Proofs of the Metric Axioms

The Metric Axioms

In section 3.2 of the paper *HyGLEAM: Hybrid General Purpose Evolutionary Algorithm and Method* submitted to SCI 2001 conference three distance measures for chromosomes AK_1 and AK_2 have been introduced, for which it must be proved that they fulfill the 4 metric axioms, which are as follows:

$$\Delta(AK_1, AK_2) = 0 \Leftrightarrow AK_1 = AK_2 \tag{A.1}$$

$$\Delta(AK_1, AK_2) \geq 0 \tag{A.2}$$

$$\Delta(AK_1, AK_2) = \Delta(AK_2, AK_1) \tag{A.3}$$

$$\Delta(AK_1, AK_3) \leq \Delta(AK_1, AK_2) + \Delta(AK_2, AK_3) \tag{A.4}$$

1. Parameter Distance

The parameter distance of two chromosomes AK_1 and AK_2, which are both not empty, $\Delta_{par}(AK_1, AK_2)$ is defined in Eq. (3.1) for all parameters, for which the upper limit og is greater than the lower one ($og_i > ug_i$), as:

<u>parameter distance:</u>

$$\Delta_{par}(AK_1, AK_2) = \frac{1}{anz}\sum_{i=1}^{anz} \frac{|param_{i,1} - param_{i,2}|}{og_i - ug_i}$$

where $param_{i,j}$: value of the i-th parameter in chromosome AK_j. The parameters are numbered in the sequence of their gene type definition.

ug_i, og_i: lower and upper bound of the range of values of the i-th parameter.

anz: number of all parameters of all genes.

Axiom A1:

For Δ_{par} two chromosomes are called equal if all their corresponding parameters are equal. As $anz > 0$ and $og_i > ug_i$, it follows from the fact, that a sum of positive addends is zero exactly when all addends are zero, for all $i = 1,...,anz$:

$$\Delta_{par}(AK_1, AK_2) = 0 \Rightarrow |param_{i,1} - param_{i,2}| = 0 \Rightarrow param_{i,1} = param_{i,2} \Rightarrow AK_1 = AK_2 \tag{A1.1}$$

$$AK_1 = AK_2 \Rightarrow param_{i,1} = param_{i,2} \Rightarrow |param_{i,1} - param_{i,2}| = 0 \Rightarrow \Delta_{par}(AK_1, AK_2) = 0 \tag{A1.2}$$

With (A1.1) and (A1.2) (A.1) is fulfilled.

Axiom A2:

From $anz > 0$, $og_i - ug_i > 0$ and $|param_{i,1} - param_{i,2}| \geq 0$ follows, that $\Delta_{par}(AK_1, AK_2) \geq 0$. Thus (A.2) is fulfilled.

Axiom A3:
A swapping of the chromosomes has an effect to the absolute value of $|param_{i,1} - param_{i,2}|$ only:

$$|param_{i,1} - param_{i,2}| = |param_{i,2} - param_{i,1}| \Rightarrow \Delta_{par}(AK_1, AK_2) = \Delta_{par}(AK_2, AK_1)$$

Thus (A.3) is fulfilled.

Axiom A4:
It must be shown that

$$\frac{1}{anz}\sum_{i=1}^{n} \frac{|param_{i,1} - param_{i,3}|}{og_i - ug_i} \le \frac{1}{anz}\sum_{i=1}^{n} \frac{|param_{i,1} - param_{i,2}|}{og_i - ug_i} + \frac{1}{anz}\sum_{i=1}^{n} \frac{|param_{i,2} - param_{i,3}|}{og_i - ug_i}$$

If this relation holds for every term of the sum, it will also hold for the sum itself. Thus, it must be shown that for every parameter $param_i$

$$\frac{|param_1 - param_3|}{og - ug} \le \frac{|param_1 - param_2|}{og - ug} + \frac{|param_2 - param_3|}{og - ug}$$

or

$$|param_1 - param_3| \le |param_1 - param_2| + |param_2 - param_3| \qquad \text{(A1.3)}$$

To show this, 6 cases must be considered:

Case1: $param_1 \le param_2 \le param_3$

$$param_3 - param_1 \le param_2 - param_1 + param_3 - param_2$$
$$\Leftrightarrow \quad param_3 - param_1 \le param_3 - param_1 \qquad \text{q.e.d.}$$

