On Convex Programming Relaxations for the Permanent

Damian Straszak[1] and Nisheeth K. Vishnoi[2]

[1,2]École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Abstract

In recent years, several convex programming relaxations have been proposed to estimate the permanent of a non-negative matrix, notably in the works of [GS02, Gur11, GS14]. However, the origins of these relaxations and their relationships to each other have remained somewhat mysterious. We present a conceptual framework, implicit in the belief propagation literature, to systematically arrive at these convex programming relaxations for estimating the permanent – *as approximations to an exponential-sized max-entropy convex program for computing the permanent*. Further, using standard convex programming techniques such as duality, we establish equivalence of these aforementioned relaxations to those based on *capacity-like* quantities studied in [Gur06, Gur09, AMOV16].

1 Introduction

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix with entries $A_{i,j}$ for $1 \leq i, j \leq n$. The permanent of A is defined as:

$$\mathrm{Per}(A) \overset{\mathrm{def}}{=} \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)},$$

where S_n is the set of all permutations over n symbols, i.e., the set of bijections $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. The matrix permanent makes its appearance in various branches of science and mathematics and algorithms to compute it are sought after, especially in theoretical computer science.

In a foundational result, Valiant [Val79] proved that it is unlikely that there is an efficient algorithm that computes the permanent of a matrix – even when the matrix is non-negative (the problem is #**P**−complete). Jerrum, Sinclair and Vigoda [JSV04], building upon a long line of work, showed that the Markov Chain Monte Carlo framework can be deployed to obtain a randomized algorithm to estimate the permanent of any non-negative matrix to within a factor of $1 + \varepsilon$ in time that is polynomial in the bit-lengths of A and $1/\varepsilon$.

A different approach to estimate the value of the permanent of a non-negative matrix has been developed – notably in the works of Linial, Samorodnitsky and Wigderson [LSW98] and Gurvits and Samorodnitsky [GS02]. Roughly, the idea in these works is to reduce the computation of the permanent of a non-negative matrices to that of a $n \times n$ doubly stochastic matrix by computing a certain *scaling*. The latter work [GS02] reveals how the scalings can be viewed as solutions to certain convex programs – leading to convex programming relaxations for the permanent. The resolution of the van der Waerden conjecture by Egorychev [Ego81] and Falikman [Fal81] – that the permanent of a doubly stochastic matrix is lower bounded by e^{-n} – resulted in a deterministic algorithm to estimate the permanent up to a multiplicative ratio of e^n. Of note here is a relaxation introduced by Gurvits [Gur06], called the *capacity* of the associated polynomial p_A (applies more generally), which allowed him to deploy the theory of hyperbolic or real-stable polynomials to derive the same result. These ideas were further developed in the works of [Gur11, GS14] where a curious convex program was presented and, it was shown using an inequality due to Schrijver [Sch98], that this program estimates the value of the permanent up to a 2^n factor.

The problem of computing the permanent has also attracted attention in statistical physics and, in particular, the belief propagation literature [WC10, Von13, CY13]. These algorithms are geared towards computing approximations to the Gibbs distribution associated to the matrix A. A bit more formally, if $G = (V, W, E)$ is the bipartite graph where $|V| = |W| = n$ and $(i, j) \in E$ if $A_{i,j} > 0$ and $\mathcal{M}$ is the set of perfect matchings in G, then the Gibbs distribution induced by A and supported on $\mathcal{M}$ is one where, for a matching $M \in \mathcal{M}$, the probability of M, denoted by, $p(M)$, is

$$p(M) \stackrel{\text{def}}{=} \frac{\prod_{(i,j) \in M} A_{i,j}}{\text{Per}(A)}.$$

The idea then is to maintain a proxy to $p(M)$, or a *belief* distribution $b : \mathcal{M} \to \mathbb{R}_{\geq 0}$ and iteratively update it according to certain rules. It is not clear when and how quickly such algorithms converge although they seem to work well in practice and for locally tree-like graphs.

Are these three seemingly distinct approaches to compute the permanent, coming from different areas, connected?

The goal of this paper is to present some old and some new results in a unified manner which seem to suggest that there may be a close connection between the latter two approaches. The starting point to our work is the well-known fact that the following convex program has $\log \text{Per}(A)$ as its optimal value.

