On the nonexistence of dimension reduction for ℓ_2^2 metrics

Avner Magen* Mohammad Moharrami *

Abstract

An ℓ_2^2 metric is a metric ρ such that $\sqrt{\rho}$ can be embedded isometrically into $\mathbb{R}^d$ endowed with Euclidean norm, and the minimal possible d is the dimension associated with ρ. A dimension reduction of an ℓ_2^2 metric ρ is an embedding of ρ into another ℓ_2^2 metric μ so that distances in μ are similar to those in ρ and moreover, the dimension associated with μ is small. Much of the motivation in investigating dimension reductions in ℓ_2^2 comes from a result of Goemans which shows that if such metrics have good dimension reductions, then they embed well into ℓ_1 spaces. This in turn yields a rounding procedure to a host of semidefinite programming with good approximation guarantees.

In this work we show that there is no dimension reduction ℓ_2^2 metrics in the following strong sense: for every function $D(n)$ and for every n there exists an n point ℓ_2^2 metric ρ such that for all embeddings of ρ into an ℓ_2^2 metric μ with distortion at most $D(n)$, the associated dimension of μ is at least $n-1$. This stands in striking contrast to the Johnson Lindenstrauss lemma which provides a logarithmic dimension reduction for ℓ_2 metrics. Further, it shows that reducing dimension in ℓ_2^2 is even harder than doing so in ℓ_1 spaces.

1 Introduction

The theory of finite metric spaces has attracted a lot of attention from algorithm designers in recent years. In fact, many substantial steps in approximation algorithms were achieved using *embeddings* of one metric space into another and estimating the *distortion* of the embedding.

We quickly review the needed background. Let $f : (X, d) \to (X', d')$ be a mapping from metric space (X, d) into metric space (X', d'). The distortion of f is the minimum D such that $\alpha \cdot d(x, y) \le d'(f(x), f(y)) \le D \cdot \alpha d(x, y)$ holds for some $\alpha \ge 0$ and for any $x, y \in X$.

One of the most useful ways embedding results are applied is in the context of Linear Program and Semidefinite Programming relaxations for combinatorial problems. By viewing optimal solutions of such relaxations as finite metric spaces and then embedding these metric spaces with low distortion into ℓ_1 one effectively obtains a rounding procedure (see [13, 7, 15]), namely, a procedure that maps a set of vectors into a $\{0, 1\}$ assignment. The groundbreaking work of Arora, Rao and Vazirani [4] used this idea to provide an improved approximation of $O(\sqrt{\log n})$ to Sparsest Cut.

Metric spaces emerging from semidefinite relaxations can be typically described as follows. Consider a finite set of point in $\mathbb{R}^p$ endowed with the *square* Euclidean distance, that is, for vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^p$ the resulting distance function is $d_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|^2$. A distance function d obtained this is called an ℓ_2^2 distance functions. If in addition d satisfies triangle inequalities, we say that d is an ℓ_2^2 *metric*, or a *Negative Type Metric*. Notice that semidefinite relaxations may enforce such (linear) constraints.

Unlike ℓ_p metrics, i.e., metrics that embed in ℓ_p space with no distortion, the class of ℓ_2^2 metrics does not inherit the structure of a host normed space. This, to a great extent, explains why analyzing such metrics proved to be notoriously hard. The aforementioned result of Arora et al. [4], while not directly about metric spaces, shows that ℓ_2^2 metrics are well embeddable into ℓ_1 and ℓ_2 in some appropriately defined average sense. The result was later extended to show that every ℓ_2^2 metric is embeddable into ℓ_1 with distortion $O(\sqrt{\log n} \cdot \log \log n)$ [3]. Finding the smallest distortion needed to embed such metrics in ℓ_1 has become an

*Department of Computer Science, University of Toronto

intriguing open question, attracting attention from both geometers as well as complexity theorists. The best lower bound known so far is due to Khot and Vishnoi [10] that show that $\Omega(\log\log n)$ distortion is required.

Another theme of interest in the theory of metric spaces is dimension reduction: to what extent can the dimension associated with a metric be reduced without changing the distances by much? Such reductions are well understood in Euclidean space. While representing the metric of n points in Euclidean space isometrically requires dimension $n-1$, much less is sufficient for near isometries. In a seminal paper [9], Johnson and Lindenstrauss show that every n-point set in ℓ_2 can be embedded into $O(\log n)$-dimensional Euclidean space with constant distortion. Alon [2] recently showed that this dimension is essentially the best possible. Exploring the possibility of a similar phenomenon in ℓ_1 spaces, Brinkman and Charikar [5] and later Lee and Naor [12] showed that there is no dimension reduction in these spaces: they exhibited an n point metric in ℓ_1 such that embedding it with constant distortion in m-dimensional ℓ_1 space is only possible for $m = n^{\Omega(1)}$.

