

Exploiting Probabilistic Independence for Permutations: Proofs

Jonathan Huang, Carlos Guestrin
 Carnegie Mellon University
 Pittsburgh, Pennsylvania 15213
 {jch1, guestrin}@cs.cmu.edu

Xiaoye Jiang, Leonidas Guibas
 Stanford University
 Stanford, California 94305
 {guibas@cs, xiaoyej}@stanford.edu

Appendix: The Littlewood-Richardson rule

Let λ be a partition of n and let p and q be positive integers such that $p + q = n$. If ρ_λ is any irreducible representation of S_n , then, restricted to permutations which lie in the subgroup $S_p \times S_q \subset S_n$, ρ_λ splits according to Maschke's theorem as a direct sum of irreducibles of $S_p \times S_q$ which take the form $\rho_\mu \otimes \rho_\nu$ (where μ and ν are partitions of p and q respectively):

$$\rho_\lambda \downarrow_{S_p \times S_q} \equiv \bigoplus_{\mu, \nu} \bigoplus_{\ell=1}^{c_{\mu, \nu}^\lambda} \rho_\mu \otimes \rho_\nu. \quad (0.1)$$

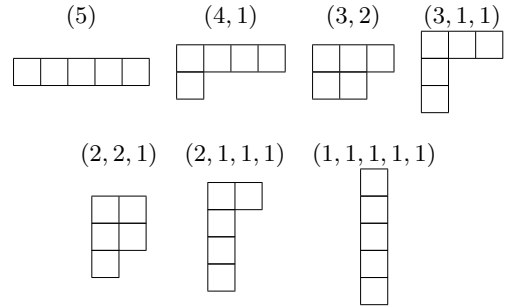
The multiplicities in the decomposition (Equation 0.1) are equivalent to the famous *Littlewood-Richardson coefficients*,¹ and in this appendix, we describe a result known as the *Littlewood-Richardson (LR) rule* which will allow us to compute the Littlewood-Richardson coefficients tractably (at least for low-order terms). There are several methods for computing these numbers (see [Knutson and Tao, 1999, Vakil, 2006], for example) but it is known ([Narayanan, 2006]) that, in general, the problem of computing the Littlewood-Richardson coefficients is $\#P$ -hard in general, and as we will see, involves enumerating the integer feasible points of a linearly constrained polytope.

The statement of the LR rule requires us to define a class of (rather complex) combinatorial objects known as the *Littlewood-Richardson tableaux*, which will be used to count the LR coefficients. We proceed by defining the LR tableaux in several stages.

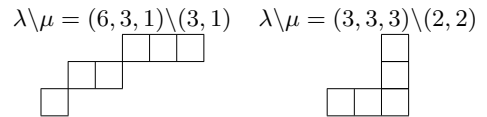
- (*Ferrers diagrams*) We can visualize a partition λ , of n , using a *Ferrers diagram* which is an array of boxes with λ_i boxes in the i^{th} row of the associated Ferrers diagram. For example, we have the

¹In most texts, the Littlewood-Richardson coefficients are defined in a slightly different way using induced representations (see [Sagan, 2001]), but the definition given in this paper is equivalent.

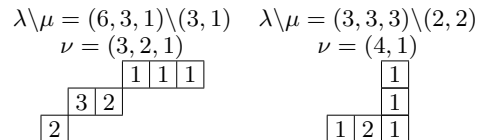
following partitions of $n = 5$ and their respective Ferrers diagrams.



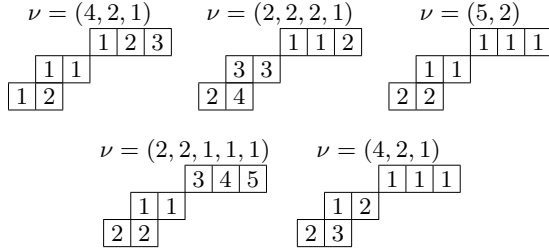
- (*Skew tableaux*) Let λ be a partition of n and μ a partition of some $p \leq n$. A *skew tableau* with shape $\lambda \setminus \mu$ is the diagram obtained by removing all boxes of the Ferrers diagram of λ which also belong to the Ferrers diagram of μ . The following are a few examples of skew tableaux and their corresponding shapes.



