



HAL
open science

Towards Quotient Barycentric Subspaces

Anna Calissano, Elodie Maignant, Xavier Penneç

► **To cite this version:**

Anna Calissano, Elodie Maignant, Xavier Penneç. Towards Quotient Barycentric Subspaces. GSI 2023: Geometric Science of Information, Aug 2023, Saint-Malo (France), France. pp.366-374, 10.1007/978-3-031-38271-0_36 . hal-04162647

HAL Id: hal-04162647

<https://inria.hal.science/hal-04162647v1>

Submitted on 15 Jul 2023




HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Towards Quotient Barycentric Subspaces

Anna Calissano¹ *, Elodie Maignant¹ , and Xavier Pennec¹ 

Université Côte d'Azur and Inria, Epione team, Sophia-Antipolis, France
[anna.calissano, elodie.maignant, xavier.pennec]@inria.fr

Abstract. Barycentric Subspaces have been defined in the context of manifolds using the notion of exponential barycenters. In this work, we extend the definition to quotient spaces which are not necessary manifolds. We define an alignment map and an horizontal logarithmic map to introduce Quotient Barycentric Subspaces (QBS). Due to the discrete group action and the quotient structure, the characterization of the subspaces and the estimation of the projection of a point onto the subspace is far from trivial. We propose two algorithms towards the estimation of the QBS and we discussed the results, underling the possible next steps for a robust estimation and their application to different data types.

Keywords: Discrete Group · Quotient Space · Barycentric Subspaces Analysis · Graph Space · Object Oriented Data Analysis.

1 Introduction

Barycentric Subspace Analysis was introduced in [10]. Given a set of data points, it aims at estimating a set of subspaces of decreasing (or increasing) dimensions, which minimizes a loss function between the data and their projection onto the subspaces. In the specific context of manifold, the author defines barycentric subspaces using the Riemannian Exponential - referred as Exponential Barycentric Subspace (EBS). The derived EBS analysis is a promising dimensionality reduction technique for two main reasons. Firstly, it differs from tangent PCA [4, 1] as it goes beyond a 1-dimensional subspace search, proposing an optimization over a flag of subspaces. Such property is useful in the context of complex data such as graphs, shapes or images, where dimensionality reduction plays a central role in terms of reducing the data complexity and interpreting the results [3, 6, 5]. Secondly, BSA can also be defined as a "within data" statistics technique: the barycentric subspaces can be parametrized by data points. Such a choice allows to visualize and interpret the variability by looking at the data points which characterize the subspaces.

Complex data such as graphs or images are considered up to symmetries in many applications, such as node permutation for graphs or reflection for shapes. These type of data are usually embedded in quotient spaces obtained by applying a discrete group action, often resulting in a non manifold. In this geometric context, different dimensionality reduction techniques have been proposed [6, 1]. To the best of our knowledge the majority of the techniques available are the equivalent of the estimation of a

* Authors equally contributed to this submission and they are alphabetically ordered.

one-dimensional subspace at the time. Motivated by both high dimensional subspace search and interpretability through reference points, we propose a definition of Quotient Barycentric Subspaces (QBS). We underline where the complexity of the problem resides and we propose two algorithms for the estimation of QBS. We showcase the performance of the algorithm in a simple example of reflection action applied to \mathbb{R}^2 . To the best of our knowledge, there are no works in the literature addressing the problem of barycentric analysis for quotient spaces which are not manifolds. Note that if the quotient space is a manifold, all the manifold statistic literature is applicable [11, 8], including exponential barycentric subspace analysis [10].

2 Quotient Barycentric Subspaces

Consider $X = \mathbb{R}^m$ a Euclidean space and \mathcal{T} a discrete group acting on X . The orbit of a point $x \in X$ is the equivalence class $[x] = \{tx, t \in \mathcal{T}\}$. We are interested by the quotient space X/\mathcal{T} , made of the collection of these equivalence classes, when the action is not free. A group action is free if the only element of the group fixing a point is the identity [9]. If the action is not free, the resulting quotient space is not a manifold, but often only a stratified space. These spaces are very common, as non-free actions appear every time there are symmetries in the data.

