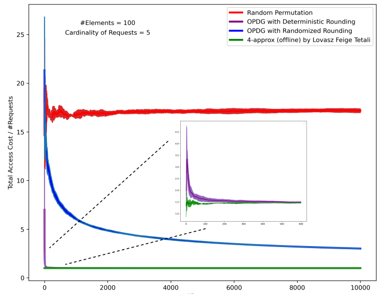
*We thank all the reviewers for their feedback, their apt comments and questions.* We will follow the comments of Reviewer 2 and 3 to ameliorate the write up and the suggestions of Reviewer 1 and 4 for experimental results.

**Reviewer 1:** The reviewer's comment on the practical efficiency of the ellipsoid method is a valid point. Unfortunately circumventing the ellipsoid method for the GMSSC seems a very hard task even for offline algorithms. The reason is that (to the best of our knowledge) there is no linear relaxation of GMSSC with polynomial description and constant integrality gap. However, for the important special case of *Min-Sum Set Cover* the subgradient can be computed via a quadratic-time algorithm. For the sake of simpler exposition we did not present it in the original draft but we will definitely add it. We have efficiently implemented in C++ both OPDG in doubly stochastic matrices and randomized /deterministic roundings. Our first experimental results (also requested by Rev. 4) reveal that the regret becomes very small quickly given random requests, while the deterministic rounding performs extremely better than its theoretical guarantees and even outperforms the randomized rounding. Finally concerning the reviewer's question on reducing the $n^{5.5}$ of the additive term, we conjecture that this is indeed possible (may be with some increase in the regret bound). Actually our experimental findings indicate that the right additive term is $n^2$. However, our current analysis is tight and novel ideas would be needed.

**Reviewer 2:** The reviewer's point on our novelty concerns the randomized rounding scheme (Algorithm 3) that comes from [36] and possibly Algorithm 4 that is based on the previous one. We just want to mention that a key contribution of the paper is combining a modification of the configuration LP in [30] (that differs form the LP in [36]) with the rounding scheme of [36] to obtain constant regret algorithms. So, we bring together the right ideas and techniques from previous work, after properly adapting them. Moreover, our deterministic rounding is novel. We find the reviewers question on optimizing the rounding in [36] for other special case of $\mathrm{K}(R)$ very interesting even for the offline case. Addressing it is, however, beyond our current scope, since our analysis concluding in Lemma 9 crucially uses the fact that $\mathrm{K}(R) = 1$.

**Reviewer 3: In the definition of the fractional access cost, the authors assume a fixed accuracy parameter, $\epsilon$, and ignore dependence of $\mathrm{FAC}_R(A)$ on $\epsilon$. What is the dependence on $\epsilon$?** The fractional access cost, $\mathrm{FAC}_R(A)$, can be defined for any value $\epsilon > 0$. The smaller the choice of $\epsilon$ is the better the regret bounds are $(2(1+\epsilon)r, 28(1+\epsilon)$ and $11.713(1+\epsilon))$ and the greater the additive term becomes $O\left(\frac{n^{5.5}}{\epsilon\sqrt{t}}\right)$. **What is the fractional access cost linear program? Why is it exponential in number of constraints? Is this the only possible formulation?** It is the linear program in Definition 2 which has exponential size since the number of different configurations is exponential in the number of elements. As already mentioned, we are not aware of any linear relaxation for GMSSC with polynomial description and constant integrality gap. **Where does the $n^4/\epsilon$ term come from?** The term $n^4/\epsilon$ plays two important roles: $i)$ it ensures an upper bound on the norm of the subgradients (which is necessary for the OPGD to run), and $ii)$ it associates the fractional access cost with the value of the (different) linear relaxation used in [6]. **Why is there a $5.03/\alpha$ in the conversion of doubly stochastic matrices to probability distribution over permutations?** The selection of $5.03$ is so as to minimize a parametric upper bound (see [36]). **Why doesn't any conversion that respects the marginals suffice?** Unfortunately, such intuitive rounding schemes do not always work. In fact there are cases where a probability distribution respecting the marginals of the doubly stochastic matrix leads to arbitrarily higher expected accessing cost than the fractional access cost of the doubly stochastic matrix. For example, for the matrix $A$ below, selecting with probability $1/2$ the permutation $\{1, 2, 3, 4, \ldots, n-2, n-1, n\}$ and with probability $1/2$ the permutation $\{n-1, n, 3, 4, \ldots, n-2, 1, 2\}$, respects the marginals. Let the request $R = \{1, 2\}$ with covering requirements $\mathrm{K}(R) = 1$. Its expected access cost is $\Theta(n)$ while its *fractional access cost* under $A$ is $\Theta(1)$. **The algorithms presented here give for the special case of $|R_t| = 1$?** The exact same bounds since they do not take into account the cardinality of the sets. However it is easy to design 2-regret algorithms for the special case where $|R_t| = 1$. **Can the relaxation of the GMSSC objective over the doubly stochastic matrices improve some known approximation bounds for the minimum linear ordering problem or its special cases?** The relaxation of the GMSSC cannot help in the more general linear ordering problem. The reason is that the dual cannot be solved in polynomial time, since the separation oracle crucially depends on the specific structure of GMSSC.




$$A = \begin{pmatrix} 1/2 & 0 & 0 & \cdots & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & \cdots & 0 & 0 & 1/2 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 1/2 & 0 & 0 & \cdots & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & \cdots & 0 & 0 & 1/2 \end{pmatrix}$$