Homework #3: PSD factorizations CSE 599I, Spring 2021
Due: Jun 2, 2021

Problem 1: Smooth PSD rank

(a) Fix $m, n \geqslant 1$, and let $C := \mathrm{conv}\left(\{uv^T : u \in \mathbb{R}_+^m, v \in \mathbb{R}_+^n\}\right)$ denote the convex hull of the set of matrices with nonnegative rank 1. Show that C contains every nonnegative $m \times n$ matrix.

(b) Define the quantity

$$\gamma_{\mathrm{psd}}(M) := \min\left\{\max_{i,j} \|A_i\|_{S_\infty}\|B_j\|_{S_1} : M_{ij} = \mathrm{Tr}(A_i B_j), 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\right\},$$

where the minimum is over all collections $A_1, \ldots, A_m, B_1, \ldots, B_n$ of *PSD matrices,* and $\|\cdot\|_{S_\infty}$ and $\|\cdot\|_{S_1}$ are the the Schatten ∞- and 1-norms, respectively. Show that for every $c > 0$, the set $\{M \in \mathbb{R}_+^{m\times n} : \gamma_{\mathrm{psd}}(M) \leqslant c\}$ is a closed convex set.

(c) One formulation of John's theorem is that for every symmetric, bounded convex body $K \subseteq \mathbb{R}^d$ with $\mathrm{Vol}_d(K) > 0$, there is an invertible linear map $F_K : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$B_{\ell_2}^d \subseteq F_K(K) \subseteq \sqrt{d} B_{\ell_2}^d,$$

where $B_{\ell_2}^d := \{x \in \mathbb{R}^d : \|x\|_2 \leqslant 1\}$ is the unit ℓ_2 ball.

Consider any two bounded sets $U, V \subseteq \mathbb{R}^d$ and define $\Delta := \max_{u\in U, v\in V} |\langle u, v\rangle|$. Let K be the convex hull of $U \cup (-U)$ and prove that the map F_K satisfies

$$\max_{u\in U, v\in V} \|F_K u\|_2 \cdot \|(F_K^{-1})^T v\|_2 \leqslant \sqrt{d}\Delta.$$

(Note that if U doesn't span $\mathbb{R}^d$, then $\mathrm{Vol}_d(K) = 0$, and one will need to pass a subspace or augment U in a way that doesn't affect the conclusion.)

(d) Use part (c) to show the following: For every $M \in \mathbb{R}_+^{m\times n}$, it holds that

$$\gamma_{\mathrm{psd}}(M) \leqslant (\mathrm{rk}_{\mathsf{psd}}(M))^{O(1)} \|M\|_{\ell_\infty},$$

where $\|M\|_{\ell_\infty} := \max_{i,j} |M_{ij}|$.

Hint: If $\mathrm{rk}_{\mathsf{psd}}(M) \leqslant d$, then you should be able to write $M_{ij} = \sum_{s,t=1}^d \langle u_{i,s}, v_{j,t}\rangle^2$ for some vectors $u_{i,s}, v_{j,t} \in \mathbb{R}^d$. Now apply (c) to the collections $U_i = \{u_{i,s} : 1 \leqslant s \leqslant d\}$, $V_j = \{v_{j,t} : 1 \leqslant t \leqslant d\}$ for each $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$.

Problem 2: Pseudodensities

For a subset $S = \{i_1, i_2, \ldots, i_k\} \subseteq [m]$ with $i_1 < i_2 < \cdots < i_k$ and $x \in \{0,1\}^n$, we use x_S to denote the vector $x_S \in \{0,1\}^{|S|}$ given by $x_S = (x_{i_1}, x_{i_2}, \ldots, x_{i_k})$, i.e., the restriction of x to the coordinates indexed by S.

(a) Recall that $\varphi : \{0,1\}^m \to \mathbb{R}$ is a degree-d pseudodensity if it holds that $\langle \varphi, \mathbf{1} \rangle = \mathbb{E}_z[\varphi(z)] = 1$, and

$$\langle \varphi, p^2 \rangle = \mathop{\mathbb{E}}_{z \in \{0,1\}^n} [\varphi(z) p(z)^2] \geqslant 0$$

whenever p is a multilinear polynomial with $\deg(p) \leqslant d$.

