Perimeter of rounded convex planar sets

László Csirmaz*

Abstract

A convex set is inscribed into a rectangle with sides a and $1/a$ so that the convex set has points on all four sides of the rectangle. By "rounding" we mean the composition of two orthogonal linear transformations parallel to the edges of the rectangle, which makes a unit square out of the rectangle. The transformation also applied to the convex set, which now has the same area, and is inscribed into a square. One would expect this transformation to decrease the perimeter of the convex set as well. Interestingly, this is not always the case. For each a we determine the largest and smallest possible increase of the perimeter.

1 Introduction

A (closed) convex set K is *inscribed* into the rectangle $ABCD$ if K has points on all sides of the rectangle. Suppose that the sides of the rectangle are $AB = a$ and $BC = 1/a$; thus its area is 1. We "round" the rectangle by squeezing in one direction, and stretching in the other one until it becomes a unit square. If we choose the coordinate axes on two sides of the rectangle, then this is achieved by the linear map $\varphi_a(x, y) = (x/a, ya)$. Now the image $K' = \varphi(K)$ is inscribed into the unit square $\varphi(ABCD)$. As K and K' have the same area, it is natural to expect that K' has smaller perimeter. Interestingly, this is not always the case. We determine the infimum and the supremum of the difference

$$\Delta_a(K) = \text{perimeter}(K) - \text{perimeter}(K')$$

as K runs over all convex sets inscribed into $ABCD$. Denote these values by $M(a)$, and $m(a)$, respectively:

$$\begin{aligned} M(a) &= \sup_K \Delta_a(K) \\ m(a) &= \inf_K \Delta_a(K). \end{aligned}$$

For values of a close to 1 the value of $m(a)$ is negative, which shows that the perimeter may indeed increase. The infimum and supremum are actually taken, and we determine the extremal sets as well.

*Central European University, Budapest

The problem arose when in a geometric construction of [1] we wanted to estimate the error term. The construction used the "rounded" image of a convex d-dimensional body, and the error term was a function of the surface of the body after applying the rounding transformation. We wished to transfer the estimate using the surface of the original body. Results of this paper show that in dimension 2 we can do it by losing a small factor only. For higher dimensions we were unable even to give estimates beyond some trivial facts. For the interested reader we recall the construction in the Appendix.

To state the result let us define the function $G_a(x)$ for positive values of the parameter a as

$$G_a(x) = \left(1 - \frac{1}{\sqrt{1+x^2}}\right) - \left(a - \frac{a^3}{\sqrt{a^4+x^2}}\right).$$

Observe that $G_a(0) = 0$. For each $a > 0$, $a \neq 1$ there exists a unique positive root of $G_a(x)$.

Theorem 1 *Let $ABCD$ be a rectangle with sides $AB = a$ and $BC = 1/a$, where $a > 1$. Let λ be the positive root of $G_{1/a}(x)$, and let λ' be the positive root of $G_a(x)$. Choose X be on the line segment AB so that $BC/BX = \lambda$ (it will always be an internal point of AB), and choose X' on BC, if possible, so that $AB/BX' = \lambda'$.*

The infimum $m(a)$ is attained by the parallelogram $AXCY$. The supremum $M(a)$ is attained by the parallelogram $AX'CY'$ if X' exists, and by the diagonal AC otherwise. The two cases are distinguished whether a is below (parallelogram) or above (diagonal) of the larger root ($\approx 3.048\ldots$) of $G_a(a^2) = 0$.

The paper is organized as follows. The next section explains our strategy to reach the result. Section 3 shows that any extremal configuration is a polygon with at most eight vertices. Section 4 enlists all cases which should be investigated. The result of sections 6 and of 7 are of independent interest: what is the case if K is a quadrangle or a triangle. Finally section 8 sums up the results.

2 How to attack the problem

Let $ABCD$ be a rectangle $ABCD$ with sides $AB = CD = a$ and $AD = BC = 1/a$, and let K be a closed convex set inscribed into the rectangle. The linear map $\varphi = \varphi_a$ defined in the previous section makes a unit square out of the rectangle. Our aim is to determine the infimum of the difference

$$\Delta_a(K) = \text{perimeter}(K) - \text{perimeter}(\varphi(K))$$

when K runs over all closed convex sets inscribed into $ABCD$. For every such K we can find an inscribed polygon K_1 such that the perimeters of K and K_1, and the perimeters of $\varphi(K)$ and $\varphi(K_1)$ are arbitrarily close simultaneously. Thus the infimum (and the supremum) of $\Delta_a(K)$ taken for inscribed convex polygons only is the same as that for all convex sets.

Inscribed convex polygons with at most n vertices form a compact ensemble, thus among them both the infimum and the supremum is taken. We refer to such polygons as *extremal* ones. Now take an extremal polygon K, and suppose that, say B is not a vertex in it. Then looking from B, K has a first point X

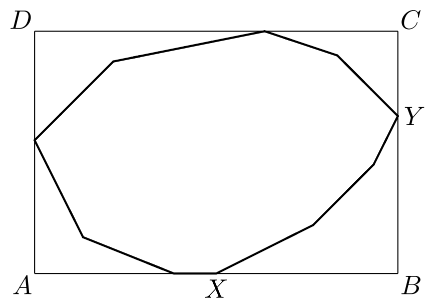


Figure 1: An inscribed convex polygon

on BA, and a first point Y on BC. The (shorter) XY arc of K is the graph of a convex function. We use methods from variational calculus to show that this arc might consist of at most two straight line segments. Further calculation shows that, in fact, it must be a single segment.

As a conclusion, an extremal polygon has at most eight vertices: at most two on each side of the rectangle. We further reduce this by showing that an extremal polygon cannot have two vertices on adjacent sides of the rectangle. Thus all such polygons have at most six vertices, consequently the infimum and the supremum of the difference $\Delta_a(K)$ is attained by such polygons.

Next we classify the candidate extremal polygons according to the number of vertices, and according whether among their vertices they have internal points of the edges of the rectangle, or not. To simplify the classification, we look at the inverse problem, namely the polygon is inscribed into the unit square, which is then distorted by the inverse transformation φ_a^{-1}. Also, we can flip the square across its diagonal, when a should be replaced by $1/a$. We investigate in detail three intriguing cases: when the polygon is a hexagon, quadrangle or triangle, and is in the most general position. Results achieved during this investigation show that in each remaining group there are at most two extremal candidates. Finally we use a computer program to pick up the extremal polygons for different values of a.

3 Reduction to 8 points

As explained in the previous section, here we consider the following scenario. The continuous convex function $f(x)$ is defined on the interval $[0, 1]$. It consist of straight line segments only, and its derivative at 0 is positive.