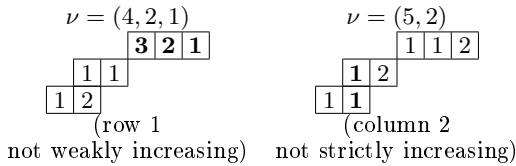
- (*Content*) As before, we will consider λ to be a partition of n and μ to be a partition of some $p \leq n$. Additionally, let ν be a partition of $q = n - p$. We say that a skew tableaux of shape $\lambda \setminus \mu$ has *content* $\nu = (\nu_1, \nu_2)$ if its boxes are filled in with ν_1 ones, ν_2 twos, and so on. To extend the previous example, we have:



- (*Semistandard tableaux*) We say that a skew tableau with shape $\lambda \setminus \mu$ and content ν is *semistandard* if its rows are *weakly increasing* (reading from left to right) and its columns are *strictly increasing* (reading from top to bottom). For example, the following are semistandard tableaux with shape $(6, 3, 2) \setminus (3, 1)$:

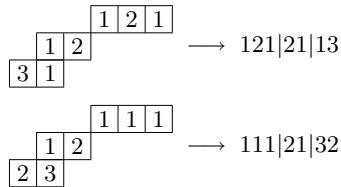


While the following are *invalid* as semistandard tableaux:



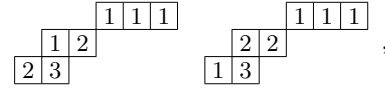
- (*Reverse lattice word constraint*) A word $w_1 \dots w_r$ is said to be a *lattice word* if, for each $s \leq r$, the subsequence $w_1 \dots w_s$ contains at least as many ones as twos, at least as many twos and threes, and so on. For example, 11123211 and 12312111 are lattice words while 1114 and 12321111 are not (in the first case because there are more fours than threes in 1114, and in the second case because there are more twos than ones in the subsequence 1232).

A skew tableau is said to satisfy the *reverse lattice word constraint* if a lattice word is obtained by reading its entries from top to bottom and from right to left (as in Hebrew). The following are two examples for skew tableaux satisfying the reverse lattice word constraint.

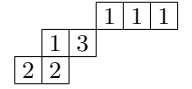


Definition 1. A skew tableaux with shape $\lambda \setminus \mu$ with content ν which is semistandard and whose row word is a lattice permutation is called a *Littlewood-Richardson tableau*.

As an example, the following are the two valid Littlewood-Richardson tableaux with shape $\lambda \setminus \mu = (6, 3, 2) \setminus (3, 1)$ and content $\nu = (4, 2, 1)$:



while the following tableau is invalid as a Littlewood-Richardson tableau since it does not satisfy the reverse lattice word constraint:



We conclude that $c_{\mu\nu}^\lambda = 2$.

Theorem 2 (Littlewood-Richardson rule). *The Littlewood-Richardson coefficient, $c_{\mu\nu}^\lambda$, is equal to the number of Littlewood-Richardson tableaux with shape $\lambda \setminus \mu$ and content ν .*

Proof. See [Sagan, 2001, James and Kerber, 1981], for example. \square

In the proofs that follow, we will rely on a useful property that follows directly from the Littlewood-Richardson rule.

Definition 3. Let λ be a partition of n and μ a partition of $p \leq n$. We say that μ is a *subpartition* of λ if for every i , $\mu_i \leq \lambda_i$.

Lemma 4. *The LR coefficient $c_{\mu\nu}^\lambda = 0$ unless both μ and ν are subpartitions of λ .*

Definition 5. Let $\lambda = (\lambda_1, \dots, \lambda_\ell)$ be a partition of n . The *height* of λ is defined to be ℓ .

For the sake of brevity, we will content ourselves to say that the set of Littlewood-Richardson tableaux for a given shape can be enumerated recursively and with complexity depending (exponentially) on the height of lambda. Finally for some concreteness and intuition, the following decompositions for low-order pairs of irreducibles hold (proofs are straightforward applications of the Littlewood-Richardson rule and are thus omitted).

