Penn State University
CSE 564 Complexity of Combinatorial Problems

Instructor: Pei Wu

Module
14

Rounding global correlations

This lecture continues our discussion of Sum-of-Squares (SoS) relaxations and how to *round* their solutions into actual assignments for constraint satisfaction problems. The focus is on *dense* Max-2-CSP instances, where the constraint graph has $\Theta(n^2)$ edges. In this regime, obtaining an *additive* approximation guarantee (e.g. within εn^2) is meaningful and often achievable with relatively low levels of the SoS hierarchy.

A low-degree SoS solution gives us *locally consistent* distributions on small sets of variables. A very tempting (but generally incorrect) idea is to sample each variable independently from its one-variable marginal. The loss compared to the SoS objective comes exactly from *global correlations*: if the pairwise marginals μ_{ij} are far from products $\mu_i \times \mu_j$, then independent rounding can destroy value. The main technique in this lecture is to measure correlations by *mutual information* and to iteratively *condition* on carefully chosen variables until the remaining variables have small average mutual information. At that point, naive independent rounding is provably good.

14.1 Dense Max-2-CSP and the SoS relaxation

A *Max-2-CSP* instance is specified by a graph $G = ([n], E)$ and, for every edge $(i, j) \in E$, a constraint function
$$\phi_{ij} : \Sigma \times \Sigma \to [0, 1].$$
An assignment is $x \in \Sigma^n$, and its value is the total satisfied weight
$$\mathrm{val}(x) \;=\; \sum_{(i,j)\in E} \phi_{ij}(x_i, x_j). \tag{14.1}$$

We write $\mathrm{OPT} = \max_{x\in\Sigma^n} \mathrm{val}(x)$. The instance is *dense* if $|E| = cn^2$ for some constant $c > 0$. In this setting, $\mathrm{OPT} = \Theta(n^2)$ and we aim for an additive εn^2 approximation.

Theorem 14.1 (Dense Max-2-CSP via SoS (informal)). *Let $q = |\Sigma|$. There is a degree $d = O(\log q/\varepsilon^2)$ SoS-based algorithm that outputs an assignment x with*
$$\mathbb{E}[\mathrm{val}(x)] \;\geq\; \mathrm{OPT} - \varepsilon|E| \;=\; \mathrm{OPT} - O(\varepsilon n^2).$$

To write an SoS relaxation, introduce indicator variables

$$Y_{i,a} \in \{0,1\} \qquad (i \in [n],\ a \in \Sigma),$$

intended to represent "$x_i = a$". A valid assignment satisfies the polynomial constraints

$$Y_{i,a}^2 - Y_{i,a} = 0 \qquad \text{(Booleanity)} \tag{14.2}$$

$$\sum_{a \in \Sigma} Y_{i,a} = 1 \qquad \text{(unique label).} \tag{14.3}$$

In this encoding, the objective (14.1) becomes the degree-2 polynomial

$$P(Y) \ = \ \sum_{(i,j)\in E} \ \sum_{a,b\in\Sigma} \phi_{ij}(a,b)\, Y_{i,a} Y_{j,b}. \tag{14.4}$$

The degree-d SoS relaxation computes the optimum $\mathrm{SDP}_d(I)$ via the following *feasibility / binary-search* formulation. For a threshold $t \in \mathbb{R}$, consider the feasibility problem: Find a degree-d p.d. μ, s.t.

$$\begin{cases} \widetilde{\mathbb{E}}[1] = 1, \\ \widetilde{\mathbb{E}}[P(Y)] \geq t, \\ \widetilde{\mathbb{E}} \text{ satisfies (14.2)–(14.3) up to degree } d. \end{cases} \tag{14.5}$$

One defines

$$\mathrm{SDP}_d(I) \ := \ \sup\big\{t \in \mathbb{R} : \text{(14.5) is feasible}\big\}.$$

The value $\mathrm{SDP}_d(I)$ can be computed (to any precision δ) by binary search on t, each iteration solving the SDP feasibility problem (14.5); this takes time $n^{O(d)} \cdot \mathrm{poly}(\log 1/\delta)$. Any real assignment $x \in \Sigma^n$ gives a feasible pseudo-expectation, which achieves $\widetilde{\mathbb{E}}[P(Y)] = \mathrm{val}(x)$; hence $\mathrm{SDP}_d(I) \geq \mathrm{OPT}$.