(1)
$$\sup \quad \sum_{M \in \mathcal{M}} b(M) \log \frac{\prod_{(i,j) \in M} A_{i,j}}{b(M)}$$
$$\text{s.t.} \quad \sum_{M \in \mathcal{M}} b(M) = 1$$
$$b(M) \geq 0 \qquad \forall M \in \mathcal{M}$$

Each variable of this program corresponds to a perfect matching in G, and the goal is to find a probability distribution over perfect matchings in G that minimizes the Kullback-Leibler divergence from the Gibbs distribution. Thus, if one could compute the optimal value of this program then one could estimate the permanent exactly. However, the number of variables in this program can be exponentially large and it is not clear how to solve it. One can take the dual of this program, however, that leads (predictably) to the same problem; see [SV14] for more details.

We show how we can bypass this barrier and construct approximate convex programs starting from this one by making assumptions on the distribution we are seeking. Key to these simplifications are some heuristics from belief propagation, which result in certain relations between these marginals that one (wishfully) expects the optimal distribution p to satisfy. Assuming these relationships, one can then write the above convex program entirely in terms of the first order marginals of the distribution b – dramatically reducing the number of variables. Of course, as a result of these simplifications, the convex programs obtained are approximations of the one above; see (3) and (6). However, as proved by Gurvits [Gur06] and Gurvits and Samorodnitsky [GS14], they give us some of the best deterministic algorithms to estimate the value of $\text{Per}(A)$. We expect that our point of view will lead to new convex programming methods for other counting problems (such as those studied in [SV16]) by leveraging on heuristic reasonings from belief propagation.

2   Relaxations via Belief Propagation

Belief propagation algorithms attempt to find the Gibbs distribution p by maintaining belief distributions b supported over $\mathcal{M}$. These relations are often arrived at by considering the case when G is a tree. In the following two subsections, we show that using two different relations, we end up with two different convex programs which, in turn, via standard techniques from convex programming allow us to recover the convex programs from [GS02, Gur09, Gur11, GS14, AMOV16]. Some of the results we present are implicit in the

papers by [WC10, Von13, CY13, CS15]. Our contribution is to present the unified picture along with some new relations.

Let us now introduce some notation to explain the ideas formally. For a probability distribution b supported on $\mathcal{M}$ we denote by $b_{i,j}(1)$ the probability that the edge (i,j) is present in a random matching drawn from b, similarly $b_{i,j}(0) = 1 - b_{i,j}(1)$ is the probability that the edge is absent. Denote $B_{i,j} = b_{i,j}(0)$, it follows in particular that

1. $\sum_{j=1}^{n} B_{i,j} = 1$ for all $1 \le i \le n$, and

2. $\sum_{i=1}^{n} B_{i,j} = 1$ for all $1 \le j \le n$.

In other words, $B = (B_{i,j})_{i,j \in [n]}$ is a doubly stochastic matrix with support determined by A. We denote this set by $\Omega_n(A)$ and the set of all doubly stochastic matrices by Ω_n.

2.1 Gurvits' first relaxation for the permanent

The starting point here is the following observation that when G is a tree and b is any distribution over perfect matchings in G, then

$$(2) \qquad b(M) = \prod_{(i,j):M_{i,j}=1} b_{i,j}(1).$$

In other words, the probability of a perfect matching in G is completely determined by the marginal probabilities of the edges and is computed as if all the events (corresponding to picking a particular edge) were independent. The reason why this is trivially true in the case of a tree is because a tree has at most one perfect matching. Therefore, all the marginals are either zero or one, depending whether a given edge belongs to the unique perfect matching or not.

The heuristic jump then occurs when one, wishfully thinking, hopes that these relations hold for all graphs – *at least when b is the Gibbs distribution p*. While incorrect, assuming that b satisfies (2), leads us to the convex program as an approximation to that in (1) in the following manner:

$$\sum_{M \in \mathcal{M}} b(M) \log \frac{\prod_{(i,j) \in E:M_{i,j}=1} A_{i,j}}{\prod_{(i,j):M_{i,j}=1} b_{i,j}(1)} = \sum_{M \in \mathcal{M}} b(M) \log \frac{\prod_{(i,j) \in E:M_{i,j}=1} A_{i,j}}{\prod_{(i,j) \in E:M_{i,j}=1} B_{i,j}}$$

$$= \sum_{M \in \mathcal{M}} b(M) \left(\sum_{(i,j) \in E:M_{i,j}=1} \log \frac{A_{i,j}}{B_{i,j}} \right).$$

