}_0)$. For $\mathbf{u}= (u_1, u_2, \ldots, u_n)$ and $\mathbf{v}= (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$, we define $< \mathbf{u}, \mathbf{v} >= u_1v_1 + u_2v_2 + \cdots u_nv_n$, this particular inner product is also written as $\mathbf{u}.\mathbf{v}$ and called the *dot product.*

2. *Continuous case*
   $V = C[a, b]$, the set of all continuous functions on the interval $[a, b]$.
   For $f(x), g(x) \in C[a, b]$, we define $< f(x), g(x) >= \int_a^b f(x).g(x)\,dx$

3 Orthogonal projection

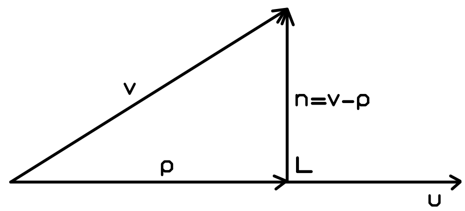

Figure 1: *orthogonal projection of* **v** *on* **u**

If **u** and **v** are vectors in a Euclidean space V, then figure 1 suggests that every vector **v** in V can be written in exactly one way as

$$\mathbf{v} = \mathbf{p} + \mathbf{n}$$

with $\mathbf{p} = k\mathbf{u}$ and **n** orthogonal to **u**. Indeed, this means that $< \mathbf{n}, \mathbf{u} >= 0$ or $< \mathbf{v} - k\mathbf{u}, \mathbf{u} >= 0$, so $k = \frac{<\mathbf{v},\mathbf{u}>}{<\mathbf{u},\mathbf{u}>}$ (if $\mathbf{u} \neq \mathbf{0}$). That particular vector **p** is called the *orthogonal projection* of **v** on **u**, we denote it by proj(**v**,**u**):

$$\text{proj}(\mathbf{v}, \mathbf{u}) = \frac{< \mathbf{v}, \mathbf{u} >}{< \mathbf{u}, \mathbf{u} >}\mathbf{u} \tag{1}$$

Examples:

1. Let $V = \mathbb{R}^3$. Find the orthogonal projection of $\mathbf{v} = (4, -1, 3)$ on $\mathbf{u} = (5, 1, 2)$ and check that $\mathbf{n} = \mathbf{v} - \text{proj}(\mathbf{v}, \mathbf{u})$ is orthogonal to **u**.
   Solution: for $\text{proj}(\mathbf{v}, \mathbf{u}) = \frac{\mathbf{v}.\mathbf{u}}{\mathbf{u}.\mathbf{u}}\mathbf{u}$, we use the function dotP (Figure 2a)

2. Let $V = C[0, 1]$. Consider the functions $e^x, x^2 \in C[0, 1]$. Find the orthogonal projection of e^x on x^2. (Figure 2b)

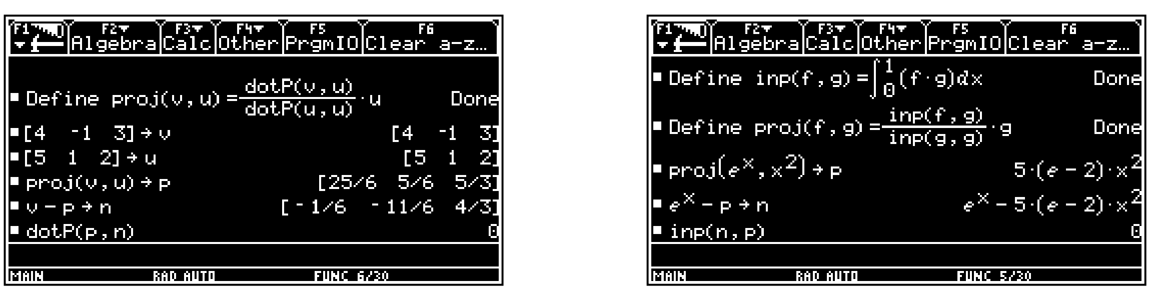

Figure 2: *orthogonal projections*

4 Least squares approximation

4.1 Definition

Let V be a Euclidean space, and W a finite-dimensional subspace of V. Every vector $\mathbf{v}$ in V can be written in exactly one way [1] as