14.2 Encoding polynomial constraints in SoS

The SoS formalism enforces algebraic constraints by requiring that they hold after multiplying by all low-degree test polynomials. The following is the guiding principle.

Fact 14.2 (How SoS encodes equalities and inequalities). *Let $\widetilde{\mathbb{E}}$ be a degree-d pseudo-expectation.*

1. *If a* true *distribution μ is supported on $\{x : P_0(x) = 0\}$, then*

$$\mathbb{E}_\mu[P_0(x)\, Q(x)] = 0 \qquad \forall Q.$$

In SoS we relax this to

$$\widetilde{\mathbb{E}}[P_0(x)\, Q(x)] = 0 \qquad \forall Q \text{ with } \deg(P_0 Q) \leq d.$$

2. *If a true distribution μ is supported on $\{x : P_1(x) \geq 0\}$, then*

$$\mathbb{E}_\mu[P_1(x)\, Q(x)^2] \geq 0 \qquad \forall Q.$$

In SoS we relax this to

$$\widetilde{\mathbb{E}}[P_1(x)\, Q(x)^2] \geq 0 \qquad \forall Q \text{ with } \deg(P_1 Q^2) \leq d.$$

Proof that these constraints can be enforced by an SDP. We work with a single variable set $x = (x_1, \dots, x_n)$ taking values in a finite alphabet Σ. Index all monomials $x^\alpha = \prod_i x_i^{\alpha_i}$ of degree at most $d/2$ by a set $\mathcal{M}$. A degree-d pseudo-expectation $\widetilde{\mathbb{E}}$ is determined by its *moment vector*

$$\mathbf{m}_\alpha \;:=\; \widetilde{\mathbb{E}}[x^\alpha], \qquad |\alpha| \le d.$$

Moment matrix (base positivity). The fundamental SDP variable is the *moment matrix* $M \in \mathbb{R}^{\mathcal{M}\times\mathcal{M}}$ defined by

$$M_{\alpha,\beta} \;:=\; \widetilde{\mathbb{E}}[x^{\alpha+\beta}], \qquad |\alpha|, |\beta| \le d/2.$$

For any polynomial $Q(x) = \sum_\alpha c_\alpha x^\alpha$ of degree at most $d/2$,

$$Q^\top M Q \;=\; \sum_{\alpha,\beta} c_\alpha c_\beta M_{\alpha,\beta} \;=\; \widetilde{\mathbb{E}}\left[\left(\sum_\alpha c_\alpha x^\alpha\right)^2\right] \;=\; \widetilde{\mathbb{E}}[Q(x)^2].$$

Hence the pseudo-expectation axiom $\widetilde{\mathbb{E}}[Q^2] \ge 0$ for all Q of degree $\le d/2$ is *exactly* the constraint $M \succeq 0$ (positive semidefiniteness). This is a standard SDP constraint.

Equality constraints (item 1). Write $P_0(x) = \sum_\gamma p_\gamma x^\gamma$. For any monomial x^β with $\deg(P_0) + |\beta| \le d$, the constraint $\widetilde{\mathbb{E}}[P_0(x) \cdot x^\beta] = 0$ becomes the linear equation

$$\sum_\gamma p_\gamma \, \mathbf{m}_{\gamma+\beta} \;=\; 0.$$

Ranging over all such β yields a finite set of *linear equality constraints* on the moment vector $(\mathbf{m}_\alpha)_{|\alpha|\le d}$. Since any polynomial Q of the appropriate degree is a linear combination of monomials, $\widetilde{\mathbb{E}}[P_0 Q] = 0$ for all such Q is equivalent to this finite set of linear equations, which is an affine constraint in the SDP.