The expression on the right hand side can be simplified further. Consider

$$\sum_{M \in \mathcal{M}} b(M) \left(\sum_{(i,j) \in E:M_{i,j}=1} \log \frac{A_{i,j}}{B_{i,j}} \right) = \sum_{(i,j) \in E} \sum_{M \in \mathcal{M}:M_{i,j}=1} b(M) \log \frac{A_{i,j}}{B_{i,j}}$$

$$= \sum_{(i,j) \in E} \log \frac{A_{i,j}}{B_{i,j}} \sum_{M \in \mathcal{M}:M_{i,j}=1} b(M)$$

$$= \sum_{(i,j) \in E} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}}.$$

Thus the relaxation (1) reduces to the problem of maximizing the above expression over $B \in \Omega_n(A)$, i.e., we arrive at the following convex program:[1]

---

[1]In (3) we switched from $\Omega_n(A)$ to Ω_n as the domain for B, however it is not hard to see that the value of the relaxation is not affected.

3

$$(3) \qquad\qquad R_E(A) \stackrel{\text{def}}{=} \sup_{B \in \Omega_n} \sum_{(i,j) \in E} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}}.$$

Using Lagrangian duality, we can now derive the relaxation considered by Gurvits. For variables $z_1, \ldots, z_n$ consider the polynomial $q_A(z) \stackrel{\text{def}}{=} \prod_{i=1}^{n} \sum_{i=1}^{n} A_{i,j} z_j$. One can show using standard techniques that the following is the dual of the above convex program.

$$(4) \qquad\qquad R_C(A) \stackrel{\text{def}}{=} \inf_{z>0} \frac{q_A(z)}{\prod_{i=1}^{n} z_i}.$$

Formally one can prove the following theorem (see Appendix A and [CS15]):

Theorem 2.1 *For all non-negative $n \times n$ matrices A, $R_C(A) = \exp(R_E(A))$.*

Since the coefficient of $\prod_{i=1}^{n} z_i$ in $q_A(z)$ is equal to $\text{Per}(A)$ it is easy to see that

$$\text{Per}(A) \le R_C(A).$$

Furthermore an important result of Gurvits [Gur06], which makes this convex program interesting, asserts that an upper bound holds:

$$R_C(A) \le e^n \cdot \text{Per}(A).$$

Note that even though we arrived at this convex program by making an incorrect assumption on the Gibbs distribution, the result is a relaxation which provably approximates the permanent up to a factor of e^n – thus yielding a *deterministic* algorithm to estimate the value of the permanent of A.

We conclude this section by showing how (3) can be reinterpreted in the language of polynomials, resulting in another relaxation for the permanent implicit in the work of [AMOV16] (see also [Gur09]). To explain their relaxation, consider any two positive matrices C and D such that their entry-wise product is equal to A, i.e., $A_{i,j} = C_{i,j} \cdot D_{i,j}$ for all $1 \le i, j \le n$. Define n^2-variate polynomials $p(x) = \prod_{i=1}^{n} \sum_{j=1}^{n} x_{i,j} C_{i,j}$ and $q(y) = \prod_{j=1}^{n} \sum_{i=1}^{n} y_{i,j} D_{i,j}$. Then the relaxation is as follows:[2]

$$R_P(A) \stackrel{\text{def}}{=} \sup_{B \in \Omega_n} \inf_{x>0,y>0} \frac{p(x) \cdot q(y) \cdot \prod_{i,j} B_{i,j}^{B_{i,j}}}{\prod_{i,j} (y_{i,j} x_{i,j})^{B_{i,j}}}.$$

We observe that this relaxation is equivalent to (3).

Theorem 2.2 *For all non-negative $n \times n$ matrices A, $R_P(A) = \exp(R_E(A))$.*

The proof can be easily deduced from the following observation.

Claim 2.3 *Let r be the following polynomial over n^2 variables $x_{i,j}$ (for $1 \le i, j \le n$)*

$$r(x) = \prod_{i=1}^{n} \sum_{j=1}^{n} M_{i,j} x_{i,j},$$

where M is any matrix with positive entries. Let $B \in \Omega_n$ be any doubly stochastic matrix, then

$$\inf_{x \in \mathbb{R}_{>0}^{n \times n}} \frac{r(x)}{\prod_{i,j} x_{i,j}^{B_{i,j}}} = \exp\left(\sum_{i,j} B_{i,j} \log \frac{M_{i,j}}{B_{i,j}} \right).$$

The proof of Claim 2.3 appears in Appendix B.

