
Collective Stability in Structured Prediction: Appendix

Ben London*

Bert Huang*

Ben Taskar†

Lise Getoor*

BLONDON@CS.UMD.EDU

BERT@CS.UMD.EDU

TASKAR@CS.WASHINGTON.EDU

GETOOR@CS.UMD.EDU

* University of Maryland, College Park, MD 20742 USA

† University of Washington, Seattle, WA 98195 USA

A. Proof of Theorem 1

Before proceeding, we recall a general form of McDiarmid's inequality.

Theorem 7 (McDiarmid, 1989, Corollary 6.10). *Let $f : \mathcal{Z}^n \rightarrow \mathbb{R}$ be a measurable function for which there exist constants $\{\alpha_i\}_{i=1}^n$ such that, for any $i \in [n]$, $\mathbf{z}_{1:i-1} \in \mathcal{Z}^{i-1}$ and $z_i, z'_i \in \mathcal{Z}$,*

$$|\mathbb{E}[f(\mathbf{Z}) \mid \mathbf{z}_{1:i-1}, z_i] - \mathbb{E}[f(\mathbf{Z}) \mid \mathbf{z}_{1:i-1}, z'_i]| \leq \alpha_i.$$

Then, for any $\epsilon > 0$,

$$\mathbb{P}\{f(\mathbf{Z}) - \mathbb{E}[f(\mathbf{Z})] \geq \epsilon\} \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n \alpha_i^2}\right).$$

Note that the above does not require independence. To prove Theorem 1, it therefore suffices to bound $\sum_{i=1}^n \alpha_i^2$. Kontorovich & Ramanan (2008, Remark 2.1) showed that, if f is c -Lipschitz with respect to the Hamming metric, then $\sum_{i=1}^n \alpha_i^2 \leq nc^2 \|\Theta_n^\pi\|_\infty^2$. (Though the published results only prove this for countable spaces, Kontorovich later extended this analysis to continuous spaces in his thesis (2007).) If f is c -Lipschitz with respect to the *normalized* Hamming metric, then $\sum_{i=1}^n \alpha_i^2 \leq c^2 \|\Theta_n^\pi\|_\infty^2 / n$, which completes the proof.

B. Proof of Corollary 1

We begin by establishing that $\mathbb{E}[L(h, \mathbf{Z}')] = \mathbb{E}[L(h, \mathbf{Z})]$. We use $l \in [m]$ to iterate over examples. Accordingly, let $Z'_{l,i}$ denote the i^{th} variable in example \mathbf{Z}'_l . Recall that each \mathbf{Z}'_l is independent and identically distributed according to $\mathbb{P}(\mathbf{Z})$. By linearity of expect-

tation, we have that

$$\begin{aligned} \mathbb{E}[L(h, \mathbf{Z}')] &= \mathbb{E}\left[\frac{1}{mn} \sum_{l=1}^m \sum_{i=1}^n \ell(Y'_{l,i}, h_i(\mathbf{X}'_l))\right] \\ &= \frac{1}{m} \sum_{l=1}^m \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \ell(Y'_{l,i}, h_i(\mathbf{X}'_l))\right] \\ &= \frac{1}{m} \sum_{l=1}^m \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \ell(Y_i, h_i(\mathbf{X}))\right] \\ &= \mathbb{E}[L(h, \mathbf{Z})]. \end{aligned}$$

To complete the proof, we simply apply Theorem 2 to $\mathbb{E}[L(h, \mathbf{Z}')]$, using the fact that $\|\Theta_{mn}^\pi\|_\infty = \|\Theta_n^\pi\|_\infty$ because the dependency matrix Θ_{mn}^π is block diagonal.

C. Proof of Lemma 1

By definition, for any $\mathbf{z}, \mathbf{z}' \in \mathcal{Z}^n$ that differ only at the i^{th} coordinate,

$$\begin{aligned} &\sum_{j=1}^n |\ell(y_j, h_j(\mathbf{x})) - \ell(y'_j, h_j(\mathbf{x}'))| \\ &= |\ell(y_i, h_i(\mathbf{x})) - \ell(y'_i, h_i(\mathbf{x}'))| \\ &\quad + \sum_{j \neq i} |\ell(y_j, h_j(\mathbf{x})) - \ell(y_j, h_j(\mathbf{x}'))|. \end{aligned}$$