Proposition 6. *Consider $n \geq 2$ and any positive integers p, q such that $p + q = n$. Then the following decomposition holds:*

$$\rho_{(n-1,1)} \downarrow_{S_p \times S_q} \equiv (\rho_{(p)} \otimes \rho_{(q)}) \oplus (\rho_{(p-1,1)} \otimes \rho_{(q)}) \oplus (\rho_{(p)} \otimes \rho_{(q-1,1)}).$$

Proposition 7. *Let $n \geq 2$ and p, q be any positive integers such that $p + q = n$. Then the following decomposition holds:*

$$\rho_{(n-2,2)} \downarrow_{S_p \times S_q} \equiv (\rho_{(p)} \otimes \rho_{(q)}) \oplus (\rho_{(p)} \otimes \rho_{(q-1,1)}) \oplus (\rho_{(p)} \otimes \rho_{(q-2,2)}) \oplus (\rho_{(p-1,1)} \otimes \rho_{(q)}) \oplus (\rho_{(p-1,1)} \otimes \rho_{(q-1,1)}) \oplus (\rho_{(p-2,2)} \otimes \rho_{(q)}),$$

except in the following exceptional boundary cases for p (the same rules apply for q):

- ($p = 3$) If $p = 3$, remove the term $\rho_{(p-2,2)} \otimes \rho_{(q)}$.
- ($p = 2$) If $p = 2$, remove the terms $\rho_{(p-1,1)} \otimes \rho_{(q)}$ and $\rho_{(p-2,2)} \otimes \rho_{(q)}$.
- ($p = 1$) If $p = 1$, remove all terms except $\rho_{(p)} \otimes \rho_{(q-1,1)}$ and $\rho_{(p)} \otimes \rho_{(q-2,2)}$.

Proposition 8. Let $n \geq 2$ and p, q be any positive integers such that $p + q = n$. Then the following decomposition holds:

$$\rho_{(n-2,1,1)} \downarrow_{S_p \times S_q} \equiv (\rho_{(p)} \otimes \rho_{(q-1,1)}) \oplus (\rho_{(p)} \otimes \rho_{(q-2,1,1)}) \oplus (\rho_{(p-1,1)} \otimes \rho_{(q)}) \oplus (\rho_{(p-1,1)} \otimes \rho_{(q-1,1)}) \oplus (\rho_{(p-2,1,1)} \otimes \rho_{(q)}),$$

except in the following exceptional boundary cases for p (the same rules apply for q):

- ($p = 2$) If $p = 2$, remove the term $\rho_{(p-2,1,1)} \otimes \rho_{(q)}$.
- ($p = 1$) If $p = 1$, remove all terms except $\rho_{(p)} \otimes \rho_{(q-1,1)}$ and $\rho_{(p)} \otimes \rho_{(q-2,2)}$.

Appendix: Proofs

Proof of Lemma 1. Define the sets: $Y = \{k : h(\sigma_i = k) > 0 \text{ for some } i \in X\}$, and $Z = \{k : h(\sigma_j = k) > 0 \text{ for some } j \in \bar{X}\}$. By construction, $h(\sigma) = 0$ unless $\sigma_X \subset Y$ (and $\sigma_{\bar{X}} = Z$), so we need only show that $|Y| = |X|$. By mutual exclusivity, $|X| \leq |Y|$ and $|\bar{X}| \leq |Z|$. We now show that $Y \cap Z = \emptyset$, which will imply that $|Y| = |X|$. Suppose that there exists some $k \in Y \cap Z$. Then by the definitions of Y and Z , there exists $i \in X$ and $j \in \bar{X}$ such that both $h(\sigma_i = k) > 0$ and $h(\sigma_j = k) > 0$. However, by mutual exclusivity, $h(\sigma_i = k, \sigma_j = k) = 0$, and by independence, we see that $h(\sigma_i = k)h(\sigma_j = k) = 0$, thus arriving at a contradiction since we assumed that neither $h(\sigma_i = k)$ nor $h(\sigma_j = k)$ is equal to zero. \square

Proof of Theorem 11. Without loss of generality, we will only consider the marginals of $(1, \dots, p)$. The Split algorithm returns the following matrix (see the proof of Proposition 7):

$$\begin{aligned} [Split(\hat{h})]_{\mu} &= \sum_{\sigma \in S_p \times S_q} h(\sigma) (\rho_{\mu} \otimes \rho_{(q)}(\sigma)), \\ &= \sum_{\sigma_p \in S_p} \left(\sum_{\sigma_q \in S_q} h(\sigma_p, \sigma_q) \right) \rho_{\mu}(\sigma_p). \end{aligned}$$

Let f be the inverse Fourier transform of $Split(\hat{h})$. By the definition of the Fourier transform, we see that for any $\sigma_p \in S_p$, $f(\sigma_p) = \sum_{\sigma_q \in S_q} h(\sigma_p, \sigma_q)$, which is the marginal probability of σ_p under the distribution h . \square

0.1 Marginal Preservation Guarantees

We now state a few properties of partitions and Littlewood-Richardson coefficients in order to prove the Join and Split guarantees.

