Collective Stability in Structured Prediction: Appendix

Ben London[*] BLONDON@CS.UMD.EDU
Bert Huang[*] BERT@CS.UMD.EDU
Ben Taskar[†] TASKAR@CS.WASHINGTON.EDU
Lise Getoor[*] 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 $f : \mathcal{Z}^n \to \mathbb{R}$ 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 $z_i, z_i' \in \mathcal{Z}$,*

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

Then, for any $\epsilon > 0$,

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

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 $\sum_{i=1}^n \alpha_i^2 \leq nc^2 \left\|\mathbf{\Theta}_n^\pi\right\|_\infty^2$. (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 $\sum_{i=1}^n \alpha_i^2 \leq c^2 \left\|\mathbf{\Theta}_n^\pi\right\|_\infty^2 / n$, 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 expectation, 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 $\left\|\mathbf{\Theta}_{mn}^\pi\right\|_\infty = \left\|\mathbf{\Theta}_n^\pi\right\|_\infty$ because the dependency matrix $\mathbf{\Theta}_{mn}^\pi$ 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,

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

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

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

Combining this with the second term, we have that

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

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 $\Phi(\mathcal{F}, \mathbf{z}) \ge \Phi(\mathcal{F}, \mathbf{z}')$. By definition, we have that

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

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

$$\mathbb{P}\left\{\Phi(\mathcal{F}, \mathbf{Z}) - \overline{\Phi}(\mathcal{F}) \ge \epsilon\right\} \le \exp\left(\frac{-2n\epsilon^2}{\beta^2 \left\|\mathbf{\Theta}_n^\pi\right\|_\infty^2} \right).$$

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 $\overline{\Phi}(\mathcal{F})$ and Jensen's inequality, we have that

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

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 $\mathbf{Z} \perp\!\!\!\perp \mathbf{Z}'$ 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,

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

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 $\varphi : \mathbb{R}^k \to \mathbb{R}$ is λ-Lipschitz under $\|\cdot\|_p$, for any $p \ge 1$, then, for any $\mathbf{z} \in \mathcal{Z}^n$,*

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

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

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

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

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

This yields

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

By induction on n, we therefore have that

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

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

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 $h : \mathcal{X}^n \to \hat{\mathcal{Y}}^n$ 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 $\tau \in [0, 1]$,

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

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

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

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 $\varphi(\omega, \dot{a}) \ge \varphi(\omega', \dot{a}')$. (If $\varphi(\omega', \dot{a}') \ge \varphi(\omega, \dot{a})$, we could state this in terms of ω'.) Using Lemma 5, we have that

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

Taking the square root completes the proof.

I. Proof of Lemma 7

Using Cauchy-Schwarz, we have that

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

because, by definition, $\left\|\mathbf{w}\right\|_2$ 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

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

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 $k(2\epsilon/\sqrt{d})^d \ge \Lambda^d$ 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 $n \ge 1$ and $p \ge 1$,*

$$\mathfrak{R}(\mathcal{F},\mathbf{Z}) \le \inf_{\epsilon} \sqrt{\frac{2\ln \mathcal{N}_p(\epsilon,\mathcal{F},\mathbf{Z})}{n}} + \epsilon,$$
$$\text{and} \quad \overline{\mathfrak{R}}_n(\mathcal{F}) \le \inf_{\epsilon} \sqrt{\frac{2\ln \mathcal{N}_p(\epsilon,\mathcal{F},n)}{n}} + \epsilon.$$

L. Proof of Lemma 9

The *ramp function* is defined as

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

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

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

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

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

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

Combining these inequalities, we have that $|\ell_\gamma(y,\hat{y}) - \ell_\gamma(y,\hat{y}')| \le (1/\gamma)\|\hat{y} - \hat{y}'\|_1$, 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 $\Delta\hat{y} \triangleq \hat{y}' - \hat{y}$,

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

which establishes the second condition. □

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.