Focusing on the first term, we have via the first admissibility condition that

$$\begin{aligned} &|\ell(y_i, h_i(\mathbf{x})) - \ell(y'_i, h_i(\mathbf{x}'))| \\ &\leq |\ell(y_i, h_i(\mathbf{x})) - \ell(y_i, h_i(\mathbf{x}'))| \\ &\quad + |\ell(y_i, h_i(\mathbf{x}')) - \ell(y'_i, h_i(\mathbf{x}'))| \\ &\leq |\ell(y_i, h_i(\mathbf{x})) - \ell(y_i, h_i(\mathbf{x}'))| + M. \end{aligned}$$

Combining this with the second term, we have that

$$\begin{aligned}
 & \sup_{h \in \mathcal{H}} \sum_{j=1}^n |\ell(y_j, h_j(\mathbf{x})) - \ell(y_j, h_j(\mathbf{x}'))| \\
 & \leq M + \sup_{h \in \mathcal{h}} \sum_{j=1}^n |\ell(y_j, h_j(\mathbf{x})) - \ell(y_j, h_j(\mathbf{x}'))| \\
 & \leq M + \lambda \sup_{h \in \mathcal{h}} \sum_{j=1}^n \|h_j(\mathbf{x}) - h_j(\mathbf{x}')\|_1 \\
 & = M + \lambda \sup_{h \in \mathcal{h}} \|h(\mathbf{x}) - h(\mathbf{x}')\|_1 \\
 & \leq M + \lambda \beta,
 \end{aligned}$$

where we have used the second admissibility condition and uniform collective stability.

D. Proof of Lemma 2

Let $\mathbf{z}, \mathbf{z}' \in \mathcal{Z}^n$ be two realizations that differ only at a single coordinate. Without loss of generality, since $|\Phi(\mathcal{F}, \mathbf{z}) - \Phi(\mathcal{F}, \mathbf{z}')| = |\Phi(\mathcal{F}, \mathbf{z}') - \Phi(\mathcal{F}, \mathbf{z})|$, assume that $\Phi(\mathcal{F}, \mathbf{z}) \geq \Phi(\mathcal{F}, \mathbf{z}')$. By definition, we have that

$$\begin{aligned}
 & |\Phi(\mathcal{F}, \mathbf{z}) - \Phi(\mathcal{F}, \mathbf{z}')| \\
 & = \left| \sup_{f \in \mathcal{F}} \{\bar{F} - F(\mathbf{z})\} - \sup_{f' \in \mathcal{F}} \{\bar{F}' - F'(\mathbf{z}')\} \right| \\
 & \leq \left| \sup_{f \in \mathcal{F}} \bar{F} - F(\mathbf{z}) - \bar{F} + F(\mathbf{z}') \right| \\
 & = \left| \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n f_i(\mathbf{z}') - f_i(\mathbf{z}) \right| \\
 & \leq \sup_{f \in \mathcal{F}} \frac{1}{n} \|f(\mathbf{z}') - f(\mathbf{z})\|_1 \leq \frac{\beta}{n}.
 \end{aligned}$$

The last inequality follows from uniform collective stability. We now have that $\Phi(\mathcal{F}, \mathbf{Z})$ satisfies the pre-conditions of [Theorem 1](#), with $c = \beta$. Recalling that $\bar{\Phi}(\mathcal{F}) = \mathbb{E}[\Phi(\mathcal{F}, \mathbf{Z})]$, we therefore have that

$$\mathbb{P} \{ \Phi(\mathcal{F}, \mathbf{Z}) - \bar{\Phi}(\mathcal{F}) \geq \epsilon \} \leq \exp \left(\frac{-2n\epsilon^2}{\beta^2 \|\Theta_n^\pi\|_\infty^2} \right).$$

Assigning δ probability to this event and solving for ϵ completes the proof.

E. Proof of Lemma 3

For the following, we use variables \mathbf{Z} and \mathbf{Z}' to distinguish between realizations of the training and testing sets respectively. Using the definition of $\bar{\Phi}(\mathcal{F})$ and

Jensen's inequality, we have that

$$\begin{aligned}
 \bar{\Phi}(\mathcal{F}) & = \mathbb{E} \left[\sup_{f \in \mathcal{F}} \mathbb{E}[F(\mathbf{Z}')] - F(\mathbf{Z}) \right] \\
 & \leq \mathbb{E} \left[\sup_{f \in \mathcal{F}} F(\mathbf{Z}') - F(\mathbf{Z}) \right].
 \end{aligned}$$