Say that a function $f : \{0,1\}^m \to \mathbb{R}$ is an *S-junta* if the value $f(x)$ only depends on x_S. Say that if $\varphi : \{0,1\}^m \to \mathbb{R}$ is a *d-local pseudodensity* if it holds that for any nonnegative S-junta $f : \{0,1\}^m \to \mathbb{R}_+$ with $|S| \leqslant d$, we have

$$\langle \varphi, f \rangle \geqslant 0.$$

Prove that if φ is a degree-d pseudodensity, then it is also a d-local pseudodensity.

(b) Show that if φ is a d-local pseudodensity, then for every $S \subseteq [m]$ with $|S| \leqslant d$, it holds that φ induces an actual probability density on $\{0,1\}^S$ in the following sense: There is a nonnegative S-junta $g_S : \{0,1\}^m \to \mathbb{R}_+$ such that for every S-junta f, we have

$$\langle \varphi, f \rangle = \mathop{\mathbb{E}}_z [g_S(z) f(z)].$$

Hint: Define $g_S(y) := \mathbb{E}_z \left[\varphi(z) \mid z_S = y_S \right]$.

(c) Define $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$, and note that $\{\chi_S : S \subseteq [m]\}$ is an orthonormal basis for the space of functions $f : \{0,1\}^m \to \mathbb{R}$. The Fourier coefficients of such a function are given by

$$\hat{f}(S) = \langle f, \chi_S \rangle = \mathop{\mathbb{E}}_{x \in \{0,1\}^m} f(x) \chi_S(x),$$

and writing f in the Fourier basis gives

$$f = \sum_{S \subseteq [m]} \hat{f}(S) \chi_S.$$

Note that if φ is d-local, then for $|S| \leqslant d$, part (b) gives

$$|\hat{\varphi}(S)| = |\mathop{\mathbb{E}}_z [g_S(z) \chi_S(z)]| \leqslant \mathop{\mathbb{E}}_z [g_S(z)] = 1,$$

where we uesd $|\chi_S(z)| \leqslant 1$.

Prove that if φ is a degree-d pseudodensity, then a stronger fact holds:

$$|\hat{\varphi}(S)| \leqslant 1, \quad \forall |S| \leqslant 2d.$$

Hint: Use the fact that $1 + \chi_A \chi_B = \frac{1}{2}(\chi_A + \chi_B)^2$ and $1 - \chi_A \chi_B = \frac{1}{2}(\chi_A - \chi_B)^2$.

(d) Suppose that $\varphi : \{0,1\}^m \to \mathbb{R}$ is a degree-d pseudodensity, and let $\varphi_{\leqslant 2d}$ denote the restriction of φ to terms of degree at most $2d$:

$$\varphi_{\leqslant 2d} := \sum_{|S| \leqslant 2d} \hat{\varphi}(S) \chi_S.$$

Show that $\varphi_{\leqslant 2d}$ is a degree-d pseuodensity and use part (c) to show that

$$\|\varphi_{\leqslant 2d}\|_\infty = \max_{z \in \{0,1\}^m} |\varphi_{\leqslant 2d}(z)| \leqslant \sum_{k=0}^{2d} \binom{m}{k}.$$

Problem 3: Pattern matrices

Recall that given a function $g : \{0,1\}^m \to \mathbb{R}_+$ and $n \geqslant m$, we defined a "pattern matrix" as follows:

$$\mathcal{L}(S,x) := g(x_S),$$

where $\mathcal{L}$ is indexed by subsets $S \subseteq [n]$ with $|S| = m$ and vectors $x \in \{0,1\}^n$.

If $\deg_{\mathsf{sos}}(g) > d$, we know there exists a degree-d pseudodensity $\varphi : \{0,1\}^m \to \mathbb{R}$ satisfying

$$\langle \varphi, g \rangle < 0.$$

Define the strictly positive quantity $\varepsilon := -\langle \varphi, g \rangle$.