We stretch the graph of f by a in the x direction, and by $1/a$ in the y

direction. The resulting graph is defined by the function

$$f_a(x) = \frac{f(x/a)}{a},$$

and, of course $f(x) = f_1(x)$. The length of the full arc of f_a is

$$I_a(f) = \int_0^a \sqrt{1 + f_a'^2(x)}\, dx = \frac{1}{a}\int_0^1 \sqrt{a^4 + f'^2(x)}\, dx.$$

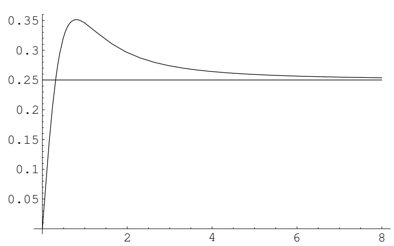
Our goal is to determine the function f so that $D_a(f) = I_a(f) - I_1(f)$ cannot be increased (or decreased) by slight modification of f. When this is the case, f is called *stationary.*

Theorem 2 *A stationary function consists of a single line segment only.*

If we prove this theorem we are done: no extremal polygon can have internal points in $ABCD$. It follows immediately from the theorem as it is stated, and by substituting $1/a$ for a.

Proof We start with two lemmas which will be used in the proof, and will be used later as well. To state the lemmas we choose a to be a positive number different from 1, and we let

$$h(x) = \frac{x}{a\sqrt{a^4 + x^2}} - \frac{x}{\sqrt{1 + x^2}}.$$



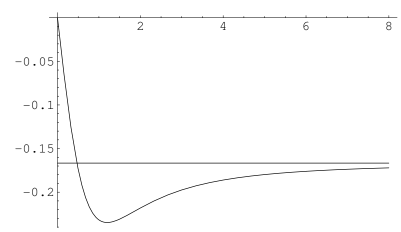


Figure 2: $h(x)$ for $a = 0.8$ and $a = 1.2$

Lemma 1 *For $0 \leq x$ the function $h(x)$ takes every value at most twice.*

Proof The claim follows from the fact that $h'(x)$ takes zero only at $x = a$, since then h is monotone in the intervals $(0, a)$ and (a, ∞). $\square$

Define the function $g(x)$ as

$$g(x) = \frac{1}{\sqrt{1 + x^2}} - \frac{a^3}{\sqrt{a^4 + x^2}}.$$

Lemma 2 *There are no two different values λ_1 and λ_2 such that $h(\lambda_1) = h(\lambda_2)$ and $g(\lambda_1) = g(\lambda_2)$ holds simultaneously.*

Proof Would this be the case, $g(\lambda_1) - ah(\lambda_1) = g(\lambda_2) - ah(\lambda_2)$ also holds. Thus we are done if we show that $g(x) - ah(x)$ is strictly increasing when $a > 1$, and strictly decreasing when $a < 1$. To this end, we compute the derivative of this difference, and check that it always has the same sign:

$$\begin{aligned} &g'(x) - ah'(x) = \\ &\quad = -\frac{x}{\left(1+x^2\right)^{3/2}} + \frac{a^3 x}{\left(a^4+x^2\right)^{3/2}} - \frac{a^4}{\left(a^4+x^2\right)^{3/2}} + \frac{a}{\left(1+x^2\right)^{3/2}} \\ &\quad = (a-x)\left(\frac{1}{\left(1+x^2\right)^{3/2}} - \frac{a^3}{\left(a^4+x^2\right)^{3/2}}\right) \\ &\quad = (a-x)\frac{\left(a^4+x^2\right)^3 - a^6\left(a+x^2\right)^3}{\left(1+x^2\right)^{3/2}\left(a^4+x^2\right)^{3/2}\left(\left(a^4+x^2\right)^{3/2} + a^3\left(a+x^2\right)^{3/2}\right)} \end{aligned}$$

The denominator is always positive, the enumerator can be factored as follows:

$$(a-x)(x^2-a^2)\big(x^4(1-a^6) + a^2x^2(1+3a^2-3a^4-a^6) + a^4(1-a^6)\big).$$

The last term has discriminant

$$\left(1+3a^2-3a^4-a^6\right)^2 - 4\left(1-a^6\right)\left(1-a^6\right) = -\left(1-a^2\right)^4\left(2+5a^2+2a^4\right)$$

which is negative, thus the last term always has the same sign as the leading factor $1-a^6$. The whole product is non-negative if $a > 1$, not positive if $a < 1$, and is zero for $x = a$ only, as was claimed. □

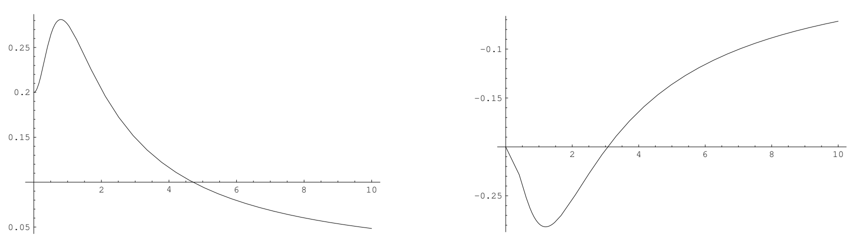


Figure 3: $g(x)$ for $a = 0.8$ and $a = 1.2$

Now we are ready to prove the theorem. First choose a function $s(x)$ so that for small enough values for ε, $f + \varepsilon s$ is still convex, moreover $s(0) = s(1) = 0$. If f is a stationary solution of the problem, then the ε-derivative of the difference $I_a(f + \varepsilon s) - I_1(f + \varepsilon s)$ must be zero at $\varepsilon = 0$. Now

$$\frac{\partial}{\partial \varepsilon} I_a(f + \varepsilon s)\bigg|_{\varepsilon=0} = \int_0^1 \frac{f'(x)s'(x)}{a\sqrt{a^4 + f'^2(x)}}\, dx$$

thus for all feasible functions s as above we must have

$$(1)\quad 0=\int_0^1\left(\frac{f'(x)}{a\sqrt{a^4+f'^2(x)}}-\frac{f'(x)}{\sqrt{1+f'^2(x)}}\right)s'(x)\,dx=\int_0^1 H(x)s'(x)\,dx.$$

Now $f'(x)$ is piecewise constant, and doesn't take the value 0. Therefore $H(x)$ is also piecewise constant. Now suppose $H(x)$ takes two different values, say h_1 on the interval (b,c) and h_2 on the interval (c,d). Define the function $s'(x)$ so that it is zero for values less than b and greater than d, takes constant negative value on (b,c), and constant positive value on (c,d) such that its integral on $(0,1)$ vanishes. Then $f+\varepsilon s$ is convex for small enough values of ε, as $f'(x)+\varepsilon s'(x)$ is increasing. However, the integral in (1) is not zero in this case.