Lemma 9. Define the partition:

$$\lambda_s^{MIN} = (n - s, \underbrace{1, \dots, 1}_s),$$

for some $0 \leq s < n$, and the set $\Lambda_s = \{\mu : \mu = (n - r, \dots) \text{ for some } r \leq s\}$. The following three statements are equivalent.

- I. $\mu \succeq \lambda_s^{MIN}$.
- II. $\mu \in \Lambda_s$.
- III. $height(\mu) \leq height(\lambda_s^{MIN}) = s + 1$.

Proof.

- ($I \rightarrow II$): If $\mu \notin \Lambda_s$ then $\mu = (n - r, \dots)$ for some $r > s$. By definition of the dominance order, we have that $\mu \triangleleft \lambda_s^{MIN}$.
- ($II \rightarrow III$): If $\mu = (n - r, \dots)$, there are at most r entries in the partition μ besides the first entry $(n - r)$. If $\mu \in \Lambda_s$, then we have $r \leq s$ meaning that there are at most s entries beyond the first and thus $height(\mu) \leq s + 1$.
- ($III \rightarrow I$): If $\mu \triangleleft \lambda_s^{MIN}$, then by definition of the dominance order, we have for each partial sum (for any i):

$$\sum_{k=1}^i \mu_k < \sum_{k=1}^i (\lambda_s^{MIN})_k$$

The height of μ is the minimum i for which $\sum_{k=1}^i \mu_k = n$. By the inequality, we see that the right side must reach n strictly before the left side does as we increase i , and thus we have that $height(\lambda_s^{MIN}) < height(\mu)$. \square

Lemma 10. Define the partitions:

$$\lambda^{MIN} = (n - s, \underbrace{1, \dots, 1}_s), \quad \mu^{MIN} = (p - k, \underbrace{1, \dots, 1}_k), \quad (0.2)$$

where $k = \min(s, p - 1)$. If λ is any partition of n such that $\lambda \succeq \lambda^{MIN}$, then for any partition μ of p which is also of subpartition of λ , we have $\mu \succeq \mu^{MIN}$.

Proof. By Lemma 9, $height(\lambda) \leq height(\lambda^{MIN}) = s + 1$. But since μ is a subpartition of λ , we also have that $height(\mu) \leq height(\lambda)$. And since μ is a partition of p ,

Proof of Proposition 7.

$$\begin{aligned}
 L \cdot \left[\widehat{f \cdot g} \right]_{\lambda} \cdot L^T &= L \cdot \left(\sum_{\sigma \in S_n} f(\sigma_p) \cdot g(\sigma_q) \rho_{\lambda}(\sigma) \right) \cdot L^T, & (\text{Def. of Fourier transform of } f \cdot g) \\
 &= \sum_{\sigma \in S_p \times S_q} f(\sigma_p) \cdot g(\sigma_q) \left(L \cdot \rho_{\lambda}(\sigma) \cdot L^T \right), & (\text{Equation 5.1, linearity}) \\
 &= \sum_{\sigma \in S_p \times S_q} f(\sigma_p) \cdot g(\sigma_q) \left(\bigoplus_{\mu, \nu}^{c_{\mu, \nu}^{\lambda}} \rho_{\mu}(\sigma_p) \otimes \rho_{\nu}(\sigma_q) \right), & (\text{Equation 5.2}) \\
 &= \bigoplus_{\mu, \nu}^{c_{\mu, \nu}^{\lambda}} \left(\sum_{\sigma \in S_p} f(\sigma_p) \rho_{\mu}(\sigma_p) \right) \otimes \left(\sum_{\sigma \in S_q} g(\sigma_q) \rho_{\nu}(\sigma_q) \right), & (\text{Bilinearity of } \otimes) \\
 &= \bigoplus_{\mu, \nu}^{c_{\mu, \nu}^{\lambda}} \left(\widehat{f}_{\mu} \otimes \widehat{g}_{\nu} \right), & (\text{Def. of Fourier transform}).
 \end{aligned}$$

