

Query By Committeeに対する近似アルゴリズム

坂田 綾香 数理・推論研究系 准教授

【Query By Committeeとは】

- 能動学習法のひとつ[1]
 - 統計モデルにとって「不確実性」が高いデータを取り入れて 学習効率を上げる
- 不確実性をCommittee memberから評価[2]
 - Committee memberに異なる属性を与えるqueryを次の入力とする

【具体例：パーセプトロン学習】

$$Y = \text{sign}(\tilde{F}x^{(0)})$$

- $Y \in \{-1, +1\}$: 観測データ
- $\tilde{F} \in \mathbb{R}^{1 \times N}$: 観測ベクトル
- $x^{(0)} \in \mathbb{R}^N$: 真のパラメータ

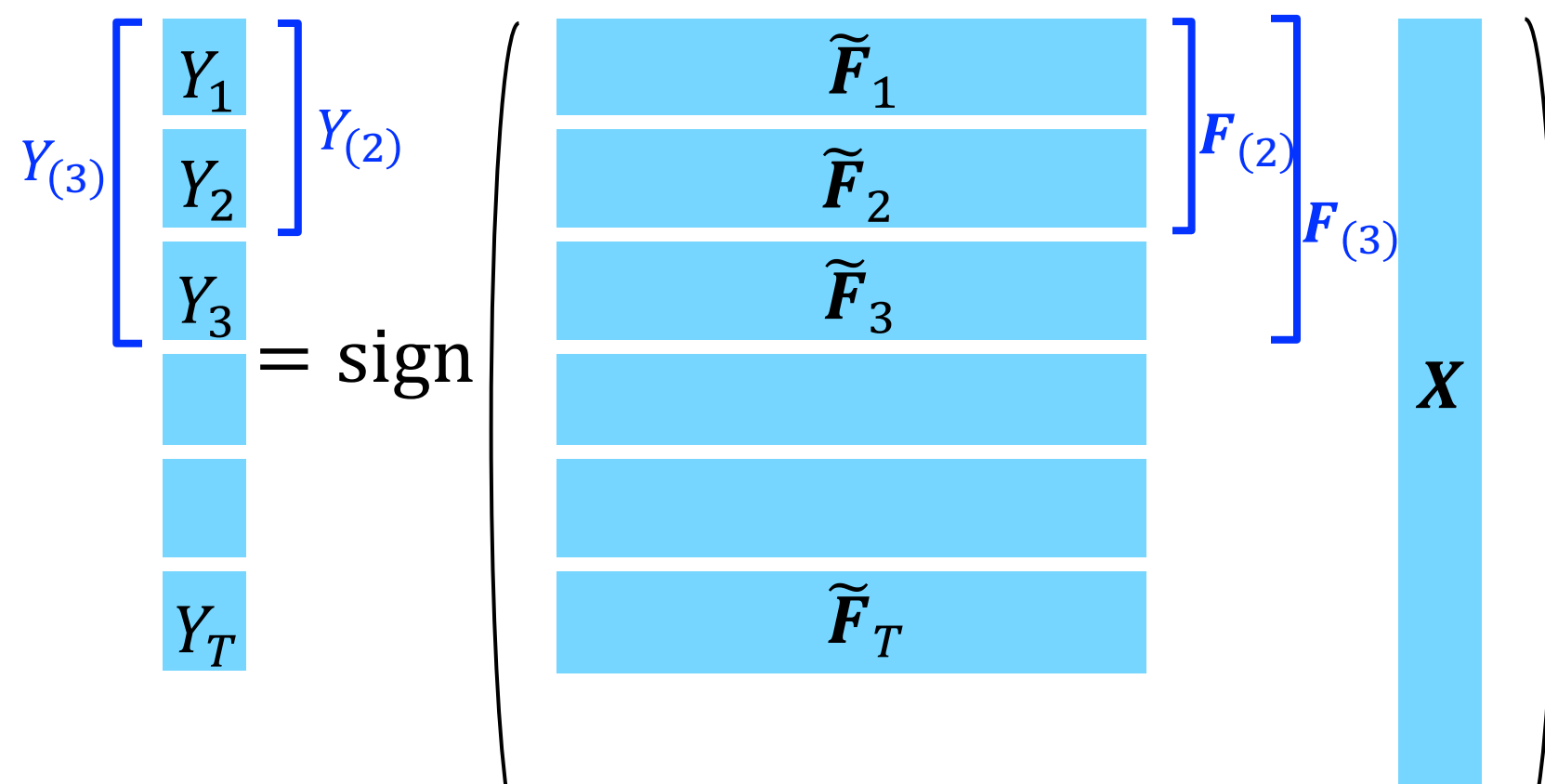
- Committee memberに対する観測が、正負半々となるような観測ベクトルをqueryとする
- その観測ベクトルと出力の組みを新たにデータとして加える

