

# Questions and Answers that somewhat lost relevance

Vasily Ilin

July 8, 2020

**Question 0.1.** Is there a relation between the error if  $\{e_1, \dots, e_k\} \subset V$ , and the error if  $\{e_1, \dots, e_k\}$  are arbitrary?

*Answer.* We will show that there is such a relationship in the special case when  $k = 1$  and all vectors in  $V$  are of length 1.

Let  $S \subset V$ , and let  $e$  be the principal vector (minimizing sum of square of distances). Let  $d(x, y)$  be the distance from  $x$  to the projection of  $x$  onto  $y$  (note that it is not a metric). Let  $v \in S$  be the closest vector to  $e$ , so that  $d(v, e) \leq d(w, e)$  for any  $w \in S$ . We will show that  $\sum_{w \in S} d(w, v)^2 \leq 4 \sum_{w \in S} d(w, e)^2 = 4\text{OPT}$ , where  $\text{OPT} = e_1(S)$  is the optimal (smallest) error.

**Lemma 0.2.** For any three unit-length vectors  $e, v, w$ , we have  $d(v, w)^2 \leq 2(d(v, e)^2 + d(e, w)^2)$ , similar to the triangle inequality.

Once we prove the lemma we will see that

$$\sum_{w \in S} d(w, v)^2 \leq 2 \sum_{w \in S} d(w, v)^2 + 2 \sum_{w \in S} d(v, e)^2 \leq 4 \sum_{w \in S} d(w, v)^2 = 4\text{OPT} \quad (1)$$

We shall now prove the lemma. First, observe that for any two vectors  $v, w$  we have

$$d(v, w)^2 = \|v\|^2 - \|\text{proj}_w(v)\|^2 = \|v\|^2 - \langle v, w \rangle^2 / \|w\|^2$$

by Pythagorean theorem. But

$$\|v\|^2 - \langle v, w \rangle^2 / \|w\|^2 = 1 - \cos^2(\gamma) = \sin^2(\gamma),$$

where  $\gamma$  is the angle between  $v$  and  $w$ . Let  $\alpha$  be the angle between  $v$  and  $e$ , and  $\beta$  be the angle between  $e$  and  $w$ . Since we do not care about the direction of vectors here, we may regard the angles between vectors as angles between lines, and restrict all angles to  $[0, \pi/2]$ . Now observe that  $\gamma \leq \alpha + \beta$ , meaning  $\sin(\gamma) \leq \sin(\alpha + \beta)$ , since  $\sin$  is increasing on  $[0, \pi/2]$ . Thus, it suffices to show  $\sin^2(\alpha + \beta) \leq 2(\sin^2(\alpha) + \sin^2(\beta))$ . In order to see that this is true

define  $f(\alpha, \beta) = 2(\sin^2(\alpha) + \sin^2(\beta)) - \sin^2(\alpha + \beta)$  and minimize it. Taking the gradient of  $f$  and equating it to zero, we see that  $4\sin(\alpha)\cos(\alpha) = 2\sin(\alpha + \beta)\cos(\alpha + \beta) = 4\sin(\beta)\cos(\beta)$ , meaning  $\sin(2\alpha) = \sin(2\beta) \implies \alpha = \beta$ , since we restricted the angles to  $[0, \pi/2]$ . Thus, minimizing  $f$  is equivalent to minimizing  $g(\alpha) = 4\sin^2(\alpha) - \sin^2(2\alpha)$ . Now note that  $g$  is continuous, and by the Intermediate Value Theorem, if it takes negative values (which we want to avoid), there exists  $\alpha^*$  with  $g(\alpha^*) = 0$ . We see that

$$g(\alpha^*) = 0 \iff (2\sin(\alpha^*) - \sin(2\alpha^*))(2\sin(\alpha^*) + \sin(2\alpha^*)) = 0 \iff 2\sin(\alpha^*) - \sin(2\alpha^*)$$

because  $\alpha^* \in [0, \pi/2]$ . We see that  $\alpha^* = 0$  is the only solution to the above, because if  $\alpha^* \neq 0$ , then  $2\sin(\alpha^*) - \sin(2\alpha^*) \iff 2 = \frac{\sin(2\alpha^*)}{\sin(\alpha^*)} = \frac{2\cos(\alpha^*)\sin(\alpha^*)}{\sin(\alpha^*)} = 2\cos(\alpha^*) \implies \alpha^* = 0$ , which is a contradiction, since we assumed that  $\alpha^* \neq 0$ . Therefore,  $g \geq 0$ , and  $f \geq 0$ , meaning  $\sin^2(\alpha + \beta) \leq 2(\sin^2(\alpha) + \sin^2(\beta))$ , as required.  $\square$