Consequently for all stationary functions f, $H(x)$ must be constant. By Lemma 1 the function $H(x)$ takes the same value at no more than two different values of $f'(x)$. It means that the graph of f might have at most two different slopes, that is, f is either a single segment, or consist of two joining segments only.

Our second task is to exclude the possibility of two segments. Suppose the two segments join at (x,y), and for the simplicity we denote $f(1)$ by v. Then the change of the arc length is

$$\begin{aligned} D(x,y) &= \sqrt{a^2x^2+\frac{y^2}{a^2}}+\sqrt{a^2(1-x)^2+\frac{(v-y)^2}{a^2}}- \\ &\quad -\sqrt{x^2+y^2}-\sqrt{(1-x)^2+(v-y)^2}, \end{aligned}$$

and it must have zero partial derivative by both x and y:

$$\begin{aligned} 0=\frac{\partial D}{\partial x} &= \frac{a^2x}{\sqrt{a^2x^2+\frac{y^2}{a^2}}}-\frac{a^2(1-x)}{\sqrt{a^2(1-x)^2+\frac{(v-y)^2}{a^2}}}- \\ &\quad -\frac{x}{\sqrt{x^2+y^2}}+\frac{1-x}{\sqrt{(1-x)^2+(v-y)^2}}, \\ 0=\frac{\partial D}{\partial y} &= \frac{y}{a^2\sqrt{a^2x^2+\frac{y^2}{a^2}}}-\frac{v-y}{a^2\sqrt{a^2(1-x)^2+\frac{(v-y)^2}{a^2}}}- \\ &\quad -\frac{y}{\sqrt{x^2+y^2}}+\frac{v-y}{\sqrt{(1-x)^2+(v-y)^2}}. \end{aligned}$$

Let the slopes be $y/x=\lambda_1$, and $(v-y)/(1-x)=\lambda_2$. Substituting these values into the above equations we get

$$\begin{aligned} 0 &= -\frac{1}{\sqrt{1+\lambda_1^2}}+\frac{1}{\sqrt{1+\lambda_2^2}}+\frac{a^3}{\sqrt{a^4+\lambda_1^2}}-\frac{a^3}{\sqrt{a^4+\lambda_2^2}} \\ 0 &= -\frac{\lambda_1}{\sqrt{1+\lambda_1^2}}+\frac{\lambda_2}{\sqrt{1+\lambda_2^2}}+\frac{\lambda_1}{a\sqrt{a^4+\lambda_1^2}}-\frac{\lambda_2}{a\sqrt{a^4+\lambda_2^2}}. \end{aligned}$$

That is, λ_1 and λ_2 satisfy $g(\lambda_1) = g(\lambda_2)$ and $h(\lambda_1) = h(\lambda_2)$, which is impossible by Lemma 2. This proves the theorem. □

4 Listing of cases

To simplify the presentation we consider the inverse problem, as was explained in section 2. We start from a unit square, and apply the inverse transformation $\varphi_a^{-1} = \varphi_{1/a}$. In this case only the sign of the difference

$$\Delta_a(K) = \text{perimeter}(K) - \text{perimeter}(\varphi_{1/a}(K))$$

changes, thus extremal polygons remain extremal ones. Also, we can flip the square across its diagonal, which amounts to substituting $1/a$ for a.

We proved in the previous section that an extremal polygon has no vertex inside the square, consequently it has at most eight vertices. Now we show that it cannot have two vertices on each of two adjacent sides. Indeed, suppose this is the case (see figure 4). Replace the cord of length z by the parallel segment of

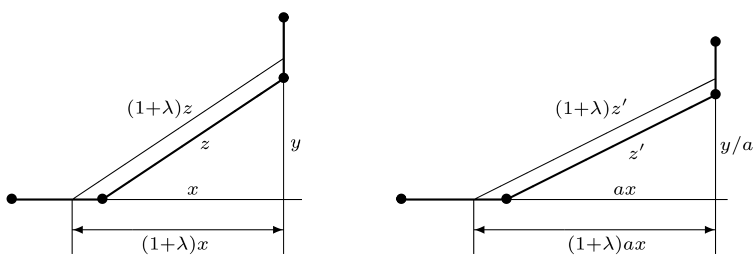


Figure 4: Vertices on two adjacent sides

length $(1+\lambda)z$ where λ has small absolute value. The perimeter of K changes by

$$(1+\lambda)z - z - \lambda x - \lambda y = \lambda(z - x - y).$$

Looking at $\varphi_a(K)$ we see that its perimeter changes by $\lambda(z' - ax - y/a)$. Thus $\Delta_a(K)$ changes by

$$\lambda(z - x - y - z' + ax - y/a).$$

As K is stationary, we should not be able to choose λ small enough so that this change takes both positive and negative values. This means that the coefficient of λ must be zero. In this case we can choose $\lambda = -1$ without affecting the difference; that is, we may drag the hypotenuse into the corner. □

Depending on how many vertices K has and how many of them are in the corners of the square, we can classify the possible extremal polygons into nineteen classes. Figure 5 shows typical element for all but one of this classification; the missing case consists of a single diagonal. The cases are denoted by one

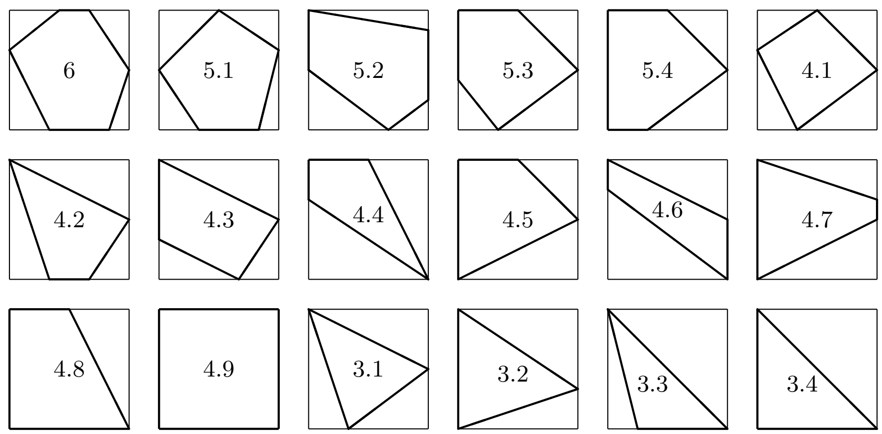


Figure 5: Classification of extremal polygons

or two digits: the first one gives the number of vertices, and the second one refers to the subcase. We investigate the configurations 6 and 4.1 in detail. The results will help us to clarify the situation in other cases as well.

