

HyperMap:高次元空間における写像 アルゴリズムとその次元縮小、クラ スタリングへの応用

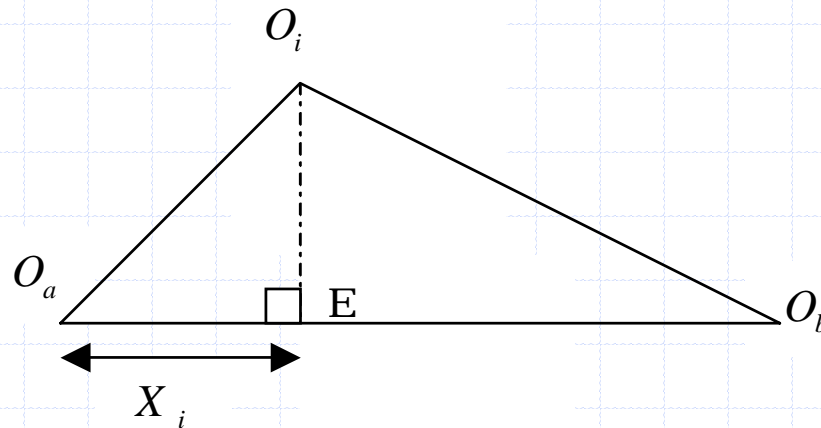
筑波大学 安際元

発表内容

- ⇒ ◆モチベーション
- ◆HyperMap射影法
- ◆クラスタリングへの応用
- ◆実験

FastMap射影法(Faloutsos, et. al. SIGMOD'95)

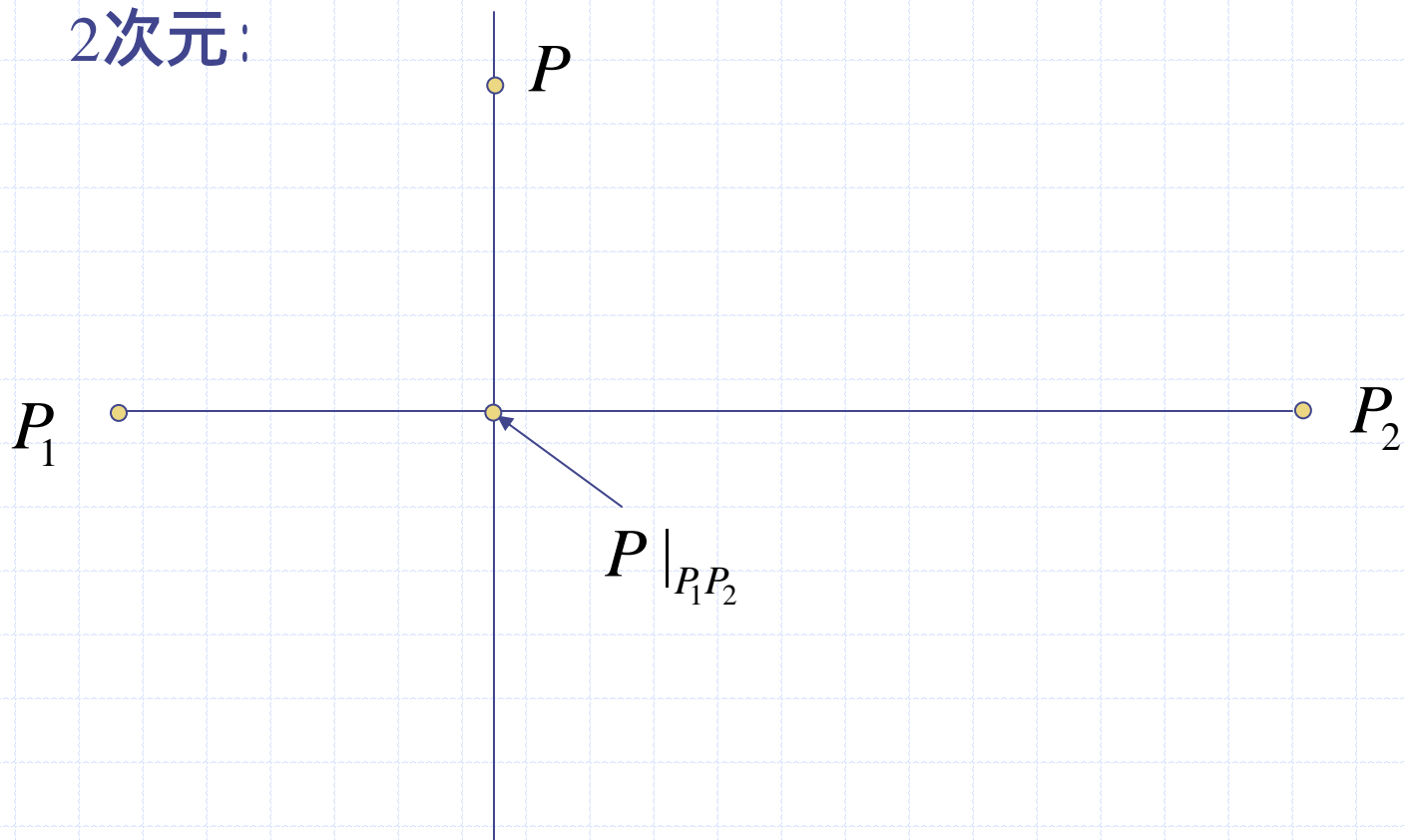
- 2個ずつピボットを選ぶことにより、射影空間が構成される
- データ座標値の求め方



