On Comparison of Two Notions in Theory of Machine Learning: Sign Sequence and Steinitz

S.A.Chobanyan, Niko Muskhelishvili Institute of Computational Mathematics, Georgian Technical University

.

1. **The main aim** of this talk is to apply our maximum inequality and transference theorem [1,2] - presented in the next section – to the following problem which is an important subtask of many problems of machine learning, scheduling theory and discrepancy theory [3-8].

Find or estimate the minimum in π the *Steinitz functional*

$$\Phi_x(\pi) = \max_{1\le k\le n} \| \sum_1^k x_{\pi(i)} \|,$$

where $x = (x_1, \dots, x_n)$ is a fixed collection of elements of a normed space X, and $\pi: \{1,\dots,n\} \to \{1,\dots,n\}$ is a permutation.

The peculiarity of the related applied problems is that *d*, the dimension of X, and the number *n* of summands are very large, so that the *brute force* idea as a rule does not work.

The problem was posed by E.Steinitz [9] who was solving the question on sum range of a conditionally convergent series in a finite dimensional space (the generalization of the famous Riemann problem). If time permits we'll come back to this and other related nice analytical problems (among them to the open so far Kolmogorov-Garsia problem on the system of almost everywhere convergence of a rearranged orthogonal sequence).

Literature: *[1] S.Chobanyan, G.Giorgobiani, Lecture Notes in Math.,1391, 1987, 33-46;[2] S.Chobanyan, Birkhauser, Prog. Probab. 35, 1994, 3-29; [3] S.Sevastyanov, Discrete Applied Math.; [4] J.Beck, V.T.Sos, Handbook in Combinatorics, v.2, Elsevier Sciernce B.V. and MIT Press, 1995; [5] N.Makai, Appl. Math. Comp., 150, 2004, 785-801; [6] L.Chobanyan, S.Chobanyan, Giorgobiani, Bull.*

Georgian National Acad. Sci., 5, 2011, 16-21; [7] L.Chobanyan, V.V.Kvaratskhelia, 9th International Conference CSIT-2013, Yerevan, Armenia, Proceedings: pp.58-60, 2013; [8]N.Harvey, S.Samadi, Workshop and Conference Proceedings 35, 1-18;2014; [9] E.Steinitz, J.Reine Angew. Math., 143 (1913), 128-175.

2. The main maximum inequalities. In this talk we apply the following two maximum inequalities to problems related to the calculation or estimation of the Steinitz functional.

Theorem 1. [1,2] ***Let*** $x_1, \dots, x_n \in X$ ***be a collection of elements of a normed space*** X *with* $\sum_1^n x_i = 0$.
Then

a. ***For any collection of signs*** $\vartheta = (\vartheta_1, \dots, \vartheta_n)$ ***and any permutation*** $\pi : \{1,\dots,n\} \to \{1,\dots,n\}$ ***there is a permutation pi star such that***

$$\max_{1\le k\le n} \| \sum_1^k x_{\pi(i)} \| + \max_{1\le k\le n} \| \sum_1^k \vartheta_i x_{\pi(i)} \| \ge$$
$$2\max_{1\le k\le n} \| \sum_1^k x_{\pi^*(i)} \|, \quad \textit{where } \pi^* \textit{ depends on both } \pi$$
and ϑ.

.

b. ***(Transference Theorem). There is a permutation*** $\sigma: \{1,\ldots,n\} \to \{1,\ldots,n\}$ ***such that***

$$\max_{1\le k\le n} \| \sum_1^k x_{\sigma(i)} \| \le \max_{1\le k\le n} \| \sum_1^k \vartheta_i x_{\sigma(i)} \|$$

for any collection of signs $\vartheta = (\vartheta_1, \ldots, \vartheta_n)$ ***.***

Literature. [1] S.Chobanyan, G.Giorgobiani, Lecture Notes in Math.,1391, 1987, 33-46;[2] S.Chobanyan, Birkhauser, Prog. Probab. 35, 1994, 3-29;

3. **The greedy algorithm is not in general the best. Given vectors** $x_1, ..., x_n \in X$ ***a greedy algorithm*** **chooses at each step a vector that minimizes the norm of the next partial sum. In other words, on step 1 it chooses an element** x_{n_1} **that has a minimum norm. On step 2 it selects an element** x_{n_2}, $x_{n_2} \neq x_{n_1}$ **such that** $\| x_{n_1} + x_{n_2} \| \leq \| x_{n_1} + x_{n_k} \|$ **for any** $n_k \neq n_1$ **, etc.**

The following example constructed by Jakub Wojtaszchik (oral communication) shows that a greedy algorithm is not in general the best one even in a two-dimensional space.

Example. Consider n groups of vectors of l_∞^2 each consisting of the three following vectors: (1,1), (2,-3), and (-3,2). Obviously, the greedy algorithm chooses at the first n steps the vectors (1,1),…, (1,1). Therefore, for the optimal permutation π_o and greedy permutation π_g we have respectively

$$\max_{1\le k\le n} \| x_{\pi_o(1)} + \ldots + x_{\pi_o(k)} \| = 3;$$

and $$\max_{1\le k\le n} \| x_{\pi_g(1)} + \ldots + x_{\pi_g(k)} \| = n+2.$$

In [1] we show that such sort of an example can be constructed in any 2-dimensional normed space.