$$\mathbf{v} = \mathbf{p} + \mathbf{n}$$

with $\mathbf{p} \in W$ and $\mathbf{n}$ orthogonal to W (i.e. orthogonal to each vector in W).
We call this vector $\mathbf{p}$ the *orthogonal projection* of $\mathbf{v}$ on W and denote it by $\text{proj}(\mathbf{v}, W)$. See figure 3.
The orthogonal projection of $\mathbf{v}$ on W is the best approximation of $\mathbf{v}$ by a vector in W, in the sense that $d(\mathbf{v}, \mathbf{p}) = \|\mathbf{v} - \mathbf{p}\|$, with $\mathbf{p} \in W$, is minimized [1] by $\mathbf{p} = \text{proj}(\mathbf{v}, W)$. As also $(d(\mathbf{v}, \mathbf{p}))^2 = \|\mathbf{v} - \mathbf{p}\|^2$ is minimized by $\mathbf{p} = \text{proj}(\mathbf{v}, W)$, we call proj($\mathbf{v}$,W) the *least squares approximation* to $\mathbf{v}$ by a vector in W.

4.2 Calculation

Choose a basis $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m)$ for W, then every vector $\mathbf{p}$ in W can be written in exactly one way as

$$\mathbf{p} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + \cdots + k_m\mathbf{b}_m \tag{2}$$

Vector $\mathbf{p}$ becomes $\text{proj}(\mathbf{v}, W)$ by expressing that $\mathbf{n} = \mathbf{v} - \mathbf{p}$ is orthogonal to W or $\mathbf{n} \perp \mathbf{b}_i$ $(i = 1, 2, \ldots m)$, this leads to a system of linear equations, called the system of *normal equations*, written in matrix form as:

$$\begin{bmatrix} <\mathbf{b}_1,\mathbf{b}_1> & <\mathbf{b}_1,\mathbf{b}_2> & \cdots & <\mathbf{b}_1,\mathbf{b}_m> \\ <\mathbf{b}_2,\mathbf{b}_1> & <\mathbf{b}_2,\mathbf{b}_2> & \cdots & <\mathbf{b}_2,\mathbf{b}_m> \\ \vdots & \vdots & \cdots & \vdots \\ <\mathbf{b}_m,\mathbf{b}_1> & <\mathbf{b}_m,\mathbf{b}_2> & \cdots & <\mathbf{b}_m,\mathbf{b}_m> \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} = \begin{bmatrix} <\mathbf{v},\mathbf{b}_1> \\ <\mathbf{v},\mathbf{b}_2> \\ \vdots \\ <\mathbf{v},\mathbf{b}_m> \end{bmatrix} \quad \text{or } A.K = B \tag{3}$$

This system has one solution $K = A^{-1}.B$, the coordinate vector of $\mathbf{p} = \text{proj}(\mathbf{v}, W)$ in (2). We prefer to find the solution with the reduced row-echelon form of the augmented matrix $C = [A|B]$ of the system of normal equations, calculated by rref(D) with the TI-92. If the basis $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m)$ for W is *orthogonal* ($\mathbf{b}_i \perp \mathbf{b}_j$ for $i \neq j$), matrix A becomes a diagonal matrix and the solution of the system can be found immediately:

$$k_i = \frac{<\mathbf{v}, \mathbf{b}_i>}{<\mathbf{b}_i, \mathbf{b}_i>} \quad (i = 1, 2, \ldots, m)$$

In this case, the least squares approximation to $\mathbf{v}$ becomes, using (1):

$$\mathbf{p} = \text{proj}(\mathbf{v}, \mathbf{b}_1) + \text{proj}(\mathbf{v}, \mathbf{b}_2) + \cdots + \text{proj}(\mathbf{v}, \mathbf{b}_m) \tag{4}$$

See figure 3, with $V = \mathbb{R}^3$ and $m = 2$.
It is possible to convert an arbitrary basis into an orthogonal basis by the *Gram-Schmidt process.* It is no longer necessary to carry out the Gram-Schmidt process, as we can solve the general system (3) of normal equations directly with the TI-92. However, an orthogonal basis has the advantage that, if we look for the best approximation of $\mathbf{v}$ by a vector in the subspace $W_j \subset W$

where W_j has the orthogonal basis $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j)$ consisting of the first j basisvectors of W, we only have to take the first j terms of (4). In order to find the best approximations to $\mathbf{v}$ by vectors in $W_1 \subset W_2 \subset \cdots \subset W_m = W$ it suffices to start with the first term in (4) and to add the next term for the next approximation, and so on.