5 The hexagonal case

There is only one possible arrangement with six vertices: two opposite sides of the square contain two vertices, and the other two contains one–one vertex. Consider two joining edges of K one of which lies on the side of the square. As we have an extremal configuration, the vertex joining these edges is in a stable position. This property is exploited further.

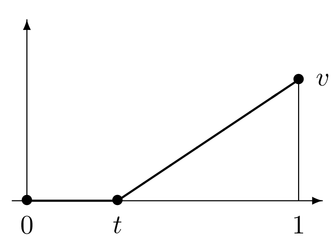


Figure 6: Two vertices on a single side

We return to the model in section 3. The two joining segments form the graph of a function $f(x)$ defined on the unit interval $[0, 1]$ as depicted on figure 6. f starts with a horizontal segment and it is zero up to t; then climbs up linearly to the value $f(1) = v$. By assumption f is stationary. It means that varying the break point t slightly, the arc length difference $D = I_a(f) - I_1(f)$ cannot both increase and decrease. In other words, this difference, as a function of t, has zero derivative.

Lemma 3 *Suppose that f is stationary. Then the slope of the second segment is determined uniquely by the value of a. Moreover the difference D, as a function of t, attains a maximal value when $a < 1$, and a minimal value when $a > 1$.*

Proof The arc length difference $D = I_a(f) - I_1(f)$, as a function of t, can be written as follows:

$$D(t) = at + \sqrt{a^2(1-t)^2 + \frac{v^2}{a^2}} - t - \sqrt{(1-t)^2 + v^2}.$$

As explained above, this function has zero derivative at t:

$$0 = D'(t) = a - \frac{a^2(1-t)}{\sqrt{a^2(1-t)^2 + \frac{v^2}{a^2}}} - 1 + \frac{1-t}{\sqrt{(1-t)^2 + v^2}}$$

Substituting λ for the slope $v/(1-t)$ we get that λ satisfies

$$1 - a = \frac{1}{\sqrt{1+\lambda^2}} - \frac{a^3}{\sqrt{a^4+\lambda^2}} = g(\lambda).$$

The derivative of g vanishes only at a, consequently g is strictly monotone for values less than a; and also for values greater than a. As $g(0) = 1 - a$, it can take this value only once, and only when $\lambda > a$.

Therefore $D'(t)$ vanishes at most in a single place only, say at t_0. This is not a zero of even order, as $g(\lambda)$ crosses the value $1-a$. (In fact, this is a single zero.) As the sign of $D'(1) = a-1$ is positive or negative depending on whether $a < 1$ or $a > 1$, at t_0 the function $D'(t)$ changes from positive to negative in the first case, and from negative to positive in the second. This means $D(t)$ has a maximum in the first case, and a minimum in the second, as was claimed. □

By the lemma the opposite sides of an optimal hexagon K are parallel, and form the same angle with the opposite sides of the rectangle. In this case the perimeter of K equals to the perimeter of the parallelogram depicted on figure 7; and the same is true for $\varphi(K)$. Consequently en extremal hexagon cannot yield better values than a parallelogram in case 4.6.

The same reasoning shows that extremal values of case 5.2 are also included in case 4.6. Case 5.4 reduces either to 4.5 or 4.8 as the "tip" on the right hand side can be moved vertically while keeping the edges starting from the tip parallel to their original direction without affecting the perimeter. Finally in case 4.7 the bottom segment can be "flipped" without affecting the perimeter, thus this case is also reduced to 4.6.

If the extremal polygon maximizes Δ_a then it might have an initial segment on a side of the square only if that side shrinks ($a < 1$), and not on side

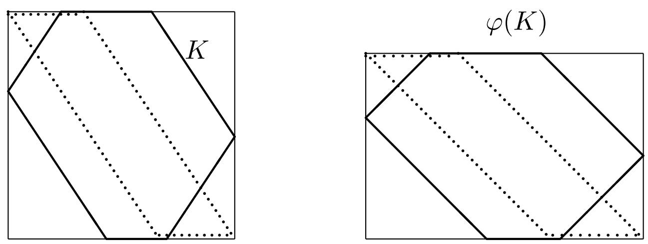


Figure 7: The hexagonal case

which expands. Similarly, if the extremal polygon minimizes Δ_a, then such a configuration may occur only on sides which expand. This shows that cases 5.3 and 4.4 cannot occur at all.

Finally, in cases 4.6 we have two possible configurations, depending on whether the parallel sides of the extremal polygon are stretched by a or by $1/a$.

6 The quadrangle case

In this section we consider the case when the extremal polygon K is a quadrangle. The most general situation is when the vertices are interior points on the sides of the square, that is, case 4.1. Picking any vertex, let us denote the slopes of the two segments starting from that vertex by λ_1 and λ_2. First we give a necessary condition these values must always satisfy.

Lemma 4 *Consider a vertex of the extremal polygon K which is an internal point on an edge of the square. The slopes λ_1 and λ_2 of the two joining segments satisfy the equation $g(\lambda_1) = g(\lambda_2)$.*

Proof As before, let the function f mimic the two joining segments, see figure 8. That is, f is defined on the unit interval, $f(0) = v_1$, $f(t) = 0$, $f(1) = v_2$, and

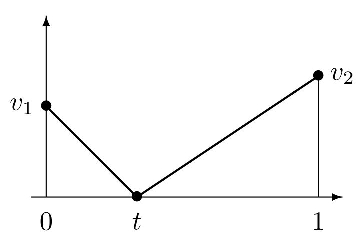


Figure 8: Vertex in an internal point

f is linear between 0 and t, as well as between t and 1. After stretching the graph, the change in the arc length is

$$D(t) = \sqrt{\frac{v_1^2}{a^2} + a^2t^2} + \sqrt{\frac{v_2^2}{a^2} + a^2(1-t)^2} - \sqrt{v_1^2 + t^2} - \sqrt{v_2^2 + (1-t)^2}.$$

This function has zero derivative at t; substituting λ_1 for v_1/t, and λ_2 for $v_2/(1-t)$ we get

$$0 = \frac{a^3}{\sqrt{a^4 + \lambda_1^2}} - \frac{a^3}{\sqrt{a^4 + \lambda_2^2}} - \frac{1}{\sqrt{1 + \lambda_1^2}} + \frac{1}{\sqrt{1 + \lambda_2^2}},$$

that is $g(\lambda_1) = g(\lambda_2)$, as was claimed. □

Lemma 5 *Suppose λ_1, λ_2, and μ_1, μ_2 are positive numbers so that $g(\lambda_1) = g(\lambda_2)$, $g(\mu_1) = g(\mu_2)$, $h(\lambda_1) = h(\mu_1)$, and $h(\lambda_2) = h(\mu_2)$. If $\lambda_1 \neq \lambda_2$, then $\mu_1 = \lambda_1$ and $\mu_2 = \lambda_2$.*