The following result due to M. Goemans presented in a workshop on methods in discrete mathematics [6] relates the themes distortion and dimension, with respect to ℓ_2^2 metrics. As such, it suggested an alternative avenue to low-distortion embeddings for such metrics. First, we need a slight restriction in the definition of the metrics spaces in question.

Definition 1 *A distance function on n points is called NEG$_{sym}$ if there are vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^p$ such that $d_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|^2$, and $\|x-y\|^2 + \|z-y\|^2 \geq \|x-z\|^2$ for all $x, y, z \in \{\pm \mathbf{v}_i\}_i$. The smallest possible p above is the dimension associated with d.*

Notice that $\mathrm{NEG}_{\mathrm{sym}}$ metrics are special cases of ℓ_2^2 metrics. We also observe that since $\|x-y\|^2 + \|z-y\|^2 - \|x-z\|^2 = 2(x-y)\cdot(z-y)$, the condition above says that $(x-y)\cdot(z-y) \geq 0$, i.e., no three points among $\{\pm \mathbf{v}_i\}_i$ span an obtuse angle. Similarly to ℓ_2^2 metrics, it is possible to optimize a linear objective functions in the distances over $\mathrm{NEG}_{\mathrm{sym}}$ metrics.

Theorem 1 (Goemans, 2000) *Every NEG$_{sym}$ metric on p dimension can be embedded into ℓ_2 with distortion $O(\sqrt{p})$.*

Notice that if one could get a logarithmic dimension reduction for $\mathrm{NEG}_{\mathrm{sym}}$ metrics á la Johnson-Lindenstrauss, then Theorem 1 would imply that such metrics are embeddable into ℓ_2 with distortion $O(\sqrt{\log n})$. In fact, it follows from [14] that applying Johnson Lindenstrauss lemma for an ℓ_2^2 metric would result in a low-dimensional ℓ_2^2 metric that cannot violate triangle inequality by a large margin. So it would seem reasonable to expect that such dimension reductions are possible. An $O(\sqrt{\log n})$ distortion achieved this way would improve the results of Arora, Lee and Naor [3], and would greatly simplify [4, 3], being based on purely geometrical principles rather than combinatorial and geometrical ones. This question of whether such dimension reductions exist was raised in a workshop on metric geometry at Texas A&M, Summer 2006 [1].

In this work we show that there is no dimension reduction in ℓ_2^2 (or even for $\mathrm{NEG}_{\mathrm{sym}}$) in a strong sense: whenever the distortion depends only on the number of points, one cannot reduce the dimension below the trivial $n-1$. Specifically, we show

Theorem 2 *For any real function $D(n)$, there exists an n-point metric space X in ℓ_2^2 (or NEG$_{sym}$) such that for every metric space Y in ℓ_2^2 that is associated with less than $n-1$ dimension, the distortion required for embedding X into Y is greater than $D(n)$.*

2 The construction

Recall that if ρ is an ℓ_2^2 metric then $\sqrt{\rho}$ is a metric that is obtained by taking points in Euclidean space where no three of them spans an obtuse angle. Therefore, dimensionality reduction in ℓ_2^2 metrics amounts to dimensionality reduction of a set of points in Euclidean space that span no obtuse angle, with the additional requirement that the image of the points do not span such angles as well.

The previous results on the minimum required dimension of a metric were due to Brinkman and Charikar [5] and to Lee and Naor [12] in the ℓ_1 case. An obvious first attempt would be to use random projections as in the Johnson Lindenstrauss lemma. Doing so certainly preserve distances approximately, and in fact allows for only small changes in angles: the sine of the angles change by an arbitrary small factor [14]; but that is not strong enough when the angles in question are close to $\pi/2$. Indeed, it is easy to see that under a random projection a right angle will become obtuse with probability 1/2. In particular, the Johnson-Lindenstrauss Lemma itself is not a good approach to the question of dimension reduction for ℓ_2^2 metrics. With this in mind, it is not surprising that the bad example we exhibit contains many right angles.

