Harvard CS229r / MIT 6.S898 October 21, 2016

Symmetry Reduction in Semidefinite Programming

Lecturer: Pablo Parrilo *Scribe: Chenyang Yuan*

1 Intoduction

In this notes, we are going to study how we can use symmetry to reduce the size of semidefinite programs. As a motivating example, let's consider finding a bound the independence number of the following symmetric graph: $V \in \{0,1\}^n$, and (u,v) is an edge if their Hamming distance $d(u,v) \leq k$. One can use a SDP to upper bound this quantity, but the size of the SDP grows as the number of vertices increase. If we want an asymptotic bound of the independence number, the problem quickly becomes very large. However, notice that the Hamming distance is invariant under permutations, we can permute the labels of vertices on the graph and the problem remains the same. This hints that we can exploit the symmetry inherent this problem to reduce its size.

Suppose $p(x)$ is a function. It is symmetric if $p(x) = p(Tx)$ under a linear transformation T. If we have two such linear transformations T and R we can compose them and $p(x)$ will also be symmetric under their composition TR. Thus we can see that these symmetries form a group structure and their interactions with $p(x)$ can be described using representation theory.

2 Representation Theory

We first recall the definition of a group:

Definition 1 (Group)**.** *A group is a set G and a binary operation $\cdot : G \times G \to G$ with the following properties:*

1. *Associativity: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$.*
2. *Existance of identity: There exists an element $e \in G$ so that $e \cdot g = g$ for all $g \in G$.*
3. *Existance of inverse: For all $g \in G$, there exists $g^{-1} \in G$ so that $g \cdot g^{-1} = e$.*

Representations of a group lets us describe group structure in terms of linear transformations and linear algebra.

Definition 2 (Group Representation)**.** *Let $V = \mathbb{C}^n$ be a vector space over the field of complex numbers and $GL(V)$ be the group of non-singular linear transformations over V.*

Then the representation of a group G on V is a group homomorphism $\rho : G \to GL(V)$ from G to $GL(V)$, such that for all $g, h \in G$:

$$\rho(g \cdot h) = \rho(g)\rho(h)$$

Example 1 (Representations of S_2)**.** We study representations of $S_2 = \{e, g\}$, the symmetric group with two elements where $g^2 = e$. We can represent this group as permutation matrices, with $\rho(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\rho(g) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. One may wonder if we need the full vector space to capture the structure of the group. This brings us to the question if there are subspaces that stay invariant under the transformations, so that if we restrict to these subspaces we have a subrepresentation of the group.

Graphically, $\rho(e)$ is the identity transformation that does nothing and $\rho(g)$ is a reflection about the line $y = x$. It is easy to see the subspaces $y = x$ and $y = -x$ are invariant under both transformations $\rho(e)$ and $\rho(g)$. Now if we restrict ourselves to the subspace $y = x$, we get the identity representation. Otherwise if we restrict to the subspace $y = -x$, we have the one-dimensional representation ρ' where $\rho'(e) = 1, \rho'(g) = -1$.

In the previous example, we saw that the vector space of the representation ρ can be decomposed into the direct sum of two smaller subspaces ($y = x$ and $y = -x$). Informally, if we can find invariant subspaces of a representation, we can decompose a representation into smaller representations. A representation that cannnt be decomposed further is called irreducible.

Definition 3 (Irreducible Representation)**.** *Given a group G and a representation $\rho : G \to GL(V)$, W is an invariant subspace if W is a subspace of V and $\rho(g)(W) \subseteq W$ for all $g \in G$. A representation ρ is irreducible if it does not have any non-trivial ($\varnothing$ or the entire vector space) invariant subspaces.*

We can define equivalence between representations:

Definition 4. *Two representations $\rho_1 : G \to GL(V_1)$ and $\rho_2 : G \to GL(V_2)$ are equivalent when there is a vector space isomorphism T so that $\rho_1(g) = T^{-1}\rho_2(g)T$ for all $g \in G$.*

The following theorem relates the size of the group to the dimensions of its irreducible representations:

Theorem 1 (Maschke)**.** *Let G be a finite group. Then it has a finite number of inequivalent irreducible representations ρ_i of dimension d_i. Furthermore, the sum of squares of dimensions is equal to the size of the group:*

$$\sum d_i^2 = |G|$$

In particular, this theorem tells us that for the previous example of representations of S_2, the only irreducible representations are the trivial and sign representations. Next we illustrate the theorem with another example:

Example 2 (Representations of S_3)**.** Let $S_3 = \{e, s, c, c^2, cs, sc\}$ be the symmetric group of three elements, where e is the identity, $s : 123 \to 213$ swaps two elements and $c : 123 \to 312$ performs a cyclic shift. The group operations are fully defined by three identities: $c^3 = e$, $s^2 = e$ and $s = csc$. Note that unlike the previous example with S_2, this group is non-abelian[1]. We first have the trivial representation ρ_T where $\rho_T(s) = \rho_T(c) = 1$. Next, notice that s and c performs an odd and even number of swaps respectively. Thus we have the sign representation ρ_A where $\rho_A(s) = -1$ and $\rho_A(c) = 1$. Finally, since S_3 is isomorphic to the dihedral group D_3 (symmetries of a triangle), we can represent s as a reflection and c as a rotation of $2\pi/3$: $s = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $c = \begin{bmatrix} \omega & 0 \\ 0 & \omega^2 \end{bmatrix}$ where $\omega^3 = 1$. The degrees of these irreducible representations are 1, 1 and 2, with $1^2 + 1^2 + 2^2 = 6 = |S_3|$.

