The Geometry of Hypothesis Testing over Convex Cones

Yuting Wei
Department of Statistics, UC Berkeley

BIRS workshop on Shape-Constrained Methods
Jan 30rd, 2018

joint work with:

Cone testing problem

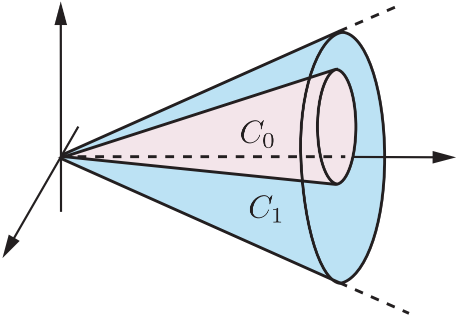
- Observation model: $y = \theta + \sigma w$, $w \sim N(0, \mathbb{I}_d)$

Cone testing problem

- Observation model: $y = \theta + \sigma w$, $w \sim N(0, \mathbb{I}_d)$
- Inference problem: $H_0 : \theta \in C_0$ v.s. $H_1 : \theta \in C_1 \backslash C_0$

Cone testing problem

- Observation model: $y = \theta + \sigma w$, $w \sim N(0, \mathbb{I}_d)$
- Inference problem: $H_0 : \theta \in C_0$ v.s. $H_1 : \theta \in C_1 \backslash C_0$
- $C_0 \subset C_1$ closed, convex cones ($\forall\ x \in C$ then $ax \in C$, $\forall\ a \geq 0$)

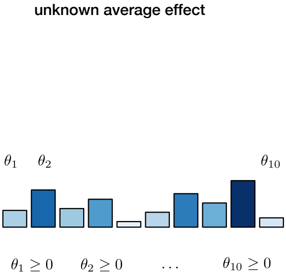


Applications

- In the fields of
  - detection of treatment effects
  - signal detection in radar processing
  - trend detection in econometrics

Example: Treatment effect detection

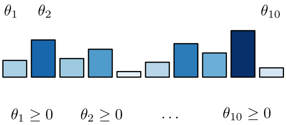
- average treatment effect of $d = 10$ different dosages of a drug
- non-negative and un-ordered space



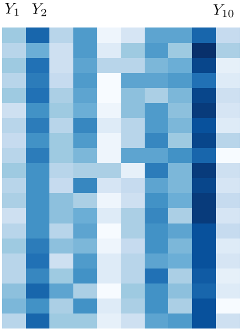
Example: Treatment effect detection

- average treatment effect of $d = 10$ different dosages of a drug
- non-negative and un-ordered space

unknown average effect



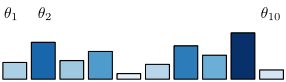
observed outcomes



Example: Treatment effect detection

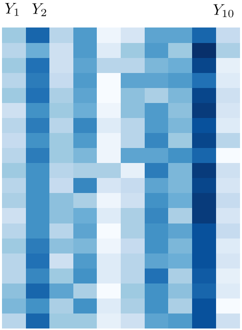
- average treatment effect of $d = 10$ different dosages of a drug
- non-negative and un-ordered space

unknown average effect



$\theta_1 \geq 0$ $\quad \theta_2 \geq 0$ $\quad \ldots$ $\quad \theta_{10} \geq 0$

observed outcomes

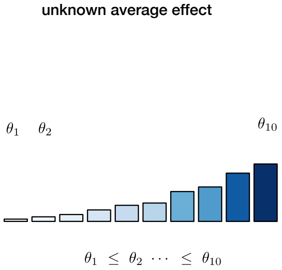


Inference

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \cdots = \theta_{10} = 0$$

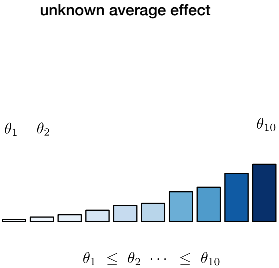
Example: Treatment effect detection

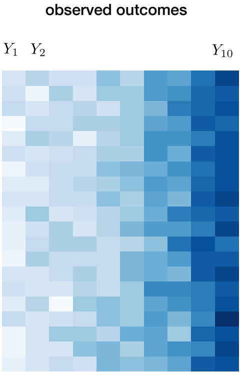
- average treatment effect of $d = 10$ different dosages of a drug
- monotonic space



