HYPERBOLICITY OF GROUPS WITH SUBQUADRATIC ISOPERIMETRIC INEQUALITY

A. YU. OL'SHANSKII

Department of Mathematics and Mechanics
Moscow University
Moscow 119899

Communicated by S. Gersten
Received 1 August 1991

AMS Mathematics Subject Classification: 20F06

The author gives a simple and coherent proof of Gromov's statement on hyperbolicity of groups with subquadratic isoperimetric inequality.

0. Consider an arbitrary finitely presented group

$$G = \langle a_1, \dots, a_k | r_1, \dots, r_l \rangle. \tag{1}$$

Define $G = F/N$, where N is the normal closure of the set of all relators r_i in the free group $F = F(a_1, \dots, a_k)$. Therefore for every word $w \in N$ (i.e., $w = 1$ in G) there is an equation

$$w = \prod_{i=1}^{m} s_i r_{j_i}^{\pm 1} s_i^{-1} \tag{2}$$

in F, where $s_i \in F$. Suppose the number $m = m(w)$ of factors in (2) is minimal for w. Since G is finitely generated, there is an integer value function f_G (depending on the presentation (1) of G), such that

$$m(w) \le f_G(|w|), \tag{3}$$

where $|w|$ is the length of the word $w \in N$.

If G is hyperbolic (sometimes the terminology "word-hyperbolic" or "negatively curved" is used), then one can choose the function f_G linear ([1], 2.3.A). In this case we say that (3) is a linear "isoperimetric inequality". For every automatic group G the function f_G is quadratic [2]. Therefore it is natural to ask what the structure is of a group for which the function f_G's subquadratic. In this case Gromov has an argument that the group G is hyperbolic ([1], 2.3.F). The argument is scattered throughout the paper, which makes reading it difficult. Here we present a simpler, shorter and connected proof for this important result.

International Journal of Algebra and Computation Vol. 1, No. 3 (1991) 281–289

Theorem 1. *Suppose $f_G(x)/x^2 \to 0$ as $x \to \infty$. Then the group G is hyperbolic.*

1. Let us begin with the definition of hyperbolicity connected with the "thickness" of geodesic triangles in the *Cayley graph* $\Gamma = C(G)$ of a group G.

Recall that the set of vertices of Γ coincides with G and vertices g and $h = ga_i^{\pm 1}$ are connected by an edge $e = (g, a_i^{\pm 1})$ with the label $\phi(e) = a_i^{\pm 1}$.

Let us endow each edge e with the metric of the unit segment $[0; 1]$, the edges (g, a_i) and (h, a_i) are to coincide as sets, if $a_i^2 = 1$ in G; and define the geodesic metric on Γ, extending metrics of all edges. (Recall that a metric space X is *geodesic* if for any pair of points $A, B \in X$ there is an isometry $\imath$ on the segment $l = [0, \rho(A, B)]$ into X, such that $\imath(0) = A$ and $\imath(\rho(A, B)) = B$. In this case $\imath(l)$ is called a *geodesic path* in X.)

By definition, an *n-gon* P in a geodesic space X is a closed broken line $p_1 p_2 \dots p_n$, where each side p_i is a geodesic path in X. The *initial point* $(p_i)_-$ and the *terminal* point $(p_i)_+$ of each path p_i are vertices of P. If $A, B, C, \dots$ are vertices of an n-gon P and sides AB, BC, $\dots$ are fixed as particular geodesic paths, then we shall write $P = ABC \dots$. The lengths of sides will be denoted by $|p_1|$, $|p_2|$, $\dots$ or $|AB|$, $|BC|$, $\dots$.

The distance function ρ is continuous, and so for every triangle $T = ABC$ there exists a point O on the side BC, such that $\rho(O, AC) = \rho(O, BC)$. Call it a *bisector* and denote $b(O) = \rho(O, AC)$. Define the *bisize* of the triangle T to be the maximum of values $b(O)$ for all bisector points O, chosen on all sides of T.

Call a group G *hyperbolic* if there exists a constant $b \geq 0$, such that *bisize* $T \leq b$ for every triangle T in the graph $C(G)$. This definition of hyperbolicity implies the standard one, using the δ-version of the ultrametric triangle inequality (because of an obvious inequality *minsize* $T \leq 2$ *bisize* T and statement 6.6.B of [1], or Proposition 21 of [3]). Thus we can use above definition at the proof of Theorem 1.