Let $c > 1$ be a constant, let $\mathbf{p_0} = -\frac{1}{2}\sum_{i=1}^{n} c^{i-1}\mathbf{e_i}$, and recursively define

$$\mathbf{p_j} = \mathbf{p_{j-1}} + c^{j-1}\mathbf{e_j}.$$

Let $X(n,c)$ be the $(n+1)$-point metric space given by the squared Euclidean distances between the points $\mathbf{p_i}$. Notice that $X(n,c)$ is simply the line metric on points spaced at intervals with lengths that increase exponentially (as power of c). Further, since all the $\mathbf{p_i}$s lie on vertices of a box centred at the origin, no three points among $\pm\mathbf{p_i}$ span an obtuse angle. Therefore $X(n,c)$ is a $\mathrm{NEG_{sym}}$ metric. We will show that there is a large enough constant c depending on D, such that in every embedding of $X(n,c)$ with distortion $\leq \sqrt{D(n)}$ into ℓ_2 the vectors $f(\mathbf{p_i}) - f(\mathbf{p_{i-1}})$ are arbitrarily close to being orthogonal. Since a set of vectors that are sufficiently close to being orthogonal must have full dimension, the dimension lower bound then follows.

Let f be an embedding from $\{\mathbf{p_0}, \mathbf{p_1}, \ldots, \mathbf{p_n}\}$ into $\mathbb{R}^d$, and let α be a positive number such that for every i, j, k

$$\|\mathbf{p_i} - \mathbf{p_j}\| \leq \|f(\mathbf{p_i}) - f(\mathbf{p_j})\| \leq \alpha\|\mathbf{p_i} - \mathbf{p_j}\|,$$

and

$$(f(\mathbf{p_i}) - f(\mathbf{p_j})) \cdot (f(\mathbf{p_k}) - f(\mathbf{p_j})) \geq 0.$$

Denote the vector $f(\mathbf{p_i}) - f(\mathbf{p_{i-1}})$ by $\mathbf{w_i}$. Much of the argument we need is captured by the following lemma bounding the angles between the vectors $\mathbf{w_i}$s.

Lemma 1 *For $i \neq j$, $|\mathbf{w_i} \cdot \mathbf{w_j}| \leq \frac{\alpha}{c-1}\|\mathbf{w_i}\|\|\mathbf{w_j}\|$*

Proof: Assume without loss of generality that $i < j$. Focusing on the points $\mathbf{p_{i-1}}$, $\mathbf{p_i}$ and $\mathbf{p_j}$ we get that

$$0 \leq (f(\mathbf{p_{i-1}}) - f(\mathbf{p_i})) \cdot (f(\mathbf{p_j}) - f(\mathbf{p_i})) = -\mathbf{w_i} \cdot \Big(\sum_{k=i+1}^{j} \mathbf{w_k}\Big),$$

therefore

$$\mathbf{w_i} \cdot \mathbf{w_j} \leq -\mathbf{w_i} \cdot \Big(\sum_{k=i+1}^{j-1} \mathbf{w_k}\Big).$$

The distortion condition implies that $\|\mathbf{w_j}\| \geq c^j$ and that $\|\mathbf{w_k}\| \leq \alpha c^k$ for every k. Therefore

$$\mathbf{w_i} \cdot \mathbf{w_j} \leq -\mathbf{w_i} \cdot \Big(\sum_{k=i+1}^{j-1} \mathbf{w_k}\Big) \leq \|\mathbf{w_i}\| \sum_{k=i+1}^{j-1} \alpha c^{k-1} \leq \alpha\|\mathbf{w_i}\|c^j/(c-1) \leq \alpha\|\mathbf{w_i}\|\|\mathbf{w_j}\|/(c-1). \tag{1}$$

To lower bound $\mathbf{w_i} \cdot \mathbf{w_j}$ we consider the angle between the same three points, with $i-1$ as the center point. We get

$$0 \leq (f(\mathbf{p_i}) - f(\mathbf{p_{i-1}})) \cdot (f(\mathbf{p_j}) - f(\mathbf{p_{i-1}})) = \mathbf{w_i} \cdot \Big(\sum_{\mathbf{k=i}}^{\mathbf{j}} \mathbf{w_k}\Big).$$