【効率的アルゴリズム】

- ベイズ推定における近似手法である近似確率伝搬法(AMP)を適用する
 - AMPは観測に相関がない時に有効であることが知られている → QBCのように、観測に相関がある場合でもAMPが使えるのかを検証する

Notation

- t 回目の観測ベクトルを \tilde{F}_t とする
- 1回目から t 回目の観測ベクトル集合を $F_{(t)}$ とする
- $F_{(t)} = [F_{(t-1)}; \tilde{F}_t]$ と表記する



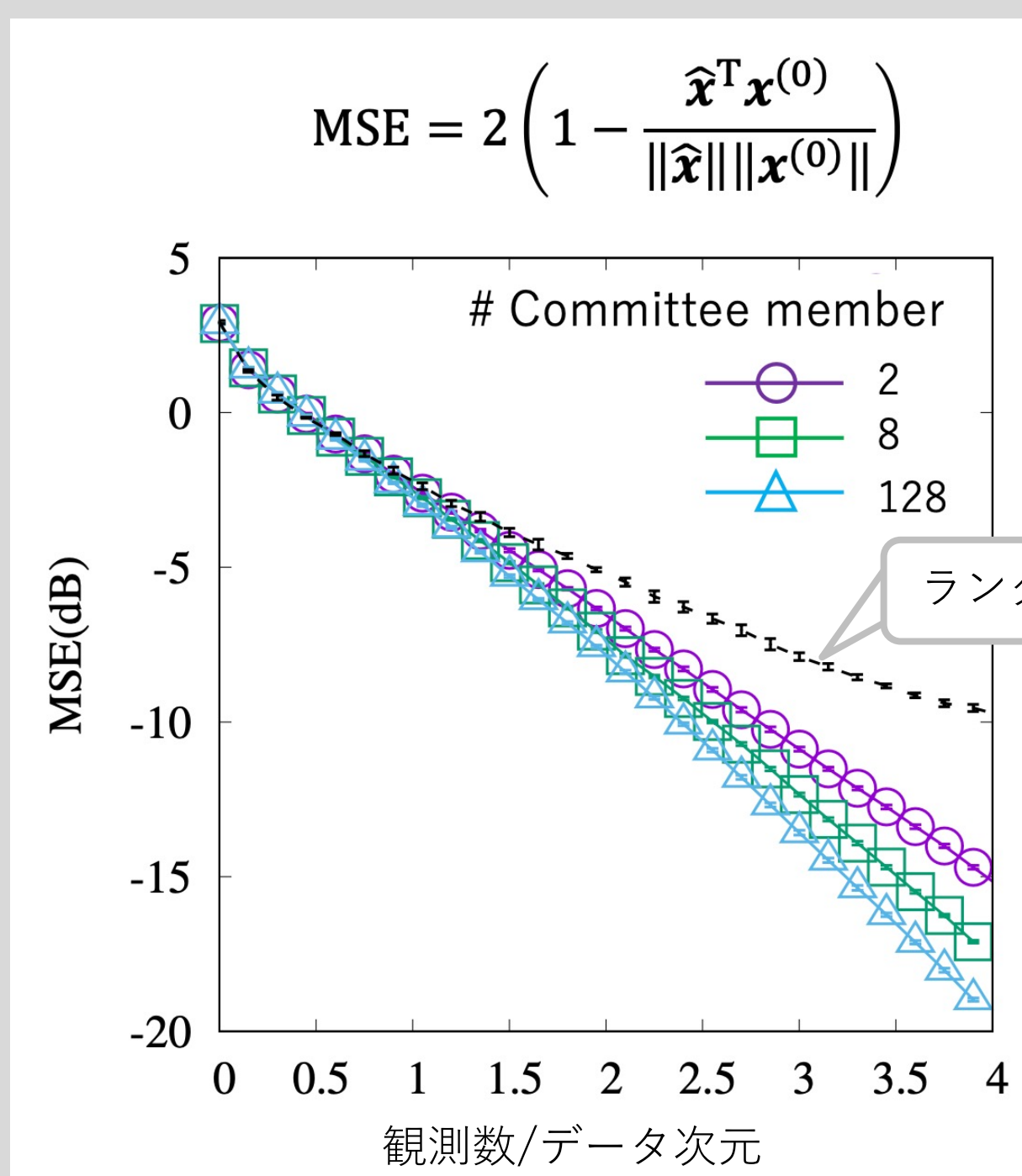
AMPによるQBCアルゴリズム

```

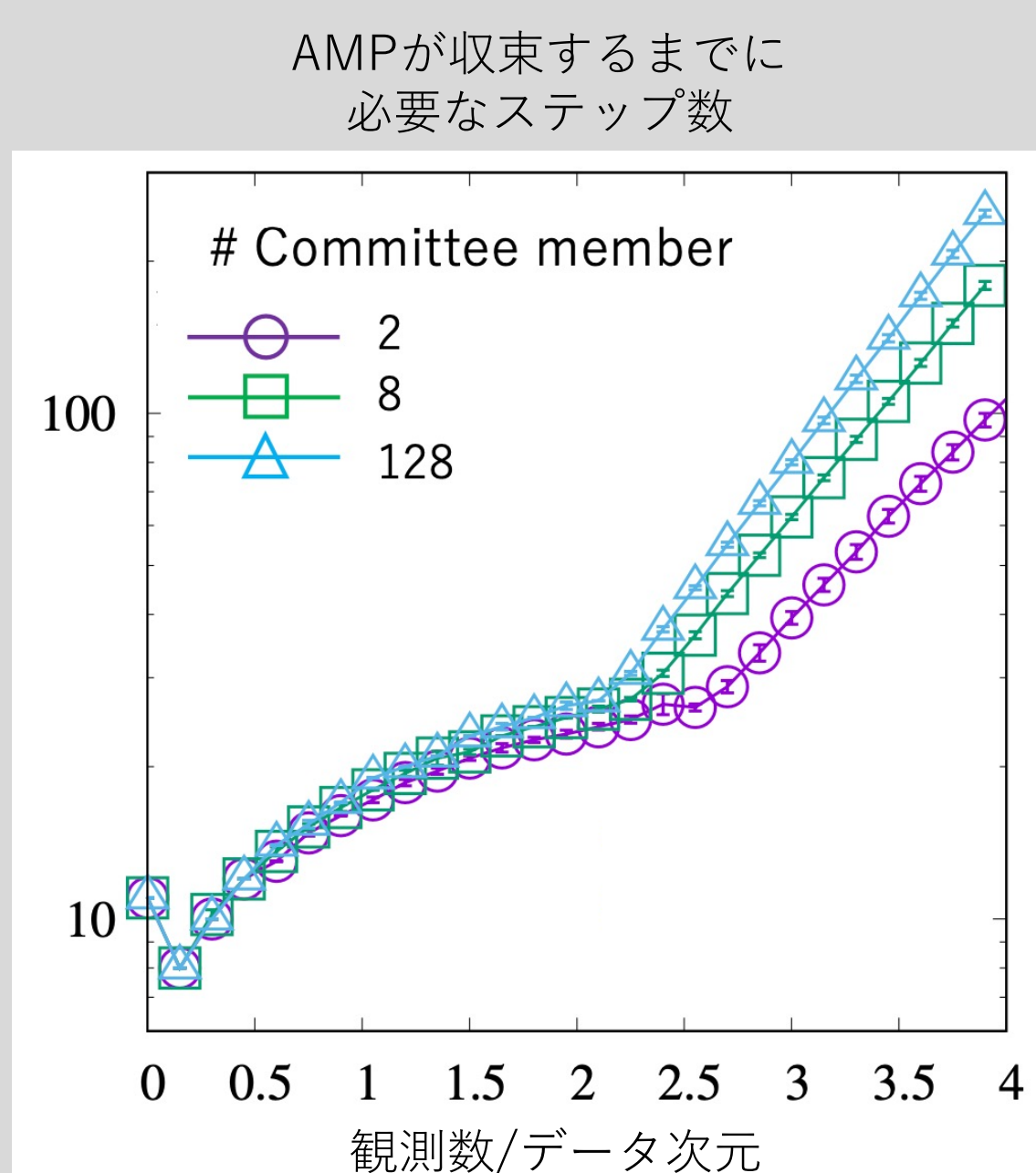
Set true signal:  $x^{(0)} \in \mathbb{R}^N$ 
Initialization:  $F_{(1)} = \tilde{F}_1 \sim P_F(\tilde{F}_1), Y_1 = \text{sign}(F_{(1)}x^{(0)})$ 

# Start
for  $t = 1, \dots, T$ 
     $\{\hat{x}_{(t)}, V_{(t)}\} \leftarrow \text{AMP}(F_{(t-1)}, Y_{(t-1)})$ 
    for  $i = 1, \dots, 2k$ 
         $x_i \sim \mathcal{N}(\hat{x}_{(t)}, V_{(t)})$ 
    end
     $\tilde{F}_t \sim P_F(\tilde{F}_t) \delta(\sum_{i=1}^{2k} \theta(\tilde{F}_t x_i) - k)$ 
     $F_{(t)} \leftarrow [F_{(t-1)}; \tilde{F}_t]$ 
     $Y_{(t)} = \text{sign}(F_{(t)}x^{(0)})$ 
end
    
```

