On Average Manhattan Distance

Inspired by How to speedrun fixing typos by carykh.

Let us consider a Manhattan-like metric on a rectangle of width w and height h. What is the average distance between two points? What is its minimum at fixed area?

Warmup

Let us first start by considering simplest boundary conditions. Distance metric is just

$$d(r_1, r_2) = |x_2 - x_1| + |y_2 - y_1| \tag{1}$$

Average distance is

$$\mathbb{E}[d] = \frac{1}{A^2} \sum_{r_1, r_2} d(r_1, r_2) = \frac{1}{w^2 h^2} \sum_{x_1, x_2 = 1}^{w} \sum_{y_1, y_2 = 1}^{h} (|x_2 - x_1| + |y_2 - y_1|) = f(w) + f(h) \tag{2}$$

where $f(z) = \frac{1}{z^2} \sum_{i,j=1}^{z} |i - j| = \frac{z^2 - 1}{3z} \simeq \frac{z}{3}$. Notice the last $\simeq$ step. I will use the following approximation from here: $w, h \gg 1$. Most answers will not change qualitatively with this approximation and the large rectangle is a natural limit of the problem. Whenever possible the full answer is provided. For clearness I am repeating the answer here

$$\mathbb{E}[d] \simeq \frac{w + h}{3} \tag{3}$$

This is minimized at $w/h = 1$, or $w = h = \sqrt{A}$.

$$\min \mathbb{E}[d] \simeq \frac{2}{3} \sqrt{A} \tag{4}$$

The full answer is $\frac{2}{3} \left(\sqrt{A} + \frac{1}{\sqrt{A}} \right)$.

Let us now consider periodic boundary conditions. As in, we can travel through each boundary to the oppposite side with zero cost. The new distance function is

$$d(r_1, r_2) = \min(|x_2 - x_1|, w - |x_2 - x_1|) + \min(|y_2 - y_1|, h - |y_2 - y_1|) \tag{5}$$

We can again compute

$$\mathbb{E}[d] = f(w) + f(h) \simeq \frac{w + h}{4} \tag{6}$$

where full $f(z) = \frac{1}{z} \lfloor \frac{z}{2} \rfloor \lceil \frac{z}{2} \rceil$. This is minimized by the same point $w/h = 1$:

$$\min \mathbb{E}[d] \simeq \frac{1}{2} \sqrt{A} \tag{7}$$

Here we will notice a few generalities about function f. This paragraph can be skipped. Notice that $w = h$ is extremum for any $f(w) + f(h)$. More than that, it is a local minimum iff $f'(\sqrt{A}) + \sqrt{A} f''(\sqrt{A}) > 0$. This is a natural result from symmetry wrt exchange $w \leftrightarrow h$ and separability of the expression. The next part is not so generous.

Text editor metric

Let us now consider the "text editor" distance. We can travel through vertical boundary, changing y by ± 1 depending on direction. We can also travel from anywhere in the top boundary to the top left corner and from anywhere in the bottom boundary to the bottom right corner. The described step is at a cost of 1.

We will call regular metric the one that allows to just travel through vertical boundary. We can say that the best distance is the minimum between regular distance, the one that takes us to bottom right corner and then to r_2 and the one that takes us to the top left corner and then to r_2.

$$d(r_1, r_2) = \min \left(d_0(r_1, r_2), d_{\text{tl}}(r_1) + d_0((1,1), r_2), d_{\text{br}}(r_1) + d_0((w,h), r_2) \right) \tag{8}$$

where d_0 is the mentioned regular distance, $d_{\text{tl}}, d_{\text{br}}$ are distances to top left and bottom right corners respectively.