Case 2: $param_1 \ge param_2 \ge param_3$

$$param_1 - param_3 \le param_1 - param_2 + param_2 - param_3$$
$$\Leftrightarrow \quad param_1 - param_3 \le param_1 - param_3 \qquad \text{q.e.d.}$$

Case 3: $param_1 \le param_2, \quad param_3 \le param_2, \quad param_1 \le param_3$

$$param_3 - param_1 \le param_2 - param_1 + param_2 - param_3$$
$$\Leftrightarrow \quad param_3 \le 2 \cdot param_2 - param_3$$
$$\Leftrightarrow \quad param_3 \le param_2 \qquad \text{q.e.d.}$$

Case 4: $param_1 \le param_2, \quad param_3 \le param_2, \quad param_3 \le param_1$

$$param_1 - param_3 \le param_2 - param_1 + param_2 - param_3$$
$$\Leftrightarrow \quad param_1 \le 2 \cdot param_2 - param_1$$
$$\Leftrightarrow \quad param_1 \le param_2 \qquad \text{q.e.d.}$$

Case 5: $param_2 \le param_1, \quad param_2 \le param_3, \quad param_3 \le param_1$

$$param_1 - param_3 \le param_1 - param_2 + param_3 - param_2$$
$$\Leftrightarrow \quad -param_3 \le -2 \cdot param_2 + param_3$$
$$\Leftrightarrow \quad param_3 \ge param_2 \qquad \text{q.e.d.}$$

Case 6: $param_2 \le param_1, \quad param_2 \le param_3, \quad param_1 \le param_3$

$$param_3 - param_1 \le param_1 - param_2 + param_3 - param_2$$
$$\Leftrightarrow \quad -param_1 \le param_1 - 2 \cdot param_2$$
$$\Leftrightarrow \quad param_1 \ge param_2 \qquad \text{q.e.d.}$$

Thus Eq. (A1.3) is proved and with this (A4) is fulfilled.

2. Positional Distance

The positional distance $\Delta_{pos}(AK_1, AK_2)$ of two chromosomes AK_1 and AK_2, the length of which is greater than 1, is defined in Eq. (3.2) and (3.3) as follows:

<u>positional distance:</u>

$$\Delta_{pos}(AK_1, AK_2) = \frac{1}{abst_{max}} \sum_{i=1}^{len} PA_{1,2}(A_i)$$

$$PA_{1,2}(A_i) = |I_1(A_i) - I_2(A_i)|$$

where *len*: length of a chromosome ($len > 1$)

$abst_{max}$: distance maximum of all genes within one chromosome. According to eq. (3.8) it depends on the length *len* of a chromosome and for $len > 1$ it is greater than 0.

$I_j(A_i)$: index of the gene *i* in chromosome *j*.

Axiom A1:
For this measure two chromosomes AK_1 and AK_2 are called equal, if the positions of corresponding genes are equal regardless of their parameter values. As $abst_{max} > 0$ and the sum of positive addends is zero exactly when all addends are zero, it follows:

$$\Delta_{pos}(AK_1, AK_2) = 0 \;\Rightarrow\; |I_1(A_i) - I_2(A_i)| = 0 \;\; \forall i = 1, \ldots, len \;\Rightarrow\; AK_1 = AK_2 \tag{A1.4}$$

$$AK_1 = AK_2 \;\Rightarrow\; |I_1(A_i) - I_2(A_i)| = 0 \;\; \forall i = 1, \ldots, len \;\Rightarrow\; \Delta_{pos}(AK_1, AK_2) = 0 \tag{A1.5}$$

With (A1.4) and (A1.5) (A.1) is fulfilled.

Axiom A2:
As the addends $PA_{1,2}(A_i)$ are according to the definition all greater or equal 0 and $abst_{max} > 0$, it can be stated that $\Delta_{pos}(AK_1, AK_2) \geq 0$. With this (A.2) is fulfilled.

Axiom A3:
Swapping the chromosomes affects only the calculation of $|I_1(A_i) - I_2(A_i)|$:

$$|I_1(A_i) - I_2(A_i)| = |I_2(A_i) - I_1(A_i)| \;\Rightarrow\; \Delta_{pos}(AK_1, AK_2) = \Delta_{pos}(AK_2, AK_1)$$

Thus (A.3) is fulfilled.

