Torus

Friday, October 8, 2021 7:19 PM

Computational problems for torus actions

$T_n(\mathbb{C}) := \{ \operatorname{diag}(t_1, \ldots, t_n) : t_i \in \mathbb{C}^\times \}$

Any (rational) group action of T_n can be described using a set weights $\Omega \subseteq \mathbb{Z}^n$. $|\Omega| = m$

V = (Laurent) polynomials with exponents from Ω

$$= \left\{ p(x_1, \ldots x_n) := \sum_{\omega \in \Omega} p_\omega \prod_{i=1}^n x_i^{\omega_i} , \ p_\omega\text{'s} \in \mathbb{C} \right\}$$

$\dim(V) = m$

T_n acts on V by scaling the variables

$$((t_1, \ldots t_n) \cdot p)(x_1, \ldots x_n) := p(t_1 x_1, \ldots, t_n x_n)$$

$$= \sum_{\omega \in \Omega} p_\omega \prod_{i=1}^n t_i^{\omega_i} \prod_{i=1}^n x_i^{\omega_i}$$

The coeff p_ω is sent to $p_\omega \prod_{i=1}^n t_i^{\omega_i}$

Question: What is the null cone for this group action?

Theorem: p is in the null cone for the above action of T_n iff $\exists \ a \in \mathbb{Z}^n$ s.t. $\langle a, \omega \rangle > 0 \ \forall \ \omega \in \operatorname{supp}(p)$.

(Hilbert - Mumford criterion for torus actions)

$\operatorname{supp}(p) := \{ \omega \in \Omega : p_\omega \neq 0 \}$

Proof: First suppose $\exists\ a \in \mathbb{Z}^n$ s.t. $\langle a, \omega \rangle > 0\ \forall\ \omega \in \mathrm{supp}(p)$

consider $(t^{a_1}, t^{a_2}, \dots, t^{a_n})$

$$(t^{a_1}, \dots, t^{a_n}) \cdot p = \sum_{\omega \in \Omega} p_\omega \prod_{i=1}^{n} t^{a_i \omega_i} \prod_i x_i^{\omega_i}$$

$$= \sum_{\omega \in \Omega} p_\omega\, t^{\langle a, \omega \rangle} \prod_i x_i^{\omega_i}$$

since $\langle a, \omega \rangle > 0\ \forall\ \omega \in \mathrm{Supp}(p)$, hence

$$\lim_{t \to 0} (t^{a_1}, \dots, t^{a_n}) \cdot p = 0$$

suppose p is in the null cone. want to show $\exists\ a \in \mathbb{Z}^n$ s.t. $\langle a, \omega \rangle > 0\ \forall\ \omega \in \mathrm{supp}(p)$.

$\exists$ a sequence $t^{(1)}, \dots, t^{(k)}, \dots$ s.t. $\lim_{k \to \infty} t^{(k)} \cdot p = 0$

$t^{(k)} := (t_1^{(k)}, \dots, t_n^{(k)})$

$\|p\|^2 = \sum_\omega |p_\omega|^2$, $\|t^{(k)} \cdot p\|_2^2$

$$= \sum_\omega |p_\omega|^2 \prod_{i=1}^{n} |t_i^{(k)}|^{2\omega_i}$$

let $x_i^{(k)} = \ln(|t_i^{(k)}|^2)$

$$= \sum_\omega |p_\omega|^2 \prod_{i=1}^{n} e^{x_i^{(k)} \omega_i}$$

$$= \sum_{\omega} |p_\omega| \prod_{i=1} e$$

$$= \sum_{\omega} |p_\omega|^2 \, e^{\langle x^{(k)}, \omega \rangle}$$

$$\lim_{k \to \infty} \left(\sum_{\omega} |p_\omega|^2 \, e^{\langle x^{(k)}, \omega \rangle} \right) = 0$$

let $\alpha := \min_{\omega \in \mathrm{supp}(p)} |p_\omega|^2$

& choose a k s.t. $\sum_{\omega} |p_\omega|^2 \, e^{\langle x^{(k)}, \omega \rangle} < \alpha$

$$\Rightarrow \quad e^{\langle x^{(k)}, \omega \rangle} < 1 \quad \forall\, \omega \in \mathrm{supp}(p)$$

$$\Rightarrow \quad \langle x^{(k)}, \omega \rangle < 0 \quad \forall\, \omega \in \mathrm{supp}(p)$$

$b := -x^{(k)}$

$\hookrightarrow \in \mathbb{R}^n$

<u>exercise</u>: if $\exists\, b \in \mathbb{R}^n$ s.t. $\langle b, \omega \rangle > 0 \ \forall\, \omega$

$\Rightarrow \exists\, a \in \mathbb{Z}^n$ s.t. $\langle a, \omega \rangle > 0 \ \forall\, \omega$

<u>Question</u>: Can we computationally decide if p is in the null cone?

Input description: describe the weights ω's $\in \Omega$ in

Input description: describe the weights ω's $\in \Omega$ in binary representation. (suppose each coordinate takes b bits)

also describe $\mathrm{supp}(p)$ explicitly.

