# Sharper Bounds for Uniformly Stable Algorithms with Stationary Mixing Process


Shi Fu[1], Yunwen Lei[2], Qiong Cao[3], Xinmei Tian[1] and Dacheng Tao[3,4]

[1]University of Science and Technology of China
[2]Hong Kong Baptist University
[3]JD Explore Academy
[4]University of Sydney



ICLR 2023

# Motivation

- **Generalization analysis** of learning algorithms often builds on a critical assumption that training examples are independently identically distributed, which is often violated in practical problems such as time series prediction.

- A widely used relaxation of the i.i.d. assumption is to assume the observations are drawn from a mixing process, where the dependency between observations weakens over time.

- In this paper, we use algorithmic stability to study the generalization performance of learning algorithms with $\psi$**-mixing data**.

# Stationary Mixing Process

Let $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ be a stationary sequence of random variables. For any $i, j \in \mathbb{N}$, let $\sigma_i^j$ denote the $\sigma$-algebra generated by the random variables $Z_k, i \leq k \leq j$.

1. $\varphi$-Mixing Sequence: For any $k \in \mathbb{N}$, the $\varphi$-mixing coefficient of $\mathbf{Z}$ is defined as

$$\varphi(k) = \sup_{n, A \in \sigma_{n+k}^{\infty}, B \in \sigma_{-\infty}^{n}} \Big| \mathrm{Pr}(A|B) - \mathrm{Pr}(A) \Big|.$$

   $\mathbf{Z}$ is said to be $\varphi$-mixing if $\varphi(k) \to 0$ as $k \to \infty$.

2. $\psi$-Mixing Sequence: For any $k \in \mathbb{N}$, the $\psi$-mixing coefficient of the stochastic process $\mathbf{Z}$ is defined as

$$\psi(k) = \sup_{n, A \in \sigma_{n+k}^{\infty}, B \in \sigma_{-\infty}^{n}} \Big| \mathrm{Pr}(A \cap B)/\mathrm{Pr}(A)\mathrm{Pr}(B) - 1 \Big|.$$

   $\mathbf{Z}$ is said to be $\psi$-mixing if $\psi(k) \to 0$ as $k \to \infty$.

3. $\varphi'$-Mixing Coefficient: Our stability analysis requires a different mixing coefficient defined as follows

$$\varphi'(k) = \sup_{n, A \in \sigma_{-\infty}^{n-k}, z_n \in \sigma_n^n, B \in \sigma_{n+k}^{\infty}} \Big| \mathrm{Pr}(z_n|A, B) - \mathrm{Pr}(z_n) \Big|.$$

# Definitions

Let $L, \gamma > 0$ and $\mu \geq 0$. Let $f : \mathcal{W} \times \mathcal{Z} \mapsto \mathbb{R}$.

**Lipschitzness:** We say $f$ is $L$-Lipschitz continuous if $|f(\mathbf{w}; z) - f(\mathbf{w}'; z)| \leq L\|\mathbf{w} - \mathbf{w}'\|$ for any $\mathbf{w}, \mathbf{w}', z$.

**Smoothness:** We say $f$ is $\gamma$-smooth if $\|\nabla f(\mathbf{w}; z) - \nabla f(\mathbf{w}'; z)\| \leq \gamma\|\mathbf{w} - \mathbf{w}'\|$ for any $\mathbf{w}, \mathbf{w}', z$.

**Convexity:** We say $f$ is $\mu$-strongly convex if for any $\mathbf{w}, \mathbf{w}', z$ we have $f(\mathbf{w}; z) - f(\mathbf{w}'; z) - \langle \mathbf{w} - \mathbf{w}', \nabla f(\mathbf{w}'; z)\rangle \geq \frac{\mu}{2}\|\mathbf{w} - \mathbf{w}'\|^2$. We say $f$ is convex if it is $\mu$-strongly convex with $\mu = 0$.

## Algorithmic Stability

Let $S = \{z_1, \ldots, z_n\}, S' = \{z'_1, \ldots, z'_n\}$ be independently drawn from sample space.