2. Define the *thickness* of an n-gon $P = p_1 \dots p_n$ in a geodesic space X as the minimal number $t = t(P)$, satisfying the following condition: for every side p_i of P and each point $O \in p_i$ there exists another side p_j (i.e., $i \neq j$), such that $\rho(O, p_j) \leq t$.

We need also an auxiliary notion of d-wide triangle. Call a triangle T *d-wide* if for some suitable labeling of its vertices

i) there exists a point $O \in BC$, such that $\rho(O, AC) \geq d$;

ii) $|AB| \leq \frac{1}{2}\rho(O, AC)$.

Lemmas 1 and 2 can be considered as exercises on the triangle inequality.

Lemma 1. *Suppose there exists a d-wide triangle in a geodesic space X. Then there exists a quadrangle P in X, such that $t = t(P) \geq d/2$ and the perimeter of P is at most $20t$.*

Proof. Choose a point O according to the above definition of a d-wide triangle ABC (Fig. 1). If necessary, d can be increased, so that one can assume that

$$d = \rho(O, AC) \geq \rho(O', AC) \tag{4}$$

for every point $O' \in BC$.

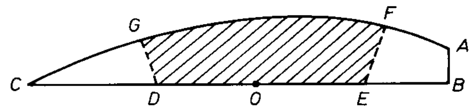

Fig. 1

Let D be the point on the segment OC such that $|OD| = 3d/2$. (If $|OC| \leq 3d/2$, then set $D = C$ by definition.) Similarly, find $E \in OB$, such that $|OE| = 3d/2$ or $E = B$.

By the choice of $O \in BC$, there are points G and F on AC, such that $|DG| \leq d$ and $|EF| \leq d$. One can put $G = C$ (or $F = A$) if $D = C$ (if $E = B$).

Note, that $\rho(O, AB) \geq |OA| - |AB| \geq d - d/2 = d/2$. In case $E \neq B$, $\rho(O, EF) \geq |OE| - |EF| \geq 3d/2 - d = d/2$. Thus, in any case $\rho(O, EF) \geq d/2$. Similarly $\rho(O, DG) \geq d/2$, and so the thickness of the quadrangle $DEFG$ is at least $d/2$, as $\rho(O, FG) \geq d$ by (4).

Finally, $|GF| \leq |GD| + |DE| + |EF| \leq d + 3d + d = 5d$, and so the perimeter of the quadrangle $DEFG$ is not greater than $10d \leq 20t$. □

Lemma 2. *Suppose a geodesic space X contains a triangle T, having bisize $T = b$, but X contains no $2b$-wide triangle. Then for some $t \geq b$ the space X contains a hexagon P, such that*

i) *the thickness $t = t(P) \geq b$;*

ii) *the perimeter of P is at most $46t$.*

Proof. According to the definition of bisize, one can choose a bisector point O on one side, say AB, of the triangle $T = ABC$ and points D and E on two other sides AC and BC (Fig. 2), such that

$$b = |OD| = \rho(O, AC) = |OE| = \rho(O, BC). \tag{5}$$

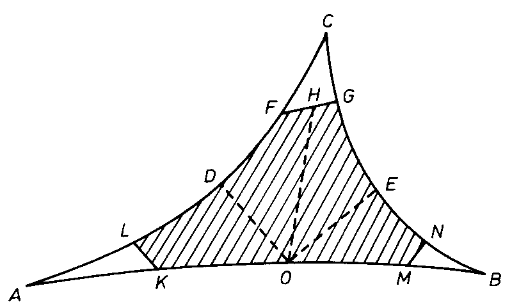

Fig. 2

Let F be the point on DC, such that $|DF| = 7b$. (Let $F = C$, if $|DC| \leq 7b$.) Note that