Now we have

$$\mathbf{w_i} \cdot \mathbf{w_j} \geq -\mathbf{w_i} \cdot \Big(\sum_{k=i}^{j-1} \mathbf{w_k}\Big)$$

and similarly to (1)

$$\mathbf{w_i} \cdot \mathbf{w_j} \geq -\mathbf{w_i} \cdot \Big(\sum_{k=i}^{j-1} \mathbf{w_k}\Big) \geq -\|\mathbf{w_i}\| \sum_{k=i}^{j-1} \alpha c^{k-1} \geq -\alpha\|\mathbf{w_i}\| c^j/(c-1) \geq -\alpha\|\mathbf{w_i}\|\|\mathbf{w_j}\|/(c-1).$$

We refer the reader to Figure 1 which illustrates the geometrical intuition of the lemma. □

Given any function $D(n)$, we set $c = n\sqrt{D(n)} + 1$. Assume that $X(n, c)$ can be embedded into an ℓ_2^2 metric with dimension less than n and distortion at most $D(n)$. Then there must be a function f from $\{\mathbf{p_0}, \ldots, \mathbf{p_n}\}$ that satisfies the conditions of Lemma 1 with $\alpha = \sqrt{D(n)}$ and the vectors $\mathbf{w_i}$s are in $\mathbb{R}^d$ with $d < n$. Let $\mathbf{w_i}' = \mathbf{w_i}/\|\mathbf{w_i}\|$ be the normalized vector $\mathbf{w_i}$. Then

$$|\mathbf{w_i}' \cdot \mathbf{w_j}'| = |\mathbf{w_i} \cdot \mathbf{w_j}|/\|\mathbf{w_i}\|\|\mathbf{w_j}\| \leq \frac{\alpha}{c-1} = \frac{\alpha}{n\sqrt{D(n)}} = 1/n.$$

But now we have a set of n unit vectors which are almost orthogonal. It is a well known fact that such a set must have full rank; for completeness we show it here. Let A be the gram matrix of $\mathbf{w_i}$'s that is $A_{i,j} = \mathbf{w_i}'\mathbf{w_j}'$. Then $A_{ii} = 1$ and $|A_{i,j}| \leq 1/n$ for $i \neq j$. Thus, $\|A - I\|_\infty < \frac{1}{n}$, and for any vector $\mathbf{x} \neq \mathbf{0}$,

$$\|A\mathbf{x}\|_\infty \geq \|\mathbf{x}\|_\infty - \|(A - I)\mathbf{x}\|_\infty \geq \|\mathbf{x}\|_\infty - (n-1)(1/n)\|\mathbf{x}\|_\infty > 0.$$

So A is nonsingular and therefore the dimension spanned by the $\mathbf{w_i}'$ (and so by their original counterparts $\mathbf{w_i}$) is n. We have therefore shown that

$$d \geq n = |X(n, c)| - 1$$

which proves Theorem 2.

3 Discussion

We have shown that dimension reduction in ℓ_2^2 metrics, as well as in NEG_{sym} metrics, is impossible in general. An interesting question the remains open is whether a relaxed notion of dimension reduction may still hold. By that we mean that the distortion of the mapping is required to be small on average. More precisely, assume without loss of generality that we deal with a mapping f that does not expand distances. Then the average distortion of f is the average distances divided by the average distances in the image of the f (see [16]). Such a notion seems especially relevant here: if there are dimension reductions in this sense, then applying Goeman's theorem we would still get low (average) distortion embeddings of ℓ_2^2 metrics into ℓ_2. This would still be enough in order to achieve a good approximation for Sparsest Cut in the uniform case (the case that is dealt with in [4]).

We currently cannot preclude the possibility of dimension reduction on average, even with as little as $O(\log n)$ dimensions. To demonstrate how different this question is from the more standard one we answer above, notice that for the example we supplied, one dimension suffices in order to achieve small average distortion. Indeed, since all distances are dominated by distances to $\mathbf{p_{n+1}}$, it is enough to map $\mathbf{p_{n+1}}$ to the origin, and all other points to a single point of distance c^{n-1} from it. Instead, we suggest considering the usual line metric with interval of equal length in order to achieve the stronger lower bound that will apply to dimension reductions on average. Interestingly, this gives rise to questions about the trace of a positive semidefinite matrix that satisfies certain linear conditions on some of its minors. Such a setting is somewhat similar in flavor to the setting of Alon's lower bound for dimension reductions in ℓ_2 [2].

Acknowledgment

We thank Michel Goemans for allowing us to include the proof of his theorem in this paper.

References

[1] Open problems raised at the concentration week on metric geometry and geometric embeddings of discrete metric spaces. http://www.math.tamu.edu/research/workshops/linanalysis/problems.html, 2006.