Theorem: $\exists$ $\mathrm{poly}(n,m,b)$ time algorithm to test if p is in the null cone.

(Linear programming)

Open question: Is this true for actions of $GL_n(\mathbb{C})$?

Recent work [Bürgisser, Doğan, Makam, Walter, Wigderson 21] extends poly time algorithms to orbit equality, orbit-closure intersection, orbit-closure containment for torus actions.

General scaling problem

Question: If p is not in the null cone, can we compute an approx of the minimum norm element in the orbit-closure.

$$\inf_{\substack{(t_1,\dots,t_n) \\ t_i \in \mathbb{C}^\times}} \sum_{\omega \in \Omega} |p_\omega|^2 \prod_{i=1}^{n} |t_i|^{2\omega_i}$$

$(t_1, \dots, t_n)$

$t_i \in \mathbb{C}^\times$

substitute $\quad x_i = \ln(|t_i|^2)$

$$\inf_{x \in \mathbb{R}^n} \sum_{\omega \in \Omega} |p_\omega|^2 \, e^{\langle x, \omega \rangle}$$

convex optimization

$$F_p(x) := \ln\left(\sum_{\omega \in \Omega} |p_\omega|^2 \, e^{\langle x, \omega \rangle} \right)$$

$$\nabla F_p(x) = \frac{\sum_{\omega \in \Omega} |p_\omega|^2 \, e^{\langle x, \omega \rangle} \, \omega}{\sum_{\omega' \in \Omega} |p_{\omega'}|^2 \, e^{\langle x, \omega' \rangle}}$$

$$\left(\nabla F_p(0) = \frac{\sum_{\omega \in \Omega} |p_\omega|^2 \, \omega}{\sum_{\omega' \in \Omega} |p_{\omega'}|^2} \right)$$

moment map

General scaling problem: Given p, $\varepsilon > 0$, find x

s.t. $\quad \| \nabla F_p(x) \|_2^2 \leq \varepsilon$.

Exercise: Look at the action of $ST_n \times ST_n$ on M_n

Exercise: Look at the action of $ST_n \times ST_n$ on M_n

$$(A, B) . M = AMB$$

Then the above framework leads to matrix scaling.

Algorithms

we'll analyze gradient descent for the general scaling problem

$$F_p(x) = \ln\Big(\sum_{\omega} |p_\omega|^2 e^{\langle x, \omega \rangle} \Big) \qquad \Big(\alpha_\omega := |p_\omega|^2$$

$$\downarrow$$

$$\ln\Big(\sum_{\omega} \alpha_\omega e^{\langle x, \omega \rangle} \Big)$$

Gradient descent: start with $x_0 = 0$

$$x_{t+1} := x_t - \eta \nabla F_p(x_t)$$

Theorem: Suppose p is not in the null cone.

$$M := \log \left(\frac{\sum_{\omega} \alpha_\omega}{\min_{\omega \in supp(p)} \alpha_\omega} \right)$$

$$N_\Omega = \max_{\omega \in \Omega} \|\omega\|_2^2$$

$\Rightarrow$ in $O\left(\frac{M N_\Omega}{\varepsilon}\right)$ iterations, gradient descent will output an x s.t. $\|\nabla F_p(x)\|_2^2 \leq \varepsilon$.

Drawback: $O(1/\varepsilon)$ dependence on error ε.

Proposition: $\underset{\text{convexity}}{0 \preceq} \nabla^2 F_p(x) \underset{\text{smoothness}}{\preceq} 2 N_\Omega I_n$

Proof: suffices to prove that $\forall$ unit vectors $u \in \mathbb{R}^n$

$$0 \leq \langle u, \nabla^2 F_p(x) u \rangle \leq 2 N_\Omega$$

$$\langle u, \nabla^2 F_p(x) u \rangle = \frac{d^2}{dt^2} F_p(x+tu) \Big|_{t=0}$$

$$F_p(x+tu) = \ln\Big(\sum_\omega \alpha_\omega e^{\langle x+tu, \omega \rangle} \Big)$$

$$= \underbrace{\ln\Big(\sum_\omega \beta_\omega e^{t\langle u, \omega\rangle} \Big)}_{g(t)}, \quad \beta_\omega := \alpha_\omega e^{\langle x, \omega \rangle}$$

$$g'(t) = \frac{\sum_\omega \beta_\omega e^{t\langle u,\omega\rangle} \langle u, \omega \rangle}{\sum_{\omega'} \beta_{\omega'} e^{t\langle u, \omega' \rangle}}$$

$$g''(t) = \frac{\sum_\omega \beta_\omega e^{t\langle u,\omega\rangle} \langle u, \omega\rangle^2}{\quad}$$

$$g''(t) = \frac{\sum_{\omega} \beta_\omega \, e \quad \langle u,\omega\rangle}{\sum_{\omega'} \beta_{\omega'} e^{t\langle u,\omega'\rangle}} - \frac{\left(\sum_{\omega\in\Omega} \beta_\omega e^{t\langle u,\omega\rangle}\langle u,\omega\rangle\right)\cdot\left(\sum_{\omega''} \beta_{\omega''} e^{t\langle u,\omega''\rangle}\langle u,\omega''\rangle\right)}{\left(\sum_{\omega'} \beta_{\omega'} e^{t\langle u,\omega'\rangle}\right)^2}$$