Literature. [1] G.Chelidze, S.Chobanyan, G.Giorgobiani, V.Kvaratskhelia, Bull. Georgian National Acad. Sci, 4, 2010, 5-7.

Let us introduce now important constants

ssc – Sign Sequence Constant and Sc - Steinitz constant:

$$\boldsymbol{ssc} = \boldsymbol{sup_x\, inf_{\vartheta}\, |x\vartheta|}$$

$$\boldsymbol{Sc} = \boldsymbol{sup_x\, inf_{\pi}\, |x_{\pi}|}$$

In the definition of the Steinitz constant supremum is taken over all collections x taken from the unit ball and summing up to zero.

In the definition of *ssc* *x* denotes a finite collection of elements from the unit ball of *X*.

In the definition of *Sc* it is additionally assumed that these elements sum up to one.

Here is an important corollary to Theorem 1b.

Theorem 2. For any fixed finite-dimensional real normed space X we have.

$$Sc \leq ssc \ .$$

Proof. According to Theorem 1 we have for any $\varepsilon > 0$:

$$Sc - \varepsilon < |x_{\pi\vartheta}| \leq ssc$$

(the first inequality holds for any for any ϑ).

4. Corollaries to the Transference theorem (Theorem 1b).

The Transference theorem allows us to get a permutation theorem given a sign theorem. Moreover, as we'll see in Section 5, if the sign algorithm is constructive, then a desired permutation can also be found constructively. As a first example we consider the classical Steinitz permutation theorem that we get from the following Grinberg-Sevostyanov sign theorem .

Theorem 2 [1,2]. *Let* ***X*** *be a normed space of dimension* d, $x_1,\ldots,x_n \in X$, $\| x_i \| \le 1$, $i = 1, \ldots, n$. *Then there exists a collection of signs* $\vartheta = (\vartheta_1,\ldots,\vartheta_n)$ *such that*

$$\max_{1\le k\le n} \| x_1\vartheta_1 + \ldots + x_k\vartheta_k \| \le 2d \text{ .}$$

The permutation version of Theorem 2 found by the Transference theorem can be stated as follows.

Corollary. *The Steinitz inequality. Let* X *be a normed space of dimension* $d,\ x_1,\ldots,x_n \in X,\ \|x_i\| \le 1,\ \ i=1,\ldots,n$ *and* $x_1+\ldots+x_n=0.$ *Then there exists a permutation* $\sigma:\{1,\ldots,n\}\to\{1,\ldots,n\}$ *such that*

$$\max_{1\le k\le n} \| x_{\sigma(1)}+\ldots+x_{\sigma(k)} \| \le 2d.$$

Remark. Steinitz [3] proved his inequality straightforwardly, however a proof through the sign version additionally allows to find the desired permutation constructively provided that the collection of signs in the sign version can be obtained constructively, by use of the Transference theorem (see Section 5).

Literature. [1] V.S.Grinberg and S.V.Sevastyanov, Funkts. Analiz I Prilozh, 14, 1980, 56-57 (in Russian); [2] I.Barani and V.S.Grinberg, Linear Algebra Appl., 41, 1981, 1-9; [3] E.Steinitz, J.Reine Ang. Mathematik, 143, 1913, 128-175, [4] J.Spencer, J.Combinatorial Theory, Ser. B, 23, 1977, 68-74.

4. Corollary to the Transference theorem. In **this section we show that the algorithm for near optimal permutation for $\Phi_x(\pi) = \max_{1\le k\le n} \| \sum_1^k x_{\pi(i)} \|$, can be reduced to the algorithm for near optimal sign. The reduction is based on the Transference theorem (Theorem 1b).**

Theorem 4. ***Let*** X ***be a normed space,*** $x_1,\dots,x_n \in X,\ \| x_i \| \le 1,\ i = 1, \dots, n,$ *and* $x_1 + \dots + x_n = 0$

Assume that for any permutation $\pi: \{1, \dots, n\} \to \{1, \dots, n\}$ ***there is an algorithm with a polynomial complexity to define*** $\vartheta = (\vartheta_1, \dots, \vartheta_n)$ ***such that***

$$\max_{1\le k\le n} \| \sum_1^k x_{\pi(i)} \vartheta_i \| \le D, \qquad (1)$$

where D *does not depend on* π.

Then for any $\varepsilon > 0$ ***there is an algorithm*** $\sigma: \{1, \dots, n\} \to \{1, \dots, n\}$ ***with a polynomial complexity to define such that***

$$\max_{1\le k\le n} \| \sum_1^k x_{\sigma(i)} \| \le D + \varepsilon \quad .$$

The complexity of the algorithm is $(C \cdot \log(n/\varepsilon)$***, where*** C ***is the complexity of the sign algorithm (for finding*** ϑ).

Literature. [1] N.Makai, Appl. Math. Comp., 150, 2004, 785-801;

[6] L.Chobanyan, S.Chobanyan, Giorgobiani, Bull. Georgian National Acad. Sci., 5, 2011, 16-21; [7] L.Chobanyan, V.V.Kvaratskhelia, 9th International Conference CSIT-2013, Yerevan, Armenia, Proceedings: pp.58-60, 2013; [8]N.Harvey, S.Samadi, Workshop and Conference Proceedings 35, 1-18;2014;