□

$\text{height}(\mu) \leq p$. Putting these inequalities together, we see that $\text{height}(\mu) \leq \min(p, s + 1)$. Finally,

$$\begin{aligned}
 \text{height}(\mu^{MIN}) &= k + 1 \\
 &= \min(s, p - 1) + 1 \\
 &= \min(p, s + 1),
 \end{aligned}$$

showing that $\text{height}(\mu) \leq \text{height}(\mu^{MIN})$. By Lemma 9 again, we conclude that $\mu \supseteq \mu^{MIN}$. □

Corollary 11. *The set of Fourier coefficients:*

$$\left\{ \widehat{f}_{\mu}, \widehat{g}_{\nu} : \begin{array}{l} \mu \text{ is a partition of } p, \\ \nu \text{ is a partition of } q, \\ \text{and } \mu, \nu \text{ are both subpartitions of } \lambda \end{array} \right\},$$

is sufficient for constructing \widehat{h}_{λ} for any partition λ , of $n = p + q$.

Proof of Theorem 8. We need to be able to construct \widehat{h}_{λ} at all partitions λ such that $\lambda \supseteq \lambda^{MIN}$. By Corollary 11, we need subpartitions μ and ν , of λ at all $\lambda \supseteq \lambda^{MIN}$. But by Lemma 4, all such subpartitions are above μ^{MIN} and ν^{MIN} with respect to the dominance ordering, respectively. □

Proof of Theorem 9. Let $\mu = (\mu_1, \mu_2, \dots, \mu_{\ell})$ be any partition of p such that $\mu \supseteq \mu^{MIN}$. By Lemma 9, we have that $\text{height}(\mu) \leq \text{height}(\mu^{MIN}) = \min(p, s + 1)$. Define the partition $\tilde{\mu} = (\mu_1 + n - p, \mu_2, \dots, \mu_{\ell})$. Two things are immediate: first, μ is a subpartition of $\tilde{\mu}$, and second, since μ is a partition of p , $\tilde{\mu}$ is a partition

of n . We also have, by Lemma 9 again, that

$$\begin{aligned}
 \text{height}(\tilde{\mu}) &= \text{height}(\mu), \\
 &\leq \min(p, s + 1), \\
 &\leq s + 1, \\
 &= \text{height}(\lambda^{MIN}),
 \end{aligned}$$

and therefore it must be the case that $\tilde{\mu} \supseteq \lambda^{MIN}$. Finally, we have that $c_{\mu, (q)}^{\tilde{\mu}} = 1$ exactly. □

0.2 Lexicographical Order Preservation Guarantees

We can extend the marginal preservation results above to hold for the finer *lexicographical* ordering on partitions.

Definition 12 (Lexicographical Comparison). Suppose $\lambda^1, \lambda^2 \vdash n$. We say that $\lambda^1 \succ \lambda^2$ if we have $\lambda_i^1 < \lambda_i^2$ at the first part i for which $\lambda_i^1 \neq \lambda_i^2$.

Lex comparisons induce a total ordering on the set of partitions of n , and we will denote the index of a particular partition λ in the lex ordering by $\text{lexindex}_n(\lambda)$. For example, the lex ordering for partitions of $n = 4$ is given by:

$$\begin{array}{cccccc}
 \text{lexindex}_n(\lambda) & 1 & 2 & 3 & 4 & 5 \\
 \lambda & (4) & (3, 1) & (2, 2) & (2, 1, 1) & (1, 1, 1, 1)
 \end{array}$$

It is known that the lexicographical ordering is a refinement of the dominance ordering.

Proposition 13. *Consider $\mu \vdash p$, $\nu \vdash q$, and $\lambda \vdash n$ (where $p + q = n$). If $\text{lexindex}_n(\lambda) < \text{lexindex}_p(\mu)$ (or similarly, if $\text{lexindex}_n(\lambda) < \text{lexindex}_q(\nu)$), then $c_{\mu, \nu}^{\lambda} = 0$.*

Before proving the proposition, we define the following operation which will allow us to compare partitions of p with partitions of n .

Definition 14. Given a partition $\mu = (\mu_1, \mu_2, \dots) \vdash p$, the partition $\mu \uparrow_p^n$ is a partition of n given by $\mu \uparrow_p^n = (\mu_1 + n - p, \mu_2, \mu_3, \dots)$. Thus μ and $\mu \uparrow_p^n$ agree on all parts except the first.

Lemma 15. Let $\mu \vdash p$. The following inequality holds: $\text{lexindex}_p(\mu) \leq \text{lexindex}_n(\mu \uparrow_p^n)$.

