Applications of Discrete Harmonic Analysis, Probabilistic Method and Linear Algebra in Fixed-Parameter Tractability and Kernelization

Gregory Gutin

Department of Computer Science
Royal Holloway, University of London

Discrete Analysis, Newton Institute, June 2011

Outline

Outline

1. Introduction
2. Using Probabilistic Method and Harmonic Analysis for k-MaxLin-AA
3. Using Linear Algebra for k-MaxLin-AA
4. More Applications of Linear Algebra, Probabilistic Method and Hypercontrative Inequality

Kernelization

- A parameterized problem Π: a set of pairs (x, k) where x is the main part and k (usually an integer) is the parameter; x is an instance (usually $k \ll |x|$).
- Example 1 (k-VertexCover): Given a graph $G = (V, E)$, decide if $\exists U \subseteq V$ s.t. every edge has a vertex in U and $|U| \leq k$. Parameter: k.

Kernelization

- A parameterized problem Π: a set of pairs (x, k) where x is the main part and k (usually an integer) is the parameter; x is an instance (usually $k \ll |x|$).
- Example 1 (k-VertexCover): Given a graph $G = (V, E)$, decide if $\exists U \subseteq V$ s.t. every edge has a vertex in U and $|U| \leq k$. Parameter: k.
- Example 2 (k-IndependentSet): Given a graph $G = (V, E)$, decide if $\exists U \subseteq V$ s.t. no edge has both vertices in U and $|U| \geq k$. Parameter: k.

Kernelization

- A parameterized problem Π: a set of pairs (x, k) where x is the main part and k (usually an integer) is the parameter; x is an instance (usually $k \ll |x|$).
- Example 1 (k-VertexCover): Given a graph $G = (V, E)$, decide if $\exists U \subseteq V$ s.t. every edge has a vertex in U and $|U| \leq k$. Parameter: k.
- Example 2 (k-IndependentSet): Given a graph $G = (V, E)$, decide if $\exists U \subseteq V$ s.t. no edge has both vertices in U and $|U| \geq k$. Parameter: k.
- Example 3 (k-MaxLin-AA): We are given a system of linear equations over $\mathbb{F}_2$: $\sum_{i \in I_j} y_i = b_j$, $j \in [m]$; $I_j \subseteq [n]$, and each equation j has a positive integral weight w_j. Decide if there is an assignment of values to y_i's s.t. the total weight of satisfied equations is at least $k + \frac{1}{2} \sum_{j=1}^{m} w_j$.

Fixed-Parameter Tractability

- A parameterized problem is fixed-parameter tractable (fpt) if it can be solved in time $f(k)|I|^{O(1)}$.
- Example: runtime $T_1 = 2^k|x|$ is often much smaller than $T_2 = |x|^k$. For $|x| = 1000,\ k = 10,\ T_1 < 1s,\ T_2$ is infeasible.
- k-VertexCover is fpt.
- k-IndependentSet is W[1]-hard and thus highly unlikely to be fpt.
- Mahajan, Raman and Sikdar (IWPEC'06 & JCSS 2009): What is the complexity of k-MaxLin-AA?

Kernelization-1

- A kernelization of Π: a poly-time algorithm that maps an instance $(x, k) \in \Pi$ to an instance $(x', k') \in \Pi$ (the kernel) such that
  - (x, k) is YES iff (x', k') is YES
  - $k' \leq h(k)$ and $|x'| \leq g(k)$ for some functions h and g.
- $g(k)$ is the size of the kernel.
- k-VertexCover has a kernel with $\leq 2k$ vertices and k^2 edges. Size $m + n = k^2 + 2k$ (quadratic).

Kernelization-2

- A decidable parameterized problem is fixed-parameter tractable iff it admits a kernelization. So, k-IndependentSet has no kernel (unless FPT=W[1]).
- Wanted: low degree polynomial-size kernels (for preprocessing).
- Many fpt problems do not have polynomial-size kernels (unless coNP$\subseteq$NP/poly).
- k-VetexCover has a poly-size kernel.
- k-MaxLin-AA has size mn. Mahajan, Raman and Sikdar (IWPEC'06 & JCSS 2009): Is there any kernel for k-MaxLin-AA?