Axiom A4:
It must be shown that

$$\frac{1}{abst_{max}} \sum_{i=1}^{len} |I_1(A_i) - I_3(A_i)| \leq \frac{1}{abst_{max}} \sum_{i=1}^{len} |I_1(A_i) - I_2(A_i)| + \frac{1}{abst_{max}} \sum_{i=1}^{len} |I_2(A_i) - I_3(A_i)|$$

If this relation holds for every term of the sum, it will also hold for the sum itself. Thus, it must be shown for every A_i:

$$|I_1(A_i) - I_3(A_i)| \leq |I_1(A_i) - I_2(A_i)| + |I_2(A_i) - I_3(A_i)|$$

As this relation was already proved for general parameters in sect.1,Eq. (A1.3), it is also true for variables>0, like the position indices. Thus (A.4) is fulfilled.

3. Difference of Gene Presence

$\Delta_{akt}(AK_1, AK_2)$ measures the presence of genes within two non empty chromosomes AK_1 and AK_2:

Difference of Gene Presence:

$$\Delta_{akt}(AK_1, AK_2) = 1 - \frac{card(A_{gem}(AK_1, AK_2))}{max(len(AK_1), len(AK_2))}$$

where $len(AK_i)$: length of chromosome AK_i
$A_{gem}(AK_i, AK_j)$: set of genes common to both chromosomes AK_i and AK_j.
A_{AKi}: set of genes of chromosome AK_i.
$card(A)$: number of elements of a set A.

Axiom A1:
For this measure two chromosomes AK_1 and AK_2 are called equal, if the two sets of their genes are equal.

$$\Delta_{akt}(AK_1, AK_2) = 0 \Rightarrow \frac{card(A_{gem})}{max(len(AK_1), len(AK_2))} = 1 \Rightarrow card(A_{gem}) = max(len(AK_1), len(AK_2)) \quad \text{(A1.6)}$$

It will be proved indirectly, that from the left equation of (A1.6) follows the equality of the chromosomes. It is assumed that both chromosomes are different:

$$\begin{aligned} assumption: &\quad card(A_{gem}) = max(len(AK_1), len(AK_2)) \Rightarrow AK_1 \neq AK_2 \\ i.e.: &\quad \exists a \in A_{AK1}: a \notin A_{AK2} \\ &\quad \Rightarrow len(AK_1) > card(A_{gem}) \\ &\quad \Rightarrow max((len(AK_1), len(AK_2)) \geq len(AK_1) > card(A_{gem})) \quad \text{contradiction!} \end{aligned}$$

The opposite case, where there is a gene in AK_2, which is not contained in AK_1, will be treated in almost the same manner. With this it is proved that:

$$\Delta_{akt}(AK_1, AK_2) = 0 \Rightarrow AK_1 = AK_2 \quad \text{(A1.7)}$$

Now it must be proved that from the identity of AK_1 and AK_2 follows, that $\Delta_{akt}(AK_1, AK_2) = 0$:

$$AK_1 = AK_2 \Rightarrow A_{AK1} = A_{AK2} = A_{gem} \Rightarrow max(len(AK_1), len(AK_2)) = card(A_{gem})$$

$$\Rightarrow \frac{card(A_{gem})}{max(len(AK_1), len(AK_2))} = 1 \Rightarrow \Delta_{akt}(AK_1, AK_2) = 0 \quad \text{(A1.8)}$$

With Eq. (A1.7) and (A1.8) (A.1) is fulfilled.

Axiom A2:

$$\Delta_{akt}(AK_1, AK_2) \geq 0 \text{ if } 1 \geq \frac{card(A_{gem})}{max(len(AK_1), len(AK_2))} \text{ or } max(len(AK_1), len(AK_2)) \geq card(A_{gem}).$$

For identical chromosomes both sides are equal and for non identical the relation holds. With this (A.2) is fulfilled.

Axiom A3:
As swapping of the chromosomes has no effect in calculating $\Delta_{akt}(AK_1, AK_2)$ according to its definition, (A.3) is fulfilled.

Axiom A4:

It must be shown that:

$$1-\frac{card(A_{AK1}\cap A_{AK3})}{max(card(A_{AK1}),card(A_{AK3}))}\le 1-\frac{card(A_{AK1}\cap A_{AK2})}{max(card(A_{AK1}),card(A_{AK2}))}+1-\frac{card(A_{AK2}\cap A_{AK3})}{max(card(A_{AK2}),card(A_{AK3}))} \qquad \text{(A1.9)}$$

For reasons of symmetry it can be assumed that $card(AK_1)\ge card(AK_3)$. Now the following three cases must be considered.