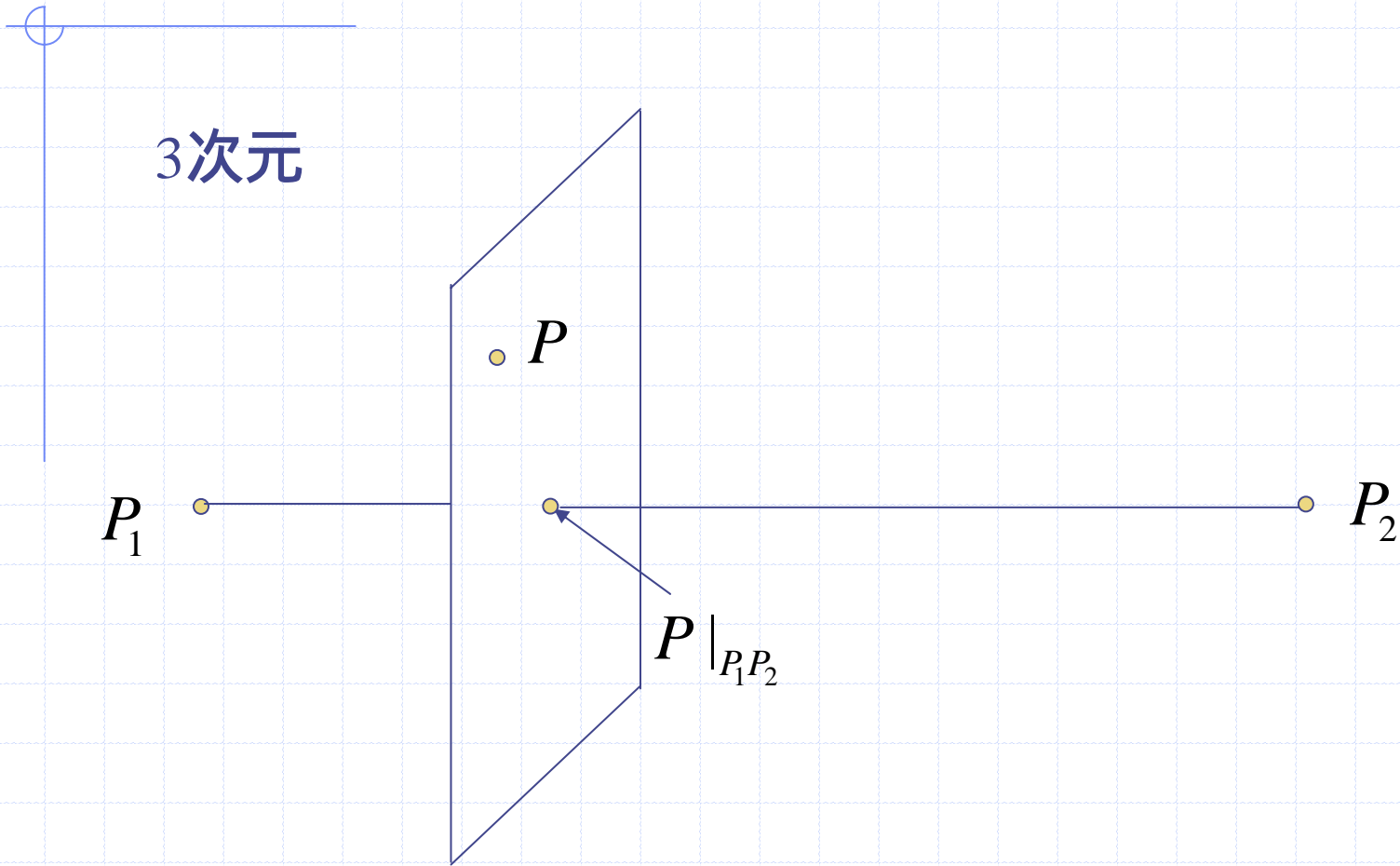
$$X_i = \frac{(D(O_a, O_b))^2 + (D(O_a, O_b))^2 - (D(O_b, O_i))^2}{2D(O_a, O_b)}$$