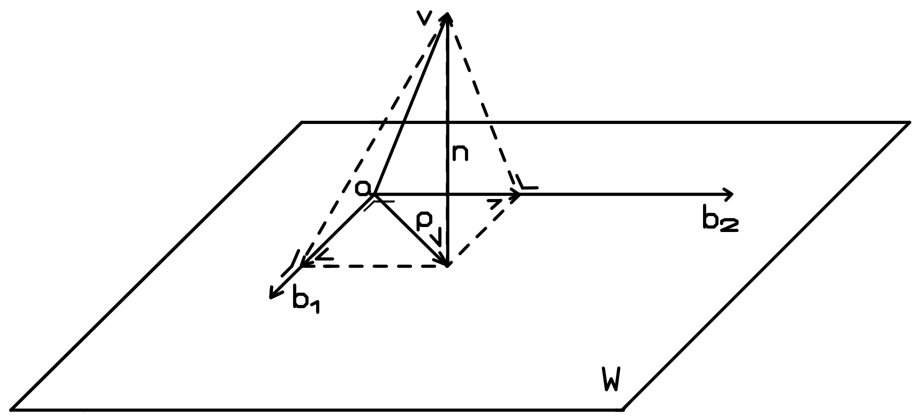

Figure 3: *orthogonal projection of* $\mathbf{v}$ *on* W*, with orthogonal basis* $(\mathbf{b}_1, \mathbf{b}_2)$.

5 The Gram-Schmidt process

Let W be a Euclidean space with basis $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m)$, then this basis can be changed into an orthogonal basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m)$ by the following procedure [1, 2]:

$$\begin{aligned}
\mathbf{e}_1 &= \mathbf{b}_1 \\
\mathbf{e}_2 &= \mathbf{b}_2 - \text{proj}(\mathbf{b}_2, \mathbf{e}_1) \\
\mathbf{e}_3 &= \mathbf{b}_3 - \text{proj}(\mathbf{b}_3, \mathbf{e}_1) - \text{proj}(\mathbf{b}_3, \mathbf{e}_2) \\
&\vdots \\
\mathbf{e}_m &= \mathbf{b}_m - \text{proj}(\mathbf{b}_m, \mathbf{e}_1) - \text{proj}(\mathbf{b}_m, \mathbf{e}_2) - \text{proj}(\mathbf{b}_m, \mathbf{e}_3) - \cdots - \text{proj}(\mathbf{b}_m, \mathbf{e}_{m-1})
\end{aligned}$$

6 Exercises

During the workshop, we will solve the following exercises. Inspiration can be found in [1, 3].

1. Let $V = C[0,1]$, the set of continuous functions on [0,1] and $W = P_2(x)$, the set of polynomials in x with degree ≤ 2.
   Find the best function $p(x) = a + b.x + c.x^2$ in $P_2(x)$ as an approximation to the function e^x in $C[0,1]$,

   - by choosing the natural basis $(1, x, x^2)$ for W and using (3)
   - by first converting the natural basis into an orthogonal basis with the Gram-Schmidt process, and using (4).

   Find the distance from e^x to the best approximation $p(x)$ and interpret graphically the meaning of "least squares approximation" by plotting the function $(e^x - p(x))^2$ on the interval $[0,1]$.

