研究進捗報告書

ミーティング日： 2021 年 6 月 28 日　　　　　　学年 D2　　　氏名 吉田　皓太郎

注意：ミーティング時には、必ず本報告書を作成し、一部を教員に提出すると共に、一部を自分用に持参して下さい。本報告書の提出がない場合、ミーティングは実施しません。また、項目１）から項目３）について未記入の箇所がある場合にも、ミーティングは実施しません。なお、本報告書は手書きでも構いません。

MTG を行うにあたってのねらい

今回の MTG では，誤差要因に関する考察を三度考察し，まとめるものである．これを通したディスカッションにより，結果の妥当性などを確認していきたいと考えている．

今週のミーティング事項について

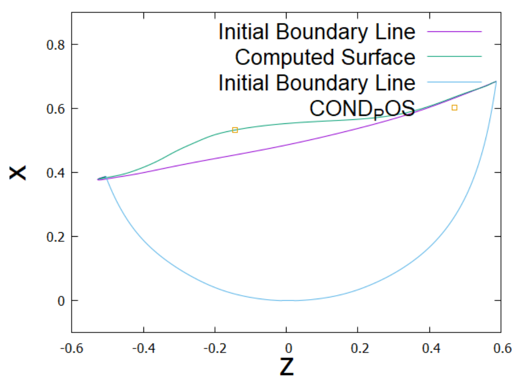
目次

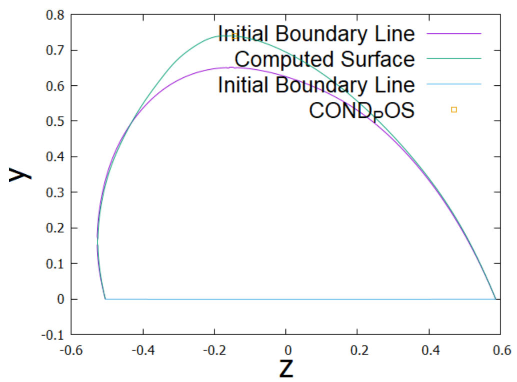
ミーティング事項の具体的な内容について

1　投稿論文の図について

現在は，2 つの円弧によって生成された曲面を対象に取り扱っていましたが，これを，ワコールから頂いた下カップの 3D データに置き換える方がよいのではと考えました．（要は簡単なデータではなく，実際にある曲面データを修正可能であると言えるため）