---

[2]In the relaxation defining $R_P(A)$ one can equivalently optimize over a larger domain for B, namely $B \in [0,1]^{n \times n}$.

2.2 Gurvits' second relaxation for the permanent

To derive the second relaxation, as before we start with a relation among a distribution b and its marginals which holds (again trivially) when G is a tree.

$$b(M) = \frac{\prod_{M_{i,j}=1} b_{i,j}(1)}{\prod_{M_{i,j}=0} b_{i,j}(0)}. \tag{5}$$

The reason this is true for trees is that for the unique perfect matching M of G for $(i,j) \in M$ we have $b_{i,j}(1) = 1$ and for $(i,j) \notin M$ we have $b_{i,j}(0) = 1$. As in the previous section, we make a jump and assume that this relation holds for any bipartite graph – at least for the case of the Gibbs distribution.

Let us simplify the convex program (1) using (5) as an assumption. We obtain that the objective of the convex program (1) can be written as

$$\sum_{M \in \mathcal{M}} b(M) \log \frac{\prod_{(i,j)\in E: M_{i,j}=1} A_{i,j}}{\prod_{(i,j)\in E: M_{i,j}=1} B_{i,j}} = \sum_{M \in \mathcal{M}} b(M) \left(\sum_{(i,j)\in E: M_{i,j}=1} \log \frac{A_{i,j}}{B_{i,j}} + \sum_{(i,j)\in E: M_{i,j}=0} \log(1 - B_{i,j}) \right).$$

The expression on the right hand side can be simplified as before to obtain

$$\sum_{M \in \mathcal{M}} b(M) \left(\sum_{(i,j)\in E: M_{i,j}=1} \log \frac{A_{i,j}}{B_{i,j}} \right) = \sum_{(i,j)\in E} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}}.$$

Similarly,

$$\sum_{M \in \mathcal{M}} b(M) \left(\sum_{(i,j)\in E: M_{i,j}=0} \log(1 - B_{i,j}) \right) = \sum_{(i,j)\in E} (1 - B_{i,j}) \log(1 - B_{i,j}).$$

This allows us to rewrite the convex program (1) in the following form:

$$R_O(A) \overset{\text{def}}{=} \sup_{B \in \Omega_n} \sum_{(i,j)\in E} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}} + \sum_{(i,j)\in E} (1 - B_{i,j}) \log(1 - B_{i,j}). \tag{6}$$

It is worth noting that even though the program (1) is convex, it does not imply that the objective in (6) is convex. In fact it is convex only when restricted to the domain Ω_n, which was observed by [Von13]. Even more remarkable is the fact that, despite us plugging in the relations between the marginals that are only guaranteed to hold when G is a tree, it was proved that the approximation guarantee is even better as for the first relaxation. Here proving that this convex program has any relation to the permanent requires some work and, indeed, Gurvits, using an inequality due to Schrijver [Sch98], proved that

$$\exp(R_O(A)) \leq \text{Per}(A).$$

Notice here that the estimate obtained from the convex program provides a *lower bound* to the permanent unlike Gurvits' first relaxation. Subsequently, Gurvits and Samorodnitsky [GS14] proved that this quantity in fact provides a sharp estimate for the permanent:

$$\text{Per}(A) \leq 2^n \cdot \exp(R_O(A)).$$

Finally, we show how another polynomial-based relaxation for the permanent that is implicit in [AMOV16] is equivalent to (6). Let, as before, C and D be two positive matrices such that their entry-wise product is

equal to A, i.e., $A_{i,j} = C_{i,j} \cdot D_{i,j}$ for all $1 \leq i, j \leq n$. Define polynomials $p(x) = \prod_{i=1}^n \sum_{j=1}^n x_{i,j} C_{i,j}$ and $q(y) = \prod_{j=1}^n \sum_{i=1}^n y_{i,j} D_{i,j}$. Consider the following

$$(7) \qquad R_Q(A) \stackrel{\text{def}}{=} \sup_{B \in \Omega_n} \inf_{x>0,y>0} \frac{p(x) \cdot q(y) \cdot \prod_{i,j} B_{i,j}^{B_{i,j}} \cdot \prod_{i,j} (1 - B_{i,j})^{1-B_{i,j}}}{\prod_{i,j} (y_{i,j} x_{i,j})^{B_{i,j}}}.$$

As in the case of the first relaxation, we have the following theorem.