結果：AMPによりQBCを近似的に解くことは可能

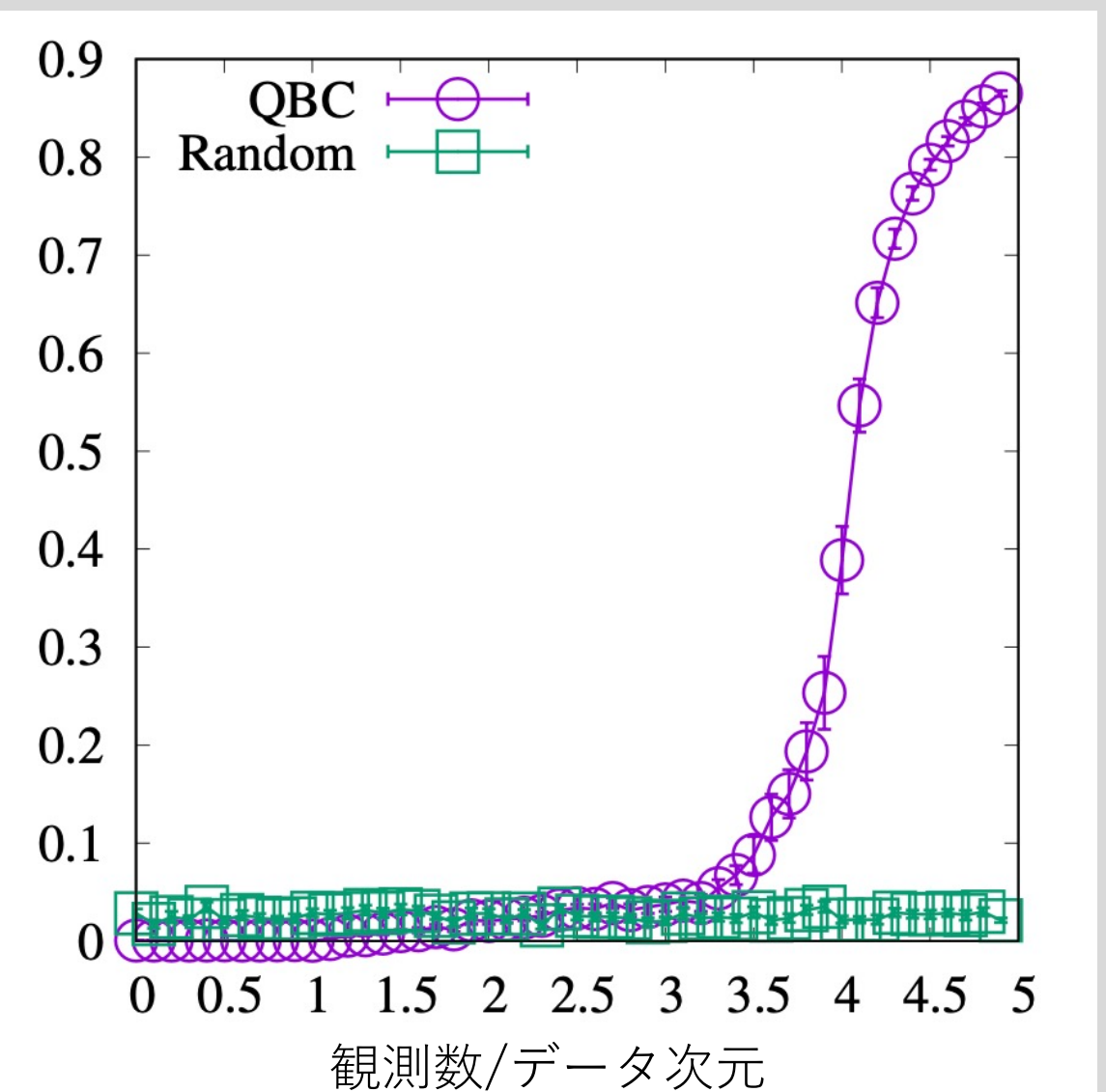


- 観測数を増やすとエラー(MSE)が減少する
 - ランダム観測よりも効率よくMSEが減る
 - Committee member数が多い方が良い



- しかし…
 - 観測数が増えていくと、AMPの収束までに必要な時間が指数関数的に長くなる
 - でも収束はする

Committee member数 = 2での第N右特異ベクトルと事後分布平均の相関係数



- その原因は、QBCにより作られる観測行列の性質にある
 - 観測数が増えていくと、観測行列の第N特異ベクトルと事後分布平均に相関が生まれる
 - この性質による収束の悪さを改善すればより効率的なアルゴリズムになると期待される

【参考文献】

[1] B. Settles, "Active Learning Literature Survey", (2009), [2] H. S. Seung, M. Opper and H. Sompolinsky, COLT (1992)