# 制約付き非凸最適化問題のための確率的 Frank-Wolfe 法とその敵対的攻撃への応用

会員　明治大学　*佐藤 尚樹　SATO Naoki
会員　明治大学　飯塚 秀明　IIDUKA Hideaki

## 1. はじめに

Frank-Wolfe 法は，制約付き最適化問題のための古典的な一次手法であり，その確率的変種である確率的 Frank-Wolfe 法は，制約集合への射影計算を必要としない特徴などから機械学習分野でよく利用されている．しかし，非凸最適化問題のための既存の収束解析は非現実的な仮定に基づいているため，特に深層学習における Frank-Wolfe 法の優れた性能を裏付ける理論は存在しない．本発表は，実用的な定数および減少学習率を有する確率的 Frank-Wolfe 法の収束解析を提供する．さらに，Frank-Wolfe 法に基づく新しい敵対的攻撃手法を提案し，提案手法が既存手法と同等程度の性能を有することを実験的に示す．

## 2. 確率的 Frank-Wolfe 法

本発表は次の制約付き非凸最適化問題を扱う．

$$\min_{\boldsymbol{\theta}\in\Omega}\left\{f(\boldsymbol{\theta}):=\frac{1}{n}\sum_{i=1}^{n}f_i(\boldsymbol{\theta})\right\} \tag{1}$$

制約付き最適化問題においては，$\|\nabla f(\boldsymbol{\theta})\|$ は収束の指標として利用できないため，代わりに次のように定義される Frank-Wolfe gap $\mathcal{G}(\boldsymbol{\theta})$ を導入する．

$$\mathcal{G}(\boldsymbol{\theta}):=\max_{\boldsymbol{v}\in\Omega}\langle \boldsymbol{v}-\boldsymbol{\theta},-\nabla f(\boldsymbol{\theta})\rangle$$

問題 (1) の最適解を $\boldsymbol{\theta}^\star$ とすると，任意の $\boldsymbol{v}\in\Omega$ に対して

$$\langle \boldsymbol{v}-\boldsymbol{\theta}^\star,\nabla f(\boldsymbol{\theta}^\star)\rangle\geq 0 \ \text{ i.e. } \mathcal{G}(\boldsymbol{\theta}^\star)\leq 0$$

が成り立つため，$\mathcal{G}(\boldsymbol{\theta})$ を小さくすることを目指す．

Frank-Wolfe 法（アルゴリズム 1）は，現在のパラメータ $\boldsymbol{\theta}_t$ と，制約集合 $\Omega$ 内で最急降下方向 $-\nabla f(\boldsymbol{\theta}_t)$ との内積を最も大きくする $\boldsymbol{v}$ との凸結合を $\boldsymbol{\theta}_{t+1}$ とするため，点列が制約集合の外に出ることがなく，制約集合への射影の計算を必要としないという特徴がある．

**アルゴリズム 1** Frank-Wolfe 法

**Require:** $\boldsymbol{\theta}_0\in\Omega,(\gamma_t)_{t\in\mathbb{N}}\subset\mathbb{R}_{++}$
  **for** $t=0$ to $T-1$ **do**
    $\boldsymbol{v}_t=\mathrm{argmax}_{\boldsymbol{v}\in\Omega}\langle \boldsymbol{v},-\nabla f(\boldsymbol{\theta}_t)\rangle$
    $\boldsymbol{d}_t=\boldsymbol{v}_t-\boldsymbol{\theta}_t$
    $\boldsymbol{\theta}_{t+1}=\boldsymbol{\theta}_t+\gamma_t\boldsymbol{d}_t$
  **end for**
    **return** $\boldsymbol{\theta}_T$

確率的 Frank-Wolfe 法（アルゴリズム 2）は，全勾配 $\nabla f(\boldsymbol{\theta}_t)$ の代わりにミニバッチ確率的勾配 $\nabla f_{\mathcal{S}_t}(\boldsymbol{\theta}_t):=\frac{1}{b}\sum_{i\in[b]}\mathsf{G}_{\xi_{t,i}}(\boldsymbol{\theta}_t)$ を使う手法である．ただし，$b$ はバッチサイズといい，$\mathsf{G}_{\xi_{t,i}}(\boldsymbol{\theta}_t)$ は $\mathbb{E}_{\xi_t}\left[\mathsf{G}_{\xi_t}(\boldsymbol{\theta}_t)\right]=\nabla f(\boldsymbol{\theta}_t)$ を満たす．

