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 \( \pi \) the Steinitz functional

\[
\Phi_x(\pi) = \max_{1 \leq k \leq n} \| \sum_{i=1}^{k} x_{\pi(i)} \|
\]

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

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 \( x_1, \ldots, x_n \in X \) be a collection of elements of a normed space \( X \) with \( \sum_1^n x_i = 0 \). Then

a. For any collection of signs \( \mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_n) \) and any permutation \( \pi: \{1, \ldots, n\} \rightarrow \{1, \ldots, n\} \) there is a permutation \( \pi^* \) such that

\[
\max_{1 \leq k \leq n} \left\| \sum_{i=1}^k x_{\pi(i)} \right\| + \max_{1 \leq k \leq n} \left\| \sum_{i=1}^k \mathcal{G}_i x_{\pi(i)} \right\| \geq 2 \max_{1 \leq k \leq n} \left\| \sum_{i=1}^k x_{\pi^*(i)} \right\|, \quad \text{where } \pi^* \text{ depends on both } \pi \text{ and } \mathcal{G}.
\]
b. (Transference Theorem). There is a permutation \( \sigma: \{1, \ldots, n\} \rightarrow \{1, \ldots, n\} \) such that

\[
\max_{1 \leq k \leq n} \left\| \sum_{i=1}^{k} x_{\sigma(i)} \right\| \leq \max_{1 \leq k \leq n} \left\| \sum_{i=1}^{k} \varrho_{i} x_{\sigma(i)} \right\|
\]

for any collection of signs \( \varrho = (\varrho_{1}, \ldots, \varrho_{n}) \).

3. The greedy algorithm is not in general the best. Given vectors \( x_1, \ldots, x_n \in X \), 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
\[
\| x_{n_1} + x_{n_2} \| \leq \| x_{n_1} + x_{n_k} \| \quad \text{for any} \quad n_k \neq n_1, \quad \text{etc.}
\]

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 $l^2_{\infty}$ 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 $\pi_o$ and greedy permutation $\pi_g$ we have respectively

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

and

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

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

Let us introduce now important constants $ssc$ – Sign Sequence Constant and $Sc$ – Steinitz constant:

$$ssc = \sup_x \inf_\theta |x\theta|$$
$$Sc = \sup_x \inf_\pi |x\pi|$$

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 $\vartheta$).
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 $d$, $x_1,...,x_n \in X$, $\|x_i\| \leq 1$, $i = 1, ..., n$. Then there exists a collection of signs $\mathcal{G} = (\mathcal{G}_1,...,\mathcal{G}_n)$ such that

$$\max_{1 \leq k \leq n} \|x_1 \mathcal{G}_1 + ... + x_k \mathcal{G}_k\| \leq 2d.$$
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 $d$, $x_1,...,x_n \in X$, $\|x_i\| \leq 1$, $i = 1, ..., n$ and $x_1 + ... + x_n = 0$. Then there exists a permutation $\sigma: \{1, ..., n\} \to \{1, ..., n\}$ such that

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

**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).
4. Corollary to the Transference theorem. In this section we show that the algorithm for near optimal permutation for
\[ \Phi_x(\pi) = \max_{1 \leq k \leq n} \| \sum_{i=1}^{k} x_{\pi(i)} \|, \]
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, \( x_1, ..., x_n \in X, \| x_i \| \leq 1, \ i = 1, ..., n, \) and \( x_1 + ... + x_n = 0. \)

Assume that for any permutation \( \pi : \{1, ..., n\} \rightarrow \{1, ..., n\} \) there is an algorithm with a polynomial complexity to define \( \mathcal{G} = (\mathcal{G}_1, ..., \mathcal{G}_n) \) such that
\[ \max_{1 \leq k \leq n} \| \sum_{i=1}^{k} x_{\pi(i)} \mathcal{G}_i \| \leq D, \]
where \( D \) does not depend on \( \pi \).

Then for any \( \varepsilon > 0 \) there is an algorithm \( \sigma : \{1, ..., n\} \rightarrow \{1, ..., n\} \) with a polynomial complexity to define such that
\[ \max_{1 \leq k \leq n} \| \sum_{i=1}^{k} x_{\sigma(i)} \| \leq D + \varepsilon. \]
The complexity of the algorithm is \( (C \cdot \log(n/\varepsilon)) \), where \( C \) is the complexity of the sign algorithm (for finding \( \mathcal{G} \)).