FastMap射影法(Faloutsos, et. al. SIGMOD'95)

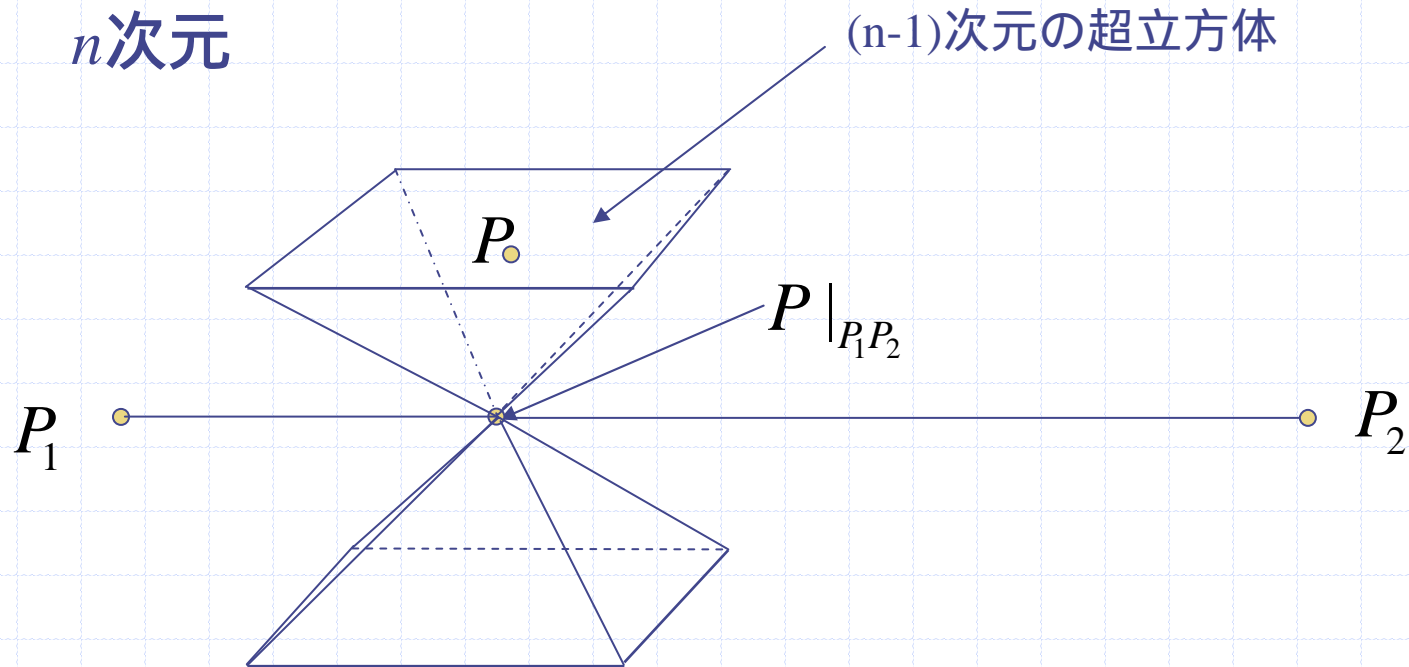
2次元:



FastMap射影法(Faloutsos, et. al. SIGMOD'95)



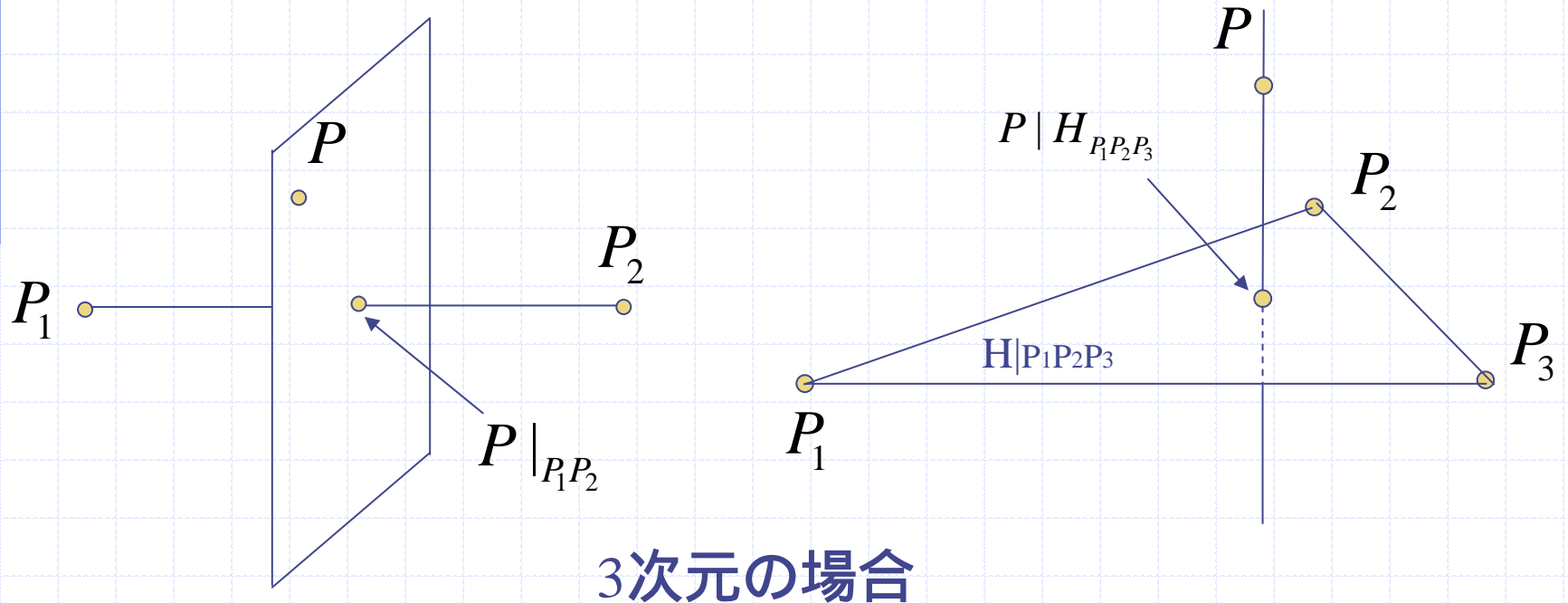
FastMap射影法(Faloutsos, et. al. SIGMOD'95)



