Analysis of aeroplane boarding via spacetime geometry and random matrix theory

2006 J. Phys. A: Math. Gen. 39 L453

(http://iopscience.iop.org/0305-4470/39/29/L01)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) L453–L459 doi:10.1088/0305-4470/39/29/L01

LETTER TO THE EDITOR

Analysis of aeroplane boarding via spacetime geometry and random matrix theory

E Bachmat[1], D Berend[1,2], L Sapir[3], S Skiena[4] and N Stolyarov[1]

[1] Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel
[2] Department of Mathematics, Ben-Gurion University, Beer-Sheva 84105, Israel
[3] Department of Management and Industrial Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel
[4] Department of Computer science, SUNY at Stony Brook, Stony Brook, NY 11794, USA

E-mail: ebachmat@cs.bgu.ac.il

Received 2 February 2006
Published 5 July 2006
Online at stacks.iop.org/JPhysA/39/L453

Abstract

We show that aeroplane boarding can be asymptotically modelled by two-dimensional Lorentzian geometry. Boarding time is given by the maximal proper time among curves in the model. Discrepancies between the model and simulation results are closely related to random matrix theory. The models can be used to explain why some commonly practised airline boarding policies are ineffective and even detrimental.

PACS numbers: 02.10.Yn, 02.40.−k, 02.50.−r, 05.40.−a, 89.40.Dd

Aeroplane boarding is a process which is experienced daily by millions of passengers worldwide. Airlines have developed various strategies in the hope of shortening boarding time, typically leading to announcements of the form 'passengers from rows 40 and above are now welcome to board the aeroplane', often heard around airport terminals. We will show how the aeroplane boarding process can be asymptotically modelled by spacetime geometry. The discrepancies between the asymptotic analysis and finite population results will be shown to be related in some cases to random matrix theory (RMT). Previously, aeroplane boarding has only been studied empirically via discrete event simulations [1–3].

We model the aeroplane boarding process as follows. Passengers $1, \ldots, N$ are represented by coordinates $X_i = (q_i, r_i)$, where q_i is the index of the passenger along the boarding queue (1st, 2nd, 3rd and so on), and r is his/her assigned row number. We rescale (q, r) to $[0, 1] \times [0, 1]$. We assume that the main cause of delay in aeroplane boarding is the time it takes passengers to organize their luggage and seat themselves, once they have arrived at their assigned row. The input parameters for our model are

u: the average amount of aisle length occupied by a passenger.
w: the distance between successive rows.
b: the number of passengers per row.

D: the amount of time (delay) it takes passengers to clear the aisle, once they have arrived at their designated row.

$p(q, r)$: the joint distribution of a passenger's row and queue joining time. $p(q, r)$ is directly affected by the airline policy and the way passengers react to the policy.

For the purposes of presentation we shall assume that u, w, b, D are all fixed. The aeroplane boarding process produces a natural partial-order relation of blocking among passengers. We say that passenger X blocks passenger Y if it is impossible for passenger Y to reach his/her assigned row before passenger X (and others blocked by X) has sat down and cleared the aisle. Aeroplane boarding functions as a peeling process for the partial order defined by the blocking relation. At first, passengers who are not blocked by any other passengers sit down. These passengers are the minimal elements in the blocking relation. In the second round, passengers who are not blocked by passengers other than those of the first round are seated and so forth. Boarding time thus coincides with the size of the longest chain in the partial order.

We assign to the boarding process with parameters $u, b, w, D, p(q, r)$ a Lorentz metric defined on the (q, r) unit square

$$\mathrm{d}s^2 = 4D^2 p(q, r)(\mathrm{d}q\,\mathrm{d}r + k\alpha(q, r)\,\mathrm{d}q^2), \tag{1}$$

where $k = bu/w$ and $\alpha(q, r) = \int_r^1 p(q, z)\,\mathrm{d}z$. There are two properties of the metric which relate it to the boarding process.

(M1) The volume form of the metric is proportional to the passenger density distribution $p(q, r)$.

(M2) The blocking partial order among passengers during the boarding process asymptotically coincides with the past–future causal relation induced by the metric on the passengers viewed as events in spacetime via their q, r coordinate representation.