この場合，展開後の二次元形状を提出する際には，何か言われるかもしれませんが．．



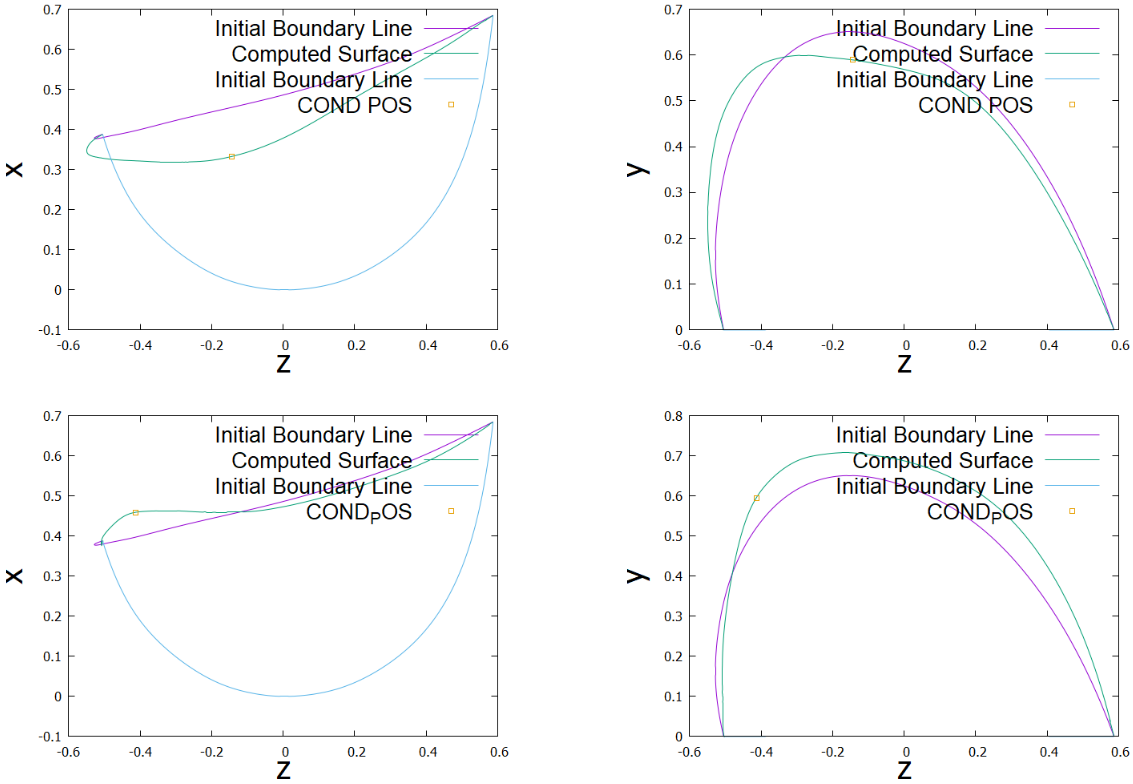


2　データ周りの話について

2.1　前回の研究会までで行ったこと

前回の考察では，主にデータ間の距離に着目し，議論を進めた．しかし，内挿および外挿的な部分が議論できていないなど，詰めるべき課題が多く残っていた．

そのため今回考えたのは，内挿と外挿の話であった．はじめは，内挿と外挿を定義するため，入力データの凸包を考え，凸



包内に出力データが入っていなければ外挿というアイデアを考えたが，計算結果として，少なくとも凸包内には全ての出力データは含まれるという結果しか得られなかったため，頓挫．

ここで，改めてテストデータに対する予測値の絶対誤差と分散をまとめたものを示す．

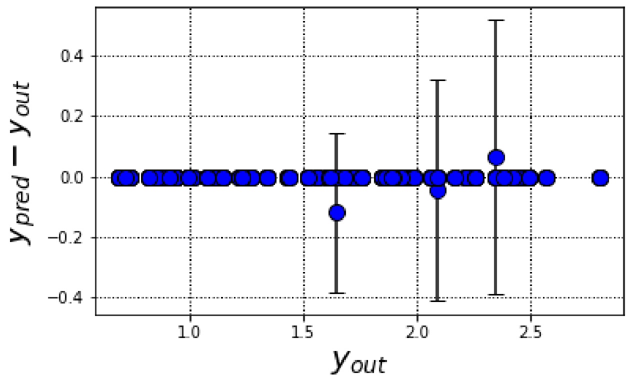


Fig. 1 Manhattan

この結果から，例え入力パラメータが異なっていても，出力値は似たような値を取ることが多い．という事象が読み取れる．このことから，1 つの仮説を導く．

2.2 仮説

このデータの意味する所を，整理すれば，

- ▼ 誤差が大きい評価値においても，似たような評価値を出力する別の入力パラメータが存在する．

となる．ここから，今回扱う問題の性質として以下を得る．

▼ 入力形状に対する評価値は，単射ではない．

単射を概説する．ある集合 $\boldsymbol{X}, \boldsymbol{Y}$ に対して対応 $f : \boldsymbol{X} \to \boldsymbol{Y}$ が与えられている場合，$\boldsymbol{X}$ の任意の 2 元 $\boldsymbol{x}, \boldsymbol{x}'$ をどのようにとっても，$f(\boldsymbol{x}) \neq f(\boldsymbol{x}')$ が成り立つことを単射であるという．