Inequality constraints (item 2). Write $P_1(x) = \sum_\gamma p_\gamma x^\gamma$ and suppose $\deg(P_1) = k$. Define the *localizing matrix* $M^{P_1} \in \mathbb{R}^{\mathcal{M}'\times\mathcal{M}'}$, where $\mathcal{M}'$ indexes monomials of degree at most $(d-k)/2$, by

$$M^{P_1}_{\alpha,\beta} \;:=\; \widetilde{\mathbb{E}}\big[P_1(x)\, x^{\alpha+\beta}\big] \;=\; \sum_\gamma p_\gamma \, \mathbf{m}_{\gamma+\alpha+\beta}.$$

Each entry is a linear function of the moment vector. For any polynomial $Q(x) = \sum_\alpha c_\alpha x^\alpha$ of degree at most $(d-k)/2$,

$$Q^\top M^{P_1} Q \;=\; \sum_{\alpha,\beta} c_\alpha c_\beta M^{P_1}_{\alpha,\beta} \;=\; \widetilde{\mathbb{E}}\big[P_1(x) \cdot Q(x)^2\big]\,.$$

Hence the condition $\widetilde{\mathbb{E}}[P_1(x)Q(x)^2] \ge 0$ for all such Q is exactly $M^{P_1} \succeq 0$, another positive-semidefiniteness constraint in the SDP.

Summary. The degree-d SoS feasibility problem (14.5) can be written explicitly as:

$$\text{find } \mathbf{m} = (\mathbf{m}_\alpha)_{|\alpha|\le d} \in \mathbb{R}^{\mathcal{M}\cup\{0\}} \quad \text{such that} \quad M(\mathbf{m}) \succeq 0, \quad M^{P_1}(\mathbf{m}) \succeq 0, \quad L(\mathbf{m}) = 0, \quad \mathbf{m}_0 = 1,$$

where M is the moment matrix, M^{P_1} is a localizing matrix for each inequality constraint, and $L(\mathbf{m}) = 0$ encodes the equality constraints and the objective threshold $\widetilde{\mathbb{E}}[P(Y)] \ge t$. The number of variables is $|\mathcal{M}| = n^{O(d)}$, and each matrix has size at most $n^{O(d)} \times n^{O(d)}$, so the SDP can be solved in time $n^{O(d)}$. □

In our Max-2-CSP relaxation, the constraints (14.2) and (14.3) are equalities, so we require the corresponding identities to hold in the sense of item (1) above.

14.3 Reweighting and pseudo-conditioning

A powerful feature of pseudo-distributions is that we can mimic basic probabilistic operations, such as reweighting and conditioning, at low degree.

Definition 14.3 (Reweighting). Let $\widetilde{\mathbb{E}}$ be a degree-d pseudo-expectation and let p be a degree-k *sum-of-squares* polynomial (so $p = \sum_r r(x)^2$) with $\widetilde{\mathbb{E}}[p] > 0$. Define the reweighted pseudo-expectation $\widetilde{\mathbb{E}}_p$ on polynomials Q of degree at most $d - k$ by

$$\widetilde{\mathbb{E}}_p[Q] \;=\; \frac{\widetilde{\mathbb{E}}[p \cdot Q]}{\widetilde{\mathbb{E}}[p]}. \tag{14.6}$$

Fact 14.4. *If $\widetilde{\mathbb{E}}$ is degree-d and p has degree k, then $\widetilde{\mathbb{E}}_p$ is a degree-$(d-k)$ pseudo-expectation.*

Proof. **Normalization.**

$$\widetilde{\mathbb{E}}_p[1] \;=\; \frac{\widetilde{\mathbb{E}}[p \cdot 1]}{\widetilde{\mathbb{E}}[p]} = \frac{\widetilde{\mathbb{E}}[p]}{\widetilde{\mathbb{E}}[p]} = 1.$$