Example 1. Consider $X = \mathbb{R}^2$ and the reflection action \mathcal{R} , giving the following equivalence relation: $(x_1, x_2) \sim (-x_1, -x_2)$. The resulting quotient space X/\mathcal{R} is not a manifold - the point $(0, 0)$ is fixed by the whole group \mathcal{R} .

Example 2. Graph Space [7, 2] is a quotient space used to study set of graphs with unlabelled nodes. It is obtained by applying permutation action to adjacency matrices. Consider a set of $n \times n$ undirected graphs represented as adjacency matrices $\{x_1, \dots, x_k\}, x_i \in X = \mathbb{R}^{n \times n}$. The space of adjacency matrices X is equipped with a Frobenious norm. If we consider the nodes to not be labelled, we can represent the unlabelled graphs as equivalence classes of permuted graphs $[x] = \{px, p \in \mathcal{P}\}$, where \mathcal{P} is the set of permutation matrices applied to the nodes, acting onto the matrices as $px = p^T xp$. Graph Space X/\mathcal{P} is a discrete quotient space, equipped with a quotient metric $d_{X/\mathcal{P}}([x], [y]) = \min_{p \in \mathcal{P}} d(x, py)$. As detailed in [3], Graph Space is not a manifold - as there are some permutations leaving the graph unchanged - making the extension of EBS not trivial.

2.1 Exponential Barycentric Subspaces

We first recall the definition of Exponential Barycentric Subspaces (EBS) introduced in [10]. Consider a Riemannian manifold \mathcal{M} equipped with a Riemannian metric on each tangent space $T_x \mathcal{M}$ and a logarithmic map $\log_x : \mathcal{M} \rightarrow T_x \mathcal{M}$.

Definition 1 (Exponential Barycentric Subspaces). *The Exponential Barycentric Subspaces (EBS) generated by the affinely independent reference points $(x_0, \dots, x_k) \in$*

\mathcal{M}^{k+1} is defined as

$$EBS(x_0, \dots, x_k) = \{x \in \mathcal{M}^*(x_0, \dots, x_k) \mid \exists w \in \mathbb{R}^{k+1} \setminus \{0\} : \sum_{i=0}^k w_i \log_x(x_i) = 0\}.$$

As detailed in [10], the definition is only valid for points x that are outside the cut locus of the reference points. In other words, EBS are defined in the tangent space $T_x\mathcal{M}$, where x belongs to EBS. This gives an implicit definition.

2.2 Quotient Barycentric Subspaces

To extend EBS to Graph Space and more generally to quotient spaces, we need to define an analogous to the logarithmic map used in the definition.

Definition 2. *The Alignment Map $a_x : X \rightarrow X$ is defined as $a_x(y) = t^*y$ where $t^* = \arg \min_{t \in T} d(x, ty)$ is the group element which optimally aligns the two points.*

The Alignment Map at a point x is only defined on $X \setminus E([x])$, where $E([x])$ is the (null measure) equidistant set [12], which consists of points which are equidistant to at least two elements of the orbit $[x]$.

Given the alignment map, we can introduce an equivalent of the horizontal logarithmic map in the context of quotient space [6, 9]:

Definition 3. *The Horizontal Logarithm $\log_x^H : X \setminus E(x) \rightarrow T_x X$ is the logarithmic map of the total space X applied to the optimally aligned points:*

$$\log_x^H(y) = \log_x(a_x(y))$$

where $\log_x : X \rightarrow T_x X$ is the logarithmic map of the total space.

As X is Euclidean, we can explicitly write the logarithmic map as $\log_x^H y = \log_x(a_x(t)) = a_x(y) - x$. Notice that due to the alignment map not being defined on the whole space, \log_x^H does not descend to a proper logarithm on the quotient space.

We can now extend the definition to the quotient space.