Example 3 (Representations of C_n)**.** Let $C_n = \{e, c, \cdots, c^{n-1}\}$ be a cyclic group of n elements where $c^n = e$. Here there are n irreducible representations, each a root of unity: $\rho_k(c) = \omega_k^c = e^{2i\pi ck/n}$.

3 Symmetry Reduction in Semidefinite Programming

Now with the tools of representation theory, we return to the problem of minimizing a symmetric polynomial $f(x)$. Can knowing the symmetries of this polynomial give us more information about the minimum values? It turns out that symmetry together with convexity can give us more information about these minimum points. For example consider $f(x)$, an even function of a single variable. If all we know is that it is symmetric about $x = 0$, we only know that if $x = c$ is a minima,p $x = -c$ is also a minima. However if we have the additional information that this function is convex, we now know that $x = 0$ must be a minima by the definition of convexity. Now we formalize this idea that if a convex function is symmetric, its solutions will also never break symmetry:

Definition 5 (Invariant Function)**.** *A function $f(x)$ is invariant with respect to representation ρ if $f(\rho(g)x) = f(x)$ for all $g \in G$.*

Definition 6 (Fixed-point Subspace)**.** *We first define a linear map $R : V \to V$ so that $R(x) = \frac{1}{|G|}\sum_{g\in G} \rho(g)x$. This is the average of images of x under a representation ρ of the elements of G. We define the fixed-point subspace to be the image of V under this linear map: $F = \{x \mid x = \rho(g)x, \forall \rho \in G\}$.*

Theorem 2. *If a function $f(x)$ is convex and is invariant with respect to ρ, then its optima lies in its fixed-point subspace.*

Proof. It is easy to see from the definitions that R maps any point x into a point in the

[1]Representation theory of non-abelian groups is much more complex than that of abelian groups. In fact, it can be shown all irreducible representations of an abelian group have dimension 1.

fixed-point subspace such that $f(R(x)) \leq f(x)$.

$$f(R(x)) = f\left(\frac{1}{|G|}\sum_{g\in G}\rho(g)x\right) \leq \frac{\sum_{g\in G} f(\rho(g)x)}{|G|} = f(x)$$

□

Now we look at a more specialized result for semidefinite programs. Recall that a semidefinite program can be written in general as:

$$\max_{X\in S_n^+\cap L} \langle C, X\rangle$$

Where S_n^+ is the cone of semidefinite matrices and $L \subseteq S^n$ is an affine subspace of symmetric $n \times n$ matrices. For most of the cases of interest, the action of a group on the space of semidefinite matrices acts on the variables that indexes a quadratic form, thus the fixed-point subspace is defined as:

$$F = \{X : X = \rho^T(g)X\rho(g), \forall g \in G\}$$

These fixed-point subspaces for semidefinite matrices have a very specific form. In particular, F will become block diagonal after a change of coordinates. This comes from us being able to write the representation ρ as a direct sum of irreducible representations. We will not formally prove this and refer readers to section 4 of [GP]. An important consequence of this transformation is that instead of the entire matrix X, we only need to prove that individual blocks are PSD, thus reducing the size of the problem.

Example 4. We consider the problem of finding the independence number of a graph $G = (V, E)$. We can write this as a semidefinite relaxation:

$$\begin{aligned} \max \langle J, X\rangle & \\ \mathrm{Tr}(X) = 1 & \\ X_{ij} = 0 \quad & \forall (i,j) \in E \\ X \succeq 0 & \end{aligned}$$

Suppose we are working on the cycle graph with n vertices. Since the graph is invariant under cyclic shifts, the fixed point subspace is the space of symmetric matrices that are invariant under simultaneous cyclic shifts of the rows and columns:

$$\begin{bmatrix} x_1 & x_n & & \cdots & x_2 \\ x_2 & x_1 & x_n & & \vdots \\ & x_2 & x_1 & \ddots & \\ \vdots & & \ddots & \ddots & x_n \\ x_n & \cdots & & x_2 & x_1 \end{bmatrix}$$

These matrices are also known as circulant matrices. In particular, the Fourier basis diagonalizes these matrices, thus we only need to consider the diagonal in this new basis, thus transforming the semidefinite constraint into a linear constraint.

Example 5. Suppose we are minimizing an even polynomial $p(x)$ of one variable. We do not know if this polynomial is convex, but the symmetry $p(-x) = p(x)$ can still be exploited. This problem is equivalent to finding the maximum λ such that $p(x) - \lambda$ is a sum of squares. Suppose $b = \begin{bmatrix} 1 & x & x^2 & \cdots & x^d \end{bmatrix}$ is a suitable basis for $p(x)$, then this can be solved by the semidefinite program $p(x) - \lambda = b^T Q b$ for some PSD matrix Q. Since $p(x)$ is even, we separate the odd and even components of b, writing $b = \begin{bmatrix} b_{\text{even}} & b_{\text{odd}} \end{bmatrix}$. If we also write Q in block diagonal form, the fixed-point subspace is given by:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}$$

which is block diagonal. Hence we can write

$$p(x) - \lambda = b_{\text{even}}^T R_1 b_{\text{even}} + x^2 b_{\text{even}}^T R_2 b_{\text{even}}$$

Now we define $q(t) = p(x^2)$. From the symmetry reduction above, we get a necessary and sufficient condition for $q(t)$ to be non-negative on $t \geq 0$: it can be written as $q(t) = f(x) + x^2 g(x)$ where f and g are sum of squares polynomials.

References

[GP] Karin Gatermann and Pablo A. Parrilo, *Symmetry groups, semidefinite programs, and sums of squares*, no. 1, 95–128.