Positivity. Let Q be a polynomial with $\deg(Q) \leq (d-k)/2$, so that in particular $\deg(Q^2) \leq d - k$. Write $p = \sum_r r^2$ where each r has degree at most $k/2$. Then rQ has degree $\deg(r) + \deg(Q) \leq k/2 + (d-k)/2 = d/2$. Hence each $(rQ)^2$ has degree at most d, and $\widetilde{\mathbb{E}}[(rQ)^2] \geq 0$ by the positivity of $\widetilde{\mathbb{E}}$. Now expand:

$$\widetilde{\mathbb{E}}_p[Q^2] \;=\; \frac{\widetilde{\mathbb{E}}[p \cdot Q^2]}{\widetilde{\mathbb{E}}[p]} \;=\; \frac{\widetilde{\mathbb{E}}\big[\big(\sum_r r^2\big)Q^2\big]}{\widetilde{\mathbb{E}}[p]} \;=\; \frac{\sum_r \widetilde{\mathbb{E}}[r^2 Q^2]}{\widetilde{\mathbb{E}}[p]} \;=\; \frac{\sum_r \widetilde{\mathbb{E}}[(rQ)^2]}{\widetilde{\mathbb{E}}[p]} \;\geq\; 0,$$

where the last step uses $\widetilde{\mathbb{E}}[(rQ)^2] \geq 0$ for each r and $\widetilde{\mathbb{E}}[p] > 0$ by assumption.

Constraint satisfaction. If $\widetilde{\mathbb{E}}$ satisfies the equality constraint $\widetilde{\mathbb{E}}[P_0 Q'] = 0$ for all Q' with $\deg(P_0 Q') \leq d$, then for any polynomial Q'' with $\deg(P_0 Q'') \leq d - k$,

$$\widetilde{\mathbb{E}}_p[P_0 Q''] = \frac{\widetilde{\mathbb{E}}[p \cdot P_0 Q'']}{\widetilde{\mathbb{E}}[p]} = 0,$$

since $\deg(p \cdot P_0 Q'') \leq k + (d-k) = d$. Thus $\widetilde{\mathbb{E}}_p$ inherits all constraints of $\widetilde{\mathbb{E}}$ up to the reduced degree $d - k$. ∎

Reweighting recovers ordinary conditioning in the Boolean setting. For a true distribution on $x_n \in \{0, 1\}$,

$$\mu \mid x_n = 1 \;\propto\; \mathbf{1}(x_n = 1)\,\mu.$$

On $\{0, 1\}$ we have $\mathbf{1}(x_n = 1) = x_n = x_n^2$, so a pseudo-version of conditioning on $x_n = 1$ is obtained by reweighting with $p(x) = x_n^2$:

$$\widetilde{\mathbb{E}}_{\mu \mid x_n = 1}[Q] \;=\; \frac{\widetilde{\mathbb{E}}_\mu[Q \cdot x_n^2]}{\widetilde{\mathbb{E}}_\mu[x_n^2]}. \tag{14.7}$$

Similarly, conditioning on $x_n = 0$ corresponds to reweighting with $(1 - x_n)^2$:

$$\widetilde{\mathbb{E}}_{\mu \mid x_n=0}[Q] \;=\; \frac{\widetilde{\mathbb{E}}_\mu[Q \cdot (1 - x_n)^2]}{\widetilde{\mathbb{E}}_\mu[(1 - x_n)^2]}. \tag{14.8}$$

In our Max-2-CSP encoding, conditioning on "$x_i = a$" is implemented by reweighting with $Y_{i,a}^2$ (which equals $Y_{i,a}$ on true assignments). Each such conditioning consumes degree 2.

14.4 Local distributions from low-degree moments

A key intuition is that a degree-d pseudo-distribution behaves like a real distribution on any set of at most $d/2$ variables.