Now define a set of Rademacher variables $\{\sigma_i\}_{i=1}^n$, and let

$$T(\sigma_i) \triangleq \begin{cases} \mathbf{Z} & \text{if } \sigma_i = 1, \\ \mathbf{Z}' & \text{if } \sigma_i = -1, \end{cases}$$

and

$$T'(\sigma_i) \triangleq \begin{cases} \mathbf{Z}' & \text{if } \sigma_i = 1, \\ \mathbf{Z} & \text{if } \sigma_i = -1. \end{cases}$$

Because $\mathbf{Z} \perp \mathbf{Z}'$ and $\mathbb{P}(\mathbf{Z}) = \mathbb{P}(\mathbf{Z}')$, it follows that $\mathbb{P}(\mathbf{Z}, \mathbf{Z}') = \mathbb{P}(T(\sigma_i) | \sigma_i) \mathbb{P}(T'(\sigma_i) | \sigma_i)$; so, by symmetry,

$$\begin{aligned}
 \bar{\Phi}(\mathcal{F}) & \leq \mathbb{E} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^n f_i(\mathbf{Z}') - f_i(\mathbf{Z}) \right] \\
 & = \mathbb{E} \left[\mathbb{E} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n f_i(T'(\sigma_i)) - f_i(T(\sigma_i)) \mid \boldsymbol{\sigma} \right] \right] \\
 & = \mathbb{E} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sigma_i (f_i(\mathbf{Z}') - f_i(\mathbf{Z})) \right] \\
 & \leq 2 \mathbb{E} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sigma_i f_i(\mathbf{Z}) \right] = 2\bar{\mathfrak{R}}_n(\mathcal{F}),
 \end{aligned}$$

which completes the proof.

F. Proof of Lemma 4

We begin with a technical lemma, which is a generalization of Talagrand's contraction lemma ([Ledoux & Talagrand, 1991](#)) to vector-valued functions and arbitrary norms.

Lemma 10. *Let \mathcal{F} be a class of functions from a domain \mathcal{Z} to \mathbb{R}^k . Let $\{\sigma_i\}_{i=1}^n$ be a set of Rademacher variables. If $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$ is λ -Lipschitz under $\|\cdot\|_p$, for any $p \geq 1$, then, for any $\mathbf{z} \in \mathcal{Z}^n$,*

$$\mathbb{E} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^n \sigma_i \varphi(f_j(z_i)) \right] \leq \lambda \sum_{j=1}^k \mathbb{E} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^n \sigma_i f_j(z_i) \right].$$

Proof. Define a function $S_n(f) \triangleq \sum_{i=1}^n \sigma_i \varphi(f(z_i))$. Conditioned on $\boldsymbol{\sigma}_{1:n-1}$, we know that there must exist

two functions $f^+, f^- \in \mathcal{H}$ such that

$$\begin{aligned}
 & \mathbb{E} \left[\sup_{f \in \mathcal{F}} S_n(f) \mid \boldsymbol{\sigma}_{1:n-1} \right] \\
 &= \mathbb{E} \left[\sup_{f \in \mathcal{F}} S_{n-1}(f) + \sigma_n \varphi(f(z_n)) \mid \boldsymbol{\sigma}_{1:n-1} \right] \\
 &= \frac{1}{2} [S_{n-1}(f^+) + \varphi(f^+(z_n))] \\
 &\quad + \frac{1}{2} [S_{n-1}(f^-) - \varphi(f^-(z_n))] \\
 &= \frac{1}{2} [S_{n-1}(f^+) + S_{n-1}(f^-) \\
 &\quad + \varphi(f^+(z_n)) - \varphi(f^-(z_n))] \\
 &\leq \frac{1}{2} [S_{n-1}(f^+) + S_{n-1}(f^-) \\
 &\quad + \lambda \|f^+(z_n) - f^-(z_n)\|_p],
 \end{aligned}$$

where the last line follows from the Lipschitz condition. For each $j \in [k]$, define a variable $s_{n,j} \triangleq \text{sgn}(f_j^+(z_n) - f_j^-(z_n))$, and note that

$$\begin{aligned}
 \|f^+(z_n) - f^-(z_n)\|_p &\leq \|f^+(z_n) - f^-(z_n)\|_1 \\
 &= \sum_{j=1}^k s_{n,j} (f_j^+(z_n) - f_j^-(z_n)).
 \end{aligned}$$

