Beyond Johnson–Lindenstrauss

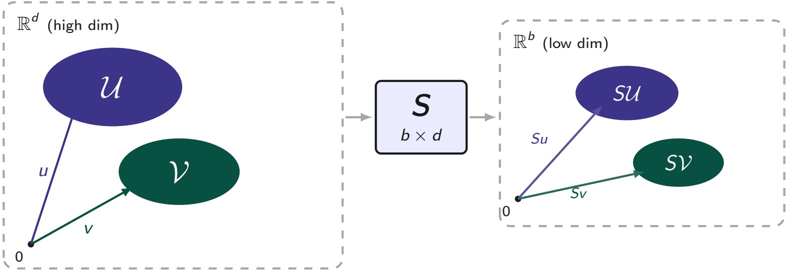
Uniform Bounds for Sketched Bilinear Forms

Rohan Deb Qiaobo Li Mayank Shrivastava Arindam Banerjee

University of Illinois, Urbana–Champaign

— problem statement —

Bound how much a sketch distorts inner products



Goal

Bound, with high probability,

$$\sup_{u \in \mathcal{U},\, v \in \mathcal{V}} \left| \langle Su, Sv \rangle - \langle u, v \rangle \right|.$$

Where this template shows up

Template: $$\sup_{u\in\mathcal{U},\, v\in\mathcal{V}} \left|u^\top S^\top S\, v - u^\top v\right|.$$

classical — $u = v$

Johnson–Lindenstrauss lemma

Preserve norms on a finite set $\mathcal{A}$:

$$\left|\, \|Su\|^2 - \|u\|^2 \,\right|, \quad u \in \mathcal{A}.$$

$\mathcal{U} = \mathcal{V} = \mathcal{A}$

Restricted isometry property

Same, but over s-sparse vectors Σ_s:

$$\left|\, \|Su\|^2 - \|u\|^2 \,\right|, \quad u \in \Sigma_s.$$

$\mathcal{U} = \mathcal{V} = \Sigma_s$

bilinear — $u \neq v$

Linear regression

Model $y_i = \beta^\top x_i + \eta_i$, $\beta \in \mathcal{B}$, $x_i \in \mathcal{X}$.
Bound $\left|\, \beta^\top S^\top S\, x - \beta^\top x \,\right|$.

Federated learning

Sketch gradient $g \in \mathcal{G}$; Hessian eigvec $h \in \mathcal{H}$.
Bound $\left|\, g^\top S^\top S\, h - g^\top h \,\right|$.

Linear bandits

Reward $r_i = {\theta_\star}^\top a_i + \eta_i$, $\theta_\star \in \Theta$, $a_i \in \mathcal{A}$.
Bound $\left|\, {\theta_\star}^\top S^\top S\, a - {\theta_\star}^\top a \,\right|$.

Random quadratic forms: from fixed A to a set $\mathcal{A}$

$$\sup_{u,v}\left|\langle Su, Sv\rangle - \langle u, v\rangle\right| \iff C_{\mathcal{A}}(\xi) = \sup_{A\in\mathcal{A}}\left|\, \|A\xi\|_2^2 - \mathbb{E}\|A\xi\|_2^2 \,\right| = \sup_{A\in\mathcal{A}}\left|\, \xi^\top A^\top A\xi - \mathbb{E}\,\xi^\top A^\top A\xi \,\right|.$$

Hanson–Wright *(fixed matrix A)*

For sub-Gaussian ξ with i.i.d. entries,

$$\mathbb{P}\left(\left|\, \|A\xi\|_2^2 - \mathbb{E}\|A\xi\|_2^2 \,\right| \geq \epsilon\right) \lesssim \exp\left(-c\min\left(\frac{\epsilon^2}{\|A^\top A\|_F^2}, \frac{\epsilon}{\|A^\top A\|_{2\to 2}}\right)\right).$$

Krahmer–Mendelson–Rauhut '14 *(uniform over a set $\mathcal{A}$)*

Let $C_{\mathcal{A}}(\xi) = \sup_{A\in\mathcal{A}}\left|\, \|A\xi\|_2^2 - \mathbb{E}\|A\xi\|_2^2 \,\right|$. Define

$$W = \gamma_2(\mathcal{A})\big(\gamma_2(\mathcal{A}) + d_F(\mathcal{A})\big) + d_F(\mathcal{A})d_{2\to 2}(\mathcal{A}), \quad V = d_{2\to 2}(\mathcal{A})\big(\gamma_2(\mathcal{A}) + d_F(\mathcal{A})\big), \quad U = d_{2\to 2}^2(\mathcal{A}).$$

$$\mathbb{P}\left(C_{\mathcal{A}}(\xi) \geq c_1 W + \epsilon\right) \leq 2\exp\left(-c_2\min\left(\frac{\epsilon^2}{V^2}, \frac{\epsilon}{U}\right)\right).$$