Proof For the sake of simplicity we assume $a < 1$, the case $a > 1$ can be handled similarly. From the proof of Lemma 2 we know that $g(x) - ah(x)$ is strictly decreasing. As $g(a) - ah(a) = 0$ it means that for $x < a$ we have $g(x) > ah(x)$, and for $x > a$ the value of $g(x)$ is below that of $ah(x)$. Also, as λ_1 and λ_2 differ, and g takes the same value at these places, one of them must be below a, and the other above a, say $\lambda_1 < a < \lambda_2$. Let λ_1' be the other place (if it exists) where $ah(\lambda_1) = ah(\lambda_1')$. Then $a < \lambda_1'$ and

$$ah(\lambda_1') = ah(\lambda_1) < g(\lambda_1) = g(\lambda_2) < ah(\lambda_2).$$

As $h(x)$ is strictly decreasing for $x > a$, this gives $\lambda_1' > \lambda_2$. Similarly, $\lambda_1 < \lambda_2' < a$ where $ah(\lambda_2') = ah(\lambda_2)$ and λ_2' and λ_2 differ.

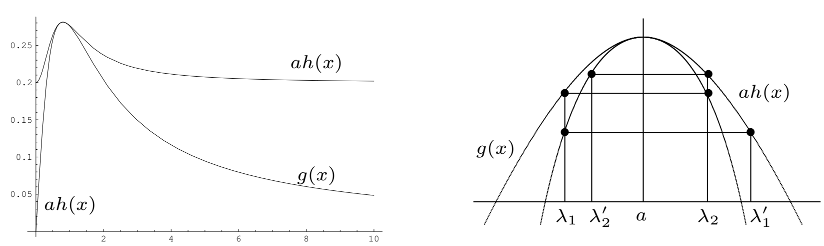


Figure 9: $g(x)$ and $ah(x)$ for $a = 0.8$

The same argument shows that $g(x)$ may take the value $g(\lambda_1')$ only for $x < \lambda_1$, and the value $g(\lambda_2')$ for $a < x < \lambda_1$.

By assumption μ_1 is equal to either λ_1 or λ_1', and μ_2 is one of λ_2 and λ_2'. We also know that $g(\mu_1) = g(\mu_2)$ which is possible only if $\mu_1 = \lambda_1$ and $\mu_2 = \lambda_2$ as required. □

Let $PQRS$ be an extremal quadrangle so that each vertex is an internal point on the corresponding side of the square. We claim that either at P and R, or at Q and S the two edges starting from a vertex share the same slope (but this can be different for the two vertices).

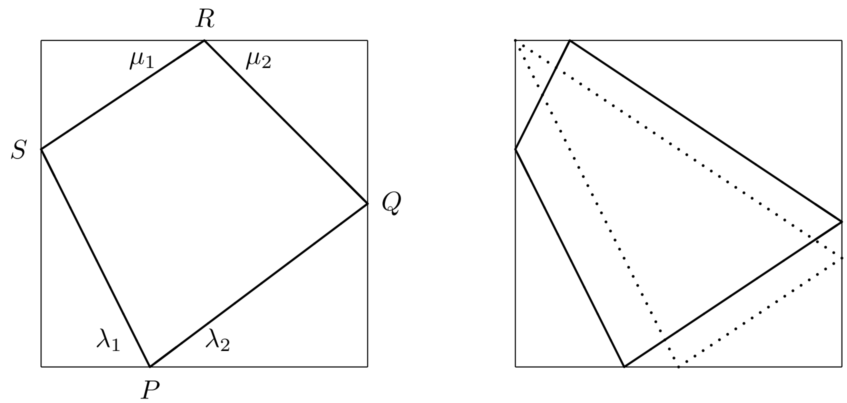


Figure 10: An extremal quadrangle

Suppose this is not the case. Let the slopes of PS and PQ be λ_1 and λ_2, and the slopes of RS and RQ be μ_1 and μ_2, respectively, and suppose $\lambda_1 \neq \lambda_2$. As $PQRS$ is an extremal quadrangle, we can apply lemma 4 for P and R yielding $g(\lambda_1) = g(\lambda_2)$ and $g(\mu_1) = g(\mu_2)$. The slopes at S are $1/\lambda_1$ and $1/\mu_1$, respectively. We can apply lemma 4 also at these points, but in this case the value of a should be be replaced by $1/a$:

$$g_{1/a}(1/\lambda_1) = g_{1/a}(1/\mu_1), \text{ and } g_{1/a}(1/\lambda_2) = g_{1/a}(1/\mu_2).$$

Now it is easy to see $g_{1/a}(1/x) = -h_a(x)$, thus $h(\lambda_1) = h(\mu_1)$ and $h(\lambda_2) = h(\mu_2)$ also holds. By lemma 5 then $\lambda_1 = \mu_1$ and $\lambda_2 = \mu_2$. This shows that at S and at Q the edges share the same slope.

In this case if we move the vertices so that the edges remain parallel to their original position, the perimeter of the rectangle does not change. Thus we can push at least one vertex into the corner, reducing this case either to 3.1 or to the case of a single diagonal.

7 The triangle case

The only remaining group with infinitely many candidates is 3.1, when two vertices of the inscribed triangle are internal points on two sides of the square. This group is, in fact, empty; however we show only that it might contain a single candidate. Our starting point is figure 8. We choose $v_2 = 1$, and plot those pairs (t, v), where the difference function

$$D(t) = \sqrt{\frac{v^2}{a^2} + a^2t^2} + \sqrt{\frac{1}{a^2} + a^2(1-t)^2} - \sqrt{v^2 + t^2} - \sqrt{1 + (1-t)^2}$$

has zero t-derivative. In section 6 we saw that this is the case exactly when $g_a(v/t) = g_a(1/(1-t)$, which can be written in an equivalent form of

$$h_{1/a}(1-t) = h_{1/a}(t/v).$$

Denote by $H(a)$ the set of pairs (t, v) in the unit square for which $h_a(1-t) = h_a(t/v)$. On figure 11 we can identify this situation twice: once for $CBAP$ from we get $(AQ, AP) \in H(1/a)$, and for $CDAQ$ from which we have $(AP, AQ) \in H(a)$. Thus in an extremal configuration the (t, v) pair should be an element of