We say $A$ is $\epsilon$-uniformly stable if for any datasets $S, \hat{S} \in \mathcal{Z}^n$ that differ by at most a single example we have

$$\sup_z \left[f(\mathbf{w}_S; z) - f(\mathbf{w}_{\hat{S}}; z)\right] \leq \epsilon.$$

# Moment Bound For $\varphi$-Mixing Sequence

## Moment Bound

Let $X_1, \ldots, X_n$ be a finite contiguous subsequence from a $\varphi$-mixing sequence. Let $Z_i$ be a function of $X_i$ with $\mathbb{E}[Z_i] = 0$ and $\Pr\{|Z_i| > \tilde{\epsilon}\} \leq 2\exp(-\tilde{\epsilon}^2/b)$. Then for any $p \geq 1$ we have

$$\Big\| \sum_{i=1}^{n} Z_i \Big\|_p \leq (9 + \log(n)) p \Delta_n \sqrt{2nb}.$$

The bound matches the existing moment bounds for i.i.d. random variables up to a logarithmic factor

# Concentration Inequality For $\varphi$-Mixing Sequence

## Concentration Inequality

Let $Z_1, \ldots, Z_n$ be a finite contiguous subsequence from a $\varphi$-mixing sequence. Let $g_1, \ldots, g_n$ be some functions $g_i : \mathcal{Z}^n \mapsto \mathbb{R}$ such that the following holds for any $i \in [n]$

- $\left|\mathbb{E}_{Z_{[n]\setminus[i]}}[g_i(Z)|Z_i]\right| \leq M$ almost surely,
- $\mathbb{E}_{Z_i}[g_i(Z)|Z_{[n]\setminus[i]}] = 0$ a.s.,
- $g_i$ is $\beta$-Lipschitz.

Then for any $p \geq 1$ we have

$$\left\| \sum_{i=1}^{n} g_i \right\|_p \leq 3M\Delta_n\sqrt{2pn} + 2^k p\beta \sum_{l=0}^{k-1}(9+l)\Delta_{2^l}^2.$$

- Our bound recovers the existing result in the i.i.d. case.
- The concentration inequality is developed for $\varphi$-mixing sequence and can be directly applied to $\psi$-mixing sequence.

# Error Decomposition

let $S^i_{i,b} = \{z_1, \ldots, z_{i-b-1}, z'_i, z_{i+b+1}, \ldots, z_{n-b}\}$. We then define the following random variables

$$g_i = \mathbb{E}_{z'_i}\big[\mathbb{E}_{z''_i}[f(\mathbf{w}_{S^i_{i,b}}; z''_i)] - f(\mathbf{w}_{S^i_{i,b}}; z_i)\big], \quad \forall i \in [n].$$

## Error Decomposition

Let $S$ be drawn from a $\psi$-mixing distribution. If the algorithm $A$ is $\beta$-uniformly stable and the loss function is bounded by $M > 0$, then

$$\big|n(F(\mathbf{w}_S) - F_S(\mathbf{w}_S))\big| \leq 2n(3b+1)\beta + nM(\varphi(b) + \varphi'(b)) + \big|\sum_{i=1}^n g_i\big|.$$

# Stability and Generalization

## General Mixing Stability Bound

Let $S$ be drawn from a $\psi$-mixing distribution. Then for any $b \in \{0, \ldots, n\}$ and any $\delta \in (0,1)$, the following inequality holds with probability at least $1-\delta$

$$\big|F(\mathbf{w}_S) - F_S(\mathbf{w}_S)\big| = O\Big(\varphi'(b) + \Delta_n \sqrt{n^{-1}\log(1/\delta)} + \beta(b + \Delta_n^2 \log^2 n \log(1/\delta))\Big).$$

- In the i.i.d. case, stability bound becomes

  $$\big|F(\mathbf{w}_S) - F_S(\mathbf{w}_S)\big| = O\big(\sqrt{n^{-1}\log(1/\delta)} + \beta \log^2 n \log(1/\delta)\big),$$

  which matches existing stability-based bounds up to a logarithmic factor.

- Our stability bound improves on previous state-of-the-art results by a factor of $O(\sqrt{n})$.

*Thank you!*