Theorem 2.4 *For all non-negative $n \times n$ matrices A, $R_Q(A) = \exp(R_O(A))$.*

This theorem can be easily deduced from Claim 2.3. Indeed, for a fixed B, x and y are separable in the objective of R_Q, hence Claim 2.3 can be applied separately to both of them, to yield the claimed result.

References

[AMOV16]   N. Anari, T. Mai, S. Oveis Gharan, and V. V. Vazirani. Nash Social Welfare for Indivisible Items under Separable, Piecewise-Linear Concave Utilities. *ArXiv e-prints*, December 2016.

[CS15]   V. Chandrasekaran and P. Shah. Relative entropy optimization and its applications. *Mathematical Programming*, pages 1–32, 2015.

[CY13]   M. Chertkov and A. B. Yedidia. Approximating the permanent with fractional belief propagation. *Journal of Machine Learning Research*, 14(1):2029–2066, 2013.

[Ego81]   G. P. Egorychev. The solution of van der Waerden's problem for permanents. *Advances in Mathematics*, 42(3):299–305, 1981.

[Fal81]   D. I. Falikman. Proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29(6):475–479, 1981.

[GS02]   L. Gurvits and A. Samorodnitsky. A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary. *Discrete & Computational Geometry*, 27:531–550, 2002.

[GS14]   L. Gurvits and A. Samorodnitsky. Bounds on the Permanent and Some Applications. In *FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 90–99, 2014.

[Gur06]   L. Gurvits. Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 417–426. ACM, 2006.

[Gur09]   L. Gurvits. On multivariate newton-like inequalities. In *Advances in Combinatorial Mathematics*, pages 61–78. Springer, 2009.

[Gur11]   L. Gurvits. Unleashing the power of Schrijver's permanental inequality with the help of the Bethe Approximation. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:169, 2011.

[JSV04]   K. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, July 2004.

[LSW98]   N. Linial, A. Samorodnitsky, and A. Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 644–652. ACM, 1998.

[Sch98]   A. Schrijver. Counting 1-factors in regular bipartite graphs. *Journal of Combinatorial Theory, Series B*, 72(1):122–135, 1998.

[SV14]   M. Singh and N. K. Vishnoi. Entropy, optimization and counting. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 50–59. ACM, 2014.

[SV16]   D. Straszak and N. K. Vishnoi. Real Stable Polynomials and Matroids: Optimization and Counting. *ArXiv e-prints*, November 2016.

[Val79]   L. G. Valiant. The complexity of computing the permanent. *Theoretical computer science*, 8(2):189–201, 1979.

[Von13]   P. O. Vontobel. The Bethe permanent of a nonnegative matrix. *IEEE Transactions on Information Theory*, 59(3):1866–1901, 2013.

[WC10]   Y. Watanabe and M. Chertkov. Belief propagation and loop calculus for the permanent of a non-negative matrix. *Journal of Physics A: Mathematical and Theoretical*, 43(24):242002, 2010.

A   Proof of Theorem 2.1

Proof: Consider an $n \times n$ non-negative matrix A and assume for simplicity that it is of full support, i.e., all the entries of A are positive. Let us first state the relaxation defining $R_E(A)$ with all constraints explicitly present.

(8)
$$
\begin{aligned}
\max_{B \in \mathbb{R}^{n \times n}} \quad & \sum_{1 \le i,j \le n} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}}, \\
\text{s.t.} \quad & \sum_{i=1}^{n} B_{i,j} = 1, && j = 1, 2, \ldots, n \\
& \sum_{j=1}^{n} B_{i,j} = 1, && i = 1, 2, \ldots, n \\
& B \ge 0.
\end{aligned}
$$

Note importantly that the objective of (8) is concave in B. We now derive the dual of the above program. Introduce Lagrangian multipliers $\alpha, \beta \in \mathbb{R}^n$ for the row and column stochasticity constraints. We arrive at the Lagrangian

$$
L(B, \alpha, \beta) = \sum_{1 \le i,j \le n} B_{i,j} \log \frac{A_{i,j}}{B_{i,j}} - \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} B_{i,j} - 1 \right) - \sum_{j=1}^{n} \beta_j \left(\sum_{i=1}^{n} B_{i,j} - 1 \right).
$$

The next step is to compute

$$
g(\alpha, \beta) = \sup_{B \ge 0} L(B, \alpha, \beta).
$$

To this end let us compute the derivative of L with respect to $B_{i,j}$.