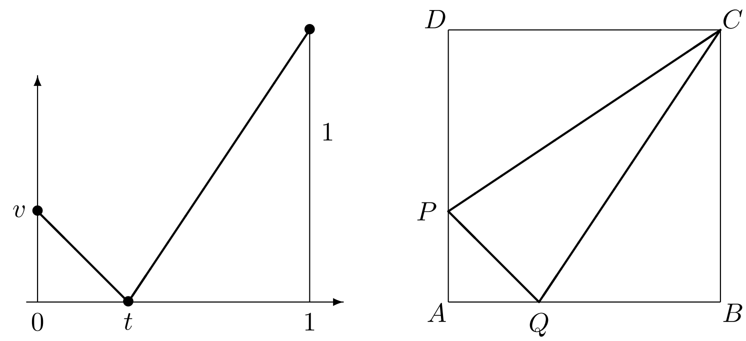


Figure 11: The triangle case

$H(a)$ as well as the mirror image of $H(1/a)$ across the diagonal. We shall see that these sets have at most a single internal point in common.

The point (t, v) is in $H(a)$ if either $1-t = t/v$, or α and β are two different values with $h(\alpha) = h(\beta)$, and $1-t = \alpha$ and $t/v = \beta$.

In the first case (t, v) is on an arc of a hyperbola, and is above the diagonal $t = v$. In the second case for each value of t we have at most one corresponding value for v, and as h is analytic, this implicit function is analytic, too. As $h(x)$ tends to $-1 + 1/a$ at the infinity, α and β can take all values above ξ, where $h(\xi) = -1 + 1/a$. One of α and β is smaller than a, and the other one is bigger, as can be seen on figure 2. Consequently the implicit function is defined on the interval $0 \le t < 1 - \xi$. This interval is empty if $\xi \ge 1$, which can only happen when $a > 1$ and $h(1) \ge 1/a - 1$, that is, when $a > 3.0491\ldots$. Otherwise the function is convex from below; we show only that it is below the tangent line drawn at $(0, 0)$.

Indeed, for a given t let the corresponding values be $\alpha(t)$ and $\beta(t)$, then it is easy to see that

$$\frac{dv}{dt} = \frac{1}{\beta(t)} + \frac{t}{\beta^2(t)} \cdot \frac{h'(\alpha(t))}{h'(\beta(t))}.$$

As t increases from 0 to $1 - \xi$, $\alpha(t)$ decreases, and $\beta(t)$ increases. As $\alpha(t)$ and $\beta(t)$ are on different sides of a, $h'(\alpha(t))$ and $h'(\beta(t))$ have different signs. Thus the derivative dv/dt is always smaller than $1/\beta(0)$, which is the slope of the tangent at $t = 0$. From here our claim follows easily. The value $\beta_0 = \beta(0)$ is the only other place where $h(\beta_0) = h(1)$, and either $\beta_0 < a < 1$, or $1 < a < \beta_0$.

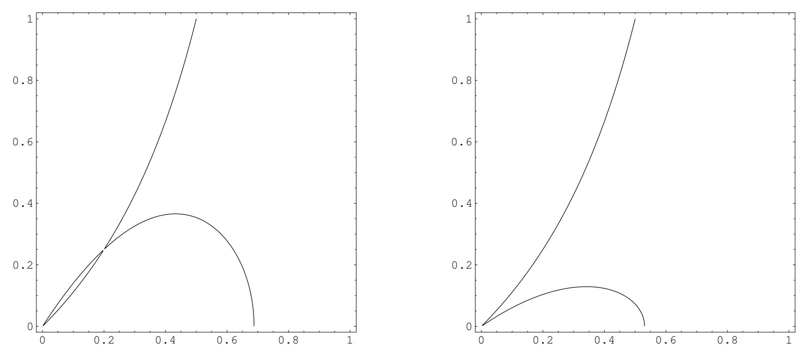


Figure 12: The set $H(a)$ for $a = 0.8$ and $a = 1.2$

Consequently, for $a > 1$ the tangent separates the curve and the hyperbola arc, while for $a < 1$ it does not, and the curve must have a point on the hyperbola (but not necessarily for $v < 1$). The point (t, v) is also on the hyperbola if, and only if, $\alpha(t) = \beta(t)$, that is, when $t = 1 - a$.

Next superpose $H(a)$ and the mirror image of $H(1/a)$ as on figure 13, and look for the common points of the two sets. The hyperbolic arcs are separated by the diagonal from $(0, 0)$ to $(1, 1)$; the curve of $H(a)$ is below the tangent line starting from $(0, 0)$ with slope $1/\beta_a(0)$, and the curve of $H(1/a)$ is above the line with slope $\beta_{1/a}(0)$. Thus $\beta_{1/a}(0) > 1/\beta_a(0)$ proves that these curves have neither common points. The only possibility left is that the hyperbolic arc from $H(1/a)$ and the curve from $H(a)$ (or vice versa if a happens to be bigger than one) intersect, as indicated on the picture. As the pair (t, v) is on the hyperbolic arc, the slopes are equal. Thus in an extremal configuration either the angles $\angle DPC$ and $\angle QPA$ are equal, or $\angle PQA$ and $\angle CQB$ are equal.

Thus we have to prove that $\beta_{1/a}(0) > 1/\beta_a(0)$. Denote these values by β' and β, respectively. Then $h_{1/a}(\beta') = h_{1/a}(1)$, that is, $g_a(1/\beta') = g_a(1)$, and, of course, $ah_a(\beta) = ah_a(1)$ (see figure 13). As in the proof of lemma 5 we use

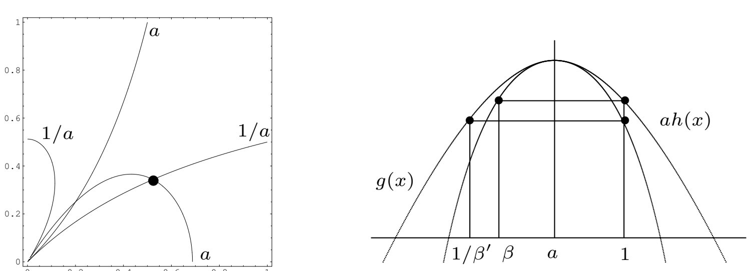


Figure 13: Superposing $H(a)$ and $H(1/a)$ for $a = 0.8$

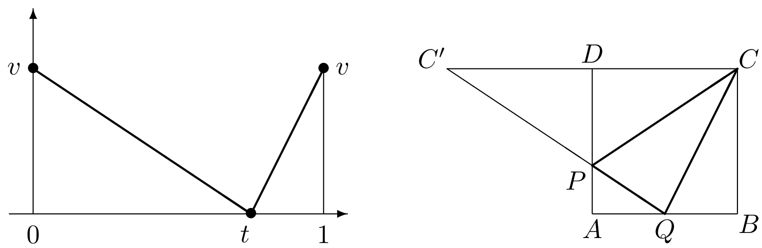


Figure 14: Triangle with equal slopes at P

that the function $g(x) - ah(x)$ is strictly decreasing for $a < 1$, yielding $1/\beta' < \beta$ immediately; or strictly increasing for $a > 1$, yielding the same relation. Thus $1/\beta < \beta'$, as was claimed.

