A choice of norm in discrete approximation*

Tomislav Marošević†

Abstract. *We consider the problem of choice of norms in discrete approximation. First, we describe properties of the standard l_1, l_2 and l_∞ norms, and their essential characteristics for using as error criteria in discrete approximation. After that, we mention the possibility of applications of the so-called total least squares and total least l_p norm, for finding the best approximation. Finally, we take a look at some nonstandard, visual error criteria for qualitative smoothing.*

Key words: *discrete approximation, l_1 norm, l_2 norm, l_∞ norm, total least squares, total least l_p norm, error criteria*

Sažetak. **Izbor norme za diskretnu aproksimaciju**. *Razmatramo problem izbora norme pri diskretnoj aproksimaciji. Prvo opisujemo svojstva standardnih normi l_1, l_2, l_∞ i njihove bitne značajke za uporabu kao kriterija greške kod diskretne aproksimacije. Zatim spominjemo mogućnost primjena tzv. potpunih najmanjih kvadrata i potpune najmanje l_p norme za nalaženje najbolje aproksimacije. Konačno navodimo i neke nestandardne, vizualne kriterije greške pri kvalitativnom glađenju.*

Ključne riječi: *diskretna aproksimacija, l_1 norma, l_2 norma, l_∞ norma, potpuni najmanji kvadrati, potpuna najmanja l_p norma, kriteriji greške*

1. Best discrete approximation problem

Approximation problem for a real, continuous function $f(x)$ on a certain interval $[a,b]$ is considered in literature (cf. [**?**], [**?**], [**?**], [**?**], [**?**]) from several aspects:

*The lecture presented at the Mathematical Colloquium in Osijek organized by Croatian Mathematical Society – Division Osijek, April 19, 1996.

†Faculty of Electrical Engineering, Department of Mathematics, Istarska 3, HR-31 000 Osijek, Croatia, e-mail: tommar@drava.etfos.hr

- a choice of a model or a form of the corresponding approximation function $F(\mathbf{a}, x)$, with an unknown vector of parameters $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$, and setting up a criterion for quality of approximation, i.e. a distance function $d(f(x), F(\mathbf{a}, x))$ in a metric space, resp. a norm $\|F(\mathbf{a}) - f\|$ in a normed linear space;
- the best approximation existence problem;
- the problem of uniqueness of best approximation;
- characterization of a solution;
- methods for solving the approximation problem.

Here we are going to consider the question of a choice of norm in approximation of a function f, especially in the case of discrete approximation, when the values of the function are given only at finitely many points (x_i, f_i), $i = 1, \ldots, m$.

A criterion for the quality of approximation is usually determined by means of norms.

Definition 1. *The l_p norm of the function f given at some finite data points set $X = \{x_i : i = 1, \ldots, m\}$, is defined by*

$$l_p(f) = \|f\|_p = \left(\sum_{i=1}^{m} |f(x_i)|^p \right)^{1/p}, \quad 1 \leq p < \infty \,.$$

For $p = \infty$ as a limit case, the l_∞ norm is defined by $\|f\|_\infty = \max_{i \in \{1,\ldots,m\}} |f(x_i)|$.

Definition 2. *The function $F(\mathbf{a}^*, x)$ is said to be the best approximation of the function f in the norm $\| \cdot \|_p$, if it holds*

$$\|F(\mathbf{a}^*) - f\|_p \leq \|F(\mathbf{a}) - f\|_p, \quad \forall \mathbf{a} \in P \subseteq R^n \,.$$

In that way, the approximation problem is reduced to the problem of minimization of the functional $S : R^n \to R$, $S(a_1, a_2, \ldots, a_n) = \|F(\mathbf{a}) - f\|_p$ (or rather, equivalently, of the functional $\|F(\mathbf{a}) - f\|_p^p$). In general, the best approximation in the l_p norm is different from the best approximation in the l_q norm ($p \neq q$). The l_p norms can be generalized by introducing the weights $(w(x_i),\ i = 1, \ldots, m)$.