$$
\frac{\partial}{\partial B_{i,j}} L(B, \alpha, \beta) = \log A_{i,j} - \log B_{i,j} - 1 - \alpha_i - \beta_j.
$$

Setting $\frac{\partial}{\partial B_{i,j}} L(B, \alpha, \beta) = 0$ gives

$$
B_{i,j} = A_{i,j} e^{-1 - \alpha_i - \beta_j}.
$$

By plugging in these values to the Lagrangian, we obtain

$$g(\alpha, \beta) = \sum_{1 \leq i,j \leq n} A_{i,j} e^{-1-\alpha_i-\beta_j} + \sum_{i=1}^{n} \alpha_i + \sum_{j=1}^{n} \beta_j.$$

Since the Slater's condition is satisfied for (8), strong duality holds and we can deduce that

$$R_E(A) = \min_{\alpha,\beta \in \mathbb{R}^n} \sum_{1 \leq i,j \leq n} A_{i,j} \cdot e^{-1-\alpha_i-\beta_j} + \sum_{i=1}^{n} \alpha_i + \sum_{j=1}^{n} \beta_j.$$

Finally, we can eliminate one set of variables due to a simple form of the dual objective $g(\alpha, \beta)$. Indeed, let us rewrite $g(\alpha, \beta)$ as

$$g(\alpha, \beta) = \sum_{i=1}^{n} e^{1+\alpha_i} \sum_{j=1}^{n} A_{i,j} e^{\beta_j} + \sum_{i=1}^{n} \alpha_i + \sum_{j=1}^{n} \beta_j.$$

We derive analytically a formula for $h(\beta) = \min_{\alpha \in \mathbb{R}^n} g(\alpha, \beta)$. Note that the problem of minimizing over α is separable and hence we just need to solve a sequence of optimization problems of the form

$$\min_{\alpha_i \in \mathbb{R}} \ e^{-1-\alpha_i} \sum_{j=1}^{n} A_{i,j} e^{-\beta_j} + \alpha_i.$$

By a simple calculation, the optimal value of the above is $\log\left(\sum_{j=1}^{n} A_{i,j} e^{-\beta_j}\right)$ and hence we arrive at

$$h(\beta) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{n} A_{i,j} e^{-\beta_j}\right) + \sum_{j=1}^{n} \beta_j.$$

Further observe that $h(\beta) = \log\left(p_A\left(e^{-\beta_1}, e^{-\beta_2}, \ldots, e^{-\beta_n}\right)\right) + \sum_{j=1}^{n} \beta_j$. Hence in particular

$$\min_{\beta \in \mathbb{R}^n} h(\beta) = \min_{x>0} \log(p_A(x)) - \sum_{j=1}^{n} \log x_j = \min_{x>0} \log \frac{p_A(x)}{\prod_{j=1}^{n} x_j}.$$

Finally we obtain

$$R_E(A) = \min_{\alpha,\beta \in \mathbb{R}^n} g(\alpha, \beta) = \min_{\beta \in \mathbb{R}^n} h(\beta) = \log R_C(A).$$

$\blacksquare$

B Proof of Claim 2.3

Proof: For any $x \in \mathbb{R}_{>0}^{n \times n}$ and any positive doubly stochastic matrix B we have

$$\prod_{i=1}^{n} \sum_{j=1}^{n} x_{ij} M_{ij} = \prod_{i=1}^{n} \sum_{j=1}^{n} B_{i,j} x_{i,j} \frac{M_{i,j}}{B_{i,j}}$$

$$\geq \prod_{i=1}^{n} \prod_{j=1}^{n} \left(x_{i,j} \frac{M_{i,j}}{B_{i,j}}\right)^{B_{i,j}}$$

$$= \prod_{1 \leq i,j \leq n} \left(\frac{x_{i,j} M_{i,j}}{B_{i,j}}\right)^{B_{i,j}}$$

8

The only inequality in the derivation above follows from applying Jensen's inequality to the function log. We obtain
$$\inf_{x \in \mathbb{R}^{n \times n}_{>0}} \frac{r(x)}{\prod_{1 \leq i,j \leq n} x_{i,j}^{B_{i,j}}} \geq \left(\frac{M_{i,j}}{B_{i,j}} \right)^{B_{i,j}}.$$

To deduce equality it remains to observe that the applied Jensen's inequality is tight whenever $x_{i,j} \frac{M_{i,j}}{B_{i,j}} = $ const. which can be easily achieved by plugging in an appropriate value of x. ∎