Fact 14.5 (Local distributions). *Let $\widetilde{\mathbb{E}}$ be a degree-d pseudo-expectation satisfying* (14.2)–(14.3)*. Fix indices $i_1, \ldots, i_k$ with $k \le d/2$ and labels $a_1, \ldots, a_k \in \Sigma$. Define*

$$\widetilde{\Pr}\big[X_{i_1} = a_1, \ldots, X_{i_k} = a_k\big] \;:=\; \widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} Y_{i_t,a_t}\right].$$

Then these numbers form a valid probability distribution on Σ^k (nonnegative and summing to 1*).*

Proof. **Nonnegativity.** Since $k \le d/2$, the product $\prod_{t=1}^{k} Y_{i_t,a_t}$ has degree $k \le d/2$. By the Booleanity constraints (14.2), $Y_{i,a}^2 = Y_{i,a}$ holds as a polynomial identity modulo the ideal, so inside the pseudo-expectation:

$$\widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} Y_{i_t,a_t}\right] \;=\; \widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} Y_{i_t,a_t}^2\right].$$

The right-hand side has degree $2k \le d$ and factors as a square $\widetilde{\mathbb{E}}\left[\left(\prod_{t=1}^{k} Y_{i_t,a_t}\right)^2\right] \ge 0$, where positivity follows from the pseudo-expectation axiom (the squared polynomial has degree $2k \le d$).

Normalization. Sum over all label tuples. The linearity of $\widetilde{\mathbb{E}}$ and the unique-label constraints (14.3) give

$$\begin{aligned}
\sum_{a_1 \in \Sigma} \cdots \sum_{a_k \in \Sigma} \widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} Y_{i_t,a_t}\right] &= \widetilde{\mathbb{E}}\left[\sum_{a_1 \in \Sigma} \cdots \sum_{a_k \in \Sigma} \prod_{t=1}^{k} Y_{i_t,a_t}\right] \\
&= \widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} \left(\sum_{a_t \in \Sigma} Y_{i_t,a_t}\right)\right] \\
&= \widetilde{\mathbb{E}}\left[\prod_{t=1}^{k} 1\right] \;=\; \widetilde{\mathbb{E}}[1] \;=\; 1,
\end{aligned}$$

where the interchange of sums and the product is valid because the i_t are distinct, so the sums over different indices act on disjoint sets of variables. ∎

In particular, if $d \ge 4$ then we obtain genuine one-variable marginals μ_i and pairwise marginals μ_{ij} (distributions on Σ and Σ^2, respectively).

14.5 Naive rounding and mutual information

Let $\widetilde{\mathbb{E}} = \widetilde{\mathbb{E}}_\mu$ be the degree-d SoS solution returned by the SDP. Assume $d \geq 4$ so that the one- and two-variable marginals are well defined. Let X_i denote the random label of variable i under these marginals:

$$\mu_i(a) \;=\; \widetilde{\mathbb{E}}[Y_{i,a}], \qquad \mu_{ij}(a,b) \;=\; \widetilde{\mathbb{E}}[Y_{i,a}Y_{j,b}].$$

Define the SoS objective value

$$\begin{aligned} \mathrm{SDP} &:= \widetilde{\mathbb{E}}\left[\sum_{(i,j)\in E}\sum_{a,b\in\Sigma} \phi_{ij}(a,b)\, Y_{i,a}Y_{j,b}\right] \\ &= \sum_{(i,j)\in E} \mathbb{E}_{(X_i,X_j)\sim\mu_{ij}}\big[\phi_{ij}(X_i,X_j)\big]. \end{aligned} \tag{14.9}$$

A *naive rounding* samples each X_i *independently* from μ_i and outputs the resulting assignment. Its expected value is

$$\mathrm{naive} \;=\; \sum_{(i,j)\in E} \mathbb{E}_{X_i\sim\mu_i,\ X_j\sim\mu_j}\big[\phi_{ij}(X_i,X_j)\big]. \tag{14.10}$$

The gap between (14.9) and (14.10) is controlled by how far μ_{ij} is from the product $\mu_i \times \mu_j$.