**アルゴリズム 2** 確率的 Frank-Wolfe 法

**Require:** $\boldsymbol{\theta}_0\in\Omega,(\gamma_t)_{t\in\mathbb{N}}\subset\mathbb{R}_{++}$
  **for** $t=0$ to $T-1$ **do**
    $\boldsymbol{v}_t=\mathrm{argmax}_{\boldsymbol{v}\in\Omega}\langle \boldsymbol{v},-\nabla f_{\mathcal{S}_t}(\boldsymbol{\theta}_t)\rangle$
    $\boldsymbol{d}_t=\boldsymbol{v}_t-\boldsymbol{\theta}_t$
    $\boldsymbol{\theta}_{t+1}=\boldsymbol{\theta}_t+\gamma_t\boldsymbol{d}_t$
  **end for**
    **return** $\boldsymbol{\theta}_T$

制約付き凸最適化問題のための確率的 Frank-Wolfe 法は理論的にも実験的にも非常によく研究されており，例えばサポートベクターマシンの訓練や回帰問題に適用されている．制約付き非凸最適化問題のための確率的 Frank-Wolfe 法の既存の収束解析には，例えば次のようなものがある [1]．

**定理 1** 関数 $f$ が $L$-平滑であるとする．このとき，学習率 $\gamma_t$ とバッチサイズ $b$ が $\gamma_t=\sqrt{\frac{C}{T}},b=T$ を満たすならば，任意の $t\in[T]$ に対して，

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\mathcal{G}(\boldsymbol{x}_t)\right]\leq\mathcal{O}\left(\frac{1}{T}\right)$$

が成り立つ．ただし，$C$ はある正の定数である．

この定理は多くの先行研究に引用されているが，学習率 $\gamma_t$ とバッチサイズ $b$ が反復回数 $T$ に依存している点が実用的でない．特に深層学習では，$T$ は非常に大きいため，$\gamma_t = \sqrt{\frac{C}{T}}$ で設定すると学習率が非常に小さくなり，点列がそれほど更新されない可能性がある．また，バッチサイズが大きすぎると，各反復でのミニバッチ確率的勾配の計算が困難になるため，$b = T$ の設定も現実的ではない．

## 3. 収束解析

本発表では，ユーザーが任意に設定できる学習率とバッチサイズを有する収束解析を提供する．

**定理 2**（定数学習率）関数 $f$ が $L$-平滑であるとする．このとき，定数学習率 $\gamma_t = \gamma$ を有する確率的 *Frank-Wolfe* 法が生成する点列を $(\boldsymbol{\theta}_t)_{t\in\mathbb{N}}$ とすると，

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\mathcal{G}(\boldsymbol{x}_t)\right] \leq \frac{\mathbb{E}[f(\boldsymbol{\theta}_0)] - \mathbb{E}[f(\boldsymbol{\theta}^\star)]}{\gamma T} + \frac{D\sigma}{\sqrt{b}} + \frac{LD^2\gamma}{2}$$

が成り立つ．ただし，$D, \sigma$ は正の定数である．

**定理 3**（減少学習率）関数 $f$ が $L$-平滑であるとする．このとき，減少学習率 $\gamma_t = (\underbrace{\gamma, \gamma, \cdots, \gamma}_{K}, \underbrace{\eta\gamma, \eta\gamma, \cdots, \eta\gamma}_{K}, \cdots, \underbrace{\eta^{P-1}\gamma, \eta^{P-1}\gamma, \cdots, \eta^{P-1}\gamma}_{K})$ を有する確率的 *Frank-Wolfe* 法が生成する点列を $(\boldsymbol{\theta}_t)_{t\in\mathbb{N}}$ とすると，

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\mathcal{G}(\boldsymbol{x}_t)\right] \leq \mathcal{O}\left(\frac{1}{T} + \frac{\sigma}{\sqrt{b}}\right)$$