Definition 4. *Consider a set of reference orbits $\mathcal{X} = \{[x_1], \dots, [x_k]\} \in X/\mathcal{T}$. An orbit $[y] \in X/\mathcal{T}$ belongs to the Quotient Barycentric Subspace $[y] \in QBS(\mathcal{X})$ if $\exists w_i \in \mathbb{R}$ such that $\sum_{i=1}^k w_i = 1$ and*

$$\sum_{i=1}^k w_i \log_y^H(x_i) = 0.$$

The above definition is implicit and corresponds to solving the following system:

$$\begin{cases} \sum_{i=1}^k w_i a_y(x_i) = y, \\ a_y(x_i) = t_i x_i, \quad t_i = \arg \min_{t \in \mathcal{T}} d(tx_i, y). \end{cases}$$

More explicitly, $[y]$ belongs to $QBS(\mathcal{X})$ if and only if its representative y can be written as a combination of some representatives of $\{[x_1], \dots, [x_k]\}$ aligned with respect to $[y]$ itself. This results in a system of interlaced equations defining both the alignment and the barycentric combination. QBS are defined using the alignment map and the logarithmic map in the total space, which is Euclidean. This strategy allows for a computation of weights $w_i \in \mathbb{R}$ in the Euclidean Space X [10]. In the sequel of this paper, we focus only on positive weights to simplify the characterization of the subspaces.

The final goal of barycentric subspace analysis is to capture data variability within subspaces of a given dimension. We define a loss function using QBS Definition 4. Consider a set $\mathcal{Y} = \{[y_1], \dots, [y_m]\}$ of orbit in the quotient space X/\mathcal{T} and a set $\mathcal{X} = \{[x_1], \dots, [x_k]\} \subset \mathcal{Y}$ of k reference points selected among our data-points. We want to minimize the following function $\mathcal{L}(\mathcal{X}) = \sum_{i=1}^m d_{X/\mathcal{T}}^2(\hat{y}_i, [y_i])$ where $\hat{y}_i \in QBS(\mathcal{X})$ is a projection on the QBS spanned by \mathcal{X} .

2.3 Characterization of the Subspaces

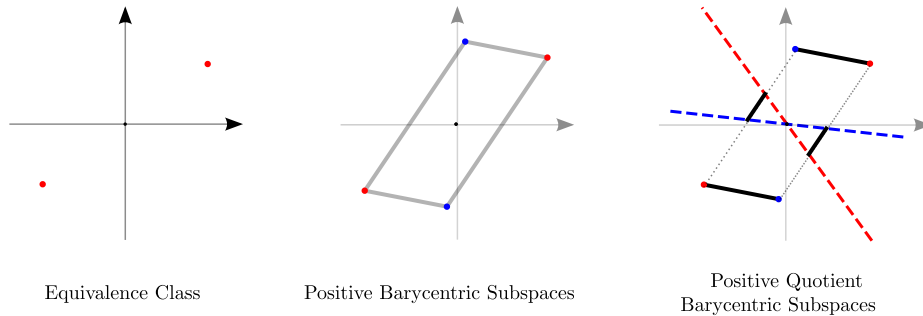


Fig. 1. Visualization of the equivalence classes in X/\mathcal{R} , the subspaces identified by two reference points, the quotient barycentric subspaces, where only the valid segments defined in Definition 4 are considered. The equidistant set of the red (resp. blue) point orbit is represented by the red (resp. blue) dashed line.

Given a set of reference points in X , the subspace identified by the orbits of the reference points results in a set of disjoint subspaces in X/\mathcal{T} . In Figure 1, we show the characterization of the QBS in X/\mathcal{R} described in Example 1. Given a data point, the resulting orbit in X/\mathcal{R} is represented as red dots in Figure 1 left. Consider two reference orbits (here the red and blue orbits), the possible positive barycentric subspaces are all the possible subspaces of dimension 1 (segments) joining two points of the different orbits - Figure 1 center. Such segments in the Euclidean space corresponds to barycentric combinations of reference points. Among such subspaces, not all points are valid

according to the QBS Definition 4. In Figure 1 right, the valid QBS are underlined in dark black, showing only the parts of the segments whose points are aligned with both reference points, as stated in Definition 4. Thus, the positive barycentric subspaces are turned into two disjoint valid segments: the *complete segments* if containing the reference orbits and the *incomplete segments* if not containing the reference orbits. The two red and blue dashed lines represent the boundary of the space where the points change alignments with respect to the reference points. In this example the identification of the valid – with respect to Definition 4 – parts of the subspaces is possible in close form due to the low dimension of both the space and the number of reference points. In more general setting, such close form description of the valid parts is not straightforward.