To establish the second property, consider passengers represented by $X = (q, r)$ and $X' = (q + \mathrm{d}q, r + \mathrm{d}r), \mathrm{d}q > 0$. Consider the time when passenger X arrives at his/her designated row. All passengers with row numbers beyond r, which are behind passenger X in the queue but in front of passenger X', will occupy aisle space behind passenger X. The number of such passengers is roughly $N\alpha\,\mathrm{d}q$. Each such passenger occupies u/w units of aisle length where we take the basic aisle length unit to be the distance between successive rows. The row difference between X and X' is $-(N/b)\,\mathrm{d}r$. We conclude that passenger X is blocking passenger X', via the passengers who are between them, roughly when $\mathrm{d}q \geqslant -\alpha k\,\mathrm{d}r$, a condition which coincides (together with $\mathrm{d}q > 0$) with the causal relation induced by the metric.

By the two main properties, we may approximate asymptotically the aeroplane boarding process by the peeling process applied to the past–future causal relation on points in the associated spacetime, sampled with respect to the volume form. By a well-known result, two-dimensional Lorentzian metrics are conformally flat; hence, after an appropriate coordinate transformation we may assume that the spacetime is given by a metric of the form

$$\mathrm{d}s^2 = r(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{2}$$

on some domain (not necessarily the unit square). In the new coordinates x, y, which are lightlike coordinates, chains in the causal relation coincide with increasing (upright) subsequences, namely, sequences of points (x_i, y_i) such that $x_i \leqslant x_j$ and $y_i \leqslant y_j$ for $i < j$. The peeling process applied to the causal relation coincides in this case with patience sorting which is a well-known card game process which optimally computes the longest increasing subsequence in a permutation [4, 5].

We define $T(X)$ as the maximal proper time (integral over $\mathrm{d}s$) of a timelike trajectory ending at $X = (q, r)$. We also define $L(\tau)$ as the length (integral over $\sqrt{-\mathrm{d}s^2}$) of the spacelike curve which is defined by the equation $T(X) = \tau$.

Using the analysis of the size of maximal increasing subsequences given in [6, 7], the two basic properties lead to the following modelling statement.

The boarding time of passenger i is approximately $\sqrt{N}T(X_i)$. In particular the total boarding time is approximately $\sqrt{N}\,\mathrm{Max}_X T(X)$, where $\mathrm{Max}_X T(X)$ is the maximal proper time of a curve in the unit square with respect to the Lorentzian metric.

Here the word approximately means that the ratio tends to 1 with probability approaching 1 as the number of passengers N tends to infinity.

We apply the statement to the analysis of boarding times. Consider first the case where the airline does not have a boarding policy, namely, passengers queue at uniformly random times; therefore, $p(q, r) = 1, \alpha = 1 - r$ and the corresponding metric is

$$\mathrm{d}s^2 = 4\,\mathrm{d}q(\mathrm{d}r + k(1 - r)\,\mathrm{d}q). \tag{3}$$

We use this model to study the effect of aeroplane design parameters such as the distance between rows and the number of passengers per row on boarding time. These parameters affect boarding time through the parameter $k = \frac{ub}{w}$. We let $d = \mathrm{Max}_X T(X)$ be the diameter of the model. Letting $A = (0, 0)$ and $B = (1, 1)$ we note that any point C in the unit square satisfies $A \prec C \prec B$, where $\prec$ denotes the causal relation induced by (3). We conclude that the maximal curve must begin at A and end at B. To find the maximal proper time curve we solve the Euler–Lagrange equation for proper time subject to the constraints of lying in the unit square and passing through A and B. Solutions to the unconstrained Euler–Lagrange (geodesic) equation are of the form

$$r = c_1\,\mathrm{e}^{2kq} + c_2\,\mathrm{e}^{kq} + 1. \tag{4}$$

For $k \leqslant \ln(2)$ the geodesic solution passing through A and B is contained in the interior of the unit square and is therefore the maximal curve. The length of the curve is

$$d(k) = 2\sqrt{\frac{\mathrm{e}^k - 1}{k}}. \tag{5}$$

For $k > \ln(2)$ the maximal curve contains a portion which runs along the q-axis until reaching the point $(1 - \ln(2)/k, 0)$. From this point the curve follows a geodesic to the point $(1, 1)$. The geodesic is tangent to the q-axis at $(1 - \ln(2)/k, 0)$ and this condition is in fact used to determine the point of departure from the boundary. We have

$$d(k) = 2\sqrt{k} + 2(1 - \ln(2))/\sqrt{k}. \tag{6}$$

We compared the above computations with simulations of our model of the boarding process. When $N = 10^6$, the length of the maximal chain matches well with the estimated values derived from the spacetime model. For $k = 0.5$, the length of the maximal chain in the simulation was 2261, while the estimated value is 2278; while for $k = 5$, the corresponding values were 4589 and 4740, respectively.