$\rho(F, CE) \leq 4b$, since

$$|DE| \leq |DO| + |OE| = 2b \tag{6}$$

and DCE cannot be a $4b$-wide triangle. (In fact, it cannot even be $2b$-wide by the condition of the lemma.) Hence the segment CE contains a point G, such that

$$|FG| \leq 4b. \tag{7}$$

The inequalities (6), (7) give us

$$|GE| \leq |GF| + |FD| + |DE| \leq 4b + 7b + 2b = 13b \tag{8}$$

and, on the other hand, if $|DF| = 7b$ (i.e., $F \neq C$), then

$$|GE| \geq |DF| - |FG| - |DE| \geq 7b - 4b - 2b = b. \tag{9}$$

Now we shall give an estimate for $|OH|$ for arbitrary $H \in FG$. On the one hand, $|AB| \leq |AD| + |DE| + |EB|$, i.e.,

$$|AB| \leq |AD| + |EB| + 2b \tag{10}$$

by (6). On the other hand,

$$(|AD| + |DF|) + (|BE| + |EG|) \leq (|AO| + |OH| + |HF|) + (|BO| + |OH| + |HG|),$$

i.e., if $F \neq C$, then in virtue of (9) and (7)

$$|AD| + |BE| + 7b + b \leq |AB| + 2|OH| + 4b. \tag{11}$$

The sum of (10) and (11) gives us

$$|OH| \geq b. \tag{12}$$

If $F = G = C$, then $H = C$, and the inequality (12) holds as well, since $|OC| \geq |OD| = b$.

Again, consider a construction choosing the point $K \in OA$, such that $|OK| = 3b$. (Let $K = A$, if $|OA| \leq 3b$.) Note, that $\rho(K, AD) \leq 2b$, as otherwise the triangle AOD would be $2b$-wide by (5). Hence, there is a point $L \in AD$, such that

$$|KL| \leq 2b. \tag{13}$$

In case $|OK| = 3b$ (i.e., $K \neq A$)

$$\rho(O, KL) \geq b \tag{14}$$

because the left term of (14) is not less than $|OK| - |KL| \geq 3b - 2b = b$ by (13). In the other case $L = K = A$, and (14) holds too, since $|OA| \geq |OD| = b$ by (5).

Similarly, segments OB and EB contain points M and N, such that

$$|OM| \leq 3b, \qquad |MN| \leq 2b, \qquad \rho(O, MN) \geq b. \tag{15}$$

The inequalities (14), (15), (12) and (5) mean that the distance from point O to the path $KLFGNM$ is equal to b, and so the thickness t of the hexagon $P = KLFGMN$ is not less than b.

To estimate the perimeter of P one can note that

$$|LD| \leq |LK| + |KO| + |OD| \leq 2b + 3b + b$$

by virtue of (13) and (5). Therefore

$$|LF| \leq 6b + 7b = 13b. \tag{16}$$

Similarly, $|EN| \leq 6b$ and by (8)

$$|GN| \leq 13b + 6b = 19b. \tag{17}$$

Finally, in view of the inequalities (15), (17), (7), (16) and (13),

$$|KM| + |MN| + |NG| + |GF| + |FL| + |LK| \leq 6b + 2b + 19b + 4b + 13b + 2b$$

$$= 46b \leq 46t. \qquad \square$$

Lemma 3. *Suppose the bisizes of triangles in a geodesic space X are unbounded. Then for any $t_0 > 0$ there exists a hexagon P in X, such that its thickness $t > t_0$ and its perimeter is not greater than* $46t$.

Proof. If X contains d-wide triangles for any $d > 0$, then the statement follows from Lemma 1, because every quadrangle can be considered as a hexagon containing two segments of length 0. If, on the other hand, X does not contain d-wide triangles for some d (and for every $d' > d$) then Lemma 2 implies desired conclusion. $\square$

3. Let us recall the definition of van Kampen diagram.

A *map* is a finite, planar, connected and simply connected 2-complex. A *diagram* Δ *over an alphabet* $A = \{a_1^{\pm 1}, \ldots, a_k^{\pm 1}\}$ is a map, such that every edge e (i.e., a 1-cell) of Δ has a label $\phi(e) \in A$, such that $\phi(e^{-1}) = \phi(e)^{-1}$. The *label* $\phi(p)$ of a path $p = e_1 \ldots e_n$ in Δ is, by definition, the word $\phi(e_1) \ldots \phi(e_n)$. We call a diagram Δ over A a *diagram over the group* G (more precisely, over the presentation (1)), if the label of the boundary path of every *face* (i.e., a 2-cell) of Δ is a cyclic permutation of some relator $r_i^{\pm 1}$.