3 Algorithms for the Estimation of QBS

Consider a set of reference orbits $\{[x_1], \dots, [x_k]\}$ and an orbit $[y]$, the estimation of the projection \hat{y} of such orbit on the barycentric subspace is far from trivial. The complexity of the problem resides in the exploration of the search space. Firstly, we need to explore all the possible subspaces identified by $\{[x_1], \dots, [x_k]\}$: the number of all the possible subspaces identified by k orbits of cardinality c is c^k . Secondly, we need to find for each subspace the valid part, which has no closed form characterization – as explained in the previous section. Remember that the validity of such parts depends on $[y]$ which is itself an orbit of dimension c . We refer to these two aspects as the combinatorial and the geometrical complexity of the given problem. As a starting point in addressing such complexity, we opted for two different algorithms: *Align to Reference Points* - algorithm 1 - and *Align to Data Point* - algorithm 2.

Algorithm 1 Align to Reference Points

Require: data point $y \in [y]$ and reference points $\{[x_1], \dots, [x_k]\}$ **Return:** \hat{y}
 Select the closest reference point $[x_i]$ and its representative x_i optimally aligned with y
 Find $x_j \in [x_j]$ optimally aligned with x_i
 Find $y \in [y]$ optimally aligned with x_i
 Find \hat{y} , the projection of y onto $BS(x_1, \dots, x_k)$ ▷ Orthogonal Projection
if \hat{y} is not optimally aligned with x_1, \dots, x_k **then**
 Set $\hat{y} = x_i$, where x_i is the closest reference point to y ▷ Closest Reference Projection
end if

Algorithm 2 Align to Data Point

Require: data point $y \in [y]$ and reference points $\{[x_1], \dots, [x_k]\}$ **Return:** \hat{y}

Randomly select a representative $y \in [y]$

For all i , Find $x_i \in [x_i]$ optimally aligned with y

Find \hat{y} , the projection of y onto $BS(x_1, \dots, x_k)$ ▷ Orthogonal Projection

if \hat{y} is not optimally aligned with x_1, \dots, x_k **then**

Set $\hat{y} = x_i$, where x_i is the closest reference point to y ▷ Closest Reference Projection

end if

4 Experiments

For the two proposed algorithms, we consider the setting described in Example 1 and we restrict the analysis to positive weights. We run two experiments:

- Experiment 1: fixing the reference orbits $\{[x_1], \dots, [x_k]\}$ and randomly sampling the data orbit $[y^i]$, $i = 1, \dots, N$ to project
- Experiment 2: fixing the data orbit $[y]$ and randomly sampling the reference orbits $\{[x_1^i], \dots, [x_k^i]\}$, $i = 1, \dots, N$ onto which project the data orbit

Such experiments aim at grasping when the algorithm is able to project a data orbit onto a valid QBS, given different configurations of reference orbits and data orbits to project. In the reflection quotient space X/\mathcal{R} , we are able to compute the true projection as we can parameterize the quotient subspaces of dimension 1. For each algorithm, we can compute the projection error as the quotient distance between $[y]$ and the estimated $[\hat{y}]$. To evaluate the performance, we measure the absolute error $\hat{\epsilon}$ between the algorithm projection error and the true projection error. In Experiment 1, we consider two reference orbits and we sample a set of $N = 10000$ orbits to project. We run several Example of Experiment 1 to better characterize the performance of the two algorithms depending on the relative position of the two reference orbits $[x_1]$ and $[x_2]$.