We now consider more realistic values of N, namely, $N = 100$ and $N = 200$. Table 1 presents the average boarding time results for 1000 computer simulations of the boarding process for several settings of k, compared with the boarding time estimate computed via the spacetime model. Upon inspection, table 1 shows that for realistic values of N, there are differences in the range of 20–60% between the asymptotic boarding time estimates computed via Lorentzian geometry and the boarding time computed via simulations of the boarding process. We note the important fact that in all cases the Lorentzian estimate is larger.

To analyse the differences between our Lorentzian estimates and the simulation results we consider L_N, the random variable representing the boarding time according to the boarding process model (the simulation results). We define the discrepancy random variable $M_N = L_N - \sqrt{N}d(k)$ which measures the difference between the boarding time and the

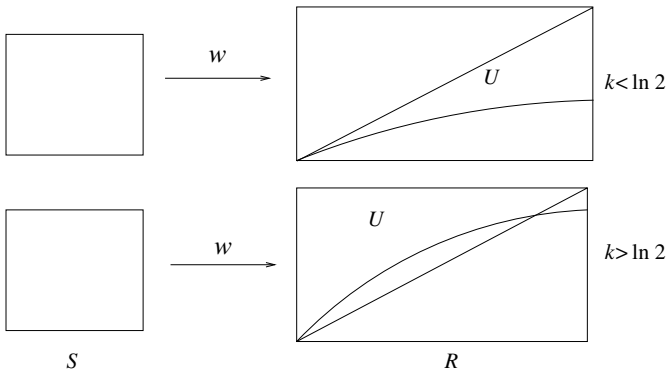

Figure 1. The image of the unit square under W.

Table 1. A comparison of spacetime model estimates with the average over 1000 simulations of boarding time results.

k	N	Simulation result	Spacetime estimate
0.5	100	18.1	22.8
0.5	200	26.8	32.2
2.0	100	23.0	32.5
2.0	200	34.7	45.8
5.0	100	29.0	47.5
5.0	200	44.9	66.9

Lorentzian estimate. The curvature of the metric given in (3) vanishes; therefore, we can apply a coordinate transformation W which changes the metric to the form

$$\mathrm{d}s^2 = 4\,\mathrm{d}x\,\mathrm{d}y. \tag{7}$$

Let S be the unit square and consider $U = W(S)$ its image under W. We refer to figure 1 for an illustration of the following arguments. We recall that any point $C \in S$ satisfies $A \prec C \prec B$ where $A = (0, 0)$ and $B = (1, 1)$. We conclude that S will be mapped into the set of points D, satisfying $W(A) \prec D \prec W(B)$ where now $\prec$ denotes the causal relation w.r.t. (7). The set of such points is a rectangle R with sides parallel to the axis and corners at $W(A)$ and $W(B)$. W maps lightlike curves w.r.t. (3) into lightlike curves w.r.t. (7), which are simply vertical and horizontal lines. It also maps the geodesics of (3) which are given by (4) into straight lines. The left, right and top sides of S are lightlike and will therefore be mapped to the left, right and top sides of R. On the other hand, the bottom of S is not lightlike nor a geodesic and will therefore be mapped to a curve within R which will be concave since by (4) the geodesics connecting points at the bottom of S lie below S. When $k \leqslant \ln(2)$ the geodesic connecting A and B lies entirely within S. The image of this geodesic is the diagonal of R and therefore $U = W(S)$ will contain the above diagonal triangle T within R. When $k > \ln(2)$, the geodesic is not contained in S and consequently the diagonal of R is not contained in U.

We can also compute the volumes of U and R as follows. Since W preserves volumes, the volume of U w.r.t. (7) equals the volume of S w.r.t. (3). The later volume can be computed directly from the volume form formula $\sqrt{-\det(g)}\,\mathrm{d}q\,\mathrm{d}r$ and we get $\mathrm{vol}(U) = 2$. The volume form formula also tells us that the volume w.r.t. (7) is twice the usual Euclidean area. Letting δx and δy denote the coordinate differences between $W(A)$ and $W(B)$, we have $\mathrm{vol}(R) = 2\delta x\delta y$. The length of the diagonal of R is $\ell(\mathrm{diag}) = 2\sqrt{\delta x\delta y}$; therefore, $\mathrm{vol}(R) = \ell(\mathrm{diag})^2/2$. On

the other hand, the diagonal is simply the image under W of the geodesic connecting A and B, whose length from (5) is $2\sqrt{\frac{e^k-1}{k}}$. We conclude that $\mathrm{vol}(R) = 2\frac{e^k-1}{k}$.