Outline

1. Introduction
2. **Using Probabilistic Method and Harmonic Analysis for k-MaxLin-AA**
3. Using Linear Algebra for k-MaxLin-AA
4. More Applications of Linear Algebra, Probabilistic Method and Hypercontrative Inequality

MaxLin-AA Reduction Rules

- Rule 1: Reduce the system such that no two equations have the same set of variables.
- Rule 2: Let $a_{ij} = 1$ if y_i is in equation j and $a_{ij} = 0$, otherwise. Reduce the system such that $\mathrm{rank} A = n$, the number of variables, over $\mathbb{F}_2$. Thus, $n \leq m$.
- After applying Rule 1 as long as possible and then Rule 2 as long as possible, we get an irreducible system (no further reductions are possible).

k-MaxLin-AA Reformulations

We consider an irreducible system.

- Each equation can be written in the 'product form': $\prod_{i \in I_j} x_i = c_j$, where $x_i = -1$ if $y_i = 1$ and $x_i = 1$ if $y_i = 0$, and $c_j = (-1)^{b_i}$.
- k-MaxLin-AA can be written in the 'function form': Let $f(x) = \sum_j d_j \prod_{i \in I_j} x_i$, where $d_j = c_j w_j$. Then the answer to k-MaxLin-AA is YES iff $\max_{x \in \{-1,1\}^n} f(x) \geq 2k$.

Strictly Above/Below Expectation Method (SABEM): Symmetric Case

- Gutin, Kim, Szeider and Yeo, IWPEC 2009 and JCSS 2011.
- Given a parameterized problem Π with parameter k.
- Apply some reduction rules.
- Introduce a random variable $\mathbf{X}$ s.t. $\mathbb{E}(\mathbf{X}) = 0$ and if $\text{Prob}(\mathbf{X} \geq k) > 0$ then the answer to Π is YES.
- If $\mathbf{X}$ is symmetric, then $\text{Prob}[\ \mathbf{X} \geq \sqrt{\mathbb{E}[\mathbf{X}^2]}\ ] > 0$.
- If $\sqrt{\mathbb{E}[\mathbf{X}^2]} \geq k$ then YES. Otherwise, $\sqrt{\mathbb{E}[\mathbf{X}^2]} < k$ and problem specific analysis is required.

SABEM: Asymmetric Case

Lemma (Alon, Gutin, Kim, Szeider, Yeo, SODA 2010)

Let $\mathbf{X}$ *be a real random variable and suppose that its first, second and forth moments satisfy* $\mathbb{E}[\mathbf{X}] = 0$, $\mathbb{E}[\mathbf{X}^2] = \sigma^2 > 0$ *and* $\mathbb{E}[\mathbf{X}^4] \leq b(\mathbb{E}[\mathbf{X}^2])^2$, *respectively. Then* $\text{Prob}[\ \mathbf{X} > \frac{\sigma}{2\sqrt{b}}\ ] > 0$.

How to check $\mathbb{E}[\mathbf{X}^4] \leq b(\mathbb{E}[\mathbf{X}^2])^2$?

Hypercontractive Inequality and its ‘Dual’

Let $f = \sum_{I\subseteq[n]} \hat{f}(I)\prod_{i\in I} x_i$, where $\hat{f}(I)$ are reals and each $x_i \in \{-1,1\}$. Assign value to each x_i randomly, uniformly and independently from the other variables. Then **f** is a random var.

Lemma (Hypercontractive Inequality (HI), Bonami, 1970)

Let $r = \max\{|I| : \ \hat{f}(I) \neq 0\}$. *Then* $\mathbb{E}[\mathbf{f}^4] \leq 9^r \mathbb{E}[\mathbf{f}^2]^2$.

Lemma (‘Dual’ HI, Gutin and Yeo, arXiv June 2011)

Let ρ *is the maximum number of appearances of a number* i *in* I *for which* $\hat{f}(I) \neq 0$. *Then* $\mathbb{E}[\mathbf{f}^4] \leq (2\rho + 1 - \frac{2\rho}{m})\mathbb{E}[\mathbf{f}^2]^2$, *where* $m = |\{I : \ \hat{f}(I) \neq 0\}|$.

For $\rho = 1$ it is tight, e.g., for $f(x) = 1 + \sum_{i=1}^n x_i$. Example: $f(x) = \prod_{i=1}^n x_i$.

