Vector-valued noise stability

Yeongwoo Hwang, Joe Neeman, Ojas Parekh, Kevin Thompson, John Wright

Borell's inequality

Take two negatively correlated Gaussians:

$$(X, Y) \sim \mathcal{N}\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}, \qquad -1 < \rho < 0.$$

Find $f : \mathbb{R}^n \to [-1, 1]$ minimizing $\mathbb{E} f(X) f(Y)$.

Take two negatively correlated Gaussians:

$$(X, Y) \sim \mathcal{N}\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}, \qquad -1 < \rho < 0.$$

Find $f : \mathbb{R}^n \to [-1, 1]$ minimizing $\mathbb{E} f(X) f(Y)$.

Theorem (Borell '85)
$f(x) = \operatorname{sgn} x_1$ *is optimal (and uniquely so, up to obvious symmetries).*

Borell's inequality

Take two negatively correlated Gaussians:

$$(X, Y) \sim \mathcal{N}\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}, \qquad -1 < \rho < 0.$$

Find $f : \mathbb{R}^n \to [-1, 1]$ minimizing $\mathbb{E} f(X) f(Y)$.

Theorem (Borell '85)
$f(x) = \operatorname{sgn} x_1$ is optimal (and uniquely so, up to obvious symmetries).

$$\mathbb{E} f(X) f(Y) \geq \frac{2}{\pi} \arcsin(\rho).$$

Borell's inequality

Take two negatively correlated Gaussians:

$$(X, Y) \sim \mathcal{N}\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}, \qquad -1 < \rho < 0.$$

Find $f : \mathbb{R}^n \to [-1, 1]$ minimizing $\mathbb{E} f(X) f(Y)$.

Theorem (Borell '85)
$f(x) = \operatorname{sgn} x_1$ is optimal (and uniquely so, up to obvious symmetries).

$$\mathbb{E} f(X) f(Y) \geq \frac{2}{\pi} \arcsin(\rho).$$

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \leq n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Borell's inequality

Take two negatively correlated Gaussians:

$$(X, Y) \sim \mathcal{N}\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}, \qquad -1 < \rho < 0.$$

Find $f : \mathbb{R}^n \to [-1, 1]$ minimizing $\mathbb{E} f(X) f(Y)$.

Theorem (Borell '85)
$f(x) = \operatorname{sgn} x_1$ *is optimal (and uniquely so, up to obvious symmetries).*

$$\mathbb{E} f(X) f(Y) \geq \frac{2}{\pi} \arcsin(\rho).$$

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \leq n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Theorem (Hwang, N., Parekh, Thompson, Wright)
$f(x) = \frac{(x_1, \ldots, x_k)}{\sqrt{x_1^2 + \cdots + x_k^2}}$ *is optimal (and uniquely so, up to obvious symmetries).*

CS interlude: Max-Cut

$G = (V, E)$ a graph, $V = \{1, \ldots, n\}$.

Maximize $\sum_{\{i,j\} \in E} (x_i - x_j)^2$ over $x \in \{\pm 1\}^n$.

$G = (V, E)$ a graph, $V = \{1, \dots, n\}$.

Maximize $\displaystyle\sum_{\{i,j\}\in E} (x_i - x_j)^2$ over $x \in \{\pm 1\}^n$.

Goemans-Williamson:

Maximize $\displaystyle\sum_{\{i,j\}\in E} |y_i - y_j|^2$ over $y \in (S^n)^n$.

Round: take $\theta \in S^n$, set $x_i = \mathbf{sgn}\langle \theta, y_i \rangle$.

CS interlude: Max-Cut

$G = (V, E)$ a graph, $V = \{1, \dots, n\}$.

Maximize $\displaystyle\sum_{\{i,j\}\in E} (x_i - x_j)^2$ over $x \in \{\pm 1\}^n$.

Goemans-Williamson:

Maximize $\displaystyle\sum_{\{i,j\}\in E} |y_i - y_j|^2$ over $y \in (S^n)^n$.

Round: take $\theta \in S^n$, set $x_i = \mathbf{sgn}\langle \theta, y_i \rangle$.

GW has approximation ratio $\displaystyle\inf_{-1<\rho<0} \frac{1 - \frac{2}{\pi}\arcsin(\rho)}{1-\rho} \approx 0.878$.