が成り立つ．ただし，$\sigma$ は正の定数である．

## 4. 敵対的攻撃

画像分類タスクにおいて，深層ニューラルネットワーク（DNN）は，本物の訓練画像に人間が感知できないほどわずかなノイズを加えた敵対的サンプルに弱いことが明らかになっている．すなわち，訓練済みの DNN に猫の画像を入力すると，DNN モデルは「猫」と出力できるが，この画像にごくわずかなノイズを加えた画像を入力すると，DNN モデルは「犬」などと誤った分類をしてしまう．このような敵対的サンプルを生成することを敵対的攻撃といい，敵対的攻撃に対して頑健なモデルのことをロバストモデルという．ロバストモデルを作るためにいくつかの手法が提案されているが，訓練時に本物の訓練画像に加えて敵対的サンプルを利用する敵対的訓練という手法が最も一般的である．このように，敵対的サンプルは敵対的訓練やロバストモデルのロバスト性の検証に利用されるため，より良い敵対的攻撃手法の開発は，より良いロバストモデルの開発のために重要である．

最後に，敵対的攻撃は制約付き非凸最適化問題であることを示す．訓練データセット $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) | \boldsymbol{x}_i \in \mathbb{R}^d, \boldsymbol{y}_i \in \mathbb{R}^c\}_{i=0}^{n-1}$ が与えられたとし，このデータセットで DNN モデル $g(\boldsymbol{x}, \boldsymbol{\theta})$ を訓練することを考える．ここで，$\boldsymbol{x}_i \in \mathbb{R}^d$ は本物の訓練画像，$\boldsymbol{y}_i$ は $\boldsymbol{x}_i$ に対応するラベル，$\boldsymbol{\theta}$ はモデル $g$ のパラメータ，$c \in \mathbb{N}$ は分類のクラス数を表している．また，$g(\boldsymbol{x})$ はモデルの予測，$f_i(g(\boldsymbol{x}_i), \boldsymbol{y}_i)$ は $i$ 番目の訓練データに対する損失関数である．損失関数 $f(\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=0}^{n-1} f_i(g(\boldsymbol{x}_i), \boldsymbol{y}_i)$ を最小化することでモデルの訓練が完了し，モデル $g$ は入力画像 $\boldsymbol{x}$ を高い精度で分類できるようになる．そのあとで入力画像 $\boldsymbol{x}$ にごくわずかなノイズを加えることで，訓練済みのモデルの予測を意図的に誤らせることを目指す．ある距離関数 $d(\cdot, \cdot)$ と正の数 $\epsilon > 0$ を使うと，敵対的サンプル $\boldsymbol{x}_{\mathrm{adv}}$ は $\boldsymbol{x}_{\mathrm{adv}} \in \{\hat{\boldsymbol{x}}_i \mid g(\hat{\boldsymbol{x}}_i) \neq \boldsymbol{y}_i, d(\hat{\boldsymbol{x}}_i, \boldsymbol{x}_i) \leq \epsilon\}$ と表せる．したがって敵対的攻撃は，次のような最適化問題として定式化される．

$$\max_{\hat{\boldsymbol{x}}\in\mathbb{R}^d} f(g(\hat{\boldsymbol{x}}), \boldsymbol{y}) \text{ s.t. } d(\hat{\boldsymbol{x}}, \boldsymbol{x}) \leq \epsilon.$$

これは問題 (1) の一例であり，制約集合 $\Omega$ は $\Omega = \{\hat{\boldsymbol{x}} \in \mathbb{R}^d \mid d(\hat{\boldsymbol{x}}, \boldsymbol{x}) \leq \epsilon\}$ となる．なお，距離関数 $d$ にはユークリッドノルムや最大値ノルムがよく利用される．このように敵対的攻撃は制約付きの非凸最適化問題として定式化できるため，Frank-Wolfe 法による敵対的攻撃が可能である．

## 参考文献


[1] S. J. Reddi, S. Sra, B. Poczos, and A. Smola, Stochastic Frank-Wolfe Methods for Nonconvex Optimization. In 54th Annual Allerton Conference on Communication, Control, and Computing, pp. 1214-1251, 2016.