Example: Treatment effect detection

- average treatment effect of $d = 10$ different dosages of a drug
- monotonic space

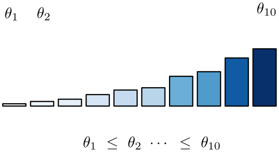




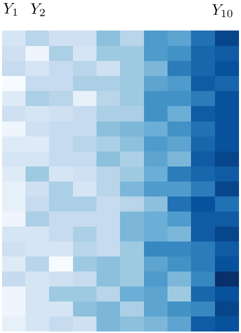
Example: Treatment effect detection

- average treatment effect of $d = 10$ different dosages of a drug
- monotonic space

unknown average effect



observed outcomes



Inference

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \cdots = \theta_{10}$$

Applications

- In the fields of
  - detection of treatment effects
  - signal detection in radar processing
  - trend detection in econometrics

Bühlmann'03, Meinshausen'03, Meyer'03, Sen and Meyer 17'

Applications

- In the fields of
  - detection of treatment effects
  - signal detection in radar processing
  - trend detection in econometrics

- More statistical models
  - linear regression $y = X\beta + w := \theta + w$
    - ★ $C_0 := \{0\}$ v.s. $C_1 := \text{range}(X)$
  - fitness of a linear model: $y = f(x_1^n) + w := \theta + w$,
    - ★ $C_0 := \{X\beta \mid \beta \in \mathbb{R}^k\}$ v.s. $C_1 :=$ convex cone

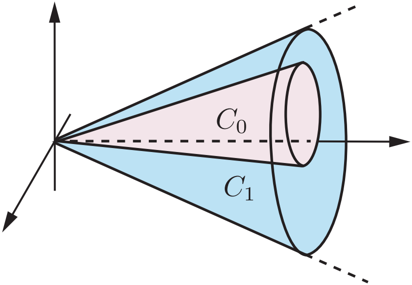
Bühlmann'03, Meinshausen'03, Meyer'03, Sen and Meyer 17'

Overview

Q: How to **solve** these constrained testing problems?

Q: How to **quantify** the hardness of a constrained testing problem?

Q: How does the hardness depend on the geometry?



Generalized Likelihood Ratio Test

- Generalized Likelihood Ratio Test (GLRT)

$$\phi_\beta(y) := \begin{cases} 1 & \text{if } T(y) \geq \beta \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{where} \quad T(y) := -2 \log \left(\frac{\sup_{\theta \in C_0} \mathbb{P}_\theta(y)}{\sup_{\theta \in C_1} \mathbb{P}_\theta(y)} \right).$$

Robertson and Wegman'78, Raubertas et al.'86, Fan et al.'01...

Generalized Likelihood Ratio Test

- Generalized Likelihood Ratio Test (GLRT)

$$\phi_\beta(y) := \begin{cases} 1 & \text{if } T(y) \geq \beta \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{where} \quad T(y) := -2 \log \left(\frac{\sup_{\theta \in C_0} \mathbb{P}_\theta(y)}{\sup_{\theta \in C_1} \mathbb{P}_\theta(y)} \right).$$

- Cone based GLRT:

$$\begin{aligned} T(y) &= \min_{\theta \in C_0} \|y - \theta\|_2^2 - \min_{\theta \in C_1} \|y - \theta\|_2^2 \\ &= \|\Pi_{C_1}(y)\|_2^2 - \|\Pi_{C_0}(y)\|_2^2 \end{aligned}$$

Robertson and Wegman'78, Raubertas et al.'86, Fan et al.'01...

Related results

- GLRT
  - different aspects of GLRT e.g. Wilks phenomenon, fitness of parametric model... [Warrack and Robertson'84, Menéndez'92, Lehmann'06, Perlman and Wu'99, Barlow 72, Lehamnn and Romano'06, Fan et al.'01, '07]

Related results

- GLRT
  - different aspects of GLRT e.g. Wilks phenomenon, fitness of parametric model... [Warrack and Robertson'84, Menéndez'92, Lehmann'06, Perlman and Wu'99, Barlow 72, Lehamnn and Romano'06, Fan et al.'01, '07]
- Cone testing
  - Raubertas, Lee & Nordheim studies the null distribution of GLRT
  - Hu and Wright characterizes GLRT equivalences for various pairs of cones