Theorem
GW-algorithm has approximation ratio of at least

$$\inf_{-1<\rho<0} \frac{1-\frac{2}{\pi}\arcsin(\rho)}{1-\rho} \approx 0.878.$$

Theorem
GW-algorithm has approximation ratio of at least

$$\inf_{-1<\rho<0} \frac{1-\frac{2}{\pi}\arcsin(\rho)}{1-\rho} \approx 0.878.$$

$$\sup_{y\in(S^n)^n} \sum_{\{i,j\}\in E} |y_i - y_j|^2 \geq \sup_{x\in\{\pm 1\}^n} \sum_{\{i,j\}\in E} (x_i - x_j)^2$$

Effect of rounding: $\rho := \langle y_i, y_j \rangle \Rightarrow |y_i - y_j|^2 = 2(1-\rho)$.

$$\mathbb{E}_\theta |\operatorname{sgn}\langle y_i, \theta\rangle - \operatorname{sgn}\langle y_j, \theta\rangle|^2 \to 2(1 - \frac{2}{\pi}\arcsin(\rho)) \text{ as } n \to \infty.$$

Theorem
GW-algorithm has approximation ratio of at least

$$\inf_{-1<\rho<0} \frac{1-\frac{2}{\pi}\arcsin(\rho)}{1-\rho} \approx 0.878.$$

$$\sup_{y\in(S^n)^n} \sum_{\{i,j\}\in E} |y_i - y_j|^2 \geq \sup_{x\in\{\pm 1\}^n} \sum_{\{i,j\}\in E} (x_i - x_j)^2$$

Effect of rounding: $\rho := \langle y_i, y_j\rangle \Rightarrow |y_i - y_j|^2 = 2(1-\rho)$.

$$\mathbb{E}_\theta |\operatorname{sgn}\langle y_i,\theta\rangle - \operatorname{sgn}\langle y_j,\theta\rangle|^2 \to 2(1 - \frac{2}{\pi}\arcsin(\rho)) \text{ as } n \to \infty.$$

Theorem *(Khot-Kindler-Mossel-O'Donnell, Mossel-O'Donnell-Oleszkiewicz)*
Doing better than 0.878 is Unique-Games-Hard.

$G = (V, E)$ a graph, $V = \{1, \dots, n\}$.

h a specific 4×4 complex matrix "anti-ferromagnetic Heisenberg interaction"

$2^n \times 2^n$ matrices:

CS interlude: Quantum Max-Cut

$G = (V, E)$ a graph, $V = \{1, \ldots, n\}$.

h a specific 4×4 complex matrix "anti-ferromagnetic Heisenberg interaction"

$2^n \times 2^n$ matrices: for $v_i, w_i \in \mathbb{C}^2$

$$h_{i,j}\big(\otimes_{k=1}^n v_k, \otimes_{k=1}^n w_k\big) = h(v_i \otimes v_j, w_i \otimes w_j) \prod_{k \neq i,j} \langle v_k, w_k \rangle$$

$$H_G = \sum_{\{i,j\} \in E} h_{i,j} \qquad \text{"anti-ferromagnetic Heisenberg model"}$$

CS interlude: Quantum Max-Cut

$G = (V, E)$ a graph, $V = \{1, \ldots, n\}$.

h a specific 4×4 complex matrix "anti-ferromagnetic Heisenberg interaction"

$2^n \times 2^n$ matrices: for $v_i, w_i \in \mathbb{C}^2$

$$h_{i,j}\left(\otimes_{k=1}^n v_k, \otimes_{k=1}^n w_k\right) = h(v_i \otimes v_j, w_i \otimes w_j) \prod_{k \neq i,j} \langle v_k, w_k \rangle$$

$$H_G = \sum_{\{i,j\} \in E} h_{i,j} \qquad \text{"anti-ferromagnetic Heisenberg model"}$$

Definition

energy of H_G: $\lambda_{\max}(H_G)$

product-state energy of H_G: $\max_{|w_1|,\ldots,|w_n|=1} H_G\left(\otimes_{i=1}^n w_i\right)$

$G = (V, E)$ a graph, $V = \{1, \ldots, n\}$.

h a specific 4×4 complex matrix "anti-ferromagnetic Heisenberg interaction"

$2^n \times 2^n$ matrices: for $v_i, w_i \in \mathbb{C}^2$

$$h_{i,j}\big(\otimes_{k=1}^n v_k, \otimes_{k=1}^n w_k \big) = h(v_i \otimes v_j, w_i \otimes w_j) \prod_{k \neq i,j} \langle v_k, w_k \rangle$$

$$H_G = \sum_{\{i,j\} \in E} h_{i,j} \qquad \text{"anti-ferromagnetic Heisenberg model"}$$

Definition
energy of H_G: $\lambda_{\max}(H_G)$

product-state energy of H_G: $\max_{|w_1|,\ldots,|w_n|=1} H_G \left(\otimes_{i=1}^n w_i\right)$

Theorem (Hwang, N., Parekh, Thompson, Wright)
Approximating these by better than ≈ 0.956 is Unique-Games hard.