If the approximating function $F(\mathbf{a}, x)$ is linear in parameters a_j, $j = 1, \ldots, n$, i.e. if $F(\mathbf{a}, x) = \mathbf{x}^T \cdot \mathbf{a}$, then it holds (cf. [**?**])

$$1 \leq p < q \leq \infty \quad \Rightarrow \quad \min_{\mathbf{a} \in R^n} \|\mathbf{X}\mathbf{a} - \mathbf{f}\|_q \leq \min_{\mathbf{a} \in R^n} \|\mathbf{X}\mathbf{a} - \mathbf{f}\|_p$$

(where $\mathbf{X}$ is a corresponding data matrix, and $\mathbf{f}$ is a vector of values of the dependent variable).

2. The l_2, l_1 and l_∞ norms

The three norms most frequently used in the practice are:

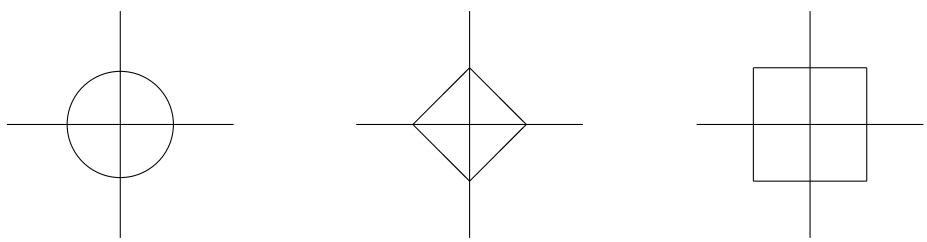
- the l_2 norm (the least squares or Euclidean norm);
- the l_1 norm (the least absolute deviations);
- the l_∞ norm (the Chebyshev norm).

The important relations among these norms are stated by the following well known theorem in the discrete case (cf. [**?**]).

Theorem 1. *For every* $\mathbf{v} \in R^m$ *it holds* $\|\mathbf{v}\|_\infty \leq \|\mathbf{v}\|_2 \leq \|\mathbf{v}\|_1$ *and*

$$\|\mathbf{v}\|_1 \leq \sqrt{m}\|\mathbf{v}\|_2 \leq m\|\mathbf{v}\|_\infty .$$

The shapes of balls in the l_2, l_1 and l_∞ norms in a normed linear space R^2 are shown in *Figure 1.*



$\|\mathbf{v}\|_2 = r$ $\|\mathbf{v}\|_1 = r$ $\|\mathbf{v}\|_\infty = r$

Figure 1. *The shapes of balls in the l_2, l_1 and l_∞ norms*

The l_2 norm. It is used traditionally and almost universally for practical applications in the approximation theory (cf. [**?**], [**?**], [**?**]). In the beginning of 19th century Legendre (1804) suggested the use of the l_2 norm for approximation of a solution to the inconsistent system of linear equations (respectively for the equivalent problem of approximation of a function which is given at the finite number of data points), and similar problem was considered by Gauss (cf. [**?**], [**?**]).

The l_2 norm is differentiable, and in the case when an approximating function $F(\mathbf{a})$ is linear with respect to parameters $\mathbf{a} = (a_1, \ldots, a_n)^T$, the approximation problem becomes a well known linear least squares problem (cf. [**?**]). Since the l_2 norm is strictly convex, in this case the best approximation exists in the l_2 norm, it is unique and depends continuously and smoothly on a function which is approximated (cf. [**?**]).

Statistical considerations show that the l_2 norm is the most suitable choice for smoothing the data in the case when additive data errors ε_i, $i = 1, \ldots, m$, have a normal distribution (i.e. $\varepsilon_i \sim N(0,\sigma^2)$, because then the influence of errors ε_i is minimized by means of use of the l_2 norm (cf. e.g. [?]).

Nonlinear least squares problems are also widely analyzed (cf. [?], [?]), and within that framework, especially the so-called separable problems (cf. [?], [?], [?]).

The l_∞ norm. One of its characteristics is that it considers only those data points where the maximal error appears. The best approximation in the l_∞ norm is obtained by minimizing the maximal distance. In 1799 Laplace suggested such a criterion (nowadays called the l_∞ norm) for approximative solving of inconsistent systems of linear equations; Fourier (1824) studied a similar problem. In the second half of 19th century Chebyshev made the first systematic analysis of this norm (because of that, the l_∞ norm is also called the Chebyshev norm, cf. [?]).