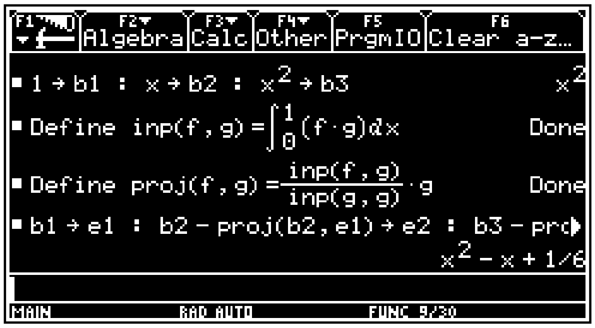

Figure 4: *Gram-Schmidt process*

2. Let V be the set of 2π periodic functions f, with f piecewise continuous on $[0, 2\pi]$. For $f, g \in V$ we define $< f(x), g(x) >= \int_0^{2\pi} f(x).g(x)\,dx$. Consider the vector space W, generated by the set $S = \{1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\}$

    - Prove that S is an orthogonal basis for W.
    - According to (4), a function $f(x)$ in V can be approximated by a *trigonometric polynomial* of order n or less:

    $$f(x) \simeq \frac{< f(x), 1 >}{< 1, 1 >} 1 + \frac{< f(x) \cos x >}{< \cos x, \cos x >}. \cos x + \cdots + \frac{< f(x), \sin nx >}{< \sin nx, \sin nx >} \sin nx$$

    For $k \to \infty$, this results in the *Fourier series* for $f(x)$. The numbers

    $$a_k = \frac{< f(x), \cos kx >}{< \cos kx, \cos kx >} \; (k = 0, 1, \ldots, n) \text{ and } b_k = \frac{< f(x), \sin kx >}{< \sin kx, \sin kx >} \; (k = 1, \ldots, n)$$

    are called *Fourier coefficients* of f.
    Find approximations to the periodic extension of the function $f(x) = x$ $(x \in [0, 2\pi])$ for $n = 1, 2, 3, 4$ and plot these functions on the interval $[0, 4\pi]$.

3. *Least squares curve fitting to data.* A practical problem is to find a relationship $y = f(x)$ between the variables x and y, when n experimentally determined points

    $$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

    are given. This relationship can be of the form

    (a) $y = a + bx$
    (b) $y = a + bx + cx^2$
    (c) $y = a. \cos x + b. \sin(2x) + c.x$
    (d) $y = a_1.f_1(x) + a_2.f_2(x) + \cdots + a_m.f_m(x)$, with $S = \{f_1(x), f_2(x), \ldots, f_m(x)\}$ a linear independent set.

    Given the form of a relationship, how can we find the curve that best fits to the data points?
    Define the discrete set $X = \{x_1, x_2, \ldots, x_n\}$, consisting of the x-coordinates of the given points. Let V be the set of all real functions on X. We can identify a function $f \in V$,

for a fixed X, with the ordered n-tuple of images $(f(x_1), f(x_2), \ldots, f(x_n))$. This allows us to identify V with $\mathbb{R}^n$ and to write:

$$f(x) = (f(x_1), f(x_2), \ldots, f(x_n))$$

The given data points can be interpreted as the function $(y_1, y_2, \ldots, y_n)$ in V, consisting of the y-coordinates of the given points. For two functions $f(x), g(x) \in V$, the inner product is defined by

$$< f(x), g(x) >= \sum_{i=1}^{n} f(x_i).g(x_i)$$

- Fit a line $p(x) = a + b.x$ to the points $(-3, 0), (1, 0), (2, 1)$.
  Put $W = \text{lin}\{1, x\}$, i.e. the linear space generated by the set $\{1, x\}$ of functions defined on $X = \{-3, 1, 2\}$. Remark that $W = \{a.1 + b.x | a, b \in \mathbb{R}\}$ is a set of functions, defined on X. The least squares approximation to the given function $(0, 0, 1)$ in $V = \mathbb{R}^3$ can now be calculated with (3). The result should be the same as the regression line calculated with the TI-92 function LinReg.
- Fit a curve $p(x) = a. \cos x + b.e^x$ to the same data points.
- Draw the data points and the approximating curves. What is the meaning of "least squares approximation"? Compute the distance from the given function to the approximating functions.