Related results

- GLRT
  - different aspects of GLRT e.g. Wilks phenomenon, fitness of parametric model... [Warrack and Robertson'84, Menéndez'92, Lehmann'06, Perlman and Wu'99, Barlow 72, Lehamnn and Romano'06, Fan et al.'01, '07]
- Cone testing
  - Raubertas, Lee & Nordheim studies the null distribution of GLRT
  - Hu and Wright characterizes GLRT equivalences for various pairs of cones
  - Dümbgen'95, Robertson'78, Warrack and Robertson'84, Brown'86, Cohan and Sackrowitz'96, Rubertas et al'86, Menéndez'91

Related results

- GLRT
  - different aspects of GLRT e.g. Wilks phenomenon, fitness of parametric model... [Warrack and Robertson'84, Menéndez'92, Lehmann'06, Perlman and Wu'99, Barlow 72, Lehamnn and Romano'06, Fan et al.'01, '07]
- Cone testing
  - Raubertas, Lee & Nordheim studies the null distribution of GLRT
  - Hu and Wright characterizes GLRT equivalences for various pairs of cones
  - Dümbgen'95, Robertson'78, Warrack and Robertson'84, Brown'86, Cohan and Sackrowitz'96, Rubertas et al'86, Menéndez'91
- Minimax testing framework
  - introduced in the seminal work of Ingster and co-authors
  - different from the Neyman-Person testing framework

Testing radius

- Test $\theta \in C_0$ v.s. $C_1 \backslash C_0$
- Uniform error for test ψ:

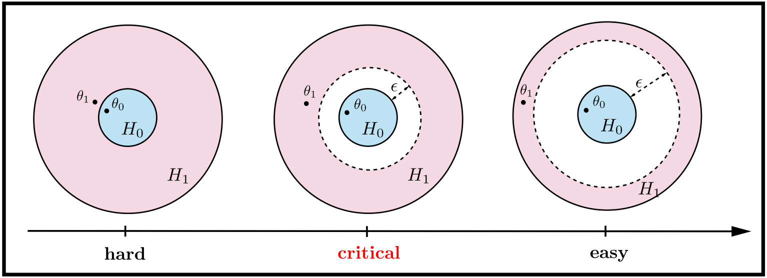
$$\text{error}(\psi,\ \epsilon) := \underbrace{\sup_{\theta \in C_0} \mathbb{E}_\theta[\psi(y)]}_{\text{type I error}} + \underbrace{\sup_{\theta \in C_1 \backslash B_2(\epsilon; C_0)} \mathbb{E}_\theta[1 - \psi(y)]}_{\text{type II error}},$$

Testing radius

- Test $\theta \in C_0$ v.s. $C_1 \backslash C_0$
- Uniform error for test ψ:

$$\text{error}(\psi,\ \epsilon) := \underbrace{\sup_{\theta \in C_0} \mathbb{E}_\theta[\psi(y)]}_{\text{type I error}} + \underbrace{\sup_{\theta \in C_1 \backslash B_2(\epsilon; C_0)} \mathbb{E}_\theta[1 - \psi(y)]}_{\text{type II error}},$$

- **Testing radius** $\epsilon_\Psi \equiv$ distance at which null/alternative are "just distinguishable" using class of tests Ψ



Minimax optimal testing radius

- Uniform error for test ψ:

$$\text{error}(\psi,\ \epsilon) := \underbrace{\sup_{\theta \in C_0} \mathbb{E}_\theta[\psi(y)]}_{\text{type I error}} + \underbrace{\sup_{\theta \in C_1 \backslash B_2(\epsilon; C_0)} \mathbb{E}_\theta[1 - \psi(y)]}_{\text{type II error}},$$

Minimax optimal testing radius

- Uniform error for test ψ:

$$\text{error}(\psi,\ \epsilon) := \underbrace{\sup_{\theta \in C_0} \mathbb{E}_\theta[\psi(y)]}_{\text{type I error}} + \underbrace{\sup_{\theta \in C_1 \setminus B_2(\epsilon; C_0)} \mathbb{E}_\theta[1 - \psi(y)]}_{\text{type II error}},$$