*Answer from [1].* Theorem 1.4 of [1] states that for any matrix  $A$  and  $\epsilon > 0$ , there exist  $k + k(k+1)/\epsilon$  rows whose span contains the rows of a rank- $k$  matrix  $\tilde{A}_k$  such that

$$\|A - \tilde{A}_k\|^2 \leq (1 + \epsilon)\|A - A_k\|^2 = (1 + \epsilon)\text{OPT}$$

$\square$

*Alternative proof.* The proof above can be simplified as follows. We will use lemma

**Lemma 0.3.** For any three unit-length vectors  $e, v, w$ , we have  $d(v, w)^2 \leq 4d^2$ , where  $d = \max\{d(v, e), d(e, w)\}$ ,

which can be proven by squaring the triangle inequality  $d(v, w) \leq d(v, e) + d(e, w)$ . The triangle inequality holds because distance  $d$  is the Hausdorff distance on the Grassmanian manifold  $\mathbf{Gr}(\mathbf{m}, \mathbf{1})$ , the 1-dimensional Grassmanian on  $\mathbb{R}^m$ . Inequality 1 then becomes

$$\sum_{w \in S} d(w, v)^2 \leq 4 \sum_{w \in S} d_w = 4 \sum_{w \in S} d(w, e)^2 = 4\text{OPT}, \quad (2)$$

where  $d_w = \max\{d(v, e), d(e, w)\}$ , and  $d_w = d(w, e)$ , by our choice of  $v$ .  $\square$

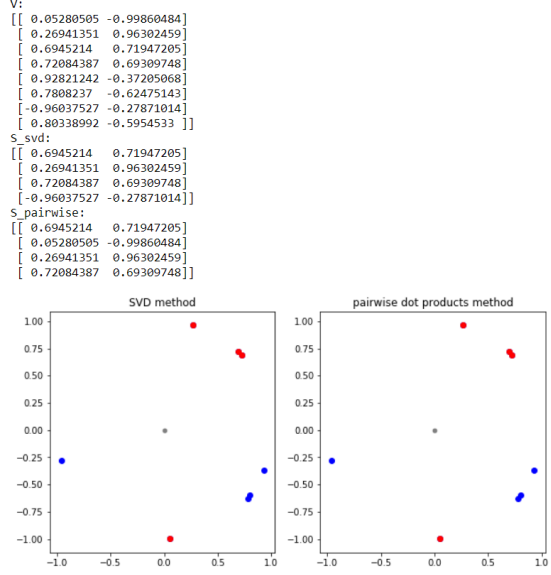
**Question 0.4.** Is the problem of minimizing  $e_1(S)$  equivalent to the problem of maximizing the sum of pairwise dot products  $\sum_{v, w \in S} \langle v, w \rangle$ ? What about squares of dot products  $\sum_{v, w \in S} \langle v, w \rangle^2$ ?

*Answer.* Neither is equivalent to minimizing  $e_1(S)$ .

The answer to the first question is definitely no. Figure 1 shows an explicit example.

An explicit example for the second question is:  $S_1 = \{(1, \epsilon), (1, -\epsilon)\}$ ,  $S_2 = \{(2\epsilon, 1)\alpha, (-2\epsilon, 1)\alpha\}$ , and  $V = S_1 \cup S_2$ , with  $\epsilon = 0.1, \alpha = 1.01$ . Then  $S_1$  maximizes sum of squares of pairwise dot products and  $S_2$  maximizes  $e_1(S)$ .  $\square$

Figure 1:



**Question 0.5.** We want to find  $S \subset V$  of size  $n$  and  $E \subset \mathbb{R}^m$  to maximize

$$\sum_{w \in S} \|\text{proj}_{\text{span}(E)}(w)\|^2.$$

If  $E$  is given, can find  $S$  easily (iterating through all  $v \in V$  and choosing  $n - k$  of them with biggest length of projection onto  $\text{proj}_{\text{span}(E)}$ ). If  $S$  is given, can approximately find  $E$  (SVD). What about both at the same time? Some kind of analogue of  $EM$ , perhaps?

*Answer.* Here is an algorithm similar to  $EM$ .

---

**Algorithm 1** FindEandS

---

- 1: Pick random  $E \subset \mathbb{R}^m$  by picking each coordinate of  $e_i$  uniformly at random from  $[-1, 1]$ , and normalizing each  $e_i$ . We can Gram-Schmidt  $E$  but it's not necessary.
  - 2: **while** true **do**:
  - 3:     Find  $S$  for  $E$  (by iterating through all  $V$  and checking  $d(v, \text{span}(E))$ )
  - 4:     Find  $E$  for this  $S$  (via SVD)
  - return**  $S, E$ .
- 

**Are there any guarantees on this?**

□

**Question 0.6** (Similar Problem). Projective clustering is the problem of finding a set of  $j$   $k$ -dimensional subspaces

$F_1, \dots, F_j$  such that

$$\sum_{p \in P} \min_i d(p, F_i)$$

for a set of points  $P \subset \mathbb{R}^m$ . This is known to be **NP**-hard [1]. The subspaces partition  $P$  into parts  $P_1, \dots, P_j$ , where  $P_i$  is the set of points closest to  $F_i$ . If  $P_i$ 's were required to be of a specific size (e.g. all of them of size  $n$ ), then solving problem ??  $N/n$  times would solve the modified projective clustering problem. Thus, most likely, problem ?? is **NP**-hard.