Figure 3 and 2 illustrate two typical situations. In Figure 3, the reference points lay close to each other’s equidistant sets. From the two boxplots, algorithm 2 performs better than algorithm 1. By looking at the error distribution in the middle panel, algorithm 1 is under-performing when $[y]$ is around the incomplete segments. The alignment to reference point strategy is not able to project onto the incomplete segments - even when the data point is closer to such segments - causing an high error. On the other hand, algorithm 2 is correctly projecting on the incomplete segments when needed. The error of algorithm 2 is only concentrated around the equidistant sets, caused by the natural difficulties in choosing an optimal alignment. Figure 2 showcase a different situation when the incomplete pieces have a small length. By always projecting onto the other two pieces, algorithm 2 covers most situations quite similarly to algorithm 1 and therefore becomes comparable in performances. We understand from this first experiment that the performance of algorithm 1 strongly depends on the relative position of the two reference points. However, algorithm 2 is overall better than algorithm 1. Experiment 2 focuses on changing reference points ($N = 10000$ configurations randomly sampled) fixing the data point $[y]$. We compare the performance of the two algorithms using two boxplots of the error distribution (Figure 4). The performance of algorithm 2 is better than for algorithm 1 even when changing reference points.

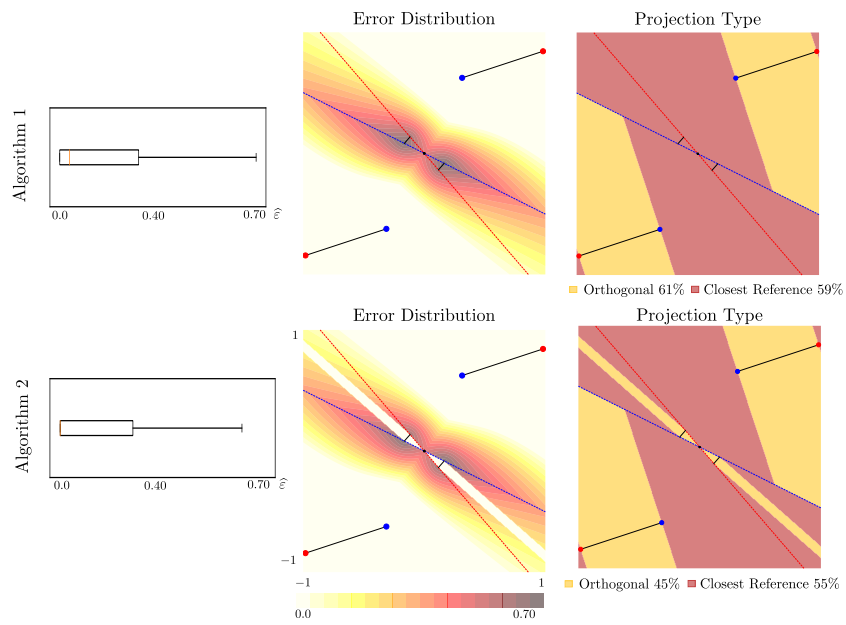


Fig. 2. Experiment 1: Example of two reference points resulting in QBS with small incomplete pieces. We plot the error distribution, the spatial distribution of the error and the spatial distribution of the type of projection. We can see that the error of Alg. 1 and Alg. 2 are similar. Alg. 2 is better performing in projecting onto the incomplete segments, which are short in this example.