- Minimax testing radius:

$$\epsilon_{\text{OPT}}(\rho) := \inf \left\{ \epsilon \mid \inf_{\psi} \text{error}(\psi,\ \epsilon) \le \rho \right\}$$

[Ingster and Suslina'12, Ermakov'91, Lepski and Spokoiny'99, Lepski and Tsybakov'00]

Minimax optimal testing radius

- Uniform error for test ψ:

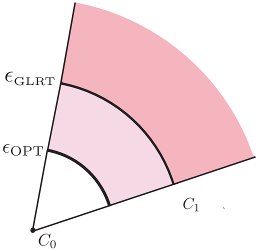
$$\text{error}(\psi,\ \epsilon) := \underbrace{\sup_{\theta \in C_0} \mathbb{E}_\theta[\psi(y)]}_{\text{type I error}} + \underbrace{\sup_{\theta \in C_1 \setminus B_2(\epsilon; C_0)} \mathbb{E}_\theta[1 - \psi(y)]}_{\text{type II error}},$$

- Minimax testing radius:

$$\epsilon_{\text{OPT}}(\rho) := \inf\left\{\epsilon \mid \inf_{\psi} \text{error}(\psi,\ \epsilon) \le \rho\right\}$$

[Ingster and Suslina'12, Ermakov'91, Lepski and Spokoiny'99, Lepski and Tsybakov'00]

- critical testing radius: ϵ_{GLRT} and ϵ_{OPT}



Main results

Theorem (W, Wainwright & Guntuboyina '17)

The GLRT testing radius satisfies

$$\epsilon^2_{GLRT} \asymp \sigma^2 \min\left\{\textit{width term},\ \textit{geometric term}\right\}.$$

Main results

Theorem (W, Wainwright & Guntuboyina '17)

The GLRT testing radius satisfies

$$\epsilon^2_{GLRT} \asymp \sigma^2 \min\left\{ \underbrace{\mathbb{E}\|\Pi_C w\|_2}_{\textit{width term}},\ \underbrace{\left(\frac{\mathbb{E}\|\Pi_C w\|_2}{\inf\limits_{\eta \in C \cap S} \langle \eta, \mathbb{E}\Pi_C w\rangle}\right)^2}_{\textit{geometric term}} \right\}.$$

- $C_0 = 0,\ \ C_1 = C$
- $w \sim N(0, \mathbb{I}_d)$
- $\Pi_C w := \arg\min_{u \in C} \|g - u\|_2$
- $S \rightarrow$ unit sphere in $\mathbb{R}^d$

supp

Main results

Theorem (W, Wainwright & Guntuboyina '17)

The GLRT testing radius satisfies

$$\epsilon^2_{GLRT} \asymp \sigma^2 \min\left\{ \underbrace{\mathbb{E}\|\Pi_C w\|_2}_{\textit{width term}},\ \underbrace{\left(\frac{\mathbb{E}\|\Pi_C w\|_2}{\inf_{\eta \in C \cap S} \langle \eta,\, \mathbb{E}\Pi_C w \rangle} \right)^2}_{\textit{geometric term}} \right\}.$$

- Width term $= \mathbb{E}\|\Pi_C w\|_2$ where $\Pi_C w := \arg\min_{u \in C} \|w - u\|_2$
- $\mathbb{E}\|\Pi_C(w)\|_2 \;=\; \mathbb{W}(C \cap S)$ where

$$\text{Gaussian width: } \mathbb{W}(A) := \mathbb{E}\Big[\sup_{u \in A} \langle u,\, w \rangle \Big]$$

Main results

Theorem (W, Wainwright & Guntuboyina '17)

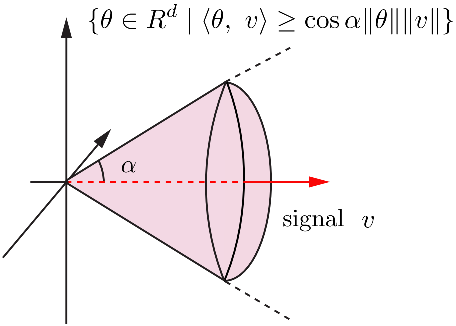
The GLRT testing radius satisfies

$$\epsilon_{GLRT}^2 \asymp \sigma^2 \min\left\{ \underbrace{\mathbb{E}\|\Pi_C w\|_2}_{\textit{width term}},\ \underbrace{\left(\frac{\mathbb{E}\|\Pi_C w\|_2}{\inf_{\eta \in C \cap S} \langle \eta,\, \mathbb{E}\Pi_C w \rangle} \right)^2}_{\textit{geometric term}} \right\}.$$