This yields

$$\begin{aligned}
 & \mathbb{E} \left[\sup_{f \in \mathcal{F}} S_n(f) \mid \boldsymbol{\sigma}_{1:n-1} \right] \\
 &\leq \frac{1}{2} \left[S_{n-1}(f^+) + \lambda \sum_{j=1}^k s_{n,j} f_j^+(z_n) \right] \\
 &\quad + \frac{1}{2} \left[S_{n-1}(f^-) - \lambda \sum_{j=1}^k s_{n,j} f_j^-(z_n) \right] \\
 &\leq \mathbb{E} \left[\sup_{f \in \mathcal{F}} S_{n-1}(f) + \lambda \sum_{j=1}^k \sigma_n s_{n,j} f_j(z_n) \mid \boldsymbol{\sigma}_{1:n-1} \right].
 \end{aligned}$$

By induction on n , we therefore have that

$$\begin{aligned}
 \mathbb{E} \left[\sup_{f \in \mathcal{F}} S_n(f) \right] &\leq \mathbb{E} \left[\sup_{f \in \mathcal{F}} \lambda \sum_{j=1}^k \sum_{i=1}^n \sigma_i s_{i,j} f_j(z_i) \right] \\
 &\leq \lambda \sum_{j=1}^k \mathbb{E} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^n \sigma_i f_j(z_i) \right],
 \end{aligned}$$

where $s_{i,j}$ disappears because of symmetry. \square

The proof of Lemma 4 follows directly from this lemma, since the second admissibility condition ensures that ℓ is λ -Lipschitz under the 1-norm. The fact

that $h : \mathcal{X}^n \rightarrow \hat{\mathcal{Y}}^n$ is irrelevant. Because Lemma 10 holds for any realization $\mathbf{z} \in \mathcal{Z}^n$, we obtain the (non-empirical) Rademacher complexity by taking the expectation over \mathbf{Z} .

G. Proof of Lemma 5

Let $\Delta a \triangleq a - \dot{a}$. By Definition 5, for any $\tau \in [0, 1]$,

$$\tau(1-\tau) \frac{\kappa}{2} \|\Delta a\|_1^2 + \varphi(\dot{a} + \tau \Delta a) - \varphi(\dot{a}) \leq \tau(\varphi(a) - \varphi(\dot{a})).$$

Since \dot{a} is, by definition, the unique minimizer of φ , it follows that $\varphi(\dot{a} + \tau \Delta a) - \varphi(\dot{a}) \geq 0$, so the above inequality is preserved when this term is dropped. Thus, dividing both sides by $\tau \kappa / 2$, we have that

$$\|\Delta a\|_1^2 \leq (1-\tau) \|\Delta a\|_1^2 \leq \frac{2}{\kappa} (\varphi(a) - \varphi(\dot{a})),$$

where the left inequality follows from the fact that $(1-\tau)$ is maximized at $\tau = 0$.

H. Proof of Lemma 6

Let $\dot{a} \triangleq \arg \min_{a \in \mathcal{A}} \varphi(\omega, a)$ and $\dot{a}' \triangleq \arg \min_{a' \in \mathcal{A}} \varphi(\omega', a')$. Without loss of generality, assume that $\varphi(\omega, \dot{a}) \geq \varphi(\omega', \dot{a}')$. (If $\varphi(\omega', \dot{a}') \geq \varphi(\omega, \dot{a})$, we could state this in terms of ω' .) Using Lemma 5, we have that

$$\begin{aligned}
 \|\dot{a}' - \dot{a}\|_1^2 &\leq \frac{2}{\kappa} (\varphi(\omega, \dot{a}') - \varphi(\omega, \dot{a})) \\
 &\leq \frac{2}{\kappa} (\varphi(\omega, \dot{a}') - \varphi(\omega', \dot{a}')) \\
 &\leq \frac{2}{\kappa} \lambda.
 \end{aligned}$$

Taking the square root completes the proof.

I. Proof of Lemma 7

Using Cauchy-Schwarz, we have that

$$\begin{aligned}
 & |E_{\mathbf{w}}(\mathbf{x}, \mathbf{a}) - E_{\mathbf{w}}(\mathbf{x}', \mathbf{a})| \\
 &= |\langle \mathbf{w}, \mathbf{f}(\mathbf{x}, \mathbf{a}) \rangle - \Psi(\mathbf{a}) - \langle \mathbf{w}, \mathbf{f}(\mathbf{x}', \mathbf{a}) \rangle + \Psi(\mathbf{a})| \\
 &= |\langle \mathbf{w}, \mathbf{f}(\mathbf{x}, \mathbf{a}) - \mathbf{f}(\mathbf{x}', \mathbf{a}) \rangle| \\
 &\leq \|\mathbf{w}\|_2 \|\mathbf{f}(\mathbf{x}, \mathbf{a}) - \mathbf{f}(\mathbf{x}', \mathbf{a})\|_2 \\
 &\leq R \|\mathbf{f}(\mathbf{x}, \mathbf{a}) - \mathbf{f}(\mathbf{x}', \mathbf{a})\|_2,
 \end{aligned}$$