Denote by $\varphi_S : \{0,1\}^n \to \mathbb{R}$ the functionals

$$\varphi_S(x) := \varphi(x_S).$$

Then we have

$$\mathop{\mathbb{E}}_{S,x}\left[\varphi_S(x)\mathcal{L}(S,x)\right] = \mathop{\mathbb{E}}_{S,x}\left[\varphi_S(x)g(x_S)\right] = \langle \varphi, g \rangle = \varepsilon < 0,$$

where the expectation is over $S \in \binom{[n]}{m}$ and $x \in \{0,1\}^n$ chosen uniformly at random.

Thus to prove that $\mathrm{rk}_{\mathsf{psd}}(\mathcal{L}) > r$, it would suffice to show that

$$\mathop{\mathbb{E}}_{S,x}\left[\varphi_S(x)\mathrm{Tr}(A(S)B(x))\right] > -\frac{\varepsilon}{2} \tag{0.1}$$

whenever $A : \binom{[n]}{m} \to \mathbb{S}_r^+$ and $B : \{0,1\}^n \to \mathbb{S}_r^+$ are functions taking values in $r \times r$ PSD matrices.

(a) **The case of low-degree squares.** Suppose that $B(x) = Q(x)Q(x)^T$, where $Q : \{0,1\}^n \to \mathbb{R}^{r\times s}$ takes values in arbitrary $r \times s$ real matrices, and $\deg(Q(x)_{ij}) \leqslant d$ for every $1 \leqslant i \leqslant r, 1 \leqslant j \leqslant s$. Show that

$$\mathop{\mathbb{E}}_{S,x}\left[\varphi_S(x)\mathrm{Tr}(A(S)B(x))\right] = \mathop{\mathbb{E}}_{S,x}\left[\varphi_S(x)\mathrm{Tr}(A(S)Q(x)Q(x)^T)\right] \geqslant 0.$$

This suggests the possibility of proving (0.1) by approximating $B(x) \approx \tilde{B}(x)$ where $\tilde{B}(x) = Q(x)Q(x)^T$ and $\deg(Q)$ is small.

(b) **The case of local selectors (extra credit!).** We could also attempt to approximate the A-side of the factorization. For this purpose, let us consider a submatrix of $\mathcal{L}$ as follows. Assume that $n = km$ so that every $x \in \{0,1\}^n = (\{0,1\}^k)^m$ is naturally partitioned into m blocks of size k.

Consider an element $\sigma \in [k]^m$ as a *selector* that chooses one bit from each block. We will abuse notation and consider such a selector as a map $\sigma : \{0,1\}^n \to \{0,1\}^m$ that takes $x \in \{0,1\}^n$ to the string $\sigma(x) \in \{0,1\}^m$ by selecting the corresponding bits from x.

Define a *d-partial selector* as an element $\alpha \in ([k] \cup \{*\})^m$ where $\alpha_i = *$ for all but d indices $1 \leqslant i \leqslant m$. For a selector $\sigma \in [k]^m$, and a partial selector α, let $\sigma_\alpha \in \{0,1\}$ be an indicator variable with $\sigma_\alpha = 1$ if σ and α agree (on all the coordinates where $\alpha_i \neq *$). Let $\mathcal{A}_d$ denote the space of d-partial selectors.

For a function $f : [k]^m \to \mathbb{R}$, define

$$\mathsf{deg}^S(f) := \min\left\{ d : f(\sigma) = \sum_{\alpha\in\mathcal{A}_d} c_\alpha \sigma_\alpha \text{ for some } \{c_\alpha\} \subseteq \mathbb{R} \right\}.$$

One can think of this as the degree of f on selectors, i.e., the minimum d such that f is in the span of functions that only look at d indices in σ.