今回の評価値の成り立ちについて考えれば，カップの囲う体積と面積を割ることで決まる評価値であり，関数の単射性が成立しないのは，自然で不都合はないと思われる．

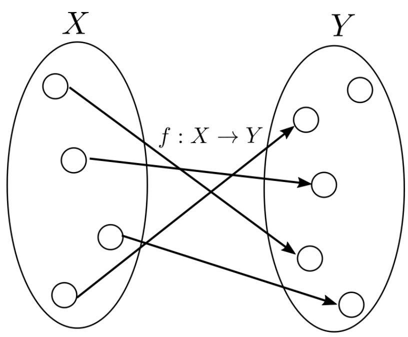


Fig. 2 単射

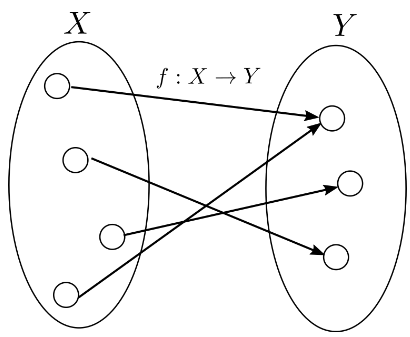


Fig. 3 単射ではない

ここから導いた仮説は

▼ 検証データにおける誤差の大きな値をとった出力値における入力値は，訓練データにおいて似たような出力値における入力値と似ていなかったため，誤差が発生した．

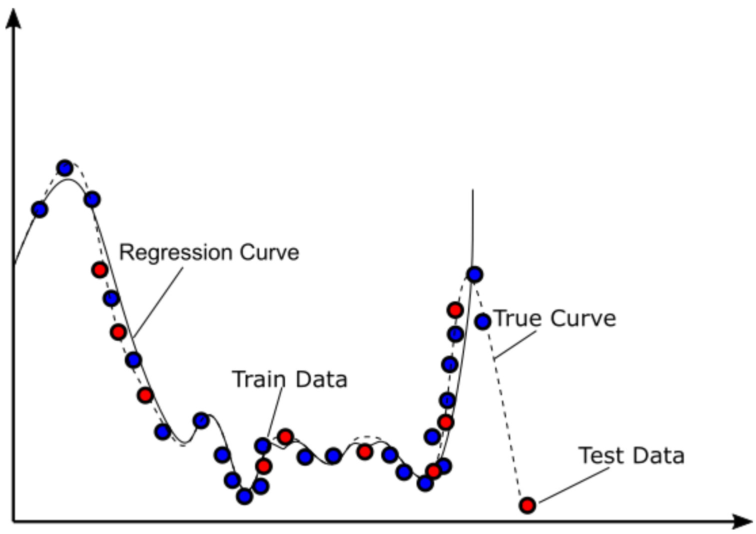


Fig. 4 外挿

例えば，図のような形を考えると，データ間の距離だけで着目すれば，似たようなデータは存在するが，このように出力を加味すれば，このデータは外挿的であり，回帰曲線を用いて計算した値と異なる可能性は高い．今回は，このことを考慮しながら，数値実験を行った．

2.3 数値実験内容

今回はこの仮説を基にして，以下のように実験を行った．以下では，訓練データ集合を $\boldsymbol{D} = (\boldsymbol{X}, \boldsymbol{y})$ と表現し，検証データ集合を $\boldsymbol{D}^* = (\boldsymbol{X}^*, \boldsymbol{y}^*)$ として表現する．

1. 検証データにおける入力パラメータ集合 $\boldsymbol{X}^*$ の全ての元 $\boldsymbol{x}^*$ に対して，近傍データセット $\boldsymbol{U}(\boldsymbol{D})$ を決定する．
2. $\boldsymbol{U}(\boldsymbol{D})$ に対する最短距離 δ を求める．この δ が大きければ大きいほど，乖離が激しく，$\boldsymbol{x}^*$ は外挿的なデータである

と判断できる．

近傍 $\boldsymbol{U}(\boldsymbol{D})$ は，以下のように定義される集合である．

$$\boldsymbol{U}(\boldsymbol{D}) = \{(\boldsymbol{x}, y) \in \boldsymbol{D} || y - y^*| \leq \varepsilon\} \tag{1}$$

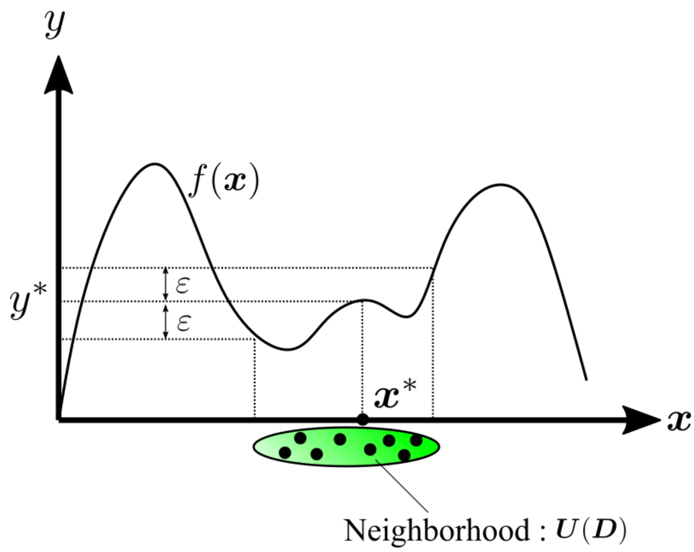


Fig. 5 Neighborhood

また，δ は以下のように定義する．

$$\delta = \min_{\boldsymbol{u} \in \boldsymbol{U}(\boldsymbol{D})} d(\boldsymbol{u}, \boldsymbol{x}^*) \tag{2}$$

今回は，$\varepsilon = 0.02$ とし，距離空間には，先月の結果から，マンハッタン距離を採用する．

これを用いて全体のパラメータを用いて計算した結果と，各 α, ω_η, D のそれぞれで分けて計算した結果を，以下に示す．

前回の結果では，データの距離が大きい部分があっても，誤差や分散が小さい例があったが，今回の場合にはそのようなことがなく，ほぼ正確に距離との相関が確認される．

個別の例では，絶対誤差と特に α の比例関係が確認される．これは，以前示したガウス過程回帰によって得られたパラメータ群における考察の結果である，α が重要なパラメータであるという事実に矛盾しない結果となった．