Claim 14.6 (Correlation loss bounded by total variation). *For each edge $(i,j) \in E$,*

$$\mathbb{E}_{(X_i,X_j)\sim\mu_{ij}}\big[\phi_{ij}(X_i,X_j)\big] - \mathbb{E}_{X_i\sim\mu_i,\ X_j\sim\mu_j}\big[\phi_{ij}(X_i,X_j)\big] \;\leq\; \|\mu_{ij} - \mu_i\times\mu_j\|_{\mathrm{TV}}\,.$$

Consequently,

$$\mathrm{naive} \;\geq\; \mathrm{SDP} \;-\; \sum_{(i,j)\in E} \|\mu_{ij} - \mu_i\times\mu_j\|_{\mathrm{TV}}\,. \tag{14.11}$$

Proof. **Per-edge bound.** Recall that for any two distributions P, Q on a finite set $\mathcal{X}$ and any function $f : \mathcal{X} \to [0,1]$,

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \;=\; \left|\sum_x f(x)(P(x)-Q(x))\right| \;\leq\; \sum_x |P(x)-Q(x)| \;=\; 2\,\|P-Q\|_{\mathrm{TV}}\,.$$

Since $\phi_{ij} \in [0,1]$ and we are bounding from below:

$$\mathbb{E}_{(X_i,X_j)\sim\mu_{ij}}[\phi_{ij}(X_i,X_j)] - \mathbb{E}_{X_i\sim\mu_i,X_j\sim\mu_j}[\phi_{ij}(X_i,X_j)] \;\leq\; \big|\mathbb{E}_{\mu_{ij}}[\phi_{ij}] - \mathbb{E}_{\mu_i\times\mu_j}[\phi_{ij}]\big| \;\leq\; \|\mu_{ij} - \mu_i\times\mu_j\|_{\mathrm{TV}}\,.$$

(The factor of 2 disappears because $f \in [0,1]$ gives a sharper bound of $\|P-Q\|_{\mathrm{TV}}$ by the dual characterization $\|P-Q\|_{\mathrm{TV}} = \sup_{f:0\leq f\leq 1}(\mathbb{E}_P[f] - \mathbb{E}_Q[f])$.)

Summing over edges. Substituting into the definitions (14.9) and (14.10):

$$\begin{aligned} \mathrm{naive} &= \sum_{(i,j)\in E} \mathbb{E}_{X_i\sim\mu_i,X_j\sim\mu_j}[\phi_{ij}(X_i,X_j)] \\ &\geq \sum_{(i,j)\in E} \big(\mathbb{E}_{\mu_{ij}}[\phi_{ij}(X_i,X_j)] - \|\mu_{ij}-\mu_i\times\mu_j\|_{\mathrm{TV}}\big) \\ &= \mathrm{SDP} - \sum_{(i,j)\in E} \|\mu_{ij}-\mu_i\times\mu_j\|_{\mathrm{TV}}\,. \square \end{aligned}$$

To turn (14.11) into an additive εn^2 bound, we need the *average* pairwise total variation distance to be $O(\varepsilon)$. This is where mutual information enters.

Fact 14.7 (Pinsker + mutual information). *For distributions P, Q on the same space,*

$$\|P - Q\|_{\mathrm{TV}}^2 \;\le\; \frac{1}{2}\,\mathrm{KL}(P\|Q).$$

In particular, for a pair $(X_i, X_j) \sim \mu_{ij}$,

$$\mathrm{KL}(\mu_{ij} \;\|\; \mu_i \times \mu_j) \;=\; \mathrm{I}(X_i; X_j).$$

Combining Claim 14.6 with Cauchy–Schwarz and Fact 14.7 gives the key bound on naive rounding. Starting from (14.11):

$$\begin{aligned}
\mathrm{SDP} - \mathrm{naive} \;&\le\; \sum_{(i,j)\in E} \|\mu_{ij} - \mu_i \times \mu_j\|_{\mathrm{TV}} \\
&\le\; |E| \cdot \left(\frac{1}{|E|} \sum_{(i,j)\in E} \|\mu_{ij} - \mu_i \times \mu_j\|_{\mathrm{TV}} \right) \\
&\le\; |E| \cdot \sqrt{\mathbb{E}_{(i,j)\in E}\Big[\, \|\mu_{ij} - \mu_i \times \mu_j\|_{\mathrm{TV}}^2 \,\Big]} \qquad (14.12)
\end{aligned}$$

