Matching on the line admits no $o(\sqrt{\log n})$-competitive algorithm

Enoch Peserico  
Università degli Studi di Padova, Italy

Michele Scquizzato  
Università degli Studi di Padova, Italy

Abstract

We present a simple proof that the competitive ratio of any randomized online matching algorithm for the line exceeds $\sqrt{\log_2(n+1)/15}$ for all $n = 2^i - 1 : i \in \mathbb{N}$, settling a 25-year-old open question.

2012 ACM Subject Classification  Theory of computation → Online algorithms

Keywords and phrases  Metric matching, online algorithms, competitive analysis


Related Version  Available at: https://arxiv.org/abs/2012.15593

Funding  Michele Scquizzato: supported, in part, by Univ. Padova grant BIRD197859/19.

Acknowledgements  We are indebted to Kirk Pruhs and the anonymous reviewers for their constructive criticism and insightful observations.

1 Online matching, on the line

In online metric matching [7, 9] $n$ points of a metric space are designated as servers. One by one $n$ requests arrive at arbitrary points of the space; upon arrival each must be matched to a yet unmatched server, at a cost equal to their distance. Matchings should minimize the ratio between the total cost and the offline cost attainable if all requests were known beforehand. A matching algorithm is $c(n)$-competitive if it keeps this ratio no higher than $c(n)$ for all possible placements of servers and requests.

It is widely acknowledged [1, 10, 14] that the line is the most interesting metric space for the problem. Matching on the line models many scenarios, like a shop that must rent to customers skis of approximately their height, where a stream of requests must be serviced with minimally mismatched items from a known store. Despite matching being specifically studied on the line since at least 1996 [8], no tight competitiveness bounds are known.

As for upper bounds, the line is a doubling space and thus admits an $O(\log n)$-competitive randomized algorithm [5]; a sequence of recent developments [1, 12, 13] yielded the same ratio without randomization. Better bounds have been obtained only by algorithms with additional power, such as that to re-assign past requests [6, 11] or predict future ones [2].

As for lower bounds, the competitive ratio is at least 4.591 for randomized algorithms and 9 for deterministic ones since the cow-path problem is a special case of matching on the line [8]. These bounds were conjectured tight [8] until a complex adversarial strategy yielded a lower bound of 9.001 for deterministic algorithms [4]. Beyond some $\Omega(\log n)$ bounds for restricted classes of algorithms [3, 10, 12], there has been no further progress on the lower-bound side before this work.
Matching on the line admits no $o(\sqrt{\log n})$-competitive algorithm

2 An $O(\sqrt{\log n})$-competitiveness bound

We prove a simple $\Omega(\sqrt{\log n})$ lower bound on the competitive ratio of randomized online matching algorithms for the line.

For any $n = 2^i + 1$ with $i \in \mathbb{N}$ consider the $[0, n+1]$ interval; for each positive integer $j \leq n$ place a server at point $j$, and place $n$ requests over $\log_2(n+1)$ rounds as follows. On the $r$th round (for $1 \leq r \leq \log_2(n+1)$) partition the interval into $(n+1)/2^r$ subintervals of length $2^r$, choose within each uniformly and independently at random an origin point, and place a request on the closest integer multiple of $2^{-n}$ breaking ties arbitrarily. “Discretizing” requests instead of directly using the corresponding origins prevents some technical difficulties – see our remark at the end.

We prove in Lemma 1 that the expected distance between the $\ell^{th}$ leftmost server and the $\ell^{th}$ leftmost origin is $O(\sqrt{\log n})$, so servers and requests can be matched with an expected offline cost $O(n \log n)$. Conversely, we prove in Lemma 2 that any online matching algorithm ALG incurs an expected $\Omega(\log n)$ cost in any given round, for a total cost $\Omega(n \log n)$. The two results can be combined to prove that on some request sequence ALG incurs $\Omega(\sqrt{\log n})$ times the offline cost.

Lemma 1. The expected distance between the $\ell^{th}$ leftmost origin and the $\ell^{th}$ leftmost server is at most $\sqrt{\log_2(n+1)} + 3$.