全体のパラメータで見ると，データの曖昧さを表現している分散の大きさと距離がほぼ完全な比例関係にあることが読み取れる．つまり，外挿的であればあるほど，得られる解の信頼度に揺らぎが生じるという結果を得た．

上記の内容は，研究会や査読に対して突っ込まれた際にも応用できるかと思いますが，いかがでしょうか．

3　次回の MTG について（終了後記載）

- ▼
- ▼

＃＃＃

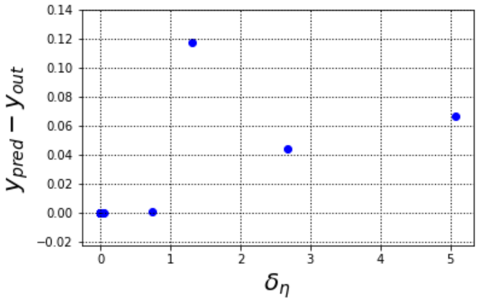


Fig. 6 ω_η

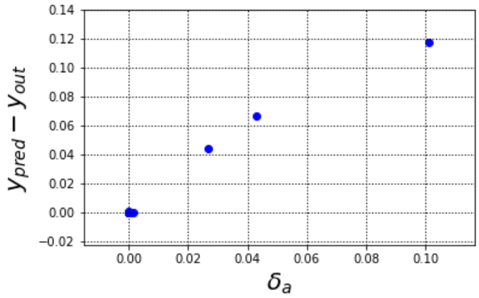


Fig. 7 α

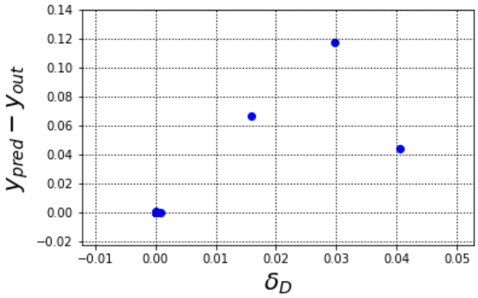


Fig. 8 D

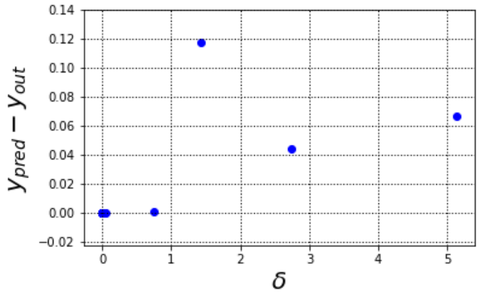


Fig. 9 All Params

Fig. 10 Regarding Abs Err

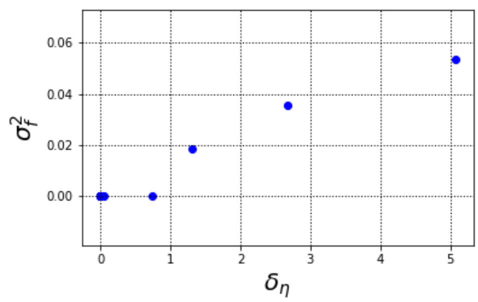


Fig. 11 ω_η

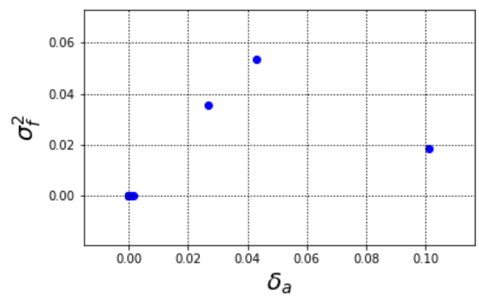


Fig. 12 α

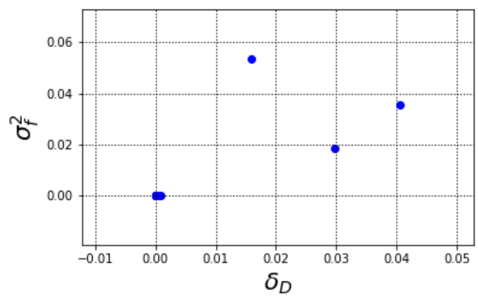


Fig. 13 D

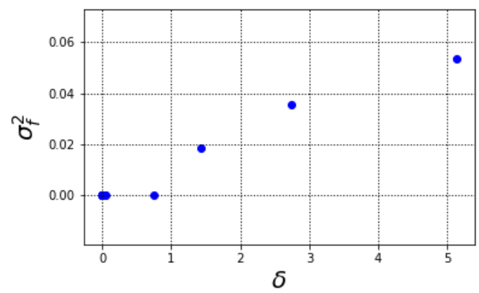


Fig. 14 All Params

Fig. 15 Regarding Vars