[2] N. Alon. Problems and results in extremal combinatorics, part i. *Discrete Math.*, 273:31–53, 2003.

[3] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut [extended abstract]. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 553–562, New York, 2005. ACM.

[4] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231 (electronic), New York, 2004. ACM.

[5] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l1. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[6] M. Goemans. Strengthening of negative type metrics. In *New Methods in Discrete Mathematics*, Alpe d'Huez, March 2000.

[7] Piotr Indyk. Algorithmic aspects of geometric embeddings. In *FOCS*, volume 42, pages 10–33, 2001.

[8] F. John. *Collected Papers*. Number 39 (2). Birkhauser, Boston, Massachusetts, January 1985.

[9] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[10] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of The 46-th Annual Symposium on Foundations of Computer Science*, 2005.

[11] P. Kumar and E. A. Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21, July 2005.

[12] J. R. Lee and A. Naor. Embedding the diamond graph in lp and dimension reduction in l1. *Geometric & Functional Analysis GAFA*, 14(4):745–747, August 2004.

[13] N. Linial, E. London, and Yu. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.

[14] A. Magen. Dimensionality reductions in ℓ_2 that preserve volumes and distance to affine spaces. *Disc. Comput. Geom.*, 38(1):139–153, 2007.

[15] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[16] Yu. Rabinovich. On average distortion of embedding metrics into the line and into l1. In *STOC*, pages 456–462, 2003.

Appendix

A Embedding of the path

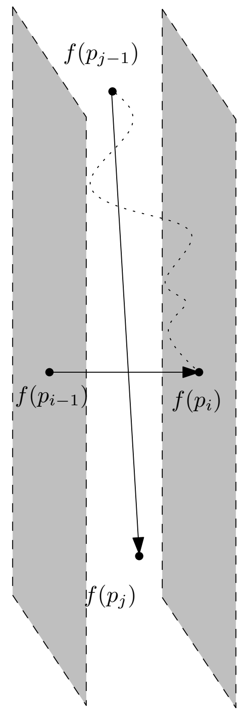

Figure 1: $f(\mathbf{p_{j-1}})$ and $f(\mathbf{p_j})$ must lie in the slab between $f(\mathbf{p_{i-1}})$ and $f(\mathbf{p_i})$; since $\mathbf{w_j} = f(\mathbf{p_j}) - f(\mathbf{p_{j-1}})$ has norm much larger than the width of the slab, $\mathbf{w_i}$ and $\mathbf{w_j}$ must be almost orthogonal.

B Goemans's theorem

In this section we present the proof of the theorem of Goemans. (the result was never published before.) We also discuss the connection of this Theorem and John's theorem[8] and the possibility of extension of the proof to subsets of ℓ_2^2 which are close to a metric.

Let d be a $\mathrm{NEG}_{\mathrm{sym}}$ metric in p dimensions and $S = \{\pm \mathbf{v}_i\}_i$, the set of vectors in the representation of d. Now, let $E = \{x | x^t Q x \leq 1\}$ be the minimum ellipsoid containing S and for any $x \in \mathbb{R}^p$, $g(x) = Q^{-1/2}x$. [1] Our goal is to show that

$$\frac{1}{2} \leq \frac{\|g(y) - g(z)\|}{\|y - z\|^2} \leq \frac{\sqrt{p}}{2}.$$

In particular, g has distortion $\sqrt{p}$.

To gain a geometric intuition for the construction, it is useful to consider the situation where S is a subset of the vertices of a box. In this case the embedding simply squares the lengths of edges a bound of $\sqrt{p}$ on the distortion for this case is readily implied. When one deals with a general S, applying $Q^{-1/2}$ is in some sense the best analogue to this squaring operation. The interesting part of the proof is to use the local condition on S in order to obtain a similar bound on the distortion.

[1] By $Q^{-1/2}$ we mean the Cholesky decomposition of Q^{-1}, namely the matrix M for which $Q^{-1} = MM^t$.

The volume of the ellipsoid E is $\det Q^{-1/2}\mathrm{Vol}(B(0,1))$ and so the optimum of the following is the matrix Q associated with the ellipsoid of minimal volume.

$$\min \log\det Q^{-1}$$

subject to:

$$x^tQx \leq 1 \text{ for all } x \in S$$

$$Q \succeq 0.$$

By the Lagrangian dual (see [11] for example) it follows that for the optimal solution $Q^{-1} = \sum_{x\in S}\lambda_x xx^t$ for some nonnegative real numbers λ_x, and moreover $\lambda_x = 0$ if $x^tQx < 1$.