As noted previously, for the metric (7), the past–future causal chains correspond to increasing sequences of points. We also note that since the spacetime points are sampled according to the volume form, they are uniformly distributed. The problem is therefore reduced to that of computing discrepancies of increasing subsequences for N uniformly distributed points in $U = W(S)$. Discrepancies have been studied in the context of increasing subsequences of uniformly distributed points in the rectangle R [8], and in the above diagonal triangle T within R [9]. Following [8] it is convenient to consider the Poissonized version where the number of chosen points is given by a Poisson distribution with parameter N. In both cases the discrepancy M_N has order of magnitude $N^{1/6}$. We therefore define the scaled discrepancy $\Delta_N = M_N/N^{1/6}$. In the case of R the scaled discrepancy is given asymptotically by the Tracy–Widom distribution F_2 [10], which measures the (scaled) discrepancy of the largest eigenvalue of an N by N matrix in the Gaussian unitary ensemble (GUE) [11], in comparison with $2\sqrt{N}$. For $N/2$ uniformly distributed points in T, the normalized discrepancy is given asymptotically by the Tracy–Widom distribution F_4 [10], which is the normalized discrepancy of the largest eigenvalue in the Gaussian symplectic ensemble (GSE). The averages for these distributions are $E(F_2) \approx -1.77$ and $E(F_4) \approx -2.3$. After Poissonization we can think of Poisson(N) uniformly distributed points in U as the restriction to U of $\mathrm{Poisson}\left(N\frac{\mathrm{vol}(R)}{\mathrm{vol}(U)}\right) = \mathrm{Poisson}\left(N\frac{e^k-1}{k}\right)$ uniformly distributed points in R.

The results of [8, 9] remain the same after de-Poissonization and can therefore be applied to the original version of our problem. When $k \leqslant \ln(2)$ we have $T \subset U \subset R$, so applying the refined estimates to $N\frac{e^k-1}{k}$ points in R we obtain

$$F_4(z) \leqslant \Pr\left(\Delta_N \leqslant \left(\frac{e^k-1}{k}\right)^{1/6} z\right) \leqslant F_2(z). \tag{8}$$

In particular, we obtain the finer asymptotic estimate

$$-2.3\left(\frac{e^k-1}{k}\right)^{1/6} \leqslant E(\Delta_N) \leqslant -1.77\left(\frac{e^k-1}{k}\right)^{1/6}. \tag{9}$$

Applying the refined estimates to the results from table 1 with $k = 0.5$ we see that the refined estimate (9) holds already for the realistic values $N = 100, 200$. Indeed, $17.7 < 18.1 < 18.8$ for $N = 100$ and $26.4 < 26.8 < 27.8$ for $N = 200$.

The discrepancy M_N experiences a phase transition in behaviour at the value $k = \ln(2)$. For example, when $k = 3$ linear regression produced a tight fit with the formula:

$$E(M_N) \sim -4.85N^{0.222}. \tag{10}$$

We do not know how to compute analytically the correct order of magnitude of M_N when $k > \ln(2)$. However, we can provide the following heuristic explanation for the change in discrepancy behaviour. When $k > \ln(2)$ the maximal curve contains a portion of the bottom edge of the unit square. The argument essentially states that the form of the metric (3) near the bottom edge forces the portion of the maximal chain along the bottom edge to have height (transversal) fluctuations which are too small to produce a small discrepancy.

Consider the length of the longest chain between the point $A = (0, 0)$ and a second point $Z = (0, z)$ at the bottom edge of the unit square. The Lorentzian estimate for the length of the sequence is $2\sqrt{k}q\sqrt{N}$. We may consider the narrow band $B_a = \{(q, r)|0 \leqslant r \leqslant N^{-a}\}$ at the bottom of the unit square. Assume that a given chain between A and Z has a non-negligible portion which is outside B_a, namely, there are numbers $0 < c < d < z$ such that all points

(q, r) in the chain with $c < q < d$ satisfy $r > N^{-a}$. On the complement of B_a consider the constant metric

$$ds^2 = 4\,dq(dr + k(1 - N^{-a})\,dq). \tag{11}$$

Any causal chain (in the complement of B_a) w.r.t. the metric (3) is also a chain w.r.t. (11). The longest chain w.r.t. (11) between (c, N^{-a}) and (d, N^{-a}) has, with high probability (w.h.p.), size at most $2\sqrt{k}(c-d)\sqrt{N} - O(N^{1/2-a})$. The chain will therefore have w.h.p. a negative discrepancy whose order of magnitude is at least $N^{1/2-a}$. We see that whenever $a < 1/3$, a maximal chain which is not essentially confined to B_a will have w.h.p. negative discrepancy whose order of magnitude is at least $N^{1/2-a} \gg N^{1/6}$. On the other hand, it follows from the computations in [12] that whenever $a > 1/6$, a chain which is confined to B_a will produce w.h.p. a negative discrepancy whose order of magnitude is larger than $N^{1/6}$. Putting these arguments together for $a = 1/4$ provides a strong heuristic evidence that the discrepancy is negative with order of magnitude larger than $N^{1/6}$.