In the practice this norm is used if the data errors are very small with respect to an approximation error (cf. [?]). Characterization of the best approximation in the l_∞ norm is described by the so-called alternating property, from which an exchange algorithm for obtaining the best l_∞ approximation is derived (cf. [?]). In discrete case, computing of an l_∞ approximation can be expressed as the linear programming problem (cf. [?]). The l_∞ norm is not strictly convex, so the best approximations are not necessarily unique.

The l_1 norm. As late as the middle of eighteenth century R. Boscovich determined a criterion by which absolute deviations of data are minimized, among all lines which pass through a centroid of data points. In 1789 Laplace gave an algebraic analysis of this problem, which was also considered by Gauss (cf. [?], [?], [?]). By approximation in the l_1 norm, all deviations of error curve, respectively of errors at data points, are equally valued regardless of whether they are close to zero or to extremal values. This criterion is suitable for use if the errors are subjected to outliers or wild points, because the magnitude of big errors does not lead to the difference of the best approximations (cf. [?]).

The theory of l_1 approximations for the finite data points sets is somewhat more different with respect to the properties of continuous l_1 approximations on the interval, what is opposite to the situation in l_2 and l_∞ approximations, where there is not a significant difference between analysis of 'continuous' and 'discrete' cases (cf. [?]).

Since the l_1 norm is not strictly convex, the best l_1 linear approximation is not necessarily unique. The problem of linear l_1 approximation can be transformed into a linear programming problem. Linear l_1 approximation (regression) is widely analyzed (cf. [?], [?], [?], [?]). Nonlinear approximation in the l_1 norm is less researched. In a discrete nonlinear case, best approximations with respect to the l_1 norm need not necessarily exist, as well as with respect to the l_2 and l_∞ norm (cf. [?], [?]).

3. New approaches

a) Total least squares method. If the given data contain additive errors in both the dependent (ε_i, $i = 1, \ldots, m$) and the independent variable (δ_i), then one can observe orthogonal distances $d_i^2 = \varepsilon_i^2 + \delta_i^2$ and their sum

$$\Phi(a_1, \ldots, a_n, \delta_1, \ldots, \delta_m) = \sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} \Big((F(\mathbf{a}, x_i + \delta_i) - f_i)^2 + \delta_i^2 \Big),$$

where it is necessary to find minimum $\min\limits_{(a_1,\ldots,a_n,\delta_1,\ldots,\delta_m)} \Phi(a_1, \ldots, a_n, \delta_1, \ldots, \delta_m)$.

The idea of total least squares can be generalized by introducing a total least l_p norm, or rather the total l_p approximation problem

$$\min_{(a_1,\ldots,a_n,\delta_1,\ldots,\delta_m)} \sum_{i=1}^{m} \Big((F(\mathbf{a}, x_i + \delta_i) - f_i)^p + \delta_i^p \Big),$$

(cf. [**?**], [**?**], [**?**]).

b) Visual error criteria. The use of standard mathematical norms in vector spaces for measuring the distance between the true curve and the estimated curve can be inappropriate from the graphic viewpoint, in respect of a visual image about distances between the curves in a plane.

Therefore, the ideas about 'qualitative smoothing' have been appearing, by which the curves (i.e. functions) are estimated through qualitative features. In such a case, a visual image of distances between the curves is taken into consideration by means of nonstandard error criteria (cf. [**?**]). The curves are viewed as the sets of points in a plane. The distance between the set G_f of points of the true curve f and the set $G_{\overline{f}}$ of points of the estimated curve $\overline{f}$ are observed. For example, one uses the Hausdorff distance defined by

$$d_H(G_f, G_{\overline{f}}) = \max\{\sup \mathcal{D}(G_f, G_{\overline{f}}), \sup(\mathcal{D}(G_{\overline{f}}, G_f)\}\,,$$

where $\mathcal{D}(G_f, G_{\overline{f}}) = \{d((x,y), G_{\overline{f}}) : (x,y) \in G_f\}$ is the set of distances from the points of the set G_f to the set $G_{\overline{f}}$, and $d((x,y),G) = \inf_{(x',y') \in G} \|(x,y) - (x',y')\|_2$ denotes the distance from the point (x,y) to the set G.