where the last step is Cauchy–Schwarz (Jensen applied to the concave square root, or equivalently $\frac{1}{m}\sum_k z_k \le \sqrt{\frac{1}{m}\sum_k z_k^2}$). By Pinsker's inequality (Fact 14.7), $\|\mu_{ij} - \mu_i \times \mu_j\|_{\mathrm{TV}}^2 \le \frac{1}{2}\mathrm{KL}(\mu_{ij}\|\mu_i \times \mu_j) = \frac{1}{2}\mathrm{I}(X_i; X_j)$. Substituting into (14.12):

$$\begin{aligned}
\mathrm{naive} &\ge \mathrm{SDP} - |E| \cdot \sqrt{\mathbb{E}_{(i,j)\in E}\Big[\, \|\mu_{ij} - \mu_i \times \mu_j\|_{\mathrm{TV}}^2 \,\Big]} \\
&\ge \mathrm{SDP} - |E| \cdot \sqrt{\tfrac{1}{2}\,\mathbb{E}_{(i,j)\in E}\big[\mathrm{I}(X_i; X_j)\big]}. \qquad (14.13)
\end{aligned}$$

Thus, if the *average mutual information*

$$A \;:=\; \mathbb{E}_{(i,j)\in E}\big[\mathrm{I}(X_i; X_j)\big]$$

satisfies $A \le 2\varepsilon^2$, then

$$\mathrm{naive} \;\ge\; \mathrm{SDP} - \varepsilon|E| \;\ge\; \mathrm{OPT} - \varepsilon|E|.$$

When A is not small, we will *condition* to reduce it.

14.6 Correlation rounding via entropy potential

We now describe the "rounding by global correlations" procedure.

Let $\mu^{(0)} = \mu$ be the degree-d pseudo-distribution returned by SoS, and set $t = 0$. At each stage we maintain: (i) a set of already fixed variables, and (ii) a conditioned pseudo-distribution $\mu^{(t)}$ (implemented via reweighting as in Section 14.3) over the remaining variables.

Rounding algorithm. Let A_t be the average mutual information among *alive* (unfixed) variables under $\mu^{(t)}$.

1. If $A_t \le \varepsilon^2$, perform naive independent rounding on all remaining variables using their one-variable marginals under $\mu^{(t)}$ and output the full assignment.
2. Otherwise, choose an alive index i such that

$$\mathbb{E}_{j \text{ alive, } j \neq i}\big[\mathrm{I}_{\mu^{(t)}}(X_i; X_j)\big] \;\ge\; \varepsilon^2.$$

(Existence follows by averaging.) Sample a label $a \sim \mu_i^{(t)}$ and fix $X_i = a$. Update $\mu^{(t+1)}$ by pseudo-conditioning on the event $Y_{i,a} = 1$.

3. Stop if $t = C \log q/\varepsilon^2$ for a sufficiently large absolute constant C; otherwise increment t and repeat.

Step (1) is justified by (14.13) applied to $\mu^{(t)}$. The point of steps (2)–(3) is to argue that we reach (1) quickly, before running out of degree.

Claim 14.8 (Termination with high probability). *With probability at least $1 - 1/C$, the rounding algorithm reaches step (1) before $t = C \log q/\varepsilon^2$.*

Proof. **Entropy potential.** Consider the potential function

$$\Phi_t \;:=\; \mathbb{E}_{j \text{ alive}} \mathrm{H}_{\mu^{(t)}}(X_j),$$

the average entropy of the alive variables. Since each X_j takes values in Σ, we have $0 \le \mathrm{H}(X_j) \le \log q$, so

$$0 \le \Phi_t \le \log q \qquad \text{and in particular} \qquad \Phi_0 \le \log q.$$

Expected drop in potential. Suppose the algorithm executes step (2) at time t and picks index i satisfying

$$\mathbb{E}_{j \text{ alive}, j \neq i} \mathrm{I}_{\mu^{(t)}}(X_i; X_j) \ge \varepsilon^2.$$

Such an index exists because if every alive i had $\mathbb{E}_{j \neq i} \mathrm{I}(X_i; X_j) < \varepsilon^2$, then averaging over alive i would give $A_t < \varepsilon^2$ (up to the asymmetry from dropping $j = i$, which contributes 0), contradicting the assumption $A_t > \varepsilon^2$.