— main result —

A bilinear Hanson–Wright for double sets

Theorem 3.1

Let $\mathcal{M} \subset \mathbb{R}^{m\times n}$, $\mathcal{N} \subset \mathbb{R}^{n\times m}$, ξ sub-Gaussian with i.i.d. entries, and

$$C_{\mathcal{M},\mathcal{N}}(\xi) = \sup_{M\in\mathcal{M},\, N\in\mathcal{N}} \left|\xi^\top M^\top N\xi - \mathbb{E}\,\xi^\top M^\top N\xi\right|.$$

$$\begin{aligned} W &= \gamma_2(\mathcal{M})\cdot\left[\gamma_2(\mathcal{N}) + d_F(\mathcal{N})\right] \;+\; \gamma_2(\mathcal{N})\cdot\left[\gamma_2(\mathcal{M}) + d_F(\mathcal{M})\right], \\ V &= \gamma_2(\mathcal{M})\, d_{2\to2}(\mathcal{N}) + \gamma_2(\mathcal{N})\, d_{2\to2}(\mathcal{M}) \;+\; \min\left\{d_F(\mathcal{M})\, d_{2\to2}(\mathcal{N}),\, d_F(\mathcal{N})\, d_{2\to2}(\mathcal{M})\right\}, \\ U &= d_{2\to2}(\mathcal{M})\, d_{2\to2}(\mathcal{N}). \end{aligned}$$

$$\mathbb{P}\left(C_{\mathcal{M},\mathcal{N}}(\xi) \geq c_1 W + \epsilon\right) \leq 2\exp\left(-c_2 \min\left(\frac{\epsilon^2}{V^2},\, \frac{\epsilon}{U}\right)\right).$$

Take-aways. *(i)* Setting $\mathcal{M} = \mathcal{N}$ recovers Krahmer et al. *(ii)* Complexities of $\mathcal{M}, \mathcal{N}$ enter *separately* — no joint set $\{M^\top N\}$. *(iii)* If one set is small, the bound is small.

— proof sketch —

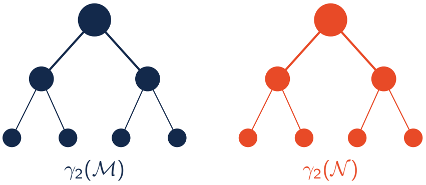
Key technique: *double chaining*

Chain twice. Two admissible sequences $\{T_r(\mathcal{M})\}$, $\{T_r(\mathcal{N})\}$. Each chain pairs its increments with the worst case of the other set.

Double-tree lemma (Thm. 3.2)

If $\{X_{(u,v)}\}$ is sub-Gaussian-Lipschitz in each coordinate,

$$\sup_{u\in\mathcal{U},\, v\in\mathcal{V}} |X_{(u,v)}| \;\lesssim\; C_u\,\gamma_{\mathbf{2}}(\mathcal{U},d) + C_v\,\gamma_{\mathbf{2}}(\mathcal{V},d).$$



bound $\le\; C\cdot\big(\gamma_2(\mathcal{M})+\gamma_2(\mathcal{N})\big)$

What the new bound buys us

01 Johnson–Lindenstrauss

Classical lemma drops out as a special case of our bilinear bound.

03 Sketched federated learning

Reuse a single sketching matrix across rounds. Convergence depends on gaussian width of gradient & loss-Hessian eigenvector sets.

02 RIP for inner products

For arbitrary $\mathcal{U}, \mathcal{V}$ on the sphere, RIP holds for

$$n \gtrsim \frac{1}{\epsilon^2}\left(\omega(\mathcal{U}) + \omega(\mathcal{V})\right)^2.$$

Gaussian-width rate; matches $\sqrt{\log|\mathcal{T}|}$ on finite sets.

04 Sketched linear regression

Solve Least Squares in the sketched space. Error bound depends on $\omega(\mathcal{B})$ and $\omega(\mathcal{X})$.

Beyond i.i.d., and a sketched bandit algorithm

Conditionally independent coordinates

Same bounds hold under a relaxed assumption that covers **CountSketch** and **SRHT** (Subsampled Randomized Hadamard Transform).

Sum of T random quadratic forms

$$\sup_{M,N} \left| \sum_{t=1}^{T} \xi_t^\top M^\top N \xi_t - \mathbb{E}[\xi_t^\top M^\top N \xi_t] \right|$$

Scales with extra $\sqrt{T}$. Needed when the sketching matrix changes per round.

Sketched linear bandits

We design compressed versions of linear bandit algorithms, where actions are projected to a lower-dimensional sketched space before applying LinUCB-style exploration.