Furthermore, one can define various so-called asymmetric error criteria and symmetric error criteria. Although these criteria seem to be good from the 'visual impression' viewpoint, in certain situations the l_2 norm has an advantage for the use due to its important optimal properties (cf. [**?**]).

References

[1] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[2] P. Bloomfield, W. L. Steiger, *Least Absolute Deviations – Theory, Applications and Algorithms*, Birkhäuser, Boston, 1983.

[3] I. Barrodale, F. D. K. Roberts, *An efficient algorithm for discrete l_1 linear approximation with linear constraints*, SIAM J. Numer. Anal. **15**(1978), 603–611.

[4] I. Barrodale, F. D. K. Roberts, C. R. Hunt, *Computing best l_p approximations by functions nonlinear in one parameter*, The Computer Journal, **13**(1970) 382–386.

[5] P. T. Boggs, R. H. Byrd, R. B. Schnabel, *A stable and efficient algorithm for nonlinear orthogonal distance regression*, SIAM J. Sci. Stat. Comput. **8**(1987), 1052–1078.

[6] E. W. Cheney, *Introduction to Approximation Theory*, Chelsea Publishing Company, New York, 1966.

[7] J. E. Dennis Jnr., *Non-linear Least Squares and Equations*, in: The State of the Art in Numerical Analysis, Jacobs, Ed., 1977., 269–309.

[8] J. E. Dennis Jnr., R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Inc., Englewood Cliffs, 1983.

[9] Y. Dodge (Ed.), *Statistical data analysis based on the L_1-norm and related methods*, 1. International Conference on Statistical Data Analysis based on the L_1-norm and Related Methods, Elsevier, Amsterdam, 1987.

[10] C. F. Gauss (translated by G. W. Stewart), *Theory of the Combination of Observations Least Subject to Errors*, SIAM, Philadelphia, 1995.

[11] G. H. Golub, C. V. Loan, *Total Least Squares*, in: Smoothing Techniques for Curve Estimation, Th. Gasser and M. Rosenblatt, Eds., Lecture Notes in Mathematics 757, Springer Verlag, Berlin, 1979, 69–76

[12] G. H. Golub, V. Pereyra, *The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems whose Variables Separate*, SIAM J. Numer. Anal. **10**(1973), 413–432.

[13] R. Gonin, *Numerical algorithms for solving nonlinear L_p-norm estimation problems: part I – a first-order gradient algorithm for well-conditioned small residual problems*, Commun. Statist. -Simula. **15**(1986), 801–813.

[14] J. S. Marron, A. B. Tsybakov, *Visual Error Criteria for Qualitative Smoothing*, Journal of the Amer. Stat. Assoc. Vol. **90**(430), 1995, 499–507.

[15] M. J. D. Powell, *Approximation Theory and Methods*, Cambridge Univ. Press, Cambridge, 1981.

[16] J. R. RICE, *The Approximations of Functions, Vol. 1 – Linear Theory*, Wesley, Reading, 1964.

[17] A. RUHE, P. A. WEDIN, *Algorithms for Separable Nonlinear Least Squares Problems*, SIAM Review **22**(1980), 318–337.

[18] H. SPÄTH, *Mathematical Algorithms for Linear Regression*, Academic Press Inc., Boston, 1992.

[19] A. TARANTOLA, *Inverse Problem Theory – Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, 1987.

[20] Y. ZI-QIANG, *On some computation methods of the nonlinear L_1 norm regression*, J. Num. Method & Comp. Appl., 1994, 18–23.

[21] G. A. WATSON, *The Numerical Solution of Total l_p Approximation Problems*, in: Numerical Analysis, D. F. Griffiths , Ed., Numerical Analysis, Lecture Notes in Mathematics 1066, Springer Verlag, Berlin, 1984, 221–238