Medians of permutations : building constraints *

Robin Milosz and Sylvie Hamel

DIRO - Université de Montréal , C. P. 6128 Succ. Centre-Ville, Montréal, Qc, Canada, H3C 3J7, {robin.milosz, sylvie.hamel}@umontreal.ca

Abstract. Given a set $\mathcal{A} \subseteq S_n$ of m permutations of [n] and a distance function d, the **median** problem consists of finding a permutation π^* that is the "closest" of the m given permutations. Here, we study the problem under the Kendall- τ distance which counts the number of pairwise disagreements between permutations. This problem has been proved to be NP-hard when $m \geq 4$, m even. In this article, we investigate new theoretical properties of \mathcal{A} that will solve the relative order between pairs of elements in median permutations of \mathcal{A} , thus drastically reducing the search space of the problem.

^{*} supported by NSERC through an Individual Discovery Grant (Hamel) and by FRQNT through a Master's scholarship (Milosz)

Appendix

Proof of Proposition ??: Let \mathcal{A} be the following set of permutations of [n]:

$$\mathcal{A} = \{\sigma_1, \sigma_2, \ldots, \sigma_a, \gamma, \pi_1, \pi_2, \ldots, \pi_a\},\$$

where each σ_i , $1 \leq i \leq a$, is a different permutation beginning with element 1 followed by element 2; each π_i , $1 \leq i \leq a$, is a different permutation ending with element 1 followed by element 2, and γ is a permutation beginning by element 2 and ending by element 1 (in γ elements 1 and 2 are thus separated by a set K of k elements). Note: n = k + 2.

So, in 2*a* permutations, element 1 is before element 2, and in only one, element 2 is before element 1. Even if the majority order for this pair of elements is 1 before 2, we can show that for an arbitrary *a*, if k > 2a - 1 then the median of \mathcal{A} will be of the form 2K1, more precisely the median will be $2\alpha 1$ where 2 is before 1 and where α is the optimal permutation of the elements of set *K* (i.e. $\pi^* = 2\alpha 1$ is a median of \mathcal{A}). The idea is that the "pressure" of the set *K* on elements 1 and 2 will be stronger than the interaction in between 1 and 2. We then choose a minimal *a* such that $\frac{2a}{2a+1} \geq s$ and k = 2a to complete the construction.

First, we observe that the cost of every element of K is identical in relation to 1 and 2 simply because the elements of K are always grouped together in the permutations of \mathcal{A} . So, in a certain permutation of \mathcal{A} , if an element $x \in K$ is on the left (resp. on the right) of 1 then all other elements of K will be on the left (resp. on the right) of 1. It goes the same way for the left/right of 2 and it applies in all permutations of \mathcal{A} .

Second, we can state that there exist at least one optimal arrangement for the elements of K based on the order they were placed in the permutations of \mathcal{A} . Let α be any optimal arrangement for the elements of K and C_{α} , its associated cost.

The two previous points are leading to this observation: a median permutation of \mathcal{A} will have the elements of K arranged following α . This is because of the fact that if a supposed median permutation have the elements of K arranged in an non-optimal way, we could make a bunch of swaps between those elements of K, omitting 1 and 2, so that their relative order will become an optimal arrangement, thus lowering the cost of the permutation.

Now, a median of \mathcal{A} can either be of the form $K_1 1 K_2 2 K_3$ or of the form $K_1 2 K_2 1 K_3$, where the K_i are possibly empty subsets of K, such that $\bigcup_{i=1}^3 K_i = K$, $K_i \bigcap K_j = \emptyset$ if $i \neq j$, $|K_i| = k_i$ and $k_1 + k_2 + k_3 = k$. The order of the elements of K is known and follows α , so the only variables to be set are the numbers of elements to be placed in K_1 , K_2 and K_3 . In other words, we have α and we need to place in 1 and 2 in an optimal way to attain the median.