$$g''(0) = \frac{\sum_{\omega} \beta_\omega \langle u,\omega\rangle^2}{\sum_{\omega'} \beta_{\omega'}} - \frac{\left(\sum_{\omega\in\Omega} \beta_\omega \langle u,\omega\rangle\right)\left(\sum_{\omega''} \beta_{\omega''}\langle u,\omega''\rangle\right)}{\left(\sum_{\omega'} \beta_{\omega'}\right)^2}$$

$$\gamma_\omega := \frac{\beta_\omega}{\sum_{\omega'} \beta_{\omega'}}$$

$$g''(0) = \sum_{\omega} \gamma_\omega \langle u,\omega\rangle^2 - \left(\sum_{\omega\in\Omega} \gamma_\omega \langle u,\omega\rangle\right)^2$$

$\sum_{\omega} \gamma_\omega = 1 \Rightarrow$ Jensen's implies $g''(0) \geq 0$

$$g''(0) \leq 2 \max_{\omega\in\Omega} \|\omega\|_2^2$$

Lemma: Suppose f is β-smooth i.e. $\nabla^2 f(x) \preceq \beta I$. Then gradient descent with step size $\eta = \frac{1}{\beta}$, returns an x s.t. $\|\nabla f(x)\|_2^2 \leq \varepsilon$, in $O\left(\beta \frac{(f(x_0) - f^*)}{\varepsilon}\right)$

returns an x s.t. $\|\nabla f(x)\|_2^2 \leq \varepsilon$, in $O\left(\beta \frac{(f(x_0) - f^*)}{\varepsilon}\right)$ iterations.

Proof: $$x_{t+1} := x_t - \frac{1}{\beta} \nabla f(x_t)$$

$\Big($ if f is β-smooth, $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|_2^2 \Big)$

exercise

$$\begin{aligned} f(x_{t+1}) &\leq f(x_t) + \langle \nabla f(x_t), x_{t+1} - x_t \rangle + \frac{\beta}{2} \|x_{t+1} - x_t\|_2^2 \\ &= f(x_t) - \frac{1}{\beta} \|\nabla f(x_t)\|_2^2 + \frac{1}{2\beta} \|\nabla f(x_t)\|_2^2 \\ &= f(x_t) - \frac{1}{2\beta} \|\nabla f(x_t)\|_2^2 \end{aligned}$$

if $\|\nabla f(x_t)\|_2^2 \geq \varepsilon$

$\Rightarrow \quad f(x_{t+1}) - f(x_t) \leq -\frac{\varepsilon}{2\beta}$

if $\|\nabla f(x_t)\|_2^2 \geq \varepsilon \quad \forall \; t \in \{0, \ldots T-1\}$

$$\Rightarrow \quad f(x_T) - f(x_0) \leq -\frac{\varepsilon T}{2\beta}$$

$$\Rightarrow \quad T \leq \frac{2\beta \left(f(x_0) - f(x_T)\right)}{\varepsilon} \leq \frac{2\beta \left(f(x_0) - f^*\right)}{\varepsilon}$$

Motivation for $\mathrm{polylog}(1/\varepsilon)$ kind of algorithms

Suppose we want to test if p is in the null cone.

Suppose we want to test if p is in the null cone.

Recall: p is not in the null cone

iff

$\forall\ x \in \mathbb{R}^n$, $\exists\ \omega \in supp(p)$ s.t. $\langle x, \omega \rangle \geq 0$

iff

$\exists\ \alpha \in \mathbb{R}^m_{\geq 0}$, $supp(\alpha) \subseteq supp(p)$ s.t. $\sum_\omega \alpha_\omega \omega = 0$

Farkas lemma or Kempf-Ness criterion for the moment map

Suppose we find x s.t.

$$\| \nabla F_p(x) \|_2^2 \leq \varepsilon$$

$$\nabla F_p(x) = \frac{\sum_\omega |p_\omega|^2 e^{\langle x, \omega \rangle} \omega}{\sum_{\omega'} |p_{\omega'}|^2 e^{\langle x, \omega' \rangle}}$$

$$\beta_\omega := \frac{|p_\omega|^2 e^{\langle x, \omega \rangle}}{\sum_{\omega'} |p_{\omega'}|^2 e^{\langle x, \omega' \rangle}}$$

$$\nabla F_p(x) = \sum_\omega \beta_\omega \omega$$

if $\| \sum_\omega \beta_\omega \omega \|_2^2 \leq \varepsilon$, how small should ε be to make sure that p is not in the null cone

Weight margin :

Definition: Let $\Omega \subseteq \mathbb{Z}^n$ be a set of weights

$$\gamma(\Omega) := \min_{\substack{\Omega' \subseteq \Omega \\ \text{s.t. } 0 \notin \operatorname{conv}(\Omega')}} \left\{ \|v\|_2 : v \in \operatorname{conv}(\Omega') \right\}$$

exp small for most torus actions

but poly small for matrix scaling