For a vector $r \in \mathbb{R}^p$ we define a parameter

$$\eta(r) = \max_{y\in S}\frac{r\cdot y}{\|r\|}$$

that essentially looks at the extremes of projections of S according to a direction. We first show that η upper bounds the Lipschitz constant of g. According to the Lagrangian conditions

$$p = Tr(Q^{-1}Q) = \sum_{x\in S}\lambda_x Tr(xx^tQ) = \sum_{x\in S}\lambda_x x^tQx = \sum_{x\in S}\lambda_x$$

Now for an arbitrary vector a,

$$\|g(a)\|^2 = a^tQ^{-1}a = \sum_{x\in S}\lambda_x(x\cdot a)^2 \leq \Big(\sum_{x\in S}\lambda_x\Big)\max_{x\in S}(x\cdot a)^2 = p\cdot\max_{x\in S}(x\cdot a)^2 = p\cdot\eta^2(a)\|a\|^2,$$

hence

$$\frac{\|g(a)\|}{\|a\|} \leq \sqrt{p}\cdot\eta(a). \tag{2}$$

Next we show that on vectors of the form $y-z$, where $y,z\in S$, η is exactly half the norm. First we show that all vectors in S have the same norm. Indeed, let $x,w\in S$, then $\|w\|^2-\|x\|^2 = (w-x)(-w-x) \geq 0$. By symmetry $\|w\| = \|x\|$. Now, let $y,z\in S$, we have

$$y\cdot(y-z) = \|y\|^2 - y\cdot z = \|z\|^2 - y\cdot z = -z\cdot(y-z). \tag{3}$$

Let x be any point in S. Triangle inequality condition implies that

$$y(y-z) \geq x(y-z) \geq z(y-z). \tag{4}$$

Using Equations (3) and (4), we get

$$\max_{x\in S}|x\cdot(y-z)| = y\cdot(y-z) = -z(y-z),$$

and thus $\max_{x\in S}|x\cdot(y-z)| = \|y-z\|^2/2$ and we get

$$\eta(y-z) = \frac{\|y-z\|}{2}. \tag{5}$$

Combining Inequality (2) and Equation (5) it follows that

$$\frac{\|g(y)-g(z)\|}{\|(y-z)\|^2} = \frac{\|g(y-z)\|}{\|y-z\|}\cdot\frac{1}{2\eta(y-z)} \leq \sqrt{p}\cdot\eta(y-z)\cdot\frac{1}{2\eta(y-z)} \leq \frac{\sqrt{p}}{2}.$$

and the expansion of g is thus bounded.

Remark 1 The fact that all the angles are not obtuse is used in order to derive Equation (5) (and only there). Relaxing this condition and allowing some angles to be as large as $\pi/2 + \epsilon$ for some $\epsilon > 0$, one can show that $\eta(y - z)$ can be arbitrarily larger than $\|y - z\|$ and the bound above would not be achieved.

We now turn to analyze the contraction. For all vectors b

$$\|b\|^2 \leq \sqrt{b^t \cdot Q^{-1} b} \sqrt{b^t Q b} = \|g(b)\| \|g^{-1}(b)\|,$$

(using Cauchy Schwartz inequality) and setting b to be $y - z$, we get

$$\frac{\|g(y) - g(z)\|}{\|(y - z)\|^2} \geq 1/\|g^{-1}(y - z)\| = 1/\sqrt{(Q^{1/2}y - Q^{1/2}z) \cdot (Q^{1/2}y - Q^{1/2}z)} \geq 1/2.$$

To put thing in broader context we relate the above embedding to a well known result of John [8] that shows that every normed space $(\mathbb{R}^p, \|\cdot\|)$ embeds linearly into Euclidean space with distortion at most $\sqrt{p}$. This is achieved by taking the minimal ellipsoid containing the unit ball of the norm $\|\cdot\|$ and showing that the scaled down version (by $\sqrt{p}$) of this ellipsoid is contained in the unit ball. The linear embedding in John's theorem is the mapping that maps the containing ellipsoid to the unit ball. Here is the connection to our setting. Let R be the convex hull of S and let $R^* = \{y | x \cdot y \leq 1 \text{ for all } x \in R\}$. It can be shown that the embedding g in Theorem 1 is precisely the embedding of John's theorem when applied to (the norm whose unit ball is) R^*.