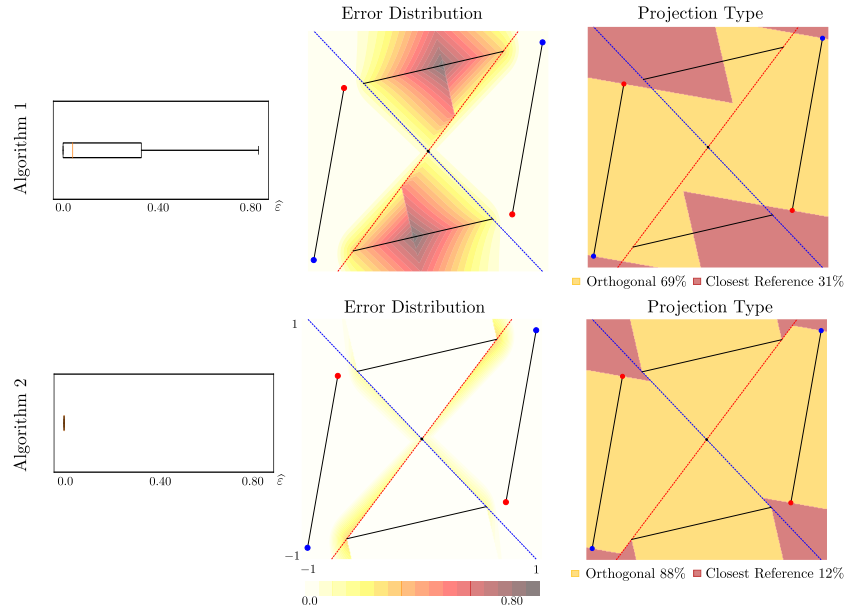


Fig. 3. Experiment 1: Example of two reference points close to each others equidistant sets. We plot the error distribution, the spatial distribution of the error and the spatial distribution of the type of projection. We can see that the error of Algo. 1 is much higher than Algo. 2. Such error is concentrated around the incomplete segments in Algo. 1 and the Equidistant sets in Algo. 2.

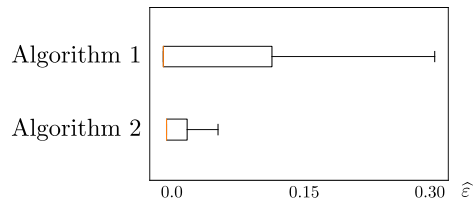


Fig. 4. Experiment 2: Comparison of the performances of algorithm 1 and algorithm 2. The boxplot are obtained by simulating a set of different reference points, keeping the data point fix. algorithm 2 is performing better.

5 Discussion

We address the extension of Exponential Barycentric Subspaces from manifolds to Quotient Spaces obtained by a discrete group action. We define Quotient Barycentric Subspaces (QBS) and we propose two different algorithms for the estimation of the projection of data-points onto QBS. The detailed analysis that we perform shows that the complexity of such projection is not only combinatorial, as could be expected, but also geometrical. The identification of the correct segment onto which project a data point depends on the reciprocal position of reference orbits, the position of the data point with respect to the segment and the orientation of the quotient orbit with respect to the subspace. As a first approach, we showed how algorithm 2 better addresses the identification of the correct subspace for projection. Further developments will include an alignment procedure tailored for subspaces rather than data points (e.g. reference points), which might allow the identification of all valid subspaces, including the incomplete segments.

Acknowledgements: We thank Gstats team for the fruitful discussions. This work was supported by the ERC grant #786854 G-Statistics from the European Research Council under the European Union’s Horizon 2020 research and innovation program and by the French government through the 3IA Côte d’Azur Investments ANR-19-P3IA-0002 managed by the National Research Agency.

Bibliography

- [1] Burcu Aydın, Gábor Pataki, Haonan Wang, Elizabeth Bullitt, James Stephen Maron, et al. A principal component analysis for trees. *The Annals of Applied Statistics*, 3(4):1597–1615, 2009.
- [2] Anna Calissano, Aasa Feragen, and Simone Vantini. Graph-valued regression: Prediction of unlabelled networks in a non-euclidean graph space. *Journal of Multivariate Analysis*, 190:104950, 2022.
- [3] Anna Calissano, Aasa Feragen, and Simone Vantini. Populations of unlabelled networks: Graph space geometry and generalized geodesic principal components. *Biometrika*, page asad024, 2023.
- [4] P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and mathematical methods in medical and biomedical image analysis*, pages 87–98. Springer, 2004.
- [5] Xiaoyang Guo, Anuj Srivastava, and Sudeep Sarkar. A quotient space formulation for generative statistical analysis of graphical data. *Journal of Mathematical Imaging and Vision*, 63:735–752, 2021.
- [6] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica*, 20(1):1–58, 2010.
- [7] B. J. Jain and K. Obermayer. Structure spaces. *Journal of Machine Learning Research*, 10(Nov):2667–2714, 2009.
- [8] D.G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

- [9] John M Lee. *Smooth manifolds*. Springer, 2013.
- [10] Xavier Pennec. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A):2711–2746, 2018.
- [11] Katie E Severn, Ian L Dryden, and Simon P Preston. Non-parametric regression for networks. *Stat*, 10(1):e373, 2021.
- [12] J. B. Wilker. Equidistant sets and their connectivity properties. *Proceedings of the American Mathematical Society*, 47(2):446–452, 1975.