Let w be a word over the alphabet A. The *van Kampen Lemma* says that w vanishes in G if and only if there exists a diagram Δ over G, such that w is the label of the boundary path of Δ. (In particular, the length $|w|$ is equal to the perimeter $|\partial\Delta|$ of Δ.) We shall consider only van Kampen diagrams with minimal number of faces. As was shown in [4] (ch. 5), the number of factors $m = m(w)$ at the equality (2) coincides with the number of faces in the van Kampen diagram Δ with the boundary label w.

The *distance* $|OO'|$ between two vertices O and O' in a diagram Δ is, by definition, the number of edges in a shortest path q in Δ, connecting O and O' (i.e., the length of the word $\phi(q)$). We call the boundary path $p = p_1 \dots p_n$ of a diagram Δ an *n-gon*, if each subpath p_i is a geodesic path in Δ.

The number $t = t(\Delta)$ is called the *thickness* of a diagram Δ with the n-gon boundary path $p = p_1 \dots p_n$, if t is the minimal number satisfying the following condition: For every path p_i and each of its vertices O there exists a vertex O' on a subpath p_j, where $j \neq i$, such that the distance $|OO'|$ in Δ is not greater than t.

Consider the following mapping γ of the set of vertices and edges of a diagram Δ with a boundary path p into the Cayley graph $\Gamma = C(G)$. Let $\gamma(p_-)$ be the unity vertex in Γ; and for an arbitrary vertex O of Δ let $\gamma(O)$ be the element of G, represented by the word $\phi(q)$, where q is a path in Δ between the vertex p_- and O. In view of the van Kampen Lemma the element $\phi(q)$ depends on O only and does not depend on q, because Δ is a simply connected diagram. For an edge e of Δ, set $\bar{e} = \gamma(e)$ to be an edge in Γ, such that $\bar{e}_- = \gamma(e_-)$ and the label $\phi(\gamma(e))$ in Γ is equal to the label $\phi(e)$ in Δ. One can extend γ onto the set of paths in Δ; so it is clear, that $|\gamma(A)\gamma(B)| \leq |AB|$ for any pair of vertices A, B of Δ.

Lemma 4. *Suppose bisizes of triangles in the Cayley graph Γ of the group G, given by* (1), *are unbounded. Then for any $t_0 > 0$ there exists a polygon diagram over G, such that its thickness $t > t_0$ and its perimeter is not greater than* $47t$.

Proof. Lemma 3 provides us a hexagon $P = p_1 \dots p_6$ in Γ, such that $t(P) > t_0 + 3/2$ and $\sum_{i=1}^{6} |p_i| \leq 46t(P)$. Note that there exists a polygon $Q = q_1 \dots q_n$ in Γ, such that $\sum_{i=1}^{n} |q_i| \leq \sum_{i=1}^{6} |p_i|$, all vertices of Q belong to $Q \cap G$ and $t(Q) \geq t(P) - 1$.

In view of the van Kampen Lemma, there exists a diagram Δ over G with boundary path $q'_1 \dots q'_n$, such that $\gamma(q'_i) = q_i$ for the above mapping γ. This diagram has the polygon boundary path $q'_1 \dots q'_n$ because the paths $q_1, \dots, q_n$ are geodesic.

Note that $t = t(\Delta) \geq t(Q) - 1/2 > t_0$ as γ does not increase distances and the distance from each point of Q to the nearest vertex of $Q \cap G$ is at most 1/2. Finally,

$$|\partial\Delta| = \sum_{i=1}^{n} |q_i| \leq \sum_{i=1}^{6} |p_i| \leq 46t(P) \leq 46(t(Q) + 1) \leq 46(t(\Delta) + 3/2) < 47t,$$

since one can suppose that $t_0 \geq 69$. $\square$

4. The following lemma repeats the author's argument in [5], lemma 10.6a (it is also related to isoperimetric inequalities).

Lemma 5. *Let $t > 0$ be a thickness of a diagram Δ with n-gon boundary path $p = p_1 \dots p_n$ over the presentation* (1). *Then the number of faces m in Δ is greater than $4t^2/M^3$, where $M = \max(|r_1|, \dots, |r_l|)$.*

Proof. By the definition of thickness, one can find a boundary geodesic segment $p = p_i$ and a vertex $O \in p$, such that $|OA| \geq t$ for every vertex $A \in p_j$ if $j \neq i$ (Fig. 3).