**Question 0.7** (Basic SVD questions). Given matrix  $A = (a_1, \dots, a_n)$  and its  $k$ -approximation  $A_k$  with (left) principal vectors  $E = e_1, \dots, e_k$ , is  $\|A - A_k\| = \sum_{i \leq n} \|a_i - \text{proj}_{\text{span}(E)}(a_i)\|^2$ ? Is  $e_1 = \frac{1}{n} \sum_{i \leq n} a_i$ ? Note that  $\text{proj}_{\text{span}(E)}(a_i) = \sum_{j \leq k} \langle a_i, e_j \rangle e_j$ .

*Answer to 0.7.* The answer to the first question above is “yes”. Let’s consider the case when  $A$  is mean-centered (mean of each row is 0), in which case  $e_1 \neq \frac{1}{n} \sum_{i \leq n} a_i = 0$ . First, consider  $k = 1$ . For a given  $i$ ,  $\|a_i - e_1 \langle a_i, e_1 \rangle\|^2 = \|a_i\|^2 - \langle a_i, e_1 \rangle^2$  (just algebra). So  $\sum_{i \leq n} \|a_i - e_1 \langle a_i, e_1 \rangle\|^2 = \sum_{i \leq n} \|a_i\|^2 - \langle a_i, e_1 \rangle^2 = \|A\|^2 - \sum_{i \leq n} \langle a_i, e_1 \rangle^2$ . We know that  $e_1$  maximizes variance in the direction  $e_1$ , which is given by  $\sum_{i \leq n} \langle a_i, e_1 \rangle^2$ , meaning  $e_1$  minimizes  $\sum_{i \leq n} \|a_i - e_1 \langle a_i, e_1 \rangle\|^2$ . I think the rest ( $k > 1$ ) follows by induction (define the new  $A$  to be  $A - A_1$ ).

The answer to the second question is “no”! For the case when  $A$  is not mean-centered, I produced an explicit counter-example (generated a random matrix and tried it). For the case when  $A$  is mean-centered,  $\frac{1}{n} \sum_{i \leq n} a_i = 0$ , and so cannot equal  $e_1$ , since  $e_1$  is non-zero. Here is a specific example: let  $A = \left( (-1, k), (1, k) \right)^T$ . Note that the mean of the two vectors is  $\mu = (0, k)$ . If  $k > 1$ , the first principal vector is  $(0, 1) = \mu / \|\mu\|$ . But if  $k < 1$ , the first principal vector is  $(1, 0)$ , orthogonal to the mean  $\mu = (0, k)$ . If  $k = 1$ , both principal vectors have equal singular values.  $\square$

**Question 0.8.** What happens to the low-rank approximation of a matrix when I replace one vector by another?

## 1 Research Summary

Evimaria Terzi and I, Vasily Ilin, are interested in detecting dense low-rank subgraphs. The setup is as follows: we are given a graph  $G = (V, E)$ , where  $V$  is a subset of  $m$ -dimensional Euclidean space  $\mathbb{R}^m$ . Given a subset  $S$  of  $V$ , one can ask for the “effective rank” of the matrix  $A$  spanned by  $S$  (matrix rows are given by  $S$ ). The notion of effective rank is formalized by requiring  $A$  to be close to its best  $k$ -rank approximation  $A_k$  in the Frobenius norm. If the effective rank  $k$  is small (less than size of  $S$ ), vectors in  $S$  are almost spanned by  $k$  vectors. The problem we are interested in is developing algorithms to detect a dense (in average degree) subgraph  $S$  of  $G$  that has small effective rank.

One possible application of this is in social (and other) networks. Consider a network of researchers, where each researcher is associated to a vector of topics of their research interests. And an edge between two researchers exists

if the two researchers have collaborated. Then we might be interested in finding "communities" of researchers who collaborate with each other, and whose topic interests align well, where "align well" is quantified as having low effective rank.

This problem is a generalization of some existing problems with known good algorithms. For example, the problem of finding the densest (by average degree) subgraph has been shown to have a simple greedy algorithm that gives a 2-approximation (along with an exact algorithm using max-flow). We hope to build on this existing work to come up with algorithms for our problem and its variants.

If you would like more information on the exact questions we have posed and our progress so far, let us know. We will be happy to share what we have so far.