We have seen that in the case of an extremal triangle PQC, either the slopes of PC and PQ are equal, or the slopes of PQ and QC are equal. Supposing that the first case holds, reflect CP across AD. The perimeter of the triangle is the same as the total length of the segments $C'Q$ and QC. Thus an extremal triangle yields an extremal two-segment graph $C'QC$, where the points C and C' are fixed, and Q runs on the side AB. The problem is the same as finding the extremal two-segment function f with $v_1 = v_1 = v = 1/2$. From lemma 4 we know, that this happens if, and only if, the slopes $\lambda_1 = v/t$, and $\lambda_2 = v/(1-t)$ satisfy $g_a(\lambda_1) = g_a(\lambda_2)$. Definitely this is the case if t is the midpoint – then the triangle degenerates to the diagonal AC. For any fixed v there are at most two other values of t, symmetrical to the midpoint, when this condition holds. This follows from the lemma below, as the condition for extremality can be written as $h_{1/a}(t/v) = h_{1/a}((1-t)/v)$ and $t/v + (1-t)/v = 1/v$ is fixed.

Lemma 6 *Suppose α and β runs over all values $0 < \alpha \leq a \leq \beta$ such that $h_a(\alpha) = h_a(\beta)$. Then $\alpha + \beta$ takes every value at most once.*

Proof Recall that $0 < \xi < a$ is the only place where $h(\xi) = h_a(\xi) = -1 + 1/a$. As the possible values of α is the interval $(\xi, a]$, the claim follows immediately, when we show that for all corresponding (α, β) pairs $h'(\alpha) + h'(\beta)$ has the same sign. Indeed, if α travels backward from a towards 0, then β will go always faster (or slower) than α. Thus their midpoint will always travel away from a (or toward the origin), and thus cannot take the same value twice.

As $h'(\alpha) = h'(\beta) = 0$ when $\alpha = \beta = a$, the sum $h'(\alpha) + h'(\beta)$ never changes sign if it does not take the value 0 anywhere else. Would this be case, we would have two different values α and β with $h(\alpha) = h(\beta)$ and $h'(\alpha) + h'(\beta) = 0$. But this is impossible if for appropriately chosen constant c, the function

$$F(x) = h(x) - c \cdot {h'}^2(x)$$

is either increasing or decreasing on the whole interval $\xi < x < +\infty$, that is, the derivative of $F(x)$ has the same sign. In what follows we argue for the case $a < 1$. The derivative of $F(x)$ is

$$F'(x) = h'(x)\,(1 - 2c \cdot h''(x))\,,$$

and $h'(x)$ takes zero at $x = a$ only: it is negative for $x < a$, and positive for $x > a$. Thus the second factor must change sign in the interval $(\xi, +\infty)$ at $x = a$ only. It implies that c must be chosen so that $1 - 2c \cdot h''(a) = 0$, and the lemma is proved if we can show that $h''(x)$ crosses the line $y = h''(a)$ only at $x = a$ in the interval $(\xi, +\infty)$. The function of $h''(x)$ is depicted on figure 15, from

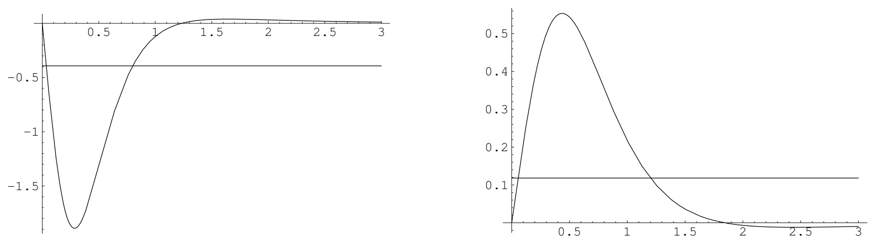


Figure 15: $h''(x)$ for $a = 0.8$ and $a = 1.2$

where it is clear, that this happens if, and only if, $h''(\xi) < h''(a)$ for $a < 1$, and if $h''(\xi) > h''(a)$ for $a > 1$.

Unfortunately we were not able to find an easy argument for this fact, but it has been checked by distinguishing several intervals for a. The check always went by finding an $\eta < \xi$ for which $h''(\eta) < h''(a)$. $\square$

We remark, that $\xi = \xi(a)$ has the asymptotical value $0.5a\sqrt{2a - 3a^2}$ when a tends to zero, and $0.5\sqrt{2a - 3}$ when a tends to infinity.

8 Conclusion

Let the sides of the rectangle $ABCD$ by $AB = CD = a$, and $BC = DA = 1/a$ with $a > 1$. The linear transformation φ_a maps $ABCD$ into the unit square. The (closed) convex set K is inscribed into $ABCD$ if K has points on all sides of the rectangle. We wanted to know how much the perimeters of K and $K' = \varphi_a(K)$ might differ, namely we were looking for the values

$$\begin{aligned} M(a) &= \sup_K \text{perimeter}(K) - \text{perimeter}(K') \\ m(a) &= \inf_K \text{perimeter}(K) - \text{perimeter}(K'). \end{aligned}$$

We have shown that the supremum and the infimum is actually taken by convex polygons, and then reduced the possibilities to members of 19 classes. By a more

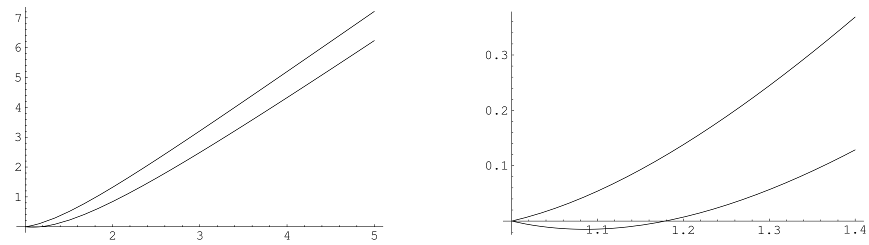


Figure 16: Maximal and minimal changes in the perimeter

detailed investigation we eliminated several of them, and showed that in the rest it is enough to consider only a finite number of candidates. For the classification we refer to figure 5; the only unlisted possibility is the diagonal.