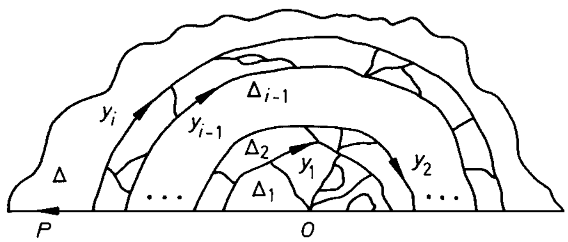

Fig. 3

Consider all faces whose boundaries contain the vertex O. We shall call them the *first level faces*. Certainly, both edges of p which contain the vertex O, belong to the boundaries of some first level faces. Let, by induction, the $(i-1)$-th level faces be defined for $i > 1$, and the union S_{i-1} of all closed faces of levels $\leq i-1$ is a connected set. ($S_0 = \{O\}$.) Define a subdiagram Δ_{i-1} of Δ such that

(i) Δ_{i-1} contains the set S_{i-1};

(ii) the boundary path z_{i-1} of Δ_{i-1} is contained in ∂S_{i-1};

(iii) z_{i-1} has the shortest length with respect to (i), (ii).

One can decompose $z_{i-1} = x_{i-1}y_{i-1}$, where x_{i-1} is the longest subpath of z_{i-1}, such that $O \in x_{i-1}$ and the initial vertex $(x_{i-1})_-$ and the terminal vertex $(x_{i-1})_+$ belong to p. Let p_{i-1} be the subpath of p such that $(p_{i-1})_- = (x_{i-1})_-$, $(p_{i-1})_+ = (x_{i-1})_+$. Suppose $|p_{i-1}| \geq 2(i-1)$ if $i - 1 < 2t/M$ (inductive hypothesis).

Then, by definition, the *i-th level* consists of all faces Π, such that $\Pi \notin \Delta_{i-1}$ and $\partial\Pi \cap y_{i-1} \neq \varnothing$. (So the set S_i is connected.)

It is clear that for every vertex $A \in y_{i-1}$ the distance $\rho(A, y_{i-2})$ is at most half of the perimeter of some face. Therefore $\rho(A, y_{i-2}) \leq M/2$, and, by induction, $|AO| \leq (i-1)M/2 < t$. Hence the path y_{i-1} contains no edges of $\partial\Delta$, by the choice of the vertex O.

If $p = up_{i-1}v$, where $|u|$, $|v| > 0$, then obviously the path p_i contains at least two edges from u and v, because $t > (i-1)M/2$, and so $|p_i| \geq |p_{i-1}| + 2 \geq 2i$. By the definition of the i-th level, the path y_i cannot contain any edge of the path y_{i-1} if $i - 1 < 2t/M$. Therefore the path y_i consists of edges belonging to boundaries of i-th level faces. As the set S_i is connected, the path y_i contains no loop by property (iii) for z_i. Hence the number m_i of i-th level faces is not less than $|y_i|/M$.

Recall that the path p is geodesic and so is its subpath p_i. Thus, $|y_i| \geq |p_i| \geq 2i$.

Consequently, $m_i \geq 2i/M$ and

$$m \geq \sum_{1 \leq i < [2t/M]+1} m_i \geq \sum_{1 \leq i < [2t/M]+1} 2i/M \geq \frac{1}{M}((2t/M)^2 + (2t/M)) > 4t^2/M^3. \quad \square$$

5. To complete the proof of Theorem 1, let us assume that the group G is not hyperbolic. This means that the bisizes of triangles in the Cayley graph $C(G)$ are unbounded. Then, by Lemma 4, for any $t_0 > 0$ there exists $t > t_0$ and a polygon diagram Δ over G, such that $t = t(\Delta)$ and $|\partial\Delta| \leq 47t$. The number of faces m in Δ is greater than $4t^2/M^3$ by Lemma 5. As we mentioned above (see Sec. 3), $m = m(w)$ for the word w written on the boundary of Δ. Therefore

$$m(w) \geq 4t^2/M^3 \geq \frac{4}{M^3}\left(\frac{|\partial\Delta|}{47}\right)^2 = 4|w|^2/47^2M^3.$$

However, this implies that

$$\overline{\lim_{x\to\infty}}\, f_G(x)/x^2 > c > 0,$$

where $c = c(M) = 1/600M^3$, because $|w| > t > t_0$. This contradicts the hypothesis of Theorem 1, and so the above assumption is false, and the theorem is proved.