Application of Symmetric Case of SABEM

Consider k-MaxLin-AA in the function form:
$f(x) = \sum_{j=1}^{m} d_j \prod_{i \in I_j} x_i$, $n \leq m$.

- Suppose that $\exists\ U \subseteq [n]$ s.t. $|U \cap I_j|$ is odd for each j.
- Then $\mathbf{f}$ is a symmetric random variable.
- By Parseval's Identity, $\mathbb{E}[\mathbf{f}^2] = \sum_{j=1}^{m} d_j^2 \geq m$.
- Thus, Prob$[\ \mathbf{f} \geq \sqrt{m}\ ] > 0$.
- If $\sqrt{m} \geq 2k$ then YES. Otherwise, $\sqrt{m} < 2k$ and $m < 4k^2$. Since $n \leq m$, we have a poly-size kernel.

Applications of Asymmetric Case of SABEM

Consider k-MaxLin-AA in the function form:
$f(x) = \sum_{j=1}^{m} d_j \prod_{i \in I_j} x_i$, $n \leq m$.

- Suppose that $r = \max\{|I_j| : j \in [m]\}$ is a constant.
- Then by HI, $\mathbb{E}[\mathbf{f}^4] \leq 9^r \mathbb{E}[\mathbf{f}^2]^2$.
- By Parseval's Identity, $\mathbb{E}[\mathbf{f}^2] = \sum_{j=1}^{m} d_j^2 \geq m$.
- By the inequality of Alon et al., Prob[$\mathbf{f} \geq \sqrt{m}/(2 \cdot 3^r)$] > 0.
- If $\sqrt{m}/(2 \cdot 3^r) \geq 2k$ then YES. Otherwise, $\sqrt{m}/(2 \cdot 3^r) < 2k$ and $m = O(k^2)$. Since $n \leq m$, we have a poly-size kernel.

We can make use of the Dual HI as well.

So Far ...

- We've been able to prove the existence of poly-size kernels but only for some special cases of k-MaxLin-AA.
- The 'asymmetric' cases can extended but will still fall far short of the general case.
- Open Question: Does k-MaxLin-AA admit a poly-size kernel?
- We do not know the answer to the question, but we can prove that k-MaxLin-AA has a kernel.
- Another approach is needed.

Outline

Algorithm $\mathcal{H}$

Introduced by Crowston, Gutin, Jones, Kim and Ruzsa (SWAT'10). Consider the 'product form.'

While the system S is nonempty do the following:

1. Choose an arbitrary equation $\prod_{i \in I} x_i = b$ and mark an arbitrary variable x_ℓ such that $\ell \in I$.
2. Mark this equation and delete it from the system.
3. Replace every equation $\prod_{i \in I'} x_i = b'$ in the system containing x_ℓ by $\prod_{i \in I \Delta I'} x_i = bb'$ (the weight of the equation is unchanged).
4. Apply Reduction Rule 1 to the system.

Lemma on Algorithm $\mathcal{H}$

Lemma (Crowston, Gutin, Jones, Kim and Ruzsa, SWAT 2010)

Let S be an irreducible system and assume that Algorithm $\mathcal{H}$ marks equations of total weight w. If $w \geq 2k$ then S is a YES*-instance of k-MaxLin-AA.*

How to choose equations to mark s.t. w is as large as possible?

Sum-Free Sets

- Let K and M be sets of vectors in $\mathbb{F}_2^n$ such that $K \subseteq M$.
- K is M-sum-free if no sum of two or more vectors in K is equal to a vector in M.

The M-sum-free lemma:

Lemma (Crowston, Gutin, Jones, Kim and Ruzsa, SWAT 2010)

Let M be a proper subset in $\mathbb{F}_2^n$ such that $\mathrm{span}(M) = \mathbb{F}_2^n$ and $\mathbf{0} \in M$. If k is a positive integer and $k+1 \leq |M| \leq 2^{n/k}$ then, in time $|M|^{O(1)}$, we can find an M-sum-free subset K of M s.t. $|K| = k+1$.