4. *Discrete Fourier transform*
It is a surprising fact that the set

$$S = \{1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin(n-1)x\}$$

of functions defined on

$$X = \{0, \frac{2\pi}{N}, 2.\frac{2\pi}{N}, 3.\frac{2\pi}{N}, \ldots, (N-1).\frac{2\pi}{N}\}$$

with $N = 2n$ an even number, is an orthogonal set of vectors in $V = \mathbb{R}^N$, so S is a basis for $\mathbb{R}^N$ [4].

Sampling an arbitrary 2π-periodic function or signal $f(x)$ in the equidistant points of X, yields a discrete signal, also denoted as $f(x)$:

$$f(x) = (f(0), f(\frac{2\pi}{N}), f(2.\frac{2\pi}{N}), f(3.\frac{2\pi}{N}), \ldots, f((N-1).\frac{2\pi}{N}))$$

which can be written in exactly one way as

$$f(x) = \hat{a}_0.1 + \hat{a}_1. \cos x + \cdots \hat{a}_n. \cos nx + \hat{b}_1. \sin x + \hat{b}_2. \sin 2x + \cdots \hat{b}_{n-1}. \sin(n-1)x \ . \quad (5)$$

The vector $(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \ldots \hat{b}_{n-1})$ is called the (real) *discrete Fourier transform* of the discrete signal $f(x)$. Taking the inner product of both sides of equation (5) with $\cos kx$ respectively $\sin kx$ results in:

$$\hat{a}_k = \frac{< f(x), \cos kx >}{< \cos kx, \cos kx >} \quad \text{and} \quad \hat{b}_k = \frac{< f(x), \sin kx >}{< \sin kx, \sin kx >}$$

with

$$\begin{aligned} < \cos kx, \cos kx > = < \sin kx, \sin kx > &= n \quad (1 \leq k < n-1) \\ < \cos kx, \cos kx > &= 2n \quad (k = 0, n) \quad [4] \end{aligned}$$

One can prove [4] that the *discrete Fouriercoefficients* $(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \ldots \hat{b}_{n-1})$ are approximations (Riemann-sums) for the corresponding "continuous" Fouriercoefficients $(a_0, a_1, \ldots, a_n, b_1, b_2, \ldots b_{n-1})$, they even become equal if the continuous function $f(x)$ is a trigonometric polynomial of order M, and $N > M$. In that case, we have:

$$\hat{a}_0 = a_0, \hat{a}_1 = a_1, \hat{b}_1 = b_1, \ldots \hat{b}_M = b_M$$

- Find the discrete Fourier transform of (1,5,3,2).
- Rediscover the formula $\cos^2 x = \frac{1}{2} + \frac{\cos 2x}{2}$ by sampling the function $\cos^2 x$ in $N = 6$ points and calculating the Fourier transform.
- Consider $f(x)$ a trigonometric polynomial of order M, then sampling $f(x)$ in N $(N > 2M)$ points guarantees the equality

$$\begin{aligned} a_0 &= \hat{a}_0 \quad \text{or, after multiplying with } 2\pi \\ \int_0^{2\pi} f(x)\,dx &= \sum_{k=0}^{N-1} f(k.\frac{2\pi}{N})\frac{2\pi}{N} \end{aligned}$$

  This yields the interesting side-effect that any Riemann-sum of the function $f(x)$, with N $(N > 2M)$ equidistant sampling points of the interval $[0, 2\pi]$, delivers the exact value of $\int_0^{2\pi} f(x)\,dx$.

  Calculate the exact value of $\displaystyle\int_0^{2\pi} \sin^2(x)\,dx$ by this way.

References

[1] ANTON, H./RORRES, C. : *Elementary Linear Algebra, Applications version*. New York: John Wiley & Sons, 1991, 6th ed.

[2] COHN, P.M. : *Algebra, Volume 1*. Chichester: John Wiley & Sons, 1974.

[3] EVANS, B./JOHNSON, J. : *Linear algebra with Derive*, New York: John Wiley & Sons, 1994.

[4] SCHNEEBELI, H.R./VOLLMER, H.R. : *Skalarprodukte-Schwingungen-Signale*, Zürich: sabe Verlag, 1998.