Remark 1. One can weaken the condition of the theorem, as one could notice, demanding just the inequality $\overline{\lim}_{x\to\infty} f_G(x)/x^2 < c$ for some positive constant $c = c(M)$. Such a constant cannot be chosen independently on M. This could be demonstrated by means of the free abelian group $G = \langle a \rangle \times \langle b \rangle$. The group G is not hyperbolic, but considering the set of commutator relators $[a, b], [a^2, b], \ldots, [a^n, b]$ for sufficiently large n, one can achieve the inequality $\overline{\lim}_{x\to\infty} f_G(x)/x^2 < \varepsilon$ for any given $\varepsilon > 0$.

Remark 2. (S. Brick) Suppose the group G given by (1) is not hyperbolic. Then one can get a little more information from the proof of the theorem: For any $t_0 \geq 1$ there exists $t \in [t_0, 2t_0]$, such that the statement of Lemma 4 holds. Therefore one can find constants C_1, C_2 (depending on M only), such that for any $x \geq C_1$ there exists a word $w \in N$ with $|w| \in [x, C_2x]$ and $m(w) > |w|^2/600M^3$. Consequently, restricting ourselves to increasing Dehn functions f_G only, we have the inequality $f_G(x) \geq kx^2$ for suitable $k = k(M) > 0$.

Remark 3. If the Cayley graph $C(G)$ contains a triangle, whose bisize is equal to b, then $f_G(x) > cx^2$ for some $x \in [b, C_2b]$ (see Remark 2). Thus, given a hyperbolic group (1), one can effectively find a bound on the bisizes of the triangles in this group.

Remark 4. The author was asked by R. Alperin, what one can deduce from the assumption

$$\lim_{x\to\infty} f_G(x)/x = 0 \tag{18}$$

for the presentation (1). It is easy to prove that (18) implies $r_i = 1$ in the free group F

for all relators r_i, i.e., $G = F$. Indeed, otherwise, the normal closure N of all relators contains a cyclically reduced word $r = a_1 u a_1$ for some u. One can suppose that the number of generators $k > 1$ and consider words $w_j = (rs)^j rs^{-j} \in N$, where $s = a_2^n$ for $n > 2|r|$. Then w_j contains no subword v, such that $|v| > 3n$ and v^{-1} is a subword of w_j also. This means that the boundary path p of any van Kampen diagram Δ_j for w_j contains no subpath q, such that $|q| > 3n$ and q^{-1} is a subpath of p as well. This implies (we leave the proof for the reader) that the number of faces m_j in Δ_j increases linearly if $j \to \infty$, and so $\overline{\lim}_{x \to \infty} f_G(x)/x > 0$.

Acknowledgments

The author was invited by J. Stallings to visit UC Berkeley for a few months in 1991. The author thanks, for their discussion, all participants of Stallings's group theoretic seminar and S. Gersten, S. Goberstein, J. McCammond, S. Margolis, J. Rhodes and M. Sapir as well.

References

1. M. Gromov, *Hyperbolic groups*, Essays in Group Theory, ed. S. M. Gersten, MSRI series, vol. 8, Springer-Verlag, 1987, pp. 75–263.
2. D. B. A. Epstein, J. W. Cannon, D. F. Holt, M. S. Paterson, and W. P. Thurston, *Word processing and group theory* (to appear).
3. E. Ghys et P. de la Harpe, *Espaces metriques hyperboliques*, Sur les groupes hyperboliques d'apres Mikhael Gromov, Birkhauser, 1991, pp. 25–43.
4. R. Lyndon and P. Schupp, *Combinatorial Group Theory*, Springer-Verlag, 1977.
5. A. Yu. Ol'shanskii, *An infinite group with subgroups of prime orders*, (in Russian; English transl. in Math. USSR Izv. **16** (1981)), Izvestia Akad. Nauk SSSR, Ser. Mat. **43** (1980), 309–321.