Main Technical Theorems

The M-sum-free lemma implies Th. 1:

Theorem (Crowston, Gutin, Jones, Kim, Ruzsa, SWAT 2010)

Let S be an irreducible system of k-MaxLin-AA and let $k \geq 1$. If $2k \leq m \leq 2^{n/(2k-1)} - 2$, then the answer to k-MaxLin-AA is YES. *Moreover, we can find a* YES*-assignment to the variables in time $m^{O(1)}$.*

Using Algorithm $\mathcal{H}$ and a depth-bounded search we can prove Th. 2:

Theorem (Crowston, Fellows, Gutin, Jones, Rosamond, Thomasse, Yeo, arXiv 2011)

There exists an $n^{2k}(nm)^{O(1)}$-time algorithm for k-MaxLin-AA.

k-MaxLin-AA is FPT

Theorem (Crowston, Fellows, Gutin, Jones, Rosamond, Thomasse, Yeo, arXiv 2011)

k-MaxLin-AA has a kernel with at most $O(k^2 \log k)$ variables.

Proof: Irreducible system with m equations and n variables; $n \leq m$. Cases:

1. $m < 2k$: $n = O(k^2 \log k)$.
2. $2k \leq m \leq 2^{n/(2k-1)} - 2$: by Th. 1, the answer is YES.
3. $m \geq n^{2k}$: by Th. 2, we can solve the problem in poly-time.
4. $2^{n/(2k-1)} - 1 \leq m \leq n^{2k} - 1$: $n^{2k} \geq 2^{n/(2k-1)}$ implying $n = O(k^2 \log k)$.

Parameterized Algorithm for k-MaxLin-AA

Theorem (Crowston, Fellows, Gutin, Jones, Rosamond, Thomasse, Yeo, arXiv 2011)

k-MaxLin-AA can be solved in time $2^{O(k \lg k)}(nm)^{O(1)}$.

Proof: Irreducible system S with m equations and n variables.
1. By the previous theorem, in time $(nm)^{O(1)}$, we either solve k-MaxLin-AA for S or get a kernel with $O(k^2 \log k)$ variables.
2. In the last case, apply the $n^{2k}(nm)^{O(1)}$-time algorithm for k-MaxLin-AA, which for $n = O(k^2 \log k)$ has runtime $2^{O(k \lg k)}m^{O(1)}$.

Outline

(k, r)-MaxLin-AA

(k, r)-Max-r-Lin-AA is k-MaxLin-AA in which every equation has at most r variables and $k + r$ is the parameter.

Lemma (Crowston, Fellows, Gutin, Jones, Rosamond, Thomasse, Yeo, arXiv 2011)

Let $M \subseteq \mathbb{F}_2^n$ s.t. $\mathrm{span}(M) = \mathbb{F}_2^n$; each vector in M contains $\leq r$ non-zero coordinates. If $n \geq r(k-1)+1$, then in time $|M|^{O(1)}$, we can find an M-sum-free subset K of M such that $|K| = k$.

Theorem (ditto)

(k, r)-Max-r-Lin-AA has a kernel with $\leq (2k-1)r$ variables.

This improves a kernel with $n \leq r(r+1)k$ by Kim and Williams (arXiv 2010).

Open Que.: Is there a poly-size kernel for (k, r)-Max-r-Lin-AA?

Max-r-Sat-AA

- CNF formula F with clauses of sizes $r_1, \ldots, r_m$ and variables $y_1, \ldots, y_n$. Let $\max_i r_i \leq r$, a constant.
- $\mathrm{sat}(F, a)$ = the number of clauses satisfied by an assignment $a: \ \{y_1, \ldots, y_n\} \to \{\mathrm{true}, \mathrm{faulse}\}$.
- Random assignment $\mathbf{a}$. $E := \mathbb{E}[\mathrm{sat}(F, \mathbf{a})] = \sum_{i=1}^{m}(1 - 2^{-r_i})$.
- Max-r-Sat-AA: Is there an assignment a s.t. $\mathrm{sat}(F, a) \geq E + k$ (k parameter)?
- Mahajan et al. (2006, 2009): What is the complexity of this problem?