Notice that you will never want to travel through the vertical boundary before going to a corner. We now need to write down the expressions $d_0, d_{\text{tl}}, d_{\text{br}}$. The ± 1 in y directions after boundary teleportation makes calculations very annoying, but it only contributes $O(1)$ to the total distance, which doesn't scale with area A. We will ignore it to get qualitative result. This is the same approximation that was used in warmup.

$$d_0(r_1, r_2) = |y_1 - y_2| + \min(|x_1 - x_2|, w - |x_1 - x_2|), \quad d_{\text{tl}}(r_1) = y_1, \quad d_{\text{br}}(r_1) = h - y_1 + 1 \tag{9}$$

Upto constant order we can simplify the expression

$$d(r_1, r_2) = \min\left(|y_1 - y_2| + \min(|x_1 - x_2|, w - |x_1 - x_2|), \min(y_1 + y_2, (h - y_1) + (h - y_2)) + \min(x_2, w - x_2)\right) \tag{10}$$

At large w, h we can replace sums with integrals. This is a natural usage of our limit $w, h \gg 1$. Dealing with integrals makes calculations in the warmup easy, and of this calculation possible.

$$\begin{aligned}&\mathbb{E}[d] \simeq \\ \simeq &\int_{[0,1]^4} dx_1 dx_2 dy_1 dy_2 \min\left(w \min(|x_2 - x_1|, 1 - |x_2 - x_1|) + h|y_2 - y_1|, w \min(x_2, 1 - x_2) + h \min(y_1 + y_2, 2 - y_1 - y_2)\right)\end{aligned} \tag{11}$$

After a some coordinate transformation we arrive at more readable

$$\mathbb{E}[d] \simeq \frac{1}{2} \int_{[0,1]^4} du\, dv\, dx\, dy\, \min(wu + 2hv, wx + 2hy) \tag{12}$$

Before reaching the exact result we can see from the previous expression that the point $w = 2h$ is special. Miraculously the integral can be evaluated exactly

$$\mathbb{E}[d] \simeq \begin{cases} \frac{w}{4} + \frac{h}{3} - \frac{w^2}{48h} + \frac{w^3}{480h^2} & \text{if } w \leq 2h \\ \frac{w}{6} + \frac{h}{4} - \frac{h^2}{6w} + \frac{h^3}{15w^2} & \text{if } w > 2h \end{cases} \tag{13}$$

Minimum is at $w/h = 2$ with the value

$$\min \mathbb{E}[d] \simeq \frac{23\sqrt{2}}{60} \sqrt{A} \tag{14}$$

As can be seen in the image the minimum is quite soft – even large variations in the ratio will not lead to significant change in the average distance. This shows that constant corrections might offer a different picture.

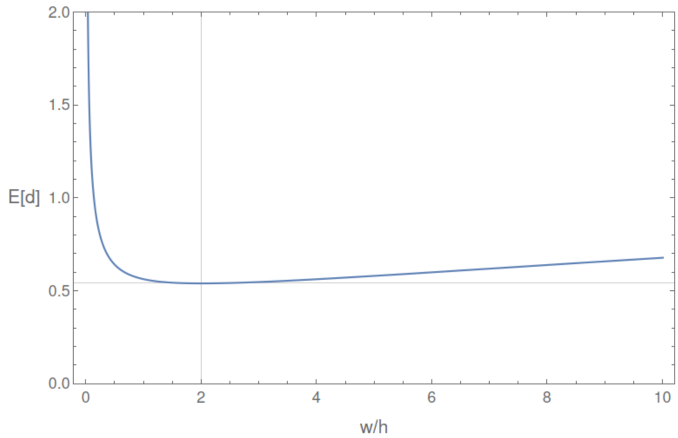

Figure 1: Average distance between points for "text editor" metric. Area is normalized to 1.

Notes

The calculations here can be improved by considering $O(1)$ corrections. We can also consider other shapes.

Some calculations here have been checked by numerics, but this does not stop errors from creeping up in parts like the definitions of distance functions.

General shapes

Here are a few answers for other shape

$$\mathbb{E}[d]_{\text{circle}} = \frac{512}{45\pi^2} r = \frac{512}{45\pi^{5/2}}\sqrt{A} \simeq 0.65\sqrt{A} \tag{15}$$

$$\mathbb{E}[d]_{\text{diamond}} = \frac{7\sqrt{2}}{15}\sqrt{A} \tag{16}$$

A general answer for simple Manhattan metric (without any boundary conditions) has been found in What is the optimal shape of a city? (Bender et al. 2004) by applying variational principle to the boundary. Equation for the boundary is differential and the formulas are quite cumbersome so we will not quote them here.