- Examples:

cone C	width term	geometric term
k-dimensional space	$\sqrt{k}$	∞
non-negative orthant	$\sqrt{d}$	d
monotone cone	$\sqrt{\log d}$	∞
ice cream cone	$\sqrt{d}$	1

Ice Cream cone



Main results

Theorem (W, Wainwright & Guntuboyina '17)

The GLRT testing radius satisfies

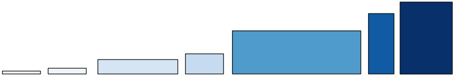
$$\epsilon_{GLRT}^2 \asymp \sigma^2 \min\Big\{ \underbrace{\mathbb{E}\|\Pi_C w\|_2}_{\textit{width term}},\ \underbrace{\Big(\frac{\mathbb{E}\|\Pi_C w\|_2}{\inf_{\eta \in C \cap S} \langle \eta,\ \mathbb{E}\Pi_C w \rangle} \Big)^2}_{\textit{geometric term}} \Big\}.$$

- Examples:

cone C	width term	geometric term
k-dimensional space	$\sqrt{k}$	∞
non-negative orthant	$\sqrt{d}$	d
monotone cone	$\sqrt{\log d}$	∞
ice cream cone	$\sqrt{d}$	1

Consequence for monotone cone

- $\mathcal{H}_0 : \theta = \theta_0, \quad$ versus $\quad \mathcal{H}_1 : \theta \in \mathcal{M}$ (monotone increasing sequence)
- θ_0 is k piece-wise constant



- optimal testing radius satisfies

$$\epsilon^2_{\mathrm{OPT}}(\theta_0, \mathcal{M}; \rho) \lesssim \sigma^2 \sqrt{k(\theta_0) \log\left(\frac{d}{k(\theta_0)}\right)}$$

idea:

$$\mathcal{H}_0 : \theta = 0, \quad \text{versus} \quad \mathcal{H}_1 : \theta \in \mathcal{T}_{\mathcal{M}}(\theta_0)$$

tangent cone: $\mathcal{T}_{\mathcal{M}}(\theta_0) := \{u \in \mathbb{R}^d \mid \exists\ t > 0 \text{ such that } \theta_0 + tu \in \mathcal{M}\}$.

Minimax lower bound

Theorem (W, Wainwright & Guntuboyina '17)

The minimax testing radius satisfies

$$\epsilon_{OPT}^2 \gtrsim \sigma^2 \min\left\{ \underbrace{\mathbb{E}\|\Pi_C w\|_2}_{\text{width term}}, \underbrace{\left(\frac{\mathbb{E}\|\Pi_C w\|_2}{\sup_{\eta \in C \cap S} \langle \eta, \mathbb{E}\Pi_C w \rangle} \right)^2}_{\text{geometric term II}} \right\}.$$

- geometric term $= \left(\frac{\mathbb{E}\|\Pi_C w\|_2}{\inf_{\eta \in C \cap S} \langle \eta, \mathbb{E}\Pi_C w \rangle} \right)^2$
- GLRT is optimal whenever

$$\inf_{\eta \in C \cap S} \langle \eta, \mathbb{E}\Pi_C w \rangle \asymp \sup_{\eta \in C \cap S} \langle \eta, \mathbb{E}\Pi_C w \rangle$$

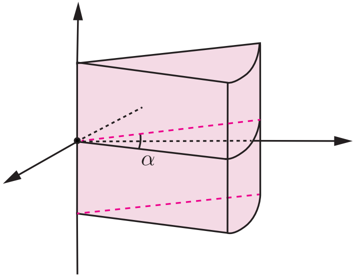
Optimal hypothesis testing

- GLRT is minimax optimal (up to constants) in all these cases,

cone C	ϵ^2_{GLRT}
k-dimensional space	$\sigma^2\sqrt{k}$
non-negative orthant	$\sigma^2\sqrt{d}$
monotone cone	$\sigma^2\sqrt{\log d}$
ice cream cone	$\sigma^2 \cdot 1$

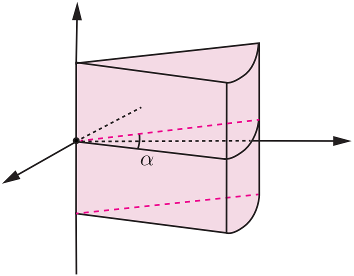
Is GLRT always optimal?