If our median is of the form $K_1 1 K_2 2 K_3$ we have

$$d_{KT}(K_1 1 K_2 2 K_3, \mathcal{A}) = \overbrace{a(2k_1 + k_2)}^{1} + \overbrace{a(k_2 + 2k_3)}^{2} + \overbrace{1(1 + k_1 + 2k_2 + k_3)}^{3} + C_{\alpha}$$
$$= 2ak + 1 + k_1 + 2k_2 + k_3 + C_{\alpha},$$

where terms 1, 2 and 3 capture the distance with respectively the permutations σ_i , $1 \leq i \leq a$, the permutations π_i , $1 \leq i \leq a$, and the permutation γ of \mathcal{A} , for the pairs (1, x), (2, x) and (1, 2), $x \in K$. The cost for the pairs (x_1, x_2) , $x_1, x_2 \in K$, is counted in C_{α} .

Following the same logic, if our median is of the form $K_1 2 K_2 1 K_3$, we have

$$d_{KT}(K_1 2 K_2 1 K_3, \mathcal{A}) = a(2k_1 + k_2 + 1) + a(k_2 + 2k_3 + 1) + 1(k_1 + k_3) + C_\alpha$$
$$= 2ak + 2a + k_1 + k_3 + C_\alpha.$$

If we try to minimize $d_{KT}(K_11K_22K_3, \mathcal{A})$, we will have to put all the k elements of K in either K_1 or K_3 because putting them in K_2 will have a doubled penalty for the cost. Doing so, the minimal cost for $K_{11}K_{22}K_3$ is $2ak+1+k+C_{\alpha}$. Now, if we try to minimize $d_{KT}(K_{12}K_{21}K_3, \mathcal{A})$, we will have to put all the k elements of K in K_2 because putting them in either K_1 or K_3 will give a penalty whereas K_2 is "silent" in the cost function. So, the minimal cost for $K_{12}K_{21}K_3$ is $2ak + 2a + C_{\alpha}$. Since it is stated in the proposition that k > 2a - 1, the minimal cost of $K_{11}K_22K_3$ is strictly greater that the minimal cost of $K_{12}K_{21}K_3$ $(2ak + 1 + k + C_{\alpha} > 2ak + 2a + C_{\alpha})$.

This gives us that median permutations π^* of \mathcal{A} are of the general form $K_1 2 K_2 1 K_3$, where K_1 and K_3 are empty. Thus π^* is of the form 2K1 or more precisely $\pi^* = 2\alpha 1$, where α be any optimal arrangement for the elements of K with respect to \mathcal{A} . Element 2 is favored to element 1 in the median, regardless of the big proportion of permutations of \mathcal{A} favoring 1 to 2, which finish the proof of non-existence of a %-majority bound.

Other tables:

Table 1: Efficiency of the Major Order theorem 1.0 on sets of uniformly distributed random permutations, from n = 8 to n = 100, m = 3 to m = 50, statistics generated over 100 000 instances for smaller n to 2000 instances for bigger n (see Table 4)

$m \backslash n$	8	10	15	20	25	30	40	50	60	80	100
3	0.4560	0.4155	0.3603	0.3339	0.3170	0.3056	0.2914	0.2840	0.2785	0.2712	0.2667
4	0.3672	0.3224	0.2595	0.2274	0.2071	0.1939	0.1768	0.1673	0.1608	0.1514	0.1470
5	0.3076	0.2591	0.1928	0.1597	0.1403	0.1272	0.1110	0.1011	0.0943	0.0859	0.0823
10	0.1449	0.1004	0.0530	0.0344	0.0251	0.0198	0.0139	0.0108	0.0089	0.0068	0.0058
15	0.0734	0.0432	0.0165	0.0085	0.0052	0.0036	0.0020	0.0014	0.0010	0.0006	0.0005
20	0.0387	0.0193	0.0054	0.0022	0.0011	0.0007	0.0003	0.0002	0.0001	0.0001	0.0000
25	0.0209	0.0088	0.0018	0.0006	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
30	0.0114	0.0041	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
35	0.0063	0.0020	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
40	0.0035	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
45	0.0020	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
50	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2: Efficiency of the Major Order theorem 2.0 on sets of uniformly distributed random permutations, from n = 8 to n = 100, m = 3 to m = 50, statistics generated over 100 000 instances for smaller n to 2000 instances for bigger n (see Table 4)