Proof. We proceed by induction on $\text{lexindex}_p(\mu)$. First, the lemma is obvious when $\text{lexindex}_p(\mu) = 1$ (when $\mu = (p)$).

Now we consider the case where $\text{lexindex}_p(\mu) > 1$. Let μ' be any partition of p with $\mu' \prec \mu$. First, we remark that $(\mu') \uparrow_p^n \prec_n \mu \uparrow_p^n$ since we are adding $n - p$ to both μ'_1 and μ_1 .

By the inductive hypothesis, we have that

$$\begin{aligned} \text{lexindex}_p(\mu') &\leq \text{lexindex}_n((\mu') \uparrow_p^n) \\ &< \text{lexindex}_n(\mu \uparrow_p^n). \end{aligned}$$

Since we have shown that $\text{lexindex}_p(\mu') < \text{lexindex}_n(\mu \uparrow_p^n)$ for every μ' such that $\mu' \prec_p \mu$, we conclude that $\text{lexindex}_p(\mu) \leq \text{lexindex}_n(\mu \uparrow_p^n)$. \square

Proof. (of Proposition 13) The assumption is that $\text{lexindex}_n(\lambda) < \text{lexindex}_p(\mu)$. By Lemma 15, it must also be the case that:

$$\text{lexindex}_n(\lambda) < \text{lexindex}_n(\mu \uparrow_p^n). \quad (0.3)$$

By definition of the lex ordering, Equation 0.3 means that there exists some i for which $(\mu \uparrow_p^n)_i < \lambda_i$ and $(\mu \uparrow_p^n)_j = \lambda_j$ for all $j < i$.

We now argue that μ cannot possibly be a subpartition of λ , which will imply that $c_{\mu\nu}^\lambda = 0$ by the Littlewood-Richardson rule. We define the following partitions:

$$\lambda_{chop} = (\lambda_i, \lambda_{i+1}, \dots), \quad \mu_{chop} = (\mu_i, \mu_{i+1}, \dots).$$

Clearly, λ_{chop} is a partition of $a = n - \sum_{j=1}^{i-1} \lambda_j$ and μ_{chop} is a partition of $b = p - \sum_{j=1}^{i-1} \mu_j$. We will prove that μ_{chop} is not a subpartition of λ_{chop} , which will then imply that μ is not a subpartition of λ .

Now we know that $(\mu \uparrow_p^n)_1 \leq \lambda_1$ (if it were greater, then the assumption in Equation 0.3 would be false). Thus, $\mu_1 + n - p \leq \lambda_1$. Or rearranging,

$$\mu_1 \leq \lambda_1 - n + p. \quad (0.4)$$

Thus, we have:

$$\begin{aligned} b &= p - \sum_{j=1}^{i-1} \mu_j \\ &= p - \mu_1 - \sum_{j=2}^{i-1} \mu_j, \\ &\geq p - (\lambda_1 - n + p) - \sum_{j=2}^{i-1} \lambda_j, \\ &\quad \text{(by Equation 0.4 and since } \mu_j = \lambda_j \text{ for} \\ &\quad \text{ } j = 2, \dots, i-1 \text{ by our definition of } i) \\ &\geq n - \sum_{j=1}^{i-1} \lambda_j, \\ &\geq a. \end{aligned}$$

To summarize, we have shown that $\lambda_{chop} \vdash a$ and $\mu_{chop} \vdash b$, where $b \geq a$, but $\mu_i < \lambda_i$. Under these conditions, it is impossible for μ_{chop} to be a subpartition of λ_{chop} . We therefore must conclude that μ is not a subpartition of λ and hence, that $c_{\mu\nu}^\lambda = 0$. \square

References

- B. Sagan. *The Symmetric Group*. Springer, 2001.
- A. Knutson and T. Tao. The honeycomb model of $gl_n(\mathbb{C})$ tensor products i: Proof of the saturation conjecture. *Journal of the American Mathematical Society*, 12(4): 1055–1090, 1999.
- R. Vakil. A geometric littlewood-richardson rule. *Annals of Math.*, (164):371–422, 2006.
- H. Narayanan. On the complexity of computing kostka numbers and littlewood-richardson coefficients. *J. Algebraic Comb.*, 24(3):347–354, 2006.
- Gordon James and Adelbert Kerber. *The Representation Theory of the Symmetric Group*. Addison-Wesley, 1981.