結論:元の空間の次元数が高くなると、同じ位置に射影されるデータ数が指数的に増える。

提案

本研究はFastMapのピボット数が2個以上に拡張する



3次元の場合

$P_1P_2P_3$ から構成される平面 $H|_{P_1P_2P_3}$ の同じ位置に射影されるデータは線になり、同じ位置に射影されるデータの可能性が減少する。

発表内容

◆モチベーション

▶ ◆HyperMap射影法

◆クラスタリングへの応用

◆実験

HyperMap射影法

- ピボットの選び方

heuristicとしてもっとも遠く離れたデータをピボットとする。軸は線でなく超平面であるため、“超軸”と呼ぶ。

- 座標値の求め方

$P_1P_2\dots P_{n+1}$ から構成される n 次元の超軸に対し、 n 個の座標値がある。

発表内容

◆モチベーション

◆HyperMap射影法

⇒ ◆クラスタリングへの応用

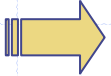
◆実験

クラスタリングへの応用

- クラスタリング手法におけるmedoidの選択は困難であり、ランダムに候補medoidを決めるのは一般的な手法である。(CLERA, CLARANS, PROCLUS, ORCLUS etc.)。
- クラスタのmedoidは離れている方がより合理的であるため、HyperMapのピボット選ぶ手法を適用した。

発表内容

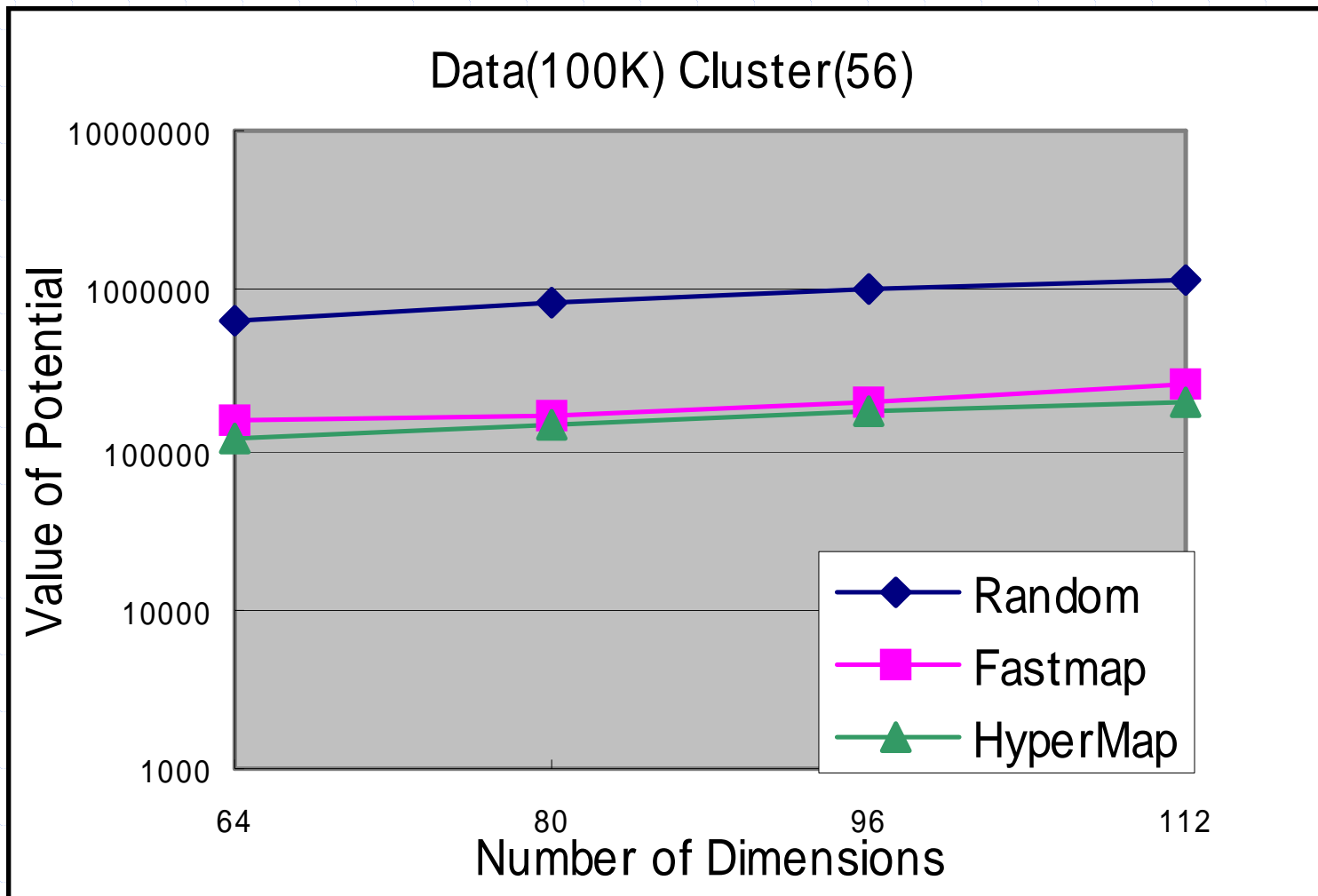
- ◆モチベーション
- ◆HyperMap射影法
- ◆クラスタリングへの応用
- ◆実験