$m \backslash n$	8	10	15	20	25	30	40	50	60	80	100
3	0.6292	0.5784	0.4981	0.4513	0.4194	0.3962	0.3651	0.3462	0.3324	0.3139	0.3020
4	0.5652	0.5333	0.4711	0.4265	0.3923	0.3661	0.3268	0.3003	0.2801	0.2504	0.2320
5	0.5966	0.5469	0.4648	0.4131	0.3761	0.3479	0.3073	0.2781	0.2555	0.2232	0.2033
10	0.5730	0.5253	0.4435	0.3886	0.3493	0.3193	0.2749	0.2439	0.2197	0.1878	0.1656
15	0.5929	0.5377	0.4453	0.3863	0.3448	0.3129	0.2679	0.2361	0.2122	0.1791	0.1561
20	0.5820	0.5310	0.4415	0.3832	0.3415	0.3097	0.2647	0.2323	0.2089	0.1750	0.1523
25	0.5979	0.5396	0.4444	0.3842	0.3415	0.3099	0.2625	0.2317	0.2074	0.1736	0.1500
30	0.5878	0.5349	0.4426	0.3828	0.3403	0.3082	0.2616	0.2293	0.2055	0.1715	0.1498
35	0.5996	0.5419	0.4455	0.3843	0.3407	0.3083	0.2612	0.2285	0.2050	0.1711	0.1481
40	0.5922	0.5374	0.4439	0.3828	0.3402	0.3070	0.2603	0.2288	0.2038	0.1707	0.1483
45	0.6016	0.5435	0.4461	0.3848	0.3405	0.3084	0.2615	0.2278	0.2038	0.1708	0.1472
50	0.5956	0.5395	0.4440	0.3833	0.3402	0.3072	0.2600	0.2284	0.2037	0.1700	0.1467

Table 3: Efficiency of the Major Order theorem 3.0 on sets of uniformly distributed random permutations, from n = 8 to n = 100, m = 3 to m = 50, statistics generated over 100 000 instances for smaller n to 2000 instances for bigger n (see Table 4)

$m \backslash n$	8	10	15	20	25	30	40	50	60	80	100
3	0.7642	0.7179	0.6345	0.5792	0.5378	0.5059	0.4608	0.4312	0.4086	0.3775	0.3563
4	0.6010	0.5757	0.5201	0.4761	0.4406	0.4127	0.3700	0.3404	0.3176	0.2833	0.2612
5	0.7381	0.6853	0.5813	0.5046	0.4470	0.4038	0.3456	0.3073	0.2793	0.2413	0.2191
10	0.6440	0.6019	0.5173	0.4515	0.4012	0.3617	0.3040	0.2649	0.2354	0.1979	0.1726
15	0.7392	0.6732	0.5445	0.4581	0.3987	0.3544	0.2946	0.2543	0.2254	0.1869	0.1612
20	0.6739	0.6262	0.5248	0.4484	0.3917	0.3486	0.2896	0.2491	0.2209	0.1819	0.1566
25	0.7463	0.6774	0.5440	0.4557	0.3939	0.3497	0.2867	0.2480	0.2186	0.1800	0.1540
30	0.6899	0.6391	0.5307	0.4496	0.3904	0.3464	0.2852	0.2448	0.2164	0.1775	0.1536
35	0.7490	0.6804	0.5467	0.4563	0.3928	0.3474	0.2846	0.2438	0.2156	0.1770	0.1517
40	0.7008	0.6473	0.5349	0.4505	0.3905	0.3448	0.2833	0.2438	0.2142	0.1764	0.1518
45	0.7522	0.6836	0.5484	0.4570	0.3924	0.3473	0.2847	0.2427	0.2141	0.1764	0.1507
50	0.7095	0.6533	0.5367	0.4518	0.3908	0.3450	0.2827	0.2431	0.2138	0.1755	0.1501

Table 4: Number of instances generated and calculated for each couple $m \setminus n$ (depending only on n)

n	8	10	15	20	25	30	40	50	60	80	100
# instances	100000	100000	100000	100000	100000	50000	20000	10000	8000	4000	2000