- Consider a Cartesian product cone: Ice Cream$_{d-1}(\alpha) \times \mathbb{R}$



Is GLRT always optimal?

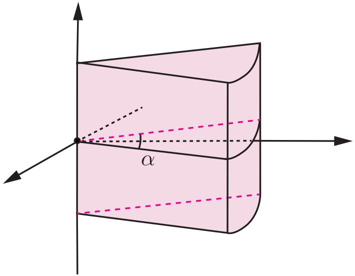
- Consider a Cartesian product cone: Ice Cream$_{d-1}(\alpha) \times \mathbb{R}$



- $\epsilon^2_{\mathrm{GLRT}} \asymp \sigma^2 \min\{\sqrt{d},\ \infty\}$
- $\epsilon^2_{\mathrm{OPT}} \asymp \sigma^2 \min\{\sqrt{d},\ 1\}$

Is GLRT always optimal?

- Consider a Cartesian product cone: $\text{Ice Cream}_{d-1}(\alpha) \times \mathbb{R}$



  - $\epsilon^2_{\text{GLRT}} \asymp \sigma^2 \min\{\sqrt{d}, \infty\}$
  - $\epsilon^2_{\text{OPT}} \asymp \sigma^2 \min\{\sqrt{d}, 1\}$

- a simple test based on $\|(Y_1, Y_d)\|_2 \rightarrow$ minimax optimal

supp

Summary

- show difficulty of testing depends on geometry
  (very different from estimation)

Summary

- show difficulty of testing depends on geometry
  (very different from estimation)

- interesting consequence for testing in monotone cone

Summary

- show difficulty of testing depends on geometry
  (very different from estimation)
- interesting consequence for testing in monotone cone
- the GLRT is NOT always optimal, and can be very poor even for simple problems

References

- **Y. Wei**, M. J. Wainwright, and A. Guntuboyina. (2017) The geometry of hypothesis testing over convex cones: Generalized likelihood tests and minimax radii. *Under revision by the Annals of Statistics*.

Thanks! Questions?

Supplementary: general cones

- cone pairs (C_0, C_1) is said to be *non-oblique* if

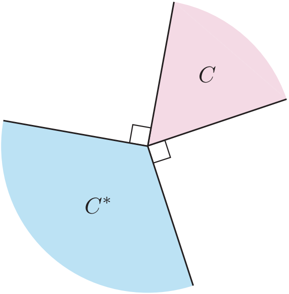
$$\Pi_{C_0}(x) = \Pi_{C_0}(\Pi_{C_1}(x)) \qquad \text{for all } x \in \mathbb{R}^d.$$

[Warrack et al.'84, Menéndez'92a,92b, Hu and Wright'94]

- $C = C_0^* \cap C_1$

- polar cone of any C:

$$C^* := \left\{ v \in \mathbb{R}^d \mid \langle v, u \rangle \leq 0 \quad \text{for all } u \in C \right\}$$



back

Supplementary: sub-optimality of GLRT

- Consider a Cartesian product cone: Ice $\text{Cream}_{d-1}(\alpha) \times \mathbb{R}$
- Consider when signal lies only in the last coordinate: $(0, \ldots, 0, \theta_d)$
- Difference on GLRT under the null and the alternative:

$$
\begin{aligned}
&\|\Pi_C(w_1, w_2, \ldots, w_d)\|_2 - \|\Pi_C(w_1, w_2, \ldots, \theta_d + w_d)\|_2 \\
=&\|(\text{proj to ice cream},\ w_d)\|_2 - \|(\text{proj to ice cream},\ \theta_d + w_d)\|_2
\end{aligned}
$$

where proj to ice cream $= \Pi_{IC}(w_1, \ldots, w_{d-1})$

- $|\theta_d| \geq \|\text{proj to ice cream}\|_2$

back