実験

- ◆ 今回の実験はクラスタリングの評価関数としてポテンシャル(potential)を利用した。ランダム、FastMap、HyperMapの手法を適用した結果を示す。

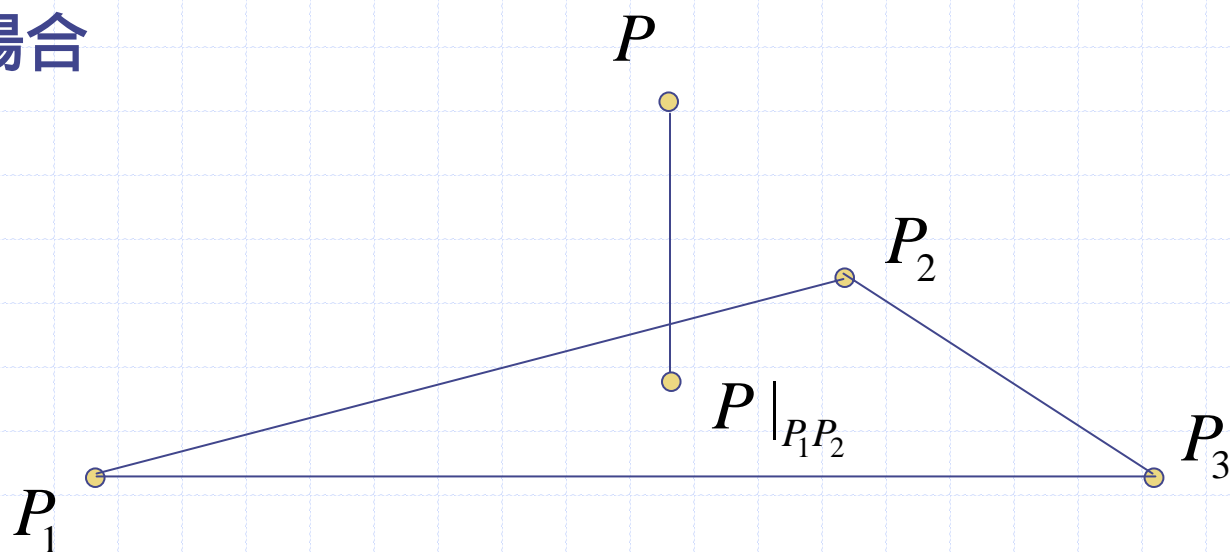
実験



ありがとうございます

提案

3次元の場合



$P_1P_2P_3$ から構成した平面 $H|_{P_1P_2P_3}$ に同じ位置に射影するデータは平面から線になり、同じ位置に射影するデータの可能性が減少する。