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \leq n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Theorem (Hwang, N., Parekh, Thompson, Wright)

$f(x) = \frac{(x_1,\dots,x_k)}{\sqrt{x_1^2+\cdots+x_k^2}}$ *is optimal (and uniquely so, up to obvious symmetries).*

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \leq n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Theorem (Hwang, N., Parekh, Thompson, Wright)
$f(x) = \frac{(x_1,\ldots,x_k)}{\sqrt{x_1^2+\cdots+x_k^2}}$ *is optimal (and uniquely so, up to obvious symmetries).*

Proof.

1. *Dimension reduction:* It suffices to solve it for $f : \mathbb{R}^k \to B_2^k$.

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \le n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Theorem (Hwang, N., Parekh, Thompson, Wright)
$f(x) = \frac{(x_1,\dots,x_k)}{\sqrt{x_1^2+\cdots+x_k^2}}$ *is optimal (and uniquely so, up to obvious symmetries).*

Proof.

1. *Dimension reduction:* It suffices to solve it for $f : \mathbb{R}^k \to B_2^k$.
2. *Shell decomposition:* It suffices to solve it for $f : S^{k-1} \to B_2^k$.

Find $f : \mathbb{R}^n \to B_2^k$ (with $k \le n$) minimizing $\mathbb{E}\langle f(X), f(Y)\rangle$.

Theorem (Hwang, N., Parekh, Thompson, Wright)
$f(x) = \frac{(x_1,\dots,x_k)}{\sqrt{x_1^2+\dots+x_k^2}}$ *is optimal (and uniquely so, up to obvious symmetries).*

Proof.

1. *Dimension reduction:* It suffices to solve it for $f : \mathbb{R}^k \to B_2^k$.
2. *Shell decomposition:* It suffices to solve it for $f : S^{k-1} \to B_2^k$.
3. *Spectral analysis:* Solve it for $f : S^{k-1} \to B_2^k$.

□

Spectral argument

Let $g : [-1, 1] \to \mathbb{R}$ be decreasing. $f : S^{k-1} \to \mathbb{R}^k$.

$$U_g f(u) := \int_{S^{k-1}} f(v) g(\langle u, v \rangle)\, d\sigma^{k-1}(v).$$

$$E f U_g f = E\langle f(X), f(Y) \rangle$$

where (X, Y) have density $g(\langle x, y \rangle)$ w.r.t. $\sigma^{k-1} \times \sigma^{k-1}$.

Spectral argument

Let $g : [-1, 1] \to \mathbb{R}$ be decreasing. $f : S^{k-1} \to \mathbb{R}^k$.

$$U_g f(u) := \int_{S^{k-1}} f(v) g(\langle u, v \rangle)\, d\sigma^{k-1}(v).$$

$$E f U_g f = E\langle f(X), f(Y)\rangle$$

where (X, Y) have density $g(\langle x, y\rangle)$ w.r.t. $\sigma^{k-1} \times \sigma^{k-1}$.

Theorem
The eigenfunctions of U_g with minimal eigenvalue are linear.

Spectral argument

Let $g : [-1, 1] \to \mathbb{R}$ be decreasing. $f : S^{k-1} \to \mathbb{R}^k$.

$$U_g f(u) := \int_{S^{k-1}} f(v) g(\langle u, v \rangle)\, d\sigma^{k-1}(v).$$

$$E f U_g f = E\langle f(X), f(Y) \rangle$$

where (X, Y) have density $g(\langle x, y \rangle)$ w.r.t. $\sigma^{k-1} \times \sigma^{k-1}$.

Theorem
The eigenfunctions of U_g with minimal eigenvalue are linear.

Corollary
Among f with $\mathbb{E}|f|^2 \leq 1$, $\mathbb{E}\langle f, U_g f \rangle$ is minimized by $f(x) = x = \frac{x}{|x|}$.

Let $g : [-1, 1] \to \mathbb{R}$ be decreasing. $f : S^{k-1} \to \mathbb{R}^k$.

$$U_g f(u) := \int_{S^{k-1}} f(v) g(\langle u, v \rangle) \, d\sigma^{k-1}(v).$$

$$E f U_g f = E\langle f(X), f(Y) \rangle$$

where (X, Y) have density $g(\langle x, y \rangle)$ w.r.t. $\sigma^{k-1} \times \sigma^{k-1}$.

Theorem
The eigenfunctions of U_g with minimal eigenvalue are linear.

Corollary
Among f with $\mathbb{E}|f|^2 \leq 1$, $\mathbb{E}\langle f, U_g f \rangle$ is minimized by $f(x) = x = \frac{x}{|x|}$.

Proof.
Shur's lemma implies that eigenfunctions are spherical harmonics, so need to compute $\frac{U_g f(v)}{f(v)}$ for spherical harmonics f. Gegenbauer polynomials + NIST Handbook. □

Shell decomposition

To handle $f : \mathbb{R}^k \to B_2^k$, it suffices to handle $f : S^{k-1} \to B_2^k$.