because, by definition, $\|\mathbf{w}\|_2$ is uniformly upper-bounded by R . Note that the features of (\mathbf{x}, \mathbf{a}) and $(\mathbf{x}', \mathbf{a})$ only differ at any clique involving node i . The number of such cliques is Q_i , which is uniformly upper-bounded by Q_G , so at most Q_G features will change. Further, since the norm of any feature function is, by

definition, uniformly upper-bounded by B , we have that

$$\begin{aligned}
 & \| \mathbf{f}(\mathbf{x}, \mathbf{a}) - \mathbf{f}(\mathbf{x}', \mathbf{a}) \|_2 \\
 &= \sqrt{\sum_{t \in \mathcal{T}} \left\| \sum_{q \in t(G)} \mathbb{1}[i \in q] (f_t(\mathbf{x}_q, \mathbf{a}_q) - f_t(\mathbf{x}'_q, \mathbf{a}_q)) \right\|_2^2} \\
 &\leq \sqrt{\sum_{t \in \mathcal{T}} \left(\sum_{q \in t(G)} \mathbb{1}[i \in q] \|f_t(\mathbf{x}_q, \mathbf{a}_q) - f_t(\mathbf{x}'_q, \mathbf{a}_q)\|_2 \right)^2} \\
 &\leq \sqrt{\left(\sum_{t \in \mathcal{T}} \sum_{q \in t(G)} \mathbb{1}[i \in q] \|f_t(\mathbf{x}_q, \mathbf{a}_q) - f_t(\mathbf{x}'_q, \mathbf{a}_q)\|_2 \right)^2} \\
 &\leq 2BQ_i \leq 2BQ_G,
 \end{aligned}$$

which completes the proof.

J. Proof of Lemma 8

We will partition $[0, \Lambda]^d$ into k hypercube *cells* with edge length $(2\epsilon/\sqrt{d})$. Using multidimensional geometry, one can show the hypercube $[0, 2\epsilon/\sqrt{d}]^d$ is inscribed in a ball of radius ϵ ; therefore, the Euclidean distance from any point in $[0, \Lambda]^d$ to the center of the nearest cell is at most ϵ . To find the value of k that ϵ -covers $[0, \Lambda]^d$, we let $k(2\epsilon/\sqrt{d})^d \geq \Lambda^d$ and solve for k .

K. Discretization Theorem

The following is a consequence of Massart's finite class lemma.

Theorem 8. *Let \mathcal{F} be a class of functions from \mathcal{Z}^n to \mathbb{R}^n . For any $n \geq 1$ and $p \geq 1$,*

$$\begin{aligned}
 \mathfrak{R}(\mathcal{F}, \mathbf{Z}) &\leq \inf_{\epsilon} \sqrt{\frac{2 \ln \mathcal{N}_p(\epsilon, \mathcal{F}, \mathbf{Z})}{n}} + \epsilon, \\
 \text{and } \bar{\mathfrak{R}}_n(\mathcal{F}) &\leq \inf_{\epsilon} \sqrt{\frac{2 \ln \mathcal{N}_p(\epsilon, \mathcal{F}, n)}{n}} + \epsilon.
 \end{aligned}$$

L. Proof of Lemma 9

The *ramp function* is defined as

$$r_\gamma(a) \triangleq \begin{cases} 1 & \text{for } a \leq 0, \\ 1 - a/\gamma & \text{for } 0 < a \leq \gamma, \\ 0 & \text{for } a > \gamma. \end{cases}$$

By definition, r_γ (hence, ℓ_γ) is bounded by $[0, 1]$; so for any $y, y' \in \mathcal{Y}$ and $\hat{y} \in \hat{\mathcal{Y}}$, $|\ell_\gamma(y, \hat{y}) - \ell_\gamma(y', \hat{y})| \leq 1$, which establishes the first admissibility condition.