Case 1: $card(AK_2)\ge card(AK_1)\ge card(AK_3)$

It must be shown that:

$$1-\frac{card(A_{AK1}\cap A_{AK3})}{card(A_{AK1})}\le 2-\frac{card(A_{AK1}\cap A_{AK2})}{card(A_{AK2})}-\frac{card(A_{AK2}\cap A_{AK3})}{card(A_{AK2})}$$

$$-1-\frac{card(A_{AK1}\cap A_{AK3})}{card(A_{AK1})}\le -\frac{card(A_{AK1}\cap A_{AK2})+card(A_{AK2}\cap A_{AK3})}{card(A_{AK2})}$$

$$\frac{card(A_{AK1}\cap A_{AK2})+card(A_{AK2}\cap A_{AK3})}{card(A_{AK2})}\le 1+\frac{card(A_{AK1}\cap A_{AK3})}{card(A_{AK1})}$$

$$card(A_{AK1}\cap A_{AK2})+card(A_{AK2}\cap A_{AK3})\le card(A_{AK2})+card(A_{AK1}\cap A_{AK3})\cdot\frac{card(A_{AK2})}{card(A_{AK1})} \qquad \text{(A1.10)}$$

The left side of relation (A1.10) is equivalent to

$$card((A_{AK1}\cup A_{AK3})\cap A_{AK2})+card((A_{AK1}\cap A_{AK3})\cap A_{AK2})$$

Because of $(A_{AK1}\cup A_{AK3})\cap A_{AK2}\subseteq A_{AK2}$ and $(A_{AK1}\cap A_{AK3})\cap A_{AK2}\subseteq A_{AK1}\cap A_{AK3}$:

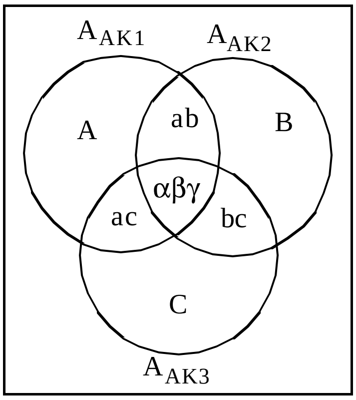
$$card(A_{AK1}\cap A_{AK2})+card(A_{AK2}\cap A_{AK3})\le card(A_{AK2})+card(A_{AK1}\cap A_{AK3})\le$$

$$card(A_{AK2})+card(A_{AK1}\cap A_{AK3})\cdot\frac{card(A_{AK2})}{card(A_{AK1})} \quad \text{because of} \quad \frac{card(A_{AK2})}{card(A_{AK1})}\ge 1$$

With this relation (A1.10) is proved.

Case 2: $card(AK_1)\ge card(AK_2)\ge card(AK_3)$

To simplify the proof the different subsets of the sets of the three chromosomes are nominated as follows:



A, B, C, ab, ac, bc and $\alpha\beta\gamma$ denote the number of elements of the corresponding subsets, e.g.:

$card\,(A_{AK1}) = A + ab + ac + \alpha\beta\gamma$ or

$card(A_{AK1}\cap A_{AK2}) = ab + \alpha\beta\gamma$ or

$card(A_{AK1}\cap A_{AK2}\cap A_{AK3}) = \alpha\beta\gamma$

Relation (A1.9) can be written as:

$$1+\frac{card(A_{AK1}\cap A_{AK3})}{max(card(A_{AK1}),card(A_{AK3}))}\ge\frac{card(A_{AK1}\cap A_{AK2})}{max(card(A_{AK1}),card(A_{AK2}))}+\frac{card(A_{AK2}\cap A_{AK3})}{max(card(A_{AK2}),card(A_{AK3}))} \qquad \text{(A1.11)}$$