After sampling $a \sim \mu_i^{(t)}$ and updating $\mu^{(t+1)} = \mu^{(t)} \mid X_i = a$, note that variable i is fixed and leaves the alive set. The expected drop in potential is

$$\begin{aligned}
\mathbb{E}_a[\Phi_t - \Phi_{t+1}] &= \mathbb{E}_{j \text{ alive}, j \neq i} \Big(\mathrm{H}_{\mu^{(t)}}(X_j) - \mathbb{E}_{a \sim \mu_i^{(t)}} \mathrm{H}_{\mu^{(t+1)}}(X_j) \Big) \\
&= \mathbb{E}_{j \text{ alive}, j \neq i} \Big(\mathrm{H}_{\mu^{(t)}}(X_j) - \mathrm{H}_{\mu^{(t)}}(X_j \mid X_i) \Big) \\
&= \mathbb{E}_{j \text{ alive}, j \neq i} \mathrm{I}_{\mu^{(t)}}(X_i; X_j) \;\ge\; \varepsilon^2,
\end{aligned}$$

using the identity $\mathrm{H}(X_j) - \mathrm{H}(X_j \mid X_i) = \mathrm{I}(X_i; X_j)$ and the law of total expectation ($\mathbb{E}_a \mathrm{H}(X_j \mid X_i = a) = \mathrm{H}(X_j \mid X_i)$).

Markov bound. Let T be the (random) number of times step (2) is executed before reaching step (1). By linearity of expectation and the per-step drop:

$$\varepsilon^2\, \mathbb{E}[T] \;\le\; \mathbb{E}[\Phi_0 - \Phi_T] \;\le\; \Phi_0 \;\le\; \log q,$$

where we used $\Phi_T \ge 0$. Hence $\mathbb{E}[T] \le \log q/\varepsilon^2$. Applying Markov's inequality:

$$\mathbb{P}\Big[T > C \frac{\log q}{\varepsilon^2}\Big] \;\le\; \frac{\mathbb{E}[T]}{C \log q/\varepsilon^2} \;\le\; \frac{1}{C}. \qquad \square$$

Degree budget. Each conditioning step (step (2)) uses a degree-2 reweighting polynomial ($Y_{i,a}^2$), so it reduces the available SoS degree by exactly 2. After T conditioning steps the remaining pseudo-distribution has degree $d - 2T$. In the worst case $T = C \log q/\varepsilon^2$ (conditioned on not exceeding the budget), so we need $d - 2C \log q/\varepsilon^2 \geq 4$, i.e. $d \geq 4 + 2C \log q/\varepsilon^2$. Taking $d = \Theta(\log q/\varepsilon^2)$ satisfies this, and the SDP with this degree can be solved in time $n^{O(d)} = n^{O(\log q/\varepsilon^2)}$, which is polynomial in n for fixed q and ε.

Putting it together. At the time step (1) is reached, the remaining pseudo-distribution $\mu^{(t)}$ has degree at least 4, so pairwise marginals are well defined. By construction, the average mutual information $A_t \leq \varepsilon^2$. Applying (14.13) to $\mu^{(t)}$ and noting that the SoS objective can only improve after conditioning (the conditioned pseudo-expectation still satisfies all constraints and gives objective value at least that of the original, since we sample conditionings from the marginals), we obtain

$$\mathbb{E}[\text{val}(\text{output})] \;\geq\; \text{SDP}_d(I) - \varepsilon|E| \;\geq\; \text{OPT} - \varepsilon|E|.$$

This completes the proof of Theorem 14.1.