Proof. Let $S_\ell$ be the $\ell^{th}$ leftmost server and $g_\ell$ be the number of origins to its left. Note that if $g_\ell$ equals respectively $\ell$ or $\ell - 1$, the $\ell^{th}$ origin is the first immediately to the left, or to the right of $S_\ell$; and since the first round placed one origin in every subinterval of size 2, such an origin is within distance 3 of $S_\ell$. By the same token, denoting by $\delta_\ell$ the quantity $|g_\ell - (\ell - \frac{\ell}{n+1})|$, the $\ell^{th}$ leftmost origin is within distance $2\delta_\ell + 3$ of $S_\ell$. Note that $\delta_\ell$ is the absolute deviation from the mean of $r_\ell$, since $r_\ell$ is the sum of $n$ independent indicator random variables each denoting whether a given origin was placed to the left of $S_\ell$, with total expectation $\frac{n}{n+1}\ell = \ell - \frac{\ell}{n+1}$ (by construction, the expected density of origins is constant throughout the main interval). At most one such variable in a given round has variance greater than 0, albeit obviously at most $1/4$: that corresponding to the origin placed in a subinterval holding $S_\ell$ strictly in its interior. Adding the individual variances we obtain that the variance of $r_\ell$, i.e. the expectation of $\delta_\ell^2$, is at most $\log_2(n+1)/4$; and since by Jensen’s inequality $E[\delta_\ell^2] \leq E[\delta_\ell]^2$, the expected distance between $S_\ell$ and the $\ell^{th}$ leftmost origin is at most $\sqrt{\log_2(n+1)} + 3$.

Lemma 2. Any randomized online matching algorithm incurs an expected cost greater than $(n+1)/12$ in each round.

Proof. Consider an origin placed uniformly at random in a subinterval of size $2^r$ during the $r$th round. Assume $m$ unmatched servers in the interior points of that subinterval divide it into $m + 1$ segments of (integer) length $d_0, \ldots, d_m$. Then the probability the corresponding request falls within a segment of length $d$ is $d/2^r$, in which case the expected distance of the request from the segment’s closer endpoint is $d/4$. Adding over all the $s_r$ segments in all the round’s subintervals, applying Jensen’s inequality, and noting that $s_r$ does not exceed the number of subintervals (i.e. $(n+1)/2^r$) plus the total number of unmatched servers (i.e. $(n+1)/2^{r-1} - 1$), the expected cost to service all requests in the round is at least:

$$\sum_{h=1}^{s_r} \frac{d_h}{4} \cdot \frac{d_h}{2^r} \geq \frac{1}{4 \cdot 2^r} s_r \left( \frac{n+1}{s_r} \right)^2 \geq \frac{1}{3(n+1)} \cdot \frac{(n+1)^2}{4 \cdot 2^r} \cdot \frac{2^r}{3(n+1)} = \frac{n+1}{12}.$$
We can then easily prove the following:

**Theorem.** The competitive ratio of any randomized online matching algorithm for the line exceeds $\sqrt{\log_2(n+1)/15}$ for all $n = 2^i - 1 : i \in \mathbb{N}$.

**Proof.** Let $C_A(\sigma)$ be the expected cost incurred by a randomized online matching algorithm ALG on a request sequence $\sigma$, and $C_O(\sigma)$ the offline cost; and let $p_\sigma$ be the probability of generating $\sigma$ through the origin-request process described earlier. Since $\forall a_i, b_i > 0$ we have that $(\sum a_i)/(\sum b_i)$ is a convex linear combination of the individual ratios $a_i/b_i$, focusing on the case $\sqrt{\log_2(n+1)/15} \geq 1$ for which $\sqrt{\log_2(n+1) + 3 + 2^{-n}} < (5/4)\sqrt{\log_2(n+1)}$:

$$\max_{\sigma \neq 0} C_A(\sigma) = \sum_{\sigma \neq 0} C_A(\sigma) p_\sigma \geq \frac{(n+1) \log_2(n+1) / 12}{n(\sqrt{\log_2(n+1) + 3 + 2^{-n}})} > \frac{\sqrt{\log_2(n+1)}}{15}.$$

**Remark:** Without discretized requests the term $\sum_{\sigma \neq 0} C_A(\sigma) p_\sigma$ in the theorem’s proof would have been an integral, potentially ill-defined (for example, if ALG serviced requests for rational points in an interval with one server and for irrational points with another).

References