Cases in the first row has been eliminated. Case 4.2 has four candidates. By lemma 3 the angles at the "base" are equal and determined uniquely. One of the two slopes of the segments at the right hand side vertex is thus fixed. Lemma 4 says that then the other slope can assume only two different values. This gives two candidates; other two come when the configuration is "flipped over" the diagonal.

Case 4.3 has also four candidates. As we have two internal vertices, two applications of lemma 4 yields four possibilities. However the slopes of the segments starting from the bottom vertex must be different. Would they be equal, we can shift it to the right without affecting perimeters, arriving at a configuration which is covered either in case 4.6 or in case 5.2.

Cases 4.4 and 4.7 were ruled out; there are four candidates in 4.5, and two in 4.6. By lemma 6 and the discussion before the lemma we have four candidates for 3.1 depending on where the slopes are equal. By the same lemma case 3.2 has also four candidates: the "peak" of the triangle can be either at the middle of the side, or at a uniquely determined point.

Altogether, for a fixed value of a the number of candidates is at most 27. We used a computer program [3] to pick up the extremal configuration among the candidates. For large values of a, not surprisingly, both $M(a)$ and $m(a)$ is a little bit less than $2a$. The maximum $M(a)$ is taken by the diagonal, and $m(a)$ is taken by the parallelogram from case 4.6; edges are part of the longer sides of the rectangle. As a gets smaller than $3.048\ldots$, the maximal difference is taken over by the "flipped" 4.6 parallelogram. From this point on both extremal cases remain parallelograms. As they are coming from case 4.6, by lemma 3 the slope of the parallelogram satisfies $g(\lambda) = 1 - a$, which is exactly the condition given in Theorem 1. For maximum the diagonal takes over the parallologram when the slope from $g(\lambda) = 1 - a$ would yield a point outside of the rectangle. This happens when $\lambda \geq a^2$, i.e. when $g(a^2) \leq 1 - a$, which proves the last claim in Theorem 1.

The values of $M(a)$ and $m(a)$ are plotted on figure 16. When a is smaller than $1.178\ldots$, then $m(a)$ becomes negative. This means that the "rounder" figure has larger perimeter; however the maximal increase is the quite small, around 0.015.

Results of the computer run show that each configuration can be locally extremal with two exceptions: the triangle case 3.1 never yields a configuration which cannot be improved by slight move of the vertices; and the same is true for case 4.3 when the second pair of slopes are equal.

Restricting ourselves for the triangle case only, extremal configurations show a greater variety. For large values of a, the maximum is taken by the diagonal. As a goes below 3, the maximum is taken by a triangle from 3.3, and when a is below 1.83, it changes to 3.2 with the third vertex at the midpoint. The minimum is always taken by a triangle from 3.2, however for $a > 2$ the third vertex is at the "tilted" position; for $a < 2$ it is the midpoint. As a goes below 1.14, the minimum becomes negative as well, but remains above -0.0124, which is slightly larger than the absolute minimum -0.015.

References

[1] L.Csirmaz, G.O.H. Katona *Geometrical Cryptography*, Proceedings of the International Workshop on Coding and Cryptopgraphy, Eds: D.Augot, P.Charpin, G.Kabatianski, INRIA, 2003 pp.101–111

[2] P.R.Scott, P.Q. Awyong, *Inequalities for convex sets*, J. Ineq. Pure and Applied Mathematics 1(1), 2000, `http://jipam.vu.edu.au/`

[3] Wolfram Research, Inc: *Mathematica*, `http://www.wolfram.com`

[4] I.M.Yaglom, V.G.Boltyanskii, *Convex Figures*, Holt, Reinhart and Winston, New York, 1961

Appendix: The background

The need of estimating the perimeter of a rounded convex body arose in connection with a construction in [1]. After several reductions we faced the following problem. Given are a d-dimensional centrally symmetric convex body K of volume 1 (the "unit ball" of some norm), and two (small) positive real numbers σ and ε as parameters. We had to construct (convex) bodies A_1, A_2, ... so that

(i) all A_i have volume $\le \sigma$,

(ii) with $B_i = A_i + \varepsilon K$, all B_i are pairwise disjoint, and are inside K,

(iii) the *utilization*, i.e. the total volume of all A_i is as large as possible.

Using Minkokwski's inequality we have derived the upper bound $\sigma(\sigma^{1/d} + \varepsilon)^{-d}$ on the utilization. For a lower bound we used the following construction. First, enclose K in a rectangular parallelepiped choosing the first two facets perpendicular to the diameter of K, and continuing recursively. This way the volume of the parallelepiped is at most $d!$. Next, use an affine transformation φ which maps this parallelepiped into a cube of the same volume. We will do the construction using $K' = \varphi K$ rather than K as φ preserves volume, Minkowski-sum, and convexity. Observe that K' can be enclosed into a cube of side length $(d!)^{1/d} \approx d/e$.

Pack the space with cubes C_i of size length $\eta = \sigma^{1/d} + 2\varepsilon(d!)^{1/d}$. A_i will be the cube of volume σ centered within C_i. Clearly $A_i + \varepsilon K' \subseteq C_i$, thus conditions (i) and (ii) hold. To get an estimate on the utilization we need to estimate how many of the C_i's are inside K'.

The diameter of C_i is $\sqrt{d}$ times of its side length. Thus either all points of C_i are inside K', or, if it has a point inside K', all points of C_i are closer to the surface of K' than $\sqrt{d}\,\eta$. As K' is convex, the total volume covered by cubes lying totally inside of K' is at least

$$1 - \eta\sqrt{d}\,\partial K'$$

where $\partial K'$ is the surface of K'. This gives the following estimate for the utilization, i.e. to the total volume of the A_i's:

$$\sigma \cdot \frac{1 - \eta\sqrt{d}\,\partial K'}{\eta^d} \approx \frac{\sigma}{(\sigma^{1/d} + c_1\varepsilon d)^d}\left(1 - \sqrt{d}(\sigma^{1/d} + c_1\varepsilon d)\partial K'\right),$$

for some constant c_1, which, up to certain error terms, agrees with the upper bound.

As K' is a kind of "rounded" version of K, namely the bounding box of K is transformed into a cube, we expected the $\partial K'$ be smaller than or equal to ∂K. By the result of this paper, this is not the case even in dimension 2. It would be very interesting to know what happens in higher dimensions, especially whether $\partial K'/\partial K$ can be bounded from below by a constant depending on the dimension only.