Shell decomposition

To handle $f : \mathbb{R}^k \to B_2^k$, it suffices to handle $f : S^{k-1} \to B_2^k$.

$$c(r, s) = \mathbb{E}\left[\langle f(X), f(Y)\rangle \,\middle|\, |X| = r, |Y| = s\right]$$
$$e(r) = \mathbb{E}\left[f(X) \,\middle|\, |X| = r\right]$$

Shell decomposition

To handle $f : \mathbb{R}^k \to B_2^k$, it suffices to handle $f : S^{k-1} \to B_2^k$.

$$c(r,s) = \mathbb{E}\left[\langle f(X), f(Y)\rangle \,\middle|\, |X| = r, |Y| = s\right]$$
$$e(r) = \mathbb{E}\left[f(X) \,\middle|\, |X| = r\right]$$

$$\mathbb{E}\langle f(X), f(Y)\rangle = \mathbb{E}c(R,S) \geq \mathbb{E}\left[\langle e(R), e(S)\rangle + \lambda_1^{R,S}\right] \geq \mathbb{E}\lambda_1^{R,S}.$$

Shell decomposition

To handle $f : \mathbb{R}^k \to B_2^k$, it suffices to handle $f : S^{k-1} \to B_2^k$.

$$c(r,s) = \mathbb{E}\left[\langle f(X), f(Y)\rangle \,\middle|\, |X| = r, |Y| = s\right]$$
$$e(r) = \mathbb{E}\left[f(X) \,\middle|\, |X| = r\right]$$

$$\mathbb{E}\langle f(X), f(Y)\rangle = \mathbb{E}c(R,S) \geq \mathbb{E}\left[\langle e(R), e(S)\rangle + \lambda_1^{R,S}\right] \geq \mathbb{E}\lambda_1^{R,S}.$$

Equality if $f(x) = \frac{x}{|x|}$.

To handle $f : \mathbb{R}^n \to B_2^k$, it suffices to handle $f : \mathbb{R}^k \to B_2^k$.

Dimension reduction

To handle $f : \mathbb{R}^n \to B_2^k$, it suffices to handle $f : \mathbb{R}^k \to B_2^k$.

Variational: either $f : \mathbb{R}^n \to B_2^k$ is “effectively k-dimensional” or it can be improved by a perturbation.

Dimension reduction

To handle $f : \mathbb{R}^n \to B_2^k$, it suffices to handle $f : \mathbb{R}^k \to B_2^k$.

Variational: either $f : \mathbb{R}^n \to B_2^k$ is “effectively k-dimensional” or it can be improved by a perturbation.

$$
\begin{aligned}
f_\theta(x) &:= f(x+\theta) \\
e(\theta) &:= \mathbb{E} f_\theta \\
J(\theta) &= \mathbb{E}\langle f_\theta(X), f_\theta(Y)\rangle
\end{aligned}
$$

Lemma

If $D_\theta e(0) = 0$ then $D^2_{\theta,\theta} J(0) \le 0$. Inequality is strict unless $f_{\alpha\theta} = f$ for all $\alpha \in \mathbb{R}$.

Corollary

If f is optimal and $D_\theta e(0) = 0$ then $f_{\alpha\theta} = f$ for all $\alpha \in \mathbb{R}$.

Dimension reduction

To handle $f : \mathbb{R}^n \to B_2^k$, it suffices to handle $f : \mathbb{R}^k \to B_2^k$.

Variational: either $f : \mathbb{R}^n \to B_2^k$ is "effectively k-dimensional" or it can be improved by a perturbation.

$$f_\theta(x) := f(x + \theta)$$
$$e(\theta) := \mathbb{E} f_\theta$$
$$J(\theta) = \mathbb{E}\langle f_\theta(X), f_\theta(Y)\rangle$$

Lemma
If $D_\theta e(0) = 0$ then $D^2_{\theta,\theta} J(0) \le 0$. Inequality is strict unless $f_{\alpha\theta} = f$ for all $\alpha \in \mathbb{R}$.

Corollary
If f is optimal and $D_\theta e(0) = 0$ then $f_{\alpha\theta} = f$ for all $\alpha \in \mathbb{R}$.

Conjecture
If $0 < \rho < 1$ *then* $\mathbb{E}\langle f(X), f(Y)\rangle$ *is maximized, among functions with* $\mathbb{E}f = 0$*, by*
$f(x) = \frac{(x_1,\ldots,x_k)}{\sqrt{x_1^2+\cdots+x_k^2}}$.

Open problems

Conjecture

If $0 < \rho < 1$ *then* $\mathbb{E}\langle f(X), f(Y)\rangle$ *is maximized, among functions with* $\mathbb{E}f = 0$*, by* $f(x) = \frac{(x_1,\ldots,x_k)}{\sqrt{x_1^2+\cdots+x_k^2}}$.

Question

What about if $\mathbb{E}f = \mu \in B_2^k$*?*

Thank you!