Max-r-Sat-AA and Pseudo-Boolean Functions

- For simplicity: each $r_i = r$.
- Let C be a clause of F with variables $y_{p_1}, \ldots, y_{p_r}$.
- $f_C(x) = 1 - \prod_{i=1}^{r}(1 + \varepsilon_{p_i} x_{p_i})$, $x_{p_i} \in \{-1, 1\}$, coef's $\varepsilon_{p_i} \in \{-1, 1\}$ and $\varepsilon_{p_i} = 1$ iff y_{p_i} (not $\bar{y}_{p_i}$) is in C. ($y_j = \mathrm{true}$ iff $x_{p_i} = -1$.)
- $f(x) = \sum_{C \in F} f_C(x)$.
- For an assignment a, we have $f(x) = 2^r[\mathrm{sat}(F, a) - E]$. Thus, $\textsc{Yes}$ iff $f(x) \geq k2^r$.

Max-r-Sat-AA Has Quadratic Kernel

- After algebraic simplification: $f(x) = \sum_{J \in \mathcal{F}} c_J \prod_{j \in J} x_i$, a Fourier expansion of f, where $|J| \leq r$ for each $J \in \mathcal{F}$.
- Use SABEM [Alon, Gutin, Kim, Szeider, Yeo, SODA 2010] to get either YES or $m = O(k^2)$.
- This can be extended to CSPs: given n variables and m Boolean formulas, each on at most r variables, define E, the average number of the formulas that can be satisfied, determine whether we can satisfy at least $E + k$ of the formulas.
- This problem is FPT. [ditto]

Betweenness

- Let $V = \{v_1, \ldots, v_n\}$ be a set of variables and let $\mathcal{C}$ be a set of m betweenness constraints of the form $(v_i, \{v_j, v_k\})$.
- Given a bijection $\alpha : V \to \{1, \ldots, n\}$, we say that a constraint $(v_i, \{v_j, v_k\})$ is satisfied if either $\alpha(v_j) < \alpha(v_i) < \alpha(v_k)$ or $\alpha(v_k) < \alpha(v_i) < \alpha(v_j)$.
- BETWEENNESS: find a bijection α satisfying the max number of constraints in $\mathcal{C}$.
- Tight Lower Bound: $m/3$, the expected number of satisfied constraints.
- Charikar, Guruswami and Manokaran (CCC'09): Approximating BETWEENNESS within factor $1/3 + \varepsilon$ is Unique-Games hard.

Betweenness-AA

- Betweenness-AA: Is there α that satisfies $\geq m/3+\kappa$ constraints? (κ is the parameter)
- Benny Chor's question in Niedermeier's book (2006): What is the parameterized complexity of Betweenness-AA?
- Reduction Rule: delete complete triples $(1,\{2,3\}),(2,\{3,1\}),(3,\{1,2\})$.
- We can introduce X required by SABEM, but ...
- It's difficult to estimate $\mathbb{E}(X^2)$, practically impossible to do $\mathbb{E}(X^4)$, but we cannot use Hypercontractive Inequality as X is not a polynomial of constant-bounded degree.

Introducing Four Bins

- An instance $(V, \mathcal{C})$, where V is the set of variables and $\mathcal{C} = \{C_1, \dots, C_m\}$ is the set of betweenness constraints.
- A function $\phi : V \to \{0, 1, 2, 3\}$ (vertices into 4 bins).
- ϕ-compatible bijections α: if $\phi(v_i) < \phi(v_j)$ then $\alpha(v_i) < \alpha(v_j)$.

Using Four Bins

- Let α be a random ϕ-compatible bijection and $\nu_p(\alpha) = 1$ if C_p is satisfied and 0, otherwise.
- Let the *weights* $w(C_p, \phi) = \mathbb{E}(\nu_p(\alpha)) - 1/3$ and $w(\mathcal{C}, \phi) = \sum_{p=1}^{m} w(C_p, \phi)$.

Lemma

If $w(\mathcal{C}, \phi) \geq \kappa$ *then* $(V, \mathcal{C})$ *is a* YES*-instance of Betweenness-AA.*

- Thus, to solve Betweenness-AA, it suffices to find ϕ for which $w(\mathcal{C}, \phi) \geq \kappa$.
- We may forget about bijections α and use SABEM !

Exploiting Four Bins

- We can get X of degree 6.
- $\mathbb{E}[\mathbf{X}^2] \geq 11m/768$.
- The rest is easy.
- Gutin, Kim, Mnich and Yeo, JCSS 2011: Betweenness-AA has an $O(\kappa^2)$-kernel.

Thank you!

- Questions?
- Comments?