The spacetime metrics given by formula (1) can be used for comparing different boarding policies. Boarding policies affect the passenger distribution function $p(q, r)$. For example, assume that the policy is to board the back third of the aeroplane, followed by the front third and finally the middle third. The corresponding distribution $p(q, r)$ is given by $p(q, r) = 3$ on the sub squares $[0, 1/3] \times [2/3, 1]$, $[1/3, 2/3] \times [0, 1/3]$, $[2/3, 1] \times [1/3, 2/3]$, and zero elsewhere. When $p(q, r) = 0$ the metric given in (1) degenerates. This problem can be overcome by setting $p(q, r) = \varepsilon$ in such regions and letting ε tend to zero. An equivalent and simpler approach is to allow the expression $dq\,dr + k\alpha(q, r)\,dq^2$ from (1) to determine a light cone at all points. The aeroplane boarding time is then given asymptotically by the maximal proper time among all timelike curves even when $p(q, r)$ vanishes in certain regions.

A comparison of the results of the spacetime computations, with the parameter $k = 4$, and the results of detailed event-driven simulations of boarding processes [2, 3], shows that the spacetime estimates are in almost complete agreement with the event-driven simulation results regarding the ranking of the different policies. This is somewhat surprising given that the trace-driven simulations in [2] take into account many details of actual boarding processes which are not considered by our model of the boarding process. The large discrepancies which we have noted previously between our own model of the boarding process and the spacetime estimates are less of a factor since when comparing boarding strategies only ratios of boarding times matter and these are less affected by the discrepancies since in all cases the spacetime models lead to overestimates. Following the analysis we conclude that

(1) the commonly practiced back to front boarding policies which board passengers from the back of the aeroplane first are ineffective for realistic values of $3 < k < 5$ and
(2) among row-dependent policies which do not severely constrain passengers, random boarding (no policy) is almost optimal.

Finally, our methods can be applied to yield new insights into several other discrete processes including patience sorting [5], polynuclear growth models (PNG) [13], kinematics of causal sets [14], maximal layers in planar domains [15] and scheduling of I/O requests to disk drives [16, 17].

Acknowledgments

We are grateful to Perci Deift, Ofer Zeitouni and Jinho Baik for very useful discussions. This research has been partially supported by a grant from the dean of natural sciences at Ben-Gurion U. The first author has been partially supported by an IBM faculty award.

References

[1] Marelli S, Mattocks G and Merry R 2000 *Boeing Aero Magazine* **1**
[2] Van Landegham H and Beuselinck A 2002 *Eur. J. Opt. Res.* **142** 294
[3] van den Briel M, Villalobos J, Hogg G, Lindemann T and Mule A 2005 *Interfaces* **35** 191
[4] Mallows C L 1973 *Bull. Inst. Math. Appl.* **9** 216
[5] Aldous D and Diaconis P 1999 *Bull. A. Math. Soc.* **36** 413
[6] Deuschel J D and Zeitouni O 1995 *Ann. Probab.* **23** 852
[7] Vershik A and Kerov S 1977 *Sov. Math. Dokl.* **18** 527
[8] Baik J, Deift P and Johansson K 1999 *J. Am. Math. Soc.* **12** 1119
[9] Baik J and Rains E 2001 *Duke J. Math.* **109** 205
[10] Tracy C A and Widom H 1994 *Commun. Math. Phys.* **159** 151
Tracy C A and Widom H 1996 *Commun. Math. Phys.* **177** 727
[11] Mehta M L 2004 *Random Matrices* (New York: Academic)
[12] Johansson K 2000 *Commun. Math. Phys.* **209** 437
[13] Prahofer M and Spohn H 2000 *Phys. Rev. Lett.* **84** 4882
[14] Bombelli L, Lee J, Meyer D and Sorkin R D 1987 *Phys Rev. Lett.* **59** 521
[15] Devroye L 1993 *Comput. Math. Appl.* **25** 19
[16] Andrews M, Bender M and Zhang L 2002 *Algorithmica* **32** 277
[17] Bachmat E 2002 *Proc. Symp. Th. of Comput. (STOC)* p 277