With the suppositions of this case and the above notation of subsets relation (A1.11), which must be proved, can be written as:

$$1 + \frac{ac + \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} \geq \frac{ab + \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} + \frac{bc + \alpha\beta\gamma}{B + ab + bc + \alpha\beta\gamma}$$

$$\frac{A + ab + ac + \alpha\beta\gamma + ac + \alpha\beta\gamma - ab - \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} \geq \frac{bc + \alpha\beta\gamma}{B + ab + bc + \alpha\beta\gamma}$$

$$(A + 2ac + \alpha\beta\gamma) \cdot (B + ab + bc + \alpha\beta\gamma) \geq (bc + \alpha\beta\gamma) \cdot (A + ab + ac + \alpha\beta\gamma)$$

$$AB + Aab + 2Bac + 2ac \cdot ab + ac \cdot bc + ac \cdot \alpha\beta\gamma + B \cdot \alpha\beta\gamma \geq bc \cdot ab \qquad \text{(A1.12)}$$

The term $AB + Aab + Bac + ac \cdot ab$ is surely smaller than the left side of relation (A1.12) as it is a subset of the first 4 addends. Thus it can be written:

$$AB + Aab + Bac + ac \cdot ab = (A + ac) \cdot (B + ab) \geq (B + bc) \cdot (B + ab) = B^2 + Bab + Bbc + bc \cdot ab \geq bc \cdot ab$$

The right relation is based on the supposition $card(AK_1) \geq card(AK_2)$ from which $A + ac \geq B + bc$ can be derived. Thus relation (A.1.11) and with it relation (A1.9) is proved for this case.

<u>Case 3:</u> $card(AK_1) \geq card(AK_3) \geq card(AK_2)$

To prove case 3 the notation of case 2 is used also. Thus and based on relation (A1.11) it must be proved that:

$$1 + \frac{ac + \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} \geq \frac{ab + \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} + \frac{bc + \alpha\beta\gamma}{C + ac + bc + \alpha\beta\gamma}$$

$$\frac{A + ab + ac + \alpha\beta\gamma + ac + \alpha\beta\gamma - ab - \alpha\beta\gamma}{A + ab + ac + \alpha\beta\gamma} \geq \frac{bc + \alpha\beta\gamma}{C + ac + bc + \alpha\beta\gamma}$$

$$(A + 2ac + \alpha\beta\gamma) \cdot (C + ac + bc + \alpha\beta\gamma) \geq (bc + \alpha\beta\gamma) \cdot (A + ab + ac + \alpha\beta\gamma)$$

$$\underline{AC + Aac + 2Cac + 2ac^2} + ac \cdot bc + \underline{C \cdot \alpha\beta\gamma + 2ac \cdot \alpha\beta\gamma} \geq bc \cdot ab + ab \cdot \alpha\beta\gamma \qquad \text{(A1.12)}$$

The sum of the two underlined parts of the left side of relation (A1.12) is surely smaller than the complete left side. For the solid underlined part it can be stated that

$$AC + Aac + 2Cac + 2ac^2 \geq AC + Aac + Cac + ac^2 = (A + ac) \cdot (C + ac) \geq (B + bc) \cdot (B + ab) \geq bc \cdot ab \qquad \text{(A1.13)}$$

The two suppositions $card(AK_1) \geq card(AK_2)$ and $card(AK_3) \geq card(AK_2)$, from which $A + ac \geq B + bc$ and $C + ac \geq B + ab$ can be derived, were used in relation (A1.13).

For the dotted underlined part it can be stated that

$$C \cdot \alpha\beta\gamma + 2ac \cdot \alpha\beta\gamma \geq C \cdot \alpha\beta\gamma + ac \cdot \alpha\beta\gamma = \alpha\beta\gamma(C + ac) \geq \alpha\beta\gamma(B + ab) \geq \alpha\beta\gamma \cdot ab \qquad \text{(A1.14)}$$

Again supposition $card(AK_3) \geq card(AK_2)$ was used. With relations (A1.14) and (A1.13) relation (A.12) is proved and by that relation (A1.9) is proved for this case.

With the proofs for all three cases (A.4) is fulfilled.

Maximum Distance

The maximum distance $dist_{max}$ results, if genes are shifted by a maximal number of positions. This can be achieved in two ways: By inverting the order of the genes and by shifting each gene a maximal way. The cases of even and odd chromosome length must be distinguished.

Maximum Distance for Chromosomes of Even Length

Fig. 1 shows the two shift scenarios for chromosomes of even length.