Assume additionally that n is even and let $\mathscr{B}_n \subseteq \{0,1\}^n$ denote the subset of *balanced strings*, i.e. those that have an equal number of zeros and ones. Suppose that $A(\sigma) = P(\sigma)P(\sigma)^T$, where $P : [k]^m \to \mathbb{R}^{r\times s}$ is such that $\mathsf{deg}^S(P(\sigma)_{ij}) \leqslant d$ for all $1 \leqslant i \leqslant r, 1 \leqslant j \leqslant s$. Show that for any $B : \mathscr{B}_n \to \mathbb{S}_r^+$,

$$\mathop{\mathbb{E}}_{\substack{\sigma\in[k]^m,\\ x\in\mathscr{B}_n}} \left[\varphi(\sigma(x))\mathrm{Tr}(A(\sigma)B(x))\right] = \mathop{\mathbb{E}}_{\substack{\sigma\in[k]^m,\\ x\in\mathscr{B}_n}} \left[\varphi(\sigma(x))\mathrm{Tr}(P(\sigma)P(\sigma)^T B(x))\right] \geqslant 0.$$

Hint: Note that as x ranges uniformly over balanced strings and σ ranges over uniformly random selectors, $\sigma(x) \in \{0,1\}^m$ is uniformly random. What's particularly nice about balanced strings is that even for any *fixed* $x_0 \in \mathscr{B}_n$, it holds that $\sigma(x_0) \in \{0,1\}^m$ is uniformly random as σ ranges over $[k]^m$.

This suggests the possibility of proving that

$$\mathop{\mathbb{E}}_{\substack{\sigma\in[k]^m,\\ x\in\mathscr{B}_n}} \left[\varphi(\sigma(x))\mathrm{Tr}(A(\sigma)B(x))\right] \geqslant -\frac{\varepsilon}{2}$$

for any $A : [k]^m \to \mathbb{S}_+^r$ by approximating $A(\sigma) \approx \tilde{A}(\sigma) = P(\sigma)P(\sigma)^T$, where $\mathsf{deg}^S(P)$ is small.

(c) **Preparing the factorization for approximation.** If $\deg(g) \leqslant 2d$, then by Problem 2(d), we can assume that

$$\|\varphi\|_\infty \leqslant m^{2d} + 1. \tag{0.2}$$

(E.g., recall that for the cut polytope, we are interested in functions g with $\deg(g) \leqslant 2$.)

This says that a lower bound like (0.1) will be stable in the sense that it will prove $\mathrm{rk}_{\mathsf{psd}}(\tilde{\mathcal{L}}) > r$ for any matrix $\tilde{\mathcal{L}}$ with

$$\left\|\mathcal{L} - \tilde{\mathcal{L}}\right\|_{L^1} = \mathop{\mathbb{E}}_{S,x} \left|\mathcal{L}(S,x) - \tilde{\mathcal{L}}(S,x)\right| \leqslant \frac{\varepsilon}{2}\left(1 + m^{2d}\right)^{-1}.$$

Using Problem 1, prove that if $\mathrm{rk}_{\mathsf{psd}}(\mathcal{L}) \leqslant r$, then for every $0 < \eta < 1$, there are functions

$$A : \binom{[n]}{m} \to \mathbb{S}_r^+, \quad B : \{0,1\}^n \to \mathbb{S}_r^+$$

satisfying the following:

(i) $\mathbb{E}_{S,x} |\mathcal{L}(S,x) - \mathrm{Tr}(A(S)B(x))| \leqslant \eta\|\mathcal{L}\|_{\ell_\infty}$

(ii) $\mathbb{E}_S A(S) = I$,

(iii) $\|A(S)\|_{S_\infty} \leqslant r^{O(1)}/\eta$ for all $S \in \binom{[n]}{m}$.

(iv) $\|B(x)\|_{S_1} \leqslant \|\mathcal{L}\|_{\ell_\infty} r^{O(1)}$ for all $x \in \{0,1\}^n$.

Hint: Given a factorization $\mathcal{L}(S,x) = \mathrm{Tr}\left(A_0(S)B_0(x)\right)$, it may be helpful to add a small multiple of the identity to each $A_0(S)$ so that $\mathbb{E}_S A_0(S)$ becomes invertible.