For $\hat{y}, \hat{y}' \in \hat{\mathcal{Y}}$, let $u \triangleq \arg \max_{y' \in \mathcal{Y}: y' \neq \hat{y}'} \langle y', \hat{y} \rangle$ and $u' \triangleq \arg \max_{y' \in \mathcal{Y}: y' \neq \hat{y}'} \langle y', \hat{y}' \rangle$. Without loss of generality, assume that $\langle y, \hat{y} \rangle - \langle u, \hat{y} \rangle \geq \langle y, \hat{y}' \rangle - \langle u', \hat{y}' \rangle$. For any $y \in \mathcal{Y}$ and $\hat{y}, \hat{y}' \in \hat{\mathcal{Y}}$, we have that

$$\begin{aligned}
 & |(\langle y, \hat{y} \rangle - \langle u, \hat{y} \rangle) - (\langle y, \hat{y}' \rangle - \langle u', \hat{y}' \rangle)| \\
 &= |\langle y, \hat{y} - \hat{y}' \rangle + \langle u', \hat{y}' \rangle - \langle u, \hat{y} \rangle| \\
 &\leq |\langle y, \hat{y} - \hat{y}' \rangle + \langle u', \hat{y}' \rangle - \langle u', \hat{y} \rangle| \\
 &= |\langle y - u', \hat{y} - \hat{y}' \rangle| \\
 &\leq \|y - u'\|_\infty \|\hat{y} - \hat{y}'\|_1 \\
 &\leq \|\hat{y} - \hat{y}'\|_1.
 \end{aligned}$$

Further, for any $a, a' \in \mathbb{R}$,

$$|r_\gamma(a) - r_\gamma(a')| \leq \left| \frac{1-a}{\gamma} - \frac{1-a'}{\gamma} \right| = \frac{1}{\gamma} |a - a'|.$$

Combining these inequalities, we have that $|\ell_\gamma(y, \hat{y}) - \ell_\gamma(y, \hat{y}')| \leq (1/\gamma) \|\hat{y} - \hat{y}'\|_1$, which establishes the second admissibility condition.

L.1. Collective Regression

In collective regression, the codomain is a bounded interval on the real number line. Since the output can always be shifted and scaled by a constant, we can assume without loss of generality that $\mathcal{Y}, \hat{\mathcal{Y}} \subseteq [0, 1]$. A standard loss function for regression is the *quadratic loss*, typically defined as $\ell_q(y, \hat{y}) \triangleq (y - \hat{y})^2$.

Lemma 11. *The quadratic loss ℓ_q is $(1, 2)$ -admissible.*

Proof. First, since both \mathcal{Y} and $\hat{\mathcal{Y}}$ are upper-bounded by 1, we have the first admissibility condition. Second, note that ℓ_q is smooth with respect to its second argument. Therefore, by the mean value theorem, there exists a $\tau \in [0, 1]$ such that, for any $y \in \mathcal{Y}$ and $\hat{y}, \hat{y}' \in \hat{\mathcal{Y}}$, with $\Delta \hat{y} \triangleq \hat{y}' - \hat{y}$,

$$\begin{aligned}
 |\ell_q(y, \hat{y}) - \ell_q(y, \hat{y}')| &= \left| \frac{\partial}{\partial \hat{y}} [\ell_q(y, \hat{y} + \tau \Delta \hat{y})](\Delta \hat{y}) \right| \\
 &= |-2(y - (\hat{y} + \tau \Delta \hat{y}))(\Delta \hat{y})| \\
 &\leq 2|y - (\hat{y} + \tau \Delta \hat{y})| |\Delta \hat{y}| \\
 &\leq 2|\Delta \hat{y}| = 2\|\hat{y} - \hat{y}'\|_1,
 \end{aligned}$$

which establishes the second condition. \square

We can thus obtain bounds on the quadratic risk for the class of TSM regressors with strongly convex regularizers, similar to how we obtained [Theorem 6](#).

References

Kontorovich, L. *Measure Concentration of Strongly Mixing Processes with Applications*. PhD thesis, Carnegie Mellon University, 2007.

Kontorovich, L. and Ramanan, K. Concentration inequalities for dependent random variables via the martingale method. *Annals of Probability*, 36(6): 2126–2158, 2008.

Ledoux, M. and Talagrand, M. *Probability in Banach Spaces: Isoperimetry and Processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, 1991.

McDiarmid, C. On the method of bounded differences. In *Surveys in Combinatorics*, volume 141 of *London Mathematical Society Lecture Note Series*, pp. 148–188. Cambridge University Press, 1989.