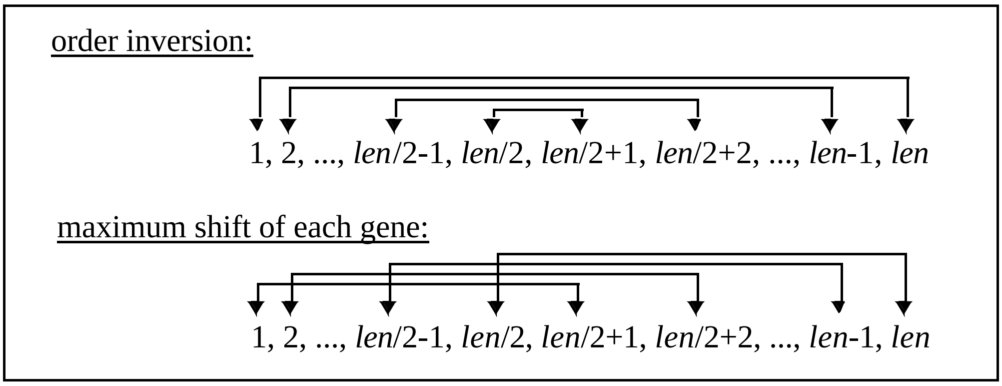


Figure 1: Two Shifting Scenarios for Chromosomes of Even Length

In case of order inversion there are *len/2* swappings of length *len-1, len-3*, … 3, 1. The distance sum $dist_{max,inv}$ of these shifts is

$$dist_{max,\,inv,\,even} = 2\sum_{i=1}^{len/2}(2i-1) = 4\sum_{i=1}^{len/2} i - len \qquad len \text{ even} \qquad (1)$$
$$= 4\frac{\frac{len}{2}\left(\frac{len}{2}+1\right)}{2} - len = len\left(\frac{len}{2}+1\right) - len = \frac{len^2}{2}$$

Shifting all genes by a maximal distance means *len* shifts of *len/2* positions, as shown in the lower part of Fig. 1. Genes of position *len/2* + 1 or higher are wrapped around the chromosome end. Their index Idx_{new} can be calculated as follows:

$$Idx_{new}(G_i) = \left(Idx(G_i) + \frac{len}{2}\right) \bmod len \qquad (2)$$

where $Idx\ (G_i)$ is the index of the gene G_i within the chromosome, also named as its position.

The resulting positional distance of all genes is *len* shifts multiplied by the length of *len/2,* which again results in $len^2/2$ as in Eq. (1).

Maximum Distance for Chromosomes of Odd Length

Fig. 2 shows the situation in the case of odd chromosome length and order inversion. When inverting the order *(len-1)/2* swappings of *len-1, len-3, …* 4, 2 positions occur, which results in the distance sum $dist_{max,inv}$ for odd *len.*

1, 2, ..., (*len*-1)/2, (*len*+1)/2, (*len*+3)/2, ..., *len*-1, *len*

Figure 2: Order Inversion for Chromosomes of Odd Length

$$dist_{max,\,inv,\,odd} = 2\sum_{i=1}^{\frac{len-1}{2}} 2i = 4\sum_{i=1}^{\frac{len-1}{2}} i = 4\,\frac{\frac{len-1}{2}\cdot\frac{len+1}{2}}{2} = 2\,\frac{len^2-1}{4} = \frac{len^2-1}{2} \qquad len \text{ odd} \qquad (3)$$

Shifting all genes by a maximum distance does not result in an equal distance for all genes due to the odd chromosome length. One can either have the shorter shifts first or the other way round. The first case results in $(len+1)/2$ shifts of distance of $(len-1)/2$ positions and $(len-1)/2$ remaining shifts of $(len+1)/2$ positions with a wrap around at the end of the chromosome. This results in a distance sum $dist_{max,shift,odd}$ of

$$dist_{max,\,shift,\,odd} = \frac{len+1}{2}\cdot\frac{len-1}{2} + \frac{len-1}{2}\cdot\frac{len+1}{2} = \frac{len^2-1}{2} \qquad len \text{ odd} \qquad (4)$$

If the longer paths are done first for reasons of symmetry the terms of the sum in Eq. (4) have to be swapped resulting in the same amount.

Thus it can be summarized:

$$\boxed{\begin{aligned} dist_{max,\,even} &= \frac{len^2}{2} \\ dist_{max,\,odd} &= \frac{len^2-1}{2} \end{aligned}}$$